HW4

侯斯存，劉宸甫，陳豪駿，尹法祥，溫翎傑

2024-10-27

目录

讀取資料：知名度與支持度

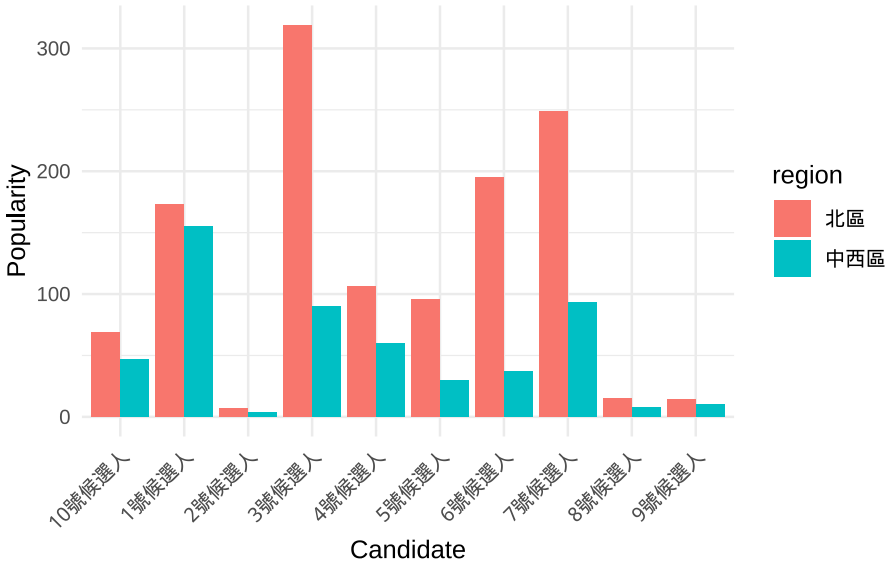
```
library(dplyr)
library(ggplot2)
library(haven)

data <- read_sav("poll.sav")
data_north<-data[which(data[,1]==1),]
data_west<-data[which(data[,1]==2),]#
data_north<-data_north[,-3]
data_west<-data_west[,-2]

north_multiple_answer<-data_north[,c(3:10)]
west_multiple_answer<-data_west[,c(3:10)]
multiple_answer<-data[,c(4:11)]
sum(north_multiple_answer == 3)
```

```
[1] 319
```

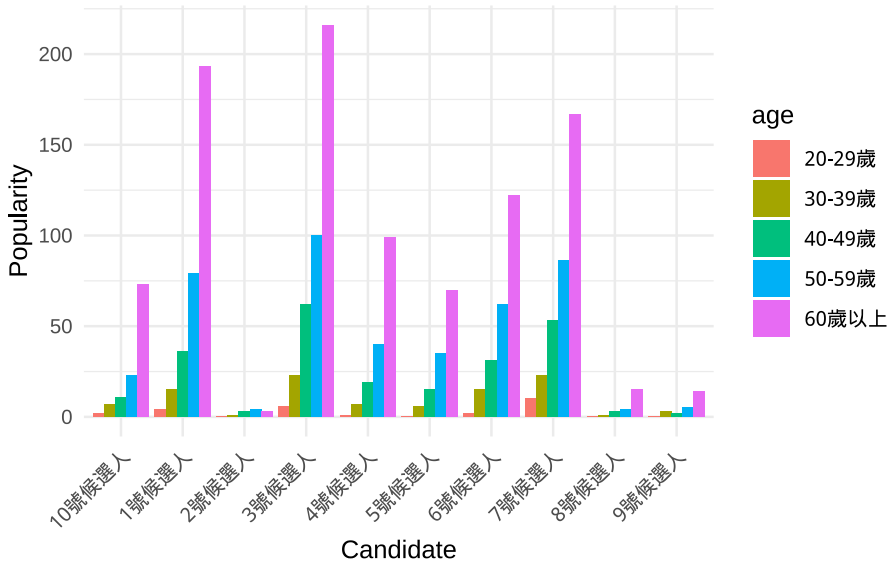
```r
#
#
candidate_ids <- 1:10
regions <- c(" ", "  ")  #
age<-c(1:5)
edu<-c(1:5)
sex<-c(1:2)
#
popularity <- data.frame(matrix(nrow = length(candidate_ids),
                                ncol = length(regions)+
                                  length(age)+
                                  length(edu)+
                                  length(sex)))
rownames(popularity) <- paste("candidate", candidate_ids, sep = "")
colnames(popularity) <- c(regions,
                          paste("age", age, sep = ""),
                          paste("education", edu, sep = ""),
                          paste("sex", sex, sep = ""))

popularity[, " "] <- sapply(candidate_ids,
function(i) sum(north_multiple_answer == i))


popularity[, "  "] <- sapply(candidate_ids,
function(i) sum(west_multiple_answer == i))

for (v in 1:5) {
  col_name <- paste("age", v, sep = "")
  popularity[, col_name] <- sapply(candidate_ids,
function(i) sum((multiple_answer == i) & (data$v6 == v)))
}

for (e in 1:5) {
  col_name <- paste("education", e, sep = "")
  popularity[, col_name] <- sapply(candidate_ids,
function(i) sum((multiple_answer == i) & (data$v7 == e)))
}

for (s in 1:2) {
  col_name <- paste("sex", s, sep = "")
  popularity[, col_name] <- sapply(candidate_ids,
function(i) sum((multiple_answer == i) & (data$v8 == s)))
}
```

```r
#
library(reshape2)
popularity$candidate<-paste("candidate",1:10,sep="")

popularity <- popularity%>%
  mutate(candidate=case_when(
    candidate == "candidate1" ~ "1  ",
    candidate == "candidate2" ~ "2  ",
    candidate == "candidate3" ~ "3  ",
    candidate == "candidate4" ~ "4  ",
    candidate == "candidate5" ~ "5  ",
    candidate == "candidate6" ~ "6  ",
    candidate == "candidate7" ~ "7  ",
    candidate == "candidate8" ~ "8  ",
    candidate == "candidate9" ~ "9  ",
    candidate == "candidate10" ~ "10  ",
    TRUE ~ as.character(candidate)
  ))

popularity_long_region <- melt(popularity,
                               id.vars = "candidate",
                               measure.vars = c(" ", "  "),
                               variable.name = "region",
                               value.name = "popularity")

#   age
popularity_long_age <- melt(popularity, id.vars = "candidate",
                            measure.vars = paste("age", 1:5, sep = ""),
                            variable.name = "age",
                            value.name = "popularity")
popularity_long_age <-popularity_long_age%>%
  mutate(age = case_when(
    age == "age1" ~ "20-29 ",
    age == "age2" ~ "30-39 ",
    age == "age3" ~ "40-49 ",
    age == "age4" ~ "50-59 ",
    age == "age5" ~ "60  ",
    TRUE ~ as.character(age)
  ))


#   education
popularity_long_education <- melt(popularity,
                                  id.vars = "candidate",
```

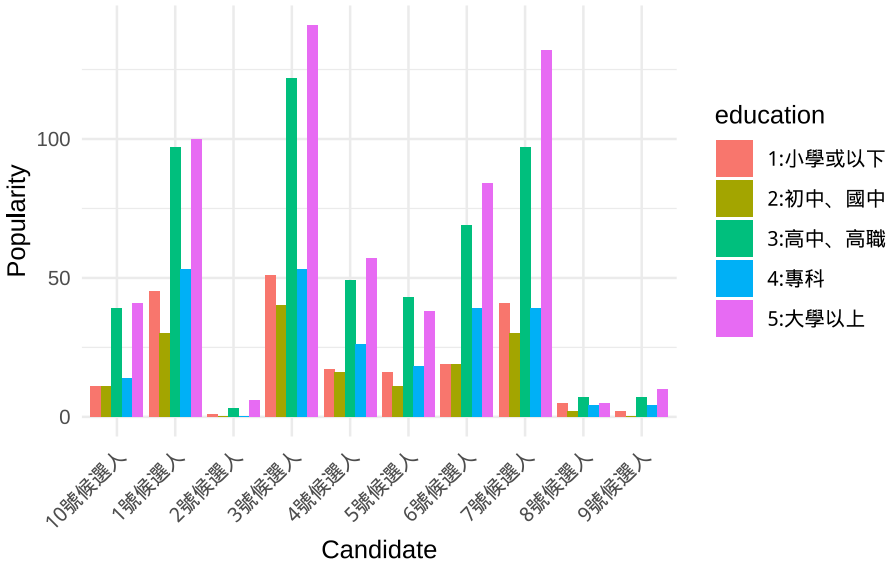
```
                                    measure.vars = paste("education", 1:5, sep = ""),
                                    variable.name = "education",
                                    value.name = "popularity")
popularity_long_education <-popularity_long_education%>%
  mutate(education = case_when(
    education == "education1" ~ "1:   ",
    education == "education2" ~ "2:   ",
    education == "education3" ~ "3:   ",
    education == "education4" ~ "4: ",
    education == "education5" ~ "5:  ",
    TRUE ~ as.character(education)
  ))

#   sex
popularity_long_sex <- melt(popularity,
                            id.vars = "candidate",
                            measure.vars = paste("sex", 1:2, sep = ""),
                            variable.name = "sex",
                            value.name = "popularity")
popularity_long_sex <-popularity_long_sex%>%
  mutate(sex = case_when(
    sex == "sex1" ~ "1: ",
    sex == "sex2" ~ "2: ",
    TRUE ~ as.character(sex)
  ))

#
ggplot(popularity_long_region,
       aes(x = candidate, y = popularity, fill = region)) +
       geom_bar(stat = "identity", position = "dodge") +
       labs(, x = "Candidate", y = "Popularity") +
       theme_minimal()+
       theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



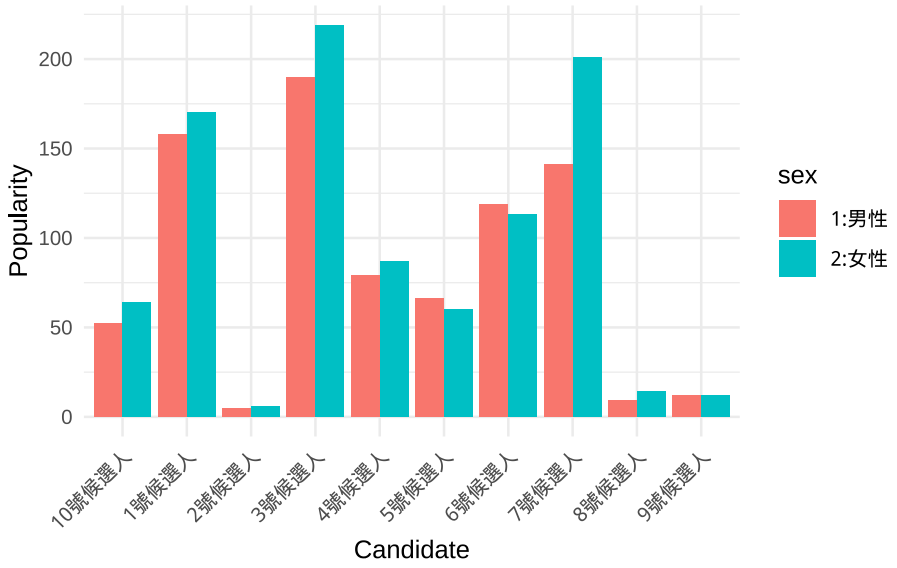

```
ggplot(popularity_long_age,
       aes(x = candidate, y = popularity, fill = age)) +
   geom_bar(stat = "identity", position = "dodge") +
   labs(, x = "Candidate", y = "Popularity") +
   theme_minimal()+
   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
ggplot(popularity_long_education,
       aes(x = candidate, y = popularity, fill = education)) +
   geom_bar(stat = "identity", position = "dodge") +
   labs(, x = "Candidate", y = "Popularity") +
   theme_minimal()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
ggplot(popularity_long_sex,
       aes(x = candidate, y = popularity, fill = sex)) +
   geom_bar(stat = "identity", position = "dodge") +
   labs(, x = "Candidate", y = "Popularity") +
   theme_minimal()+
   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

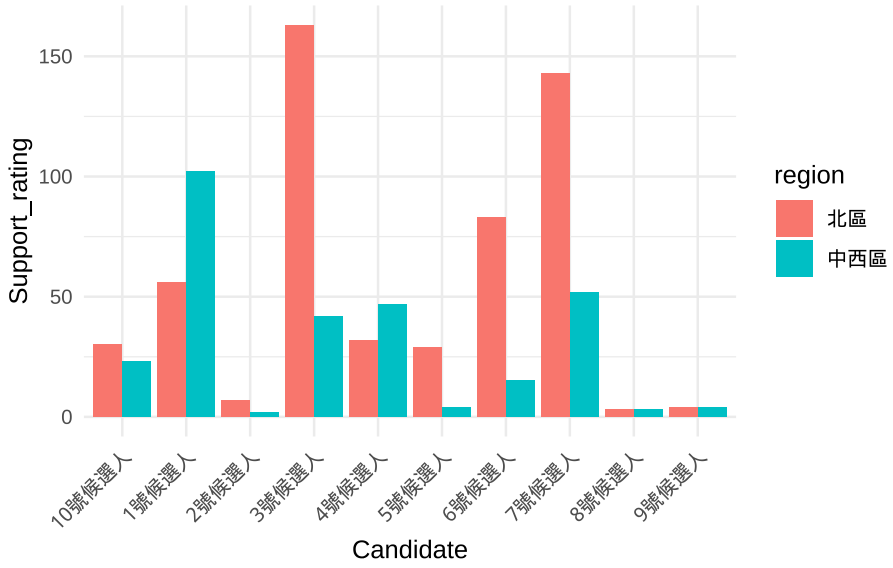


#Support rating

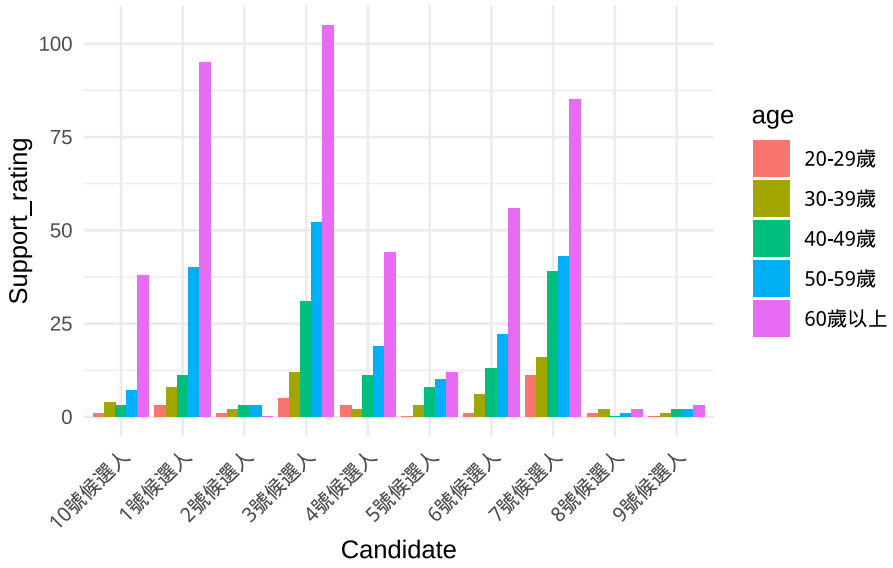
```
candidate_ids <- 1:10
regions <- c(" ", "  ")
age<-c(1:6)
edu<-c(1:5)
sex<-c(1:2)
#
Support_rating <- data.frame(matrix(nrow = length(candidate_ids),
                                    ncol = length(regions)+
                                      length(age)+
                                      length(edu)+
                                      length(sex)))
rownames(Support_rating) <- paste("candidate", candidate_ids, sep = "")
colnames(Support_rating) <- c(regions,paste("age", age, sep = ""),
                              paste("education", edu, sep = ""),
                              paste("sex", sex, sep = ""))

Support_rating[, " "] <- sapply(candidate_ids ,
function(i) sum(data_north$v5 == i))


Support_rating[, "  "] <- sapply(candidate_ids,
function(i) sum(data_west$v5 == i))
```

```r
for (v in 1:5) {
  col_name <- paste("age", v, sep = "")
  Support_rating[, col_name] <- sapply(candidate_ids,
function(i) sum((data$v5 == i) & (data$v6 == v)))
}

for (e in 1:5) {
  col_name <- paste("education", e, sep = "")
  Support_rating[, col_name] <- sapply(candidate_ids,
function(i) sum((data$v5 == i) & (data$v7 == e)))
}

for (s in 1:2) {
  col_name <- paste("sex", s, sep = "")
  Support_rating[, col_name] <- sapply(candidate_ids,
function(i) sum((data$v5 == i) & (data$v8 == s)))
}

Support_rating$candidate<-paste("candidate",1:10,sep="")
Support_rating <- Support_rating%>%
 mutate(candidate=case_when(
    candidate == "candidate1" ~ "1  ",
    candidate == "candidate2" ~ "2  ",
    candidate == "candidate3" ~ "3  ",
    candidate == "candidate4" ~ "4  ",
    candidate == "candidate5" ~ "5  ",
    candidate == "candidate6" ~ "6  ",
    candidate == "candidate7" ~ "7  ",
    candidate == "candidate8" ~ "8  ",
    candidate == "candidate9" ~ "9  ",
    candidate == "candidate10" ~ "10  ",
    TRUE ~ as.character(candidate)
  ))

#
Support_rating_long_region <- melt(Support_rating, id.vars = "candidate",
                                   measure.vars = c(" ", "  "),
                                   variable.name = "region",
                                   value.name = "Support_rating")


Support_rating_long_age <- melt(Support_rating, id.vars = "candidate",
                                measure.vars = paste("age", 1:5, sep = ""),
                                variable.name = "age",
```

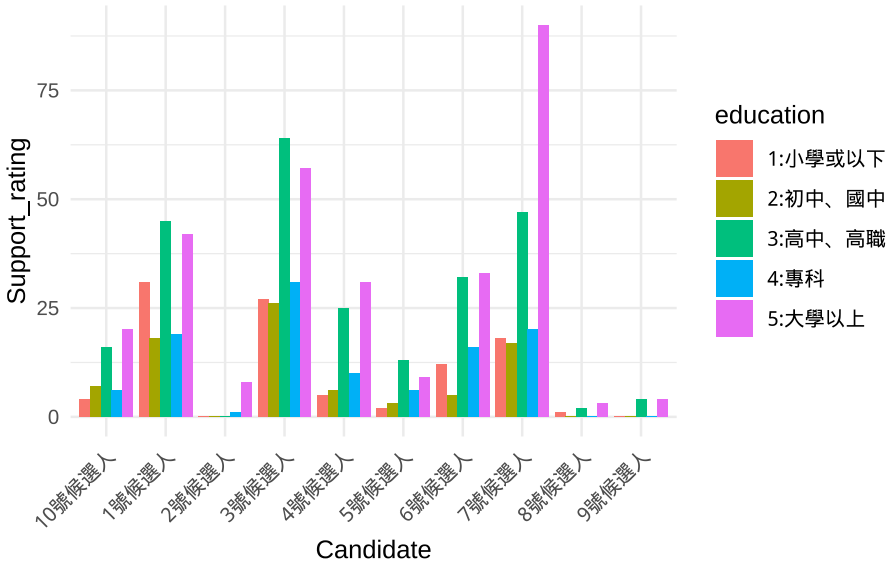
```
                              value.name = "Support_rating")
Support_rating_long_age<- Support_rating_long_age%>%
  mutate(age = case_when(
    age == "age1" ~ "20-29 ",
    age == "age2" ~ "30-39 ",
    age == "age3" ~ "40-49 ",
    age == "age4" ~ "50-59 ",
    age == "age5" ~ "60  ",
    TRUE ~ as.character(age)
  ))


Support_rating_long_education <- melt(Support_rating, id.vars = "candidate",
                                      measure.vars = paste("education", 1:5, sep = ""),
                                      variable.name = "education",
                                      value.name = "Support_rating")
Support_rating_long_education <-Support_rating_long_education%>%
  mutate(education = case_when(
    education == "education1" ~ "1:   ",
    education == "education2" ~ "2:   ",
    education == "education3" ~ "3:   ",
    education == "education4" ~ "4: ",
    education == "education5" ~ "5:  ",
    TRUE ~ as.character(education)
  ))

Support_rating_long_sex <- melt(Support_rating, id.vars = "candidate",
                                measure.vars = paste("sex", 1:2, sep = ""),
                                variable.name = "sex",
                                value.name = "Support_rating")
Support_rating_long_sex <- Support_rating_long_sex%>%
  mutate(sex = case_when(
    sex == "sex1" ~ "1: ",
    sex == "sex2" ~ "2: ",
    TRUE ~ as.character(sex)
  ))

ggplot(Support_rating_long_region,
       aes(x = candidate, y = Support_rating, fill = region)) +
   geom_bar(stat = "identity", position = "dodge") +
   labs(, x = "Candidate", y = "Support_rating") +
   theme_minimal()+
   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



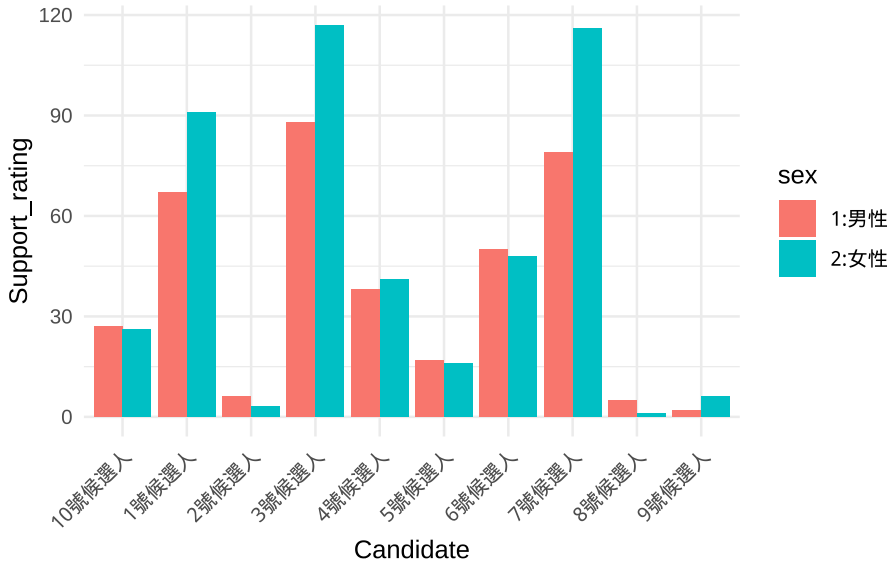

```
ggplot(Support_rating_long_age,
       aes(x = candidate, y = Support_rating, fill = age)) +
       geom_bar(stat = "identity", position = "dodge") +
       labs(, x = "Candidate", y = "Support_rating") +
       theme_minimal()+
       theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



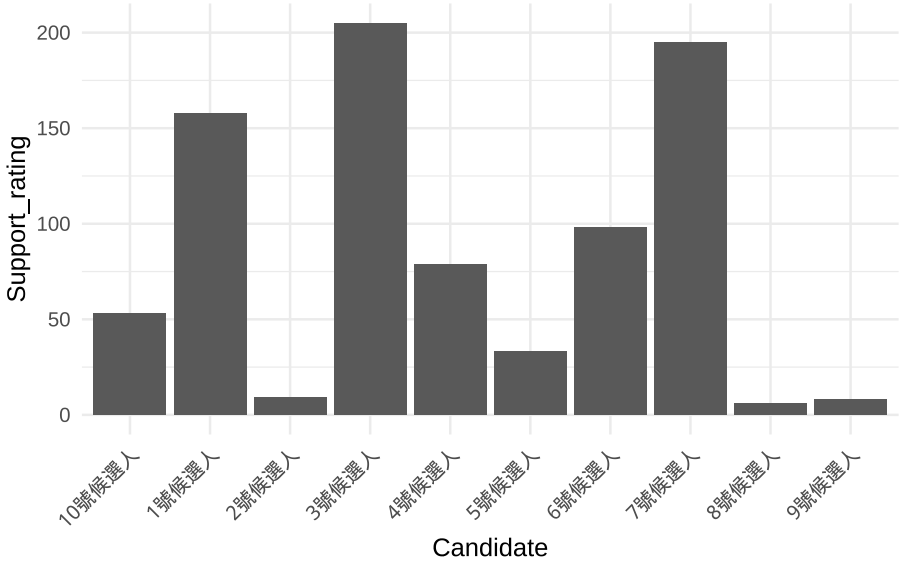

```
ggplot(Support_rating_long_education,
       aes(x = candidate, y = Support_rating, fill = education)) +
       geom_bar(stat = "identity", position = "dodge") +
       labs(, x = "Candidate", y = "Support_rating") +
       theme_minimal()+
       theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
ggplot(Support_rating_long_sex,
       aes(x = candidate, y =Support_rating, fill = sex)) +
       geom_bar(stat = "identity", position = "dodge") +
       labs(, x = "Candidate", y = "Support_rating") +
       theme_minimal()+
       theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
total_supporting<-data.frame(matrix(nrow = length(candidate_ids), ncol = 1))
colnames(total_supporting)<-"Support_rating"
total_supporting$candidate<-c("1  ",
                              "2  ",
                              "3  ",
                              "4  ",
                              "5  ",
                              "6  ",
                              "7  ",
                              "8  ",
                              "9  ",
                              "10  ")
total_supporting[, "Support_rating"] <- sapply(candidate_ids ,

ggplot(total_supporting, aes(x = candidate, y =Support_rating)) +
   geom_bar(stat = "identity", position = "dodge") +
   labs(, x = "Candidate", y = "Support_rating") +
   theme_minimal()+
   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



1、知名度與支持度分析

1. 整體來看，不論是知名度或支持度，2,8,9號候選人都很沒有競爭力。

   依照地區的知名度來看，北區最高是3號候選人，中西區最高是1號。
   而在中西區，3,4,7,10號候選人的知名度相差不大。

2. 高齡大多知曉3號候選人，而年輕族群則較為熟識7號候選人。
3. 教育水準較高的選民較多比例會選擇3號候選人，反之較低者會傾向選7號候選人。
4. 支持度最高為3號候選人，其次為7號，但差異不大。

2、三號候選人選舉策略

觀察

知名度

1. 3號候選人在中西區的知名度低於1,7號候選人
2. 在年輕選民中（20-40間），3號候選人的知名度低於7號候選人，尤其是在20-30歲的區段。
3. 3號候選人在各個教育程度的群眾中皆較為知名，但與7號候選人的差距在高教育程度的選民時減小。

支持度

4. 教育水準較低的選民較傾向3號候選人，反之較高者會傾向選7號候選人。

5. 整體而言，3號候選人的支持度略微高於7。

結論

認為3號候選人若想在市議員選舉中增加當選的勝算，策略上應該以7號候選人作為主要競爭對手。
除了穩住自己的基本盤（高齡、低教育程度）以外，應該也向年輕選民和高教育程度者多多表現自己，提升支持度。
且因為調查限制為20歲，忽略了選舉時剛獲得選舉權的首投族，以此調查的趨勢來看，7號候選人可能有潛藏的票倉，需要注意。

3、第3號候選人支持率的預測模式

```
#   logistic
dt<-data[,c("v1","v5","v6","v7","v8")]
colnames(dt)<-c("region","support","age","education","sex")
dt$region<-as.factor(dt$region)
dt[,"support"]<-dt[,"support"]==3
dt<-dt[-which(dt[,"age"]==6),]#
dt<-dt[-which(dt[,"education"]==95),]#
dt$age <- factor(dt$age)
dt$education<-factor(dt$education)
dt$sex<-as.factor(dt$sex)
model1<-glm(dt$support~dt$region+dt$age+dt$education+dt$sex,
            data=dt,
            family = binomial(link = "logit"))
summary(model1)
```

```
Call:
glm(formula = dt$support ~ dt$region + dt$age + dt$education +
    dt$sex, family = binomial(link = "logit"), data = dt)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.08096    0.55148  -3.773 0.000161 ***
dt$region2    -0.72945    0.18312  -3.983 6.79e-05 ***
dt$age2        0.30874    0.56826   0.543 0.586914
dt$age3        0.45620    0.51624   0.884 0.376852
dt$age4        0.37840    0.50210   0.754 0.451064
dt$age5        0.05231    0.49757   0.105 0.916275
dt$education2  0.54130    0.30091   1.799 0.072035 .
dt$education3  0.37288    0.25697   1.451 0.146771
dt$education4  0.39770    0.30565   1.301 0.193206
dt$education5 -0.06315    0.28283  -0.223 0.823305
dt$sex2       -0.06044    0.15613  -0.387 0.698664
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1225.3  on 1600  degrees of freedom
Residual deviance: 1192.7  on 1590  degrees of freedom
AIC: 1214.7

Number of Fisher Scoring iterations: 5
```

```r
#Step method
model1.step <- step(model1,direction = c("both"))
```

```
Start:  AIC=1214.75
dt$support ~ dt$region + dt$age + dt$education + dt$sex

               Df Deviance    AIC
- dt$age        4   1197.1 1211.1
- dt$sex        1   1192.9 1212.9
<none>              1192.8 1214.8
- dt$education  4   1201.7 1215.7
- dt$region     1   1210.3 1230.3

Step:  AIC=1211.08
dt$support ~ dt$region + dt$education + dt$sex

               Df Deviance    AIC
- dt$sex        1   1197.2 1209.2
<none>              1197.1 1211.1
- dt$education  4   1206.5 1212.5
+ dt$age        4   1192.8 1214.8
- dt$region     1   1215.7 1227.7

Step:  AIC=1209.15
dt$support ~ dt$region + dt$education

               Df Deviance    AIC
<none>              1197.2 1209.2
- dt$education  4   1206.8 1210.8
+ dt$sex        1   1197.1 1211.1
+ dt$age        4   1192.9 1212.9
- dt$region     1   1215.8 1225.8
```

```r
summary(model1.step)
```

```
Call:
glm(formula = dt$support ~ dt$region + dt$education, family = binomial(link = "logit"),
    data = dt)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -2.0600     0.2073  -9.937  < 2e-16 ***
dt$region2     -0.7481     0.1826  -4.096  4.2e-05 ***
dt$education2   0.6172     0.2959   2.086   0.0370 *
dt$education3   0.5075     0.2445   2.075   0.0379 *
dt$education4   0.6118     0.2829   2.163   0.0306 *
dt$education5   0.1581     0.2472   0.640   0.5224
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1225.3  on 1600  degrees of freedom
Residual deviance: 1197.2  on 1595  degrees of freedom
AIC: 1209.2

Number of Fisher Scoring iterations: 5
```

模型觀察

Logistic : Step Method

透過Step method，我們知道模型中education2,3,4與region2為顯著。
因模型預設以每個變數的 1 為基底，我們得到以下結論：

- region2(中西區)的係數約為-0.7481。
  代表選民來自中西區時，支持三號候選人的可能性約為北區選民的 $e^{-0.4781} = 0.473$ 倍
- education2(初中、國中)的係數約為0.6172。
  代表選民最高學歷為初中或國中時，支持三號候選人的可能性約為最高學歷為小學或以下的 $e^{0.6172} = 1.855$ 倍
- education3(高中、高職)的係數約為0.5075。
  代表選民最高學歷為高中或高職時，支持三號候選人的可能性約為最高學歷為小學或以下的 $e^{0.5075} = 1.661$ 倍
- education4(專科)的係數約為0.6118。
  代表選民最高學歷為專科時，支持三號候選人的可能性約為最高學歷為小學或以下的

$e^{0.6118} = 1.844$ 倍