



# R語言實作

## 台北市房價分析



徐雋堯

# 報告大綱



前情提要：



台北市房價探索



資料定義



資料檢索  
&  
清理



資料分析

研究  
三  
步  
驟



報告要點：

資料來源

➤ 網頁擷取

程式簡要

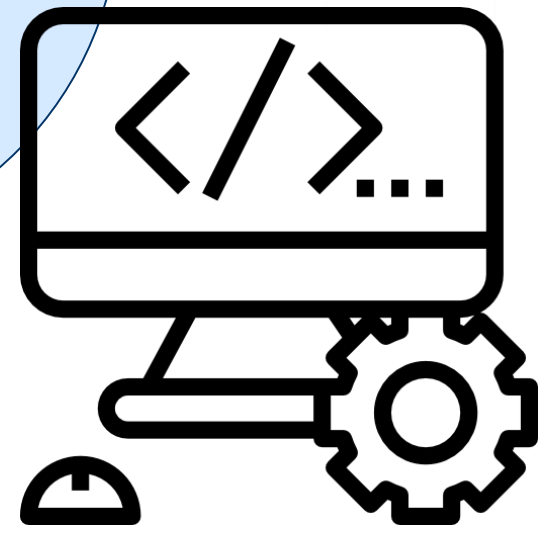
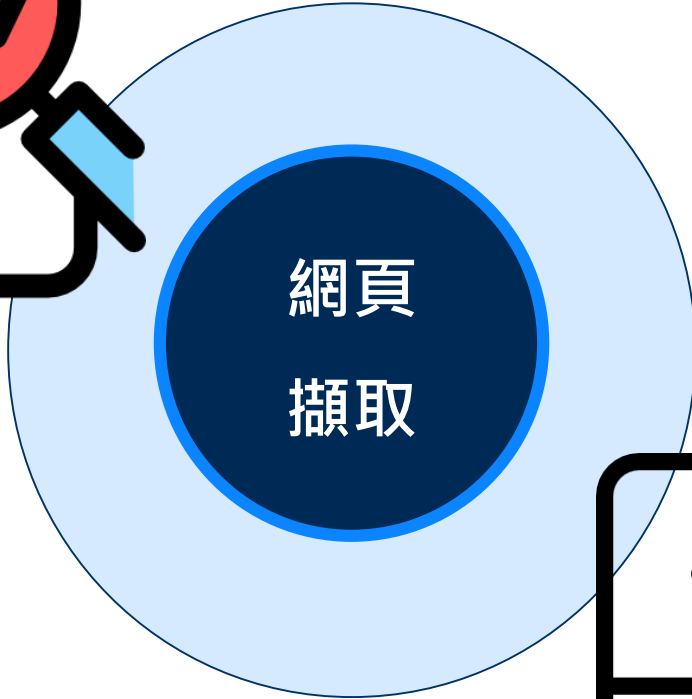
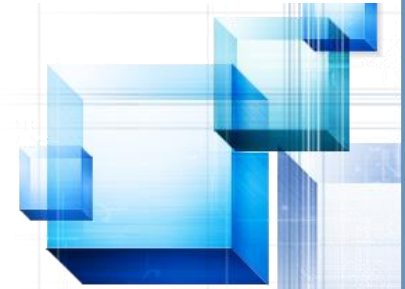
資料清理

➤ 方式運用

資料分析

➤ 敘述性統計

➤ 視覺化



# 資料來源-網頁說明



The screenshot shows the homepage of the 信義房屋 (Xinyi Real Estate) website. At the top is a dark navigation bar with the 信義房屋 logo and links for 買屋 (Buy), 賣屋 (Sell), 社區 (Community), 成交行情 (Transaction Market), 新聞百科 (News & Encyclopedia), 租屋 (Rent), 居家生活 (Home Life), 門市據點 (Branch Locations), 人才招聘 (Talent Recruitment), and 企業社會責任 (Corporate Social Responsibility). The main banner features a blurred background image of a modern building and text announcing a 2017 award: "榮獲 2017遠見雜誌「五星服務獎」房屋仲介類 首獎" (Awarded 2017 Far Vision Magazine 'Five Star Service Award' Real Estate Brokerage Category First Prize). Below this is the slogan "為了家，還可以更好。" (For home, it can be even better.) followed by a "看影片" (Watch Video) link with a play button icon. A search section below the banner includes tabs for 中古屋 (Old House), 新建案 (New Development), 海外置產 (Overseas Investment), and 居家服務 (Home Service). The 新建案 tab is selected. There is a search input field with the placeholder text "請輸入路街、社區、關鍵字或物件編號" (Please enter street, community, keyword, or property number), a "進階篩選" (Advanced Filter) dropdown, and a red "搜尋" (Search) button. At the bottom of the search section are three icons with labels: a location pin for "地圖搜尋" (Map Search), a train for "捷運搜尋" (MRT Search), and a house for "社區搜尋" (Community Search).



目前全台買、賣房資料量最大房仲平台

# 資料來源-擷取目標



共有 10,149 間房屋

預設排序



店長推薦



天母國小靜巷豪宅 信義房屋品牌、台灣松下營造

台北市士林區天母北路 電梯華廈、成屋

30.56建坪 20.98主+陽 1年 3/7樓

2房1廳2衛0室

近公園

近市場

近學校

店長推薦



正天母學區小豪宅 台灣松下營造、天母小豪宅

台北市士林區天母北路 電梯華廈、成屋 其他車位

18.14建坪 12.64主+陽 1年 2/7樓

1房1廳1衛1室

近公園

近市場

近學校

店長推薦



美日僑全新三房 信義房屋品牌打造、台灣松下營造

台北市士林區天母北路 電梯華廈、成屋 其他車位

45.39建坪 31.20主+陽 1年 6/7樓

3房2廳2衛0室

近公園

近市場

近學校

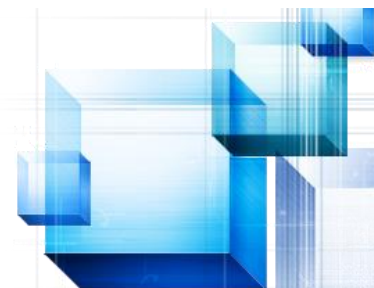
擷取目標

5,549萬

預約看屋



# 資料來源-程式提要



```
for (i in 1:3) {  
  for_htr <- paste("http://buy.sinyi.com.tw/list/Taipei-city/",  
    i,  
    ".html",  
    sep = "")
```

運用for迴圈，連續擷取網頁

```
  siyi <- read_html(for_htr) # 將網頁所有資料存入siyi  
  siyi_name <- html_nodes(siyi, ".item_title") %>% html_text()  
  siyi_local <- html_nodes(siyi, ".detail_line1") %>% html_text()  
  siyi_price <- html_nodes(siyi, ".price_new") %>% html_text()  
  siyi_detail <- html_nodes(siyi, ".detail_line1+ .detail_line2") %>% html_text()  
  siyi_room <- html_nodes(siyi, ".detail_line2+ .detail_line2") %>% html_text()
```

我們需要的資料ex：  
房型、價格等...

```
  siyi_df <- data.frame(  
    siyi_name = siyi_name,  
    siyi_local = siyi_local,  
    siyi_price = siyi_price,  
    siyi_detail = siyi_detail,  
    siyi_room = siyi_room
```

將我們所需的資料存成  
「資料框架」

```
)  
  siyi_house <- rbind(siyi_house, siyi_df) # 將資料一筆一筆過入觀察值  
  Sys.sleep(runif(1,2,4)) # 擷取程式停機時間，降低被認為機器人的可能性  
}
```

View(siyi\_house) # 檢視資料

# 資料來源-資料輸出(1/2)



55 石牌捷運黃金透入  
56 敦南國民景觀8樓  
57 松捷鋼骨三房車位  
58 國泰怡園公園美居  
59 統一大廈方正美屋  
60 兆雄小山田高樓戶

siyi\_local

1

台北市士林區天母北路\n  
電梯華廈、成屋\n

2

\n  
台北市士林區天母北路\n  
電梯華廈、成屋\n  
其他車位\n

3

\n  
台北市士林區天母北路\n  
電梯華廈、成屋\n  
其他車位\n

4



R語言輸出型式



# 資料來源-資料輸出(2/2)



A	B	C	D	E
siyi_name	siyi_local	siyi_price	siyi_detail	siyi_room
信義謙石起	台北市士林區天母北路	5,549萬	45.39建坪	3房
	電梯華廈、成屋、新成屋		31.2主+陽	2廳
			0.5年	2衛
	其他車位		6/7樓	0室



Excel 輸出型式

```
write.table(siyi_house, file = "/yyy/xxx.csv") # 輸出csv格式
```

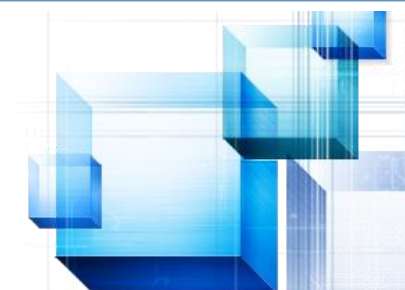
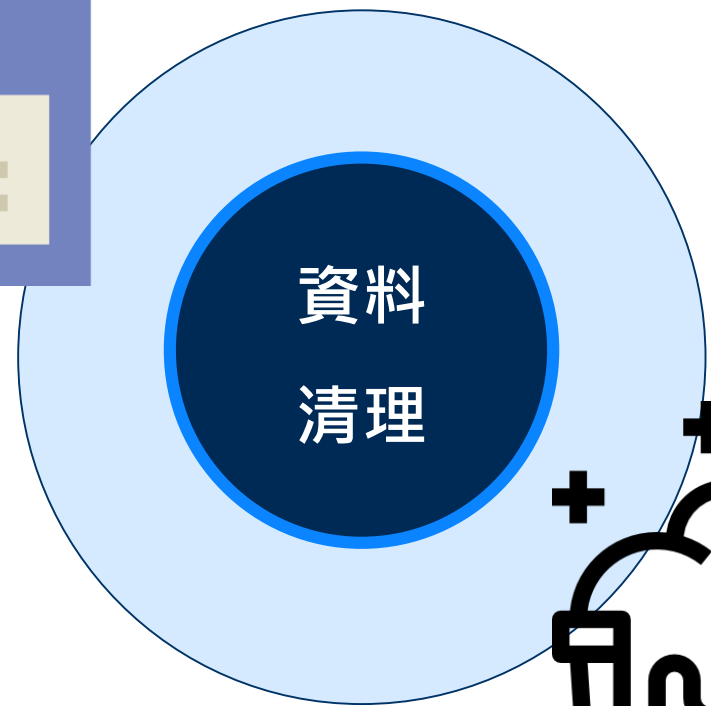
45.39建坪

31.2主+陽

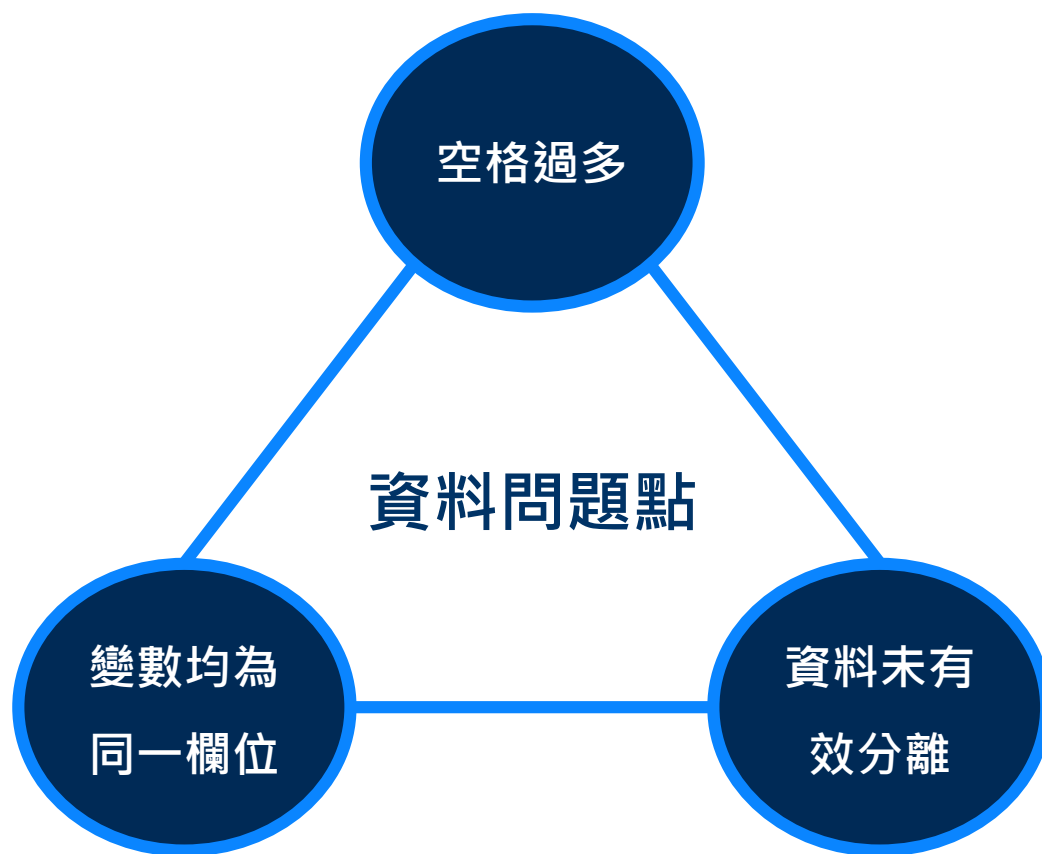
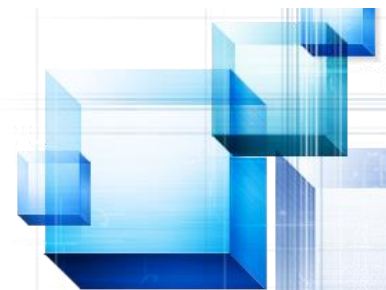
6/7樓

0.5年





# 資料來源-問題



# 資料清理-RawData(1/2)



資料『表面上』的平靜 ~~~~~

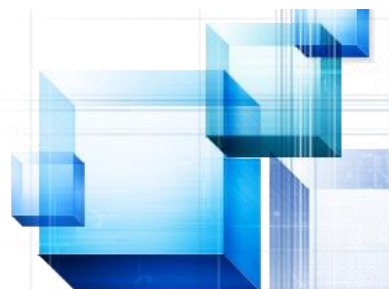
siyi_name	siyi_local	siyi_price	siyi_detail	siyi_room
1 信義謙石超優三房	台北市士林區天母北路 電梯華廈、成屋、新成屋 其他車位	5,549萬	45.39建坪 31.2主+陽 0.5年 6/7樓	3房 2廳 2衛 0室
2 信義謙石小豪宅	台北市士林區天母北路 電梯華廈、成屋、新成屋 其他車位	2,047萬	18.14建坪 12.64主+陽 0.5年 2/7樓	1房 1廳 1衛 1室
3 信義謙石2房豪宅	台北市士林區天母北路 電梯華廈、成屋、新成屋 其他車位	3,558萬	30.56建坪 20.98主+陽 0.5年 3/7樓	2房 1廳 2衛 0室
4 石牌精美電梯美屋	台北市北投區石牌路一段 電梯華廈、成屋	1,580萬	22.71建坪 13.46主+陽 4.4年 2/7樓 凱安植璜	1房 2廳 1衛 0室 (加蓋格局: 2房 0廳 1衛 0室)
5 捷運高樓MVP	台北市內湖區金湖路 電梯大樓、成屋 坡平車位	4,050萬	57.08建坪 31.66主+陽 12.3年 14/14樓 大漢MVP	3房 2廳 2衛 0室
6 大安兩房崗石建築	台北市大安區光復南路 電梯大樓、成屋、新成屋	2,840萬	22.65建坪 15.4主+陽 6/14樓	2房 2廳 1衛 0室
7 南京捷運大器三房	台北市松山區南京東路五段 公寓、成屋	2,588萬	39.48建坪 39.48主建物 50.4年 4/4樓	3房 2廳 2衛 0室
8 大安精緻全新兩房	台北市大安區光復南路 電梯大樓、成屋、新成屋	2,888萬	23.23建坪 15.81主+陽 5/14樓	2房 1廳 1衛 0室
9 紀念堂4房大露台	台北市中正區寧波西街 電梯大樓、成屋	6,988萬	80.83建坪 44.51主+陽 20.5年 7/12樓 中正寧波	4房 2廳 2衛 1室
10 A I T一樓美寓	台北市內湖區康寧路一段 多樓層、成屋	2,750萬	64.58建坪 24.99主建物 37年 B1-1/5樓	2房 2廳 1衛 2室 (加蓋格局: 1房 0廳 2衛 0室)
11 大戶移居河濱生活	台北市文山區樟新街 電梯華廈、成屋	4,120萬	71.09建坪 48.32主+陽 3.3年 4/9樓	3房 2廳 2衛 0室
12 面公園三房十車位	台北市文山區木柵路一段 電梯華廈、成屋 其他車位	2,478萬	49.65建坪 33.41主+陽 2年 2/5樓	3房 2廳 2衛 0室
13 中山世紀美學家	台北市中山區中山北路二段 電梯大樓、成屋 坡平車位	17,350萬	157.68建坪 88.34主+陽 1.7年 9/14樓	3房 2廳 5衛 0室
14 南京雙捷公園豪邸	台北市中山區吉林路 電梯華廈、成屋 其他車位	6,500萬	58.74建坪 34.13主+陽 4.6年 7/7樓 易居	3房 2廳 2衛 0室
15 湯自慢溫泉小品	台北市北投區奇岩路 電梯華廈、成屋	678萬	13.65建坪 8.25主+陽 5.1年 3/5樓	0房 1廳 2衛 0室
16 天璽正大兩房新屋	台北市中山區農安街 電梯大樓、成屋 坡平車位	3,900萬	65.66建坪 34.21主+陽 1.3年 3/14樓	2房 1廳 2衛 0室
17 空間大滿足歐夏蕾	台北市中山區林森北路 電梯大樓、成屋	1,188萬	12.34建坪 7.89主建物 12.3年 5/14樓	1房 1廳 1衛 0室 (加蓋格局: 1房 0廳 0衛 0室)

Showing 1 to 18 of 8,250 entries

# 資料清理-RawData(2/2)



資料 ~~ 實際 ~~ 『真面目!!』



60 兆雄小山田高樓戶

1 siyi\_local

1

台北市士林區天母北路\n  
電梯華廈、成屋\n

2

\n

台北市士林區天母北路\n  
電梯華廈、成屋\n

\n

其他車位

\n

3

\n

台北市士林區天母北路\n  
電梯華廈、成屋\n

\n

其他車位

\n

4

R

~

R

~



# 資料清理-工具說明(1/2)



## { R 字串套件 }

- 開發者：Hadley Wickham
- 字串的查詢、取代、替換...等。相關處理功能 !!
- 可結合「正規表示法」運用



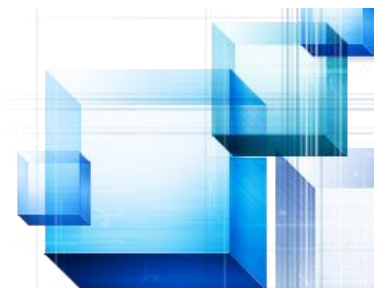
/[\\w.\_%+-]+@[\\w.-]+\\. [a-zA-Z]{2,4}/

Match a single period      Match at least two

## { 正規表示法 }

- 用來描述、符合一系列某種句法規則的字串
- 通常被用來檢索、替換符合某種模式的文字

# 資料清理-語法提示



`str_trim()` : 移除空白

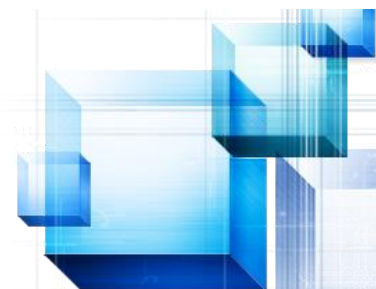
`str_sub(string, start = ,end = )` : 根據位置取值

`str_extract(string, pattern )` : 可用正規表示法取值

`str_replace(string, pattern, replacement )` : 代換字串



# 資料清理-正規表示法



\d : 數字

\D : 非數字

\w : 數字、字母、底線

\W : 非 \w

[0-9]+ : 整數

[0-9]+\.[0-9]+ : 小數點

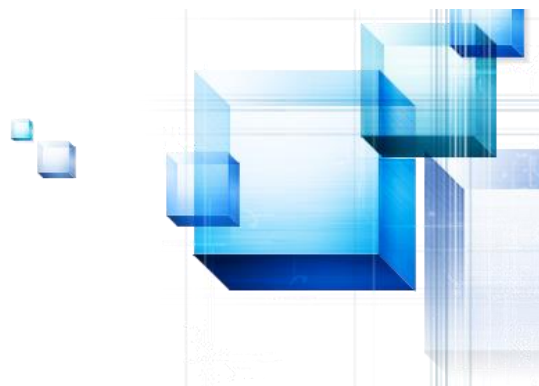
^ : 字串開始位置

\ : 跳脫字元





# 資料清理-程式碼簡要



# ---- 行政區切割 ----

`siyi$siyi_local` # 資料檢視後，發現資料空白處很多，會影響清理效率

`siyi$siyi_local <- str_trim(siyi$siyi_local)` # 將資料空白字刪除

`siyi$行政區 <- str_sub(siyi$siyi_local, start = 4, end = 6)`

# 將空白處刪除後，可以發現行政區資訊，是在變數中的第4~6位置，

# 運用此函數將資料取出，另建立新變數

# ----- 路段切割 -----

# `str_extract` 此函數其中的引數`pattern`可以正則表示法

`siyi$路段 <- str_extract(siyi$siyi_local,`  
`pattern = "^台北市\\w{0,30}")`

# 將`siyi_local`運用正則表示法處理

# `^`：符號代表起始值

# `\\w`：意思代表字母與數字

# `pattern = "^台北市\\w{0,30}"`

# 上述意思為：以台北市開頭的字串，取0~30位置。

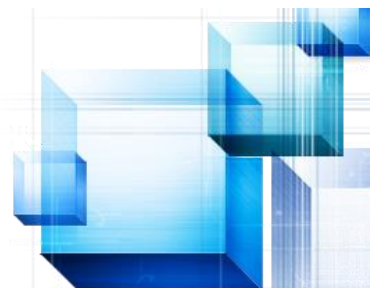
# 故可包括所有路段長知不一樣的資訊

`siyi$路段 <- str_sub(siyi$路段,7,)`

# 只需行政區資訊，因此再用`str_sub`鎖定第7以後的位置

siyi_name	siyi_local	行政區	路段
信義謙石超優三房	台北市士林區天母北路 電梯華廈、成屋、新成屋 其他車位	士林區	天母北路
信義謙石小豪宅	台北市士林區天母北路 電梯華廈、成屋、新成屋 其他車位	士林區	天母北路
信義謙石2房豪宅	台北市士林區天母北路 電梯華廈、成屋、新成屋 其他車位	士林區	天母北路
石牌精美電梯美屋	台北市北投區石牌路一段 電梯華廈、成屋	北投區	石牌路一段
捷運高樓MVP	台北市內湖區金湖路 電梯大樓、成屋 坡平車位	內湖區	金湖路
大安兩房崗石建築	台北市大安區光復南路 電梯大樓、成屋、新成屋	大安區	光復南路
南京捷運大器三房	台北市松山區南京東路五段 公寓、成屋	松山區	南京東路五段
大安精緻全新兩房	台北市大安區光復南路 電梯大樓、成屋、新成屋	大安區	光復南路
紀念堂4房大露台	台北市中正區寧波西街 電梯大樓、成屋	中正區	寧波西街
A I T一樓美寓	台北市內湖區康寧路一段 多樓層、成屋	內湖區	康寧路一段
大戶移居河濱生活	台北市文山區樟新街 電梯華廈、成屋	文山區	樟新街
面公園三房十車位	台北市文山區木柵路一段 電梯華廈、成屋 其他車位	文山區	木柵路一段
中山世紀美學家	台北市中山區中山北路二段 電梯大樓、成屋 坡平車位	中山區	中山北路二段

# 資料清理-程式碼簡要



```
## ---- 房型 ---- ##
```

```
siyi$siyi_local <- str_trim(siyi$siyi_local)
```

```
siyi$siyi_local # 資料檢視
```

```
siyi$房型 <- str_extract(siyi$siyi_local, pattern = "\\w{0,}、")
```

```
# 從資料檢視中發現，我們所需要資料的字串接有"、"符號，因此在
```

```
# 表示法上可以用"、"前的所有字串進行查找
```

```
siyi$房型 <- str_replace(siyi$房型, "、", "")
```

```
# 將所要的資料找出之後，再將"、"用空白取代
```

	siyi_name	siyi_local	房型
1	信義謙石超優三房	台北市士林區天母北路 電梯華廈、成屋、新成屋 其他車位	電梯華廈
2	信義謙石小豪宅	台北市士林區天母北路 電梯華廈、成屋、新成屋 其他車位	電梯華廈
3	信義謙石2房豪宅	台北市士林區天母北路 電梯華廈、成屋、新成屋 其他車位	電梯華廈
4	石牌精美電梯美屋	台北市北投區石牌路一段 電梯華廈、成屋	電梯華廈
5	捷運高樓MVP	台北市內湖區金湖路 電梯大樓、成屋 坡平車位	電梯大樓
7	南京捷運大器三房	台北市松山區南京東路五段 公寓、成屋	公寓
9	紀念堂4房大露台	台北市中正區寧波西街 電梯大樓、成屋	電梯大樓
0	A I T一樓美寓	台北市內湖區康寧路一段 多樓層、成屋	多樓層
1	大戶移居河濱生活	台北市文山區樟新街 電梯華廈、成屋	電梯華廈
2	面公園三房十車位	台北市文山區木柵路一段 電梯華廈、成屋 其他車位	電梯華廈
3	中山世紀美學家	台北市中山區中山北路二段 電梯大樓、成屋 坡平車位	電梯大樓
4	南京雙捷公園豪邸	台北市中山區吉林路 電梯華廈、成屋 其他車位	電梯華廈
5	湯自慢溫泉小品	台北市北投區奇岩路 電梯華廈、成屋	電梯華廈
6	天璽正大兩房新屋	台北市中山區農安街 電梯大樓、成屋 坡平車位	電梯大樓
7	空間大滿足歐夏蕾	台北市中山區林森北路 電梯大樓、成屋	電梯大樓
8	高樓景觀小資臻品	台北市中山區林森北路 電梯大樓、成屋	電梯大樓

# 資料清理-程式碼簡要



## ----- 房、廳、衛資料篩選 -----

# str\_sub 只能用位置選出變數中的值

# str\_extract 可用正規表示法，選出變數中的值

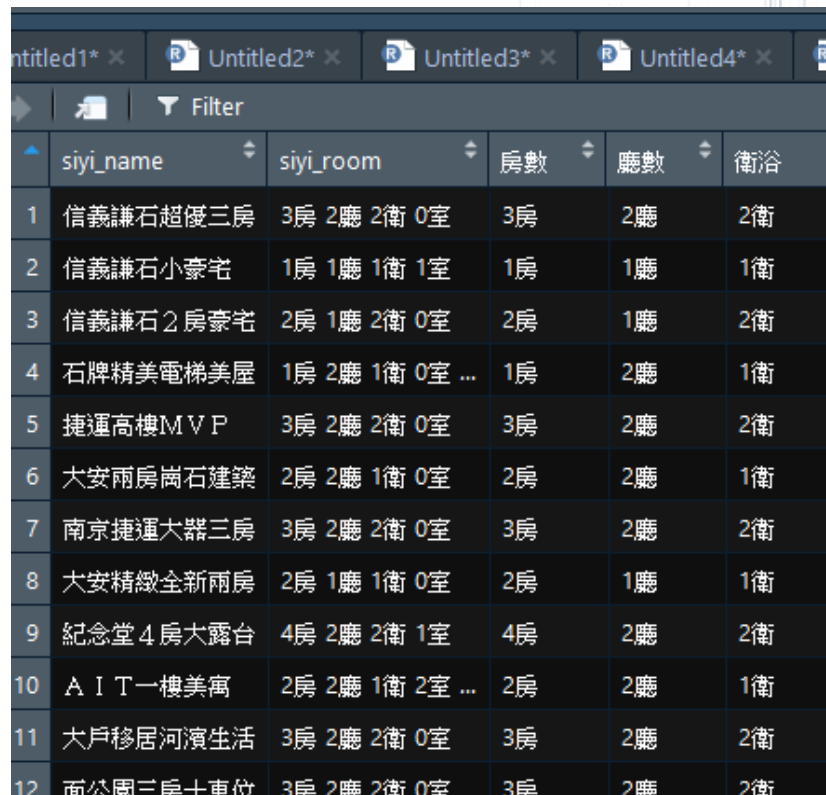
siyi\$房數 <- str\_extract(siyi\$siyi\_room, pattern = "[0-9]房") # 篩出房數

siyi\$廳數 <- str\_extract(siyi\$siyi\_room, pattern = "[0-9]廳") # 篩出廳數

siyi\$衛浴 <- str\_extract(siyi\$siyi\_room, pattern = "[0-9]衛") # 篩出衛浴

# 相關數值，都是在數字之後接上：房、廳、衛

# 因此在正規表示法上，直接用[0-9]接所要資訊即可



	siyi_name	siyi_room	房數	廳數	衛浴
1	信義謙石超優三房	3房 2廳 2衛 0室	3房	2廳	2衛
2	信義謙石小豪宅	1房 1廳 1衛 1室	1房	1廳	1衛
3	信義謙石 2 房豪宅	2房 1廳 2衛 0室	2房	1廳	2衛
4	石牌精美電梯美屋	1房 2廳 1衛 0室 ...	1房	2廳	1衛
5	捷運高樓MVP	3房 2廳 2衛 0室	3房	2廳	2衛
6	大安兩房崗石建築	2房 2廳 1衛 0室	2房	2廳	1衛
7	南京捷運大器三房	3房 2廳 2衛 0室	3房	2廳	2衛
8	大安精緻全新兩房	2房 1廳 1衛 0室	2房	1廳	1衛
9	紀念堂 4 房大露台	4房 2廳 2衛 1室	4房	2廳	2衛
10	A I T 一樓美寓	2房 2廳 1衛 2室 ...	2房	2廳	1衛
11	大戶移居河濱生活	3房 2廳 2衛 0室	3房	2廳	2衛
12	面公園三房+車位	3房 2廳 2衛 0室	3房	2廳	2衛

# 資料清理-程式碼簡要



```
siyi$siyi_detail <- str_trim(siyi$siyi_detail) # 空白處刪除
```

```
siyi$建坪 <- str_extract(siyi$siyi_detail, pattern = "[0-9]+建坪|[0-9]+.[0-9]+建坪")
```

```
# 因為「坪數」前的數字資料有小數點的資料和未有小數點的資料，
```

```
# 在表示法上，須留意用"|" (或)，將聯集二種可能性的組合，進行篩選
```

```
siyi$建坪 <- str_replace(siyi$建坪, "建坪", "") # 將"建坪"以"空白"取代
```

```
class(siyi$建坪) # 檢視建坪變數型態
```

```
siyi$建坪 <- as.numeric(siyi$建坪) # 將建坪資料變為「數值型資料」
```

```
siyi$屋齡 <- str_extract(siyi$siyi_detail, pattern = "[0-9]+年|[0-9]+.[0-9]+年")
```

```
# 比照坪數方式處理
```

```
siyi$屋齡 <- str_replace(siyi$屋齡, "年", "")
```

```
siyi$屋齡 <- as.numeric(siyi$屋齡)
```

siyi_name	siyi_detail	建坪	屋齡
信義謙石超優三房	45.39建坪 31.2主+陽 0.5年 6/7樓	45.39	0.5
信義謙石小豪宅	18.14建坪 12.64主+陽 0.5年 2/7樓	18.14	0.5
信義謙石2房豪宅	30.56建坪 20.98主+陽 0.5年 3/7樓	30.56	0.5
石牌精美電梯美屋	22.71建坪 13.46主+陽 4.4年 2/7樓 凱安植璞	22.71	4.4
捷運高樓MVP	57.08建坪 31.66主+陽 12.3年 14/14樓 大漢MVP	57.08	12.3
南京捷運大器三房	39.48建坪 39.48主建物 50.4年 4/4樓	39.48	50.4
紀念堂4房大露台	80.83建坪 44.51主+陽 20.5年 7/12樓 中正寧波	80.83	20.5
A I T一樓美寓	64.58建坪 24.99主建物 37年 B1-1/5樓	64.58	37.0
大戶移居河濱生活	71.09建坪 48.32主+陽 3.3年 4/9樓	71.09	3.3
面公園三房十車位	49.65建坪 33.41主+陽 2年 2/5樓	49.65	2.0
中山世紀美學家	157.68建坪 88.34主+陽 1.7年 9/14樓	157.68	1.7
南京雙捷公園豪邸	58.74建坪 34.13主+陽 4.6年 7/7樓 易居	58.74	4.6
湯自慢溫泉小品	13.65建坪 8.25主+陽 5.1年 3/5樓	13.65	5.1
天璽正大兩房新屋	65.66建坪 34.21主+陽 1.3年 3/14樓	65.66	1.3
空間大滿足歐夏蕾	12.34建坪 7.89主建物 12.3年 5/14樓	12.34	12.3

# 資料清理-程式碼簡要



```
siyi$總樓層 <- str_extract(siyi$siyi_detail, pattern = "\\d+樓")
```

```
# 樓層資料必需區分為「總樓層」以及「出售樓層」
```

```
# 總樓層資料因數字後接著「樓」的文字，
```

```
# 表示法上用：\\d(所有數字) + 樓，篩選出資訊即可
```

```
siyi$總樓層 <- str_replace(siyi$總樓層, "樓", "")
```

```
siyi$總樓層 <- as.numeric(siyi$總樓層)
```

```
# 須轉換成數值資料，以利後續分析計算
```

```
siyi$出售樓層 <- str_extract(siyi$siyi_detail,  
                             pattern = "\\d+/")
```

```
# /(斜線之前)的數值
```

```
siyi$出售樓層 <- str_replace(siyi$出售樓層, "/", "")
```

```
siyi$出售樓層 <- as.numeric(siyi$出售樓層)
```

```
siyi$總價_萬 <- str_replace(siyi$siyi_price, "萬", "")
```

```
siyi$總價_萬 <- str_replace(siyi$總價, ",", "")
```

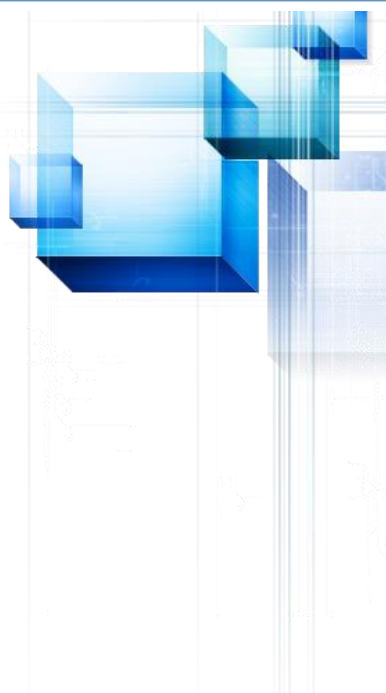
```
siyi$總價_萬 <- as.numeric(siyi$總價_萬)
```

```
siyi_clean <- na.omit(siyi)
```

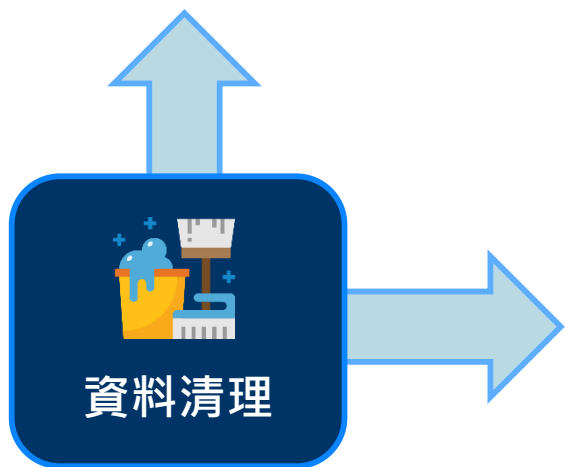
	siyi_name	siyi_detail	總樓層	出售樓層	總價_萬
1	信義謙石超優三房	45.39建坪 31.2主+陽 0.5年 6/7樓	7	6	5549
2	信義謙石小豪宅	18.14建坪 12.64主+陽 0.5年 2/7樓	7	2	2047
3	信義謙石2房豪宅	30.56建坪 20.98主+陽 0.5年 3/7樓	7	3	3558
4	石牌精美電梯美屋	22.71建坪 13.46主+陽 4.4年 2/7樓 凱安植璜	7	2	1580
5	捷運高樓MVP	57.08建坪 31.66主+陽 12.3年 14/14樓 大漢MVP	14	14	4050
7	南京捷運大器三房	39.48建坪 39.48主建物 50.4年 4/4樓	4	4	2588
9	紀念堂4房大露台	80.83建坪 44.51主+陽 20.5年 7/12樓 中正寧波	12	7	6988
10	A I T一樓美寓	64.58建坪 24.99主建物 37年 B1-1/5樓	5	1	2750
11	大戶移居河濱生活	71.09建坪 48.32主+陽 3.3年 4/9樓	9	4	4120
12	面公園三房十車位	49.65建坪 33.41主+陽 2年 2/5樓	5	2	2478
13	中山世紀美學家	157.68建坪 88.34主+陽 1.7年 9/14樓	14	9	17350
14	南京雙捷公園豪邸	58.74建坪 34.13主+陽 4.6年 7/7樓 易居	7	7	6500
15	湯自慢溫泉小品	13.65建坪 8.25主+陽 5.1年 3/5樓	5	3	678
16	天璽正大兩房新屋	65.66建坪 34.21主+陽 1.3年 3/14樓	14	3	3900
17	空間大滿足歐夏蕾	12.34建坪 7.89主建物 12.3年 5/14樓	14	5	1188
18	高樓景觀小資臻品	9.22建坪 6.41主建物 35.5年 11/12樓	12	11	520



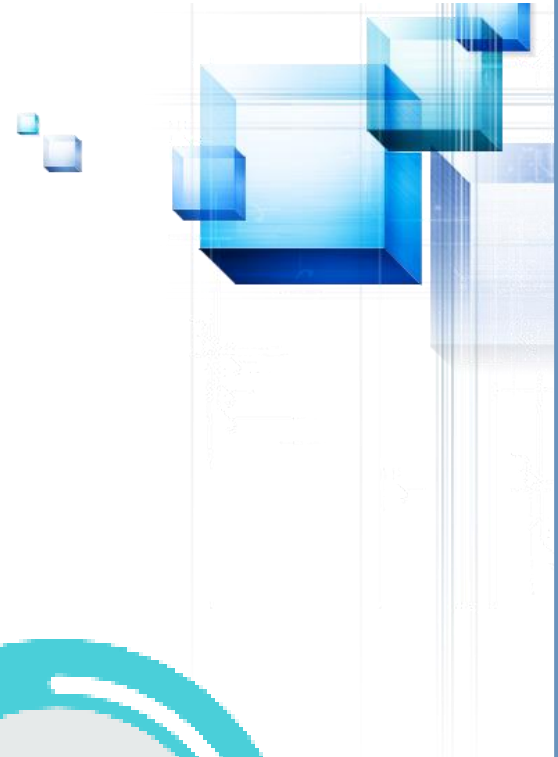
# 資料清理-處理後資料



	行政區	路段	房型	房數	廳數	衛浴	建坪	屋齡	總樓層
1	士林區	天母北路	電梯華廈	3房	2廳	2衛	45.39	0.5	7
2	士林區	天母北路	電梯華廈	1房	1廳	1衛	18.14	0.5	7
3	士林區	天母北路	電梯華廈	2房	1廳	2衛	30.56	0.5	7
4	北投區	石牌路一段	電梯華廈	1房	2廳	1衛	22.71	4.4	7
5	內湖區	金湖路	電梯大樓	3房	2廳	2衛	57.08	12.3	14
7	松山區	南京東路五段	公寓	3房	2廳	2衛	39.48	50.4	4
9	中正區	寧波西街	電梯大樓	4房	2廳	2衛	80.83	20.5	12
10	內湖區	康寧路一段	多樓層	2房	2廳	1衛	64.58	37.0	5
11	文山區	樟新街	電梯華廈	3房	2廳	2衛	71.09	3.3	9
12	文山區	木柵路一段	電梯華廈	3房	2廳	2衛	49.65	2.0	5
13	中山區	中山北路二段	電梯大樓	3房	2廳	5衛	157.68	1.7	14
14	中山區	吉林路	電梯華廈	3房	2廳	2衛	58.74	4.6	7
15	北投區	奇岩路	電梯華廈	0房	1廳	2衛	13.65	5.1	5
16	中山區	農安街	電梯大樓	2房	1廳	2衛	65.66	1.3	14



	siyi_name	行政區	路段	房型	房數	廳數	衛浴	建坪	屋齡	總樓層	出售樓層	總價_萬
1	信義謙石超優三房	士林區	天母北路	電梯華廈	3房	2廳	2衛	45.39	0.5	7	6	5549
2	信義謙石小豪宅	士林區	天母北路	電梯華廈	1房	1廳	1衛	18.14	0.5	7	2	2047
3	信義謙石2房豪宅	士林區	天母北路	電梯華廈	2房	1廳	2衛	30.56	0.5	7	3	3558
4	石牌精美電梯美屋	北投區	石牌路一段	電梯華廈	1房	2廳	1衛	22.71	4.4	7	2	1580
5	捷運高樓MVP	內湖區	金湖路	電梯大樓	3房	2廳	2衛	57.08	12.3	14	14	4050
7	南京捷運大器三房	松山區	南京東路五段	公寓	3房	2廳	2衛	39.48	50.4	4	4	2588
9	紀念堂4房大露台	中正區	寧波西街	電梯大樓	4房	2廳	2衛	80.83	20.5	12	7	6988
10	A I T一樓美寓	內湖區	康寧路一段	多樓層	2房	2廳	1衛	64.58	37.0	5	1	2750
11	大戶移居河濱生活	文山區	樟新街	電梯華廈	3房	2廳	2衛	71.09	3.3	9	4	4120
12	面公園三房十車位	文山區	木柵路一段	電梯華廈	3房	2廳	2衛	49.65	2.0	5	2	2478
13	中山世紀美學家	中山區	中山北路二段	電梯大樓	3房	2廳	5衛	157.68	1.7	14	9	17350
14	南京雙捷公園豪邸	中山區	吉林路	電梯華廈	3房	2廳	2衛	58.74	4.6	7	7	6500
15	湯自慢溫泉小品	北投區	奇岩路	電梯華廈	0房	1廳	2衛	13.65	5.1	5	3	678
16	天璽正大兩房新屋	中山區	農安街	電梯大樓	2房	1廳	2衛	65.66	1.3	14	3	3900
17	空間大滿足歐夏蕾	中山區	林森北路	電梯大樓	1房	1廳	1衛	12.34	12.3	14	5	1188



資料  
分析





# 資料分析-敘述性統計



```
> ## EDA 資料探索 ----
```

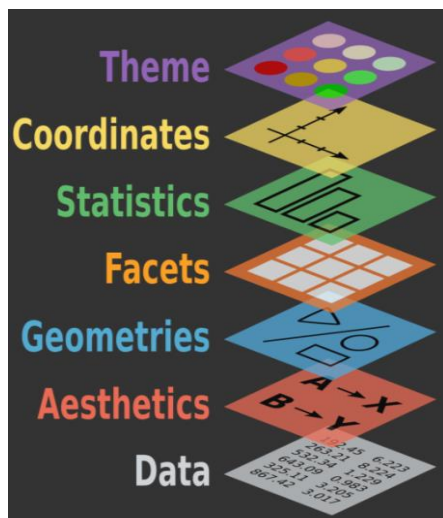
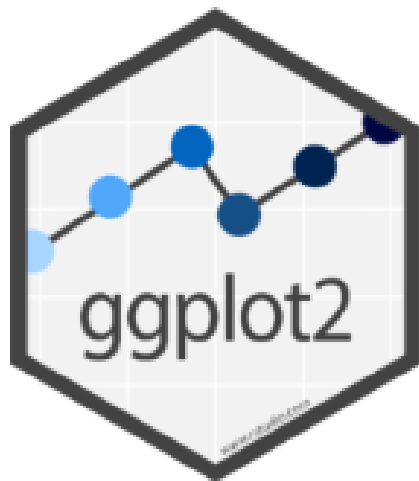
```
> summary(siyi[,c(8,9,12)])
```

建坪		屋齡		單價_萬	
Min.	: 3.38	Min.	: 0.4	Min.	: 21.34
1st Qu.	: 25.32	1st Qu.	: 8.4	1st Qu.	: 58.26
Median	: 35.90	Median	: 24.0	Median	: 72.15
Mean	: 45.07	Mean	: 23.7	Mean	: 80.27
3rd Qu.	: 55.16	3rd Qu.	: 36.6	3rd Qu.	: 91.95
Max.	: 712.64	Max.	: 106.0	Max.	: 732.75

```
> summary(siyi[,c(4,15,16)])
```

房型		說明		備註	
公寓	: 1850	1樓特	: 760	1F	: 760
其它1	: 205	一般	: 6904	中樓層	: 3456
其它2	: 124	店面	: 57	低樓層	: 2188
透天厝	: 305	特殊	: 88	高樓層	: 1405
電梯大樓	: 3078				
電梯華廈	: 2247				

# 資料分析-繪圖文法



```
ggplot(data, aes(x = 房型, y = 價格))
```

底圖



```
geom_boxplot()
```

盒鬚圖

OR 

```
geom_density()
```

機率密度圖

OR 

```
geom_point()
```

散佈圖

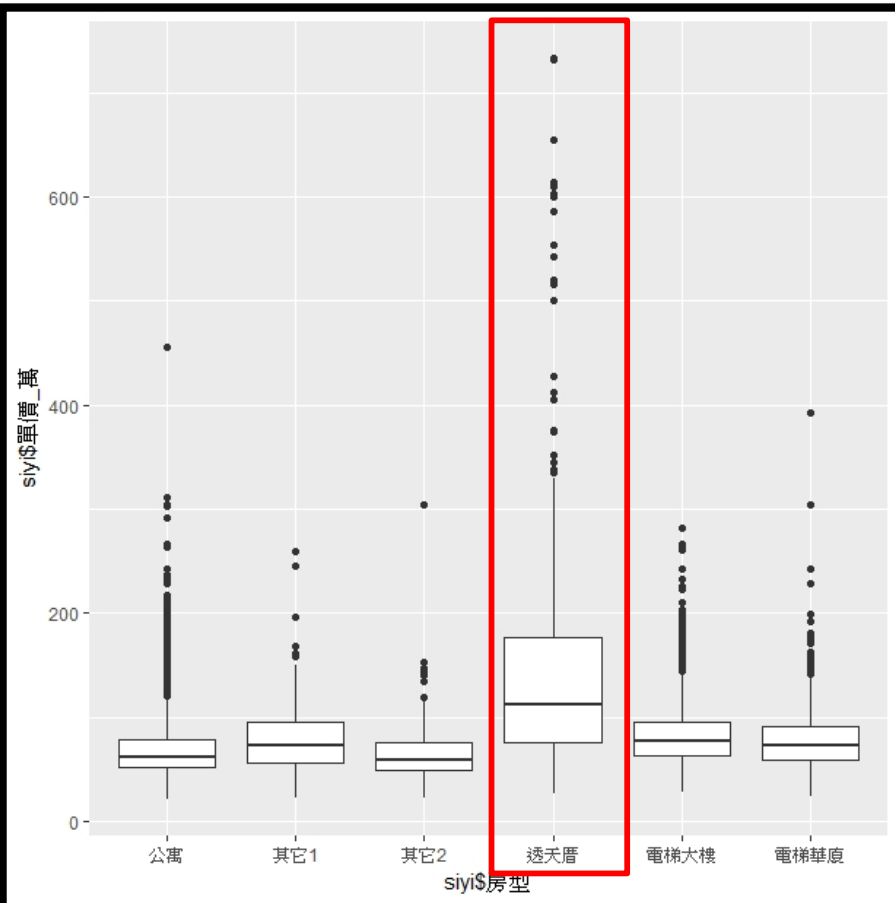
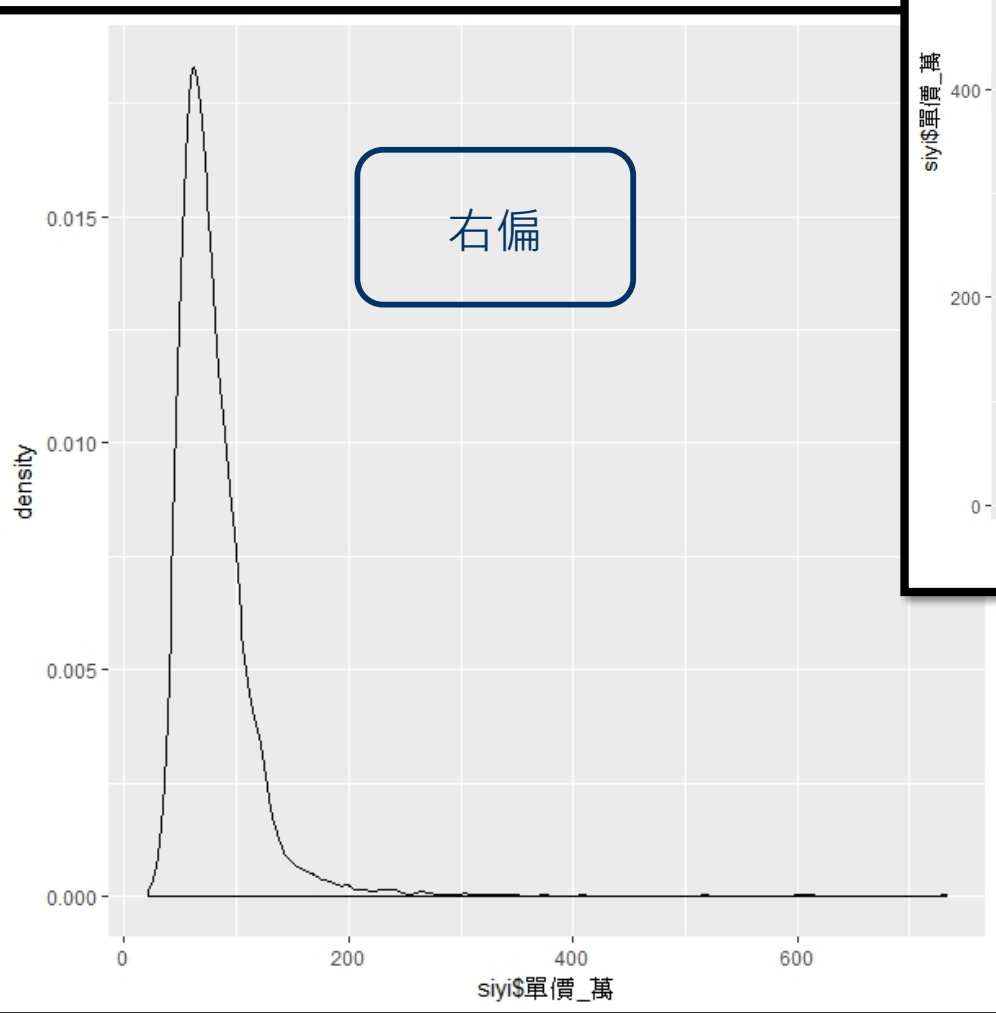


```
facet_grid()
```

分面

# 資料分析-全部資料

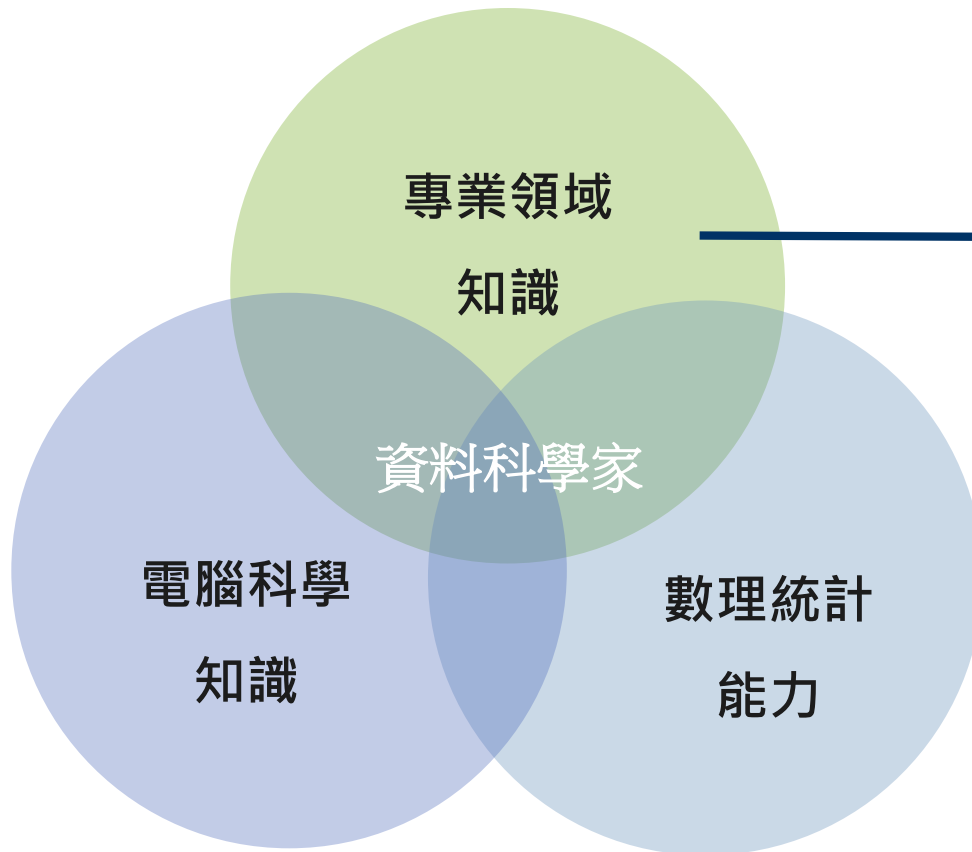
```
ggplot(siyi, aes(x = siyi$單價_萬)) + geom_density()
```



```
ggplot(siyi, aes(x = siyi$房型, y = siyi$單價_萬)) +  
  geom_boxplot()
```

➤ 干擾值太多

# 資料分析



特殊物件

- 一樓
- 店面
- 透天厝

# 資料分析-篩選後



```
> summary(siyi2[,c(8,9,12)])
```

建坪	屋齡	單價_萬
Min. : 5.14	Min. : 0.40	Min. : 21.34
1st Qu.: 24.96	1st Qu.: 7.90	1st Qu.: 57.64
Median : 35.58	Median : 21.90	Median : 70.14
Mean : 43.81	Mean : 22.17	Mean : 74.25
3rd Qu.: 53.48	3rd Qu.: 35.80	3rd Qu.: 88.09
Max. : 712.64	Max. : 59.90	Max. : 157.79

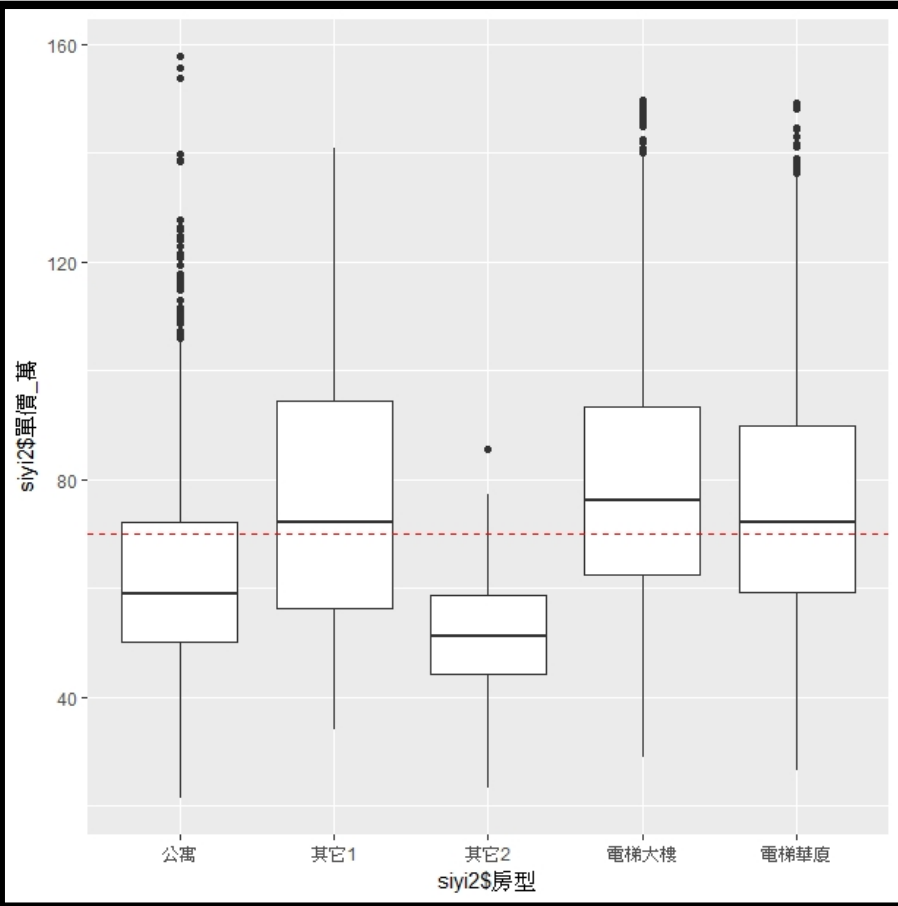
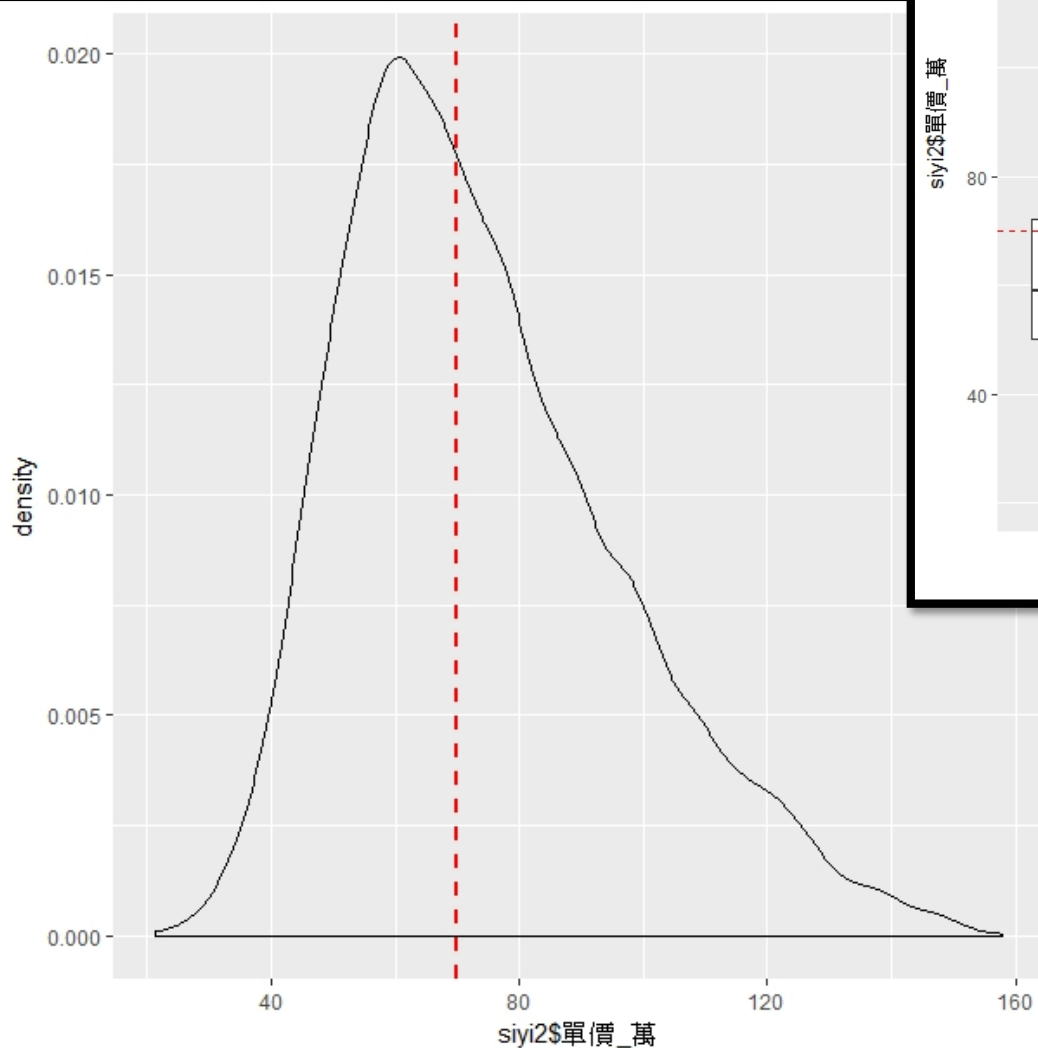
```
> |
```

```
> summary(siyi2[,c(4,15,16)])
```

房型	說明	備註
公寓 : 1520	1樓特: 0	1F : 0
其它1 : 139	一般 : 6654	中樓層: 3343
其它2 : 48	店面 : 0	低樓層: 1947
透天厝 : 0	特殊 : 0	高樓層: 1364
電梯大樓: 2932		
電梯華廈: 2015		

```
> |
```

# 資料分析-一般居住



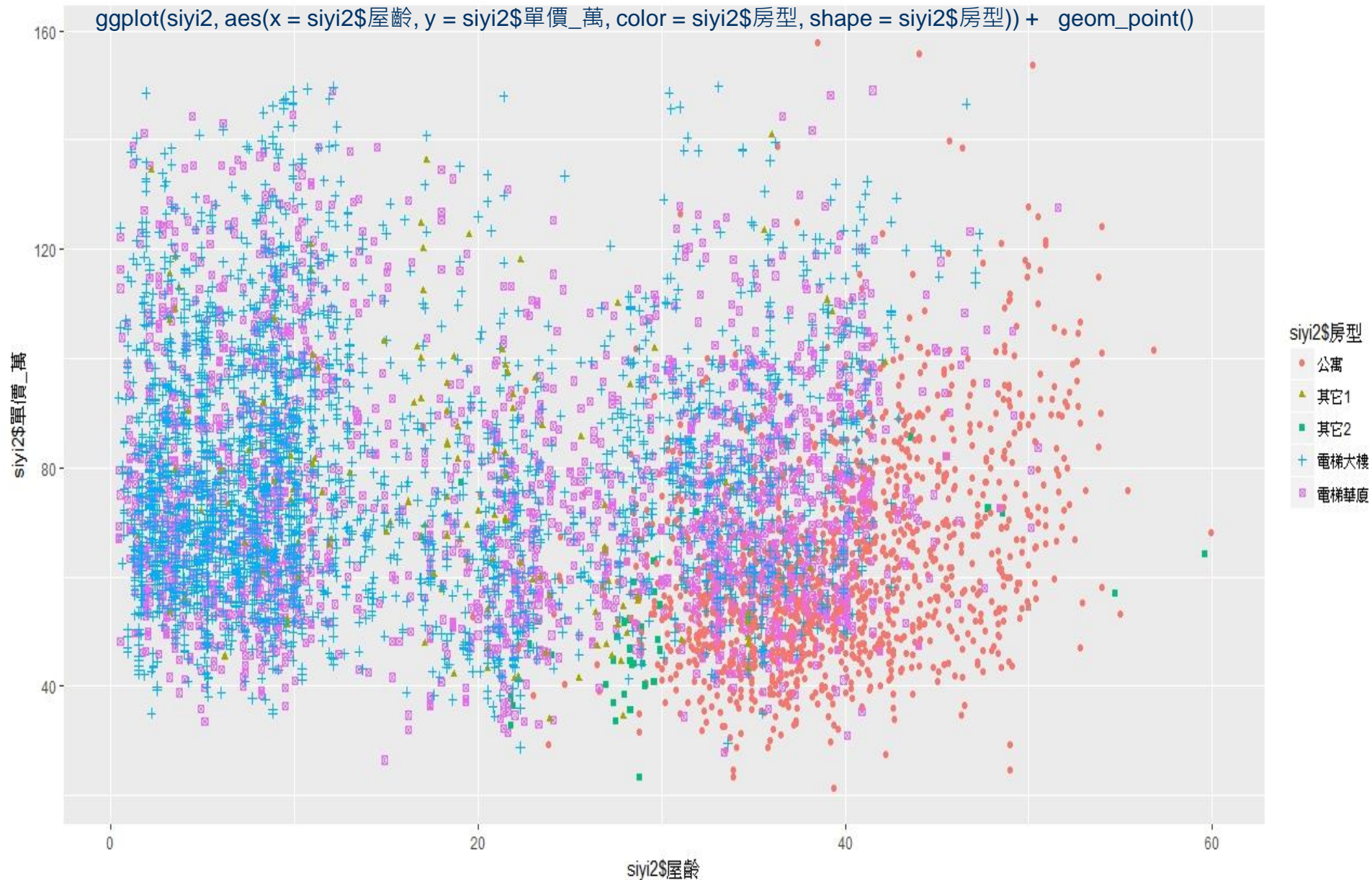
➤ 資料較穩定



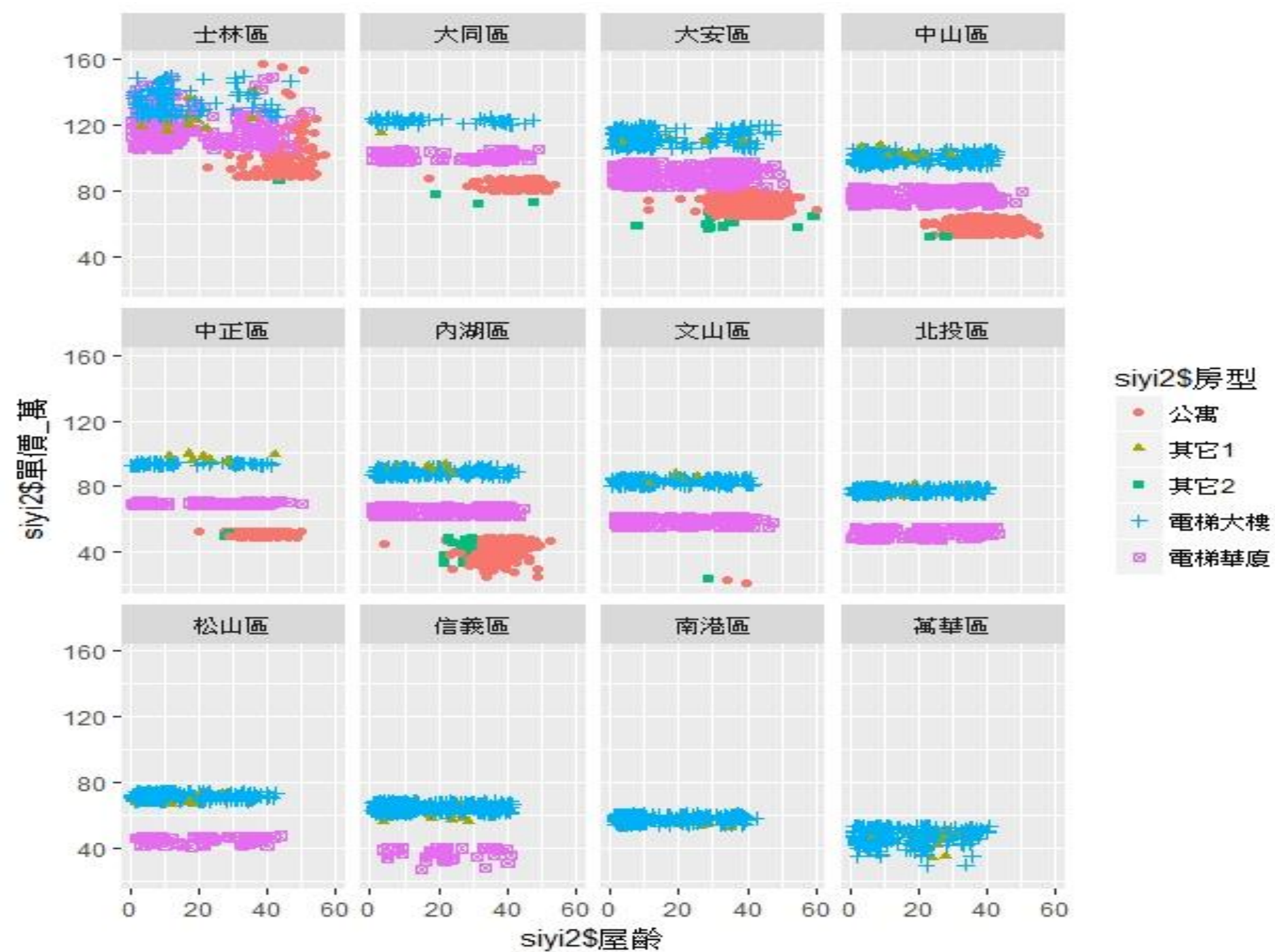
# 資料分析



```
ggplot(siyi2, aes(x = siyi2$屋齡, y = siyi2$單價_萬, color = siyi2$房型, shape = siyi2$房型)) + geom_point()
```







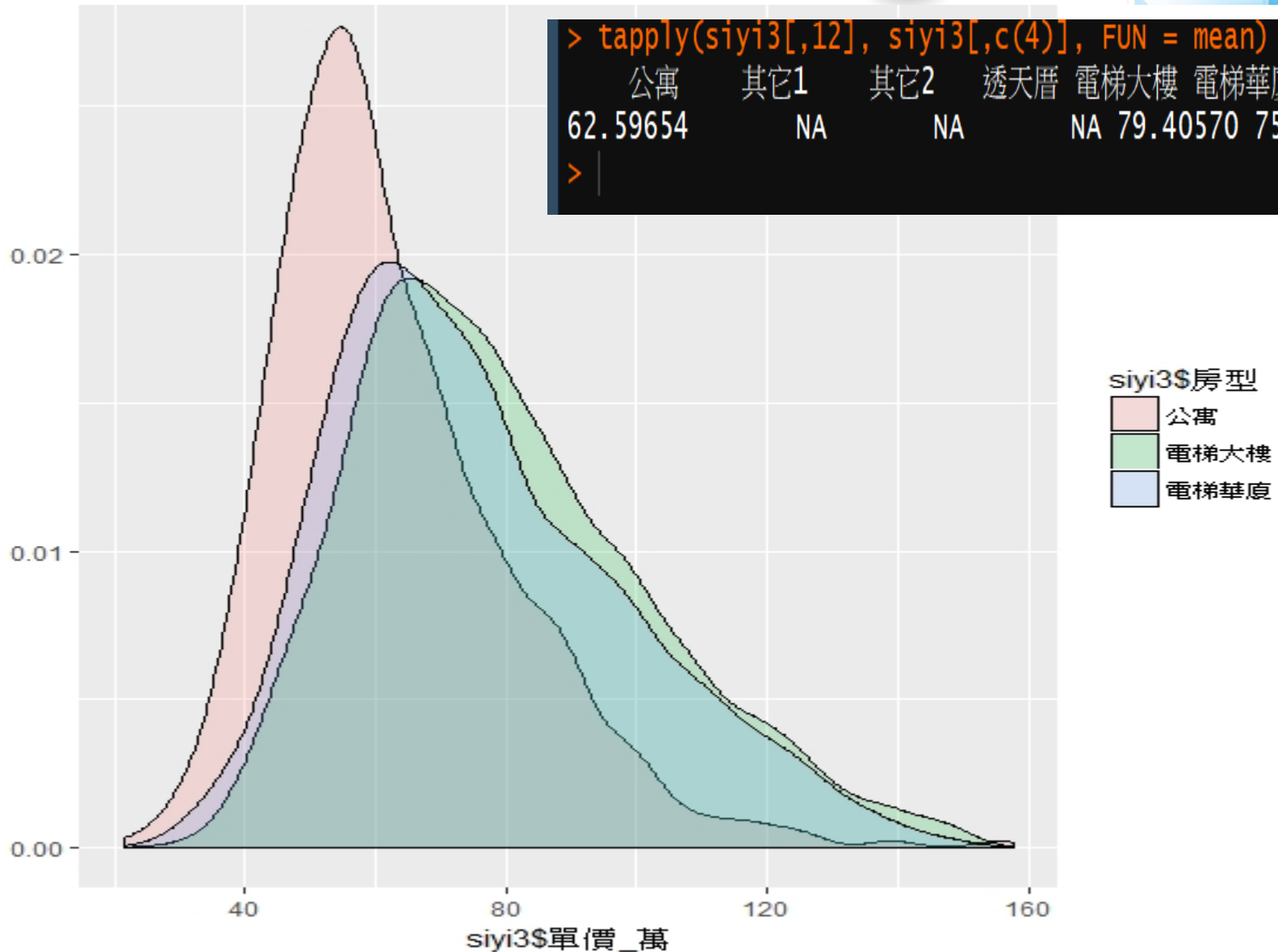
# 資料分析

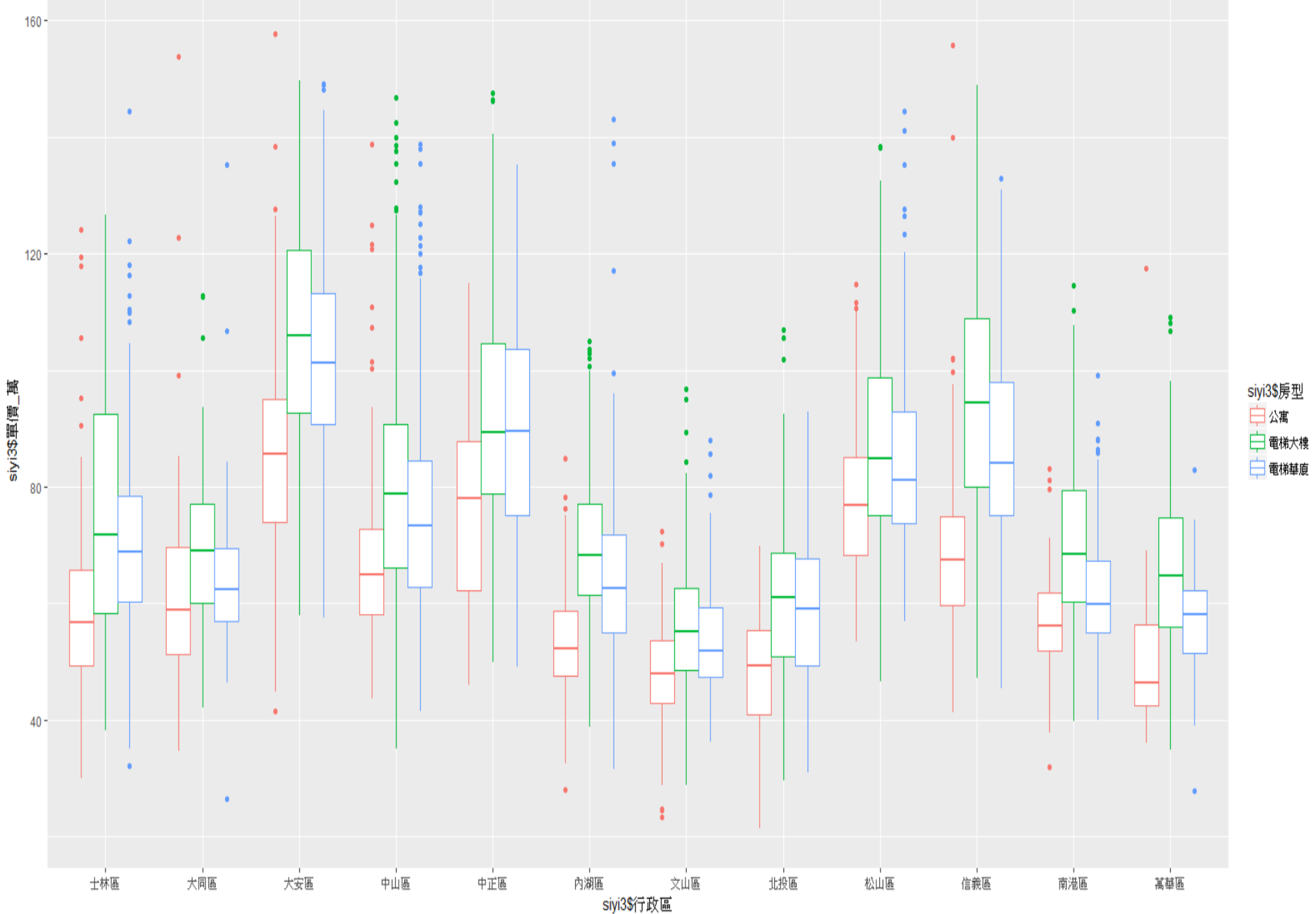


```
> tapply(siyi3[,12], siyi3[,c(4)], FUN = mean)
```

公寓	其它1	其它2	透天厝	電梯大樓	電梯華廈
62.59654	NA	NA	NA	79.40570	75.91288

```
> |
```





```
ggplot(siyi3, aes(x = siyi3$行政區, y = siyi3$單價_萬, color = siyi3$房型)) + geom_boxplot()
```

# 資料分析-各區價格



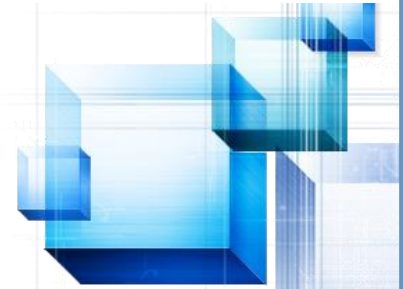
```
> tapply(siyi3[,12], siyi3[,c(2,4)], FUN = mean)
```

房型

行政區	公寓	其它1	其它2	透天厝	電梯大樓	電梯華廈
士林區	58.49611	NA	NA	NA	74.32616	70.67504
大同區	62.21175	NA	NA	NA	68.94522	64.70256
大安區	84.90795	NA	NA	NA	106.77969	102.42001
中山區	68.06405	NA	NA	NA	80.55334	76.26758
中正區	76.40744	NA	NA	NA	92.50525	89.46200
內湖區	53.04117	NA	NA	NA	69.44350	64.01728
文山區	47.74206	NA	NA	NA	56.12113	53.79838
北投區	48.34040	NA	NA	NA	61.01717	58.46944
松山區	78.31383	NA	NA	NA	87.11057	85.94482
信義區	69.20063	NA	NA	NA	95.41456	86.88159
南港區	56.71316	NA	NA	NA	70.68535	62.52429
萬華區	49.72176	NA	NA	NA	65.09550	56.44131

```
> |
```

# 資料分析-地圖概念



點資料 (SpatialPointsDataFrame)

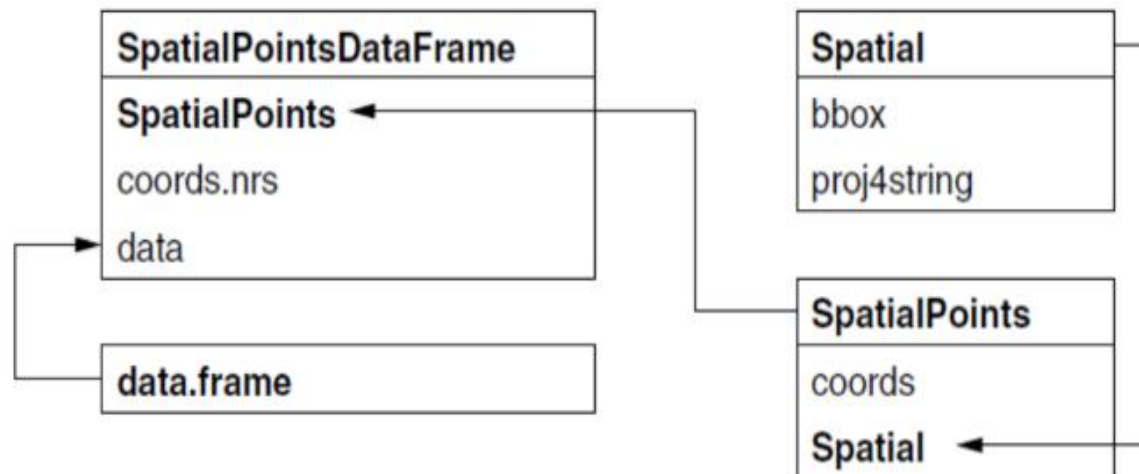


Fig. 2.2 Spatial points classes and their slots; *arrows* show subclass extensions

# 資料分析-地圖概念



線資料與多邊形 (SpatialLines and SpatialPolygons)

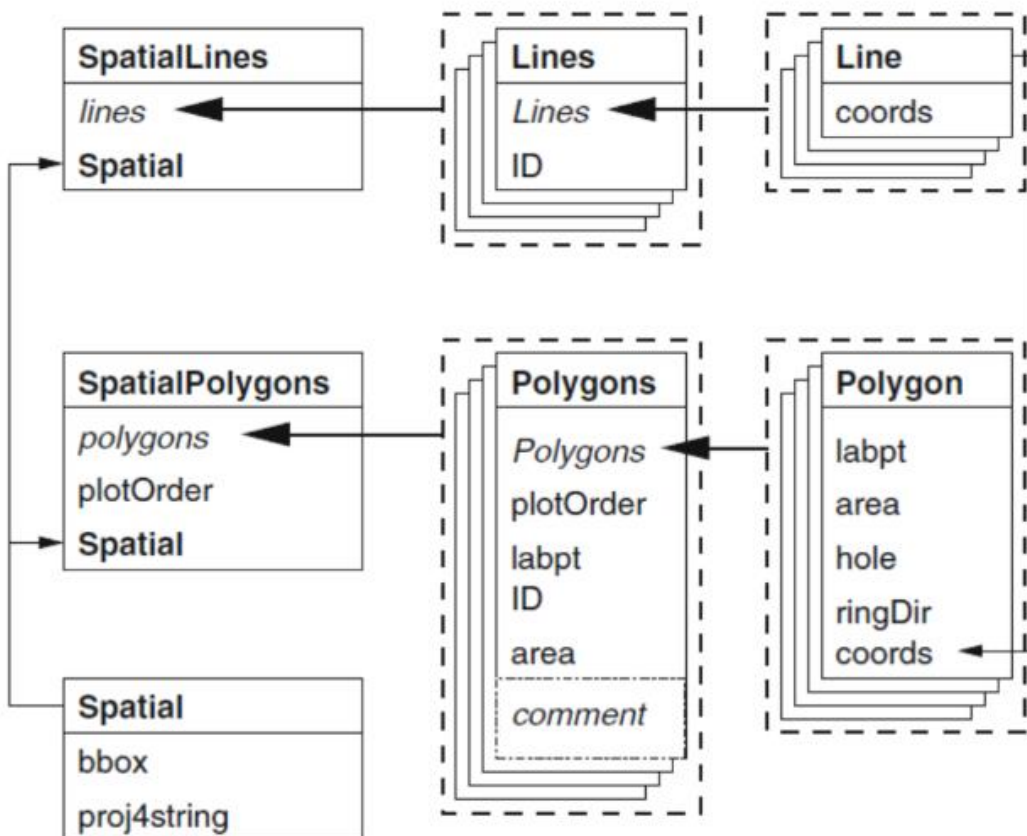
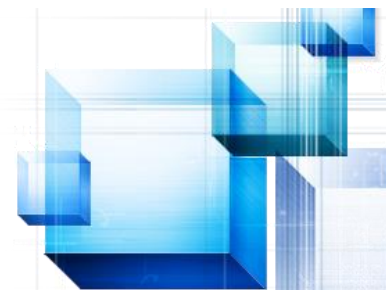


Fig. 2.4 SpatialLines and SpatialPolygons classes and slots; *thin arrows* show sub-class extensions, *thick arrows* the inclusion of lists of objects

# 資料分析-地圖概念



➤ 政府資料開放平臺  
SHP格式

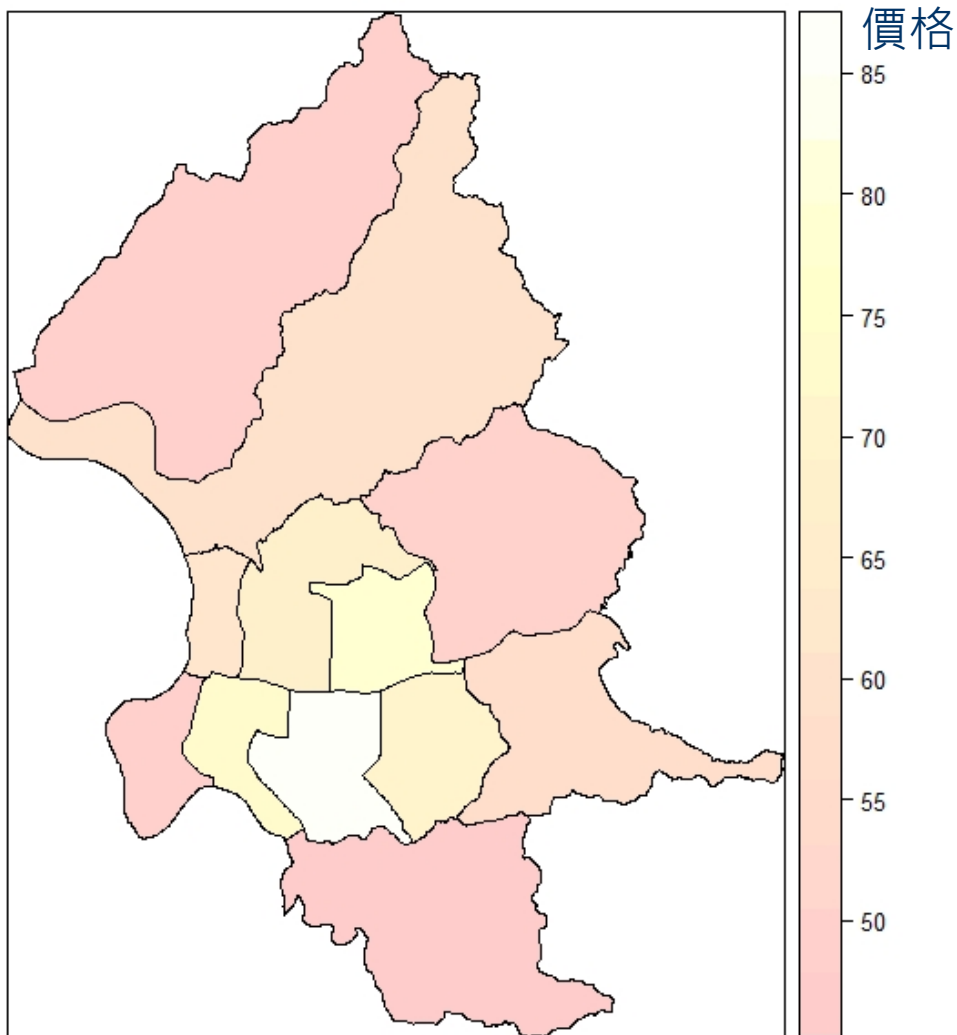
➤ 鄉鎮市區界線  
Polygons

➤ 選出台北市行政區  
圖層資訊

➤ 房價資訊放入圖層  
mapping



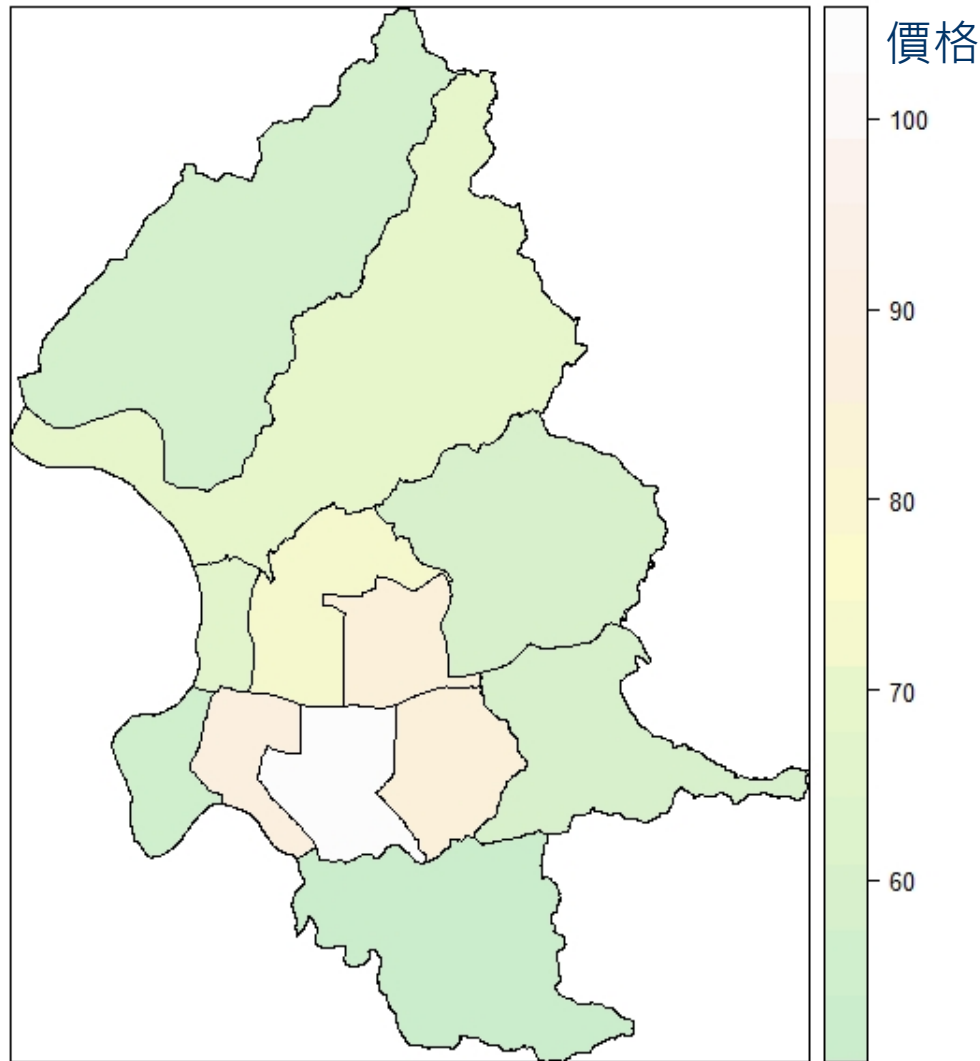
# 資料分析-公寓



公寓價格

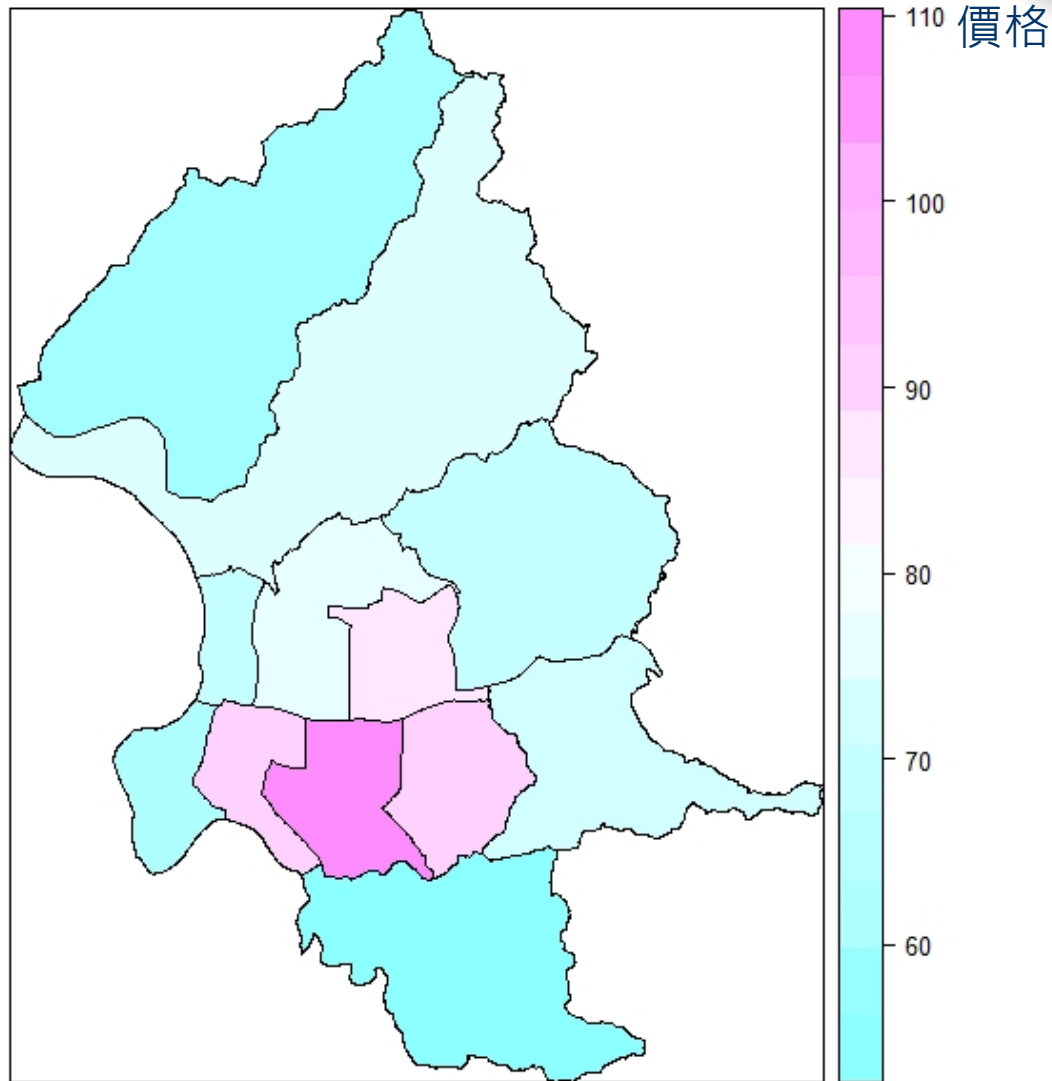
```
dev.new();spplot(x, c("公寓"), col.regions=heat.colors(20,alpha = 1(0.2)))
```

# 資料分析-電梯華廈



```
dev.new();spplot(x, c("電梯華廈"), col.regions=terrain.colors(20,alpha = I(0.2)))
```

# 資料分析-電梯大樓



```
dev.new();spplot(x, c("電梯大樓"),col.regions=cm.colors(20,alpha = I(0.9)))
```

# 資料分析-小結

## ➤ 產品價格而言

電梯大樓 > 電梯華廈 > 公寓

## ➤ 地區價格而言

大安、信義、中正價格 --- 均較高

北投、文山價格 --- 均較低

## ➤ 房地產仍以『地段』為王

以公寓的BOXPLOT來看，雖均價較低，但最高價仍超過另外二種房型，原因在於，早期建案所在地多為現在的黃金地段。

EX：瑞安街、仁愛路門牌



The image features a collection of blue, three-dimensional geometric shapes, primarily cubes and rectangular prisms, arranged in a dynamic, overlapping fashion on the left side. These shapes vary in size and are rendered with a gradient from light blue to dark blue, giving them a sense of depth and volume. A circular icon with a blue arrow pointing to the right is positioned near the center, overlapping the text. The background is white with faint, light blue horizontal and vertical lines, creating a grid-like pattern. The overall aesthetic is clean, modern, and tech-oriented.

# Thank You !

[www.themegallery.com](http://www.themegallery.com)