熱區分析 Hot spot analysis (Localized spatial analysis)

授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw

授課大綱:熱區分析的空間統計方法

- Host-spot analysis (for polygon data)
  - Local Moran’s I index
  - Local G-statistic (Gi*)
- Issues of multiple testing for hot-spot analysis
  - Bonferroni correction
  - False discovery rate (FDR)

Textbook Chapter

TEXT_Local.Stat.pdf

Chapter 8

Local Statistics

CHAPTER OBJECTIVES

In this chapter, we:

- Explain the concepts underlying the emerging array of *local statistics*
- Account for the relatively late arrival of local statistics on the spatial analytic scene
- Review the various approaches that can be used to construct *localities* for the development of local statistics
- Discuss how the popular Getis-Ord family of G statistics are calculated and interpreted
- Outline the local version of Moran's I statistic
- Explain why inference based on local statistics is challenging and describe current approaches to dealing with the difficulties
- Provide an overview of the increasingly popular method *geographically weighted regression*
- Explain how many other spatial analysis methods can be considered as local statistics even if this was not the intent behind their original development

Chap 8: Local Statistics

8.1 Introduction: Think geographically, measure locally

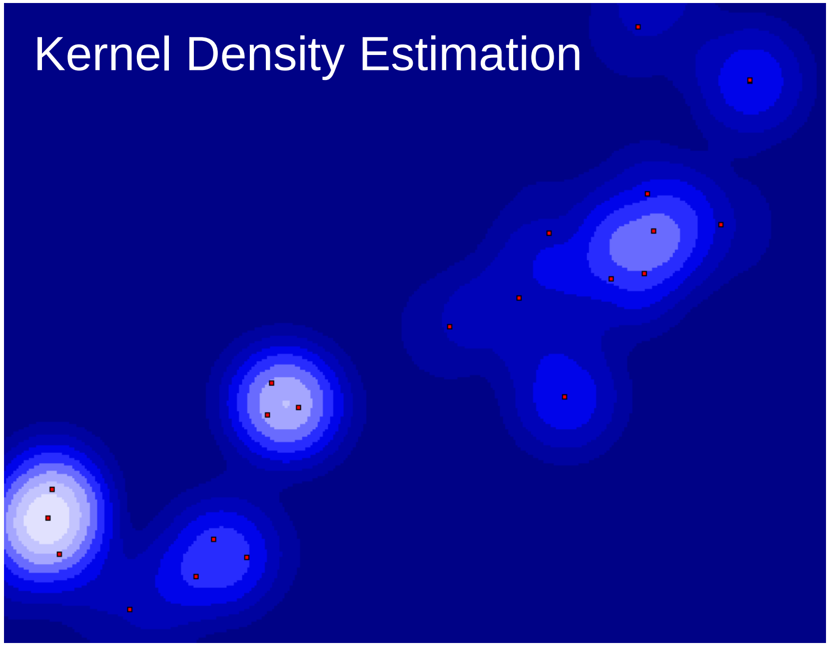
8.2 Defining the local

8.3 An example

8.4 Inference with local statistics

Identifying hot spots

Point data



Polygon data

Recap: Global Moran's I

相關係數

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

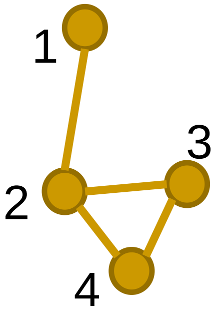
自相關係數
Moran's I

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where

- N is the number of cases
  X is the mean of the variable
  X_i is the variable value at a particular location
  X_j is the variable value at another location
  W_{ij} is a weight indexing location of i relative to j

- Applied to a continuous variable for polygons or points



$$\begin{array}{c c} & \begin{array}{cccc} \#1 & \#2 & \#3 & \#4 \end{array} \\ \begin{array}{c} \#1 \\ \#2 \\ \#3 \\ \#4 \end{array} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(y_i-\bar{y})(y_j-\bar{y}) / \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}}$$

$$\frac{\left[\begin{array}{l}(x1-\bar{x})(x2-\bar{x}) + \\ (x2-\bar{x})(x1-\bar{x}) + (x2-\bar{x})(x3-\bar{x}) + (x2-\bar{x})(x4-\bar{x}) + \\ (x3-\bar{x})(x2-\bar{x}) + (x3-\bar{x})(x4-\bar{x}) + \\ (x4-\bar{x})(x2-\bar{x}) + (x4-\bar{x})(x3-\bar{x})\end{array}\right] \Big/ 8}{\dfrac{(x1-\bar{x})^2+(x2-\bar{x})^2+(x3-\bar{x})^2+(x4-\bar{x})^2}{4}}$$

Using row-standardized spatial weights

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$I = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$I = \frac{\sum_i \sum_j \frac{w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{s_i \times s_j}}{\sum_i \frac{(x_i - \bar{x})^2}{s_i \times s_i}}$$

轉換原始分數成為 Z 分數的公式為：

$$Z = \frac{X_i - \overline{X}}{S}$$

$$\Downarrow$$

$$I = \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}$$

1. Local Moran's I
(Local Indicator of Spatial Association, LISA)

$$I = \frac{\sum_i \boxed{\sum_j w_{ij} z_i z_j}}{\sum_i z_i^2}$$

⬇

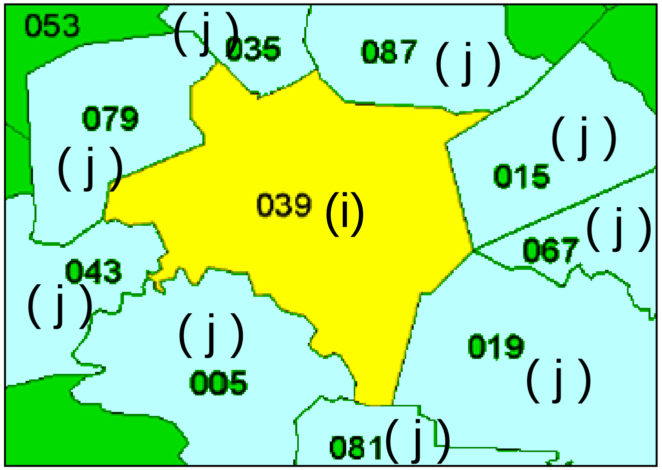
$$I_i = \frac{\sum_j w_{ij} z_i z_j}{\boxed{\sum_i z_i^2}}$$

a constant for each i

Local Moran's I (LISA)

$$I_i = z_i \boxed{\sum_j w_{ij} z_j}$$

$$z_i = (x_i - \bar{x}) / \delta$$



- High LISA value
  - Cluster of similar values (can be high or low)
- Low LISA value
  - Cluster of dissimilar values

Test of Statistical Significance

The z_{I_i}-score for the statistics are computed as:

$$z_{I_i} = \frac{I_i - \mathrm{E}[I_i]}{\sqrt{\mathrm{V}[I_i]}} \tag{3}$$

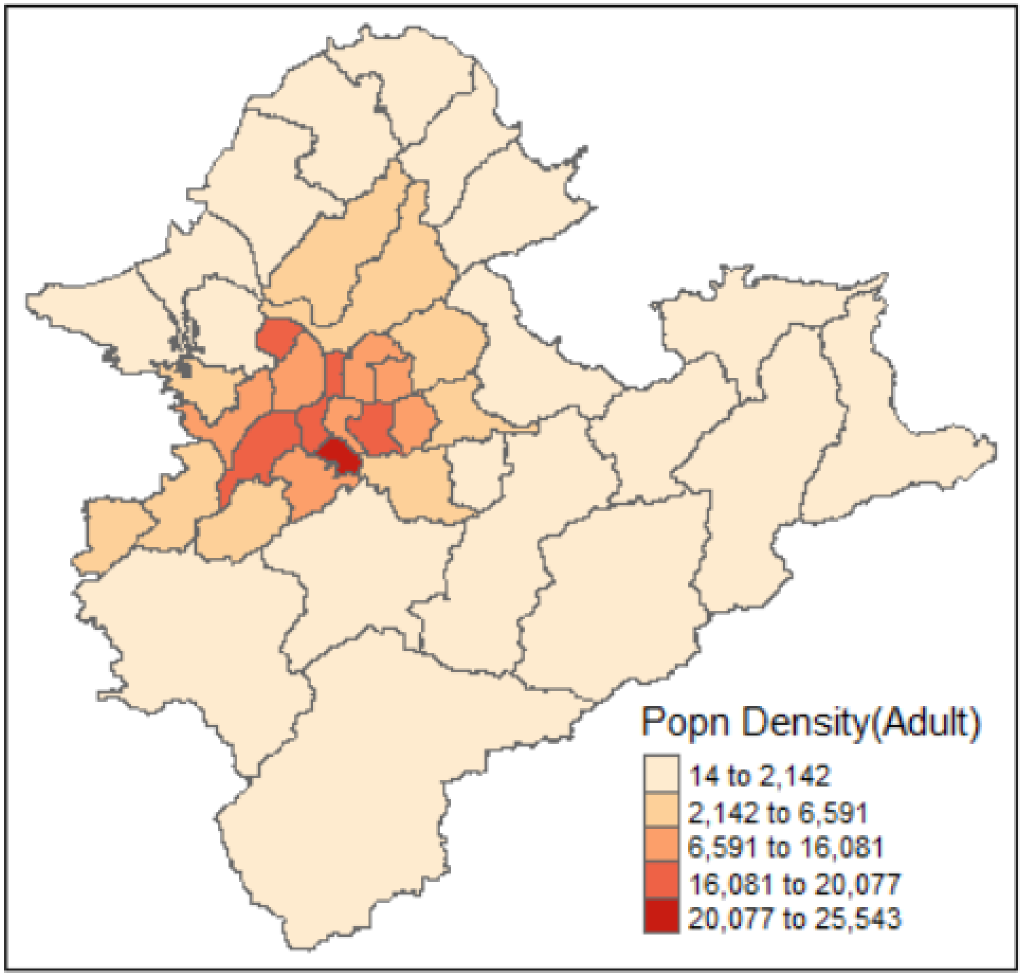
where:

$$\mathrm{E}[I_i] = -\frac{\sum\limits_{j=1, j \neq i}^{n} w_{ij}}{n - 1} \tag{4}$$

$$\mathrm{V}[I_i] = \mathrm{E}[I_i^2] - \mathrm{E}[I_i]^2 \tag{5}$$

Lab 1: Local Moran's I

年齡15-64的人口密度 (/km^2)



Lab 1: Local Moran's I in R

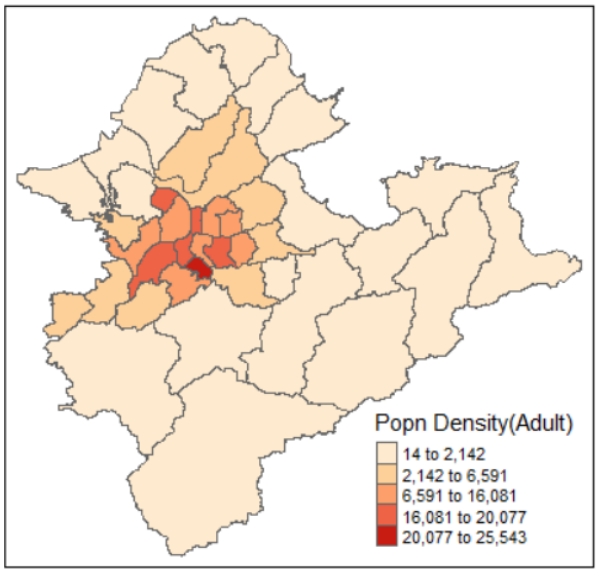
using localmoran() function

```
LISA.Popn <- localmoran(Density, TWN_nb_w, zero.policy=T)

NorthTW_sf$z.li <- LISA.Popn[,4]
NorthTW_sf$pvalue <- LISA.Popn[,5]
```

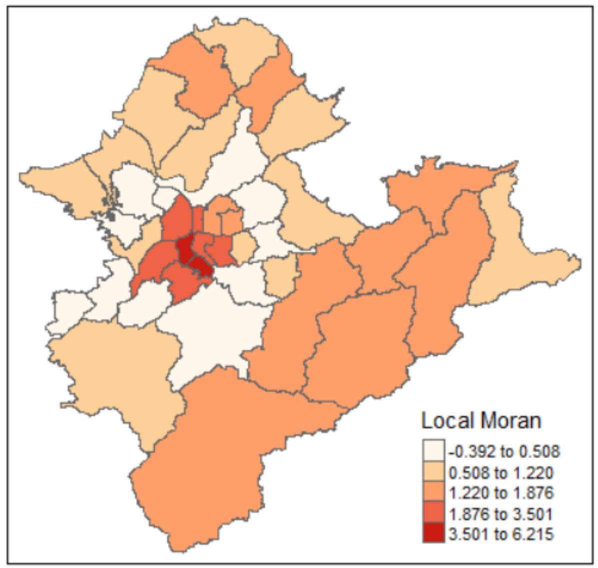
```
> LISA.Popn <- localmoran(Density, TWN_nb_w, zero.policy=T)
> LISA.Popn
             Ii   E.Ii     Var.Ii         Z.Ii    Pr(z > 0)
221 0.668654386 -0.025 0.17429750  1.66148935 4.830760e-02
222 0.552267439 -0.025 0.22386865  1.22005786 1.112215e-01
223 1.122969659 -0.025 0.17429750  2.74969696 2.982520e-03
224 0.520540764 -0.025 0.14125006  1.45155294 7.331298e-02
225 1.283663684 -0.025 0.14125006  3.48203973 2.488049e-04
226 1.275561091 -0.025 0.17429750  3.11519460 9.191180e-04
227 2.310784759 -0.025 0.14125006  6.21496221 2.566850e-10
```

R Lab: Local Moran's I in R

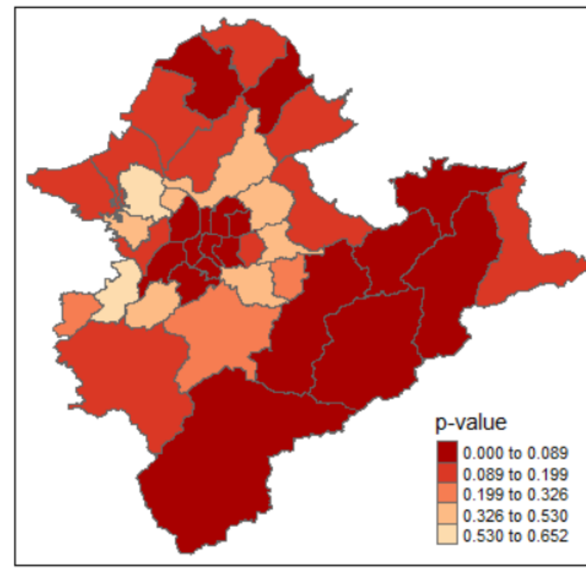
Population Density



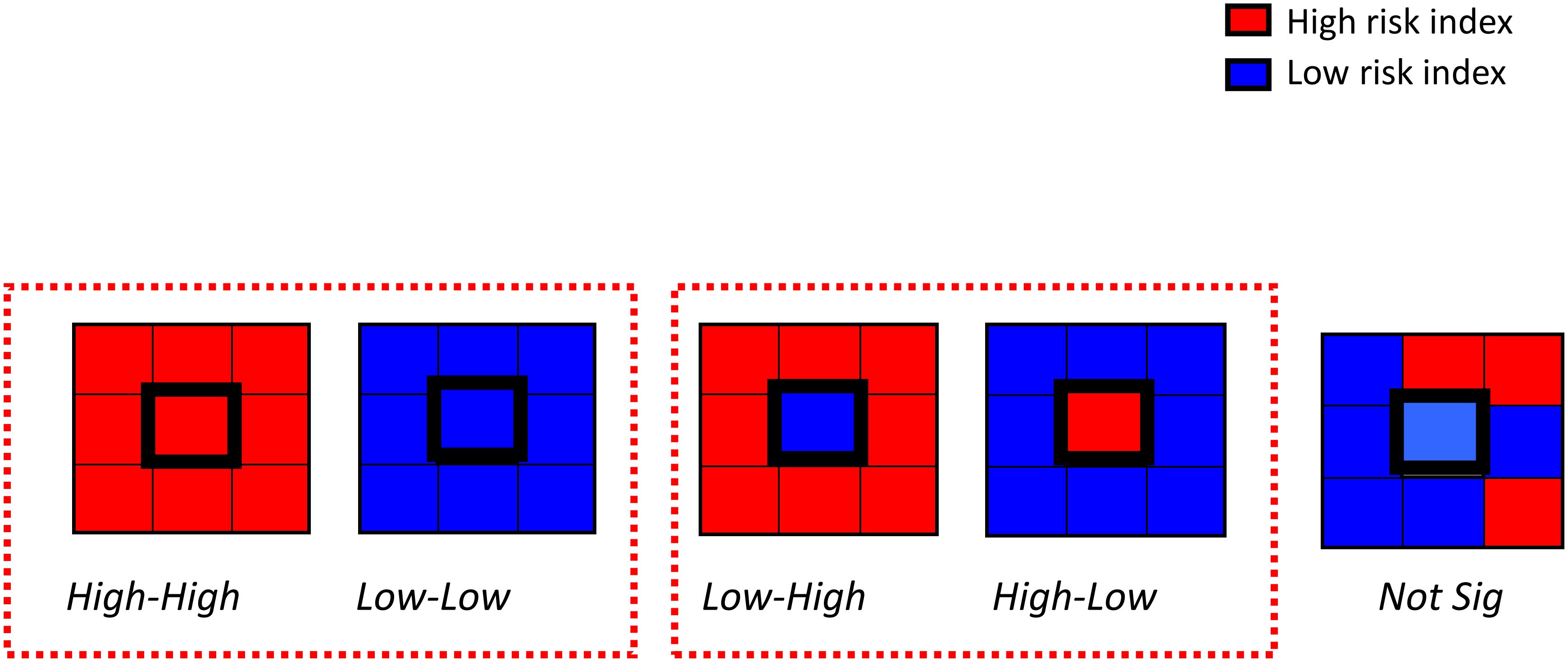
Local Moran (z-score for LISA)



p-value



Local Moran's I (LISA)



High LISA value:
Cluster of similar values

Low LISA value:
Cluster of dissimilar values

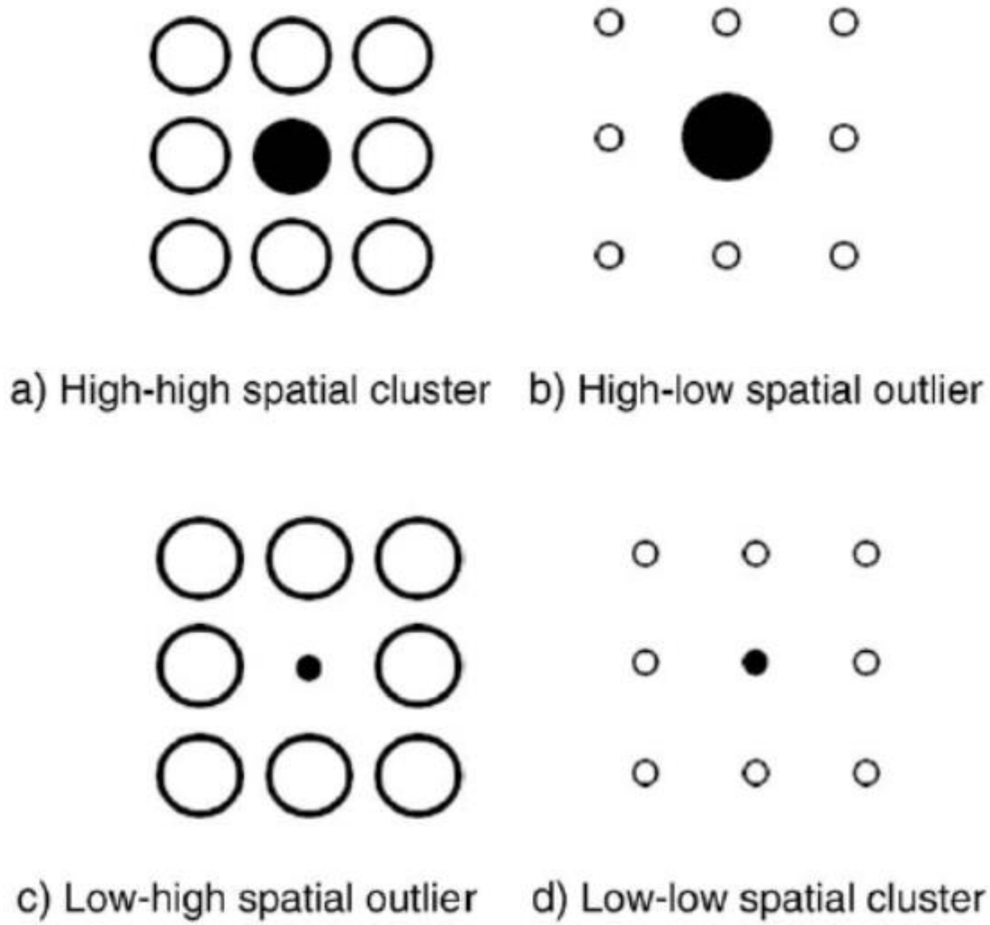
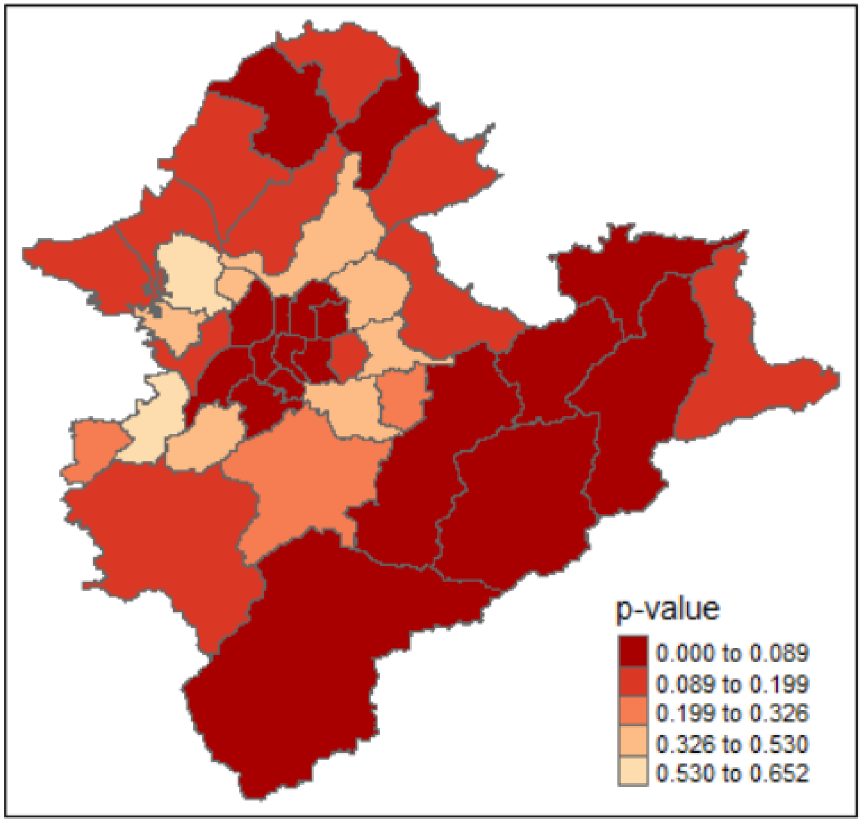


Fig. 1 – Sketch figure showing the relationship of a location and its neighbourhood: a) and d) spatial cluster; b) and c) spatial outlier; a) and b) hot spots; c) and d) cool spots.

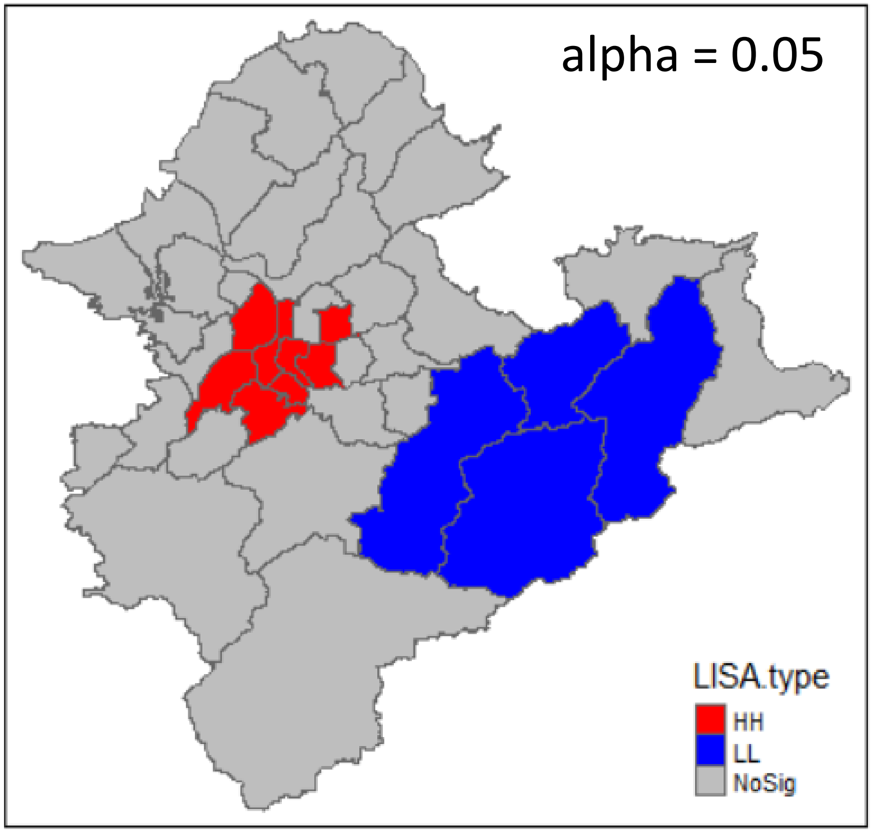
Zhang et al., 2008

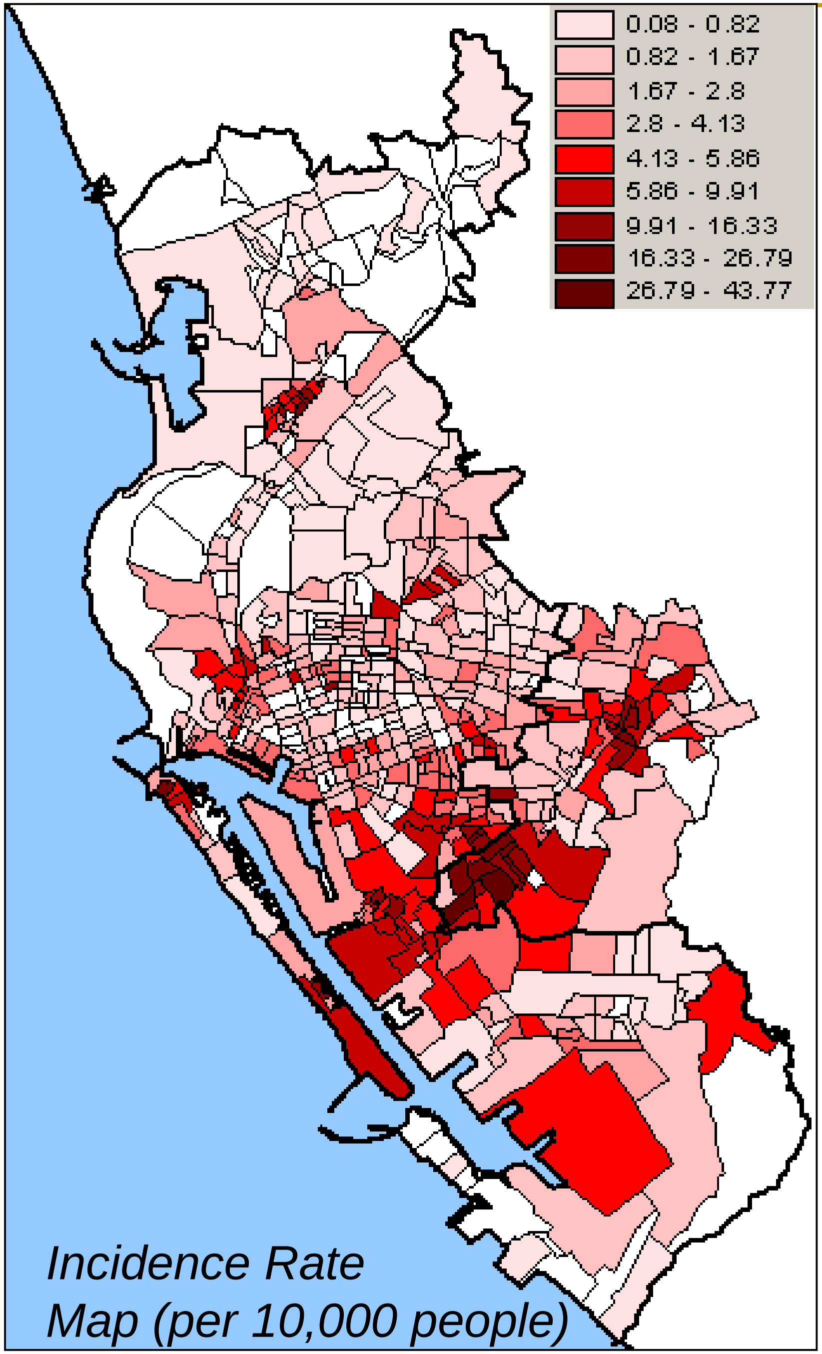
R Lab: Local Moran's I

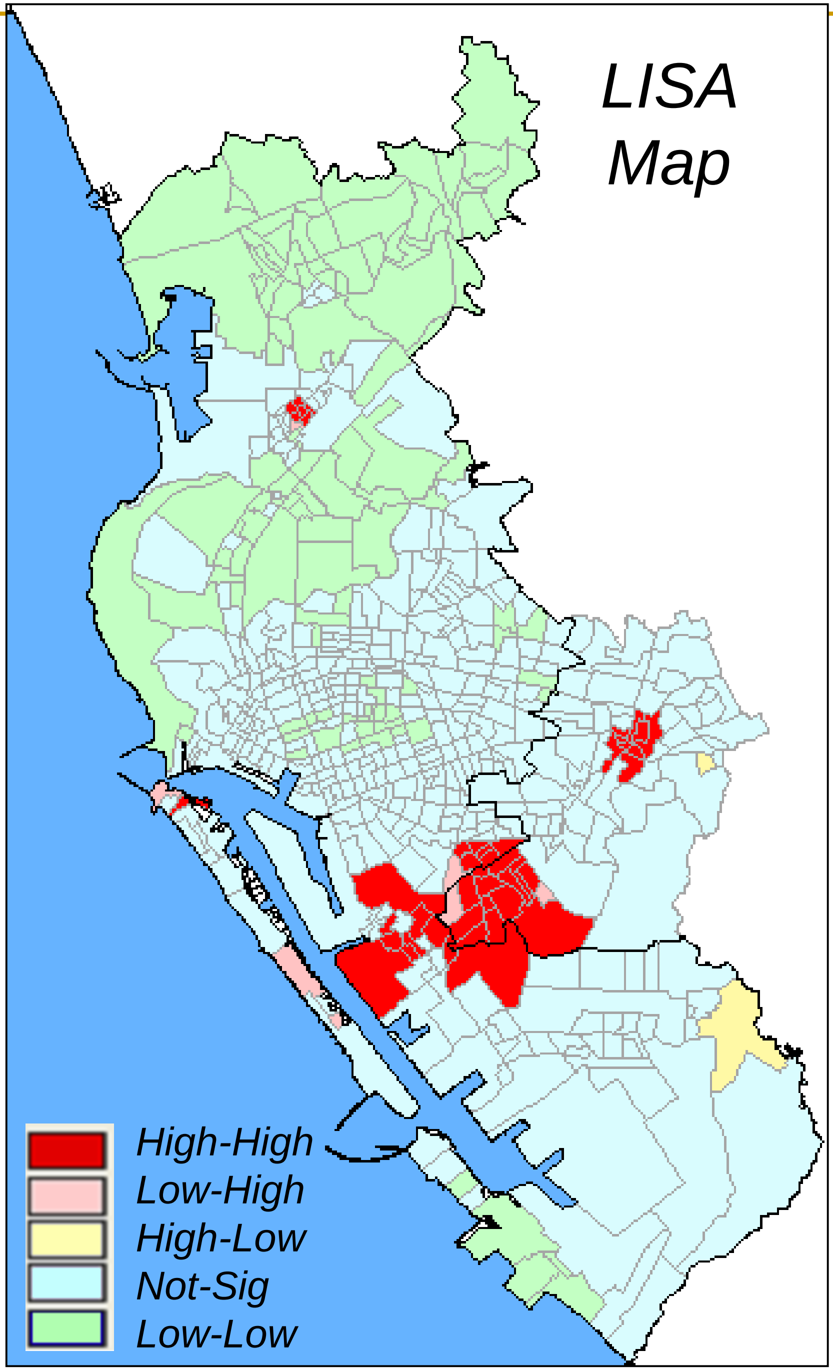
p-value of Local Moran



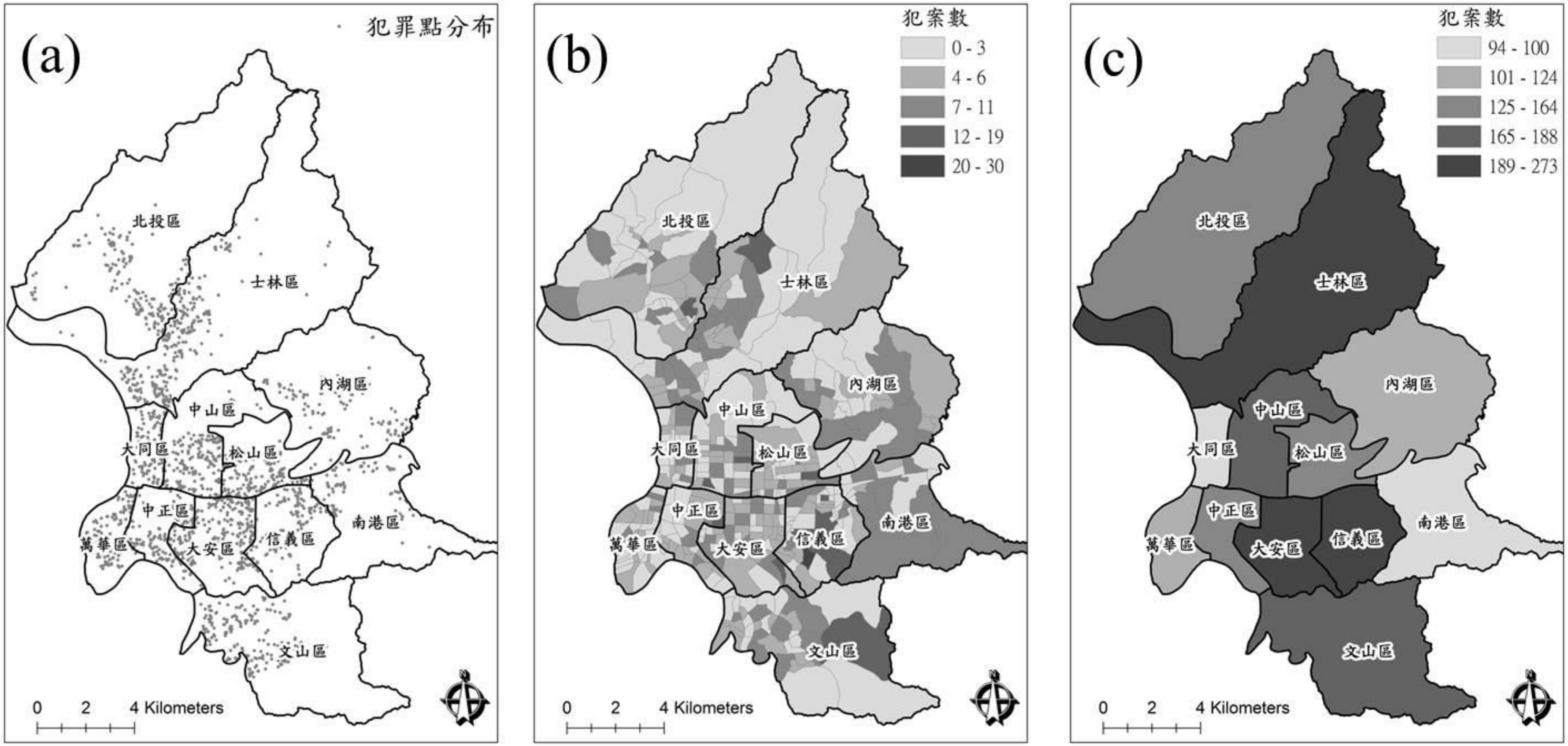
LISA Cluster Map




0.08 - 0.82
0.82 - 1.67
1.67 - 2.8
2.8 - 4.13
4.13 - 5.86
5.86 - 9.91
9.91 - 16.33
16.33 - 26.79
26.79 - 43.77
Incidence Rate
Map (per 10,000 people)

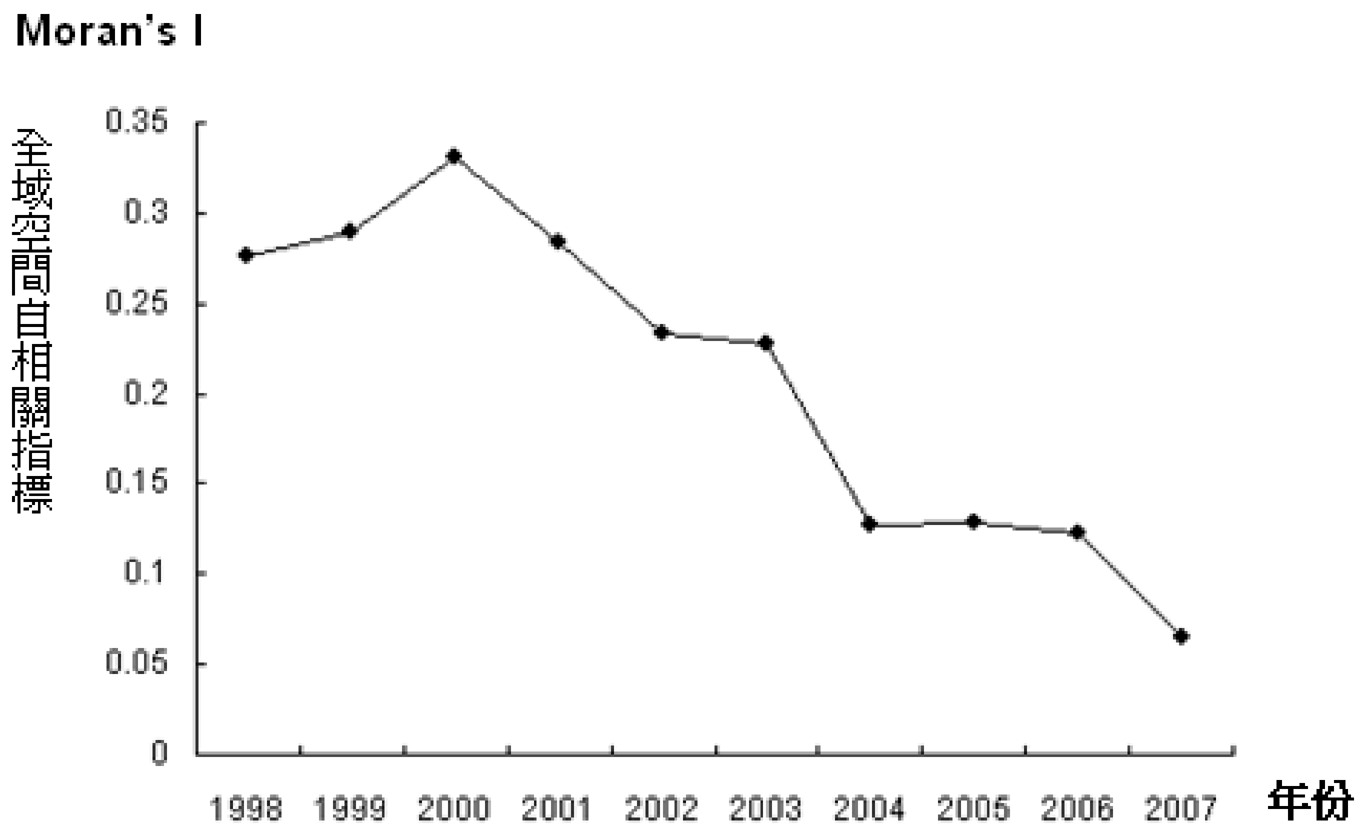

LISA
Map
High-High
Low-High
High-Low
Not-Sig
Low-Low

應用範例：
1998-2007年台北市住宅竊盜犯罪趨勢分析

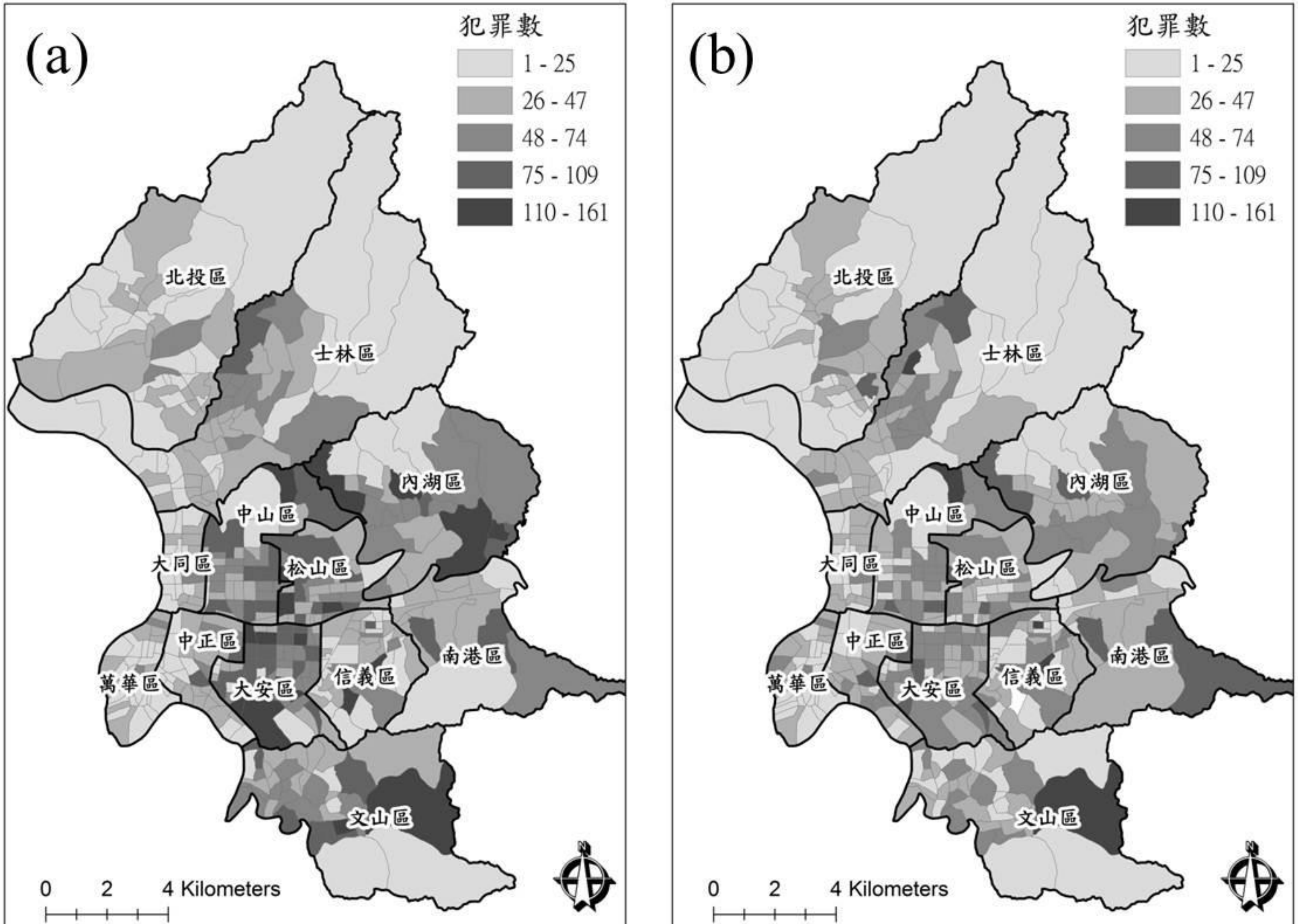


溫在弘等 (2010)，犯罪地圖繪製與熱區分析方法及其應用：以1998~2007年臺北市住宅竊盜犯罪為例，地理研究 52:43-63

Temporal Trend of Global Moran's I

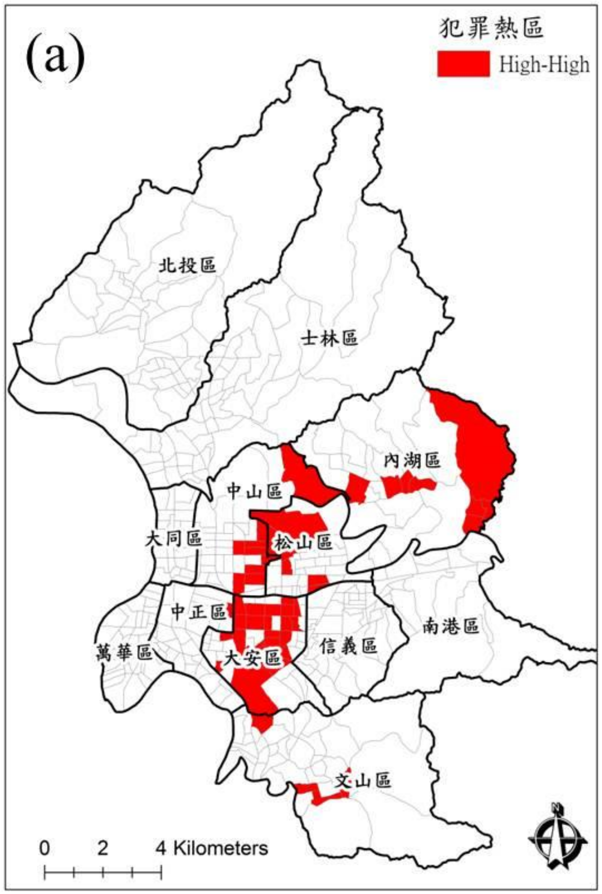


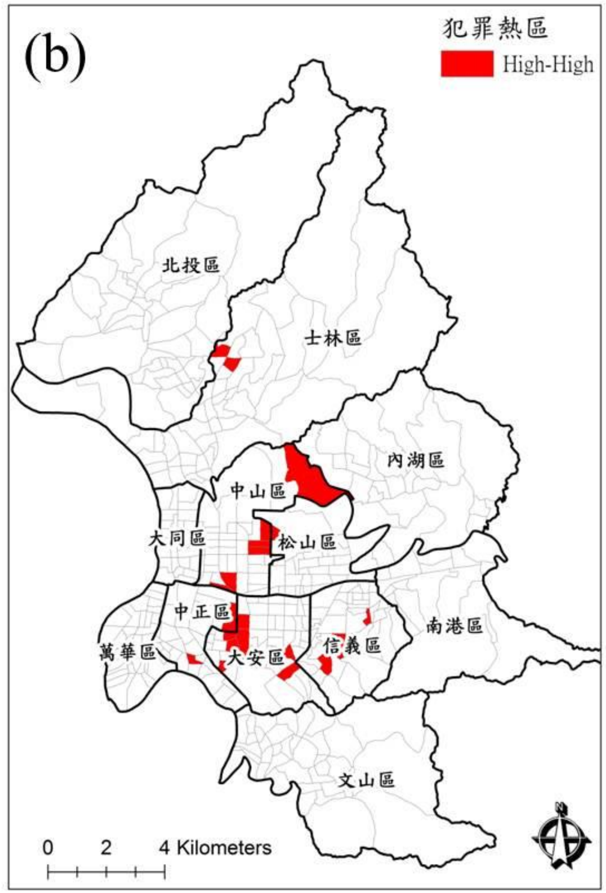
(a). 1998-2002 vs. (b). 2003-2007



Local Moran's I (LISA): H-H hot-spots

(a). 1998-2002 vs. (b). 2003-2007





參數設定：以正方格四交點相鄰的Queen型態為相鄰定義，紅色的地區亦即表示在0.05的統計顯著水準下，犯罪趨勢顯著呈現地理群聚的區域

Recap: General G-statistic

- Moran's I & Geary's C Ratio 無法區別

  " hot spots" or "cold spots"

- *Spatial Concentration* method
- Definition

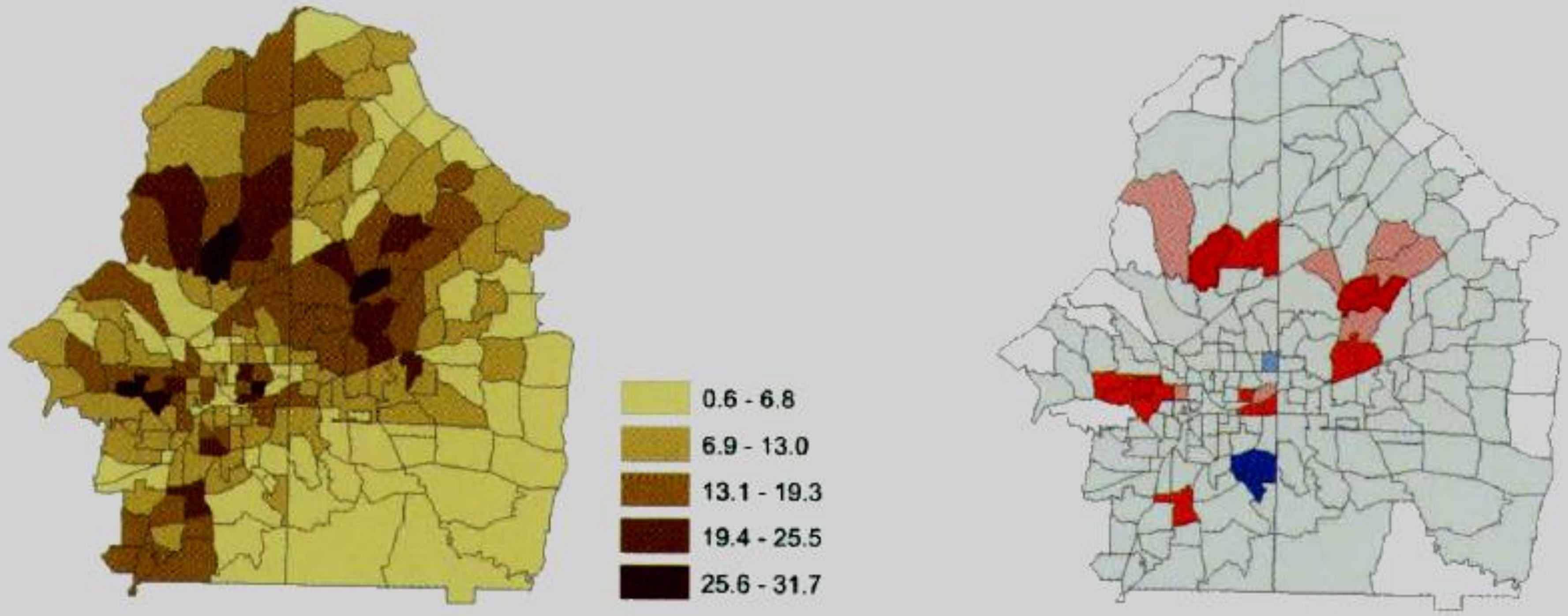
$$G(d) = \frac{\sum \sum w_{ij}(d) x_i x_j}{\sum \sum x_i x_j}$$

d : neighborhood distance

W_{ij} : 1 if it is within d, 0 otherwise

- Calculation of G must begin by identifying a neighborhood distance within which cluster is expected to occur

2. Local Analysis of G-statistic: Identifying spatial concentration with low and high values



Percent age 65 and over, by census tract (left). The map on the right shows clusters of tracts having a high percentage of seniors (orange) and tracts having a significantly lower percentage than their neighbors (blue).

Source: Chapter 4 Identifying clusters
The ESRI Guide to GIS Analysis, Volume 2

Local G-statistic

Two versions of the local G-statistic

There are two versions of this statistic, both developed by Art Getis and Keith Ord. In one version, the value of the target feature itself is not included in the equation. This is the Gi statistic. You'd use the Gi statistic if you're interested in the effect of the target feature on what's going on around it. This would be the case if you're interested in the dispersion of a particular phenomenon from the target feature to the surrounding area over time. Getis and Ord, for example, used Gi to track the dispersion of AIDS to counties surrounding San Francisco County over the course of several years. They wanted to see if the intensity of clustering of AIDS cases in counties surrounding San Francisco increased over time and the distance at which the clustering peaked. See the references at the end of this chapter for more on the Gi statistic.

In the other version, called Gi* (pronounced G-i-star), the value of the target feature is included. If you're interested in finding hot spots or cold spots, you'd use Gi*—you'll want to include the value of the target feature since its value contributes to the occurrence of the cluster.

Source: Chapter 4 Identifying clusters
The ESRI Guide to GIS Analysis, Volume 2

Getis-Ord Local G Statistic

Neighborhood Definition

$$G_i(d) = \frac{\sum_j w_{ij}(d) x_j}{\sum_j x_j}; j \neq i$$

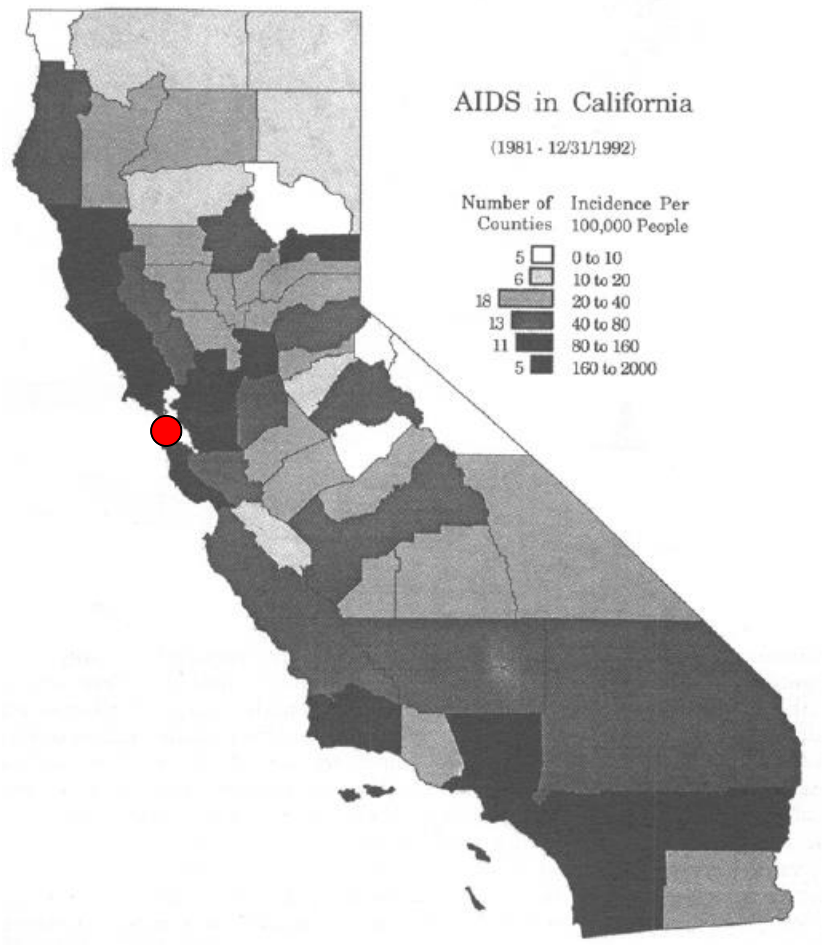
- The Gi statistic **excludes** the value at i from the summation and is used for spread or diffusion studies
- the Gi* **includes** the value at i in the summation (for all j) and is most often used for studies of clustering

Smaller ← Gi(d) → *Larger*

Cluster of low values Mean *Cluster of high values*

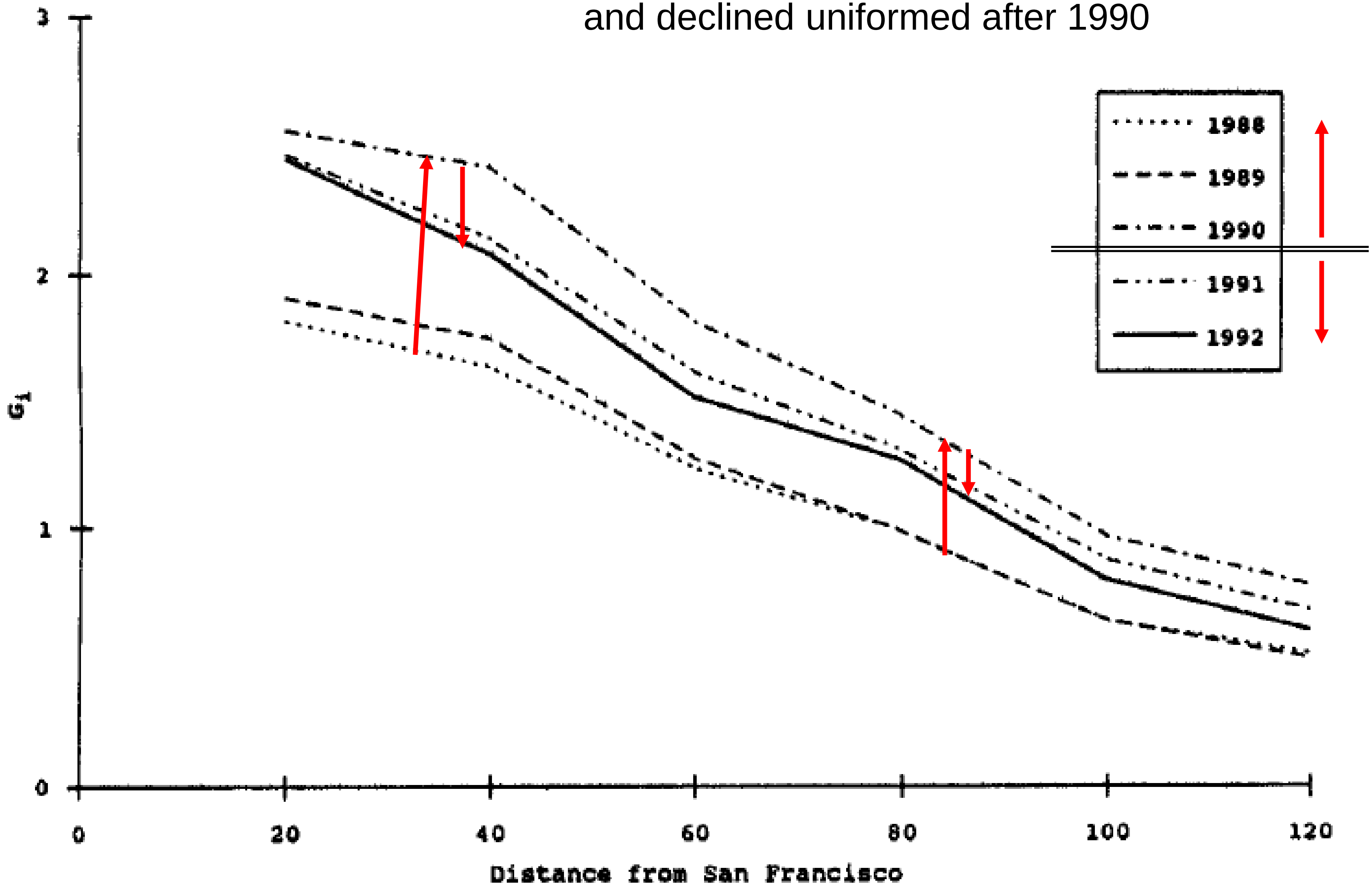
Source: Chapter 4 Identifying clusters
The ESRI Guide to GIS Analysis, Volume 2

Example: Diffusion of AIDS in California



Source: Ord and Getis,1995

The rate of AIDS cases increased uniformly over the area of clustering from 1988-1990 and declined uniformed after 1990



Source: Ord and Getis,1995

Testing the statistical significance of Gi*

The Z-score for Gi*

The expected Gi* value is subtracted from the observed Gi*....

$$Z(G_i^*) = \frac{G_i^* - E(G_i^*)}{\sqrt{Var(G_i^*)}}$$

....and the difference divided by the square root of the variance

The expected Gi* value

The weights (w_{ij}) at a distance (d) are summed....

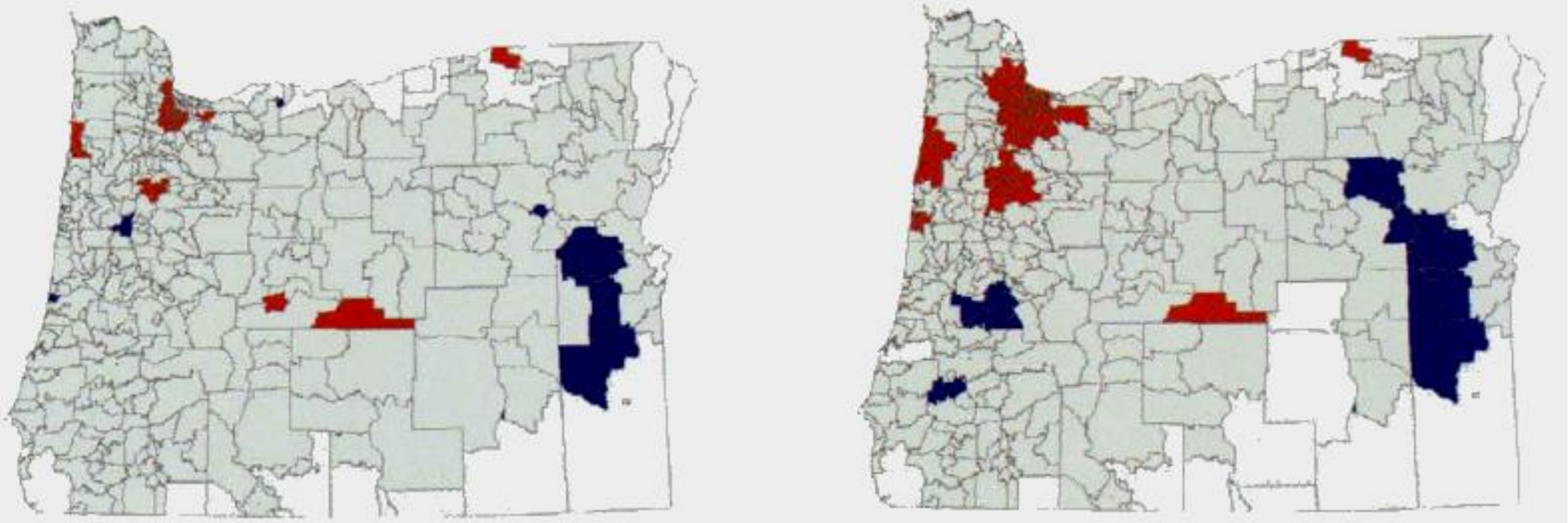
$$E(G_i^*) = \frac{\sum_j w_{ij}(d)}{n-1}$$

....and divided by the number of features (n), minus one

Source: Chapter 4 Identifying clusters
The ESRI Guide to GIS Analysis, Volume 2

Neighborhood Definition

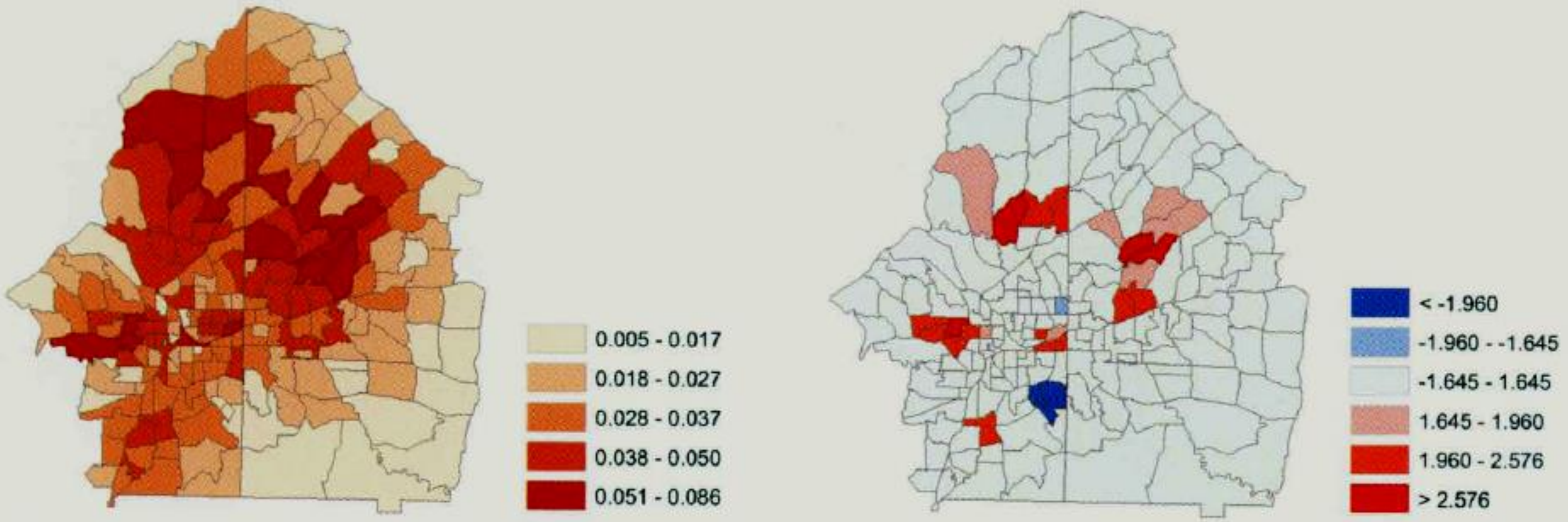
- Distance-based neighborhood.



Clusters of ZIP Codes having high numbers of people more likely (orange) or less likely (blue) to buy pet supplies. Using a distance of five miles (left map), the clusters are smaller and more localized. Using a distance of 20 miles (right) creates larger, regional clusters.

Source: Chapter 4 Identifying clusters
The ESRI Guide to GIS Analysis, Volume 2

Mapping the result of Gi* values



Census tracts color coded by Gi values (left) and Z-scores, calculated from percent age 65 and over*

Source: Chapter 4 Identifying clusters
The ESRI Guide to GIS Analysis, Volume 2

R Lab: Using R packages to calculate Gi*

using **localG()** function

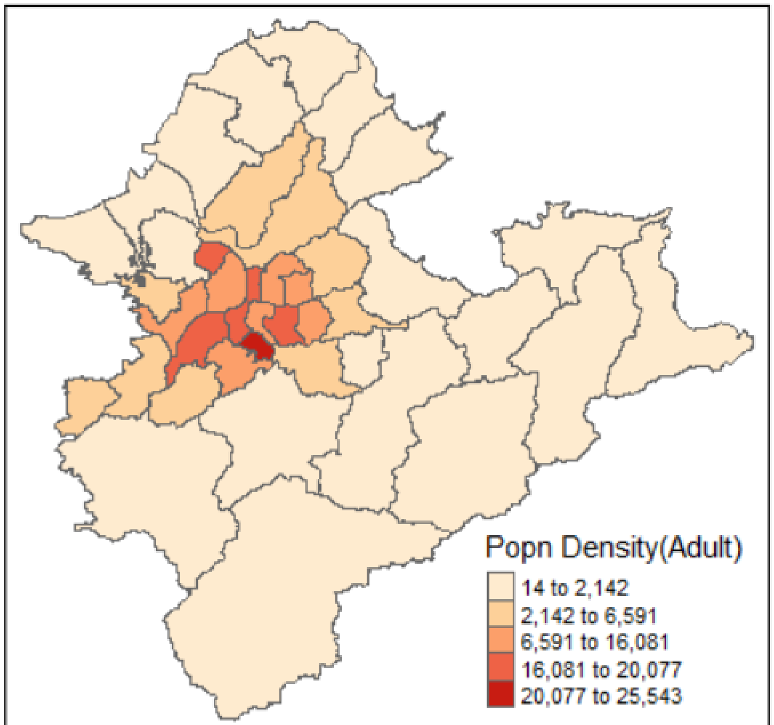
```
TWN_nb_in <- include.self(TWN_nb)

TWN_nb_in_w <- nb2listw(TWN_nb_in, zero.policy=T)

LG <- localG(Density, TWN_nb_in_w)
```

Standardized Gi* values

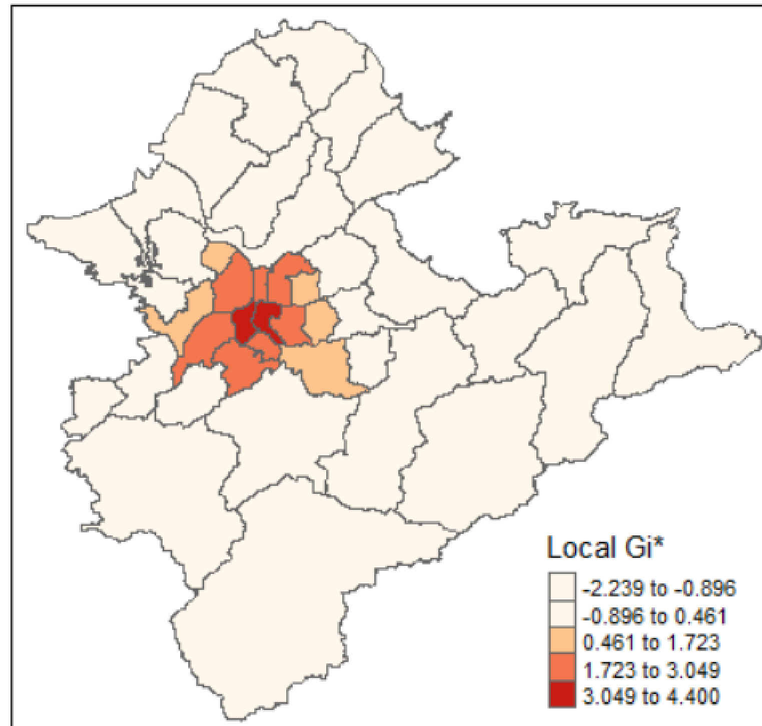
```
> LG
 [1]  1.7233446  1.5358246  2.2955770  2.2575385  3.5680055  2.4243832  4.4002452  1.5935843 -0.2011367  0.1873412
[11]  0.2039434 -1.7879950  2.6015773  3.0486377  2.4917352  2.9034051  1.1389338 -0.8962848  0.4607859 -0.9741356
[21] -1.4210311 -1.4498460 -1.6708723 -1.8162000  0.4273101  1.0063072  0.4332790 -0.3465971 -1.3485541 -1.0850525
[31] -2.2394723 -2.0938802 -1.7500732 -1.5259770 -1.6278155 -2.1806525 -2.0660263 -1.5440033 -1.9083605 -1.4421784
[41] -1.9553767
attr(,"gstari")
[1] TRUE
attr(,"call")
localG(x = Density, listw = TWN_nb_in_w)
attr(,"class")
[1] "localG"
```

R Lab: Mapping Standardized Gi* values

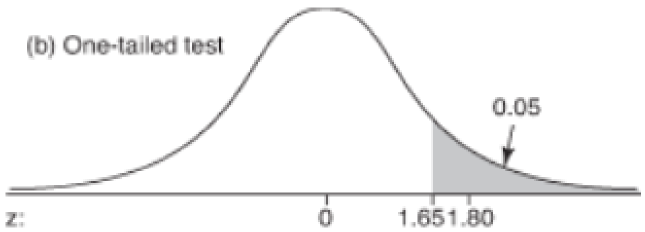
Population



Standardized Gi* values

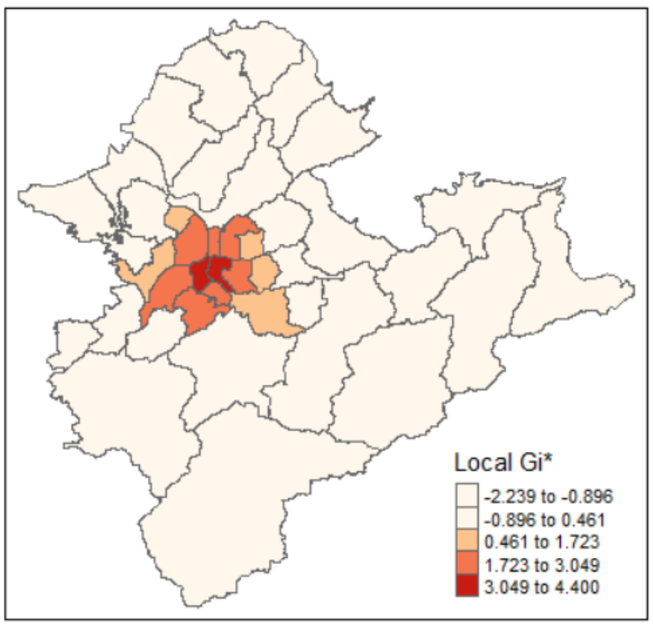


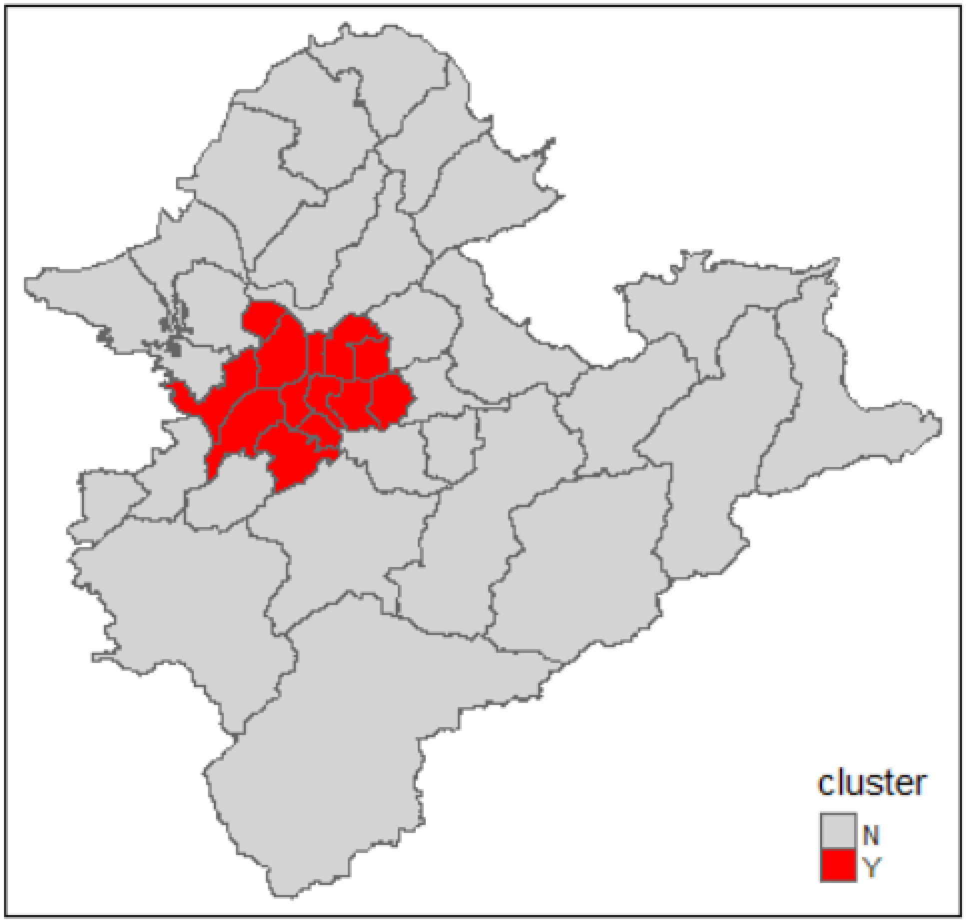
R Lab: Mapping Significant Hot-spots



alpha = 0.05

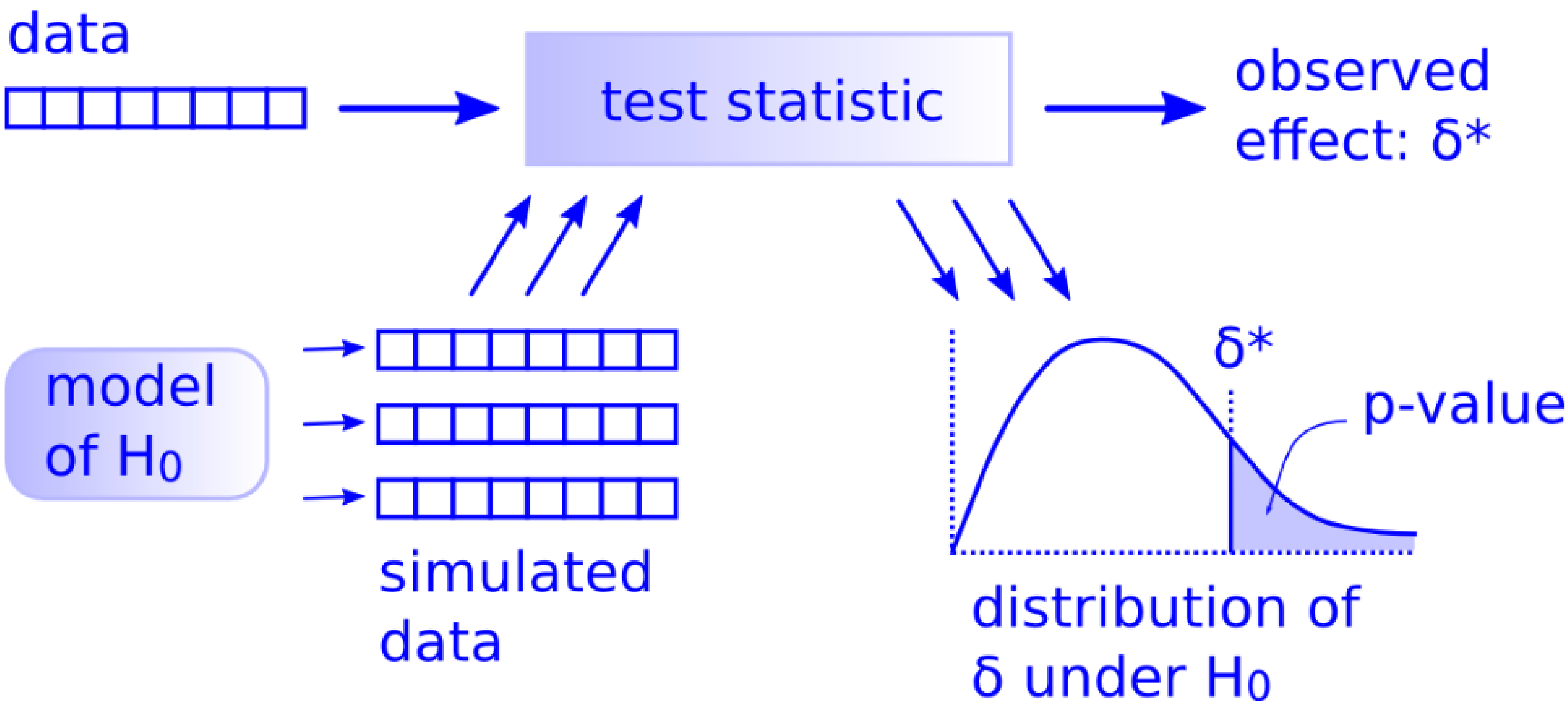
Standardized Gi* values





熱區的統計顯著性校正

Controlling the False Discovery Rate (FDR)



Recap: Type I Error and Significance Level

- **Type I error** (false positive 偽陽性):

  the incorrect rejection of a true null hypothesis

- **Significance level** (or rate of Type I error):

  the probability of rejecting the null hypothesis

  given that it is true.

Recap: Global vs. Local Measures in Spatial Analysis

- **Global measures** consider all available locations simultaneously, utilizing a single statistic that summarizes the spatial pattern.
- **Local measures** represent the association between each location and its neighbors based on defined distances.

  One statistic is provided for each location, facilitating the identification of clusters, testing of stationarity assumptions, and inference about distances over which spatial association occurs.

Statistical Issues for Local Measures

- Local statistics rely on tests of spatial association for each location in the data, and the issues of multiple comparisons and spatial dependency are the concerns when assessing their **significance**

Multiple Comparisons (or Multiple Testing)

Setting the significance level = 0.05 (rate of Type I Error)

REAL

A total of polygons (N) = 1050

TRUE Random = 1000 (null)

TRUE Clusters = 50 (reject null)

Spatial Statistical Results

A total of polygons (N) = 1050

Accept null hypothesis (n1 = 950):

Random = 1000 * 0.95 = 950

Reject null hypothesis (n2 = 100):

Clusters (Type I error) = 1000 * 0.05 = 50

TRUE Clusters = 50 (true positive)

The local analysis identifies 100 polygons as clusters!

However, **HALF** of them are WRONG !!!

How to make sure the rate of Type I Error = 0.05

$$0.05 = 1 - (1 - \alpha)^{n}$$

$$= 1 - (1 - \alpha)^{100}$$

$$\alpha = 0.000153$$

Adjusting for multiple comparisons (cont'd)

Sidak Correction (1967, 1968, 1971)

The Sidak correction controls for **the overall probability of type I error,** but with critical values appraised at a level $1-(1-\alpha)^{1/n}$. Therefore, a test is considered significant when $p \leqq p_{critical} = 1-(1-\alpha)^{1/n}$

Weakness: usually produce **conservative** results.

Adjusting for multiple comparisons

The Bonferroni method

It evaluates the significance of the test statistics at a critical probability value ($p_{critical}$) set equal to α/n, where α is the overall type I error rate for the data.

All test statistics whose probability values (p) satisfy the condition $p \leqq p_{critical} = \alpha / n = p_{BON}$ are considered significant (null hypothesis is rejected)

Weakness: usually produce **conservative** results.

Rationale for alternative approach

- **Ignoring the issue of multiple testing** would imply spending a large amount of human and financial resources *unnecessarily and inefficiently*.
- **The extremely conservative methods** would result in a *major failure* to curb the spread of the disease (or crime events).

Adjusting for multiple comparisons: The false discovery rate (FDR)

Benjamini and Hochberg (1995)

step-by-step procedure:

Assume that there are *m* hypotheses to be tested

(1) order the test statistics p-values (p_i)

in ascending order ($p_1 \leqq p_2 \leqq . . . \leqq p_m$);

(2) starting from p_m find the first p_i for which $p_i \leqq p_{critical} = (i/m)\, \alpha$;

(3) regard all tests as significant for which $p_i \leqq p_{critical} = (i/m)\, \alpha = p_{FDR}$.

Comparisons of different corrections

Year	Unadjusted	Correcting for multiplicity*			Correcting for multiplicity and spatial dependence[†]		Recovery ratio[‡]
		Bonferroni	Sidak	FDR	Bonferroni	Sidak	
1987 (N = 740)							
$G_i^*(d)$							
Accept null	387	665	664	436	650	649	
Reject null	353	75	76	304	90	91	
Cluster high rates	142	20	21	116	28	28	0.824
Cluster low rates	211	55	55	188	62	63	
$p_{critical}$	0.025	0.0000338	0.0000342	0.0100272	0.0000616	0.0000624	
$z_{critical}$	± 1.95996	± 3.98469	± 3.98169	± 2.32533	± 3.83958	± 3.83648	
$G_i(d)$							
Accept null	392	675	675	437	659	659	
Reject null	348	65	65	303	81	81	
Cluster high rates	136	16	16	116	21	21	0.841
Cluster low rates	212	49	49	187	60	60	
$p_{critical}$	0.025	0.0000338	0.0000342	0.0101780	0.0000616	0.0000624	
$z_{critical}$	± 1.95996	± 3.98469	± 3.98169	± 2.31972	± 3.83958	± 3.83648	
Moran's I_i							
Accept null	439	740	740	479	740	740	
Reject null	301	0	0	261	0	0	
Cluster high rates	100	0	0	83	0	0	0.867
Cluster low rates	201	0	0	178	0	0	
$p_{critical}$	0.05	0.0000676	0.0000693	0.0220000	0.0001232	0.0001264	
$z_{critical}$	1.64485	3.81691	3.81061	2.01409	3.66588	3.65936	

Comparisons of different corrections (cont'd)

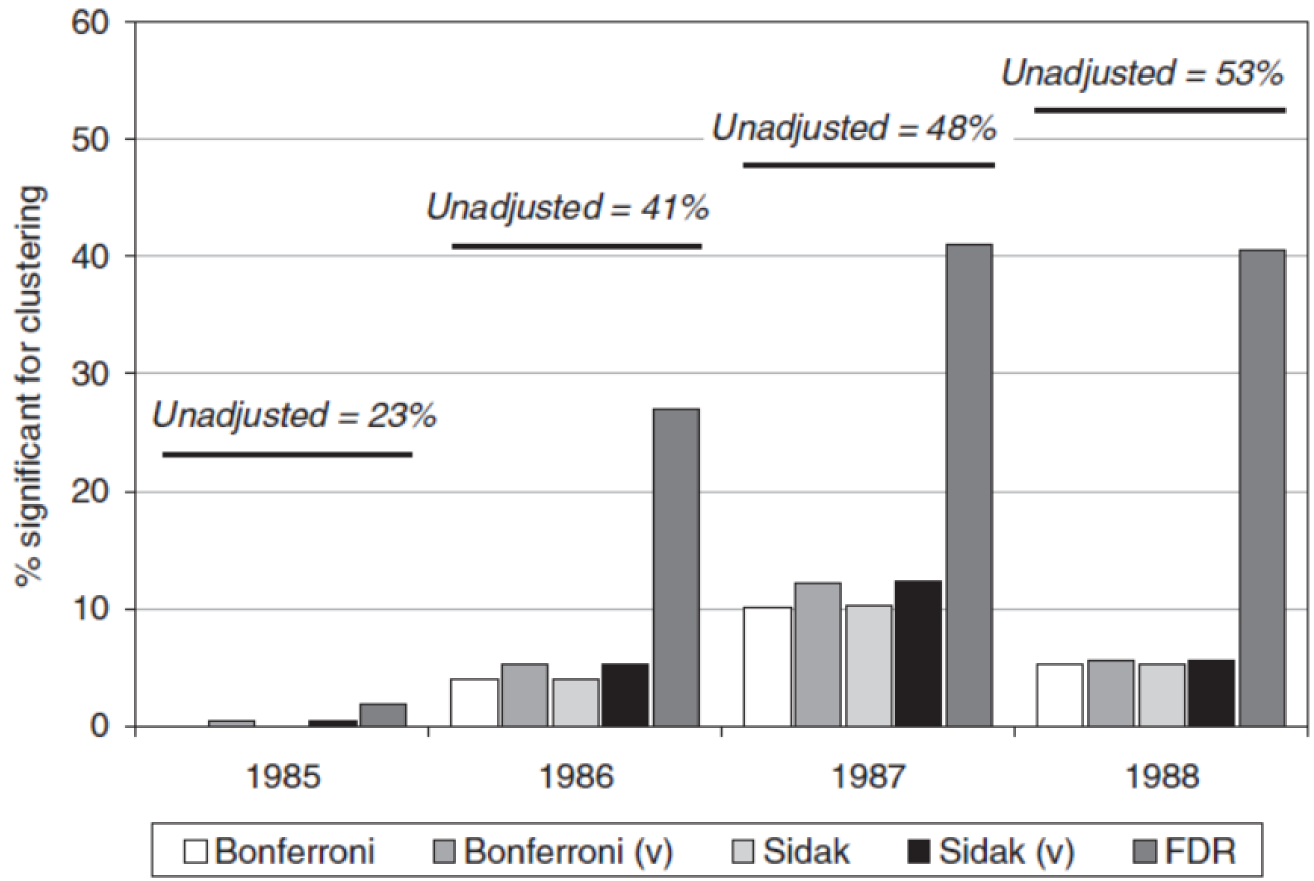
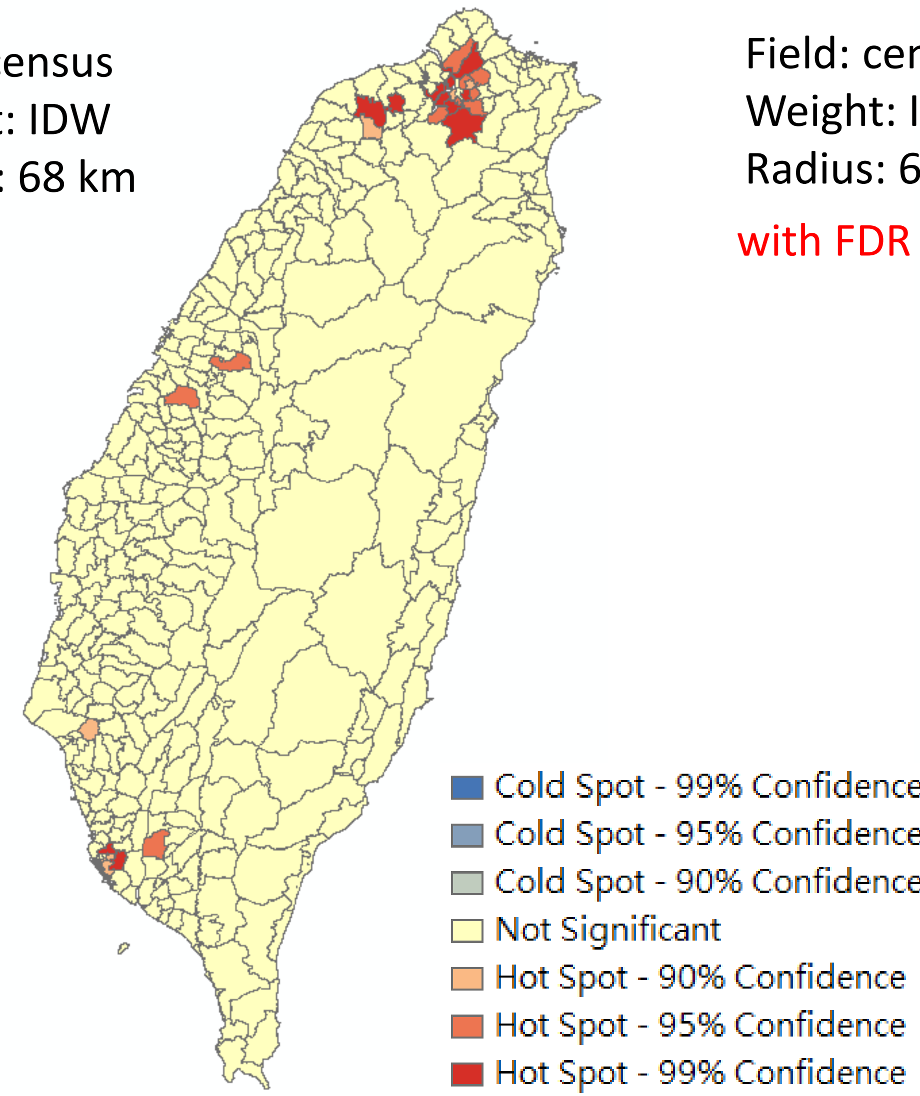


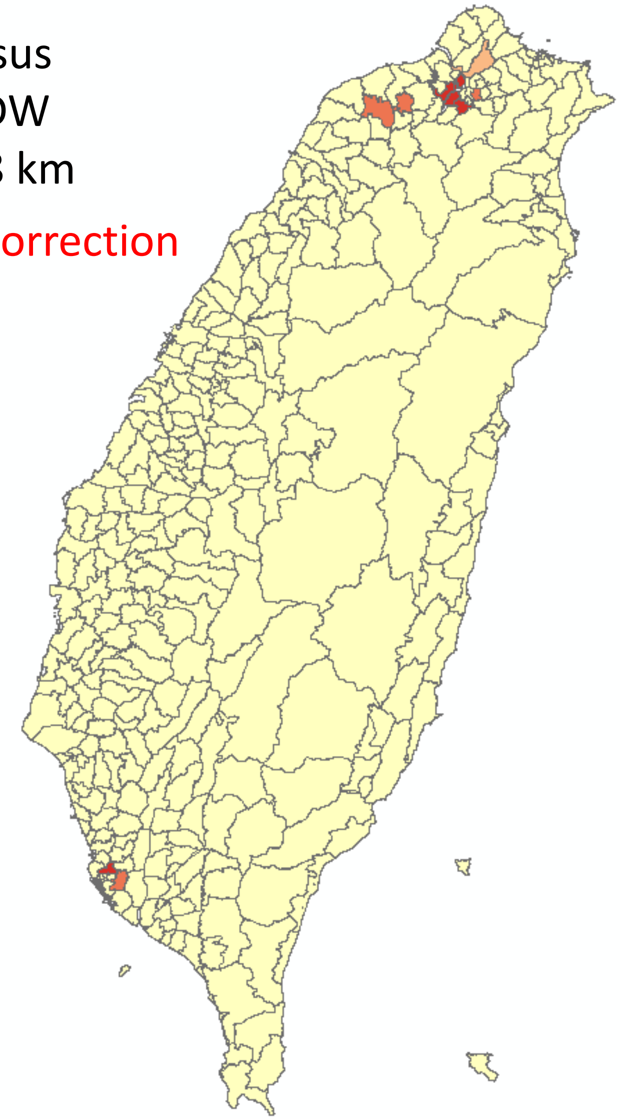
Figure 2. Percentage of plots that tested significant for clustering, according to the $G_i^*(d)$ statistic and different control procedures—Machadinho (1985/95).

Result: the hotspots of population using Gi*(d)

Field: census
Weight: IDW
Radius: 68 km



Field: census
Weight: IDW
Radius: 68 km
with FDR correction



Hotspots with/without FDR correction

with FDR correction

Gi_IDW

OID	OBJECTID	SOURC	CENSUS	GiZScore	GiPValue	Gi_Bin
55	56	55	516468	6.020918	0	3
46	47	46	401619	4.501931	0.000007	3
43	44	43	384051	4.269309	0.00002	3
63	64	63	376584	4.170142	0.00003	3
228	229	228	357536	3.912712	0.000091	3
220	221	220	338361	3.661243	0.000251	3
218	219	218	329913	3.548686	0.000387	3
262	263	262	322678	3.451571	0.000557	3
31	32	31	315818	3.367603	0.000758	3
29	30	29	292096	3.050694	0.002283	3
62	63	62	272500	2.790864	0.005257	3
34	35	34	254521	2.553956	0.010651	2
35	36	35	253920	2.546308	0.010887	2
36	37	36	247904	2.465707	0.013674	2
38	39	38	237530	2.331119	0.019747	2
45	46	45	231938	2.255883	0.024078	2
352	353	352	231129	2.239411	0.025129	2
49	50	49	229383	2.225159	0.026071	2
1	2	1	221815	2.116672	0.034288	2
32	33	32	216043	2.047546	0.040604	2
179	180	179	215245	2.028891	0.042469	2
37	38	37	205031	1.901785	0.057199	1
40	41	40	204024	1.890186	0.058733	1
232	233	232	203001	1.868696	0.061665	1
233	234	233	199535	1.823772	0.068186	1
103	104	103	198372	1.805552	0.070988	1
219	220	219	194521	1.757798	0.078782	1
93	94	93	185752	1.638691	0.101278	0
294	295	294	177796	1.534813	0.12483	0

< 0.01

< 0.05

< 0.1

Gi_IDW_FDR

OID	OBJECTID	SOURCE_ID	CENSUS	GiZScore	GiPValue	Gi_Bin
55	56	55	516468	6.020918	0	3
46	47	46	401619	4.501931	0.000007	3
43	44	43	384051	4.269309	0.00002	3
63	64	63	376584	4.170142	0.00003	3
228	229	228	357536	3.912712	0.000091	3
220	221	220	338361	3.661243	0.000251	2
218	219	218	329913	3.548686	0.000387	2
262	263	262	322678	3.451571	0.000557	2
31	32	31	315818	3.367603	0.000758	2
29	30	29	292096	3.050694	0.002283	1
62	63	62	272500	2.790864	0.005257	0
34	35	34	254521	2.553956	0.010651	0
35	36	35	253920	2.546308	0.010887	0
36	37	36	247904	2.465707	0.013674	0
38	39	38	237530	2.331119	0.019747	0
45	46	45	231938	2.255883	0.024078	0
352	353	352	231129	2.239411	0.025129	0
49	50	49	229383	2.225159	0.026071	0
1	2	1	221815	2.116672	0.034288	0
32	33	32	216043	2.047546	0.040604	0
179	180	179	215245	2.028891	0.042469	0
37	38	37	205031	1.901785	0.057199	0
40	41	40	204024	1.890186	0.058733	0
232	233	232	203001	1.868696	0.061665	0
233	234	233	199535	1.823772	0.068186	0
103	104	103	198372	1.805552	0.070988	0
219	220	219	194521	1.757798	0.078782	0
93	94	93	185752	1.638691	0.101278	0
294	295	294	177796	1.534813	0.12483	0

FDR correction

$$p_i \leq p_{\text{critical}} = (i/m)\alpha = p_{\text{FDR}}.$$

	A	B	C	F	G	H	I	J	K	L	M	N	O
1	OBJECTID	SOURCE_	**CENSUS**	**GiZScore**	**GiPValue**	**Gi_Bin**	Rank	**P_FDR_0.1**	**P_FDR_0.05**	**P_FDR_0.01**	**Sig_0.1**	**Sig_0.05**	**Sig_0.01**
2	56	55	516468	6.021	**0.000000002**	**3**	1	0.000277008	0.000138504	2.77008E-05	Sig	Sig	**Sig**
3	47	46	401619	4.502	**0.000006734**	**3**	2	0.000554017	0.000277008	5.54017E-05	Sig	Sig	**Sig**
4	44	43	384051	4.269	**0.000019608**	**3**	3	0.000831025	0.000415512	8.31025E-05	Sig	Sig	**Sig**
5	64	63	376584	4.170	**0.000030441**	**3**	4	0.001108033	0.000554017	0.000110803	Sig	Sig	**Sig**
6	229	228	357536	3.913	**0.000091265**	**3**	5	0.001385042	0.000692521	0.000138504	Sig	Sig	**Sig**
7	221	220	338361	3.661	**0.000250994**	**3**	6	0.00166205	0.000831025	0.000166205	Sig	**Sig**	FALSE
8	219	218	329913	3.549	**0.000387159**	**3**	7	0.001939058	0.000969529	0.000193906	Sig	**Sig**	FALSE
9	263	262	322678	3.452	**0.000557333**	**3**	8	0.002216066	0.001108033	0.000221607	Sig	**Sig**	FALSE
10	32	31	315818	3.368	**0.000758246**	**3**	9	0.002493075	0.001246537	0.000249307	Sig	**Sig**	FALSE
11	30	29	292096	3.051	**0.002283134**	**3**	10	0.002770083	0.001385042	0.000277008	**Sig**	FALSE	FALSE
12	63	62	272500	2.791	**0.005256756**	**3**	11	0.003047091	0.001523546	0.000304709	FALSE	FALSE	FALSE
13	35	34	254521	2.554	**0.010650667**	**2**	12	0.0033241	0.00166205	0.00033241	FALSE	FALSE	FALSE
14	36	35	253920	2.546	**0.010886894**	**2**	13	0.003601108	0.001800554	0.000360111	FALSE	FALSE	FALSE
15	37	36	247904	2.466	**0.013674312**	**2**	14	0.003878116	0.001939058	0.000387812	FALSE	FALSE	FALSE
16	39	38	237530	2.331	**0.019747089**	**2**	15	0.004155125	0.002077562	0.000415512	FALSE	FALSE	FALSE
17	46	45	231938	2.256	**0.024077970**	**2**	16	0.004432133	0.002216066	0.000443213	FALSE	FALSE	FALSE
18	353	352	231129	2.239	**0.025129188**	**2**	17	0.004709141	0.002354571	0.000470914	FALSE	FALSE	FALSE
19	50	49	229383	2.225	**0.026070581**	**2**	18	0.00498615	0.002493075	0.000498615	FALSE	FALSE	FALSE
20	2	1	221815	2.117	**0.034287671**	**2**	19	0.005263158	0.002631579	0.000526316	FALSE	FALSE	FALSE
21	33	32	216043	2.048	**0.040604468**	**2**	20	0.005540166	0.002770083	0.000554017	FALSE	FALSE	FALSE
22	180	179	215245	2.029	**0.042469389**	**2**	21	0.005817175	0.002908587	0.000581717	FALSE	FALSE	FALSE
23	38	37	205031	1.902	**0.057199259**	**1**	22	0.006094183	0.003047091	0.000609418	FALSE	FALSE	FALSE
24	41	40	204024	1.890	**0.058733043**	**1**	23	0.006371191	0.003185596	0.000637119	FALSE	FALSE	FALSE
25	233	232	203001	1.869	**0.061665141**	**1**	24	0.006648199	0.0033241	0.00066482	FALSE	FALSE	FALSE
26	234	233	199535	1.824	**0.068186495**	**1**	25	0.006925208	0.003462604	0.000692521	FALSE	FALSE	FALSE
27	104	103	198372	1.806	**0.070988331**	**1**	26	0.007202216	0.003601108	0.000720222	FALSE	FALSE	FALSE
28	220	219	194521	1.758	**0.078781940**	**1**	27	0.007479224	0.003739612	0.000747922	FALSE	FALSE	FALSE
29	94	93	185752	1.639	0.101277593	0	28	0.007756233	0.003878116	1.07427E-06	FALSE	FALSE	FALSE
30	295	294	177796	1.535	0.124829823	0	29	0.008033241	0.00401662	1.11264E-06	FALSE	FALSE	FALSE

Using p.adjust() function

p.adjust

Adjust P-Values For Multiple Comparisons

Given a set of p-values, returns p-values adjusted using one of several methods.

Keywords htest

Usage

```
p.adjust(p, method = p.adjust.methods, n = length(p))

p.adjust.methods
# c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY",
#   "fdr", "none")
```

Arguments

p numeric vector of p-values (possibly with `NA` s). Any other R object is coerced by `as.numeric` .

method correction method. Can be abbreviated.

Mapping p-values (localmoran)

```
> LISA.Popn <- localmoran(Density, TWN_nb_w, zero.policy=T)
> LISA.Popn
             Ii   E.Ii     Var.Ii       Z.Ii  Pr(z > 0)
221 0.668654386 -0.025 0.17429750 1.66148935 4.830760e-02
222 0.552267439 -0.025 0.22386865 1.22005786 1.112215e-01
223 1.122969659 -0.025 0.17429750 2.74969696 2.982520e-03
224 0.520540764 -0.025 0.14125006 1.45155294 7.331298e-02
225 1.283663684 -0.025 0.14125006 3.48203973 2.488049e-04
226 1.275561091 -0.025 0.17429750 3.11519460 9.191180e-04
227 2.310784759 -0.025 0.14125006 6.21496221 2.566850e-10
```

Mapping p-values (FDR correction)

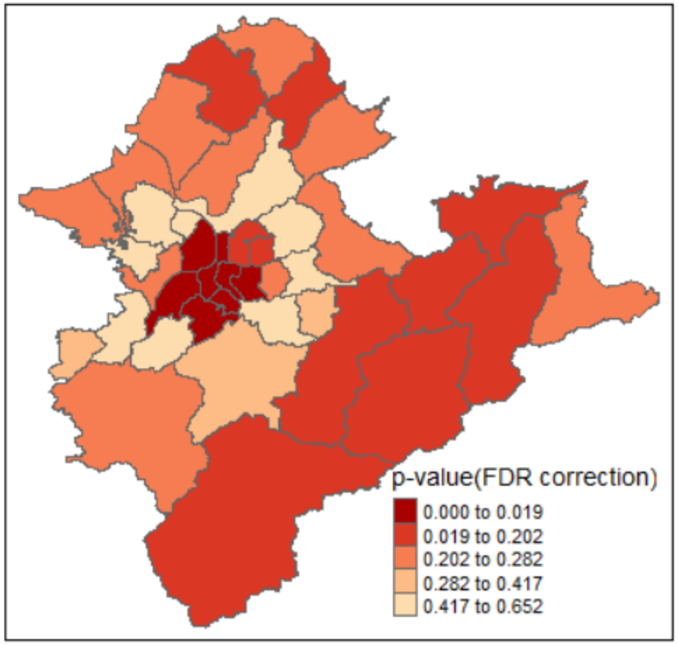
LISA.Popn <- localmoran(Density, TWN_nb_w, zero.policy=T)

NorthTW_sf$pvalue.adj <- p.adjust(LISA.Popn[,5], method="fdr")

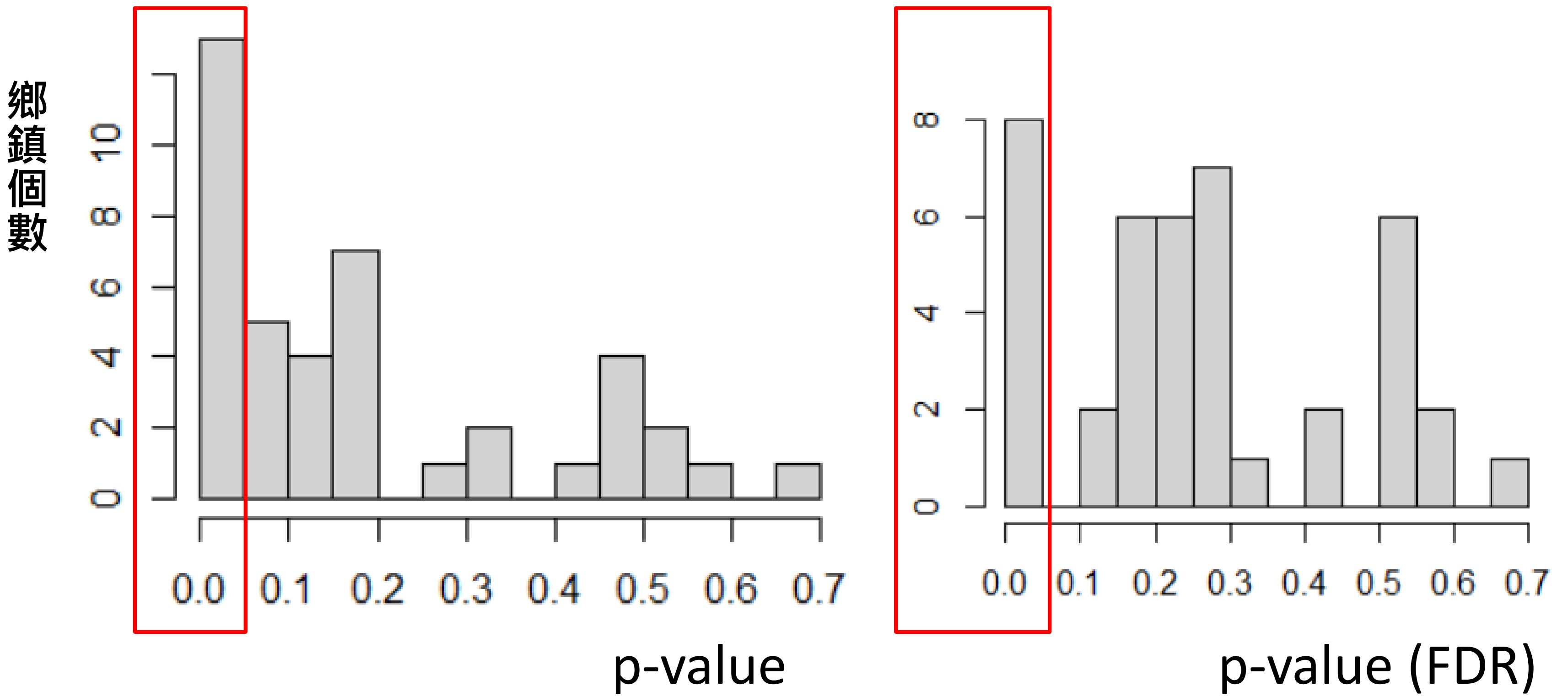
pvalue = (i/41) x pvalue.adj

→ pvalue * (41/i) = pvalue.adj

```
> min(NorthTW_sf$pvalue.adj)
[1] 1.052409e-08
> min(NorthTW_sf$pvalue) * 41
[1] 1.052409e-08

> max(NorthTW_sf$pvalue.adj)
[1] 0.6524963
> max(NorthTW_sf$pvalue)
[1] 0.6524963
```



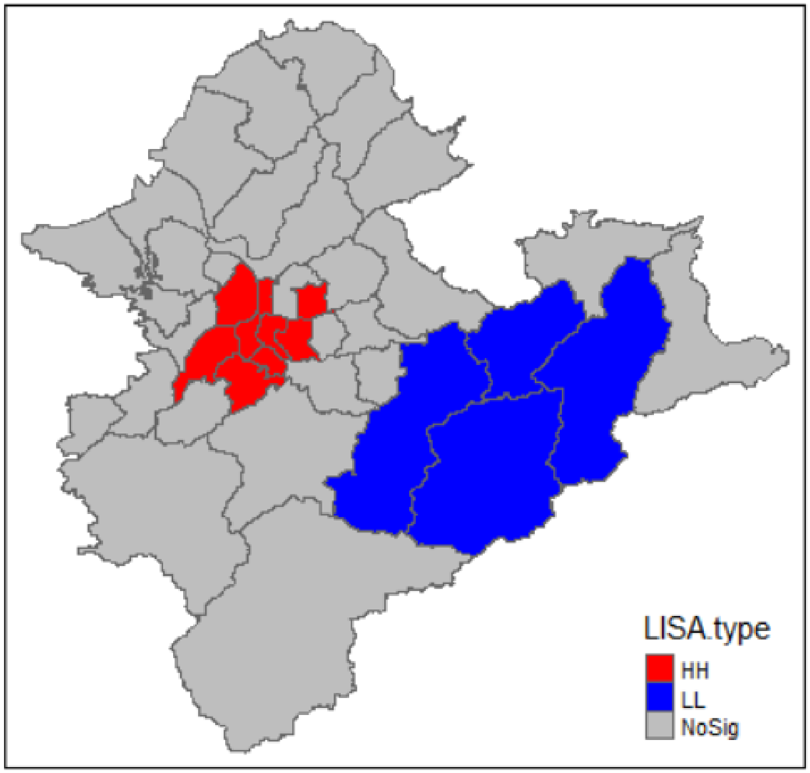
Comparisons of p-values before/after correction



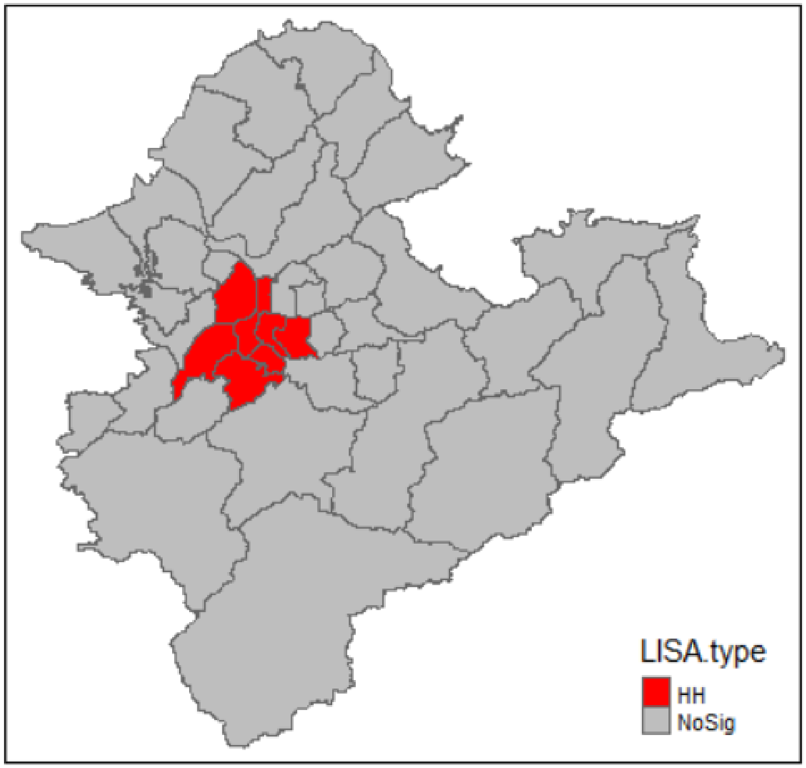
Mapping LISA Maps

alpha = 0.05

LISA map



LISA map (FDR correction)



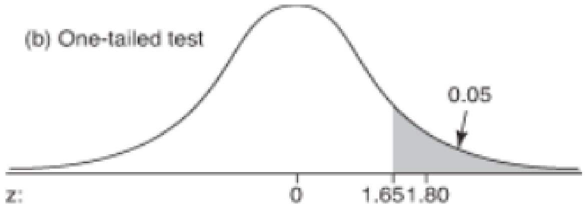
Using localg() function for adjustment

spdep (version 1.1-7)

localG: G and Gstar local spatial statistics

Description

The local spatial statistic G is calculated for each zone based on the spatial weights object used. The value returned is a Z-value, and may be used as a diagnostic tool. High positive values indicate the posibility of a local cluster of high values of the variable being analysed, very low relative values a similar cluster of low values. For inference, a Bonferroni-type test is suggested in the references, where tables of critical values may be found (see also details below).

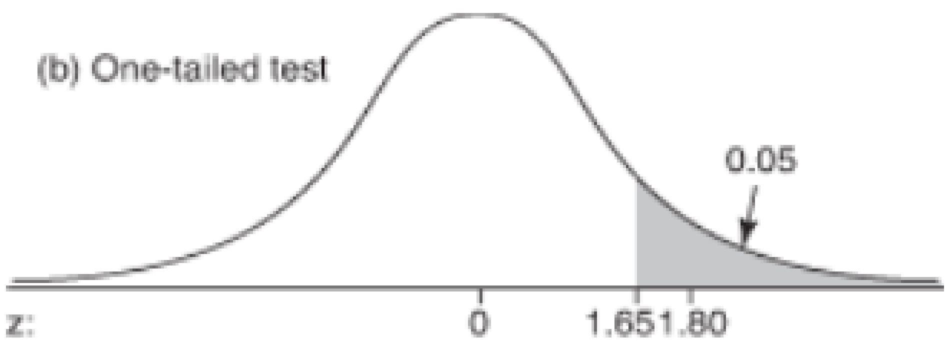


The critical values of the statistic under assumptions given in the references for the 95th percentile are for n=1: 1.645, n=50: 3.083, n=100: 3.289, n=1000: 3.886.

The Bonferroni correction method

The critical values of the statistic under assumptions given in the references for the 95th percentile are for n=1: 1.645, n=50: 3.083, n=100: 3.289, n=1000: 3.886.

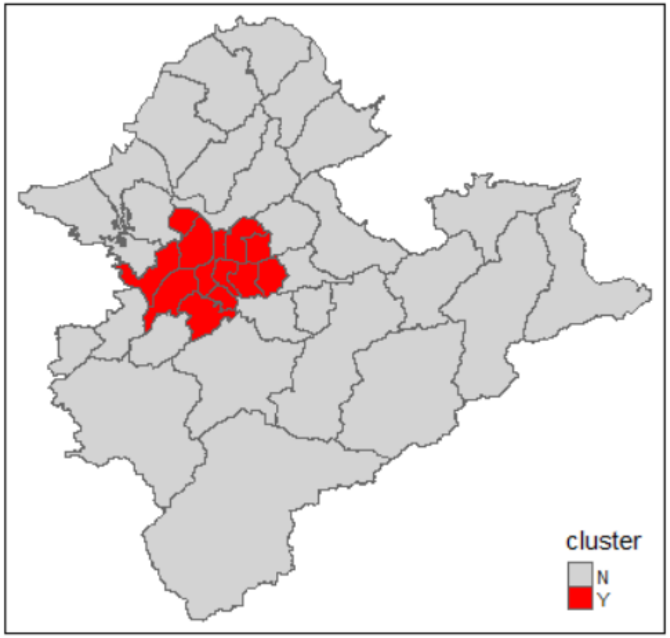
```
> qnorm(1-0.05, 0, 1)
[1] 1.644854
> qnorm(1-0.05/50, 0, 1)
[1] 3.090232
> qnorm(1-0.05/100, 0, 1)
[1] 3.290527
> qnorm(1-0.05/1000, 0, 1)
[1] 3.890592
```



Mapping Significant Hot-spots (Gi*)

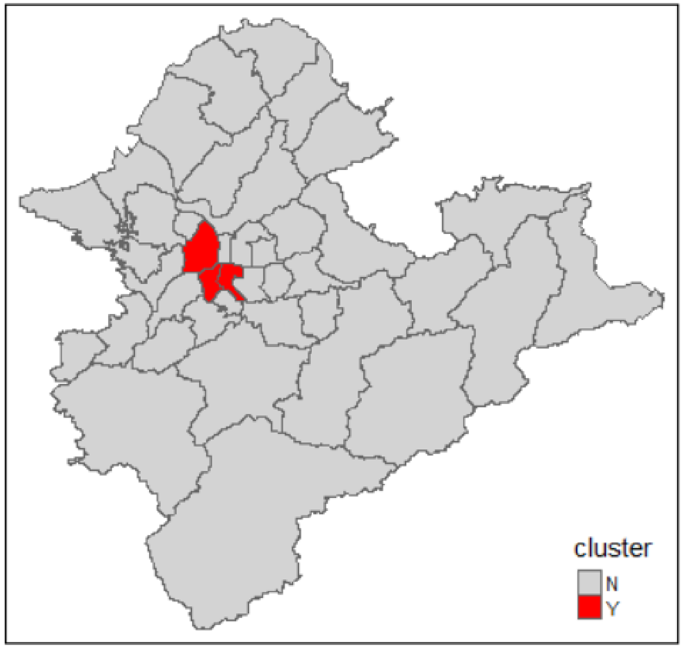
alpha = 0.05

$Z^* = 1.65$



Bonferroni correction
alpha = 0.05/41

$Z^* = 3.03$



本週實習

資料: Popn_TWN2.shp

- 定義: 高齡人口密度 (> 65 / total)
- 繪製台灣鄉鎮高齡人口比例的主題地圖:
  - 原始數值
  - LISA map (p-value < 0.05) (區分 HH, HL, LH, LL)
  - Standardized Gi * values (p-value < 0.05)
    (區分 cluster, non-cluster)
  - 比較LISA進行FDR校正前後的H-H熱區分布
  - 比較Gi*進行Bonferroni校正前後的熱區分布