取消文化之現象分析

高嘉妤、柯堯琙、吳承恩、趙友誠

2024-11-15

Table of contents

report 的重要頁碼

52: 網路瘾誘與脫序行為之子題說明

92: 資料人口結構與母群人口結構比較表

281: 各題目之測量概念

處理後資料簡介

原始資料維度: rows×columns = 1004×207

原始的資料有 207 個變數，代表問卷中所有的問題選項 (包含複選題以及注意力偵測題等等)。

我們首先移除與分析無關的變數:

8 題: 大部分的人都有透過網路接觸名人的資訊或討論 (只有四個人沒有)，所以決定移除。

9 題: 即時通訊軟體多為聯繫熟人或工作上使用，較難顯現是否有突破同溫層的現象，所以決定移除。

疫情相關題目 (12~15): 關心的題目 (28,29) 時間範圍較廣，並不只局限於疫情期間，所以決定移除。

```
library(haven)          #read sav file
library(labelled)       #remove attribute of sav data
library(Hmisc)          #describe
library(showtext)       #show zw-tw in ggplot2
library(dplyr);library(ggplot2);library(MASS)
library(rlang)          #for building function
#DB.sav <-read_sav("DisruptiveBehavior.sav")
#write.csv(DB.sav,file= "DisruptiveBehavior.csv", row.names= FALSE)
DB.csv <-read.csv("DisruptiveBehavior.csv")[,-c(1:4)]
showtext_auto() #render 的 ggplot 可以顯示中文

# 移除注意力偵測題
DB.csv[,match("q21a_1", colnames(DB.csv)):match("q21a_6_text", colnames(DB.csv))] <- NULL
DB.csv$q37a <- NULL
DB.csv$rq21a <- NULL
DB.csv$rq37a <- NULL
DB.csv$r <- NULL # 基本資料與第一波網調是否相符

# 移除 q8,q9
DB.csv[, match("q8_1", colnames(DB.csv)):match("q8_90", colnames(DB.csv))] <- NULL
DB.csv[, match("q9_1",colnames(DB.csv)):match("q9_90",colnames(DB.csv))] <- NULL

# 移除疫情相關的問題 (12 題到 15 題)
DB.csv[,match("q12_1", colnames(DB.csv)):match("q15_03_1", colnames(DB.csv))] <- NULL
```

接著在對一些題目進行細部的選項討論:

人口結構變數處理

年齡: 移除出生年的資訊，將 rrq2 的年齡分層變數重新命名"q2_rr"。

出生地: 其他類別歸在一類 (24)。但是類別有點多，考慮對人口結構表格中的分類方式 (北北基宜、桃竹苗等區分)，還沒做。

教育程度: 重新劃分為四個等級 (1: 高中及以下, 2: 專科, 3: 大學, 4: 研究所)，劃分參考人口結構表格的分類方式。

```
# 第二題 (出生年) 改成年齡的區段
DB.csv$q2 <- DB.csv$qrq2
DB.csv$qrq2 <- NULL
DB.csv$q2_rr <- DB.csv$rrq2
DB.csv$rrq2 <- NULL

# 把第三題 (出生地) 的其他類別歸為一類
DB.csv$q3_other <- NULL

# 第四題沒有人選其他
DB.csv$q4_88_text <- NULL

# 教育程度重新劃分為四個等級
DB.csv$q4[DB.csv$q4<=8] <- 1
DB.csv$q4[DB.csv$q4!=1 & DB.csv$q4<=15] <- 2
DB.csv$q4[DB.csv$q4>2 & DB.csv$q4<=19] <- 3
DB.csv$q4[DB.csv$q4>3] <- 4
```

其他變數的更動

6、7 題: 時間統一單位 (分)

10 題: 改成” 使用幾個與 yt 名人討論相關的社群媒體”，因為有些社群媒體不會造成抵制名人行為，例如:Pinterest,Linkedin,+ 其他類 Pixiv,Mobile01,Komica,MeWe 跟名人相關的討論比較少，所以決定簡化選項; 巴哈姆特, 巴哈姆特場外休憩區兩個則要計算。

11 題: 改成” 有無使用 YT, Twitch, 或 bilibili”(1: 有使用,0: 沒有使用)，原因與第十題類似。

```
# 時間統一單位（分）
DB.csv$q6 <- DB.csv$q6_h*60+DB.csv$q6_m
DB.csv$q7 <- DB.csv$q7_h*60+DB.csv$q7_m
DB.csv$q6_h <- NULL; DB.csv$q6_m <- NULL
DB.csv$q7_h <- NULL; DB.csv$q7_m <- NULL

# 整理第十題
DB.csv$q10_4 <- NULL
DB.csv$q10_10 <- NULL
DB.csv$q10_90 <- NULL
DB.csv$q10_88[DB.csv$q10_88_text!=" 巴哈姆特場外休憩區"&DB.csv$q10_88_text!=" 巴哈姆特"] <- NA
DB.csv$q10_88_text <- NULL
DB.csv$q10 <- apply(DB.csv[,c("q10_1", "q10_2", "q10_3", "q10_5", "q10_6", "q10_7", "q10_8", "q10_9","q1
                    1, function(row){sum(!is.na(row))})
DB.csv[,c("q10_1", "q10_2", "q10_3", "q10_5", "q10_6", "q10_7", "q10_8", "q10_9", "q10_88")] <- NULL

# 整理第十一題
DB.csv$q11_2 <- NULL
DB.csv$q11_3 <- NULL
DB.csv$q11_4 <- NULL
DB.csv$q11_5 <- NULL
DB.csv$q11_6 <- NULL
DB.csv$q11_8 <- NULL
DB.csv$q11_90 <- NULL
DB.csv$q11_88[DB.csv$q11_88_text!="bilibili"] <- NA
DB.csv$q11_88_text <- NULL
DB.csv$q11 <- apply(DB.csv[,c("q11_1", "q11_7")],
                    1, function(row){sum(!is.na(row))})
DB.csv[,c("q11_1", "q11_7","q11_88")] <- NULL
```

16 題 ~19 題 (惡搞行為): 將每個類別補 0(變成 1,0)，再創建一個標籤變數 q1719_label(1: 至少有一個惡搞行為,0: 都沒有)。

```
DB.csv$q16 <- NULL
DB.csv$q18 <- NULL
DB.csv$q17_01[is.na(DB.csv$q17_01)|DB.csv$q17_01==2] <- 0
DB.csv$q17_02[is.na(DB.csv$q17_02)|DB.csv$q17_02==2] <- 0

DB.csv$q19_01[is.na(DB.csv$q19_01)|DB.csv$q19_01==2] <- 0
DB.csv$q19_02[is.na(DB.csv$q19_02)|DB.csv$q19_02==2] <- 0

DB.csv$q1719_label <- apply(
  DB.csv[,match("q17_01",colnames(DB.csv)):match("q19_02",colnames(DB.csv))],
  MARGIN = 1,
  function(row){
    return(paste0(row,collapse = ""))
```

```
  })
unique(DB.csv$q1719_label)
```

```
[1] "0000" "1101" "1100" "1000" "0100" "1110" "1111" "0101" "0001"
```

```
DB.csv$q1719_label <- ifelse(DB.csv$q1719_label=="0000", 0, 1)
```

第二十二題 ~ 二十六題: 參考碩士論文: 台灣消費者抵制行為之研究 —以台商親中言論衍生之抵制為例 (https://www.airitilibrary.com/Article/Detail/U0004-G0107932056) 之做法，將相同大主題的 ordinal 主觀評分加總作為該主題程度的分數。

這裡的分數要不要用加總的? 跟哪些分數要加在一起要討論一下，我覺得 25 的幾題跟 20 題那邊的蠻像的。

22 題 (看見他人網路攻擊行為 (網路使用環境)): 分數越高越常看到環境中其他人的攻擊。

23 題 (自己的網路攻擊行為): 分數越高代表自己的攻擊性越高

24 題 (回聲室效應): 分數越高則較常突破同溫層或是媒體識讀素養較高

25 題 (網路攻擊接受性): 分數越高越覺得網路上的攻擊行為 OK。但是 25 題的第三題

26 題 (推測對他人之攻擊意圖): 受訪者對網路攻擊行為的看法，分數越高代表受訪者越覺得網路攻擊行為容易引起他人的攻擊性。

```
DB.csv$q22 <- rowSums(DB.csv[,c("q22_01_1", "q22_02_1", "q22_03_1", "q22_04_1", "q22_05_1")])
DB.csv$q23 <- rowSums(DB.csv[,c("q23_01_1", "q23_02_1", "q23_03_1", "q23_04_1", "q23_05_1")])
DB.csv$q24 <- rowSums(DB.csv[,c("q24_01_1", "q24_02_1", "q24_03_1", "q24_04_1", "q24_05_1")])
DB.csv$q25 <- rowSums(DB.csv[,c("q25_01_1", "q25_02_1", "q25_03_1", "q25_04_1")])
DB.csv$q26 <- rowSums(DB.csv[,c("q26_01_1", "q26_02_1", "q26_03_1")])
DB.csv[,match("q22_01_1",colnames(DB.csv)):match("q26_03_1",colnames(DB.csv))] <- NULL
```

38 題 ~42 題 (最後一題)

38 題: 心理幸福感（表現自尊）的評分，將 (生活滿意度、社會滿意度) 加總

40 題: 國民黨偏好 $0 \sim 100 \rightarrow 1 \sim 5$

41 題: 民進黨偏好 $0 \sim 100 \rightarrow 1 \sim 5$

42 題: 反台獨程度 $1 \sim 10$

```
DB.csv$q38 <- rowSums(DB.csv[,c("q38_01_1", "q38_02_1")])
DB.csv$q38_01_1 <- NULL
DB.csv$q38_02_1 <- NULL

DB.csv$q40 <- cut(DB.csv$q40_1,
                        breaks = c(0, 20, 40, 60, 80, 100),
                        labels = c(1, 2, 3, 4, 5),
                        right = TRUE)
DB.csv$q40[is.na(DB.csv$q40)] <- 1

DB.csv$q41 <- cut(DB.csv$q41_1,
                       breaks = c(0, 20, 40, 60, 80, 100),
                       labels = c(1, 2, 3, 4, 5),
                       right = TRUE)
DB.csv$q41[is.na(DB.csv$q41)] <- 1

DB.csv$q41_1 <- NULL
DB.csv$q40_1 <- NULL
```

對抵制行為相關問題變數之處理

針對 28,29 進行細部討論後，針對選項進行合理的歸類。

首先對其他類進行歸類:

28 題 (做過的抵制行為): 其他類 (16 個人有填) 分到前三類或是設 0: 沒有抵制行為。

29 題 (抵制原因): 原先將其他類裡面有出現的不當發言歸為第五類，後續覺得”不當發言”可以與”不道德、不正當或不合法行為”合併，“有不同的政治意識型態或價值觀”可以與”不表態支持重要的社會議題”合併，最後 29 題剩下三個類別 + 沒有抵制行為的 000

```
# 處理 28 的選項
DB.csv$q28_5 <- NULL
q28.manipulation <- function(row){
  # 亂回答的要把其他抵制行為的問題回答（28-36）也移除
  delete.term <- c(" 會破壞我對他（她）的形象",
                   " 從來都不關注",
                   " 若名人不自我反省就會抵制，但是通常名人都會願意出來面對錯誤",
                   " 未來此人所說的話均會產生疑問",
                   " 用選票來抵制",
                   " 很多時候都是立場不同、換位思考一下後，就可以消弭一些爭議。",
                   " 看看就好",
                   " 沒意見",
                   " 看看就好，自己會有自己的判斷")
  # 要移除 q28_4 標籤的
  amend.term <- c(" 指正他的錯誤",
                  " 拒買相關商品",
                  " 與親朋好友說明事實真相",
                  " 要看是什麼原因決定一時間這麼做還是永久")
  if(row[5] %in% delete.term){row <- c(rep(NA,4),"",rep(NA,5),"",rep(NA,7))}
  else if(row[5] %in% amend.term){row[4:5] <- c(NA,"")}
  return(row)
}

DB.csv[,match("q28_1",colnames(DB.csv)):match("q36_1",colnames(DB.csv))] <- as.data.frame(
  t(apply(DB.csv[,match("q28_1",colnames(DB.csv)):match("q36_1",colnames(DB.csv))],
        1,
        q28.manipulation))
)
# 要歸類的要一個一個看歸在哪類
DB.csv[DB.csv$q28_4_text==" 每個人有合法的言論自由，我只會拒絕觀看有問題違法的影片，不會一竿子打翻一條船。",
       c('q28_2','q28_4','q28_4_text')] <- c(1,NA,"")

DB.csv[DB.csv$q28_4_text==" 減少看他們的發文或影片",  c('q28_2','q28_4','q28_4_text')] <- c(1,NA,"")

DB.csv[DB.csv$q28_4_text==" 轉發相關的指正或譴責文章",c('q28_3','q28_4','q28_4_text')] <- c(1,NA,"")

DB.csv$q28_4 <- NULL
DB.csv$q28_4_text <- NULL
```

```
# 處理 29 的選項
#29 的第五選項改定義為 錯誤資訊、不當言論
q29.manipulation <- function(row){
  # 亂回答的要把其他抵制行為的問題回答（28-36）也移除
  delete.term <- c(" 道不同不相為謀不理他們",
                   " 沒有此情況",
```

```
                    "不會抵制",
                    "我沒有特別抵制過呢",
                    "從來沒有",
                    "不明白指的是什麼",
                    "已讀",
                    "不理他們",
                    "不予置評",
                    "無",
                    "不會做無聊的事情",
                    "目前沒有",
                    "不曾",
                    "沒遇過要抵制的事",
                    "沒有",
                    "沒有抵制過")
  # 要被歸類到第五類（不當發言、錯誤資訊）的
  class5 <- c("錯誤資訊",
              "發表錯誤資訊且不更改",
              "指鹿為馬，不實言論，刻意誤導輿論方向。",
              "不當發言",
              "謾罵",
              "誤導",
              "散播不正確消息且不認錯",
              "對動物議題留下錯誤言論，對疫情走向發出錯誤言論（去年康健發文說嬰幼兒不會染疫，被我指正，卻不改
              "假名人之姿發表利已損害公眾利益的言論，企圖影響他人判斷的言論者。",
              "")
  if(row[9] %in% delete.term){row <- c(rep(NA,8),"",rep(NA,7))}
  else if(row[9] %in% class5){row[9] <- ""}
  return(row)
}

DB.csv[,match("q28_1",colnames(DB.csv)):match("q36_1",colnames(DB.csv))] <- as.data.frame(
  t(apply(DB.csv[,match("q28_1",colnames(DB.csv)):match("q36_1",colnames(DB.csv))],
        1,
        q29.manipulation))
)
# 要歸類的要一個一個看歸在哪類
DB.csv[DB.csv$q29_5_text=="過於私人或主觀意識的回答會讓我反感進而抵制收看",
       c('q29_2','q29_5','q29_5_text')] <- c(1,NA,"")

DB.csv[DB.csv$q29_5_text=="味全黑心油事件",
       c('q29_5','q29_5_text')] <- c(NA,"")

DB.csv[DB.csv$q29_5_text=="說謊話（至少是我覺得他在說謊），做錯事不負責還甩鍋給別人。",
  c('q29_5','q29_5_text')] <- c(NA,"")

DB.csv[DB.csv$q29_5_text=="有些事情的看法 做法不同",
       c('q29_2','q29_5','q29_5_text')] <- c(1,NA,"")

DB.csv[DB.csv$q29_5_text=="違反當初自己宣揚的理念",
       c('q29_4','q29_5','q29_5_text')] <- c(1,NA,"")

DB.csv[
  DB.csv$q29_5_text=="泛指公眾人物沒有責任表態但有義務不支持通稱反人類行為，私領域不要太誇張都沒差",
```

```
  c('q29_4','q29_5','q29_5_text')] <- c(1,NA,"")

DB.csv$q29_5_text <- NULL
```

NA 補 0 是在這個階段處理完其他類 (文字) 之後才做。

因為想要做的方向有兩個: “甚麼原因會造成有抵制行為?”, “甚麼原因會影響抵制行為的程度”，所以在最後建立 q28_YN 的二元變數。

```
# 處理完其他類之後先把 NA 補 0
DB.csv <- as.data.frame(
  apply(DB.csv,2,function(col){
    col <- as.numeric(col)
    col[is.na(col)] <- 0
    return(col)
}))

DB.csv$q29_2 <- ifelse(DB.csv$q29_2 | DB.csv$q29_3, 1,0)
DB.csv$q29_3 <- ifelse(DB.csv$q29_4 | DB.csv$q29_5, 1,0)
DB.csv$q29_4 <- NULL
DB.csv$q29_5 <- NULL

# 有無抵制行為 (1: 有,0: 沒有)
DB.csv$q28_YN[DB.csv$q28_1 | DB.csv$q28_2 | DB.csv$q28_3] <- 1
DB.csv$q28_YN[!(DB.csv$q28_1 | DB.csv$q28_2 | DB.csv$q28_3)] <- 0

# 重新調整欄位 index
#colnames(DB.csv)
#colnames(DB.csv)[c(1:2,29,3:5,30:33,6:9,34,10:11,35:39,12,43,13:25,40,26,41:42,27,28)]
DB.csv <- DB.csv[,c(1:2,29,3:5,30:33,6:9,34,10:11,35:39,12,43,13:25,40,26,41:42,27,28)]
for(i in c(1:5,7:42)){
  DB.csv[,i] <- as.integer(DB.csv[,i])
}
```

Table 1: 變數解釋

Variables	Explanation	remark
q1	性別	1: 男性, 2: 女性
q2	年齡	
q2_rr	年齡分層	1:18~29, 2:30~39, 3:40~49, 4:50~59, 5:60~69, 6:70+
q3	出生縣市	1~19: 台灣的縣市 (資料沒有連江、澎湖、金門), 24: 其他
q4	教育程度	1: 高中及以下, 2: 專科, 3: 大學, 4: 研究所
q5_1	週平均上網天數	
q6	上網分鐘 (工作、學習)	
q7	上網分鐘 (娛樂、休閒)	
q10	使用幾個與名人討論相關的社群媒體	
q11	是否使用 YT,Twitch 或 bilibili	
q17_01	是否參與過: 不傷害、騙人	1: 是,0: 否
q17_02	是否參與過: 不傷害、不騙人	1: 是,0: 否
q19_01	是否參與過: 傷害、騙人	1: 是,0: 否
q19_02	是否參與過: 傷害、不騙人	1: 是,0: 否

Variables	Explanation	remark
q1719_label	是否至少有參與過一種網路惡搞	1: 是,0: 否
q20_01_1	主動激化傾向	
q20_02_1	主動激化傾向	
q22	他人攻擊傾向	
q23	自己攻擊傾向	
q24	回聲室效應	
q25	被攻擊的接受度	
q26	推測他人攻擊意圖	
q27_1	抵制意圖	
q28_YN	是否採取過抵制行為	
q28_1	採取過: 取消關注	
q28_2	採取過: 拒絕觀看	
q28_3	採取過: 在網路上留言或發文指責	
q29_1	抵制的原因: 歧視特定國家、種族或性別	
q29_2	抵制的原因: 有不同的政治意識型態或價值觀	
q29_3	抵制的原因: 做出不道德、不正當或不合法行為	
q30_1	抵制行為的有效程度	
q31_1	抵制前的同理心	
q32_1	抵制行為的對名人的傷害程度	
q33_1	抵制行為的對自己的重要程度	
q34_1	抵制成本	
q35_1	抵制規模感知	
q36_1	抵制的社會壓力	
q38	心理幸福感	不滿意 2~5 滿意
q39_1	生活品質	不快樂 1~5 快樂
q40	國民黨喜好程度	不喜歡 0~5 喜歡
q41	民進黨喜好程度	不喜歡 0~5 喜歡
q42_1	意識形態	0~10: 台獨 ~ 統一
weight	人口結構修正權重	

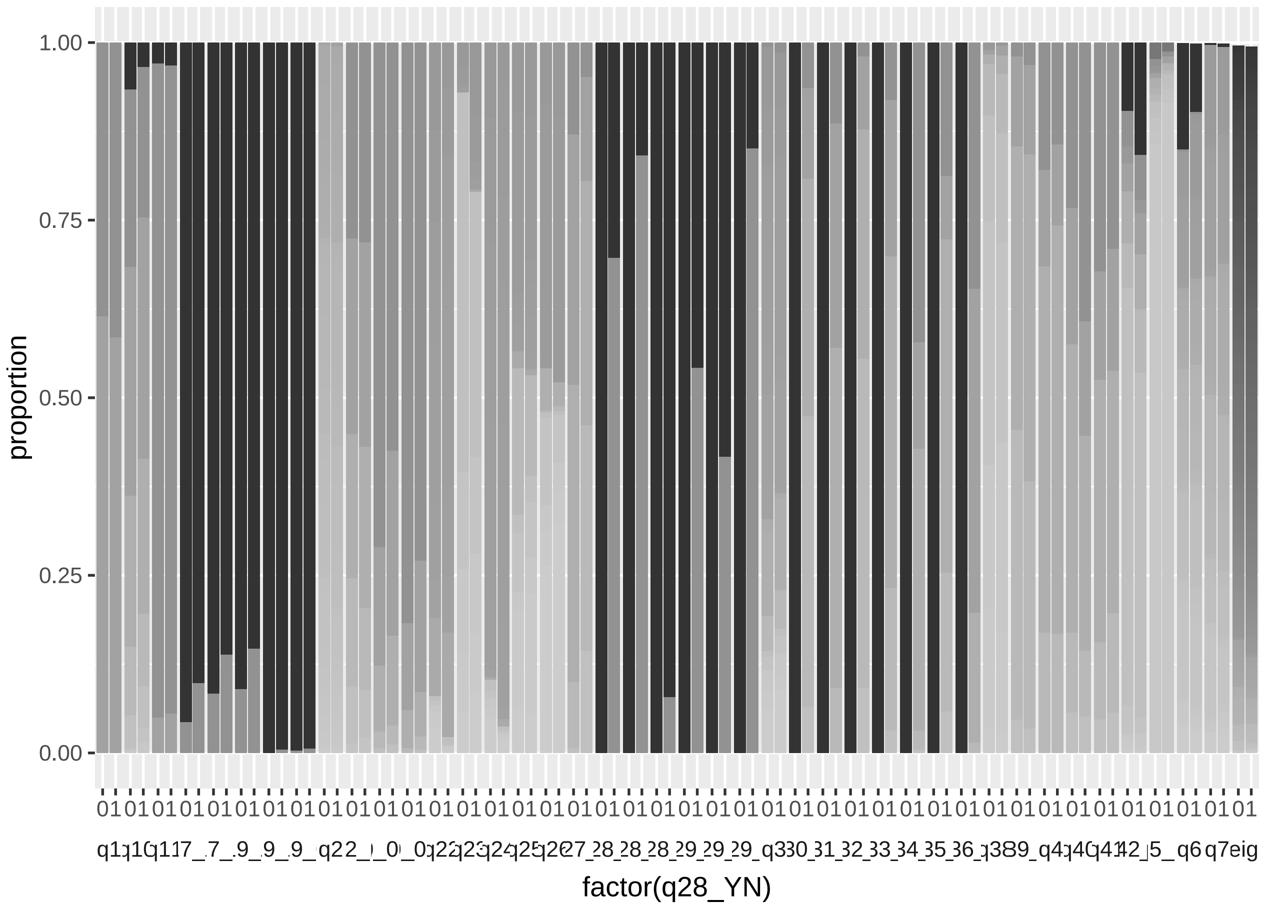
資料視覺化

describe

```
latex(describe(DB.csv),title="",file="")
```

對各變數依 **q28_YN** 二元變數畫比例圖

```
#test code chunk
Barplot.p(myCount_q28(DB.csv, colnames(DB.csv)[-match("q28_YN",colnames(DB.csv))]))
```



Logistic and Decision tree and PCA and XGboost

```
glm_log <- glm(
  factor(q28_YN)~
    factor(q1)+
    factor(q2_rr)+
    factor(q3)+
    factor(q4)+
    q5_1+
    q6+
    q7+
    q10+
    factor(q11)+
    q1719_label+
    factor(q20_01_1)+
    factor(q20_02_1)+
    q22+ q23+ q24+ q25+ q26+
    factor(q27_1), family = binomial, data = DB.csv, weights = weight)

summary(stepAIC(glm_log, direction = 'both', trace = 0))
```

```
Call:
glm(formula = factor(q28_YN) ~ factor(q3) + q6 + q10 + factor(q20_01_1) +
```

```
    q22 + q23 + factor(q27_1), family = binomial, data = DB.csv,
    weights = weight)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.2550617  1.3646272   0.920 0.357724
factor(q3)2       -3.7016521  1.3915703  -2.660 0.007813 **
factor(q3)3       -4.0078695  1.3978406  -2.867 0.004141 **
factor(q3)4       -3.3160304  1.4209972  -2.334 0.019617 *
factor(q3)5       -2.6750755  1.7868612  -1.497 0.134372
factor(q3)6       -4.0853317  1.4637853  -2.791 0.005256 **
factor(q3)7       -2.5793886  1.4483397  -1.781 0.074924 .
factor(q3)8       -2.9539964  1.5137013  -1.952 0.050997 .
factor(q3)9       -2.3762308  1.4112700  -1.684 0.092229 .
factor(q3)10      -4.6413021  1.3959327  -3.325 0.000885 ***
factor(q3)11      -3.3791212  1.4064383  -2.403 0.016279 *
factor(q3)12      -3.5726507  1.4190294  -2.518 0.011813 *
factor(q3)13      -1.1007775  1.7007817  -0.647 0.517491
factor(q3)14      -2.5463636  1.4007548  -1.818 0.069087 .
factor(q3)15      -2.8243291  1.3964057  -2.023 0.043117 *
factor(q3)16      -3.7307116  1.3972777  -2.670 0.007585 **
factor(q3)17      -3.9543060  1.4349890  -2.756 0.005858 **
factor(q3)18      -5.5456206  1.4900729  -3.722 0.000198 ***
factor(q3)19      -4.7039317  1.7470998  -2.692 0.007093 **
factor(q3)24      -3.5900582  1.5169602  -2.367 0.017952 *
q6                -0.0009015  0.0004410  -2.044 0.040942 *
q10               -0.1549530  0.0713357  -2.172 0.029843 *
factor(q20_01_1)2  0.4819075  0.2364915   2.038 0.041576 *
factor(q20_01_1)3  0.6411000  0.2886885   2.221 0.026369 *
factor(q20_01_1)4 -1.3036924  0.4845737  -2.690 0.007137 **
factor(q20_01_1)5 -1.4066257  1.0237434  -1.374 0.169441
q22                0.1163618  0.0257926   4.511 6.44e-06 ***
q23                0.1100196  0.0450363   2.443 0.014569 *
factor(q27_1)2     0.0640050  0.3419154   0.187 0.851507
factor(q27_1)3     0.9976101  0.3357019   2.972 0.002961 **
factor(q27_1)4     2.6922838  0.4006163   6.720 1.81e-11 ***
factor(q27_1)5     4.4761444  0.9611724   4.657 3.21e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1261.72  on 1003  degrees of freedom
Residual deviance:  891.64  on  972  degrees of freedom
AIC: 823.17

Number of Fisher Scoring iterations: 6
```

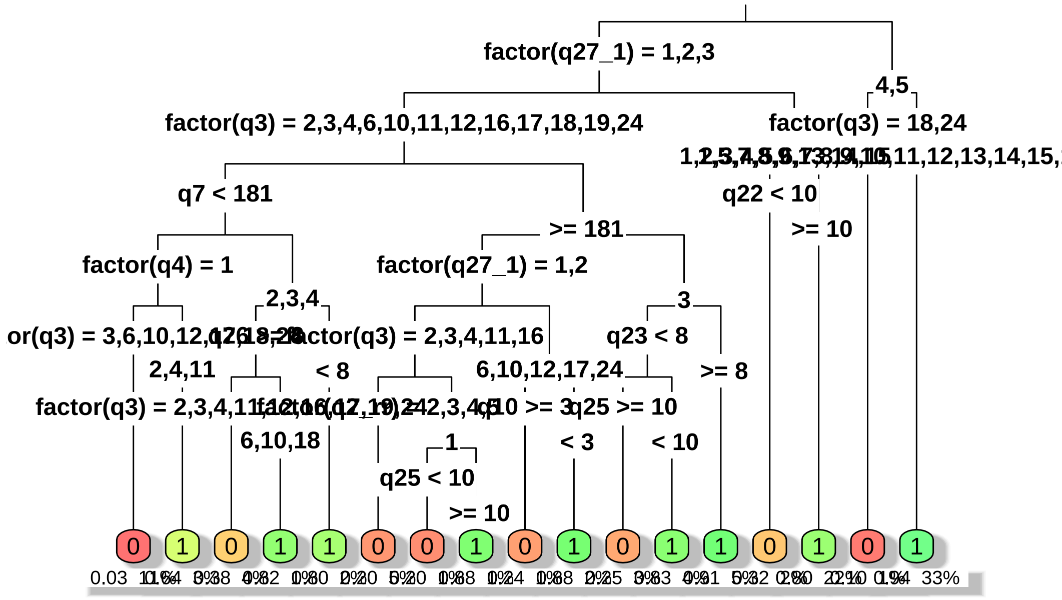
```r
summary(glm_log)
```

```
Call:
glm(formula = factor(q28_YN) ~ factor(q1) + factor(q2_rr) + factor(q3) +
    factor(q4) + q5_1 + q6 + q7 + q10 + factor(q11) + q1719_label +
    factor(q20_01_1) + factor(q20_02_1) + q22 + q23 + q24 + q25 +
```

```
    q26 + factor(q27_1), family = binomial, data = DB.csv, weights = weight)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.159e+00  1.560e+00   0.743 0.457640
factor(q1)2        2.124e-02  2.031e-01   0.105 0.916715
factor(q2_rr)2    -1.994e-01  3.169e-01  -0.629 0.529191
factor(q2_rr)3     1.484e-02  3.402e-01   0.044 0.965220
factor(q2_rr)4     3.438e-01  3.670e-01   0.937 0.348893
factor(q2_rr)5     3.879e-01  3.939e-01   0.985 0.324773
factor(q2_rr)6     4.254e-01  4.632e-01   0.918 0.358369
factor(q3)2       -3.787e+00  1.412e+00  -2.683 0.007300 **
factor(q3)3       -4.221e+00  1.424e+00  -2.964 0.003033 **
factor(q3)4       -3.451e+00  1.443e+00  -2.392 0.016764 *
factor(q3)5       -2.847e+00  1.838e+00  -1.549 0.121321
factor(q3)6       -4.094e+00  1.486e+00  -2.754 0.005883 **
factor(q3)7       -2.874e+00  1.478e+00  -1.944 0.051839 .
factor(q3)8       -3.083e+00  1.543e+00  -1.998 0.045718 *
factor(q3)9       -2.560e+00  1.433e+00  -1.786 0.074121 .
factor(q3)10      -4.708e+00  1.418e+00  -3.321 0.000898 ***
factor(q3)11      -3.535e+00  1.430e+00  -2.472 0.013447 *
factor(q3)12      -3.629e+00  1.450e+00  -2.504 0.012296 *
factor(q3)13      -1.234e+00  1.735e+00  -0.711 0.476914
factor(q3)14      -2.879e+00  1.427e+00  -2.017 0.043646 *
factor(q3)15      -3.043e+00  1.420e+00  -2.142 0.032159 *
factor(q3)16      -3.836e+00  1.418e+00  -2.705 0.006839 **
factor(q3)17      -3.999e+00  1.462e+00  -2.736 0.006217 **
factor(q3)18      -5.922e+00  1.521e+00  -3.893 9.90e-05 ***
factor(q3)19      -4.803e+00  1.779e+00  -2.700 0.006943 **
factor(q3)24      -3.630e+00  1.539e+00  -2.358 0.018375 *
factor(q4)2        2.531e-01  2.937e-01   0.862 0.388904
factor(q4)3        5.722e-02  2.377e-01   0.241 0.809787
factor(q4)4       -1.108e-03  3.662e-01  -0.003 0.997587
q5_1               4.566e-03  5.412e-02   0.084 0.932761
q6                -8.912e-04  4.891e-04  -1.822 0.068433 .
q7                 9.483e-04  6.086e-04   1.558 0.119206
q10               -1.234e-01  8.101e-02  -1.523 0.127705
factor(q11)1      -3.173e-01  4.021e-01  -0.789 0.430041
factor(q11)2      -7.046e-01  6.963e-01  -1.012 0.311533
q1719_label        2.762e-01  3.919e-01   0.705 0.480986
factor(q20_01_1)2  5.214e-01  2.725e-01   1.913 0.055690 .
factor(q20_01_1)3  4.923e-01  3.977e-01   1.238 0.215777
factor(q20_01_1)4 -1.104e+00  5.346e-01  -2.064 0.038979 *
factor(q20_01_1)5 -1.457e+00  1.084e+00  -1.344 0.178961
factor(q20_02_1)2  1.784e-03  3.321e-01   0.005 0.995714
factor(q20_02_1)3  3.907e-01  5.117e-01   0.763 0.445223
factor(q20_02_1)4 -1.051e+00  7.558e-01  -1.391 0.164195
factor(q20_02_1)5  1.140e+01  6.001e+02   0.019 0.984837
q22                1.391e-01  3.466e-02   4.012 6.03e-05 ***
q23                1.190e-01  5.058e-02   2.353 0.018617 *
q24               -1.675e-02  3.689e-02  -0.454 0.649852
q25               -5.686e-03  2.508e-02  -0.227 0.820622
q26               -3.543e-03  4.226e-02  -0.084 0.933177
factor(q27_1)2    -2.557e-02  3.599e-01  -0.071 0.943367
```

```
factor(q27_1)3     9.567e-01  3.512e-01   2.724 0.006450 **
factor(q27_1)4     2.634e+00  4.171e-01   6.315 2.71e-10 ***
factor(q27_1)5     4.458e+00  9.767e-01   4.564 5.02e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1261.72  on 1003  degrees of freedom
Residual deviance:  879.15  on  951  degrees of freedom
AIC: 852.73

Number of Fisher Scoring iterations: 13
```

```r
#Decision tree
library(rpart)
library(rpart.plot)

tree_model <- rpart(
  factor(q28_YN)~
    factor(q1)+
    factor(q2_rr)+
    factor(q3)+
    factor(q4)+
    q5_1+
    q6+
    q7+
    q10+
    factor(q11)+
    q1719_label+
    factor(q20_01_1)+
    factor(q20_02_1)+
    q22+ q23+ q24+ q25+ q26+
    factor(q27_1),data = DB.csv, method = "class",weights = weight)

rpart.plot(tree_model,
           type = 3,
           extra = 106,
           under = TRUE,
           faclen = 0,
           fallen.leaves = TRUE,
           box.palette = "RdYlGn",
           shadow.col = "gray",
           cex = 0.75)
```




```
# glmnet and xgboost
library(glmnet)
x <- model.matrix(factor(q28_YN)~
    factor(q1)+
    factor(q2_rr)+
    factor(q3)+
    factor(q4)+
    q5_1+
    q6+
    q7+
    q10+
    factor(q11)+
    q1719_label+
    factor(q20_01_1)+
    factor(q20_02_1)+
    q22+ q23+ q24+ q25+ q26+
    factor(q27_1), data = DB.csv)[, -1]
y <- as.factor(DB.csv$q28_YN)
glmnet_model <- cv.glmnet(x, y, family = "binomial", alpha = 1)
glmnet_model
```

```
Call:  cv.glmnet(x = x, y = y, family = "binomial", alpha = 1)

Measure: Binomial Deviance

      Lambda Index Measure       SE Nonzero
min 0.006669    31   1.061 0.01955      34
1se 0.029547    15   1.077 0.01715       7
```

```
library(xgboost)
xgb_data <- xgb.DMatrix(data = x, label = as.numeric(y) - 1, weight = DB.csv$weight)
xgb_model <- xgboost(data = xgb_data, objective = "binary:logistic", nrounds = 100)
```

```
[1] train-logloss:0.574666
[2] train-logloss:0.493215
```

```
[3] train-logloss:0.429208
[4] train-logloss:0.389516
[5] train-logloss:0.358326
[6] train-logloss:0.330614
[7] train-logloss:0.314755
[8] train-logloss:0.299092
[9] train-logloss:0.284489
[10]    train-logloss:0.270964
[11]    train-logloss:0.261739
[12]    train-logloss:0.249105
[13]    train-logloss:0.241648
[14]    train-logloss:0.232529
[15]    train-logloss:0.225461
[16]    train-logloss:0.216504
[17]    train-logloss:0.213187
[18]    train-logloss:0.210046
[19]    train-logloss:0.205544
[20]    train-logloss:0.201221
[21]    train-logloss:0.195001
[22]    train-logloss:0.192773
[23]    train-logloss:0.188282
[24]    train-logloss:0.185185
[25]    train-logloss:0.180532
[26]    train-logloss:0.175191
[27]    train-logloss:0.169632
[28]    train-logloss:0.162943
[29]    train-logloss:0.160096
[30]    train-logloss:0.155941
[31]    train-logloss:0.152498
[32]    train-logloss:0.150827
[33]    train-logloss:0.148601
[34]    train-logloss:0.146105
[35]    train-logloss:0.142850
[36]    train-logloss:0.138865
[37]    train-logloss:0.137224
[38]    train-logloss:0.135617
[39]    train-logloss:0.134019
[40]    train-logloss:0.132901
[41]    train-logloss:0.128400
[42]    train-logloss:0.126397
[43]    train-logloss:0.123882
[44]    train-logloss:0.119471
[45]    train-logloss:0.117321
[46]    train-logloss:0.115881
[47]    train-logloss:0.114528
[48]    train-logloss:0.112168
[49]    train-logloss:0.110848
[50]    train-logloss:0.110056
[51]    train-logloss:0.109196
[52]    train-logloss:0.107770
[53]    train-logloss:0.105331
[54]    train-logloss:0.104177
[55]    train-logloss:0.103222
[56]    train-logloss:0.102752
```

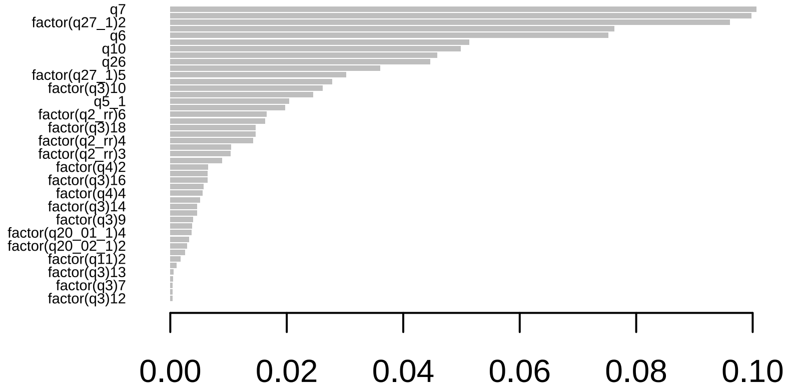
```
[57]    train-logloss:0.100455
[58]    train-logloss:0.099342
[59]    train-logloss:0.097922
[60]    train-logloss:0.096223
[61]    train-logloss:0.095029
[62]    train-logloss:0.093646
[63]    train-logloss:0.091576
[64]    train-logloss:0.088959
[65]    train-logloss:0.087818
[66]    train-logloss:0.087059
[67]    train-logloss:0.085210
[68]    train-logloss:0.083371
[69]    train-logloss:0.081275
[70]    train-logloss:0.079933
[71]    train-logloss:0.079209
[72]    train-logloss:0.078521
[73]    train-logloss:0.078028
[74]    train-logloss:0.076624
[75]    train-logloss:0.074534
[76]    train-logloss:0.072216
[77]    train-logloss:0.070940
[78]    train-logloss:0.070147
[79]    train-logloss:0.069122
[80]    train-logloss:0.068415
[81]    train-logloss:0.067022
[82]    train-logloss:0.065554
[83]    train-logloss:0.064586
[84]    train-logloss:0.063931
[85]    train-logloss:0.063062
[86]    train-logloss:0.061561
[87]    train-logloss:0.060931
[88]    train-logloss:0.060578
[89]    train-logloss:0.059746
[90]    train-logloss:0.058945
[91]    train-logloss:0.058250
[92]    train-logloss:0.057910
[93]    train-logloss:0.057316
[94]    train-logloss:0.056885
[95]    train-logloss:0.056308
[96]    train-logloss:0.055865
[97]    train-logloss:0.054884
[98]    train-logloss:0.053853
[99]    train-logloss:0.052588
[100]   train-logloss:0.052071
```

```r
importance_matrix <- xgb.importance(model = xgb_model)
importance_matrix
```

```
             Feature         Gain        Cover    Frequency
              <char>        <num>        <num>        <num>
 1:               q7 0.1006014818 0.0869011388 0.1144475921
 2:              q22 0.0997652266 0.0879794214 0.1014164306
 3:   factor(q27_1)2 0.0960891679 0.0313989583 0.0147308782
 4:              q24 0.0762530783 0.0696095066 0.0787535411
 5:               q6 0.0751986631 0.1091082918 0.1184135977
```

```
 6:              q25 0.0513265858 0.0610713404 0.0770538244
 7:              q10 0.0498453929 0.0366651821 0.0611898017
 8:              q23 0.0457890618 0.0616189312 0.0657223796
 9:              q26 0.0445914564 0.0415649448 0.0634560907
10:    factor(q27_1)4 0.0360202125 0.0544773962 0.0209631728
11:    factor(q27_1)5 0.0301873688 0.0585528482 0.0158640227
12:    factor(q27_1)3 0.0278206632 0.0141968799 0.0215297450
13:      factor(q3)10 0.0261380393 0.0177619363 0.0073654391
14:      factor(q3)15 0.0245024601 0.0095432797 0.0084985836
15:             q5_1 0.0204062232 0.0202677462 0.0169971671
16:       factor(q4)3 0.0197054761 0.0106482746 0.0300283286
17:    factor(q2_rr)6 0.0165076672 0.0211964073 0.0079320113
18:       factor(q1)2 0.0162385112 0.0116488761 0.0266288952
19:      factor(q3)18 0.0146431542 0.0248320604 0.0062322946
20:      factor(q11)1 0.0146411486 0.0085891738 0.0050991501
21:    factor(q2_rr)4 0.0142578982 0.0049844483 0.0090651558
22:       factor(q3)2 0.0104549392 0.0119665244 0.0096317280
23:    factor(q2_rr)3 0.0103919175 0.0084545116 0.0164305949
24:    factor(q2_rr)2 0.0089299687 0.0079951363 0.0147308782
25:       factor(q4)2 0.0065179947 0.0058619459 0.0090651558
26:    factor(q2_rr)5 0.0064266914 0.0063681736 0.0073654391
27:      factor(q3)16 0.0063710811 0.0065533432 0.0033994334
28: factor(q20_01_1)2 0.0057090976 0.0130485028 0.0084985836
29:       factor(q4)4 0.0055326696 0.0061001275 0.0107648725
30: factor(q20_01_1)5 0.0051199941 0.0005388135 0.0005665722
31:      factor(q3)14 0.0046382541 0.0084937609 0.0033994334
32:       q1719_label 0.0045868711 0.0088767414 0.0062322946
33:       factor(q3)9 0.0039125798 0.0185257093 0.0062322946
34: factor(q20_01_1)3 0.0037028945 0.0114151481 0.0062322946
35: factor(q20_01_1)4 0.0036669837 0.0021630422 0.0011331445
36: factor(q20_02_1)3 0.0031811496 0.0064089829 0.0039660057
37: factor(q20_02_1)2 0.0028894211 0.0019694675 0.0045325779
38:       factor(q3)3 0.0025281345 0.0045682236 0.0033994334
39:      factor(q11)2 0.0017416238 0.0094845525 0.0033994334
40:       factor(q3)4 0.0010497378 0.0028209978 0.0022662890
41:      factor(q3)13 0.0005684266 0.0059322212 0.0016997167
42:      factor(q3)11 0.0004391047 0.0016507497 0.0011331445
43:       factor(q3)7 0.0004077576 0.0027097133 0.0016997167
44:      factor(q3)17 0.0003570660 0.0023370553 0.0011331445
45:      factor(q3)12 0.0003467037 0.0031395131 0.0016997167
             Feature         Gain        Cover    Frequency
```

```r
xgb.plot.importance(importance_matrix)
```




```r
index.q28_1 <- match("q28_1",colnames(DB.csv))
index.q28_3 <- match("q28_3",colnames(DB.csv))
index.q29_1 <- match("q29_1",colnames(DB.csv))
index.q29_3 <- match("q29_3",colnames(DB.csv))

q28.label <- as.factor(apply(
  DB.csv[,index.q28_1:index.q28_3],
  MARGIN = 1,
  function(row){
    return(paste0(row,collapse = ""))
  }))
unique(q28.label)
```

```
[1] 000 100 010 111 110 101 011 001
Levels: 000 001 010 011 100 101 110 111
```

```r
q29.label <- as.factor(apply(
  DB.csv[,index.q29_1:index.q29_3],
  MARGIN = 1,
  function(row){
    return(paste0(row,collapse = ""))
  }))
unique(q29.label)
```

```
[1] 000 111 011 101 010 100 001 110
Levels: 000 001 010 011 100 101 110 111
```

```r
q2829.label <- as.factor(apply(
  DB.csv[,c(index.q28_1:index.q28_3,index.q29_1:index.q29_3)],
  MARGIN = 1,
  function(row){
    return(paste0(row,collapse = ""))
  }))
unique(q2829.label)
```

```
 [1] 000000 100111 010011 111111 110101 110010 010100 110111 110001 010101
[11] 100001 100101 010111 110110 010001 100010 101001 100011 011001 110011
[21] 111110 010010 111100 111011 110100 011011 100100 111101 011010 100110
[31] 010110 001010 111001 011101 011100 001111 101011 111010 001011 001001
[41] 001101 101101 101010 011111
44 Levels: 000000 001001 001010 001011 001101 001111 010001 010010 ... 111111
```

```r
unique(q28.label)
```

```
[1] 000 100 010 111 110 101 011 001
Levels: 000 001 010 011 100 101 110 111
```

```
unique(q29.label)
```

```
[1] 000 111 011 101 010 100 001 110
Levels: 000 001 010 011 100 101 110 111
```

```
unique(q2829.label)
```

```
 [1] 000000 100111 010011 111111 110101 110010 010100 110111 110001 010101
[11] 100001 100101 010111 110110 010001 100010 101001 100011 011001 110011
[21] 111110 010010 111100 111011 110100 011011 100100 111101 011010 100110
[31] 010110 001010 111001 011101 011100 001111 101011 111010 001011 001001
[41] 001101 101101 101010 011111
44 Levels: 000000 001001 001010 001011 001101 001111 010001 010010 ... 111111
```

預期 28 題有選三 (發文等抵制行為) 的抵制程度較高

法一: 1,0 法二: 選項一二合併 vs. 有選三 (11,10,01,00)

第 29 題:

分成: 1 自己, 23 至少選一, 45 至少選一

抵制程度 ~ 其他因素關聯分析

Canonical analysis and PCA

```
library(FactoMineR)
library(factoextra)
boycott <- subset(DB.csv, q28_YN == 1)
# 缺失值轉 0

boycott[,c('q30_1','q32_1','q35_1')] <- lapply(boycott[, c('q30_1','q32_1','q35_1')], as.numeric)
y <-boycott[,c('q30_1','q32_1','q35_1')]


boycott$q28_1_2 <- ifelse(boycott$q28_1==1 | boycott$q28_2==1,1,0)
boycott$q29_1_2_inter<-boycott$q29_1*boycott$q29_2
boycott$q29_1_3_inter<-boycott$q29_1*boycott$q29_3
boycott$q29_2_3_inter<-boycott$q29_3*boycott$q29_2


boycott[, c("q28_1_2","q28_3","q29_1","q29_2","q29_3")] <- lapply(boycott[, c("q28_1_2","q28_3","q29_1",
y <-boycott[,c('q30_1','q32_1','q35_1')]


x <-boycott[,c("q2","q4","q6","q7","q10","q11","q1719_label","q20_01_1","q20_02_1","q22","q23","q24","q2

cca <-cancor(x,y)
# 典型相關係數
cca$cor
```

```
[1] 0.5494695 0.3032716 0.2165462
```

```
# 最大典型相關係數為 0.47，且第一典型變數主要由  q29_3 和 q33_1 和 q35_1 貢獻組成
x_lodings <-cor(x,as.matrix(x)%*% cca$xcoef)
```

```r
y_lodings <-cor(y,as.matrix(y)%*% cca$ycoef)
x_lodings[,c(1,2)]
```

```
                    [,1]        [,2]
q2             -0.3567708  0.06366864
q4              0.3903483  0.19621373
q6              0.2661164 -0.12424438
q7              0.2581490 -0.22226125
q10             0.3477882  0.13376953
q11             0.2730098  0.15786977
q1719_label     0.2702486  0.17157008
q20_01_1        0.1242671  0.38053922
q20_02_1        0.1159655  0.35664324
q22             0.5875827  0.06772881
q23             0.3120231  0.30296822
q24             0.4655742  0.01215483
q25             0.1346612  0.17063059
q26             0.5507047 -0.02770697
q29_1           0.3350883  0.16896076
q29_2          -0.1750344  0.25280512
q29_3           0.3191556 -0.08512301
q31_1           0.3285206 -0.61397509
q33_1           0.5558310  0.06822685
q34_1           0.1193952 -0.40146088
q36_1           0.3705888  0.01899097
q29_1_2_inter   0.1343088  0.28563763
q29_1_3_inter   0.3953651  0.11727137
q40            -0.3481424 -0.31980172
q42_1          -0.3563221 -0.21345380
```

```r
y_lodings[,c(1,2)]
```

```
           [,1]       [,2]
q30_1 0.4999811 -0.1601825
q32_1 0.3001813 -0.9443830
q35_1 0.9637951  0.1623855
```

```r
# 第一典型變數與 q22,q33_1 高度相關,q2(負),q4,q10,q23,q24,q26,q29_1,q29_3,q31_1,q36_1,q29_1_3_inter 中度相
# 第一典型變數與 q35_1 高度相關,q30_1 中度相關
# 越常看到別人在網路上的攻擊行為，抵制行為程度越高。如果認為抵制行為很重要，抵制程度也會比較高。抵制程度與抵制

# 自我相關係數
round((colSums(x_lodings^2)[1:2]/4),4)
```

```
[1] 0.7328 0.3660
```

```r
round((colSums(y_lodings^2)[1:2]/4),4)
```

```
[1] 0.3172 0.2360
```

```r
# 典型相關係數平方
num<-round(cca$cor^2,4)[1:2]

round((colSums(x_lodings^2)[1:2]/4)*num,4)
```

```
[1] 0.2212 0.0337
```

```r
round((colSums(y_lodings^2)[1:2]/4)*num,4)
```

```
[1] 0.0958 0.0217
```

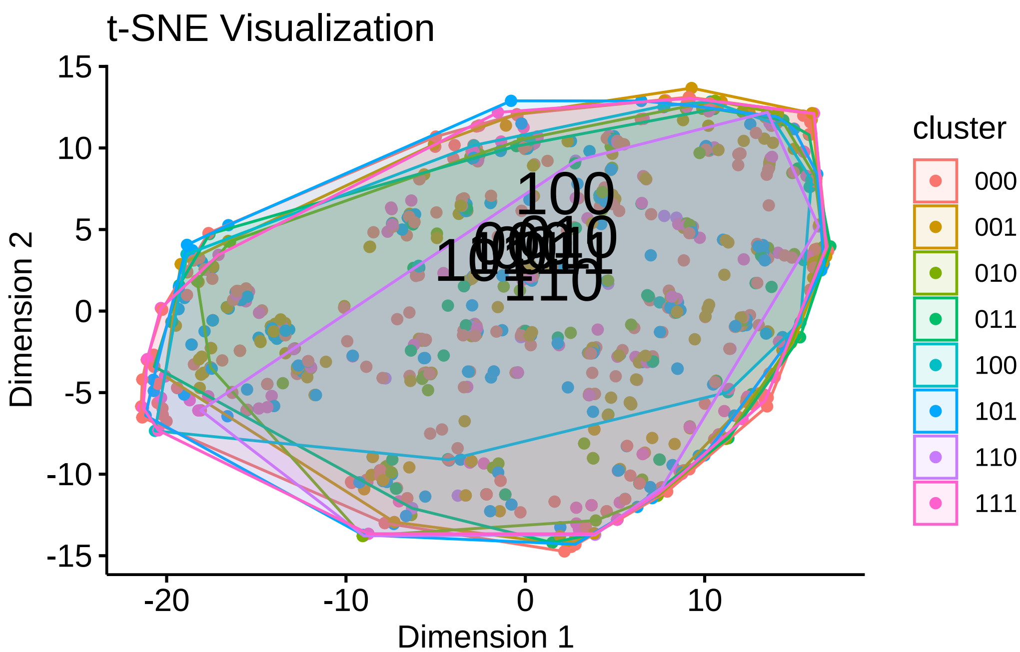
```r
# 第一典型變數能解釋約 9.67% 的預測變數變異、7.42% 的準則變數變異
```

t-SNE visualization

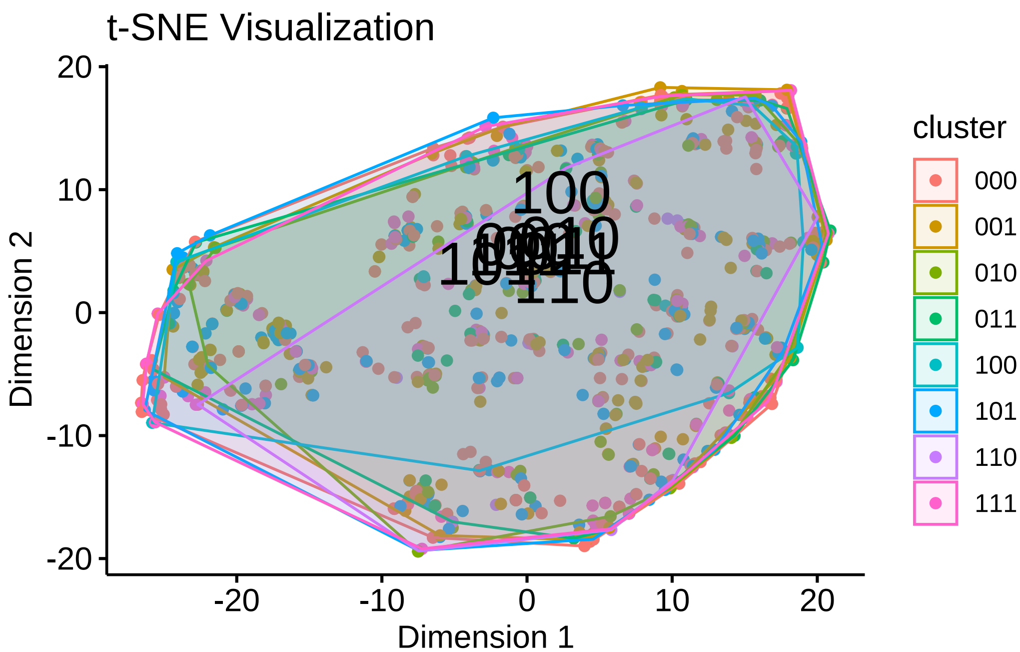
```r
library(Rtsne)
library(ggpubr)
library(magrittr)
tSNE.2d <- function(X, label_vector,seed=123, theta = 0.5, perplexity = 30, size = 1.5, num_threads = 1L
  label <- apply(X[,label_vector],1,paste0, collapse = "")
  X <- X[,-which(colnames(X) %in% label_vector)]
  set.seed(seed)
  tsne_result <-  X%>%
    Rtsne(dims=2,theta=theta,perplexity=perplexity,num_threads=num_threads)%>%
    {.$Y}%>%
    {colnames(.) <- c("V1", "V2"); .} %>%
    as_tibble(.name_repair = "unique")%>%
    mutate(cluster=label)
  pic <- ggscatter(data = tsne_result, x = "V1", y = "V2",
                   color = "cluster",
                   title = "t-SNE Visualization",
                   xlab = "Dimension 1", ylab = "Dimension 2",
                   size = size,
                   ellipse = TRUE,
                   ellipse.type = "convex",
                   repel = TRUE) +
            scale_color_discrete()+
            geom_text(data = tsne_result %>%
                        group_by(cluster) %>%
                        summarize(V1 = mean(V1), V2 = mean(V2), .groups = 'drop'),
                      aes(x = V1, y = V2, label = cluster),
                      vjust = -1,
                      size = 8,
                      color = "black")+
            theme(legend.position = "right")
  return(pic)
}
```

tuning parameters

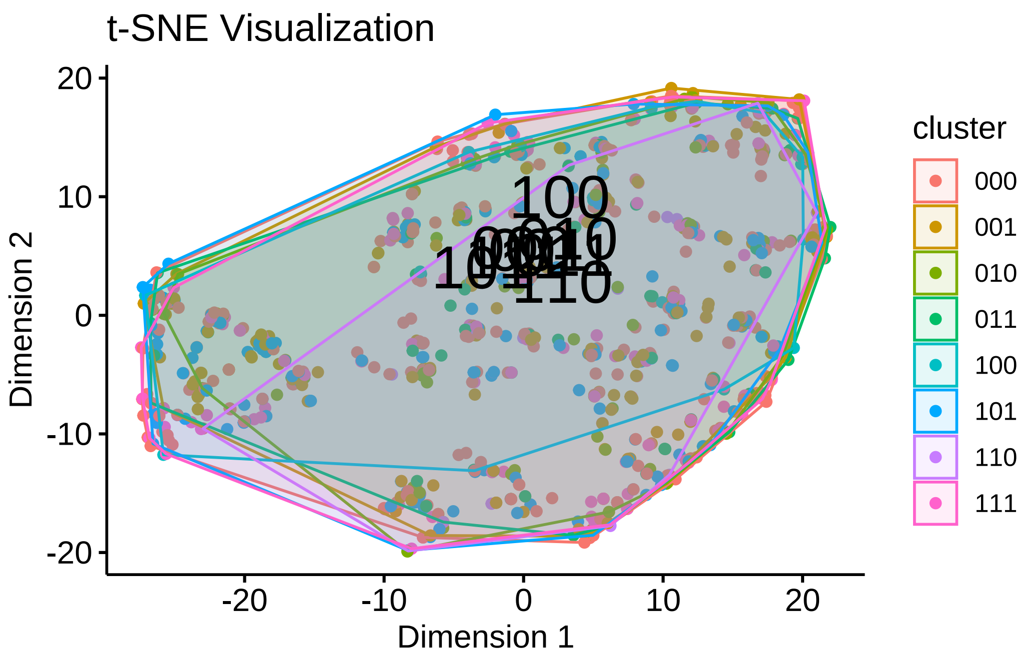
```r
tSNE.2d(X = DB.csv,
        label_vector = c("q29_1", "q29_2", "q29_3"),
        seed = 2012,
        theta = 0.3, perplexity=70,
        size=1.5, num_threads=4L)
```




```
tSNE.2d(X = DB.csv,
        label_vector = c("q29_1", "q29_2", "q29_3"),
        seed = 2012,
        theta = 0.3, perplexity=50,
        size=1.5, num_threads=4L)
```




```
tSNE.2d(X = DB.csv,
        label_vector = c("q29_1", "q29_2", "q29_3"),
        seed = 2012,
        theta = 0.1, perplexity=50,
        size=1.5, num_threads=4L)
```


t-SNE Visualization
20
10
0
-10
-20
Dimension 2
-20
-10
0
10
20
Dimension 1
100
010
001
011
101
111
110
cluster
000
001
010
011
100
101
110
111

try UMAP