



Popisná statistika bodových dat

Modelování prostorového uspořádání bodů

Kernel density

jaro 2023

Lukáš Herman

herman.lu@mail.muni.cz

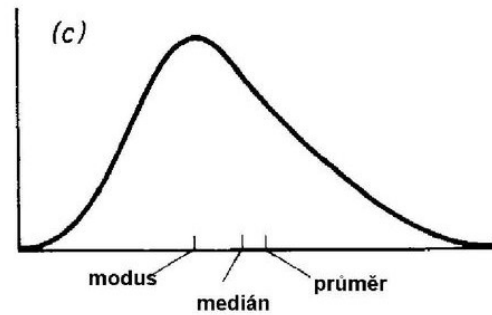


Popisná statistika bodových objektů

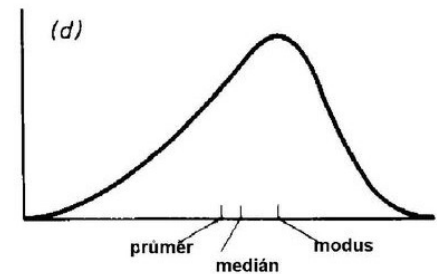
- Charakteristiky polohy
- Charakteristiky rozptylu
- Charakteristiky asymetrie
- Charakteristiky špičatosti



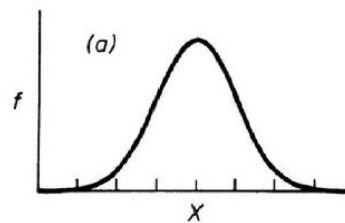
Pozitivní šikmost:



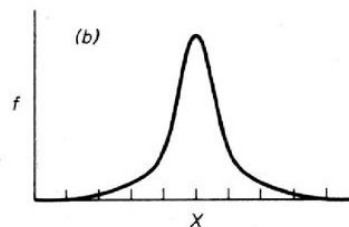
negativní šikmost:



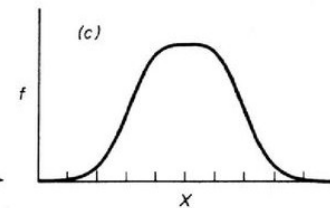
Normální špičatost:



nadnormální:



podnormální:





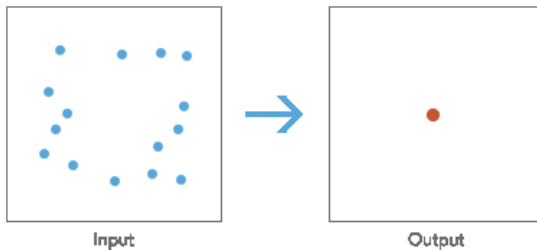
Charakteristiky polohy

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

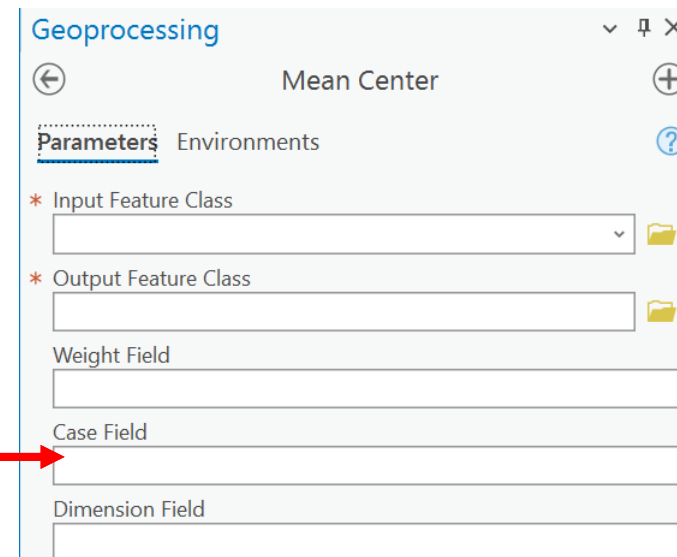


Průměrný střed

- 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

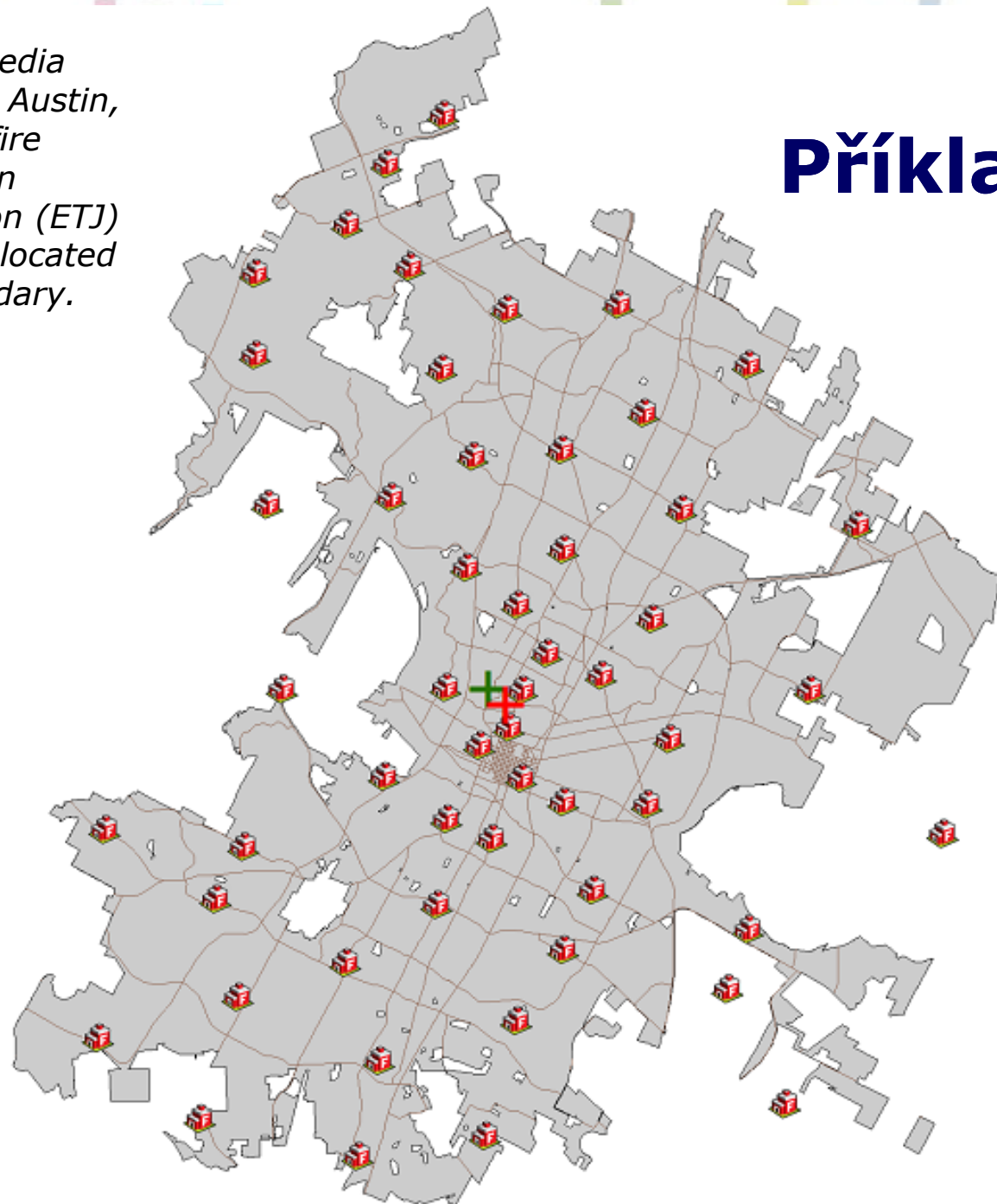


The mean center and media center of fire stations in Austin, Texas. Note that some fire stations are in the Austin extraterritorial jurisdiction (ETJ) area, and therefore are located outside of the city boundary. Data source: data.AustinTexas.gov

Příklad

Legend

- + Median Center
- + Mean Center
- 🚒 Fire Stations
- Major Roads
- ▭ Austin City Boundary





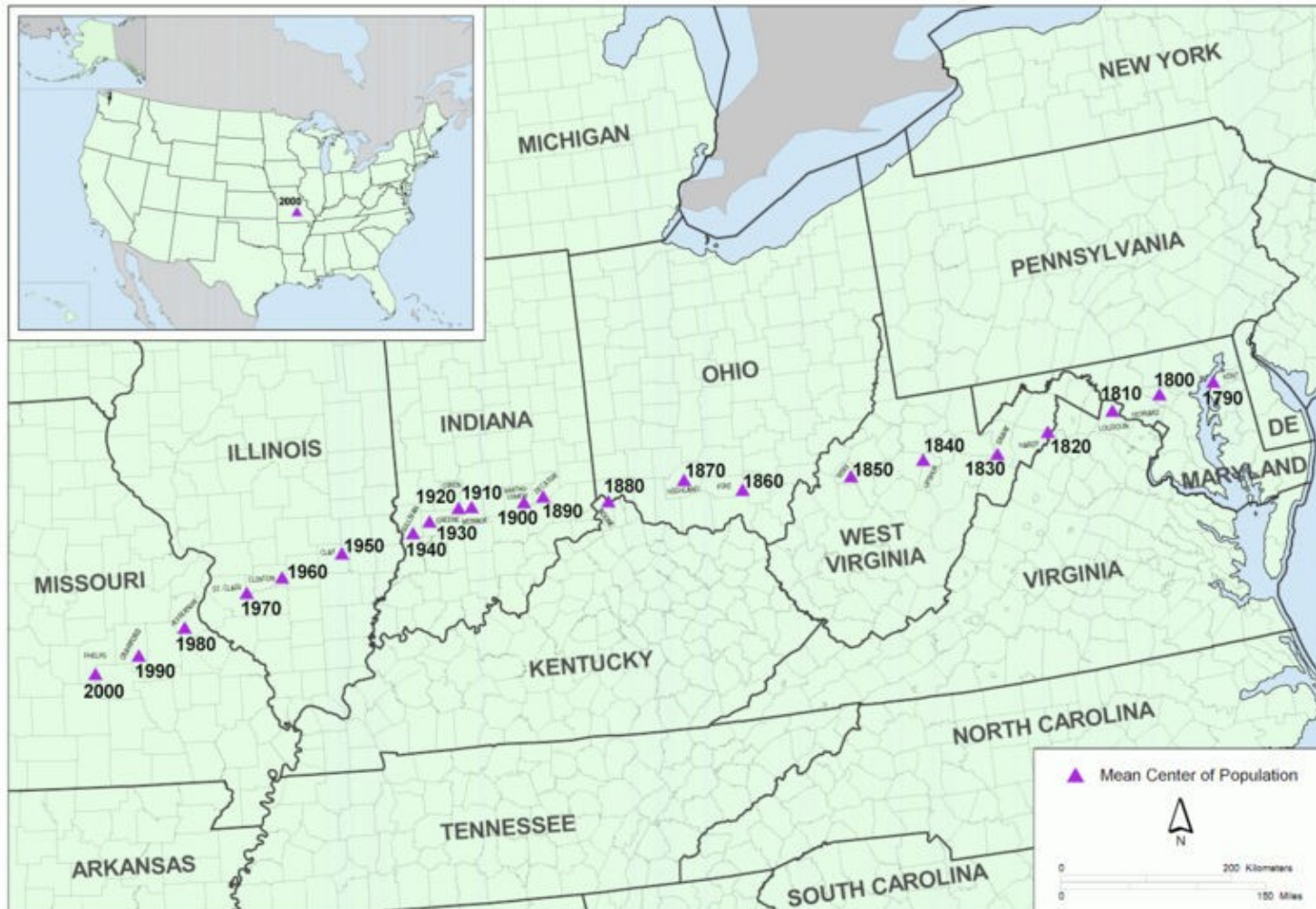
Vážený průměrný střed

- 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 ArcGIS Geoprocessing environment showing the 'Mean Center' tool. The tool's parameters are listed in a table-like structure. The 'Weight Field' parameter is highlighted with a red circle. The 'Input Feature Class' and 'Output Feature Class' fields are empty. The 'Case Field' and 'Dimension Field' fields are also empty.

Parameter	Value
* Input Feature Class	
* Output Feature Class	
Weight Field	
Case Field	
Dimension Field	

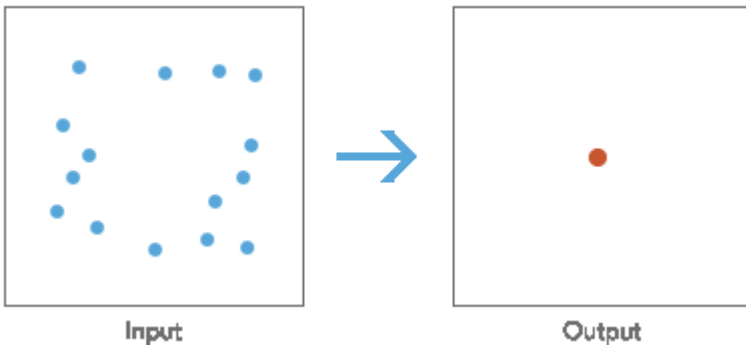
Mean Center of Population for the United States: 1790 to 2000



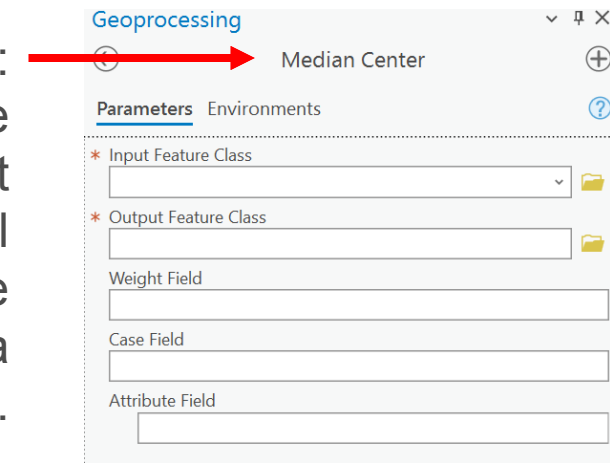


Mediánový střed

- 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.



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 typy jevů.
- Mohou být použity i pro studium dynamiky jevů (př.: *různé kružnice pro jeden jev v různých časových horizontech*)

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

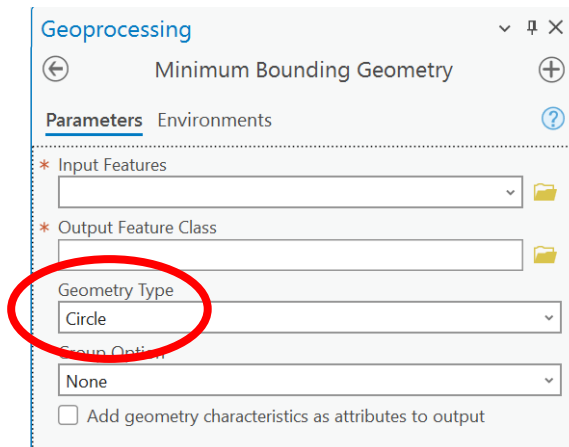
Weight Field

Case Field



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: $R = \sqrt{(P/\pi)}$

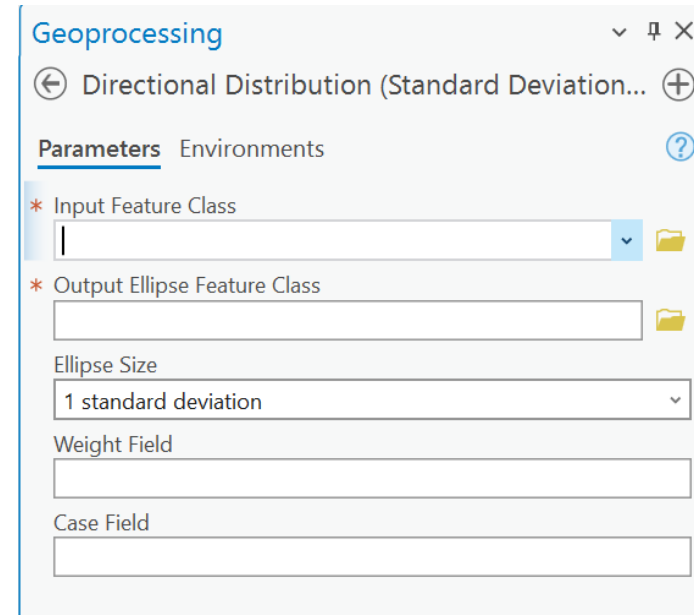
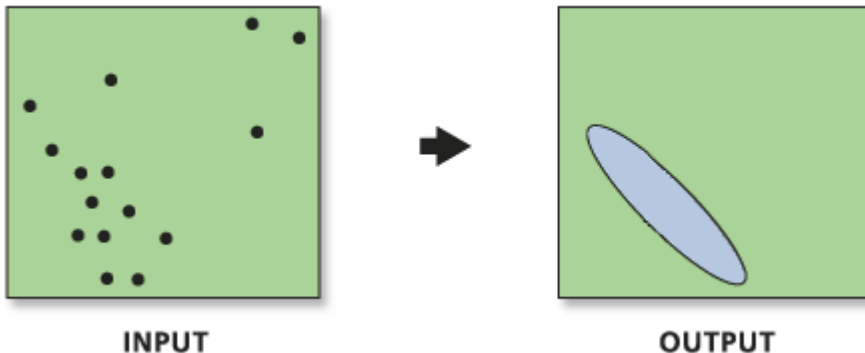


Směrodatná elipsa odchylek

- 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






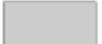


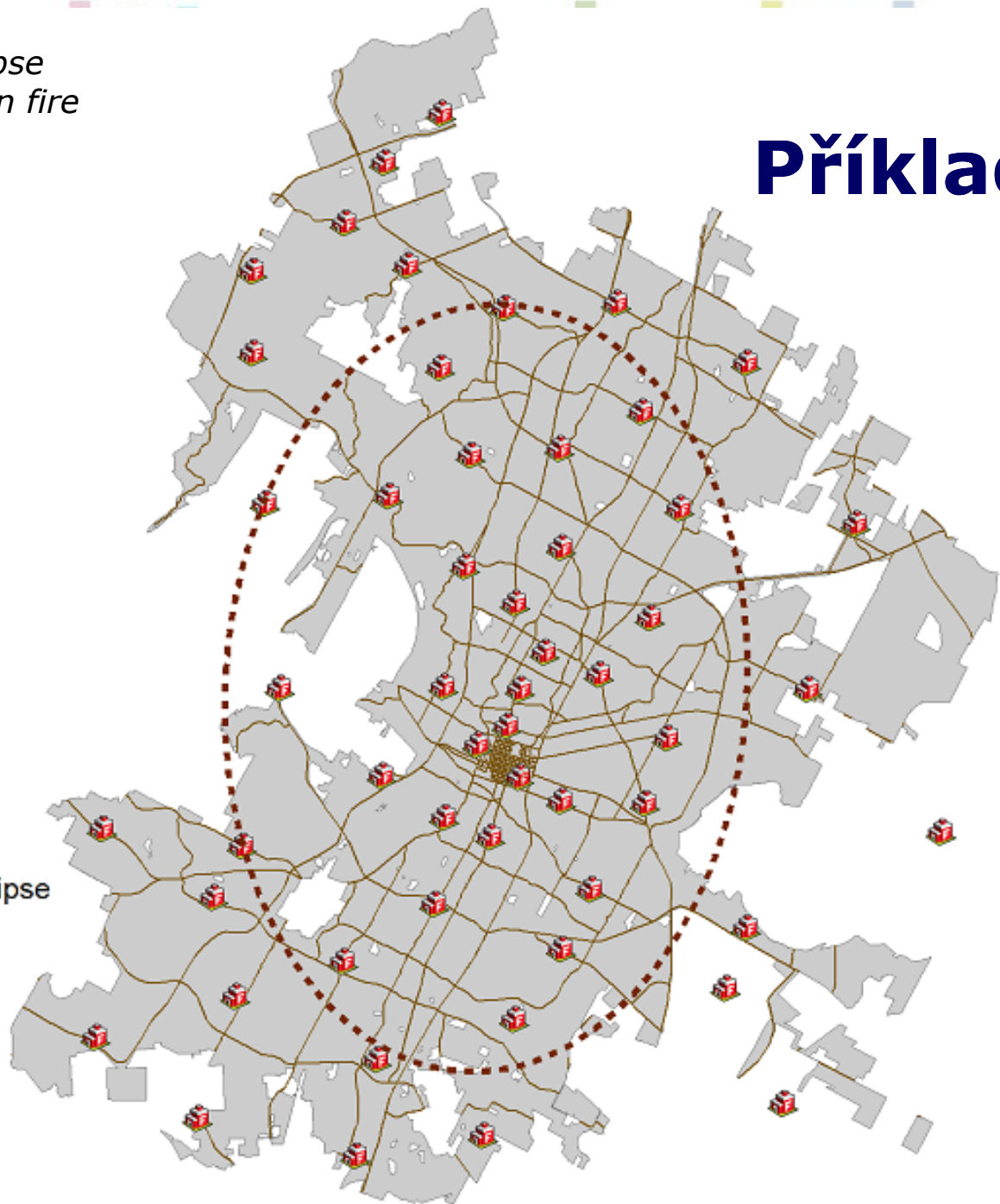
Standard deviation ellipse
created based on Austin fire
station locations

Data source:
data.AustinTexas.gov

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

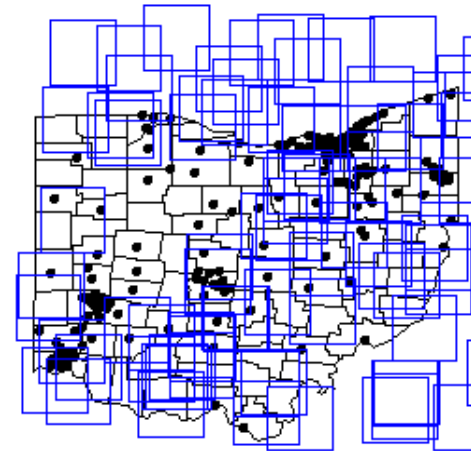
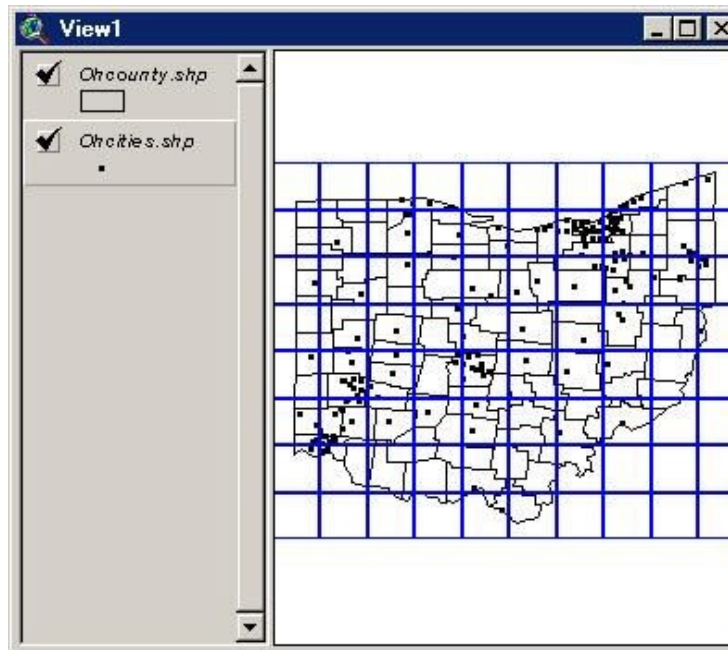


Metody statistického popisu bodů – obecně

- 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.



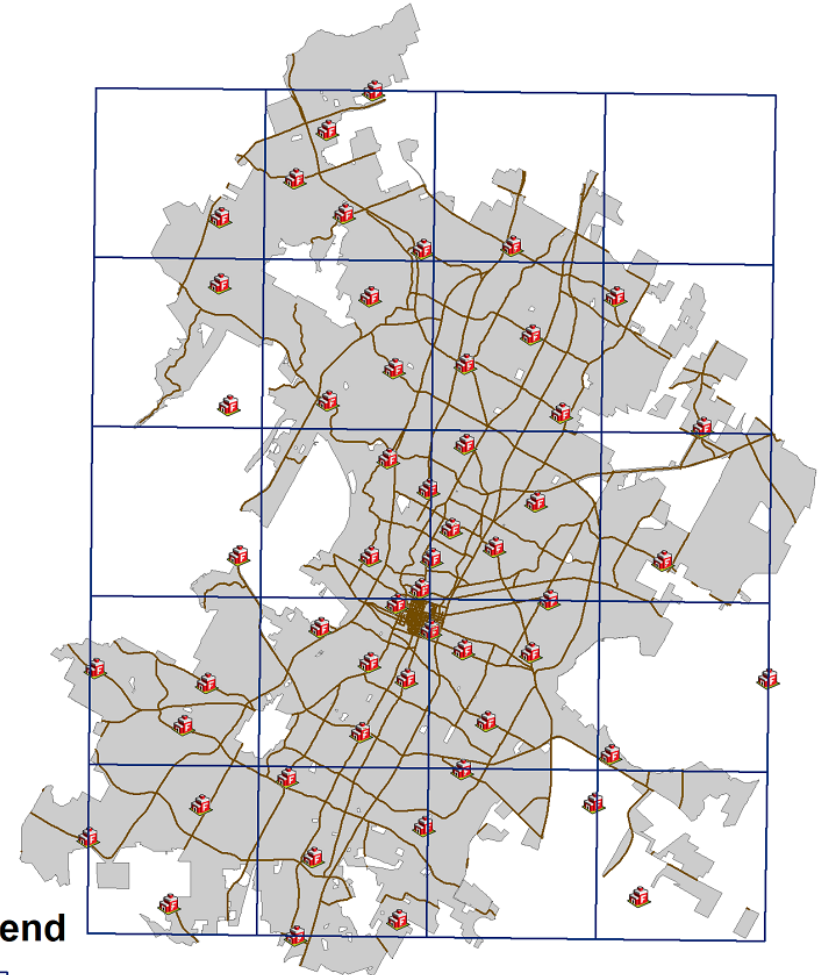


Analýza kvadrátů





- 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 km², 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: $D_\alpha = \frac{1,36}{\sqrt{m}} \quad D_\alpha = 1,36 \sqrt{\frac{m_1 + m_2}{m_1 \cdot m_2}}$
6. Je-li vypočtená hodnota D větší než kritická hodnota D_α , 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 difference
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

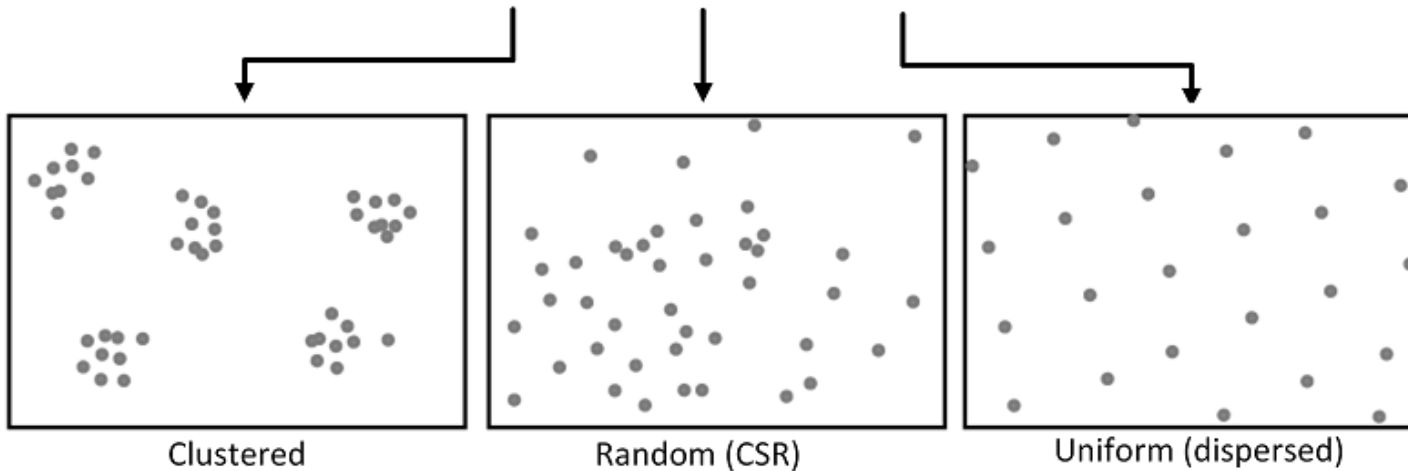
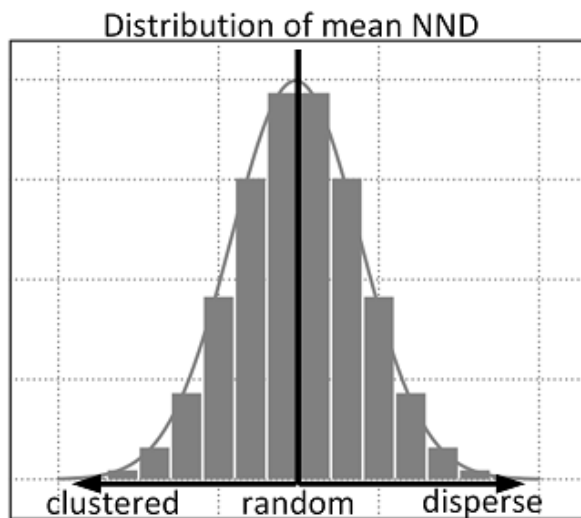
Materiály předmětu
Z6101 Základy
geostatistiky



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).

Metoda nejbližšího souseda





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).



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

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.

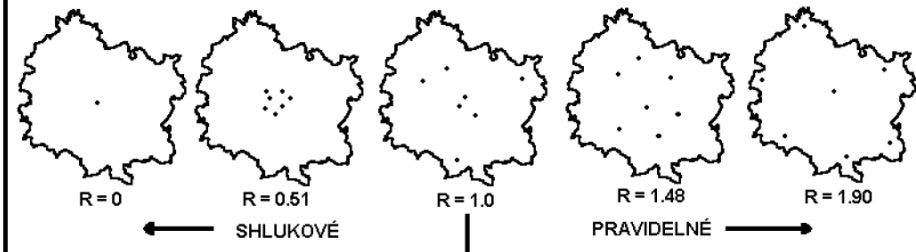
Hodnotu r_{exp} zjistíme ze vztahu:

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

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

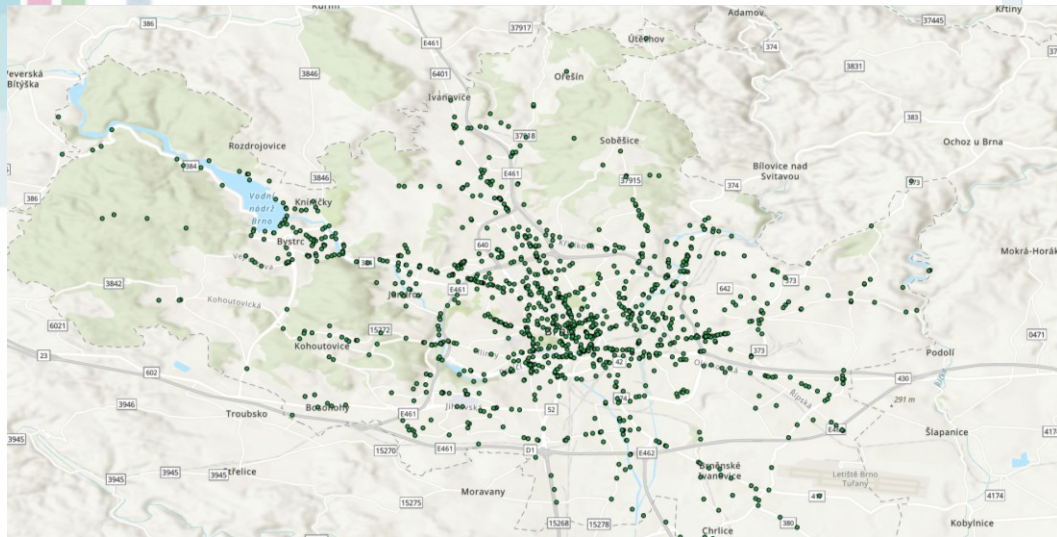
* Input Feature Class

Distance Method
Euclidean

Generate Report

Area

Metoda nejbližšího souseda



Average Nearest Neighbor Summary

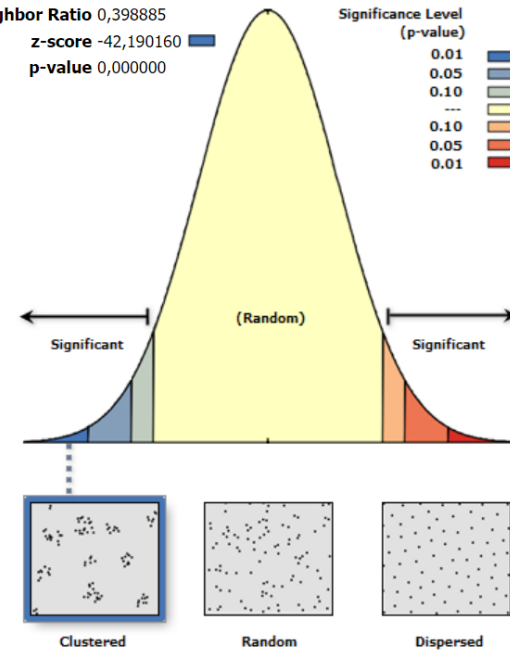
Nearest Neighbor Ratio 0,398885
 z-score -42,190160
 p-value 0,000000

Significance Level (p-value)
 0.01
 0.05
 0.10

 0.10
 0.05
 0.01

Critical Value (z-score)
 < -2.58
 -2.58 - -1.96
 -1.96 - -1.65

 1.65 - 1.96
 1.96 - 2.58
 > 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 vizualni srovnani prostoroveho rozloženi 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 studovanem 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

Spatial Autocorrelation Report

Moran's Index 0,069576

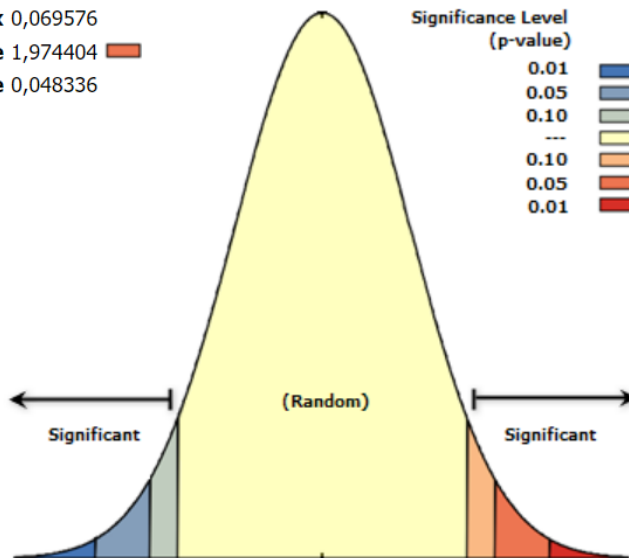
z-score 1,974404

p-value 0,048336

Significance Level
(p-value)

Critical Value
(z-score)

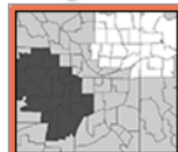
0.01	< -2.58
0.05	-2.58 - -1.96
0.10	-1.96 - -1.65
---	-1.65 - 1.65
0.10	1.65 - 1.96
0.05	1.96 - 2.58
0.01	> 2.58



Dispersed



Random



Clustered

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.

autokorelace

Geoprocessing

Spatial Autocorrelation (Global Moran's I)

Parameters

Input Feature Class: cyklo_nehody_cykl_neh_xytablepoint_spa1

Input Field: hmotna_sko

Generate Report

Conceptualization of Spatial Relationships: **Inverse distance**

Standardization: None

Distance Band or Threshold: Distance

Global Moran's I Summary

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



Metoda jádrových odhadů

- Hlavní metodou pro identifikaci anomálních lokalit, které bývají často nazývány jako **hot spots**, je metoda **jádrových odhadů** (kernel density estimation) či **jádrového vyhlazení**.
- **Jaká je hlavní nevýhoda??**
 - Základním nedostatkem je subjektivita v intepretaci výsledků.
 - Stejná podkladová data mohou být zobrazena značně rozdílně jen s využitím rozdílného nastavení metody a způsobu zobrazení.
 - *Použité parametry je vhodné uvést.*
 - Z tohoto důvodu je potřeba zvýraznit statisticky významné výsledky.



Metoda jádrových odhadů

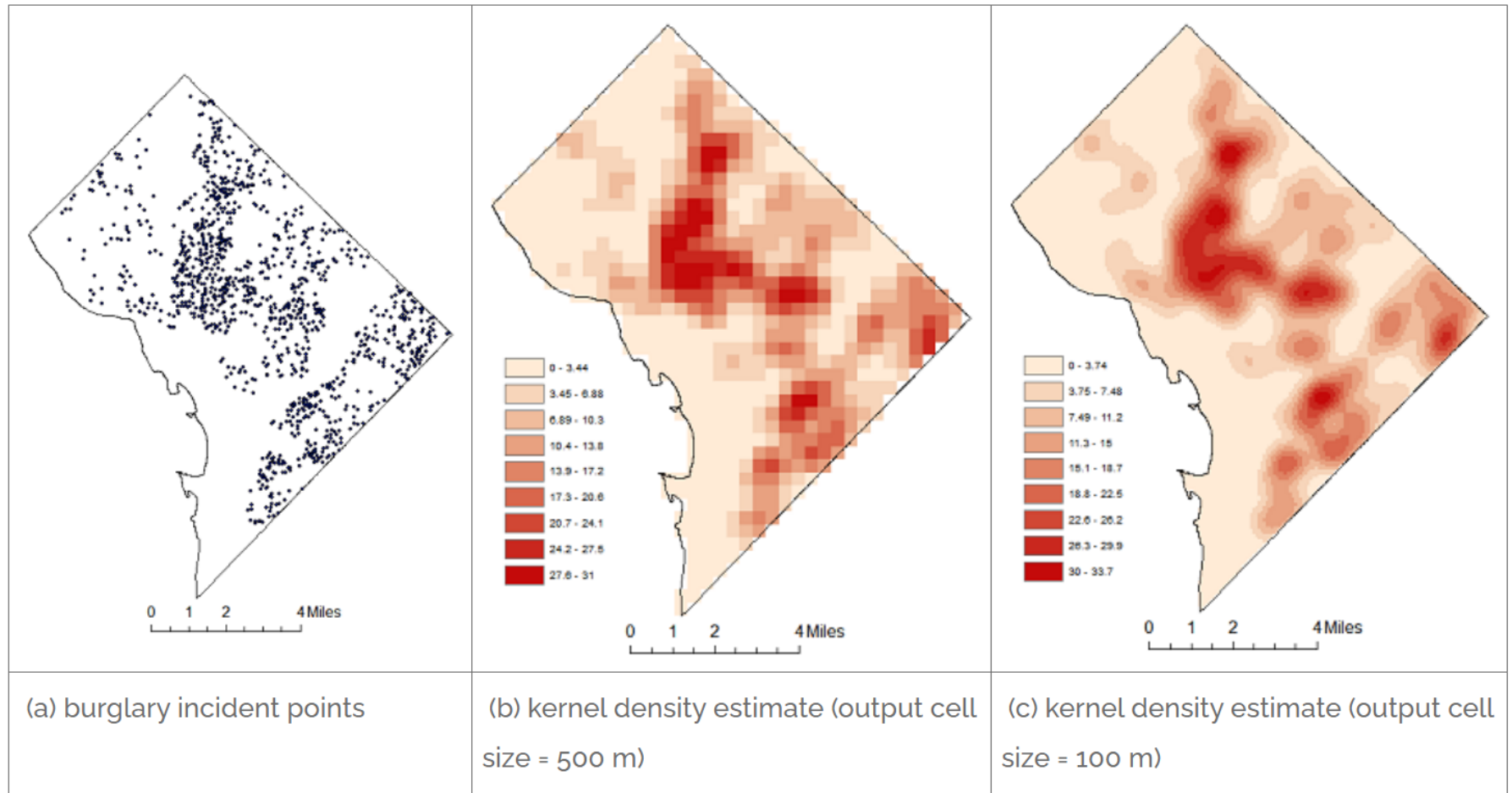


Figure 2a - 2c. Kernel density estimates of burglary incidents in Washington, D.C., 2018. Source: author.



Předpoklady užití metody

- **Není vhodná pro zobrazení rozsáhlých území.**
- **Vhodná pro mapy větších měřítek (obce či jejich části).**
- **Není doporučena pro větší územní celky (okres, kraj, ČR) → toto záleží na zobrazovaném jevu**
- **Neexistuje také žádná hranice pro minimální počet událostí v oblasti.**
- **Doporučujeme však brát v potaz počet bodů a plochu analyzované oblasti. Pokud je oblast menší, je možné pracovat i s menším počtem událostí.**
- **V případě malých počtů na větší ploše použití jádrového vyhlazení není doporučeno.**



Krok I – PŘEDZPRACOVÁNÍ DAT

- Základní podmínkou - kvalitní data.
- Nutné se zaměřit na:
 - správnost a přesnost souřadnicového určení polohy,
 - časové určení,
 - tematické určení.
- Rozlišit případy, kdy již záznam deliktu obsahuje souřadnice, od těch, kde je poloha vyjádřena pouze adresou či jiným referencováním.
- Pokud jsou body lokalizovány na jedno místo, tak zde vznikají umělé shluky, které mylně identifikují lokalitu jako anomální.
 - Řešení – náhodné rozmístění událostí podél/uvnitř lokalizovaného objektu.



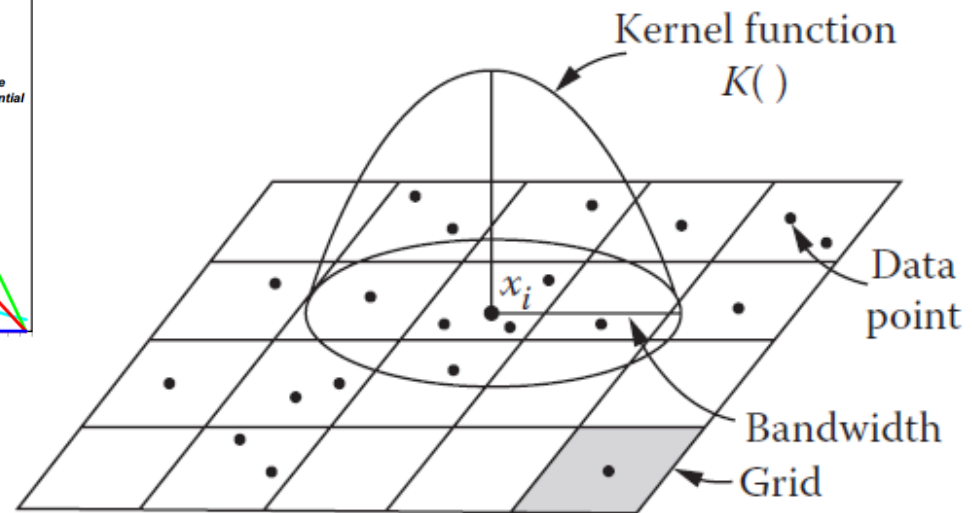
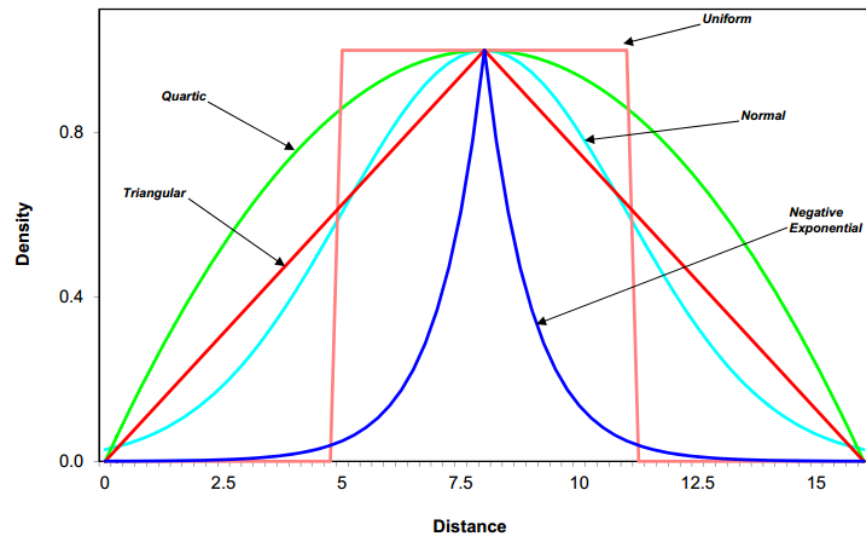
Krok II – VOLBA METODY

- KDE? **v celé ploše území vs výskyt omezen pouze na jisté části území.**
- jádrové odhady **plošné** (2D) a jednorozměrné (1D), modelující výskyt **pouze na liniích.**
- Obecně metoda jádrových odhadů přiřazuje **každému bodu v mapě odhad intenzity na základě vzdálenosti k ostatním událostem.** Nemůžeme však tuto intenzitu počítat pro každý bod, jelikož těch je nekonečně mnoho, a tak je analyzované území proloženo čtvercovým gridem a intenzity jsou počítány pro centroidy jednotlivých buněk.
- V prvním kroku je potřeba vybrat metodu jádrového odhadu:
 - **Jednoduchý**
 - **Duální**
- Dále je nutné volit mezi jádrovým odhadem s dosahem:
 - **Fixní**
 - **Adaptivní**



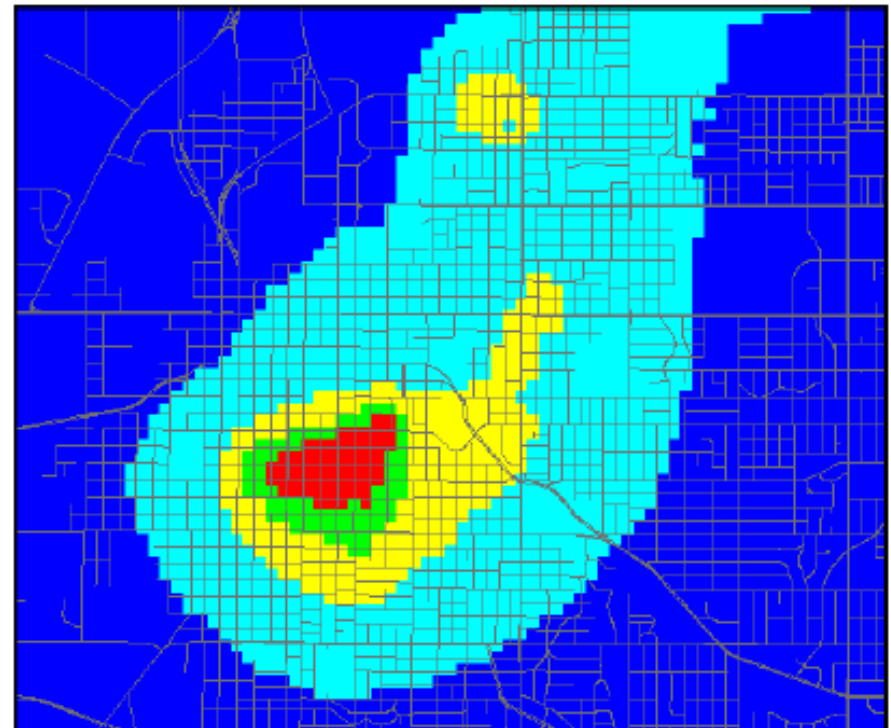
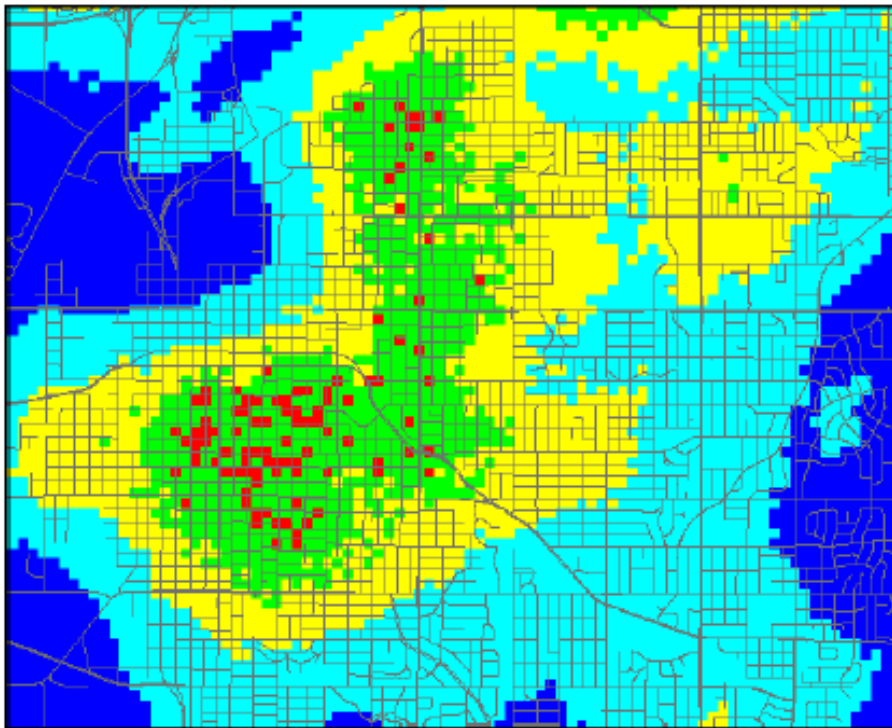
Krok III – NASTAVENÍ VYHLAZOVACÍ FUNKCE

- šest různých vyhlazovacích funkcí: normální, rovnoměrná, kvartická, kuželová, kvadratická a záporná exponenciální.
- nejčastěji se využívá kvartická funkce,



Závislost na zvolené vyhlazovací funkci

Trojúhelníková vs. Gausova (normální)





Velikost buňky

- GRID = nezbytné správně zvolit jeho prostorové rozlišení.
- Velikost buňky tohoto gridu ovlivňuje získané výsledky z pohledu detailnosti a také velikosti souboru.
- nehraje na přesnost výsledků tak důležitou roli, jako další dva parametry.
- Jak stanovit?
 - MBR (**kratší strana/150**).
 - ČR – města a obce velikost buňky 50 m, minimum 10 m.
 - **Výjimky?**

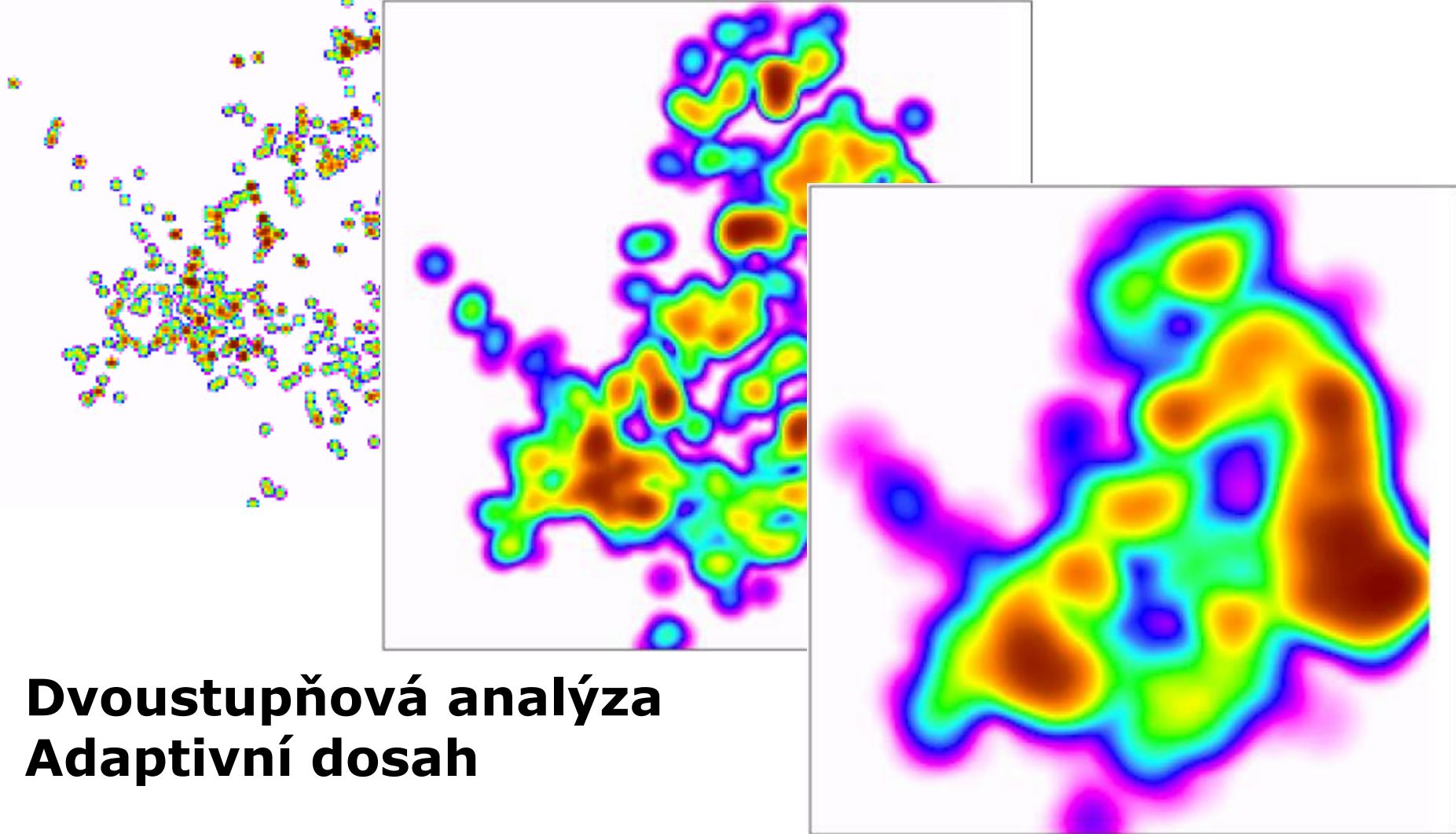


Dosah (šířka pásma)

- Pro výsledky jádrových odhadů je klíčová především volba dosahu vyhlazovací funkce. Neexistuje žádné obecné pravidlo, jak určit nejvhodnější hodnotu dosahu.
- Vždy záleží na prostorové distribuci bodů, typu události a měřítku – závislost dosahu konkrétního trestného činu.
- **Explorace** (vývoj území) vs. **Identifikace anomálií** (hot spots).

Druh kriminality	Druh kernelu	pásma (m)
Loupež	Kvartická	200
Výtržnictví	Kvartická	200
Úmyslné ublížení na zdraví	Kvartická	200
Krádeže motorových vozidel	Válcová	100
Krádeže věcí z aut	Válcová	100

50 – 200 – 400 m rozsah



**Dvoustupňová analýza
Adaptivní dosah**



PROVĚŘENÍ STATISTICKÉ VÝZNAMNOSTI

- Výstup = grid s intenzitami událostí, sám o sobě neposkytuje informaci o výskytu statisticky významných oblastí a jeho interpretace je **velmi subjektivní**.
- Nejpoužívanějším postupem pro hodnocení výsledků jádrových odhadů je **Getis-Ord G_i^* index**.
- Pro výpočet G_i^* doporučeno použít **topologické okolí definované pohybem královny prvního řádu**.
Doporučujeme zobrazit jen statisticky významné výsledky na hladině významnosti **nejméně 95 %**.
- Následně **hranici těchto významných shluků zobrazit** spolu s výsledky jádrového vyhlazení a vyznačit v tomto výstupu hranice těchto statisticky významných anomálních oblastí.





Getis-Ord G_i^*

- Ukazatel významnosti shluku.
- G_i^* statistika vrací pro každý prvek v datové sadě tzv. Z skóre (z-score).
- Statisticky významné pozitivní z-score = čím větší, tím je intenzivnější shluk vysokých hodnot (hot spot).
- Statisticky významné negativní Z skóre = čím menší Z skóre, tím intenzivnější shluk nízkých hodnot (cold spot).



GI a GI* statistika

1	1	1	5	0	0	0	1	0	0	0	0	0	3	2	
0	3	0	0	6	1	0	1	1	0	0	0	0	1	3	
5	0	0	0	0	1	9	5	0	0	3	0	0	1	0	1
1	4	0											0	0	2
1	0	2	0	0	2	1	2						1	5	0
3	5	0											0	1	0
0	0	1	1	6	6	2	2						1	2	0
0	2	2											0	0	2
0	0	3	6	12	9	2	2						0	2	2
1	2	4											3	0	2
4	4	1	1	2	1	1	1						1	6	4
1	1	0											6	1	0
0	0	0	0	1	0	1	3						4	0	0
1	1	0	0	0	0	0	0	2	0	0	13	0	0	0	0
0	0	0	1	1	0	0	0	1	4	6	0	2	0	0	0
0	8	2	6	0	0	0	4	3	1	4	7	0	0	0	0

- Každá buňka má jednoznačnou hodnotu.
- Nulová hypotéza:
- **Není žádný vztah mezi hodnotami počtu trestných činů v buňce a v jejím okolí, a to až do vzdálenosti d měřené ve všech směrech. Srovnáno se sumou hodnot na celém studovaném území.**



GI a GI* statistika

Srovnání lokálního s globálním

- Existuje lokální prostorová asociace?
- **Hodně vysokých hodnot v blízkosti buňky.**
 - G_i^* hodnoty budou pozitivní pro všechny buňky
- **Hodně nízkých hodnot pohromadě**
 - G_i^* hodnoty budou negativní pro všechny buňky
- Příklad: Pro hodnotu 9 v centru vzorku platí:
 - G_i^* value = 4.1785
 - *G_i^* hodnota je pozitivní*
- V relativním porovnání (lokální vs. globální) se jedná o hodně buněk s vysokou hodnotou jevu.

0	0	2	1	2
1	6	6	2	2
6	12	9	2	2
1	2	1	1	1
0	1	0	1	3



GI a GI* statistika

- G_i^* výsledky jsou Z skóre
- Z skóre indikují umístění dané hodnoty v datové sadě vzhledem k průměru, standardizované s ohledem na směrodatnou odchylku (standard deviation).
 - **$Z = 0$ odpovídá průměru**
 - **$Z < 0$ méně než průměr**
 - **$Z > 0$**
- Z skóre používáno pro určení prahu spolehlivosti a zhodnocení statistické významnosti.

GI a GI* statistika

Statistická významnost

Z-skóre hodnoty pro úrovně (hladiny) statistické významnosti:

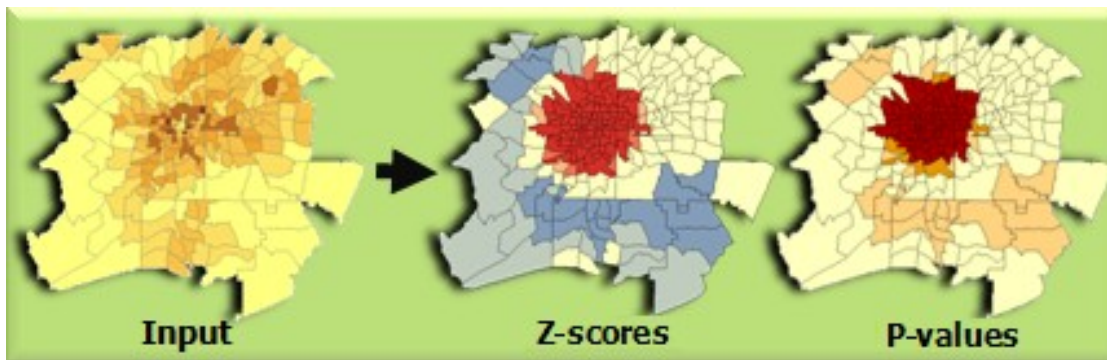
- 90% ≥ 1.645
 - 95% ≥ 1.960
 - 99% ≥ 2.576
 - 99.9% ≥ 3.291
- Univerzální Z skóre bez ohledu na jevu, umístění, velikosti území...
 - Příklad:
 - G_i^* hodnota = 4.1785
 - Větší než 99.9% významnost!





Jak to udělat prakticky?

- Conversion Tools > From Raster > **Raster to Point**
- Data Management Tools > Sampling > **Create Fishnet.**
 - Vlastnosti identické jako v rastru (velikost buněk, rozsah)!
- Data Management Tools > Features > **Feature To Polygon**
 - Propojení bodů a polygonů (Fishnet), případně použít Spatial Join
- Spatial Statistics > ... > **Hot Spot Analysis (Getis-Ord Gi*)**



Geoprocessing

Hot Spot Analysis (Getis-Ord Gi*)

Parameters Environments

* Input Feature Class

* Input Field

* Output Feature Class

Conceptualization of Spatial Relationships

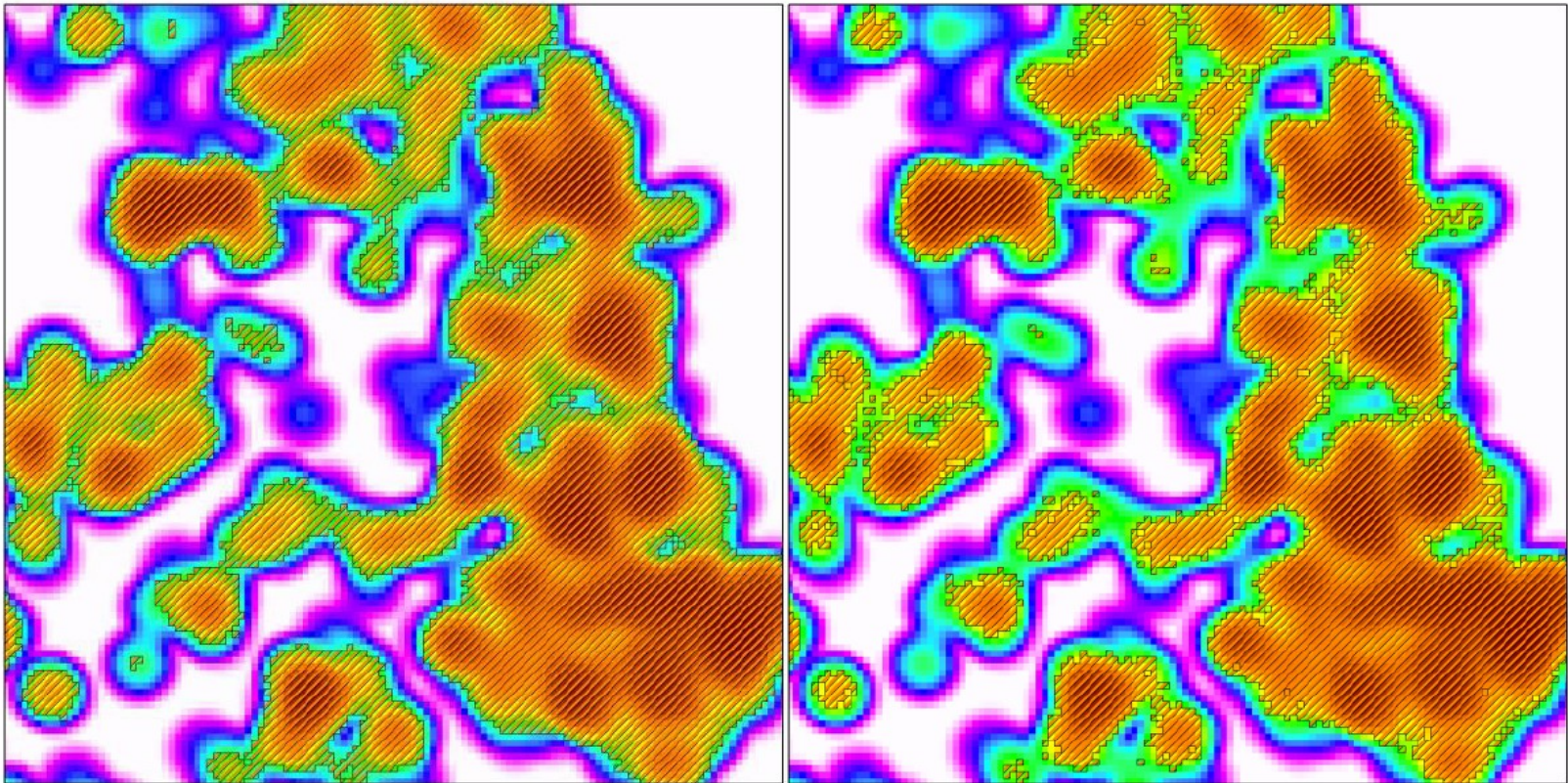
Contiguity edges corners

Self Potential Field

Apply False Discovery Rate (FDR) Correction

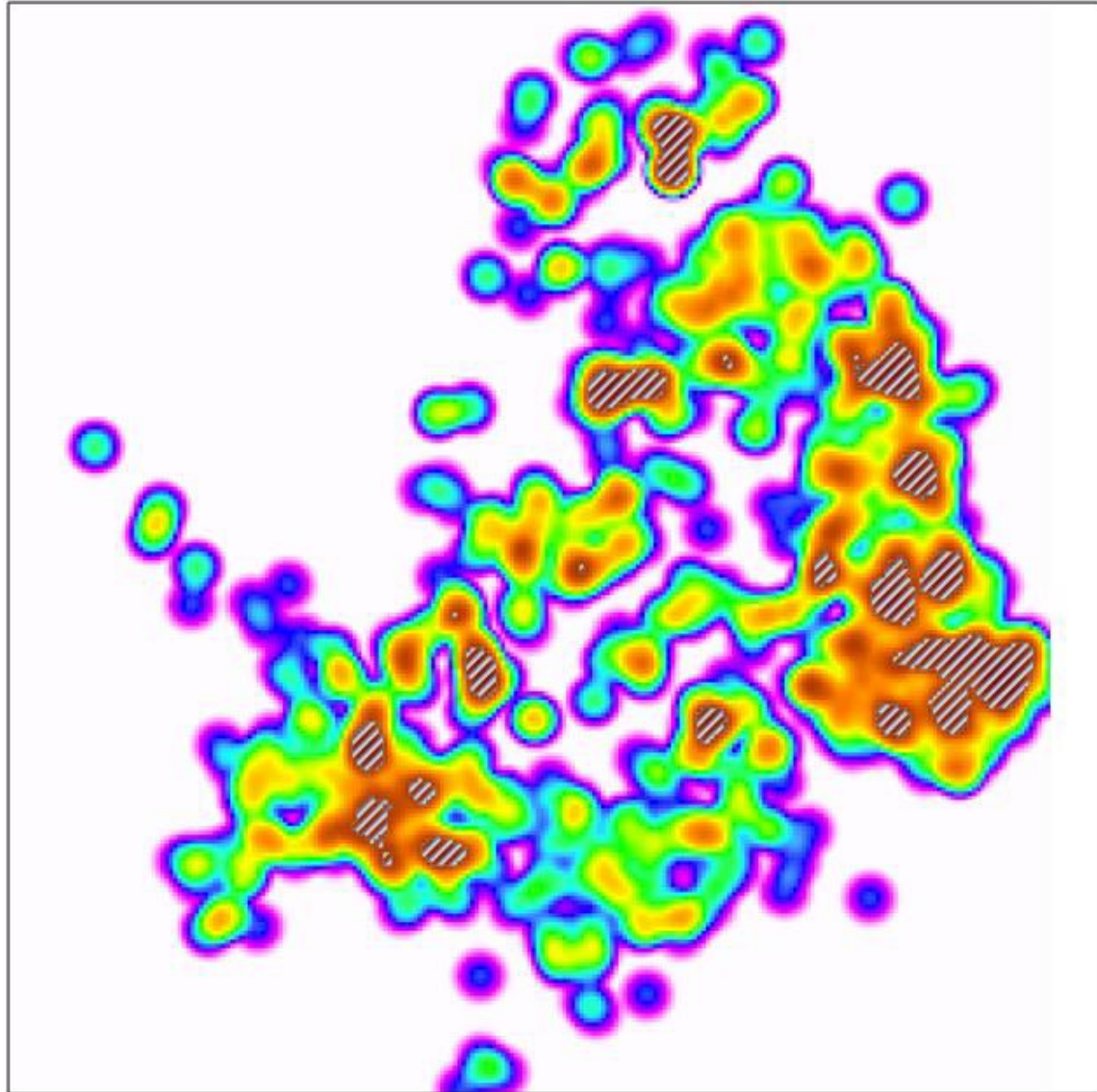
Statistická významnost

- Finální výsledky zobrazující statistické výsledky na hladině významnosti 95 % (vlevo) a 99 % (vpravo).



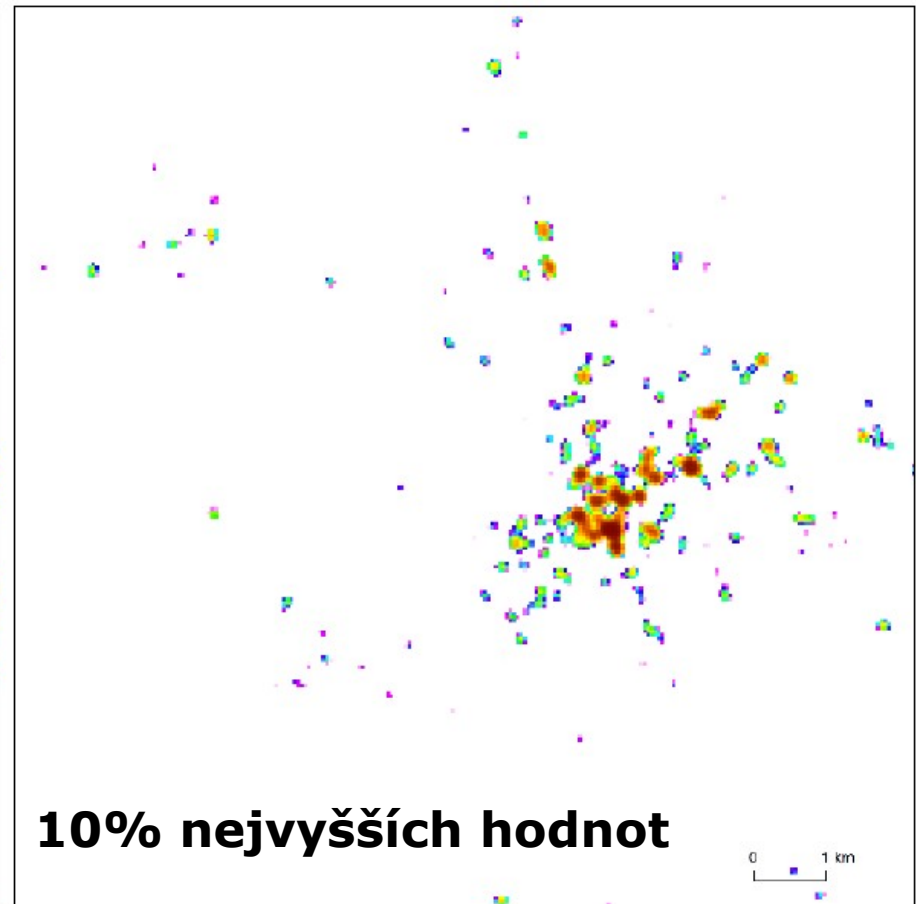
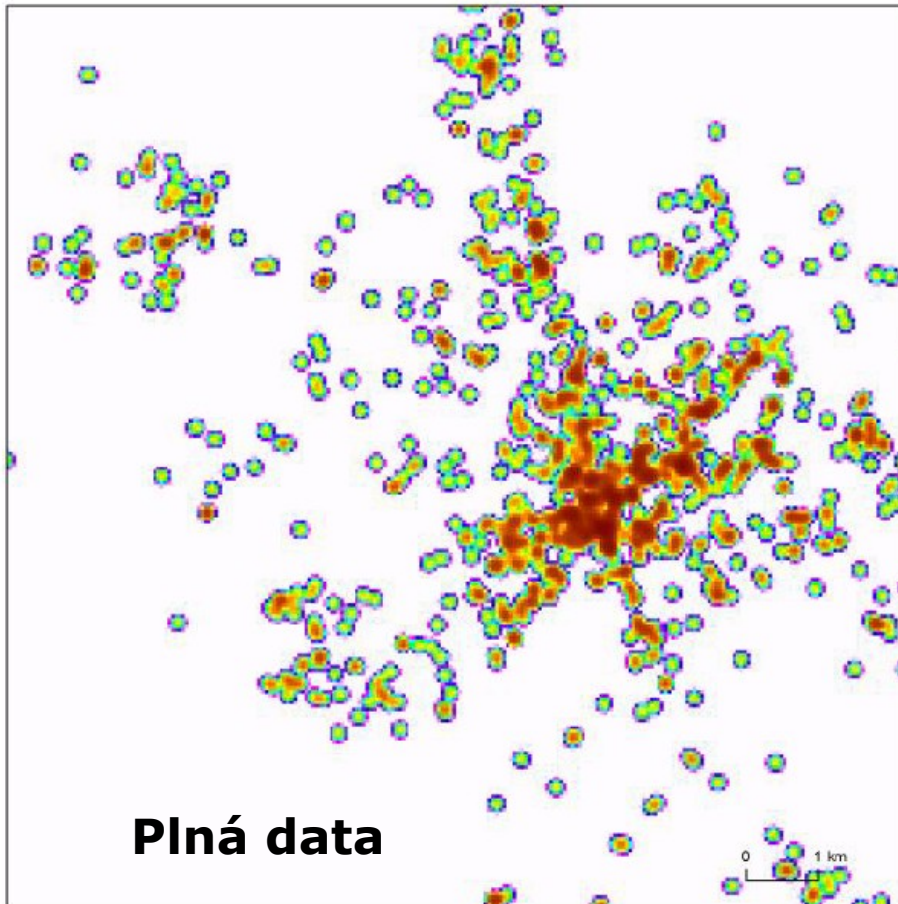
Statistická významnost

- **Jak zlepšit zacílení na významné oblasti?**
- Testovat statistickou významnost jen na nejvyšších hodnotách.
- Kombinovaný postup, z výsledku jádrového vyhlazení vybereme jen 20 % nejvyšších hodnot a z těchto hodnot vybereme jen statisticky významné výsledky metodou G_i^* .



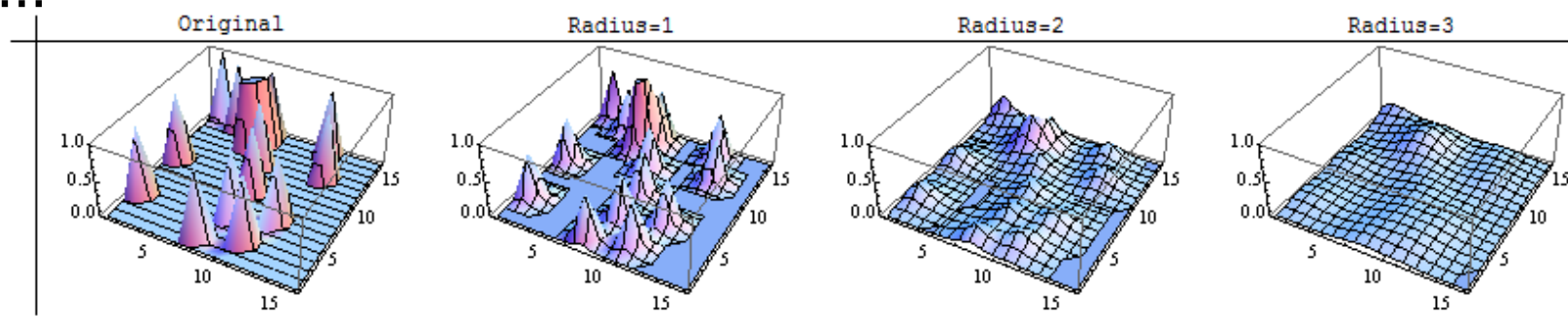
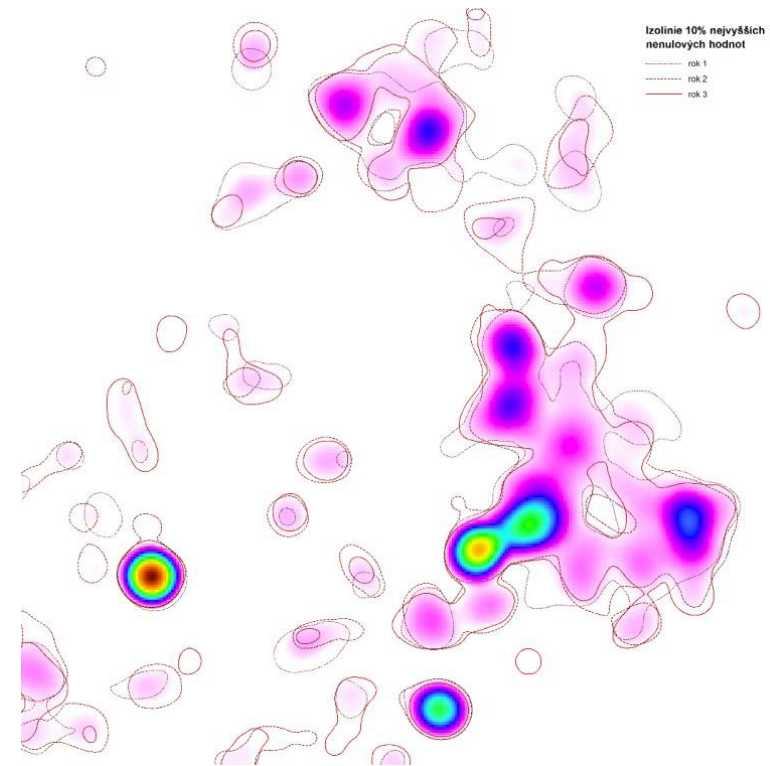
POSTPROCESSING A VIZUALIZACE

- **Vizuální omezení** – podpora rozhodování dle zadání a uživatelské skupině.



Vizualizace – „alternativní“

- **Metody zobrazení**
 - Vícebarevné
 - 3D
 - Izolinie
- **Škály**
- **Podklad**
 - Topografická mapa
 - Ortofoto
 - ...



Vizualizace – empirické testování

- Nétek & Slezáková (2018): Implementation of Heat Maps in Geographical Information System – Exploratory Study on Traffic Accident Data. <https://doi.org/10.1515/geo-2018-0029>
 - Most of the users chose bright colors with a **negative feeling**, such as **red**, for traffic accident visualization.
 - The best settings for transparency was identified to be around **50%**.
 - The final questions were about map **readability based on radius**. This setting is **tied to mapscale** but follows a common trend throughout the research.

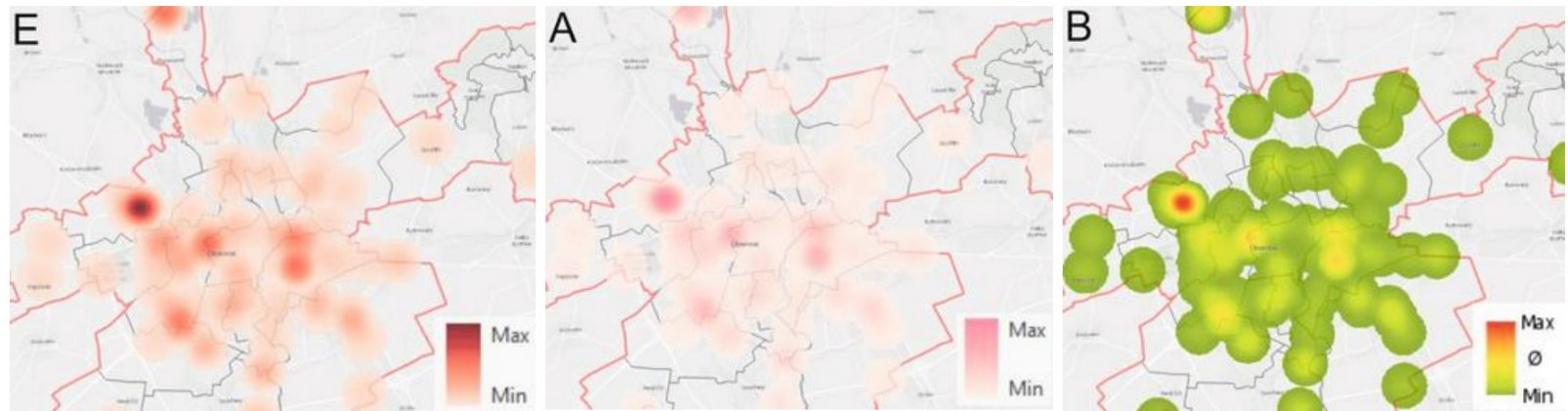
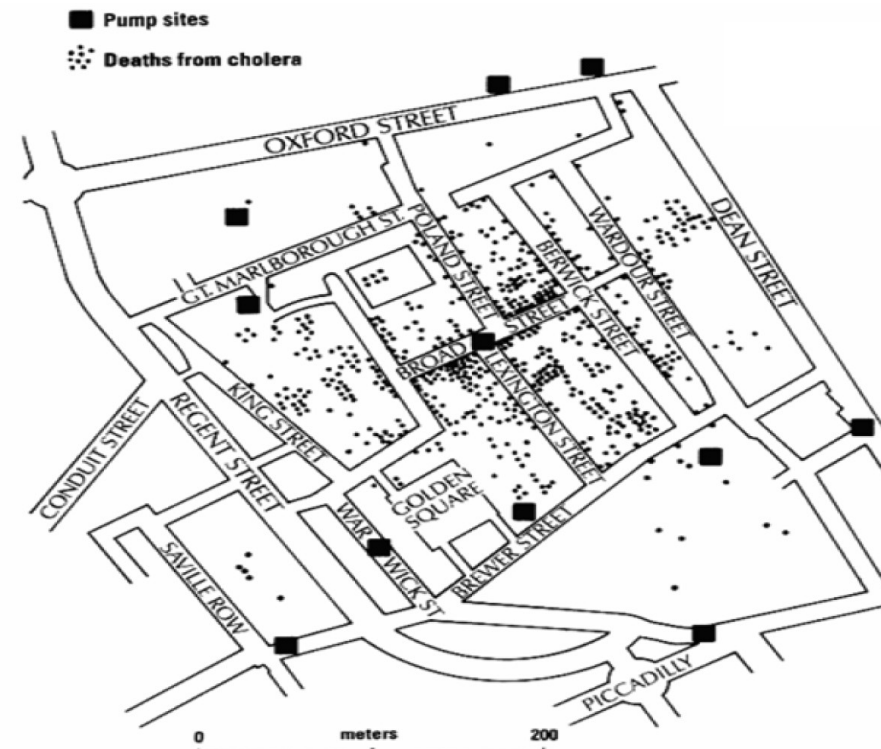


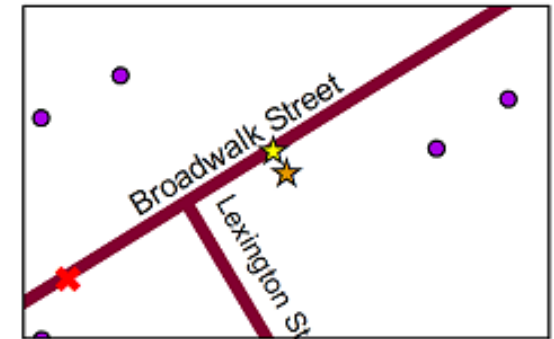
Figure 10: Three most preferred color ranges according to carto group

Ještě jeden příklad...

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

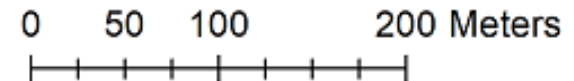


Ještě jeden příklad...

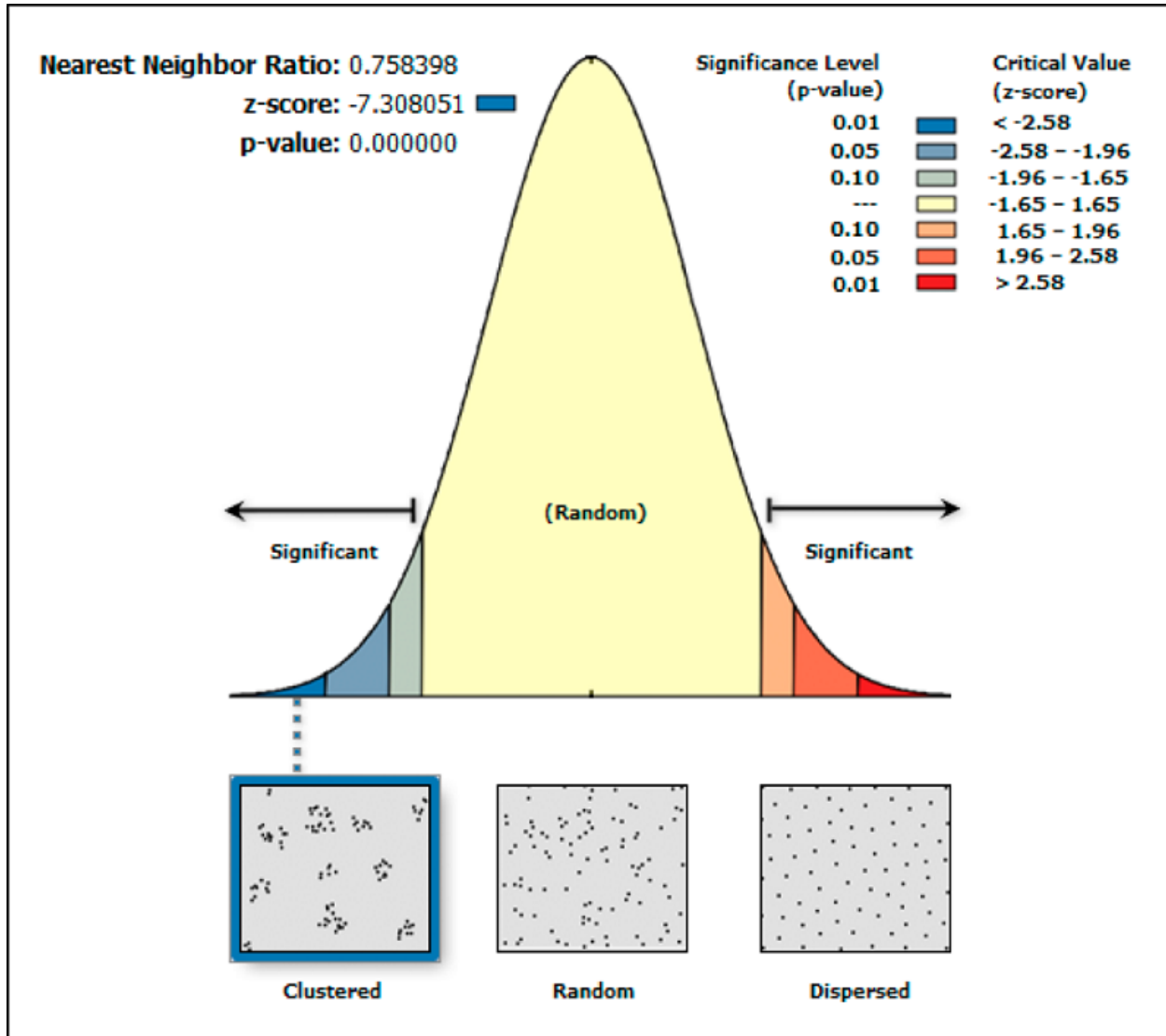


Legend

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

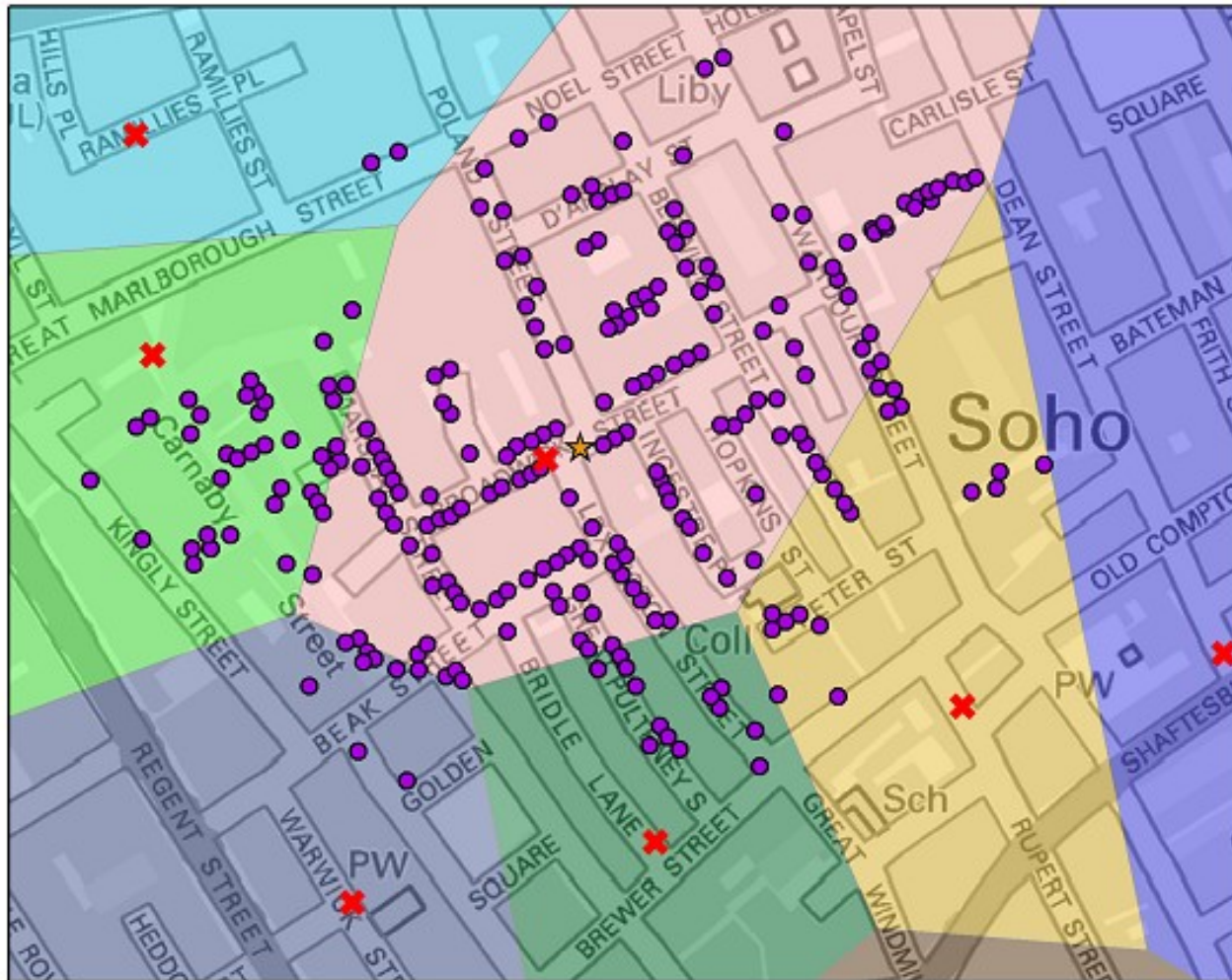


Ještě jeden příklad...





Ještě jeden příklad...

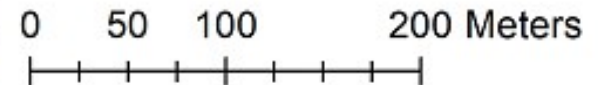


Legend

- ✖ Pumps
- Cholera Death
- ★ Mean Center

Pump ID

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





Ještě jeden příklad...

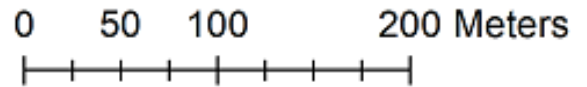


Legend

- ✕ 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
- Materiály předmětu Z6101 Základy geostatistiky
- A odkazy přímo na slajdech