世代之爭爭什麼？

從探索的角度
發掘以問卷調查資料
進行意義探勘的潛力

劉正山

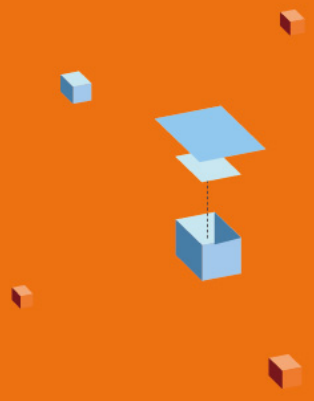
中山大學政治學研究所 教授

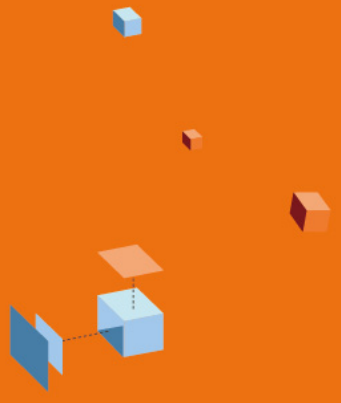
Director, Smilepoll.tw

中研院 調查研究專題中心

1

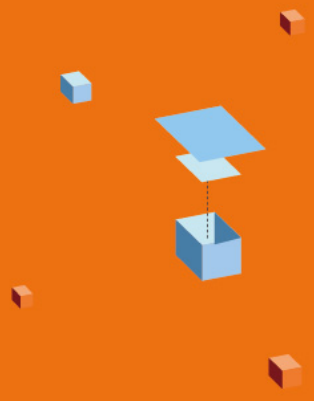
話說...
一年多前的自我對話

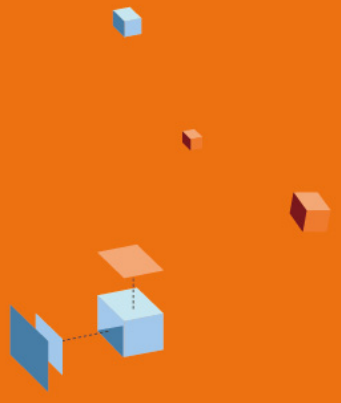




2

研究問題與探索





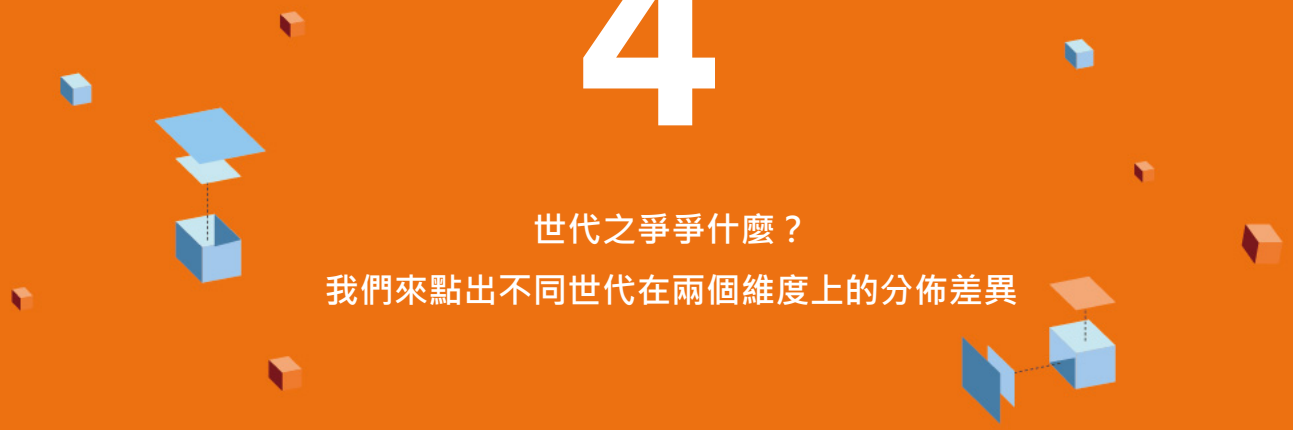
3

MCA 方法帶來的新視野

4

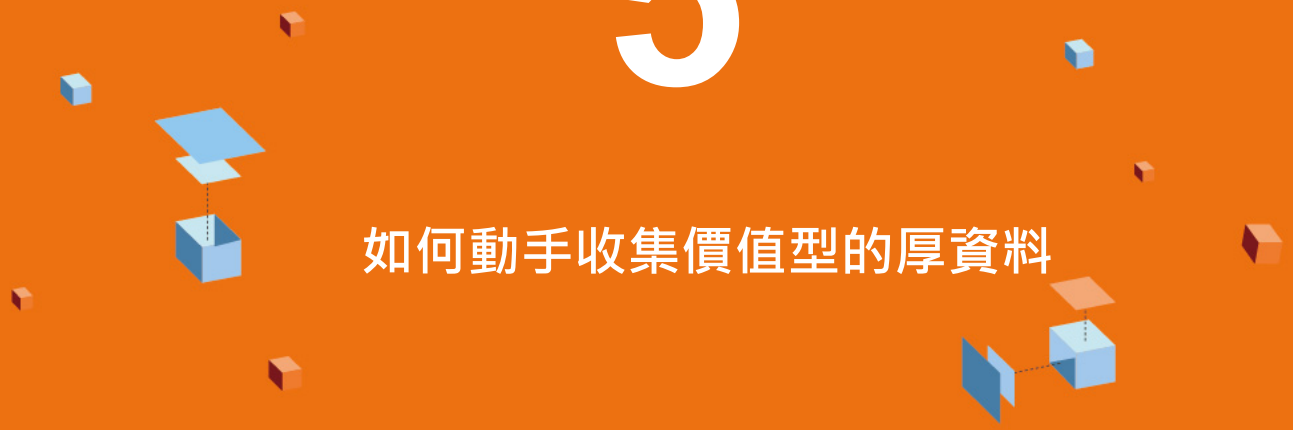
世代之爭爭什麼？

我們來點出不同世代在兩個維度上的分佈差異



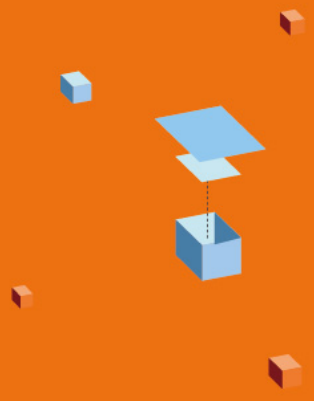
5

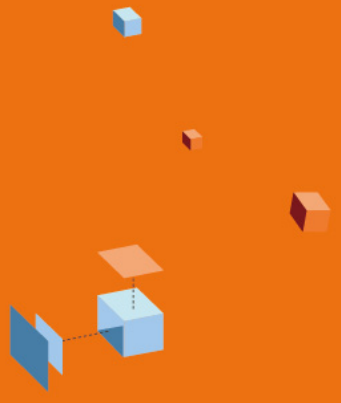
如何動手收集價值型的厚資料



1

話說...
一年多前的自我對話





大數據分析的探索精神，小數據的擁有者沒有嗎？

(OF COURSE YES; WE HAVE IT.)

做實證的社會科學家，理應也能做做不同於描述和假設檢定的事。

調查資料正在貶值中 ?!

- 一般市場與民意調查只被拿來做簡單描述分析；在學術界則被拿來作理論與假設的檢定。
- 隨機抽樣的樣本，獲取成本很高（面訪>電訪）；
- 商業上的焦點團體與立意抽樣等方法，因為樣本少而和大數據相形失色。
- 問卷題（多是類別型變數）看似只能做做描述統計或兩兩之間的相關分析；技術含金量有限。

March 2016. Google watched how people use a phone in a van for over an hour at a time. Goal: complete interviewing 500 people.

WELL...

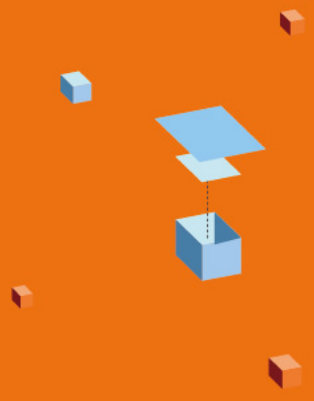
Google 拿質性訪問來確認大數據中看見的樣貌。但這並不算是正視問卷調查資料用於意義開發的潛力。

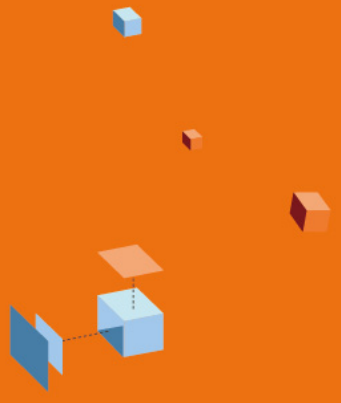
只要問了好問題，並運用探索工具MCA，民調市調資料與大數據同樣珍貴。

我們需要有能讓資料分析者發從調查資料中掘出消費者、選民的價值和偏好組合的探索工具。

2

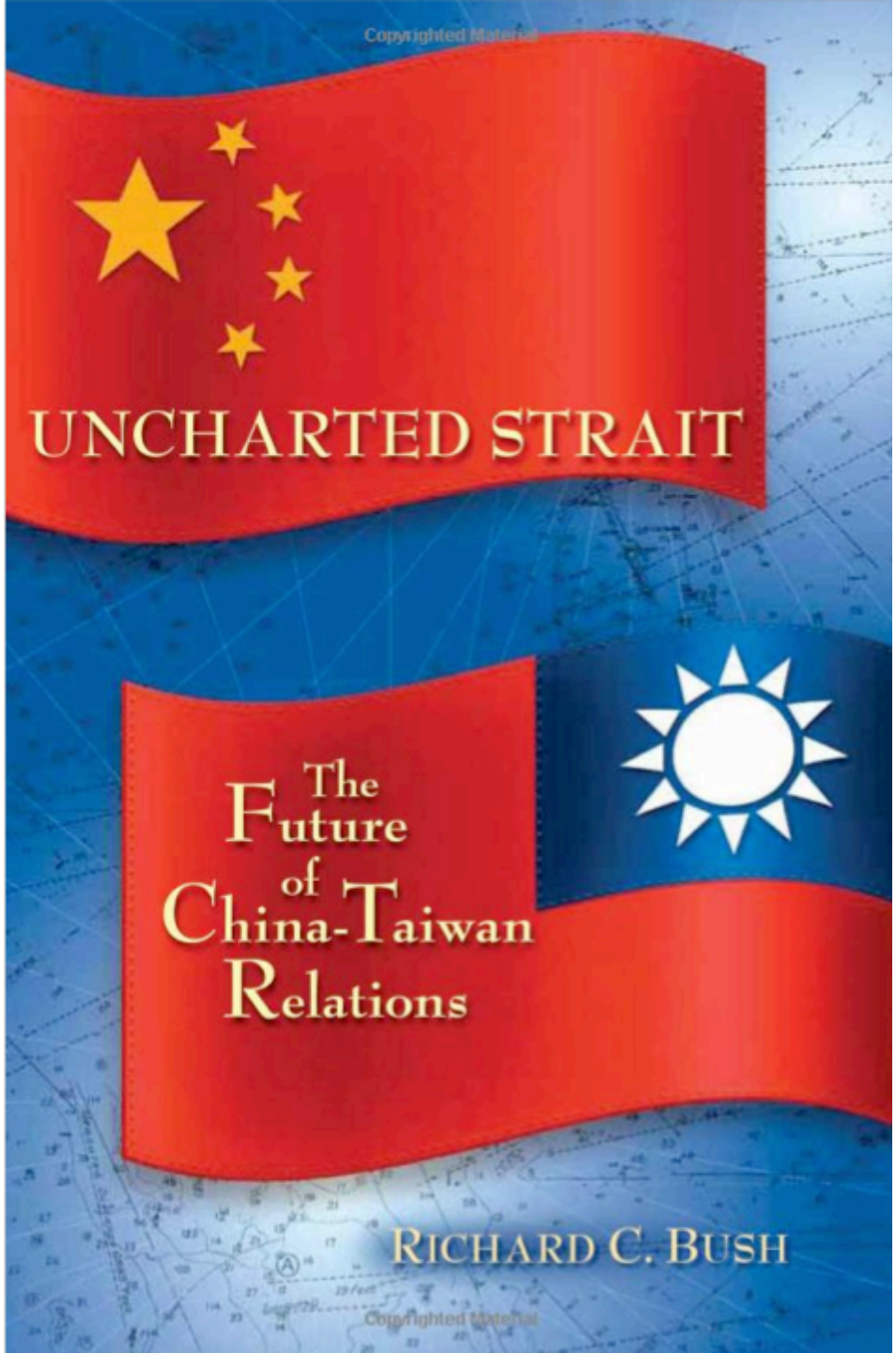
研究問題與探索





目前常用的政治認同的概念與測量

- 國家：國號台灣或中華民國？
- 民族/族群: 台灣人、中國人、都是？
- 兩岸的未來：統一、台灣獨立、維持現狀？
- 還有許多相關的測量題，例如條件統獨、疆域、歷史記憶等


Copyrighted Material
UNCHARTED STRAIT
The Future of China-Taiwan Relations
RICHARD C. BUSH
Copyrighted Material

台灣（已）是民族-國家？

概念與測量的瓶頸：民族？國家？nation-state?

我們在測量民族與國家的時候，手上的測量工具是準確的嗎？

我們知道世代之間的政治傾向上不同，到底怎麼呈現出來才能解答那些是真相，那些是誤解？

我們用於釐清概念與測量之間連結的努力，其實還不夠。

看似客觀的研究但充滿了主觀的預設

例如：如何測量國家認同？

「台灣人/中國人/都是」還是「統/獨/維持現狀」還是...

研究目的

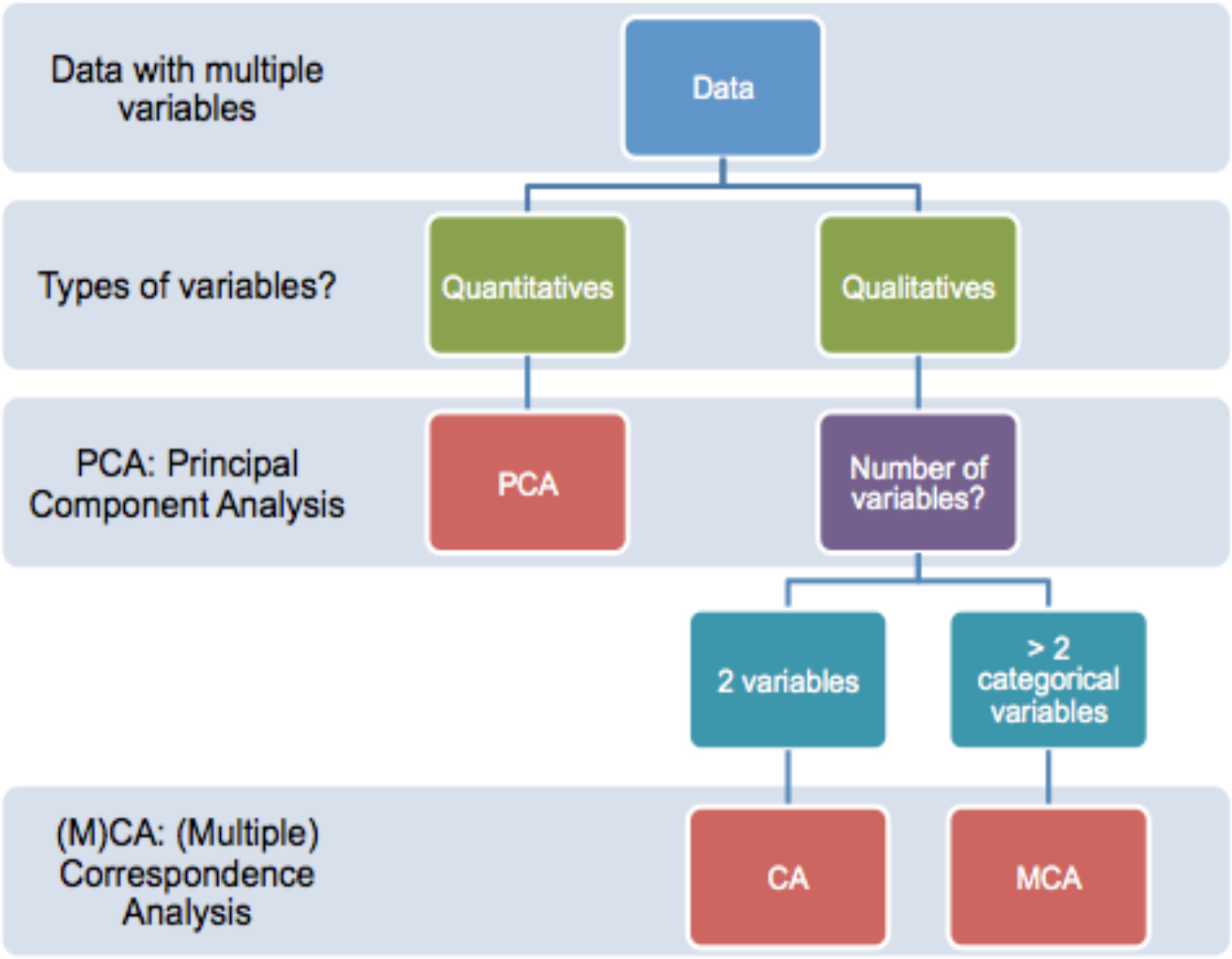
- 由下而上「探索」測量題背後的概念
- 試著釐清這些測量題背後的概念
- 透過這些概念重新檢視我們的選民認同分佈
- 本研究先選擇年齡/世代來觀察

3

MCA 方法帶來的新視野

多重對應分析

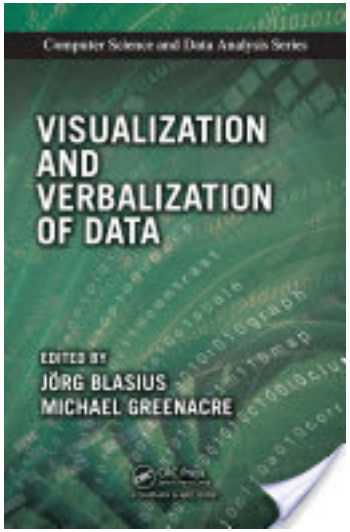
- Multiple Correspondence Analysis（MCA）早在二戰前就出現在歐洲，但其潛力目前尚未受到社會科學的重視。2000左右介紹進美國之後，已經應用在**語言學**的研究中，成為該學門中的重要研究方法（Glynn, et al., 2014; Glynn, & Robinson, 2014）。商管學門也已在使用，但並未在國內形成氣候。
- 最近五年則因為R語言及套件的開發，使這個由法國學者為開發主力的方法經由專書及多個套件的出版得以在全球資料分析者之間傳開。

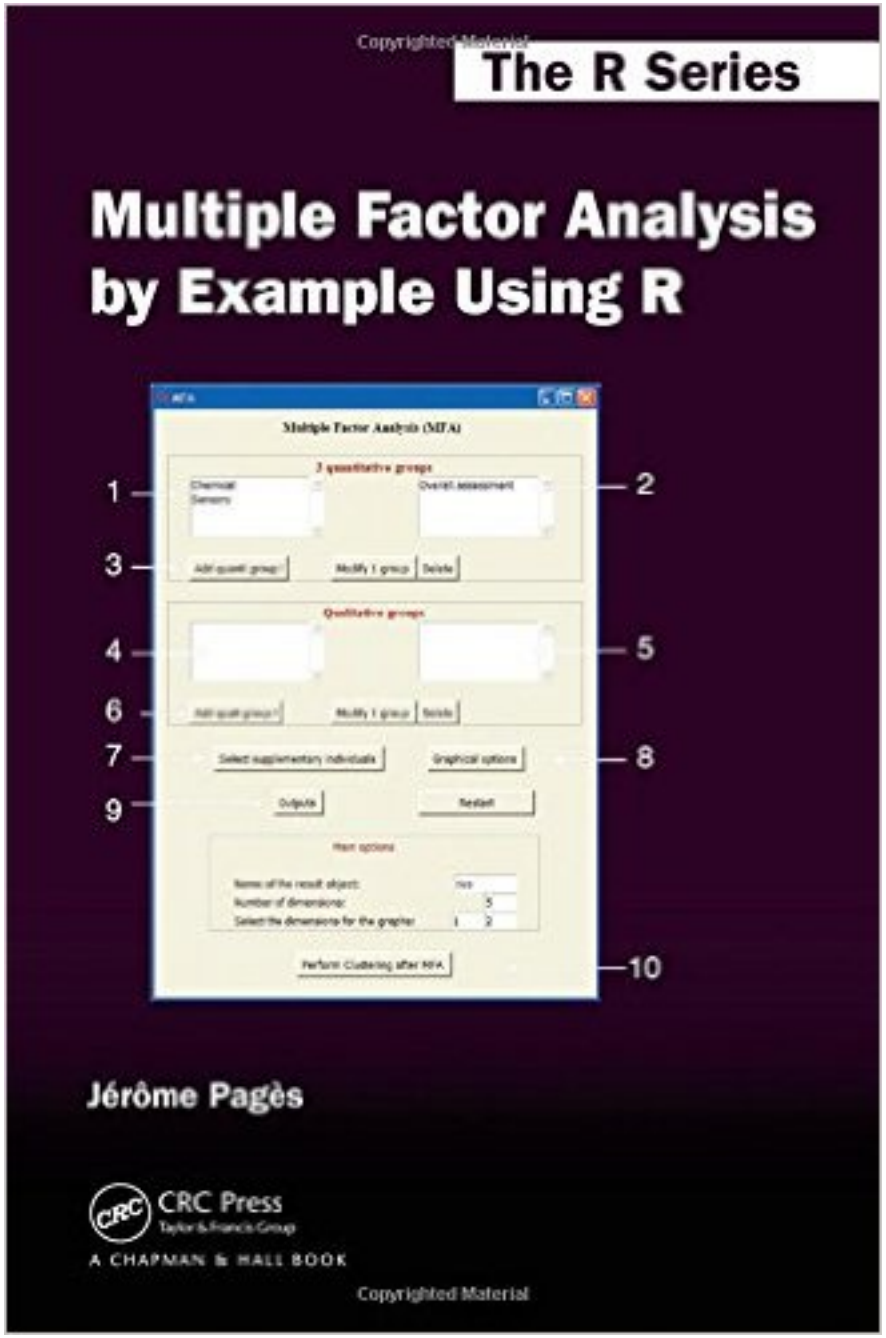


R packages for the analyses: **FactoMineR** (PCA, CA, MCA); **ade4** (PCA, CA, MCA); **stats** (PCA); **ca** (CA); **MASS** (CA)

Use **factoextra** to easily extract and visualize the results


MULTIPLE
CORRESPONDENCE
ANALYSIS
Brigitte LeRoux
Henry Rouanet
Series: Quantitative Applications
in the Social Sciences
163
SAGE


Computer Science and Data Analysis Series
VISUALIZATION
AND
VERBALIZATION
OF DATA
EDITED BY
JÖRG BLASIUS
MICHAEL GREENACRE


Copyrighted Material
The R Series
Multiple Factor Analysis
by Example Using R
Jérôme Pagès
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK
Copyrighted Material

MCA最特別的地方

讓問卷題的分析可以像因素分析一樣，選項之間的關係（不只有題目之間的關係！）可以重新整併出樣貌。

運用MCA於研究、行銷、服務

- 在更短時間內掌握民眾的行為圖像；
- 發掘出資料背後更豐富的意義
- 若大數據分析或大小數據一起來，如虎添翼。

資料與方法

- 中央研究院社會學研究所執行收集的面訪資料：傅仰止、章英華、杜素豪、廖培珊主持的「台灣社會變遷基本調查計畫第六期第四次：國家認同組」。
- 面訪調查於2013年9月22日至12月10日執行，於2014年2月釋出，N=1,952。[有代表性！]
- 這筆資料包含了當前學界所認可的國家認同測量題，如「台灣人/中國人認同」、兩岸關係偏好，亦包含了民族認同題組、條件統獨題組等。

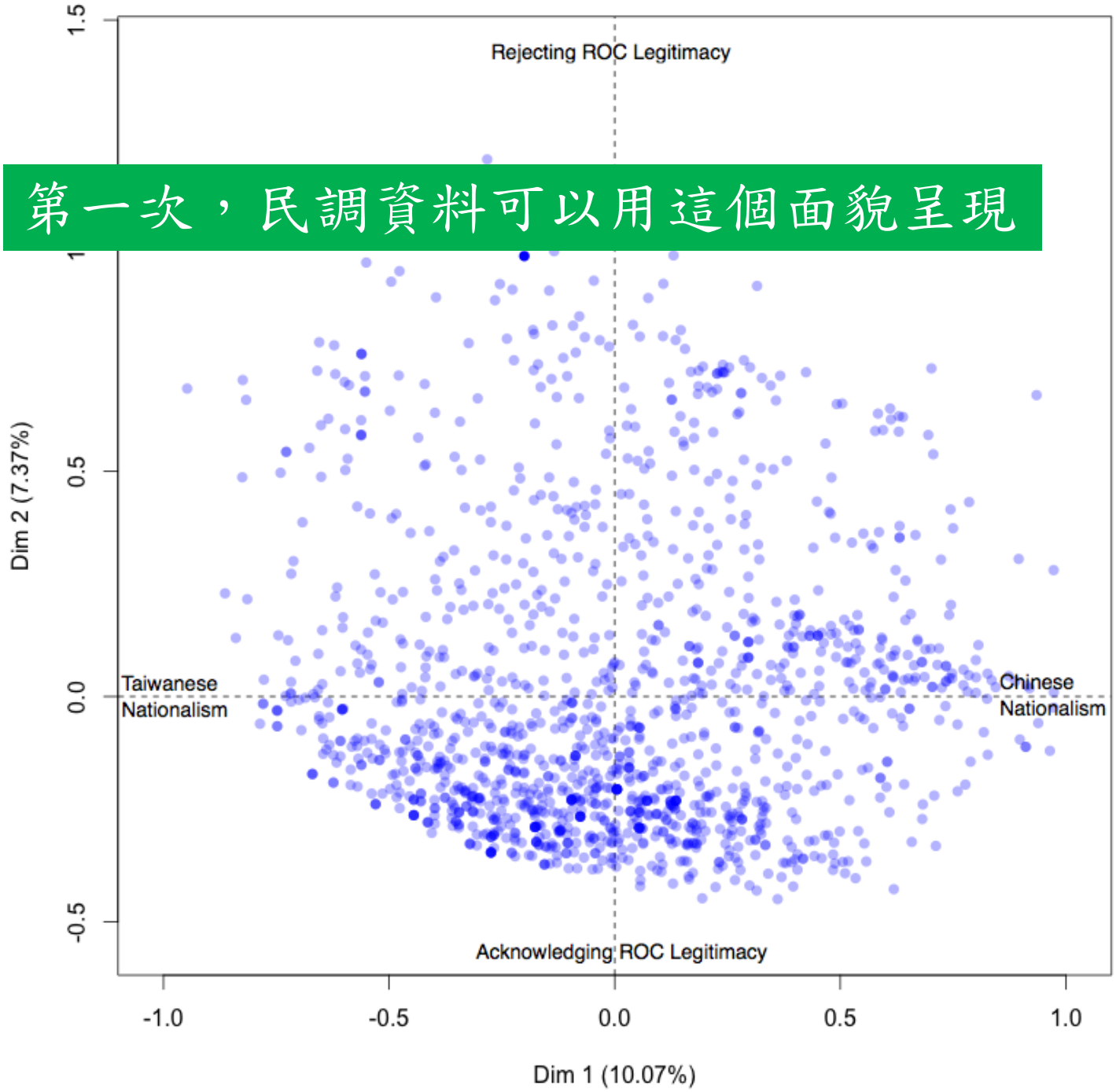
世代的切分與定義

- 第一世代（出生於1931年前）：1949年前後見證了台灣族群的對立；
- 第二世代（1932與1953年之間出生）在1949與1971年間見證了外交困境；
- 第三世代（1954與1968年生），在1986至1996年間見證了台灣經濟的起飛；
- 第四世代（1979至1989年間出生）於1986年至1996年間見證了學運及民主化
- 第五世代（1979至1988年間出生）經歷了1996年台海飛彈危機及政黨輪替；
- 第六世代（1989年之後出生）經歷了第二、三次政黨輪替及太陽花學運。

選出與政治認同最相關的30題

- 當前學界所認可的國家認同測量題，如「台灣人/中國人/都是」、兩岸關係偏好、民族認同題組
- 特別是條件統獨題組：
  - 「有人認為，如果台灣獨立不會引起戰爭，就應該宣佈獨立。請問您同不同意？」
  - 「有人認為，如果大陸在經濟、社會、政治方面的發展跟台灣差不多，兩岸就應該統一。請問您同不同意？」
  - 完整題組

第一次，民調資料可以用這個面貌呈現



在此例中被分析的問卷題（共30題）

- 如果有人問您的祖國是哪裡，請問您會怎麼回答？
- 請問您覺得下列這些歷史事件是不是很重要，要讓下一代永遠記得？
- 目前社會上有人會說自己是台灣人，有人會說自己是中國人，也有人會說兩者都是。請問您認為自己是台灣人、中國人還是兩者都是
- 對於未來台灣與中國大陸的關係，有人主張台灣獨立，也有人主張與大陸統一。請問您比較贊成哪一種主張？

75.請問您認為，我們國家的土地範圍應該包括哪些地方？（提示卡 23）

□(01)台灣　□(02)台灣、澎湖　□(03)台灣、澎湖、金門、馬祖

□(04)台灣、澎湖、金門、馬祖、港澳　□(05)台灣、澎湖、金門、馬祖、港澳、中國大陸

76.請問您覺得我們的國家現在應該叫什麼名字比較合乎您的看法？（提示卡 24）

□(01)中華民國　□(02)中華民國在台灣　□(03)台灣

□(04)台灣共和國　□(05)中國台灣　□(06)中華人民共和國

□(07)其他，請說明：_____

89.關於台灣社會文化的現象，請問您同不同意以下各種說法或想法？（提示卡 27）

	非常同意	同意	既不同意也不反對	不同意	非常不同意
(a)中華民族本來就包含很多族群，不應該分離	□(01)	□(02)	□(03)	□(04)	□(05)
(b)面對外來勢力時，台灣人應該有「自己當家作主」的自覺與決心	□(01)	□(02)	□(03)	□(04)	□(05)
(c)現在的台灣文化已經不能再說是中國文化的一部分	□(01)	□(02)	□(03)	□(04)	□(05)
(d)台灣是個小而美的國度，未來也都會繼續維持下去	□(01)	□(02)	□(03)	□(04)	□(05)
(e)台灣人的祖先就是黃帝，我們要繼承這樣的血統與歷史	□(01)	□(02)	□(03)	□(04)	□(05)
(f)在台灣長久居住或成長的人們應該一起發展出自己的新民族	□(01)	□(02)	□(03)	□(04)	□(05)
(g)台灣人很優秀，各行各業都有人才在世界上有很成功的表現	□(01)	□(02)	□(03)	□(04)	□(05)
(h)作為華夏子孫，我們在國際上應該盡力將中華文化發揚光大	□(01)	□(02)	□(03)	□(04)	□(05)
(i)不管台灣發生任何問題，我都一定會挺它到底，絕對不會想要移民到國外	□(01)	□(02)	□(03)	□(04)	□(05)

58.請您用 0 至 10 分來表示您自認為是台灣人的程度，10 分表示「完全是台灣人」，0 分表示「完全不是台灣人」。請問您會選幾分？**（提示卡 19）**

完全不是台灣人										完全是台灣人
□(0)	□(1)	□(2)	□(3)	□(4)	□(5)	□(6)	□(7)	□(8)	□(9)	□(10)

59.請您用 0 至 10 分來表示您自認為是中國人的程度，10 分表示「完全是中國人」，0 分表示「完全不是中國人」。請問您會選幾分？**（提示卡 20）**

完全不是中國人										完全是中國人
□(0)	□(1)	□(2)	□(3)	□(4)	□(5)	□(6)	□(7)	□(8)	□(9)	□(10)

```
> install.packages("FactoMineR")
> install.packages("devtools")
> devtools::install_github("kassambara/factoextra")
```

```
> library(FactoMineR)
> library(factoextra)
> library(dplyr)
```

```
> load("tscs2013.rda")
```

```
> tscs2013forMCA <- select(tscs2013,
+                          c(# 核心變數 (core vars)
+                            gen.1, gen.2, gen.3, gen.4, gen.5, # 世代
+                            v15r, #「祖國」是哪裡
+                            v54ar, v54br, v54cr, v54dr,#最有承傳價值的歷史事件
+                            v57r, #台灣人/既是台灣人也是中國人/其他
+                            v61r, # 統獨立場
+                            v76r, # 國號
+                            v89ar, v89br, v89cr, v89dr,
+                            v89er, v89fr, v89gr, v89hr, v89ir, # 民族－國家
+
+                            # quantatative supplementary vars
+                            v58r, # 自認台灣人程度
+                            v59r, # 自認中國人程度
+                            # v84ar,  # 去大陸次數（1－6）
+
+                            #qualitative supplementary vars
+                            sex,
+                            college, # 大專教育程度
+                            camp, # 政黨傾向
+                            v71ar, # 中華民族包含台灣原住民
+                            v71er, # 中華民族包含台灣居民
+                            v75r  #  國家領土範圍
+                            ))
```

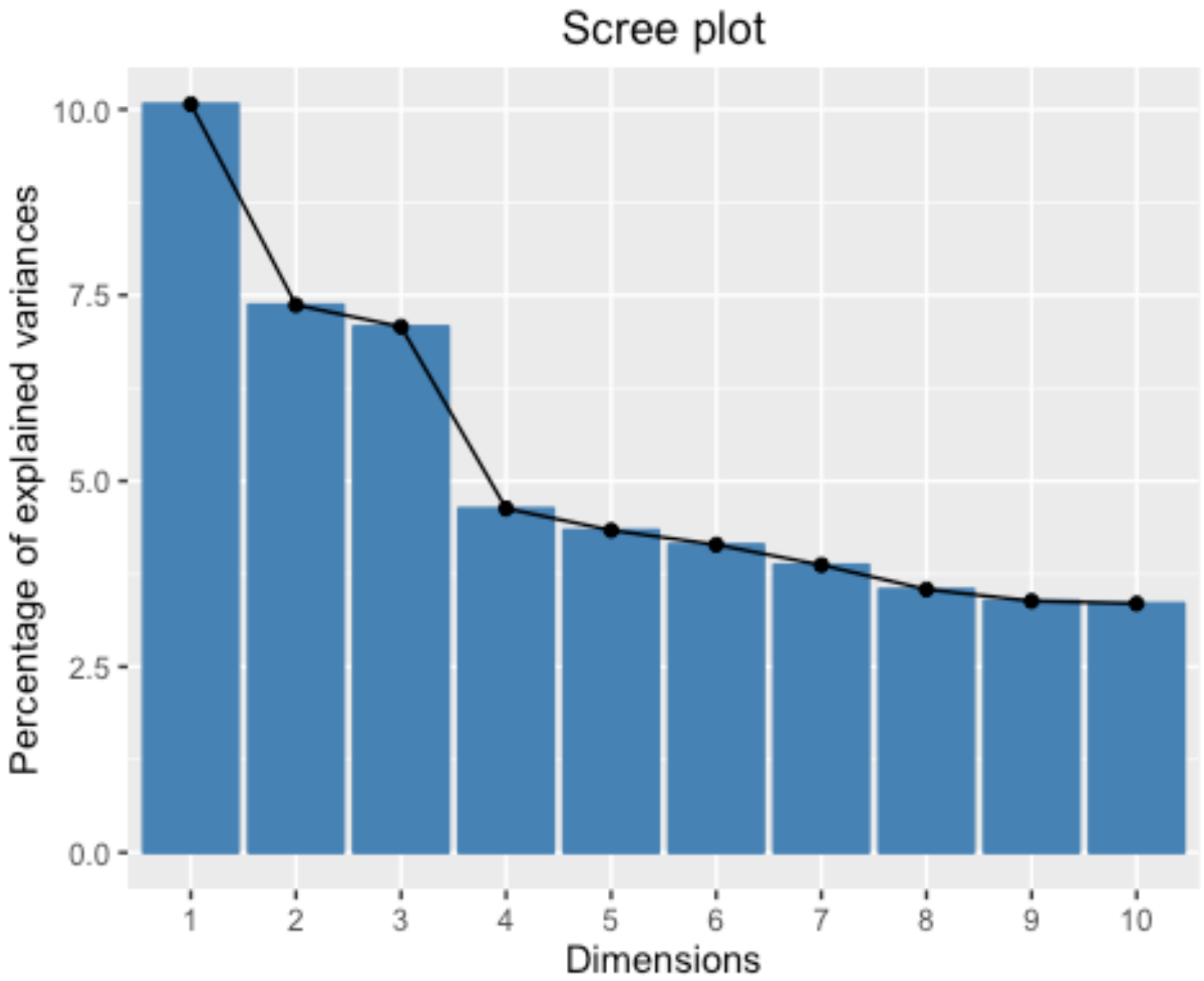
```
> # 將無效值剔除（list-wise deletion）。
> tscs2013forMCA.nona <- na.omit(tscs2013forMCA)
```

```
> nrow(tscs2013forMCA.nona)
[1] 1496

> names(tscs2013forMCA.nona)
```

```
> res<-MCA(tscs2013forMCA.nona, ncp=10,
           quanti.sup=c(23,24),
           quali.sup=25: 30, graph= F) #ncp 10個維次
```

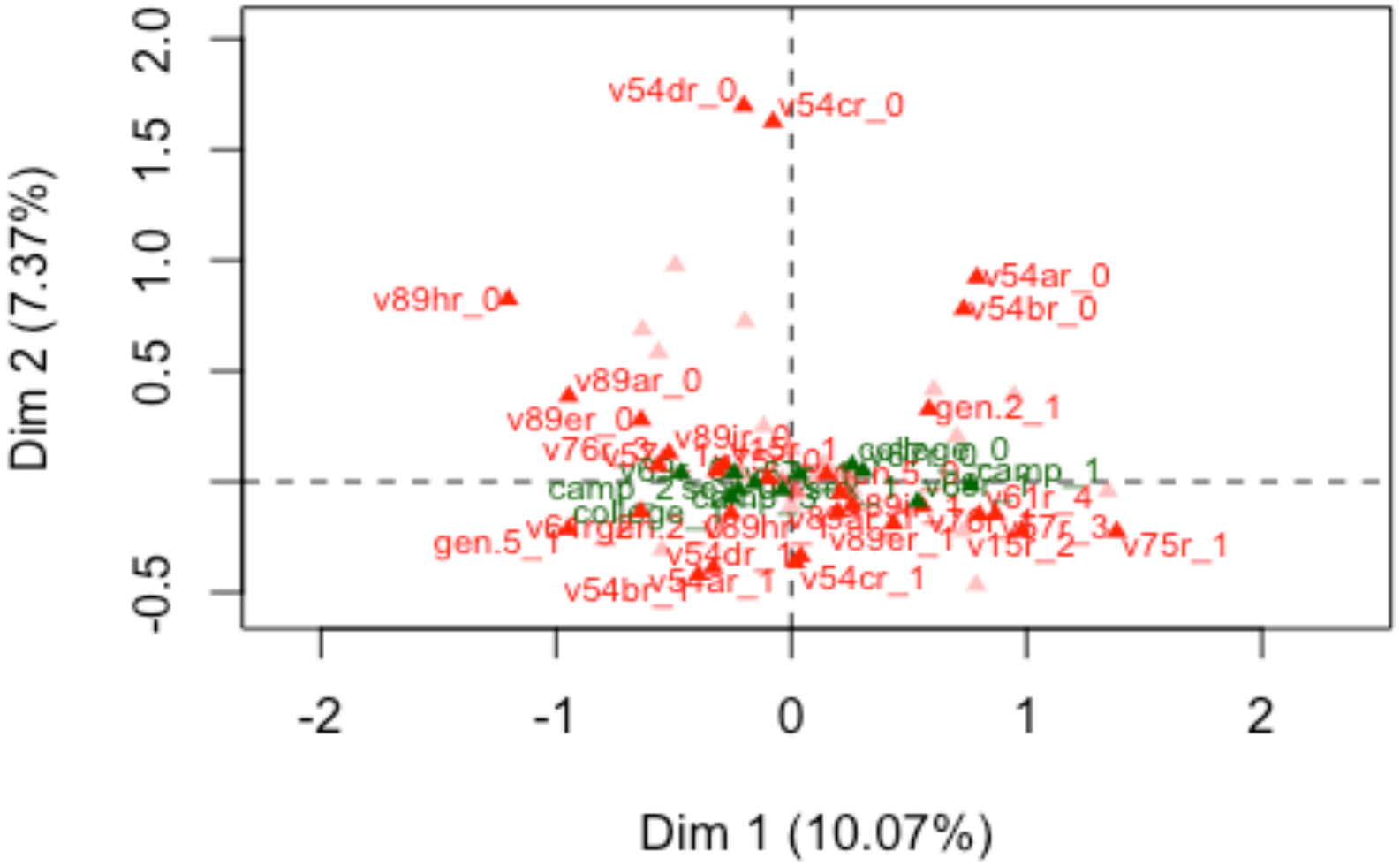
```
> fviz_screeplot(res, ncp=10)
```

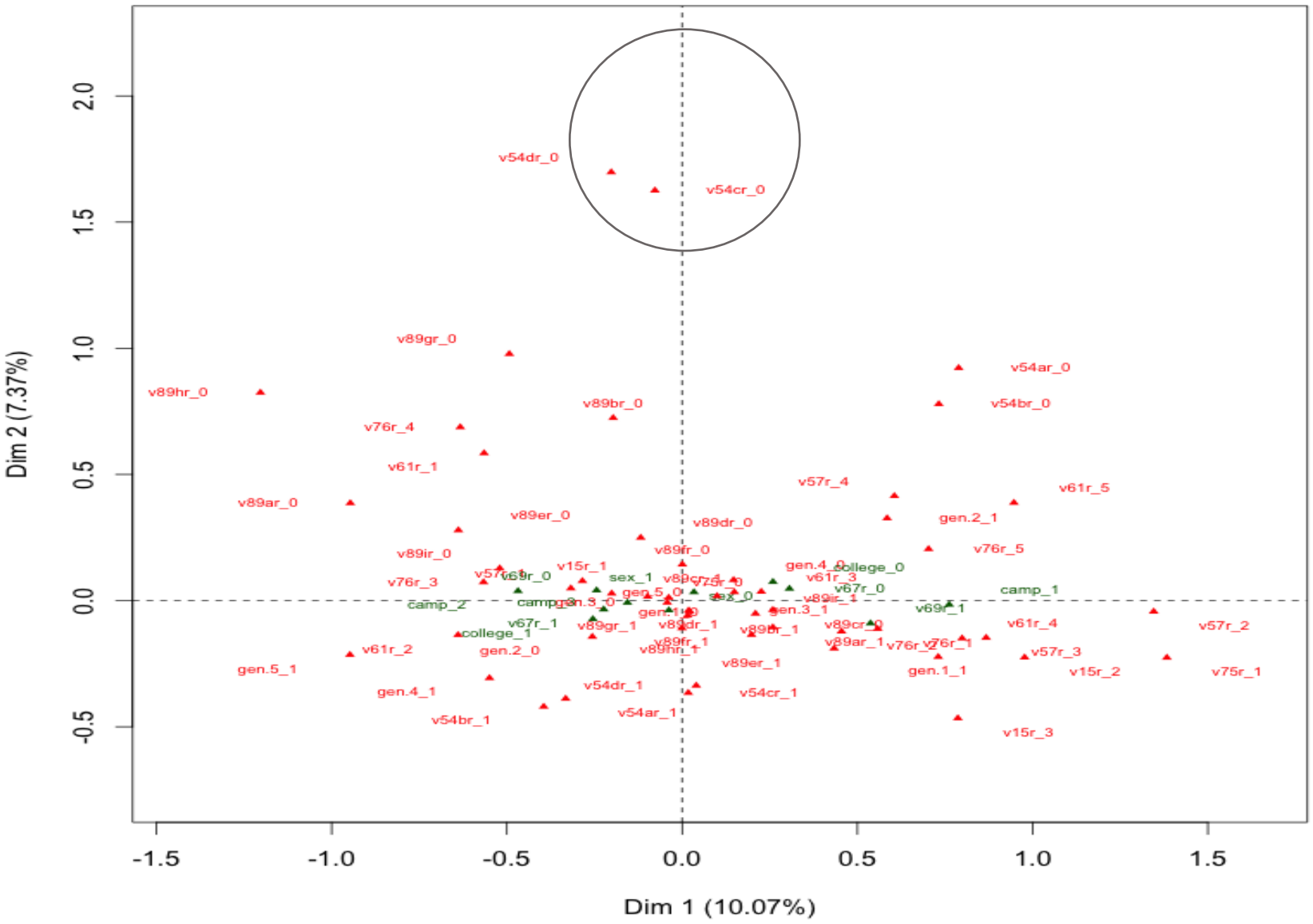


將其中最重要變數類別（選項）的組合挑出：

```
> plot(res, axes=c(1, 2), new.plot=TRUE,
       col.var="red",
       col.ind="black",
       col.ind.sup="black",
       col.quali.sup="darkgreen",
       col.quanti.sup="blue",
       label=c("var"), cex=0.7,
       selectMod = "cos2 30",  #共52個選項組合
       invisible=c("ind", "quali.sup"),
       xlim=c(-1.2,1.2),
       ylim=c(-0.6,2),
       autoLab = "yes",
       # title="Top 30 Critical Elements on the MCA Factor Map")
       title="")
```

顯示最重要變數的組合

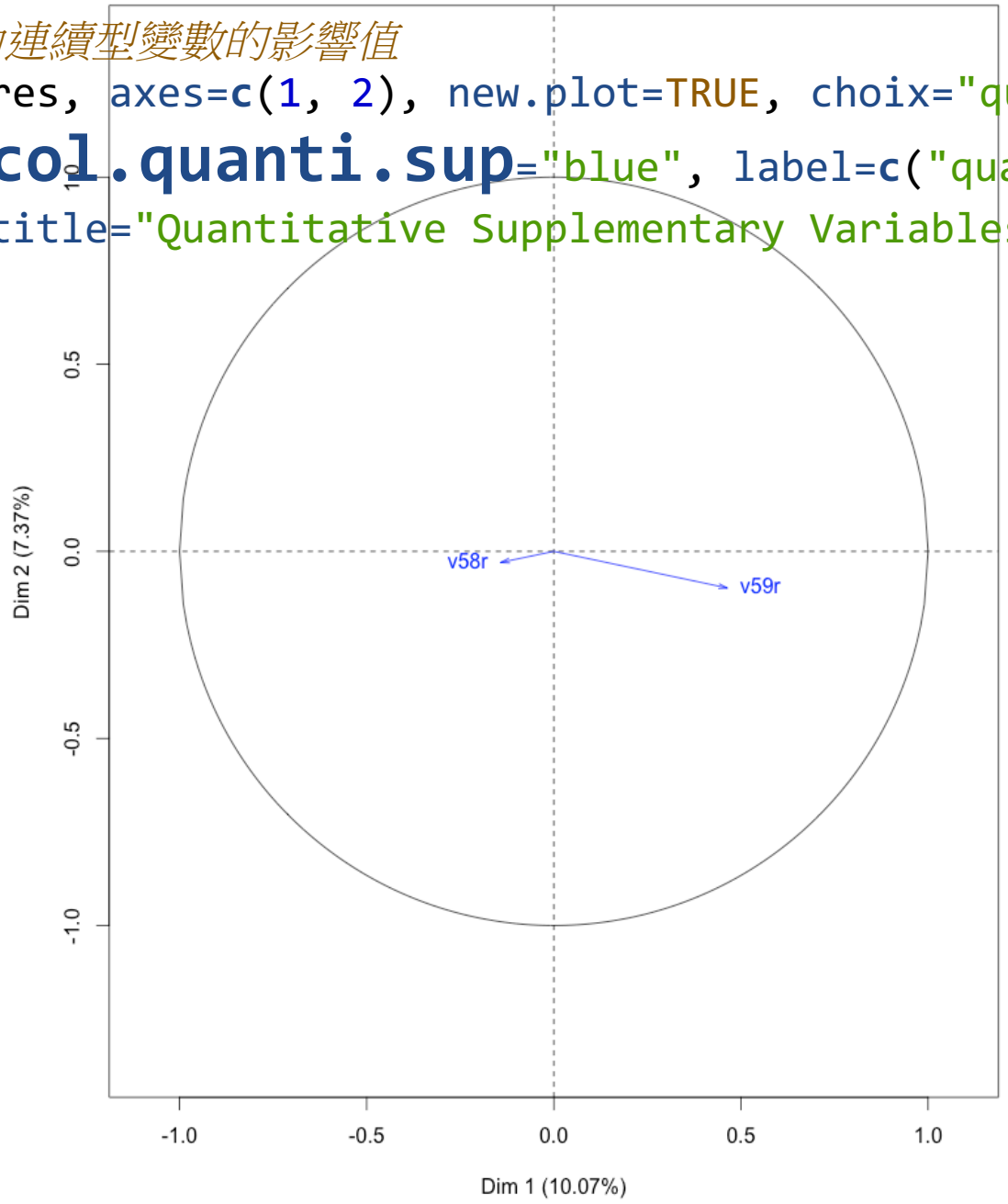

Dim 2 (7.37%)
2.0
1.5
1.0
0.5
-0.5
-2
-1
0
1
2
Dim 1 (10.07%)
v54dr_0
v54cr_0
v89hr_0
v54ar_0
v54br_0
v89ar_0
v89er_0
gen.2_1
camp_1
v61r_4
v75r_1
v54dr_1
v54ar_1
v54cr_1


Dim 2 (7.37%)
Dim 1 (10.07%)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
v54dr_0
v54cr_0
v89gr_0
v89hr_0
v89br_0
v54ar_0
v54br_0
v76r_4
v61r_1
v89ar_0
v89er_0
v57r_4
v61r_5
gen.2_1
v89dr_0
v89ir_0
v89fr_0
v76r_5
v15r_1
v57r_0
v89cr_0
v61r_3
v67r_0
v76r_3
sex_1
gen.5_0
sex_0
camp_1
camp_2
gen.3_0
gen.3_1
v69r_1
v67r_1
v89gr_1
v89dr_1
v89br_1
v89cr_1
v61r_4
v57r_2
college_1
v89fr_1
v89ar_1
v76r_2
v76r_1
v57r_3
v61r_2
gen.2_0
v89hr_1
gen.1_1
v89er_1
v15r_2
v75r_1
gen.5_1
v54dr_1
v54cr_1
gen.4_1
v54ar_1
v54br_1
v15r_3

注意：構成第二維次（Y軸）的因素

- 第二維次的代表題：「請問您覺得下列這些歷史事件是不是很重要，要讓下一代永遠記得？」
  - 「推翻滿清，建立中華民國」(v54c) 與
    「八年對日抗戰勝利」(v54d) 一組；
  - 「二二八事件」（v54ar）與
    「美麗島事件、黨外民主運動」（v54br）一組

```
> # 輔助連續型變數的影響值
> plot(res, axes=c(1, 2), new.plot=TRUE, choix="quanti.sup",
+      col.quanti.sup="blue", label=c("quanti.sup"),
+      title="Quantitative Supplementary Variables")
```



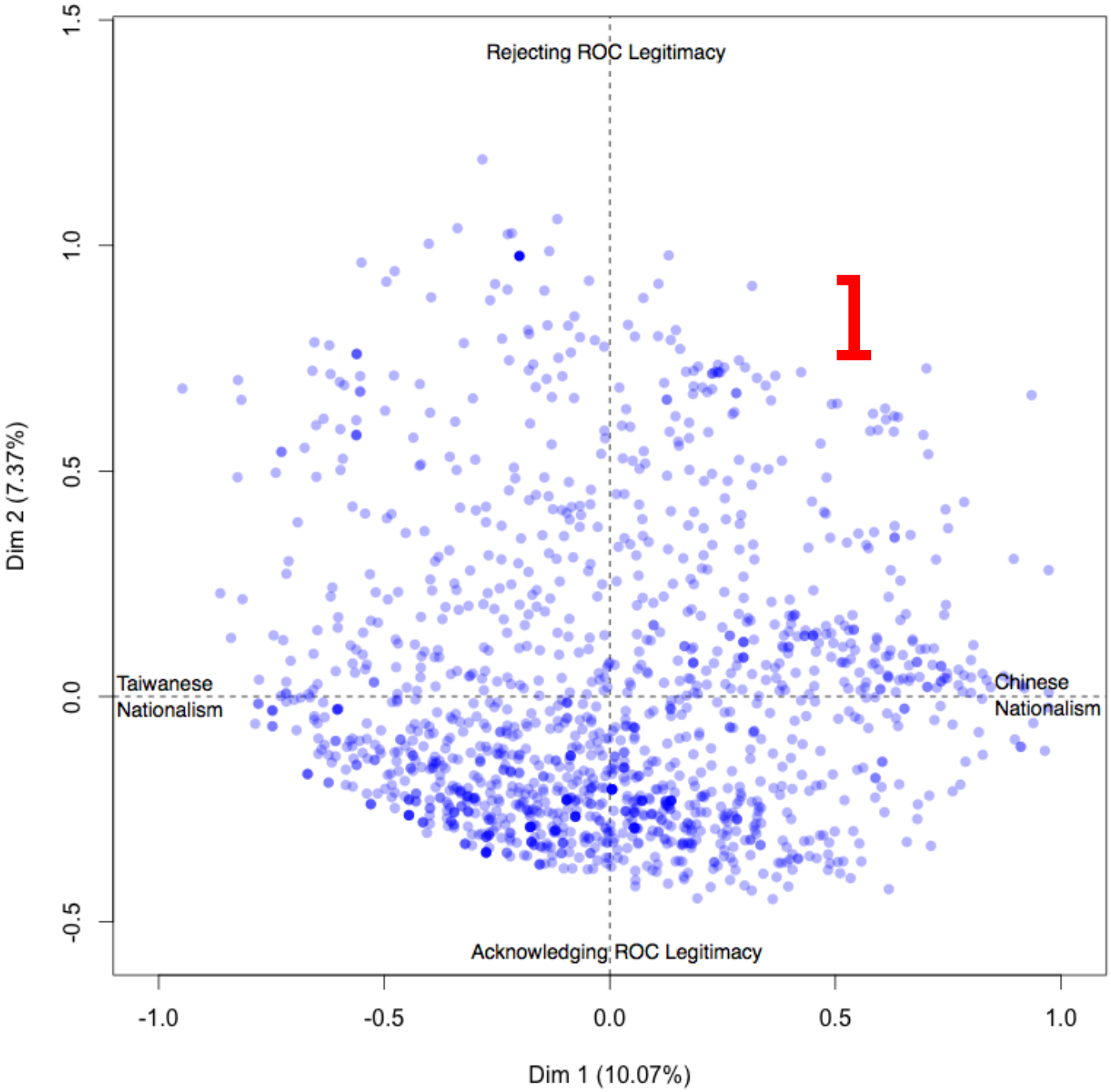
受訪者在兩個維度的分佈

```
> plot(res, axes=c(1, 2), new.plot=TRUE, choix="ind",
+       col.var="red", col.quali.sup="darkgreen",
+       label=c("var"),
+       xlim=c(-1,1),
+       selectMod ="cos2 15", select="cos2 1",
+       invisible=c("quali.sup", "var"),
+       )
```

接下來，為概念命名

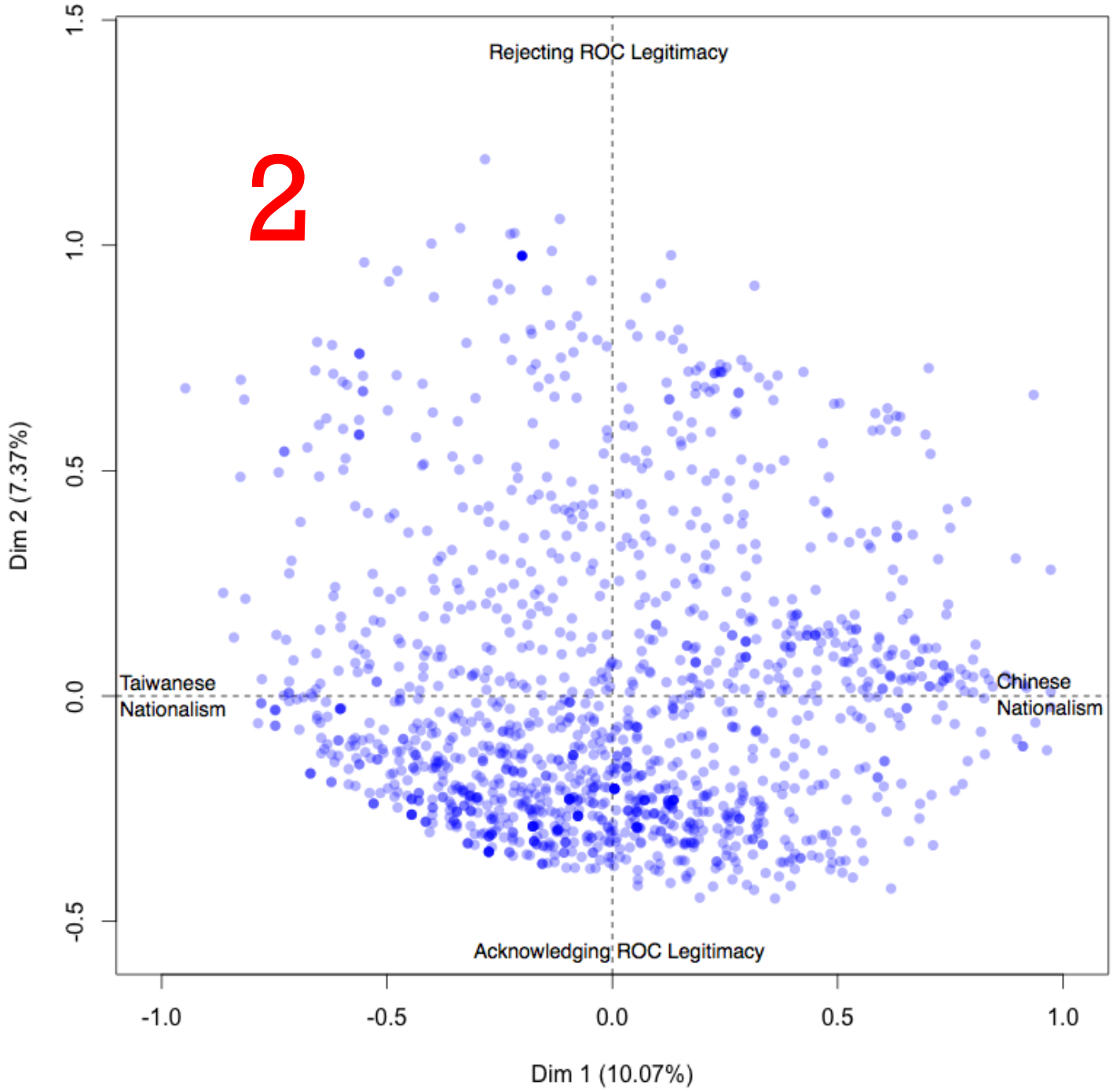
第一軸線的代表概念：**民族認同**（中華民族或台灣民族）

第二軸線的代表概念：**國家正當性**（接受中華民國與否）


Rejecting ROC Legitimacy
Acknowledging ROC Legitimacy
Taiwanese Nationalism
Chinese Nationalism
1
Dim 2 (7.37%)
Dim 1 (10.07%)
1.5
1.0
0.5
0.0
-0.5
-1.0
-0.5
0.0
0.5
1.0

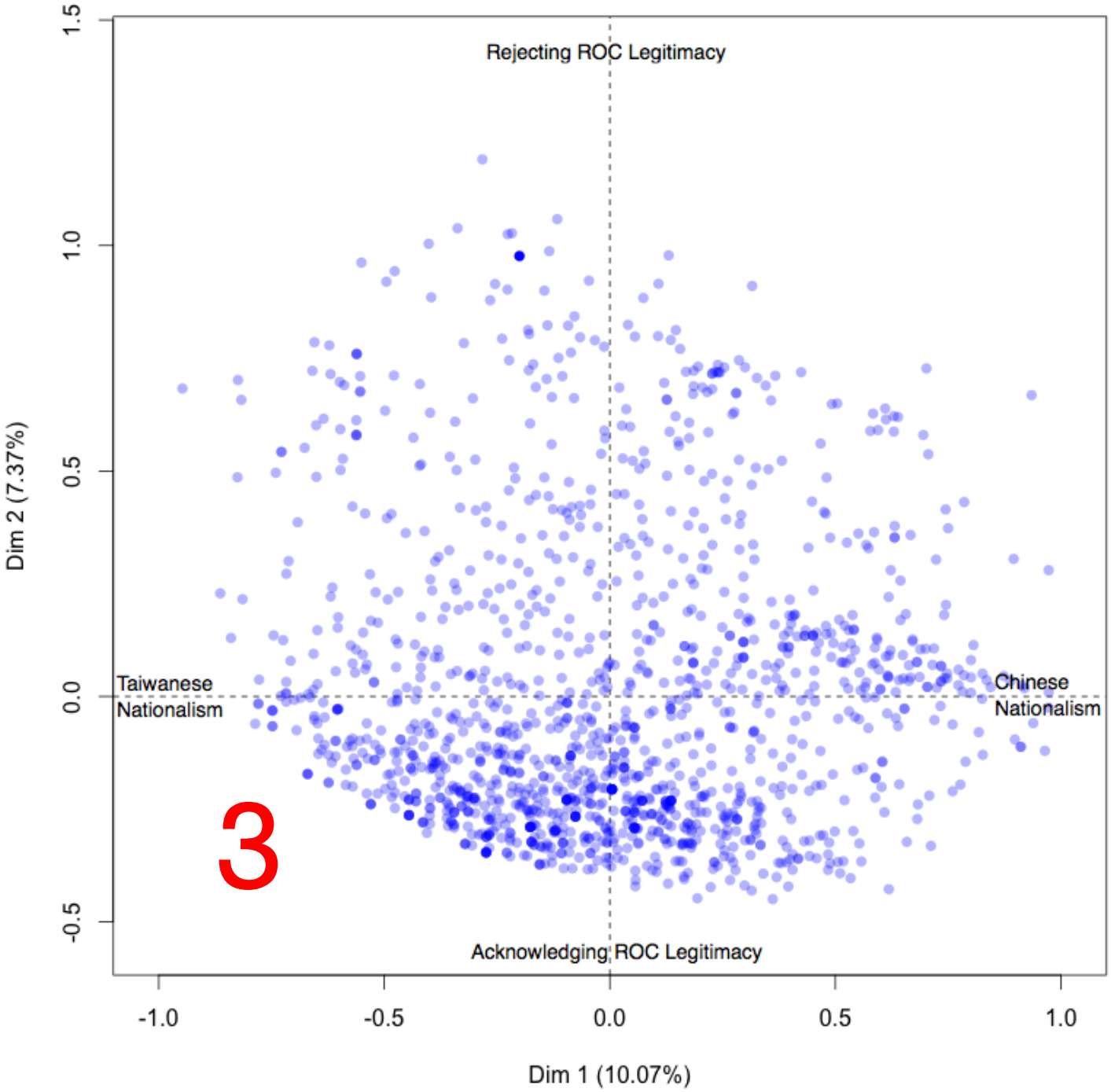
位於第一象限的民眾特徵：

- 第二世代
- 政黨傾向為藍營 [不在第四象限？]
- 不認為「二二八事件」是重要歷史事件
- 不認為「美麗島事件、黨外民主運動」是重要歷史事件
- 無大專教育程度
- 男性


2
Rejecting ROC Legitimacy
Acknowledging ROC Legitimacy
Taiwanese Nationalism
Chinese Nationalism
Dim 2 (7.37%)
Dim 1 (10.07%)
1.5
1.0
0.5
0.0
-0.5
-1.0
-0.5
0.0
0.5
1.0

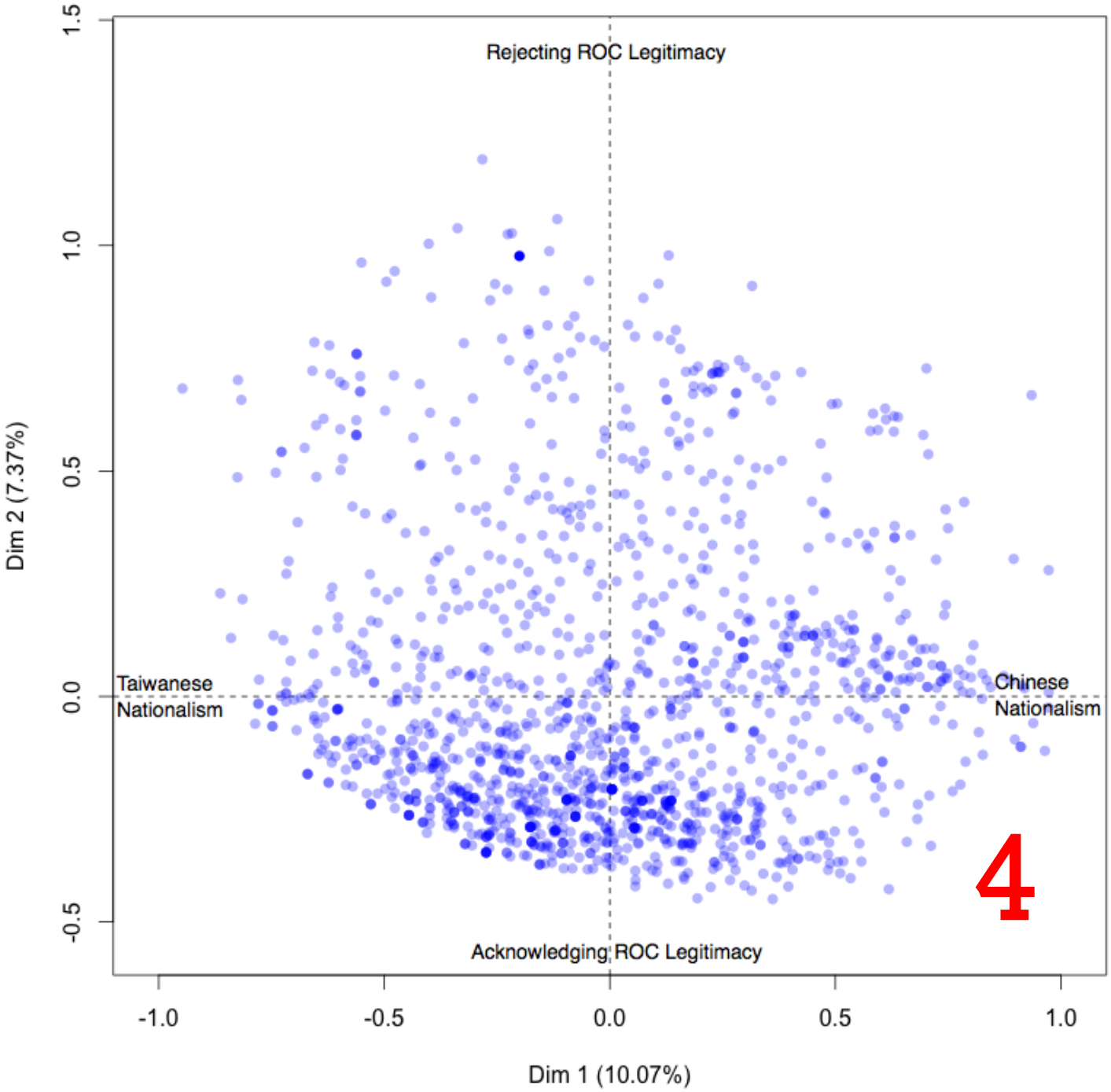
位於第二象限的民眾特徵:

- 政黨傾向為綠營以及「中間 / 不表態/其他」
- 認為自己的祖國是台灣（不是中華民國、中國或其他）
- 認為自己是台灣人（不是中國人亦非都是）
- 認為國土不包含中國大陸
- 認為國家現在名字應該叫作台灣
- 不同意「中華民族本來就包含很多族群，不應分離」
- 不同意「台灣人的祖先就是黃帝，我們要繼承這樣的血統與歷史」
- 不同意「作為華夏子孫，我們在國際上應該盡力將中華文化發揚光大」
- 不同意「不管台灣發生任何問題，我都一定會挺它到底，絕對不會想要移民到國外」


Rejecting ROC Legitimacy
Acknowledging ROC Legitimacy
Taiwanese Nationalism
Chinese Nationalism
Dim 1 (10.07%)
Dim 2 (7.37%)
1.5
1.0
0.5
0.0
-0.5
-1.0
-0.5
0.0
0.5
1.0
3

位於第三象限的民眾特徵:

- 第五世代 [不在第二象限]
- 有大專學歷
- 「維持現狀，以後走向獨立」
- 如果台灣獨立不會引起戰爭，就應該宣佈獨立
- 二二八事件、美麗島事件及黨外民主運動算是歷史上的重要、值得永遠被記得的事件
- 如果大陸在經濟、社會、政治方面的發展跟台灣差不多，兩岸也不應該統一


Rejecting ROC Legitimacy
Acknowledging ROC Legitimacy
Taiwanese Nationalism
Chinese Nationalism
Dim 1 (10.07%)
Dim 2 (7.37%)
-1.0
-0.5
0.0
0.5
1.0
1.5
-0.5
0.0
0.5
1.0
4

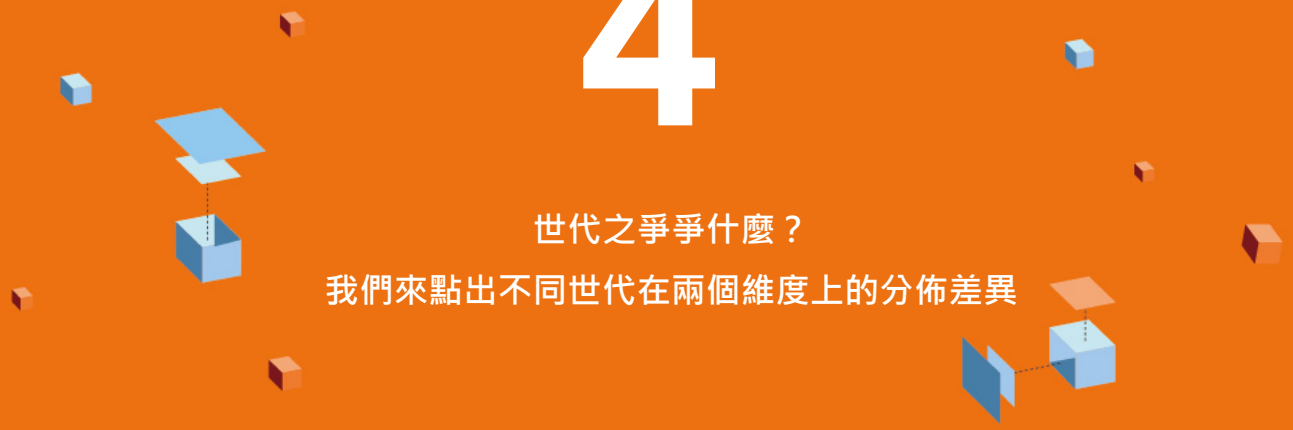
位於第四象限的民眾特徵:

- 自己是台灣人也是中國人
- 國家現在叫作中華民國比較適合
- 中華民國是祖國
- 兩岸維持現狀，以後走向統一
- 「推翻滿清，建立中華民國」與「八年對日抗戰勝利」很重要，要讓下一代永遠記得。
- 「台灣人的祖先就是黃帝，我們要繼承這樣的血統與歷史」
- 「中華民族本來就包含很多族群，不應該分離」
- 「不管台灣發生任何問題，我都一定會挺它到底，絕對不會想要移民到國外」
- 「作為華夏子孫，我們在國際上應該盡力將中華文化發揚光大」
- 即使台灣獨立不會引起戰爭，也不該宣佈獨立。

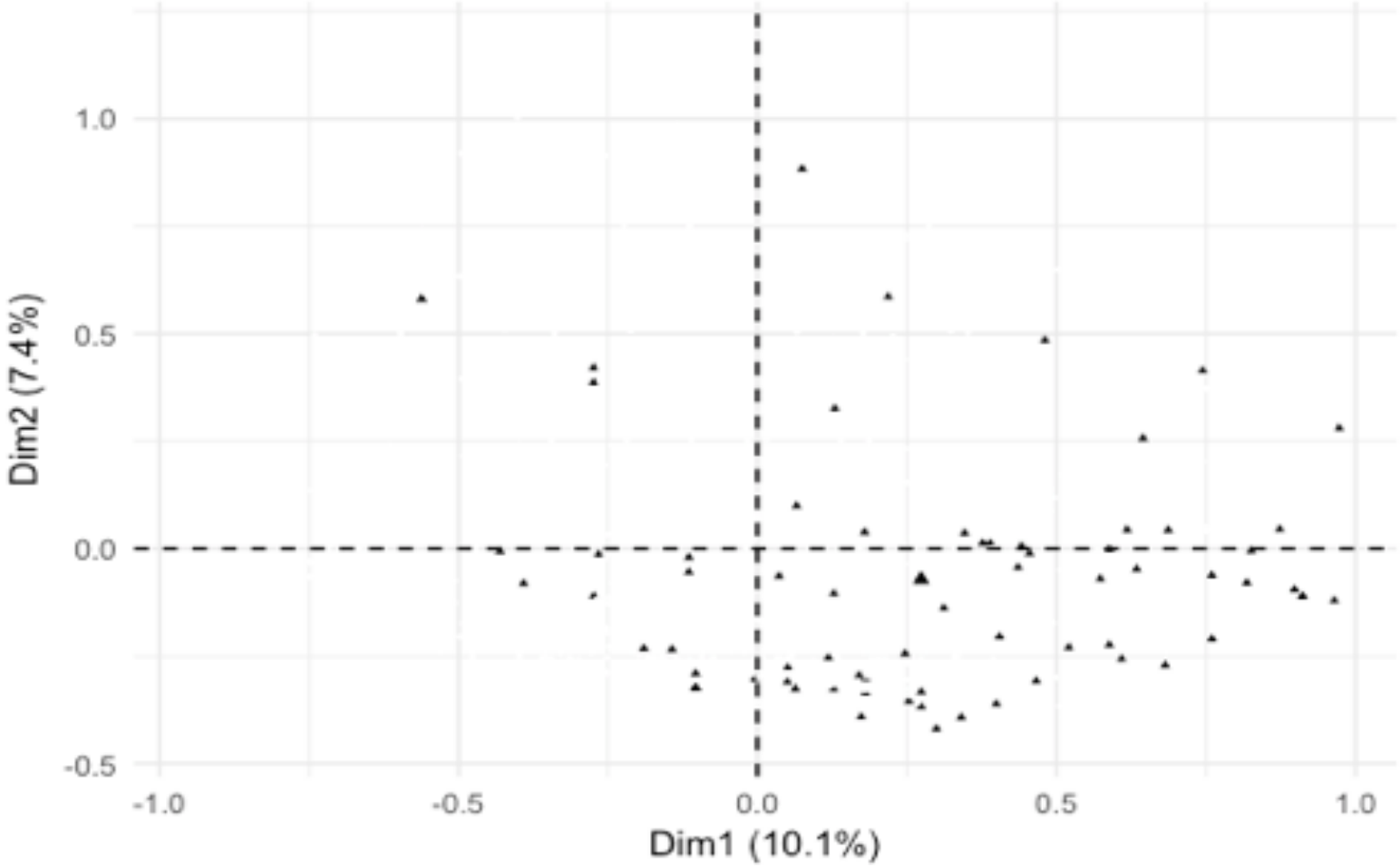
4

世代之爭爭什麼？

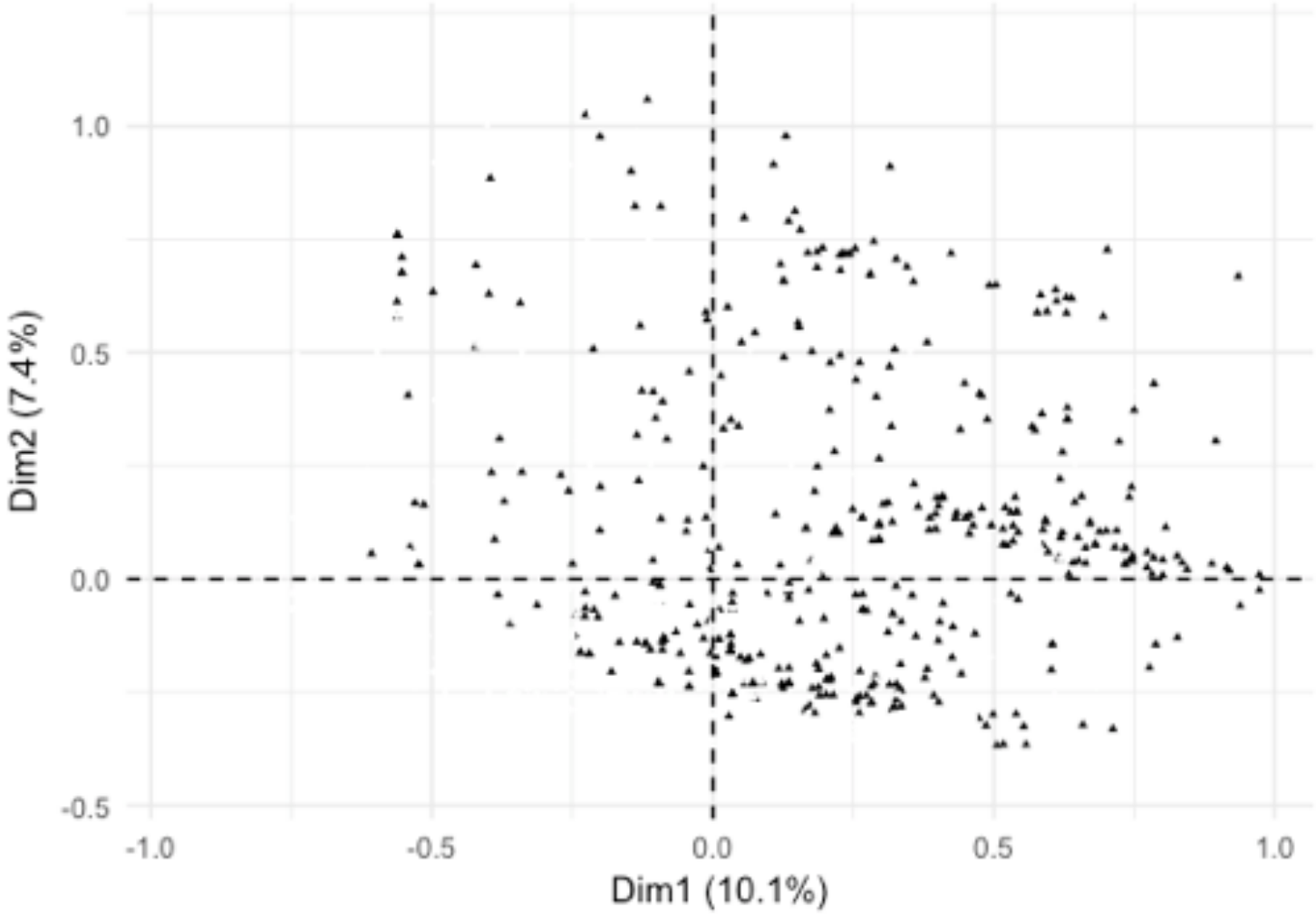
我們來點出不同世代在兩個維度上的分佈差異



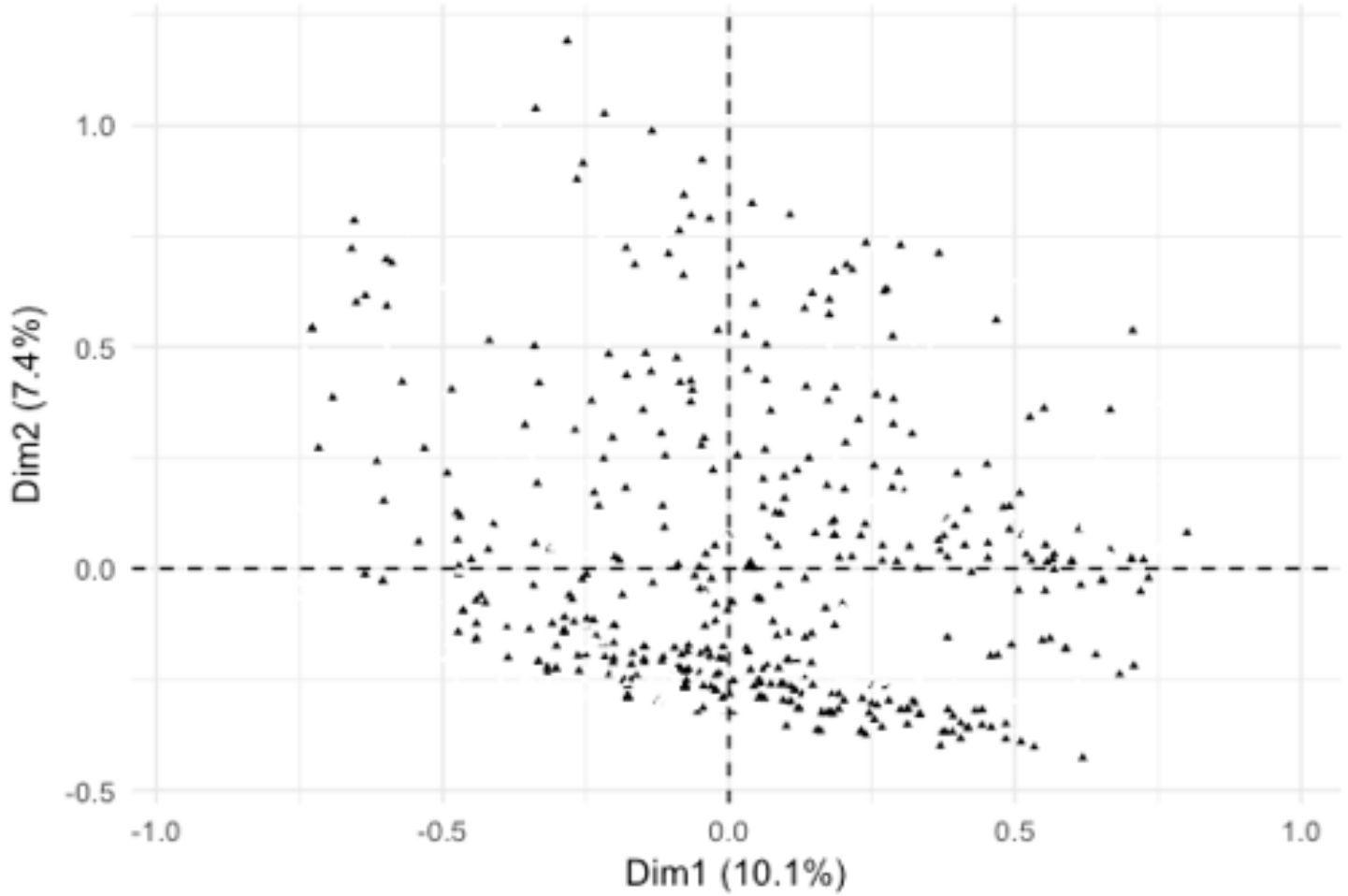
第一世代



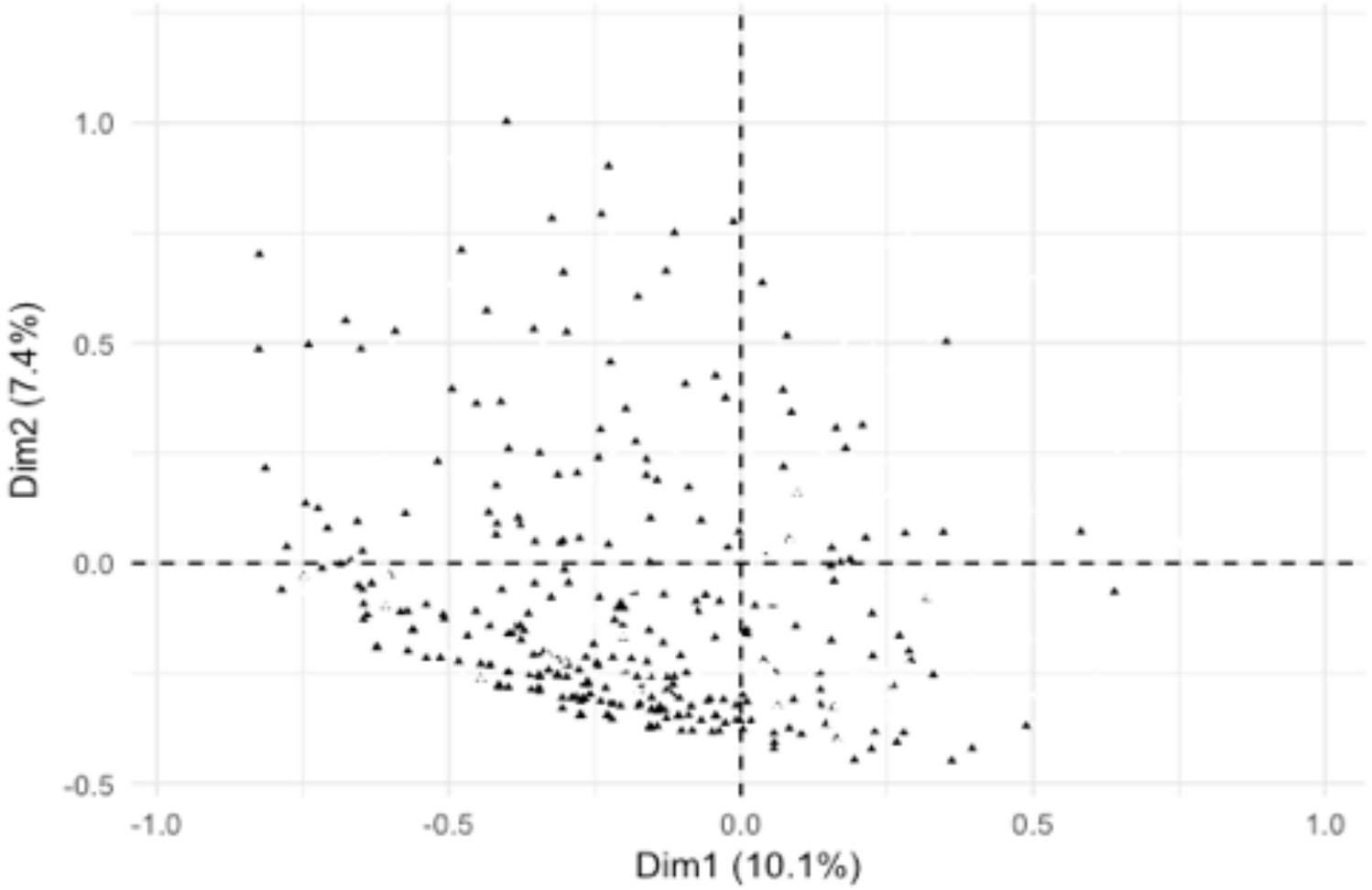
第二世代



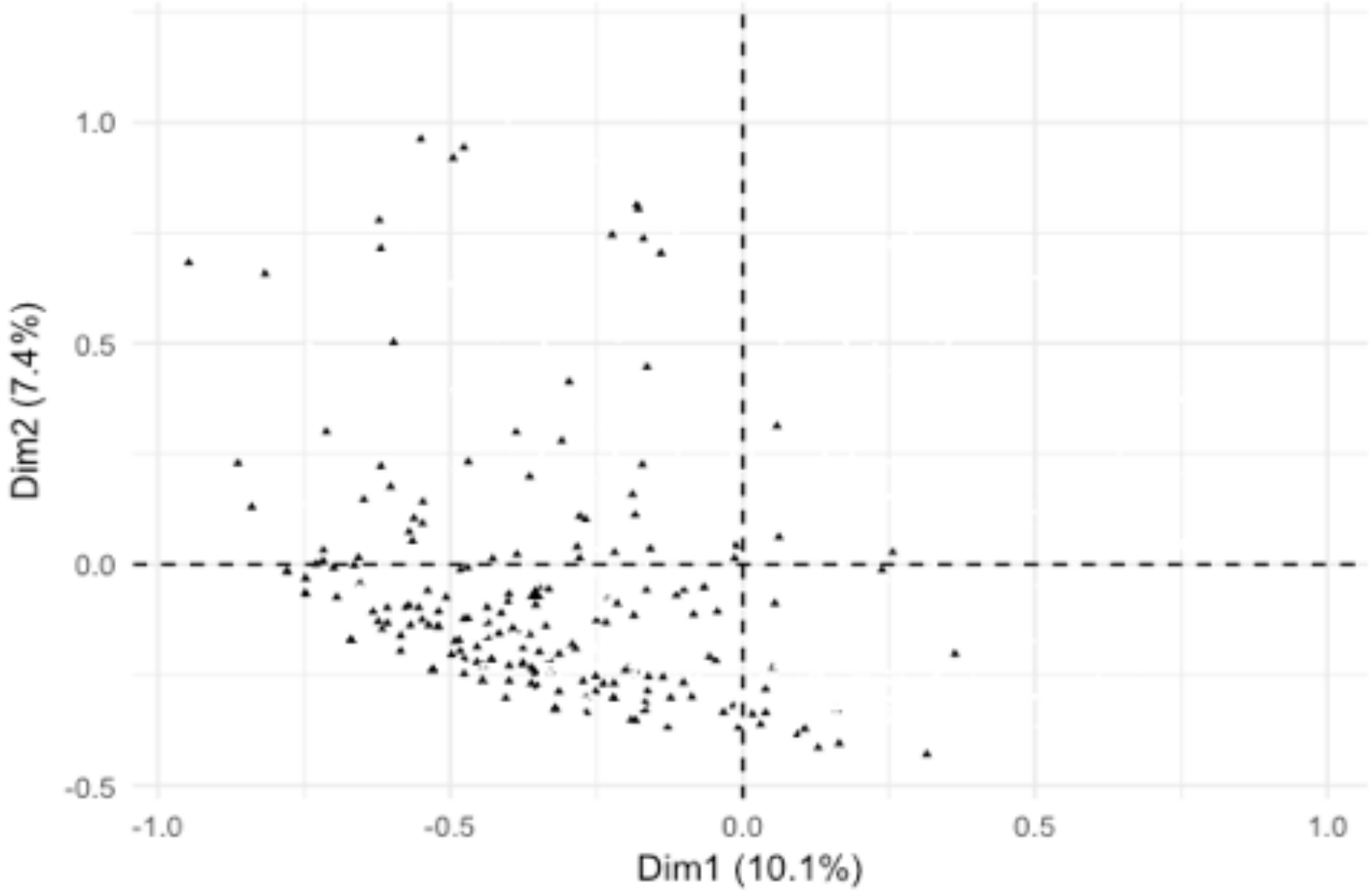
第三世代



第四世代

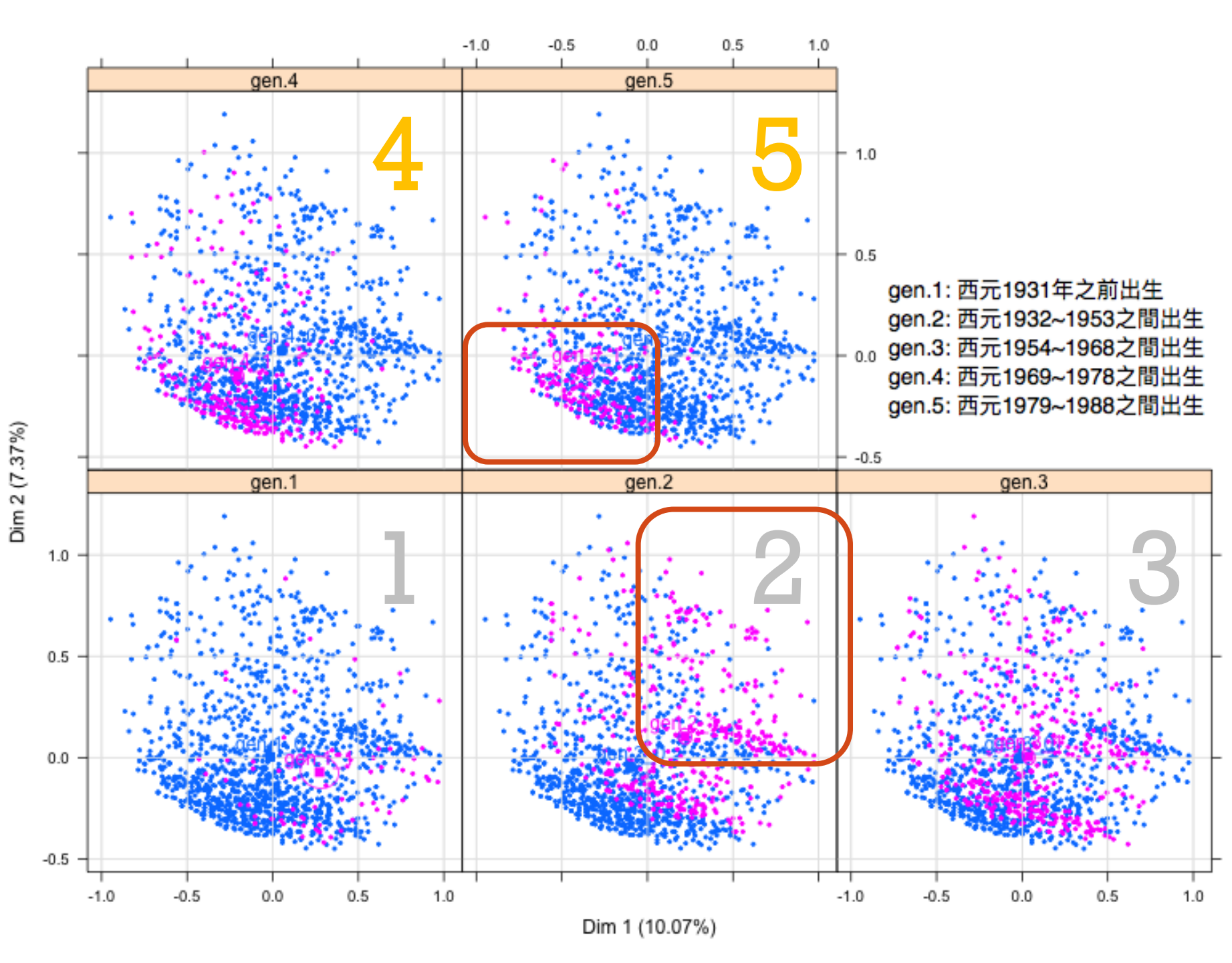


第五世代



世代分佈的差異

```
> library(factoextra)
> plotellipses(res,
                keepvar =
                c(“gen.1”,“gen.2”,
                  "gen.3","gen.4","gen.5"))
```


gen.4
gen.5
gen.1
gen.2
gen.3
4
5
1
2
3
gen.1: 西元1931年之前出生
gen.2: 西元1932~1953之間出生
gen.3: 西元1954~1968之間出生
gen.4: 西元1969~1978之間出生
gen.5: 西元1979~1988之間出生
Dim 1 (10.07%)
Dim 2 (7.37%)
-1.0
-0.5
0.0
0.5
1.0

你能看見什麼？

- 從人數的分佈來看（桃紅色的點）
- 從每個世代的所在位置來看
- 這些點所代表的都是每個選民不太容易移動的認同與立場。
- 不同的認同結構是造成選民、媒體、及政治人物所說出來的話差異的原因。

以這種樣貌所呈現出的民調資料具有不同於賽馬民調的潛力

- 若能解讀這張圖，你就看得出
  - 2014年太陽花學運的社會氣氛、
  - 2014年縣市長選舉、
  - 甚至是2016年選民大致在想什麼、選票在那裡，以及為什麼政黨推出的競選策略。


看圖猜猜看：
什麼是最能爭取到多數選民接受的政治語言？
Dim 2 (7.37%)
1.0
0.5
0.0
-0.5
Taiwanese
Nationalism
Chinese
Nationalism

維·持·現·狀

「這些多數選民，認同什麼？」

A: 台灣人的中華民國 + 中國人的中華民國。

第二世代已開始懷疑他們曾經認定的中華民國，而年輕世代（第五世代以後）已重新定義中華民國。

用傳統民調分析方法，要看出這件事可不容易啊。

小結：

- 當前所使用的主要的用來測量「國家認同」的題目，包括台灣人/中國人、統獨立場、條件統獨，乃至國號選擇等題目，多屬於同一個維次「民族」的概念。
- 「對中華民國正當性的認同」是個與民族認同分立的概念。
- 傳統的統獨題在本研究的三十道題目中，既不算是民族認同這個概念的主要構成因子，也無法對應到任何民族認同之外的概念。吳乃德（2005）所倡議的條件統獨題組則如預期，反映了受訪者的民族認同。
- 在2013年時「中華民國的正當性」並未明顯消退，但已浮現「一個中華民國，不同世代各自表述」的樣貌。

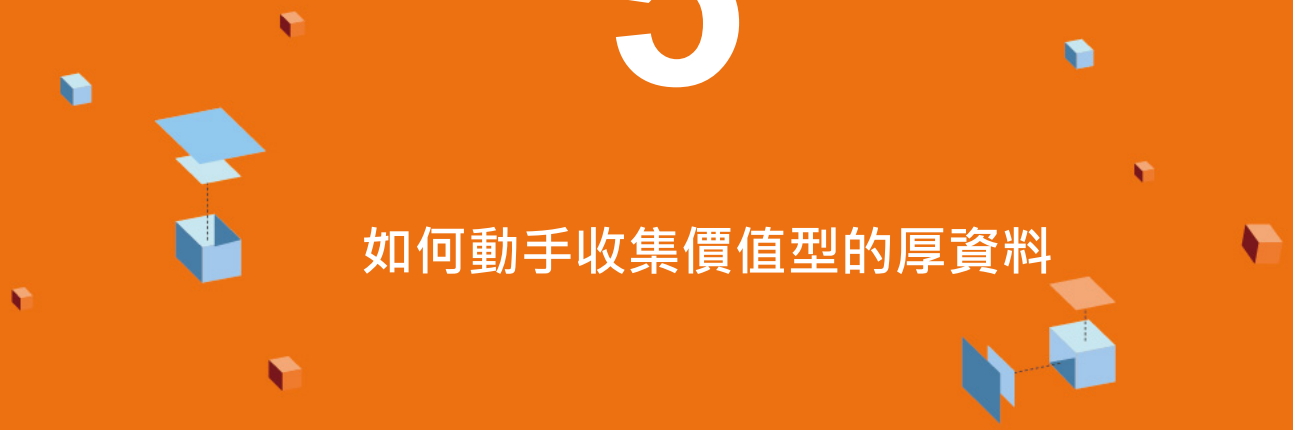
- 第一、二世代或可以說是「天然統」（結合中華民族主義的中華民國史觀認同），
- 但第五世代的並不算是「天然獨」，因為就認同中華民國的正當性對他們來說並不弱於對其於他世代。
- 「台灣人的中華民國」（而非台灣人的台灣國）是第四、五世代最鮮明的國家認同觀，與第二世代「中華民國是全體中國人的中華民國」對比強烈。

展望

- 收集第六世代的資料；
  - 國內選舉調查多以「合格選民」（20歲）為訪問對象，因此本研究無法觀察到更年輕的民眾。經查該筆資料中並無1989年後出生（即受訪時24歲以下）的受訪者。
- 需要取得2014年到2016年之間，的資料，分析各個世代經歷學運以及第三次政黨輪替時認同變化的樣貌。
- 兩個維次的命名及標籤問題（ 本研究目前還無法確切為第二維次貼上準確的標籤，只能以「中華民國正當性」暫稱 ）
- 期待更多（有創意的）測量題加入潛在概念探索的行列。

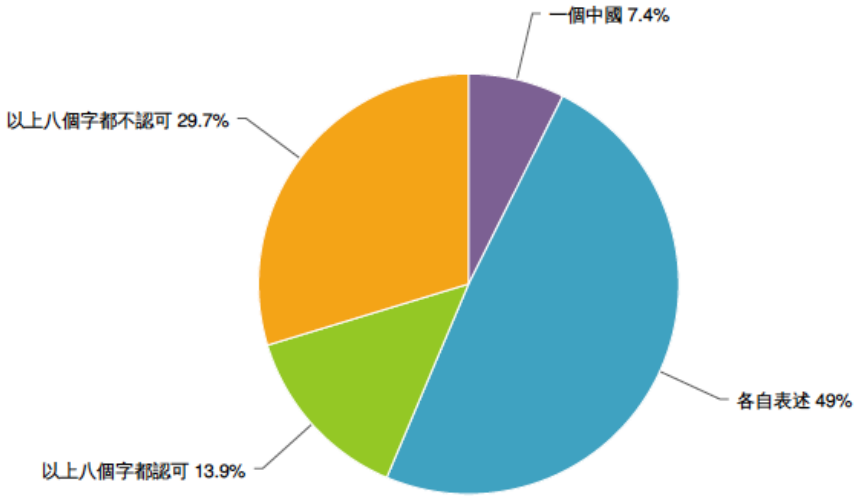
5

如何動手收集價值型的厚資料



傳統的長條圖和圓餅圖

15.「九二共識」中的「一個中國、各自表述」八個字中您最認可的內涵是什麼？



Value	Percent	Count
一個中國	7.4%	61
各自表述	49.0%	404
以上八個字都認可	13.9%	115
以上八個字都不認可	29.7%	245
	Total	825

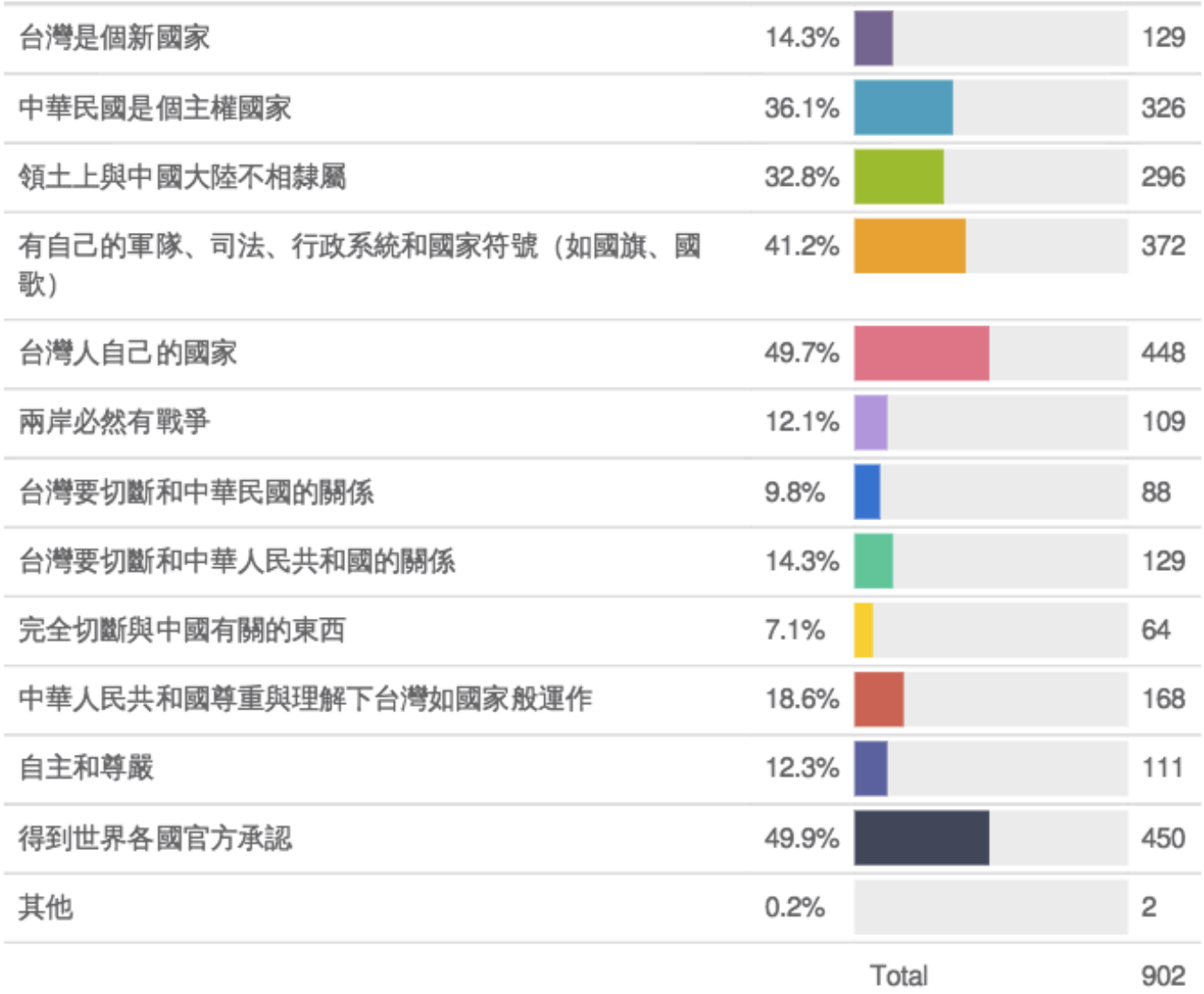
Statistics	
Sum	2,194.0
Average	2.7
StdDev	1.0
Max	4.0

透過問受訪者更深刻的問題，我們可以從調查資料中發掘更多的可能樣貌。

一般民調市調會偏重於詢問行為面及偏好的問題，但我們還可以問出更多關於價值觀的問題。

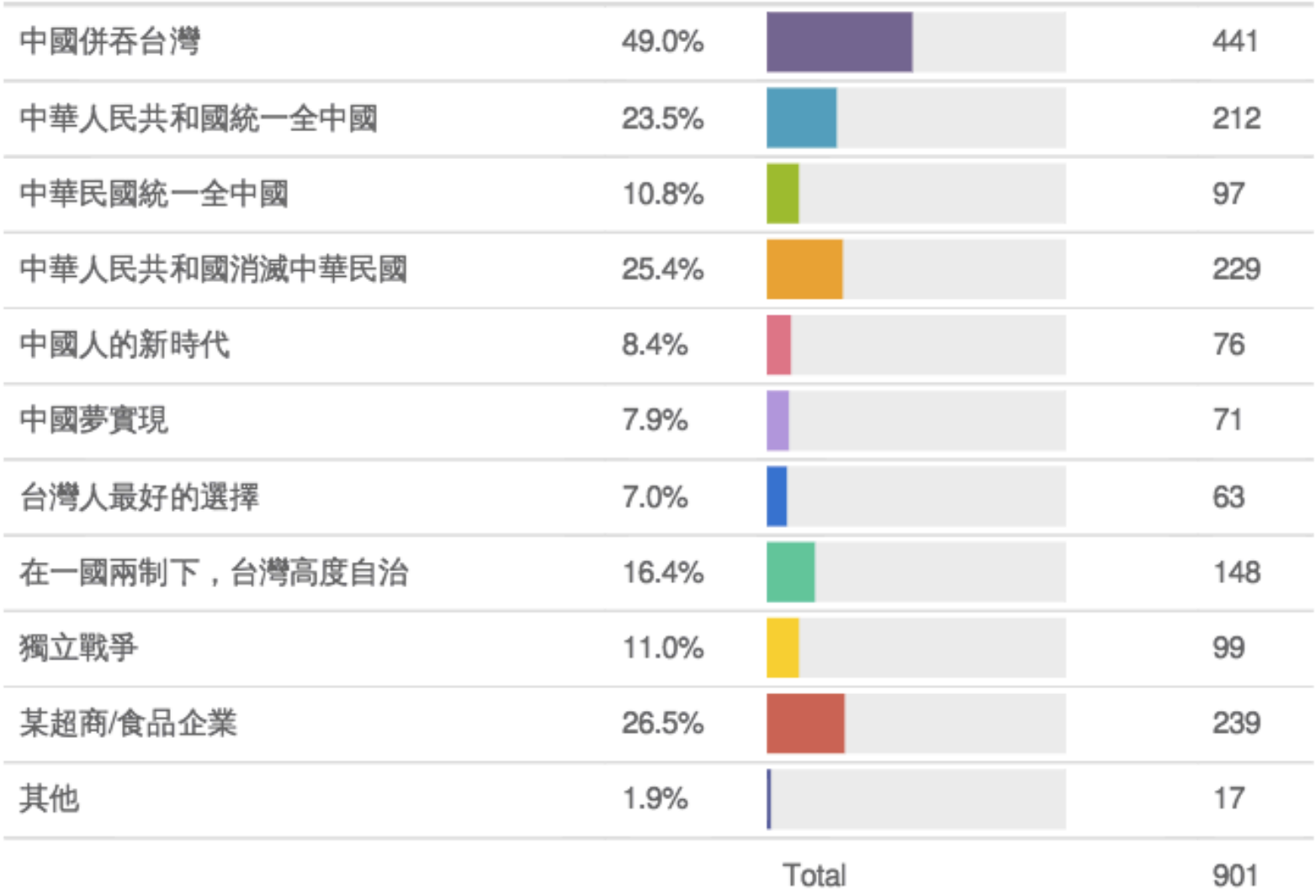
你有想過，台灣民眾對於「獨立」的定義有很多種，而且很可能沒有什麼共識嗎？

19. 說到獨立，您最先想到什麼？請從下面挑出最符合這個想法的選項。（可複選）



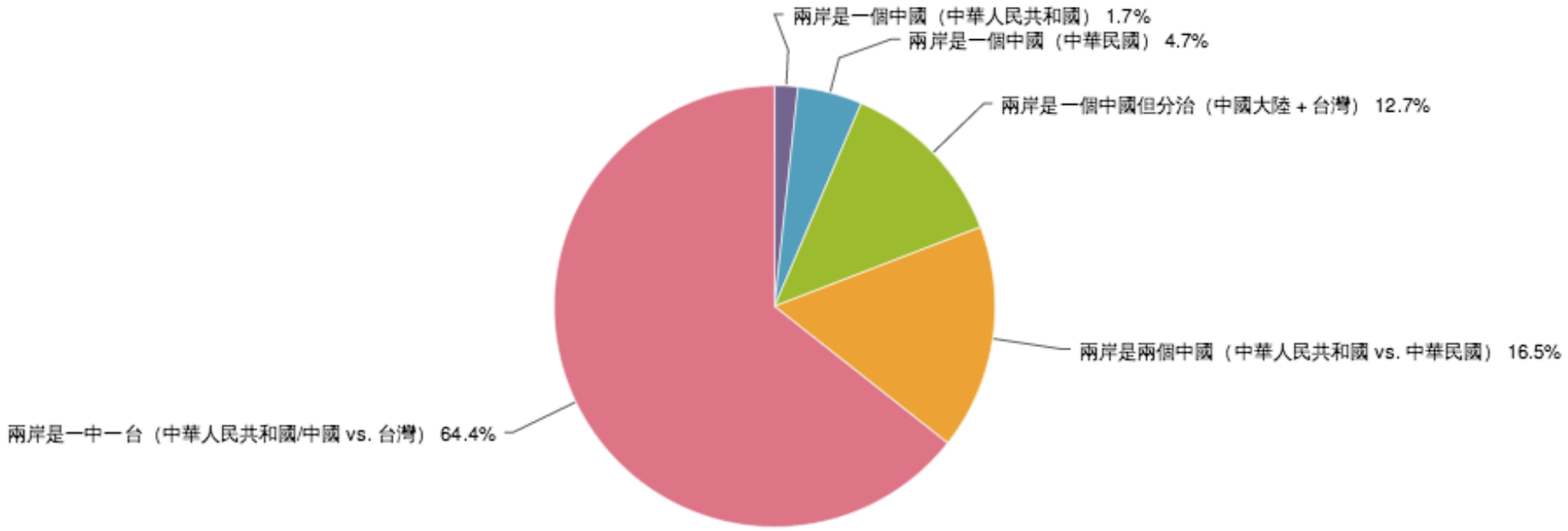
台灣是個新國家	14.3%	129
中華民國是個主權國家	36.1%	326
領土上與中國大陸不相隸屬	32.8%	296
有自己的軍隊、司法、行政系統和國家符號（如國旗、國歌）	41.2%	372
台灣人自己的國家	49.7%	448
兩岸必然有戰爭	12.1%	109
台灣要切斷和中華民國的關係	9.8%	88
台灣要切斷和中華人民共和國的關係	14.3%	129
完全切斷與中國有關的東西	7.1%	64
中華人民共和國尊重與理解下台灣如國家般運作	18.6%	168
自主和尊嚴	12.3%	111
得到世界各國官方承認	49.9%	450
其他	0.2%	2
	Total	902

20. 說到統一，您最先想到什麼？請從下面挑出最符合這個想法的選項。（可複選）



中國併吞台灣	49.0%		441
中華人民共和國統一全中國	23.5%		212
中華民國統一全中國	10.8%		97
中華人民共和國消滅中華民國	25.4%		229
中國人的新時代	8.4%		76
中國夢實現	7.9%		71
台灣人最好的選擇	7.0%		63
在一國兩制下，台灣高度自治	16.4%		148
獨立戰爭	11.0%		99
某超商/食品企業	26.5%		239
其他	1.9%		17
		Total	901

25. 關於兩岸關係的說法很多種，那一個最符合您的想法？



Value	Percent	Count
兩岸是一個中國（中華人民共和國）	1.7%	14
兩岸是一個中國（中華民國）	4.7%	39
兩岸是一個中國但分治（中國大陸 + 台灣）	12.7%	105
兩岸是兩個中國（中華人民共和國 vs. 中華民國）	16.5%	136
兩岸是一中一台（中華人民共和國/中國 vs. 台灣）	64.4%	531
	Total	825

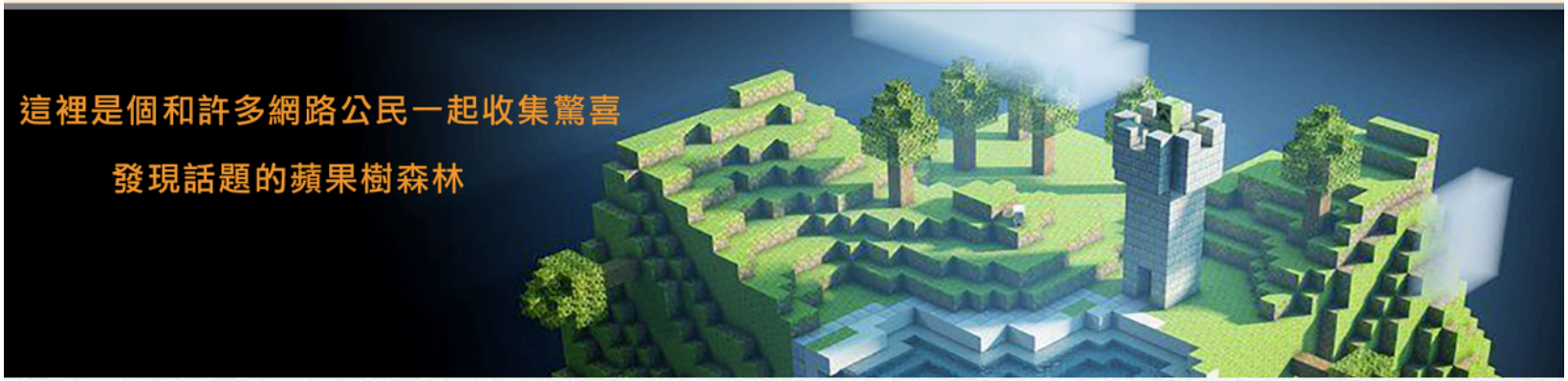
Statistics	
Sum	3,606.0
Average	4.4
StdDev	1.0
Max	5.0

看懂了之後，
你的發問可以天馬行空繼續下去

你會發現，問卷調查其實是可以讓你打造出專屬於自己研究領域厚資料礦脈的神器。

微笑小熊 調查小棧

首頁 會員登入 佈告欄 成果筆記 調查資料庫 關於 FB粉絲頁 RSS訂閱



最新消息

[新問卷:最熟悉的人PARTII] 爸爸，是男人最溫柔的...

[新問卷:最熟悉的人PartII] 爸爸，是男人最溫柔的名稱，對每個人來說，爸爸在心中都有一定的地位，但也有媽媽、姐姐..等其他家人也扮演著爸爸的角色。在父親節的前夕，無

進行中的訪問

7/15最熟悉的人PARTII | 博愛座博愛?

最新問卷：最熟悉的人 part II 17

2016-07-15

爸爸，是男人最溫柔的名稱，對每個

「在小熊身上得到更多其他人對社會的觀點，也改變自己不以自身想法為事件之看法。」 by 石凱云 (2016.05)

贊助單位

微笑小熊調查小棧

[smilepoll.tw]

我們的政治科學+資訊管理+行銷管理團隊
致力於發問及厚資料意義探勘的訓練及應用

~歡迎學術, 產學及官學合作 & 歡迎跨領域新星申請中山政治學研究所~

littlesmilebear@gmail.com

打造自己的社群網調平台的好處

- 資料科學家從資料聆聽者（被動爬梳挖來或買來的數據）轉換為資料創造者（主動收集到被研究對象價值和偏好）。
- 降低資料雜訊及更快速的決策。
- 形成社群後可以創造定群追蹤樣本（panel data），產生變數的合併帶來的巨大價值。
- 先以小數據作初探（pilot stud），之後再啟動隨機電話抽樣，將大幅增加推論力度。
- 初探階段便可以進行隨機分派實驗（A/A前測、A/B對照），找出意義和印證想法。

6

可否用電話及網路資料進行探索？
可以看到什麼？效果如何？

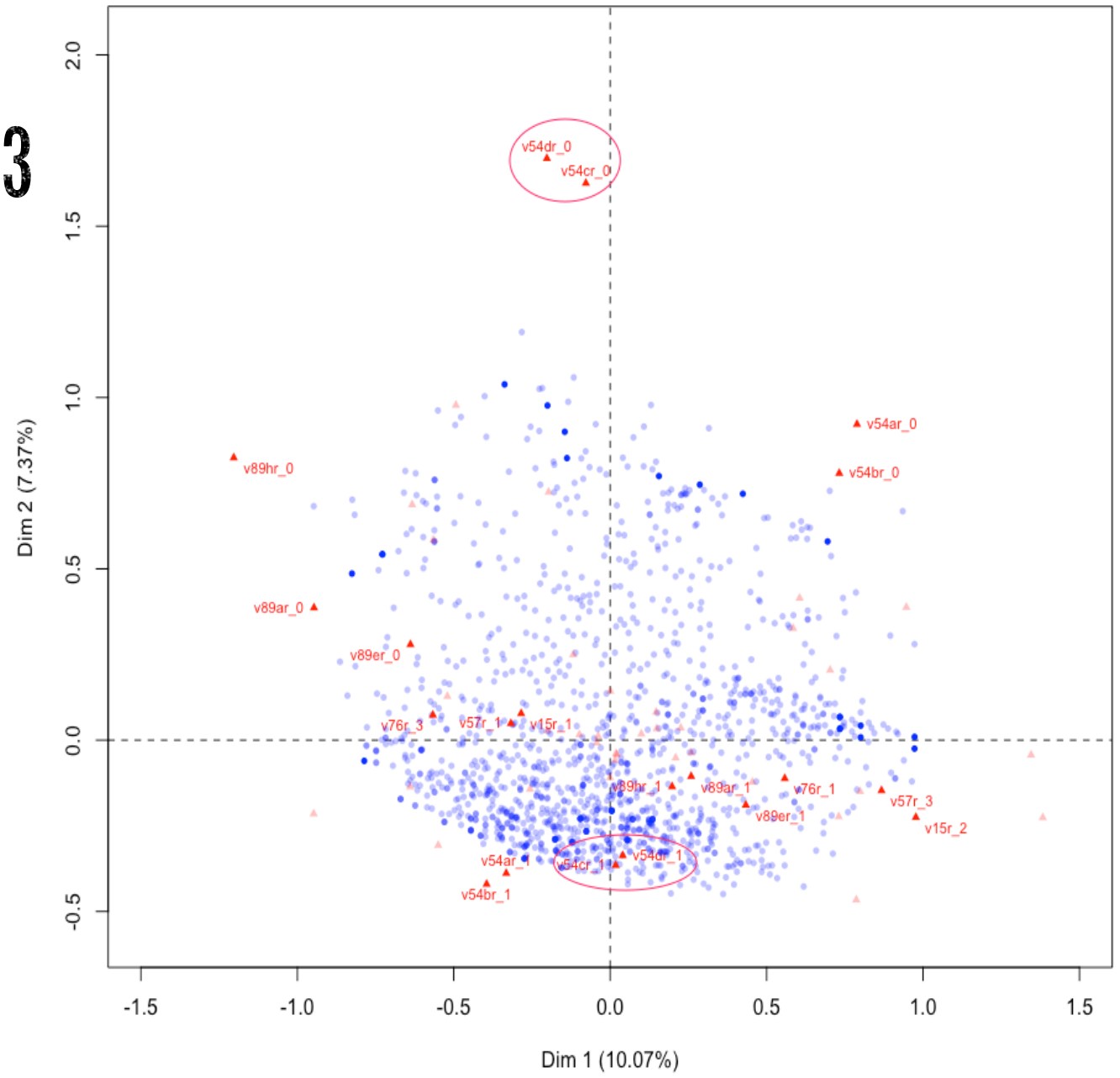
DATA SETS

- F2F Survey: Taiwan Social Change Survey 2013 (n=1,952) -- 上述使用的面訪資料
- CATI Telephone survey 2015 (n=1,100)
- Web panel 2015-2016 (n=468)

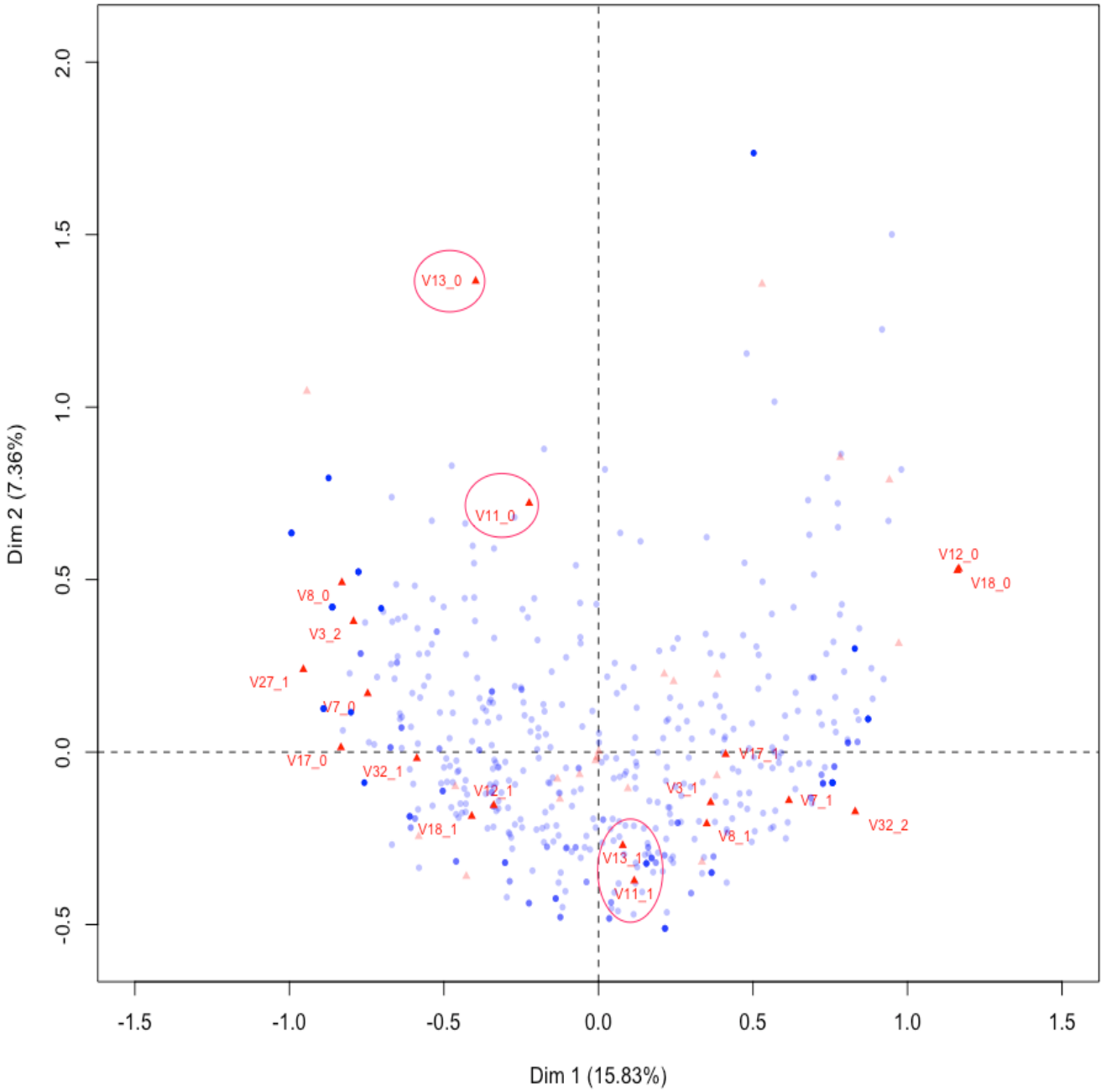
大多數問卷題測到的是民族認同，但還是可以看到新的概念-測量連結

有待我們進一步探索與打磨

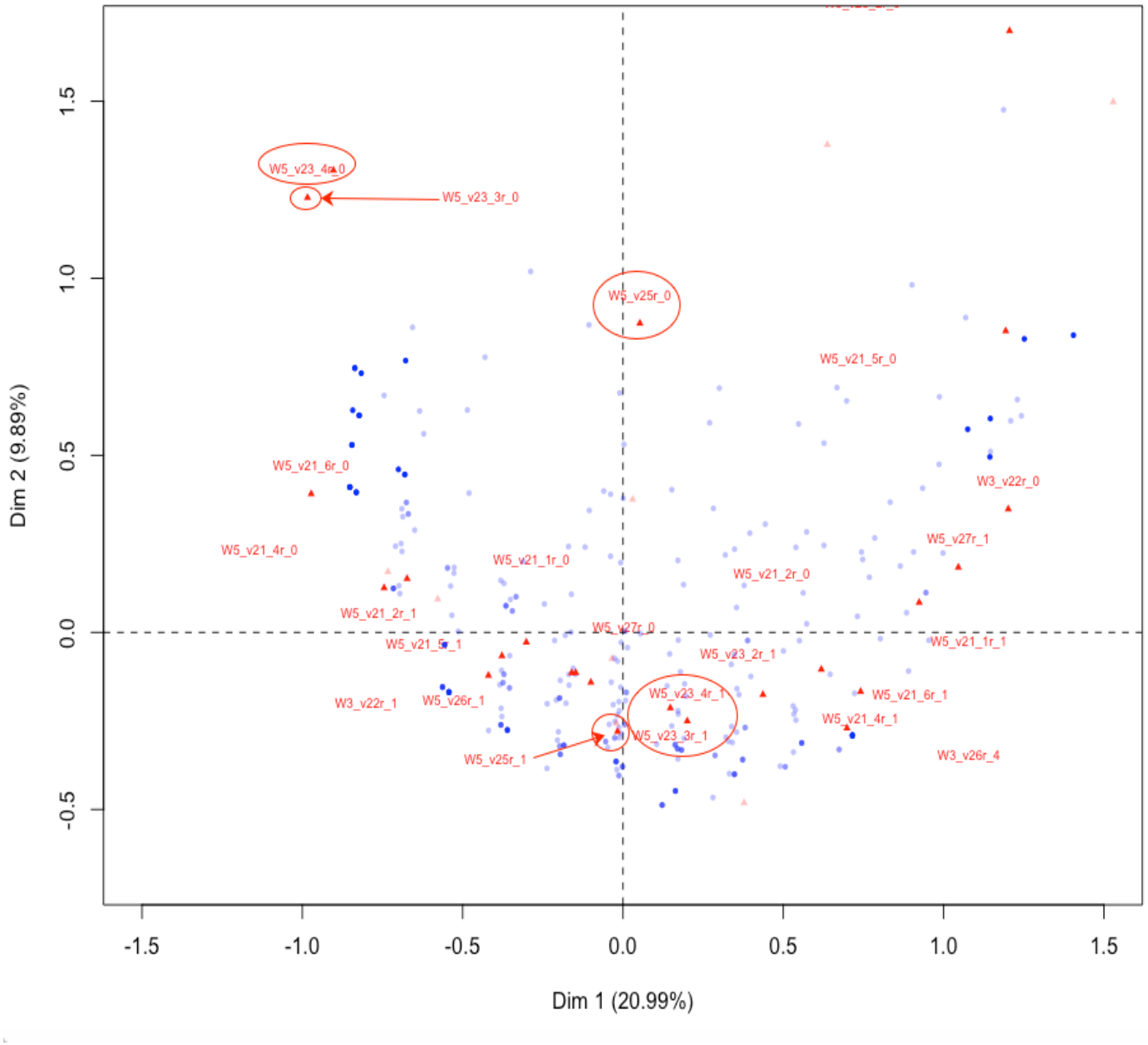
TSCS2013


v54dr_0
v54cr_0
v54ar_0
v54br_0
v89hr_0
v89ar_0
v89er_0
v76r_3
v57r_1
v15r_1
v89hr_1
v89ar_1
v76r_1
v57r_3
v89er_1
v15r_2
v54ar_1
v54cr_1
v54dr_1
v54br_1
Dim 2 (7.37%)
Dim 1 (10.07%)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

ID2015


V13_0
V11_0
V12_0
V18_0
V8_0
V3_2
V27_1
V7_0
V17_0
V32_1
V12_1
V18_1
V3_1
V17_1
V7_1
V32_2
V8_1
V13_1
V11_1
Dim 2 (7.36%)
Dim 1 (15.83%)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

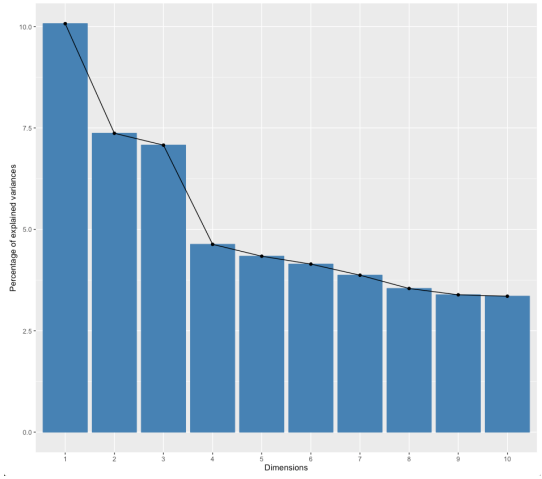
Webpanel 2015-2016

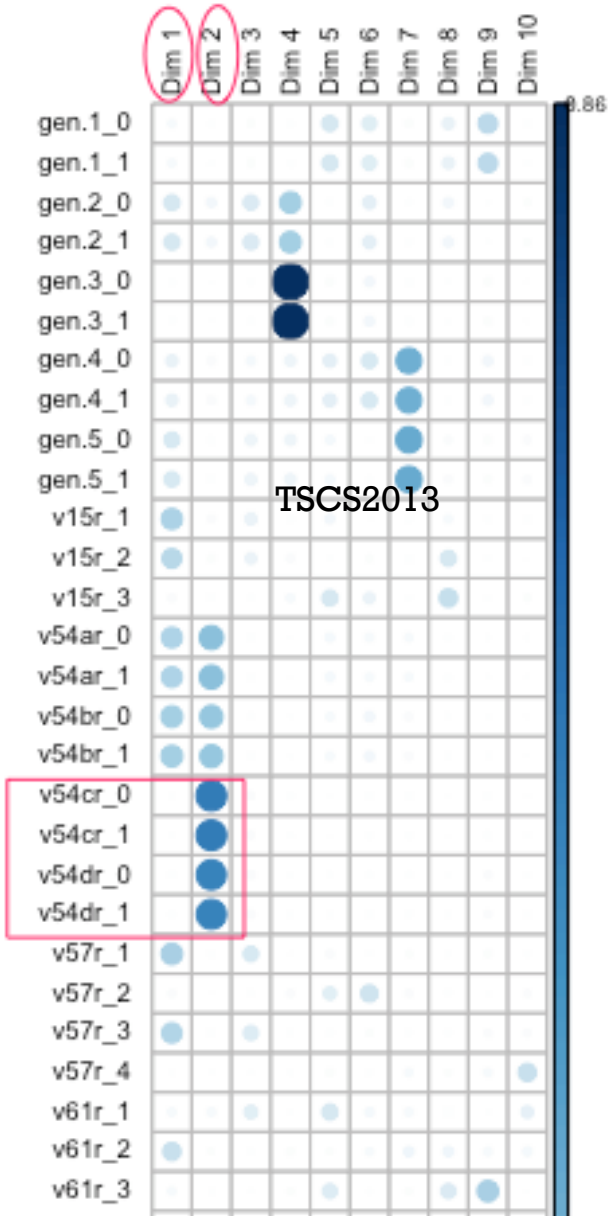


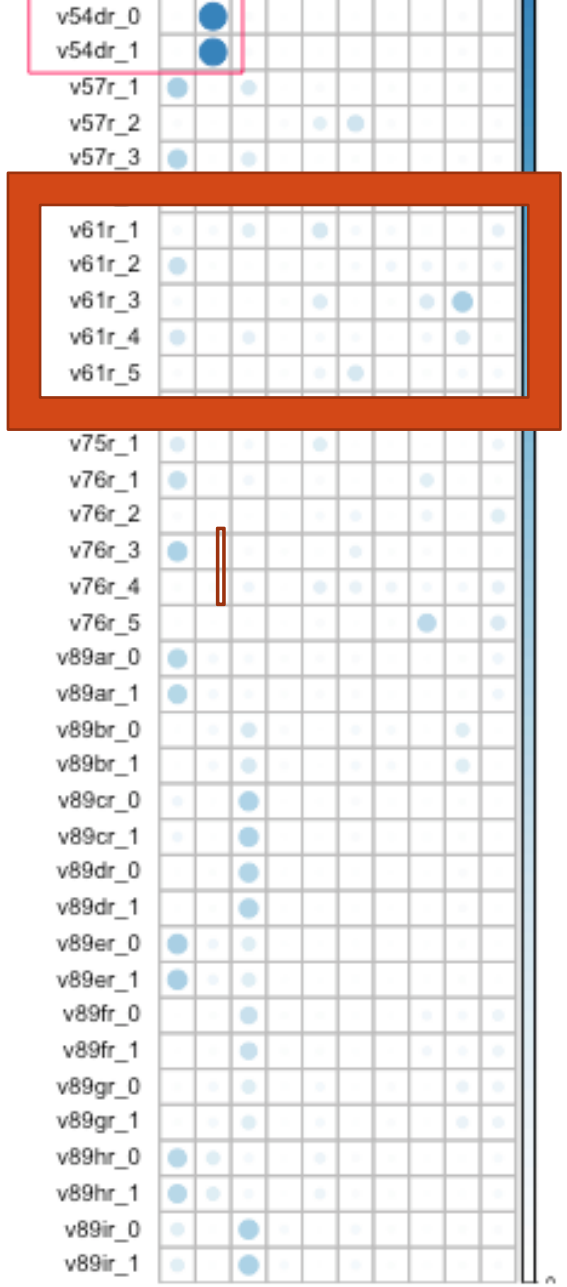
我還看到了當前統獨題的測量困境

The commonly used simple "unification/independence" question can NOT be grouped into any of the top 10 dimensions.

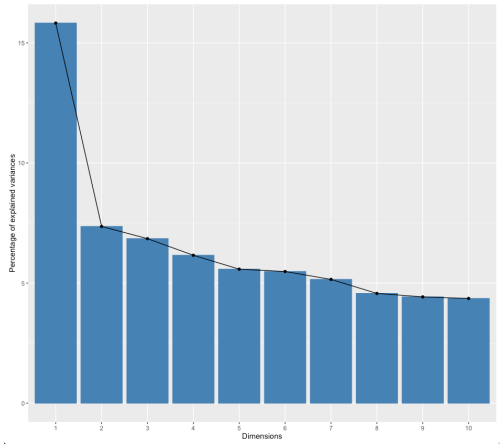
TSCS2013

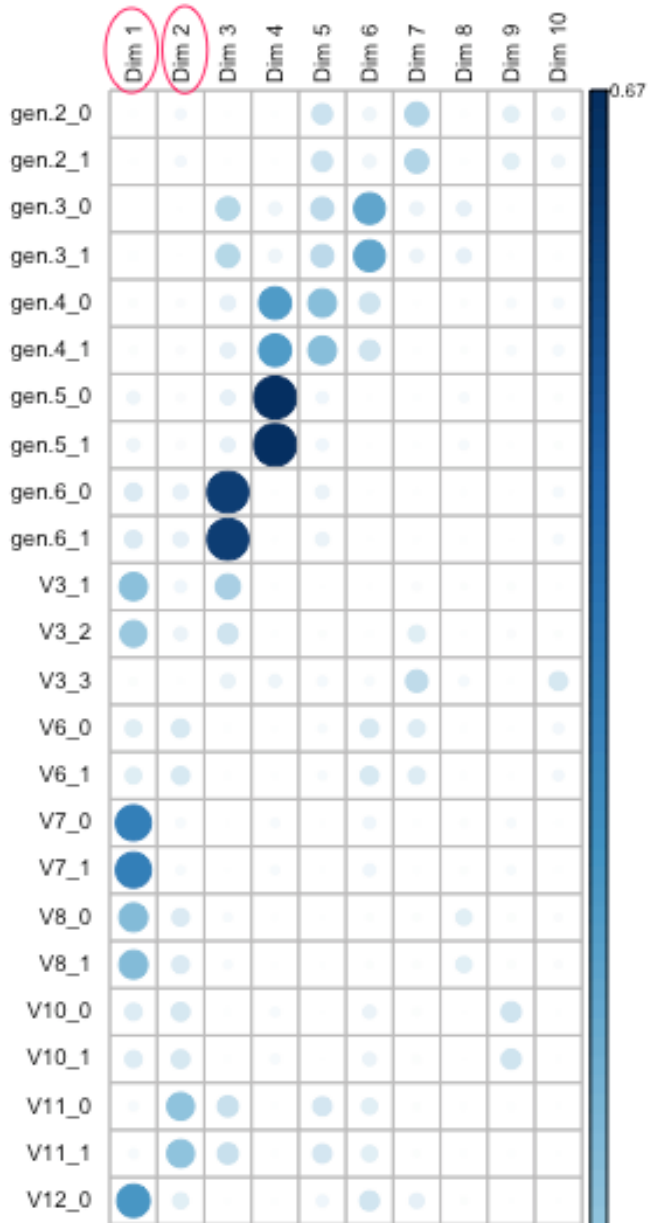

Percentage of explained variances
Dimensions

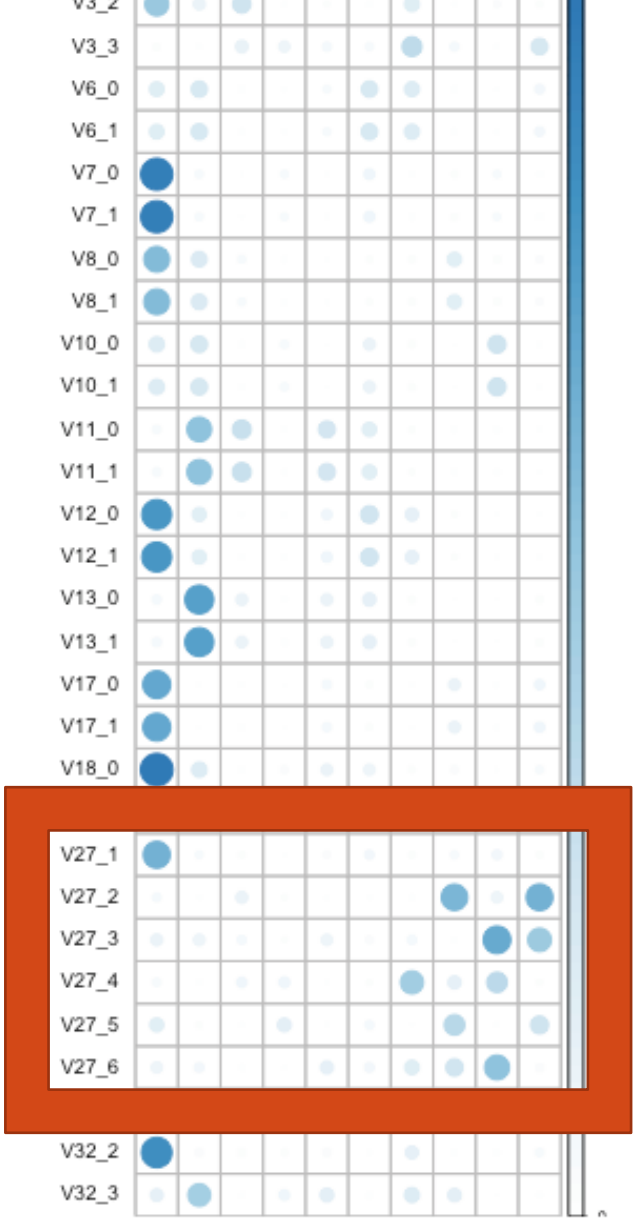

Dim 1
Dim 2
Dim 3
Dim 4
Dim 5
Dim 6
Dim 7
Dim 8
Dim 9
Dim 10
0.86
gen.1_0
gen.1_1
gen.2_0
gen.2_1
gen.3_0
gen.3_1
gen.4_0
gen.4_1
gen.5_0
gen.5_1
v15r_1
v15r_2
v15r_3
v54ar_0
v54ar_1
v54br_0
v54br_1
v54cr_0
v54cr_1
v54dr_0
v54dr_1
v57r_1
v57r_2
v57r_3
v57r_4
v61r_1
v61r_2
v61r_3
TSCS2013


v54dr_0
v54dr_1
v57r_1
v57r_2
v57r_3
v61r_1
v61r_2
v61r_3
v61r_4
v61r_5
v75r_1
v76r_1
v76r_2
v76r_3
v76r_4
v76r_5
v89ar_0
v89ar_1
v89br_0
v89br_1
v89cr_0
v89cr_1
v89dr_0
v89dr_1
v89er_0
v89er_1
v89fr_0
v89fr_1
v89gr_0
v89gr_1
v89hr_0
v89hr_1
v89ir_0
v89ir_1

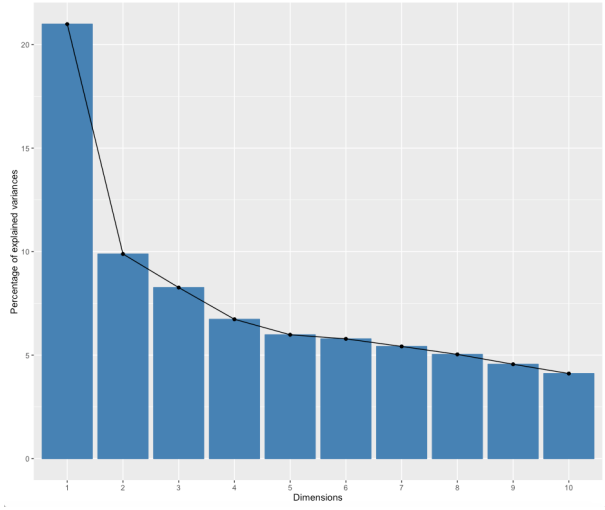
ID2015

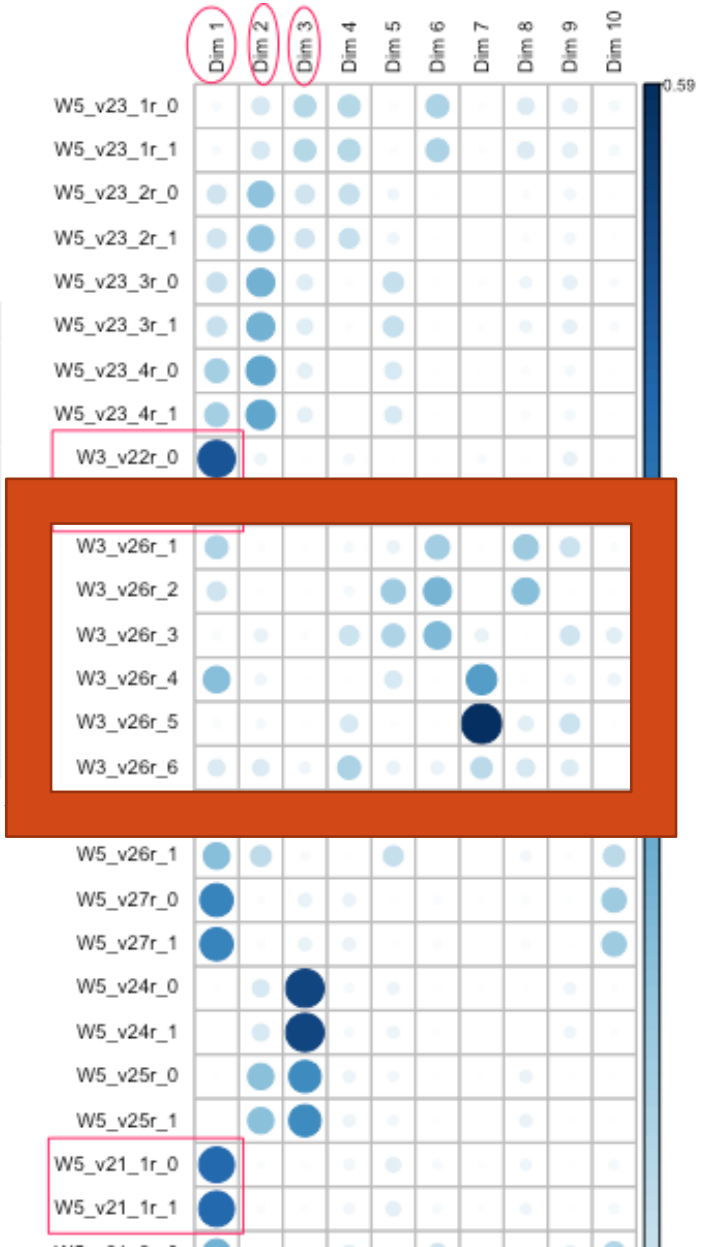

Percentage of explained variances
Dimensions


Dim 1
Dim 2
Dim 3
Dim 4
Dim 5
Dim 6
Dim 7
Dim 8
Dim 9
Dim 10
0.67
gen.2_0
gen.2_1
gen.3_0
gen.3_1
gen.4_0
gen.4_1
gen.5_0
gen.5_1
gen.6_0
gen.6_1
V3_1
V3_2
V3_3
V6_0
V6_1
V7_0
V7_1
V8_0
V8_1
V10_0
V10_1
V11_0
V11_1
V12_0


V3_3
V6_0
V6_1
V7_0
V7_1
V8_0
V8_1
V10_0
V10_1
V11_0
V11_1
V12_0
V12_1
V13_0
V13_1
V17_0
V17_1
V18_0
V27_1
V27_2
V27_3
V27_4
V27_5
V27_6
V32_2
V32_3

WEBPANEL 2015-2016





野人獻曝 敬請多多指教

I do hope this thick data approach and the application of MCA are more than just interesting to you.

非常感謝全球R社群的奉獻，以及國內資料科學社群的努力！

劉正山 csliu@mail.nsysu.edu.tw

THICK DATA

(APPROACH)

資料科學中的厚資料視野

FB: thickdatabarbor/資料吼

參考資料

- Blasius, J., & Greenacre, M. (Eds.). (2014). *Visualization and Verbalization of Data*. CRC Press.
- Husson, F., Le, S., & Pages, J. (2010). *Exploratory Multivariate Analysis by Example Using R* (1 edition). CRC Press.
- Pagès, J. (2014). *Multiple Factor Analysis by Example Using R* (1 edition). Boca Raton: Chapman and Hall/CRC.
- Pasek, J., Jang, S. M., Cobb, C. L., Dennis, J. M., & Disogra, C. (2014). Can marketing data aid survey research? Examining accuracy and completeness in consumer-file data. *Public Opinion Quarterly*, *78*(4), 889–916.
- Roux, B. L., & Rouanet, H. (2009). *Multiple Correspondence Analysis*. SAGE Publications.

同場
加映

從探索的角度重新發掘民調市
調資料在意義探勘上的潛力

新一代的「厚」資料收集流程

- 一：以探索的角度設計問卷 [關鍵 & 最難]
- 二：收集資料（面訪、電話、網路）
- 三：描述資料
- 四：分析、視覺化 & 判讀（說故事）[新!]

動機

市調與行銷的資料科學家，除了「描述」和「解釋」，現在開始，可以加上「探索」

發問

將幾想知道的面向轉為題組，是的，聽起來簡單。
but 你真的是那個能夠指出國王新衣的好奇寶寶嗎？

分析

用MCA發掘關聯，你從小數據看到的樣貌，跟使用大數據分析所看見的，會產生高度互補效果。

詮釋

盯著客觀的資料分析結果，把你看見的故事和意義說出來。
這必需要回到你對於自己問的問題瞭解的程度，以及自己專業領域訓練的視野。

結語：

LET'S THINK AGAIN: DATA MINING FOR WHAT?

PATTERNS & MEANINGS!

小數據的意義探勘可以是資料科學重要的一環。它將是社科人及民調/市調專業者踏入資料科學領域的彩虹橋，也將是資料科學吸納更多調查及傳播專業人才的磁石。

當資料取得及技術變得平民化，發掘意義的能力和訓練，將變得與技術能力的訓練一樣重要。

只是，這種抽取出意義的能力往往是經歷過專業訓練以及對產業及世界的觀察與思考（以及人生起伏）之後才累積出的能力。基本上可以透過閱讀及學術訓練取得。高階經理人尤其需要這種訓練與能力。

要注意的是，除了呈現分析結果這個步驟之外，整個研究過程非常主觀。而這正是大數據分析的知識論立場，無可厚非。

因此，若要讓開拓性的研究途徑成為資料科學的一環，資料科學家必須保有科學家**open to challenge**的精神，虛心地確保每一個分析環節及結果詮釋都透明，並接受社群的檢驗及論辯。換言之，本講所呈現的，還未達到所謂的「真相」或「事實」。真相是逐漸被「逼進」而顯示出來的。沒有人能一步到位，或是一次就宣稱拿到了聖杯。