空間自相關
Spatial Autocorrelation

授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw

期末書面報告(總成績10%)：相關規定說明

- 以「組」為單位繳交，同一組別將評定相同分數。
- 繳交期限：**6/03 (週一), 2:00pm**
- 繳交格式：1-2頁(A4)，PDF格式 + R Markdown file (.html)
- 繳交內容，應包括以下項目：

  1. 動機與問題、資料說明(來源)、分析方法、結果與討論
  2. 分析過程與結果：R Markdown動態檔案 (.html)

- 書面報告的參考範例：

  http://homepage.ntu.edu.tw/~r07228005/1072SA/final_demo.pdf

- 入選標準：分析方法的正確運用與解讀，呈現方式不限，

  能引起老師與助教的注意，想聽到口頭簡報

入選組別的期末口頭報告

- 獎勵對象：由老師和助教挑選書面報告較佳的(至多) 7組
- 公告入選組別：6/04 (週二) 24:00pm前
- 上傳簡報資料：6/10 (週一) 2:00pm前
- 進行口頭簡報：6/10 (週一) 上課時間
- 每組口頭報告 15 min，形式不限。
- 由老師/助教 (50%)、學長姐 (30%) 與修課同學 (20%) 共同評分
- 個人的評論文字意見內容，作為**該週的實習成績**

入選組別的獎勵方式

- **第1名**：學期成績 A+
- **第2名**：學期成績 A (或期末考 + 80)
- **第3名**：學期成績 A- (或期末考 + 70)
- **入選獎勵 (其餘入選組別)**：期末考 + 60
- **人氣加分獎勵 (所有入選組別)**：期末考 + 20

  6/14 (週五) 6:00pm前，**FB 貼文: 按讚 >60 +分享 >5 + 留言 >3**

  (該篇貼文需tag 助教，且不得出現任何暗示請求按讚等文字)
- **精闢評論獎勵 (未入選組別，至多10位)**：期末考 + 20
- 以上加分獎勵可從缺

本週課程：空間自相關的觀念與計算

Moran's I coefficient

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

N: no. of spatial units

$w_{i,j}$: a matrix of spatial weights

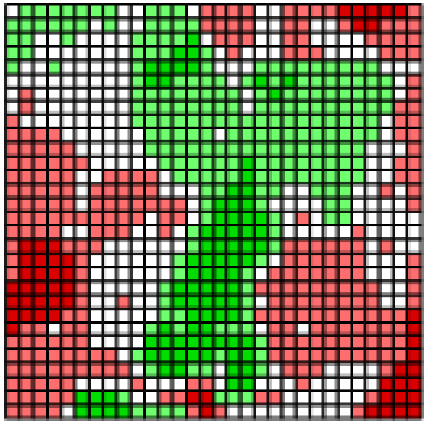
$W = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$ (sum of all $w_{i,j}$)

Global and Local Measures

- ***Global*** Measures
  - A single value which applies to the entire data set
    - The same pattern occurs over the entire geographic area
    - An average for the entire area
- ***Local*** Measures
  - A value calculated for <u>each</u> observation unit
    - Different patterns or processes may occur in different parts of the region
    - A unique number for each location

Global Analysis Methods 全域分析的方法

- **Point data *without* attributes**
  - Quantrat Analysis
  - Nearest Neighbor Methods
    - K-order Nearest Neighborhood Analysis (NNA), G and F functions
  - Ripley's K-function: K(d) and L(d)
- **Point/Polygon data *with* attributes**
  - Definition of Neighborhoods or Spatial Structures
  - Spatial Autocorrelation Index
    - **Moran's I** and Geary's C Ratio
  - Spatial Concentration Index
    - **General G-statistic**

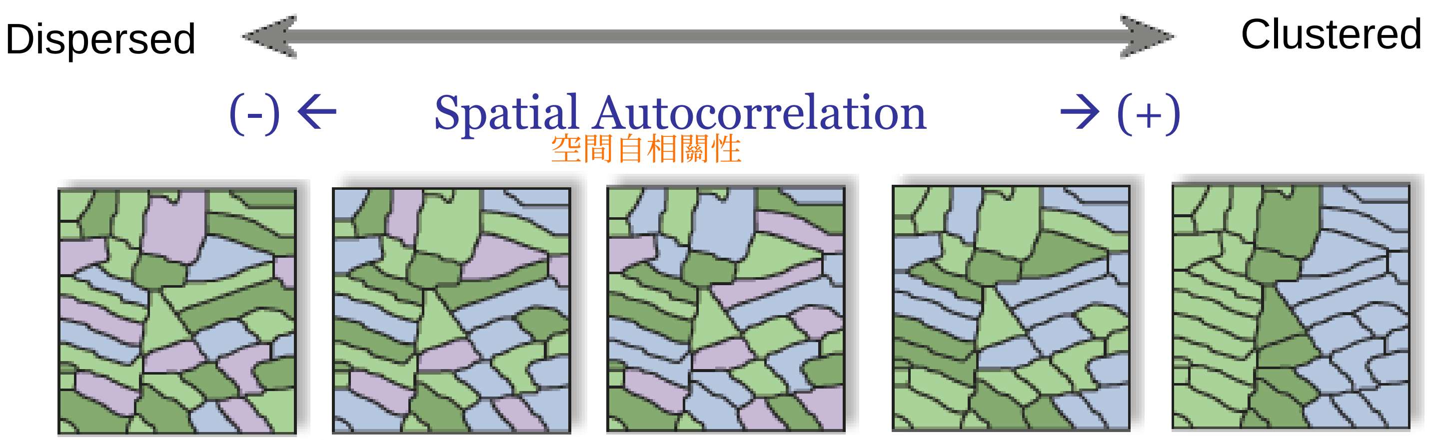


空間關連性與相依性
Spatial Relationship and Dependency

Tobler's *First Law of Geography* (1970)：

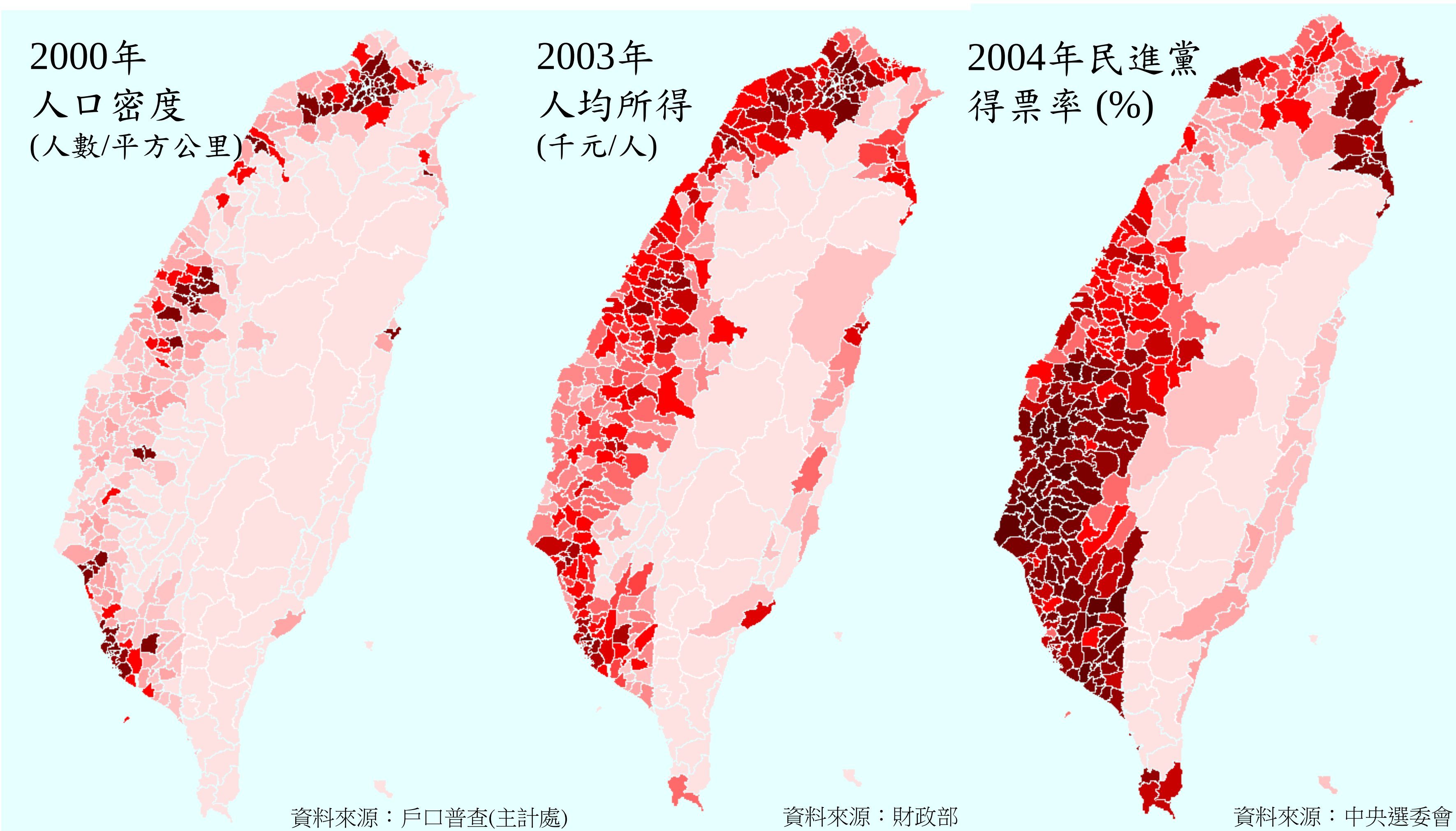
Everything is related to everything else,

but near things are more related than distant things.



Tobler W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2): 234-240.

無所不在的空間自相關（空間相依的特性）

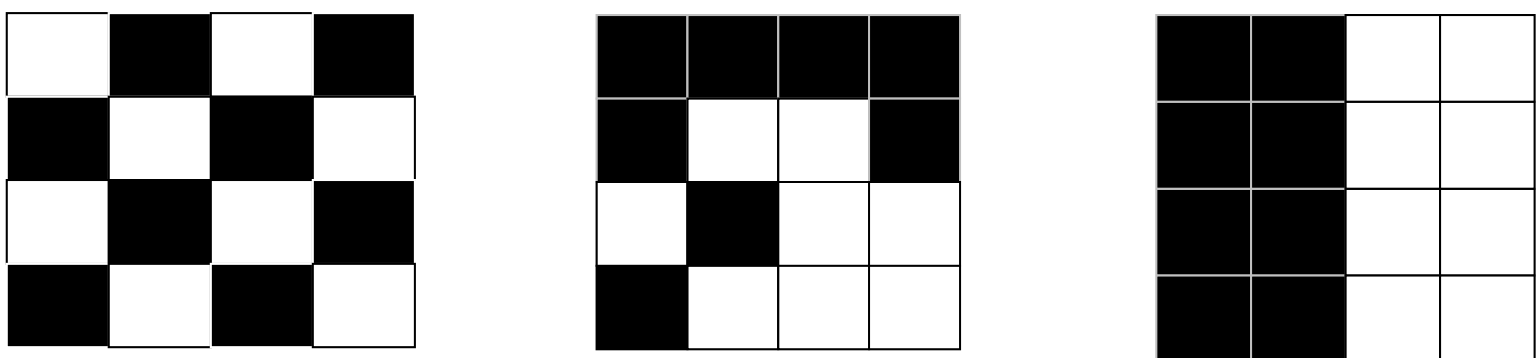


Definition of Spatial Autocorrelation

定義：去量測自己跟鄰居屬性相似性，相似性愈高愈正相關，反之亦然

- Measurement of the ***similarity of attributes*** among spatial units within their ***neighborhood.***

(-) ← Spatial Autocorrelation → (+)



Dispersed *Random* *Clustered*

How we define the Neighborhood ?

1. Spatial adjacency
   - Physically contacted with each others
2. Distances between the centroids

1. Spatial Adjacency

- Rook’s

  判斷誰是鄰居:看跟誰有共同的邊

  - Units that shares common boundary with length greater than zero

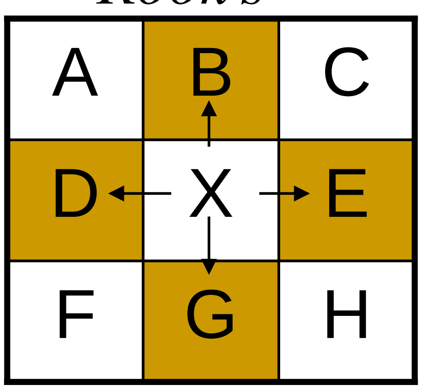
- Queen’s

  判斷鄰居：除了共同的邊之外也看共同的點

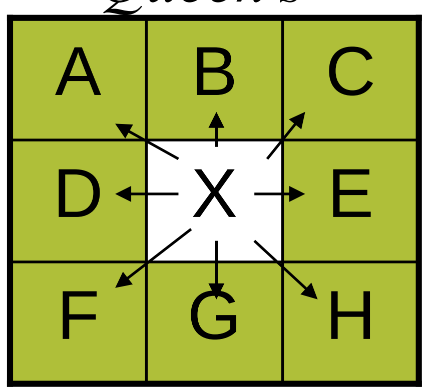
  - Units that have common vertex are also included

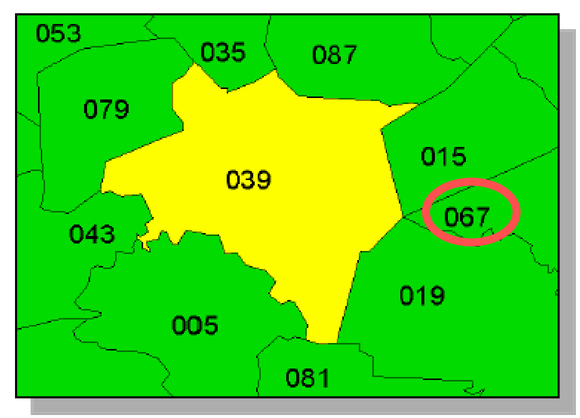
    (e.g. unit 067 in figure below)

Rook’s

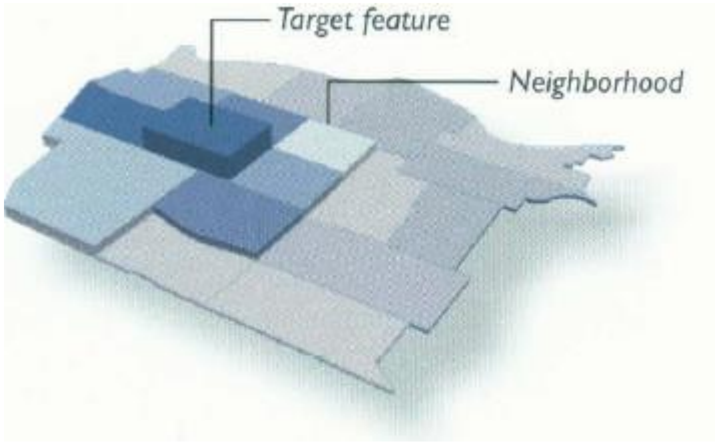


Queen’s

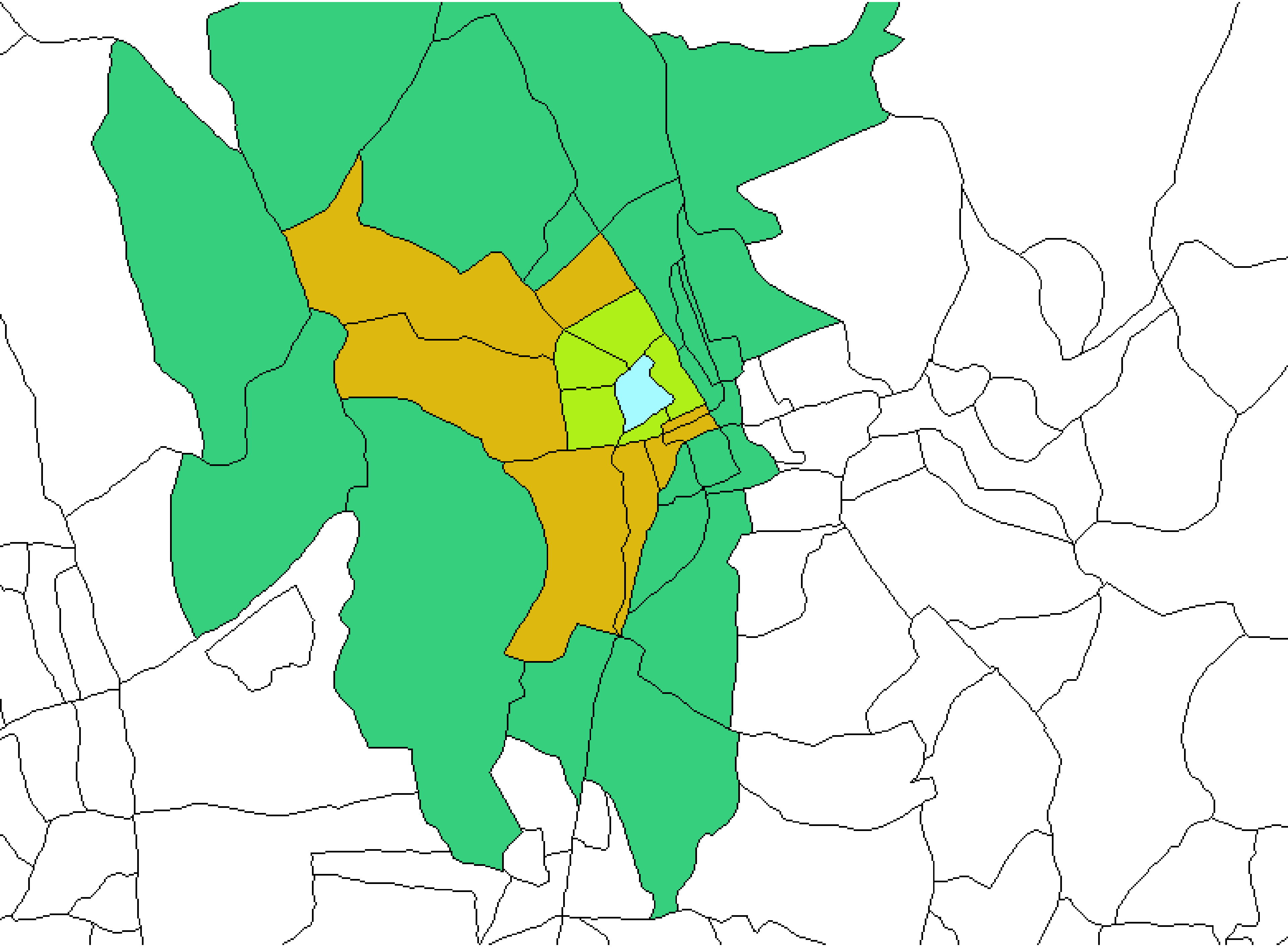




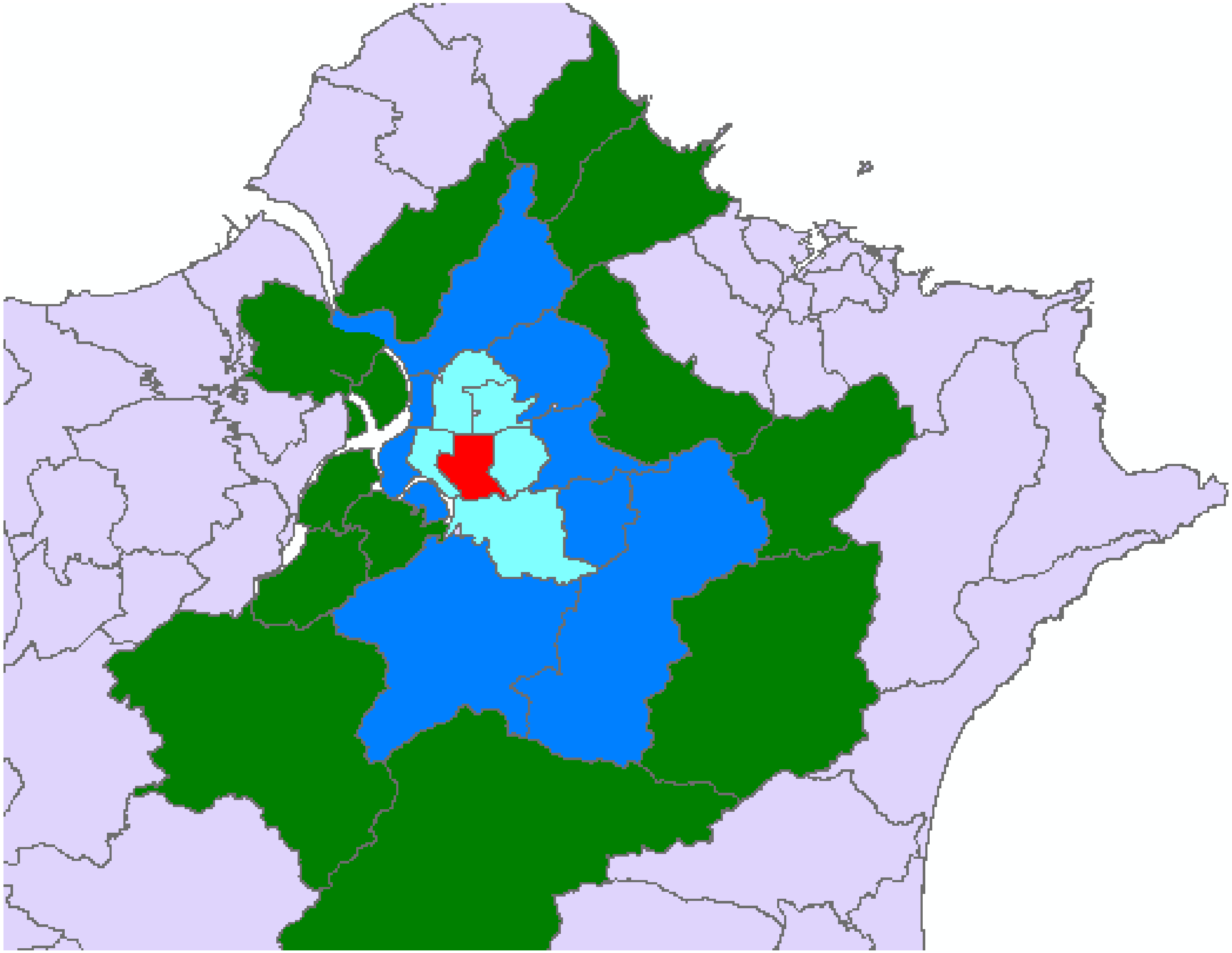
Order of Neighborhood



- 1st order
  - Immediate neighbor
  - Defined by Rook's or Queen's criteria
- 2nd order 鄰居的鄰居
- Higher order 鄰居的鄰居的鄰居的.....*N



台北市大安區的三階鄰近鄉鎮



Binary Connectivity Matrix

- Symmetrical $C_{ij} = C_{ji}$
- Values on diagonal are zeros
- Raw sum $C_i = \Sigma C_{ij}$
  - The number of neighbors of unit I

- Same as connectivity matrix for network
- Not efficient for large numbers of objects
  - Redundant storage
  - Mostly zeros
  - Another way : Sparse Matrix (稀疏矩陣)

Sparse matrix

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	*Sum*
長安里	文化里	中心里	中庸里	豐年里	溫泉里	大同里	清江里								7
中心里	泉源里	開明里	林泉里	中庸里	長名里	溫泉里									6
中庸里	中和里	開明里	文化里	智仁里	中心里	長名里									6
開明里	中和里	泉源里	智仁里	中心里	中庸里										5
文化里	智仁里	中庸里	豐年里	長名里	大同里										5
大同里	文化里	豐年里	長名里	清江里	中央里										5
中和里	泉源里	開明里	智仁里	中庸里											4
泉源里	中和里	開明里	林泉里	中心里											4
智仁里	中和里	開明里	文化里	中庸里											4
溫泉里	林泉里	中心里	長名里	清江里											4
清江里	長名里	溫泉里	大同里	中央里											4
林泉里	泉源里	中心里	溫泉里												3
豐年里	文化里	大同里													3
中央里	大同里	清江里													2

Binary Connectivity Matrix 有相鄰是1，沒相鄰是0

ID	中和里	泉源里	開明里	文化里	智仁里	林泉里	中心里	中庸里	豐年里	長名里	溫泉里	大同里	清江里	中央里	Sum
中和里	0	1	1	0	1	0	0	1	0	0	0	0	0	0	4
泉源里	1	0	1	0	0	1	1	0	0	0	0	0	0	0	4
開明里	1	1	0	0	1	0	1	1	0	0	0	0	0	0	5
文化里	0	0	0	0	1	0	0	1	1	1	0	1	0	0	5
智仁里	1	0	1	1	0	0	0	1	0	0	0	0	0	0	4
林泉里	0	1	0	0	0	0	1	0	0	0	1	0	0	0	3
中心里	0	1	1	0	0	1	0	1	0	1	1	0	0	0	6
中庸里	1	0	1	1	1	0	1	0	0	1	0	0	0	0	6
豐年里	0	0	0	1	0	0	0	0	0	1	0	1	0	0	3
長安里	0	0	0	1	0	0	1	1	1	0	1	1	1	0	7
溫泉里	0	0	0	0	0	1	1	0	0	1	0	0	1	0	4
大同里	0	0	0	1	0	0	0	0	1	1	0	0	1	1	5
清江里	0	0	0	0	0	0	0	0	0	1	1	1	0	1	4
中央里	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2

Sparse matrix 稀疏矩陣的儲存方式

Index						
0	0	0	0	0	0	0
1	0	3	0	0	0	0
2	0	0	0	6	0	0
3	0	0	9	0	0	0
4	0	0	0	0	12	0

Sparse Matrix

5	6	4
1	1	3
2	3	6
3	2	9
4	4	12

Line #1: 這個矩陣是5X6矩陣，非零元素有4個

Line #2：記錄其位置的列索引、行索引與儲存值

Stochastic Matrix

- Equally weighted for neighbors
  - Wij = Cij / Ci
- 又稱做 Row-standardized matrix (列標準化矩陣)
  - 考慮每一個相鄰的object 的影響量

把這個里的相鄰[鄰居個數除以鄰居個數，除以多少其實就是空間的權重，除的越多代表鄰居越多，影響力會被分散的就越多

Binary Connectivity Matrix

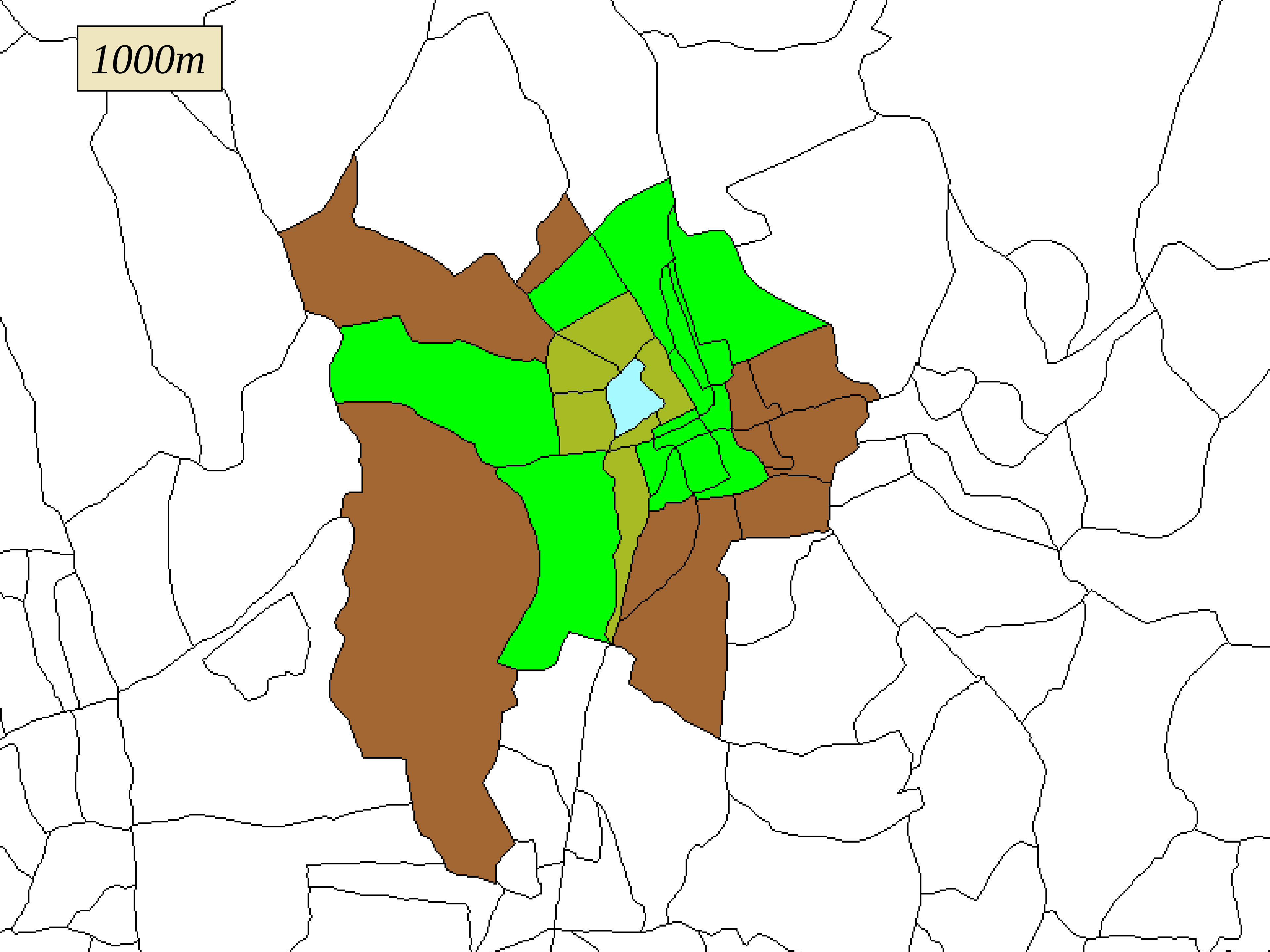
ID	中和里	泉源里	開明里	文化里	智仁里	林泉里	中心里	中庸里	豐年里	長名里	溫泉里	大同里	清江里	中央里	Sum
中和里	0	1	1	0	1	0	0	1	0	0	0	0	0	0	4
泉源里	1	0	1	0	0	1	1	0	0	0	0	0	0	0	4
開明里	1	1	0	0	1	0	1	1	0	0	0	0	0	0	5
文化里	0	0	0	0	1	0	0	1	1	1	0	1	0	0	5
智仁里	1	0	1	1	0	0	0	1	0	0	0	0	0	0	4
林泉里	0	1	0	0	0	0	1	0	0	0	1	0	0	0	3
中心里	0	1	1	0	0	1	0	1	0	1	1	0	0	0	6
中庸里	1	0	1	1	1	0	1	0	0	1	0	0	0	0	6
豐年里	0	0	0	1	0	0	0	0	0	1	0	1	0	0	3
長安里	0	0	0	1	0	0	1	1	1	0	1	1	1	0	7
溫泉里	0	0	0	0	0	1	1	0	0	1	0	0	1	0	4
大同里	0	0	0	1	0	0	0	0	1	1	0	0	1	1	5
清江里	0	0	0	0	0	0	0	0	0	1	1	1	0	1	4
中央里	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2

Stochastic Weighted Matrix

ID	中和里	泉源里	開明里	文化里	智仁里	林泉里	中心里	中庸里	豐年里	長名里	溫泉里	大同里	清江里	中央里
中和里	0.00	0.25	0.25	0.00	0.25	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00
泉源里	0.25	0.00	0.25	0.00	0.00	0.25	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
開明里	0.20	0.20	0.00	0.00	0.20	0.00	0.20	0.20	0.00	0.00	0.00	0.00	0.00	0.00
文化里	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.20	0.20	0.20	0.00	0.20	0.00	0.00
智仁里	0.25	0.00	0.25	0.25	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00
林泉里	0.00	0.33	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.00	0.00	0.00
中心里	0.00	0.17	0.17	0.00	0.00	0.17	0.00	0.17	0.00	0.17	0.17	0.00	0.00	0.00
中庸里	0.17	0.00	0.17	0.17	0.17	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.00	0.00
豐年里	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.00	0.00
長安里	0.00	0.00	0.00	0.14	0.00	0.00	0.14	0.14	0.14	0.00	0.14	0.14	0.14	0.00
溫泉里	0.00	0.00	0.00	0.00	0.00	0.25	0.25	0.00	0.00	0.25	0.00	0.00	0.25	0.00
大同里	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.20	0.20	0.00	0.00	0.20	0.20
清江里	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.25	0.25	0.00	0.25
中央里	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.50	0.00

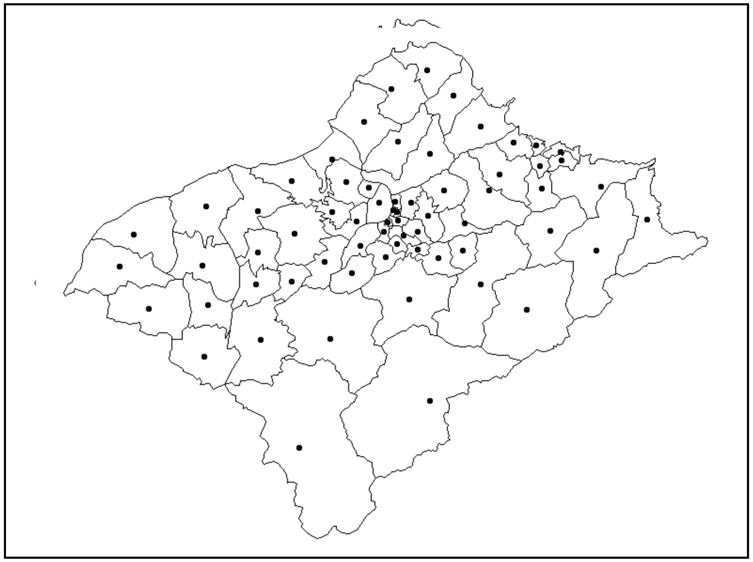
2. Distances

- Distance decay
  - $W_{ij} = 1 / d_{ij}$
  - $W_{ij} = 1 / d^{2}_{ij}$


1000m

Centroid Distances

- Distance between centroids
- Centroid : geometric center of the polygon
  - Affected by the shape of the polygon
  - May be located outside the polygon



Spatially Weighted Matrix

Using Centroid Distances

ID	中和里	泉源里	開明里	文化里	智仁里	林泉里	中心里	中庸里	豐年里	長安里	溫泉里	大同里	清江里	中央里
中和里	0	1359	1602	2229	1948	2334	1913	2142	3149	2530	2763	3169	3238	3375
泉源里	1359	0	2033	2966	2657	1737	2002	2602	3716	2824	2764	3618	3337	3758
開明里	1602	2033	0	974	684	1477	506	585	1692	930	1224	1633	1645	1812
文化里	2229	2966	974	0	309	2335	1309	607	952	932	1519	1073	1607	1316
智仁里	1948	2657	684	309	0	2103	1075	478	1204	903	1444	1265	1644	1496
林泉里	2334	1737	1477	2335	2103	0	1029	1747	2683	1678	1310	2461	1869	2508
中心里	1913	2002	506	1309	1075	1029	0	741	1790	829	856	1640	1381	1762
中庸里	2142	2602	585	607	478	1747	741	0	1115	450	966	1049	1209	1238
豐年里	3149	3716	1692	952	1204	2683	1790	1115	0	1005	1480	324	1203	522
長安里	2530	2824	930	932	903	1678	829	450	1005	0	595	816	760	935
溫泉里	2763	2764	1224	1519	1444	1310	856	966	1480	595	0	1208	581	1214
大同里	3169	3618	1633	1073	1265	2461	1640	1049	324	816	1208	0	882	247
清江里	3238	3337	1645	1607	1644	1869	1381	1209	1203	760	581	882	0	782
中央里	3375	3758	1812	1316	1496	2508	1762	1238	522	935	1214	247	782	0

Nearest Distances (較不常用)

- 兩個Polygon之間最短的點的距離
  - Will be zero for adjacent polygons

ID	中和里	泉源里	開明里	文化里	智仁里	林泉里	中心里	中庸里	豐年里	長安里	溫泉里	大同里	清江里	中央里
中和里	0	0	0	51	0	837	403	0	606	464	723	615	934	1075
泉源里	0	0	0	920	849	0	0	686	1236	731	653	1122	1095	1301
開明里	0	0	0	148	0	319	0	0	529	166	304	477	572	773
文化里	51	920	148	0	0	878	299	0	0	0	567	0	549	481
智仁里	0	849	0	0	0	879	322	0	405	260	600	415	753	870
林泉里	837	0	319	878	879	0	0	535	986	234	0	609	463	691
中心里	403	0	0	299	322	0	0	0	471	0	0	269	269	477
中庸里	0	686	0	0	0	535	0	0	89	0	209	89	407	558
豐年里	606	1236	529	0	405	986	471	89	0	0	583	0	547	67
長安里	464	731	166	0	260	234	0	0	0	0	0	0	0	169
溫泉里	723	653	304	567	600	0	0	209	583	0	0	70	0	179
大同里	615	1122	477	0	415	609	269	89	0	0	70	0	0	0
清江里	934	1095	572	549	753	463	269	407	547	0	0	0	0	0
中央里	1075	1301	773	481	870	691	477	558	67	169	179	0	0	0

Spatially Weighted Matrix

Using Nearest Distances

Spatial Weights Matrix Approaches

Neighborhood Definition

- *Rook's Definition*
- *Queen's Definition*

Spatial Weights Matrix

- *Binary Connective Matrix*
- *Stochastic or Row Standardized Weights Matrix*
- *Centroid Distances*
- *Nearest Distances*

教科書的研讀教材（不用繳交心得作業，但列入考試範圍）

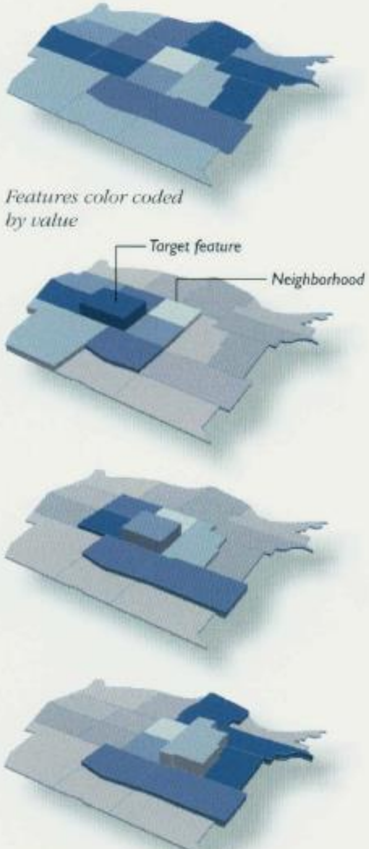
TEXT_Spatial_Weights.pdf

Defining spatial neighborhoods and weights

Several of the methods discussed in chapters 3 and 4 show you how to analyze patterns and clusters of feature values. These methods look at both the difference between the values of features and the spatial relationship between the features (distance or other measure).

Specifically, the GIS compares the value of a feature (the "target") to the values of neighboring features. It then moves to the next feature and does the same thing, and so on, for all the features in the study area. In order to do this, the GIS requires that you define the area surrounding each target feature within which feature values are compared—termed the "neighborhood"—and the nature of the spatial relationship between features. The GIS then assigns weights to each feature pair to specify whether the two features are in each other's neighborhoods, and to represent the spatial relationship between the features.

You define the neighborhood based on the interaction between features. Features might influence each other—for example, the value



Features color coded by value

Each feature in turn is

Measuring Spatial Autocorrelation

- Spatial weighting W_{ij}
  - Contiguity [binary or row-standardized]
    - **Common Border**
  - Distance [centroids or nearest]
    - **Distance band**
    - **K^{th}-nearest neighbors** 每一個都找最近的五個

1. Index of Spatial Autocorrelation: 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

$\bar{x}$ is the mean of the variable

x_i is the variable value at a particular location I

w_{ij} is a **spatial weight indexing** location of i relative to j

$W = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$ (sum of all $w_{i,j}$)

- Applied to a **continuous variable** for polygons or points

這個公式是怎麼想出來的？

皮爾森相關係數 Pearson's correlation coefficient (r):

共變異數 cov (X,Y) 的觀念

共變異數越大，x,y趨勢越相似，越小則差異越大。如果共變異數是負的，代表x,y的變化是相反的關係(X越大Y越小)

X,Y的共變異數

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

$$\boxed{\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})/n}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/n}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}}}$$

平均每個變異數離平均的距離

這個公式是怎麼想出來的？[續]

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})/n}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/n}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}}$$

Neighboring values of variable y replace those of x

把 xi 替換成 <u>yi 的鄰居</u> (c_{ij}.yi) 描述我跟鄰居的相關性

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

這個公式是怎麼想出來的？[續]

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

$$\Downarrow$$

指標越大代表我跟鄰居變化的趨勢越像

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

Spatial Autocorrelation: Moran's I Statistic

Product of the deviation from the mean for all pairs of adjacent regions (w_{ij}=1)

$$I = \frac{n\sum_{i=1}^{n}\sum_{i=j}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\left(\sum\sum_{i\neq j} w_{ij}\right)}$$

Essentially a measure of variance across the regions

Sum of the weights (count of all adjacent pairs)

- n = number of regions
- w_{ij} = measure of spatial proximity between region i and j

Moran's I Interpretations

- Similar to correlation coefficient, range between ± 1.0
  - 0 indicates no spatial autocorrelation, approximate technically it is $-1/(n-1)$
  - Highly auto-correlated, if I is closed to 1 or -1
  - Sign of values indicate negative/positive autocorrelation
- Can be used as index for dispersion/random/cluster patterns
  - 0: random
  - Positive : more toward clustering
  - Negative: more toward dispersion/uniform

Significance Tests for Moran's I

- Z-score: $Z = (I - E(I)) / S_{Err}(I)$
  - I: Moran's I of sample
  - E(I): Expected value of I ; $E(I) = -1/(n-1)$
  - S_{Err}: Standard error
    - Depend on if free or non-free sampling is used 兩種sampling 的方法
- $\alpha = 0.05$, Critical Z value = $\pm$ 1.96
  - will be ± 1.645 for $\alpha = 0.1$
- 檢定是否為達到顯著差異
  - Ho:無差異（隨機分佈）
  - At $p < 0.05$, Reject Ho if $|Z| > 1.96$

$$E_N(I) = E_R(I) = \frac{-1}{n-1}$$

$$VAR_N(I) = \frac{(n^2 S_1 - nS_2 + 3W^2)}{W^2(n^2-1)} - [E_N(I)]^2$$

$$VAR_R(I) = \frac{n[(n^2-3n+3)S_1 - nS_2 + 3W^2]}{(n-1)(n-2)(n-3)W^2} - \frac{k[(n^2-n)S_1 - nS_2 + 3W^2]}{(n-1)(n-2)(n-3)W^2} - [E_R(I)]^2,$$

where

free sampling
(with replacement.)
(normality)
像排列組合的取後放回

non-free sampling
(without replacement)
(randomization)
取後不放回(較常用)

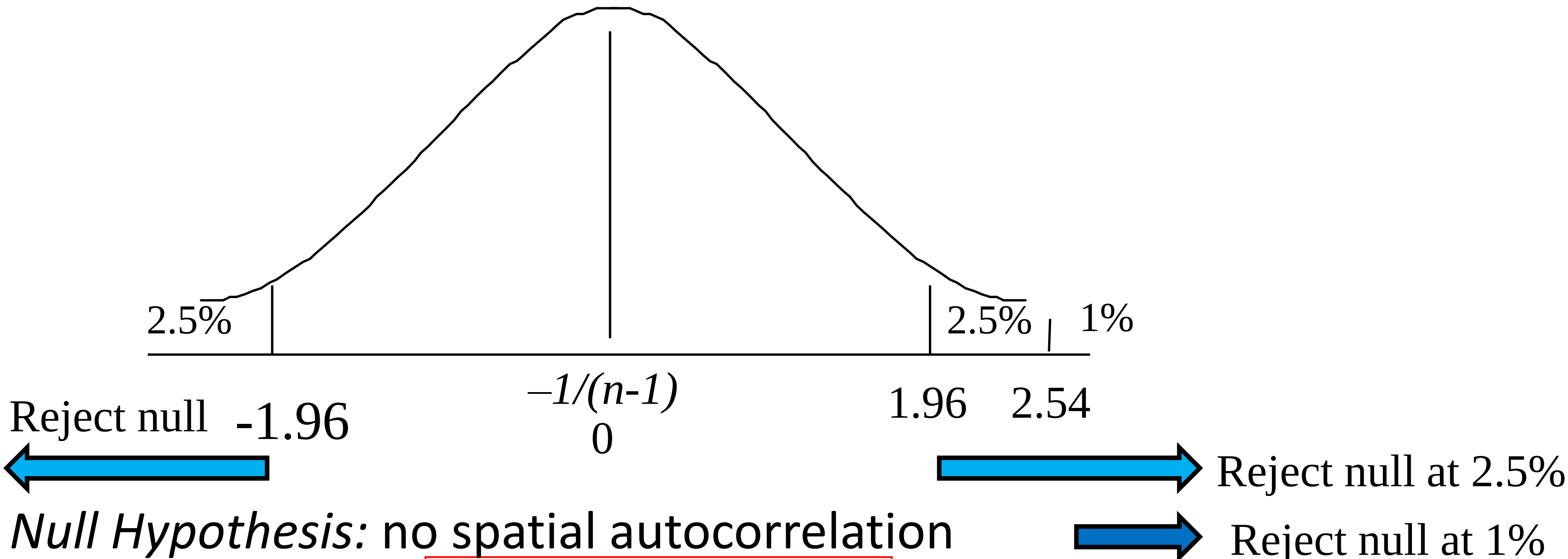
$$W = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$$

$$S_1 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij} + w_{ji})^2}{2}$$

$$S_2 = \sum_{i=1}^{n}(w_{i.} + w_{.i})^2$$

$$k = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2}.$$

Test Statistic for Normal Frequency Distribution



Null Hypothesis: no spatial autocorrelation

*Moran's I = 0 ** technically –1/(n-1)*

Alternative Hypothesis: spatial clustering exists

*Moran's I > 0（單尾檢定）

Reject *Null Hypothesis* if Z test statistic > 1.96 (or < -1.96)

---less than a 2.5% chance that, in the population, there is no spatial clustering

---97.5% confident that spatial clustering exits

Monte-Carlo Significance Test

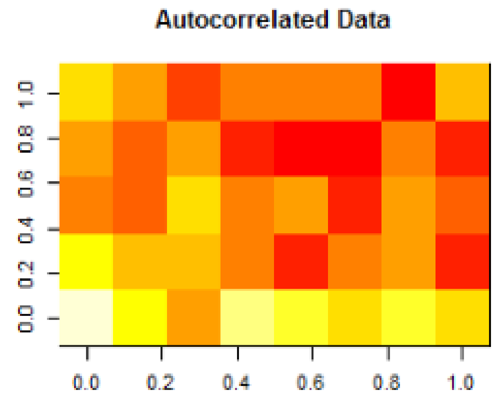
例如檢定人口密度低的地方周邊區域人是否口密度也很低。360人在一塊塊地上，用蒙地卡羅方式就是把地分成好幾塊重新排列多次

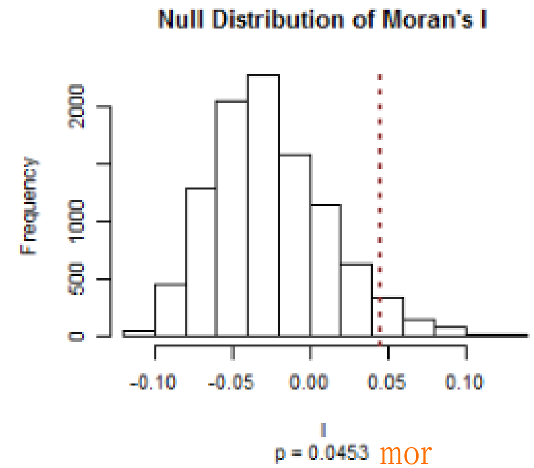
- **Permutation test (排列檢定)**
  - The null hypothesis is that the data were determined and then assigned to their spatial locations at random.
  - The alternative is that the assignment to each location depended on the assignment at that location's neighbors.
  - The permutation test does not randomize over the possible sets of data values--it considers them given--
    but **conditional on the data observed**, considers all possible ways of reassigning them to the locations.

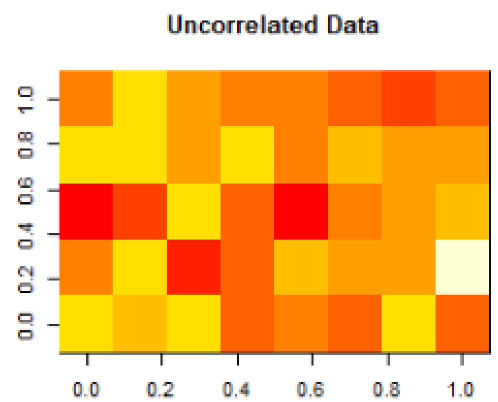
Monte-Carlo Significance Test (cont'd)

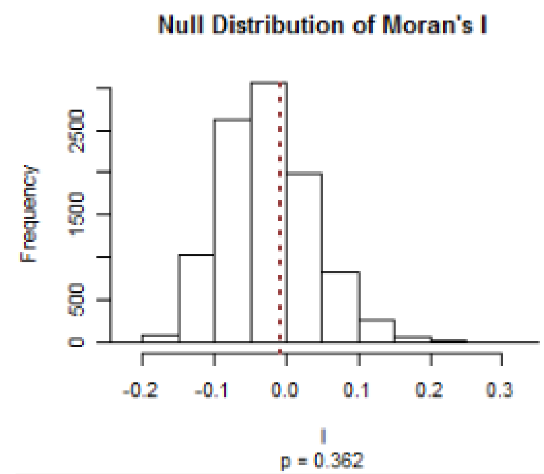
- **Permutation test (排列檢定)**
- Such a reassignment is a *permutation.* For *n* data points, there **are n! = n x (n-1) x (n-2) x ... x (2) x (1)** permutations.
  For *n* much larger than 10 or so, that's too many to generate.
- There usually is no simple analytical expression for the full permutation distribution.
- Accordingly, we typically resort to sampling from the set of all permutations at random, giving them all equal weight. The distribution of the autocorrelation statistic in a sufficiently large sample (usually involving at least 500 permutations) approximates the true distribution.

Examples


Autocorrelated Data
0.0
0.2
0.4
0.6
0.8
1.0

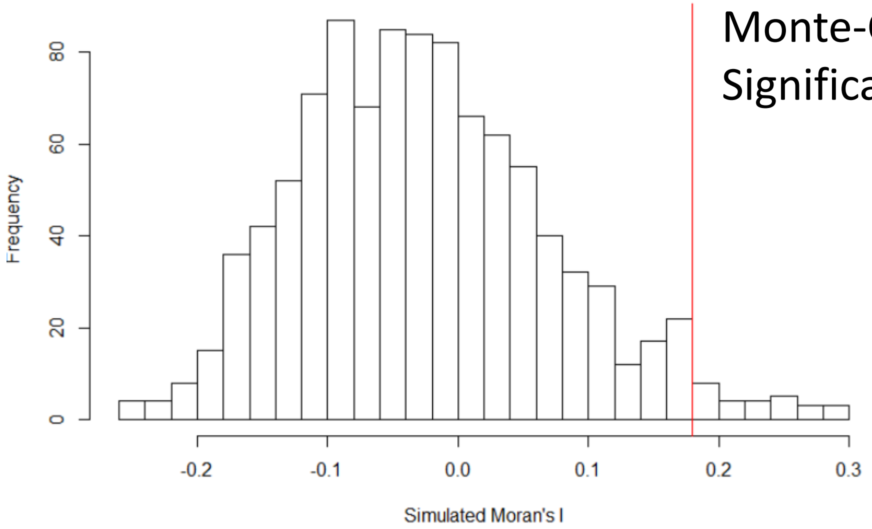

Null Distribution of Moran's I
Frequency
0
500
1000
2000
-0.10
-0.05
0.00
0.05
0.10
I
p = 0.0453
mor


Uncorrelated Data
0.0
0.2
0.4
0.6
0.8
1.0


Null Distribution of Moran's I
Frequency
0
500
1500
2500
-0.2
-0.1
0.0
0.1
0.2
0.3
I
p = 0.362

Output in R: Moran's I statistic

```
> M<-moran.test(Popn, listw=TWN_nb_w, zero.policy=T); M

        Moran I test under randomisation

data:  Popn
weights: TWN_nb_w

Moran I statistic standard deviate = 2.1678, p-value = 0.01509
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
      0.181359104      -0.025000000       0.009062094
```



Monte-Carlo simulation
Significance Test

Moran.test()

moran.test {spdep} R Documentation

Moran's I test for spatial autocorrelation

Description

Moran's test for spatial autocorrelation using a spatial weights matrix in weights list form. The assumptions underlying the test are sensitive to the form of the graph of neighbour relationships and other factors, and results may be checked against those of `moran.mc` permutations.

Usage

```
moran.test(x, listw, randomisation=TRUE, zero.policy=NULL,
 alternative="greater", rank = FALSE, na.action=na.fail, spChk=NULL, a
```

`randomisation`	variance of I calculated under the assumption of randomisation, if FALSE normality

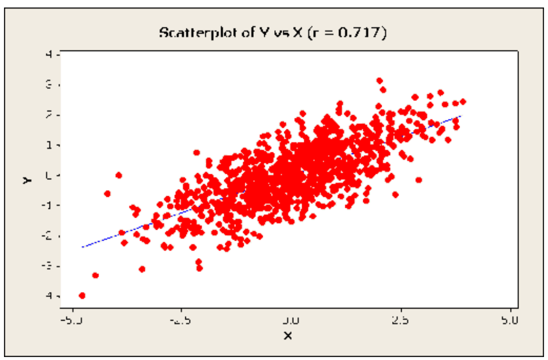
Concept of Moran Scatter Plots

Scatter Plot

Moran Scatter Plot

Two variables

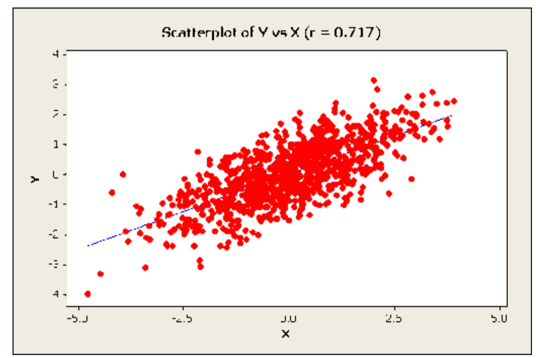
Income



Education

Only one variable

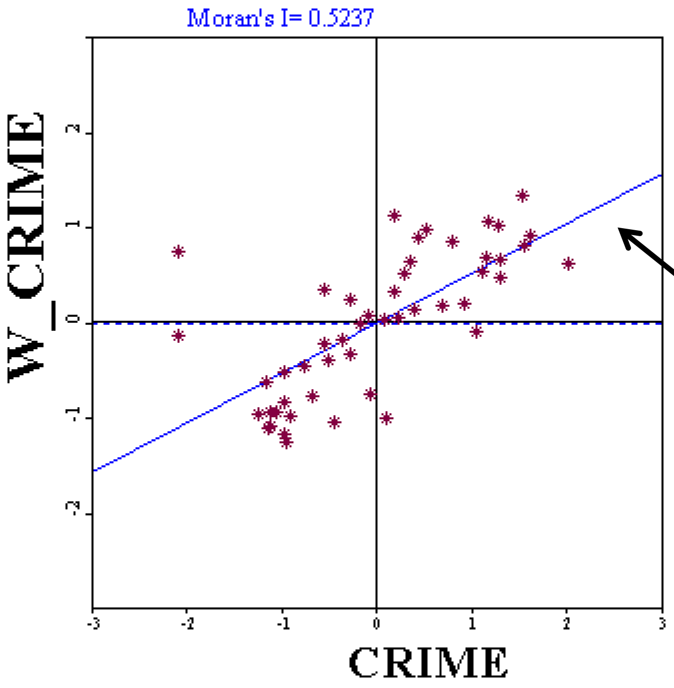
Crime in nearby area

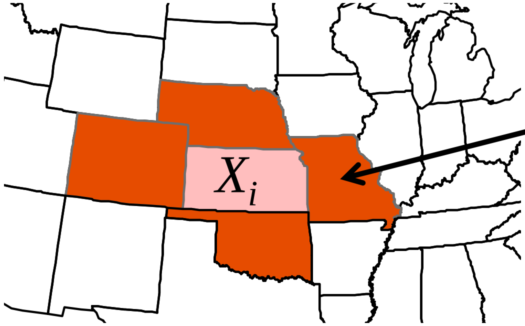


Crime Rate

Moran Scatter Plots

Moran’s I can be interpreted as the correlation between variable, X, and the “spatial lag” of X formed by averaging all the values of X for the neighboring polygons.





Lag X_i is average of these

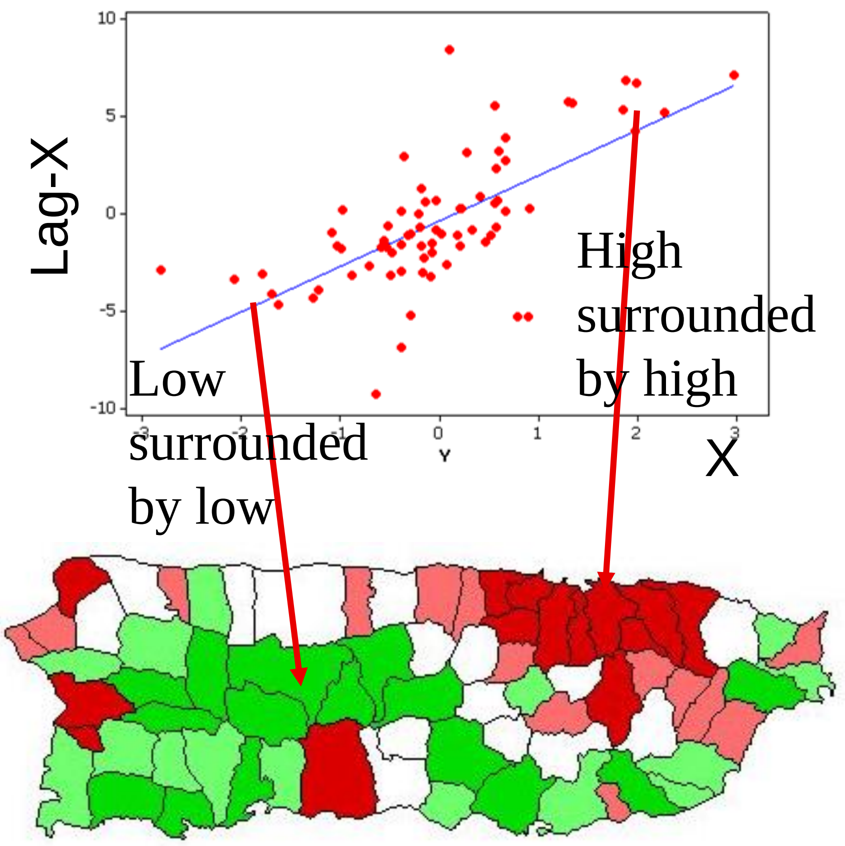
Least squares “best fit” line to the points.

The <u>slope</u> of this *regression line* is Moran’s I

Moran Scatter Plot: example

- Scatter plot of **X** vs. **Lag-X**
- The slope of the regression is Moran's *I*

Moran's $I = 0.49$



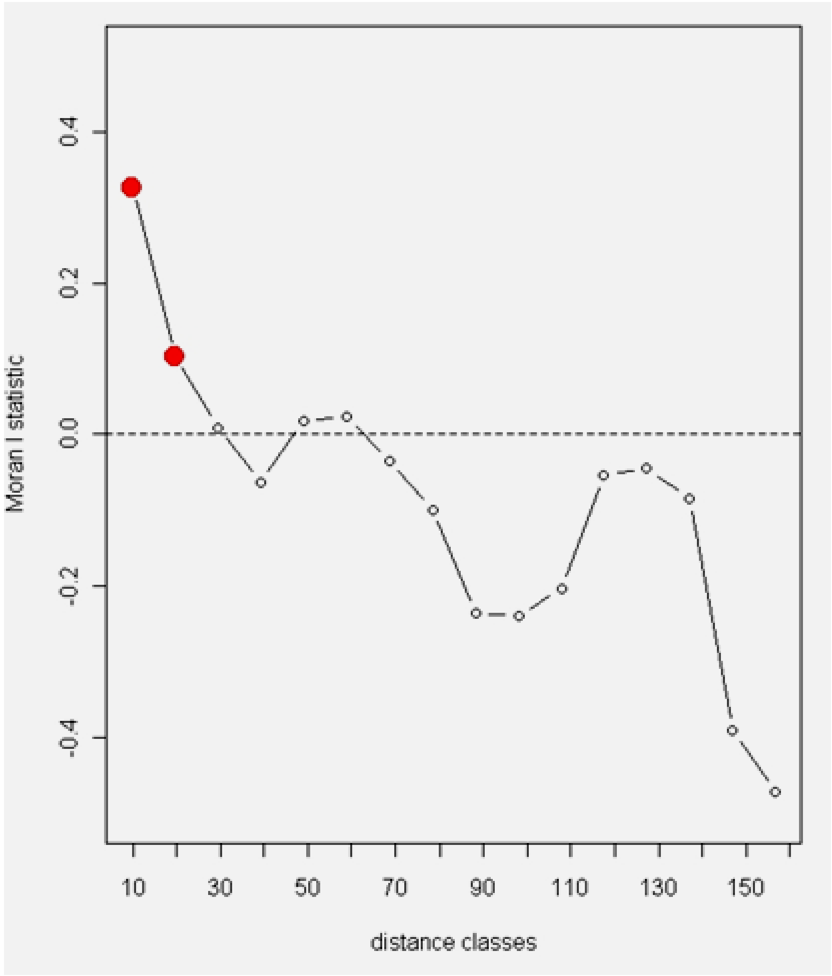
Population density

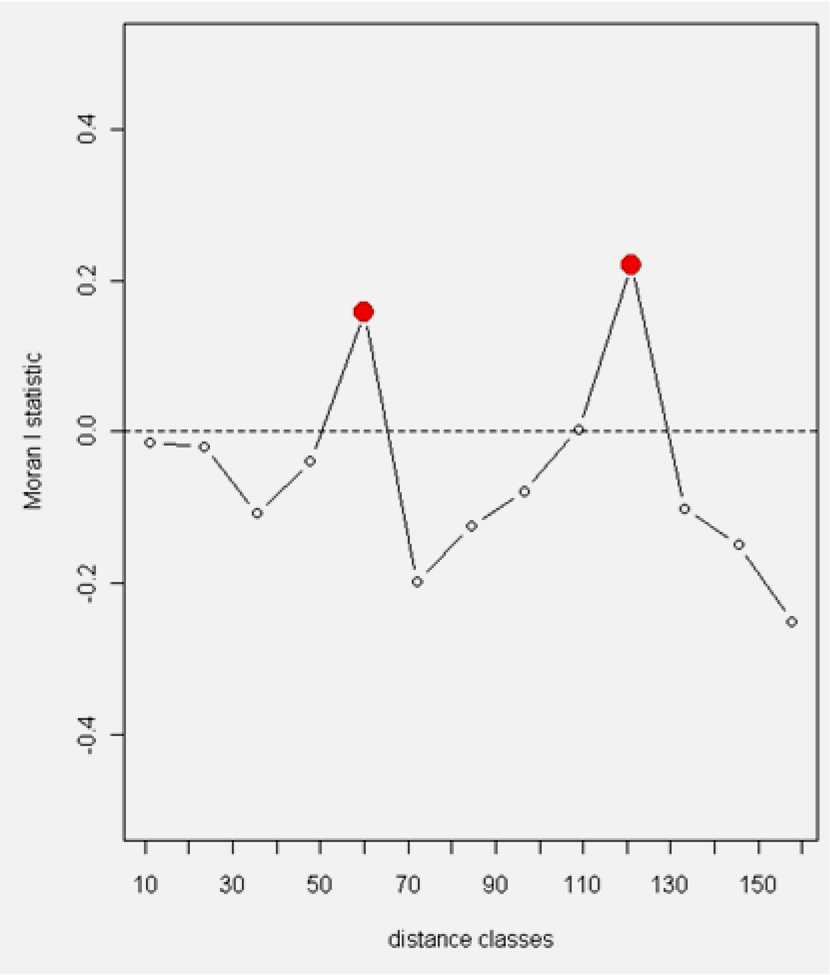
Moran Correlograms

不同半徑下Moran I的變化

Correlogram: plot distance on X-axis against correlation coefficient on Y-axis

可能是城市跟城市之間的影響(例如疾病爆發)





Getis-Ord General G-statistic

- Moran's I 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

Getis-Ord General G-Statistic

- General G-statistic **can distinguish between hot/cold spots**. It identifies **spatial concentrations**.
  - G is relatively large if high values cluster together
  - G is relatively low if low values cluster together
- G statistic is interpreted relative to its expected value
  - > E(G) → potential "***hot spot***"
  - < E(G) → potential "***cold spot***"
  - = E(G) → no spatial association
- 所謂之larger/smaller不能單從值的大小判斷，需以 **Z test statistic** 來檢定差異的統計顯著性。

Significance Test for Getis-Ord General G

- Statistical Significance Test

$$Z = \frac{G - E(G)}{S_{Err(G)}}$$

- Expected G : $E(G) = \frac{W}{n(n-1)}$; where $W = \sum_i \sum_j w_{ij}(d),$

- Standard Error will depend on the sampling method (free / non-free)

Getis-Ord General G-statistic in R

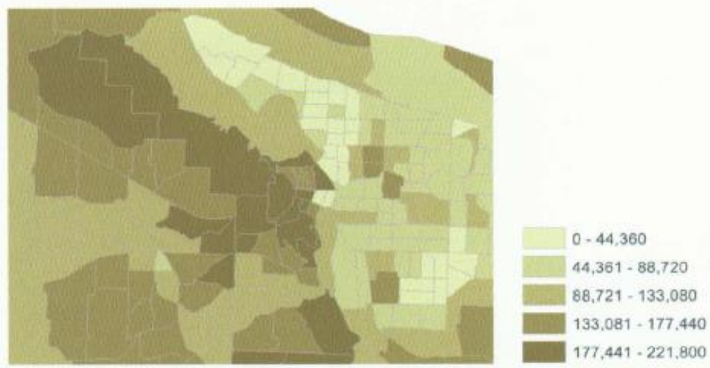
```
> G<-globalG.test(Popn, listw=TWN_ran1_wb); G

        Getis-Ord global G statistic

data:  Popn
weights: TWN_ran1_wb

standard deviate = 3.4804, p-value = 0.0002504
alternative hypothesis: greater
sample estimates:
Global G statistic        Expectation           Variance
      0.6216802233       0.5402439024       0.0005474983
```

教科書的研讀教材（列入考試範圍）

TEXT_Pattern.of.Feature.Value.pdf

Measuring the spatial pattern of feature values

In addition to measuring the pattern formed by the locations of features, you can also measure patterns of attribute values associated with features, such as the pattern formed by median house values. These methods reveal whether similar values tend to occur near each other, or whether high and low values are interspersed.



Median house value by census tract.

The idea behind measuring patterns of feature values

Measuring the spatial pattern of feature values is based on the notion that things near each other are more alike than things far apart, an idea often attributed to geographer Waldo Tobler. The idea is consistent with our

實習：介紹 R package: spdep

Spatial Dependence: Weighting Schemes, Statistics and Models

spdep: Spatial Dependence: Weighting Schemes, Statistics and Models

A collection of functions to create spatial weights matrix objects from polygon 'contiguities', from point patterns by distance and tessellations, for summarizing these objects, and for permitting their use in spatial data analysis, including regional aggregation by minimum spanning tree; a collection of tests for spatial 'autocorrelation', including global 'Morans I', 'APLE', 'Gearys C', 'Hubert/Mantel' general cross product statistic, Empirical Bayes estimates and 'Assunção/Reis' Index, 'Getis/Ord' G and multicoloured join count statistics, local 'Moran's I' and 'Getis/Ord' G, 'saddlepoint' approximations and exact tests for global and local 'Moran's I'; and functions for estimating spatial simultaneous 'autoregressive' ('SAR') lag and error models, impact measures for lag models, weighted and 'unweighted' 'SAR' and 'CAR' spatial regression models, semi-parametric and Moran 'eigenvector' spatial filtering, 'GM SAR' error models, and generalized spatial two stage least squares models.

spdep 重要函數

- Spatial Neighbors
  - Contiguity: QUEEN vs. ROOK poly2nb(); nb2mat()
  - K-nearest Neighbors (KNN) knn2nb(); knearneigh(coords, k=2)
  - Distance-based dnearneigh()
- From Spatial Neighbors to ListW (Weighting matrix)
  - nb2listw()
- Spatial Autocorrelation
  - Mapping the attribute **GISTools::** choropleth()
  - Moran's I Statistic moran.test()
  - Monte-Carlo simulation moran.mc()
  - Moran correlogram sp.correlogram()
  - Moran Scatter Plot moran.plot()
  - Getis-Ord General G Statistic globalG.test()

實習

資料： **Popn_TWN2.shp**

- 台灣鄉鎮市區**<u>人口密度</u>**的空間型態分析
  - 計算以下統計量與繪製圖表，說明其參數設定，並解釋其意義。包括：Moran’s I coefficient, Monte-Carlo simulation, Moran scatter plot, Correlogram, and General G statistic.
  - 利用以下三種不同的空間鄰近定義，計算Moran’s I coefficient，比較其數值的差異，並討論可能的原因。

    Spatial Neighbors: Contiguity; K-nearest Neighbors (KNN); Distance-based

作業

- 共三題 (以Contiguity定義鄰近)　　資料：**Popn_TWN2.shp**
  - 繪製各鄉鎮的鄰居數的直方圖。
  - 找出台灣本島最多鄰居的鄉鎮是哪一個? (TOWN_ID)
  - 繪製台灣各鄉鎮的1st-order鄰居人口密度的面量圖。