HW4_1

陳凱騫

2024-10-28

第一題

```
library(foreign)
library(Hmisc)
```

```
Attaching package: 'Hmisc'
```

```
The following objects are masked from 'package:base':
```

```
    format.pval, units
```

```
library(table1)
```

```
Attaching package: 'table1'
```

```
The following objects are masked from 'package:Hmisc':
```

```
    label, label<-, units
```

```
The following objects are masked from 'package:base':
```

```
    units, units<-
```

```
library(DataExplorer)
library(knitr)
library(dplyr)
```

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:Hmisc':
```

```
    src, summarize
```

```
The following objects are masked from 'package:stats':
```

```
    filter, lag
```

```
The following objects are masked from 'package:base':
```

```
    intersect, setdiff, setequal, union
```

```
#Sys.setlocale("LC_ALL","Chinese (Traditional)_Taiwan.950")
```

進行資料讀取與基本資料解析

```
data <- read.spss("poll.sav")
```

```
re-encoding from CP950
```

```
Warning in read.spss("poll.sav"): Undeclared level(s) 91 added in variable:
v4_1
```

```
Warning in read.spss("poll.sav"): Undeclared level(s) 91 added in variable: v5
```

```
data <- data.frame(data)
```

```
latex(describe(data),file="", caption.placement = "top")
```

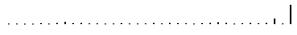
data
15 Variables 1671 Observations

v1

n	missing	distinct
1671	0	2

Value		
Frequency	1107	564
Proportion	0.662	0.338

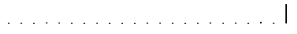
v2



n	missing	distinct
1671	0	36

```
lowest :
highest:
```

v3



n	missing	distinct
1671	0	23

```
lowest :
highest:
```

v4_1

n	missing	distinct
1671	0	12

Value	(1)	(10)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	91	
Frequency	328	11	5	214	43	27	38	47	4	1	14	939
Proportion	0.196	0.007	0.003	0.128	0.026	0.016	0.023	0.028	0.002	0.001	0.008	0.562

v4_2

n	missing	distinct
1671	0	10

Value	(2)	(3)	(4)	(5)	(6)	(7)
Frequency	6	189	59	32	75	99
Proportion	0.004	0.113	0.035	0.019	0.045	0.059

Value	(8)	(9)	(10)
Frequency	2	4	15	1190
Proportion	0.001	0.002	0.009	0.712

v4_3

n	missing	distinct
1671	0	9

Value	(3)	(4)	(5)	(6)	(7)	(8)
Frequency	6	60	36	61	91	1
Proportion	0.004	0.036	0.022	0.037	0.054	0.001

Value	(9)	(10)
Frequency	2	19	1395
Proportion	0.001	0.011	0.835

v4_4

n	missing	distinct
1671	0	8

Value	(4)	(5)	(6)	(7)	(8)	(9)
Frequency	4	28	41	52	3	4
Proportion	0.002	0.017	0.025	0.031	0.002	0.002

Value	(10)
Frequency	20	1519
Proportion	0.012	0.909

v4_5

n	missing	distinct
1671	0	7

Value	(5)	(6)	(7)	(8)	(9)	(10)
Frequency	3	14	38	4	3	15
Proportion	0.002	0.008	0.023	0.002	0.002	0.009

Value	
Frequency	1594
Proportion	0.954

v4_6

n	missing	distinct
1671	0	6

Value	(6)	(7)	(8)	(9)	(10)	
Frequency	3	12	6	7	20	1623
Proportion	0.002	0.007	0.004	0.004	0.012	0.971

v4_7

n	missing	distinct
1671	0	5

Value	(7)	(8)	(9)	(10)	
Frequency	3	2	3	12	1651
Proportion	0.002	0.001	0.002	0.007	0.988

v4_8

n	missing	distinct
1671	0	3

Value	(8)	(10)	
Frequency	1	4	1666
Proportion	0.001	0.002	0.997

v5

n	missing	distinct
1671	0	13

Value	(1)	(10)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Frequency	158	53	9	205	79	33	98	195	6	
Proportion	0.095	0.032	0.005	0.123	0.047	0.020	0.059	0.117	0.004	

Value	(9)	91								
Frequency	8	10	269	548						
Proportion	0.005	0.006	0.161	0.328						

v6

n	missing	distinct
1671	0	6

Value	20-29	30-39	40-49	50-59	60
Frequency	52	94	201	336	946	42
Proportion	0.031	0.056	0.120	0.201	0.566	0.025

v7

n	missing	distinct
1671	0	6

Value						
Frequency	292	165	431	198	520	65
Proportion	0.175	0.099	0.258	0.118	0.311	0.039

v8

```
     n    missing    distinct
  1671          0           2

Value
Frequency    682     989
Proportion 0.408   0.592
```

將"遺漏值或跳答"改為遺漏值，並以回答數(v4的總和)作為政治熱衷程度(以新變 量"v4"代表)。 v2與v3可以合併為一個變數，因為都是區底下的里，且不會出現交集。

```
data[data=="    "] <- NA
v4 <- rowSums(!is.na(subset(data,select=c(v4_1,v4_2,v4_3,v4_4,v4_5,v4_6,v4_7,v4_8))))
data_new <- cbind(subset(data,select=-c(v4_1,v4_2,v4_3,v4_4,v4_5,v4_6,v4_7,v4_8)),v4)
v9 <- coalesce(data_new$v2, data_new$v3)
data_new <- cbind(subset(data_new,select=-c(v2,v3)),v9)
data_new <- data_new %>%
  select(v1,v9,v4,v5,v6,v7,v8)
# head(data_new,n=20)
```

進行table 1 繪製。

```
table1(~.|v1,data_new)
```

```
Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	北區	中西區	Overall
	(N=1107)	(N=564)	(N=1671)
v9			
北華里	14 (1.3%)	0 (0%)	14 (0.8%)
元寶里	12 (1.1%)	0 (0%)	12 (0.7%)
中樓里	4 (0.4%)	0 (0%)	4 (0.2%)
東興里	8 (0.7%)	0 (0%)	8 (0.5%)
正覺里	35 (3.2%)	0 (0%)	35 (2.1%)
公園里	10 (0.9%)	0 (0%)	10 (0.6%)
成功里	39 (3.5%)	0 (0%)	39 (2.3%)
大港里	70 (6.3%)	0 (0%)	70 (4.2%)
大和里	34 (3.1%)	0 (0%)	34 (2.0%)
大豐里	38 (3.4%)	0 (0%)	38 (2.3%)
賢北里	25 (2.3%)	0 (0%)	25 (1.5%)
開元里	24 (2.2%)	0 (0%)	24 (1.4%)
長勝里	12 (1.1%)	0 (0%)	12 (0.7%)
振興里	19 (1.7%)	0 (0%)	19 (1.1%)
合興里	32 (2.9%)	0 (0%)	32 (1.9%)
力行里	18 (1.6%)	0 (0%)	18 (1.1%)
重興里	16 (1.4%)	0 (0%)	16 (1.0%)
仁愛里	23 (2.1%)	0 (0%)	23 (1.4%)
北門里	40 (3.6%)	0 (0%)	40 (2.4%)
永祥里	21 (1.9%)	0 (0%)	21 (1.3%)
小北里	41 (3.7%)	0 (0%)	41 (2.5%)
大光里	33 (3.0%)	0 (0%)	33 (2.0%)
大興里	29 (2.6%)	0 (0%)	29 (1.7%)
長興里	12 (1.1%)	0 (0%)	12 (0.7%)
華德里	32 (2.9%)	0 (0%)	32 (1.9%)
福德里	42 (3.8%)	0 (0%)	42 (2.5%)
成德里	57 (5.1%)	0 (0%)	57 (3.4%)
和順里	39 (3.5%)	0 (0%)	39 (2.3%)

	北區	中西區	Overall
立人里	12 (1.1%)	0 (0%)	12 (0.7%)
雙安里	28 (2.5%)	0 (0%)	28 (1.7%)
文成里	37 (3.3%)	0 (0%)	37 (2.2%)
元美里	31 (2.8%)	0 (0%)	31 (1.9%)
文元里	42 (3.8%)	0 (0%)	42 (2.5%)
不知道	161 (14.5%)	84 (14.9%)	245 (14.7%)
拒答	17 (1.5%)	18 (3.2%)	35 (2.1%)
遺漏值或跳答	0 (0%)	0 (0%)	0 (0%)
郡王里	0 (0%)	19 (3.4%)	19 (1.1%)
赤嵌里	0 (0%)	25 (4.4%)	25 (1.5%)
法華里	0 (0%)	14 (2.5%)	14 (0.8%)
西湖里	0 (0%)	51 (9.0%)	51 (3.1%)
西賢里	0 (0%)	42 (7.4%)	42 (2.5%)
城隍里	0 (0%)	20 (3.5%)	20 (1.2%)
開山里	0 (0%)	14 (2.5%)	14 (0.8%)
南美里	0 (0%)	24 (4.3%)	24 (1.4%)
永華里	0 (0%)	8 (1.4%)	8 (0.5%)
小西門里	0 (0%)	15 (2.7%)	15 (0.9%)
五條港里	0 (0%)	9 (1.6%)	9 (0.5%)
兌悅里	0 (0%)	29 (5.1%)	29 (1.7%)
藥王里	0 (0%)	34 (6.0%)	34 (2.0%)
淺草里	0 (0%)	18 (3.2%)	18 (1.1%)
府前里	0 (0%)	6 (1.1%)	6 (0.4%)
南廠里	0 (0%)	19 (3.4%)	19 (1.1%)
西和里	0 (0%)	43 (7.6%)	43 (2.6%)
大涼里	0 (0%)	22 (3.9%)	22 (1.3%)
光賢里	0 (0%)	35 (6.2%)	35 (2.1%)
南門里	0 (0%)	15 (2.7%)	15 (0.9%)
v4			
Mean (SD)	1.68 (1.30)	1.55 (1.18)	1.63 (1.26)
Median [Min, Max]	1.00 [1.00, 8.00]	1.00 [1.00, 8.00]	1.00 [1.00, 8.00]
v5			
(1號)	56 (5.1%)	102 (18.1%)	158 (9.5%)
(10號)	30 (2.7%)	23 (4.1%)	53 (3.2%)
(2號)	7 (0.6%)	2 (0.4%)	9 (0.5%)
(3號)	163 (14.7%)	42 (7.4%)	205 (12.3%)
(4號)	32 (2.9%)	47 (8.3%)	79 (4.7%)
(5號)	29 (2.6%)	4 (0.7%)	33 (2.0%)
(6號)	83 (7.5%)	15 (2.7%)	98 (5.9%)
(7號)	143 (12.9%)	52 (9.2%)	195 (11.7%)
(8號)	3 (0.3%)	3 (0.5%)	6 (0.4%)
(9號)	4 (0.4%)	4 (0.7%)	8 (0.5%)
91	7 (0.6%)	3 (0.5%)	10 (0.6%)
無反應	185 (16.7%)	84 (14.9%)	269 (16.1%)
尚未決定	365 (33.0%)	183 (32.4%)	548 (32.8%)
v6			
20-29歲	35 (3.2%)	17 (3.0%)	52 (3.1%)
30-39歲	70 (6.3%)	24 (4.3%)	94 (5.6%)
40-49歲	144 (13.0%)	57 (10.1%)	201 (12.0%)
50-59歲	231 (20.9%)	105 (18.6%)	336 (20.1%)
60歲及以上	602 (54.4%)	344 (61.0%)	946 (56.6%)
不知道/拒答	25 (2.3%)	17 (3.0%)	42 (2.5%)

	北區	中西區	Overall
遺漏值或跳答	0 (0%)	0 (0%)	0 (0%)
v7			
小學或以下	187 (16.9%)	105 (18.6%)	292 (17.5%)
初中、國中	105 (9.5%)	60 (10.6%)	165 (9.9%)
高中、高職	300 (27.1%)	131 (23.2%)	431 (25.8%)
專科	124 (11.2%)	74 (13.1%)	198 (11.8%)
大學以上	357 (32.2%)	163 (28.9%)	520 (31.1%)
拒答	34 (3.1%)	31 (5.5%)	65 (3.9%)
遺漏值或跳答	0 (0%)	0 (0%)	0 (0%)
v8			
男性	457 (41.3%)	225 (39.9%)	682 (40.8%)
女性	650 (58.7%)	339 (60.1%)	989 (59.2%)
遺漏值或跳答	0 (0%)	0 (0%)	0 (0%)

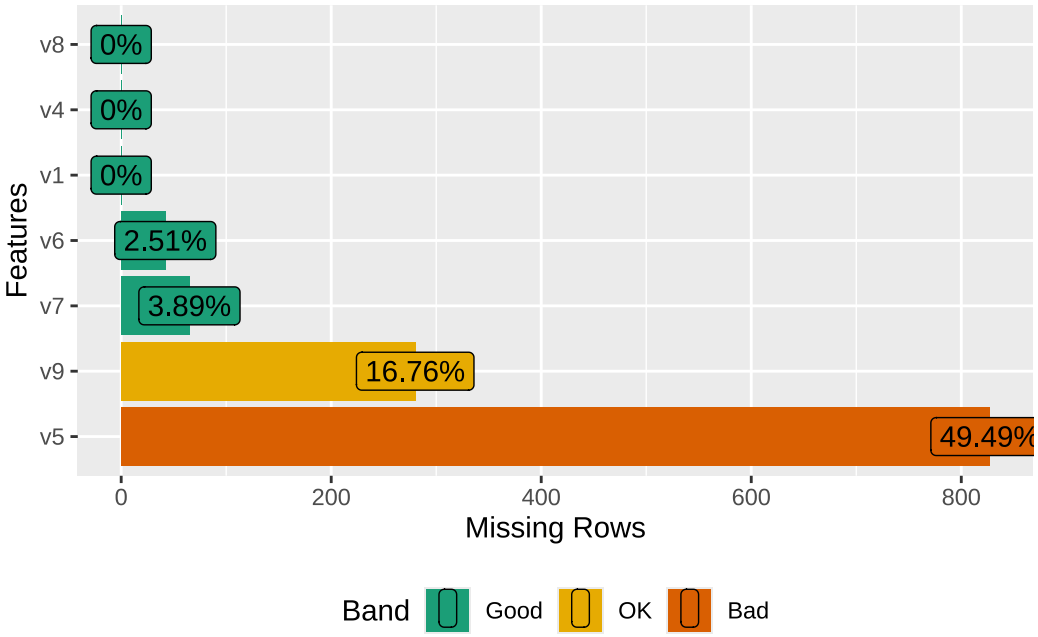
遺失值，先定義每一個變量的遺失定義(使用summary/table找出可能的遺失選項)。

```
discrip <- data.frame(variable=c("v1( ) v4(  ) v8( )","v9( )","v5( )",
                                 "v6( )","v7( )"),
                      discription=c("    ","    ","     91",
                                    "    "," "))
  #table(data_new$v5)
kable(discrip,caption="  ")
```

表 2: 遺失選項

variable	discription
v1(區)、v4(熱衷度)、v8(性別)	不會出現遺失值
v9(里)	不知道、拒答
v5(支持)	無反應、尚未決定、91
v6(年紀)	不知道/拒答
v7(學歷)	拒答

```
data_new[data_new$v9==""  "|data_new$v9==" ",]$v9 <- NA
data_new[data_new$v5==""  "|data_new$v5==""  "|data_new$v5==91,]$v5 <- NA
data_new[data_new$v6==""    ",]$v6<- NA
data_new[data_new$v7==" ",]$v7 <- NA
plot_missing(data_new)
```



分析所有候選人的支持者分析。

第二題

```
data_nom <- data_new[!is.na(data_new$v5),]
table1(~.|v5,data_nom)
```

```
Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	(1號) (N=158)	(10號) (N=53)	(2號) (N=9)	(3號) (N=205)	(4號) (N=79)	(5號) (N=33)	(6號) (N=98)	(7號) (N=195)	(8號) (N=6)	(9號) (N=8)	Overall (N=844)
v1											
北區	56 (35.4%)	30 (56.6%)	7 (77.8%)	163 (79.5%)	32 (40.5%)	29 (87.9%)	83 (84.7%)	143 (73.3%)	3 (50.0%)	4 (50.0%)	550 (65.2%)
中西區	102 (64.6%)	23 (43.4%)	2 (22.2%)	42 (20.5%)	47 (59.5%)	4 (12.1%)	15 (15.3%)	52 (26.7%)	3 (50.0%)	4 (50.0%)	294 (34.8%)
拒答/戶籍不在臺南市以上幾區/家中無有投票權的人	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
遺漏值或跳答	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
v9											
北華里	2 (1.3%)	0 (0%)	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	1 (1.0%)	2 (1.0%)	0 (0%)	0 (0%)	6 (0.7%)
元寶里	2 (1.3%)	1 (1.9%)	0 (0%)	2 (1.0%)	0 (0%)	0 (0%)	1 (1.0%)	1 (0.5%)	0 (0%)	0 (0%)	7 (0.8%)
中樓里	1 (0.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.1%)
東興里	1 (0.6%)	1 (1.9%)	0 (0%)	1 (0.5%)	1 (1.3%)	0 (0%)	1 (1.0%)	1 (0.5%)	0 (0%)	0 (0%)	6 (0.7%)

	(1號)	(10號)	(2號)	(3號)	(4號)	(5號)	(6號)	(7號)	(8號)	(9號)	Overall
正覺里	2 (1.3%)	3 (5.7%)	0 (0%)	8 (3.9%)	1 (1.3%)	0 (0%)	4 (4.1%)	1 (0.5%)	1 (16.7%)	0 (0%)	20 (2.4%)
公園里	0 (0%)	0 (0%)	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	3 (3.1%)	1 (0.5%)	0 (0%)	0 (0%)	5 (0.6%)
成功里	1 (0.6%)	3 (5.7%)	0 (0%)	7 (3.4%)	0 (0%)	0 (0%)	3 (3.1%)	9 (4.6%)	0 (0%)	0 (0%)	23 (2.7%)
大港里	3 (1.9%)	0 (0%)	0 (0%)	23 (11.2%)	2 (2.5%)	2 (6.1%)	8 (8.2%)	5 (2.6%)	0 (0%)	0 (0%)	43 (5.1%)
大和里	2 (1.3%)	0 (0%)	0 (0%)	5 (2.4%)	1 (1.3%)	0 (0%)	3 (3.1%)	7 (3.6%)	0 (0%)	0 (0%)	18 (2.1%)
大豐里	1 (0.6%)	0 (0%)	0 (0%)	9 (4.4%)	3 (3.8%)	1 (3.0%)	1 (1.0%)	4 (2.1%)	0 (0%)	0 (0%)	19 (2.3%)
賢北里	5 (3.2%)	2 (3.8%)	0 (0%)	2 (1.0%)	1 (1.3%)	0 (0%)	3 (3.1%)	5 (2.6%)	0 (0%)	0 (0%)	18 (2.1%)
開元里	0 (0%)	1 (1.9%)	0 (0%)	5 (2.4%)	0 (0%)	0 (0%)	3 (3.1%)	2 (1.0%)	0 (0%)	0 (0%)	11 (1.3%)
長勝里	0 (0%)	1 (1.9%)	0 (0%)	1 (0.5%)	2 (2.5%)	1 (3.0%)	2 (2.0%)	2 (1.0%)	0 (0%)	0 (0%)	9 (1.1%)
振興里	1 (0.6%)	0 (0%)	0 (0%)	2 (1.0%)	0 (0%)	0 (0%)	1 (1.0%)	4 (2.1%)	0 (0%)	0 (0%)	8 (0.9%)
合興里	1 (0.6%)	0 (0%)	0 (0%)	4 (2.0%)	0 (0%)	0 (0%)	3 (3.1%)	3 (1.5%)	0 (0%)	0 (0%)	11 (1.3%)
力行里	3 (1.9%)	1 (1.9%)	0 (0%)	1 (0.5%)	1 (1.3%)	2 (6.1%)	0 (0%)	2 (1.0%)	0 (0%)	0 (0%)	10 (1.2%)
重興里	1 (0.6%)	0 (0%)	0 (0%)	2 (1.0%)	0 (0%)	0 (0%)	0 (0%)	3 (1.5%)	1 (16.7%)	0 (0%)	7 (0.8%)
仁愛里	0 (0%)	0 (0%)	0 (0%)	2 (1.0%)	0 (0%)	1 (3.0%)	3 (3.1%)	2 (1.0%)	0 (0%)	0 (0%)	8 (0.9%)
北門里	3 (1.9%)	0 (0%)	0 (0%)	4 (2.0%)	1 (1.3%)	4 (12.1%)	6 (6.1%)	4 (2.1%)	0 (0%)	0 (0%)	22 (2.6%)
永祥里	0 (0%)	0 (0%)	0 (0%)	1 (0.5%)	0 (0%)	5 (15.2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (0.7%)
小北里	2 (1.3%)	1 (1.9%)	1 (11.1%)	5 (2.4%)	1 (1.3%)	1 (3.0%)	3 (3.1%)	9 (4.6%)	0 (0%)	0 (0%)	23 (2.7%)
大光里	0 (0%)	1 (1.9%)	0 (0%)	2 (1.0%)	3 (3.8%)	3 (9.1%)	9 (9.2%)	1 (0.5%)	0 (0%)	0 (0%)	19 (2.3%)
大興里	1 (0.6%)	0 (0%)	1 (11.1%)	6 (2.9%)	1 (1.3%)	0 (0%)	1 (1.0%)	5 (2.6%)	0 (0%)	1 (12.5%)	16 (1.9%)
長興里	0 (0%)	0 (0%)	0 (0%)	5 (2.4%)	0 (0%)	0 (0%)	1 (1.0%)	2 (1.0%)	0 (0%)	0 (0%)	8 (0.9%)
華德里	1 (0.6%)	1 (1.9%)	0 (0%)	7 (3.4%)	3 (3.8%)	0 (0%)	1 (1.0%)	8 (4.1%)	0 (0%)	0 (0%)	21 (2.5%)
福德里	1 (0.6%)	0 (0%)	0 (0%)	4 (2.0%)	1 (1.3%)	1 (3.0%)	1 (1.0%)	12 (6.2%)	0 (0%)	1 (12.5%)	21 (2.5%)
成德里	5 (3.2%)	2 (3.8%)	0 (0%)	8 (3.9%)	0 (0%)	0 (0%)	2 (2.0%)	12 (6.2%)	1 (16.7%)	1 (12.5%)	31 (3.7%)
和順里	7 (4.4%)	1 (1.9%)	0 (0%)	3 (1.5%)	2 (2.5%)	2 (6.1%)	4 (4.1%)	3 (1.5%)	0 (0%)	0 (0%)	22 (2.6%)
立人里	0 (0%)	1 (1.9%)	1 (11.1%)	2 (1.0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	5 (0.6%)

	(1號)	(10號)	(2號)	(3號)	(4號)	(5號)	(6號)	(7號)	(8號)	(9號)	Overall
雙安里	0 (0%)	1 (1.9%)	0 (0%)	1 (0.5%)	2 (2.5%)	0 (0%)	1 (1.0%)	4 (2.1%)	0 (0%)	0 (0%)	9 (1.1%)
文成里	1 (0.6%)	1 (1.9%)	0 (0%)	7 (3.4%)	1 (1.3%)	3 (9.1%)	2 (2.0%)	7 (3.6%)	0 (0%)	1 (12.5%)	23 (2.7%)
元美里	2 (1.3%)	2 (3.8%)	1 (11.1%)	11 (5.4%)	1 (1.3%)	1 (3.0%)	0 (0%)	5 (2.6%)	0 (0%)	0 (0%)	23 (2.7%)
文元里	0 (0%)	3 (5.7%)	1 (11.1%)	8 (3.9%)	1 (1.3%)	0 (0%)	3 (3.1%)	4 (2.1%)	0 (0%)	0 (0%)	20 (2.4%)
不知道	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
拒答	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
遺漏值或跳答	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
郡王里	1 (0.6%)	1 (1.9%)	0 (0%)	1 (0.5%)	3 (3.8%)	0 (0%)	1 (1.0%)	2 (1.0%)	0 (0%)	0 (0%)	9 (1.1%)
赤嵌里	7 (4.4%)	1 (1.9%)	0 (0%)	1 (0.5%)	1 (1.3%)	0 (0%)	3 (3.1%)	3 (1.5%)	0 (0%)	0 (0%)	16 (1.9%)
法華里	3 (1.9%)	1 (1.9%)	0 (0%)	1 (0.5%)	3 (3.8%)	0 (0%)	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	9 (1.1%)
西湖里	10 (6.3%)	0 (0%)	1 (11.1%)	5 (2.4%)	4 (5.1%)	1 (3.0%)	1 (1.0%)	9 (4.6%)	0 (0%)	1 (12.5%)	32 (3.8%)
西賢里	4 (2.5%)	3 (5.7%)	1 (11.1%)	4 (2.0%)	3 (3.8%)	0 (0%)	2 (2.0%)	2 (1.0%)	1 (16.7%)	0 (0%)	20 (2.4%)
城隍里	4 (2.5%)	2 (3.8%)	0 (0%)	0 (0%)	2 (2.5%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8 (0.9%)
開山里	3 (1.9%)	0 (0%)	0 (0%)	2 (1.0%)	1 (1.3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (0.7%)
南美里	6 (3.8%)	1 (1.9%)	0 (0%)	0 (0%)	2 (2.5%)	0 (0%)	1 (1.0%)	2 (1.0%)	0 (0%)	0 (0%)	12 (1.4%)
永華里	1 (0.6%)	0 (0%)	0 (0%)	0 (0%)	1 (1.3%)	0 (0%)	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	3 (0.4%)
小西門里	2 (1.3%)	0 (0%)	0 (0%)	1 (0.5%)	1 (1.3%)	0 (0%)	2 (2.0%)	1 (0.5%)	0 (0%)	0 (0%)	7 (0.8%)
五條港里	0 (0%)	0 (0%)	0 (0%)	2 (1.0%)	3 (3.8%)	0 (0%)	0 (0%)	2 (1.0%)	0 (0%)	0 (0%)	7 (0.8%)
兌悅里	3 (1.9%)	2 (3.8%)	0 (0%)	2 (1.0%)	4 (5.1%)	0 (0%)	1 (1.0%)	5 (2.6%)	0 (0%)	0 (0%)	17 (2.0%)
藥王里	11 (7.0%)	3 (5.7%)	0 (0%)	1 (0.5%)	3 (3.8%)	0 (0%)	0 (0%)	6 (3.1%)	0 (0%)	0 (0%)	24 (2.8%)
淺草里	2 (1.3%)	0 (0%)	0 (0%)	1 (0.5%)	1 (1.3%)	0 (0%)	0 (0%)	1 (0.5%)	1 (16.7%)	0 (0%)	6 (0.7%)
府前里	2 (1.3%)	0 (0%)	0 (0%)	0 (0%)	1 (1.3%)	0 (0%)	1 (1.0%)	0 (0%)	0 (0%)	0 (0%)	4 (0.5%)
南廠里	6 (3.8%)	1 (1.9%)	0 (0%)	1 (0.5%)	2 (2.5%)	0 (0%)	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	11 (1.3%)
西和里	7 (4.4%)	2 (3.8%)	0 (0%)	8 (3.9%)	3 (3.8%)	0 (0%)	1 (1.0%)	4 (2.1%)	0 (0%)	0 (0%)	25 (3.0%)
大涼里	6 (3.8%)	2 (3.8%)	0 (0%)	1 (0.5%)	1 (1.3%)	0 (0%)	0 (0%)	3 (1.5%)	0 (0%)	0 (0%)	13 (1.5%)

	(1號)	(10號)	(2號)	(3號)	(4號)	(5號)	(6號)	(7號)	(8號)	(9號)	Overall
光賢里	6 (3.8%)	0 (0%)	0 (0%)	5 (2.4%)	3 (3.8%)	0 (0%)	1 (1.0%)	3 (1.5%)	0 (0%)	2 (25.0%)	20 (2.4%)
南門里	4 (2.5%)	1 (1.9%)	0 (0%)	0 (0%)	2 (2.5%)	0 (0%)	0 (0%)	2 (1.0%)	0 (0%)	0 (0%)	9 (1.1%)
Missing	21 (13.3%)	6 (11.3%)	2 (22.2%)	19 (9.3%)	6 (7.6%)	5 (15.2%)	10 (10.2%)	16 (8.2%)	1 (16.7%)	1 (12.5%)	87 (10.3%)
v4											
Mean (SD)	1.66 (1.35)	2.15 (1.54)	1.89 (1.17)	1.86 (1.34)	1.95 (1.35)	2.52 (1.99)	2.09 (1.62)	1.90 (1.34)	1.83 (2.04)	1.50 (0.756)	1.91 (1.43)
Median [Min, Max]	1.00 [1.00, 8.00]	1.00 [1.00, 7.00]	1.00 [1.00, 4.00]	1.00 [1.00, 7.00]	1.00 [1.00, 6.00]	2.00 [1.00, 8.00]	1.00 [1.00, 8.00]	1.00 [1.00, 7.00]	1.00 [1.00, 6.00]	1.00 [1.00, 3.00]	1.00 [1.00, 8.00]
v6											
20-29歲	3 (1.9%)	1 (1.9%)	1 (11.1%)	5 (2.4%)	3 (3.8%)	0 (0%)	1 (1.0%)	11 (5.6%)	1 (16.7%)	0 (0%)	26 (3.1%)
30-39歲	8 (5.1%)	4 (7.5%)	2 (22.2%)	12 (5.9%)	2 (2.5%)	3 (9.1%)	6 (6.1%)	16 (8.2%)	2 (33.3%)	1 (12.5%)	56 (6.6%)
40-49歲	11 (7.0%)	3 (5.7%)	3 (33.3%)	31 (15.1%)	11 (13.9%)	8 (24.2%)	13 (13.3%)	39 (20.0%)	0 (0%)	2 (25.0%)	121 (14.3%)
50-59歲	40 (25.3%)	7 (13.2%)	3 (33.3%)	52 (25.4%)	19 (24.1%)	10 (30.3%)	22 (22.4%)	43 (22.1%)	1 (16.7%)	2 (25.0%)	199 (23.6%)
60歲及以上	95 (60.1%)	38 (71.7%)	0 (0%)	105 (51.2%)	44 (55.7%)	12 (36.4%)	56 (57.1%)	85 (43.6%)	2 (33.3%)	3 (37.5%)	440 (52.1%)
不知道/拒答	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
遺漏值或跳答	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Missing	1 (0.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	2 (0.2%)
v7											
小學或以下	31 (19.6%)	4 (7.5%)	0 (0%)	27 (13.2%)	5 (6.3%)	2 (6.1%)	12 (12.2%)	18 (9.2%)	1 (16.7%)	0 (0%)	100 (11.8%)
初中、國中	18 (11.4%)	7 (13.2%)	0 (0%)	26 (12.7%)	6 (7.6%)	3 (9.1%)	5 (5.1%)	17 (8.7%)	0 (0%)	0 (0%)	82 (9.7%)
高中、高職	45 (28.5%)	16 (30.2%)	0 (0%)	64 (31.2%)	25 (31.6%)	13 (39.4%)	32 (32.7%)	47 (24.1%)	2 (33.3%)	4 (50.0%)	248 (29.4%)
專科	19 (12.0%)	6 (11.3%)	1 (11.1%)	31 (15.1%)	10 (12.7%)	6 (18.2%)	16 (16.3%)	20 (10.3%)	0 (0%)	0 (0%)	109 (12.9%)
大學以上	42 (26.6%)	20 (37.7%)	8 (88.9%)	57 (27.8%)	31 (39.2%)	9 (27.3%)	33 (33.7%)	90 (46.2%)	3 (50.0%)	4 (50.0%)	297 (35.2%)
拒答	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
遺漏值或跳答	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Missing	3 (1.9%)	0 (0%)	0 (0%)	0 (0%)	2 (2.5%)	0 (0%)	0 (0%)	3 (1.5%)	0 (0%)	0 (0%)	8 (0.9%)
v8											
男性	67 (42.4%)	27 (50.9%)	6 (66.7%)	88 (42.9%)	38 (48.1%)	17 (51.5%)	50 (51.0%)	79 (40.5%)	5 (83.3%)	2 (25.0%)	379 (44.9%)
女性	91 (57.6%)	26 (49.1%)	3 (33.3%)	117 (57.1%)	41 (51.9%)	16 (48.5%)	48 (49.0%)	116 (59.5%)	1 (16.7%)	6 (75.0%)	465 (55.1%)

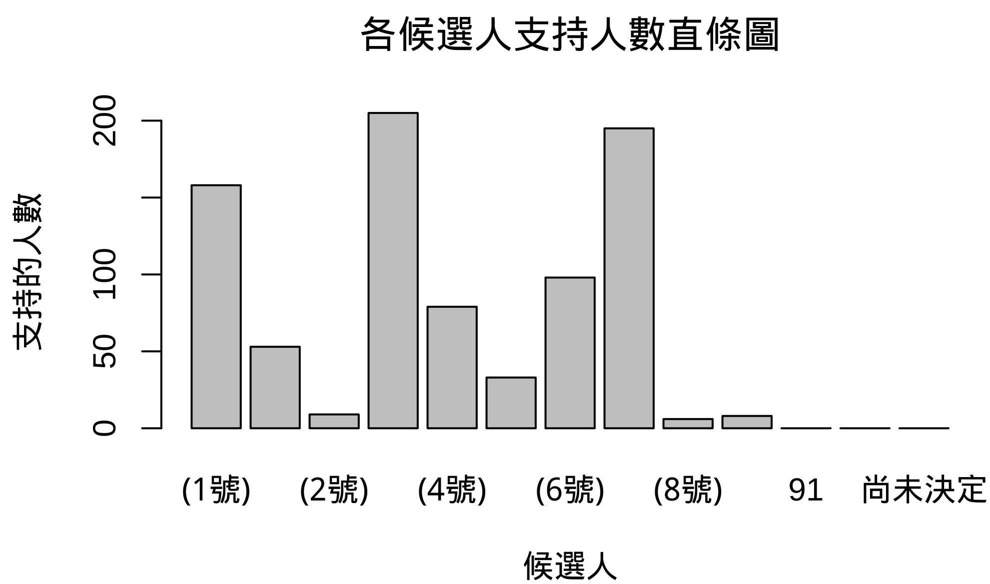
	(1號)	(10號)	(2號)	(3號)	(4號)	(5號)	(6號)	(7號)	(8號)	(9號)	Overall
遺漏值或跳答	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

以直條圖表示支持人數與支持率。

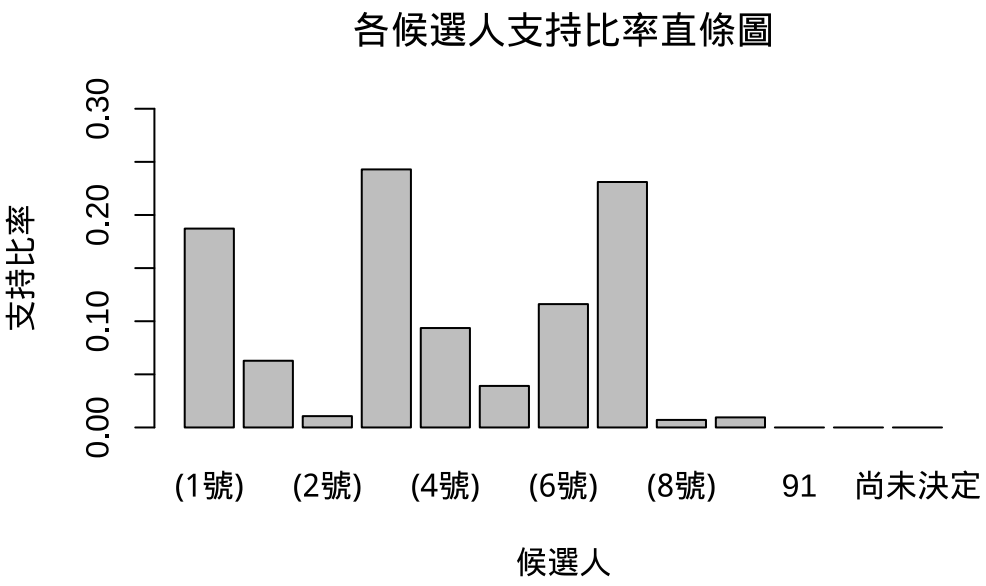
```
data5 <- data_new
data5[is.na(data5$v5),] <- " "
```

```
Warning in `[<-.factor`(`*tmp*`, iseq, value = c(" ", " ", " ", :
invalid factor level, NA generated
Warning in `[<-.factor`(`*tmp*`, iseq, value = c(" ", " ", " ", :
invalid factor level, NA generated
Warning in `[<-.factor`(`*tmp*`, iseq, value = c(" ", " ", " ", :
invalid factor level, NA generated
Warning in `[<-.factor`(`*tmp*`, iseq, value = c(" ", " ", " ", :
invalid factor level, NA generated
Warning in `[<-.factor`(`*tmp*`, iseq, value = c(" ", " ", " ", :
invalid factor level, NA generated
Warning in `[<-.factor`(`*tmp*`, iseq, value = c(" ", " ", " ", :
invalid factor level, NA generated
```

```
barplot(table(data5$v5),xlab="  ",ylab="   ",main = "       ")
```




```
  #
num <- length(data_nom$v5)
barplot(table(data_nom$v5)/num,ylim=c(0,0.3),xlab="  ",ylab="  ",
        main="      ")
```



第三題

```
library(showtext)

# Enable showtext to use system fonts
showtext_auto()
# install.packages("foreign")
library(foreign, help, pos = 2, lib.loc = NULL)
data1 <- read.spss("poll.sav")
```

```
re-encoding from CP950
```

```
Warning in read.spss("poll.sav"): Undeclared level(s) 91 added in variable:
v4_1
```

```
Warning in read.spss("poll.sav"): Undeclared level(s) 91 added in variable: v5
```

```
levels(data1$v4_1)
```

```
 [1] "(1 )"        "(10 )"       "(2 )"        "(3 )"        "(4 )"
 [6] "(5 )"        "(6 )"        "(7 )"        "(8 )"        "(9 )"
[11] "91"           "  "       "   "
```

```
class(data1$v4_1)
```

```
[1] "factor"
```

```
levels(data1$v5)
```

```
 [1] "(1 )"    "(10 )"   "(2 )"    "(3 )"    "(4 )"    "(5 )"
 [7] "(6 )"    "(7 )"    "(8 )"    "(9 )"    "91"       "  "
[13] "   "
```

```
data1 <- as.data.frame(data1)
class(data1$v4_1)
```

```
[1] "factor"
```

清洗資料

了解每個variables中有那些結果，並把無確切答案的設成NA

```
levels(data1$v1)
```

```
[1] " "
[2] "  "
[3] " /      /     "
[4] "    "
```

```
levels(data1$v2)
```

```
 [1] "  "        "  "        "  "        "  "        "  "
 [6] "  "        "  "        "  "        "  "        "  "
[11] "  "        "  "        "  "        "  "        "  "
[16] "  "        "  "        "  "        "  "        "  "
[21] "  "        "  "        "  "        "  "        "  "
[26] "  "        "  "        "  "        "  "        "  "
[31] "  "        "  "        "  "        "  "        " "
[36] "    "
```

```
levels(data1$v3)
```

```
 [1] "  "        "  "        "  "        "  "        "  "
 [6] "  "        "  "        "  "        "  "        "   "
[11] "   "      "  "        "  "        "  "        "  "
[16] "  "        "  "        "  "        "  "        "  "
[21] "  "        " "          "    "
```

```
levels(data1$v6)
```

```
[1] "20-29 "       "30-39 "       "40-49 "       "50-59 "       "60   "
[6] "    "   "    "
```

```
levels(data1$v7)
```

```
[1] "   "   "   "   "   "   "   " "          "   "
[6] " "          "    "
```

```
levels(data1$v8)
```

```
[1] " "          " "          "    "
```

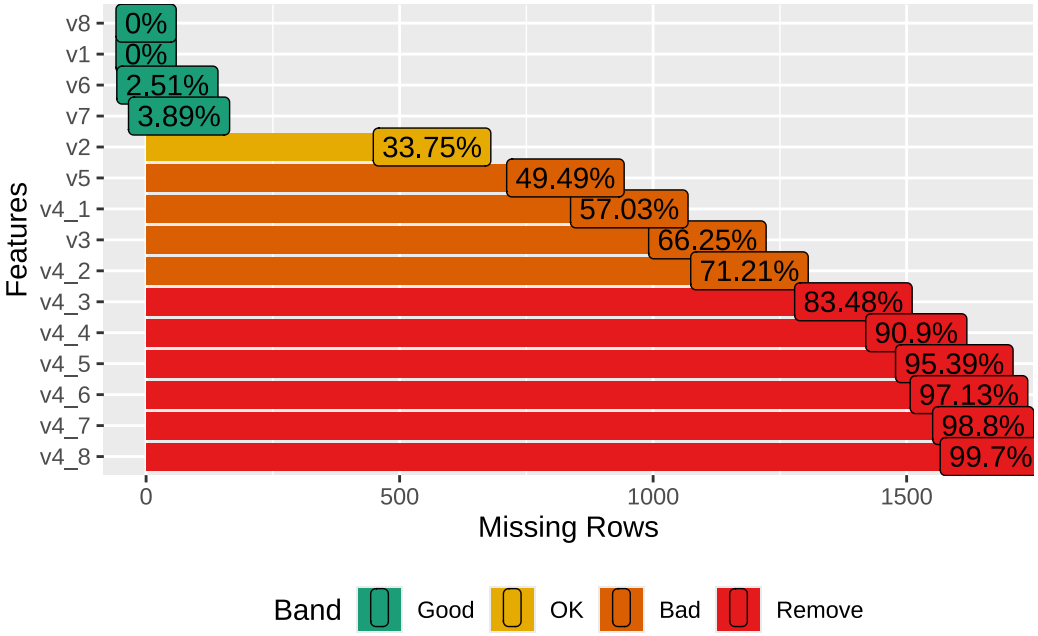
```
data1$v1 <- replace(data1$v1, data1$v1 %in% c(" /      /     ", "    "), NA)
data1$v2 <- replace(data1$v2, data1$v2 %in% c("    "), NA)
data1$v3 <- replace(data1$v3, data1$v3 %in% c("    "), NA)

#       "  ", "91", "    "
data1$v4_1 <- replace(data1$v4_1, data1$v4_1 %in% c("  ", "91", "    "), NA)
data1$v4_2 <- replace(data1$v4_2, data1$v4_2 %in% c("  ", "91", "    "), NA)
data1$v4_3 <- replace(data1$v4_3, data1$v4_3 %in% c("  ", "91", "    "), NA)
data1$v4_4 <- replace(data1$v4_4, data1$v4_4 %in% c("  ", "91", "    "), NA)
data1$v4_5 <- replace(data1$v4_5, data1$v4_5 %in% c("  ", "91", "    "), NA)
data1$v4_6 <- replace(data1$v4_6, data1$v4_6 %in% c("  ", "91", "    "), NA)
data1$v4_7 <- replace(data1$v4_7, data1$v4_7 %in% c("  ", "91", "    "), NA)
data1$v4_8 <- replace(data1$v4_8, data1$v4_8 %in% c("  ", "91", "    "), NA)

#       "  ", "91", "   "
```

```
data1$v5 <- replace(data1$v5, data1$v5 %in% c("  ", "91", "  "), NA)
head(data1$v4_1)
```

```
[1] <NA>  (1 ) (4 ) (3 ) <NA>  <NA>
13 Levels: (1 ) (10 ) (2 ) (3 ) (4 ) (5 ) (6 ) (7 ) (8 ) (9 ) ...
```

```
data1$v6 <- replace(data1$v6, data1$v6 %in% c("    ","    "), NA)
data1$v7 <- replace(data1$v7, data1$v7 %in% c(" ","    "), NA)
data1$v8 <- replace(data1$v8, data1$v8 %in% c("    "), NA)
```

```
######## missing data
# install.packages("DataExplorer")

library(Hmisc)
library(DataExplorer)
plot_missing(data1)
```




```
v5 <- as.character(data1$v5)
describe(v5)
```

```
v5
       n  missing distinct
     844      827       10

Value      (10 )  (1 )  (2 )  (3 )  (4 )  (5 )  (6 )  (7 )  (8 )
Frequency      53    158      9    205     79     33     98    195      6
Proportion 0.063  0.187  0.011  0.243  0.094  0.039  0.116  0.231  0.007

Value       (9 )
Frequency      8
Proportion 0.009
```

```
#   v5
#    v5
describe(data1$v5)
```

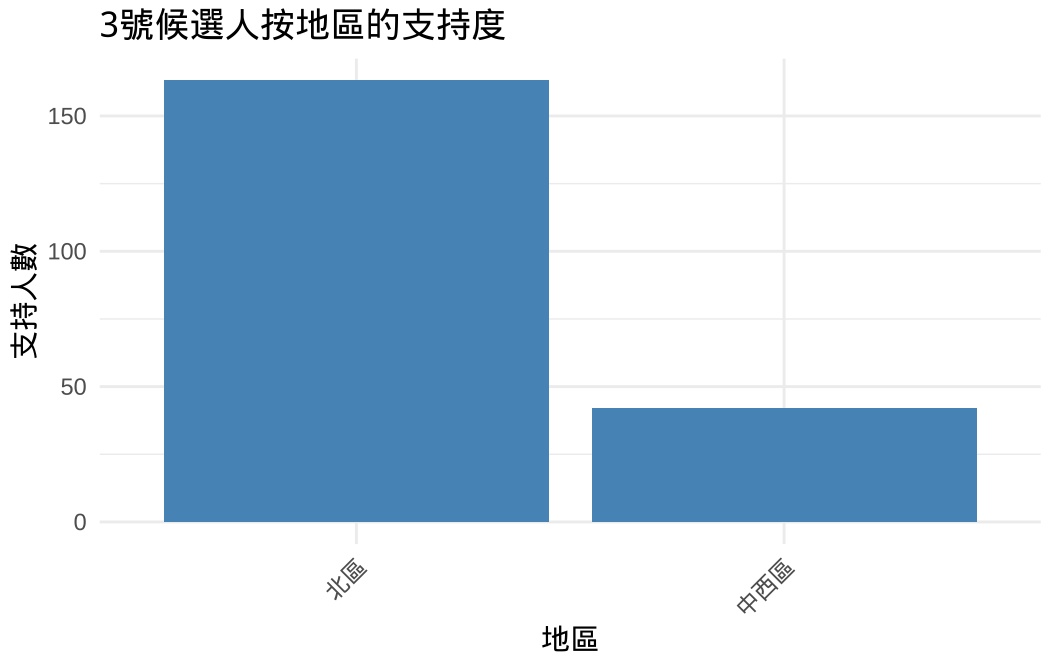
```
data1$v5
       n  missing distinct
     844      827       10

Value       (1 ) (10 )  (2 )  (3 )  (4 )  (5 )  (6 )  (7 )  (8 )
Frequency    158     53      9    205     79     33     98    195      6
Proportion 0.187  0.063  0.011  0.243  0.094  0.039  0.116  0.231  0.007

Value       (9 )
Frequency      8
Proportion 0.009
```

從上圖可了解除了v4以外，其餘缺失值並不多。

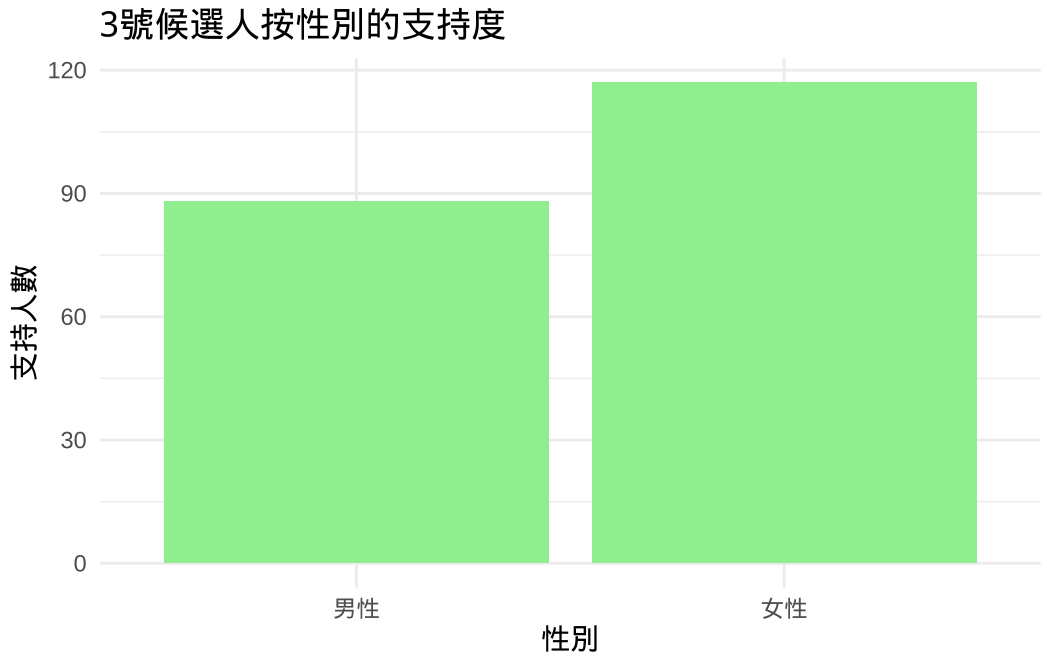
支持3號的資料

接著我們用支持3號的資料和不支持3號的資料做卡方檢定，看在地區、性別等等變數上面是否有顯著關聯。

```
library(dplyr)
library(knitr)
library(ggplot2)
# Filter the data for candidate 3
candidate_3 <- data1 %>% filter(v5 == "(3 )")

# Analyze support by region
support_by_region <- candidate_3 %>% count(v1)

# Analyze support by age group
support_by_age <- candidate_3 %>% count(v6)

# Analyze support by gender
support_by_gender <- candidate_3 %>% count(v8)

# Print the results
#   ggplot2
ggplot(support_by_region, aes(x = v1, y = n)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "3      ",
       x = " ",
       y = "   ") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  #       X
```

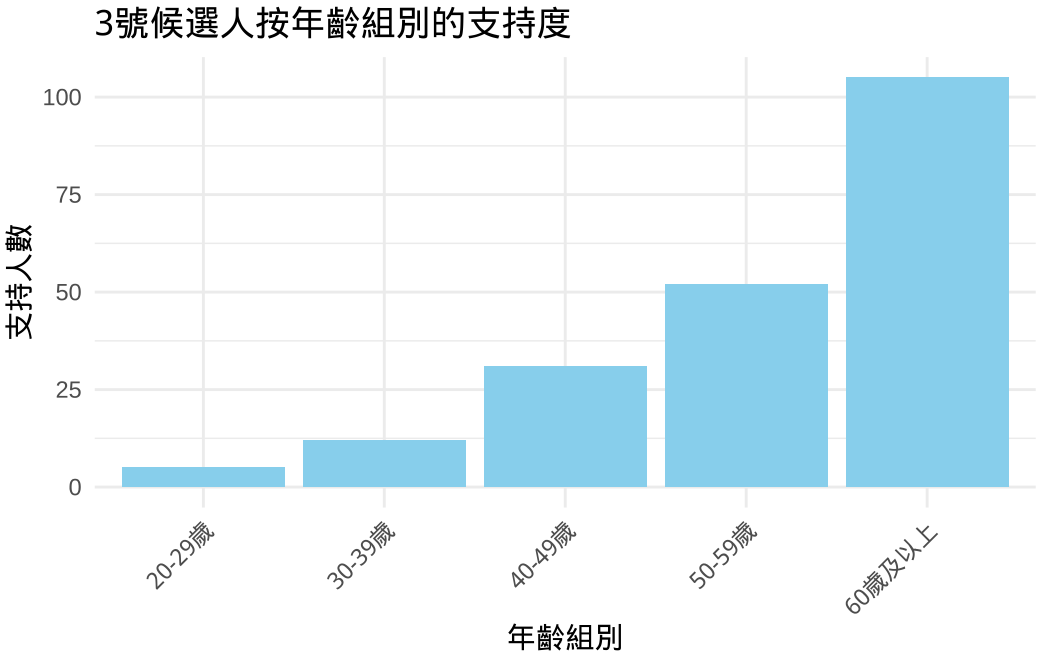



```
#   ggplot2
ggplot(support_by_gender, aes(x = v8, y = n)) +
  geom_bar(stat = "identity", fill = "lightgreen") +
  labs(title = "3      ",
       x = " ",
       y = "   ") +
  theme_minimal()
```



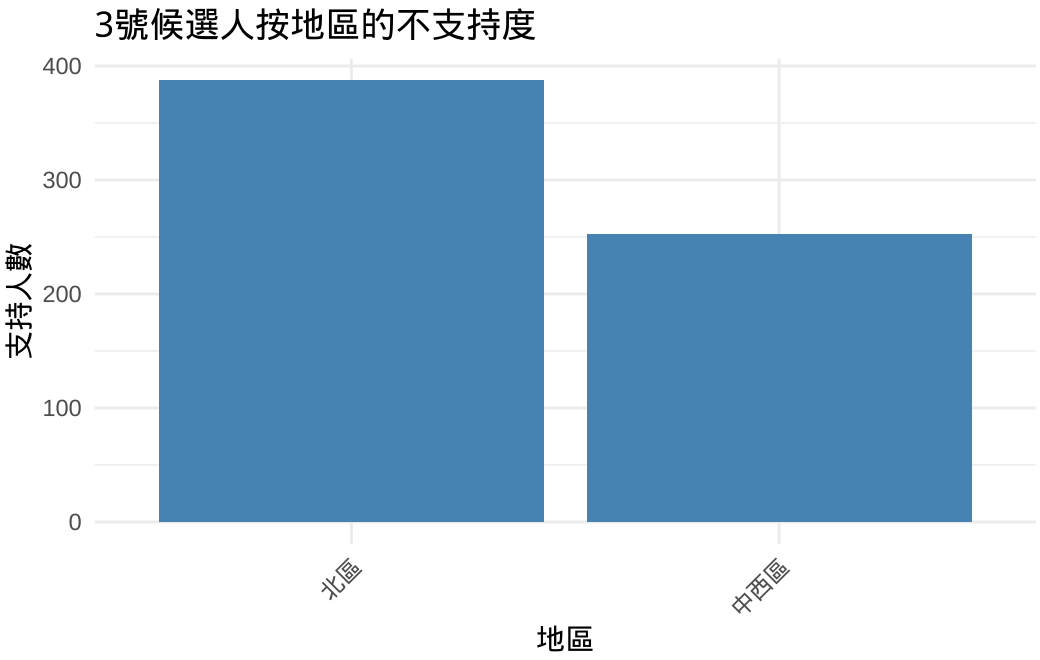

```
#   ggplot2
ggplot(support_by_age, aes(x = v6, y = n)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "3       ",
```

```
       x = "   ",
       y = "   ") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

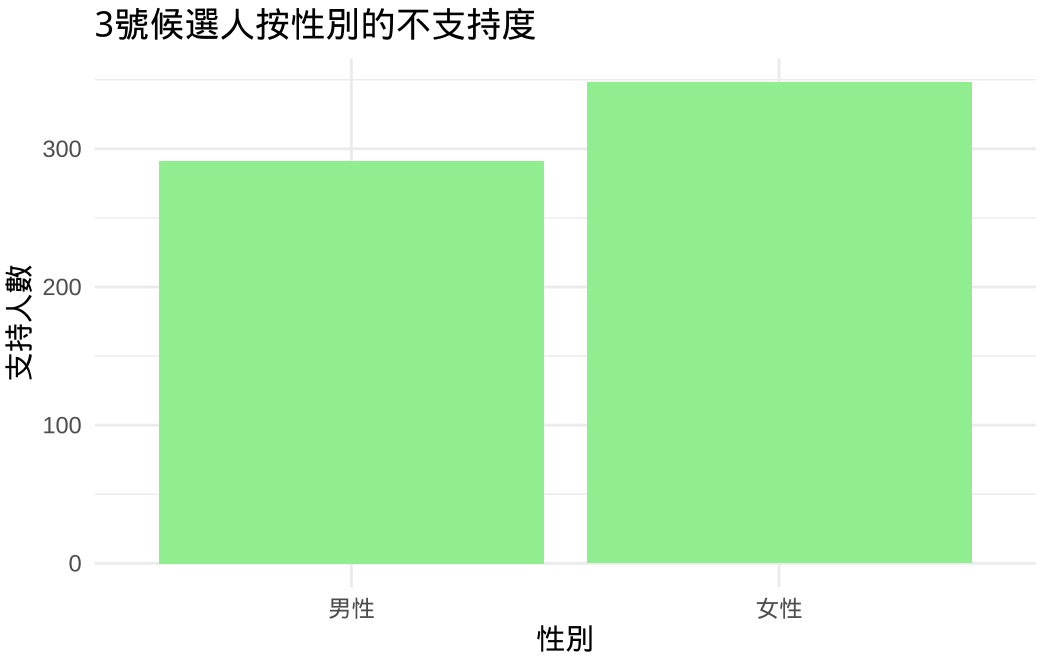


不支持3號的資料

```
Notcandidate_3 <- data1 %>% filter(v5 != "(3 )")

# Analyze support by region
Nsupport_by_region <- Notcandidate_3 %>% count(v1)

# Analyze support by age group
Nsupport_by_age <- Notcandidate_3 %>% count(v6)

# Analyze support by gender
Nsupport_by_gender <- Notcandidate_3 %>% count(v8)

# Print the results
#   ggplot2
ggplot(Nsupport_by_region, aes(x = v1, y = n)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "3       ",
       x = " ",
       y = "   ") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  #        X
```

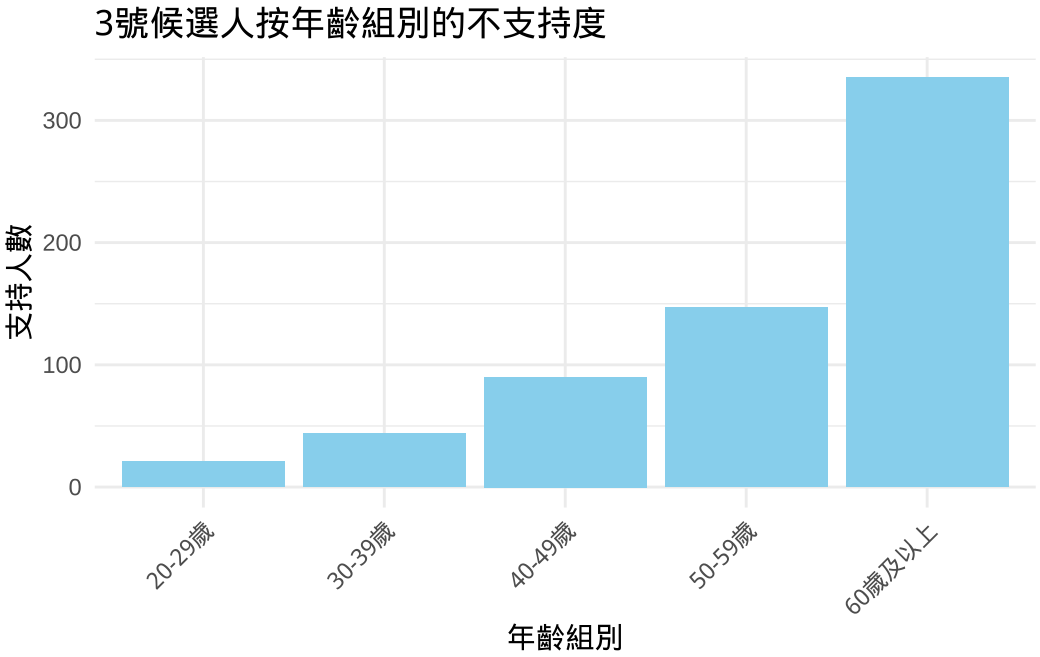



```
#   ggplot2
ggplot(Nsupport_by_gender, aes(x = v8, y = n)) +
  geom_bar(stat = "identity", fill = "lightgreen") +
  labs(title = "3       ",
       x = " ",
       y = "   ") +
  theme_minimal()
```




```
#   ggplot2
ggplot(Nsupport_by_age[-6,], aes(x = v6, y = n)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "3       ",
```

```
      x = "   ",
      y = "   ") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



卡方檢定

地區的卡方檢定:

```
#
#  3
support_region_table <- table(candidate_3$v1)
Nsupport_region_table <- table(Notcandidate_3$v1)

#
region_table <- rbind(support_region_table, Nsupport_region_table)

#  NA
region_table <- region_table[,c(-3,-4)]
#
chi_square_region <- chisq.test(region_table)
print(chi_square_region)
```

```
    Pearson's Chi-squared test with Yates' continuity correction

data:  region_table
X-squared = 23.722, df = 1, p-value = 1.113e-06
```

```
print(chi_square_region$observed)
```

```
support_region_table   163     42
Nsupport_region_table  387    252
```

```
print(chi_square_region$expected)
```

```
support_region_table  133.59  71.40995
Nsupport_region_table 416.41 222.59005
```

年齡的卡方檢定:

```
#  3
support_age_table <- table(candidate_3$v6)
Nsupport_age_table <- table(Notcandidate_3$v6)

#
age_table <- rbind(support_age_table, Nsupport_age_table)
age_table <- age_table[,c(-6,-7)]
#
chi_square_age <- chisq.test(age_table)
print(chi_square_age)
```

```
    Pearson's Chi-squared test

data:  age_table
X-squared = 1.1344, df = 4, p-value = 0.8888
```

```
print(chi_square_age$observed)
```

```
                  20-29  30-39  40-49  50-59  60
support_age_table        5      12      31      52        105
Nsupport_age_table      21      44      90     147        335
```

```
print(chi_square_age$expected)
```

```
                     20-29  30-39   40-49    50-59  60
support_age_table   6.330166 13.6342 29.45962  48.45012   107.1259
Nsupport_age_table 19.669834 42.3658 91.54038 150.54988   332.8741
```

性別的卡方檢定:

```
#  3
support_gender_table <- table(candidate_3$v8)
Nsupport_gender_table <- table(Notcandidate_3$v8)

#
gender_table <- rbind(support_gender_table, Nsupport_gender_table)
gender_table <- gender_table[,-3]
#
chi_square_gender <- chisq.test(gender_table)
print(chi_square_gender)
```

```
    Pearson's Chi-squared test with Yates' continuity correction

data:  gender_table
X-squared = 0.32925, df = 1, p-value = 0.5661
```

```
print(chi_square_gender$observed)
```

```
support_gender_table    88  117
Nsupport_gender_table  291  348
```

```
print(chi_square_gender$expected)
```

```
support_gender_table   92.05569 112.9443
Nsupport_gender_table 286.94431 352.0557
```

Result

根據卡方檢定的結果，我們可以為3號候選人的競選策略提出以下建議：

1. 地區差異

- **地區：**
  - 卡方檢定結果顯示，3號候選人在不同地區的支持度存在顯著差異（p-value = 1.113e-06，顯著性水平0.05以下）。具體而言，3號候選人在北區（163人）相較於中西區（42人）擁有更高的支持度。
  - **結論：3號候選人應該重點在中西區加強拉票，因為這個區域的支持度相對較低，存在較大的增長潛力。**

2. 年齡差異

- **年齡組別：**
  - 卡方檢定結果顯示，3號候選人支持者的年齡分佈與非支持者的年齡分佈沒有顯著差異（p-value = 0.8888）。這表明3號候選人的支持在不同年齡段較為平均。
  - **結論：年齡層次的拉票策略可以相對平均，並不需要針對特定年齡群體。但如果要進一步加強，可以考慮針對50歲以上群體（105人支持），因為這個群體是目前支持度最高的。**

3. 性別差異

- **性別：**
  - 性別的卡方檢定結果顯示，支持3號候選人的性別分佈與非支持者的性別分佈無顯著差異（p-value = 0.5661）。這意味著性別並不是影響3號候選人支持的關鍵因素。
  - **結論：拉票策略不需要特別針對性別差異，可以在男性和女性之間均勻展開競選活動。**

總結建議：

- **加強在中西區的競選活動**：由於3號候選人在中西區的支持率較低，應該更多地在該地區進行宣傳和動員選民。
- **維持對50歲及以上群體的關注**：雖然年齡分佈沒有顯著差異，但50歲以上群體對3號候選人的支持度較高，因此可以重點鞏固這部分選民。
- **平衡男性和女性的拉票活動**：性別不是關鍵因素，因此可以均衡地向男性和女性選民傳遞訊息。

這樣的策略能夠幫助3號候選人最大化競選資源的效果，提升支持率。

地圖

使用地圖進行競選策略規劃。

```
library(sf)
```

```
Linking to GEOS 3.12.1, GDAL 3.8.4, PROJ 9.3.1; sf_use_s2() is TRUE
```

```
#
map <- st_read("VILLAGE_NLSC_1130807.shp")
```

```
Reading layer `VILLAGE_NLSC_1130807' from data source
  `C:\Users\user\Desktop\Textbook\4-1\stat_consult\hw4\hw4\hw4\VILLAGE_NLSC_1130807.shp'
  using driver `ESRI Shapefile'
Simple feature collection with 7953 features and 11 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: 114.3593 ymin: 10.37135 xmax: 124.5612 ymax: 26.38528
Geodetic CRS:  GCS_TWD97[2020]
```

```
map1 <- map[map$COUNTYNAME=="  ",]
map1 <- map1[map1$TOWNNAME==" "|map1$TOWNNAME=="  ",]
```

```
#       dataframe
#
candidate_3_data <- data[data$v5=="(3 )",]
num <- data.frame(table(c(candidate_3_data$v2,candidate_3_data$v3)))
num <- num[c(-33,-48,-10),]
colnames(num) <- c("VILLNAME", "Freq")
map1 <- merge(map1,num,by="VILLNAME",all=T)
map1[is.na(map1)] <- 0
```

```
#
library(ggplot2)
describe(map1$Freq)
```

```
map1$Freq
       n  missing distinct     Info     Mean      Gmd      .05      .10
      56        0       12    0.969    3.321    3.849     0.00     0.00
     .25      .50      .75      .90      .95
    1.00     2.00     5.00     7.50     8.75

Value          0     1     2     3     4     5     6     7     8    11    16
Frequency     10    15     9     2     4     6     1     3     3     1     1
Proportion 0.179 0.268 0.161 0.036 0.071 0.107 0.018 0.054 0.054 0.018 0.018

Value         23
Frequency      1
Proportion 0.018

For the frequency table, variable is rounded to the nearest 0
```

```
map_N <- map1[map1$TOWNNAME==" ",]
map_MW <- map1[map1$TOWNNAME=="  ",]
```

由兩區地圖可以看出，三號候選人在西北方有較高的支持度，尤其是在大港里。在東南部分支持者較少。

```
#
coord <- as.data.frame(st_centroid(map1))
```

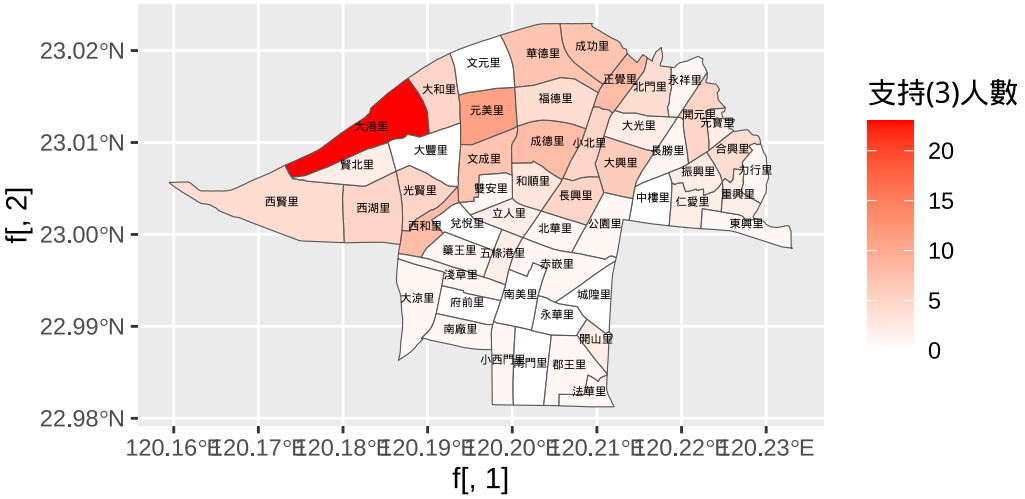
```
Warning: st_centroid assumes attributes are constant over geometries
```

```
f <- st_coordinates(st_centroid(map1))
```

```
Warning: st_centroid assumes attributes are constant over geometries
```

```
ggplot(data = map1) +
  geom_sf(aes(fill = Freq)) +
  scale_fill_gradient(low="white",high="red",name = " (3) ",limits=c(0,23))+
  geom_text(data = coord, aes(x = f[,1], y = f[,2], label = coord$VILLNAME),
          size = 1.5, color = "black", vjust =0)
```

```
Warning: Removed 3 rows containing missing values or values outside the scale range
(`geom_text()`).
```



從單一北區地圖可以看到北區有許多人支持三號候選人。

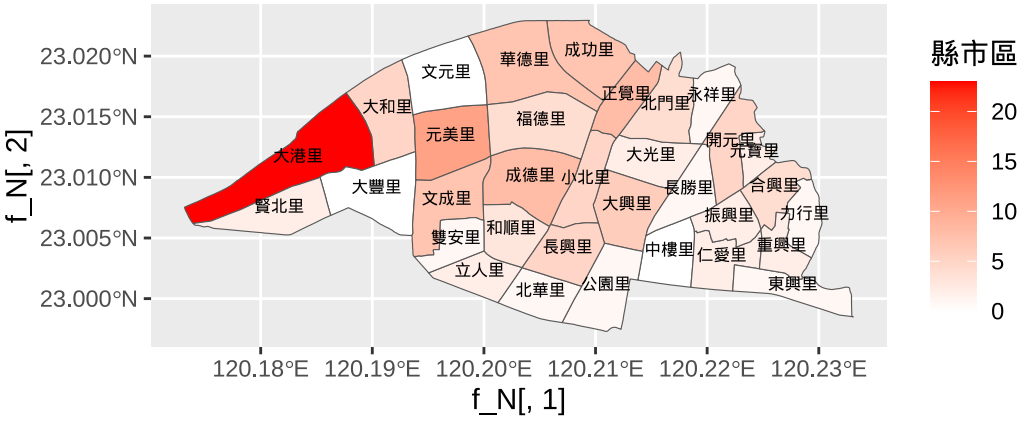
```
#
coord_N <- as.data.frame(st_centroid(map_N))
```

```
Warning: st_centroid assumes attributes are constant over geometries
```

```
f_N <- st_coordinates(st_centroid(map_N))
```

```
Warning: st_centroid assumes attributes are constant over geometries
```

```
ggplot(data = map_N) +
    geom_sf(aes(fill = Freq)) +
    scale_fill_gradient(low="white",high="red",name = "  ",limits=c(0,23))+
    geom_text(data = coord_N, aes(x = f_N[,1], y = f_N[,2], label = coord_N$VILLNAME),
          size = 2.2, color = "black", vjust =0)
```



從中西區地圖來看，在中西區支持三號候選人的人較少，如果想要增加影響力可以考慮往中西區進行宣傳。

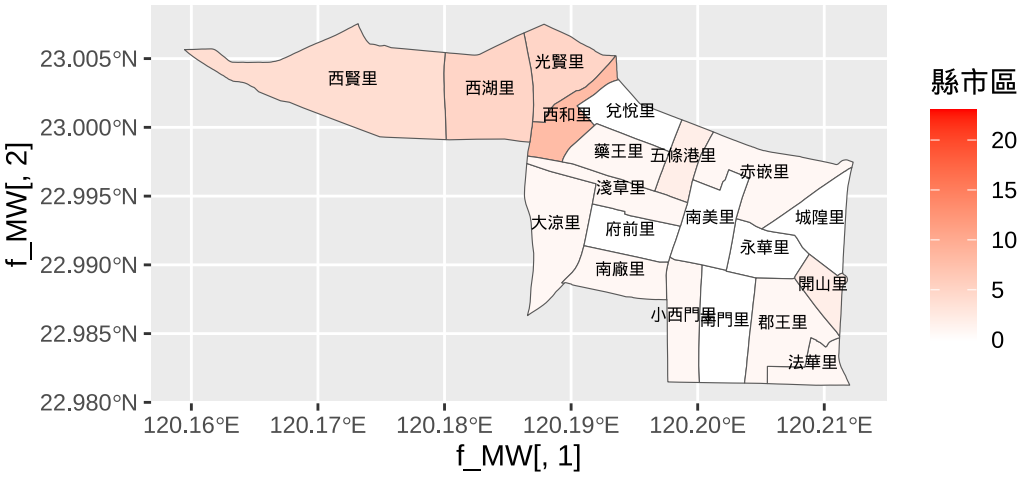
```
#
coord_MW <- as.data.frame(st_centroid(map_MW))
```

```
Warning: st_centroid assumes attributes are constant over geometries
```

```
f_MW <- st_coordinates(st_centroid(map_MW))
```

```
Warning: st_centroid assumes attributes are constant over geometries
```

```
ggplot(data = map_MW) +
  geom_sf(aes(fill = Freq)) +
  scale_fill_gradient(low="white",high="red",name = "  ",limits=c(0,23))+
  geom_text(data = coord_MW, aes(x = f_MW[,1], y = f_MW[,2], label = coord_MW$VILLNAME),
          size = 2.2, color = "black", vjust =0)
```



第四題

處理v4資料

```
v4 <- data1[,c(4,5,6,7,8,9,10,11)]
head(v4)
```

```
   v4_1   v4_2  v4_3   v4_4 v4_5 v4_6 v4_7 v4_8
1  <NA>   <NA>  <NA>   <NA> <NA> <NA> <NA> <NA>
2 (1 )  (4 ) (7 ) (10 ) <NA> <NA> <NA> <NA>
3 (4 )   <NA>  <NA>   <NA> <NA> <NA> <NA> <NA>
4 (3 ) (10 )  <NA>   <NA> <NA> <NA> <NA> <NA>
5  <NA>   <NA>  <NA>   <NA> <NA> <NA> <NA> <NA>
6  <NA>   <NA>  <NA>   <NA> <NA> <NA> <NA> <NA>
```

```
#     NA    0   NA    1
political_engagement <- v4 %>%
  mutate_at(vars(v4_1:v4_8), ~ifelse(is.na(.), 0, 1))

#
political_engagement$engagement_score <- rowSums(political_engagement)

#
head(political_engagement)
```

```
  v4_1 v4_2 v4_3 v4_4 v4_5 v4_6 v4_7 v4_8 engagement_score
1    0    0    0    0    0    0    0    0                0
2    1    1    1    1    0    0    0    0                4
3    1    0    0    0    0    0    0    0                1
4    1    1    0    0    0    0    0    0                2
5    0    0    0    0    0    0    0    0                0
6    0    0    0    0    0    0    0    0                0
```

```
data2 <- cbind(data1,political_engagement$engagement_score)
data2 <- data2[,c(-4,-5,-6,-7,-8,-9,-10,-11)]
head(data2)
```

```
      v1     v2       v3     v5         v6         v7   v8
1      <NA>        <NA> 60
2      <NA>     (7 ) 60
3      <NA>       (6 ) 60
4            <NA>  (7 ) 60
5            <NA>   <NA> 60
6      <NA>      (10 ) 60
  political_engagement$engagement_score
1                                     0
2                                     4
3                                     1
4                                     2
5                                     0
6                                     0
```

```
#
mean_engagement <- mean(political_engagement$engagement_score)
var_engagement <- var(political_engagement$engagement_score)
print(mean_engagement)
```

```
[1] 1.063435
```

```
print(var_engagement)
```

```
[1] 2.55765
```

```
#
if (var_engagement > mean_engagement) {
  print("          ")
}
```

```
[1] "          "
```

```
#         MASS
if (var_engagement > mean_engagement) {
  library(MASS)
  glm_negbin <- glm.nb(political_engagement$engagement_score ~ v6 + v7 + v8 + v1, data = data2)
  summary(glm_negbin)
}
```

```
Attaching package: 'MASS'

The following object is masked from 'package:dplyr':

    select
```

```
Call:
glm.nb(formula = political_engagement$engagement_score ~ v6 +
    v7 + v8 + v1, data = data2, init.theta = 0.6258750942, link = log)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.07825    0.30518  -3.533 0.000411 ***
v630-39      0.77007    0.31503   2.444 0.014506 *
v640-49      0.89173    0.29142   3.060 0.002213 **
v650-59      1.01004    0.28303   3.569 0.000359 ***
v660    0.93793    0.27922   3.359 0.000782 ***
v7     0.25893    0.16467   1.572 0.115855
v7     0.50668    0.13195   3.840 0.000123 ***
v7         0.51789    0.15992   3.239 0.001201 **
v7      0.50471    0.14010   3.603 0.000315 ***
v8        -0.19371    0.08241  -2.350 0.018749 *
v1       -0.14018    0.08591  -1.632 0.102753
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6259) family taken to be 1)

    Null deviance: 1531.0  on 1600  degrees of freedom
Residual deviance: 1485.5  on 1590  degrees of freedom
  (70 observations deleted due to missingness)
AIC: 4584.9

Number of Fisher Scoring iterations: 1

              Theta:  0.6259
          Std. Err.:  0.0473

 2 x log-likelihood:  -4560.9200
```

根據結果，您的負二項回歸模型在解釋受訪者的「政治熱衷程度」上提供了一些顯著的發現，並表明過度離散存在，這使得負二項回歸成為更合適的模型。

模型結果解釋：

1. 截距（Intercept）：

- `Estimate = -1.07825`，表示在所有自變數（年齡、教育程度、性別、地區）都是參考組的情況下，預期的政治熱衷程度的對數值為 -1.07825，轉換成原始單位後相當於參與度較低。

2. 年齡（v6）：

- **30-39歲** (`Estimate = 0.77007, p = 0.014506`)：這個年齡組的政治熱衷程度顯著高於參考組（<30歲），且與其他年齡組相比，較年輕族群政治熱衷程度低於中年族群。
- **40-49歲** (`Estimate = 0.89173, p = 0.002213`)：這個年齡組對政治的熱衷程度顯著高於參考組。
- **50-59歲** (`Estimate = 1.01004, p = 0.000359`)：該年齡組顯示出顯著更高的政治熱衷程度。
- **60歲及以上** (`Estimate = 0.93793, p = 0.000782`)：同樣顯示出高於參考組的政治參與度，且非常顯著。

3. 教育程度（v7）：

- **初中、國中**：不顯著 (p = 0.115855)，但高於參考組（小學及以下）。
- **高中、高職** (Estimate = 0.50668, p = 0.000123)：顯示出顯著高於低教育程度的政治熱衷程度。
- **專科** (Estimate = 0.51789, p = 0.001201)：比低教育程度顯著更高的政治熱衷程度。
- **大學以上** (Estimate = 0.50471, p = 0.000315)：政治參與顯著較高。

4. 性別（v8）：

- **女性** (Estimate = -0.19371, p = 0.018749)：性別對政治熱衷程度有顯著影響，女性的參與度顯著低於男性。

5. 地區（v1）：

- **中西區** (Estimate = -0.14018, p = 0.102753)：地區對政治熱衷程度的影響不顯著（p值略大於0.05），因此可能與其他地區相比，差異並不顯著。

模型診斷：

- 模型的**AIC**為 4584.9，這是一個衡量模型擬合優劣的指標，數值越小代表模型越好。
- **Null deviance** 與 **Residual deviance** 表示模型擬合的情況，較小的 Residual deviance 表明模型能夠較好地解釋數據。
- **Theta = 0.6259** 表示該負二項回歸的過度離散參數，數值越低表明數據的過度離散越大。

結論：

從結果來看，年齡、教育程度和性別對政治熱衷程度有顯著影響。具體來說，年齡越大，受訪者越傾向於對政治有更高的參與度；教育程度越高的人也傾向於更多地參與政治；而女性在這項研究中表現出較低的政治參與度。地區變數（中西區）對政治熱衷程度的影響不顯著。