

# 世代之爭爭什麼？

使用R套件探索民意調查資料的意義

劉正山

中山大學政治學研究所 教授

Director, Smilepoll.tw

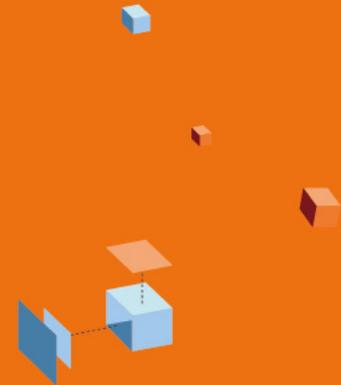
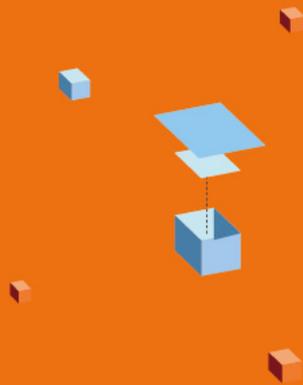
@ R Taiwan 2016

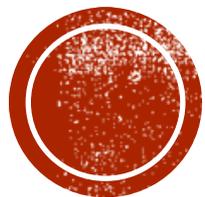
2016.12.23



# 1

話說...  
一年前的自我對話





大數據分析的探索精神，  
小數據的擁有者沒有嗎？

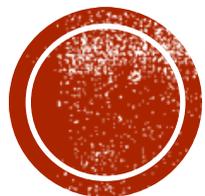
**(OF COURSE YES; WE HAVE IT.)**

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）看似只能做描述統計或兩兩之間的相關分析；技術含金量有限。





只要問了**好問題**，並運用探索工具**MCA**，  
民調市調資料將與大數據同樣珍貴。

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

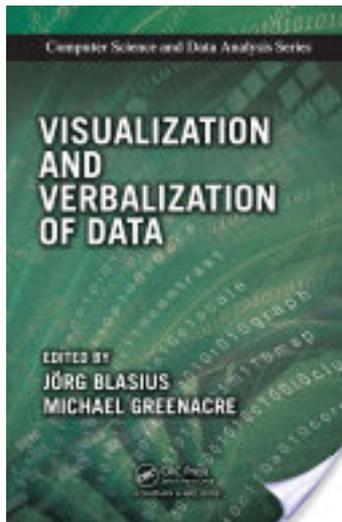
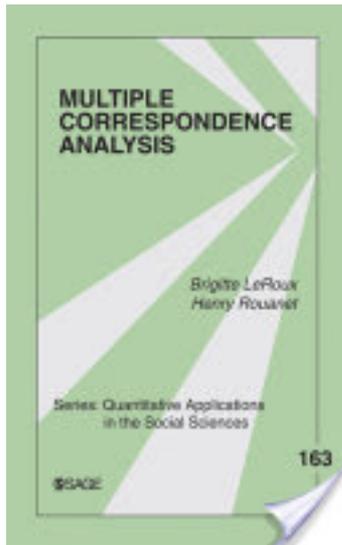
## MCA 方法帶來的新視野



# 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。





Copyrighted Material

# The R Series

## Multiple Factor Analysis by Example Using R

Multiple Factor Analysis (MFA)

3 quantitative groups

Chemical Sensors

Overall assessment

Add quantitative group

Modify 1 group

Delete

Qualitative groups

Add qualitative group

Modify 1 group

Delete

Select supplementary individuals

Graphical options

Outputs

Restart

Main options

Name of the result object:	res
Number of dimensions:	5
Select the dimensions for the graph:	1 2

Perform Clustering after MFA

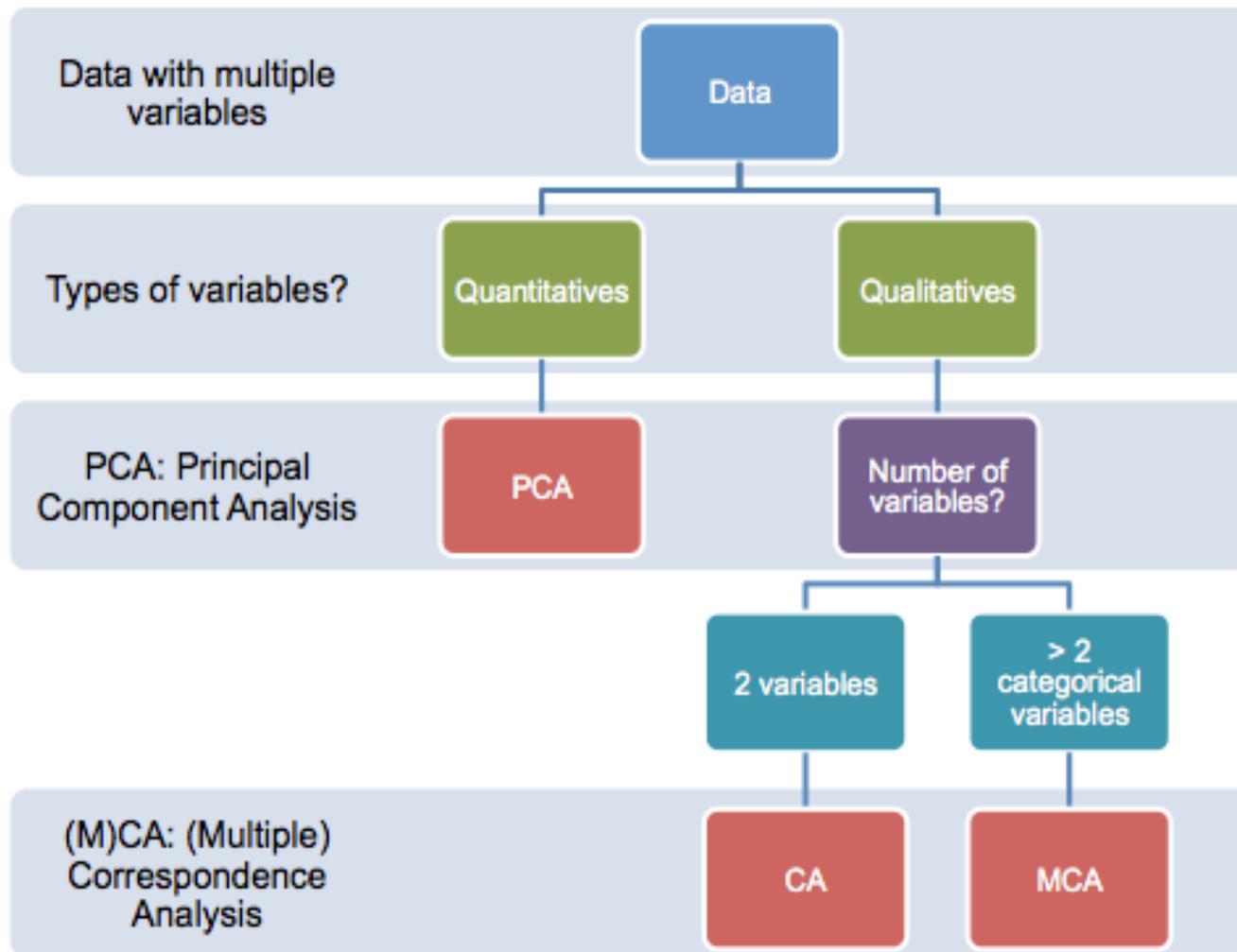
Jérôme Pagès

CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

Copyrighted Material





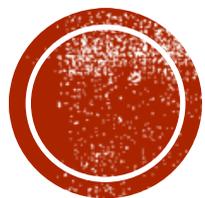
R packages for the analyses: **FactoMineR** (PCA, CA, MCA); **ade4** (PCA, CA, MCA); **stats** (PCA); **ca** (CA); **MASS** (CA)

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



# 研究問題：

民眾看似多重的民族、國家、政黨認同之間，能夠說清楚它們之間的關聯嗎？



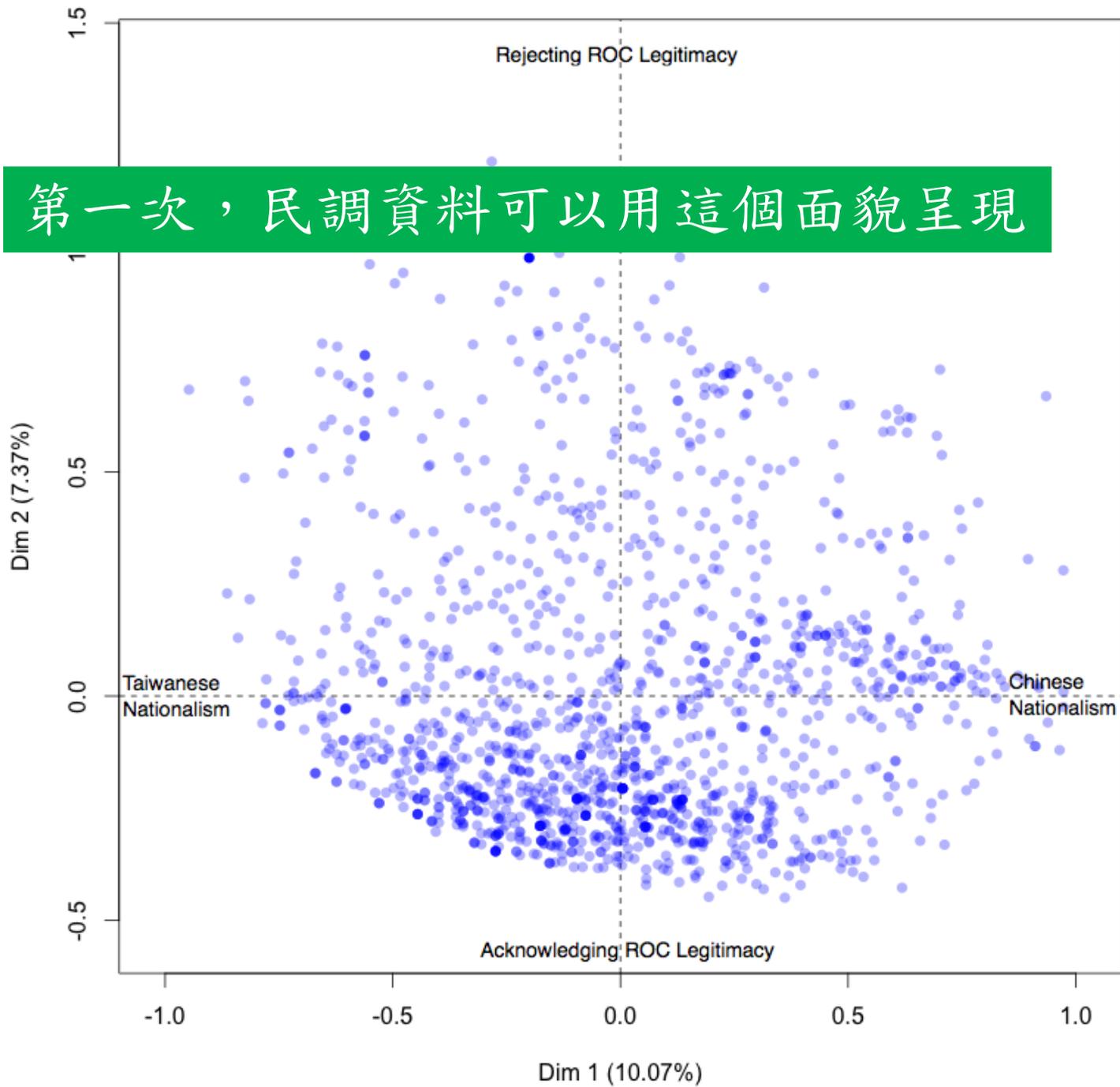
我們知道世代之間的政治傾向上不同，到底怎麼呈現出來才能解答那些是真相，那些是誤解？

# 拿國內權威資料來分析

- 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。



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	<input type="checkbox"/> (01)	<input type="checkbox"/> (02)	<input type="checkbox"/> (03)	<input type="checkbox"/> (04)	<input type="checkbox"/> (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	<input type="checkbox"/> (01)	<input type="checkbox"/> (02)	<input type="checkbox"/> (03)	<input type="checkbox"/> (04)	<input type="checkbox"/> (05)
(c)現在的台灣文化已經不能再說是中國文化的一部分	<input type="checkbox"/> (01)	<input type="checkbox"/> (02)	<input type="checkbox"/> (03)	<input type="checkbox"/> (04)	<input type="checkbox"/> (05)
(d)台灣是個小而美的國度，未來也都會繼續維持下去	<input type="checkbox"/> (01)	<input type="checkbox"/> (02)	<input type="checkbox"/> (03)	<input type="checkbox"/> (04)	<input type="checkbox"/> (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	<input type="checkbox"/> (01)	<input type="checkbox"/> (02)	<input type="checkbox"/> (03)	<input type="checkbox"/> (04)	<input type="checkbox"/> (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	<input type="checkbox"/> (01)	<input type="checkbox"/> (02)	<input type="checkbox"/> (03)	<input type="checkbox"/> (04)	<input type="checkbox"/> (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	<input type="checkbox"/> (01)	<input type="checkbox"/> (02)	<input type="checkbox"/> (03)	<input type="checkbox"/> (04)	<input type="checkbox"/> (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	<input type="checkbox"/> (01)	<input type="checkbox"/> (02)	<input type="checkbox"/> (03)	<input type="checkbox"/> (04)	<input type="checkbox"/> (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	<input type="checkbox"/> (01)	<input type="checkbox"/> (02)	<input type="checkbox"/> (03)	<input type="checkbox"/> (04)	<input type="checkbox"/> (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)
```

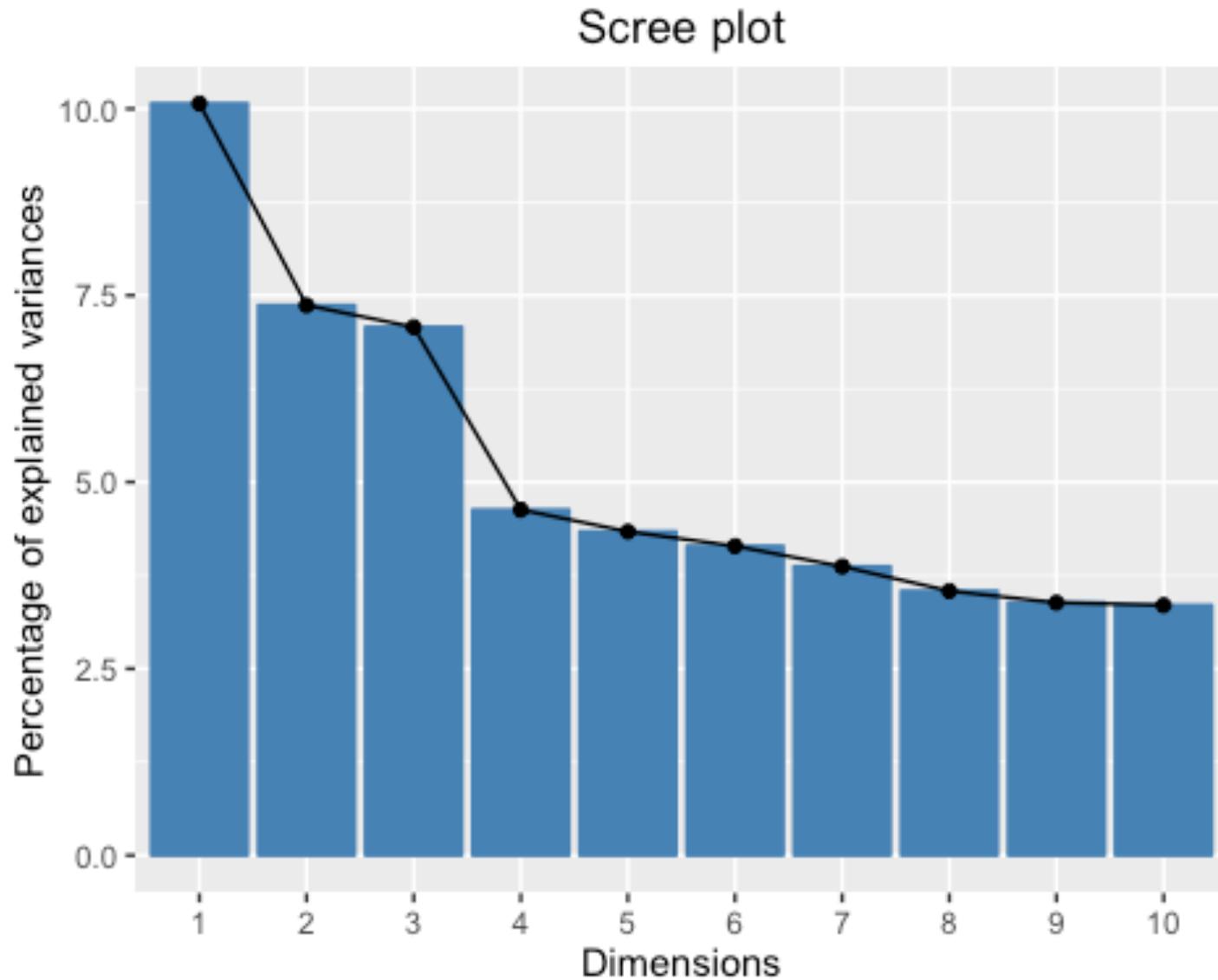
```
> nrow(tscs2013forMCA.nona)  
[1] 1496
```

```
> 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_screplot(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.quant.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="")
```





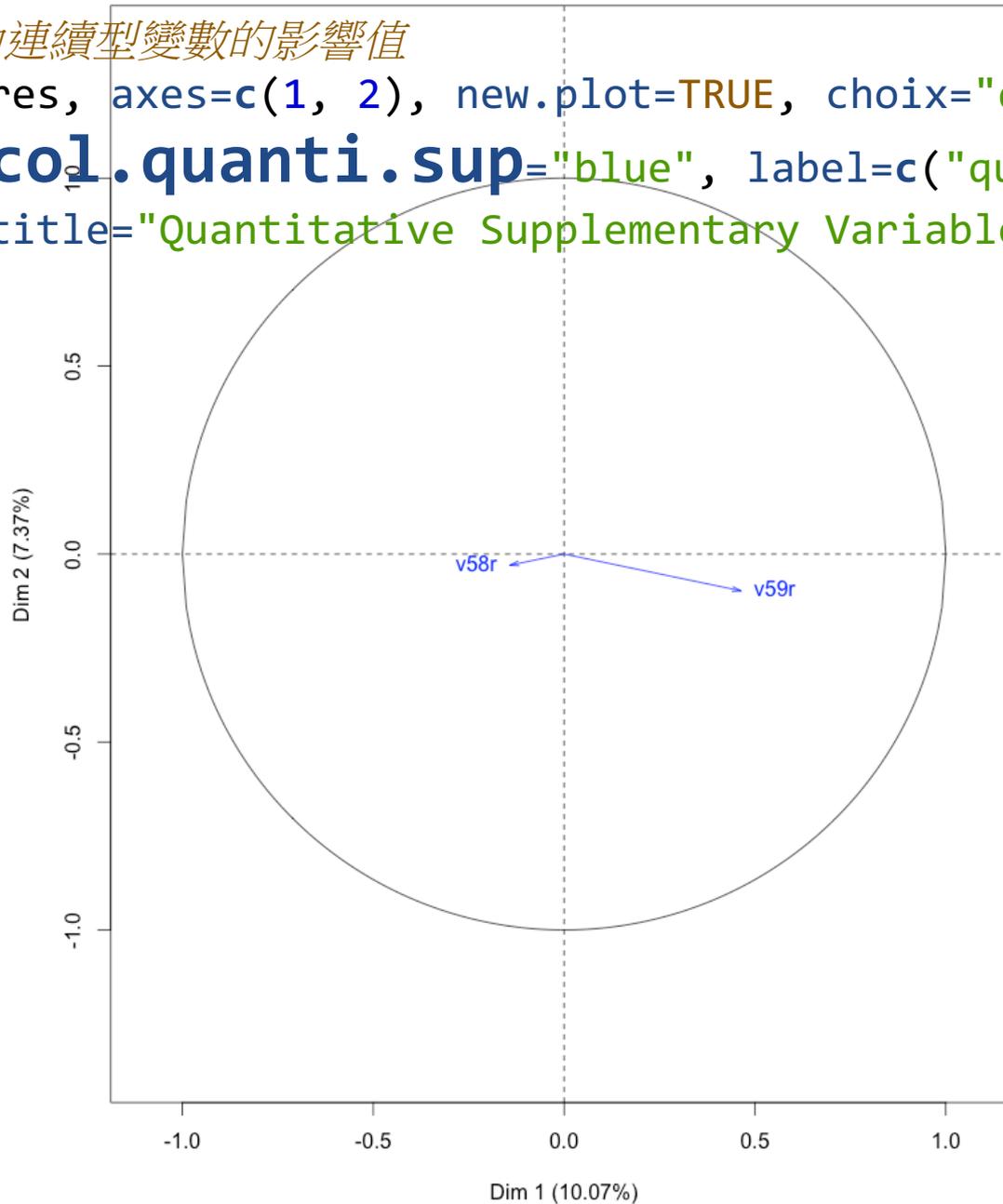


## 注意：構成第二維次（Y軸）的因素

- 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

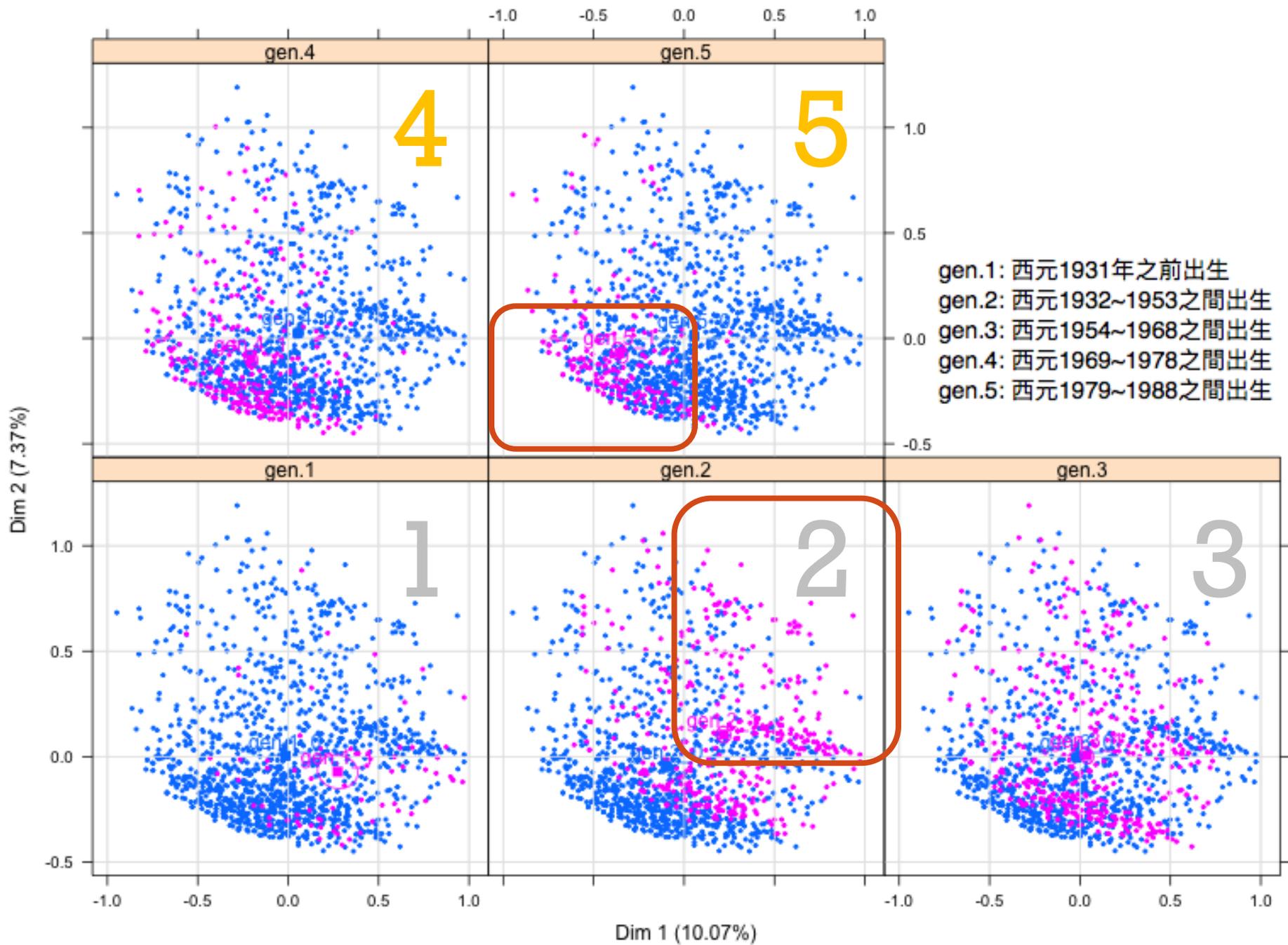
第一軸線的代表概念：**民族認同**（中華民族或台灣民族）

第二軸線的代表概念：**國家正當性**（接受中華民國與否）

## 世代分佈的差異

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





# 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。



# 小數據也能預測 & 協助決策

- 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。

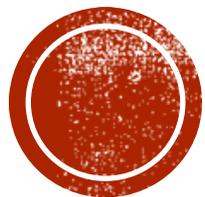




再看一眼，猜猜看：  
什麼是最能爭取到多數選民接受的政治語言？

維·持·現·狀





# 你說

「這去年我就知道了，沒什麼了不起」

以上的所有資訊，都是用**2013**年的資料就可以挖掘出來的的意義哦。

若我們**2016**年再收集一次資料，再來分析一次，你又會看見什麼？要不要一起來解答「太平”礁“事件對台灣民心的衝擊」呢？

Dim 2 (7.37%)

1.0

0.5

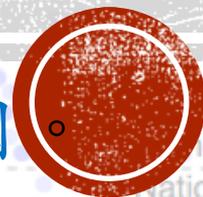
0.0

-0.5

那麼問難一點的題目，我問：  
「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**

要如何團結（或分裂）台灣民眾，方法和策略已經浮現。



Chinese Nationalism

Acknowledging PRC Legitimacy

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。



# 舉一反三

現在，若我們能夠將這整套**MCA**方法用於瞭解其他民眾的行為...

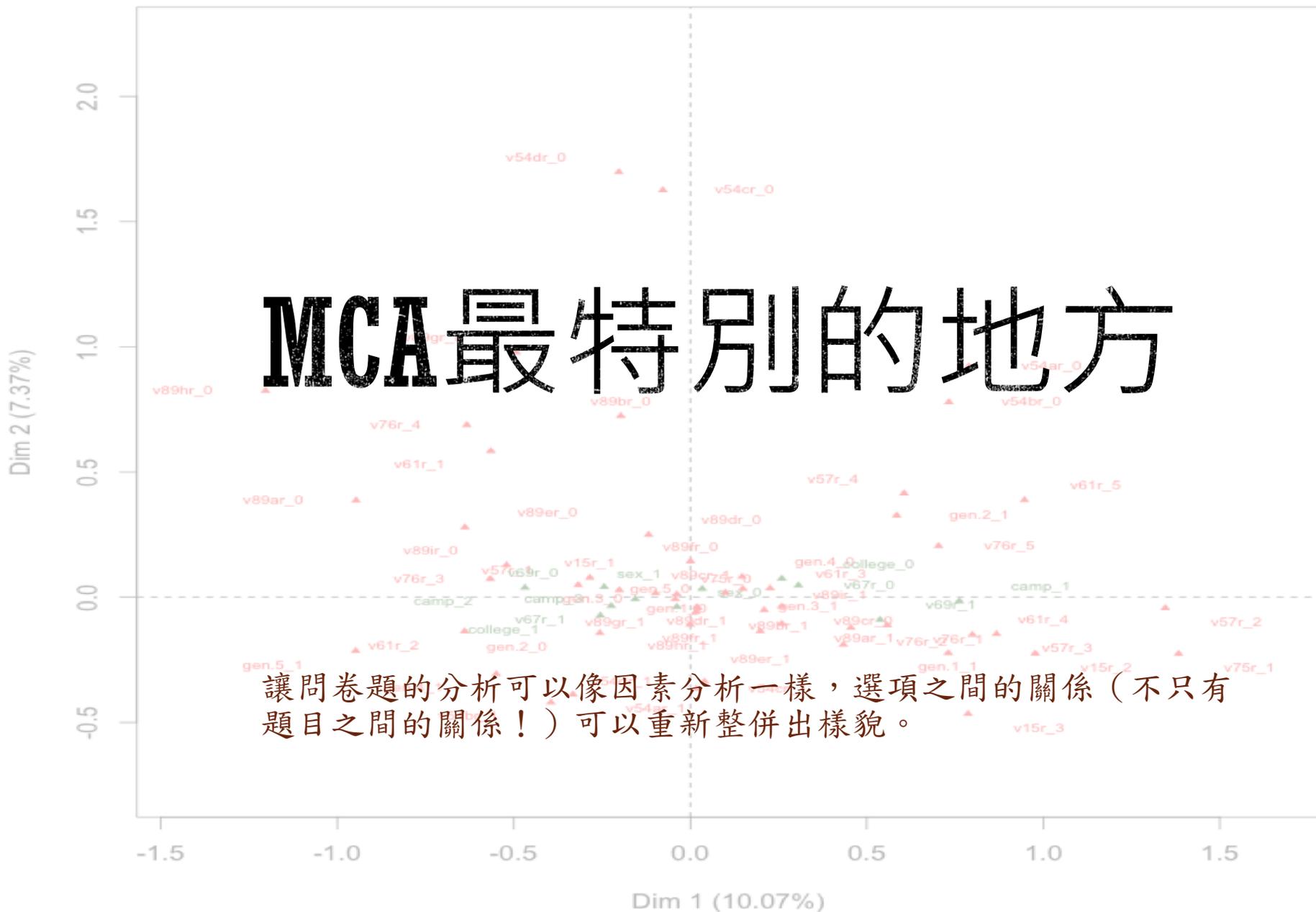




那我們就發 (現新大陸) 了 😊

# 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。



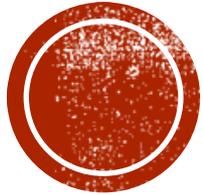


# 3

重新看待民調市調資料的價值



結語：

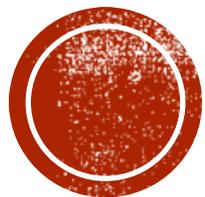


**LET'S THINK AGAIN:  
DATA MINING FOR WHAT?**

# PATTERNS & MEANING!

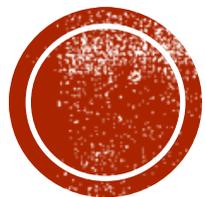
小數據的意義探勘可以是資料科學重要的一環。它將是社科人及民調/市調專業者踏入資料科學領域的彩虹橋，也將是資料科學吸納更多調查及傳播專業人才的磁石。





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。

# THICK DATA (APPROACH)

資料科學中的 **厚資料** 視野

FB: [thickdatabarbor/](https://www.facebook.com/thickdatabarbor/) **資料吼**



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](mailto:csliu@mail.nsysu.edu.tw)



# 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.



# 資料檔及講義

- 範例資料檔(rda)下載：<http://jmp.sh/pvLZWYE>
- R操作講義下載：<http://jmp.sh/UHJd7OV>



# 同場 加映

如何動手收集價值型的厚資料



# 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。



這裡是個和許多網路公民一起收集驚喜  
發現話題的蘋果樹森林



最新消息

[新聞卷:最熟悉的人PARTII] 爸爸，是男人最溫柔的...

[新聞卷: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)



# 微笑小熊調查小棧

[ [smilepoll.tw](http://smilepoll.tw) ]

~歡迎學術, 產學及官學合作 &  
歡迎跨領域新星申請中山政研究所~

