



# ggplot 與 資料處理

空間分析 2020.03.08  
TA 杜承軒

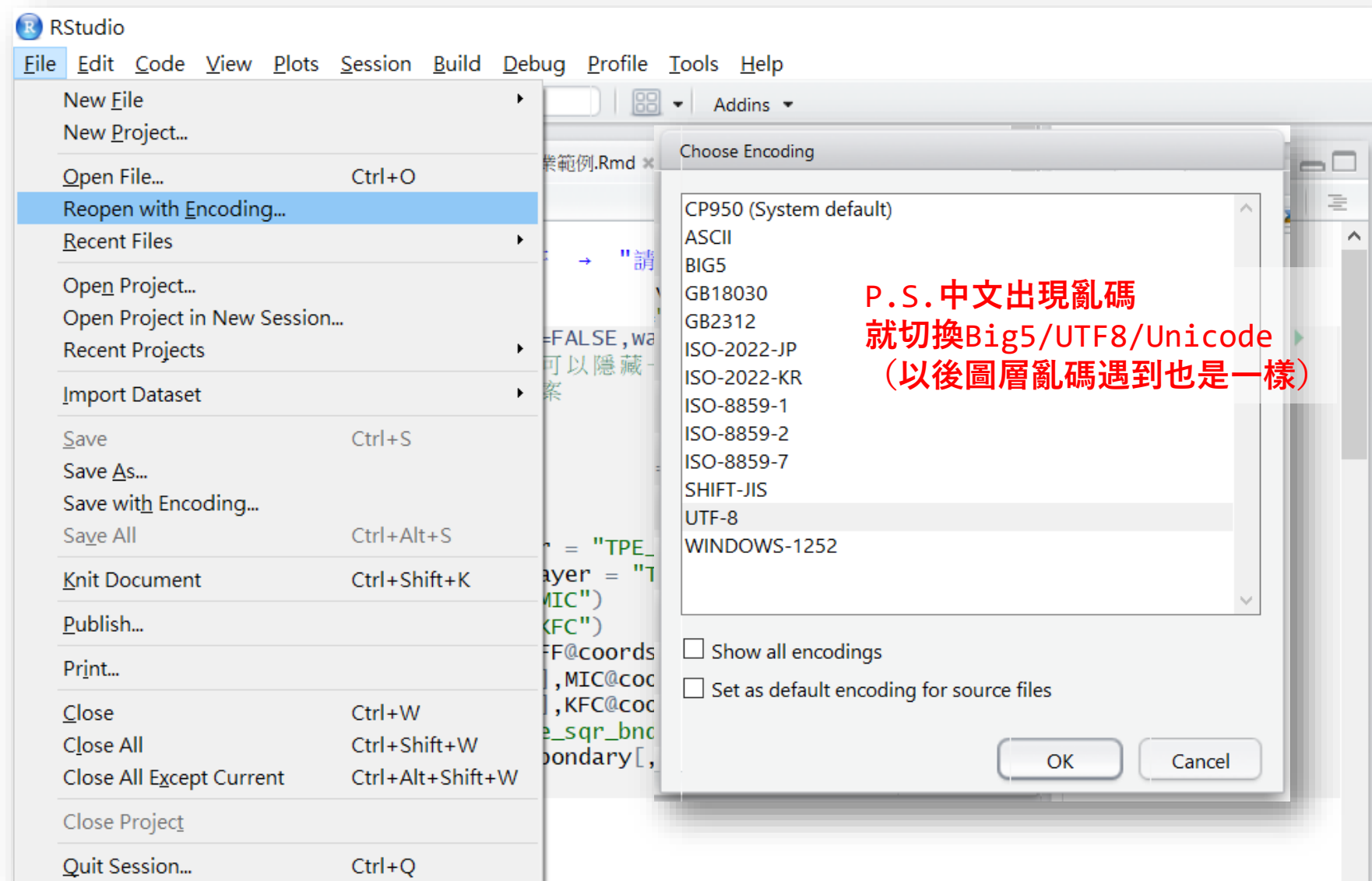


Tpe\_Fastfood  
加入行政區、行政區ID

## RMarkdown 程式碼繳交格式

請搭配 RMarkdown範例.zip

P.S. 範例中的 .Rmd檔 如果亂碼，請依以下步驟操作



## RMarkdown 程式碼繳交格式

路徑請盡量  
避免使用中文  
(善用捷徑)

第一個R段落可以先呼叫library和讀檔案  
為了隱藏不需要的訊息可以使用 `results/message/warning`

```
```\{r results='hide', message=FALSE, warning=FALSE}  
library(GISTools);library(rgdal);library(splancs)  
setwd("D:/1082SA/Data")  
tpe = readOGR(dsn = ".", layer = "Vill",encoding="utf8",verbose = F)  
```
```

讀完shapefile  
不會再回傳訊息



```
#上面的results/message/warning可以隱藏一些  
#可以在這個段落跑library和讀檔案  
library(GISTools)  
library(rgdal)  
library(splancs)  
setwd("D:/1072SA/Data") #路徑請盡量不使用  
#讀取資料與格式轉換  
tpe=readOGR(dsn = ".", layer = "Taipei_V  
FF=readOGR(dsn = ".", layer = "Tpe_Fastf
```

```
MIC=subset(FF,FF@data$STORE=="MIC")  
KFC=subset(FF,FF@data$STORE=="KFC")  
FF.pt=as.points(FF@coords[,1],FF@coords[  
MIC.pt=as.points(MIC@coords[,1],MIC@coor  
KFC.pt=as.points(KFC@coords[,1],KFC@coor
```



```
library(splancs)
```

```
## Warning: package 'splancs' was built under R version 3.4.4
```

```
##  
## Spatial Point Pattern Analysis Code in S-Plus  
##  
## Version 2 - Spatial and Space-Time analysis
```

```
setwd("D:/1072SA/Data") #路徑請盡量不使用中文  
#讀取資料與格式轉換  
tpe=readOGR(dsn = ".", layer = "Taipei_Vill", encoding="utf8")
```

```
## OGR data source with driver: ESRI Shapefile  
## Source: ".", layer: "Taipei_Vill"  
## with 456 features  
## It has 9 fields  
## Integer64 fields read as strings: CENSUS FASTFOOD
```

## RMarkdown 程式碼繳交格式

- > Q1.
- > 題號請用">"當前綴來標示

換行最後請空2~4格(讓它辨識你要換行)  
或打<br/>(支援html語法)

\*\*\*

- > Q2. 題目通常都是畫圖和解釋

```
```{r}
#回答的時候請保留程式碼和結果以利批改
plot(tpe, col="#DDDDDD")
pointmap(MIC.pt, pch=21, bg="yellow", add=T)
pointmap(KFC.pt, pch=21, bg="red", add=T)
```
```

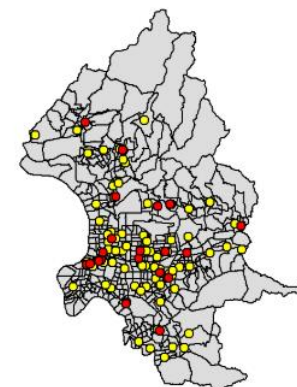
※不需要留不必要的資訊(e.g. 列出data的內容)

Q1.以下有些RMarkdown的教學  
題號請用">"當前綴來標示

一題結束可以用分隔線隔開

Q2.題目通常都是畫圖和解釋

```
#回答的時候請保留程式碼以利批改
plot(tpe, col="#DDDDDD")
pointmap(MIC.pt, pch=21, bg="yellow", add=T)
pointmap(KFC.pt, pch=21, bg="red", add=T)
```



## 隨堂小考

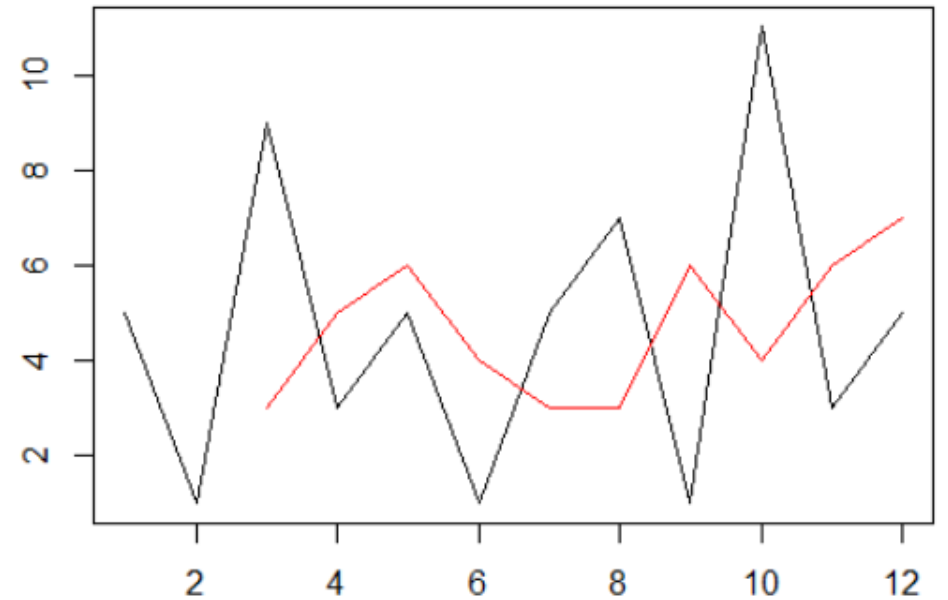
函數：MA

引數：(vector) x、(numeric) n

```
MA=function(x,n=3){  
  m=length(x)  
  if(n>=m) return("error")  
  SMA=c()  
  for(i in (n+1):m)  
    SMA=c(SMA,mean(x[(i-n):(i-1)]))  
  plot(c(1:m), x, type = 'l',ann=F)  
  lines(c((n+1):m), SMA, col=2)  
  return(SMA)  
}
```

- ❶ 計算並回傳移動平均
- ❷ 預設  $n = 3$
- ❸  $n \geq x$  的長度，中止並回傳 "error"
- ❹ 畫出原始數列(黑)和移動平均(紅)

```
> value=c(5,1,9,3,5,1,5,7,1,11,3,5)  
> MA(x=value,n=2)  
[1] 3 5 6 4 3 3 6 4 6 7
```



## 隨堂小考

- ① 移動平均
- ② 預設  $n = 3$
- ③ error中止
- ④ 繪圖

MA=function(x,n=3){

m=length(x)  
if(n>=m) { return("error") }

①

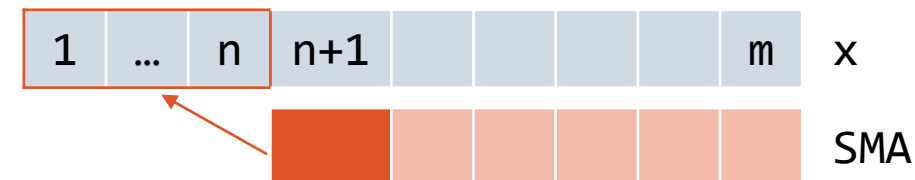
SMA=c()  
for(i in (n+1):m) {  
    SMA=c(SMA,mean(x[(i-n):(i-1)]))  
    現在在第i個位置，取出前1~n個  
}

plot(c(1:m), x, type = 'l', col='black')  
lines(c((n+1):m), SMA, col='red')

return(SMA)

}

③



或

SMA=c()  
for(i in 1:(m-n)) {  
    SMA[i]=mean(x[i:(i+n-1)])  
    第i筆資料，加總從i開始的n個  
}

④

- **factor → number**

問題：將數字辨識成factor格式

(x) 直接轉成numeric

(o) 先轉成character再轉成numeric

```
> fee
[1] 1245 2567 432 135 742
Levels: 135 432 742 1245 2567
> as.numeric(fee)
[1] 4 5 2 1 3
> as.numeric(as.character(fee))
[1] 1245 2567 432 135 742
```

- **xtabs()：樞紐分析表**

xtabs(sum~group)

i.e. xtabs(~TOWN+STORE)

↑  
沒有參數→代表count

- **aggregate(欄位, by=group, FUN=mean)**



# R資料處理

- 選取：
  1. 邏輯判斷
  2. which
  3. 直接使用欄位對位選取
  4. `subset(data, 判斷式)`

常用表格資料

逐列偵測

`%in%`：判斷是否在其中

```
mic      1 3 5 8 9 4 6
kfc      4 7 3 7 2 1 5
mic>5    F F F T T F T
x=which(mic>5)           4 5 7
kfc[x]           7 2 5
kfc[mic>5]  → 7 2 5
subset(kfc,mic>5) → 7 2 5
```

```
fast[mic>5,]
subset(fast,mic>5)
```

|   | mic | kfc |
|---|-----|-----|
| 4 | 8   | 7   |
| 5 | 9   | 2   |
| 7 | 6   | 5   |

```
> fast
  mic kfc
1   1   4
2   3   7
3   5   3
4   8   7
5   9   2
6   4   1
7   6   5
```

- 選取出密度大於10的圖徵：

```
index = TPE$dens > 10
high = TPE[index,]
low = TPE[!index,]

high = subset(TPE, index)
```

readOGR：向量資料使用readOGR讀取，在R中的格式為Spatial\*DataFrame，簡稱sp

setwd("D:/1082SA/Data") #設定路徑

TPE=**readOGR**(dsn = ".", layer = "Vill", encoding="utf8", use\_iconv=T ,**verbose=F**)  
或

TPE=readOGR(dsn = "Vill.shp", encoding="utf8", verbose=F)

"/" 當前資料夾

"/" 當前資料夾的上層

```
setwd("D:/1082SA")
```

```
TPE=readOGR(dsn = "Data", layer = "Vill")
```

- TPE@data

屬性工作表（格式data.frame）

可用\$呼叫欄位：TPE@data\$ID（直接 **TPE\$ID** 也可以）

- TPE@proj4string or proj4string(TPE)

CRS arguments:

+proj=tmerc +lat\_0=0 +lon\_0=121 +k=0.9999 +x\_0=250000 +y\_0=0

+ellps=GRS80 +units=m +no\_defs

- FastFood@coords or coordinates(FastFood)

點資料的x,y座標

- poly.areas(TPE)

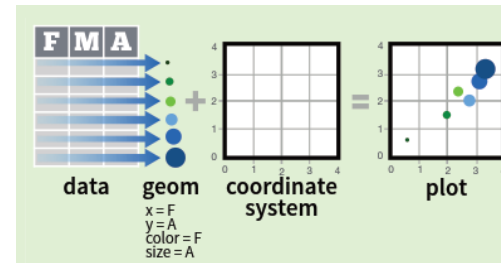
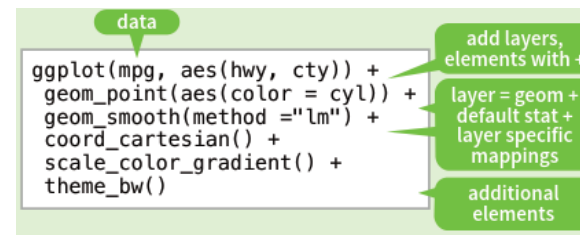
面資料的面積

# ggplot

請搭配 [ggplot2-cheatsheet.pdf](#)

```
ggplot(data, aes(x=..., y=...)) +  
  #放在也geom中可以
```

```
geom_xxx( ) +  
scale_xxx_xxx( ) +  
labs() + #座標軸  
theme() #主題
```



### One Variable

#### Continuous

`a <- ggplot(mpg, aes(hwy))`

- `a + geom_area(stat = "bin")`  
x, y, alpha, color, fill, linetype, size  
b + geom\_area(aes(y = ..density..), stat = "bin")
- `a + geom_density(kernel = "gaussian")`  
x, y, alpha, color, fill, linetype, size, weight  
b + geom\_density(aes(y = ..county..))
- `a + geom_dotplot()`  
x, y, alpha, color, fill
- `a + geom_freqpoly()`  
x, y, alpha, color, linetype, size  
b + geom\_freqpoly(aes(y = ..density..))
- `a + geom_histogram(binwidth = 5)`  
x, y, alpha, color, fill, linetype, size, weight  
b + geom\_histogram(aes(y = ..density..))

#### Discrete

`b <- ggplot(mpg, aes(fl))`

- `b + geom_bar()`  
x, alpha, color, fill, linetype, size, weight

### Two

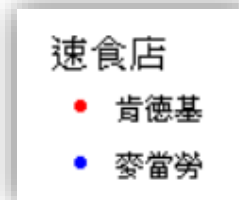
#### Continuous X, Continuous Y

`f <- ggplot(mpg, aes(cty, hwy))`

- `f + geom_blank()`
- `f + geom_jitter()`  
x, y, alpha, color, fill, shape, size
- `f + geom_point()`  
x, y, alpha, color, fill, shape, size
- `f + geom_quantile()`  
x, y, alpha, color, linetype, size, weight
- `f + geom_rug(sides = "bl")`  
alpha, color, linetype, size
- `f + geom_smooth(model = lm)`  
x, y, alpha, color, fill, linetype, size, weight
- `f + geom_text(aes(label = cty))`  
x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

# scale

```
scale_color_manual("速食店",  
  values=c("red", "blue"),  
  labels = c("肯德基", "麥當勞"))
```



※調整顏色、樣式、分類、圖例名稱……

## Scales

Scales map data values to the visual values of an aesthetic. To change a mapping, add a new scale.

`(n <- d + geom_bar(aes(fill = fl)))`

scale\_ aesthetic to adjust prepackaged scale to use scale specific arguments

`n + scale_fill_manual(values = c("skyblue", "royalblue", "blue", "navy"), limits = c("d", "e", "p", "r"), breaks = c("d", "e", "p", "r"), name = "fuel", labels = c("D", "E", "P", "R"))`

range of values to include in mapping title to use in legend/axis labels to use in legend/axis breaks to use in legend/axis

## Color and fill scales (Continuous)

`o <- c + geom_dotplot(aes(fill = ..x..))`

- `o + scale_fill_distiller(palette = "Blues")`
- `o + scale_fill_gradient(low="red", high="yellow")`
- `o + scale_fill_gradient2(low="red", high="blue", mid = "white", midpoint = 25)`
- `o + scale_fill_gradientn(colours=topo.colors(6))`  
Also: `rainbow()`, `heat.colors()`, `terrain.colors()`, `cm.colors()`, `RColorBrewer::brewer.pal()`

`s <- ggplot(mpg, aes(fl, fill = drv))`

- `s + geom_bar(position = "dodge")`  
Arrange elements side by side
- `s + geom_bar(position = "fill")`  
Stack elements on top of one another, normalize height
- `e + geom_point(position = "jitter")`  
Add random noise to X and Y position of each element to avoid overplotting
- `e + geom_label(position = "nudge")`  
Nudge labels away from points
- `s + geom_bar(position = "stack")`  
Stack elements on top of one another

## xtabs 函數 樞紐分析表

xdata=xtabs(~TOWN+STORE,data=data) → xdata=data.frame(xdata)

| TOWN | STORE |     |
|------|-------|-----|
|      | KFC   | MIC |
| 士林區  | 2     | 8   |
| 大同區  | 1     | 3   |
| 大安區  | 2     | 11  |
| 中山區  | 4     | 9   |
| 中正區  | 2     | 8   |

→資料型別是xtabs(table)  
格式也無法使用

|   | TOWN | STORE | Freq |
|---|------|-------|------|
| 1 | 士林區  | KFC   | 2    |
| 2 | 大同區  | KFC   | 1    |
| 3 | 大安區  | KFC   | 2    |
| 4 | 中山區  | KFC   | 4    |
| 5 | 士林區  | MIC   | 8    |
| 6 | 大同區  | MIC   | 3    |
| 7 | 大安區  | MIC   | 11   |
| 8 | 中山區  | MIC   | 9    |

→ggplot可以  
用的資料格式

## reshape2 套件

Q. 本來資料就是data.frame了？

|   | 科系 | 男生 | 女生 |
|---|----|----|----|
| 1 | 地理 | 13 | 14 |
| 2 | 大氣 | 21 | 6  |
| 3 | 地質 | 7  | 13 |

melt()

dcast()

|   | 科系 | 性別 | 人數 |
|---|----|----|----|
| 1 | 地理 | 男生 | 13 |
| 2 | 大氣 | 男生 | 21 |
| 3 | 地質 | 男生 | 7  |
| 4 | 地理 | 女生 | 14 |
| 5 | 大氣 | 女生 | 6  |
| 6 | 地質 | 女生 | 13 |

```
melt(data,  
      id.vars = "科系",  
      variable.name = "性別",  
      value.name = "人數")  
  
dcast(data,  
       formula = 科系~性別,  
       value.var = "人數")
```

## 用ggplot 畫多邊形

- 以TOWN這欄來合併畫圖

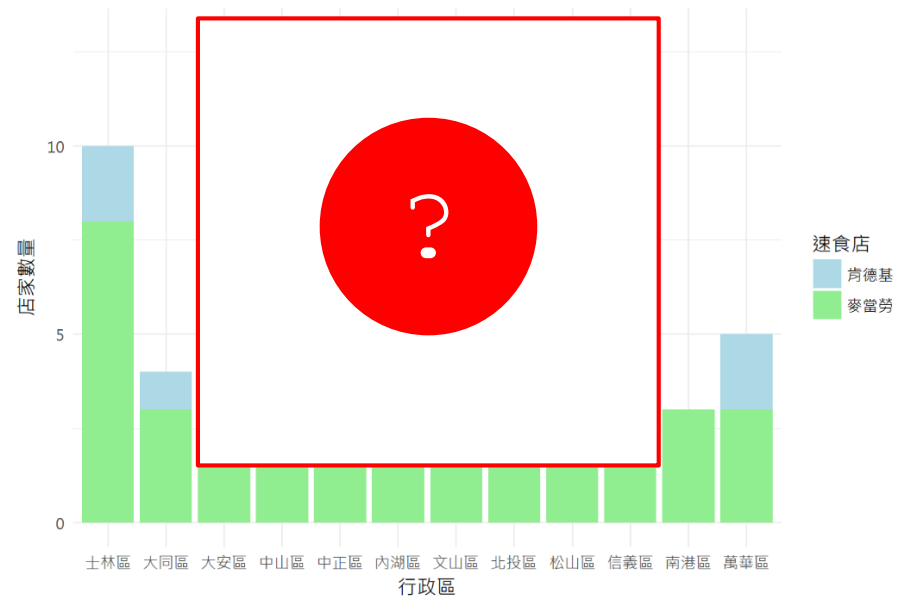
```
TOWN.f = fortify(Vill, region="TOWN")
```

```
TOWN.f = merge(TOWN.f, Vill@data,  
               by.x = "id", by.y = "TOWN") #只是把資料對回去
```

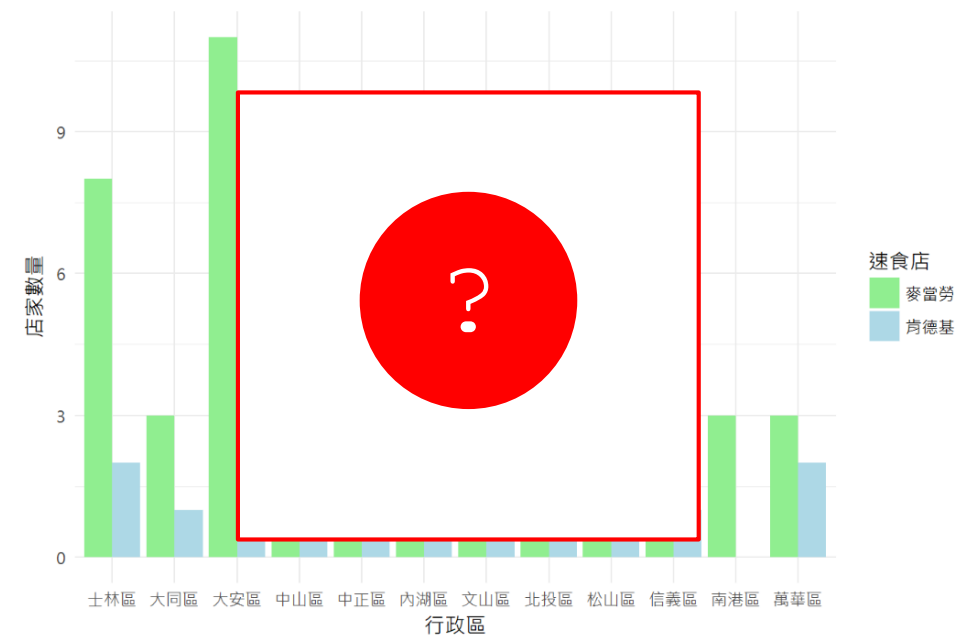
```
ggplot()+  
  geom_polygon(data = TOWN.f,  
              aes(x=long, y = lat, group = group),  
              fill="khaki1", color="black")+  
  coord_fixed(1.0)
```

# Lab 1 參考圖

台北市各行政區的麥當勞與肯德基家數



台北市各行政區的麥當勞與肯德基家數



台北市速食店99年銷售規模的空間分布

