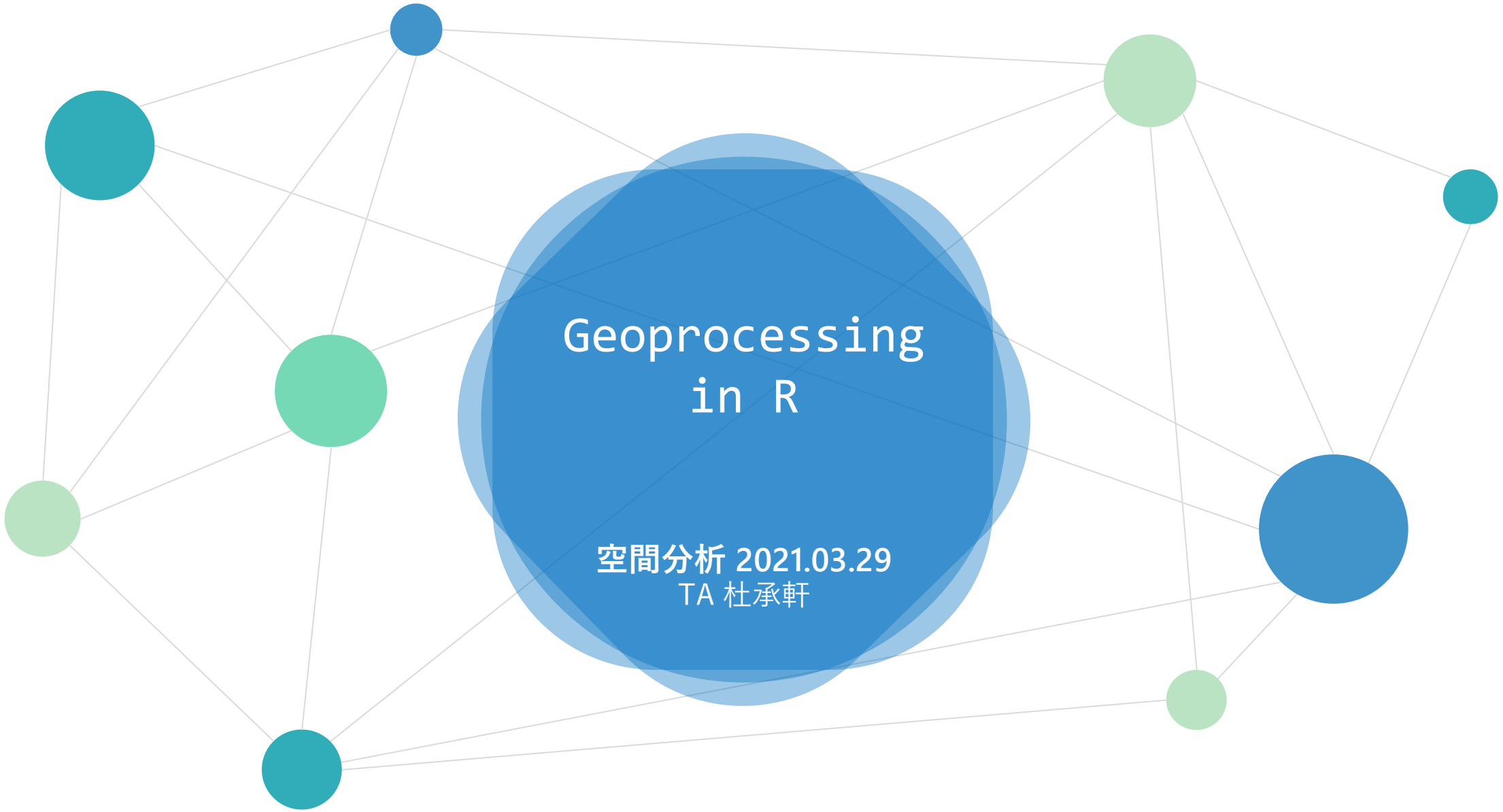


Geoprocessing in R

空間分析 2021.03.29

TA 杜承軒

Lab 3

1. 利用**村里淹水面積比例**，計算**洪災影響人數**。
2. 依照**行政區(TOWN/TOWN_ID)彙總統計**。

※注意座標格式不同要先轉換

圖資

- **Taipei_Vill.shp**
- **flood.shp**

CRS 轉換

```
TPE = st_transform(TPE, 3826)
TPE = st_transform(TPE, st_crs(TWN))
```

- EPSG: 3826
  TWD97 / TM2 zone 121
  https://epsg.io/3826
- EPSG: 4326
  WGS 84
  https://epsg.io/4326

Boundary Box

st_bbox()

```
> st_bbox(TPE)
     xmin      ymin      xmax      ymax
296094.4 2761518.3  317198.9 2789179.6
```

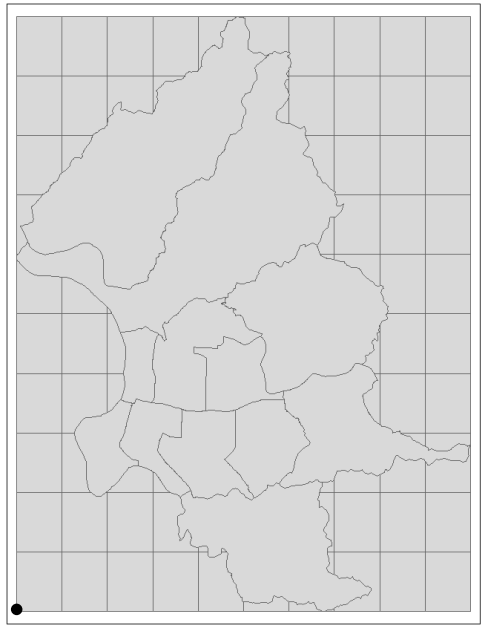
Fishnet

st_make_grid(sf, cellsize, offset, n)

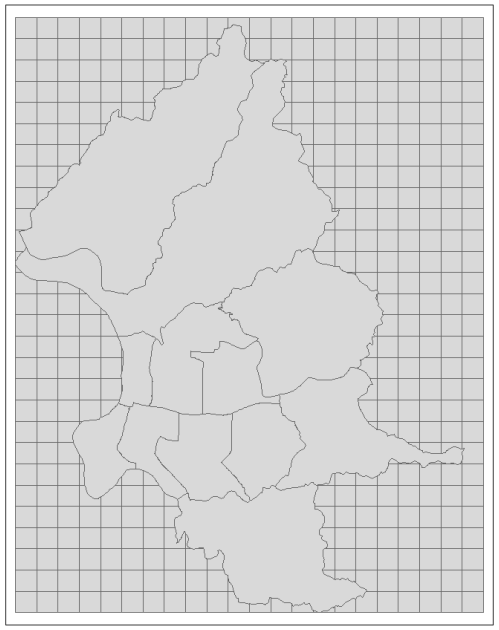
default

- $\text{cellsize} = \left(\frac{X_{\max} - X_{\min}}{n_x}, \frac{Y_{\max} - Y_{\min}}{n_y}\right)$
- $\text{offset} = (X_{\min}, Y_{\min})$
- $\text{n} = (10, 10)$

st_make_grid(TPE)



st_make_grid(TPE,1000)



Convert to sf object

st_sf(grid)
st_sf(grid,ID=1:length(grid))

%>%

```
TPE %>% st_bbox  # 同 st_bbox(TPE)
```

Dissolve

```
TOWN=VILL %>% group_by(TOWN) %>% summarise
```

↑

對哪個欄位進行群組

```
TOWN=VILL %>% group_by(TOWN) %>% summarise(POP=sum(CENSUS))
```

新欄位名稱　運算函數 e.g. length　對舊欄位做分組運算

	VILLAGE	TOWN	CENSUS
1	群英里	大安區	7438
2	虎嘯里	大安區	3852
3	臥龍里	大安區	2027
4	龍淵里	大安區	10389
5	龍門里	大安區	4696
6	大學里	大安區	8974
7	芳和里	大安區	6018
8	黎元里	大安區	5714
9	黎孝里	大安區	7865
10	黎和里	大安區	5042
11	建安里	大安區	5260
12	建倫里	大安區	6602

A tibble: 12 x 1

	TOWN
*	<chr>
1	士林區
2	大同區
3	大安區
4	中山區
5	中正區
6	內湖區
7	文山區
8	北投區
9	松山區
10	信義區
11	南港區
12	萬華區

A tibble: 12 x 2

	TOWN	POP
*	<chr>	<dbl>
1	士林區	290688
2	大同區	130991
3	大安區	314361
4	中山區	230990
5	中正區	163447
6	內湖區	286696
7	文山區	274903
8	北投區	257783
9	松山區	210648
10	信義區	229605
11	南港區	121640
12	萬華區	194278

Intersection

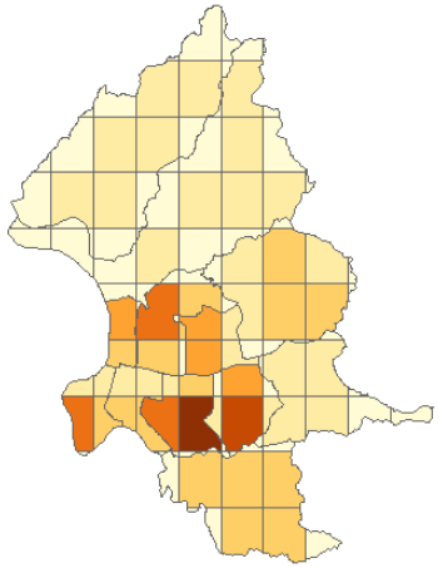
面×面→面

計算區塊人口

```
st_intersection(TOWN,grid)
```

```
TOWN$area = st_area(TOWN)
TG = st_intersection(TOWN,grid)
```

```
TG$area.in = st_area(TG)
TG$pop.in = TG$POP*TG$area.in/TG$area
TG$pop.in = round(TG$pop.in)
```



? as.integer vs. round

```
# A tibble: 126 x 4
```

	TOWN	POP	area	ID
*	<chr>	<dbl>	[m^2]	<int>
1	文山區	274903	31219238	5
2	文山區	274903	31219238	6
3	文山區	274903	31219238	7
4	文山區	274903	31219238	8
5	萬華區	194278	7480598	12
6	大安區	314361	11331460	14
7	中正區	163447	7408349	14
8	文山區	274903	31219238	14
9	大安區	314361	11331460	15
10	文山區	274903	31219238	15

TOWN　　grid

```
# A tibble: 126 x 4
```

	POP	area	area.in	pop.in
	<dbl>	[m^2]	[m^2]	<int>
1	274903	31219238	1301261.50	11458
2	274903	31219238	4778710.75	42079
3	274903	31219238	4985635.47	43901
4	274903	31219238	1601433.58	14101
5	194278	7480598	76277.68	1981
6	314361	11331460	64120.28	1778
7	163447	7408349	71121.51	1569
8	274903	31219238	1728049.21	15216
9	314361	11331460	263316.82	7305
10	274903	31219238	5423815.18	47759

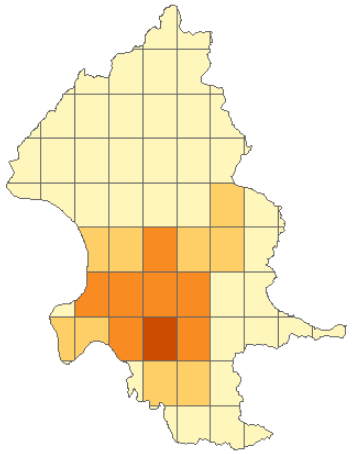
```
# ... with 116 more rows
```

Link Data

TG → grid

```
NEW.G=TG%>% group_by(ID) %>% summarise(POP=sum(pop.in))
```

```
# A tibble: 68 x 2
     ID   POP
 * <int> <int>
1     5 11458
2     6 42079
3     7 43901
4     8 14101
5    12  1981
```

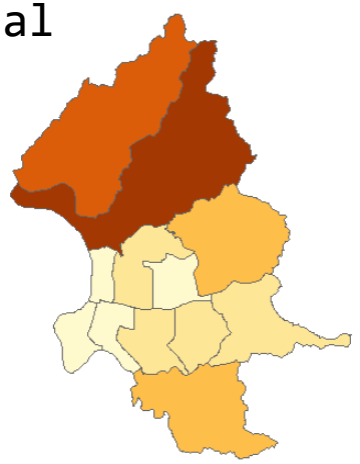


```
grid$POP[NEW.G$ID] = NEW.G$POP
```

TG → TOWN

```
NEW.T =TG%>% group_by(TOWN) %>% summarise(total=length(TOWN))
TOWN$count[NEW.T$TOWN] = NEW.T$total
```

```
# A tibble: 12 x 2
   TOWN   total
 * <chr>  <int>
1 士林區    26
2 大同區     4
3 大安區     8
4 中山區     9
```



HW3

Q：A區連鎖密度平均是否顯著高於B區

1. 虛無假設與對立假設

   $H_0: \mu_1 - \mu_2 \leq 0 \quad H_1: \mu_1 - \mu_2 > 0 \quad (\alpha = 0.05)$

2. 計算統計量

   兩母體平均差 → t檢定

3. 比較 $p - value$ 與 α (或觀察信賴區間範圍) *注意單雙尾

4. 拒絕虛無假設？

   $p - value < \alpha \quad \rightarrow \quad$ 拒絕H_0

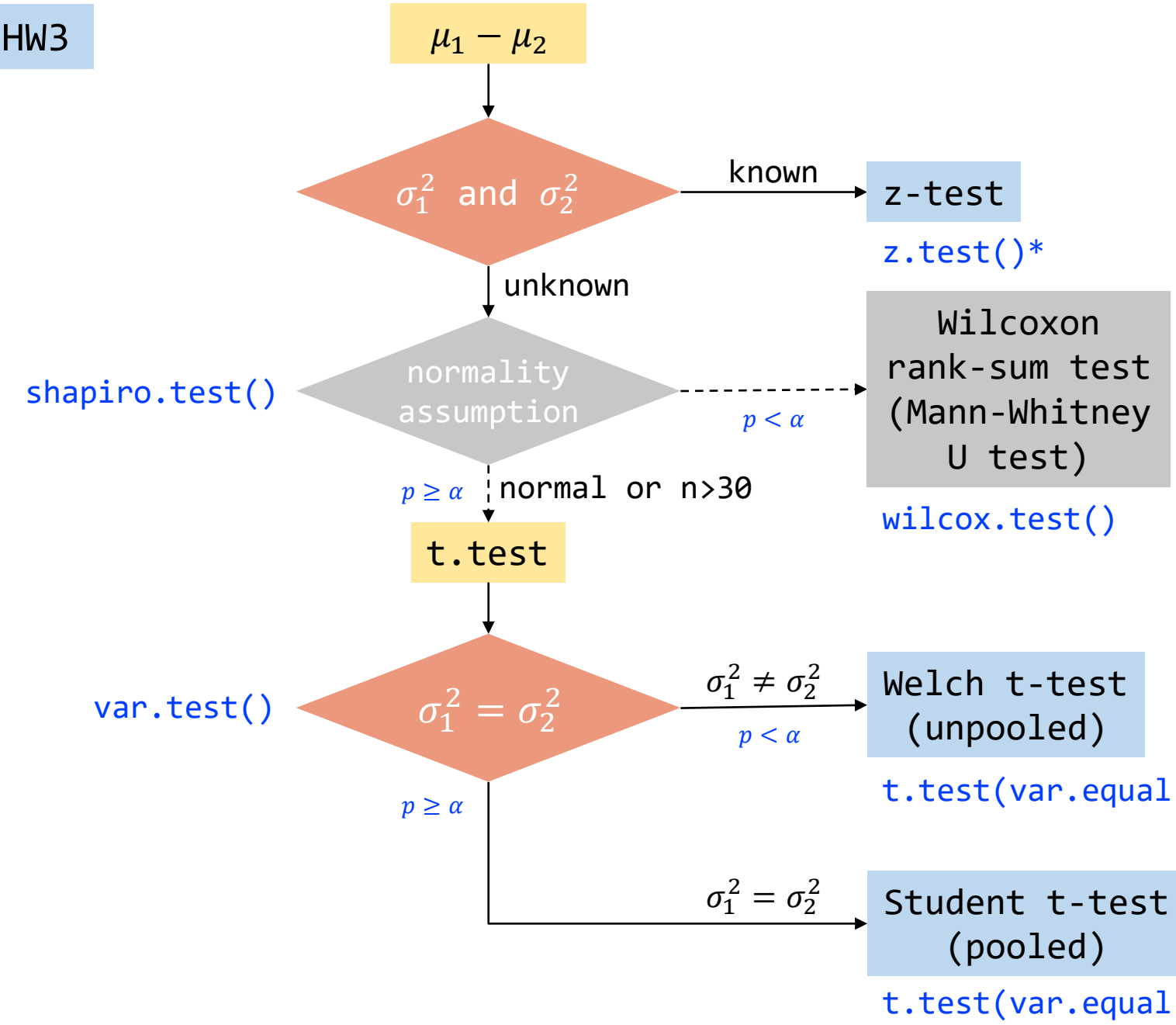
   $p - value \geq \alpha \quad \rightarrow \quad$ 接受H_0

5. 結論

```
> t.test(A,B,alternative="greater",paired=F,var.equal=T,conf.level=0.95)
```

HW3

* Hypothesis Testing of $\mu_1 - \mu_2$




```
> wilcox.test(A,B,alternative="greater",exact = F)

        Wilcoxon rank sum test with continuity correction

data:  A and B
W = 225.5, p-value = 0.6946
alternative hypothesis: true location shift is greater than 0
```

```
> t.test(A,B,alternative="greater",var.equal = F)

        Welch Two Sample t-test

data:  A and B
t = -0.61475, df = 42.502, p-value = 0.729
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -2.646442       Inf
sample estimates:
mean of x mean of y
 6.923077  7.631579
```

```
> t.test(A,B,alternative="greater",var.equal = T)

        Two Sample t-test

data:  A and B
t = -0.59458, df = 43, p-value = 0.7224
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -2.711675       Inf
sample estimates:
mean of x mean of y
 6.923077  7.631579
```

HW3

* Hypothesis Testing of $\mu_1 - \mu_2$ when σ_1^2 and σ_2^2 are unknown.

Assume $\sigma_1^2 = \sigma_2^2 = S_p^2$ → pooled

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2), \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Assume $\sigma_1^2 \neq \sigma_2^2$ → unpooled

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(df), \qquad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

* $\sigma_1^2 = \sigma_2^2$?

$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$ $(\alpha = 0.05)$

兩母體變異數比 → F檢定

```
> var.test(A,B)
```

$$F = \frac{s_1^2}{s_2^2} \sim F\,(n_1 - 1, n_2 - 1)$$

```
> var.test(A,B)

        F test to compare two variances

data:  A and B
F = 1.5307, num df = 25, denom df = 18, p-value = 0.3553
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6144293 3.5788918
sample estimates:
ratio of variances
          1.530676
```

- P.S. 當$n_1 = n_2$時→ 兩者沒差
- P.S. 一說為0.5<s_1/s_2<2用pooled，其餘用unpooled

HW4
(Mid 1)

HW4 批改標準：該題全符合才給分

1. (1)列表呈現
10% (2)四個數字正確

```
##      type stores seats
## 1 private    ANS  4860
## 2  public     82   ANS
```

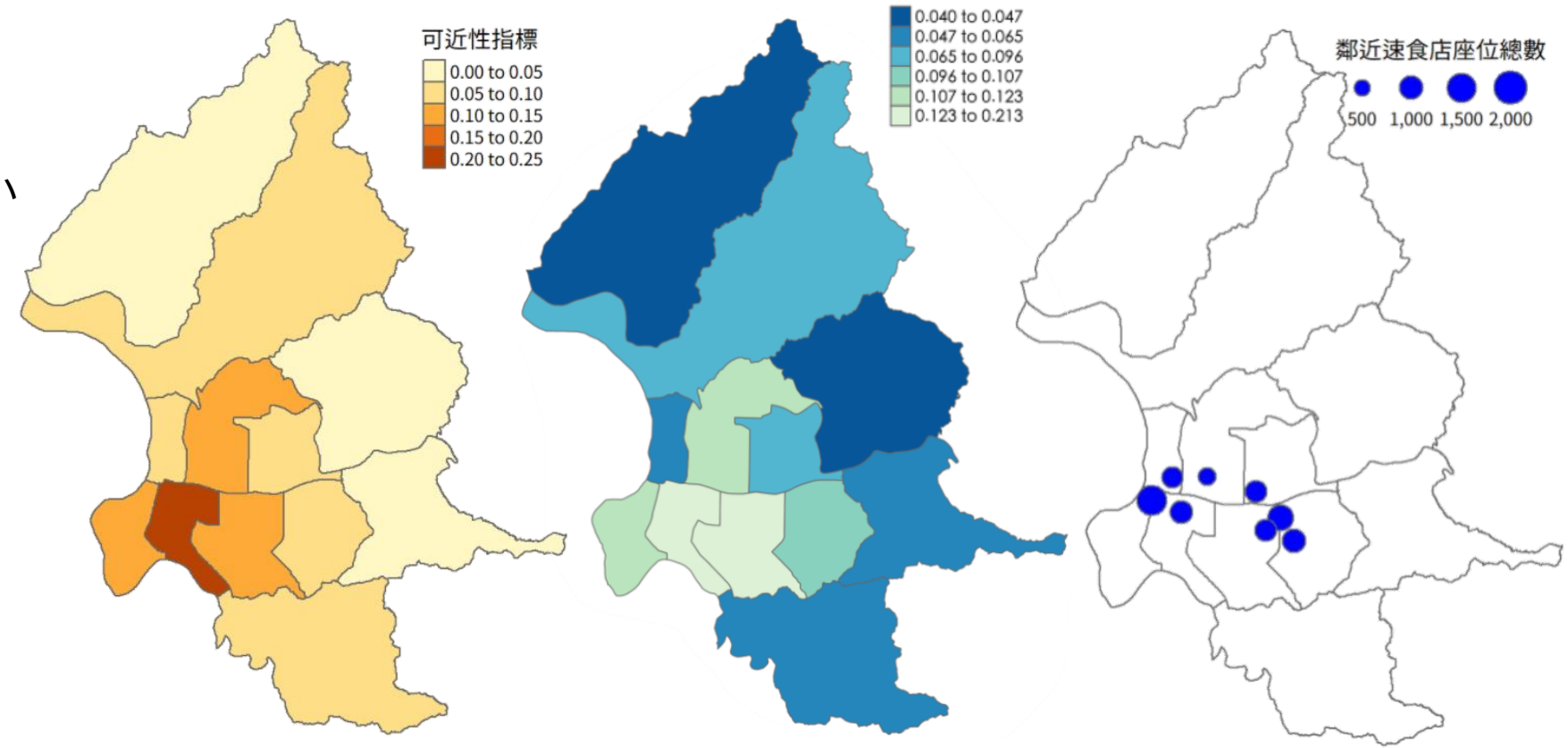
```
## type      store  seat
##   private   105 11759
##   public    246 26320
```

```
#補充：which.max
x=SC$store
ids=c()
for(i in 1:8){
  id=which.max(x)
  x[id]=NA
  ids=c(ids,id)
}
all(ids==order(SC$store,decreasing=T)[1:8])
```

```
## [1] TRUE
```

2. (1)數值正確(落在0.039~0.213之間)
10% (2)面量圖顏色相對深淺關係
*xtabs()要注意順序

3. (1)8間小學的位置
20% (2)泡泡圖調整大小
*善用order()



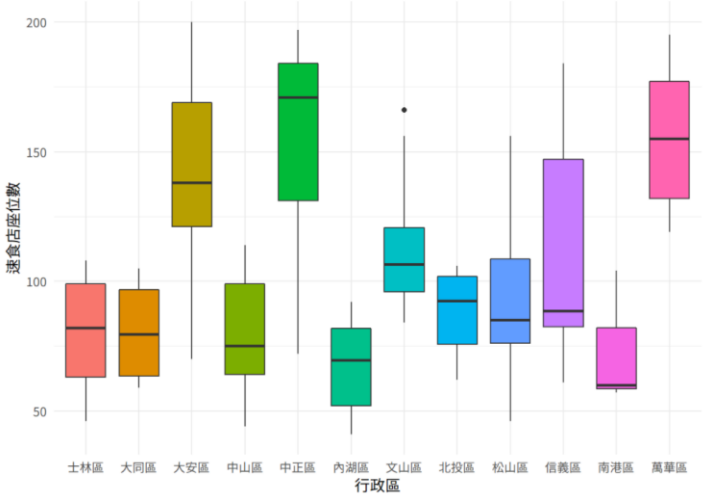
HW4
(Mid 1)

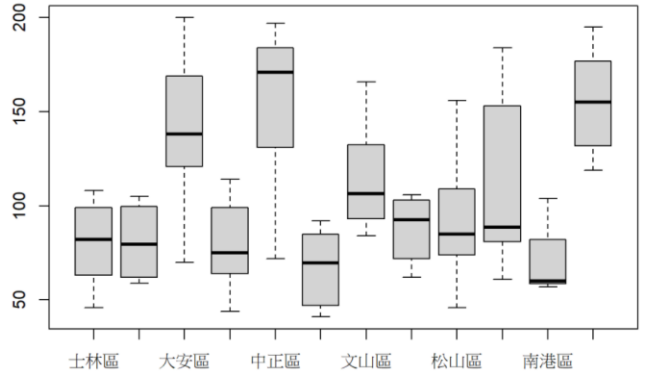
HW4 批改標準：該題全符合才給分

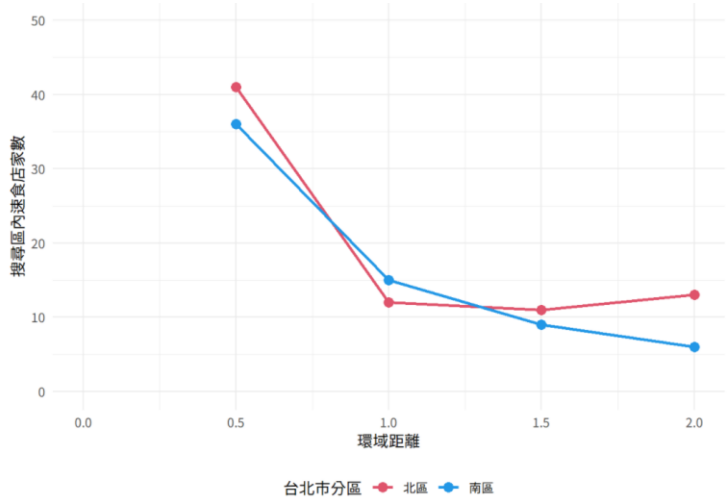
20% 4-1. 盒狀圖正確
10% 4-2. (1)假設檢定步驟
(2)列出ANOVA報表

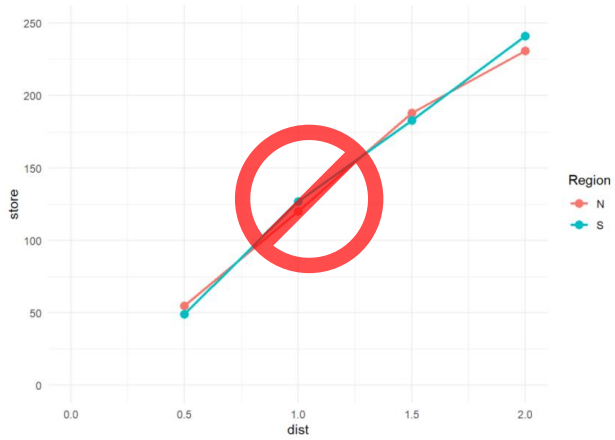
20% 5-1.折線圖正確
10% 5-2. (1)假設檢定步驟
(2)列出卡方報表

*避免手動重複計算









假設檢定步驟：

- H0 & Ha
- 方法、統計量
- p-value
- 結果
- 結論