

取消文化之現象分析

高嘉好、柯堯城、吳承恩、趙友誠

2024-11-15

Table of contents

report 的重要頁碼	1
處理後資料簡介	1
人口結構變數處理	2
其他變數的更動	2
對抵制行為相關問題變數之處理	5
資料視覺化	8
describe	8
對各變數依 q28_YN 二元變數畫比例圖	8
Logistic and Decision tree and PCA and XGboost	9
Canonical analysis and PCA	18
t-SNE visualization	19
tuning parameters	20

report 的重要頁碼

52: 網路癮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$q37a <- NULL
DB.csv$rrq21a <- NULL
DB.csv$rrq37a <- NULL
DB.csv$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","q10_88")],
                    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$q20 <- rowSums(DB.csv[,c('q20_01_1', 'q20_02_1')])
DB.csv$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("q20_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 0～100 → 1～5

41 題: 民進黨偏好 0～100 → 1～5

42 題: 反台獨程度 1～10

```
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，對疫情走向發出差誤言論（去年康健發文說嬰幼兒不會染疫，被我指正，卻不改",
            " 假名人之姿發表利己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('q1','q2','q2_rr','q3','q4','q5_1','q6','q7','q10','q11','q17_01','q17_02','q19_01',
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	

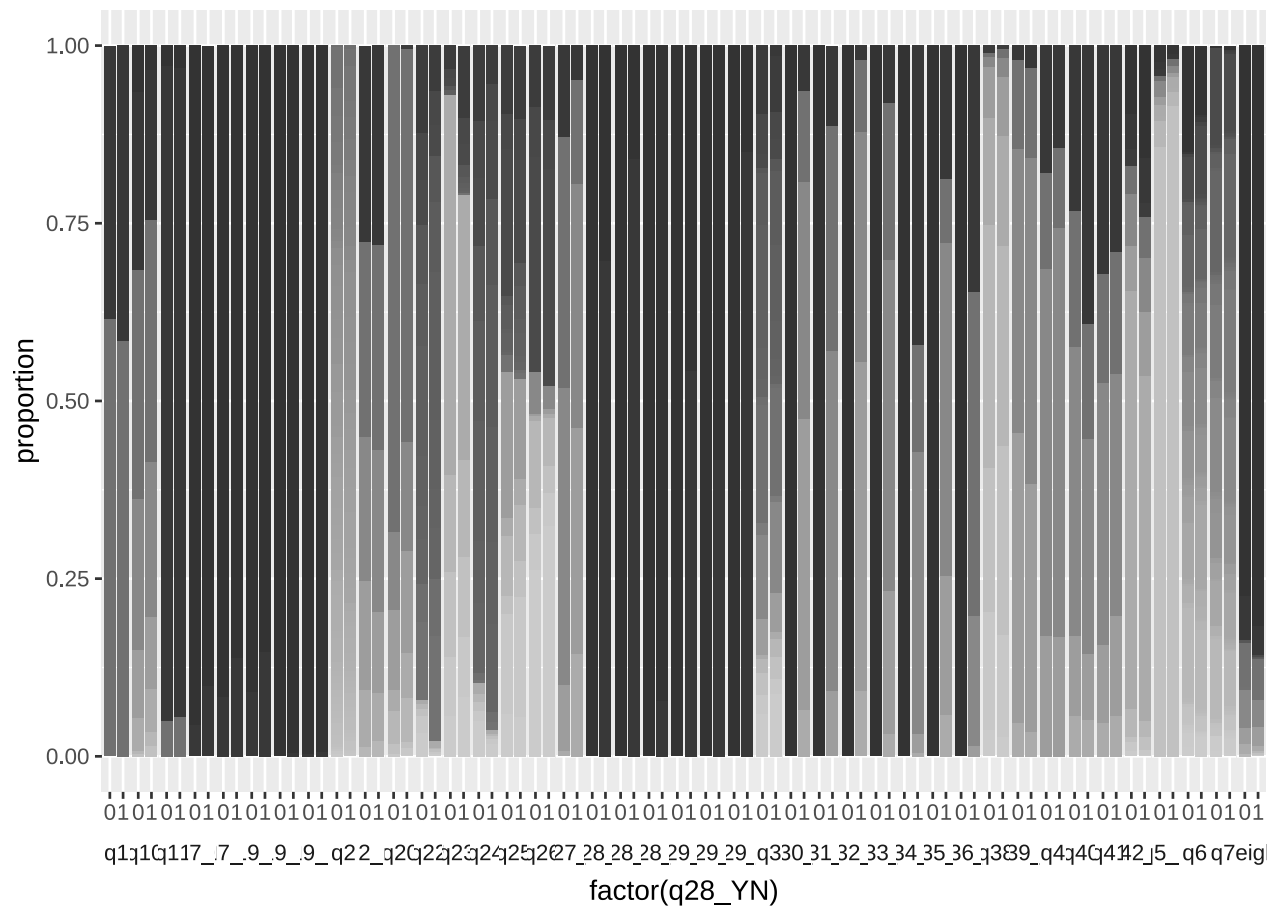
資料視覺化

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+
    q20+
    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(q1) + factor(q2_rr) + factor(q3) +
     q6 + q7 + q10 + q22 + q26 + factor(q27_1), family = binomial,
```

```
data = DB.csv, weights = weight)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.797e+01	2.054e+03	0.009	0.993017	
factor(q1)2	5.861e-01	3.324e-01	1.763	0.077852	.
factor(q2_rr)2	-2.536e+00	6.831e-01	-3.713	0.000205	***
factor(q2_rr)3	-2.015e+00	6.550e-01	-3.076	0.002099	**
factor(q2_rr)4	-1.203e+00	6.517e-01	-1.846	0.064881	.
factor(q2_rr)5	-1.332e+00	6.865e-01	-1.940	0.052336	.
factor(q2_rr)6	-8.401e-01	6.759e-01	-1.243	0.213895	
factor(q3)2	-1.971e+01	2.054e+03	-0.010	0.992341	
factor(q3)3	-2.090e+01	2.054e+03	-0.010	0.991878	
factor(q3)4	-1.937e+01	2.054e+03	-0.009	0.992475	
factor(q3)5	-2.523e+00	2.839e+03	-0.001	0.999291	
factor(q3)6	-1.987e+01	2.054e+03	-0.010	0.992280	
factor(q3)7	-1.763e+01	2.054e+03	-0.009	0.993149	
factor(q3)8	-1.278e+00	3.100e+03	0.000	0.999671	
factor(q3)9	-1.769e+01	2.054e+03	-0.009	0.993128	
factor(q3)10	-2.145e+01	2.054e+03	-0.010	0.991668	
factor(q3)11	-1.980e+01	2.054e+03	-0.010	0.992306	
factor(q3)12	-2.001e+01	2.054e+03	-0.010	0.992225	
factor(q3)13	3.928e-01	3.844e+03	0.000	0.999918	
factor(q3)14	-1.858e+01	2.054e+03	-0.009	0.992780	
factor(q3)15	-1.837e+01	2.054e+03	-0.009	0.992863	
factor(q3)16	-2.063e+01	2.054e+03	-0.010	0.991986	
factor(q3)17	-2.001e+01	2.054e+03	-0.010	0.992226	
factor(q3)18	-2.355e+01	2.054e+03	-0.011	0.990851	
factor(q3)19	-2.046e+01	2.054e+03	-0.010	0.992052	
factor(q3)24	-1.982e+01	2.054e+03	-0.010	0.992299	
q6	-1.653e-03	8.773e-04	-1.884	0.059504	.
q7	2.244e-03	1.027e-03	2.186	0.028814	*
q10	-2.701e-01	1.350e-01	-2.001	0.045401	*
q22	1.390e-01	5.462e-02	2.544	0.010950	*
q26	1.280e-01	6.388e-02	2.004	0.045045	*
factor(q27_1)2	-2.264e-01	5.863e-01	-0.386	0.699401	
factor(q27_1)3	1.325e+00	5.683e-01	2.331	0.019763	*
factor(q27_1)4	2.784e+00	6.201e-01	4.489	7.16e-06	***
factor(q27_1)5	1.872e+01	1.201e+03	0.016	0.987567	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 725.89 on 196 degrees of freedom
Residual deviance: 391.65 on 162 degrees of freedom
AIC: 461.65

Number of Fisher Scoring iterations: 17

```
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 +
q20 + q22 + q23 + q24 + q25 + q26 + factor(q27_1), family = binomial,
data = DB.csv, weights = weight)

```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.758e+01	2.032e+03	0.009	0.99310
factor(q1)2	7.702e-01	3.528e-01	2.183	0.02901 *
factor(q2_rr)2	-2.292e+00	7.623e-01	-3.007	0.00264 **
factor(q2_rr)3	-1.872e+00	7.413e-01	-2.526	0.01154 *
factor(q2_rr)4	-9.433e-01	7.676e-01	-1.229	0.21911
factor(q2_rr)5	-1.110e+00	7.929e-01	-1.400	0.16146
factor(q2_rr)6	-1.005e+00	8.192e-01	-1.227	0.21994
factor(q3)2	-1.952e+01	2.032e+03	-0.010	0.99234
factor(q3)3	-2.083e+01	2.032e+03	-0.010	0.99182
factor(q3)4	-1.923e+01	2.032e+03	-0.009	0.99245
factor(q3)5	-2.449e+00	2.815e+03	-0.001	0.99931
factor(q3)6	-1.968e+01	2.032e+03	-0.010	0.99227
factor(q3)7	-1.795e+01	2.032e+03	-0.009	0.99295
factor(q3)8	-1.119e+00	3.082e+03	0.000	0.99971
factor(q3)9	-1.757e+01	2.032e+03	-0.009	0.99310
factor(q3)10	-2.113e+01	2.032e+03	-0.010	0.99170
factor(q3)11	-1.981e+01	2.032e+03	-0.010	0.99222
factor(q3)12	-1.947e+01	2.032e+03	-0.010	0.99235
factor(q3)13	5.614e-01	3.838e+03	0.000	0.99988
factor(q3)14	-1.837e+01	2.032e+03	-0.009	0.99279
factor(q3)15	-1.838e+01	2.032e+03	-0.009	0.99278
factor(q3)16	-2.061e+01	2.032e+03	-0.010	0.99191
factor(q3)17	-1.971e+01	2.032e+03	-0.010	0.99226
factor(q3)18	-2.353e+01	2.032e+03	-0.012	0.99076
factor(q3)19	-2.055e+01	2.032e+03	-0.010	0.99193
factor(q3)24	-1.974e+01	2.032e+03	-0.010	0.99225
factor(q4)2	6.319e-01	6.312e-01	1.001	0.31677
factor(q4)3	3.666e-01	7.891e-01	0.465	0.64220
factor(q4)4	1.745e+01	2.396e+03	0.007	0.99419
q5_1	5.371e-02	7.666e-02	0.701	0.48359
q6	-1.624e-03	9.530e-04	-1.704	0.08838 .
q7	2.052e-03	1.060e-03	1.936	0.05290 .
q10	-2.972e-01	1.471e-01	-2.021	0.04331 *
factor(q11)1	-4.335e-01	6.887e-01	-0.629	0.52907
factor(q11)2	1.621e+01	3.608e+03	0.004	0.99642
q1719_label	6.929e-03	8.330e-01	0.008	0.99336
q20	1.506e-04	1.113e-01	0.001	0.99892
q22	1.153e-01	6.242e-02	1.847	0.06479 .
q23	2.843e-02	8.976e-02	0.317	0.75144
q24	-9.947e-03	6.527e-02	-0.152	0.87888
q25	2.617e-02	4.389e-02	0.596	0.55100
q26	1.358e-01	6.951e-02	1.954	0.05069 .
factor(q27_1)2	-2.017e-01	6.323e-01	-0.319	0.74970
factor(q27_1)3	1.409e+00	6.119e-01	2.302	0.02132 *
factor(q27_1)4	2.838e+00	6.616e-01	4.290	1.79e-05 ***
factor(q27_1)5	1.883e+01	1.190e+03	0.016	0.98737

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

(Dispersion parameter for binomial family taken to be 1)

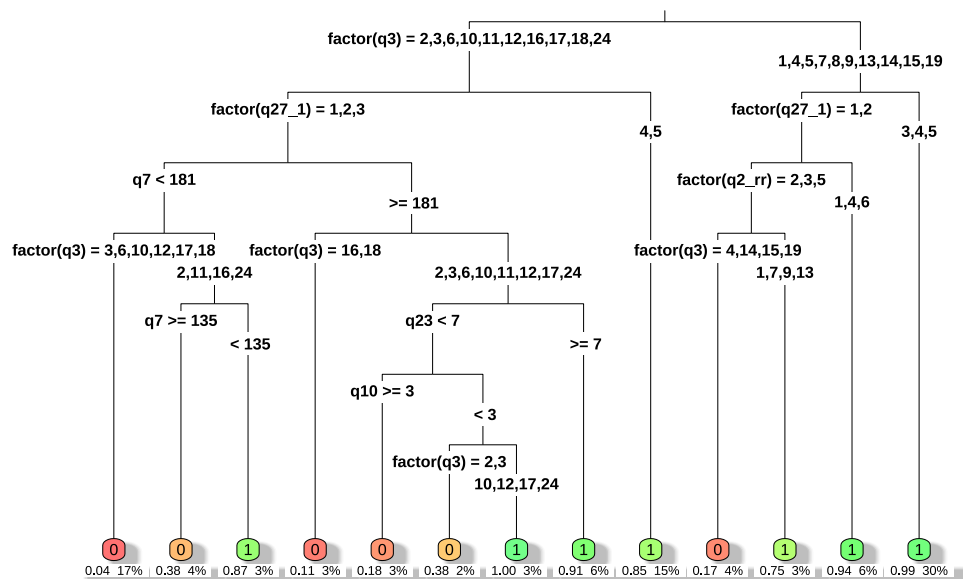
Null deviance: 725.89 on 196 degrees of freedom
Residual deviance: 385.64 on 151 degrees of freedom
AIC: 477.64

Number of Fisher Scoring iterations: 17

```
#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+
    q20+
    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 = 0.5)
```



```
# 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+
  q20+
  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	0.00667	31	1.054	0.03094	31
1se	0.03243	14	1.081	0.02689	6

```
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:0.539179
[2] train-logloss:0.446868
[3] train-logloss:0.356501
```

[4] train-logloss:0.312915
[5] train-logloss:0.260324
[6] train-logloss:0.227397
[7] train-logloss:0.204813
[8] train-logloss:0.183880
[9] train-logloss:0.167727
[10] train-logloss:0.147810
[11] train-logloss:0.141517
[12] train-logloss:0.129310
[13] train-logloss:0.115569
[14] train-logloss:0.106285
[15] train-logloss:0.096666
[16] train-logloss:0.090032
[17] train-logloss:0.079723
[18] train-logloss:0.074775
[19] train-logloss:0.070098
[20] train-logloss:0.065889
[21] train-logloss:0.061435
[22] train-logloss:0.058541
[23] train-logloss:0.055210
[24] train-logloss:0.051497
[25] train-logloss:0.048502
[26] train-logloss:0.046019
[27] train-logloss:0.043975
[28] train-logloss:0.042014
[29] train-logloss:0.040287
[30] train-logloss:0.038859
[31] train-logloss:0.037325
[32] train-logloss:0.035684
[33] train-logloss:0.034608
[34] train-logloss:0.033543
[35] train-logloss:0.031997
[36] train-logloss:0.030778
[37] train-logloss:0.030024
[38] train-logloss:0.029081
[39] train-logloss:0.028285
[40] train-logloss:0.027309
[41] train-logloss:0.026686
[42] train-logloss:0.025934
[43] train-logloss:0.025354
[44] train-logloss:0.024719
[45] train-logloss:0.023937
[46] train-logloss:0.023463
[47] train-logloss:0.022957
[48] train-logloss:0.022582
[49] train-logloss:0.022220
[50] train-logloss:0.021586
[51] train-logloss:0.021112
[52] train-logloss:0.020739
[53] train-logloss:0.020360
[54] train-logloss:0.019965
[55] train-logloss:0.019579
[56] train-logloss:0.019236
[57] train-logloss:0.018902

```

[58] train-logloss:0.018626
[59] train-logloss:0.018236
[60] train-logloss:0.017910
[61] train-logloss:0.017653
[62] train-logloss:0.017423
[63] train-logloss:0.017136
[64] train-logloss:0.016947
[65] train-logloss:0.016656
[66] train-logloss:0.016348
[67] train-logloss:0.016209
[68] train-logloss:0.016001
[69] train-logloss:0.015843
[70] train-logloss:0.015702
[71] train-logloss:0.015487
[72] train-logloss:0.015368
[73] train-logloss:0.015200
[74] train-logloss:0.015045
[75] train-logloss:0.014945
[76] train-logloss:0.014790
[77] train-logloss:0.014693
[78] train-logloss:0.014562
[79] train-logloss:0.014471
[80] train-logloss:0.014337
[81] train-logloss:0.014155
[82] train-logloss:0.013982
[83] train-logloss:0.013878
[84] train-logloss:0.013770
[85] train-logloss:0.013673
[86] train-logloss:0.013563
[87] train-logloss:0.013418
[88] train-logloss:0.013293
[89] train-logloss:0.013200
[90] train-logloss:0.013131
[91] train-logloss:0.013036
[92] train-logloss:0.012955
[93] train-logloss:0.012871
[94] train-logloss:0.012763
[95] train-logloss:0.012670
[96] train-logloss:0.012606
[97] train-logloss:0.012514
[98] train-logloss:0.012432
[99] train-logloss:0.012370
[100] train-logloss:0.012311

```

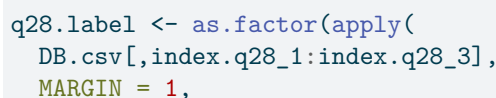
```

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

```

	Feature <char>	Gain <num>	Cover <num>	Frequency <num>
1:	factor(q27_1)2	0.1128759012	0.0590628213	0.032407407
2:	q7	0.0977061430	0.0702257340	0.125000000
3:	factor(q3)10	0.0936159549	0.0567736714	0.019675926
4:	q25	0.0824015984	0.0910425124	0.127314815
5:	q22	0.0692224553	0.0829205730	0.087962963
6:	q23	0.0572483615	0.0517309947	0.060185185

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




```
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
```

```
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
```

```
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
```

```
table(q28.label)
```

```
q28.label
000 001 010 011 100 101 110 111
301  8 195  10  98  6 355  31
```

```
table(q29.label)
```

```
q29.label
000 001 010 011 100 101 110 111
301 189  63  70  24 197  18 142
```

```
table(q2829.label)
```

```
q2829.label
000000 001001 001010 001011 001101 001111 010001 010010 010011 010100 010101
  301      2      3      1      1      1      68      31      24      9      35
010110 010111 011001 011010 011011 011100 011101 011111 100001 100010 100011
   5     23     2     2     2     1     2     1     37     10     5
100100 100101 100110 100111 101001 101010 101011 101101 110001 110010 110011
   8     21     3     14     2     1     2     1     74     15     35
110100 110101 110110 110111 111001 111010 111011 111100 111101 111110 111111
   5    133     6    87     4     1     1     1     4     4     16
```

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","q22","q23","q24","q25","q26","q29_1",

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

```
[1] 0.5484909 0.3032351 0.2136809
```

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

	[,1]	[,2]
q2	-0.3584664	0.06572691
q4	0.3911103	0.19573708
q6	0.2667750	-0.12495481
q7	0.2588039	-0.22304586
q10	0.3492110	0.13223616
q11	0.2739053	0.15702499
q1719_label	0.2710477	0.17087136
q20	0.1326107	0.40533758
q22	0.5888084	0.06673522
q23	0.3130277	0.30211707
q24	0.4663135	0.01167462

```

q25          0.1352137  0.17011910
q26          0.5519676 -0.02886869
q29_1        0.3359433  0.16826048
q29_2       -0.1750348  0.25276601
q29_3        0.3191000 -0.08470257
q31_1        0.3283443 -0.61374395
q33_1        0.5565760  0.06789128
q34_1        0.1197939 -0.40217525
q36_1        0.3712339  0.01853303
q29_1_2_inter 0.1347307  0.28539696
q29_1_3_inter 0.3961883  0.11665606
q40         -0.3486927 -0.31966715
q42_1       -0.3573373 -0.21257935

```

```
y_loadings[,c(1,2)]
```

```

      [,1]      [,2]
q30_1 0.4960180 -0.1574662
q32_1 0.2988692 -0.9441093
q35_1 0.9648598  0.1609171

```

```

# 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 中度相關
# 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。抵制程度與抵制力

```

```

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

```

```
[1] 0.7330 0.3386
```

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

```
[1] 0.3166 0.2355
```

```

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

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

```

```
[1] 0.2205 0.0312
```

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

```
[1] 0.0952 0.0217
```

```
# 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

```

library(Rtsne)
library(ggpubr)
library(magrittr)
tSNE.2d <- function(X, label_vector, seed=123, theta = 0.5, perplexity = 30, size = 1.5, num_threads = 1) {
  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)
}

```

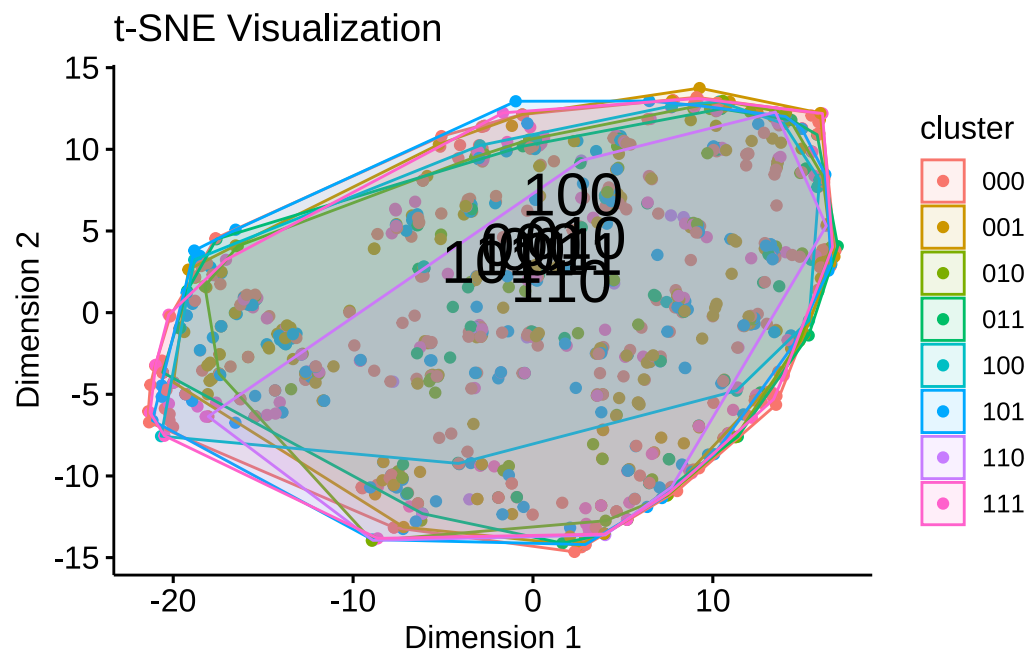
tuning parameters

參數待微調

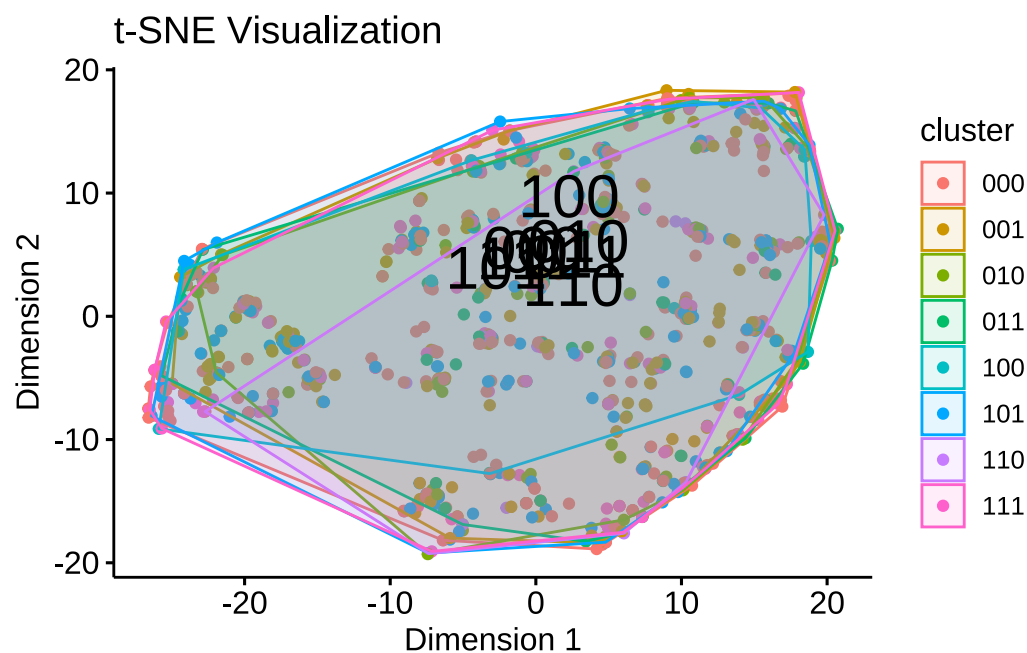
```

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)
```

