點型態分析：密度分析法
Point Pattern Analysis: Density-based Methods

https://ceiba.ntu.edu.tw/1092Geog2017_

授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw

Spatial Point Clustering

- Spatial point clustering
- *Distance-based* methods (summary functions)
  - G(d), and F(d) Functions
  - Ripley's K(d) function
- *Density-based* methods
  - Smoothing: Kernel density estimation (KDE)
    - Kernel density estimation
    - Optimal bandwidth for kernel smoothing

課程大綱

heat map; density map

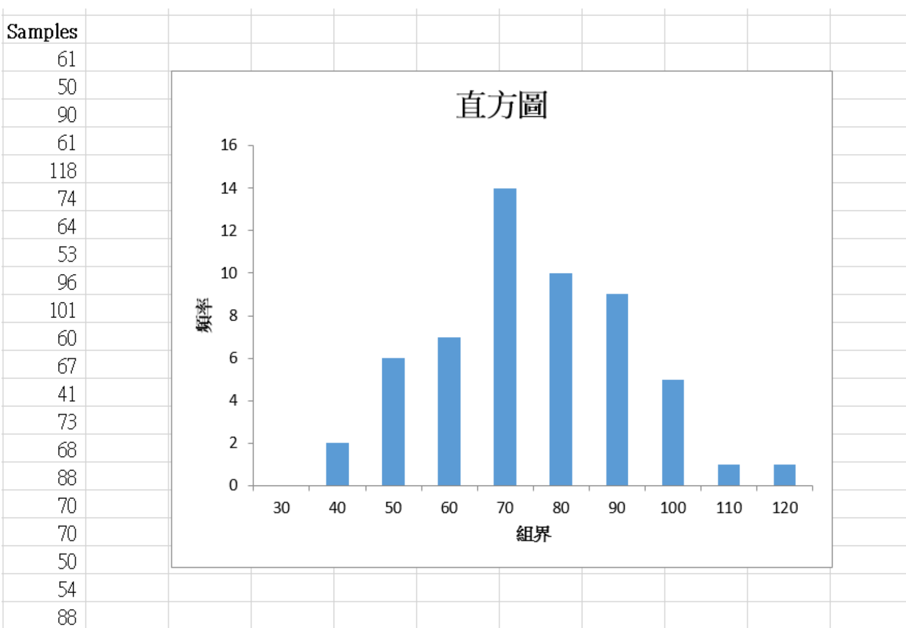
- *Density-based* methods
  - Kernel density estimation (KDE)
    - Kernel density estimation
    - Optimal bandwidth for kernel smoothing
  - Dual KDE
    - Considering underlying population
    - Detecting change over time

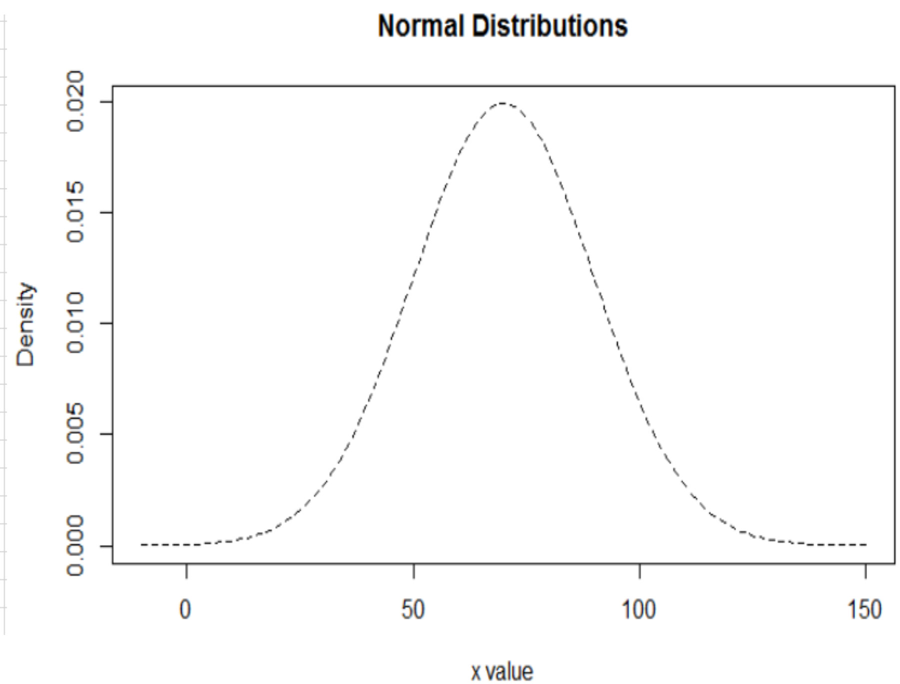
Probability Density Function (PDF)

Samples, n = 55

Population: norm(70, 20)

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi}\hat{\sigma}} e^{(x-\hat{\mu})/2\hat{\sigma}^2}, \quad x \in I\!R$$





Kernel Density Estimation (KDE)

STATISTICS IN MEDICINE, VOL. 14, 2335–2342 (1995)

NON-PARAMETRIC ESTIMATION OF SPATIAL VARIATION IN RELATIVE RISK

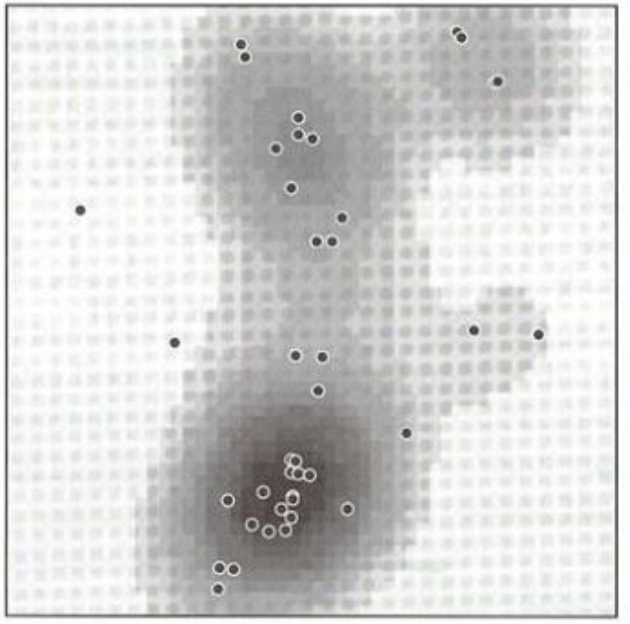
JULIA E. KELSALL AND PETER J. DIGGLE

Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, U.K.

SUMMARY

We consider the problem of estimating the spatial variation in relative risks of two diseases, say, over a geographical region. Using an underlying Poisson point process model, we approach the problem as one of density ratio estimation implemented with a non-parametric kernel smoothing method. In order to assess the significance of any local peaks or troughs in the estimated risk surface, we introduce pointwise tolerance contours which can enhance a greyscale image plot of the estimate. We also propose a Monte Carlo test of the null hypothesis of constant risk over the whole region, to avoid possible over-interpretation of the estimated risk surface. We illustrate the capabilities of the methodology with two epidemiological examples.

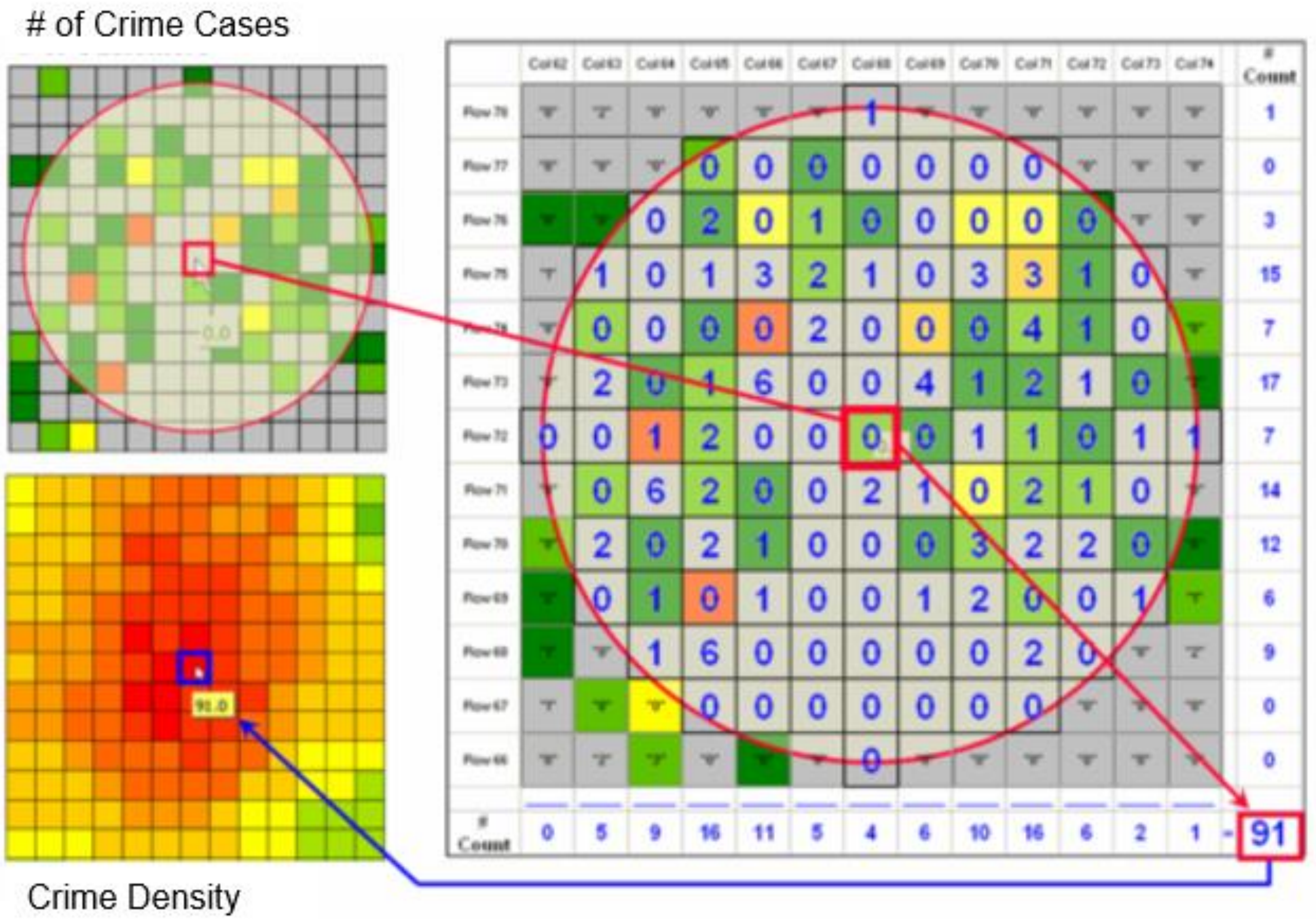
Density method



Intensity of Point P

$$\hat{\lambda}_p = \frac{no.[S \in C(p,r)]}{\pi r^2}$$

KDE Calculation Procedures



Source: Joseph K. Berry, W.M. Keck (2009)

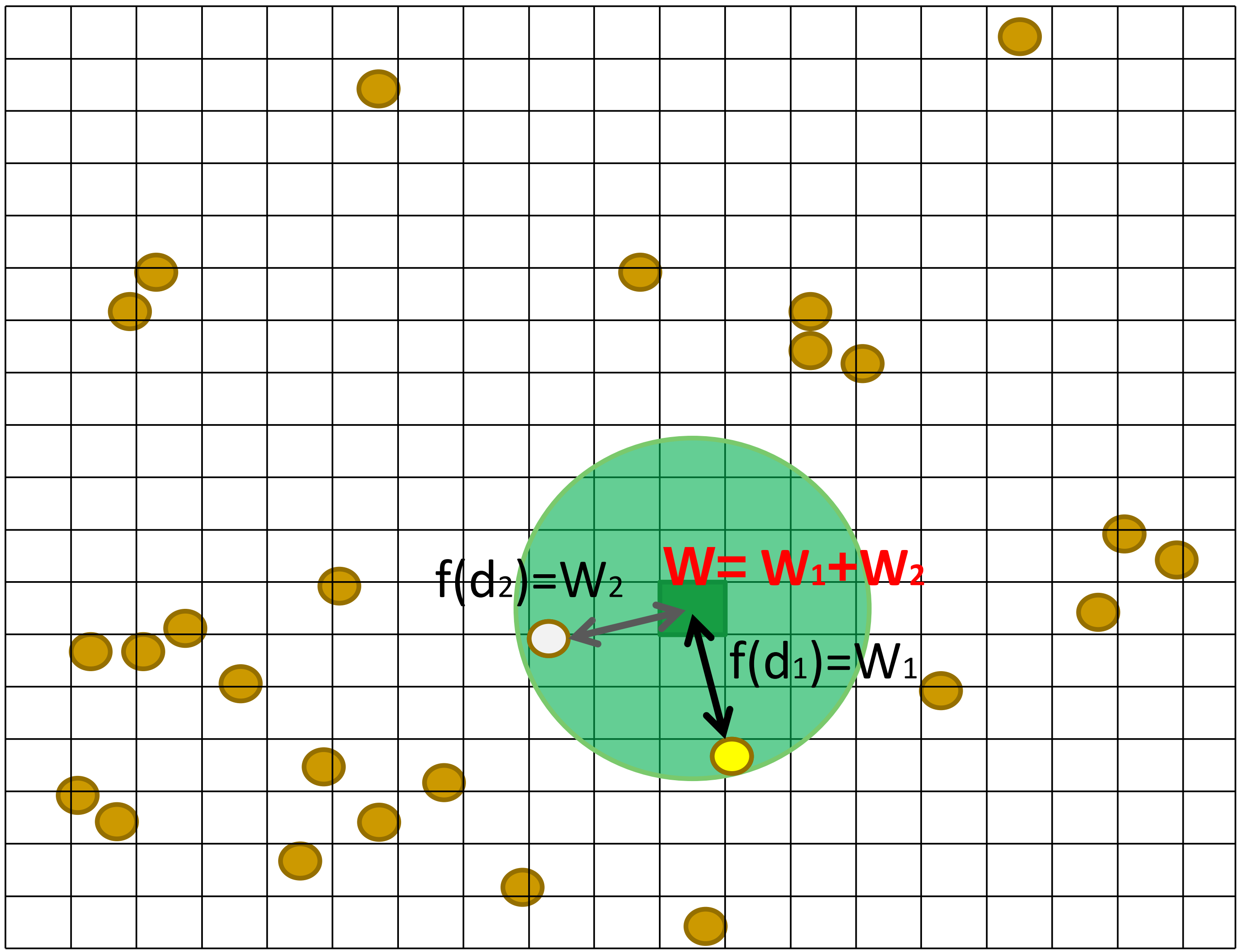
Step 1: 研究區域建立均勻網格 (cell / grid)

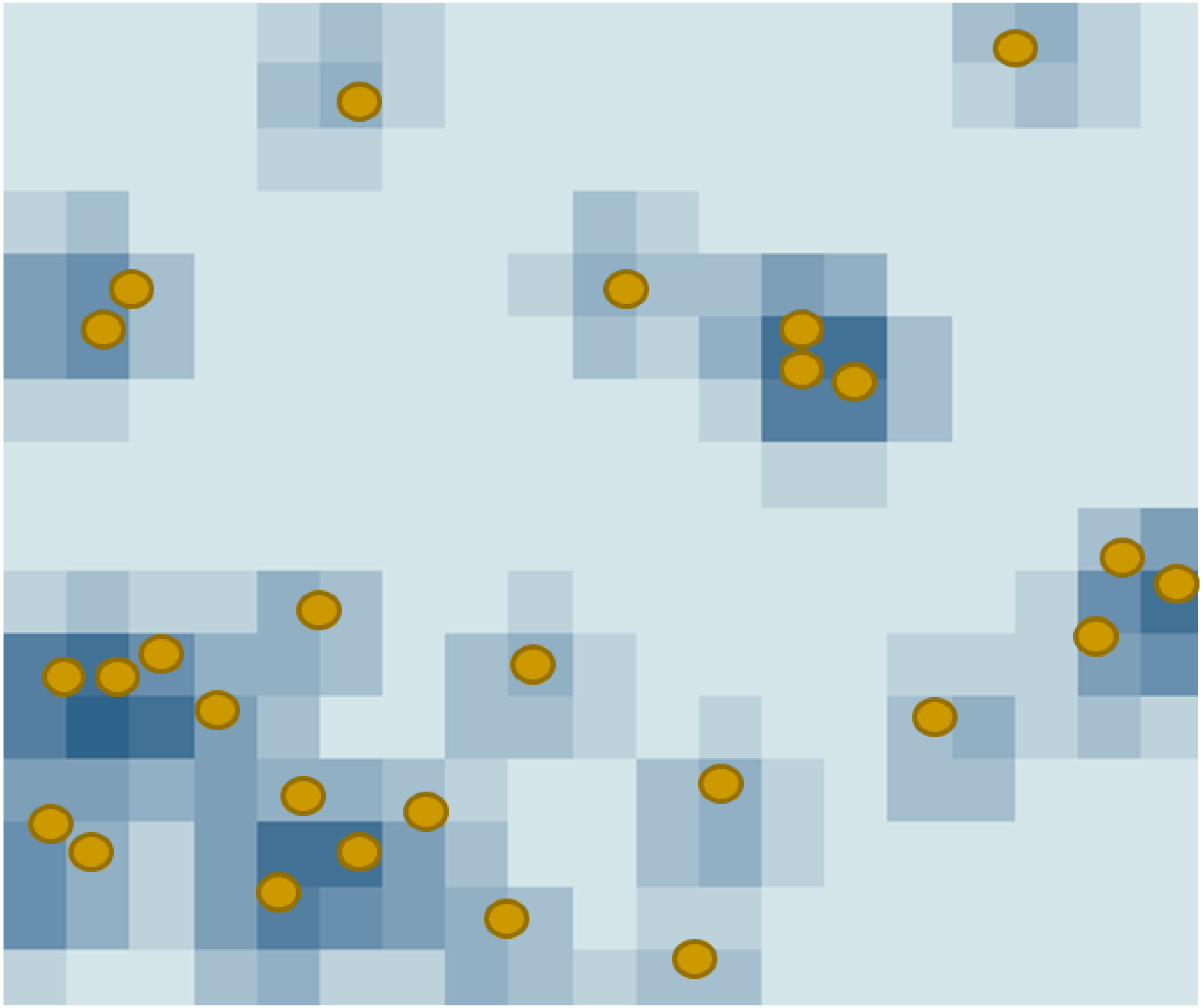
Step 2: 設定搜尋半徑 (bandwidth, h)

Step 3: 選擇核密度函數 (Kernel function)

density estimate at (x,y)

$$\hat{f}(x, y) = \frac{1}{nh^2}\sum_{i=1}^{n} K(d_i/h)$$





Kernel Density Estimation

$$\hat{f}(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{d_{i,(x,y)}}{h}\right)$$

where $\hat{f}(x, y)$ is the estimated density value at location (x, y), n is the total number of event points under concern (e.g., disease cases), h is a measure of the window width and is called kernel bandwidth (e.g., for a circular kernel it is the radius of the circle), $d_{i,(x,y)}$ is the distance between event point i and location (x, y), and K is a density function characterizing how the contribution of point i varies as a function of $d_{i,(x,y)}$.

Kernel Function: Normal distribution

Probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$\mu =$ mean of x
$\sigma =$ standard deviation of x
$\pi \approx 3.14159 \ldots$
$e \approx 2.71828 \ldots$

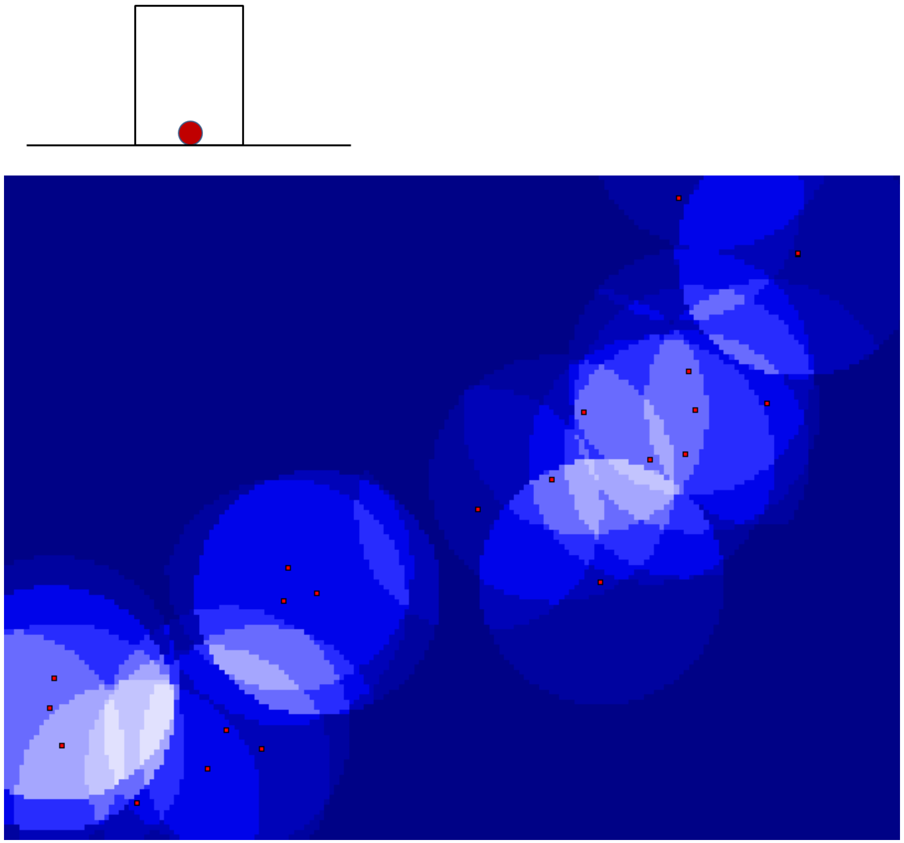
$g(x_j)$ is the density of cell *j*

$$g(x_j) = \sum_{i=1}^{N} \left[\boxed{KW_i I_i} \frac{1}{h^2 2\pi} e^{-\frac{d_{ij}^2}{2h^2}} \right]$$

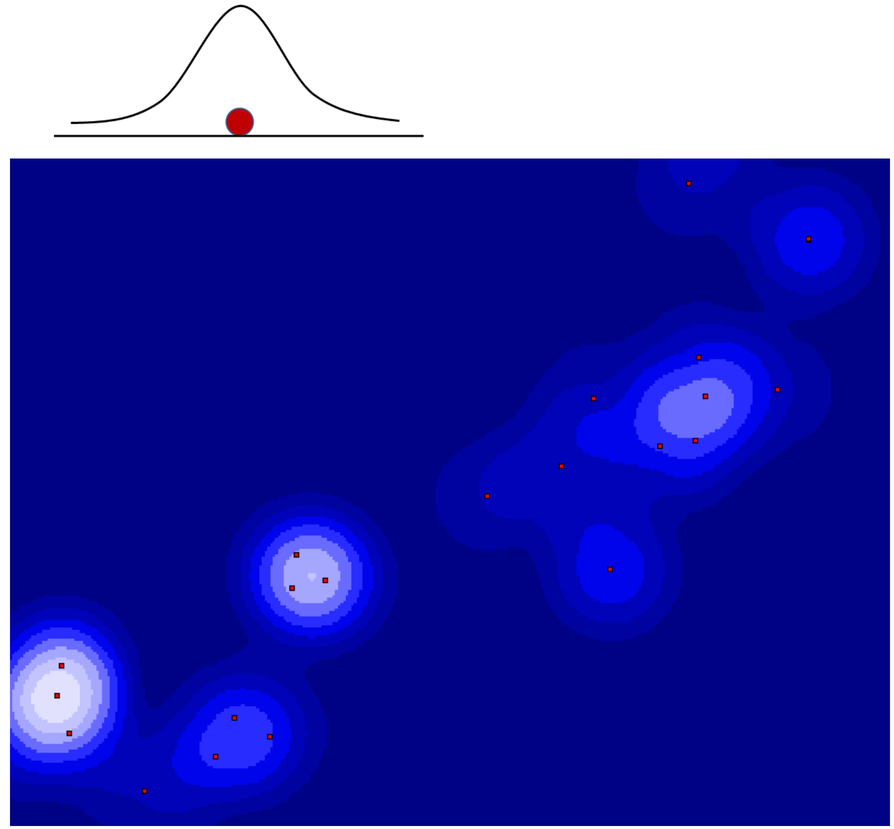
KW: weight

I: indicator function

Different Kernel Functions

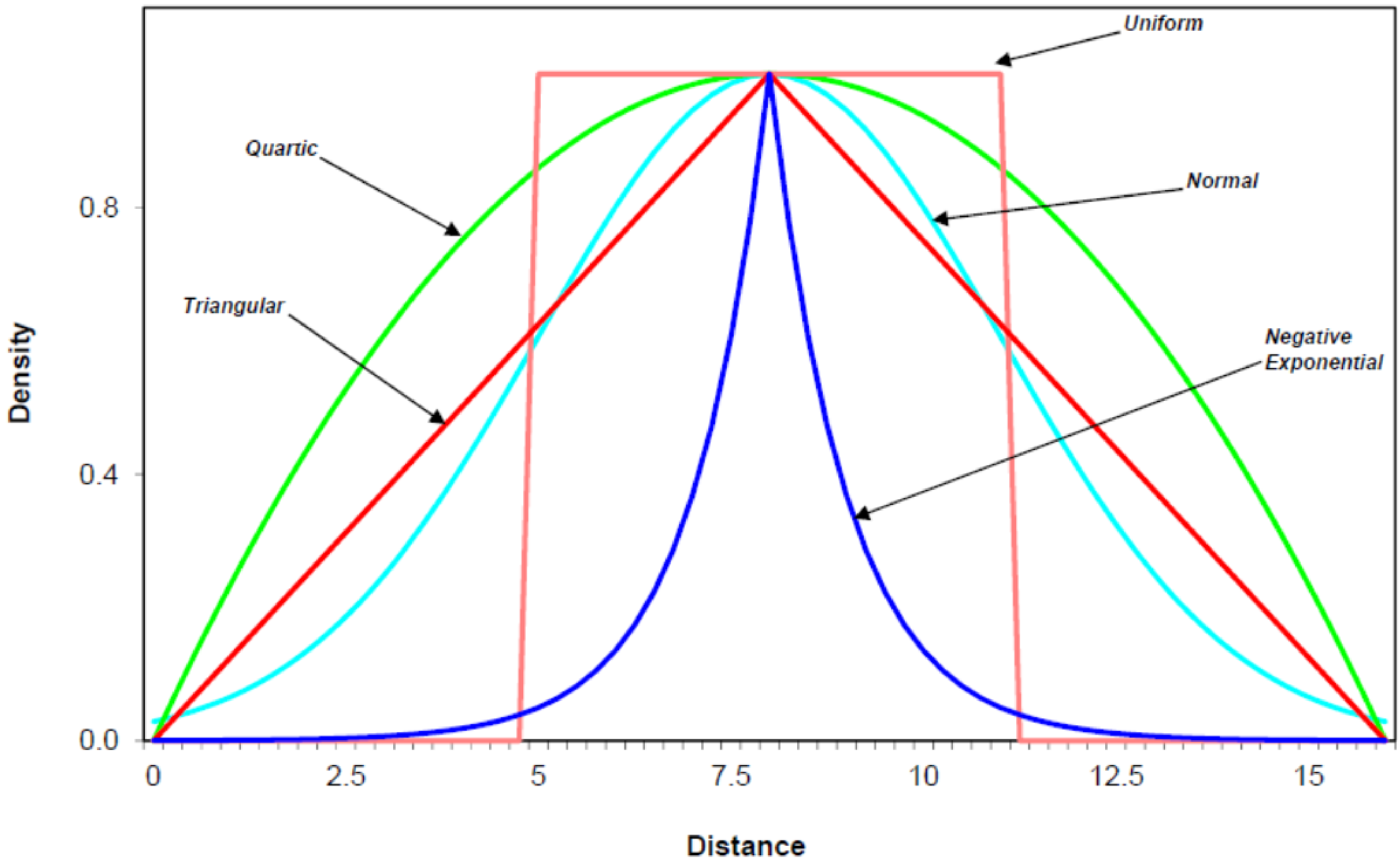


Uniform Distribution



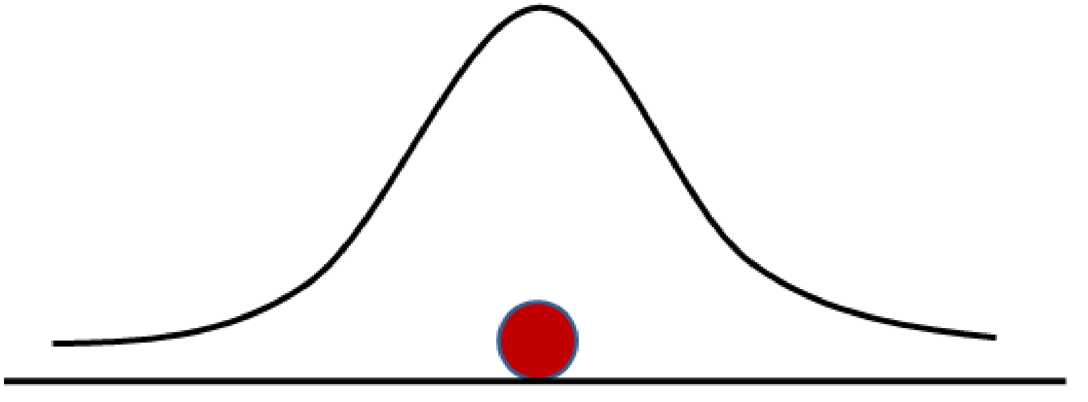
Normal Distribution

Choosing Appropriate Kernel Functions



KDE Parameters

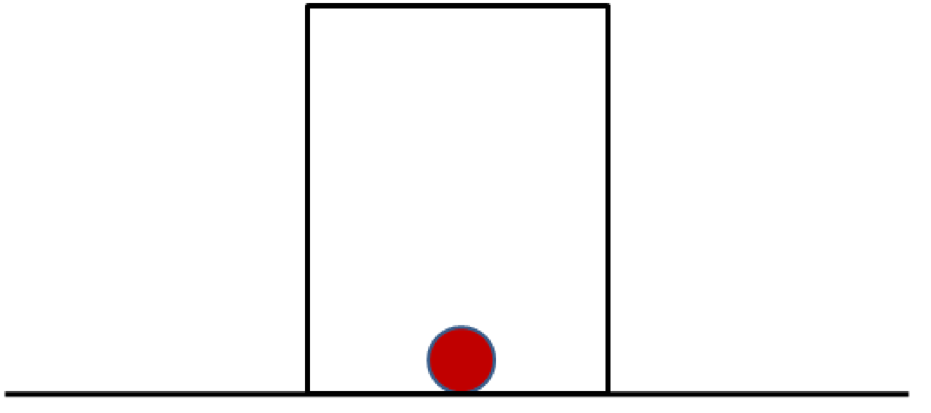
- Choosing appropriate kernel functions
  - Normal (bell curve)
    - peaks & declines rapidly
    - No defined radius; continues across entire grid



Source: *CrimeStat III Workbook*

KDE Parameters

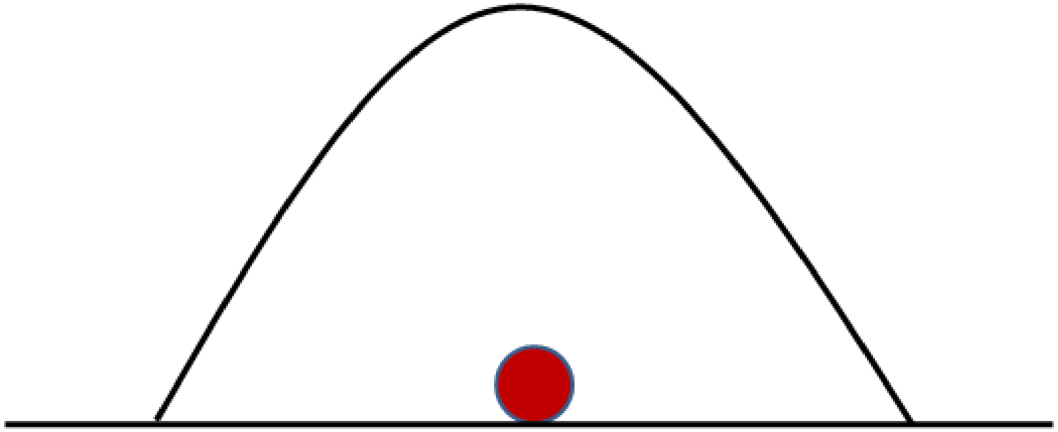
- Choosing appropriate kernel functions
  - Uniform (flat) distribution
    - Represented by cylinder; all points in radius equal



Source: *CrimeStat III Workbook*

KDE Parameters

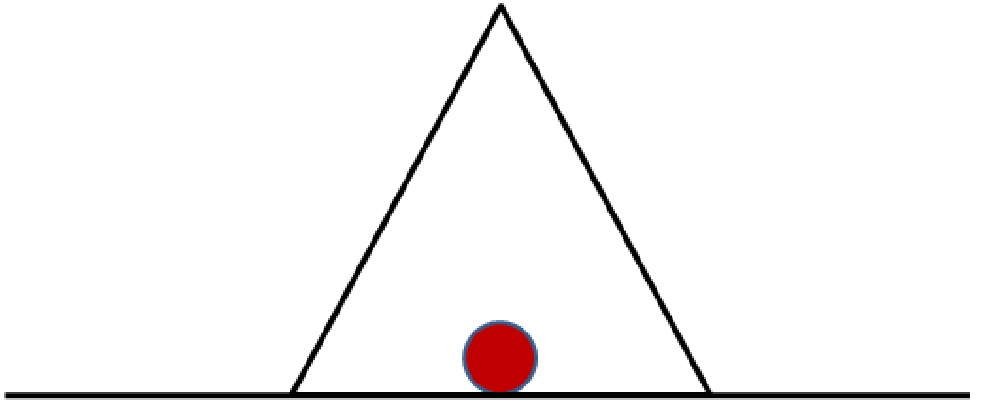
- Choosing appropriate kernel functions
  - Quartic (spherical) distribution
    - Gradual curve; density highest over point; falls to limit of radius



Source: *CrimeStat III Workbook*

KDE Parameters

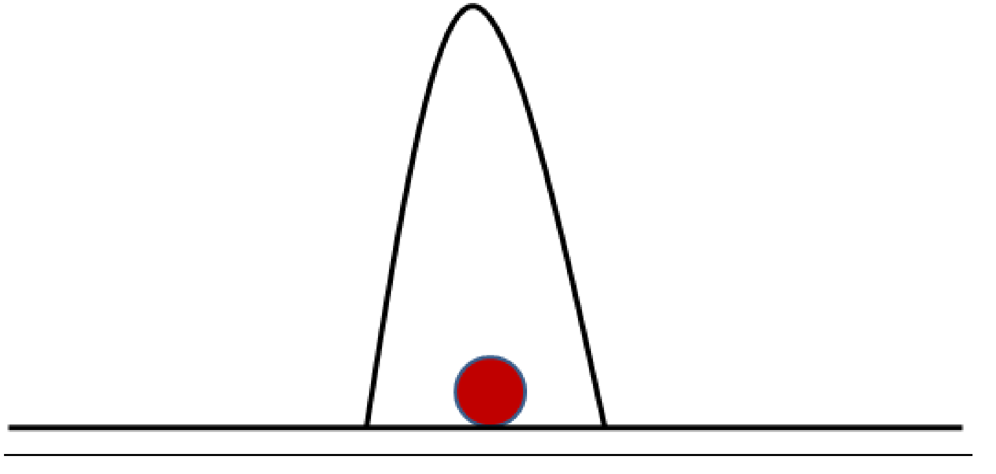
- Choosing appropriate kernel functions
  - Triangular (conical) distribution
    - Peaks above the point; falls off in a linear manner to edges of radius



Source: *CrimeStat III Workbook*

KDE Parameters

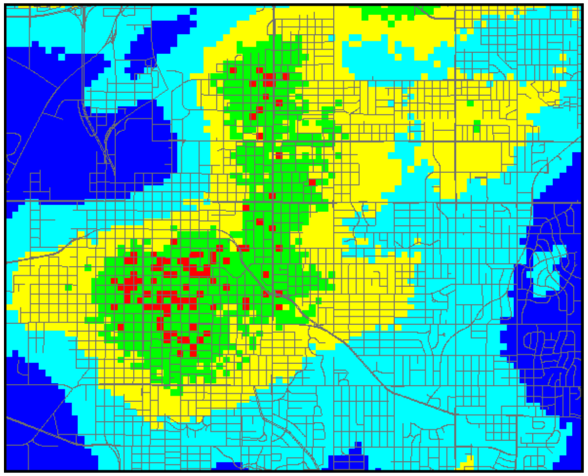
- Choosing appropriate kernel functions
  - Negative exponential distribution
    - Curve that falls off rapidly from the peak to a specified radius



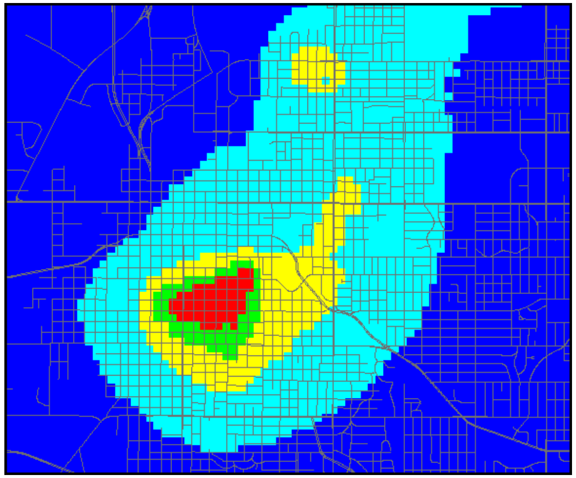
Source: *CrimeStat III Workbook*

KDE Parameters

- Each method will produce different results
  - Triangular & negative exponential produce many small hot and cold spots
  - Quartile, uniform and normal distribution functions smooth data more



Negative exponential



Normal Distribution

Source: *CrimeStat III Workbook*

Incident Type	Interval	Interpolation Method	Reasoning
Residential burglaries (住宅竊盜)	1 mile	Moderately dispersed: **quartic or uniform**	Some burglars choose particular houses, but most cruise neighborhoods looking for likely targets. A housebreak in any part of a neighborhood transfers risk to the rest of the neighborhood.
Domestic violence (家庭暴力)	0.1 mile	Tightly focused: **negative exponential**	Domestic violence occurs among specific individuals and families. Incidents at one location do not have much chance of being contagious in the surrounding area.
Commercial robberies (商業搶劫)	2 miles	Focused: **triangular or negative exponential**	A commercial robber probably chooses to strike in a commercial area, and then looks for preferred targets (banks, convenience stores) within that area. The wide area may thus be at some risk, but the brunt of the weight should remain with the particular target that has already been struck.
Thefts from vehicles (汽機車偷竊)	0.25 mile	Dispersed: **uniform**	If a parking lot experiences a lot of thefts from vehicles, your GIS will probably geocode them at the center of the parcel. This method ensures that the risk disperses evenly across the parcel and part of the surrounding area (which probably makes sense)—but not too far, since we know that parking lots tend to be hot spots for specific reasons.

Source: *CrimeStat III Workbook*

Choosing Optimal Bandwidth: Fixed vs. Adaptive Intervals

- Depending on the type of kernel estimate used, this interval has a slightly different meaning.
- For the **normal kernel function**, the bandwidth is the standard deviation of the normal distribution.
- For the **uniform, quartic, triangular, or negative exponential kernels**, the bandwidth is the radius of the search area

Choosing Optimal Bandwidth:
1. Suggested settings

Bowman and Azzalini (1997) and Scott 1992

$$h_x = \sigma_x \left(\frac{2}{3n} \right)^{\frac{1}{6}} \tag{6.2}$$

where σ_x is the standard deviation of the x_i. A similar formula exists for h_y, replacing σ_x with σ_y the standard deviation of the y_i The central KDE in Figure 6.2 is based on choosing h_x and h_y using this method.

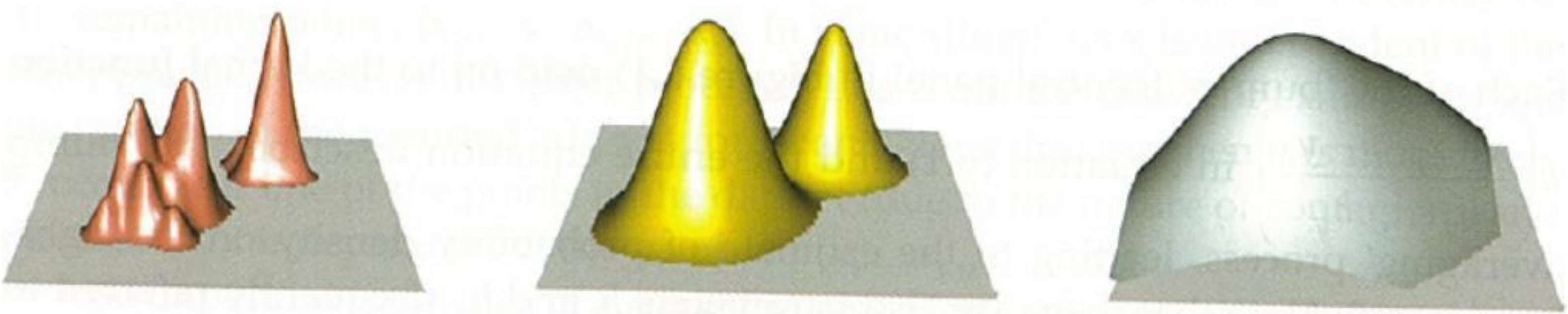


Figure 6.2 Kernel density estimation bandwidths: h_x and h_y too low (left); h_x and h_y appropriate (centre); h_x and h_y too high (right)

Choosing Optimal Bandwidth
2. Estimating range of influence:
K-order nearest neighbor analysis

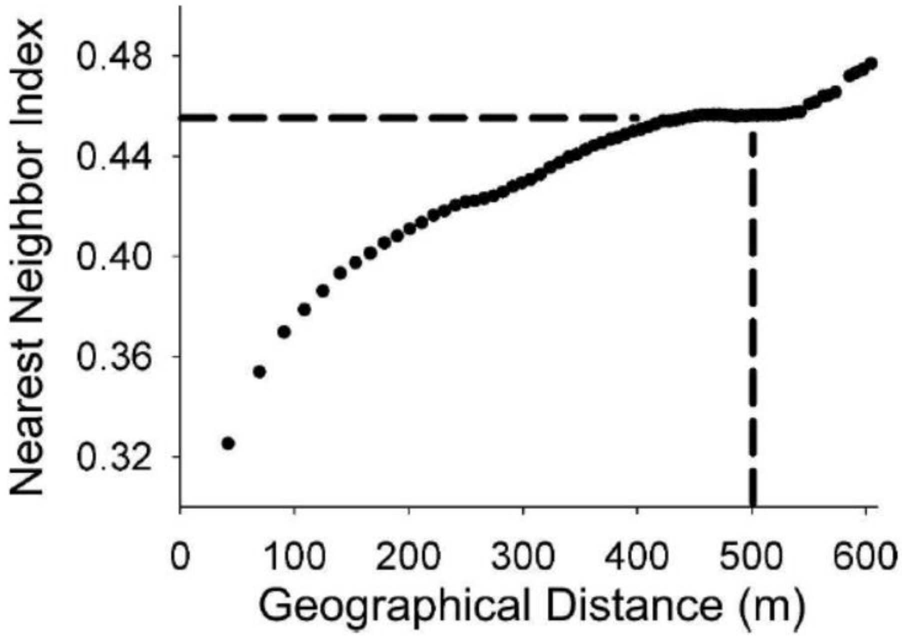


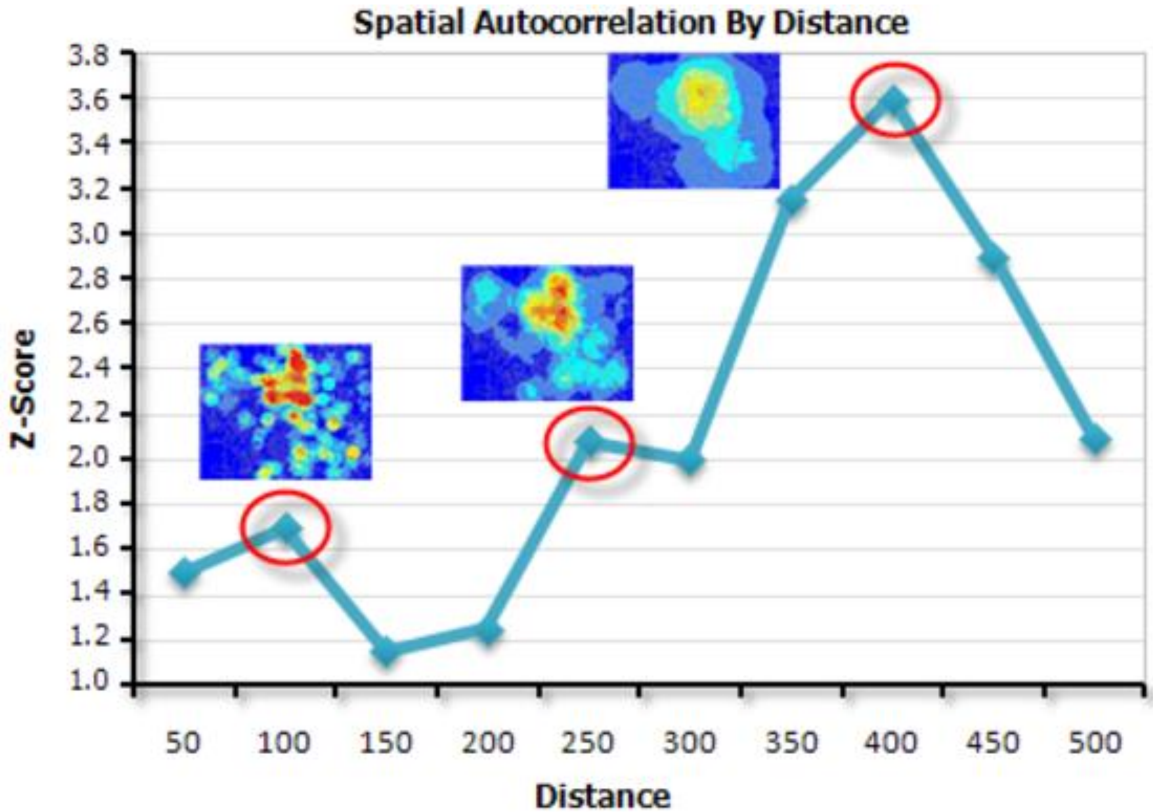
FIGURE 2. The K-order nearest neighbor analysis showed the relationship between the nearest neighbor index and observed distance (m) of the Kth nearest neighbor. A 500-m bandwidth was selected from a plateau (450–550 m).

Choosing Optimal Bandwidth
2. Estimating range of influence: Incremental spatial autocorrelation

Measures spatial autocorrelation for a series of distances and optionally creates a line graph of those distances and their corresponding z-scores. Z-scores reflect the intensity of spatial clustering, and statistically significant peak z-scores indicate distances where **spatial processes promoting clustering are most pronounced**. These peak distances are often appropriate values to use for tools with a Distance Band or Distance Radius parameter.

http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/incremental-spatial-autocorrelation.htm

Incremental spatial autocorrelation



Z-score peaks reflect distances where the spatial processes promoting clustering are most pronounced.

http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/incremental-spatial-autocorrelation.htm

Choosing Optimal Bandwidth
2. Estimating range of influence: User-defined

Previous studies of **the moving distance of humans and vectors** (Harrington et al., 2005; Maciel-De-Freitas, Codeco, & Lourenco-de-Oliveira, 2007; Wen et al., 2012) were used to establish dengue risk maps, with a searching radius of 1500 m and vector risk maps with a searching radius of 300 m.

Choosing Optimal Bandwidth: Adaptive Intervals

- Density estimation and adaptive bandwidths: A primer for public health practitioners, *International Journal of Health Geographics* 2010, 9:39.
- Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds, *International Journal of Geographical Information Science* 2010, 24:5.

補充參考

核密度圖 (kernel density map)

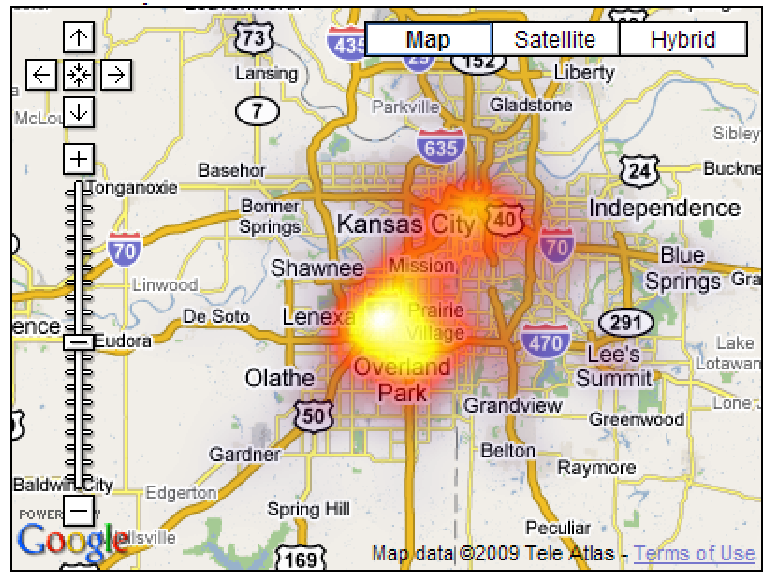
HeatMapAPI
Create Your Own Heat Maps

Create your own heat maps using HeatMapAPI. Integrate heat map images into Google Maps or other GIS systems. Heat maps are rendered real-time.

Home FREE API Key Uses Features Sample Code Support

Welcome to HeatMapAPI.com

What is a Heat Map?
A heat map is a graphical representation of two dimensional data (X, Y and Value) on a two dimensional surface using colors. A map a heat map is a representation of density of latitude and longitude points on a map. HeatmapAPI lets you creates heat map images you can overlay on your maps in your software or website.



Why Heat Maps?
Heat maps help your website visitors or users quickly visualize density. Most maps show a sea of points or themed polygons, however being able to visualize the density of those points makes it much easier to understand, especially when using colors as shown using a heat map.

Create Your Own Heat Maps
We give you an API that enables you to create your own heat maps on your maps, in your application. We offer a limited but free API you can use, or you can upgrade for unlimited data points or optionally run the API on your server.

Why HeatMapAPI?
Your heat map is rendered real-time. Its a single image that can be overlayed onto just about any GIS system that can overlay an image. Because the images are rendered real-time, you can change the data on the fly, and this makes it very powerful.

http://www.heatmapapi.com/Default.aspx

小結

- 需要針對不同的資料型態，選擇合適的方法。
- 運用各種統計分析方法亦需瞭解其各種方法的基本假設、分析模式的侷限性以及分析結果的正確解讀。
- 透過各種資料型態與不同統計方法的分析，在各種方法間之相互比較，若都有共同的分析結果，較能合理的反映出該地區確實具有地理群聚現象。

Dual Kernel: Measuring Change over Time

Using Dual Kernel Density Estimation To Examine Changes in Voucher Density Over Time

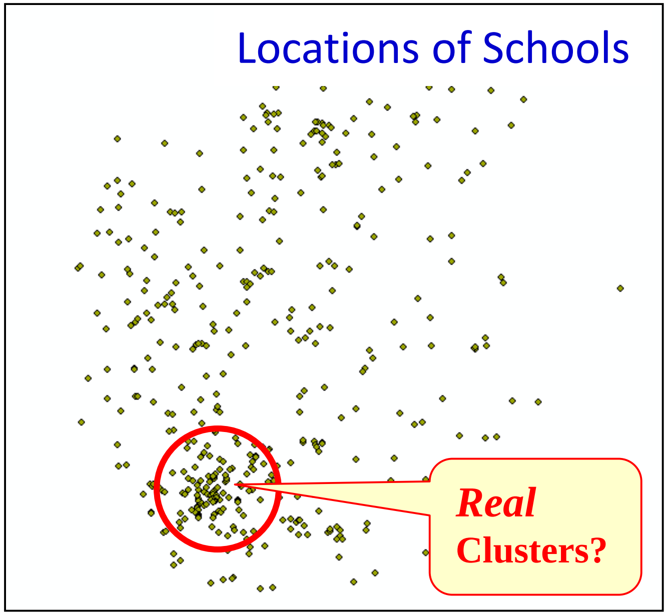
Ron Wilson
U.S. Department of Housing and Urban Development

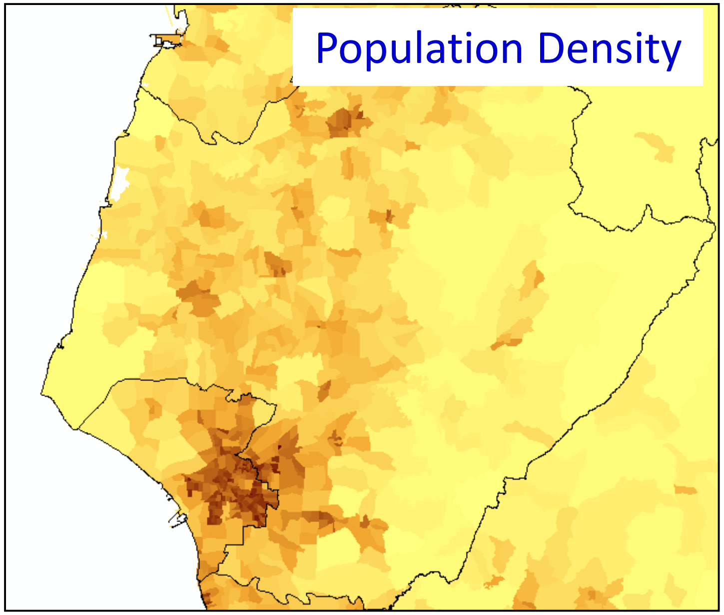
The views expressed in this article are those of the author and do not represent the official positions or policies of the Office of Policy Development and Research or the U.S. Department of Housing and Urban Development.

Abstract

The measurement of participants in the Housing Choice Voucher Program across time is an important analytical step toward understanding their settlement patterns, particularly whether they concentrate or deconcentrate. Many analyses of voucher-holder settlement patterns employ some areal unit in which counts are divided by unit area to calculate a density. This approach has methodological problems and produces less-than-accurate results because it does not directly measure the locations of voucher holders. In this article, I show how to apply a technique, known as Dual Kernel Density Estimation, to measure directly the concentration of voucher-holder locations to produce more accurate results about where voucher holders have concentrated and deconcentrated over time.

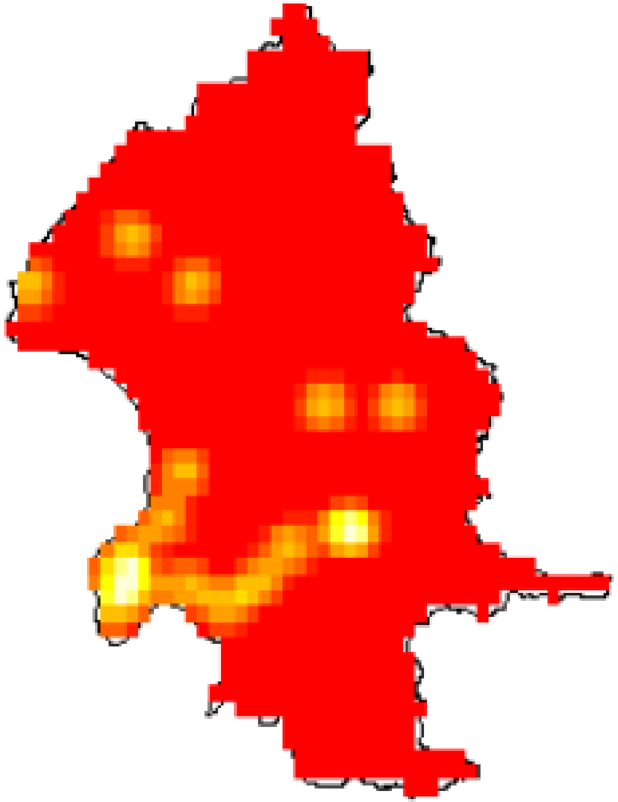
Dual Kernel: Underlying Population-adjusted Kernel Densities



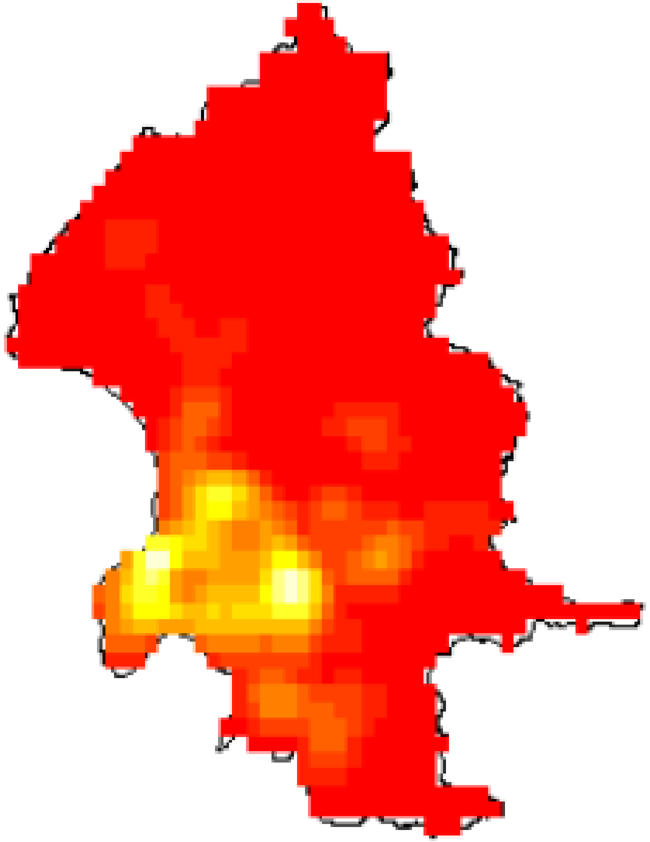


Comparing two densities

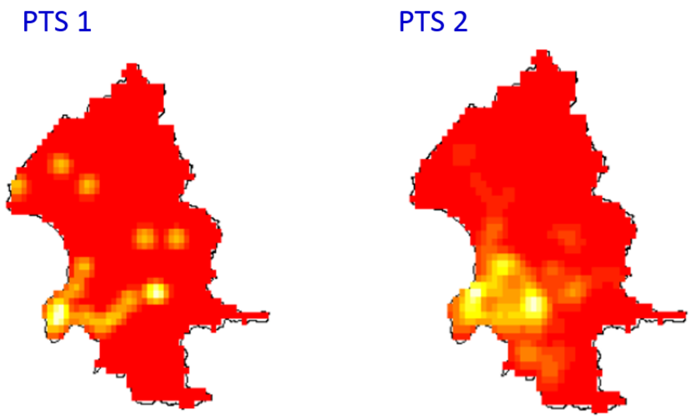
PTS 1

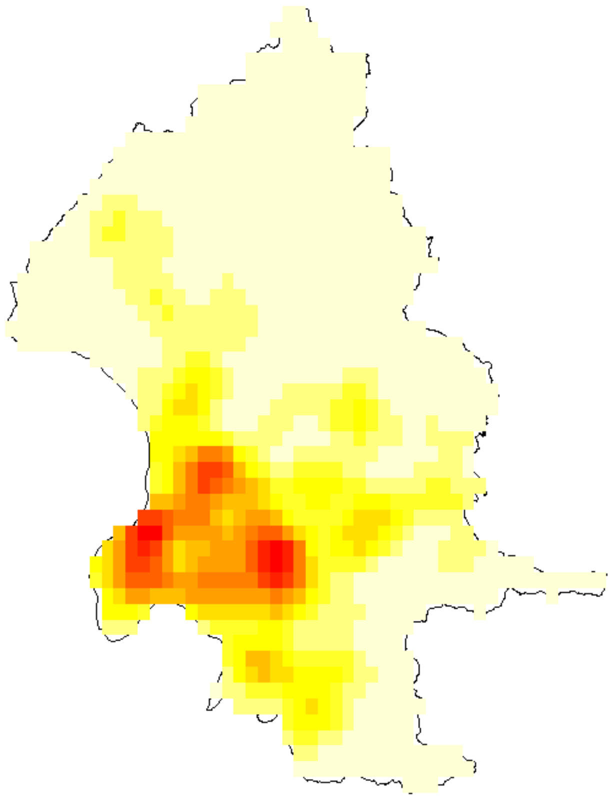


PTS 2



Raster Algebra：PTS 1 – PTS 2





Dual KDE 的應用

[1] Detecting change over time

Dual Kernel Density Estimation as a Method for Describing Spatio-Temporal Changes in the Upper Austrian Food Retailing Market

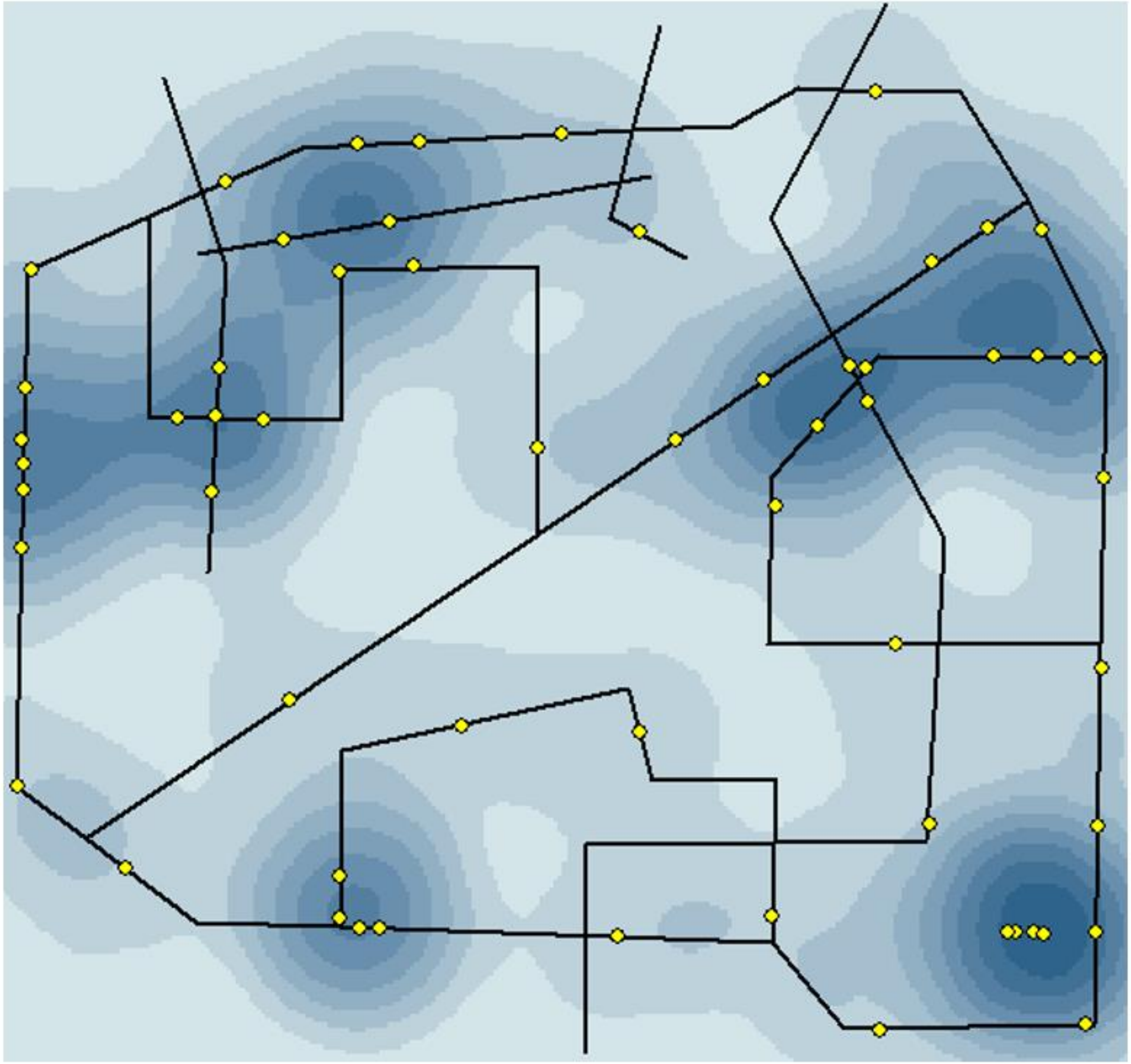
Eva Maria Jansenberger and Petra Staufer-Steinnocher
Department of Economic Geography & Geoinformatics
Vienna University of Economics and Business Administration
Rossauer Lände 23/1
1090 Vienna, Austria

[2] Underlying population-adjusted densities

GeoJournal
October 2015, Volume 80, Issue 5, pp 711–720 | Cite as

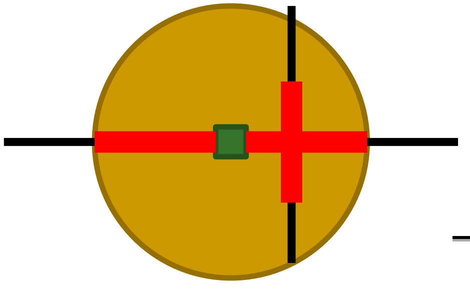
Using a dual kernel density estimate as a preliminary evaluation of the spatial distribution of diagnosed chronic kidney disease (CKD) in Edo State, Nigeria

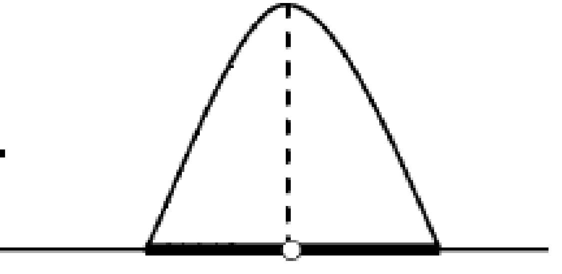
延伸應用：

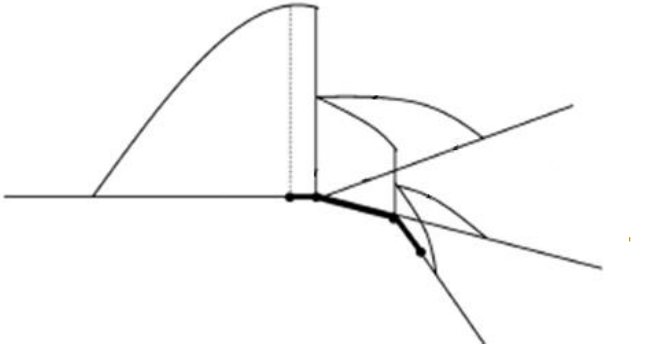


延伸應用：網絡空間的核密度估計簡介

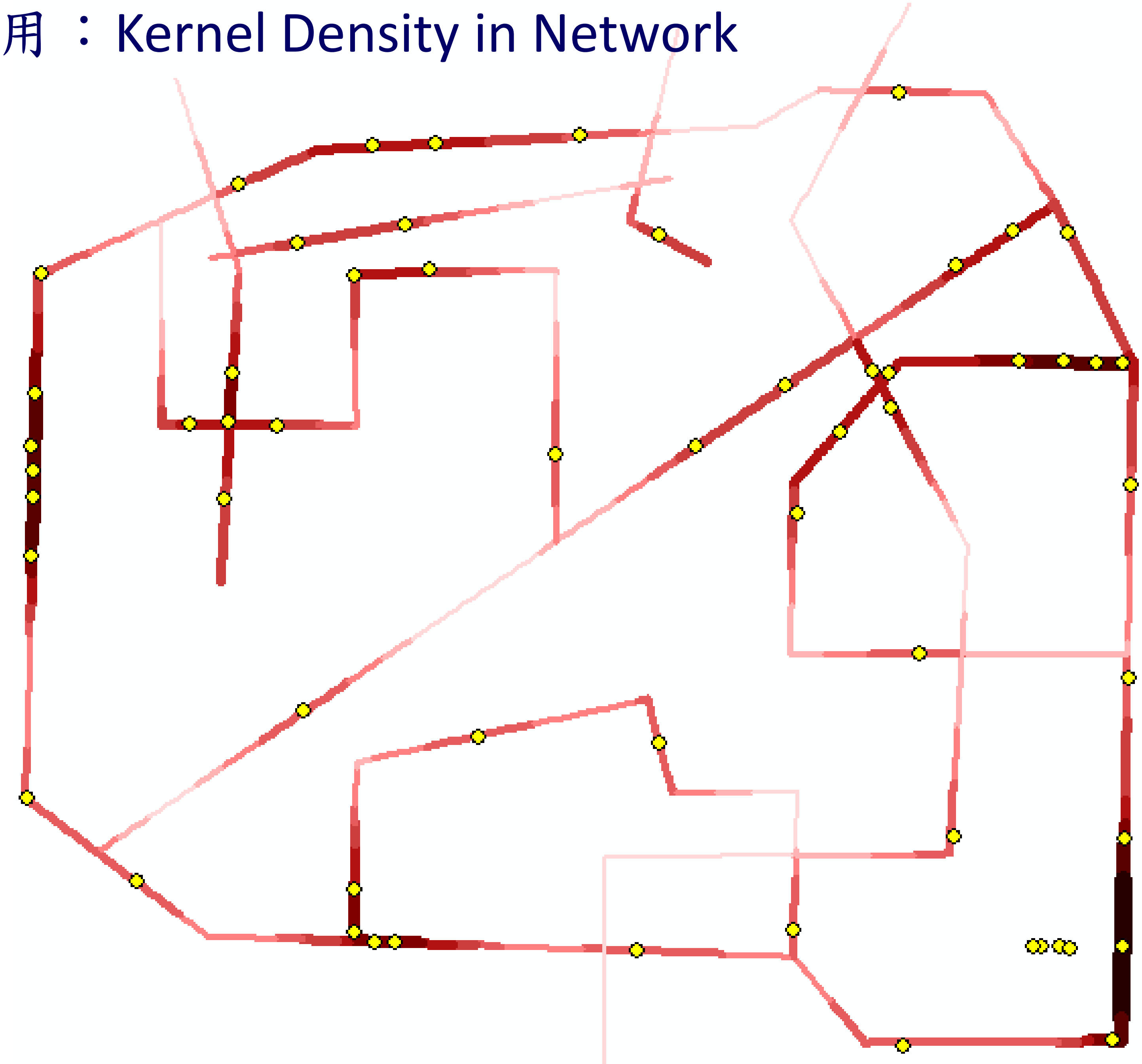
- 原理與平面空間的核密度估計類似
- 相異點：
  1. 網絡空間 vs. 平面空間
  2. 最短路徑 vs. 直線距離
  3. 岔路的處理



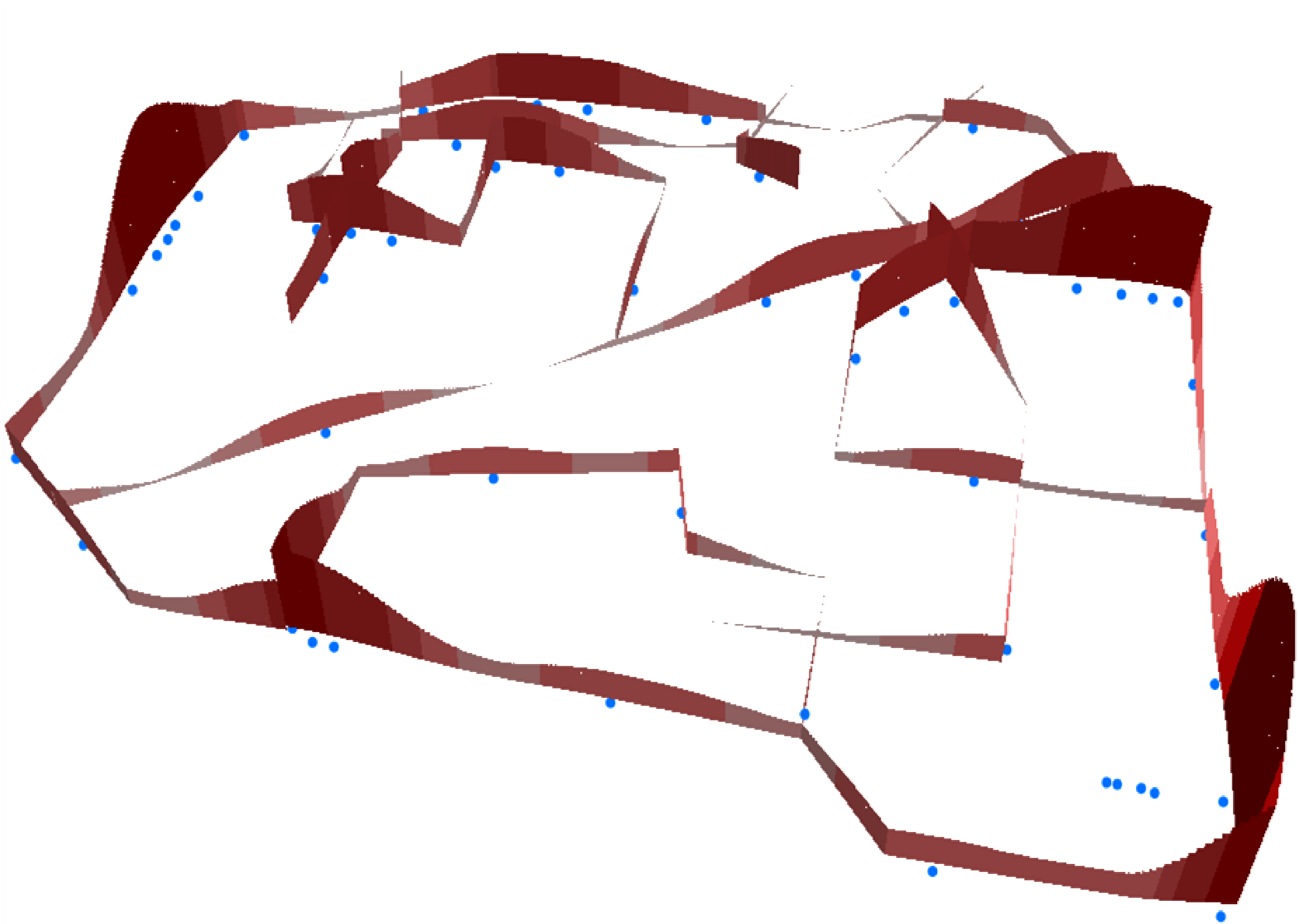




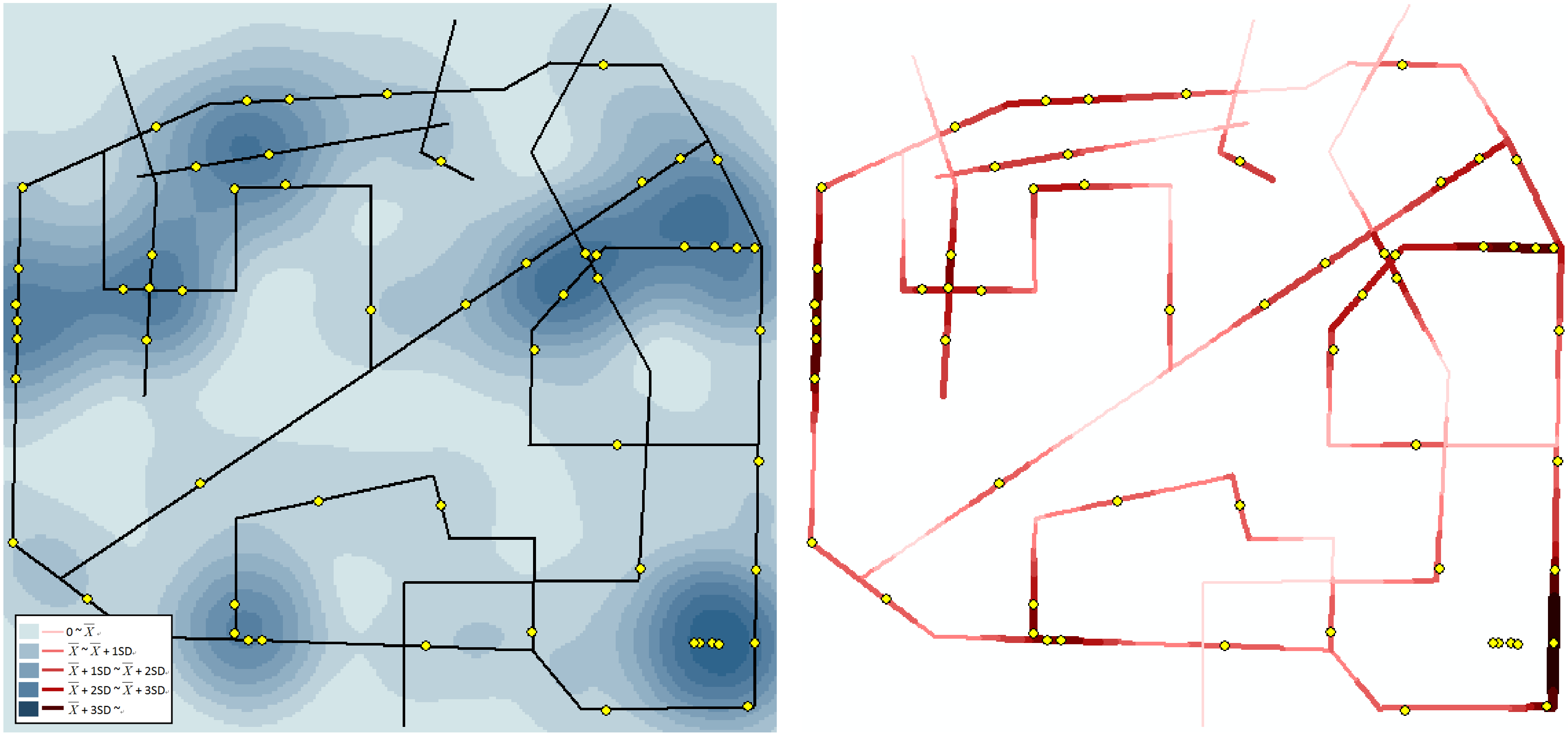
延伸應用：Kernel Density in Network



延伸應用：Kernel Density in Network



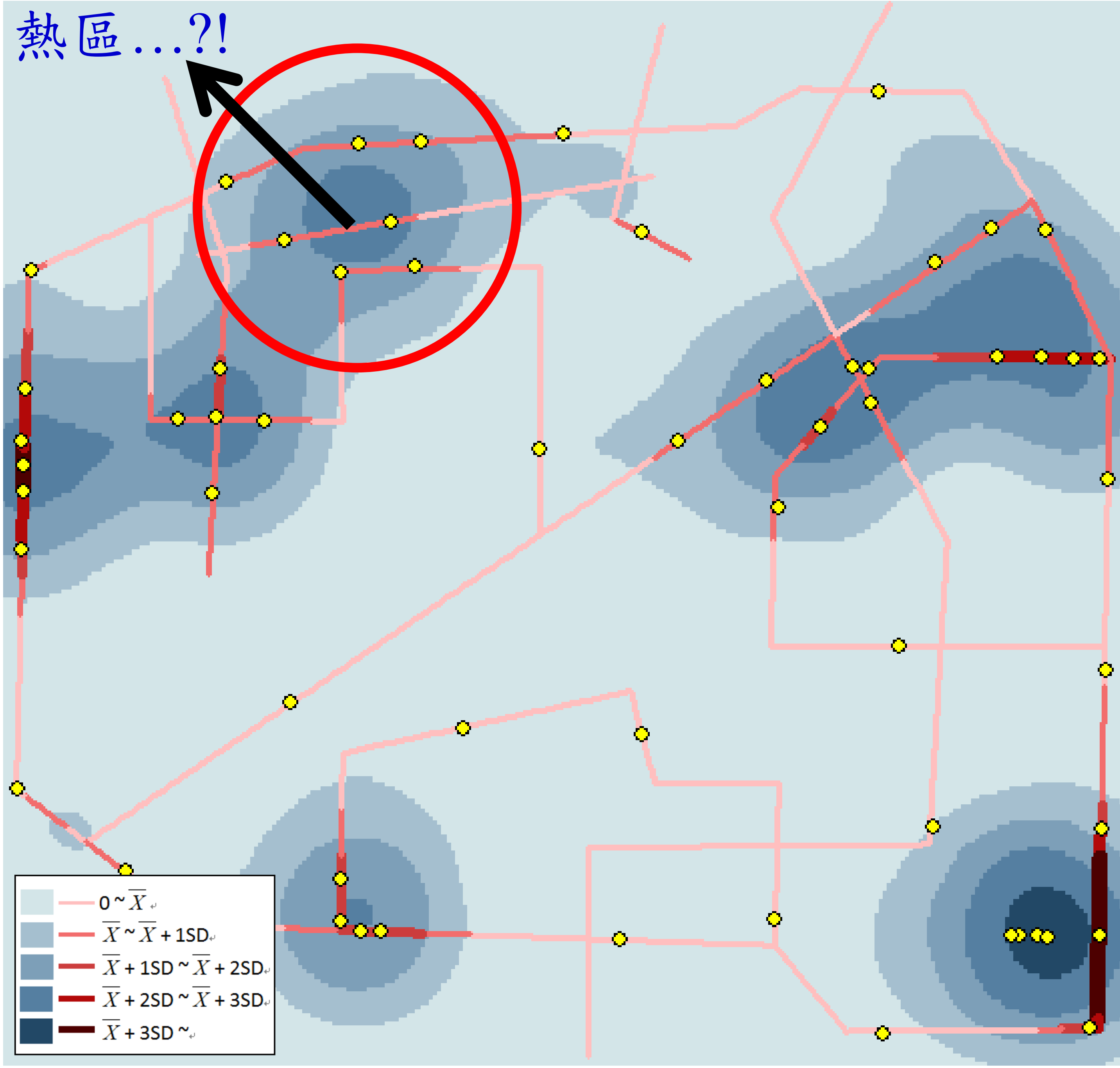
延伸應用：Kernel Density in Network



在哪些情況下，

網絡空間的核密度估計能更適當地表現出真實的群聚現象？

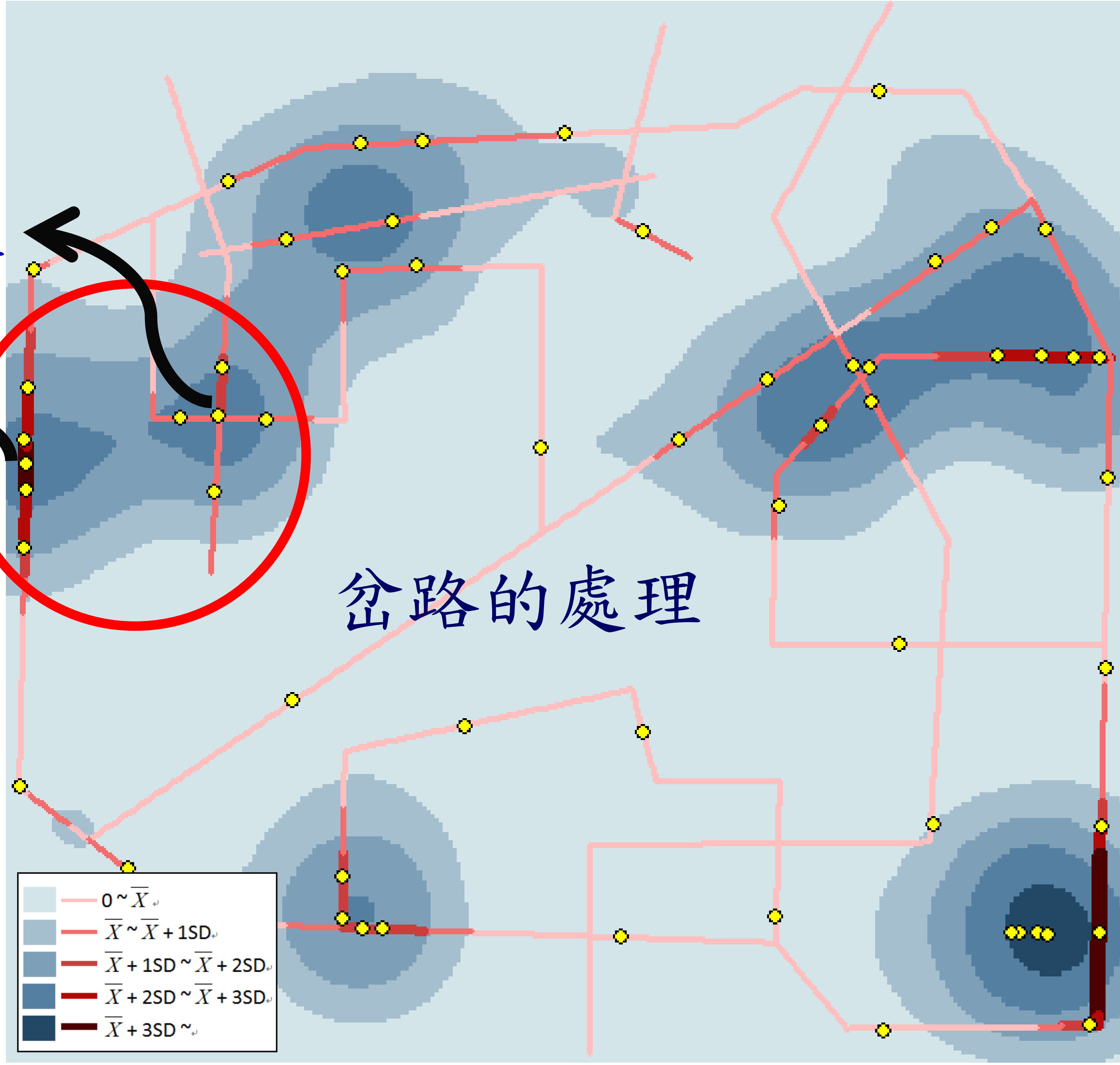
在哪些情況下，平面空間的核密度估計比較恰當？



熱區...?!
0 ~ $\overline{X}$
$\overline{X}$ ~ $\overline{X}$ + 1SD
$\overline{X}$ + 1SD ~ $\overline{X}$ + 2SD
$\overline{X}$ + 2SD ~ $\overline{X}$ + 3SD
$\overline{X}$ + 3SD ~

岔路的處理

皆為熱區



0 ~ $\overline{X}$
$\overline{X}$ ~ $\overline{X}$ + 1SD
$\overline{X}$ + 1SD ~ $\overline{X}$ + 2SD
$\overline{X}$ + 2SD ~ $\overline{X}$ + 3SD
$\overline{X}$ + 3SD ~

R 實作

SpatialKDE 0.6.2 | Get started | Reference | Changelog

SpatialKDE

Github actions: R-CMD-check passing | Build Windows Binaries and Create Release passing | Pkgdown - build and deploy website passing | Lifecycle stable

Travis build - R-devel: build passing

Cran Status: CRAN 0.6.2 | downloads 8293

R package to calculate spatial KDE. Inspired by the tool Heatmap tool from QGIS. Help for Heatmap tool can be found here, the help is for older version of the tool, but the window of the tool looks relatively the same.

Documentation

Available here.

https://jancaha.github.io/SpatialKDE/index.html

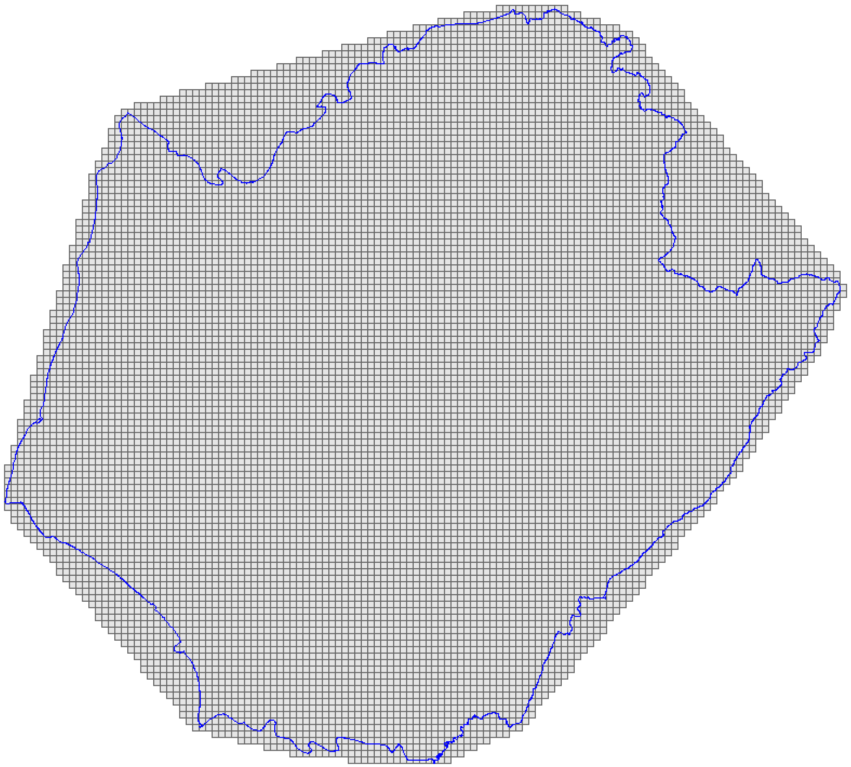
R code: Using Kde() function

```r
kde(
  points,
  band_width,
  decay = 1,
  kernel = c("quartic", "uniform", "triweight", "epanechnikov", "triangular"),
  scaled = FALSE,
  weights = c(),
  grid,
  cell_size
)
```

Step 1: Generating grids

台南市邊框

```
grid1 <- create_grid_rectangular(TN_BND, cell_size = 500)
```



Step 2: Kernel density estimation

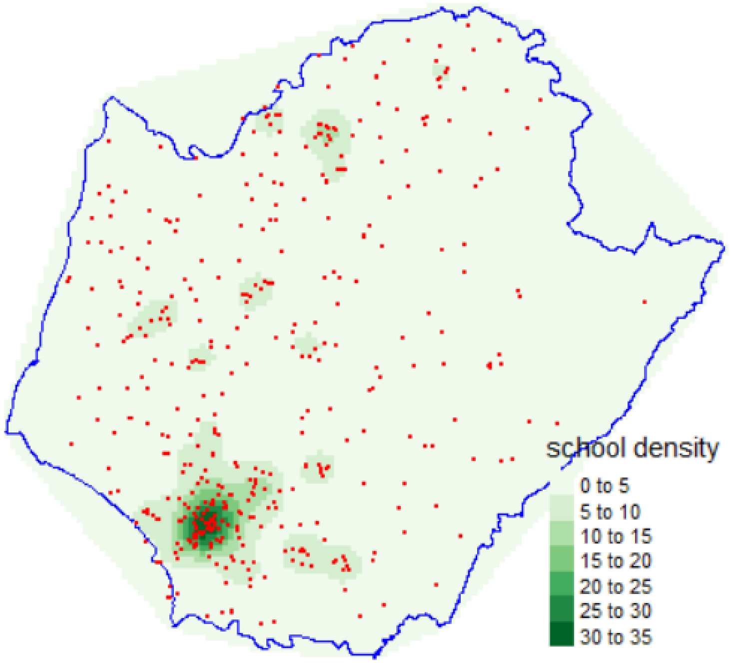
學校點資料

```
TN_KDE <- kde(schools_sf, band_width = 4000, grid = grid1)
```

```
> class(TN_KDE)
[1] "sf"         "data.frame"
```

Step 3: Mapping

```
TN_KDE_lyr <- tm_shape(TN_KDE) +
        tm_polygons( col = "kde_value",
                palette = "Greens",
                border.alpha=0,
                title = "school density")
```



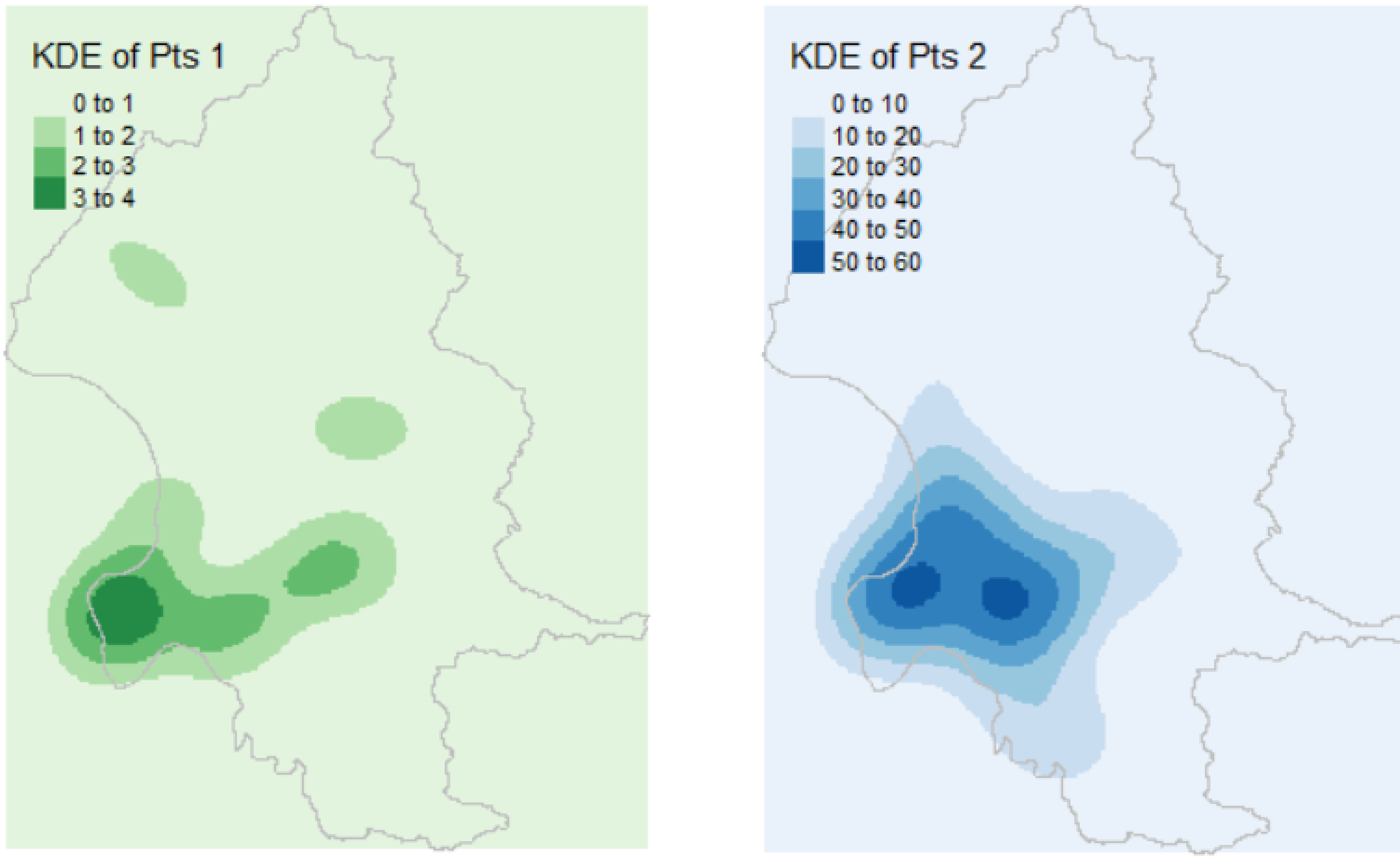
R code: Raster format

```
raster_grid <- create_raster(TPE_BND, cell_size = 100)

TPE_Pts1 <- kde(pts1_sf, band_width = 3000,
                         grid = raster_grid)

TPE_KDE1_lyr <- tm_shape(TPE_Pts1)
                        + tm_raster(palette = "Greens",
                                    title = "KDE of Pts 1")
```

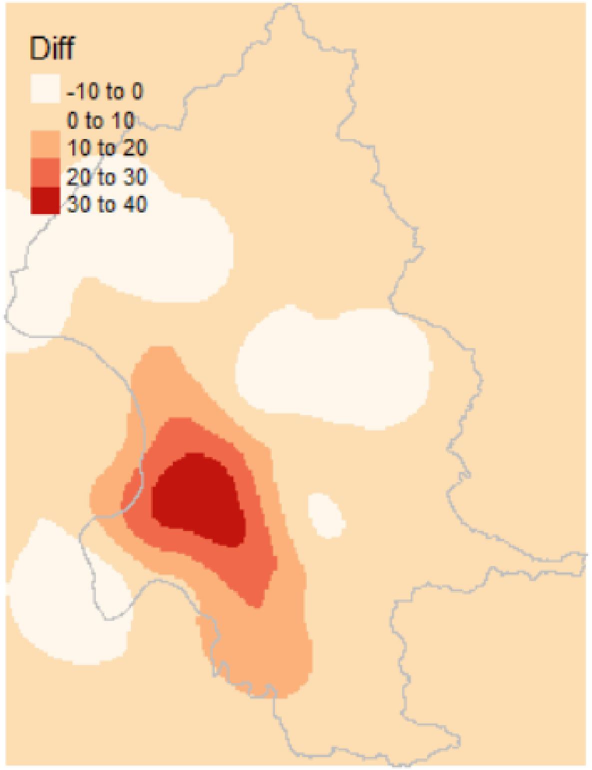
R code: Raster format



R code: Raster Algebra

```
TPE_Diff<- TPE_Pts2 - 10 * TPE_Pts1

TPE_Diff_lyr<- tm_shape(TPE_Diff)
                + tm_raster(palette = "OrRd", title = "Diff")
```



本週實習

- 利用KDE方法，分別呈現台灣南部地區(嘉義、台南、高雄與屏東縣市)的「媽祖」與「觀音菩薩」寺廟密度地圖。

本週作業-1

- 透過 KDE方法 ，設定網格大小為 100x100公尺，核密度函數為 quartic ，搜尋半徑為 1000公尺，找出哪些學校周圍的速食店市場獨占性最高 ？（即該地方附近的速食店多但學校數量少）

- (1) 依上述題幹，繪製 Dual KDE 地圖 。
- (2) 列出速食店獨占性最高的三間學校的名稱（Hint: 學校所在網格之 KDE數值 ）。

本週作業-2

- 論文研讀心得

https://www.mdpi.com/2413-8851/2/3/91

urban science

Open Access Article

Measuring Urban Renewal: A Dual Kernel Density Estimation to Assess the Intensity of Building Renovation—Case Study in Lisbon

by Filipe J. S. Brandão [1,*], Ricardo M. Correia [1] and Alexandra Paio [2]

[1] ISTAR-IUL, Instituto Universitário de Lisboa (ISCTE-IUL) Avenida das Forças Armadas, 1649-026 Lisboa, Portugal

[2] ISTAR-IUL and VITRUVIUS FABLAB and DINAMIA-IUL, Instituto Universitário de Lisboa (ISCTE-IUL), Avenida das Forças Armadas, 1649-026 Lisboa, Portugal

[*] Author to whom correspondence should be addressed.

Urban Sci. **2018**, *2*(3), 91; https://doi.org/10.3390/urbansci2030091