



# **Popisná statistika bodových dat**

## **Statistický popis prostorového uspořádání bodů**

## **Statistická analýza plošných jevů**

podzim 2023

**Lukáš Herman**

[herman.lu@mail.muni.cz](mailto:herman.lu@mail.muni.cz)



- O čem článek (v kostce) je?
- Co je prostorová autokorelace?



Scientific  
Research  
Publishing

Open Journal of Civil Engineering, 2017, 7, 208-221

<http://www.scirp.org/journal/ojce>

ISSN Online: 2164-3172

ISSN Print: 2164-3164

## A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord $G_i^*$ Statistic

Azad Abdulhafedh\*

Department of Civil and Environmental Engineering, University of Missouri, Columbia, MO, USA

Email: asa8cd@mail.missouri.edu

How to cite this paper: Abdulhafedh, A. (2017) A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord  $G_i^*$  Statistic. *Open Journal of Civil Engineering*, 7, 208-221. <https://doi.org/10.4236/ojce.2017.72013>

Received: February 13, 2017

Accepted: June 3, 2017

Published: June 6, 2017

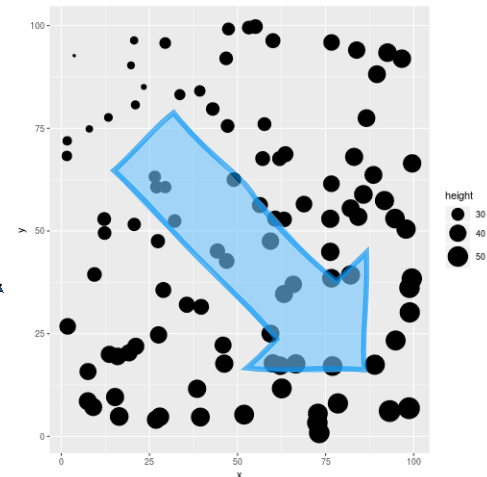
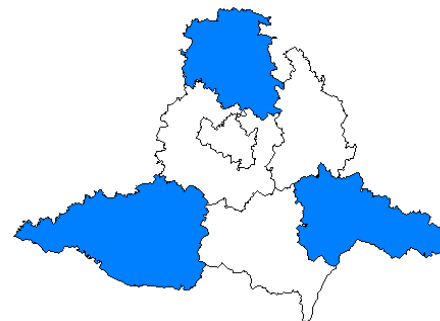
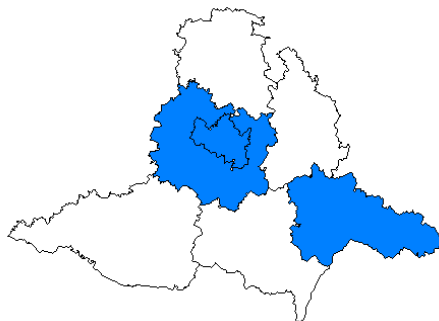
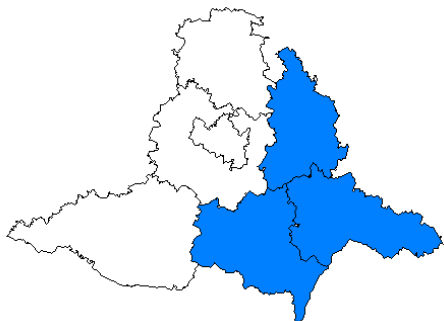
Copyright © 2017 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

### Abstract

Spatial autocorrelation is a measure of the correlation of an observation with other observations through space. Most statistical analyses are based on the assumption that the values of observations are independent of one another. Spatial autocorrelation violates this assumption, because observations at nearby locations are related to each other, and hence, the consideration of spatial autocorrelations has been gaining attention in crash data modeling in recent years, and research have shown that ignoring this factor may lead to a biased estimation of the modeling parameters. This paper examines two spatial autocorrelation indices: Moran's Index; and Getis-Ord  $G_i^*$  statistic to measure the spatial autocorrelation of vehicle crashes occurred in Boone County roads in the state of Missouri, USA for the years 2013-2015. Since each index can identify different clustering patterns of crashes, therefore this paper introduces a

# Prostorová autokorelace

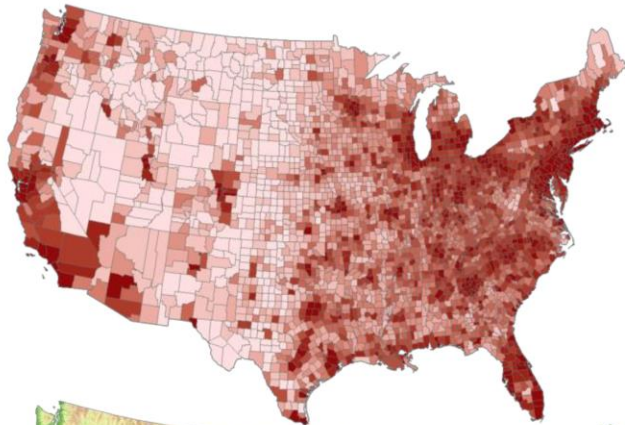
- Hodnoty atributů spolu korelují v závislosti na jejich vzájemné poloze.
- To může být v důsledku podobných přirozených (přírodních) podmínek (*např. produkce zemědělských podniků*) či v důsledku přirozené spojitosti jevů.
- Příklad – okresy Jihomoravského kraje: **pozitivní** prostorové autokorelace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



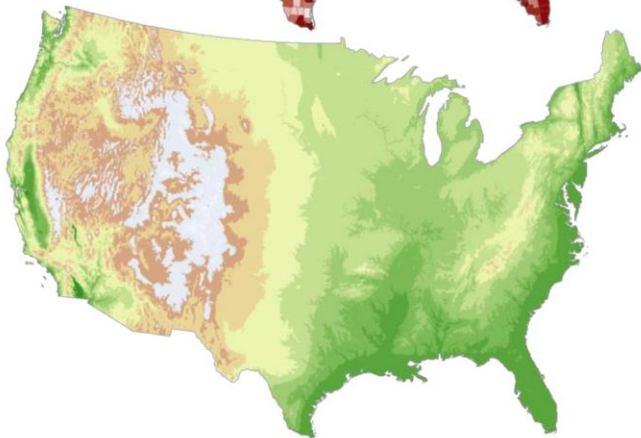
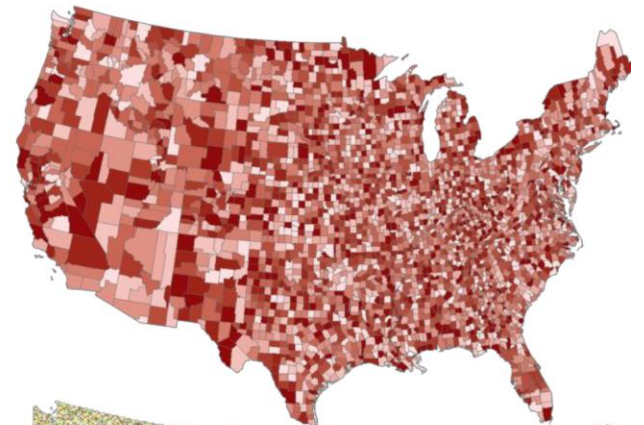


# Prostorová autokorelace

If features were randomly distributed ...



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



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





- Jaké příklady prostorových statistik/analýz článek popisuje?
- V čem spočívá „novost“ představené hybridní metody?



Scientific  
Research  
Publishing

Open Journal of Civil Engineering, 2017, 7, 208-221

<http://www.scirp.org/journal/ojce>

ISSN Online: 2164-3172

ISSN Print: 2164-3164

## A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord $G_i^*$ Statistic

Azad Abdulhafedh\*

Department of Civil and Environmental Engineering, University of Missouri, Columbia, MO, USA

Email: asa8cd@mail.missouri.edu

How to cite this paper: Abdulhafedh, A. (2017) A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord  $G_i^*$  Statistic. *Open Journal of Civil Engineering*, 7, 208-221. <https://doi.org/10.4236/ojce.2017.72013>

Received: February 13, 2017

Accepted: June 3, 2017

Published: June 6, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

### Abstract

Spatial autocorrelation is a measure of the correlation of an observation with other observations through space. Most statistical analyses are based on the assumption that the values of observations are independent of one another. Spatial autocorrelation violates this assumption, because observations at nearby locations are related to each other, and hence, the consideration of spatial autocorrelations has been gaining attention in crash data modeling in recent years, and research have shown that ignoring this factor may lead to a biased estimation of the modeling parameters. This paper examines two spatial autocorrelation indices: Moran's Index; and Getis-Ord  $G_i^*$  statistic to measure the spatial autocorrelation of vehicle crashes occurred in Boone County roads in the state of Missouri, USA for the years 2013-2015. Since each index can identify different clustering patterns of crashes, therefore this paper introduces a

- Odhalili jste nějaké specifikum z hlediska geometrie použitých prostorových dat?
- Odhalili jste v článku nějaký problém (nedostatek, chybu)?



Open Journal of Civil Engineering, 2017, 7, 208-221  
<http://www.scirp.org/journal/ojce>  
ISSN Online: 2164-3172  
ISSN Print: 2164-3164

## A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord $G_i^*$ Statistic

Azad Abdulhafedh\*

Department of Civil and Environmental Engineering, University of Missouri, Columbia, MO, USA  
Email: asa8cd@mail.missouri.edu

How to cite this paper: Abdulhafedh, A. (2017) A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord  $G_i^*$  Statistic. *Open Journal of Civil Engineering*, 7, 208-221.  
<https://doi.org/10.4236/ojce.2017.72013>

Received: February 13, 2017

Accepted: June 3, 2017

Published: June 6, 2017

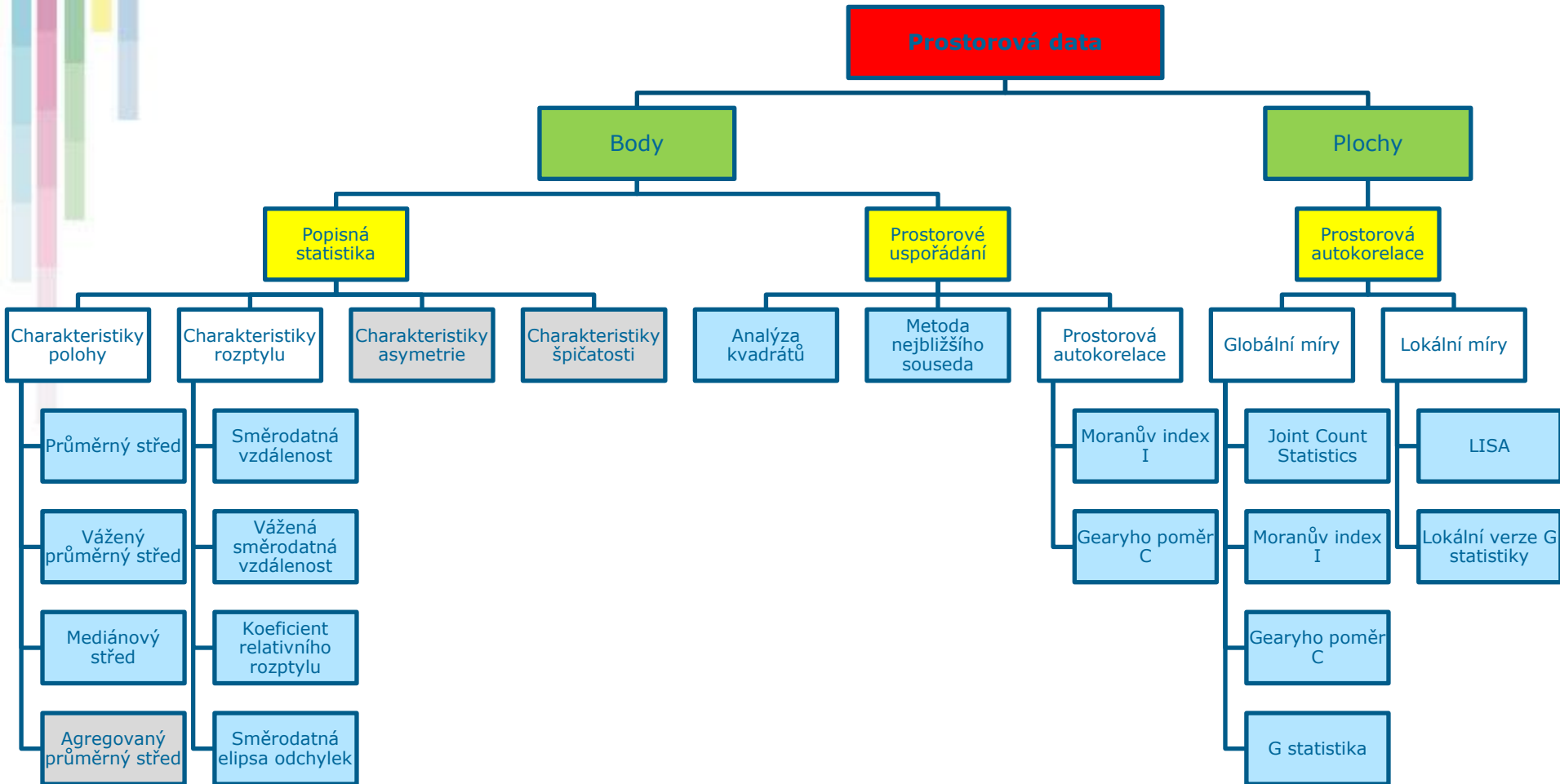
Copyright © 2017 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

### Abstract

Spatial autocorrelation is a measure of the correlation of an observation with other observations through space. Most statistical analyses are based on the assumption that the values of observations are independent of one another. Spatial autocorrelation violates this assumption, because observations at nearby locations are related to each other, and hence, the consideration of spatial autocorrelations has been gaining attention in crash data modeling in recent years, and research have shown that ignoring this factor may lead to a biased estimation of the modeling parameters. This paper examines two spatial autocorrelation indices: Moran's Index; and Getis-Ord  $G_i^*$  statistic to measure the spatial autocorrelation of vehicle crashes occurred in Boone County roads in the state of Missouri, USA for the years 2013-2015. Since each index can identify different clustering patterns of crashes, therefore this paper introduces a



# Prostorová statistiky bodových a plošných dat



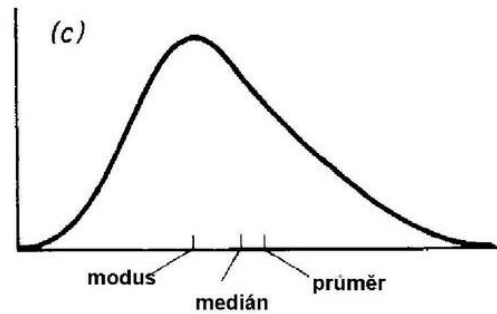


# Popisná statistika bodových objektů

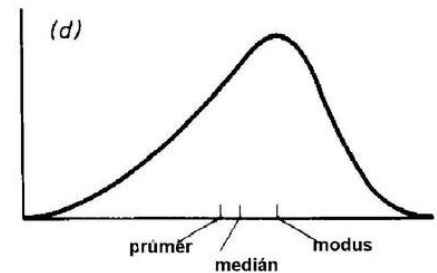
- Charakteristiky polohy
- Charakteristiky rozptylu
- Charakteristiky asymetrie
- Charakteristiky špičatosti



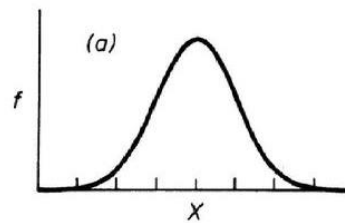
Pozitivní šikmost:



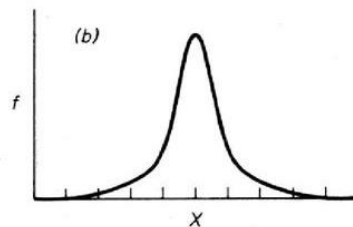
negativní šikmost:



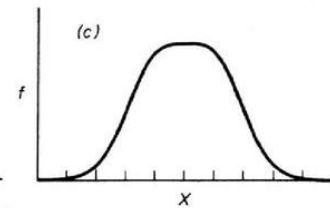
Normální špičatost:



nadnormální:



podnormální:







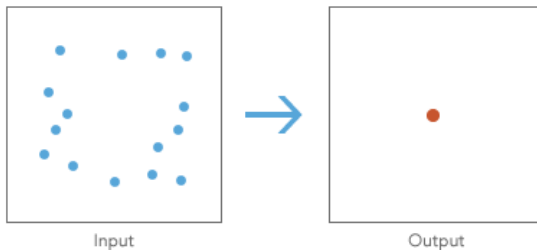
# Charakteristiky polohy

- **Průměrný střed (mean center)**
- **Vážený průměrný střed (weighted mean center)**
- Agregovaný průměrný střed
- **Mediánový střed (median center)**

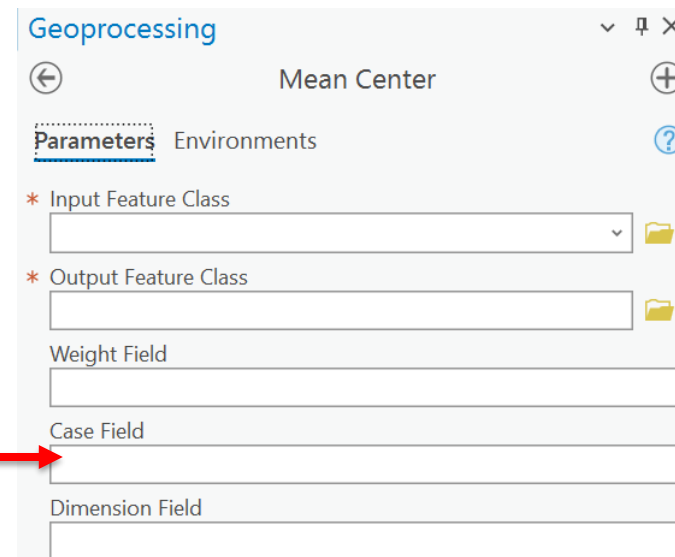


# Průměrný střed

- Průměrný střed leží na průměru souřadnic X a Y.
- Vzorec:
  - $X_{\text{průměr}} = \text{Součet hodnot X souřadnic všech bodů} / \text{počet bodů}$
  - $Y_{\text{průměr}} = \text{Součet hodnot Y souřadnic všech bodů} / \text{počet bodů}$
- Má stejné nevýhody jako aritmetický průměr – je to především citlivost na extrémní hodnoty.
  - Například v případě shlukového uspořádání bodů průměrný střed dobře nereprezentuje množinu bodů



Rozčlenění bodů  
na kategorie

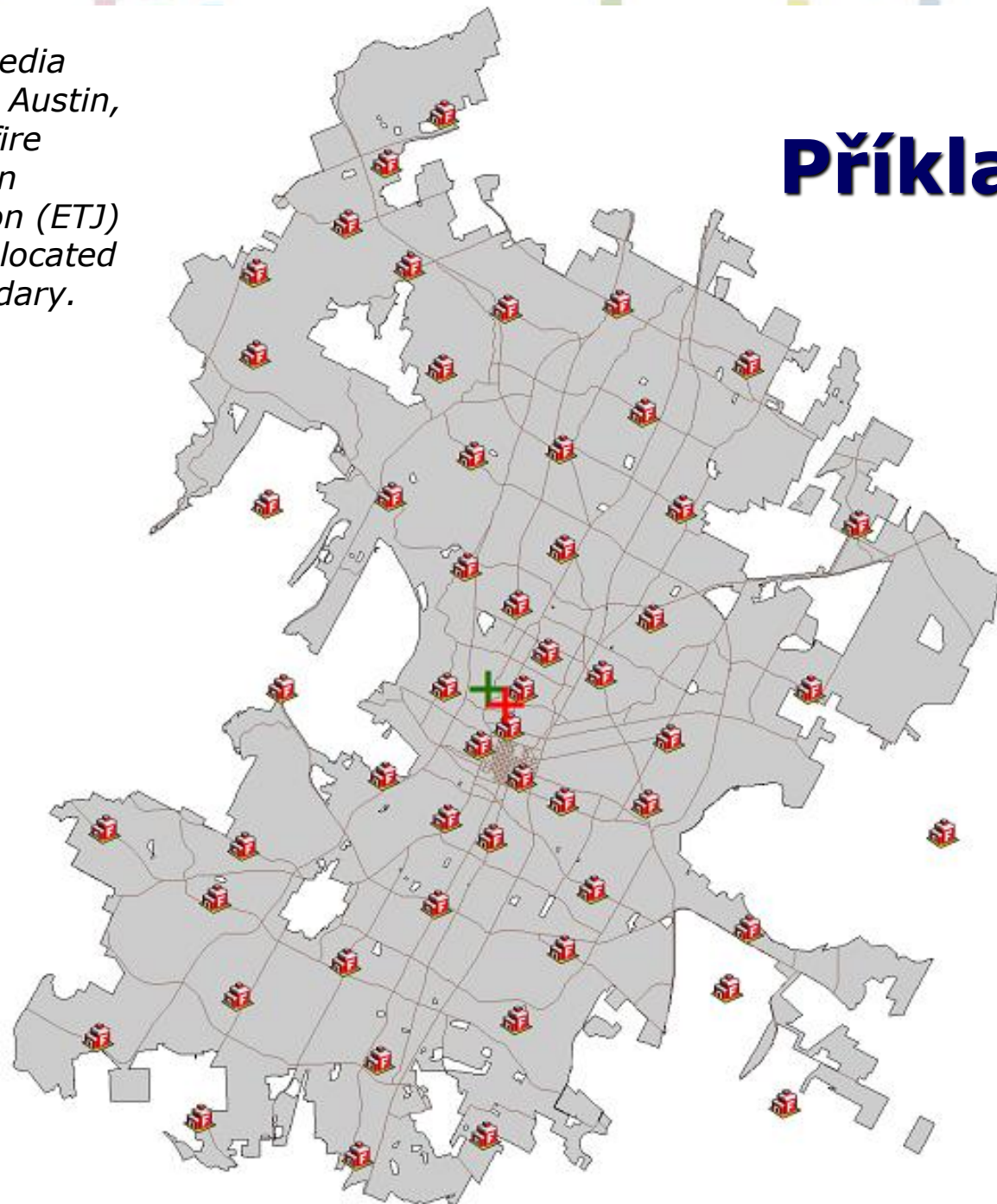


*The mean center and media center of fire stations in Austin, Texas. Note that some fire stations are in the Austin extraterritorial jurisdiction (ETJ) area, and therefore are located outside of the city boundary. Data source: [data.AustinTexas.gov](http://data.AustinTexas.gov)*

# Příklad

## Legend

-  Median Center
-  Mean Center
-  Fire Stations
-  Major Roads
-  Austin City Boundary





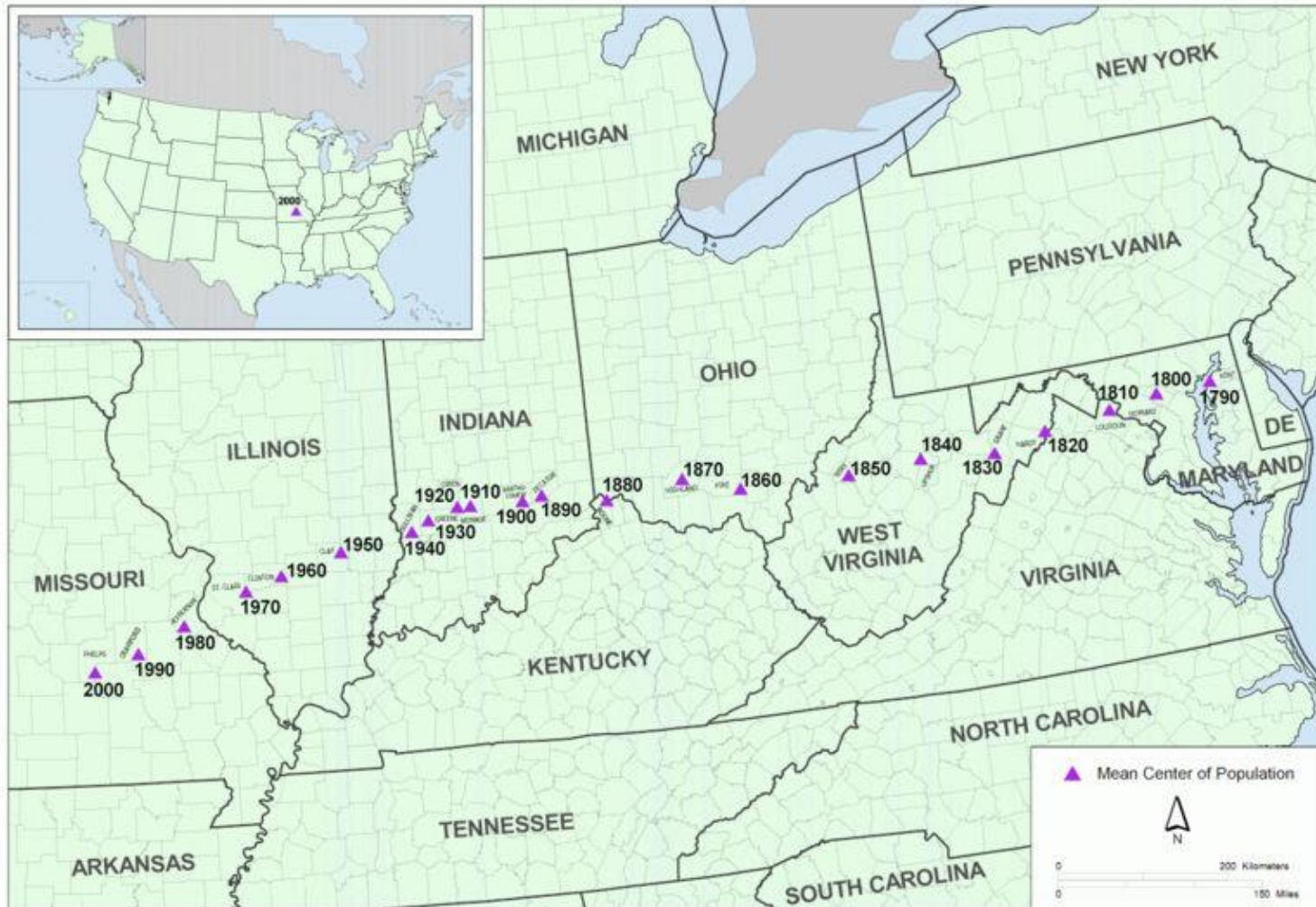
# Vážený průměrný střed

- Používá se v případě výskytu více událostí/objektů na stejném místě.
- Pak má každý bod váhu přímo úměrnou počtu událostí/objektů na tomto místě.
- *Např.: při výpočtu prostorového průměru několika měst bude průměrný střed dávat realističtější představu o centrální tendenci, jestliže ho budeme vážit počtem obyvatel jednotlivých měst*

A screenshot of the 'Mean Center' tool in a Geoprocessing environment. The window title is 'Geoprocessing' and the tool name is 'Mean Center'. Below the title bar, there are navigation icons (back, forward, search, help) and tabs for 'Parameters' and 'Environments'. The 'Parameters' tab is active. The interface includes several input fields:

- 'Input Feature Class': A dropdown menu with a folder icon to its right.
- 'Output Feature Class': A text input field with a folder icon to its right.
- 'Weight Field': A text input field, which is circled in red in the image.
- 'Case Field': A text input field.
- 'Dimension Field': A text input field.

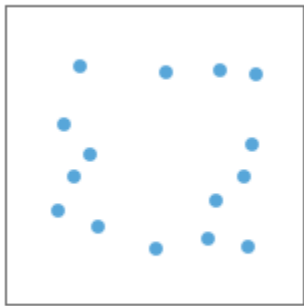
Mean Center of Population for the United States: 1790 to 2000



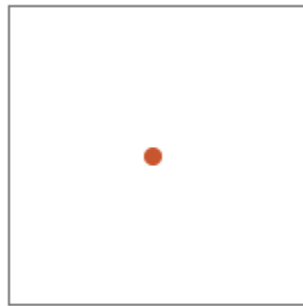


# Mediánový střed

- najdeme medián na ose X a Y a vedeme z nich linie kolmé na směr osy. Takto definovaný „medián ze souřadnic“ ale nemusí odpovídat mediánu souboru bodů, protože distribuce nemusí být mezi kvadranty vyrovnaná.
- (UK): Mediánový střed je střed, kterým se studovaná plocha dělí do čtyř kvadrantů, z nichž každý obsahuje stejný počet bodů.
- (US): Mediánový střed jako střed vyžadující minimální (nejkratší) cestu. Tj. celková vzdálenost z mediánového středu do každého z bodů je minimální. Jinak řečeno – cesta z jakéhokoliv jiného místa do všech bodů oblasti bude delší než cesta z mediánového středu.

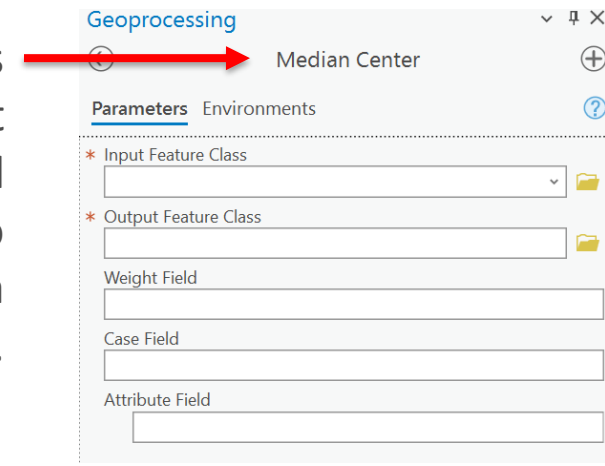


Input



Output

**ArcGIS Pro:** Identifies the location that minimizes overall Euclidean distance to the features in a dataset.





# Vlastnosti charakteristik polohy

- Průměrný střed minimalizuje sumu čtverců vzdáleností
- Mediánový střed minimalizuje sumu vzdáleností – jeho interpretace je jednodušší
- Nejčastěji se využívá váženého mediánového středu (demografie)
  - Příklad: *srovnání vývoje osídlení v čase*
- Charakteristiky polohy bez uvedení charakteristik rozptylu mají malou vypovídací schopnost a mohou být zavádějící



# Charakteristiky rozptylu

- **Směrodatná vzdálenost  
(standard distance circle)**
- **Vážená směrodatná vzdálenost  
(weighted standard distance)**
- **Koeficient relativního rozptylu  
(coefficient of relative dispersion)**
- **Směrodatná elipsa odchylek  
(standard deviational ellipse)**





# Směrodatná vzdálenost

- Směrodatná vzdálenost je nejčastěji používána ve formě kružnice kolem průměrného středu (*Standard distance circle*), jejíž poloměr je právě hodnota směrodatné vzdálenosti.
- Tyto kružnice nám dávají představu o rozptylu hodnot kolem střední hodnoty pro jednotlivé jevů.

A polygon feature class that will contain a circle polygon for each input center. These circle polygons graphically portray the standard distance at each center point.

Geoprocessing Standard Distance

Parameters Environments

\* Input Feature Class

\* Output Standard Distance Feature Class

Circle Size

1 standard deviation

Weight Field

Case Field



# Směrodatná vzdálenost

- Mohou být použity i pro studium dynamiky jevů
  - př.: *různé kružnice pro jeden jev v různých časových horizontech*
- Směrodatná vzdálenost (standard distance) **je absolutní** mírou – je problematické její použití k porovnání několika souborů
- Vhodnější jsou míry **relativní**

## Vážená směrodatná vzdálenost

\* Output Standard Distance Feature Class

Circle Size

1 standard deviation

Weight Field

Case Field

The screenshot shows a software interface for calculating weighted standard distance. The 'Weight Field' label and its corresponding input box are circled in red. The 'Circle Size' dropdown menu is set to '1 standard deviation'. The 'Output Standard Distance Feature Class' field is empty and has a folder icon to its right. The 'Case Field' field is also empty.

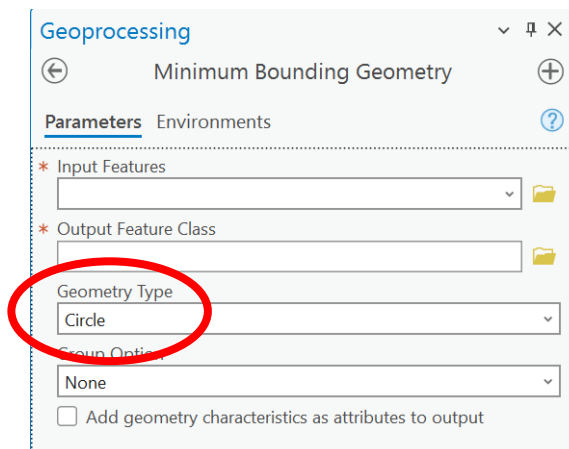


# Koeficient relativního rozptylu

- Poměr směrodatné vzdálenosti a poloměru kruhu se stejnou plochou jakou má studovaná oblast.
- Řeší problém použití absolutní míry směrodatné vzdálenosti.
- Je-li oblast různě velká (ohraničená), vznikají zavádějící hodnoty.
- K získání relativní míry při studiu variability obyvatelstva se někdy používá poloměr země nebo státu místo poloměru kruhu se stejnou plochou jakou má studovaná oblast.

$$CRD = 100 * \frac{S_d}{A_k} = 100 * \frac{S_d}{\sqrt{\frac{R}{\pi}}} = 100 * S_d * \sqrt{\frac{\pi}{R}}$$

poloměr z plochy kruhu:  $R = \sqrt{(P/\pi)}$



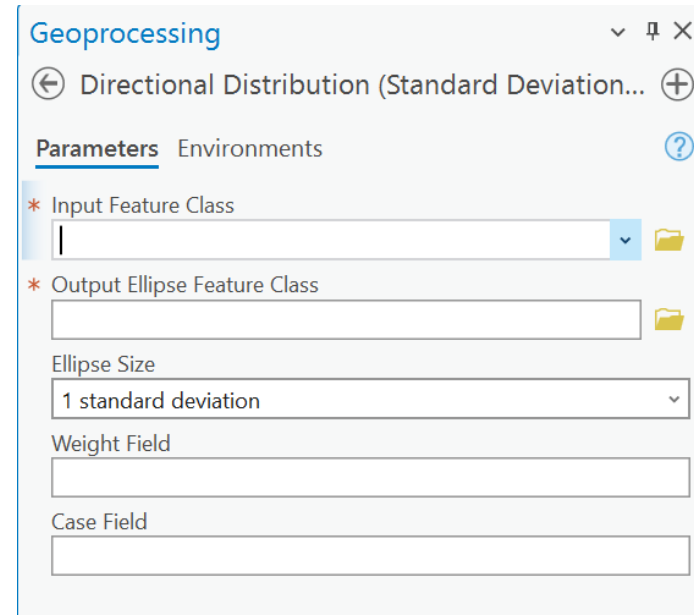
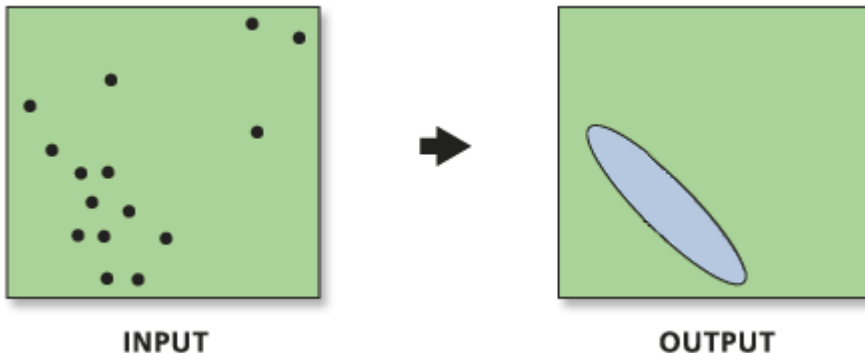


# Směrodatná elipsa odchylek

- V mnoha případech může vykazovat prostorové rozdělení jevů určité rysy směrovosti (directional bias), např.:
  - *rozdělení míst nejčastějších dopravních nehod podél dálnice.*
- V tomto případě se použití kružnice jako míry rozptylu hodnot jeví jako nevhodné.
- Jako logické rozšíření směrodatné kružnice odchylek se může jevit použití směrodatné elipsy odchylek. Tuto elipsu popisují tři atributy:
  - úhel rotace
  - směrodatná odchylka podél hlavní osy elipsy
  - směrodatná odchylka podél vedlejší osy elipsy
- Maximální rozptyl bude orientován v souladu s hlavní osou elipsy.

# Směrodatná elipsa odchylek

- Př.:
  - Množství kontaminující látky ve vzorku studní může indikovat trend jejich šíření
  - Porovnání velikosti, tvaru resp. překryvu elips k porovnání změn v rozšiřování etnik či rostlinných resp. živočišných společenstev
  - Epidemiologie – vystižení hlavního trendu šíření onemocnění v populaci





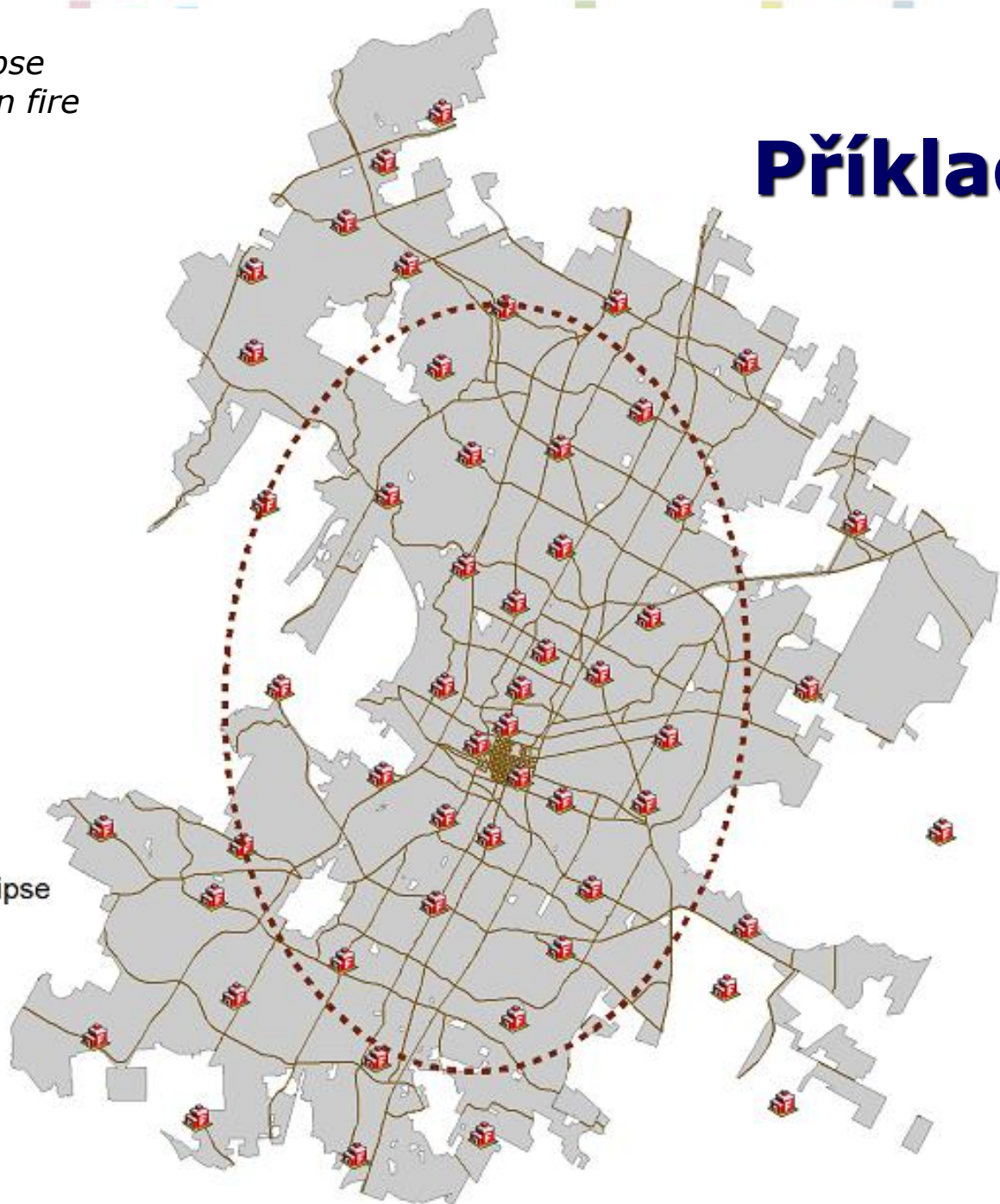
Standard deviation ellipse  
created based on Austin fire  
station locations

Data source:  
[data.AustinTexas.gov](http://data.AustinTexas.gov)

# Příklad

## Legend

-  Standard Deviation Ellipse
-  Fire Stations
-  Major Roads
-  Austin City Boundary





# Poznámky k deskripci bodů

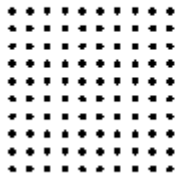
- hustota bodů v ploše (počet/plocha =  $n/R$ ),
- charakteristiky založené na vzdálenosti mezi body či na relativních vzdálenostech jako je např.  $d_i/d_{max}$ .
- použití – porovnávání (např. v čase)
- při výpočtech v relativně malých oblastech používáme euklidovskou geometrii, protože se v nich neprojeví zakřivení Země.
- uvedené míry mohou být aplikovány i na plochy.
  - **Jakým způsobem?**



# Základní metody statistického popisu prostorového uspořádání bodů

- **Analýza kvadrátů** – testujeme, zda rozmístění bodů v ploše je náhodné či nikoliv.
- **Metoda nejbližšího souseda** – porovnává průměrnou vzdálenost mezi nejbližšími sousedy pole bodů k teoretickému rozmístění.

Regular



Random



Clustered



- **Prostorová autokorelace** – měří jak podobné či nepodobné jsou hodnoty atributů sousedních bodů.



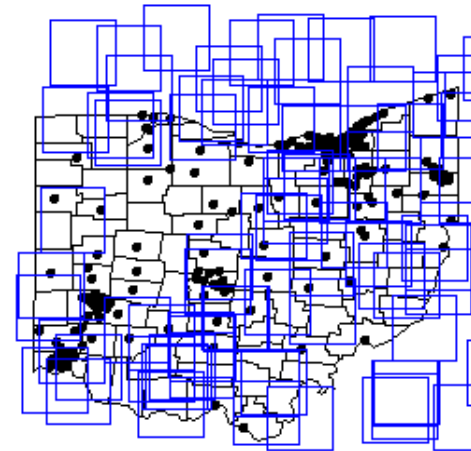
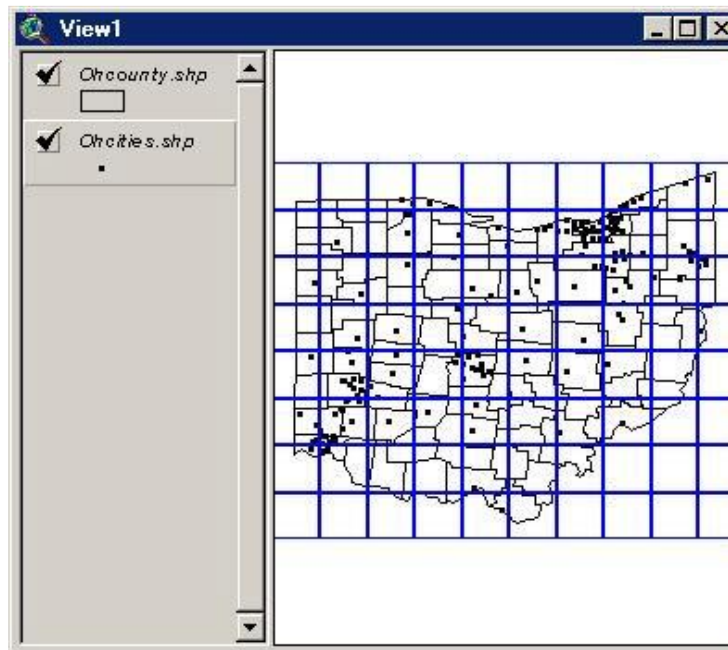


# Metody statistického popisu bodů – obecně

- Rozmístění bodů v prostoru je výsledkem určitých procesů či vhodných podmínek (lokace měst je výsledkem působení faktorů jako reliéf, přírodní zdroje, komunikace, atd.)
- Cílem studia prostorového rozmístění bodů je zjistit:
  - jak daleko má konkrétní rozmístění objektů k rozmístění teoretickému
  - jak se liší rozmístění bodů ve dvou různých oblastech
  - jak se mění rozmístění bodů v rámci jedné oblasti v čase.
- Statisticky prokázaný výskyt určitého prostorového uspořádání může být základem pro zjišťování příčin, které vedly k pozorovanému uspořádání.
- Problémy:
  - měřítko
  - rozsah studované oblasti
  - kartografická projekce

# Analýza kvadrátů

- Je založena na hodnocení změn hustoty bodů v prostoru.
- Je porovnáváno, zda rozmístění bodů v prostoru je náhodné, či má blíže k uspořádání shlukovému či pravidelnému.
- Studovaná plocha je rozdělena pravidelnou sítí na buňky a je zjištěn počet bodů v každé buňce.

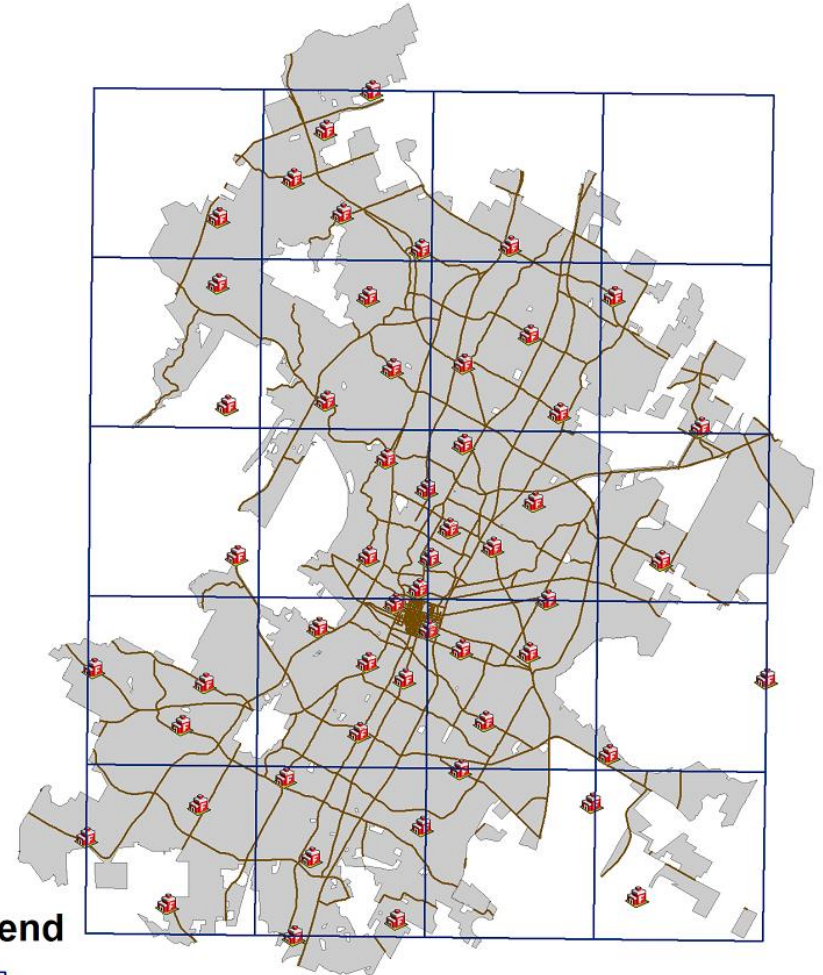


# Analýza kvadrátů





- Je analyzováno rozdělení četností buněk s určitým počtem bodů.
- Toto rozdělení je porovnáváno s náhodným rozdělením četností.
  - **Extrémně shlukové uspořádání** – většina bodů v jedné či několika málo buňkách.
  - **Extrémně pravidelné** – ve všech buňkách přibližně stejně
- Buňky se označují jako kvadráty a nemusí jít o čtverce, ale např. i o kruhy či šestiúhelníky – je to dáno empirií.
- V rámci jedné analýzy však tvar a velikost buněk musí být konstantní.

# Analýza kvadrátů

- Optimalní velikost kvadrátů (QS)
  - $QS = (2 \cdot A) / n$
  - A: plocha studované oblasti
  - n: počet analyzovaných bodů.
- Velikost strany vhodného kvadrátu
  - $\sqrt{(2A/n)}$



## Legend

-  Quadrats
-  Fire Stations
-  Major Roads
-  City Boundary

*Study area has been divided into 4\*5 uniformly shaped quadrats of 81 km<sup>2</sup>, and the top left quadrat has a density of 1/81.*

## Praktický postup testování výsledků analýzy kvadrátů

1. ( $H_0$ ) - neexistuje statisticky významný rozdíl (je-li rozdíl malý, může být výsledkem náhody, čím je větší, s tím větší pravděpodobností náhodný není, ale je statisticky významný).

2. Zvolíme hladinu významnosti  $\alpha = 0,05$

3. Vypočteme kumulované četnosti

4. Vypočteme testovací kritérium:  $D = \max|O_i - E_i|$

5. Vypočteme kritickou hodnotu:  $D_\alpha = \frac{1,36}{\sqrt{m}}$   $D_\alpha = 1,36 \sqrt{\frac{m_1 + m_2}{m_1 \cdot m_2}}$

6. Je-li vypočtená hodnota D větší než kritická hodnota  $D_\alpha$ , potom rozdíl mezi oběma uspořádáními je statisticky významný.

# Analýza kvadrátů

## Testování výsledků analýzy kvadrátů K-S testem

Počet měst v každém čtverci	Zjištěné rozdělení	Relativní četnosti	Kumulativní četnosti	Pravidelné rozdělení	Relativní četnosti	Kumulativní četnosti	Absolutní diference
0	36	0,450	0,450	0	0,000	0,00	0,45
1	17	0,213	0,663	26	0,325	0,33	0,34
2	10	0,125	0,788	26	0,325	0,65	0,14
3	3	0,038	0,825	26	0,325	0,98	0,15
4	2	0,025	0,850	2	0,025	1,00	0,15
5	2	0,025	0,875	0	0,000	1,00	0,13
6	1	0,013	0,888	0	0,000	1,00	0,11
7	1	0,013	0,900	0	0,000	1,00	0,10
8	1	0,013	0,913	0	0,000	1,00	0,09
9	1	0,013	0,925	0	0,000	1,00	0,08
10	1	0,013	0,938	0	0,000	1,00	0,06
11	1	0,013	0,950	0	0,000	1,00	0,05
12	1	0,013	0,963	0	0,000	1,00	0,04
13	1	0,013	0,975	0	0,000	1,00	0,03
14	1	0,013	0,988	0	0,000	1,00	0,01
28	1	0,013	1,000	0	0,000	1,00	0,00
164	0	0,000	1,000	0	0,000	1,00	0,00

Testovací kritérium:

$D = 0,45$

Kritická hodnota pro  $\alpha = 0,05$ :

$D_\alpha = 0,2115$

Zamítáme nulovou hypotézu - rozdělení měst se statisticky významně liší od rozdělení pravidelného

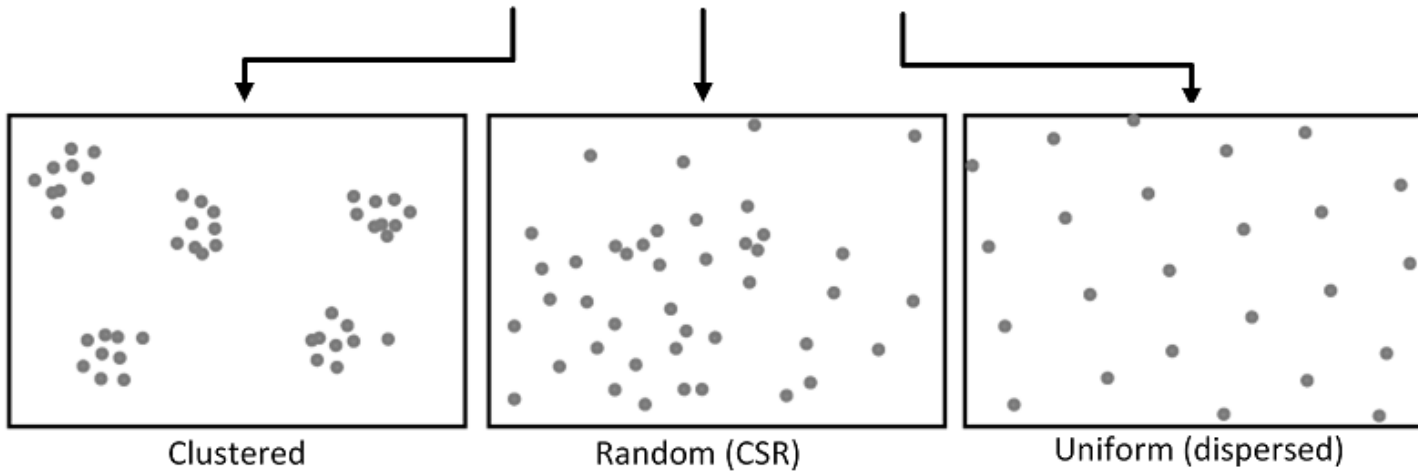
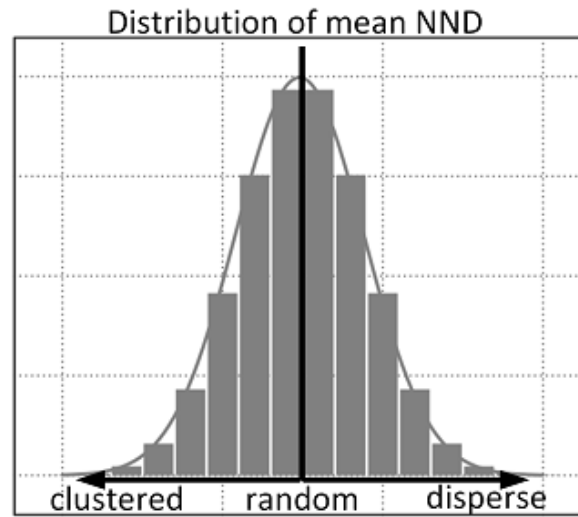
Materiály předmětu  
Z6101 Základy  
geostatistiky



# Metoda nejbližšího souseda

- NEAREST NEIGHBOUR ANALYSIS
- Metoda analýzy kvadrátů je založena na konceptu hustoty (počet bodů v ploše)
- **Metoda analýzy nejbližšího souseda** je naopak založena na konceptu vzdálenosti (spacing – plocha připadající na bod).
- Metoda analýzy nejbližšího souseda je založena na porovnání pozorované průměrné vzdálenosti mezi nejbližšími sousedy a této průměrné vzdálenosti u známého (teoretického) prostorového uspořádání (pravidelného či náhodného).
- K testování, zda má určité rozložení bodů v ploše jistý vzorek lze využít R statistiku (*R - randomness*).

# Metoda nejbližšího souseda





# Metoda nejbližšího souseda

## R statistika

Určí se jako poměr mezi pozorovanou a očekávanou průměrnou vzdáleností nejbližších sousedů v určité oblasti:

$$R = \frac{r_{obs}}{r_{exp}}$$

Hodnotu  $r_{obs}$  zjistíme tak, že určíme vzdálenost mezi daným bodem a všemi jeho sousedy. Dále najdeme nejkratší vzdálenost – tedy nejbližšího souseda. Tento proces se opakuje pro všechny body. Ze všech nejkratších vzdáleností se vypočte průměr.

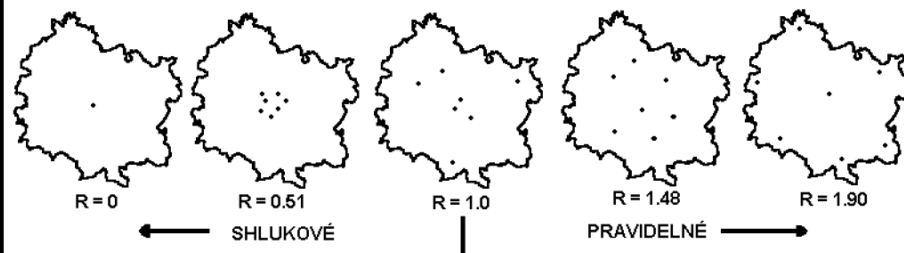
Hodnotu  $r_{exp}$  zjistíme ze vztahu:

$$r_{exp} = \frac{1}{2\sqrt{n/A}}$$

## Interpretace hodnot R statistiky

Čím je hodnota  $R < 1$ , tím více se prostorové rozložení bodů blíží rozložení shlukovému ( $r_{obs} < r_{exp}$ ).

Čím je hodnota  $R > 1$ , tím více se prostorové rozložení bodů blíží rozložení pravidelnému ( $r_{obs} > r_{exp}$ ).



R = 0      zcela shlukové uspořádání

R = 1      náhodné uspořádání

R = 2,149      zcela pravidelné uspořádání



## Geoprocessing

### Average Nearest Neighbor

Parameters Environments

\* Input Feature Class

Distance Method

Euclidean

Generate Report

Area

# Metoda nejbližšího souseda

## Average Nearest Neighbor Summary

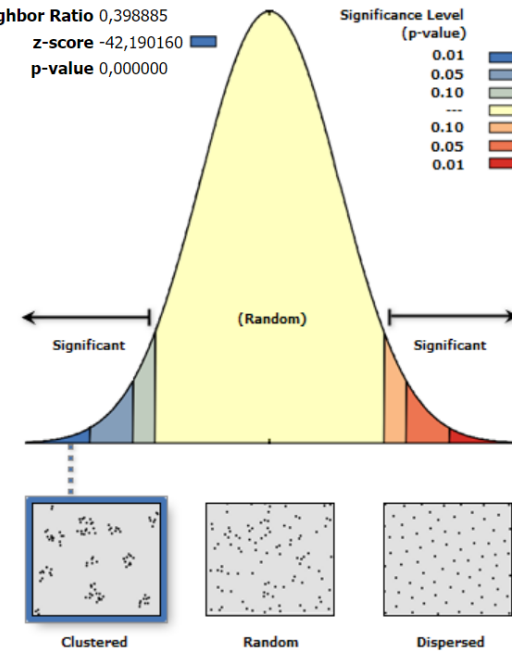
Nearest Neighbor Ratio 0,398885

z-score -42,190160

p-value 0,000000

Significance Level  
(p-value)

Significance Level (p-value)	Critical Value (z-score)
0.01	< -2.58
0.05	-2.58 - -1.96
0.10	-1.96 - -1.65
---	-1.65 - 1.65
0.10	1.65 - 1.96
0.05	1.96 - 2.58
0.01	> 2.58



Given the z-score of -42.19016, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

## Average Nearest Neighbor Summary

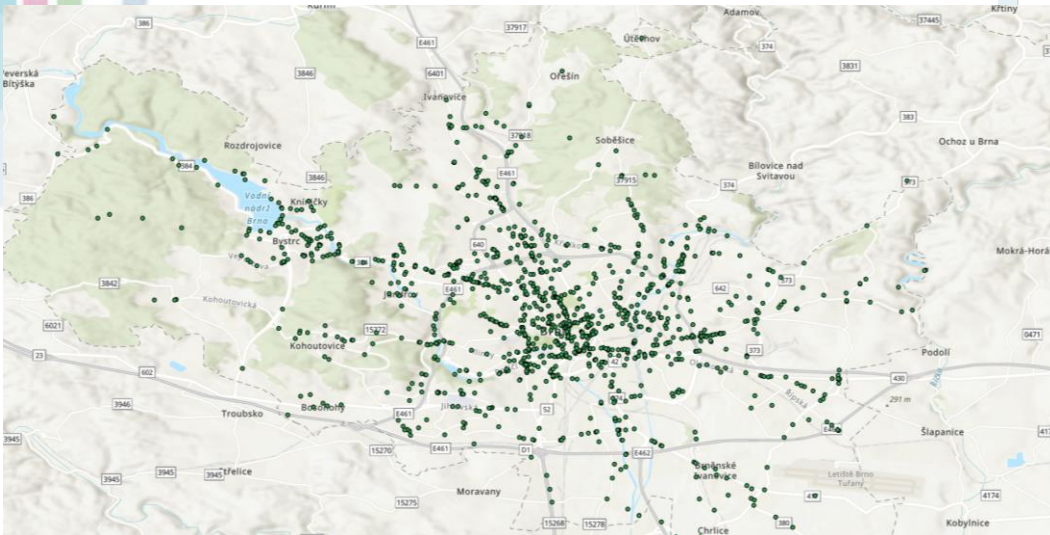
**Observed Mean Distance** 98,1936 Meters

**Expected Mean Distance** 246,1702 Meters

**Nearest Neighbor Ratio** 0,398885

**z-score** -42,190160

**p-value** 0,000000



## Average Nearest Neighbor Summary

Observed Mean Distance	98,193552
Expected Mean Distance	246,170249
Nearest Neighbor Ratio	0,398885
z-score	-42,190160
p-value	0,000000



# Metoda nejbližšího souseda

- Nelze spolehat na vizuální srovnání prostorového rozložení ani na vypočtenou hodnotu R. Ta by měla být doplněna hodnotou  $Z_R$  (Z skóre) pro ověření statistické významnosti pozorovaného rozdílu.
- Výsledky jsou vysoce citlivé k měřítku (lokální vs. regionální)
- V závislosti na studovaném jevu musí být věnována pozornost vymezení studované plochy (administrativní či přirozené hranice).

Pomocí směrodatné chyby lze vypočítat standardizovanou hodnotu (Z-score):

$$Z_R = \frac{r_{obs} - r_{exp}}{SE_r} \quad SE_r = \frac{0,26136}{\sqrt{n^2/A}}$$



# Prostorová autokorelace – koncepce

- Jak analýza kvadrátů tak analýza vzdálenosti nejbližšího souseda pracují pouze s polohou bodů.
- Nerozlišují body podle hodnot jejich atributů.
- Oba parametry (polohu i atributy) hodnotí prostorová autokorelace (SA) – je tedy metodou vhodnější.
- **Východiska prostorové autokorelace:** Většina jevů se v prostoru mění spojitě. Blízké body budou mít i podobné hodnoty studovaného jevu a naopak.
  - *First law of geography – Tobler (1970)*



# Prostorová autokorelace

- Mezi nejpoužívanější koeficienty prostorové autokorelace náleží:
  - **Gearyho poměr C** (Geary's Ratio)
  - **Moranův index I** (Moran's I)
- Lze jich využít pro intervalová a poměrová data.

Rozdíly mezi oběma indexy jsou dány způsobem výpočtu rozdílů mezi hodnotami atributu. Obor hodnot, kterých mohou oba indexy nabývat se tedy také liší, jak uvádí následující tabulka:

Prostorové uspořádání	Gearyho poměr C	Moranův index I
Shlukové uspořádání, sousední body vykazují podobné hodnoty	$0 < C < 1$	$I > E(I)$
Náhodné uspořádání, body nevykazují znaky podobnosti	$C \sim 1$	$I \sim E(I)$
Pravidelné uspořádání, sousední body vykazují rozdílné charakteristiky	$1 < C < 2$	$I < E(I)$

kde  $E(I) = (-1)/(n-1)$  je očekávaná hodnota indexu



- **Jak autor popisuje p-hodnotu (p-value) a Z-skóre? Respektive jaká je mezi nimi souvislost?**



Scientific  
Research  
Publishing

Open Journal of Civil Engineering, 2017, 7, 208-221

<http://www.scirp.org/journal/ojce>

ISSN Online: 2164-3172

ISSN Print: 2164-3164

## A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord $G_i^*$ Statistic

Azad Abdulhafedh\*

Department of Civil and Environmental Engineering, University of Missouri, Columbia, MO, USA

Email: asa8cd@mail.missouri.edu

How to cite this paper: Abdulhafedh, A. (2017) A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord  $G_i^*$  Statistic. *Open Journal of Civil Engineering*, 7, 208-221. <https://doi.org/10.4236/ojce.2017.72013>

Received: February 13, 2017

Accepted: June 3, 2017

Published: June 6, 2017

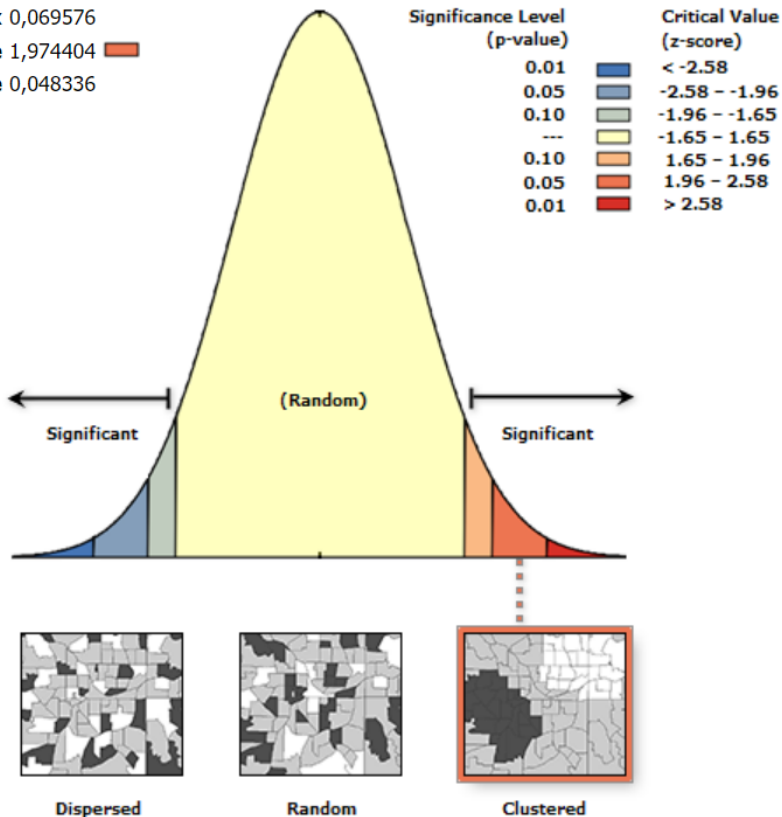
Copyright © 2017 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

### Abstract

Spatial autocorrelation is a measure of the correlation of an observation with other observations through space. Most statistical analyses are based on the assumption that the values of observations are independent of one another. Spatial autocorrelation violates this assumption, because observations at nearby locations are related to each other, and hence, the consideration of spatial autocorrelations has been gaining attention in crash data modeling in recent years, and research have shown that ignoring this factor may lead to a biased estimation of the modeling parameters. This paper examines two spatial autocorrelation indices: Moran's Index; and Getis-Ord  $G_i^*$  statistic to measure the spatial autocorrelation of vehicle crashes occurred in Boone County roads in the state of Missouri, USA for the years 2013-2015. Since each index can identify different clustering patterns of crashes, therefore this paper introduces a

# Spatial Autocorrelation Report

Moran's Index 0,069576  
 z-score 1,974404  
 p-value 0,048336



Given the z-score of 1.974404, there is a less than 5% likelihood that this clustered pattern could be the result of random chance.

## Global Moran's I Summary

<b>Moran's Index</b>	0,069576
<b>Expected Index</b>	-0,000851
<b>Variance</b>	0,001272
<b>z-score</b>	1,974404
<b>p-value</b>	0,048336

# autokorelace

Geoprocessing

Spatial Autocorrelation (Global Moran's I)

Parameters

Input Feature Class: cyklo\_nehody\_cykl\_neh\_xytabletpoint\_spa1

Input Field: hmotna\_sko

Generate Report

Conceptualization of Spatial Relationships: **Inverse distance**

Standardization: None

Distance Band or Threshold: Distance

na mesic t den  
 11 září pátek

# 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



# Prostorová autokorelace

- 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)



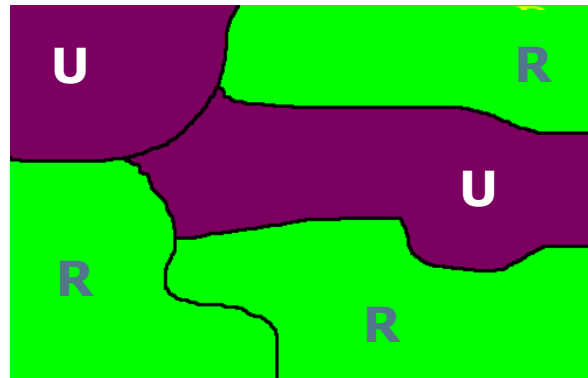


# Míry prostorové autokorelace areálů

- **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)

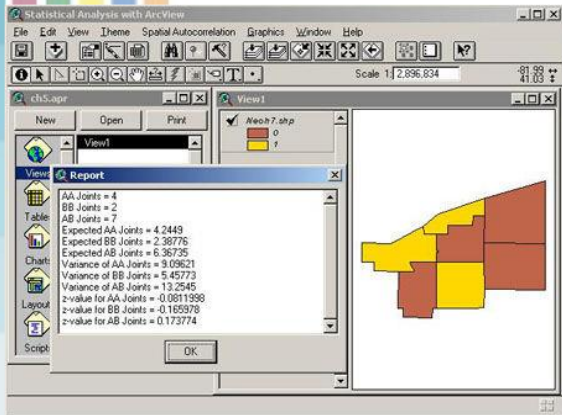
- Touto metodou lze zjistit, zda uspořádání ploch, které mohou nabývat binárních hodnot vykazuje prvky náhodnosti.
- Tedy zda existuje pozitivní (clustered pattern) či negativní (random pattern) prostorová autokorelace.



- 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

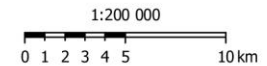


# Joint count statistics (JCS)



## 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

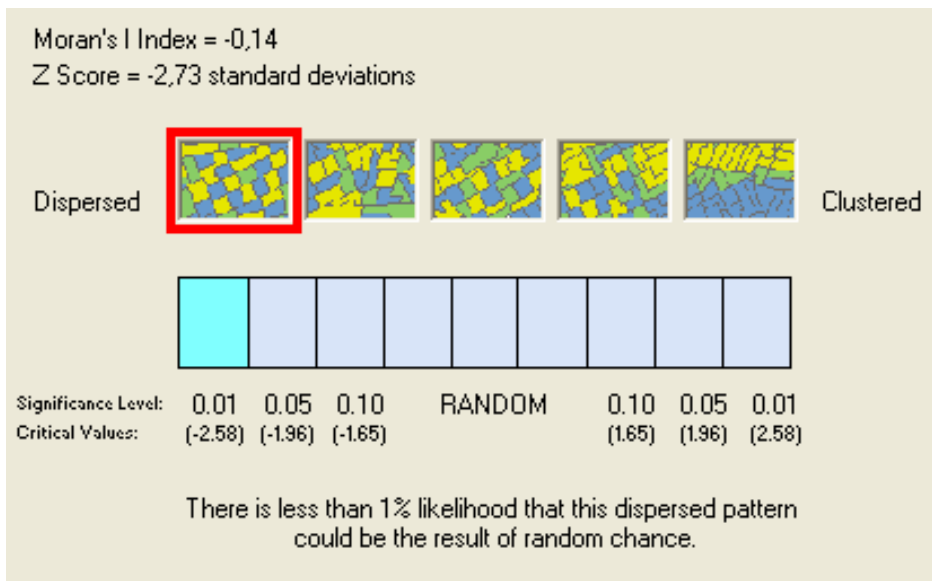


# Indexy pro hodnocení prostorové autokorelace plošných jevů

- **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  $\alpha=0,05$ .



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



# Moranův (I) index

- 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

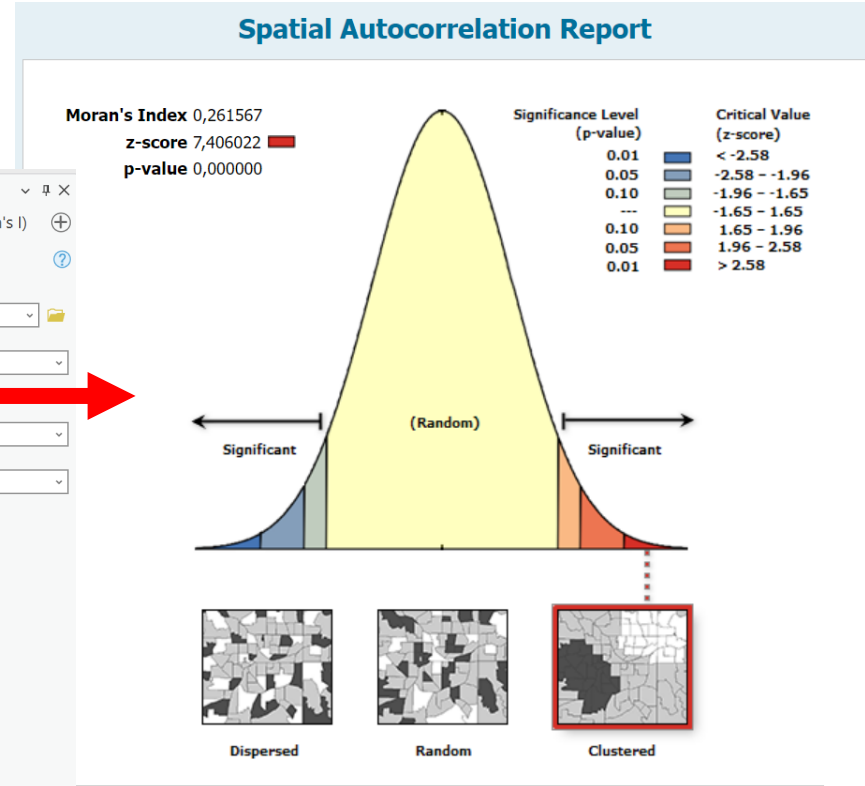
Input Feature Class  
 brno\_obyv\_20-9-2021

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**

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



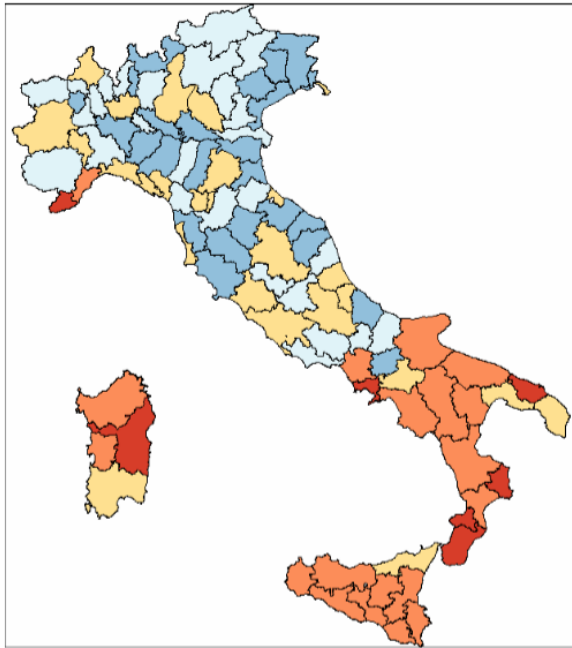
# Gearyho poměr C

- 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.

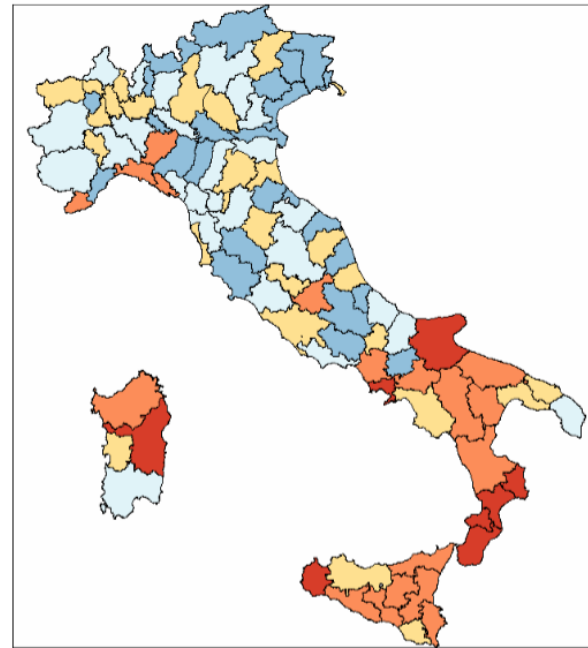
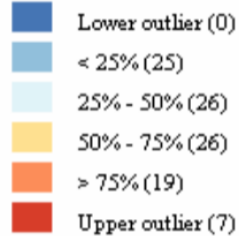
1999

2003

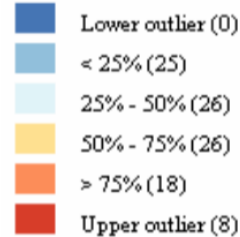
# Porovnání



BoxMap (Hinge=1.5) : MURDER\_99



BoxMap (Hinge=1.5) : MURDER\_03



[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





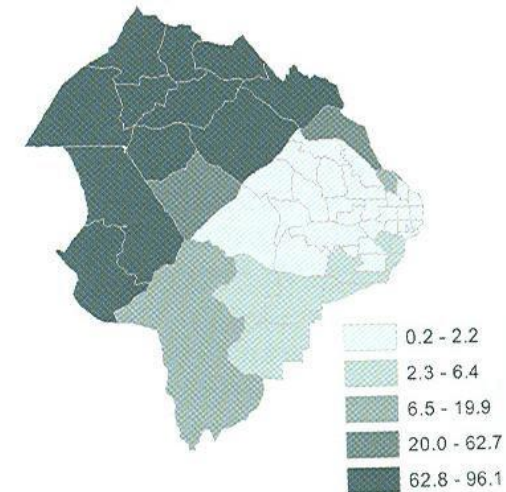
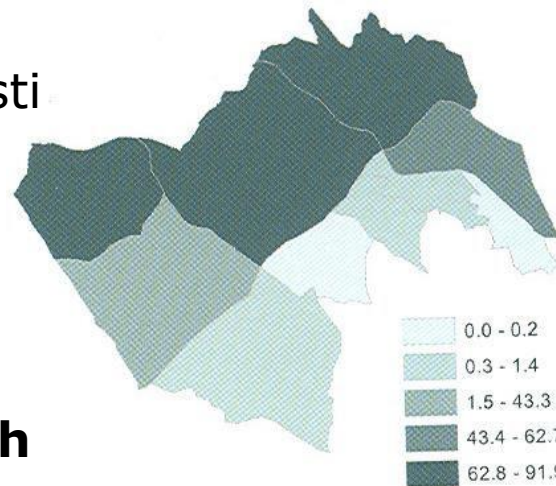
# Omezení globálních měr I, C

- 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 tak, aby posuzovaná plocha měla **alespoň jednoho** 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	Gauga	Cuyahoga	Trumbull	Summit	Portage	Ashtabula	Lake
Gauga	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



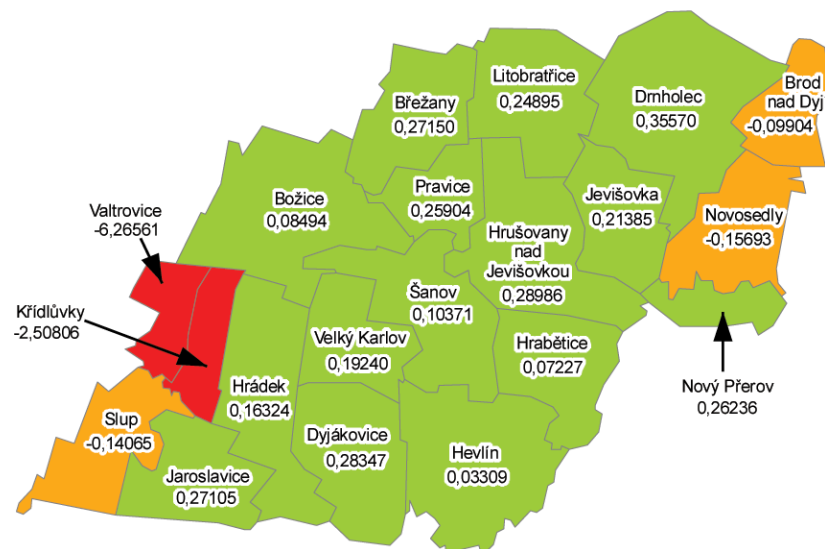
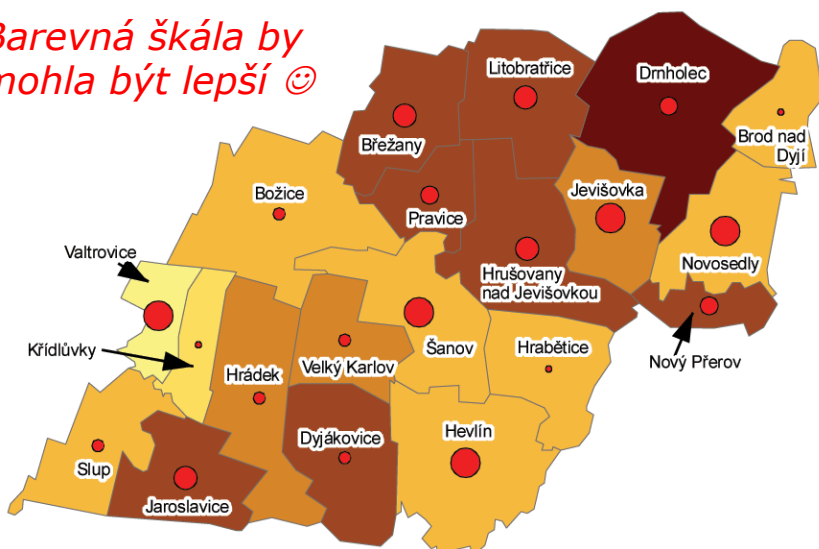
# Lokální statistiky prostorové autokorelace

- 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.

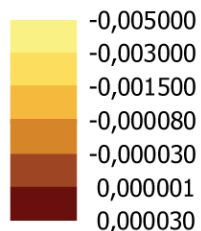


# Lokální Moranův index

Barevná škála by mohla být lepší 😊

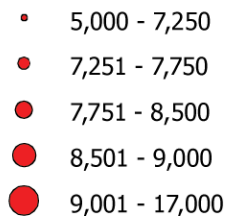


## Moranův index



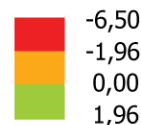
## Podíl

### Poměr obyvatel starších 64 let



## Z skóre

### Moranova indexu



hranice obce a příslušné z skóre

Lukáš HERMAN, 4. ročník KART  
Geografický ústav PĚF MU Brno  
Vytvořeno v květnu 2010 v Brně  
Zdroj dat: ČSÚ  
Souřadnicový systém: S-42

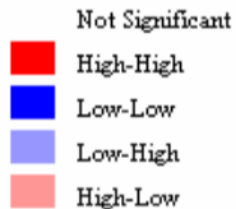
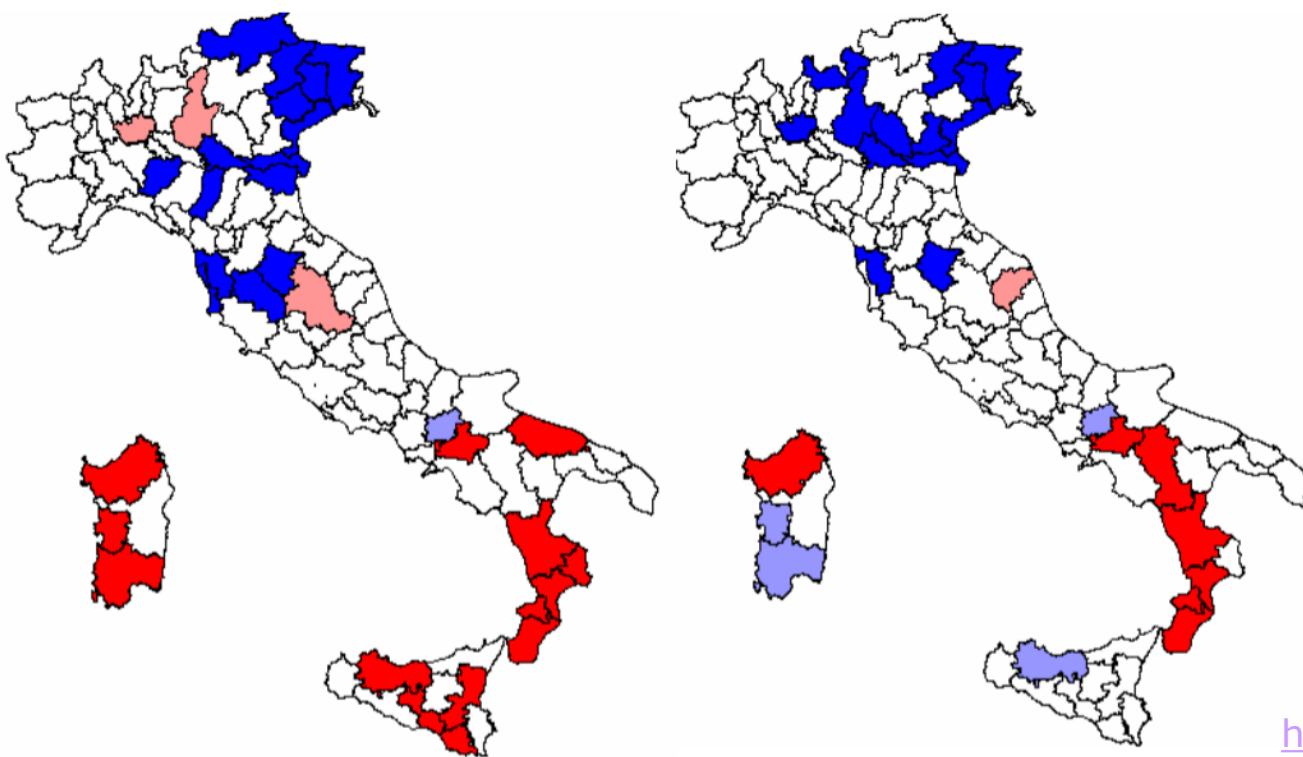


# Lokální statistiky prostorové autokorelace

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](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

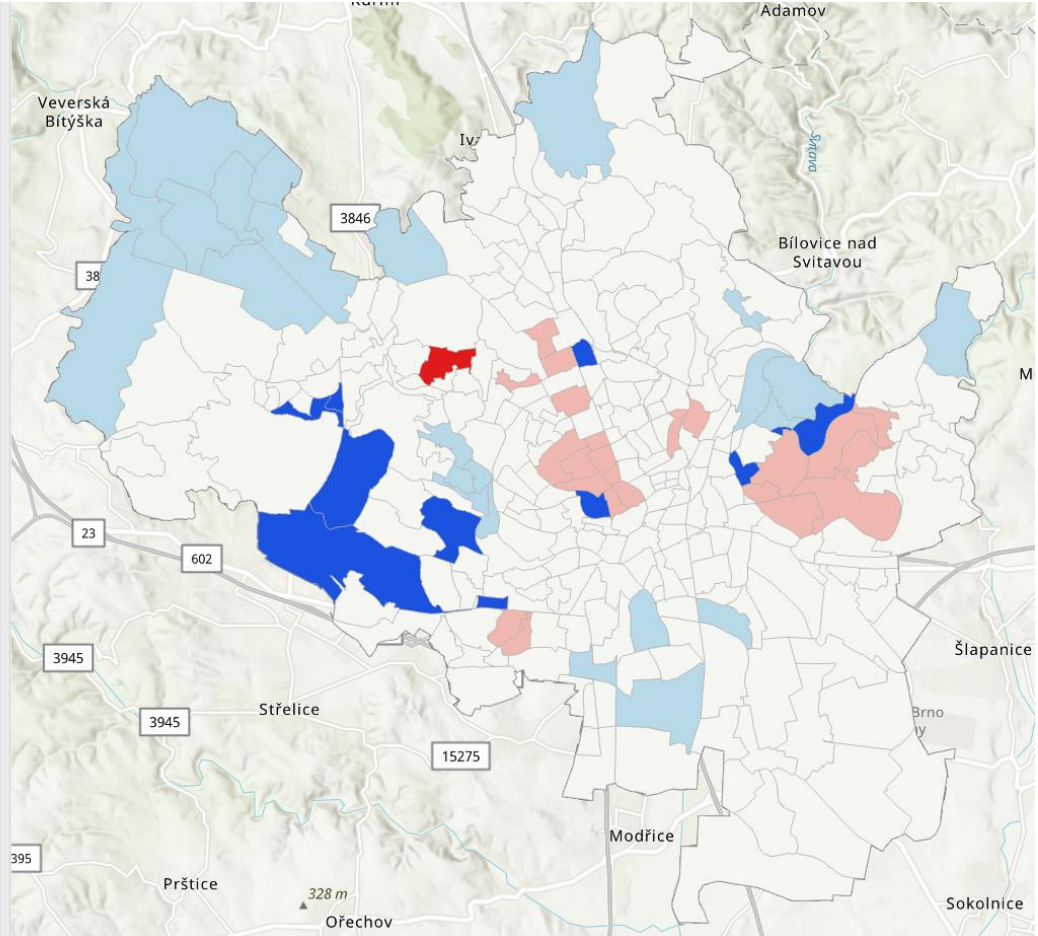
## ArcGIS Pro: Cluster and Outlier Analysis (Anselin Local Moran's I)



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

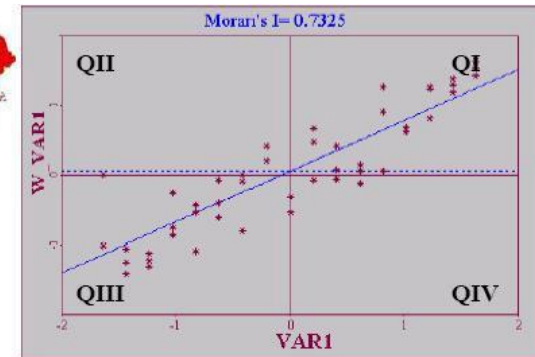
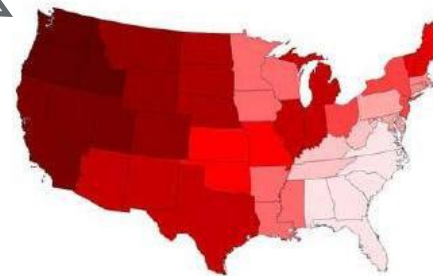
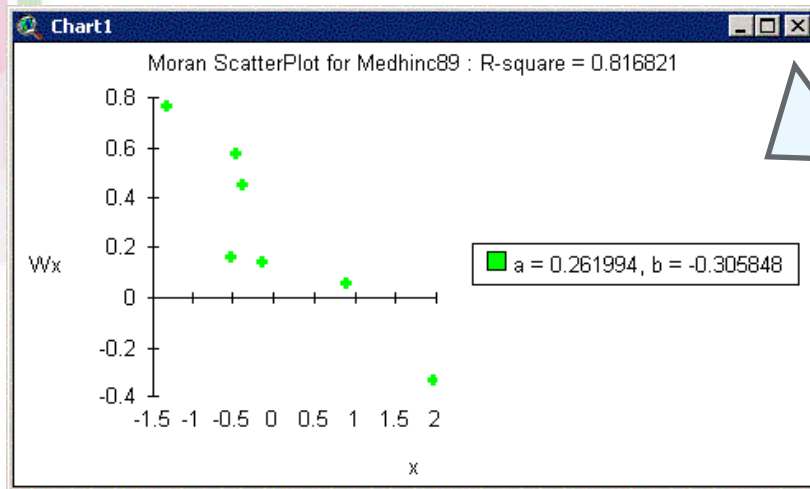


# Moranovo korelační pole

- 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:

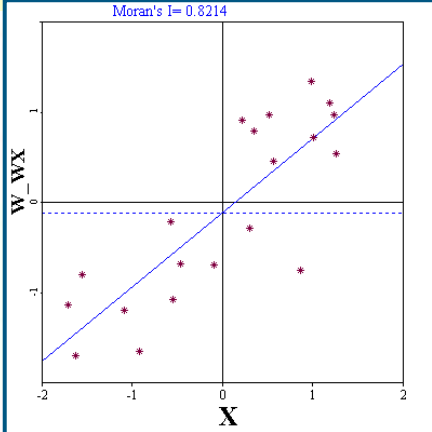
- high-high, low-low (2. nebo 3. kvadrant) = spatial clusters
- high-low, low-high (1. nebo 4. kvadrant) = spatial outliers



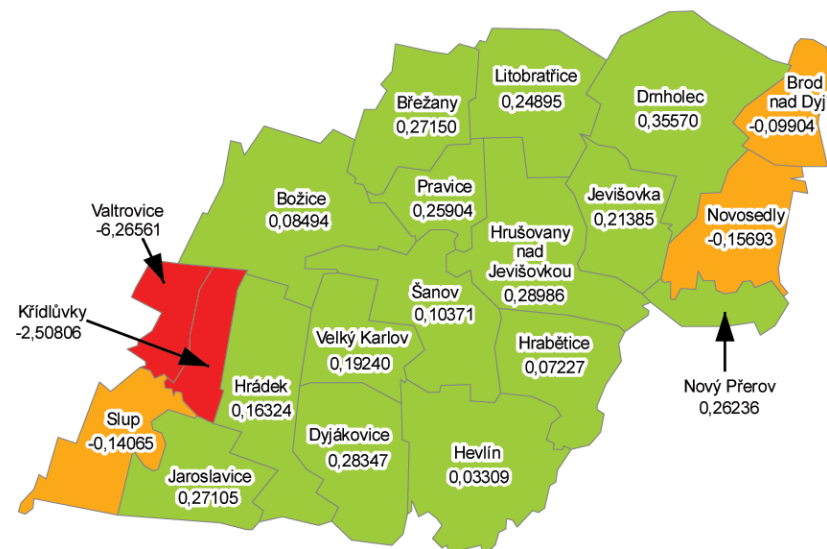
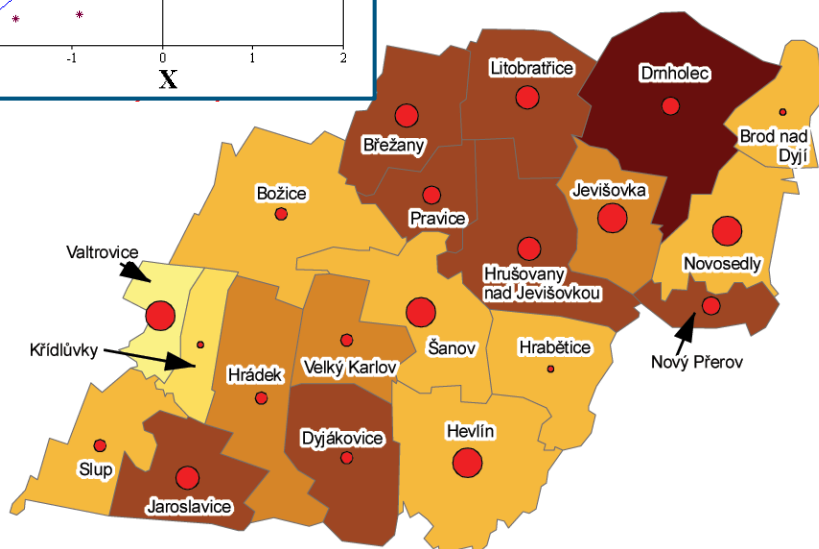
Hodnota Moranova indexu indikuje slabou negativní prostorovou autokorelaci (celky s vysokou hodnotou studovaného atributu jsou blízko celků s nízkými hodnotami).

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

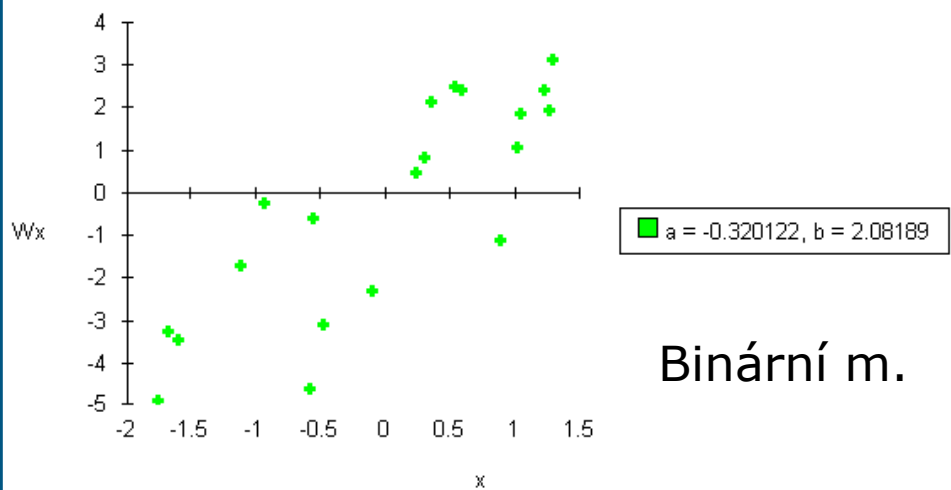
Moran's I = 0.8214



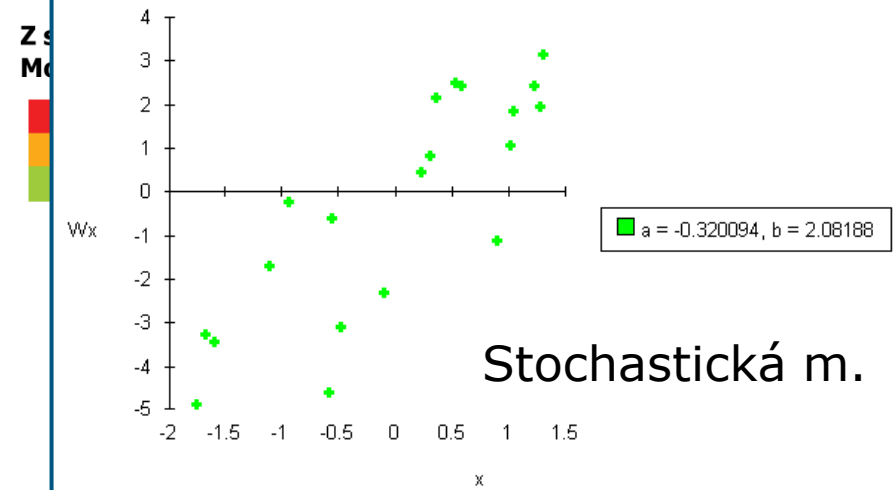
# Moranovo korelační pole



Moran ScatterPlot for Obce\_id : R-square = 0.685529



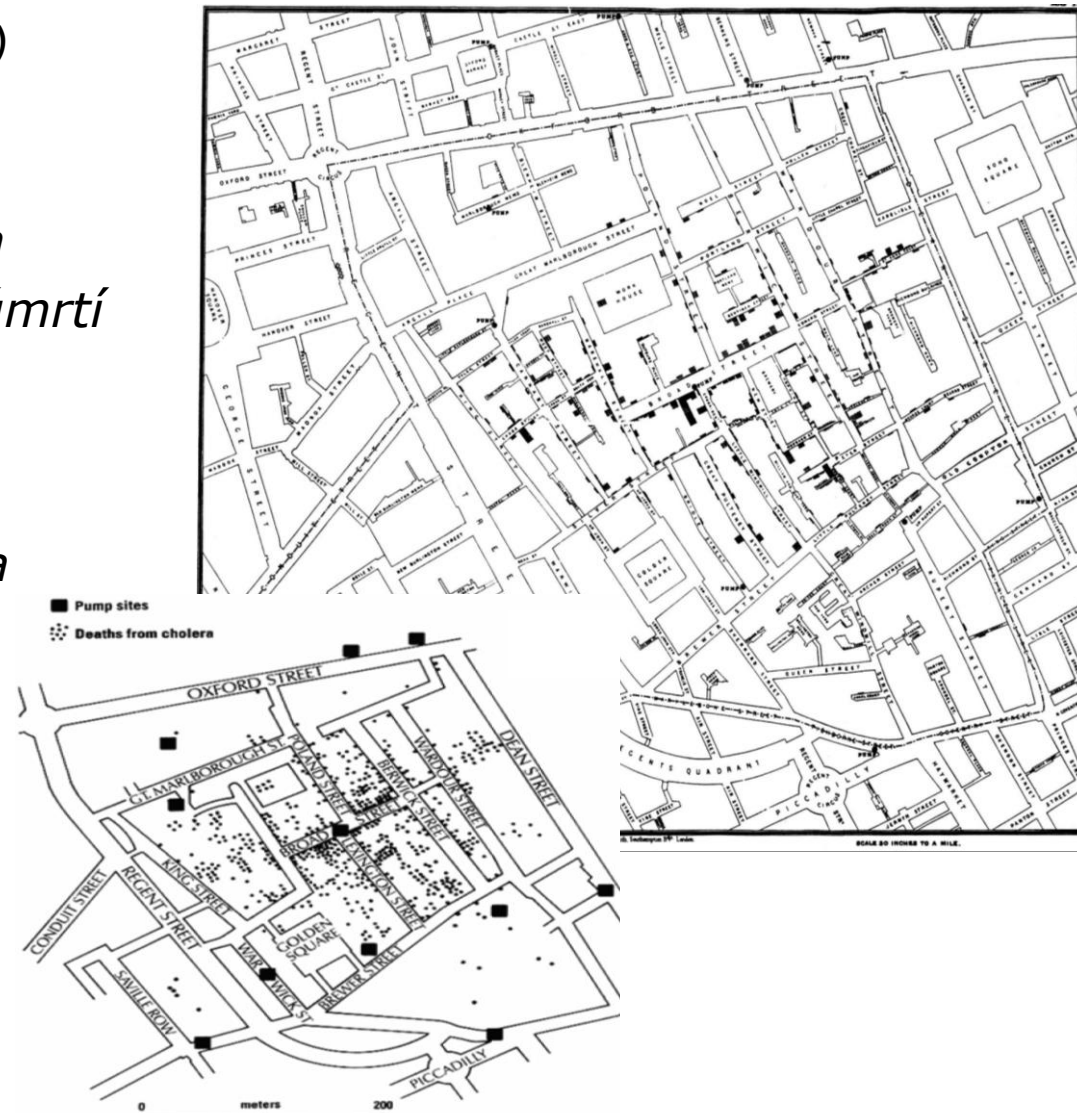
Moran ScatterPlot for Obce\_id : R-square = 0.685533



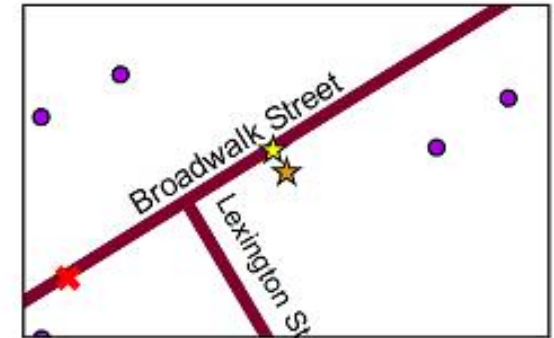


# Ještě jeden příklad...

- *John Snow (1812 – 1858)*
  - *Syn farmáře, asistent chirurga a lékárník*
  - *1854 – mapa Londýna*
    - *Výskyt cholery = úmrtí*
    - *Měřítko 1:2000*
    - *Pomocí prostorové analýzy došlo k identifikaci ohniska nákazy, které bylo možné následně omezit*



# Ještě jeden příklad...



## Legend

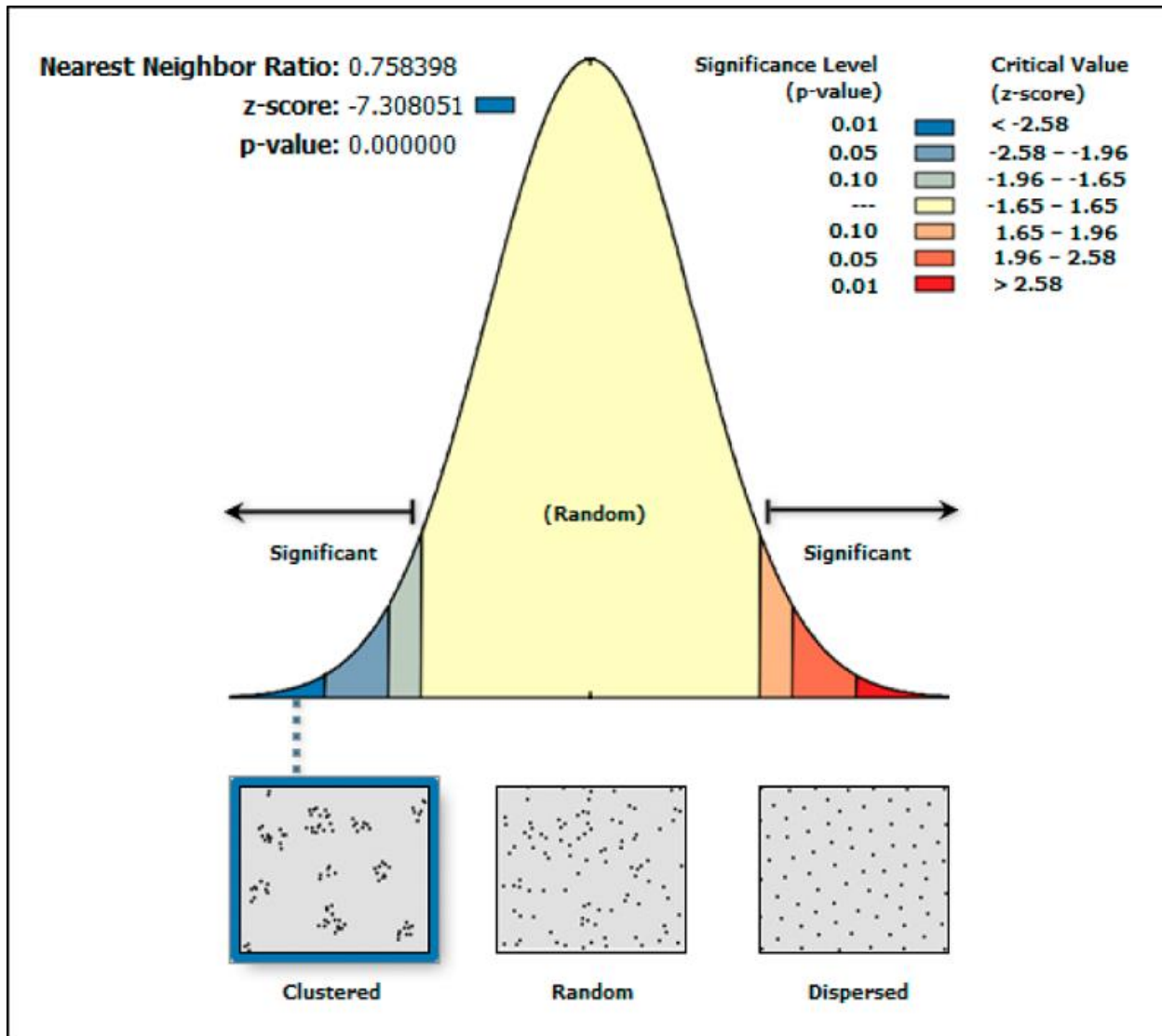
- ✕ Pumps
- Cholera Death
- ★ Mean Center
- ★ Median Center
- 1 Standard Deviation
- 2 Standard Deviation
- Streets



0 50 100 200 Meters

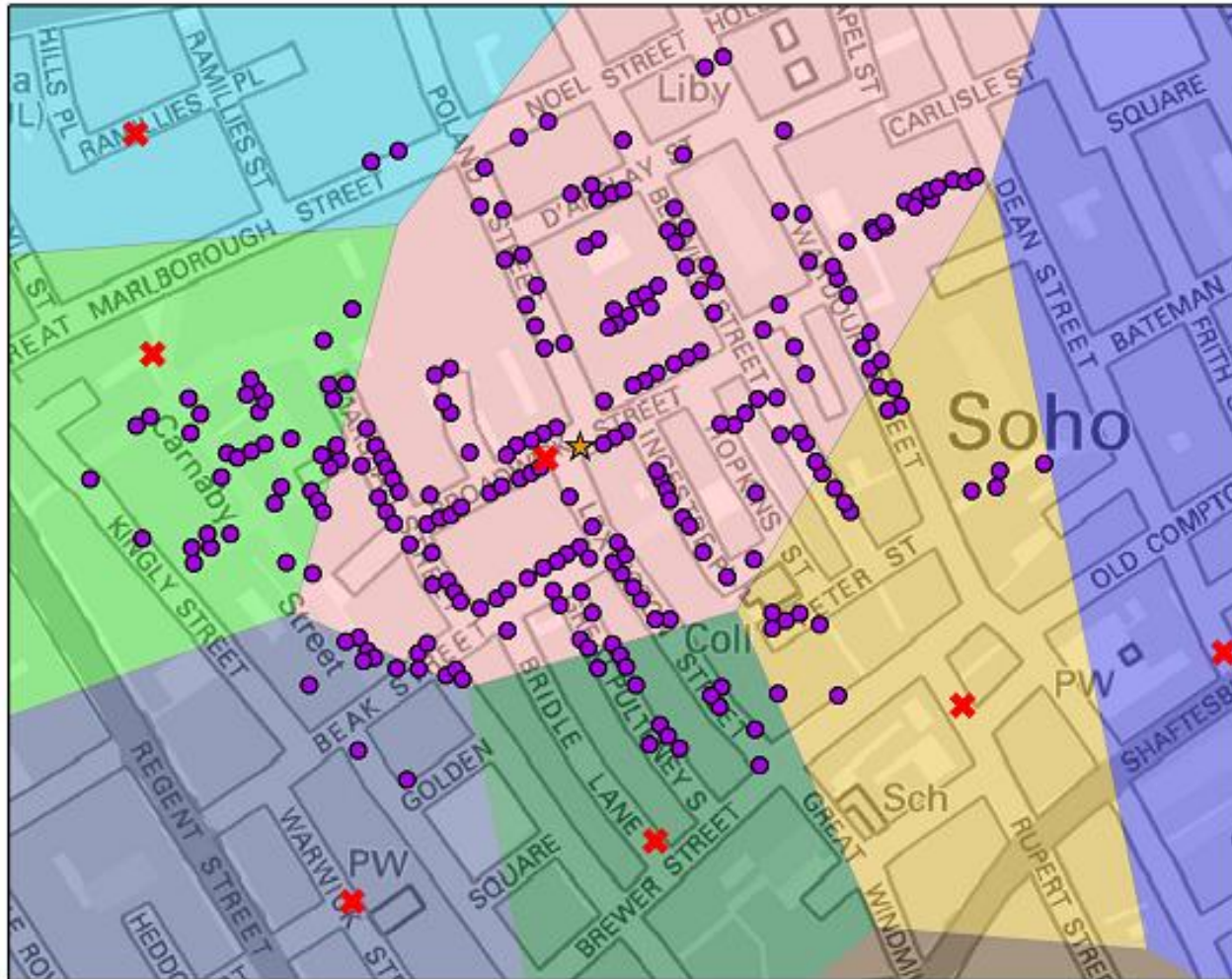


# Ještě jeden příklad...





# Ještě jeden příklad...



## Legend

- ✖ Pumps
- Cholera Death
- ★ Mean Center

## Pump ID

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7





# Ještě jeden příklad...

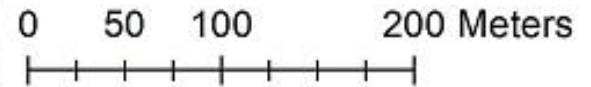


## Legend

- ✖ Pumps
- Cholera Death
- ★ Mean Center

## Kernel Density

- Very Low
- Low
- Medium
- High
- Very High



## Zdroje:

- <https://gistbok.ucgis.org/bok-topics/kernels-and-density-estimation>
- <https://gistbok.ucgis.org/bok-topics/point-pattern-analysis>
- Ivan, I., Horák, J. (2015): Metodika identifikace anomálních lokalit kriminality pomocí jádrových odhadů. Dostupné z: [http://gisak.vsb.cz/GIS\\_Ostrava/GIS\\_Ova\\_2016/sbornik/papers/gis\\_2016568b7fa9bf442.pdf](http://gisak.vsb.cz/GIS_Ostrava/GIS_Ova_2016/sbornik/papers/gis_2016568b7fa9bf442.pdf)
- 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](https://www.youtube.com/watch?v=GWRh_dq3U_U)