



MUNI

GIS4SG

Mapování a modelování kriminality



podzim 2020

Petr Kubíček

kubicek@geogr.muni.cz

**Laboratory on Geoinformatics and Cartography (LGC)
Institute of Geography
Masaryk University
Czech Republic**



CRIME MAPPING AND ANALYSIS



Podstata prediktivního modelování

- Doposud jsme se zabývali problémem, jak počítač „vidí“ geografická data prostřednictvím popisných (deskriptivních) technik a vytváří z nich oblasti s určitými vlastnostmi.
- Další logický krok je použití „**prediktivních – předpovědních**“ technik k vytvoření **extrapolačních map předvídajících budoucí podmínky**.
- Využití v řadě oblastí:
 - **Predikce kriminality.**



The role of 'place' in crime

Two key considerations (Spencer Chainey)

- Crime has an inherent **geographical quality**
- Crime is **not randomly distributed**



Crime has an inherent geographical quality

The four dimensions of crime:

- **Legal** (a law must be broken).
- **Victim** (someone or something has to be targeted).
- **Offender** (someone has to do the crime).
- **Spatial** (it has to happen at a place - somewhere, in space and time).



Crime is not randomly distributed

If crimes were random:

- Equal chance of them happening anywhere at anytime.

But crime is not randomly distributed

- Concentrated into places of activity
 - Crime hotspots
- Series follow geographic patterns
 - Serious and volume crime



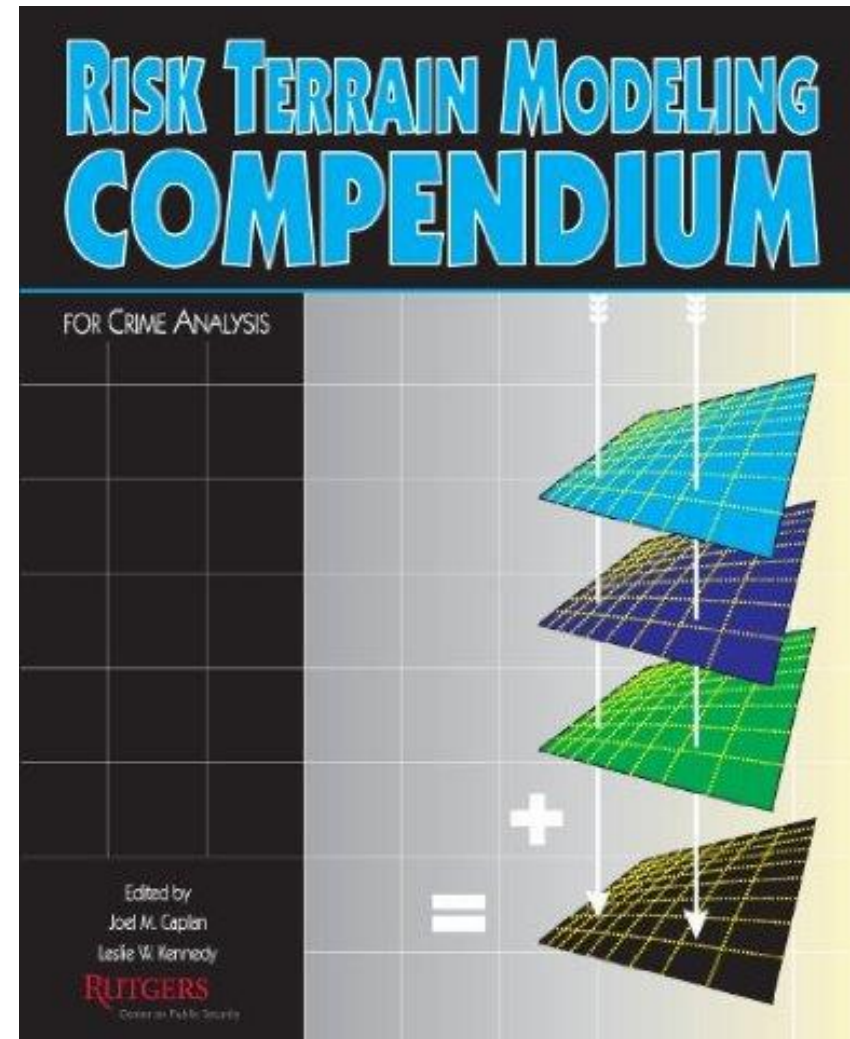
Predictive Crime Analysis

- **WHAT?**
- „Predictive policing in the context of place is the use of **historical data** to create a **spatiotemporal forecast** of **crime hot spots**.
- **WHY?**
- that will be the **basis for police resource allocation** decisions with the expectation that having officers at the proposed place and time **will deter or detect criminal activity.**“



Risk Terrain Modeling Prediction

- Risk terrain modeling (RTM) is an **approach to risk assessment** in which separate **map layers** representing the influence and intensity of a **crime risk factor** at every place throughout a geography is created in a geographic information system (GIS).
- Map layers are combined to produce a **composite “risk terrain” map** with values that account for all risk factors at every place throughout the geography.
- Available in PDF – ask your lecturer 😊





RTM steps

1. Select an outcome **event** of particular interest
2. Choose a study **area**
3. Choose a time **period**
4. Obtain **base maps** of your study area
5. Identify **aggravating** and **mitigating** factors related to the outcome event
6. **Select** particular **factors** to include in the RTM
7. **Operationalize** the spatial influence of factors to risk map layers
8. **Weight** risk map layers relative to one another
9. **Combine** risk map layers to form a composite map
10. **Finalize** the risk terrain map to **communicate** meaningful and actionable information.



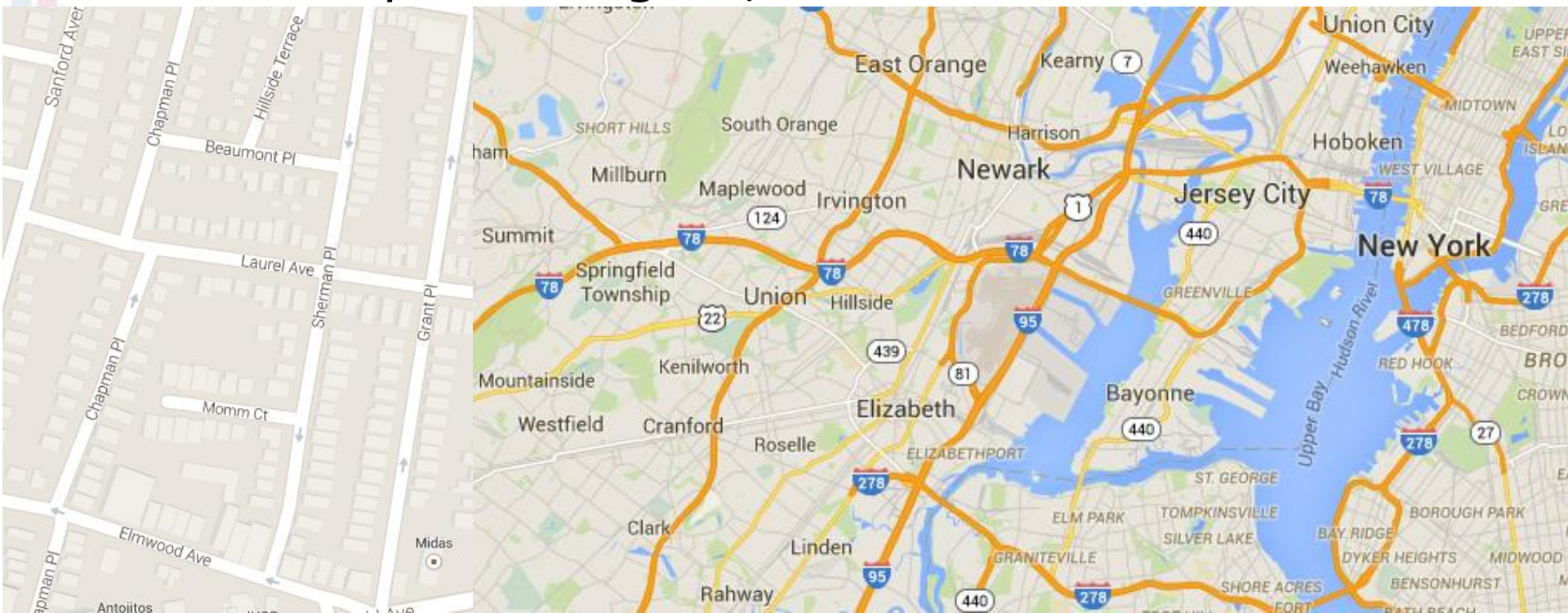
Step 1 -2

1. Select an outcome **event** of particular interest

Gun shooting incidents.

2. Choose a study **area on which risk terrain maps will be created.**

The Township of Irvington, NJ.





Step 3

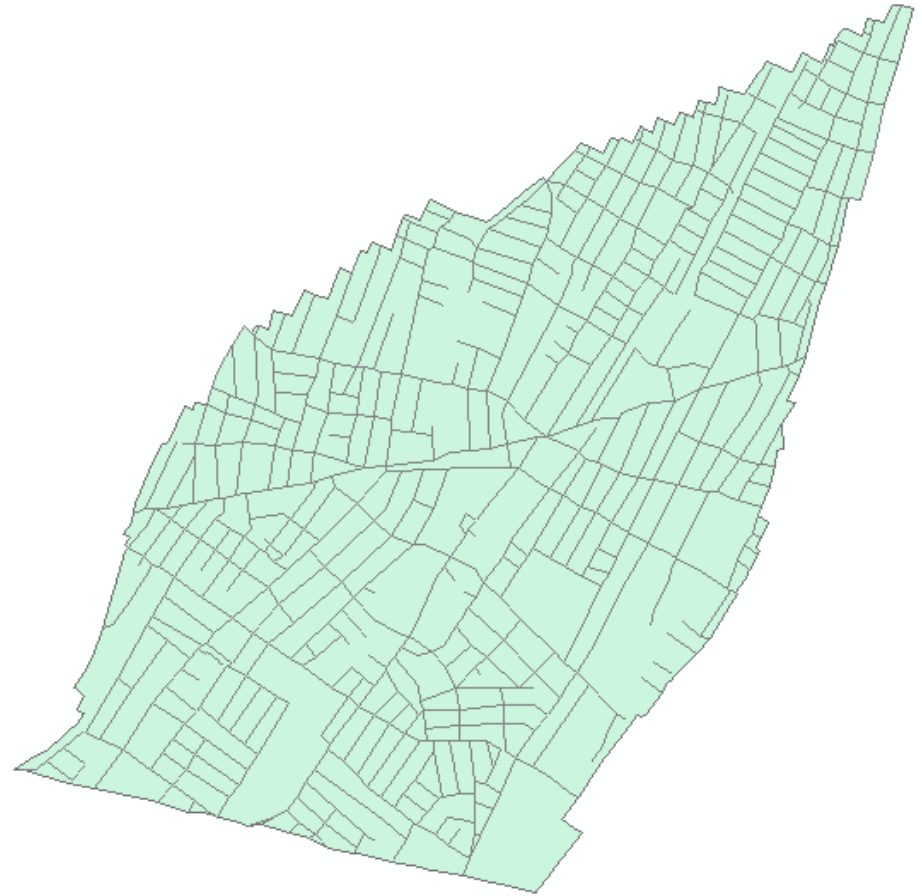
STEP 3: Choose a time period to create risk terrain maps for.

- Six month time period: January 1 to June 30.
- It is expected that this time period will adequately assess the place-based risk of shootings during the next 6-month time period (July 1 to December 31).
- **Data availability and comparability ?? Is it really justifiable and valid for the Czech Republic?**



Step 4

- ***STEP 4: Obtain base maps of your study area.***
- Two base maps were obtained from Census 2000 TIGER/Line Shapefiles:
 - 1) Polygon shapefile of the Township and
 - 2) **Street centerline** shapefile for the Township.





Step 5

STEP 5: Identify aggravating and mitigating risk factors that are related to the outcome event.

- Three **aggravating factors** were identified based on a ***review of empirical literature***:
 - dwellings of known gang members (**habitual offenders**),
 - locations of **retail business infrastructure** (bars, strip clubs, bus stops, check cashing outlets, pawn shops, fast food restaurants, and liquor stores),
 - locations of **drug arrests** (places, where the police action happened).



Step 6

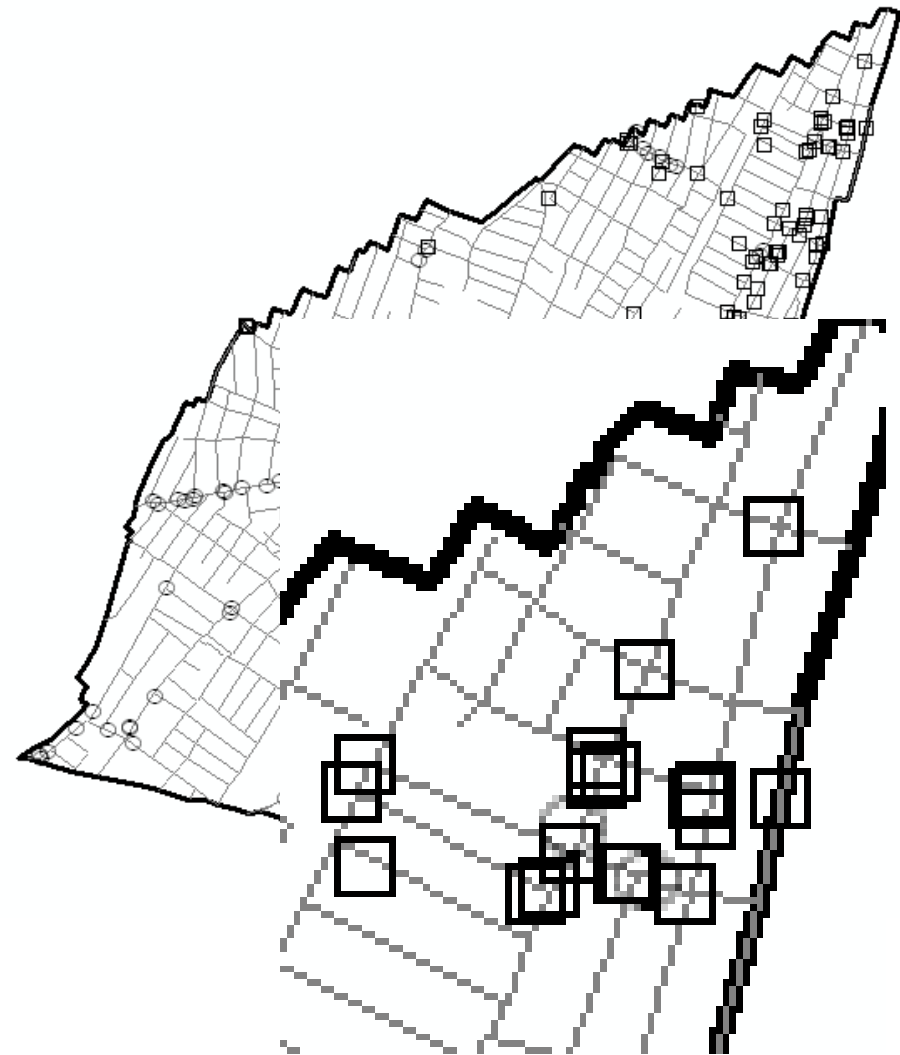
- ***STEP 6: Select particular risk factors to include in the risk terrain model.***
- All three risk factors identified in Step 5 will be included.
- Raw data in tabular form (i.e. Excel spreadsheets) was provided by the Township police and the many **datasets they maintain, validate and update regularly to support internal crime analysis and police investigations.**
- Attributes + **addresses** + time stamps + ??
- **State of the art of the investigation including the punishment and legal procedure.**



- ***STEP 7: Operationalize risk factors to risk map layers.***

- The tabular data was geocoded to street centerlines of Irvington to create point features representing:
 - the locations of gang members' **residences** (hidden on the map to protect the gang members),
 - retail **business outlets**,
 - and **drug arrests**, respectively as three separate map layers.

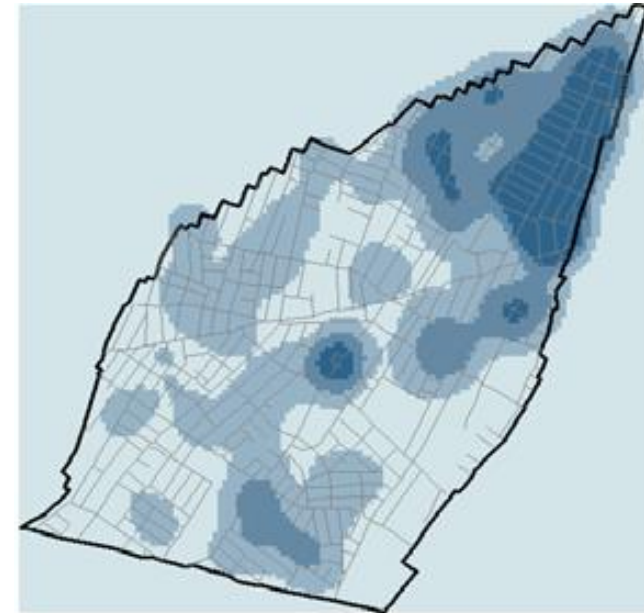
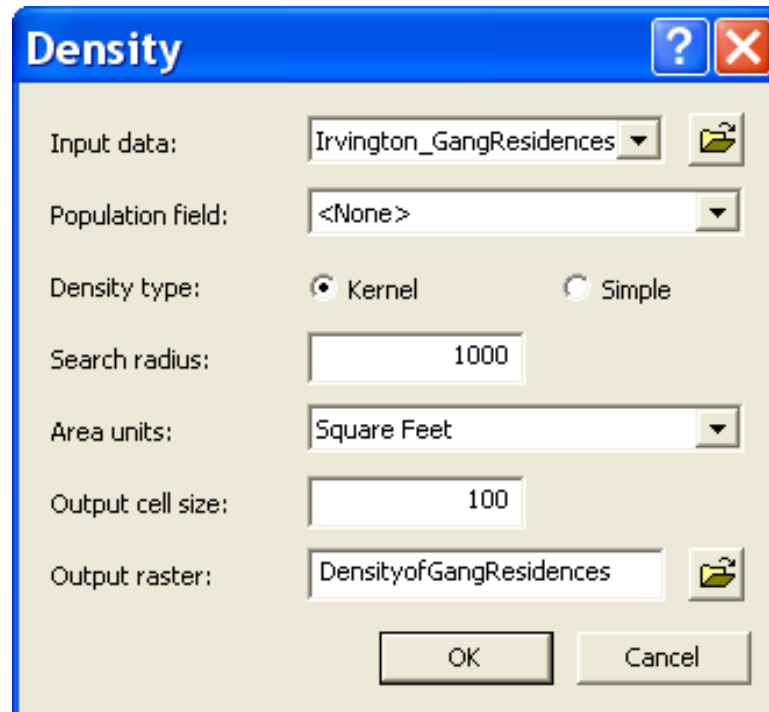
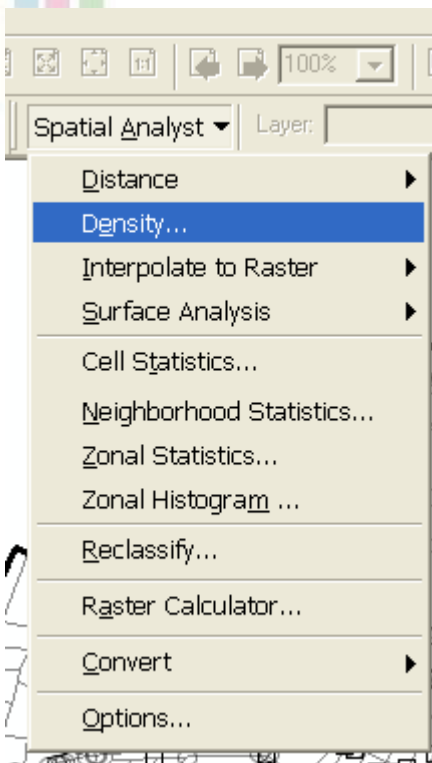
Step 7





Step 7a – gang member residence

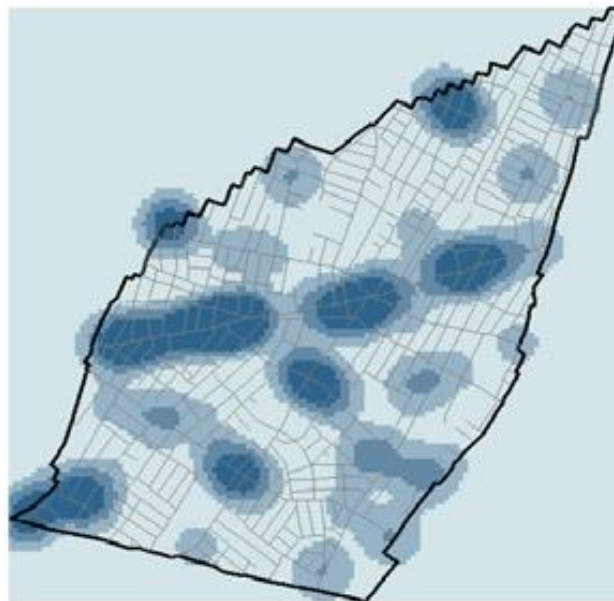
The spatial influence of the “gang members’ residences” risk factor was operationalized as: “Areas with **greater concentrations of gang members residing will increase the risk of those places having shootings.**” So, a **density map** was created from the points of gang members’ residences.





Step 7b - infrastructure

- The spatial influence of the “infrastructure” risk factor was operationalized as:
- “**High concentrations** of bars, strip clubs, bus stops, check cashing outlets, pawn shops, fast food restaurants, and liquor stores **will increase the risk** of those dense places having shootings.”

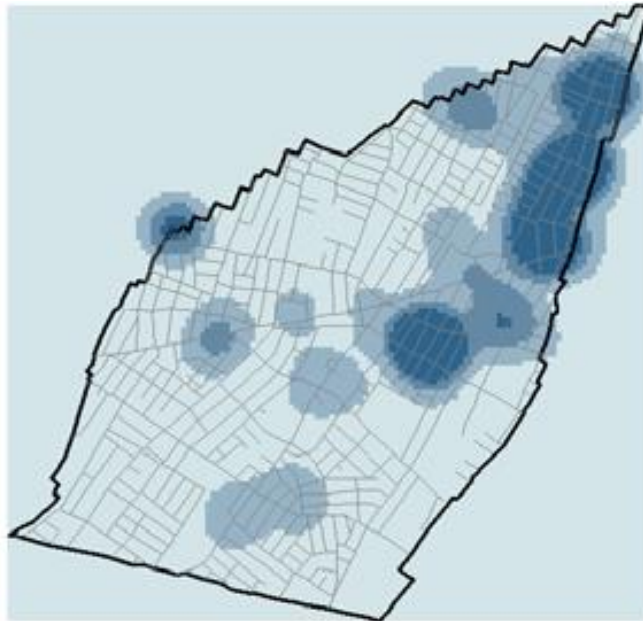




Step 7C – the drug arrest

the “drug arrest” risk factor was operationalized as:

- “Areas with **high concentrations** of drug arrests **will be at a greater risk for shootings** because these arrests create new ‘open turf’ that other drug dealers fight over to control.”





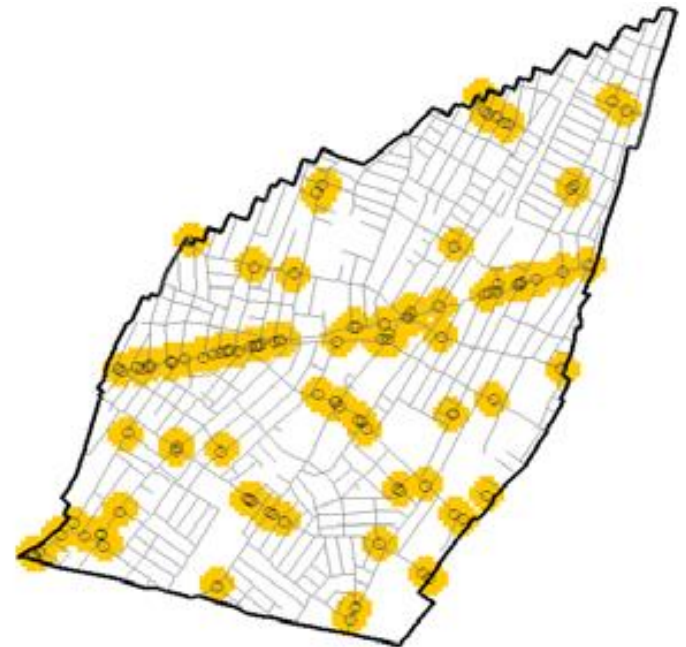
Step 7 – map density method details

- **Kernel density** values were calculated for each of the risk map layers so that *points lying near the center of a cell's search area would be weighted more heavily than those lying near the edge*, in effect smoothing the distribution of values.
- Cells within each density map layer were **classified into four groups according to standard deviational breaks**. The dark blue colored cells had values in the **top five percent** of the distribution and were considered the **"highest risk"** places.



Step 7d – distance from infrastructure

- The spatial influence of the “infrastructure” risk factor was also operationalized as:
- “The **distance of one block**, or about 350ft (app. 100 m), from a facility poses the greatest risk of shootings because **victims** are often **targeted** when **arriving** at or **leaving** the establishment.”

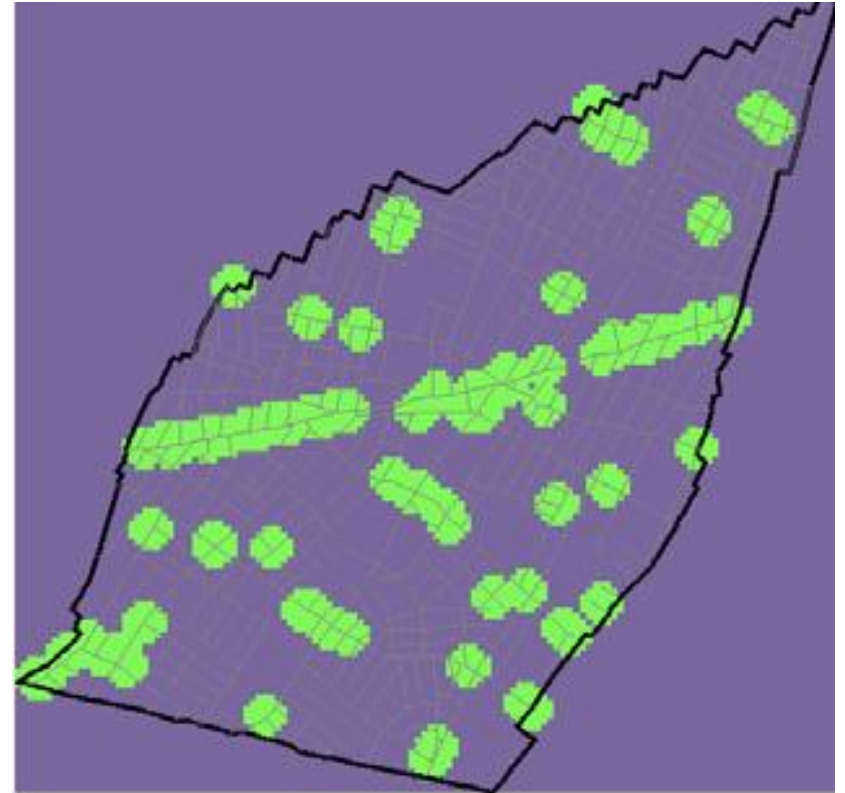
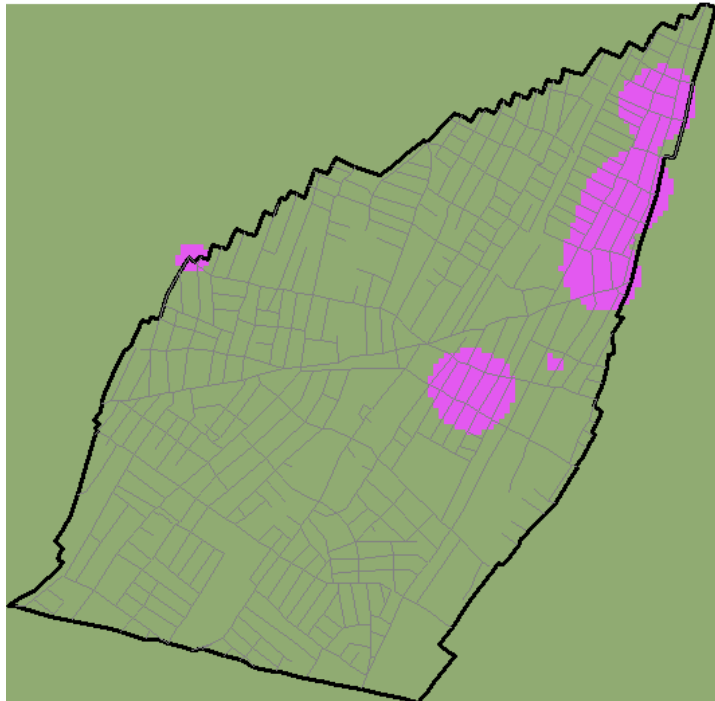
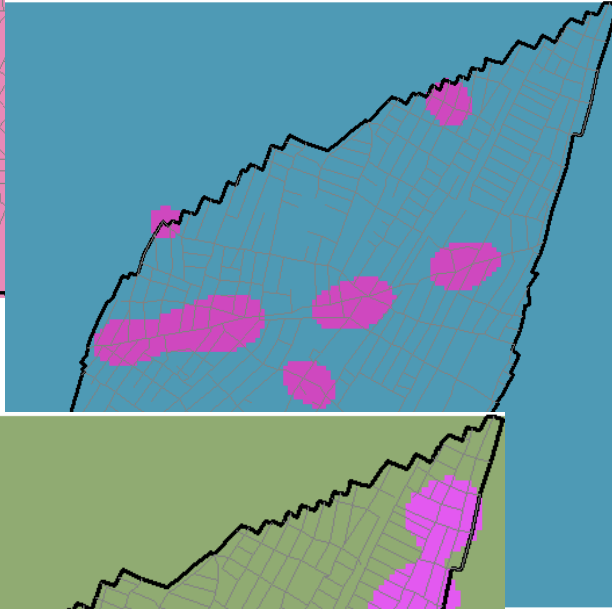
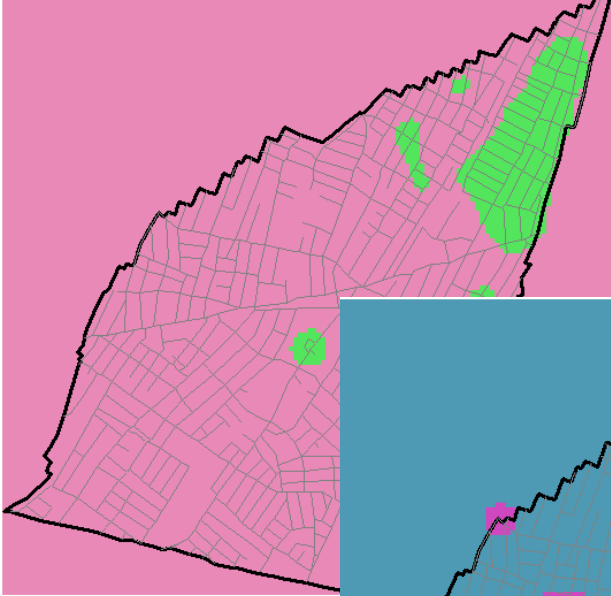




7e – final operationalization

- **We** are only interested in knowing where places are the most at risk for shootings, so we used a **binary-valued schema** to designate the “**highest risk**” places across all four risk map layers.
- The highest risk places of each risk map layer, respectively, will be given a value of “1”; all other places will be given a value of “0”.
- All risk factors are operationalized as **aggravating factors**, so these values will **remain positive**.

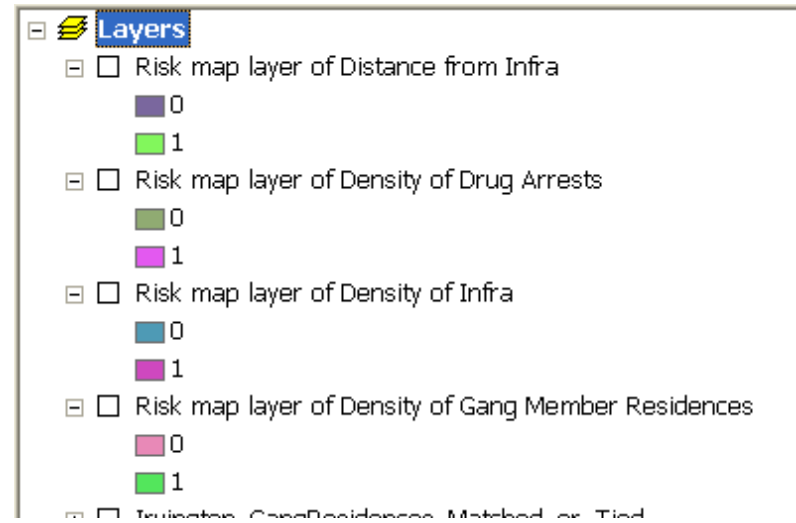
Step 7 - reclassification





Step 7 – final comparison

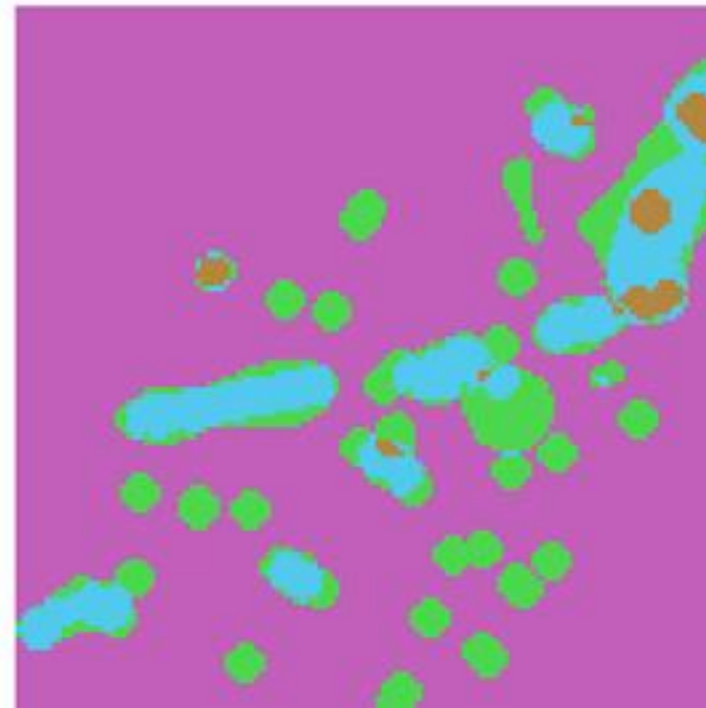
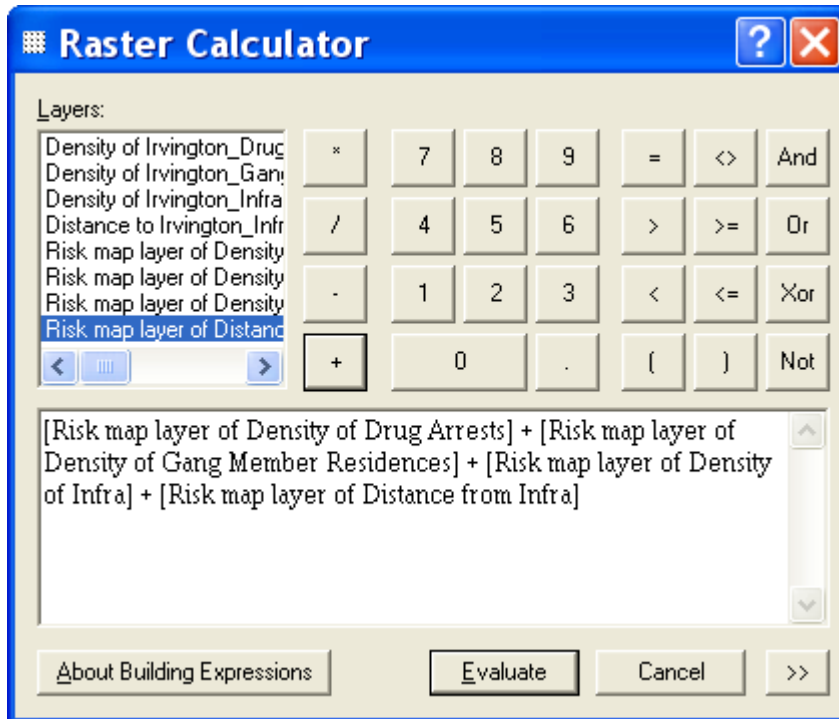
- We now have **four (final) risk map layers, operationalized from three risk factors.**
- **Binary** reclassification – 0 – 1
- The cells of different map layers are the same size and were classified in a standard way, the risk **map layers can be summed together** to form a **composite risk terrain map.**





Step 8 + 9 - *Inter Risk Map Layer Weighting and CRTM*

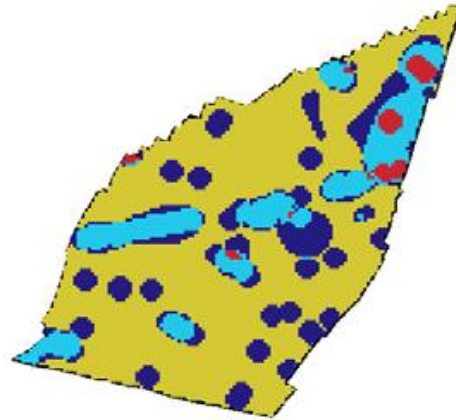
All risk **map layers** will carry equal weights to produce an **un-weighted risk terrain model**. It is assumed, for example, that being in a place with a high concentration of drug arrests **poses the same risk** of having a shooting as being in a place with a high concentration of gang member residences. Unless we know better 😊 !!





STEP 10 - Finalize the Risk Terrain Map to Communicate Meaningful Information.

- Clip our risk terrain map to the boundary of Irvington.

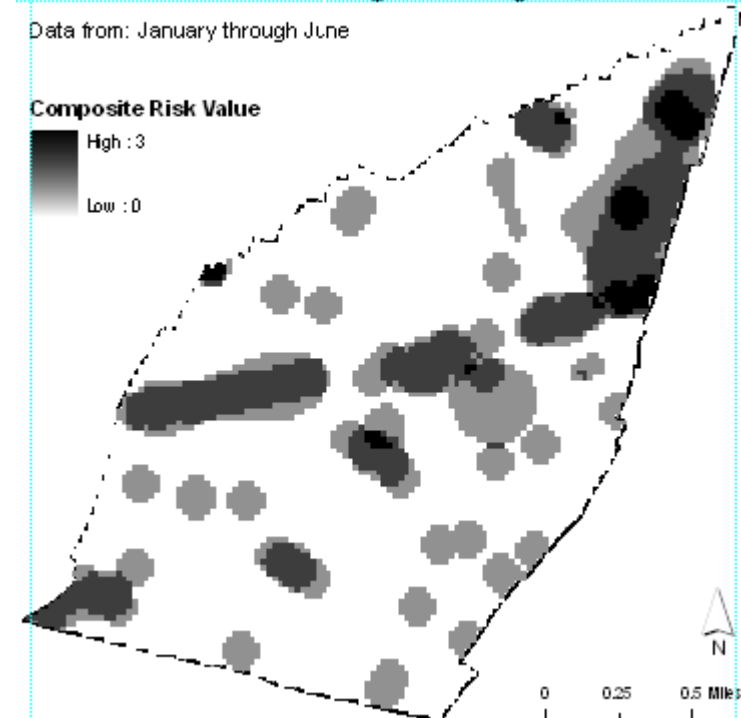
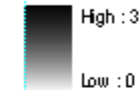


- produce a final map with shades of grey and layout.

Risk Terrain for Shootings in Irvington, NJ

Data from: January through June

Composite Risk Value





Step 10 – make the risk count

- convert the risk terrain map from raster to vector we can (still using the regular structure converted to square polygons):
- **count the number of shootings that actually occur in the high-risk areas during the subsequent time period;**
- calculate the **square area** of the highest risk areas (i.e., places with a composite risk value of 3);



Step 10 – make the risk count

- Select all street segments within these areas to inform police commanders about where patrols might be increased.
- Operationalise the command and control on the day by day basis.

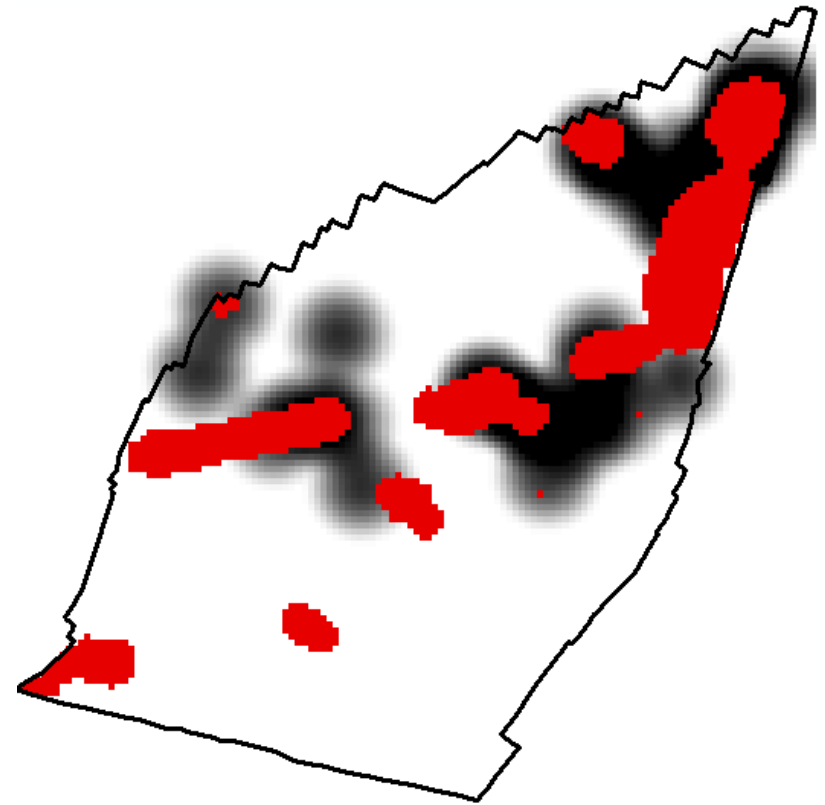
The screenshot shows the ArcMap interface with a map of street segments. The 'Layers' panel on the left lists various layers, including 'Composite Risk Value' and 'Irvington_Roads_Clippped'. The 'Selected Attributes of Irvington_Roads_Clippped' table is displayed at the bottom, showing a list of street segments with their attributes.

| FID | Shape | TLID | FNODE | MNODE | LENGTH | FENAME | FETYPE | FEDIRS | CFCC | FRADDL | TOADDL | FRADDR | TOADDR | ZIPL | ZIPR | CFCC1 | CFCC2 | SOURCE | COUNTRY |
|-----|----------|----------|-------|-------|---------|----------|--------|--------|------|--------|--------|--------|--------|------|------|-------|-------|--------|---------|
| 432 | Polyline | 63464488 | 8695 | 8434 | 0.14957 | 18th Ave | A41 | 376 | 438 | 395 | 399 | 07111 | 07111 | A | A4 | A | A | ESSEX | |
| 436 | Polyline | 63464492 | 8775 | 8695 | 0.04408 | 18th Ave | A41 | 354 | 374 | 335 | 353 | 07111 | 07111 | A | A4 | A | A | ESSEX | |
| 877 | Polyline | 63465432 | 8842 | 8775 | 0.03914 | 18th Ave | A41 | 344 | 352 | 343 | 353 | 07111 | 07111 | A | A4 | A | A | ESSEX | |
| 878 | Polyline | 63465433 | 8925 | 8842 | 0.0476 | 18th Ave | A41 | 328 | 342 | 327 | 341 | 07111 | 07111 | A | A4 | A | A | ESSEX | |
| 447 | Polyline | 63464503 | 8396 | 8127 | 0.14332 | 19th Ave | A41 | 171 | 235 | 172 | 234 | 07111 | 07111 | A | A4 | A | A | ESSEX | |
| 470 | Polyline | 63464527 | 8480 | 8396 | 0.04208 | 19th Ave | A41 | 161 | 169 | 156 | 170 | 07111 | 07111 | A | A4 | A | A | ESSEX | |
| 472 | Polyline | 63464529 | 8547 | 8460 | 0.0546 | 19th Ave | A41 | 141 | 159 | 140 | 154 | 07111 | 07111 | A | A4 | A | A | ESSEX | |
| 489 | Polyline | 63464553 | 8573 | 8522 | 0.10016 | 21st St | A41 | 372 | 410 | 371 | 411 | 07111 | 07111 | A | A4 | A | A | ESSEX | |
| 840 | Polyline | 63465009 | 8477 | 8490 | 0.02137 | 21st St | A41 | 0 | 0 | 413 | 417 | 07111 | 07111 | A | A4 | A | A | ESSEX | |



RTM validation

- **Comparison with the subsequent time period (June 1 – December 31) – high risk RTM classes and hot spot analysis of actual shooting accidents.**
- About 50% (15 out of 31) of the shootings during the subsequent time period (July 1 to December 31) happened in these high-risk cluster areas.





Things to remeber

- **Remember**, risk terrain modeling is only a *tool for spatial risk assessment*; it is not the solution to crime problems.
- You (the analyst) give **value and meaning to RTM**, so be innovative in your thinking about risk factors and how risk terrain maps can be applied to police operations.



PODSTATA METODY JÁDROVÝCH ODHAŮ

**Metodika idnentifikace anomálních lokalit
kriminality (Horák a kol. 2015)**



Metodika identifikace anomálních lokalit kriminality pomocí jádrových odhadů (Horák a kol. 2015)

- **Cíl** - doporučit standardizovaný postup využívání **metody plošných jádrových odhadů** pro identifikaci anomálních lokalit kriminality.
- Po krocích správně **připravit data**, nastavit a provést potřebné **analýzy** a zajistit **dosažení** vhodného **výsledku**.
- **Doporučuje** použití jednotlivých **variant** metody, optimalizaci jednotlivých parametrů pro jednorázová i opakovaná řešení.



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 - subjektivita v interpretaci výsledků.
- Stejná podkladová data mohou být zobrazena značně rozdílně jen s využitím rozdílného nastavení metody a způsobu zobrazení.
- Z tohoto důvodu je potřeba zvýraznit statisticky významné výsledky.



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).**
- **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.



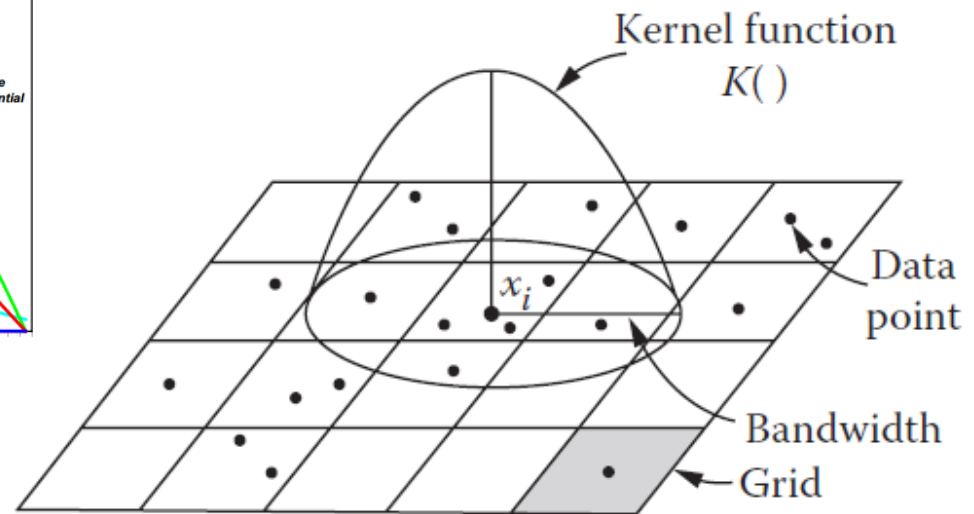
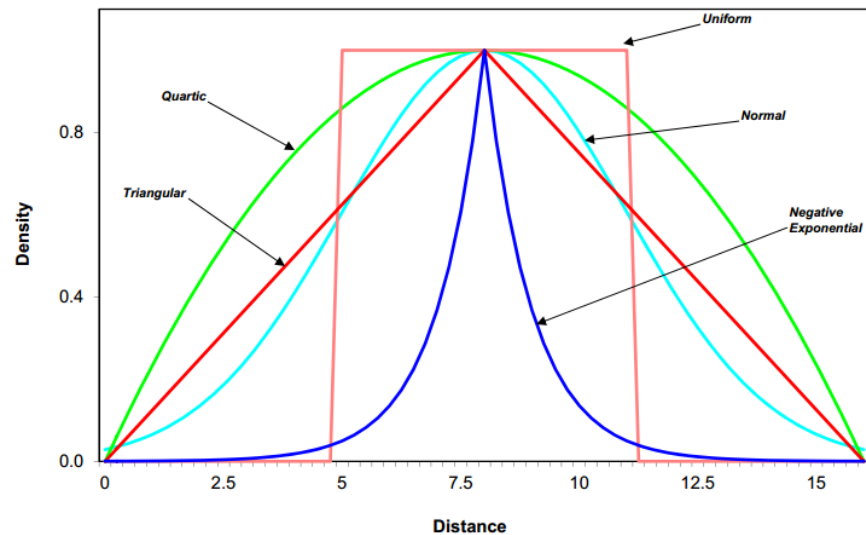
Krok II - VOLBA METODY

- 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**



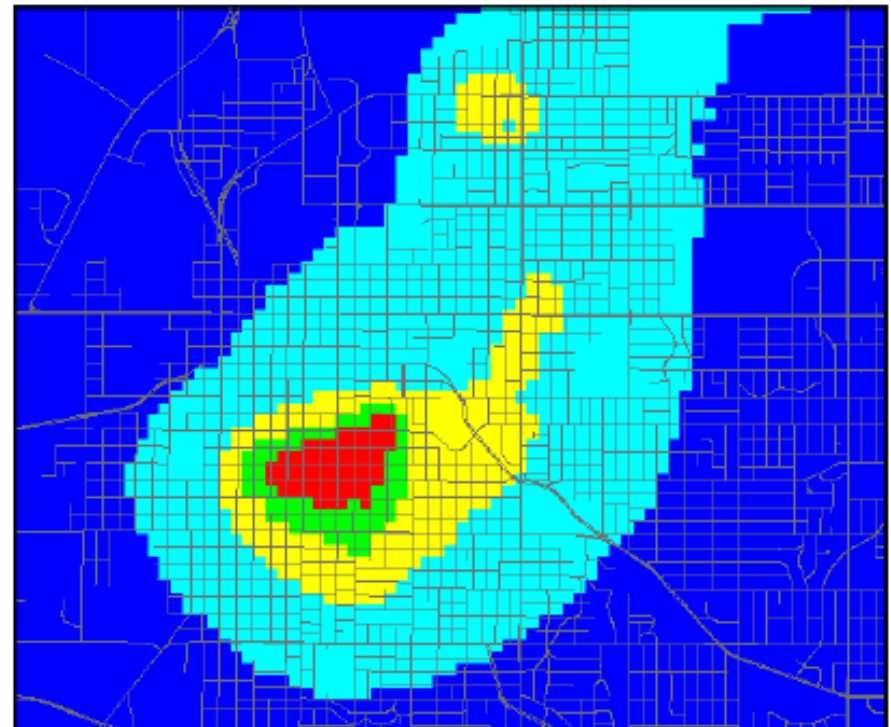
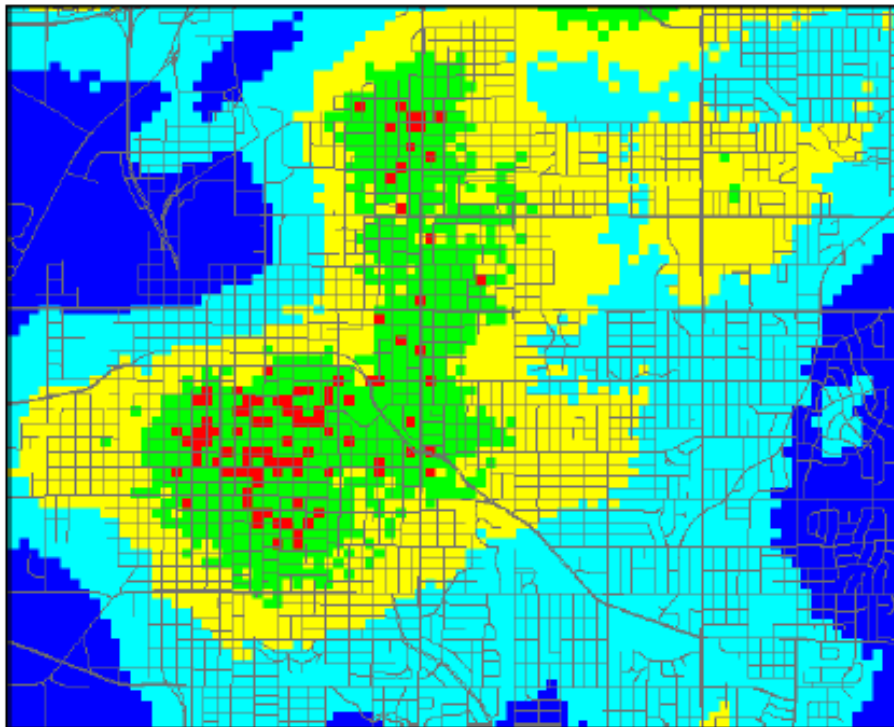
Krok III - VOLBA NASTAVENÍ - Vyhlazovací funkce

- 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. Min=10 m.**
- **Výjimky?**

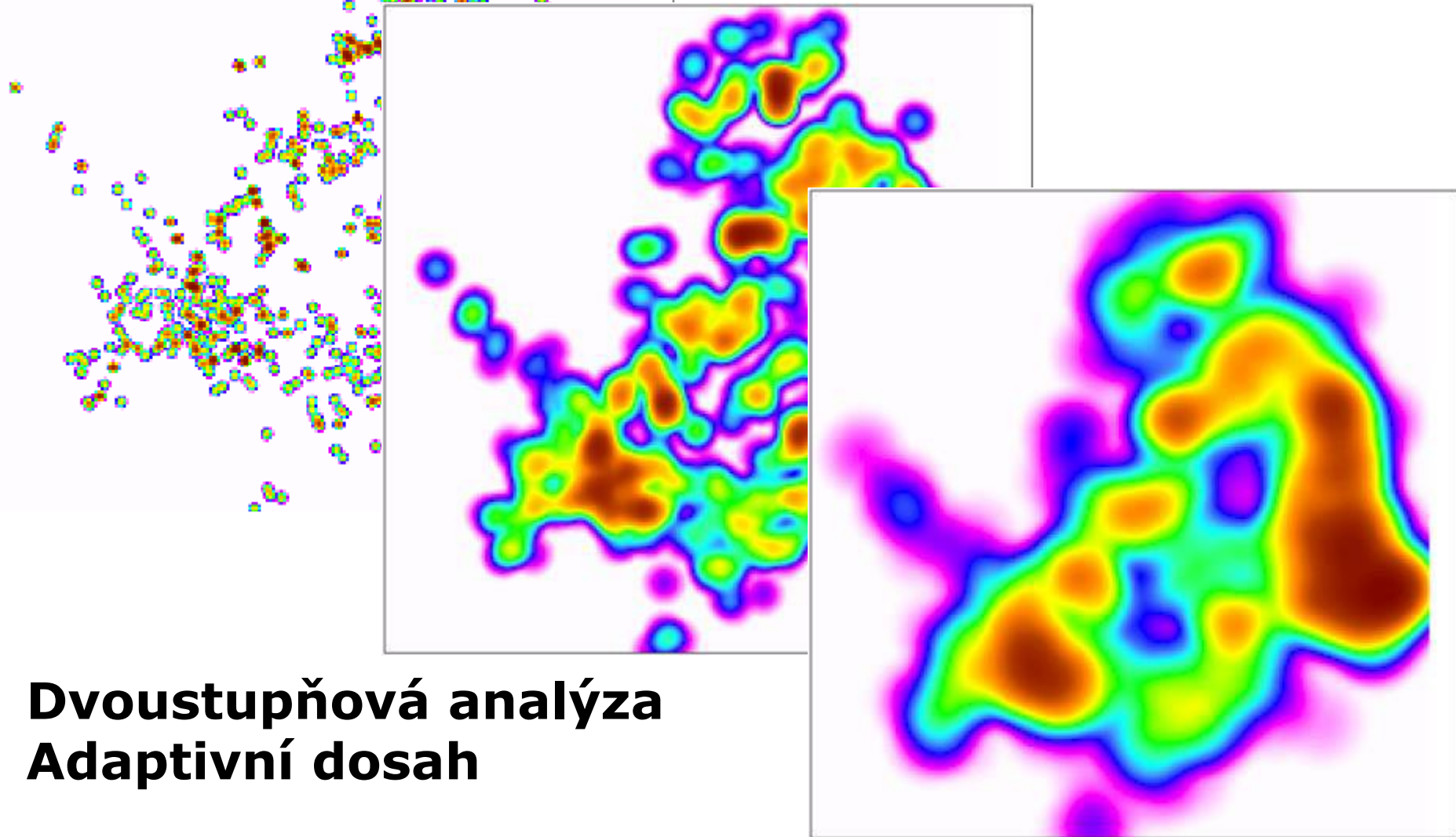


Dosah (šířka pásma)

- 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-score.
- Statisticky významné pozitivní z-score = čím větší, tím je intenzivnější shluk vysokých hodnot (hot spot).
- Statisticky významné negativní z-score =, čím menší z-score, 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.**
- **GI* hodnoty budou pozitivní pro všechny buňky**
- **Hodně nízkých hodnot pohromadě**
- **GI* hodnoty budou negativní pro všechny buňky**
- Příklad: Pro hodnotu 9 v centru vzorku platí:

GI* value = 4.1785

- *GI* hodnota je pozitivní*
- V relativní porovnání (lokální vs. Globální) *se jedná o hodně buněk s vysokou hodnotou trestného činu.*
- **Jaké jsou míry??**

| | | | | |
|---|----|---|---|---|
| 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 score
- Z scores 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 score používáno pro určení prahu spolehlivosti a zhodnocení statistické významnosti.

GI a GI* statistika

Statistická významnost

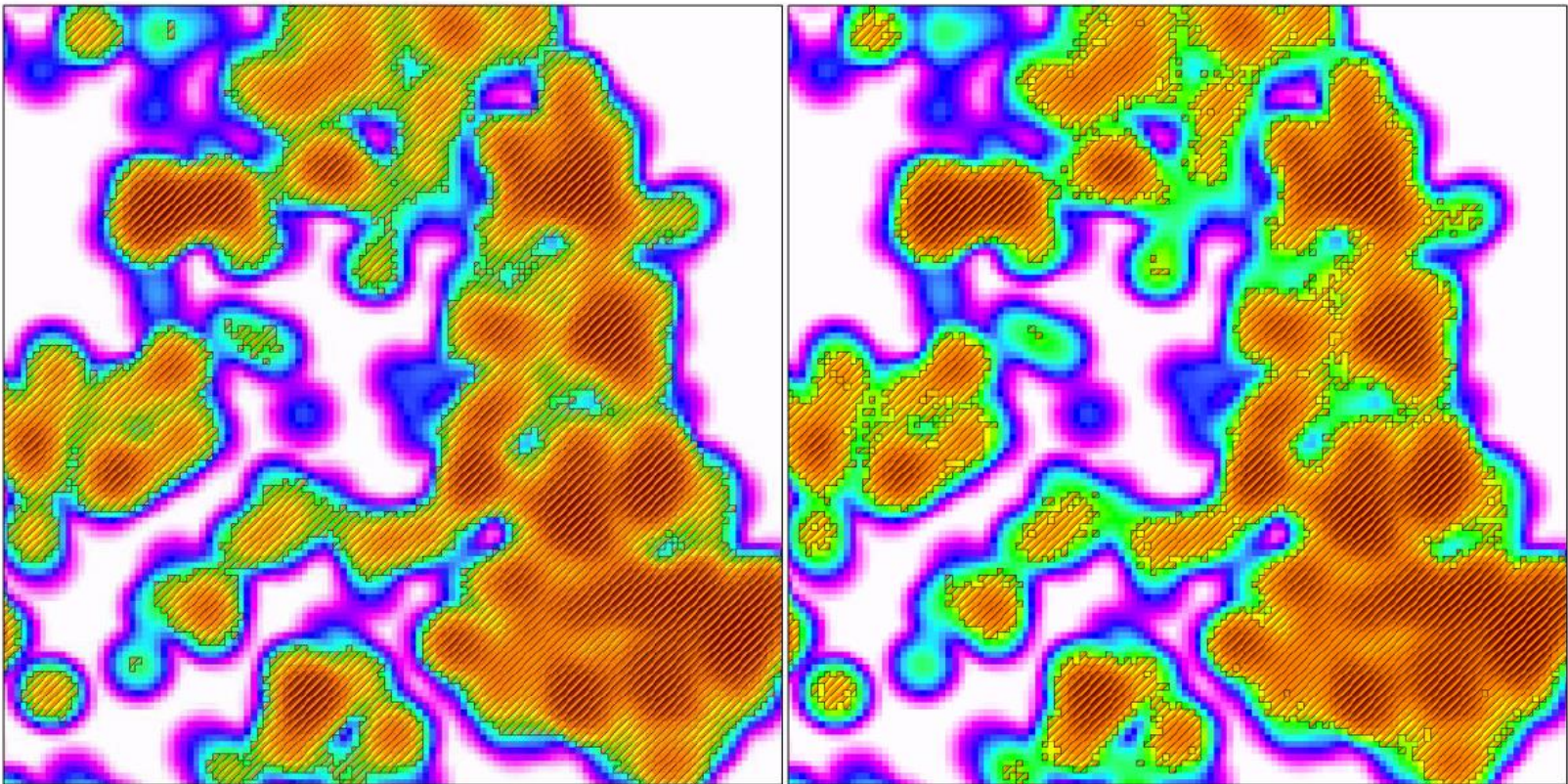
Z score hodnoty pro úrovně statistické významnosti:

- – 90% significant: ≥ 1.645
- – 95% significant: ≥ 1.960
- – 99% significant: ≥ 2.576
- – 99.9% significant: ≥ 3.291 (shluk trestné činnosti)
- **Univerzální Z score bez ohledu na typ trestné činnosti, umístění, velikosti území...**
- Příklad:
- G_i^* hodnota = 4.1785
- Větší než 99.9% významnost!



Statistická významnost

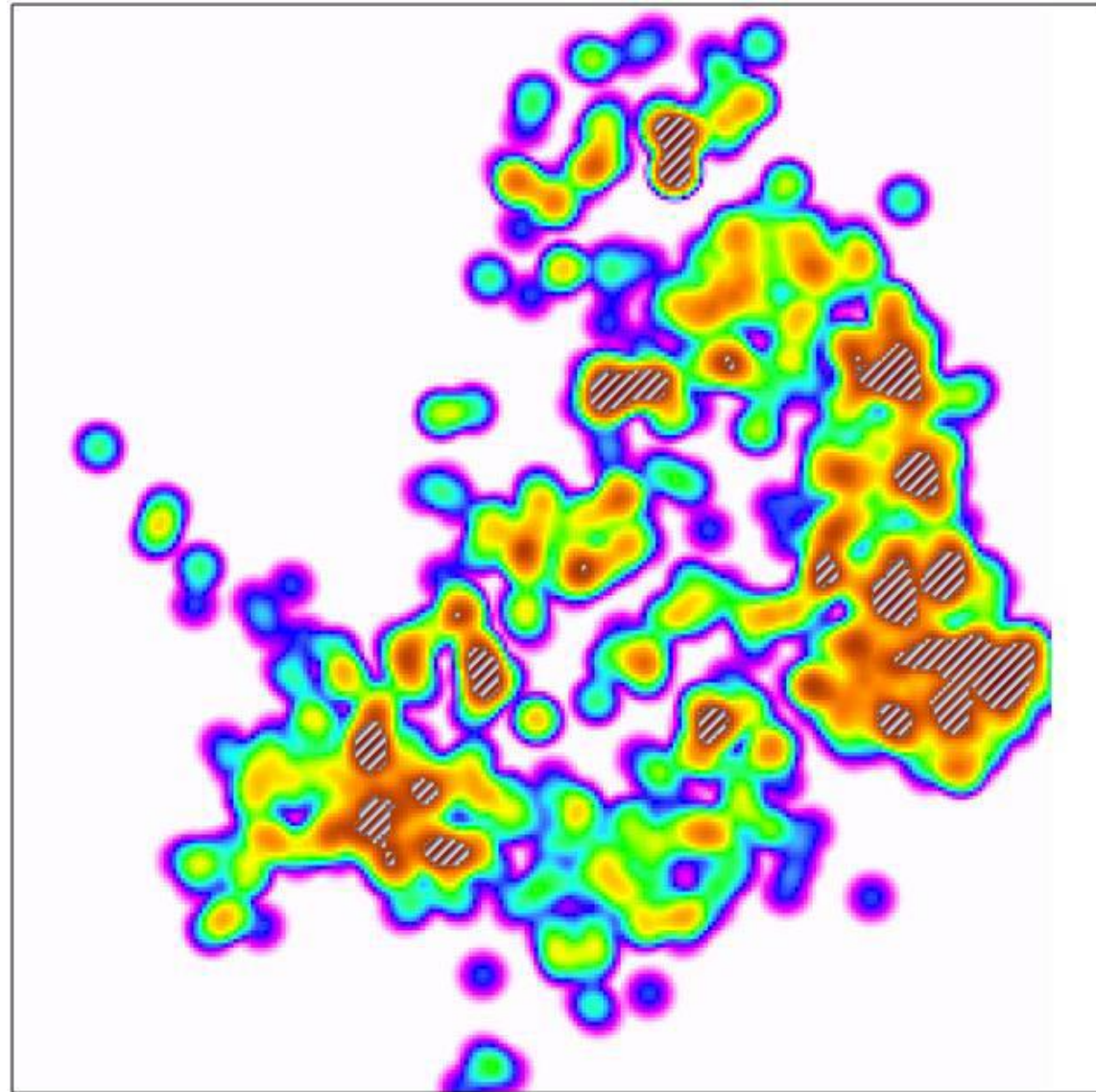
- Finální výsledky zobrazující statistické výsledky na hladině významnosti 95 % (vlevo) a 99 % (vpravo).



- Stačí to?? Kde je problém?

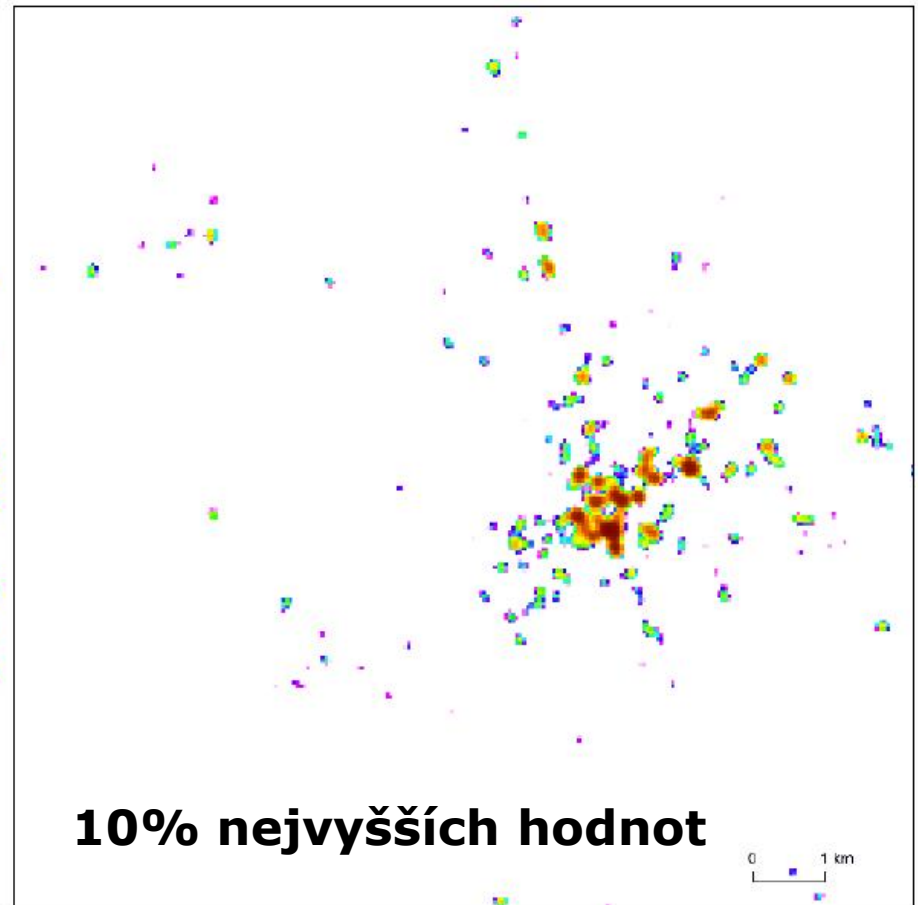
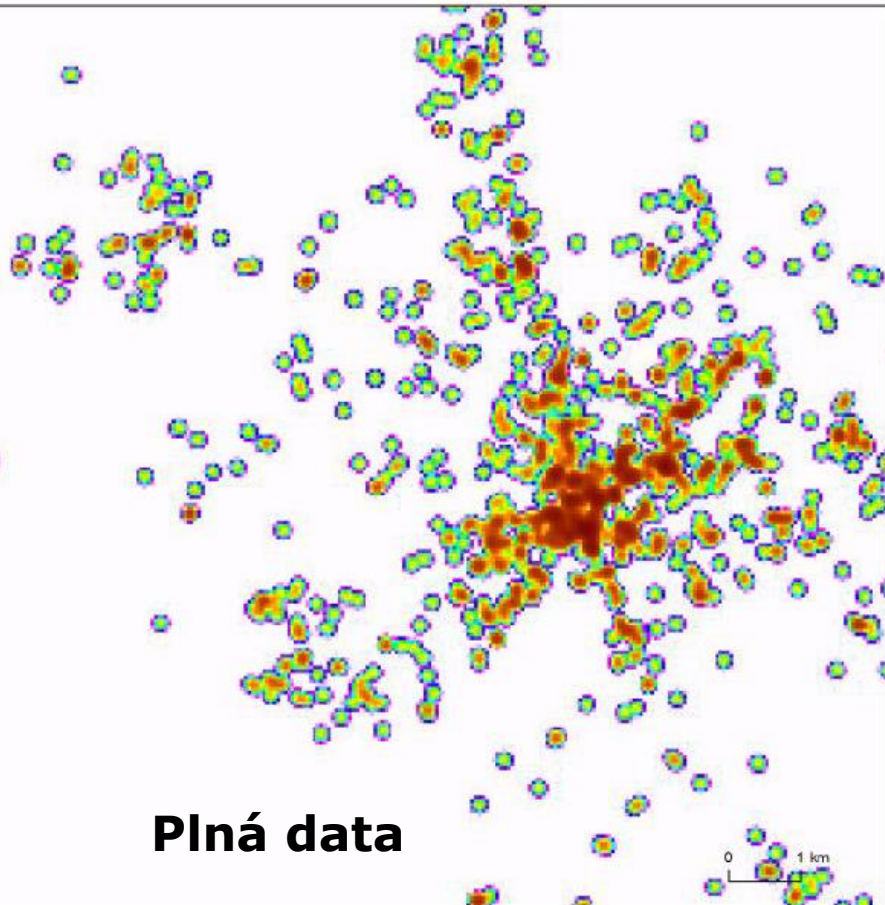
Statistická významnost

- **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^* .



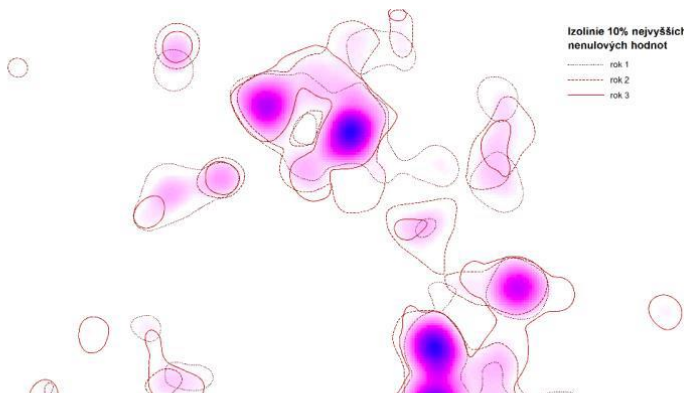
POSTPROCESSING A VIZUALIZACE

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



Vizualizace (alternativní)

- metody zobrazení – vícebarevné, trojrozměrné a izoliniové.
- Škály, podklad (topo), ortofoto.



Original

Radius=1

Radius=2

Radius=3

