GIS4SG

Mapování a modelování kriminality II

podzim 2017

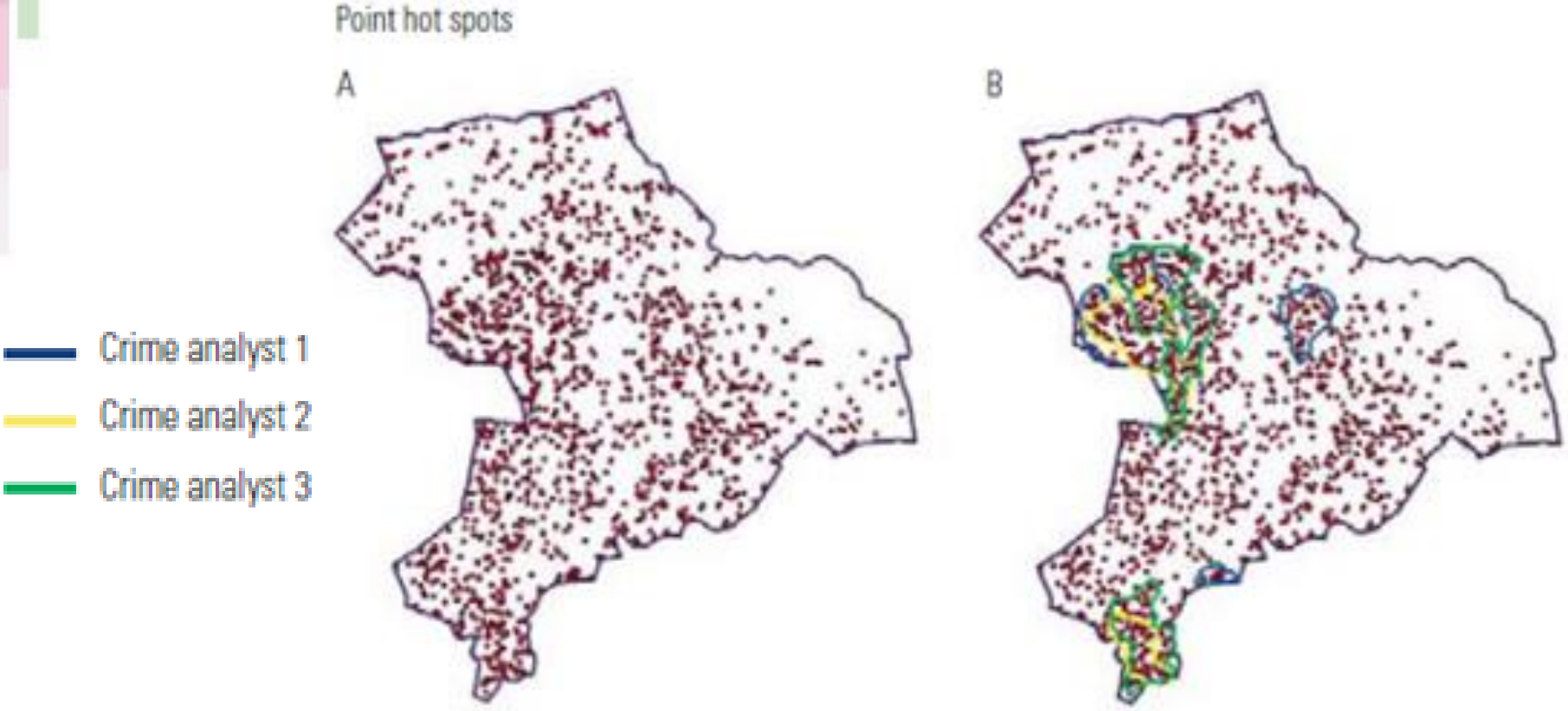
Petr Kubíček

kubicek@geogr.muni.cz

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

Crime mapping techniques - Point mapping

- The most common approach for displaying geographic patterns of crime is point mapping
- Interpret spatial patterns and hot spots in the crime point data can be difficult .

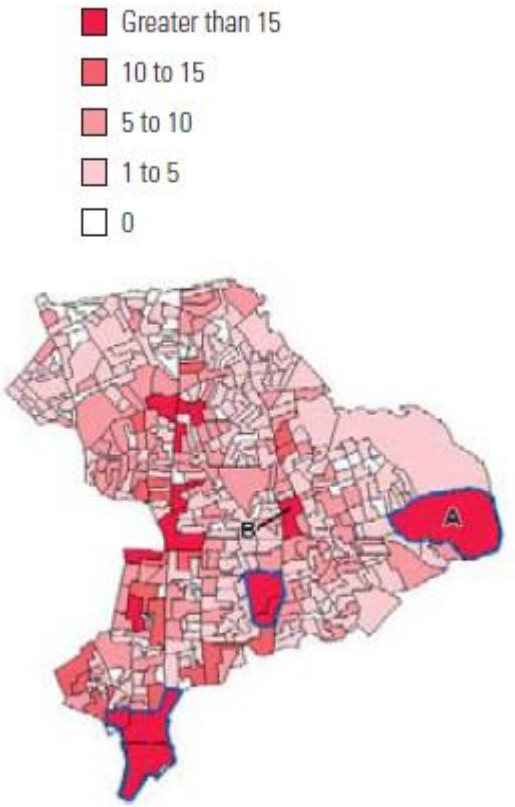


Point and graduated symbols

- Point maps do have their application for:
  - mapping individual events of crime,
  - small volumes of crime,
  - and repeat locations through the use of graduating symbol sizes
- less effective for identifying hot spots of crime, particularly from large data volumes.

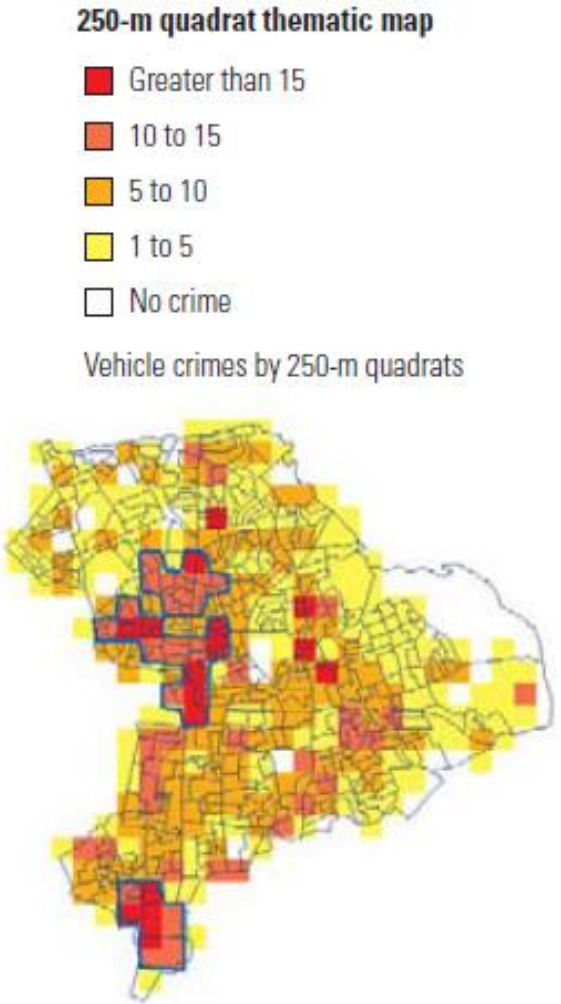
Thematic mapping of geographic boundaries

- A popular technique for representing any spatial distribution .
- Geographic boundaries usually are defined administrative or political areas such as census blocks, polling districts, wards, or borough boundaries.
- Due to the **varying size** and shape of most geographic boundaries, thematic shading can **mislead** the audience in identifying where **the spatial cluster of crime** may exist.



Quadrat thematic mapping – raster based analysis

- **Use of uniform grid.**
- Thematic value:
  - a count of crimes per grid cell - SUM.
  - a density value calculated from the count and cell area.
- Uniformity - **loss of spatial detail** within each quadrat and across quadrat boundaries. This can lead to **problems of inaccurate interpretation**.



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 intepretaci 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

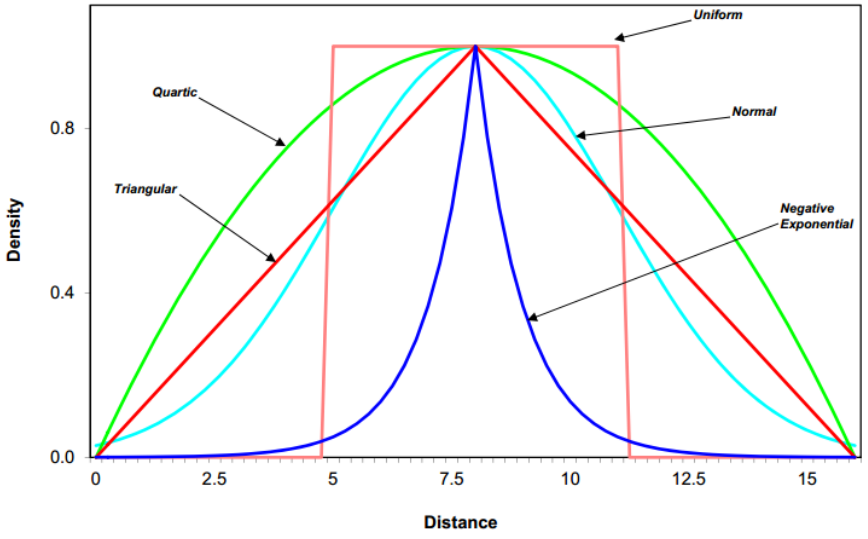
- Základní podmínkou - kvalitní data.
- Nutné se zaměřit na:
  - správnost a přesnost souřadnicového určení polohy,
  - časové určení,
  - tematické určení.
- Rozlišit případy, kdy již záznam deliktu obsahuje souřadnice, od těch, kde je poloha vyjádřena pouze adresou či jiným referencováním.
- Pokud 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.

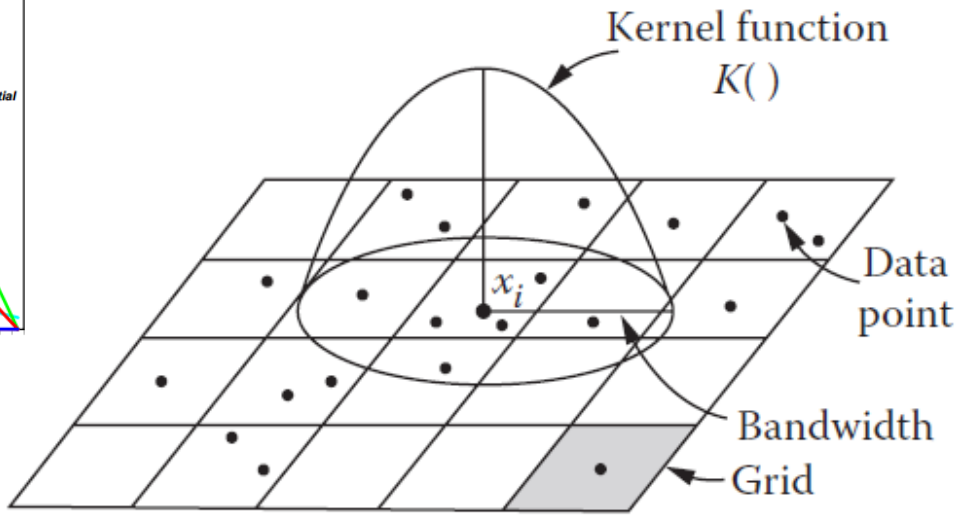
Krok II - VOLBA METODY

- 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

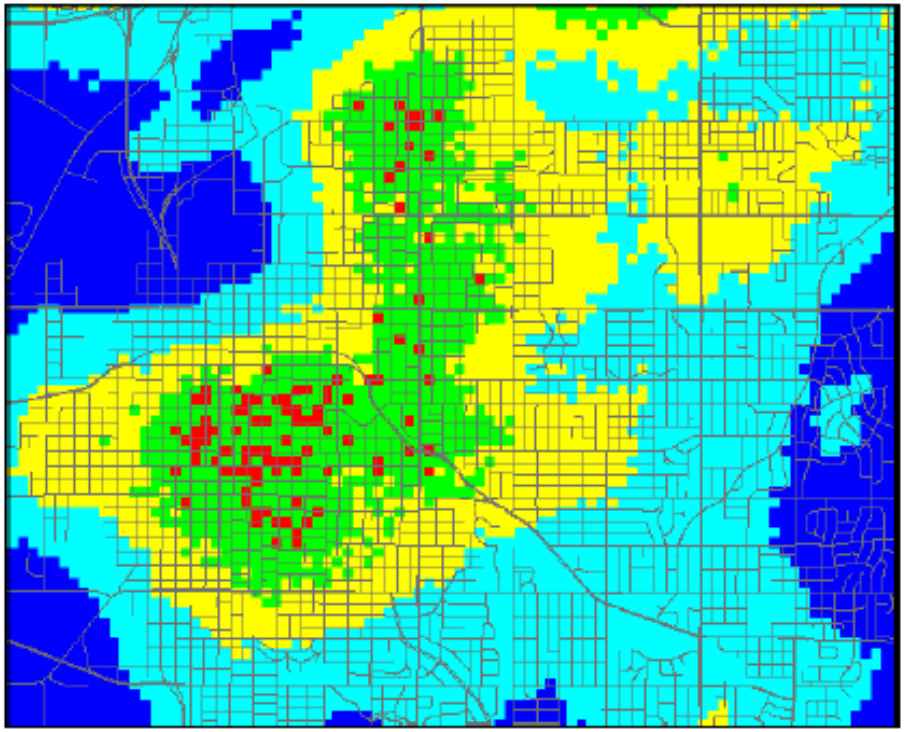
- šest různých vyhlazovacích funkcí: normální, rovnoměrná, kvartická, kuželová, kvadratická a záporná exponenciální.
- nejčastěji se využívá kvartická funkce,

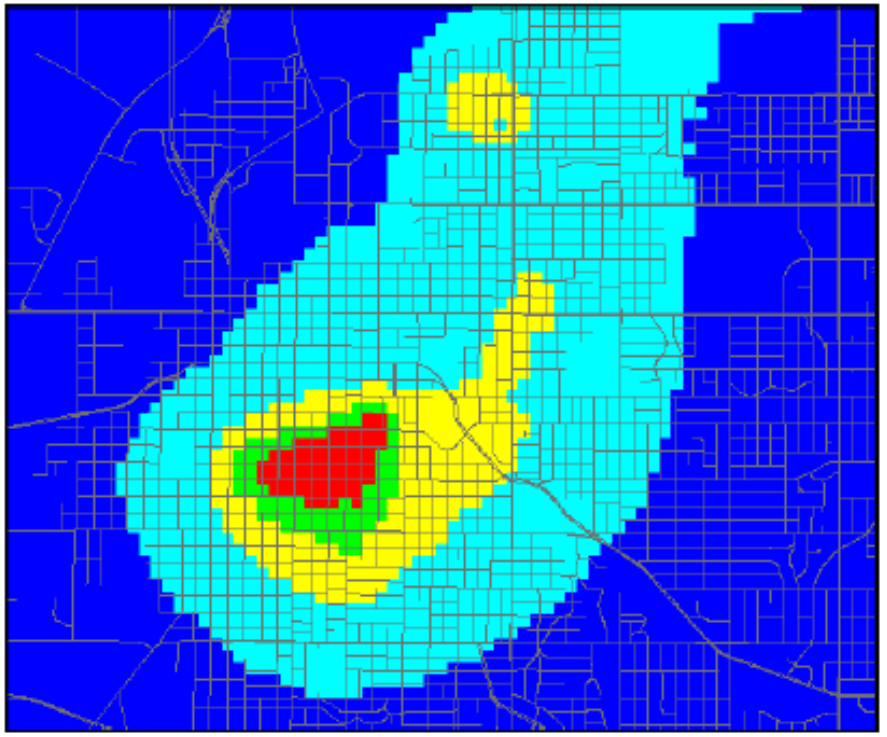




Závislost na zvolené vyhlazovací funkci

Trojúhelníková vs. Gausova (normální)





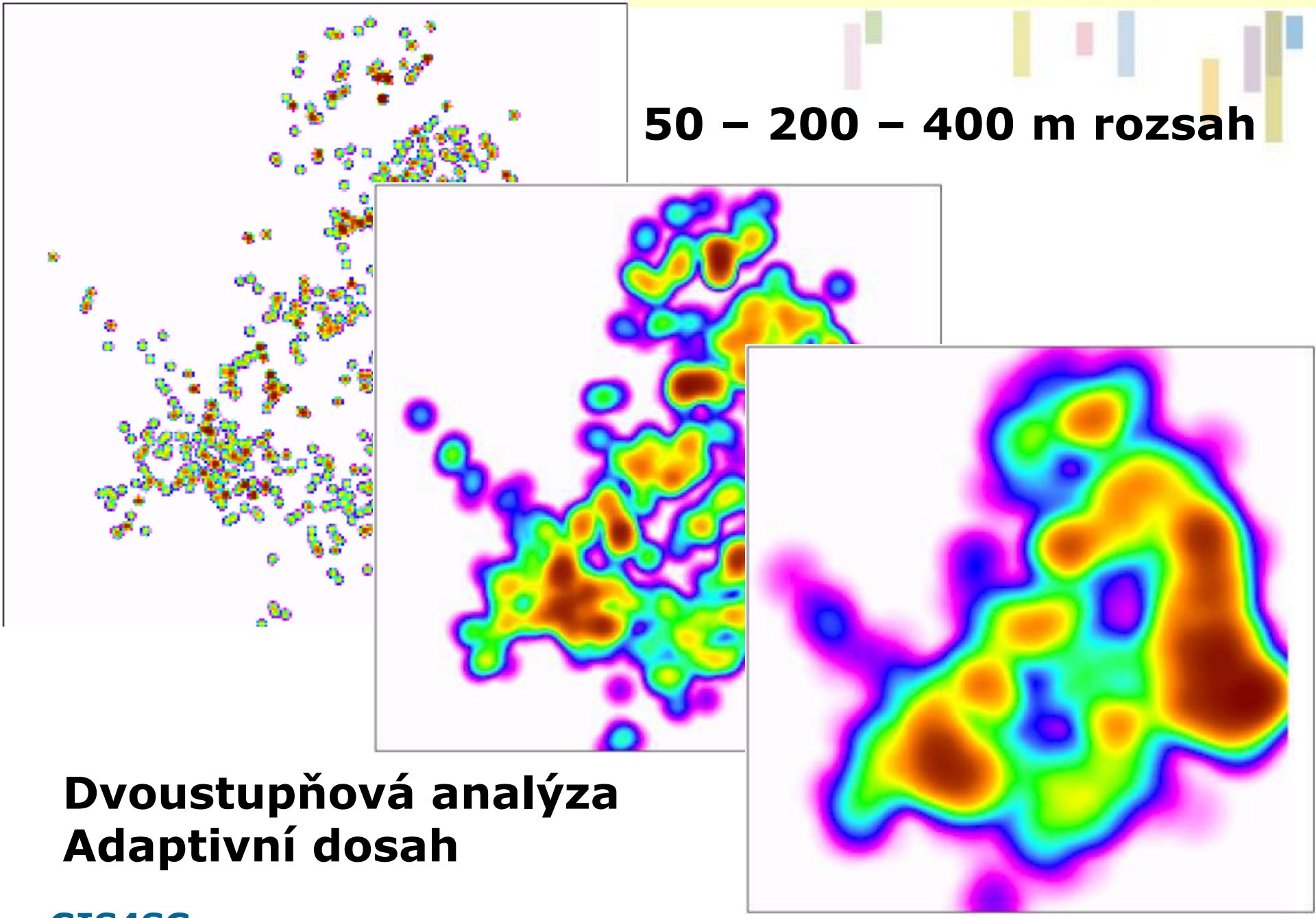
Velikost buňky

- 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	pásmo (m)
Loupež	Kvartická	200
Výtržnictví	Kvartická	200
Úmyslné ublížení na zdraví	Kvartická	200
Krádeže motorových vozidel	Válcová	100
Krádeže věcí z aut	Válcová	100



Dvoustupňová analýza
Adaptivní dosah

GIS4SG

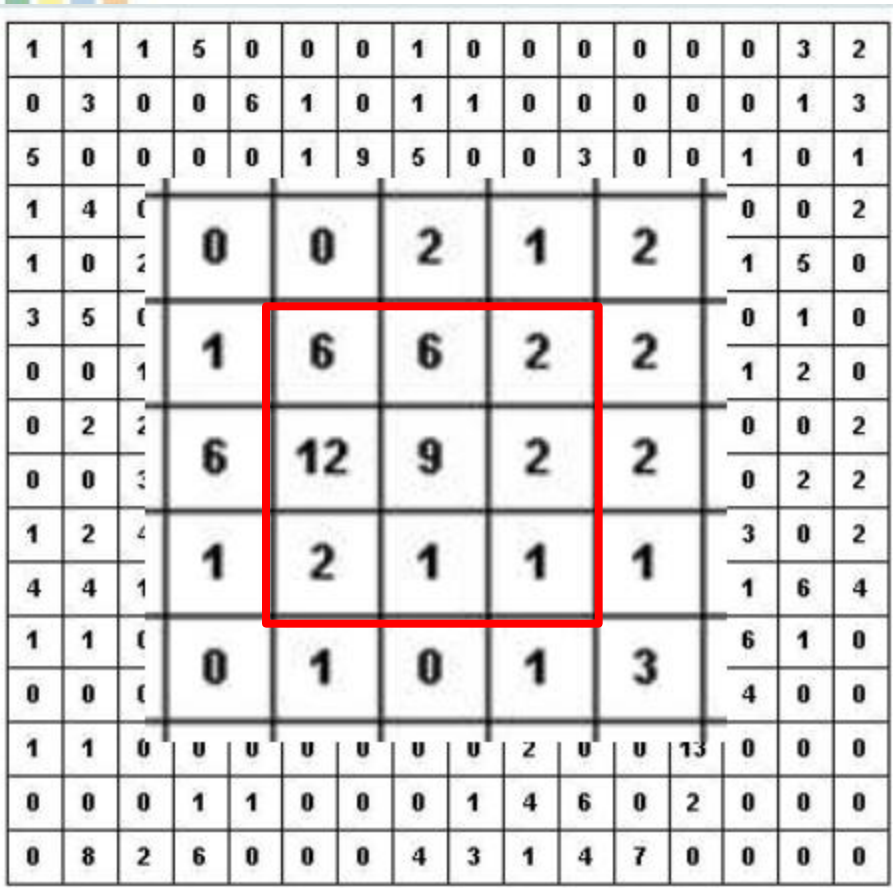
Krok IV - PROVĚŘENÍ STATISTICKÉ VÝZNAMNOSTI

- 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 Gi* index**.
- Pro výpočet Gi* 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 GI*

- Ukazatel významnosti shluku.
- Gi* 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



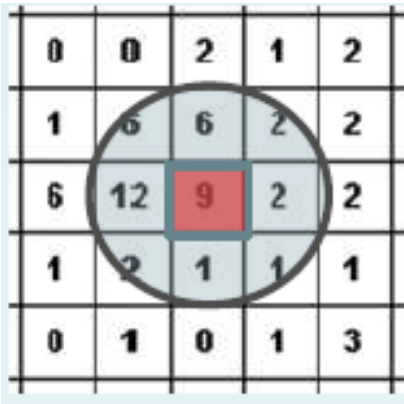
- Každá buňka má jednoznačnou hodnotu.
- Nulová hypotéza:
- **Není žádný vztah mezi hodnotami počtu trestných činů v buňce a v jejím okolí, a to až do vzdálenosti d měřené ve všech směrech. Srovnáno se sumou hodnot na celém studovaném území.**

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 negatvní pro všechny buňky**
- Příklad: Pro hodnotu 9 v centru vzorku platí:

  Gi* value = 4.1785
- *Gi* hodnota je pozitivní*
- V realativní porovnání (lokální vs. Globální) *se jedná o hodně buněk s vysokou hodnotou trestného činu.*
- ***Jaké jsou míry??***



GI a GI* statistika

- Gi* 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

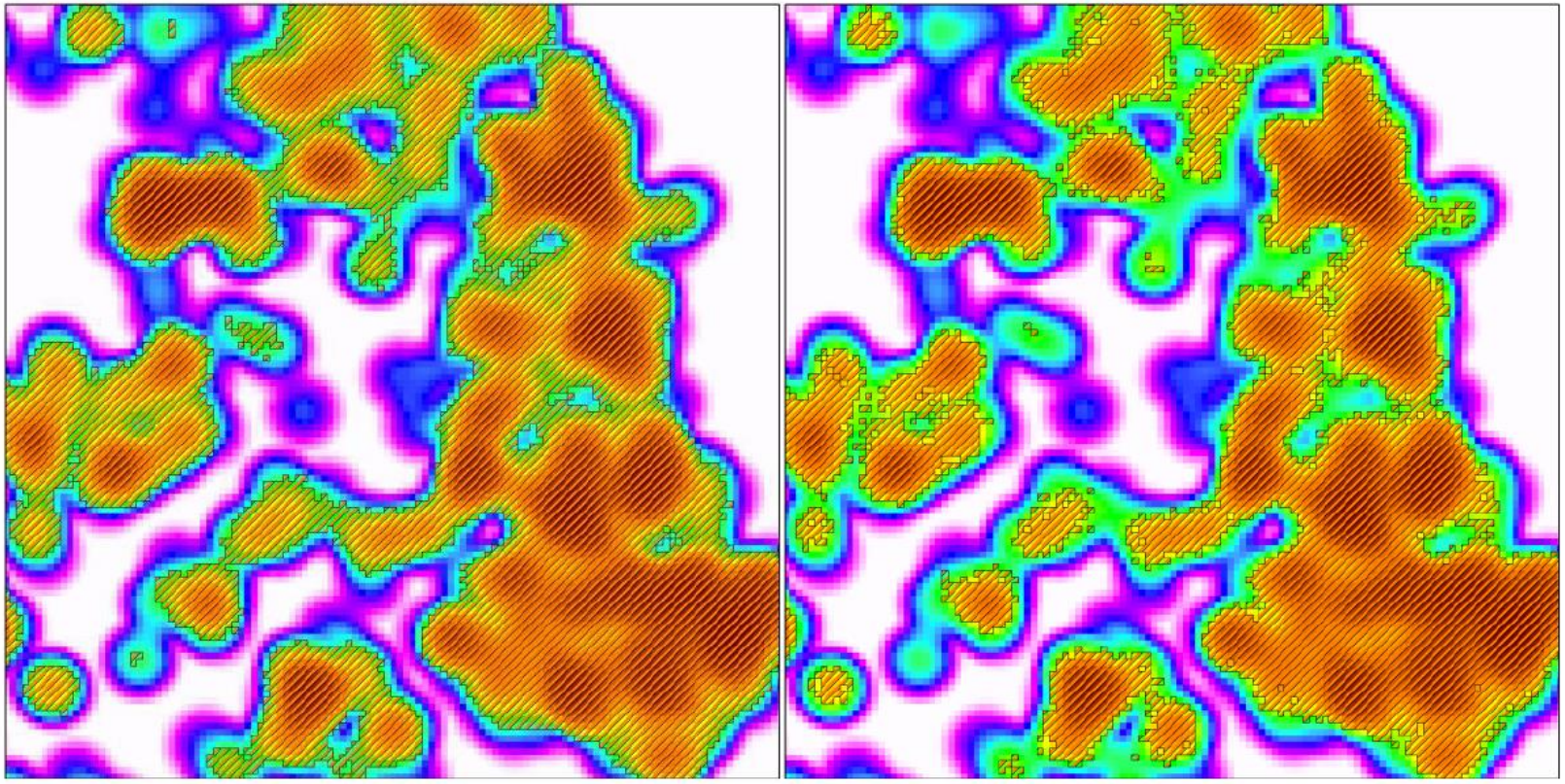
Statistická významnost

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:
- Gi* hodnota = 4.1785
- Větší než 99.9% významnost!

Statistická významnost

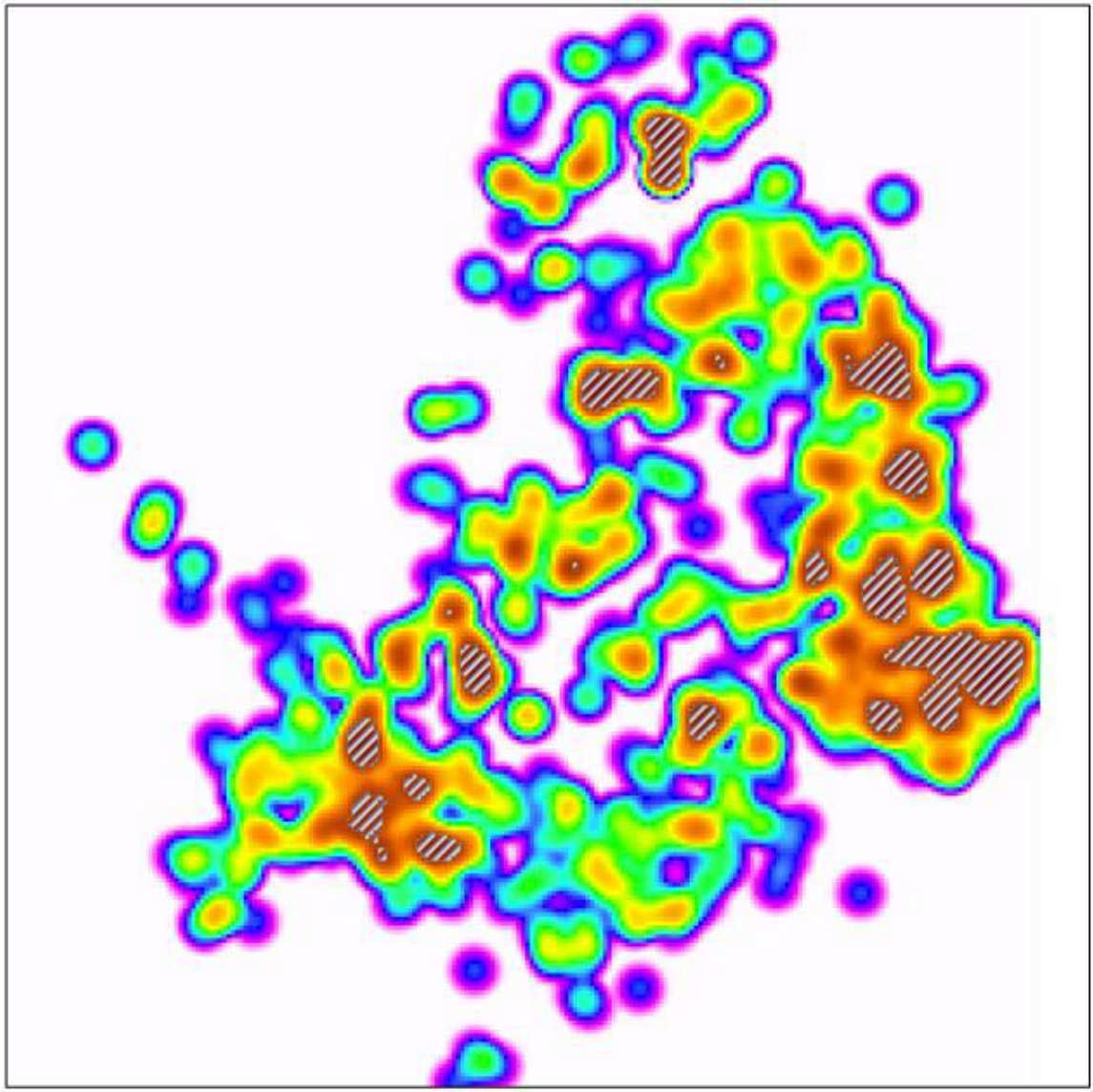
- Finální výsledky zobrazující statistické výsledky na hladině významnosti 95 % (vlevo) a 99 % (vpravo).



- Stačí to?? Kde je problém?

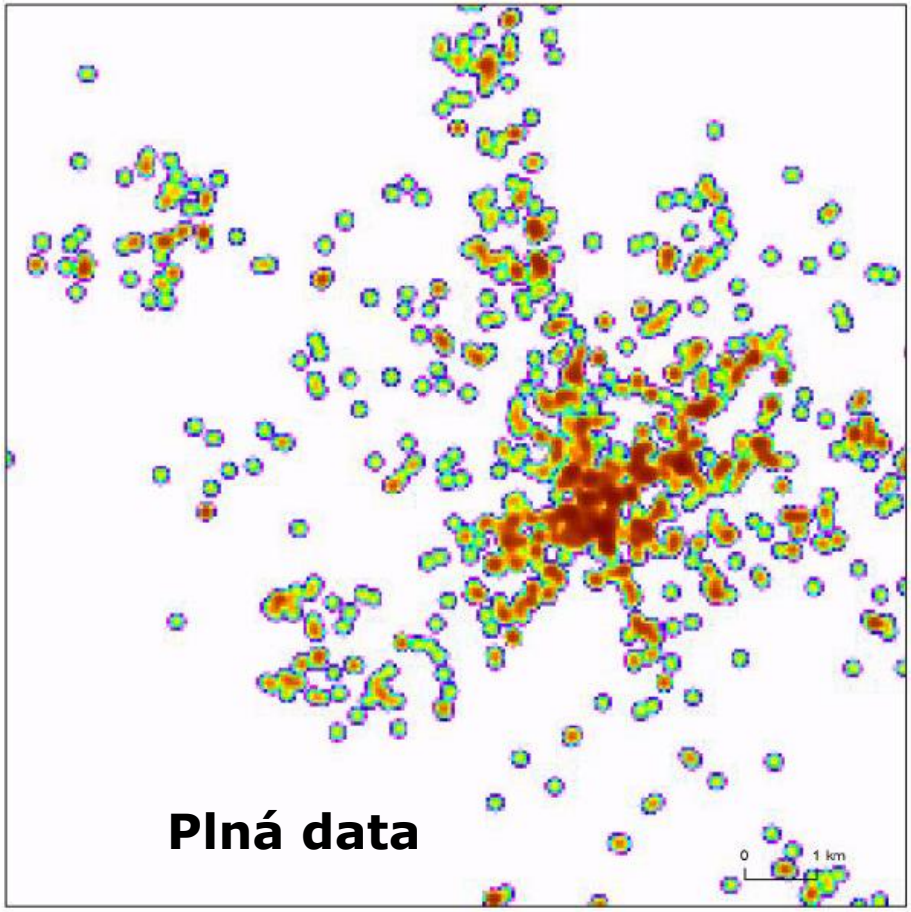
Statistická významnost

- **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 Gi*.

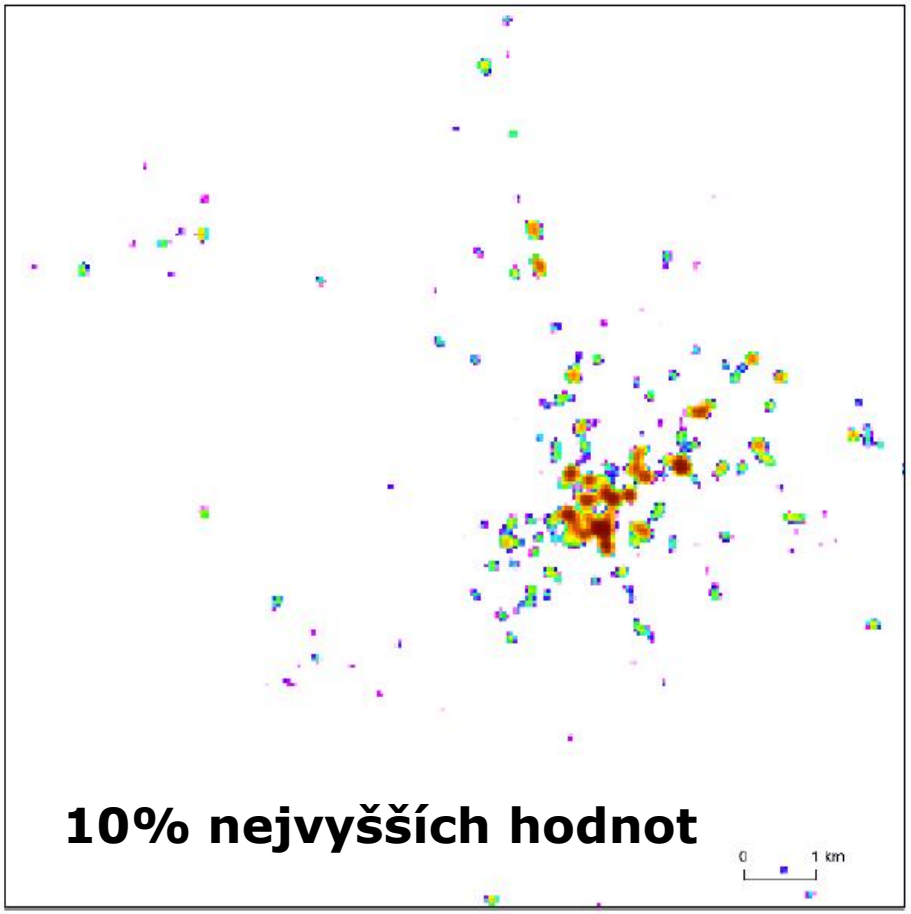


POSTPROCESSING A VIZUALIZACE

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



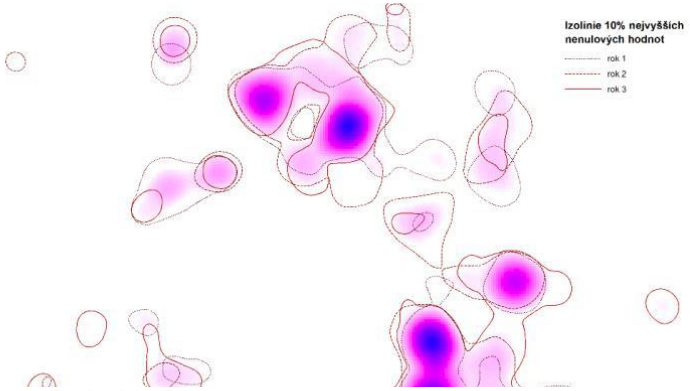
Plná data

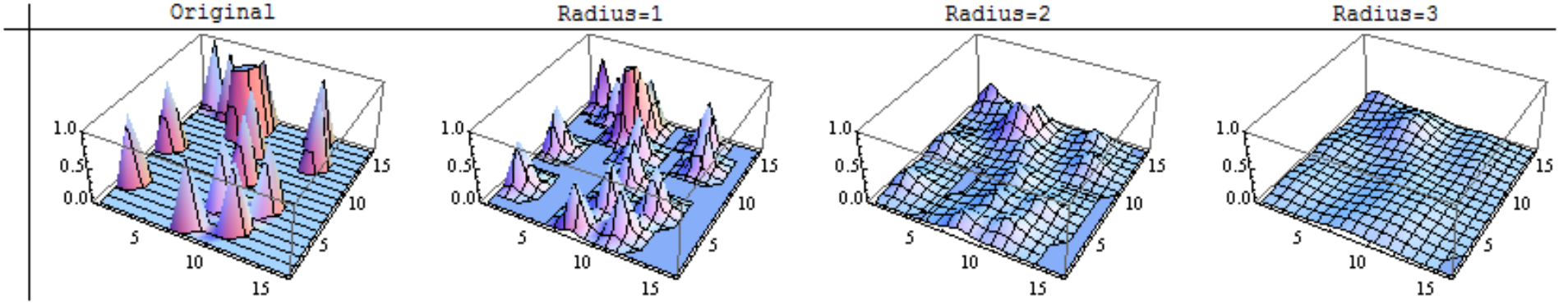


10% nejvyšších hodnot

Vizualizace (alternativní)

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





Variations in time

- Each hot spot map considered in this lecture accounts only for a **specific snapshot period** in time.
- New areas of research are beginning to **explore space-time interaction** .
- These methods aim to reveal whether certain types of crime display temporal hot spots in particular areas (e.g., crime **hot spots that emerge only on certain days of the week**).
- The creation of **crime hot spot animations** to visualize space and time interaction.

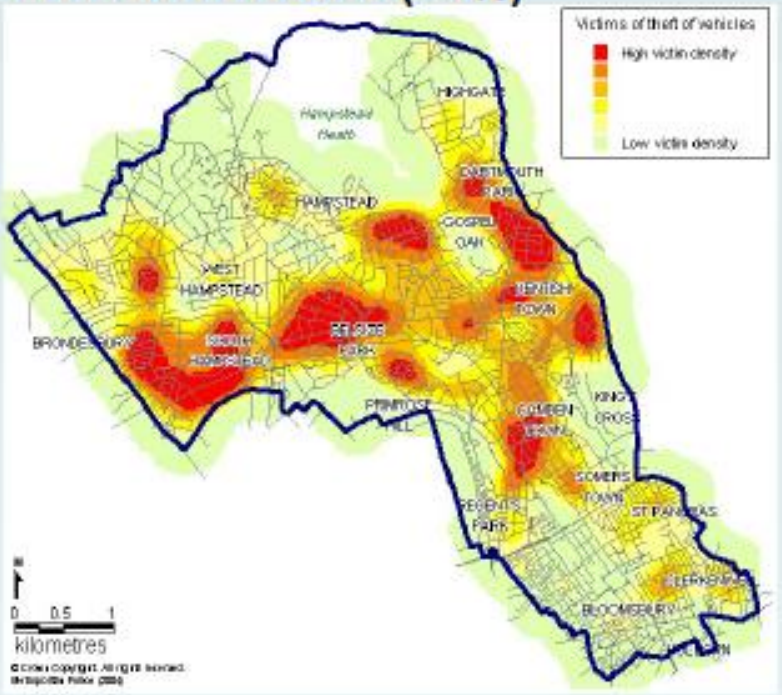
Crime analysis - example

Analysing vehicle crime in central London:

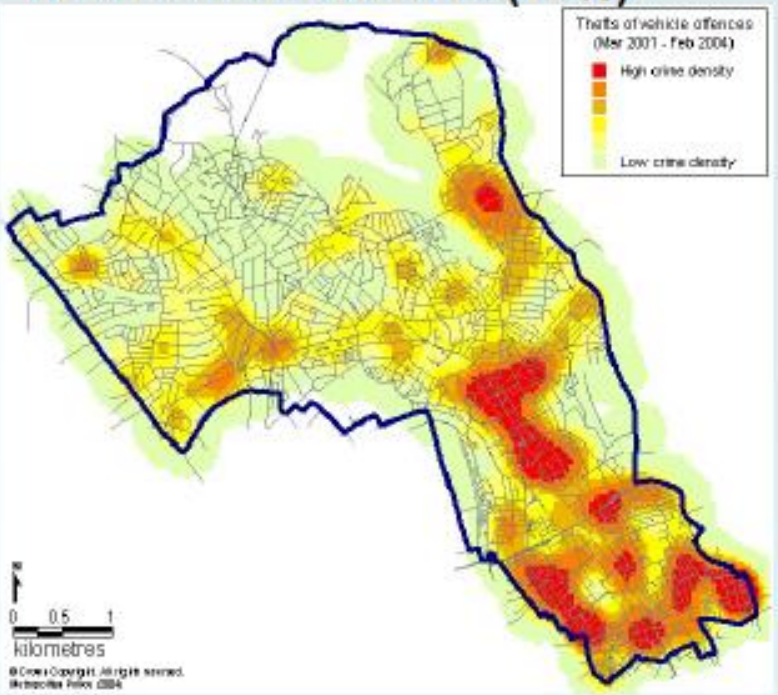
- Hypothesis: "We think it relates mainly to **local residents** having their **cars** stolen **at night**" (The Police)
- Crime analysis involves ***breaking the problem apart*** and exploring the specifics of the problem
- We have a series of questions that we can turn into hypotheses
- Explore 'place' across these
- Helping to explain the problem

Locals vs visitors

Camden victims (58%)



Non-Camden victims (42%)



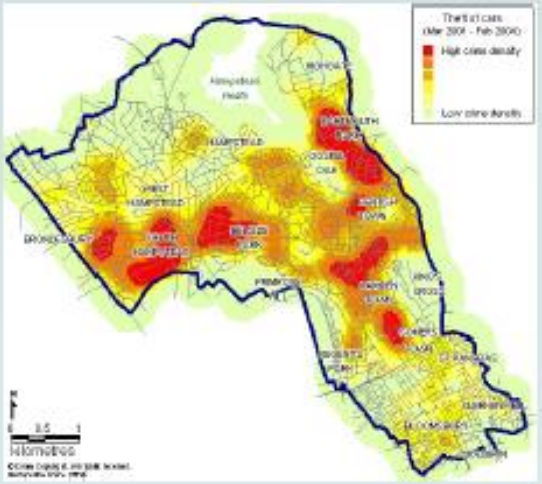
Vehicle statistics

of Crime Sc

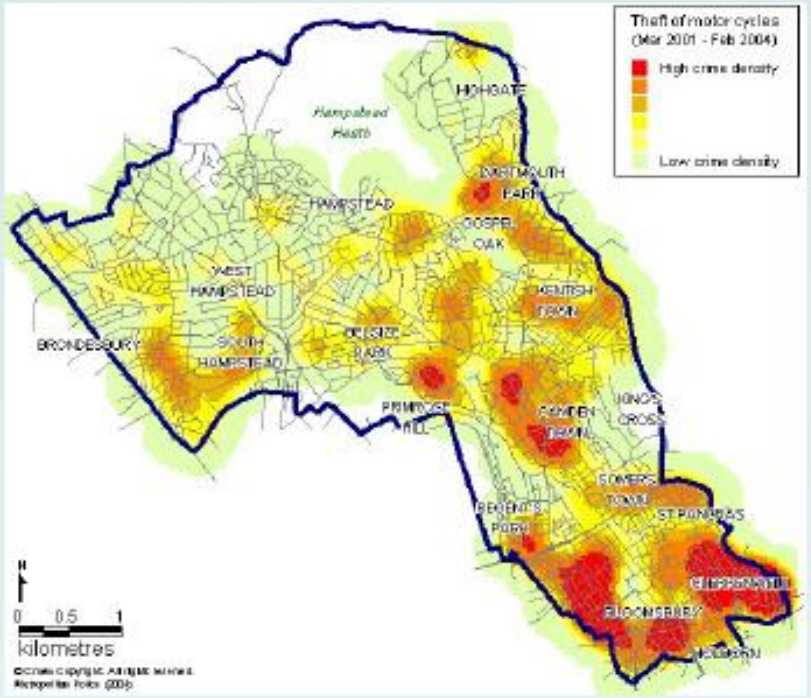
Type of vehicles stolen

Vehicle type description	Offences	%
Hatchback	1258	21.7%
Saloon	1433	24.7%
Estate	220	3.8%
People carrier	45	0.8%
Convertible	120	2.1%
Sports	42	0.7%
4 X 4's	4	0.1%
Moped or scooter	1494	25.8%
Motor cycle	755	13.0%
Motor caravan	11	0.2%
Van	274	4.7%
Other	50	0.9%
Not known	23	0.4%

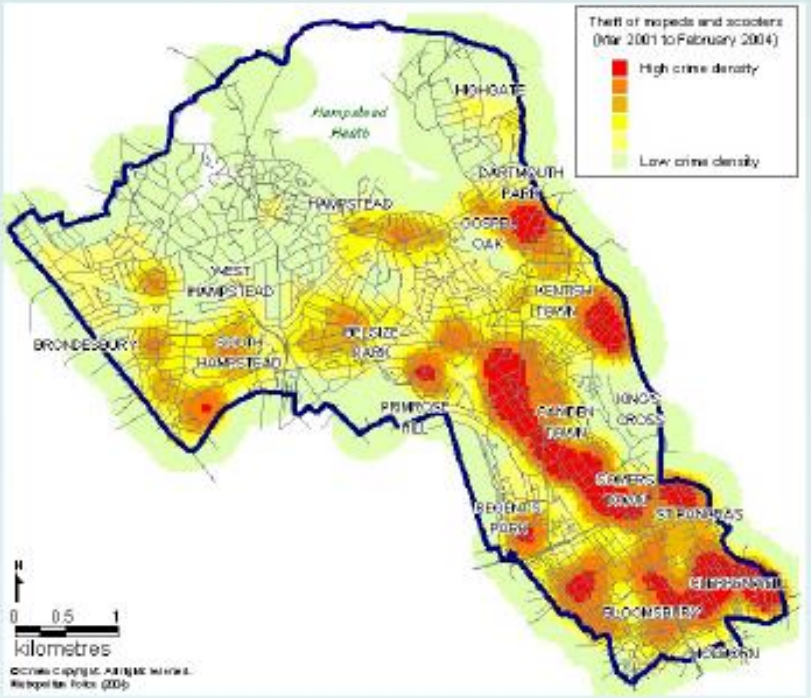
Cars



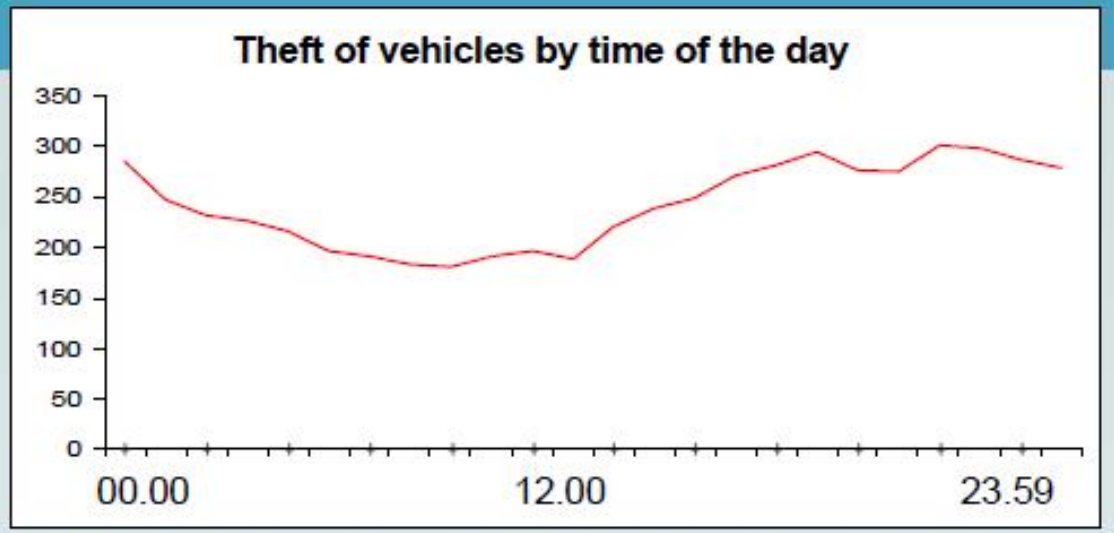
Motorbikes

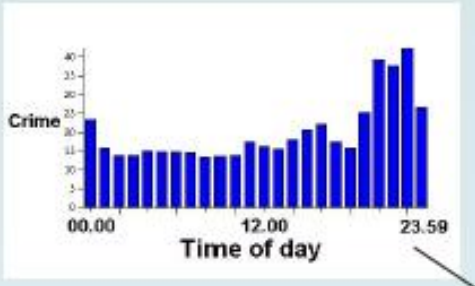


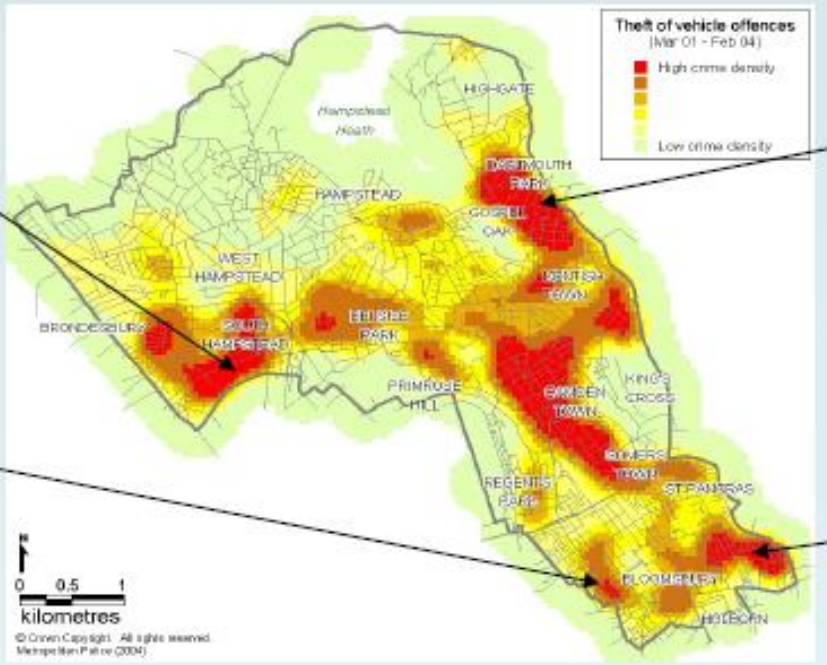
Scooters and mopeds

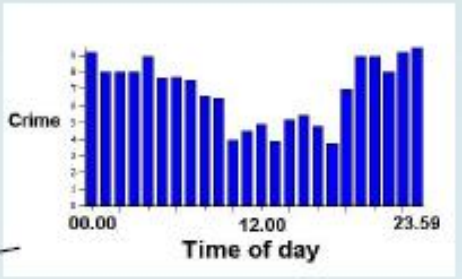


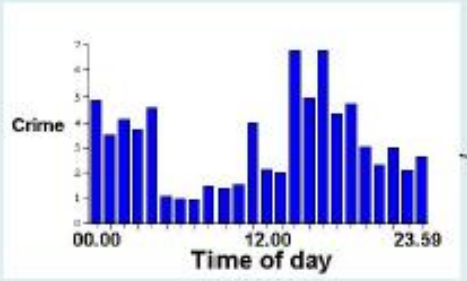
Place: space and time

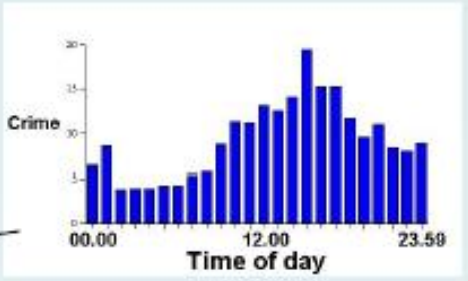






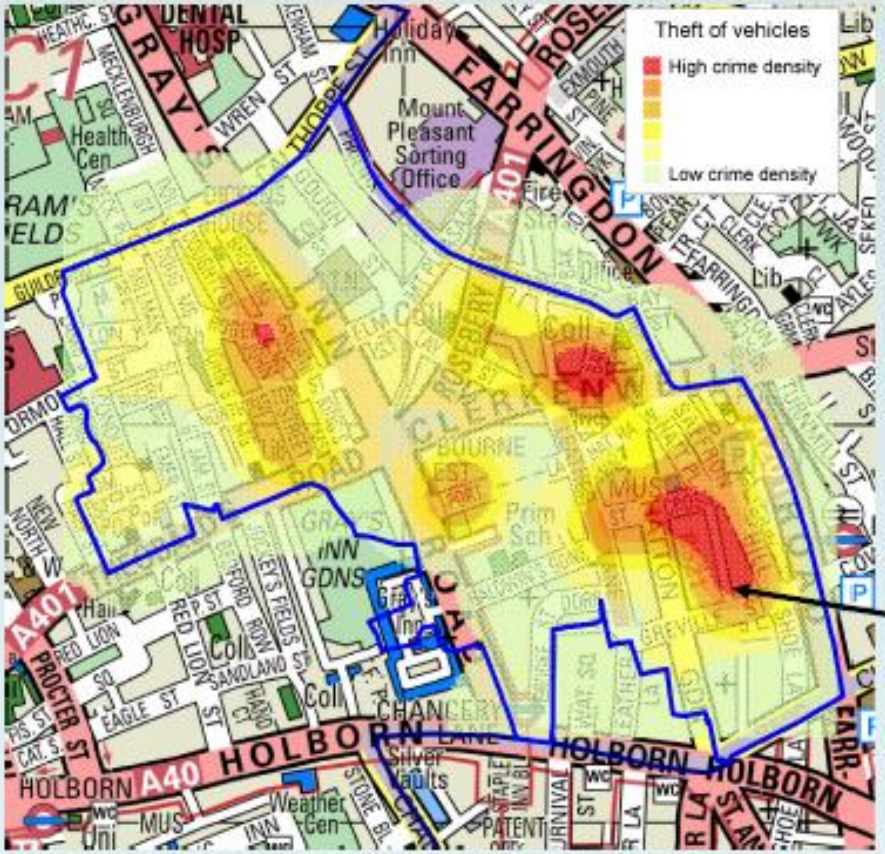






Detail view

Vehicle type	Camden	Clerkenwell (n)	Clerkenwell(%)
Car	51%	41	18%
Sports or convertible	3%	5	2%
Scooter or moped	26%	95	42%
Motor cycle	13%	70	31%
Van	5%	3	1%
Other	2.0%	10	4%
Not known	0.5%	0	0%



So it's not all to do with residents having their cars stolen at night …