

從政治科學接軌資料科學： 社科學子如何透過工具升級心態

劉正山

國立中山大學 政治學研究所

2018.11.5 @國防大學政戰學院



從政治科學看 資料科學

重點是政治、資料、還是科學？

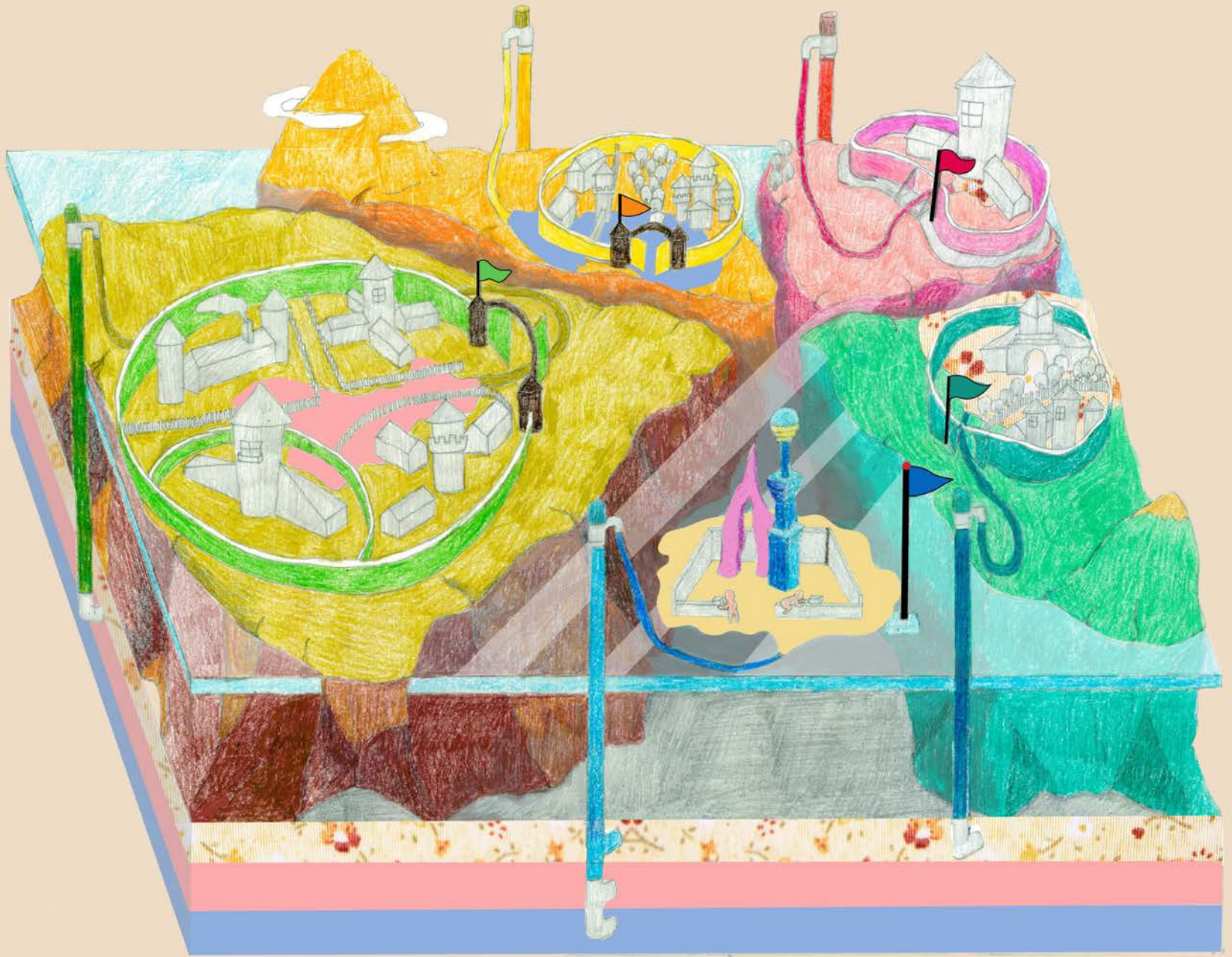
本講大綱

- 鳥瞰政治科學的方法論
- 資料科學興起對政治學門的挑戰及機會
- 探索的方法及工具
- 從工具的學習來提升心態
- 結語及答問



鳥瞰政治科學的方法論

實證主義者、實存主義者、詮釋主義者，還是實用主義者？

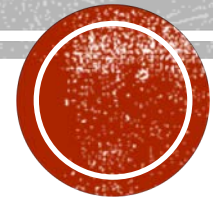


Illustrated by Prof. Chengshan (Frank) LIU

WHAT IS “BIG DATA”?

5Vs: Big volume, velocity, variety, veracity, and value.

Honestly, this term has gone out of fashion.



政治學者談「大數據」時用那些工具與方法？

- In our field "data-driven" and "method-driven" research works are labelled as "big data" studies.
- Methods that are associated with "big data"
 - Text-mining (文本探勘) ,
 - data-mining (資料探勘) ,
 - automatic content analysis (自動內容分析) ,
 - computer-assisted text analysis (電腦輔助文本分析) ,
 - automatic annotation (自動附記) ,
 - sentiment analysis (情緒分析) ,
 - geographic information system (地理資訊系統)
 - network analysis (網絡分析) 等等。



主題一：資訊工具在社科(政治)應用概論

- 2010. “A Method of Automated Nonparametric Content Analysis for Social Science.”
- 2012. “Social Science Research Methods in Internet Time.
- 2014. “Restructuring the Social Sciences: Reflections from Harvard’s Institute for Quantitative Social Science.”
- 2015. “Computer-Assisted Text Analysis for Comparative Politics.”
- 2015. “No! Formal Theory, Causal Inference, and Big Data Are Not Contradictory Trends in Political Science.”
- 2015. “We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together.”
- 2015. “Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites.”
- 2016. “Machine Translation: Mining Text for Social Theory.”



主題二：公共言論趨勢之辨識或追蹤

- 2008. “Recognizing Citations in Public Comments.”
- 2008. “Parsing, Semantic Networks, and Political Authority Using Syntactic Analysis to Extract Semantic Relations from Dutch Newspaper Articles.”
- 2008. “Good News or Bad News? Conducting Sentiment Analysis on Dutch Text to Distinguish Between Positive and Negative Relations.”
- 2008. “Media Monitoring by Means of Speech and Language Indexing for Political Analysis.”
- 2012. “Media Coverage in Times of Political Crisis: A Text Mining Approach.”
- 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.”
- 2014. “Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data.”
- 2017. “Critical News Reading with Twitter? Exploring Data-mining Practices and their Impact on Societal Discourse.”



其他主題 (三~五)

■ 主題三：政治立場的辨識/追蹤

- 2003. “Extracting Policy Positions from Political Texts Using Words as Data.”
- 2008. “A Scaling Model for Estimating Time-series Party Positions from Texts.”
- 2014. “Scaling Politically Meaningful Dimensions Using Texts and Votes.”
- 2015. “Quantifying Social Media’s Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook.”

■ 主題四：政治言論的管制策略

- 2013. “How Censorship in China Allows Government Criticism but Silences Collective Expression.”
- 2013. Media Commercialization & Authoritarian Rule in China.
- 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument."

■ 主題五：公共政策形成之探討

- 2005. “Using Geographic Information Systems to Study Interstate Competition.”
- 2014. “‘Big Data’ in Research on Social Policy.”
- 2015. “Analyzing Big Data: Social Choice and Measurement.”



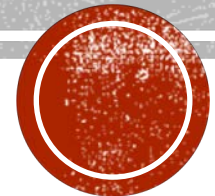
其他主題（六~八）

- 主題六：政治言論的語意分析
 - 2008. “Automatic Annotation of Semantic Fields for Political Science Research.”
 - 2015. “Uncovering Social Semantics from Textual Traces: A Theory Driven Approach and Evidence from Public Statements of US Members of Congress.”
- 主題七：政治選舉的運用
 - 2014. “Political Campaigns and Big Data.”
 - 2017. “The Pulse of the People: Can internet data outdo costly and unreliable polls in predicting election outcomes?”
- 主題八：國際關係研究
 - 2012. “Richardson in the Information Age: Geographic Information Systems and Spatial Data in International Studies.”



WHY (NOT) BIG DATA?

Your epistemological and methodological stances and attitudes toward methods decide how you evaluate (if not disdain) “big data”.

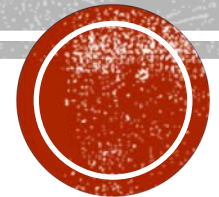


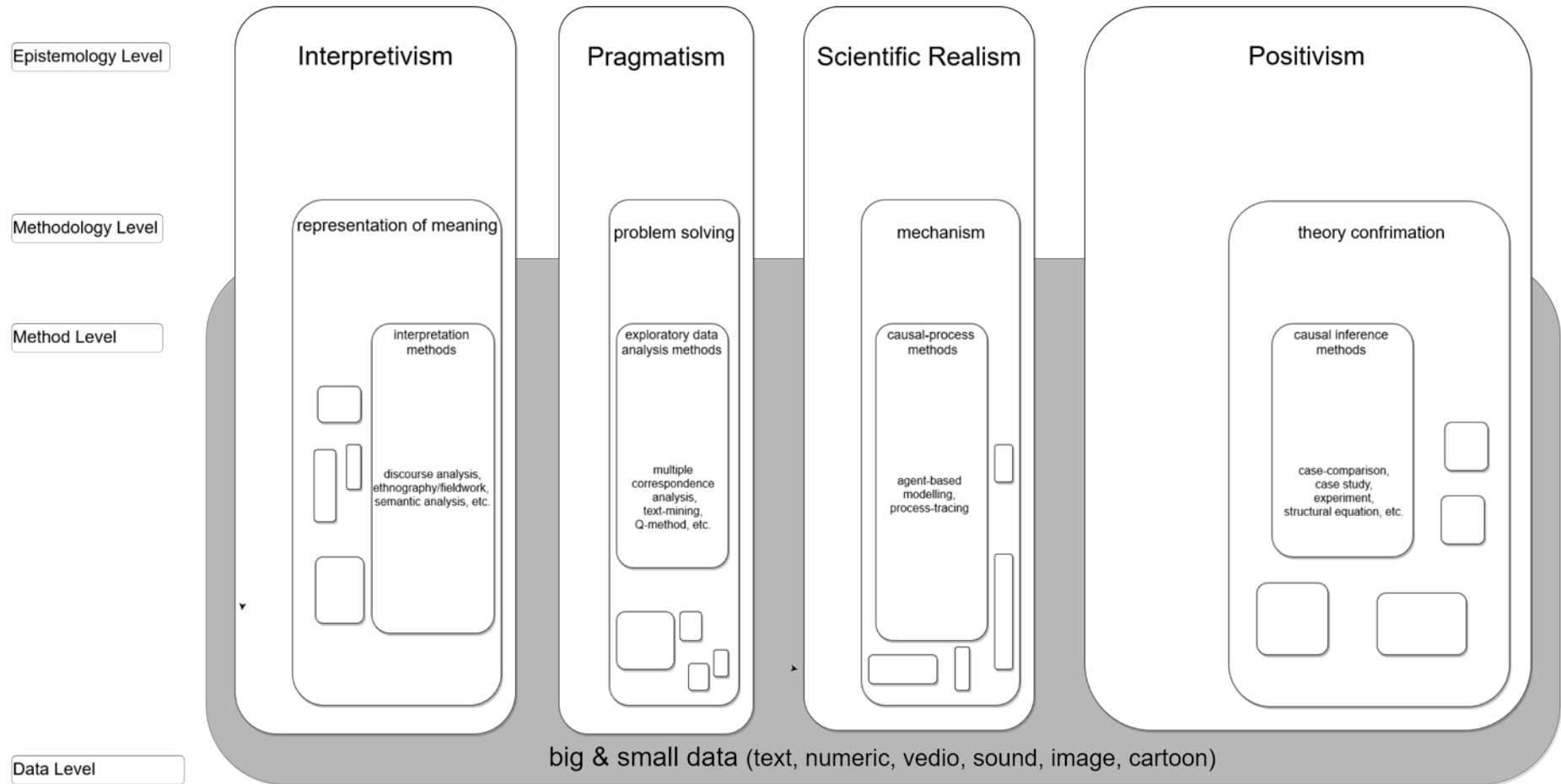
資料科學興起對 政治科學的挑戰及機會

問題意識造成學術內戰

知識論：「怎麼樣才算知道？」

方法論：「要怎麼樣才能求得答案？」



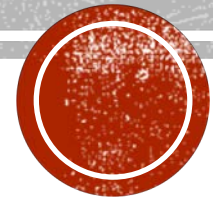


Illustrated by Prof. Chengshan (Frank) LIU

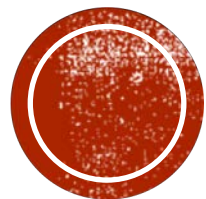


FROM BIG DATA TO DATA SCIENCE

“Data science is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to data mining.”



~ Wikipedia

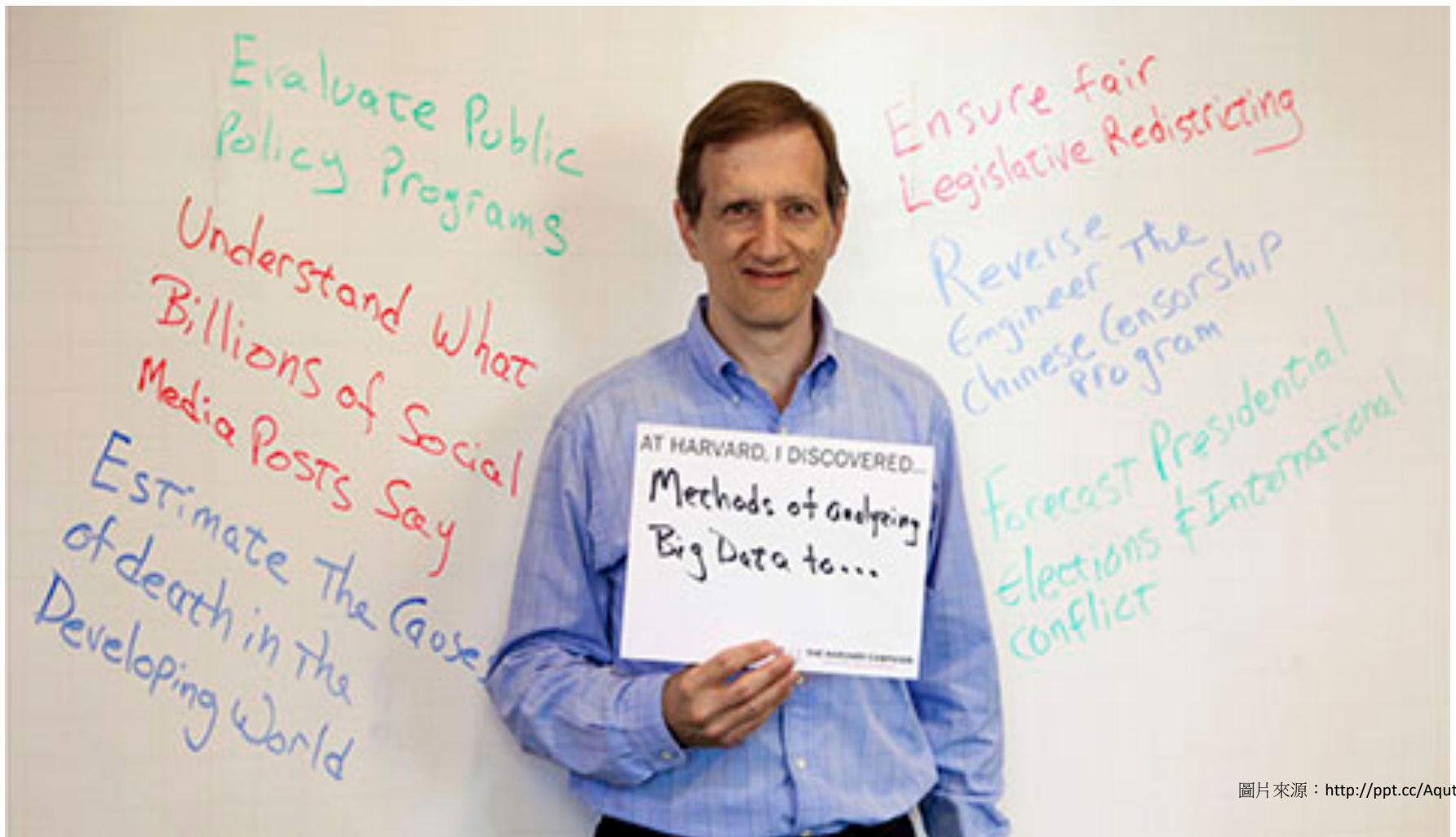


實證主義者正往實 驗之路大步邁進

因果？

代表性？

1. Bartels, L. M. (2014). Remembering to forget: A note on the duration of campaign advertising effects. *Political Communication*, 31(4), 532–544. <https://doi.org/10.1080/10584609.2014.956198>
2. Boudreau, C., & MacKenzie, S. A. (2014). Informing the electorate? How party cues and policy information affect public opinion about initiatives. *American Journal of Political Science*, 58(1), 48–62. <https://doi.org/10.1111/ajps.12054>
3. Driscoll, J., & Maliniak, D. (2016). Did Georgian voters desire military escalation in 2008? Experiments and observations. *Journal of Politics*, 78(1), 265–280. <https://doi.org/10.1086/683340>
4. Feldman, L., Myers, T. A., Hmielowski, J. D., & Leiserowitz, A. (2014). The mutual reinforcement of media selectivity and effects: Testing the reinforcing spirals framework in the context of global warming. *Journal of Communication*, 64(4), 590–611. <https://doi.org/10.1111/jcom.12108>
5. Isani, M., & Silverman, D. (2016). Foreign policy attitudes toward islamic actors: An experimental approach. *Political Research Quarterly*, 69(3), 571–582. <https://doi.org/10.1177/1065912916654988>
6. Kohama, S., Inamasu, K., & Tago, A. (2017). To denounce, or not to denounce: Survey experiments on diplomatic quarrels. *Political Communication*, 34(2), 243–260. <https://doi.org/10.1080/10584609.2016.1200700>
7. Mason, L. (2016). A cross-cutting calm: How social sorting drives affective polarization. *Public Opinion Quarterly*, 80(S1), 351–377. <https://doi.org/10.1093/poq/nfw001>
8. Messing, S., & Westwood, S. J. (2012). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 0093650212466406. <https://doi.org/10.1177/0093650212466406>
9. Persson, M., & Solevid, M. (2014). Measuring political participation: Testing social desirability bias in a web-survey experiment. *International Journal of Public Opinion Research*, 26(1), 98–112. <https://doi.org/10.1093/ijpor/edt002>
10. Santoso, L. P., Stein, R., & Stevenson, R. (2016). Survey experiments with Google Consumer Surveys: Promise and pitfalls for academic research in social science. *Political Analysis*, 24(3), 356–373. <https://doi.org/10.1093/pan/mpw016>
11. Tedesco, J. C. (2007). Examining Internet interactivity effects on young adult political information efficacy. *American Behavioral Scientist*, 50(9), 1183–1194. <https://doi.org/10.1177/0002764207300041>
12. Tomz, M. R., & Weeks, J. L. P. (2013). Public opinion and the democratic peace. *American Political Science Review*, 107(4), 849–865. <https://doi.org/10.1017/S0003055413000488>
13. Zhu, H., & Huberman, B. A. (2014). To switch or not to switch: Understanding social influence in online choices. *American Behavioral Scientist*, 58(10), 1329–1344. <https://doi.org/10.1177/0002764214527089>



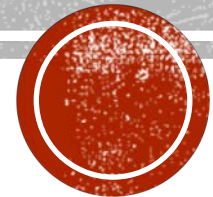
圖片來源：<http://ppt.cc/Aqutw>

- 國外關於大數據應用於政治學研究的出版以**Gary King**為主帥。其他文獻也大都或多或少受過**Gary King**所帶領的研究群之影響與啟發，儼然成為**Gary King**學派。Gary King在哈佛大學社會科學量化研究院（Institute for Quantitative Social Science, IQSS）中，鑽研如何使用不同的研究方法與量化工具推進社會科學研究。

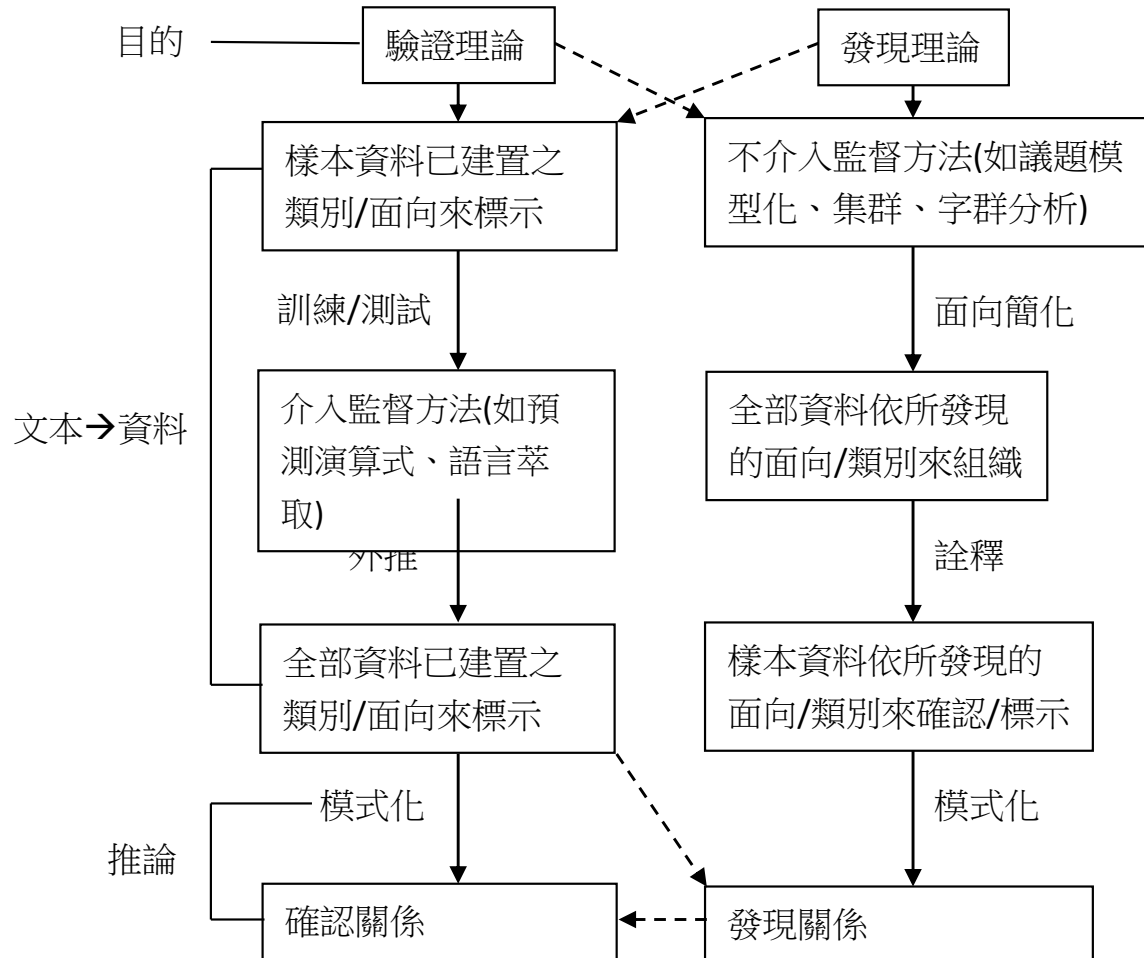


KING'S PURPOSES OF EMBRACING BIG DATA

- Evaluate public policy
- understand what social posts say
- estimate the causes of death,
- ensure fair legislative redistricting,
- reverse engineer Chinese government's censorship program,
- forecast elections and international conflict



HOW POSITIVISTS LOOK AT “BIG DATA”?

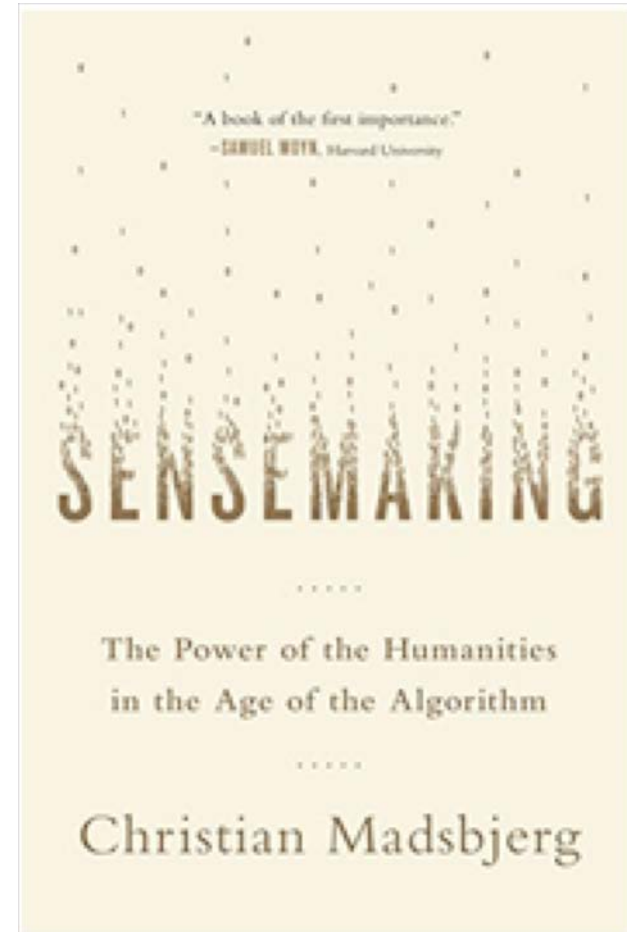
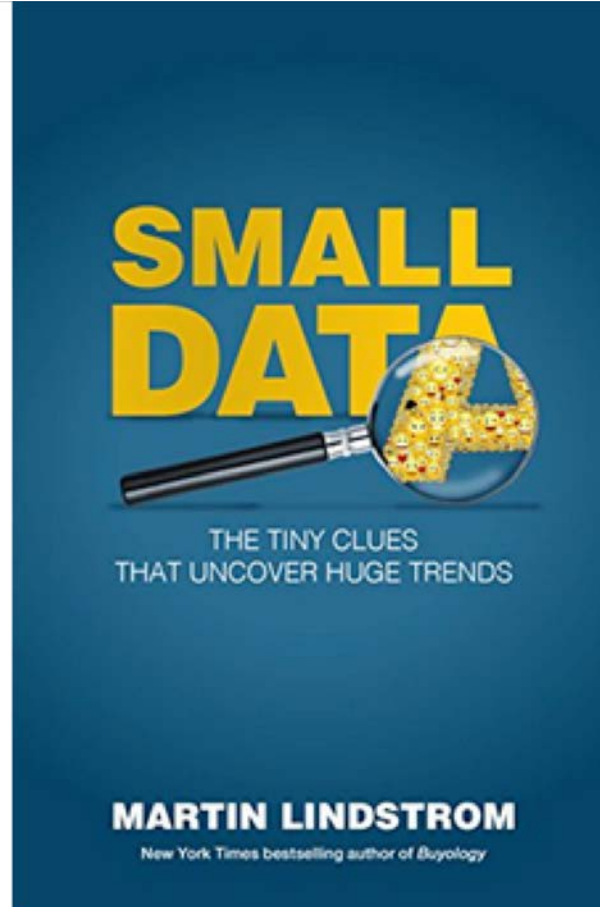
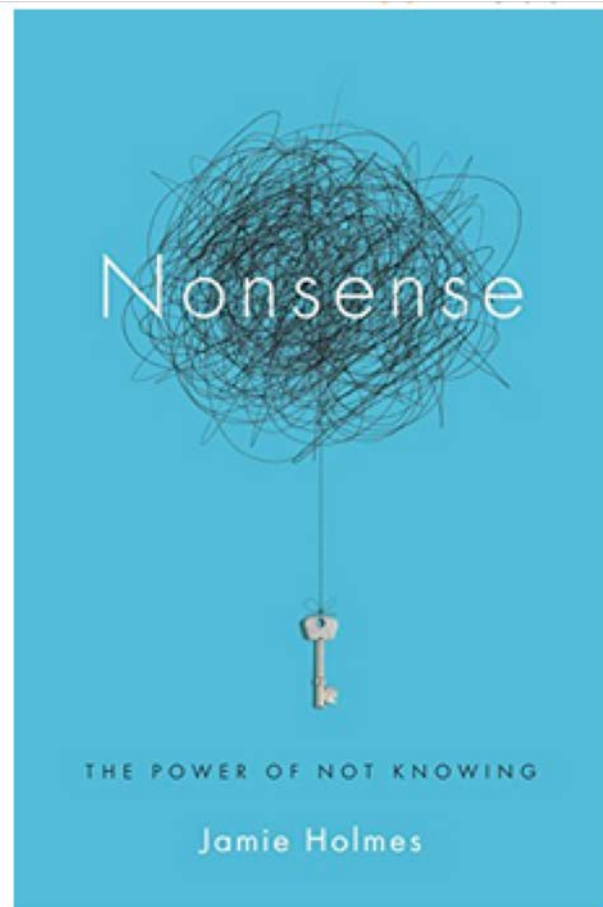




March 2016. Google watched how people use a phone in a van for over an hour at a time. Goal: complete interviewing 500 people.



REFLECTIONS FROM THE HUMANITIES



Holmes, J. (2015). *Nonsense: The Power of Not Knowing* (First Edition). New York: Crown Publishers. 《無知的力量》

Lindstrom, M. (2016). *Small Data: The Tiny Clues That Uncover Huge Trends*. New York City: St. Martin's Press. 《小數據獵人》

Madsbjerg, C. (2017). *Sensemaking: The Power of the Humanities in the Age of the Algorithm*. New York, NY: Hachette Books.

《演算法下的行銷優勢》



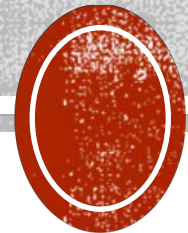
調和？



就是要由我們來讓
資料科學更有人性

**DATA SCIENCE FOR
EXTRACTING FACTS AND
DISCOVERING MEANING**

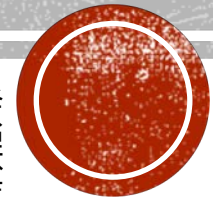
fact vs. truth vs.
reality vs.
knowledge



資料輔助的意義織造

DATA-ASSISTED MEANING NETTING

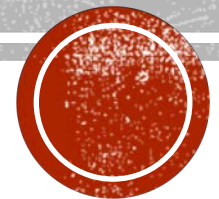
大數據的實作告訴我們，既然知識目的是探索。那就專注在在發現，而不（必）在驗證。資料數據可用於發現關聯，更可用於探勘意義。不妨先辨識自己有興趣的概念或面向（什麼價值、什麼行為、什麼態度？），再透過資料進行探索。一面辨識出不同價值、態度、行為之間的可能關係，一面與自己的預期關係進行對話。最後再來進行意義的詮釋。



Let's make our exploration DAMN right.

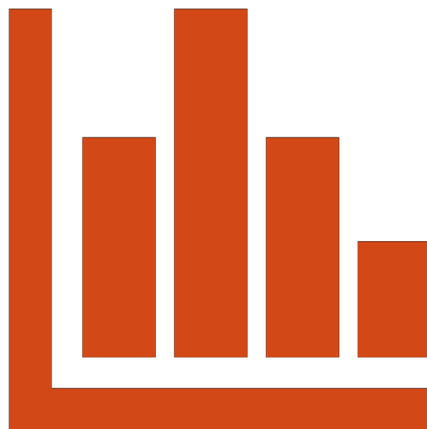
DAMN的方法論

不妨先辨識自己有興趣的概念或面向（什麼價值、什麼行為、什麼態度？），再透過資料進行探索。一面辨識出不同價值、態度、行為之間的可能關係，一面與自己的預期關係進行對話。最後再來進行意義的詮釋。



探索的方法及工具

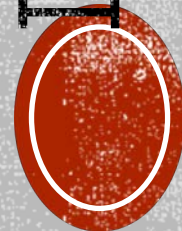
有了方法論的視野 你才能知道怎麼看待工具、方法及（大）數據

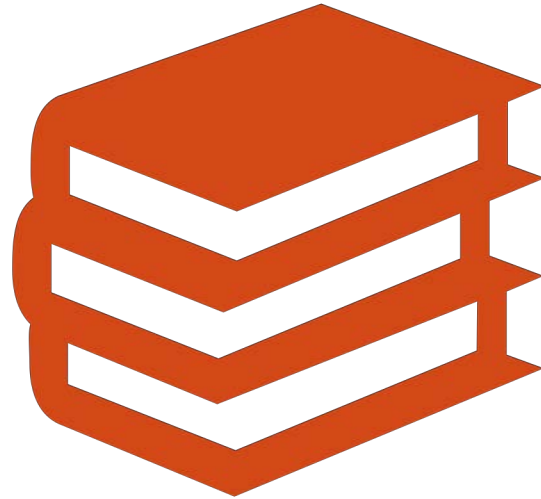


探索式資料分析很復古

Factor analysis & Explorative data analysis 早就有了

但以前使用類別型資料的民調同行 想都別想

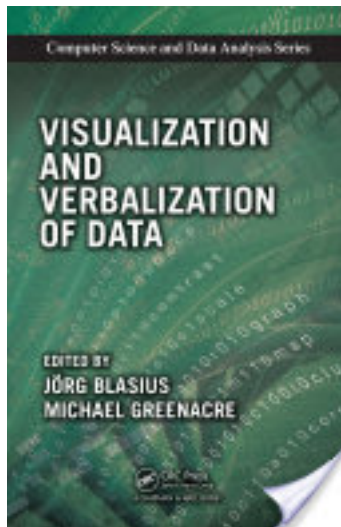
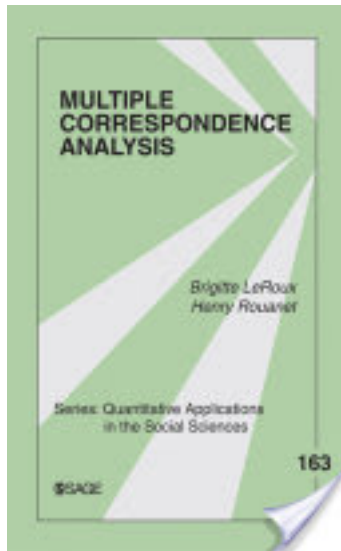




民調結合探索式資料分析 卻也可以很創新

Lakatos, Z. (2015). Traditional values and the Inglehart constructs. *Public Opinion Quarterly*, 79(S1), 291–324.
<https://doi.org/10.1093/poq/nfv005>





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

1 2 3 4 5 6 7 8 9 10

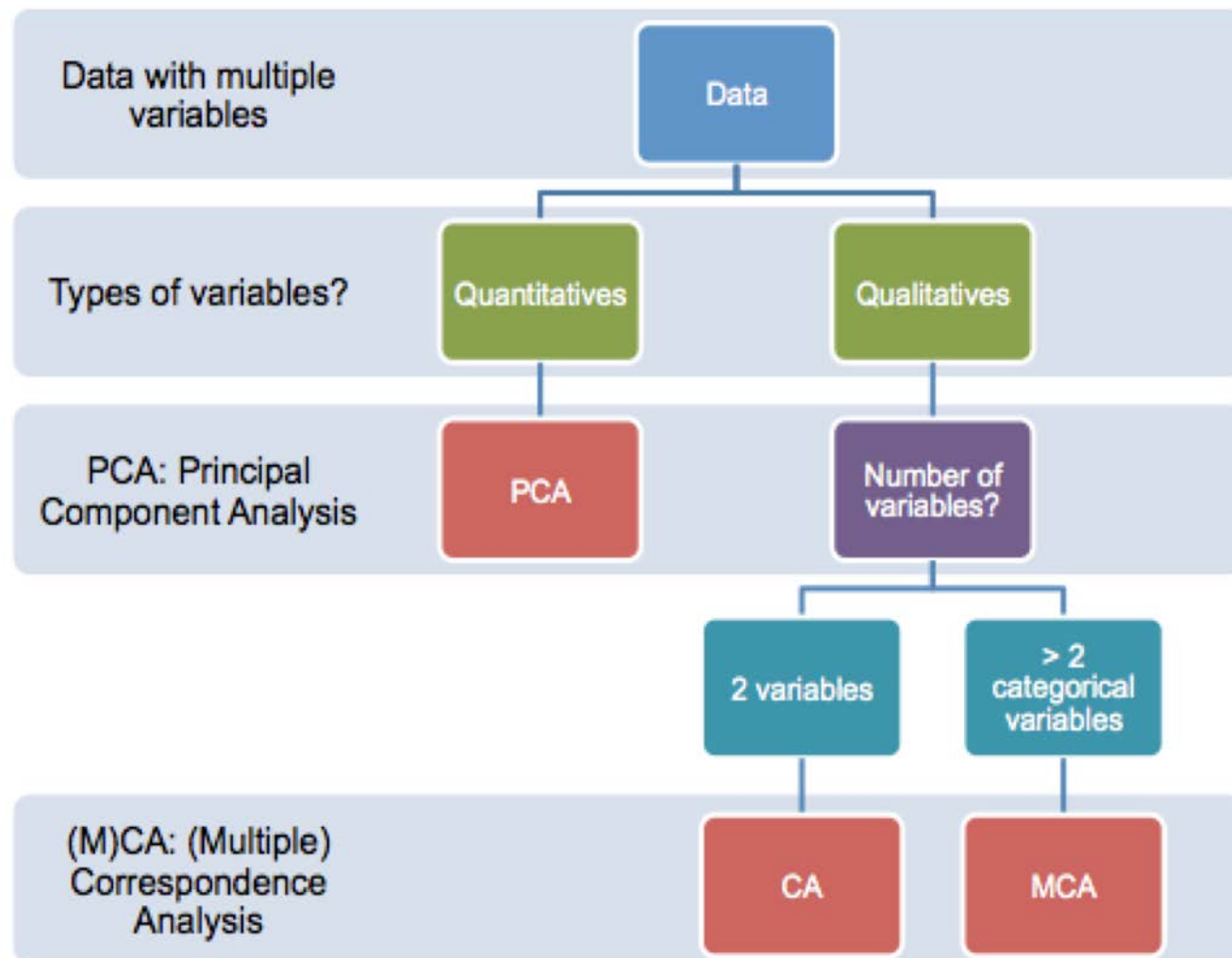
Jérôme Pagès

CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK

Copyrighted Material





(M)CA: (Multiple) Correspondence Analysis

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

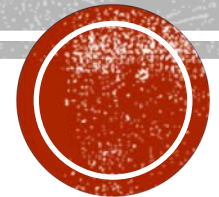
Use **factoextra** to easily extract and visualize the results



用民調做探索式的資料分析

發掘世代之間的差異：

以 18 歲前後經歷重大政治事件作為切點，分出五個世代。

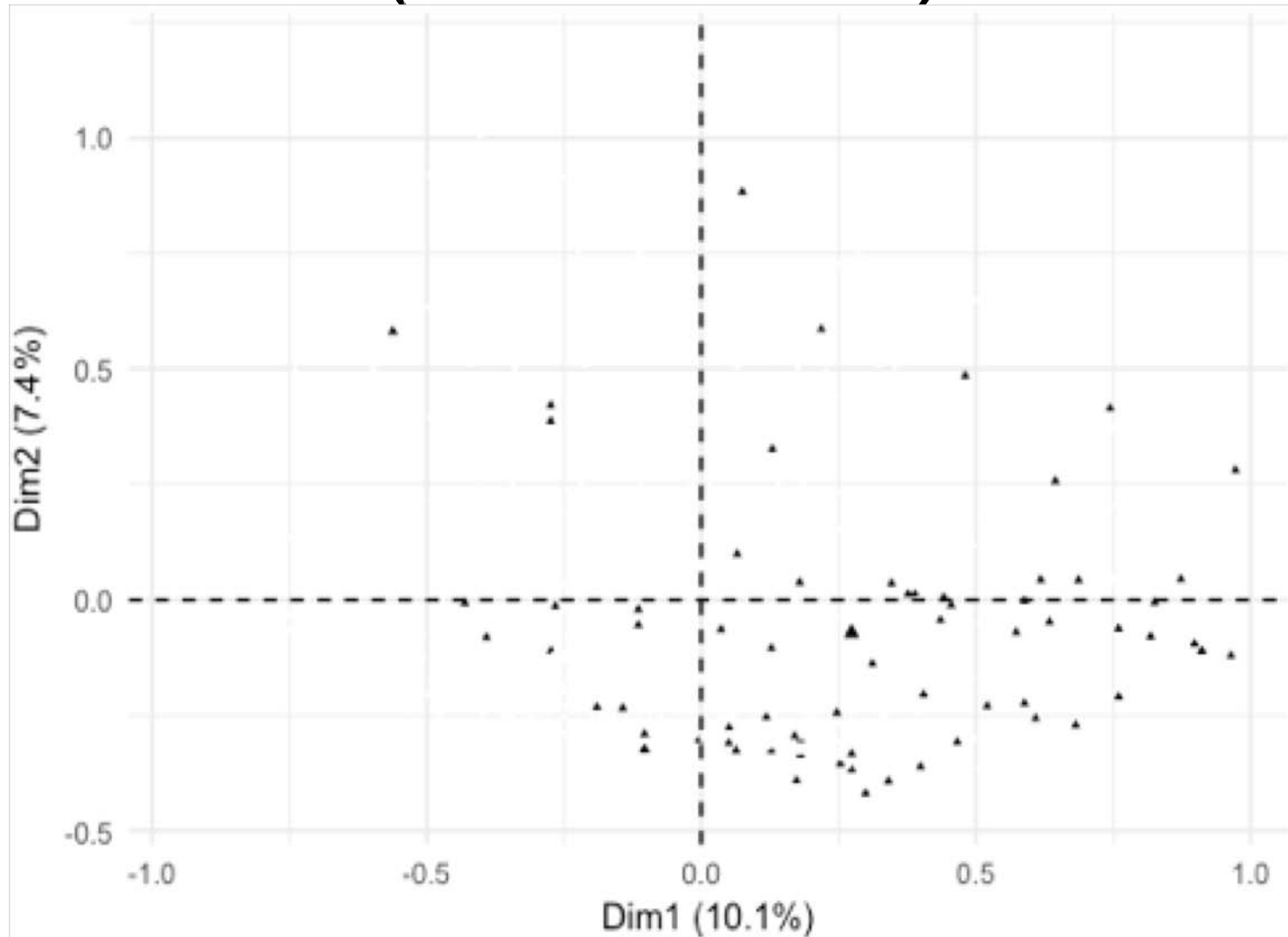


DATA SETS

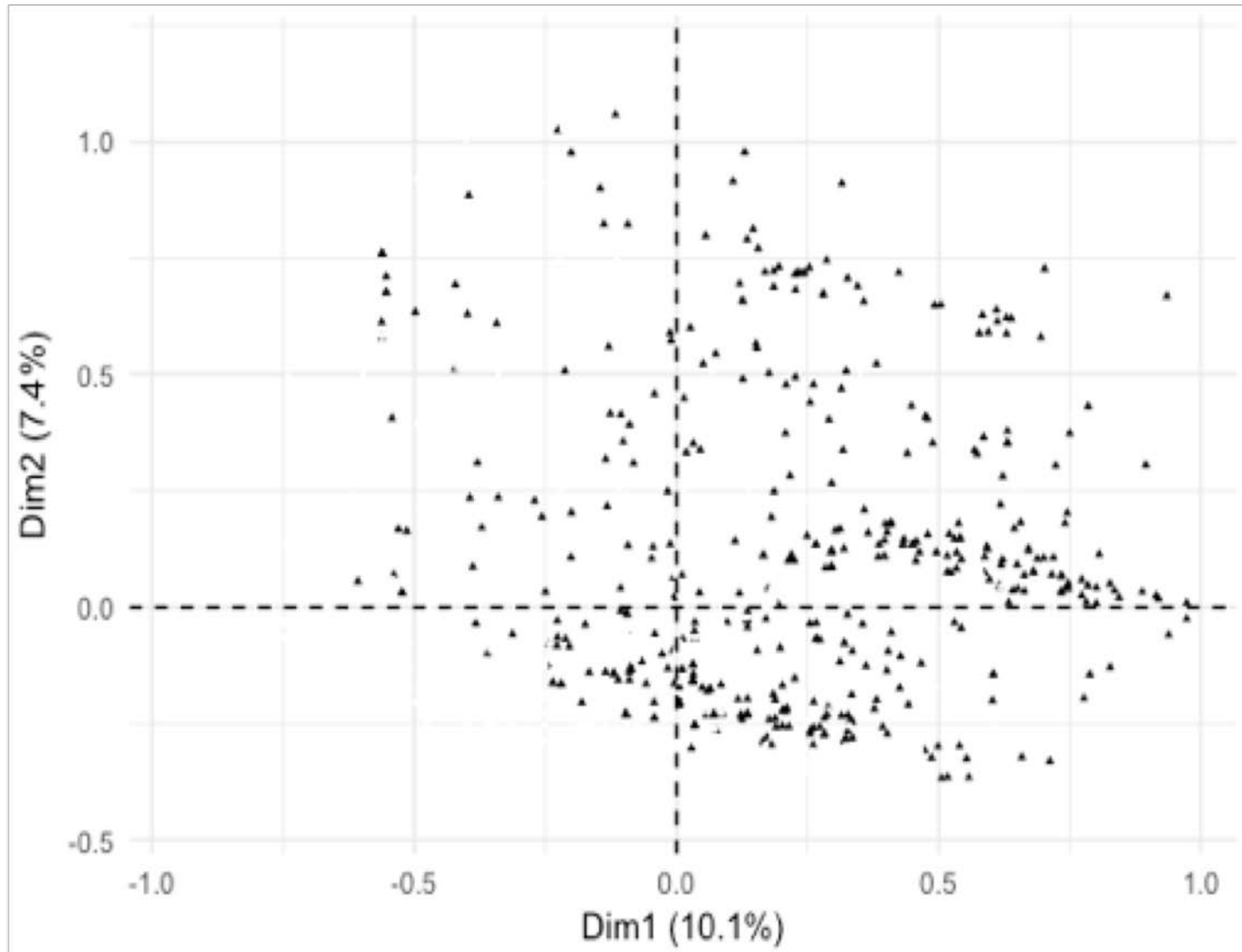
- F2F Survey:
Taiwan Social Change Survey (tscs) 2013 (n=1,952)
- CATI Telephone survey 2015 (n=1,100)
- Web panel 2015-2016 (n=468)



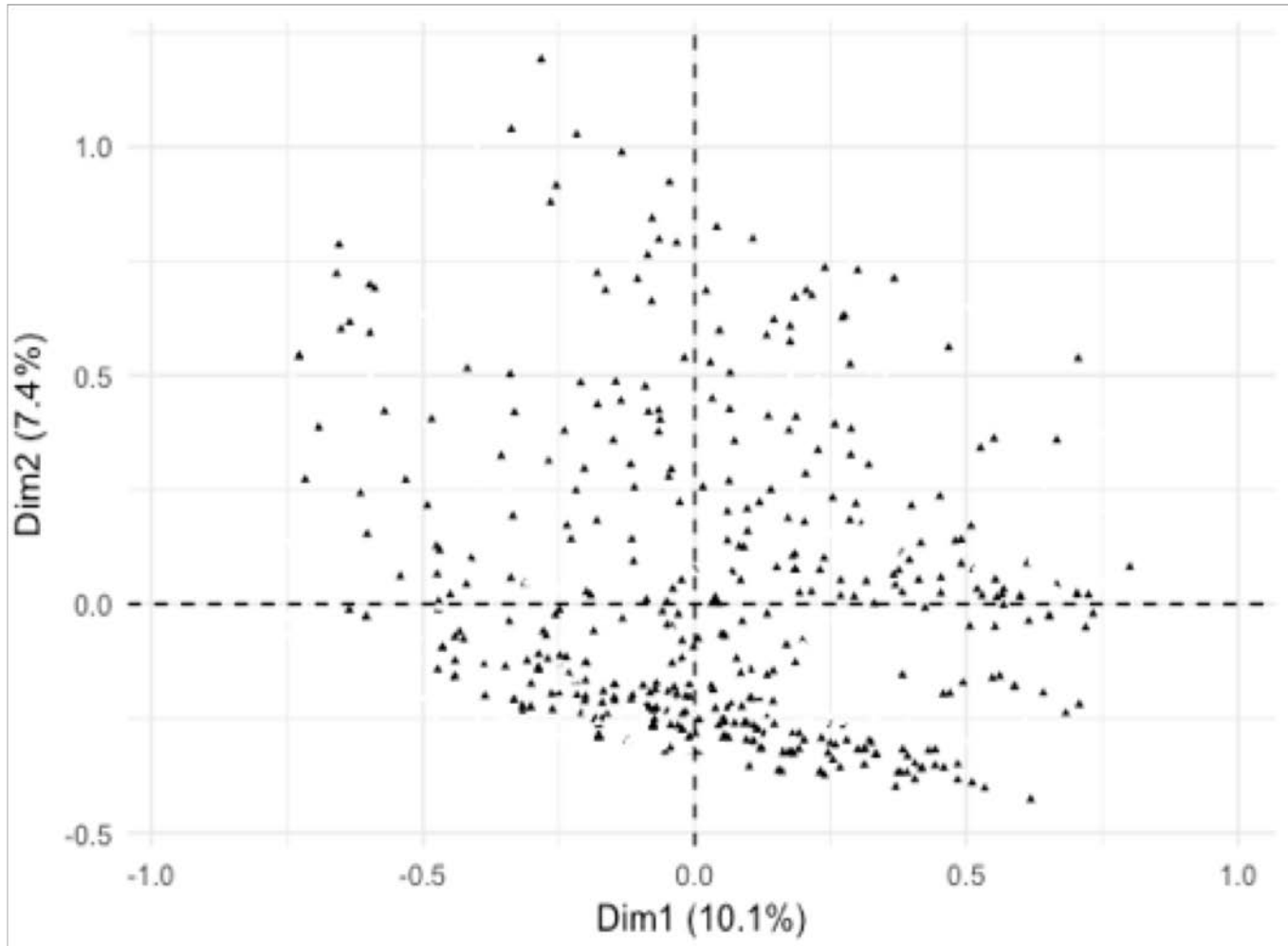
第一世代 (1931前出生)



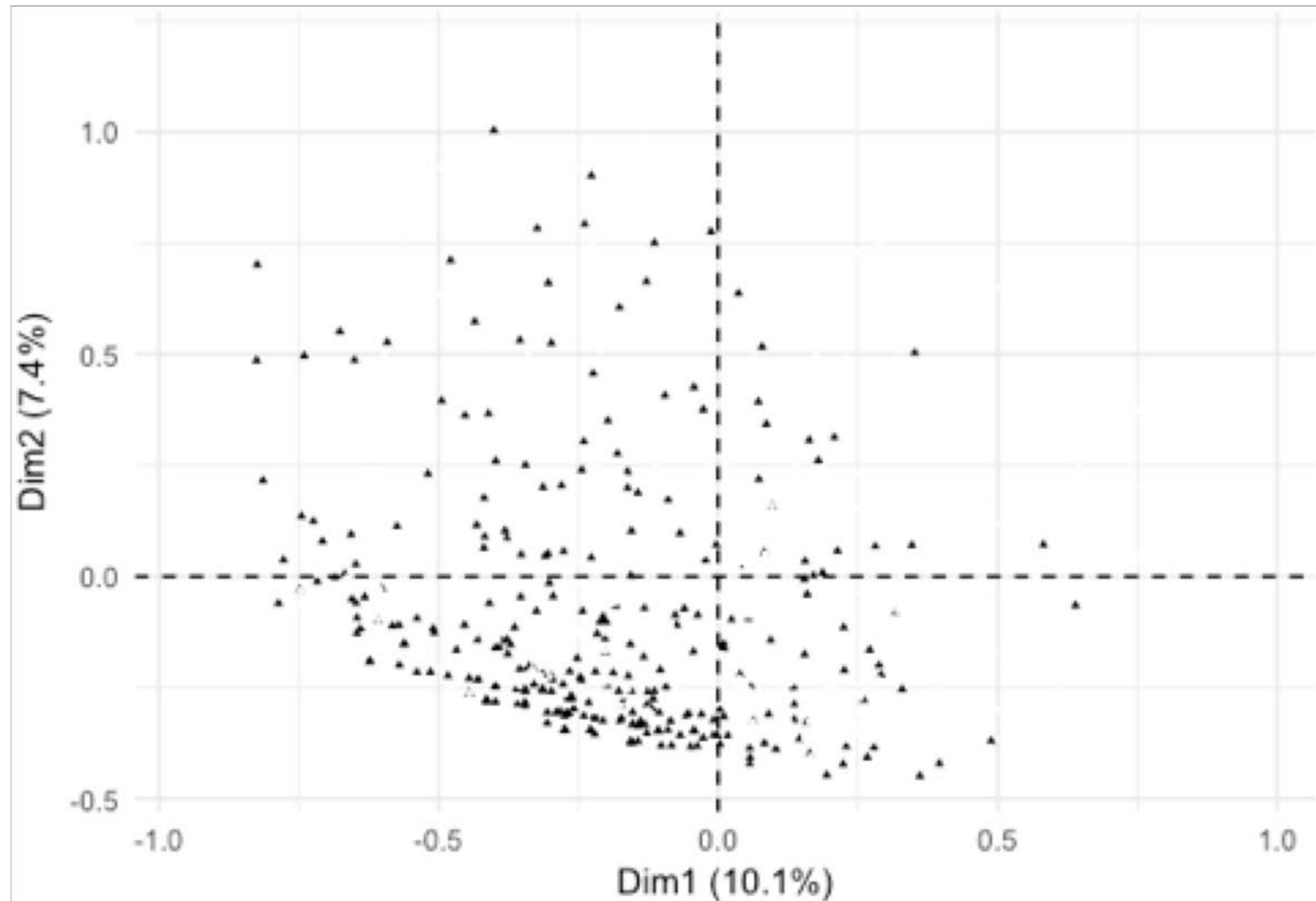
第二世代 (1932~1953出生)



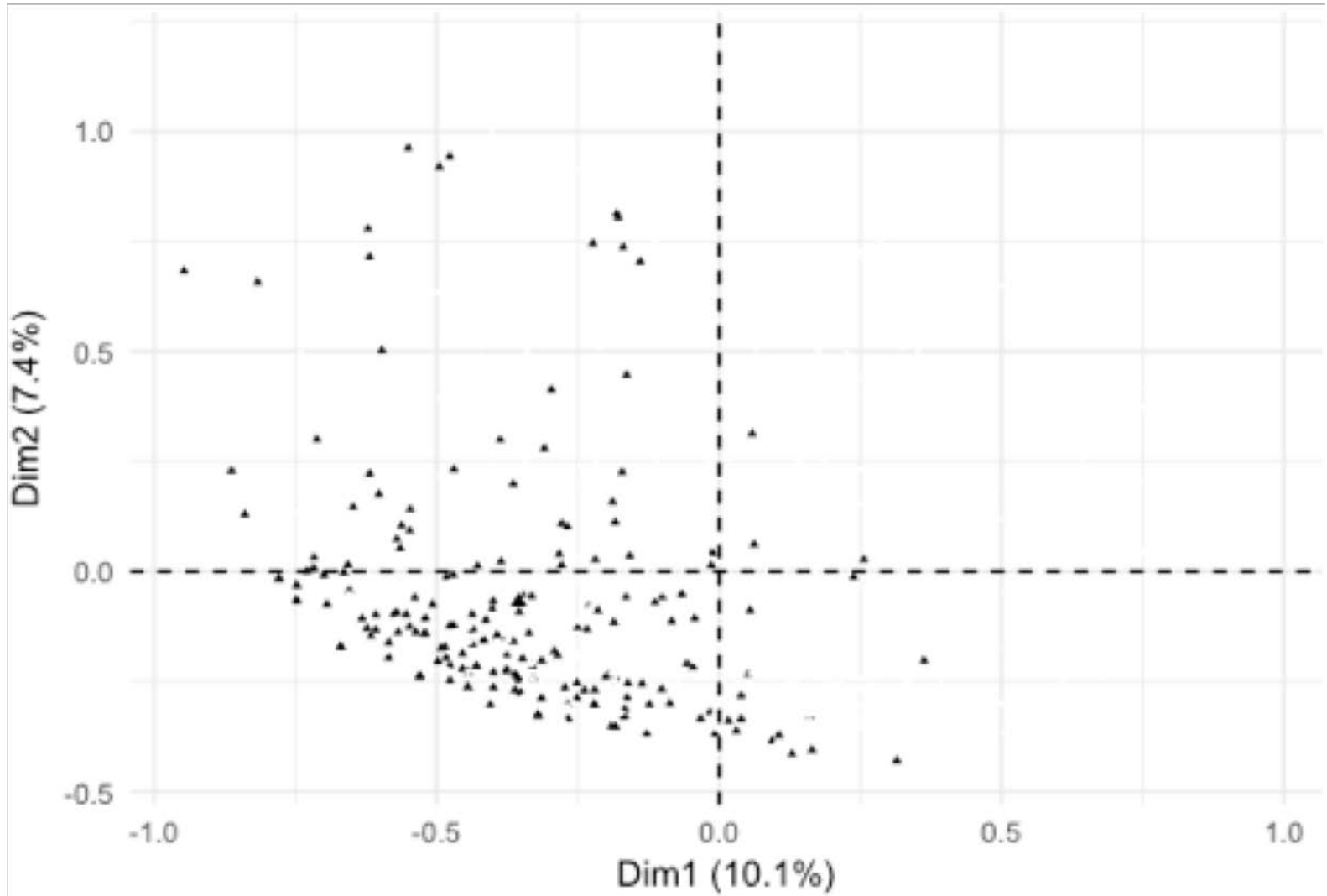
第三世代 (1954~1968出生)

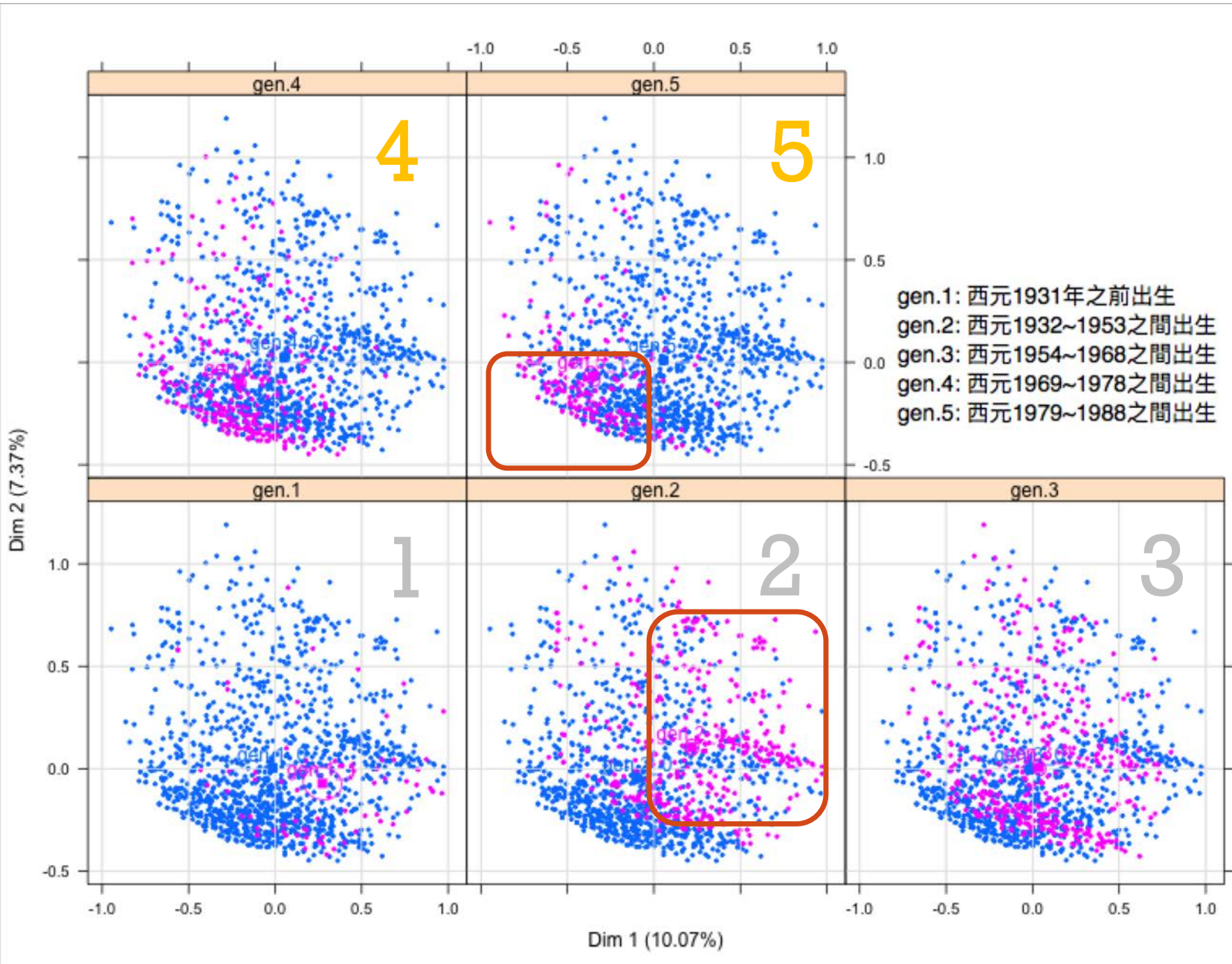


第四世代 (1969~1978出生)



第五世代 (1979~1988出生)





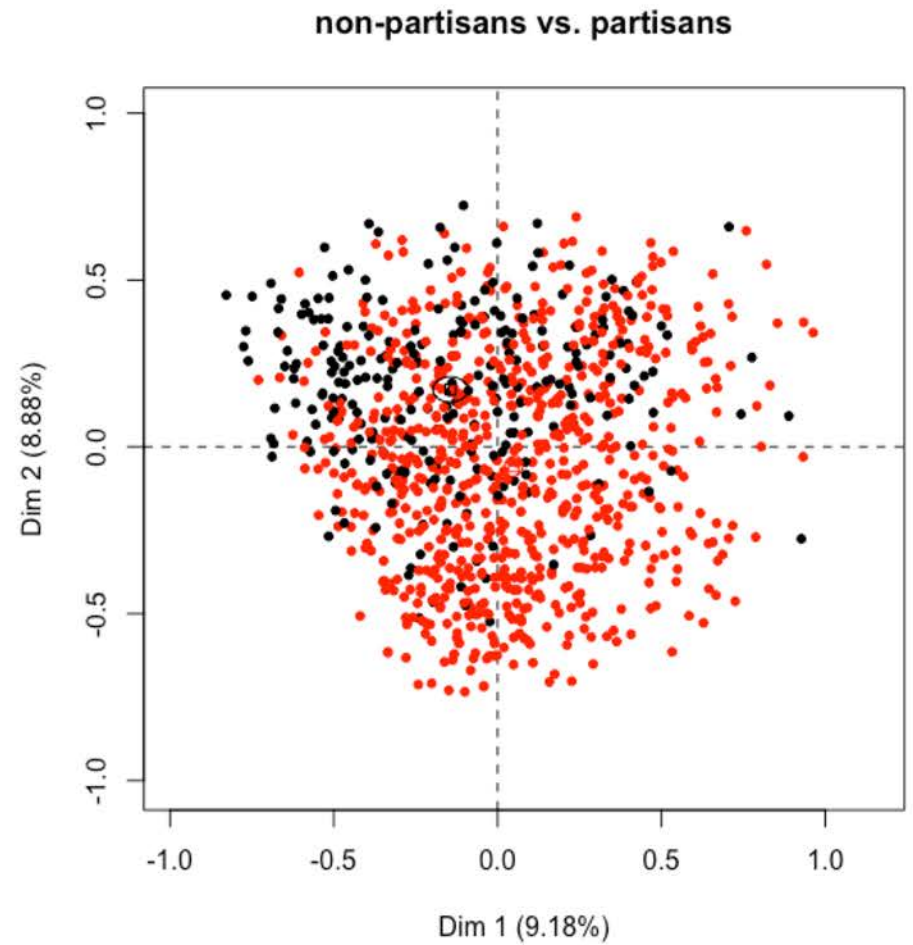
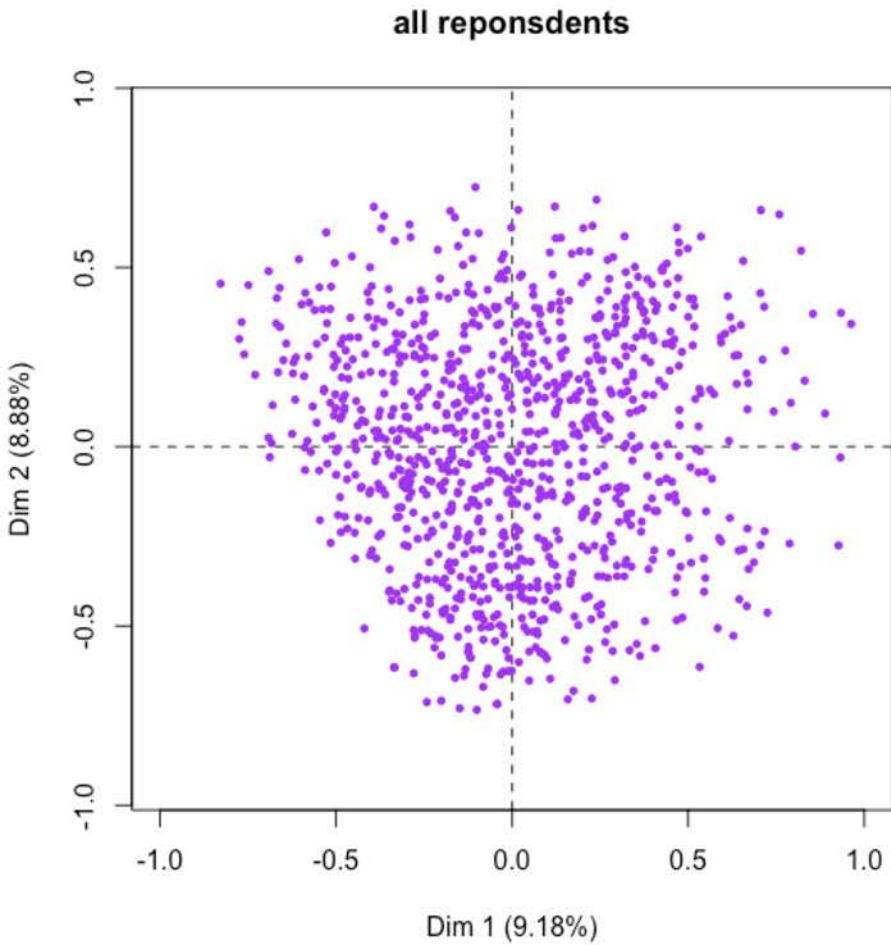
資料

- Taiwan Election and Democracy Studies 2016
- Data Collection Period: **2017.1.17 ~ 4.28**
- N=1,690
- \$\$\$: > NTD 1,000,000

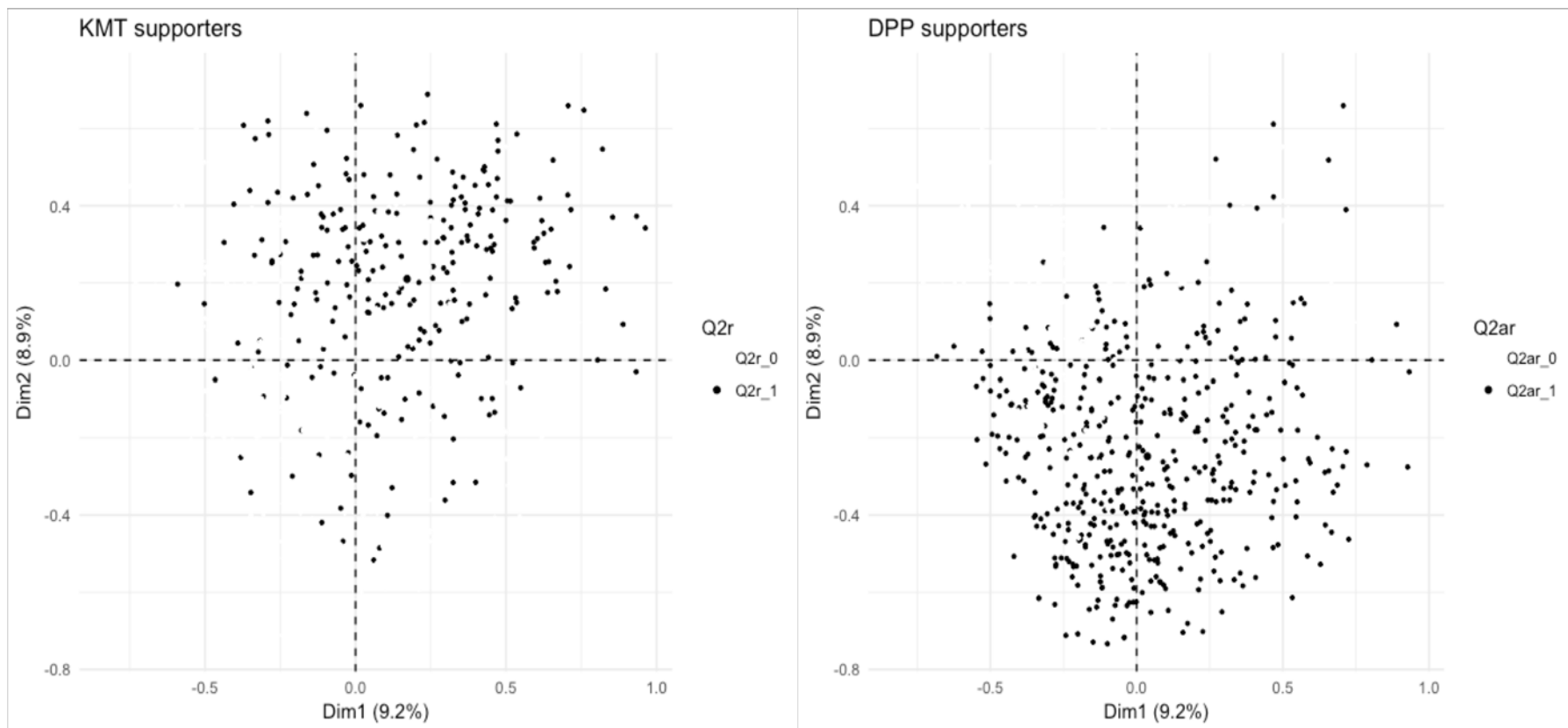


```
load("teds2016e.rda")
teds2016eforMCA <- select(teds2016e,
  c(# 核心變數 (core vars)
    A1r, # 算常接觸媒體=1; otherwise=0
    B1r, # 對政治算有興趣=1; otherwise=0
    D1r, # 覺得對政府沒有影響力=1; otherwise=0
    D3r, # 覺得政治太複雜=1; otherwise=0
    D2r, # 覺得政治人物不管我們的想法=1; otherwise=0
    D4r, # 覺得政府作的決定大多時候是對的=1; otherwise=0
    D5r, # 覺得政府常浪費稅金=1; otherwise=0
    D6r, # 覺得政府會優先考量公共利益=1; otherwise=0
    D7r, # 覺得自己算是懂政治=1; otherwise=0
    E3r, # 覺得大多數政治人物是可信的=1; otherwise=0
    E4r, # 覺得政治人物(才)是國家的主要問題=1; otherwise=0
    E5r, # 我們的國家需要政治強人=1; otherwise=0
    E6r, # 重要政策應該由人民而非政治人物來做=1; otherwise=0
    E7r, # 政治人物只在乎權貴者的利益=1; otherwise=0
    G1r, # 覺得過去一年台灣經濟變好=1; 沒變=2; 變差=3
    G2r, # 覺得未來一年台灣經濟會變好=1; 沒變=2; 會變差=3
    H1r, # 投票是種責任=1; otherwise=0
    H3, # 民主應優先於其他型態的政權=1; otherwise=0
    H5r, # 算是滿意台灣的民主=1; otherwise=0
    H6r, # 透過投票可以改變現況=1; otherwise=0
    P1, # 族群認同-台灣人=1; 都是=2; 中國人=3
    P2r, # 兩岸關係會變和諧=2; 不變=1; 變緊繃=3
    Q2r, # 11分量尺：表對國民黨算有好感(>=6); otherwise=0
    Q2ar, # 11分量尺：表對民進黨算有好感(>=6); otherwise=0
    partisanship # 自覺有(些)政黨傾向=1; otherwise=0
  ))
```

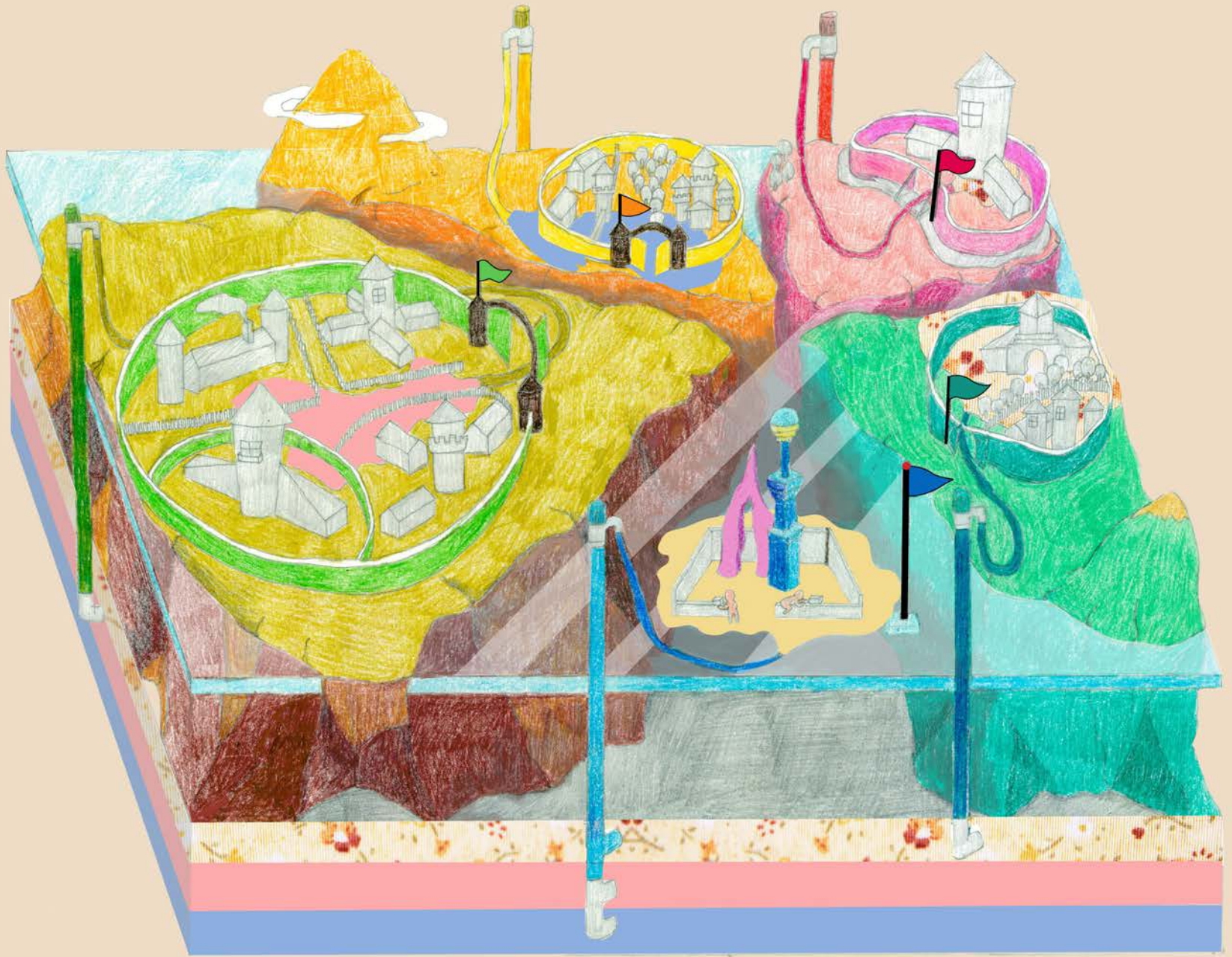

無政黨支持傾向者的樣貌



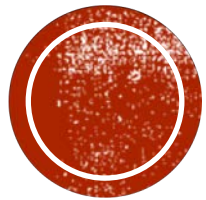
藍綠支持者的樣貌



從工具的學習來提升心態



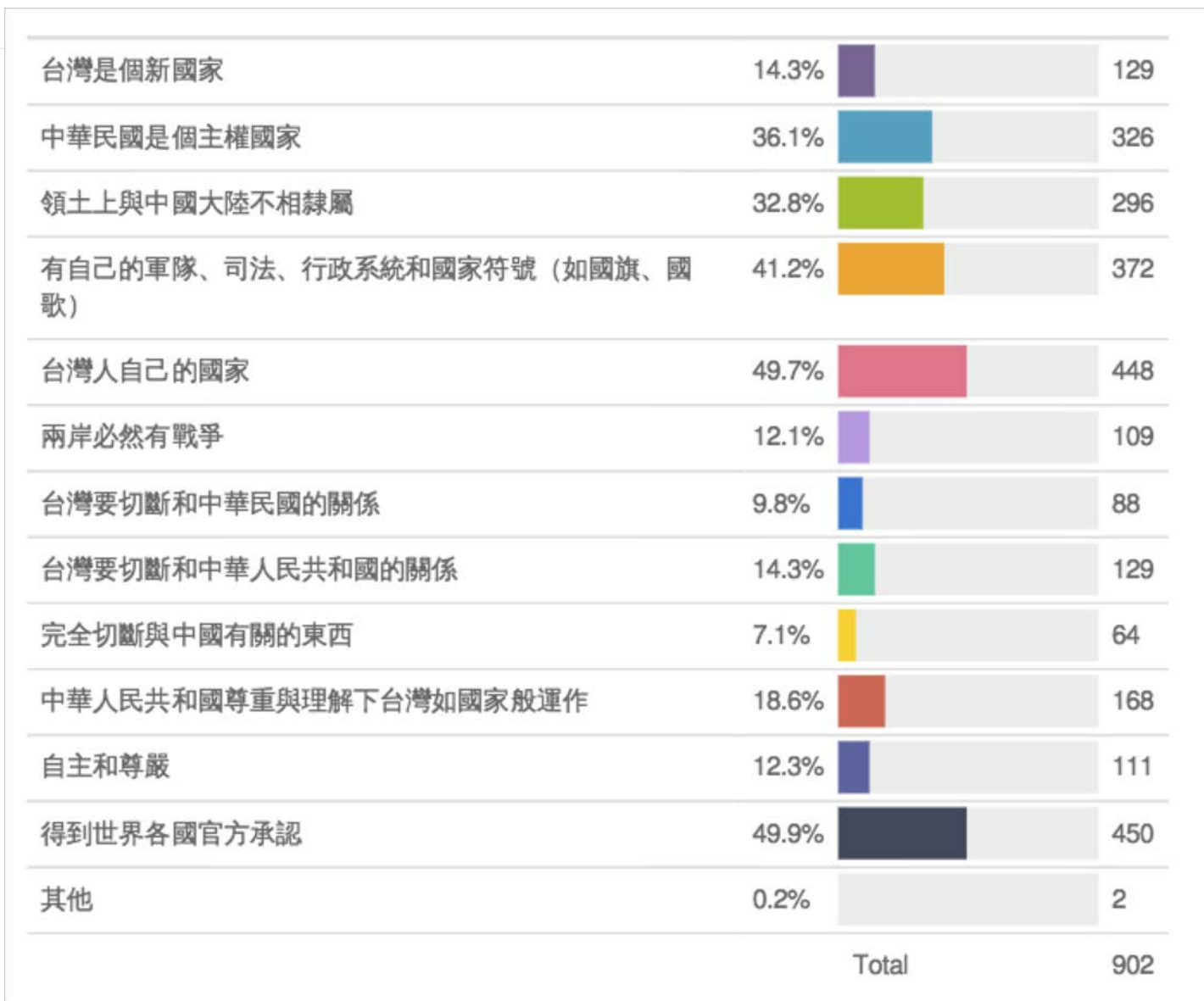
Illustrated by Prof. Chengshan (Frank) LIU



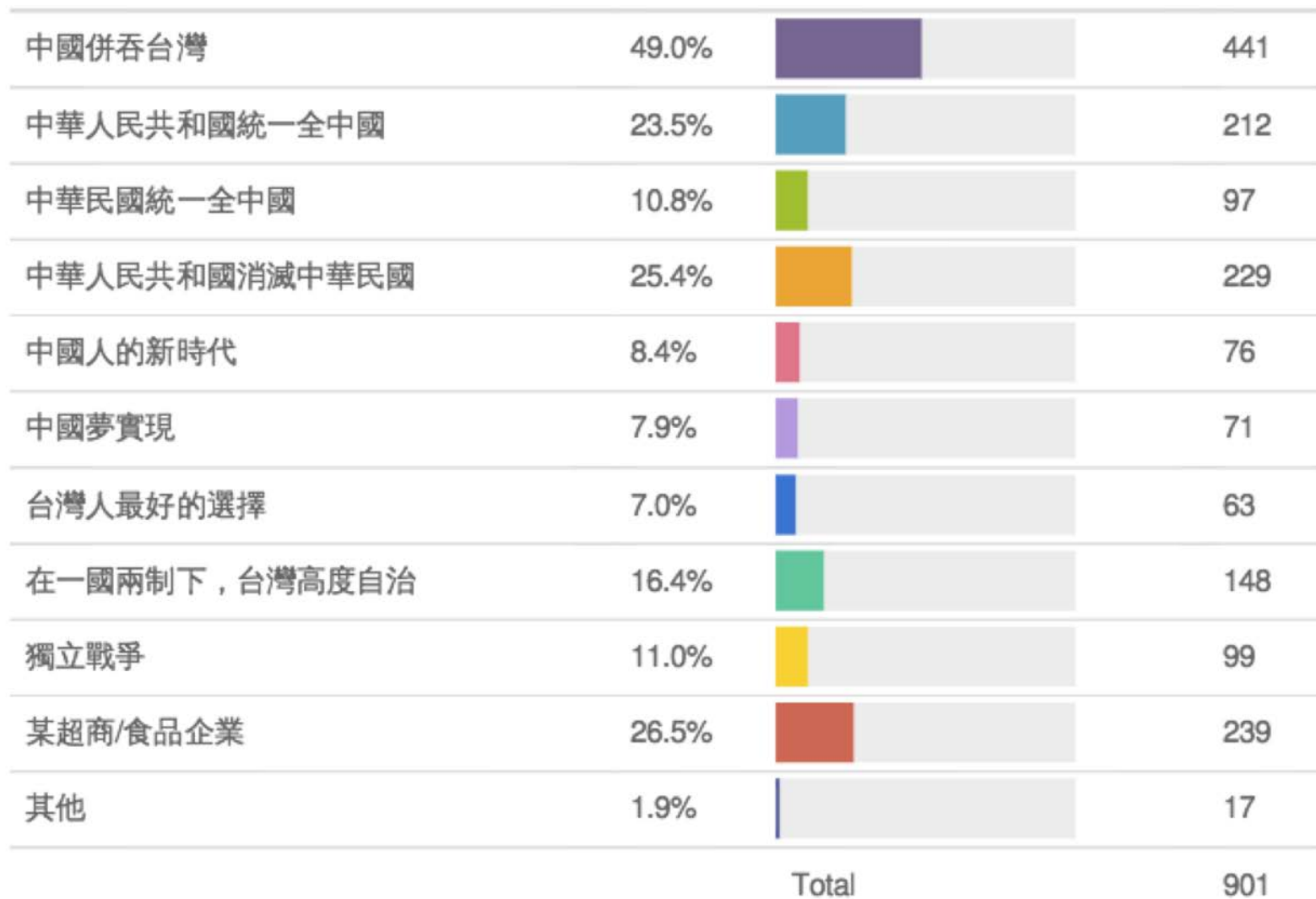
一般民調市調只會問出偏重於
行為及偏好的問題。從**DAMN**觀
點來看，我們還可以問出更多
關於價值觀的問題。

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？

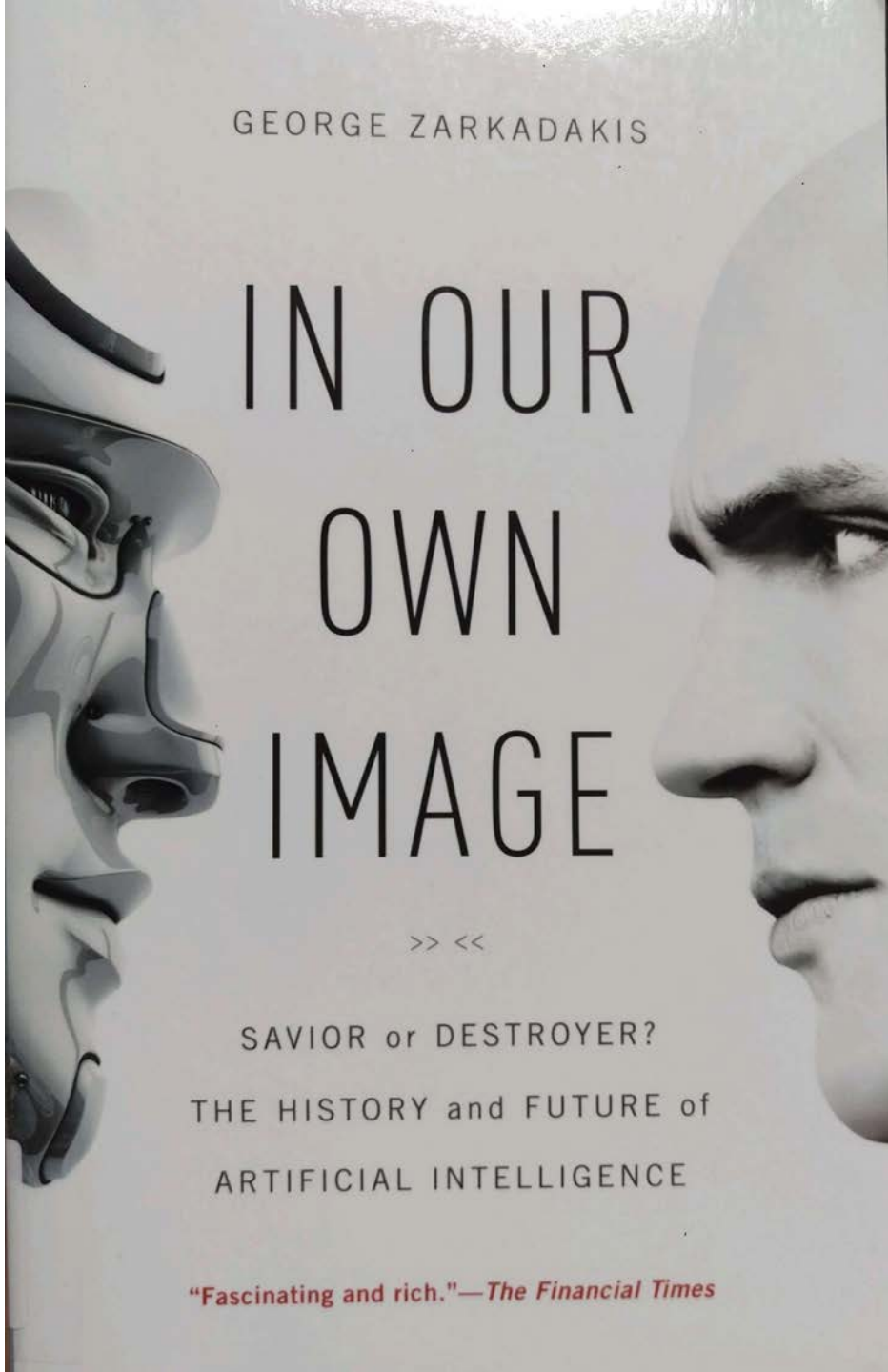
19. 說到獨立，您最先想到什麼？請從下面挑出最符合這個想法的選項。（可複選）



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）



GEORGE ZARKADAKIS



IN OUR OWN IMAGE

>> <<

SAVIOR or DESTROYER?
THE HISTORY and FUTURE of
ARTIFICIAL INTELLIGENCE

"Fascinating and rich."—The Financial Times



(大)數據呈現的 也許是真相但可能是假象

真相無法用不是手段上的量化**vs.**質化，或是大數據**vs.**小數據所能找到
關於社會的真相多半來自研究者心中資料 - 意義之間的對話

我們如何抗拒 對「確定感」的追求？

科學家與預言家相像：都在「幫助我們確定一件事」。

民眾相信自己所喜歡的。

所以擁有詮釋權力的資料科學家

將具備比一般民眾更可能為惡的本錢及能力。

學術 VS. 魔術

經驗本身不見得能被拿來當作詮釋的材料，所以資料科學登場。但科學家/學者所作的詮釋又如何可信？

民眾需要知道提供資訊的人是否已經善盡了他們的責任。

結語

- 大數據已成為資料科學的一環
- 政治科學必須迎接資料分析的新時代
- (大小) 數據相關技術已愈來愈平民化，因此社科學子早已沒有什麼理由與籍口不去接觸
- 擁抱資料科學時不應只心儀於技術、程式與方法 (而忘掉了科學)
- 從資料中發現有意義的樣貌，以及透過豐富背景知識來賦予資料意義，才是政治科學的強項
- 政治科學學子必須從方法論的視野強化對資料的掌握及意義的詮釋
- 無論從實證主義出發還是從詮釋主義出發，都可以看到大數據帶來的機會



心廣且虛
手勤且黑

這是今天帶給大家互相勸勉的最重要八個字

參考資料 : MAKING SENSE OF DATA

- Blackburn, S. (2012). *What Do We Really Know? The Big Questions in Philosophy*. London: Quercus.
- Cohen, L. H. (2013). *I don't know: In Praise of Admitting Ignorance*. New York: Riverhead Books.
- Holmes, J. (2015). *Nonsense: The Power of Not Knowing* (First Edition). New York: Crown Publishers.
- Madsbjerg, C. (2017). *Sensemaking: The Power of the Humanities in the Age of the Algorithm*. New York, NY: Hachette Books.
- Sesno, F., & Blitzer, W. (2017). *Ask More: The Power of Questions to Open Doors, Uncover Solutions, and Spark Change*. New York: AMACOM.
- Zarkadakis, G. (2016). *In Our Own Image: Savior or Destroyer? The History and Future of Artificial Intelligence* (1 edition). Pegasus Books.



其他參考資料

- 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.



中山政研

宏觀世局 · 精準決策

網路調查 / 文字探勘 / 社群媒體
思考架構 / 個案探討 / 歷史縱深

Politics Matters

全新課程設計 深廣你的決策視角

民眾態度價值輿情分析
兩岸三地政局走向分析
歐洲政經及投資環境評估
世界金融與能源的政經分析
全球政經思潮及投資決策展望
中國政局及東南亞投資條件分析



☎ 07-5252000 #5551
✉ polisci.nsysu@gmail.com
👉 www.ips.nsysu.edu.tw

