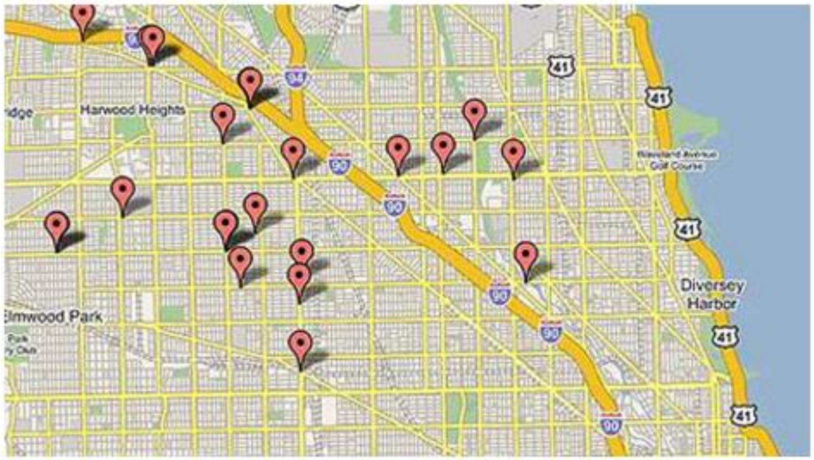
Describing a Point Pattern

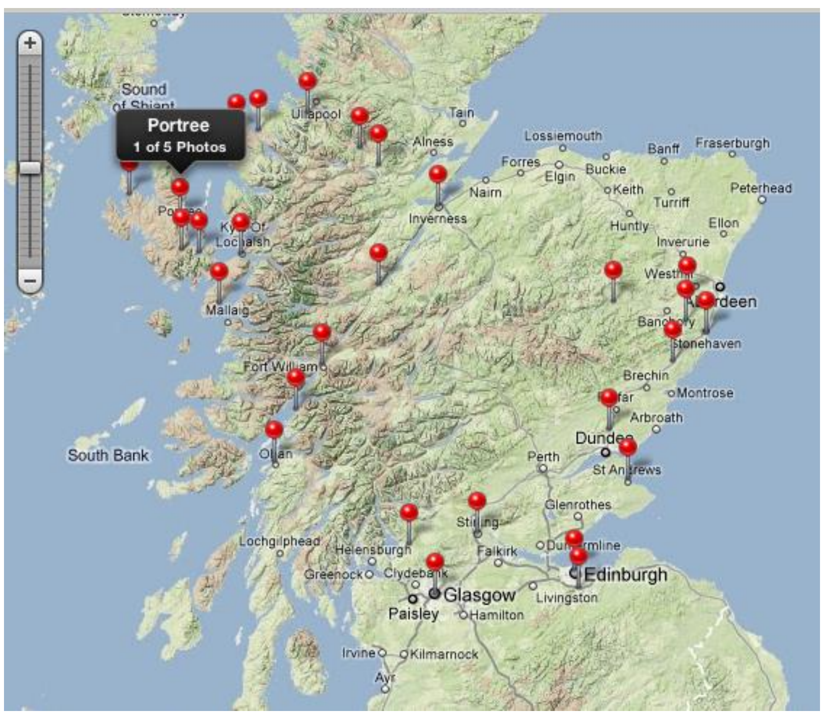
https://ceiba.ntu.edu.tw/1092Geog2017_

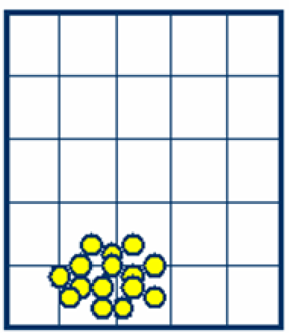
授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw

The Era of Geotagging: What's the Next?

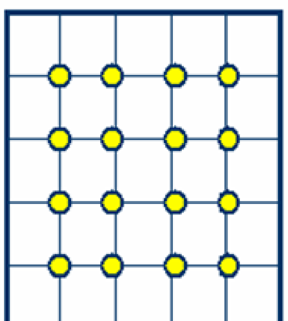


Flickr: pixeljones

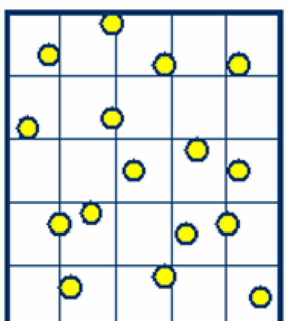




(a) Clustering



(b) Dispersion/Uniform



(c) Random

Spatial Point Pattern Analysis

- Measuring Geographic Distribution
  - Spatial Mean & Standard Distance
  - Spatial Median
  - Standard Deviational Ellipse

Exploratory Analysis:

Descriptive Statistics

- Analyzing ***Global*** Patterns
  - Quandrat Analysis
  - Nearest Neighbor Methods
  - Ripley's K Function

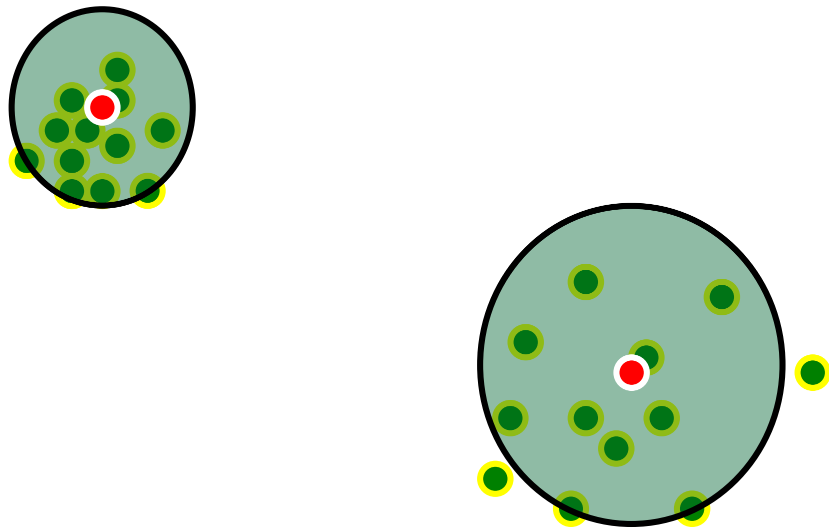
Confirmatory Analysis:

Inferential Statistics

Issues of Point Pattern Description

- ***WHAT*** are the centrality & dispersion of a set of points data.
- ***WHY*** mapping the centrality & dispersion.
- ***HOW*** to measure the centrality & dispersion.

Why mapping the centrality & dispersion of a set of points



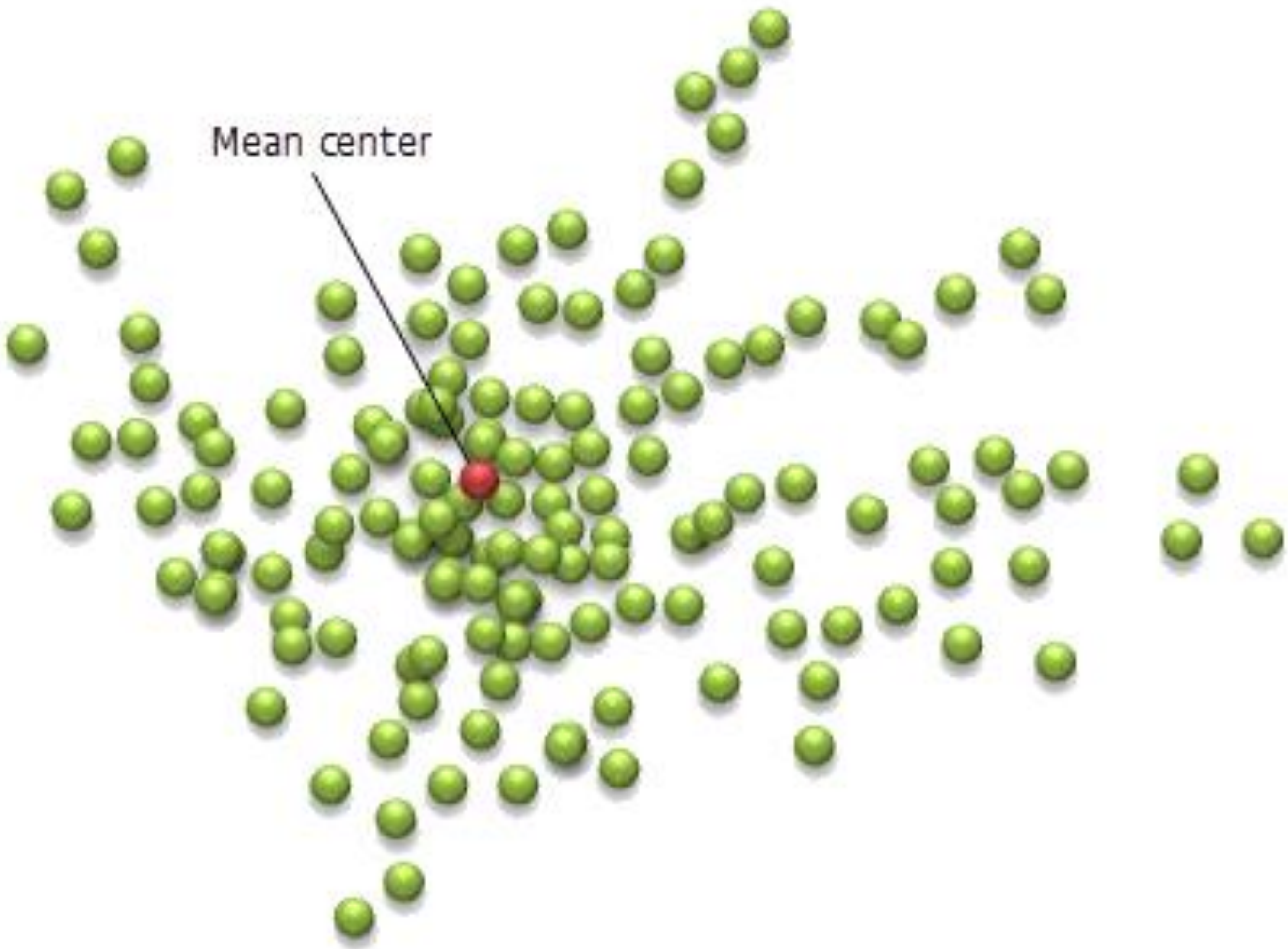
Measures of Centrality

- Mean Center $$(x_{mc}, y_{mc}) = \left(\frac{\sum x_i}{n}, \frac{\sum y_i}{n}\right)$$
- Weighted mean center $$(x_{wmc}, y_{wmc}) = \left(\frac{\sum w_i x}{\sum w_i}, \frac{\sum w_i y_i}{\sum w_i}\right)$$
- Median Center

1. **Center of Minimum Distance**
   - Euclidean distance / Manhattan distance
2. **As the region is divided into 4 quadrants with equal numbers of data point**

- Weighted median center

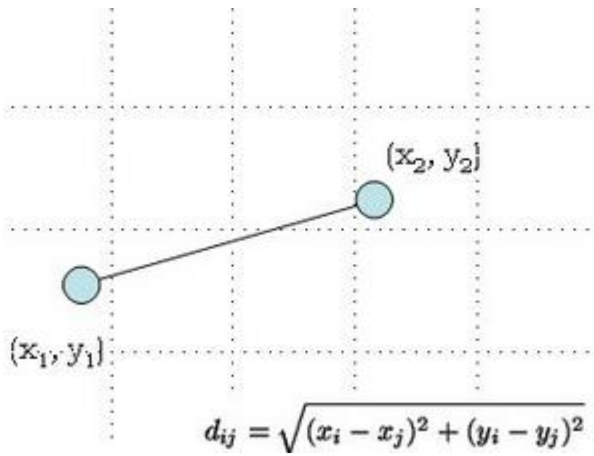
Mean Center



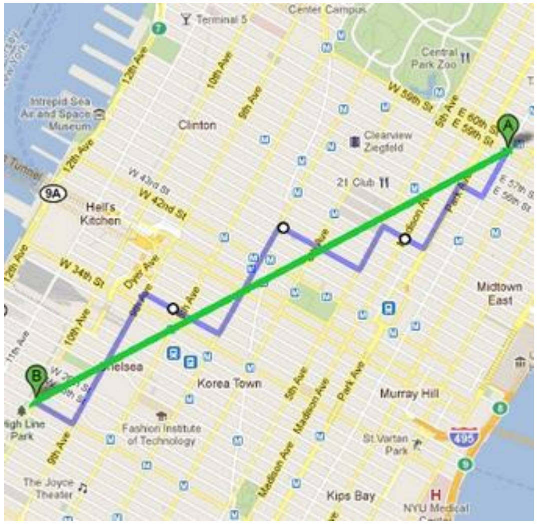
Median Center

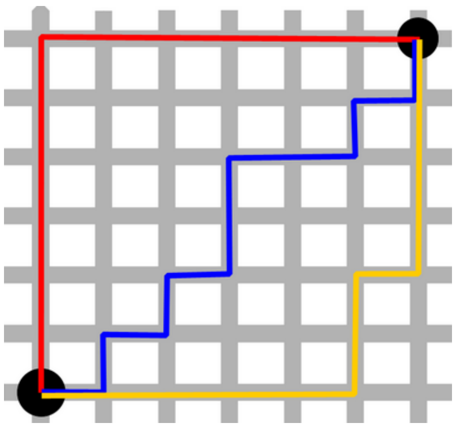
$$\textit{Minimize } \sum_i \quad d_i^t = \sqrt{(X_i - X^t)^2 + (Y_i - Y^t)^2}$$

Euclidean distance
歐幾里得距離

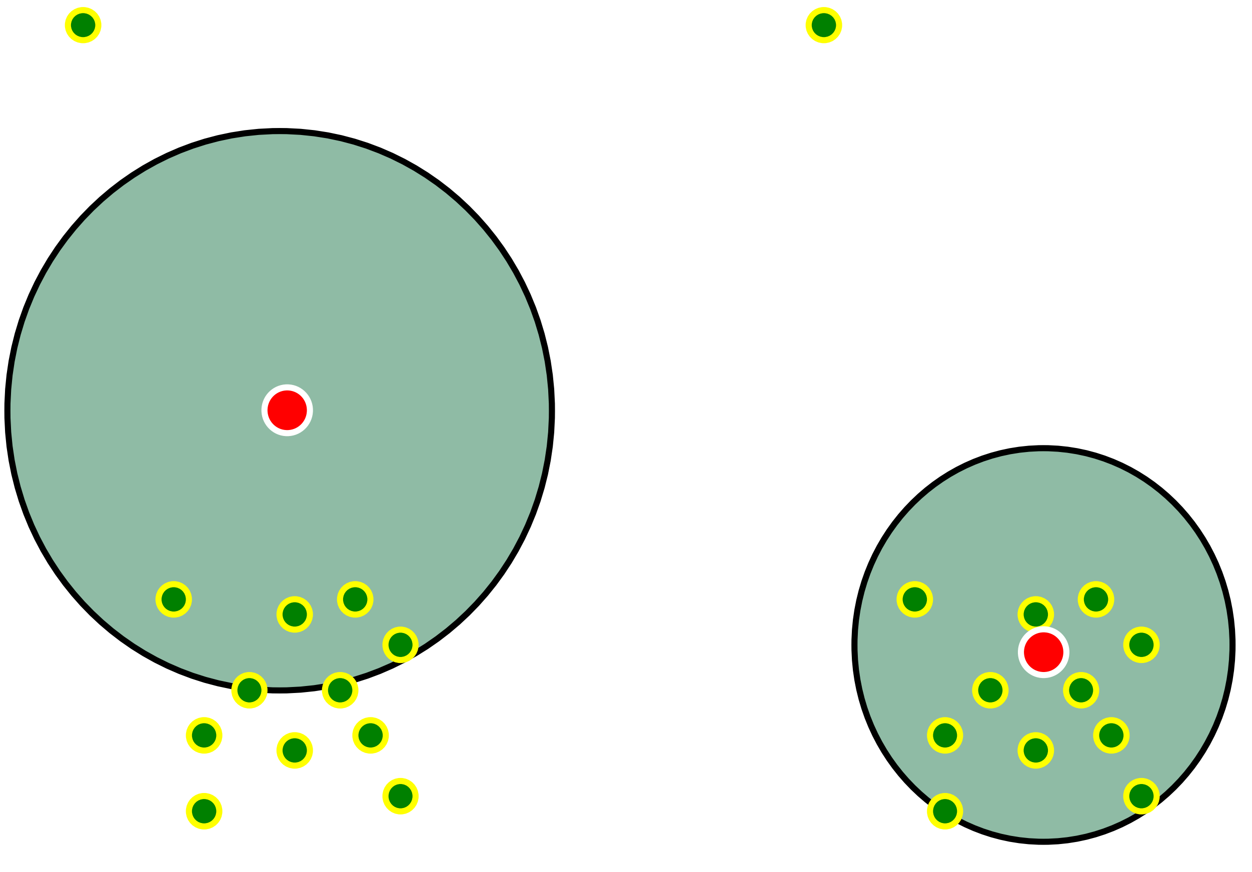


Manhattan distance
曼哈頓距離

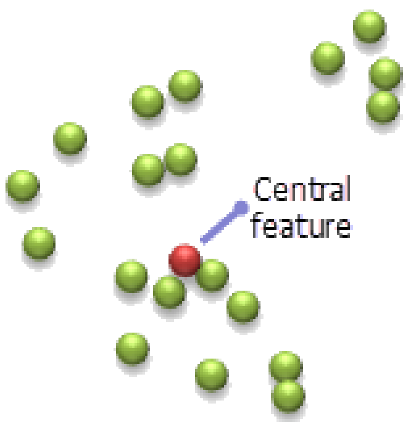




Mean Center vs. Median Center



Central Feature



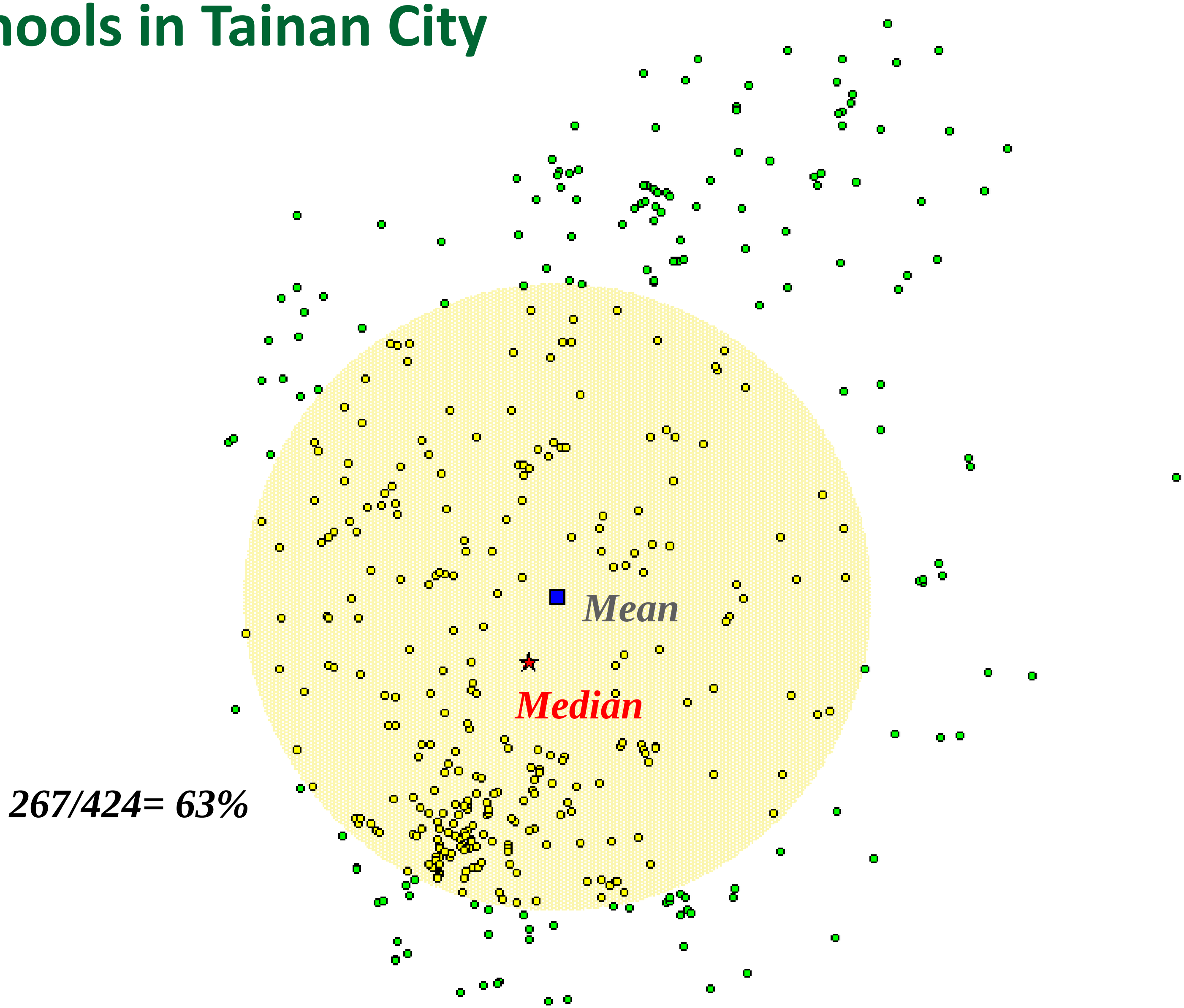
	OBJECTID	Shape	Name
	1	Point	A
	2	Point	B
	3	Point	C
	4	Point	D
	5	Point	E
	6	Point	F
	7	Point	G
	8	Point	H
	9	Point	I
	10	Point	J
	11	Point	K
	12	Point	L

	OBJECTID	Shape	Name
▸	1	Point	G

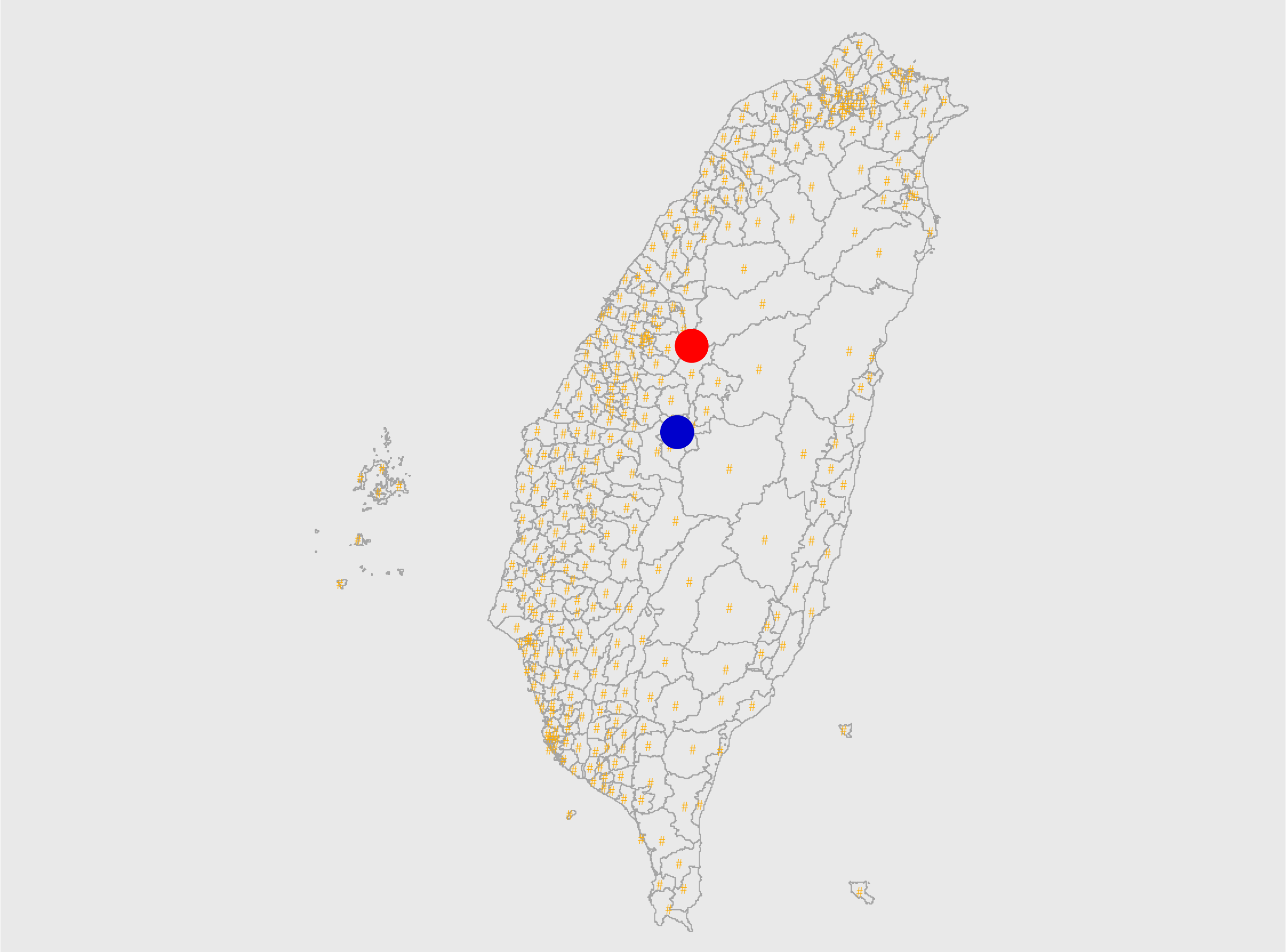
Summary

Center	What it represents	What it's good for
Mean	The average x-coordinate and average y-coordinate for all features in the study area	Tracking changes or comparing distributions
Median	The x,y coordinate having the shortest distance to all features in the study area	Finding the most accessible location
Central feature	The feature having the shortest total distance to all other features in the study area	Finding the most accessible feature

Demo: Schools in Tainan City



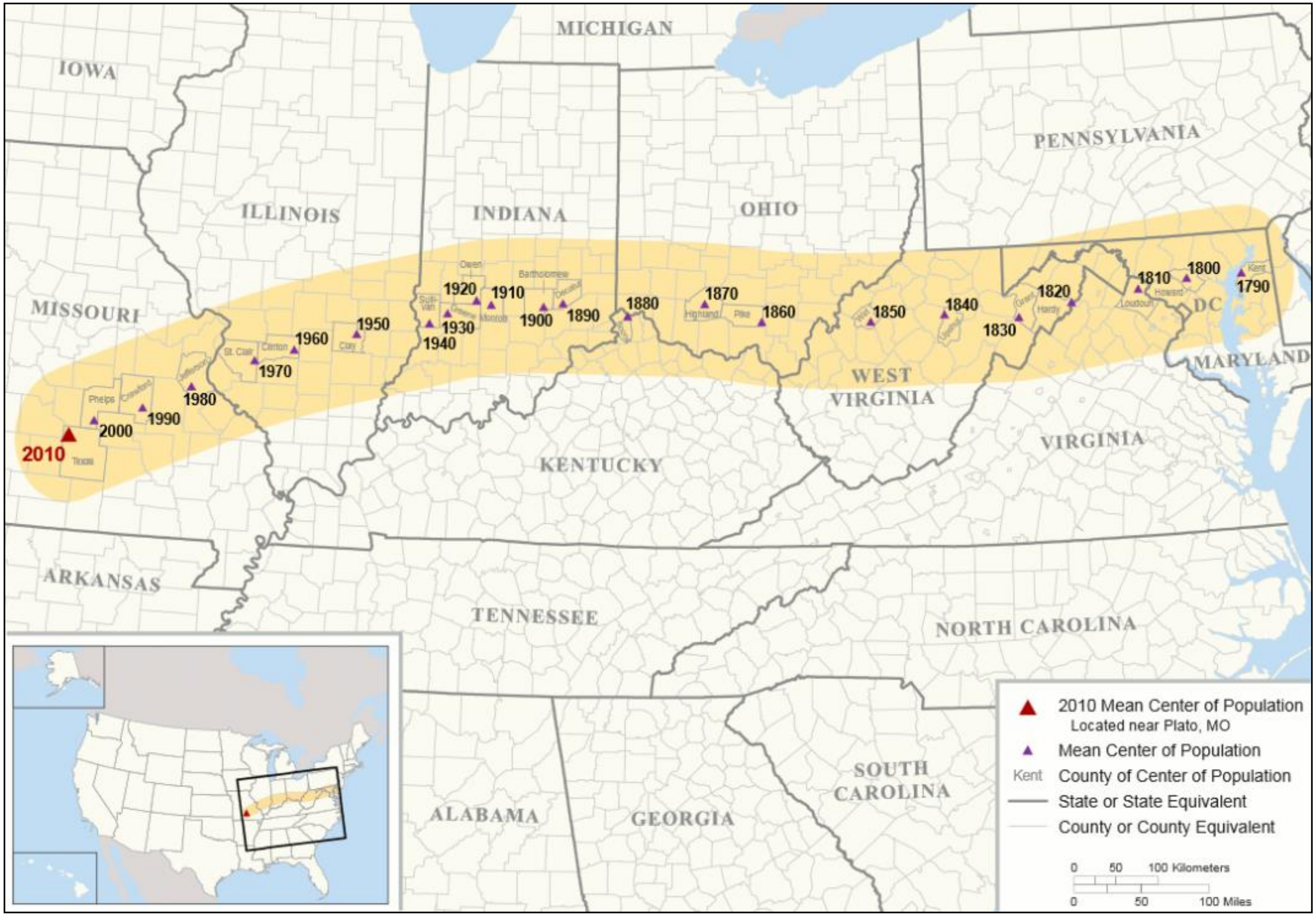
Demo: Geographic Center (weighted by population)



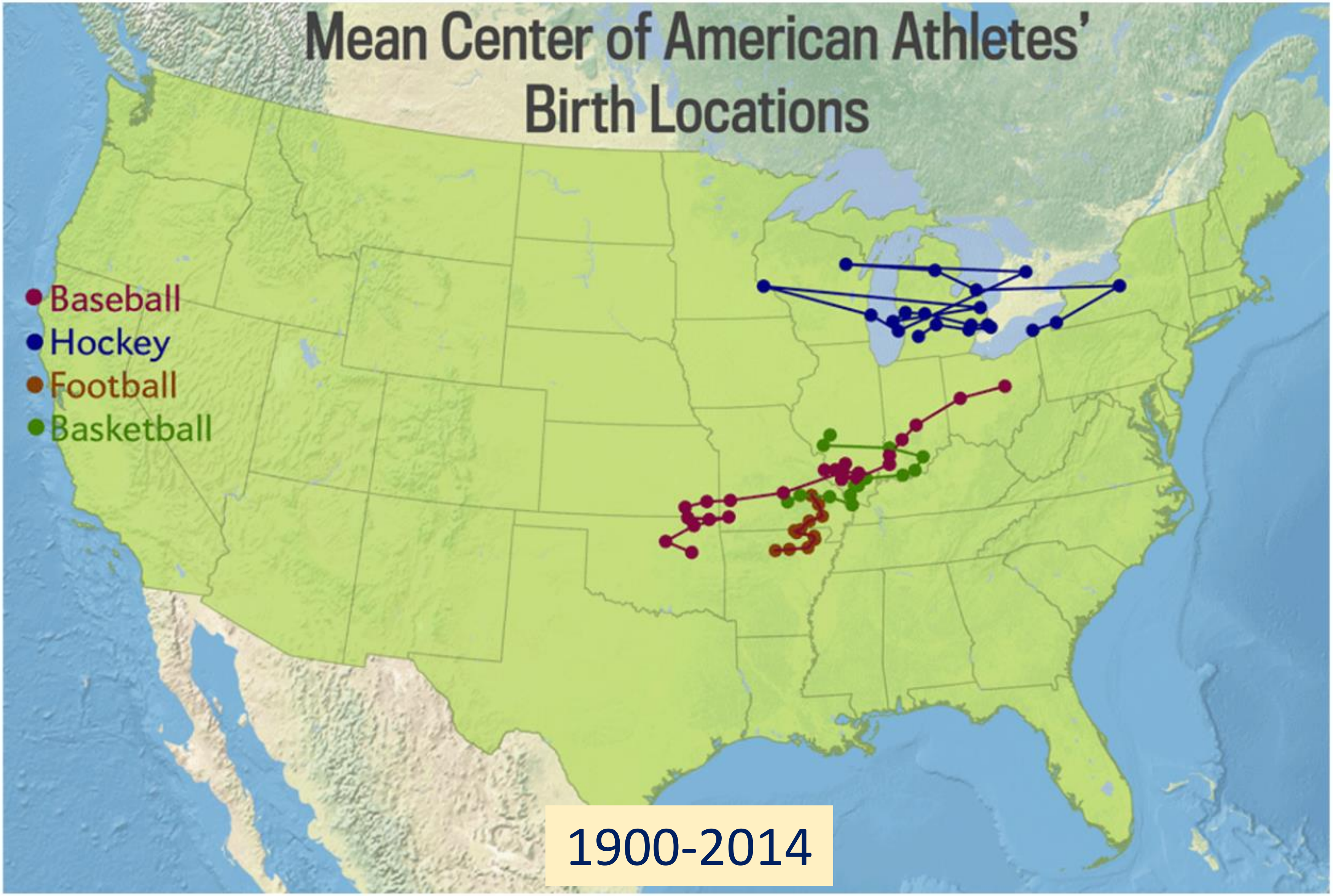
Position of the Geographic Center of Area, Mean and Median Centers of Population: 2010

https://www2.census.gov/geo/pdfs/reference/cenpop2010/centerpop_geographic2010.pdf

Mean centers of the United States population



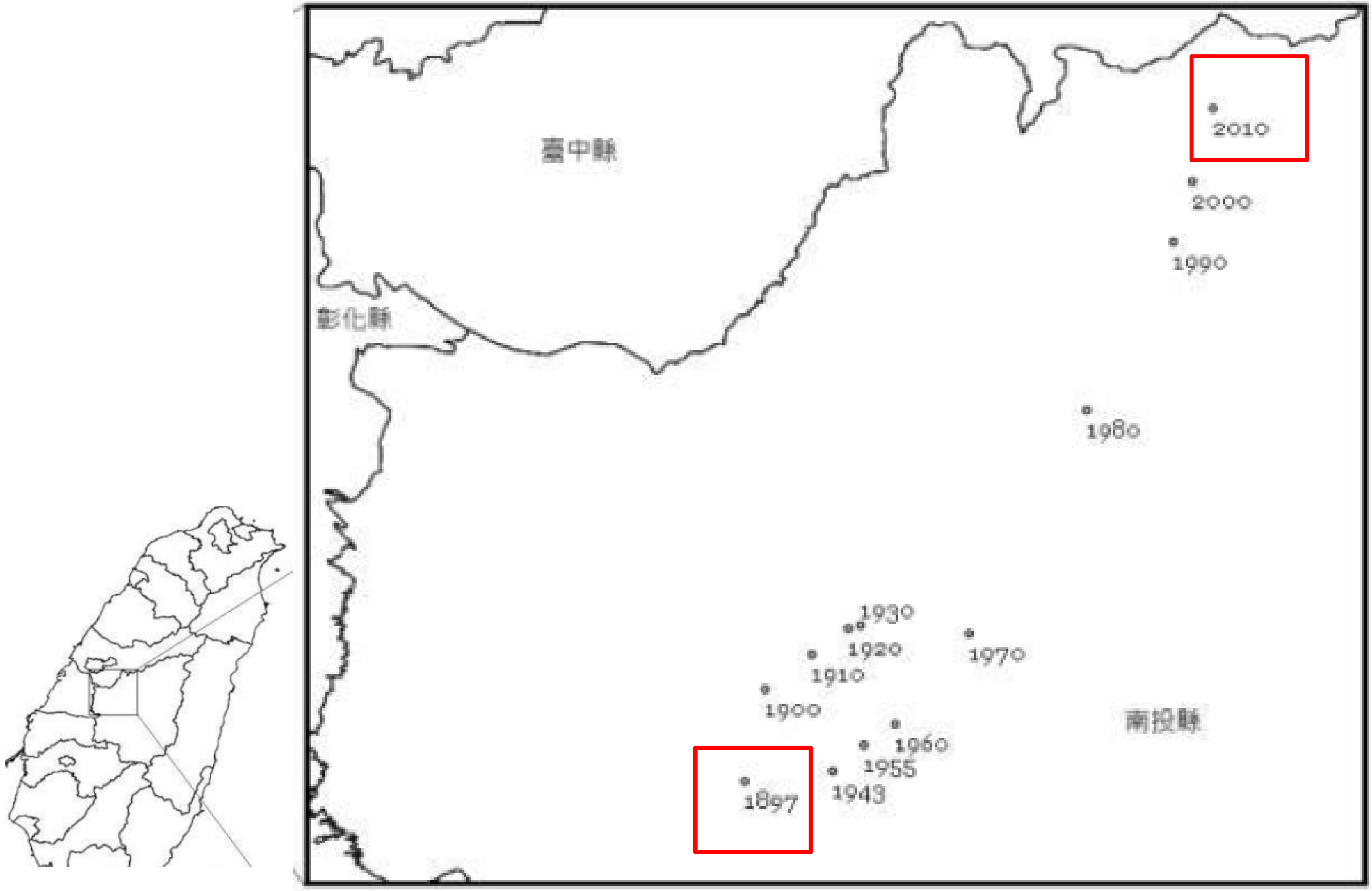
Map showing changes to the mean center of population for the United States, 1790–2010 (US Census Bureau)



Data from sports-reference.com. Created by Ben Blatt

http://www.slate.com/articles/sports/culturebox/2014/10/baseball_player_map_a_new_u_s_map_based_on_where_baseball_players_were_born.html

台灣人口中心的變遷：1897-2010



Measures of Dispersion

- Standard Distance

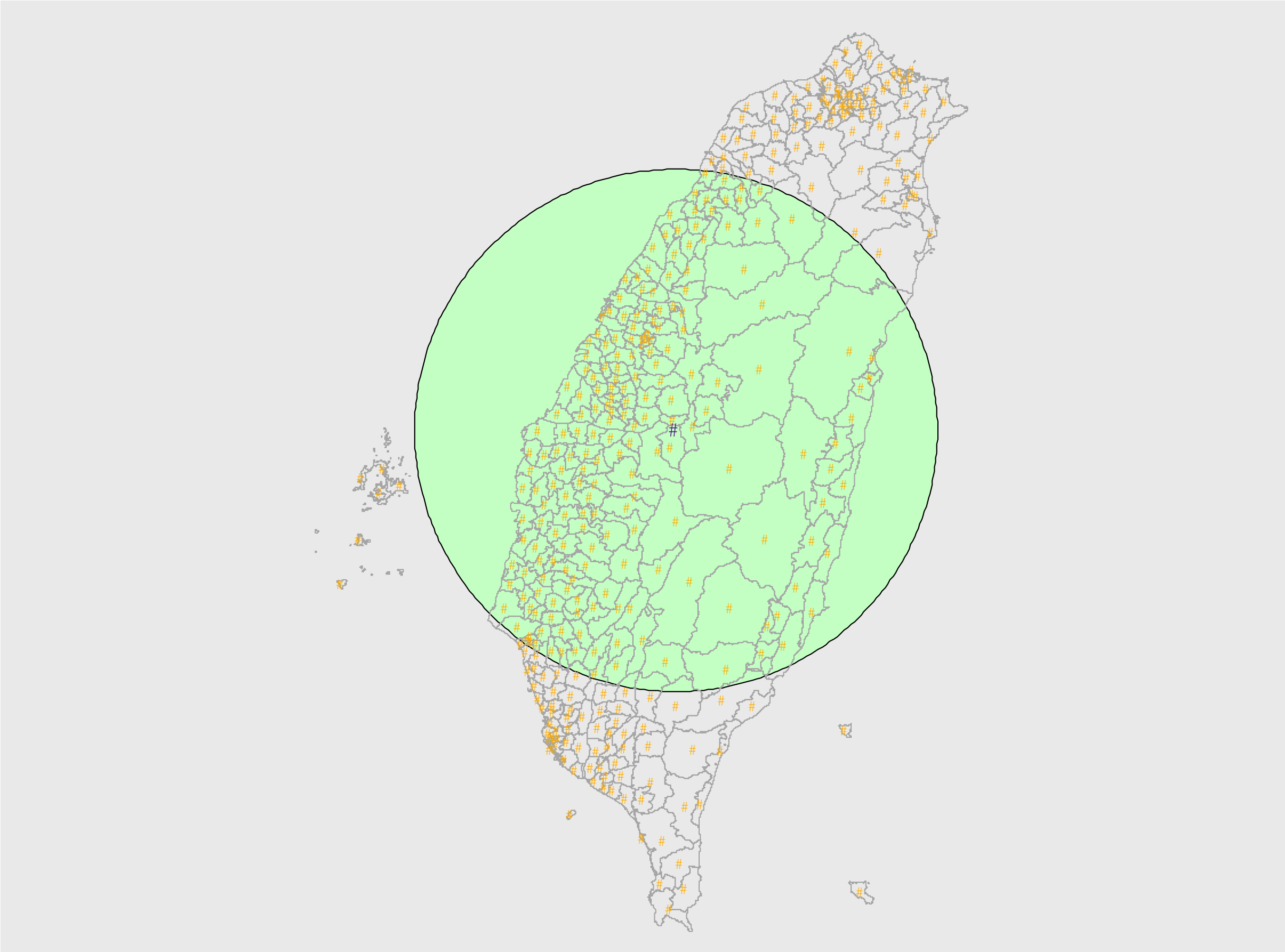
$$SD = \sqrt{\frac{\sum (x_i - x_{mc})^2 + \sum (y_i - y_{mc})^2}{n}}$$

  - **A spatial analogy of Std. Dev. to describe the distribution of points around the mean center**
  - **Does not capture any directional bias and capture the shape of the distribution**
- Weighted Std. Distance

$$SD = \sqrt{\frac{\sum w_i (x_i - x_{mc})^2 + \sum w_i (y_i - y_{mc})^2}{\sum w_i}}$$

- Standard Deviational Ellipse
  - gives dispersion in two dimensions

Standard Distance

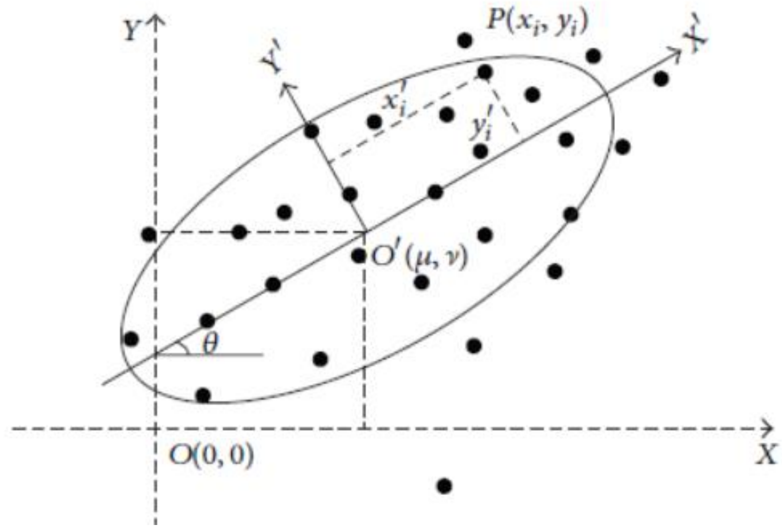


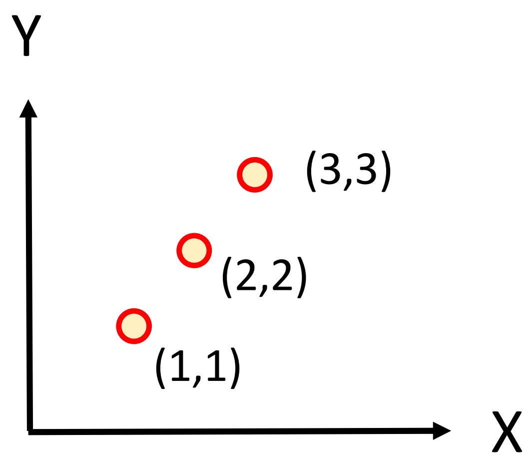
Standard Deviational Ellipse

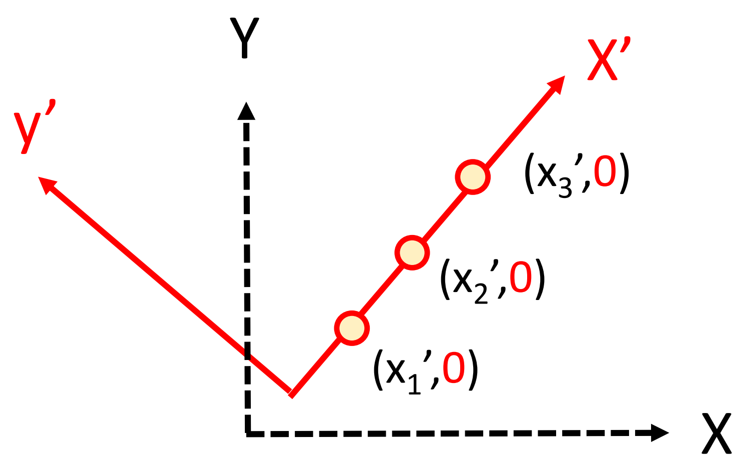
長(短)軸的長度
(標準差)

$$\sigma_x = \sqrt{2}\sqrt{\frac{\sum_{i=1}^{n}(\tilde{x}_i \cos\theta - \tilde{y}_i \sin\theta)^2}{n}}$$

$$\sigma_y = \sqrt{2}\sqrt{\frac{\sum_{i=1}^{n}(\tilde{x}_i \sin\theta + \tilde{y}_i \cos\theta)^2}{n}}$$







When the $\sigma_{y'}$ value obtains the minimum value in the rotated coordinate system, the rotated angle θ is the direction of scattered points. Then it can get the minimum value through calculating the derivative of $\sigma_{y'}$ for (12)

$$\frac{d\sigma_{y'}}{d\theta} = \frac{1}{n\sigma_{y'}}\left[\sum_{i=1}^{n}\overline{x_i}^2 \sin\theta\cos\theta - \sum_{i=1}^{n}\overline{x_i}\,\overline{y_i}\left(\cos^2\theta - \sin^2\theta\right) - \sum_{i=1}^{n}\overline{y_i}^2 \sin\theta\cos\theta\right]. \tag{14}$$

Solving for θ,

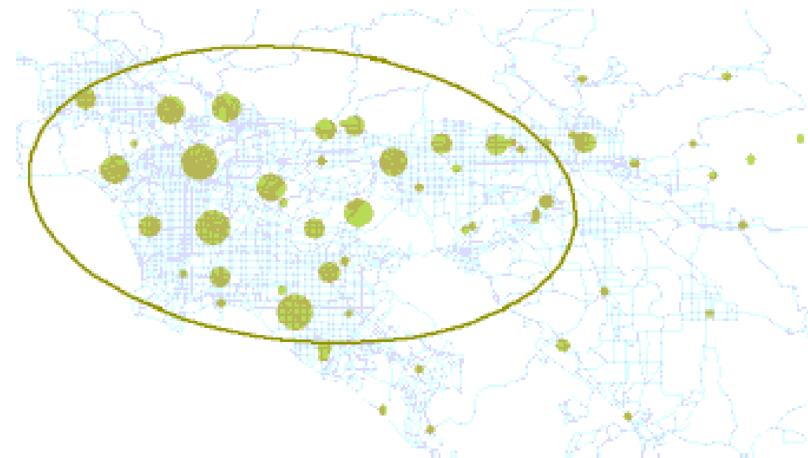
$$\tan\theta = \frac{\left(\sum_{i=1}^{n}\overline{x_i}^2 - \sum_{i=1}^{n}\overline{y_i}^2\right) \pm \sqrt{\left(\sum_{i=1}^{n}\overline{x_i}^2 - \sum_{i=1}^{n}\overline{y_i}^2\right)^2 + 4\left(\sum_{i=1}^{n}\overline{x_i}\,\overline{y_i}\right)^2}}{2\sum_{i=1}^{n}\overline{x_i}\,\overline{y_i}}. \tag{15}$$

Standard Deviational Ellipse

X-Y軸的標準差

$$SDE_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n}}$$

$$SDE_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n}}$$



旋轉角度

$$\tan\ \theta = \frac{A + B}{C}$$

$$A = \left(\sum_{i=1}^{n} \tilde{x}_i^2 - \sum_{i=1}^{n} \tilde{y}_i^2\right)$$

$$B = \sqrt{\left(\sum_{i=1}^{n} \tilde{x}_i^2 - \sum_{i=1}^{n} \tilde{y}_i^2\right)^2 + 4\left(\sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i\right)^2}$$

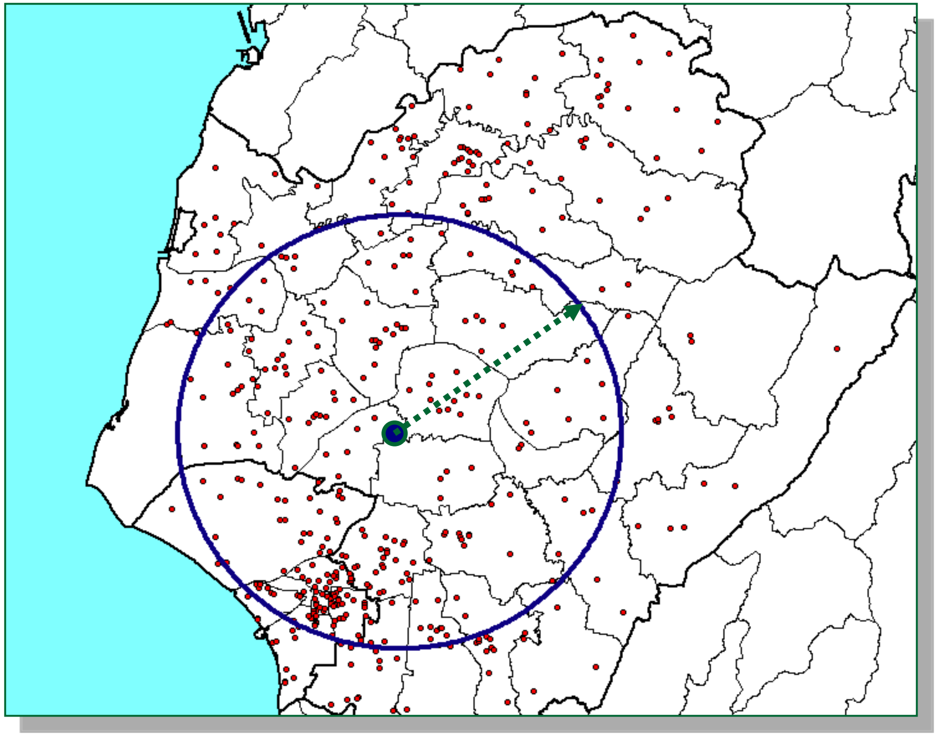
$$C = 2\sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i$$

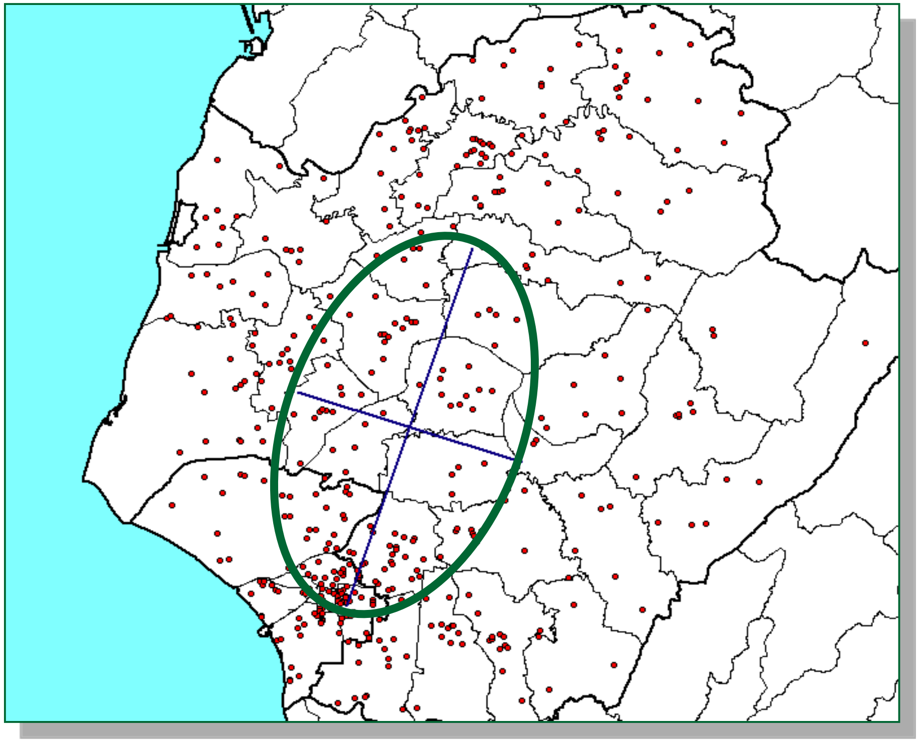
長(短)軸的長度
(標準差)

$$\sigma_x = \sqrt{2}\sqrt{\frac{\sum_{i=1}^{n}(\tilde{x}_i \cos\ \theta - \tilde{y}_i \sin\ \theta)^2}{n}}$$

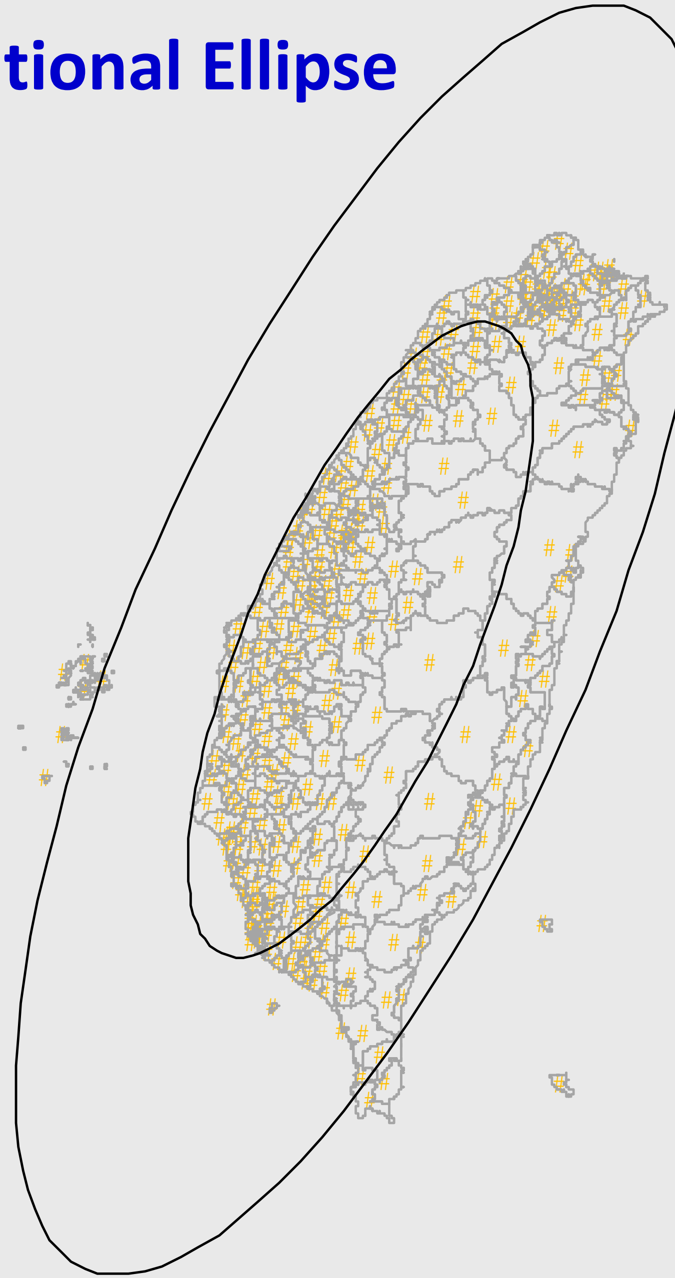
$$\sigma_y = \sqrt{2}\sqrt{\frac{\sum_{i=1}^{n}(\tilde{x}_i \sin\ \theta + \tilde{y}_i \cos\ \theta)^2}{n}}$$

Std Distance vs. Std Deviational Ellipse

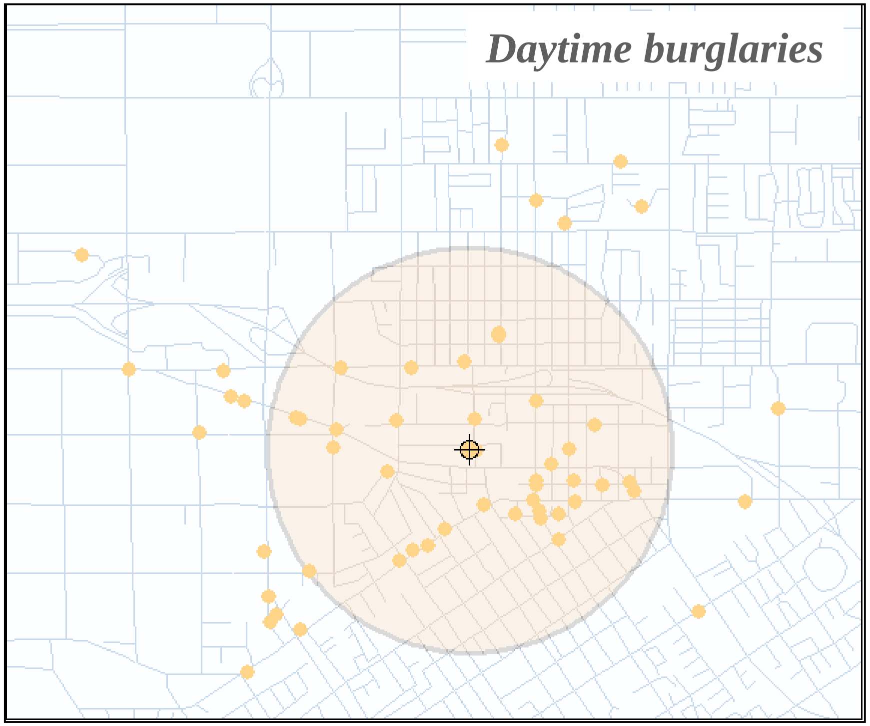


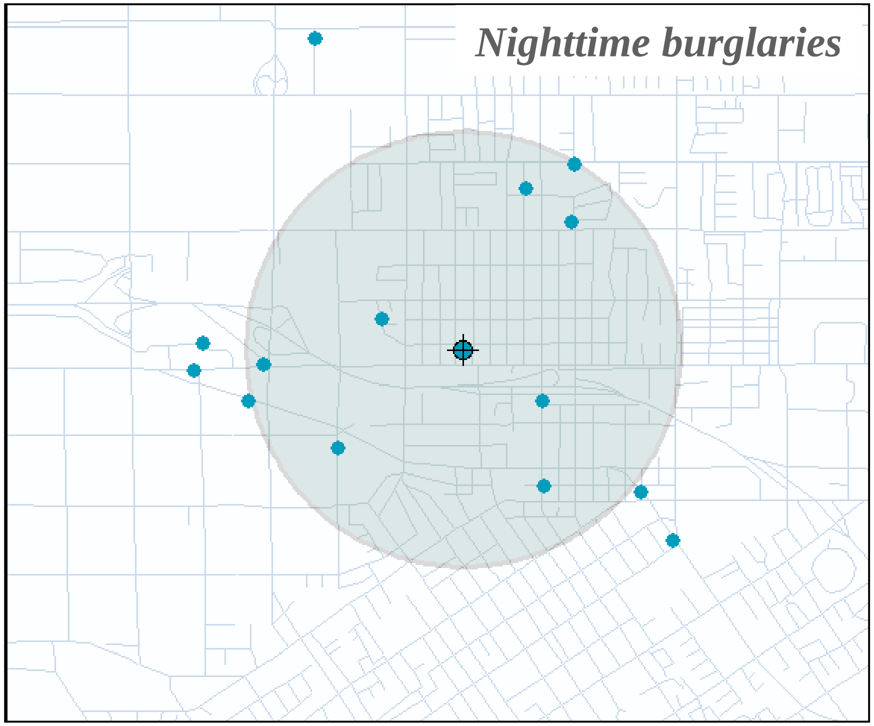


Standard Deviational Ellipse



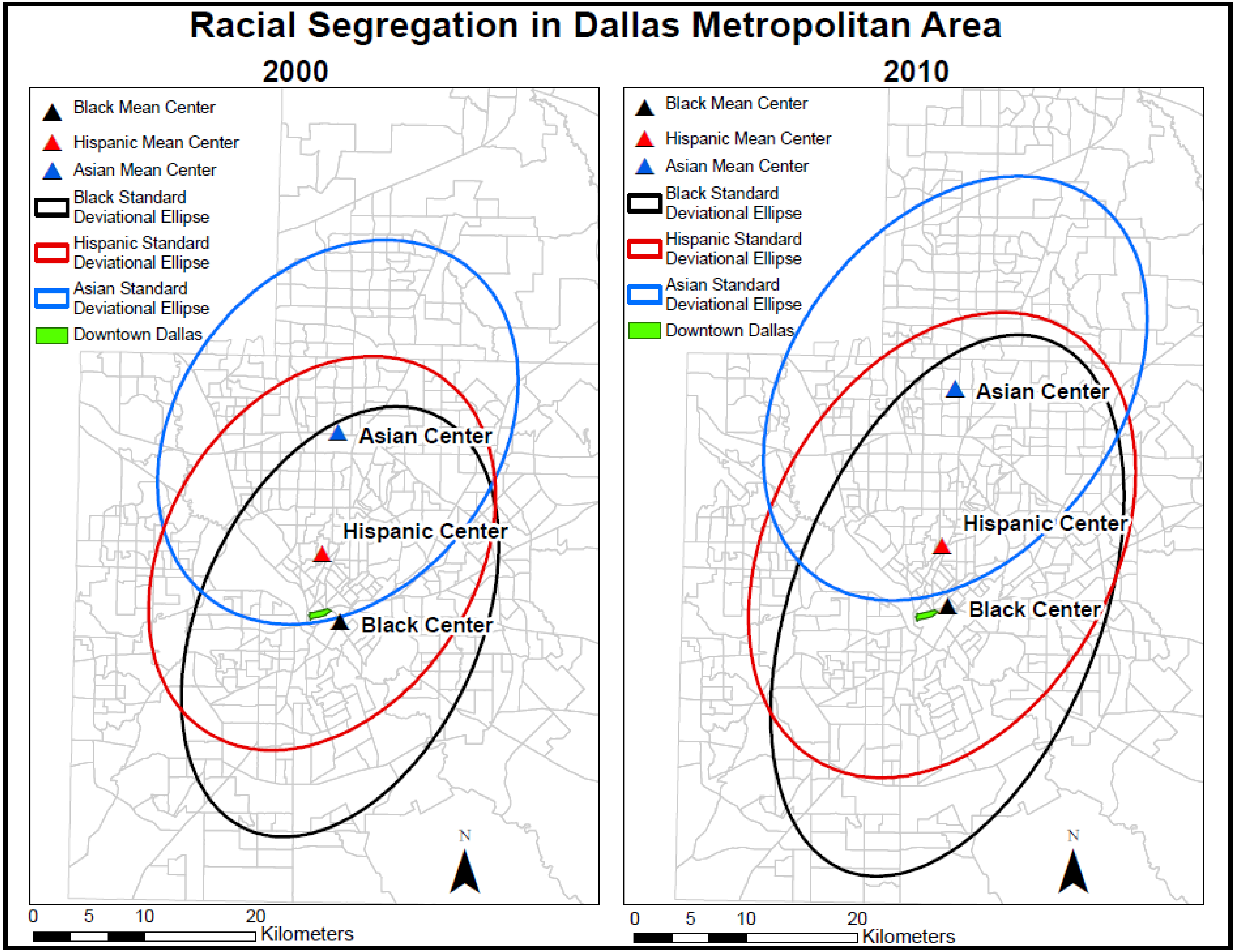
Comparisons of spatial distributions





1. Daytime burglaries tend to be slightly more concentrated

2. The differences in the locations

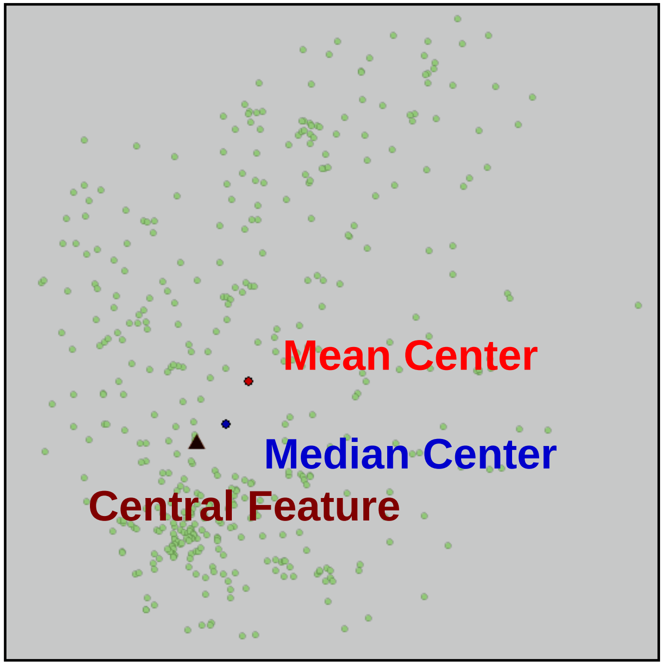


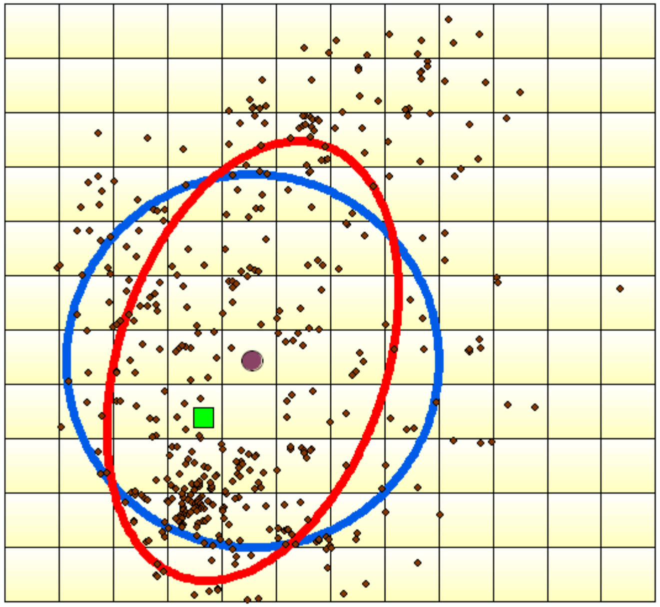
Projection: USA Contiguous Albers Equal Area Conic
Data Source: US Census Tracts

Produced by Heather Cormier on February 5, 2014

Demo：Describing a Point Pattern

- 提供圖資(schools.shp)：
  - 台南地區學校分佈
- 分析方法：
  - Spatial Distribution
    - Spatial Mean Center
    - Median Center
    - Central Feature
    - Standard Distance
    - Standard Deviational Ellipse





Sample Data: schools_sf

	AREA	PERIMETER	SCHOOL_	SCHOOL_ID	NAME	X_coor	Y_coor	ID1	NEAR_FID	NEAR_DIST	geometry
1	0	0	38	39	土城子國小分校	156258.5	2550982	39	29	3293.07480	c(156258.5, 2550982)
2	0	0	81	82	鎮海國小	160475.1	2546815	82	18	668.74035	c(160475.09375, 2546815)
3	0	0	82	83	幼稚園	162749.6	2545007	83	26	327.91552	c(162749.59375, 2545007)
4	0	0	86	87	慈幼高工	171221.4	2543562	87	4	139.67050	c(171221.40625, 2543562)
5	0	0	87	88	母佑幼稚園	171159.1	2543437	88	3	139.67050	c(171159.09375, 2543437)
6	0	0	88	89	牧群幼稚園	171164.3	2543261	89	4	176.07689	c(171164.296875, 2543261)
7	0	0	89	90	復興國中	171283.6	2542581	90	197	666.11873	c(171283.59375, 2542581)
8	0	0	93	94	南興國小	163112.5	2552910	94	12	1562.81303	c(163112.5, 2552910)
9	0	0	98	99	瀛海高中	168313.2	2548684	99	69	756.48794	c(168313.203125, 2548684)
10	0	0	100	101	安和幼稚園	168512.3	2547675	101	69	673.41490	c(168512.296875, 2547675)
11	0	0	101	102	安順國小	169022.0	2550011	102	79	211.31126	c(169022, 2550011)
12	0	0	108	109	學甲托兒所	161381.1	2553384	109	12	240.38739	c(161381.09375, 2553384)

0. Data Preparation

```
# Generating no. of student in each school
schools_sf$Students<-as.integer (runif(424,100,1000))

# Generating school type: cluster vs. isolation
for (i in 1: 424) {

  if (schools_sf$NEAR_DIST[i]< 500) {
    schools_sf$type[i]<- "Cluster"
  } else schools_sf$type[i]<- "Isolation"

}

index<- schools_sf$type == "Cluster"
school_cluster <- schools_sf[index,]

length(school_cluster)
schools.c_lyr <- tm_shape(school_cluster)+tm_dots(col="red", size= 0.1) + tm_layout(frame = F)

School_df <- data.frame(x=schools_sf$X_coor, y=schools_sf$Y_coor,
                        type=schools_sf$type, students=schools_sf$Students)
```

產生連續數值的屬性

產生類別數值的屬性

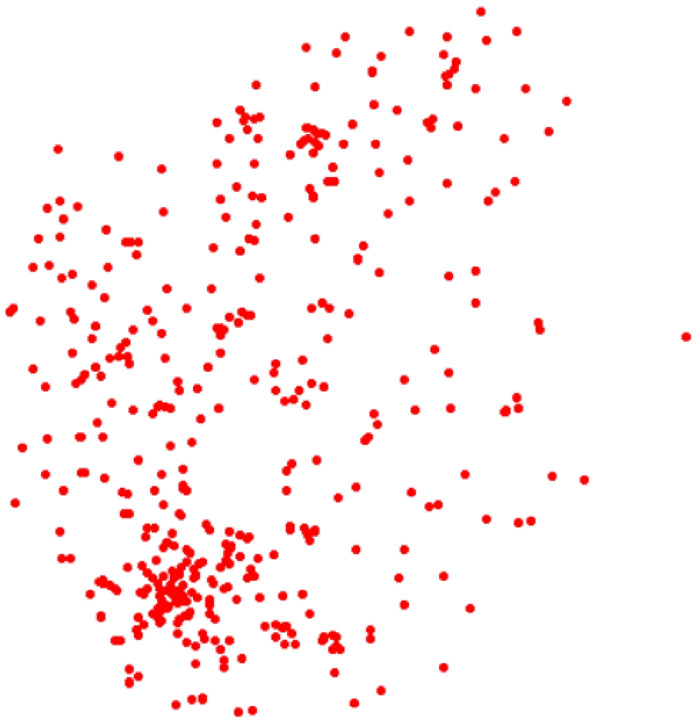
擷取特定屬性的圖資

0. Data Preparation: School_df

```
> School_df
          x       y      type students
1  156258.5 2550982 Isolation      506
2  160475.1 2546815 Isolation      405
3  162749.6 2545007   Cluster      661
4  171221.4 2543562   Cluster      605
5  171159.1 2543437   Cluster      960
6  171164.3 2543261   Cluster      120
7  171283.6 2542581 Isolation      898
8  163112.5 2552910 Isolation      720
9  168313.2 2548684 Isolation      473
10 168512.3 2547675 Isolation      427
11 169022.0 2550011   Cluster      459
12 161381.1 2553384   Cluster      155
```

Comparisons

schools_sf



school_cluster

R packages: aspace

aspace-package {aspace} R Documentation

A collection of functions for estimating centrographic statistics and computational geometries for spatial point patterns

Description

A collection of functions for computing centrographic statistics (e.g., standard distance, standard deviation ellipse, standard deviation box) for observations taken at point locations. Separate plotting functions have been developed for each measure. Users interested in writing results to ESRI shapefiles can do so by using results from aspace functions as inputs to the convert.to.shapefile and write.shapefile functions in the shapefiles library. The aspace library was originally conceived to aid in the analysis of spatial patterns of travel behaviour (see Buliung and Remmel, 2008). Major changes in the current version include (1) removal of dependencies on several external libraries (e.g., gpclib, maptools, sp), (2) the separation of plotting and estimation capabilities, (3) reduction in the number of functions, and (4) expansion of analytical capabilities with additional functions for descriptive analysis and visualization (e.g., standard deviation box, centre of minimum distance, central feature).

Details

Package:	aspace
Type:	Package
Version:	3.2
Date:	2012-08-08
License:	GPL (>= 2.0)

1. Mean and Median Centers

```
Mean.Center <- mean_centre (id=1, points=School_df[,1:2])

Median.Center <- median_centre (id=1, points=School_df[,1:2])
```

建立 sf 格式的point data

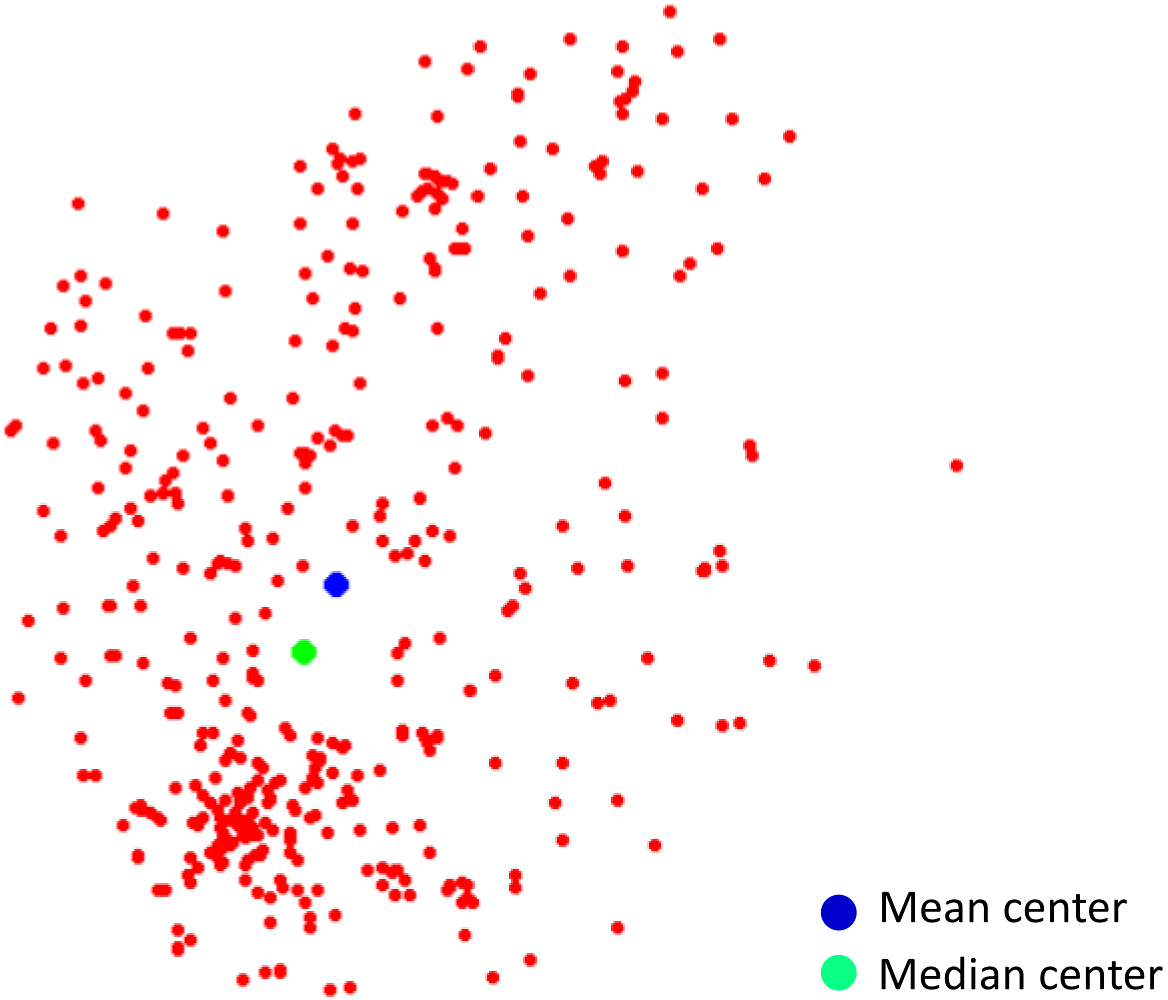
```
Median.Center_sfg = st_point(c(Median.Center[,2],Median.Center[,3]))
Median.Center_sfc = st_sfc(Median.Center_sfg)
Median.Center_sf <- st_sf(Median.Center_sfc)
st_crs(Median.Center_sf) <- st_crs(schools_sf)
tm_shape(Median.Center_sf)+tm_dots(col="green", size= 0.5)
```

補充： %>% 運算元

```
Mean.Center_sfg = st_point(c(Mean.Center[,2],Mean.Center[,3]))
Mean.Center_sfc = st_sfc(Mean.Center_sfg)
Mean.Center_sf <- st_sf(Mean.Center_sfc)
```

```
library(magrittr)
Mean.Center.Coor <- c(Mean.Center[,2],Mean.Center[,3])
Mean.Center_sf <- Mean.Center.Coor %>% st_point %>% st_sfc %>% st_sf
```

Comparisons



2. Weighted Mean Center

```
W.Mean.Center <- mean_centre (id=1, weighted=TRUE, weights=School_df$students,
                              points=School_df[,1:2])

W.Mean.Center.Coor <- c(W.Mean.Center[,2],W.Mean.Center[,3])
W.Mean.Center_sf <- W.Mean.Center.Coor %>% st_point %>% st_sfc %>% st_sf
st_crs(W.Mean.Center_sf) <- st_crs(schools_sf)
tm_shape(W.Mean.Center_sf)+tm_dots(col="green", size= 0.5)
```

3. SDD: Standard Distance

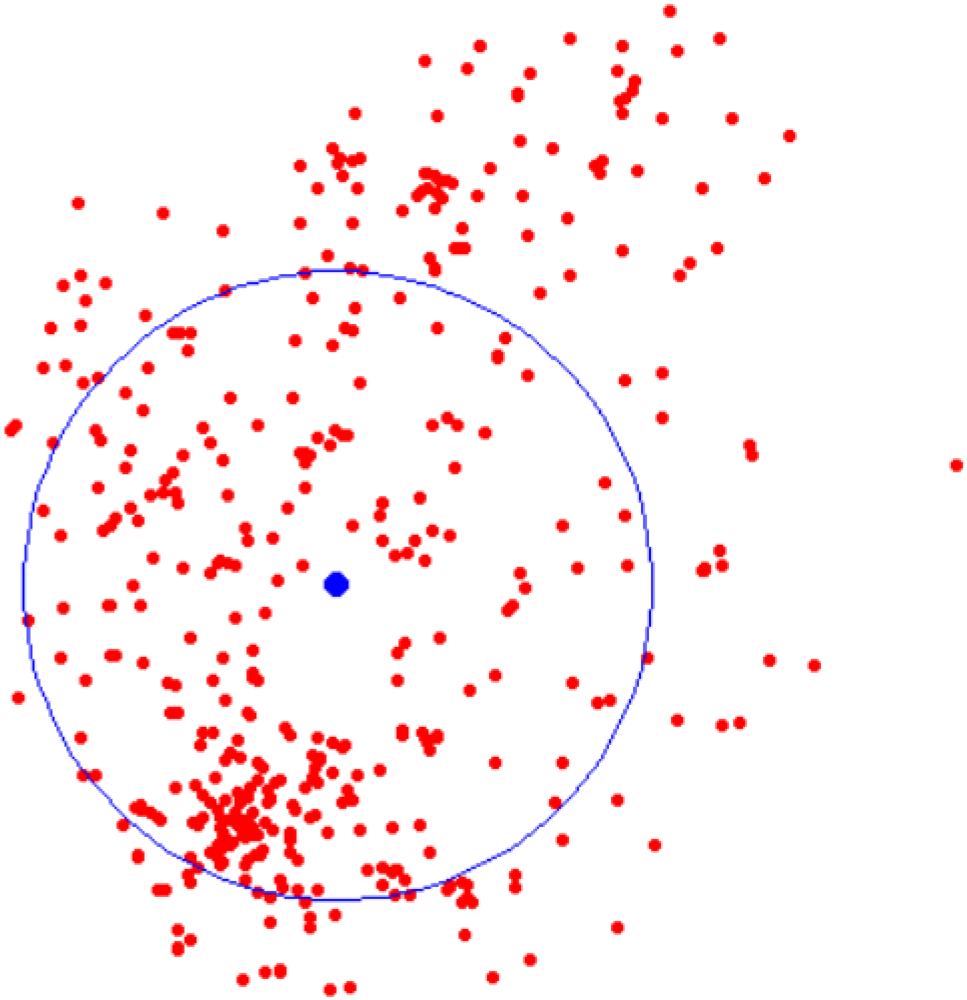
school.SDD<- calc_sdd(id=1, points=School_df[,1:2])

school.SDD	1 obs. of 7 variables
id : num 1	
calccentre: logi TRUE	
weighted : logi FALSE	
CENTRE.x : num 173824	
CENTRE.y : num 2557130	
SDD.radius: num 17225	
SDD.area : num 9.32e+08	

Mapping Standard Distance: R code

```
center.x<- school.SDD$CENTRE.x
center.y<- school.SDD$CENTRE.y
center.coor <- c(center.x, center.y)
center_sf<- center.coor %>% st_point %>% st_sfc %>% st_sf
st_crs(center_sf) <-  st_crs(schools_sf)

rad<- school.SDD$SDD.radius
SD_sf<- st_buffer(center_sf, rad)
SD_lyr <- tm_shape(SD_sf) + tm_borders(col = "blue") +
            tm_layout(frame = F)
Center_lyr <-  tm_shape(center_sf) + tm_dots(col="blue", size= 0.5)

schools_lyr+ Center_lyr+ SD_lyr
```

Mapping Standard Distance（results）



Weighted Standard Distance

```
school.SDD2<- calc_sdd(id=1, points = School_df[,1:2],
                        weighted = TRUE, weights=School_df$students)
```

```
school.SDD2      1 obs. of 7 variables
  id : num 1
  calccentre: logi TRUE
  weighted : logi TRUE
  CENTRE.x : num 173848
  CENTRE.y : num 2557236
  SDD.radius: num 17425
  SDD.area : num 9.54e+08
```

Export SDD to Shapefile

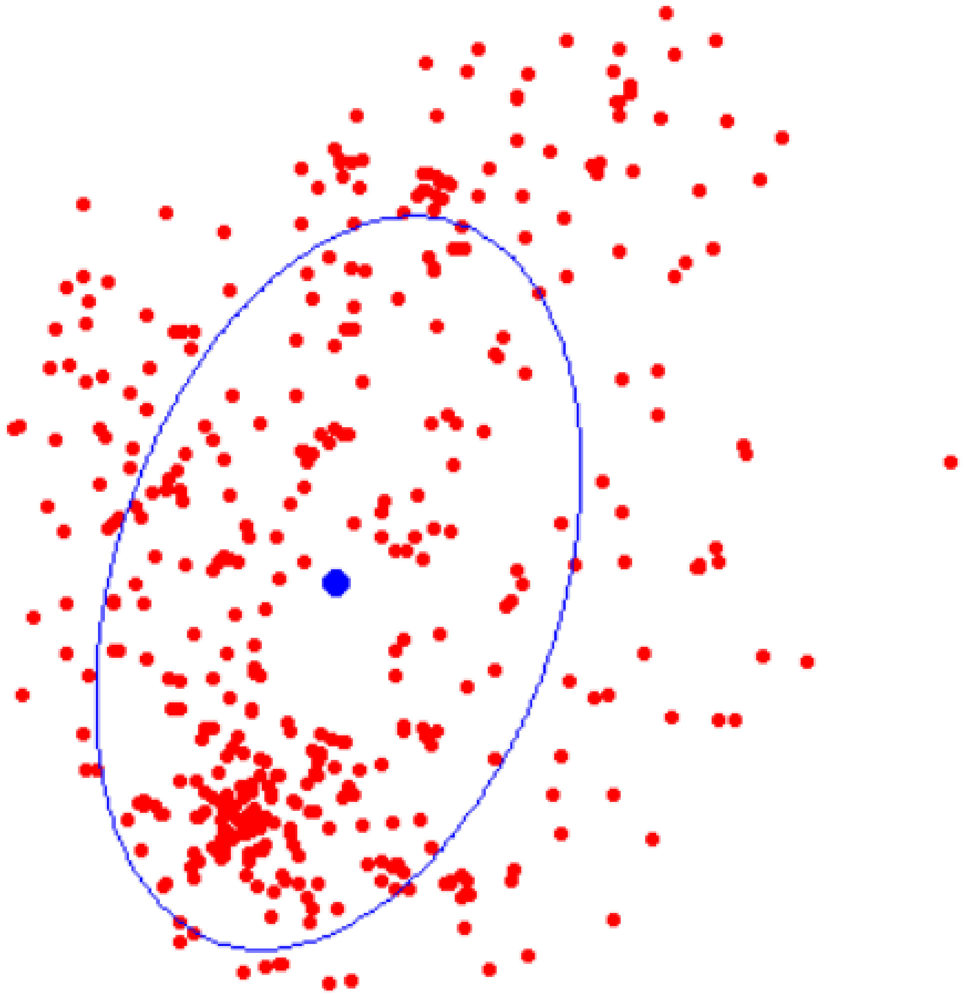
```
# ESRI Shapefile type 1=point, 3=polyLine, 5=polygon
shp <- convert.to.shapefile( sddloc, sddatt, "id", 5)

write.shapefile(shp, "SDD_Shape", arcgis=T)


# reading shp and mapping
SDD_Shape_sf <- st_read("SDD_Shape.shp")
st_crs(SDD_Shape_sf) <-  st_crs(schools_sf)
schools_lyr+ tm_shape(SDD_Shape_sf) +
                  tm_borders(col = "blue")+tm_layout(frame = F)
```

4. SDE: Standard Deviational Ellipse

school.SDE <- calc_sde(id=1, points=School_df[,1:2])

sdeatt	1 obs. of 19 variables
id : num 1	
CALCCENTRE : logi TRUE	
weighted : logi FALSE	
CENTRE.x : num 173824	
CENTRE.y : num 2557130	
Sigma.x : num 12377	
Sigma.y : num 20982	
Major : chr "SigmaY"	
Minor : chr "SigmaX"	
Theta : num 18.4	
Eccentricity : num 0.807	
Area.sde : num 8.16e+08	
TanTheta : num 0.332	
SinTheta : num 0.315	
CosTheta : num 0.949	
SinThetaCosTheta: num 0.299	
Sin2Theta : num 0.0993	
Cos2Theta : num 0.901	
ThetaCorr : num 18.4	

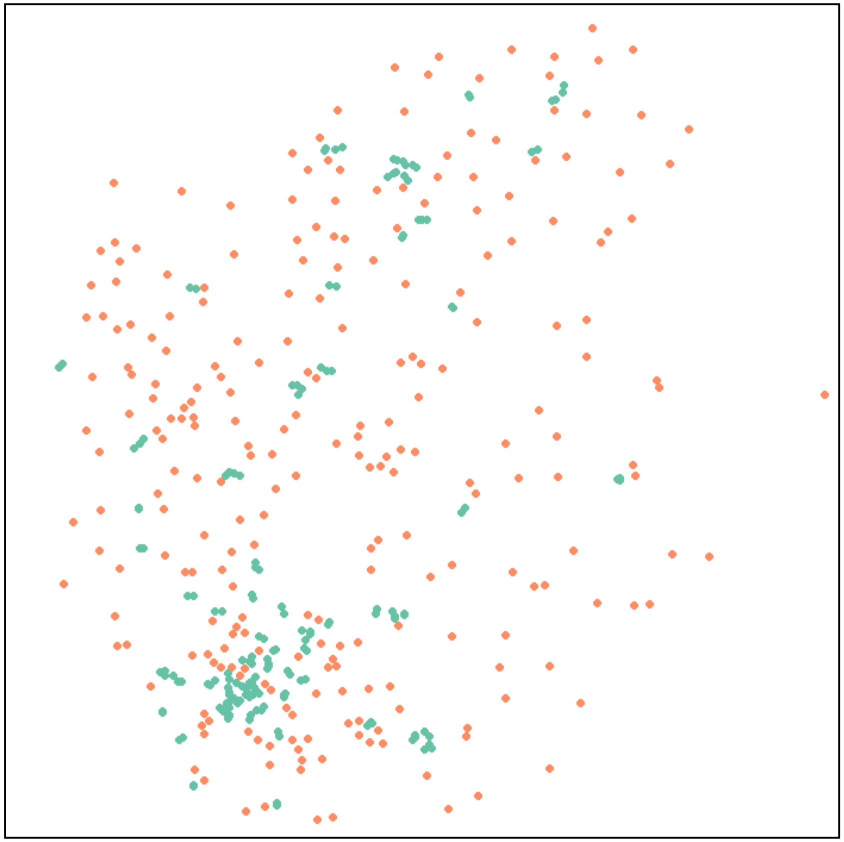
Mapping Standard Deviational Ellipse

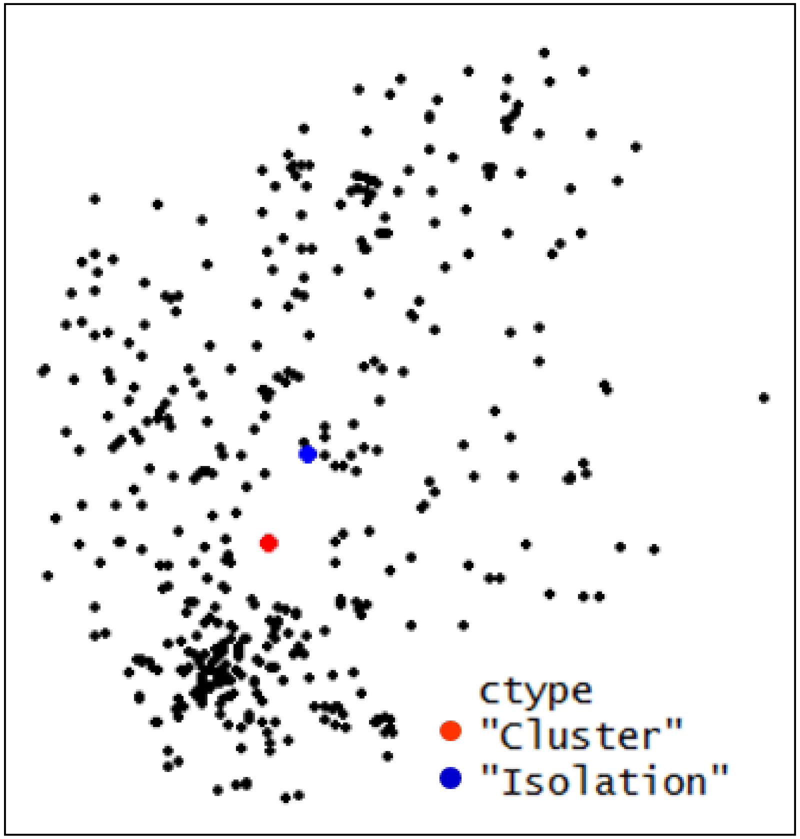


5. Central Feature

```
school.CF<- CF(id=1, points=School_df[,1:2])

CF.x<- school.CF$CF.x
CF.y<- school.CF$CF.y
CF.coor <- c(CF.x, CF.y)
CF_sf <- CF.coor %>% st_point %>% st_sfc %>% st_sf
st_crs(CF_sf) <-  st_crs(schools_sf)

CF_lyr <- tm_shape(CF_sf) + tm_dots(col = "blue", size= 0.5) +
            tm_layout(frame = F)

schools_lyr + CF_lyr
```

6. Mean centers by grouping attributes





Mean centers by grouping attributes

```
type <- schools_sf$type
newid<- unique(type)

xx<-vector(); yy<-vector(); ctype<-vector()
for (i in 1:2){
  index<-(type == newid[i])
  newschool<-schools_sf[index, ]
  xcoor<-newschool$X_coor
  ycoor<-newschool$Y_coor
  newschool.mc <- mean_centre(id=1, points=cbind(xcoor, ycoor))
  xx[i]<-newschool.mc$CENTRE.x
  yy[i]<-newschool.mc$CENTRE.y
  ctype[i]<-newid[i]
}
```

Mean centers by grouping attributes (cont'd)

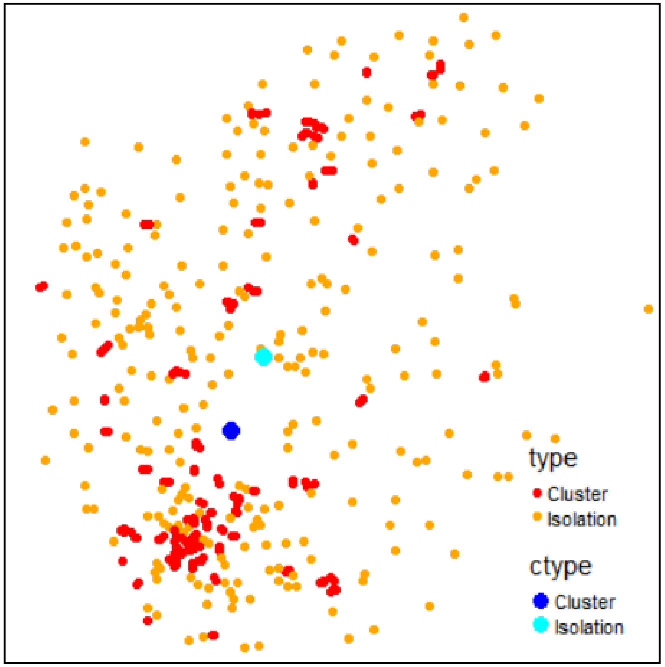
```
newcenterxy <- data.frame(xx,yy, ctype)
New_sf <- st_as_sf(newcenterxy , coords=c("xx","yy"))
st_crs(New_sf) <-  st_crs(schools_sf)
```

```
> newcenterxy
        xx      yy     ctype
1 174955.9 2559734 Isolation
2 172181.3 2553352   Cluster
```

```
> New_sf
Simple feature collection with 2 features and 1 field
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: 172181.3 ymin: 2553352 xmax: 174955.9 ymax: 2559734
Projected CRS: Transverse Mercator
      ctype                 geometry
1 Isolation POINT (174955.9 2559734)
2   Cluster POINT (172181.3 2553352)
```

Mean centers by grouping attributes (cont'd)

```
New_lyr <- tm_shape(New_sf) +
tm_dots(col="ctype", palette=c(Cluster='blue', Isolation='cyan'), size= 0.5)
+ tm_layout(frame = F)

schools_lyr2 <- tm_shape(schools_sf) +
tm_dots("type", palette=c(Cluster='red', Isolation='orange'), size= 0.1) +
tm_layout(frame = F)

schools_lyr2 + New_lyr
```



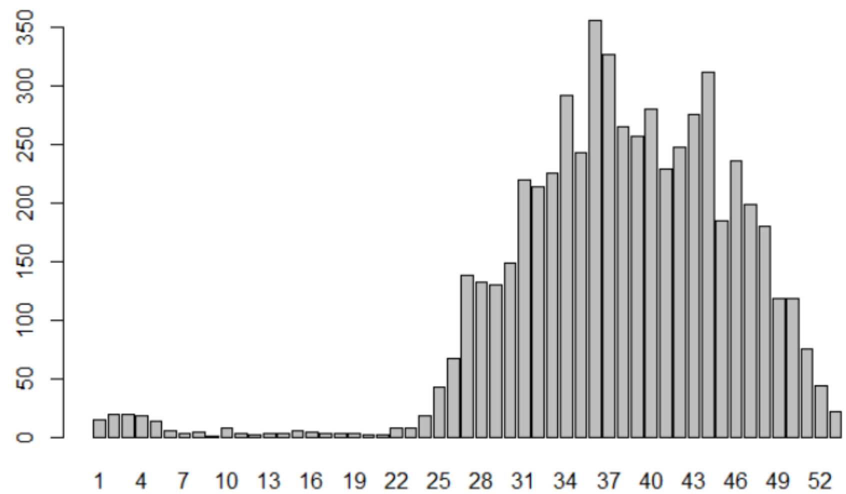
實習：描述疾病擴散的時空趨勢

- 圖資 **point_event.shp**
  - 登革熱病例分佈 (僅為實習為目的，不是真實資料)

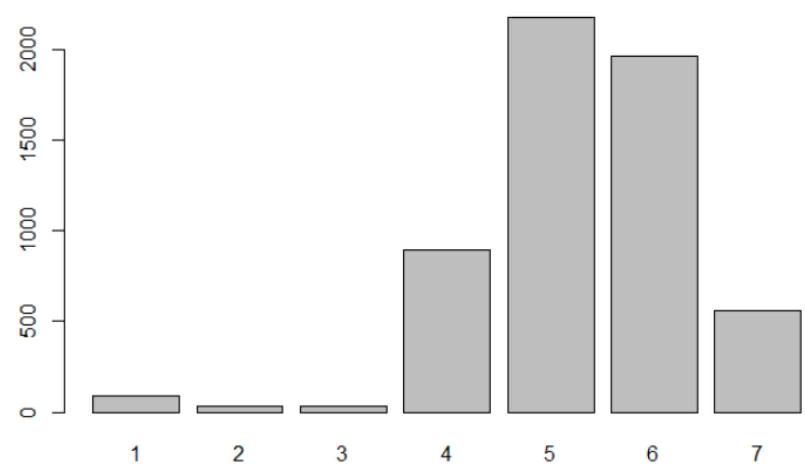
```
> Point.Attr
   WEEK      X       Y AGE KEYID coords.x1 coords.x2 optional
1    23 328891 2730820  50     1    328891   2730820     TRUE
2     3 325963 2738218  59     2    325963   2738218     TRUE
3    34 320039 2775558  26     3    320039   2775558     TRUE
4    42 313465 2611430  71     4    313465   2611430     TRUE
5    22 312964 2772731  57     5    312964   2772731     TRUE
6     4 312207 2774350  39     6    312207   2774350     TRUE
7    35 311334 2658377  32     7    311334   2658377     TRUE
8    29 310825 2775259  41     8    310825   2775259     TRUE
9    42 309045 2771607  30     9    309045   2771607     TRUE
10   40 308939 2607900  13    10    308939   2607900     TRUE
```

Task 1: Exploring **temporal trends** in different time-scales

By week



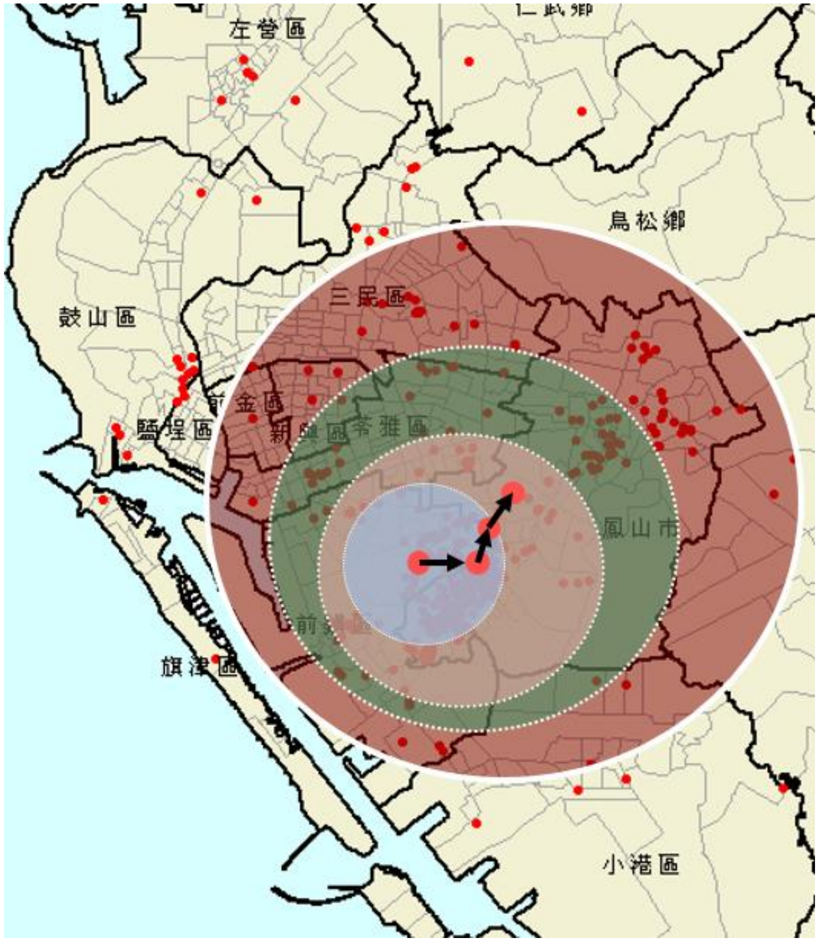
By period



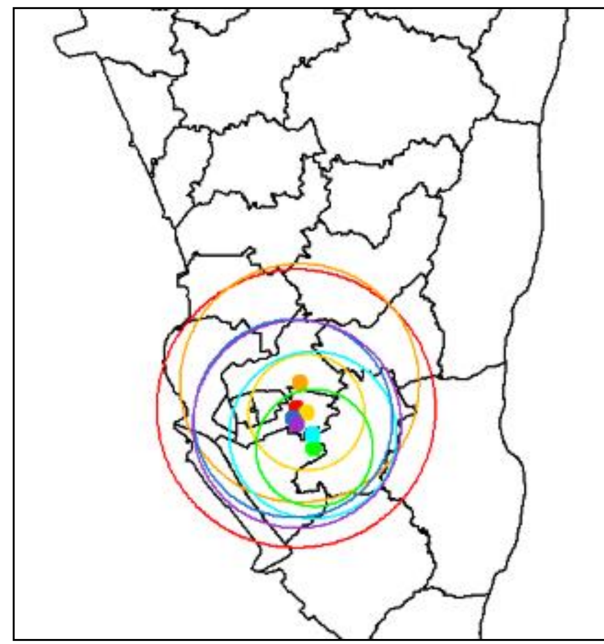
week 01~07: period 1
week 08~15: period 2
week 16~23: period 3
week 24~31: period 4
week 32~39: period 5
week 40~47: period 6
week 48~55: period 7

Task 2: Exploring **spatial trends** in different periods

擷取高雄地區的登革熱病例分佈



參考結果

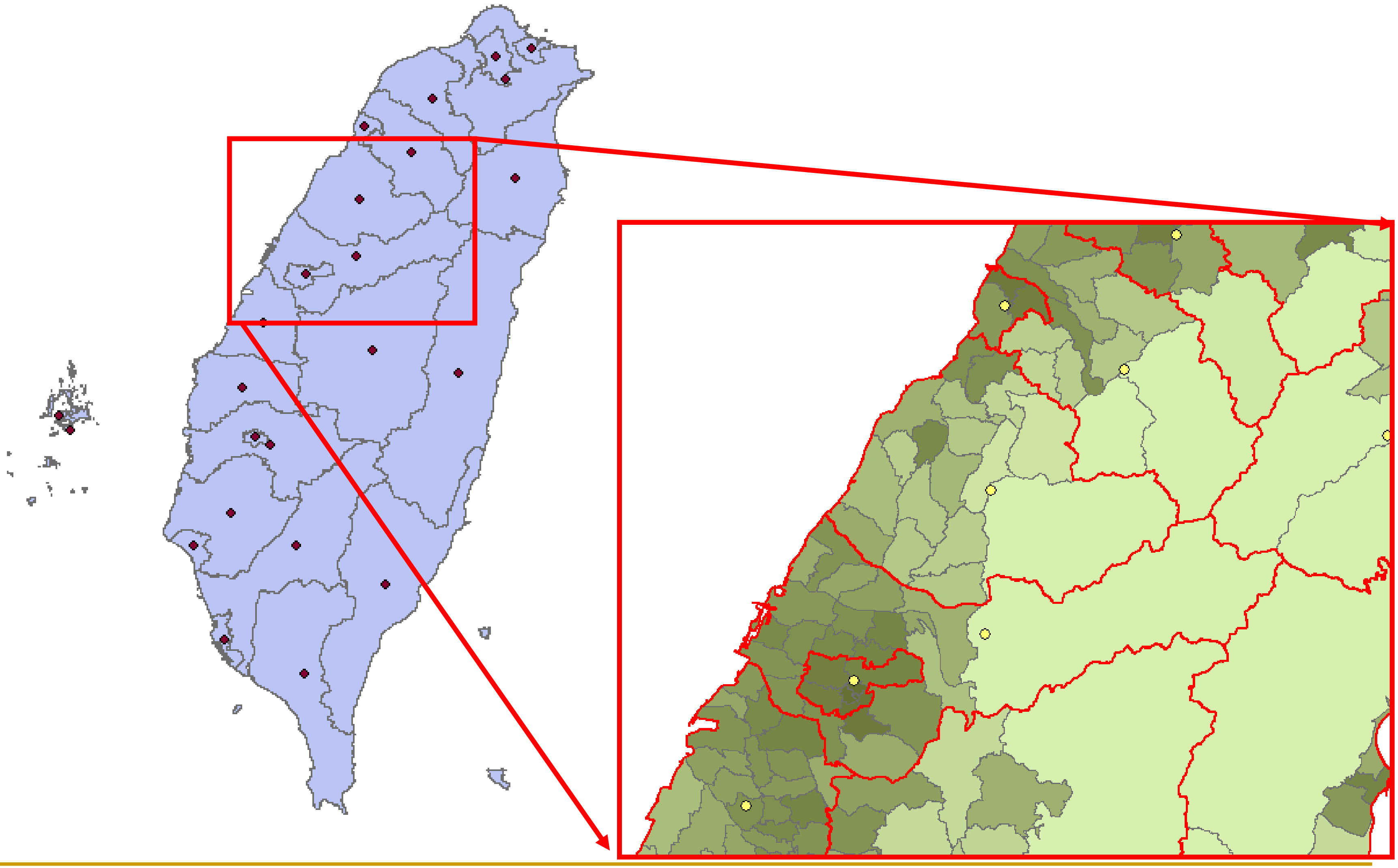


本週作業：計算各縣市人口中心點

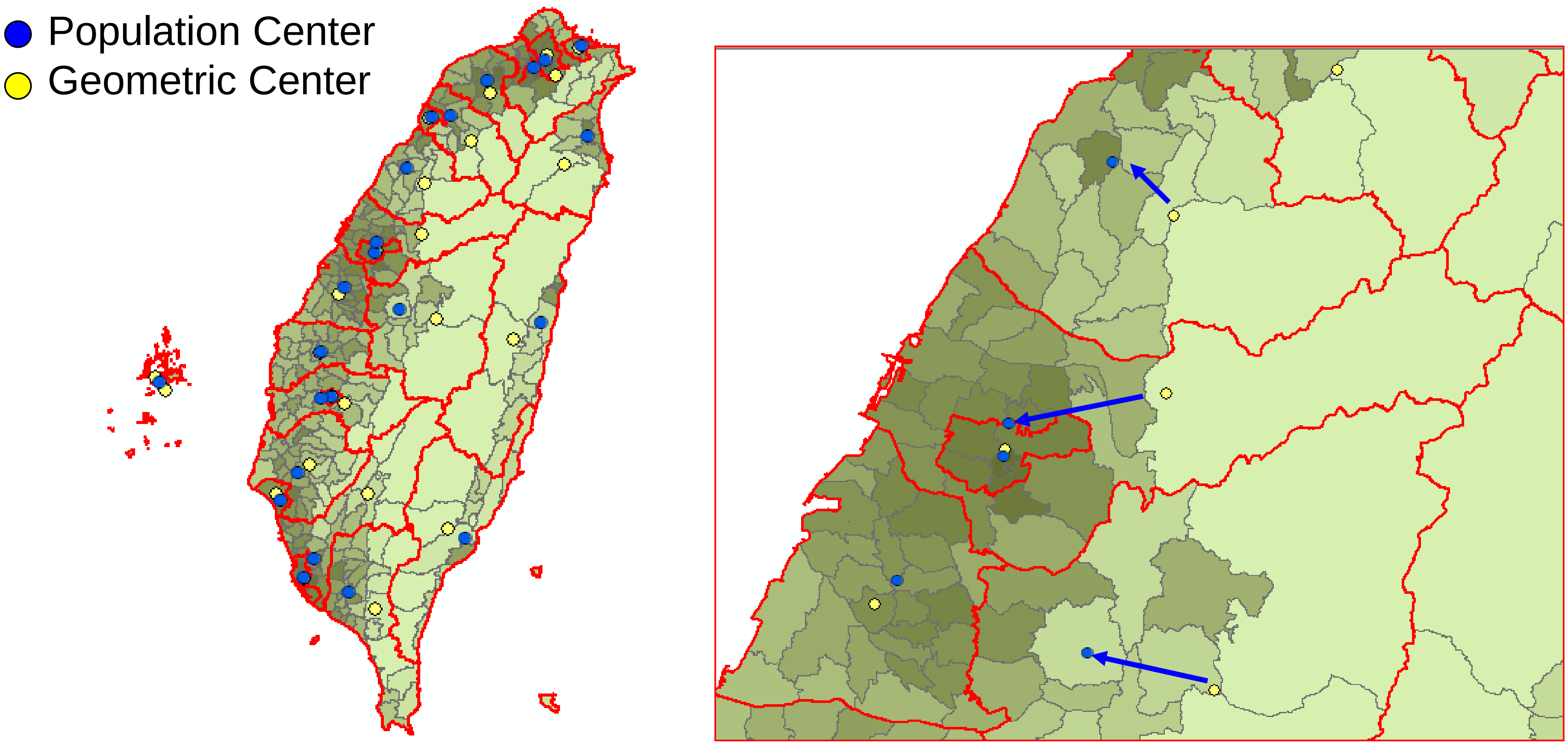
- 圖資：
  - 台灣鄉鎮人口 **Popn_TWN2.shp**
- 範例內容:
  - 產生台灣縣市邊界，並計算各縣市幾何中心
  - 利用台灣鄉鎮人口加權方式，產生各縣市的人口中心點
  - 比較幾何中心點與人口中心點的差異

繳交期限：4/26 下午2:00

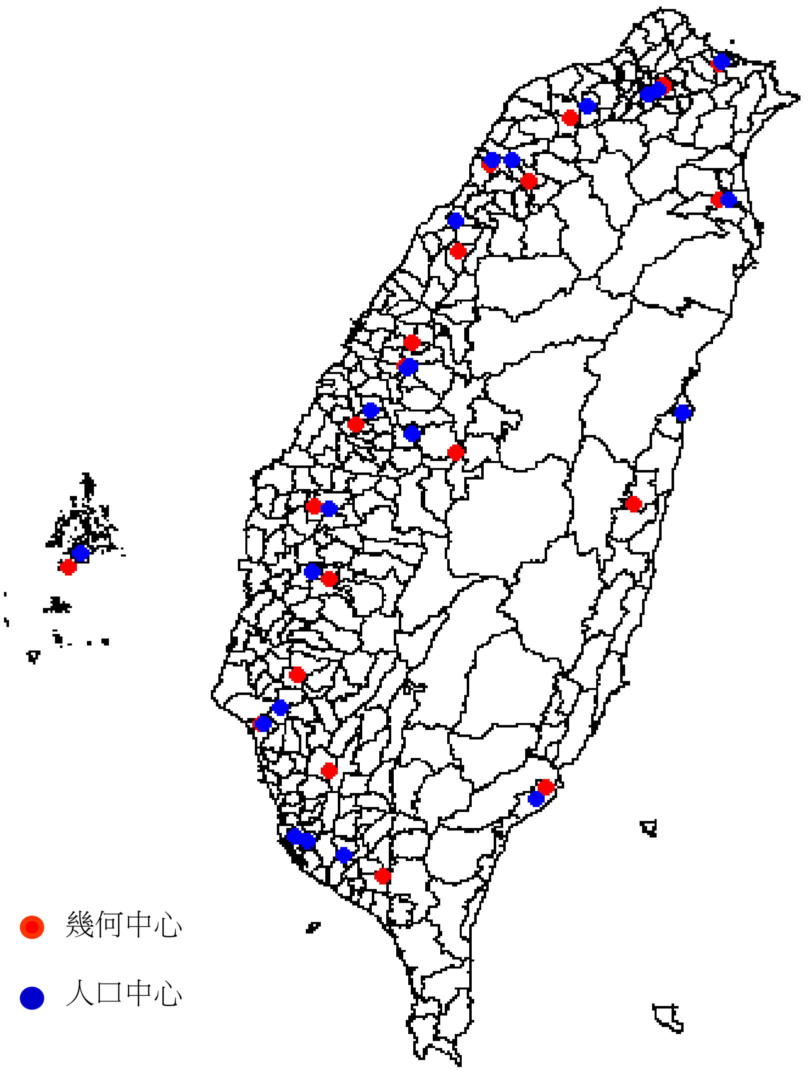
產生台灣各縣市的幾何中心點



利用各鄉鎮人口加權，產生各縣市的人口中心點



預期結果：比較各縣市的人口中心與幾何中心



提醒：4/19 期中考 (open-book)

Midterm Exam

- 考試範圍：
  - **Cartography**: projected coordinate system; thematic mapping
  - **GIS operations**: fishnet; buffer zones; intersections; data join; data summarized by attributes or locations
  - **Basic statistics**: probability distributions; inferential statistics
  - **R Programming**: user-defined functions
  - **Describing point patterns**: centrality and dispersions