從抽樣理論到應用GIS/GPS克服跨國調查中的抽樣挑戰

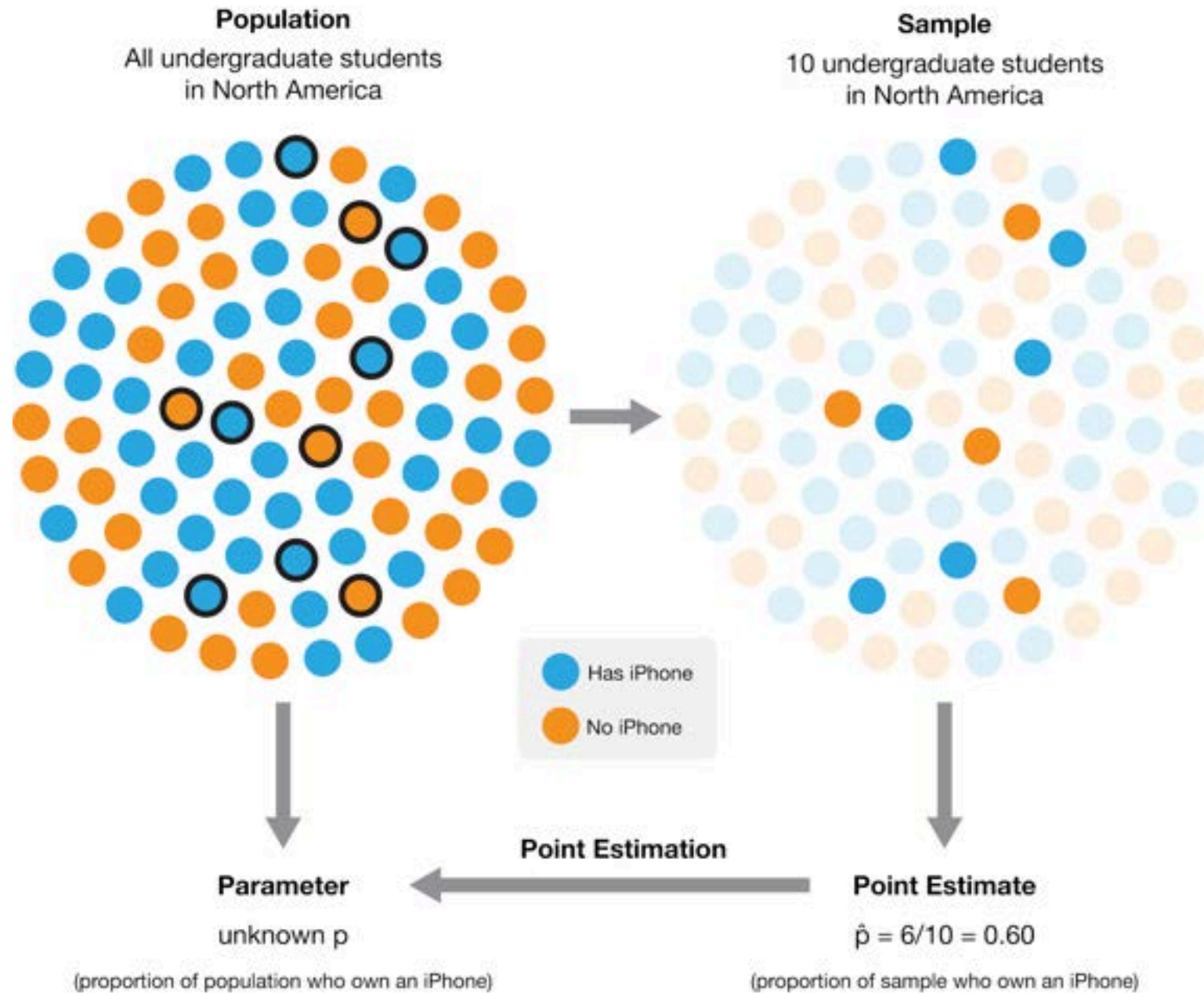## 以2018-20亞洲民主動態調查為例

黃忠偉

助研究員
胡佛東亞民主研究中心 | 亞洲民主動態調查
國立臺灣大學

調查? 抽樣? 抽樣調查！?

- 調查研究(Survey Research)：敘述或解釋目標群體的特徵
  - 選舉預測、市場開發、估算族群大小、流行病影響程度
- 為何要抽樣?
  - "Sampling is done because you usually *cannot gather data from the entire population* . Even in relatively small populations, the data may be needed *urgently*, and including everyone in the population in your data collection may take too long." (London School of Hygiene and Tropical Medicine, 2009)

Population
All undergraduate students in North America

Sample
10 undergraduate students in North America

Has iPhone
No iPhone

Point Estimation

Parameter
unknown p

(proportion of population who own an iPhone)

Point Estimate
$\hat{p} = 6/10 = 0.60$

(proportion of sample who own an iPhone)

Figures are from: Tiffany Timbers, Trevor Campbell, Melissa Lee, Joel Ostblom, Lindsey Heagy (2022) Data Science: A First Introduction (Python Edition). 3

如何抽樣?

- 街頭民調 v.s. 傳統民調

- 評估抽樣過程的兩個指標：
  - 公正性
  - 代表性

Figure 3. A Random, Representative Sample from the Population

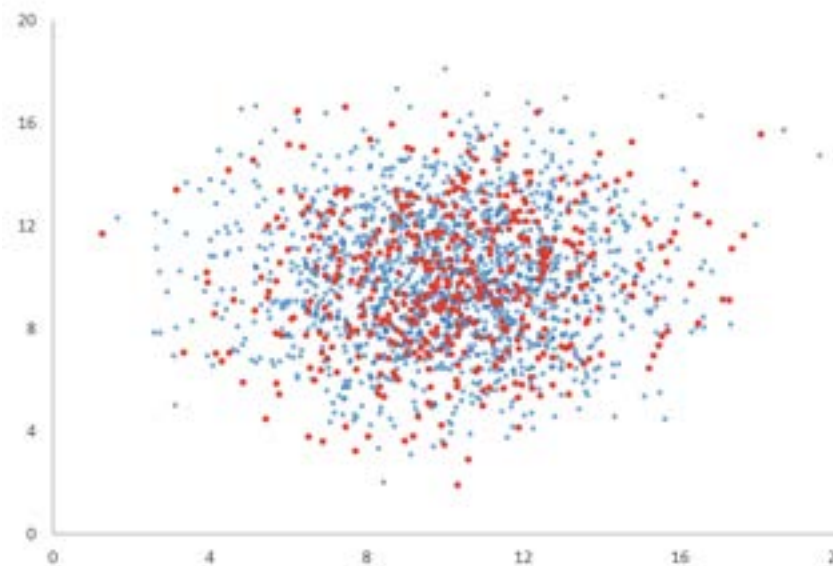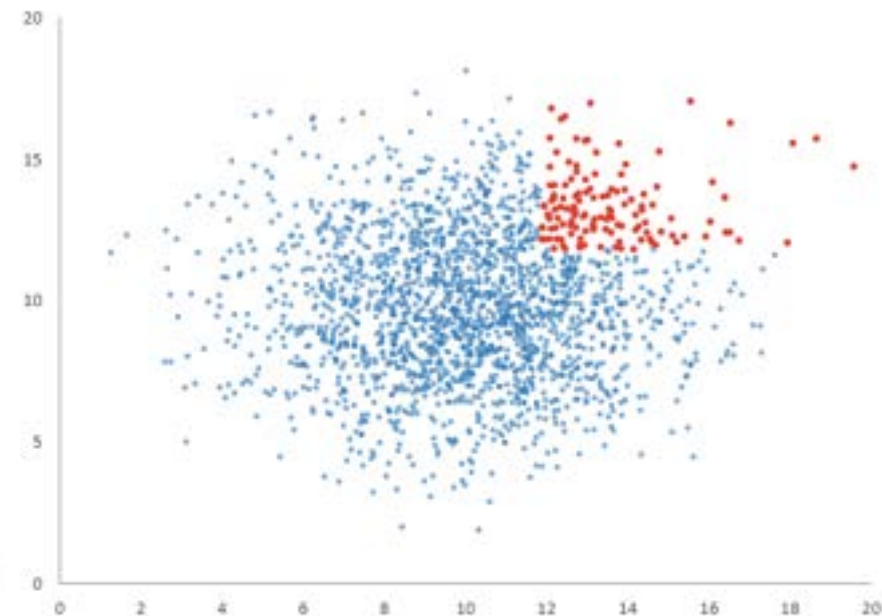Figure 4. A Non-random, Non-representative Sample from the Population

簡單隨機抽樣：色球比例的推估

設想一盒不同顏色的十顆球，其中紅白兩色分別有7、3顆。所以我們決定抽出n顆球來**推論**各色球所佔的比例。現令我們抽出的紅、白球球數分別為x和$n-x$（抽出不放回），此時：

- 抽出球數的經驗資料可讓我們推估出：
  - $P(紅球 = x; n) = \frac{x}{n}; P(白球 = n - x; n) = 1 - \frac{x}{n}$
- 但以邏輯推論（排列組合）的觀點，不同抽出球數之下紅白球數的比例為：
  - $P(紅球 = x; n) = \frac{C_x^7 C_{n-x}^3}{\sum\limits_{i=0}^{n} C_i^7 C_{n-i}^3}; P(白球 = n - x; n) = 1 - \frac{C_x^7 C_{n-x}^3}{\sum\limits_{i=0}^{n} C_i^7 C_{n-i}^3}$

簡單隨機抽樣的公正性

- 以下是把所有抽出球數n和抽中紅球數x帶入上述公式的結果：

紅球數 x	抽出球數 n 組合數(組合配對 $x_1x_2x_3x_4$)									
	1	2	3	4	5	6	7	8	9	10
0	3(0,1)	3(0,2)	1(0,3)							
1	7(1,0)	21(1,1)	21(1,2)	7(1,3)						
2		21(2,0)	63(2,1)	63(2,2)	21(2,3)					
3			35(3,0)	105(3,1)	105(3,2)	35(3,3)				
4				35(4,0)	105(4,1)	105(4,2)	35(4,3)			
5					21(5,0)	63(5,1)	63(5,2)	21(5,3)		
6						7(6,0)	21(6,1)	21(6,2)	7(6,3)	
7							1(7,0)	3(7,1)	3(7,2)	1(7,3)

註：抽出紅球的平均期望值不管抽出球數 n 值皆為 0.7
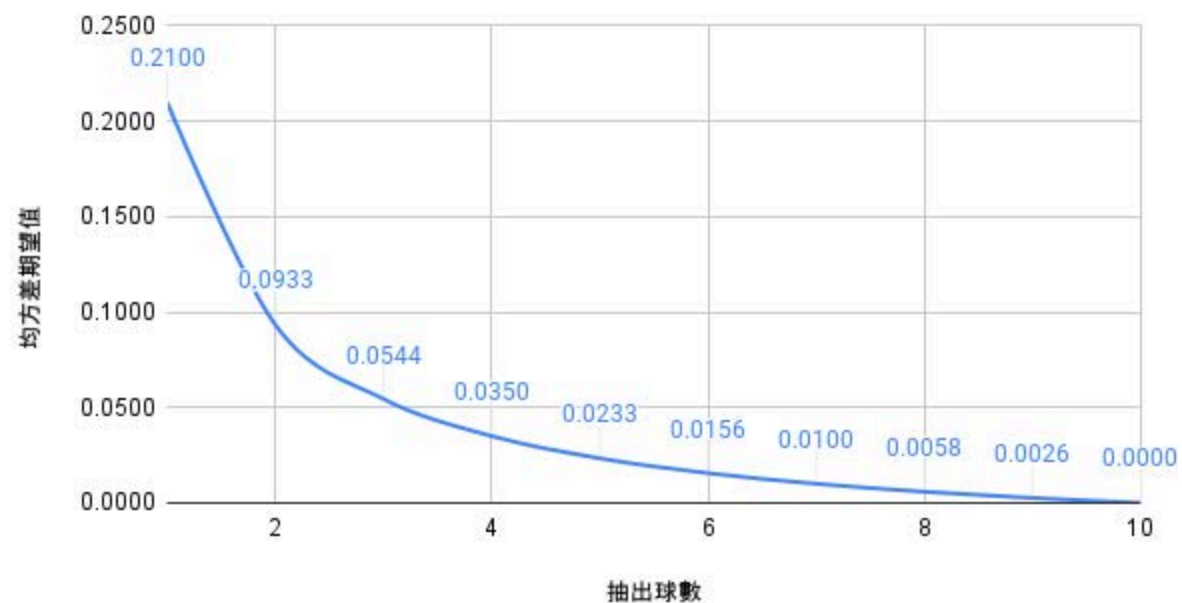
- 公正性=期望值 $E[X] = x_1p_1 + x_2p_2 + \ldots + x_kp_k$

簡單隨機抽樣的代表性

- 抽出**一**球中紅球數值和的機率：$x = 0(\frac{3}{10}), 1(\frac{7}{10})$

- 抽出**兩**球中紅球數值和的機率：$x = 0(\frac{3}{45}), 1(\frac{21}{45}), 2(\frac{21}{45})$

- 抽出**三**球中紅球數值和的機率：$x = 0(\frac{1}{120}), 1(\frac{21}{120}), 2(\frac{63}{120}), 3(\frac{35}{120})$
  ...

- 抽出**十**球中紅球數值和的機率：$x = 7(\frac{1}{1})$

- 樣本數增加，代表性提升？利用均方差(mean squared error)來計算吧!

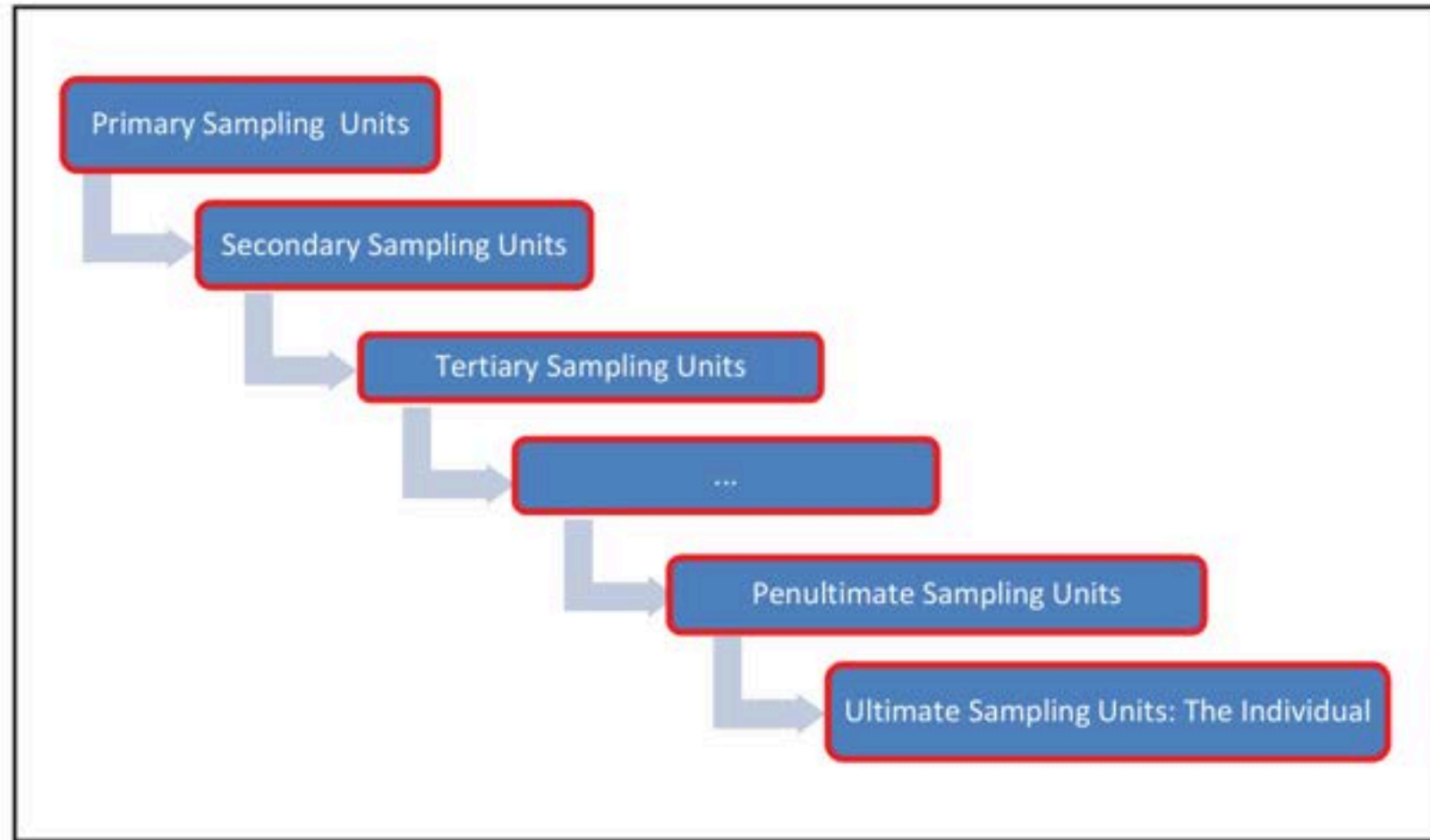| 機率	紅球個數								均方差期望值	邊際遞減值
抽出球數	0	1	2	3	4	5	6	7		
1	0.300	0.700							0.2100	
2	0.067	0.467	0.467						0.0933	0.1167
3	0.008	0.175	0.525	0.292					0.0544	0.0389
4		0.033	0.300	0.500	0.167				0.0350	0.0194
5			0.083	0.417	0.417	0.083			0.0233	0.0117
6			0.167	0.500	0.300	0.033			0.0156	0.0078
7				0.292	0.525	0.175	0.008		0.0100	0.0056
8					0.467	0.467	0.067		0.0058	0.0042
9						0.700	0.300		0.0026	0.0032
10							1.000		0.0000	0.0026

縱軸:均方差期望值, 橫軸:抽出球數

如何求得不偏母體參數：理想與實務

- 以「機率抽樣」(probability sampling)來得到母體參數的不偏估計(unbiased estimation)
  - 母體內每一個體被隨機選出的機率**大於**0（至少有機會被抽到）
  - 被選到的機率**已知**（知道每一個人被抽到的機率）
- 好抽樣清冊(sampling frame)做好調查
  - 窮盡母體抽樣單位的「名單」，確保每一個抽樣單位都有大於零的中選機會
  - 從該「名單」抽出樣本的同時，也知道每個抽樣單位的中選機率

大型面訪調查的機率抽樣：probability proportional to size (pps) sampling

Figures are from: Peter Lance, John Spencer, Aiko Hattori (2014) GIS and Sampling, USAID, PEPFAR, MEASURE Evaluation.
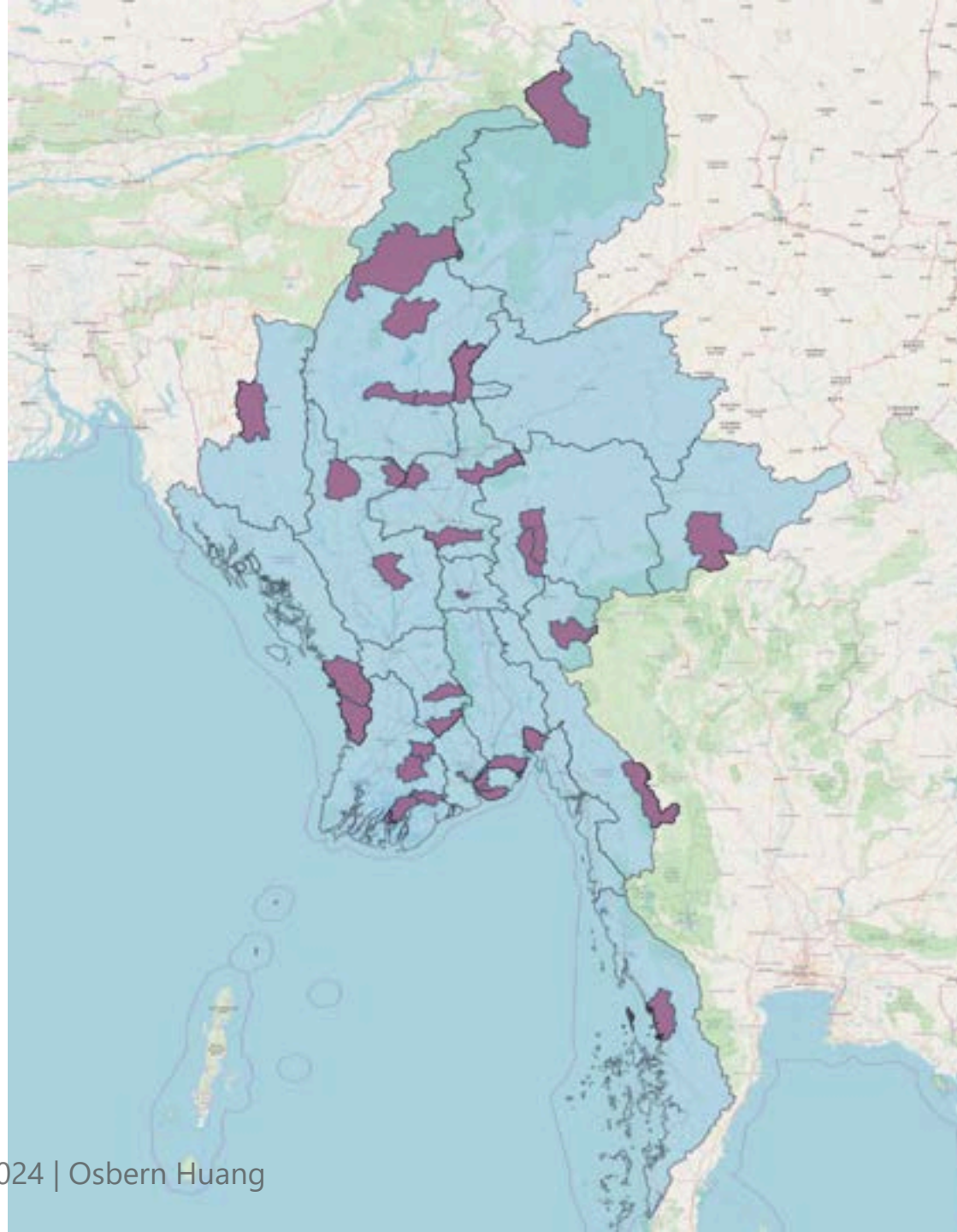
Figure 5. Multistage Sampling

Primary Sampling Units

Secondary Sampling Units

Tertiary Sampling Units

...

Penultimate Sampling Units

Ultimate Sampling Units: The Individual

ASIAN BAROMETER

- 2001 第一波: 8國 ⇒ 2021 第五波: 17國
- PPS的操作範例:

中國大陸	台灣
PSU: 省市 (125) SSU: 縣市 (2) TSU: 鄉鎮街道 個人：地址做戶中選樣*	PSU: 選區 (28) SSU: 村里 (3) 個人- 戶口資料

*利用Kish表進行戶中選樣

Why PPS?

Difficulties Surfacing...

- In most places, **electoral registers and postcode addresses** were used as sampling frames (Lynn & Taylor, 1995).

- However, people **moves** (Landry & Shen, 2005; Chen et al., 2018)
  - Migrants, temporary or seasonal workers, students, etc.

- Even worse, sometimes government officials are **reluctant to share household registration data**.
  - Or, they simply **don't have it**.

Objective of the "Mission"

- Emergency problem-solving-oriented project.
- **Without...**
  - Electoral/household registers and postcode addresses in **China and Myanmar**.
  - Long-term guarantee for the survey permission from the GAD or government officials.
- *No reliable sampling frame* for the last stage of the PPS sampling.
- With emerging resources of GIS data, we have a chance to utilize it for survey sampling.

RECIPES FOR WAFFLES

Abbreviations
- t.—teaspoon
- T.—Tablespoon
- C.—Cup

No. 1—

2 eggs	2 C. flour
2 C. milk	2 t. baking powder
2 t. sugar	1 t. salt

5 T. melted butter

Beat eggs thoroughly. Sift flour and baking powder. Add milk to eggs, then add dry in-gredients and **beat mixture thoroughly. Then add butter.**

No. 2—

3 eggs	1 ½ C. flour
2 t. sugar	2 C. milk
2 t. baking powder	½ t. salt

5 T. melted butter

Beat yolks and whites separately. Add bak-ing powder to flour and sift. Add milk to beat-en yolks, add dry ingredients. **Add melted butter, then fold in beaten whites.**

SOUR MILK WAFFLES

No. 3—

2 eggs	⅔ t. baking powder
2 C. flour	¼ t. salt
1 t. soda	2 C. sour milk

5 T. melted butter

Beat eggs, add milk, add sifted flour with soda and baking powder. Add other dry in-gredients. **Add melted butter.**

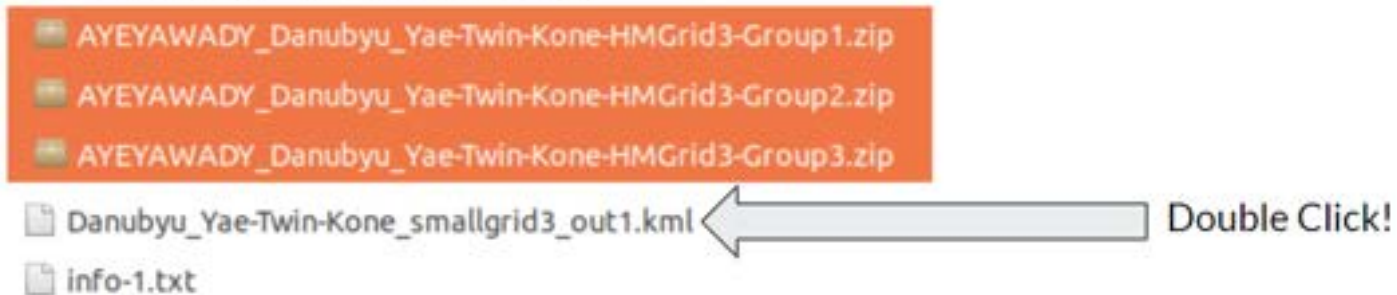(over)

Ready, steady, cook.

1. Download UN-adjusted gridded population estimates from Worldpop.

2. Batch download administrative boundary `.shp` files using `curl` from OSM boundaries.

3. To segment Worldpop estimates into sub-administrative files with `GDAL` function in QGIS.

4. Grid the sub-administrative population file and compute the population count (half-minute grid).

5. Repeat the process to produce corresponding smaller grids (100m x 100m).

6. Review and verify the results.

Image: RecipeCurio.com

16

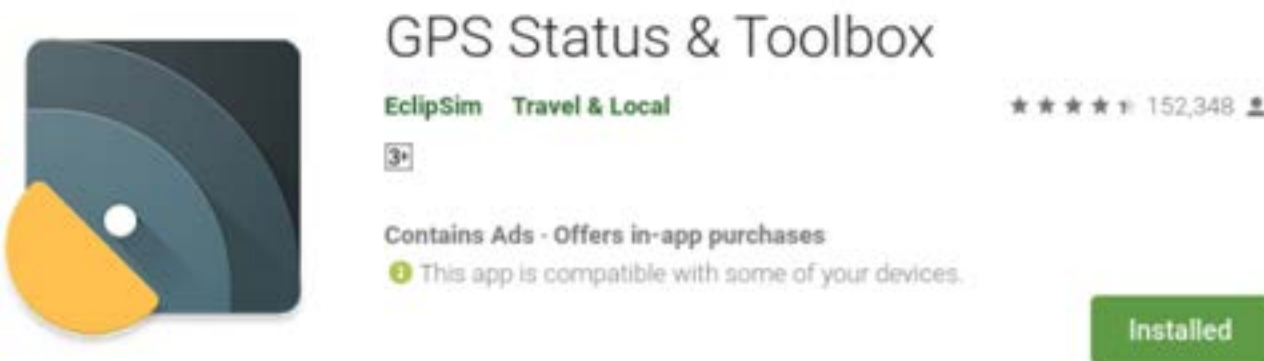# Fieldwork Supplies for Supervisors/Interviewers

In every TSU, our fieldwork supervisors and interviewers will be provided with:

Open GIS Sample Files

Taking "AYEYAWADY-Danubyu-Yae-Twin-Kone" as an example, after unzipping the file, there are three folders, with two files in each folder:
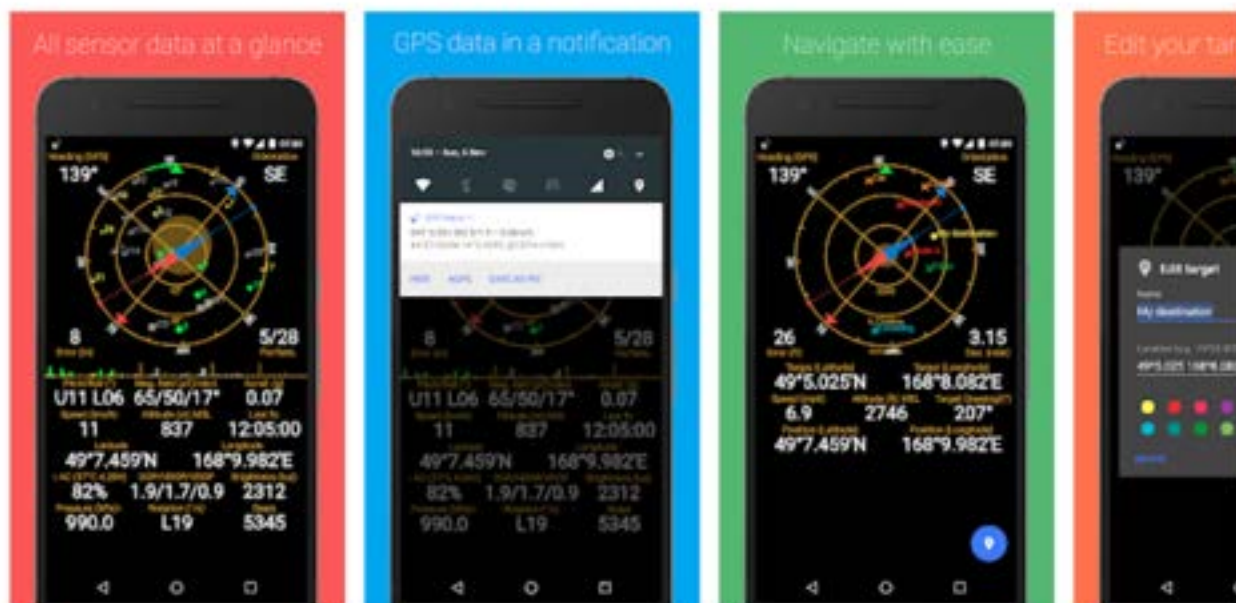
- AYEYAWADY_Danubyu_Yae-Twin-Kone-HMGrid3-Group1.zip
- AYEYAWADY_Danubyu_Yae-Twin-Kone-HMGrid3-Group2.zip
- AYEYAWADY_Danubyu_Yae-Twin-Kone-HMGrid3-Group3.zip
- Danubyu_Yae-Twin-Kone_smallgrid3_out1.kml ⟵ Double Click!
- info-1.txt

-->

Tools Utilized On-Site

- **GPS devices (basic smartphone works!)** for precise location tracking.
- **Google Earth/Maps** for navigational guidance.
- **Pen and paper** for documentation.

Image: RecipeCurio.com

18

Test for Sample Representativeness

China

- Education: Highlighting those with College education and above.

	Sample	Population	
No college degree	81.48	81	P>.05
College ⇑	18.52	19	

- Party membership status

	Sample	Population	
CCP member	11.58	9	P<.01
Not CCP member	88.42	91	

Myanmar

- Education: Highlighting those with College education and above.

	Sample	Population	
No college degree	90.17	90	P>.05
College ⇑	9.83	10	

- Buddhism Religion

	Sample	Population	
Buddhism	86.78	88	P>.05
Not Buddhism	13.22	12	

Understanding Myanmar Through the Lens of the Asian Barometer Survey:

Democratic Norms and Protest Participation

Osbern Huang

Assistant Research Fellow,
Asian Barometer Survey / Hu Fu Centre
for East Asian Democracy
National Taiwan University

2007　　　　**2010**　　　　**2015**　　　　**2020**

Questions

1. Under the NLD intellectual government, **have the people of Myanmar become more inclined towards a democratic system?**

2. Is there enough **political culture/democratic values** to support democratic transition?

Indicators We Used To Understand Democratic Support in Myanmar

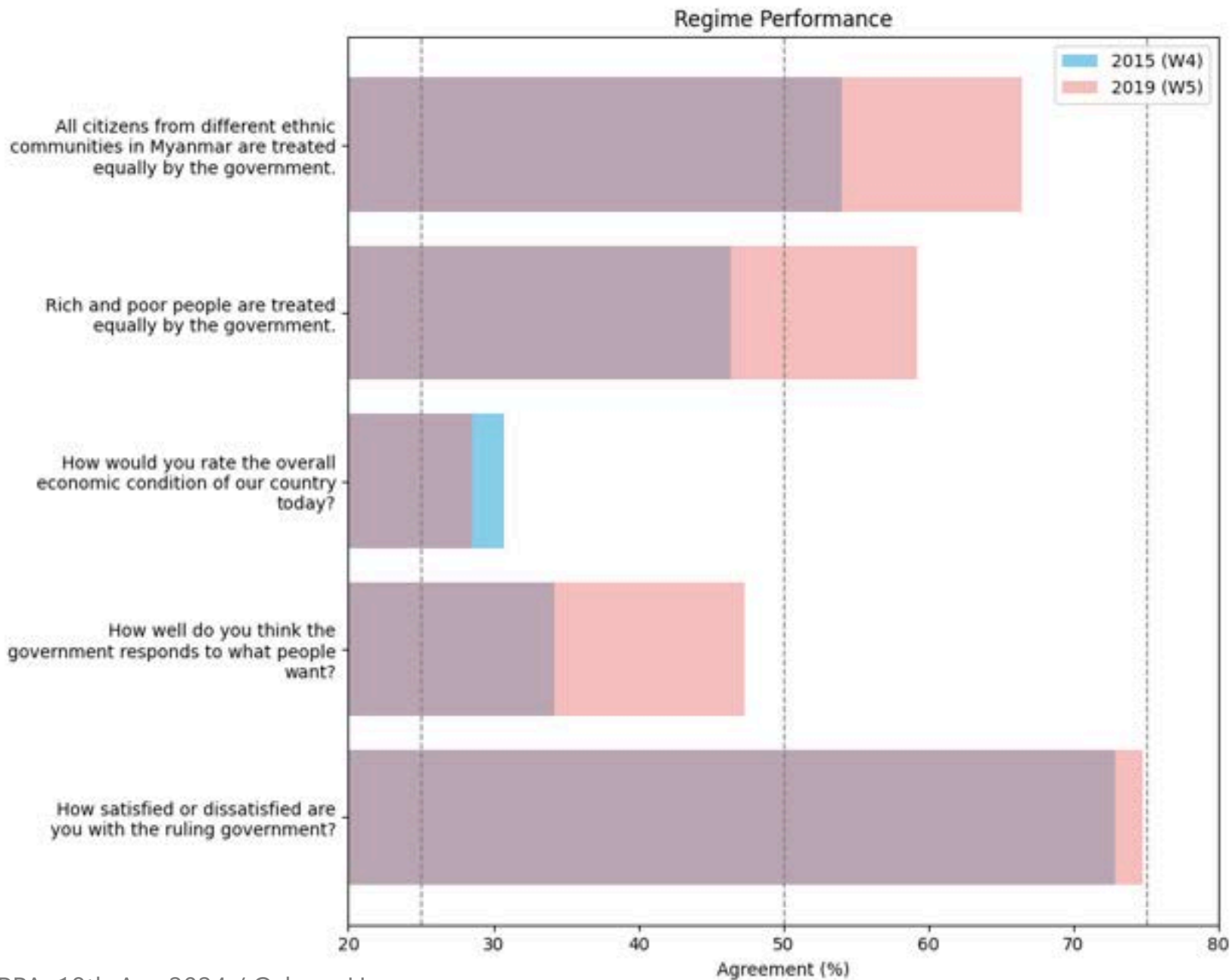**Support for the current *democratic* regime**

- Regime Performance

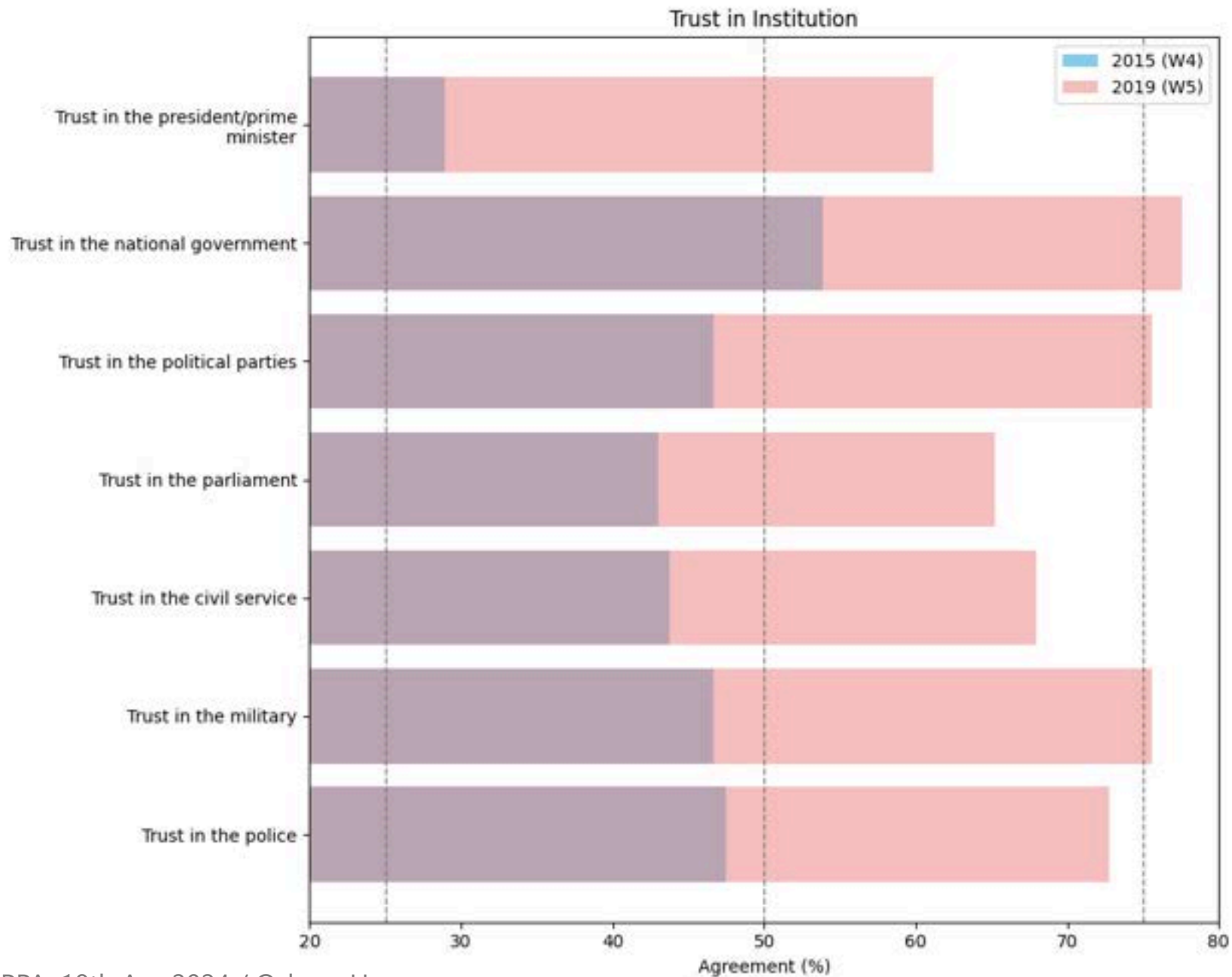- Trust in Political Institution

- Regime Support

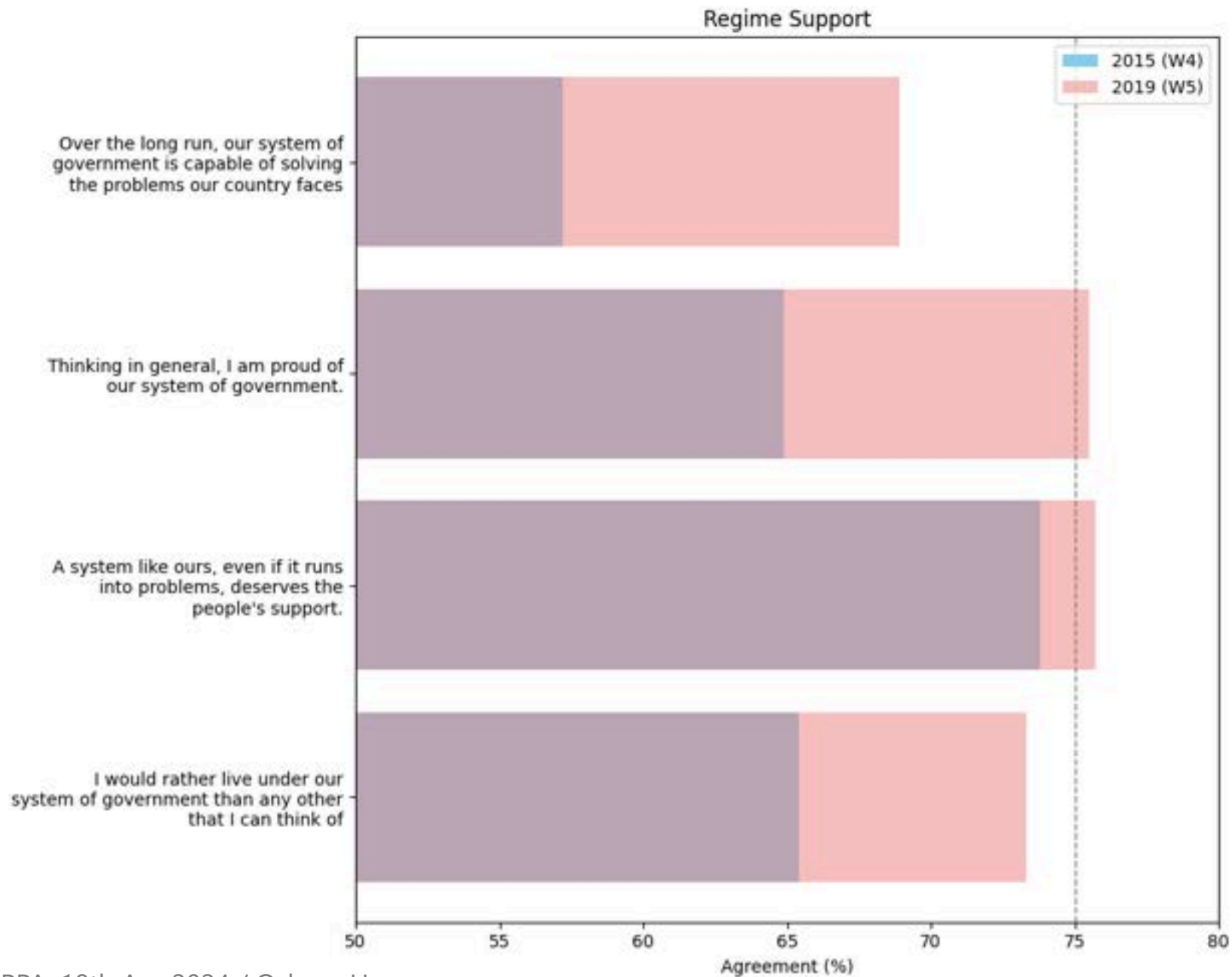**Recognition of the *values of liberal democracy***

- Detach From Authoritarian Alternatives

- Liberal Democratic Values

Support for the current democratic regime

- Regime Performance

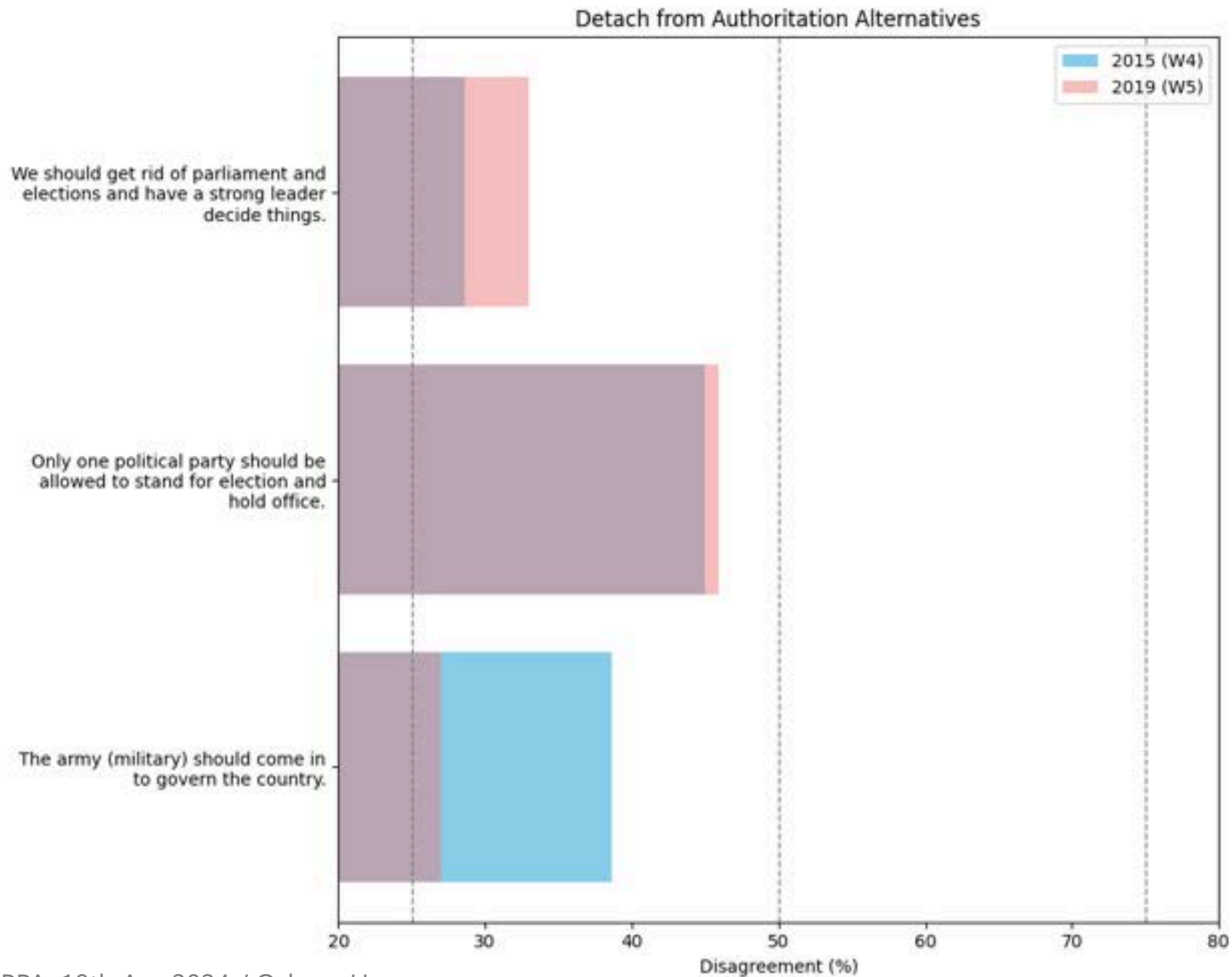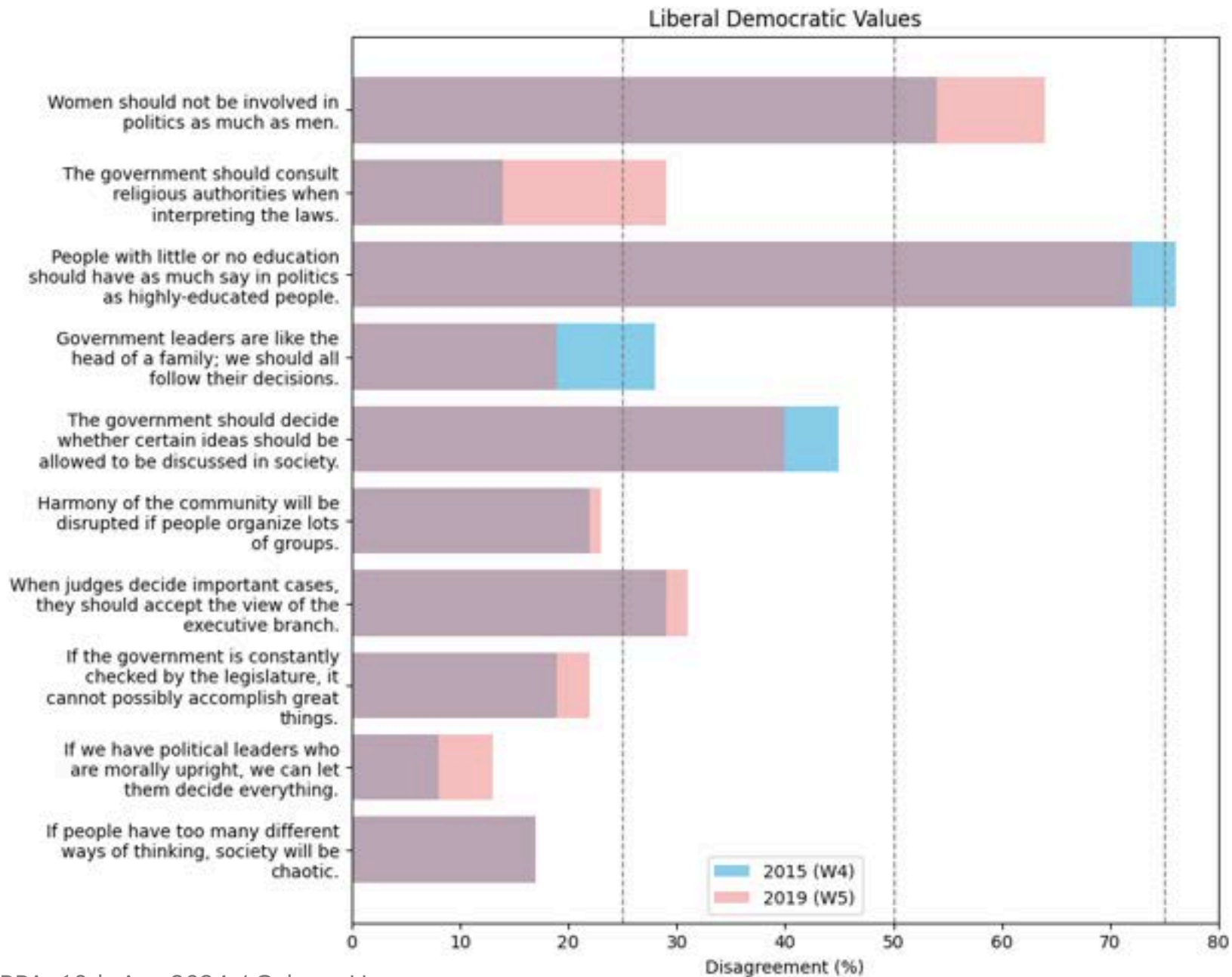- Trust in Political Institution

- Regime Support

Regime Performance

Trust in Institution

Regime Support

Recognition of the values of liberal democracy

- Detach From Authoritarian Alternatives

- Liberal Democratic Values

Detach from Authorization Alternatives

Liberal Democratic Values

Quick Summary to 2015/2019 Comparison

1. Trust in government institutions and regime performance has greatly risen under a "well-performing" regime, reinforcing **support for the current political system**.

2. However, individual democratic values, including liberal-democratic ideals and a shift away from authoritarian alternatives, have **not notably improved**. Instead, the patriarchal concept of political values has been reinforced.

2021: The end of democratic experiment

On February 1, the military detains Aung San Suu Kyi and other NLD leaders, seizing power and sparking international condemnation and widespread protests (Civil Disobedience Movement, CDM).

The Unprecedented Social Mobilization in Myanmar

- Massive public outrage against military coup

- Public demonstrations, protests, civil disobedience, general strikes

- Unity across diverse segments of society, including ethnic and religion

- Despite dangers and uncertainties, movements persist

- Shared aspiration for democracy

Image from ANU Myanmar Research Centre

35

Historical Precursors of Democracy Struggles

Historical Movements

- 8888 Uprising (1988) against military dictatorship
  - Emergence of Aung San Suu Kyi and her significance
- Saffron Revolution (2007) led by Buddhist monks
  - Broader struggle for political freedom

Violence and Resilience

- Military regime's violent responses
- International condemnation and casualties
- Creation of underground civil society

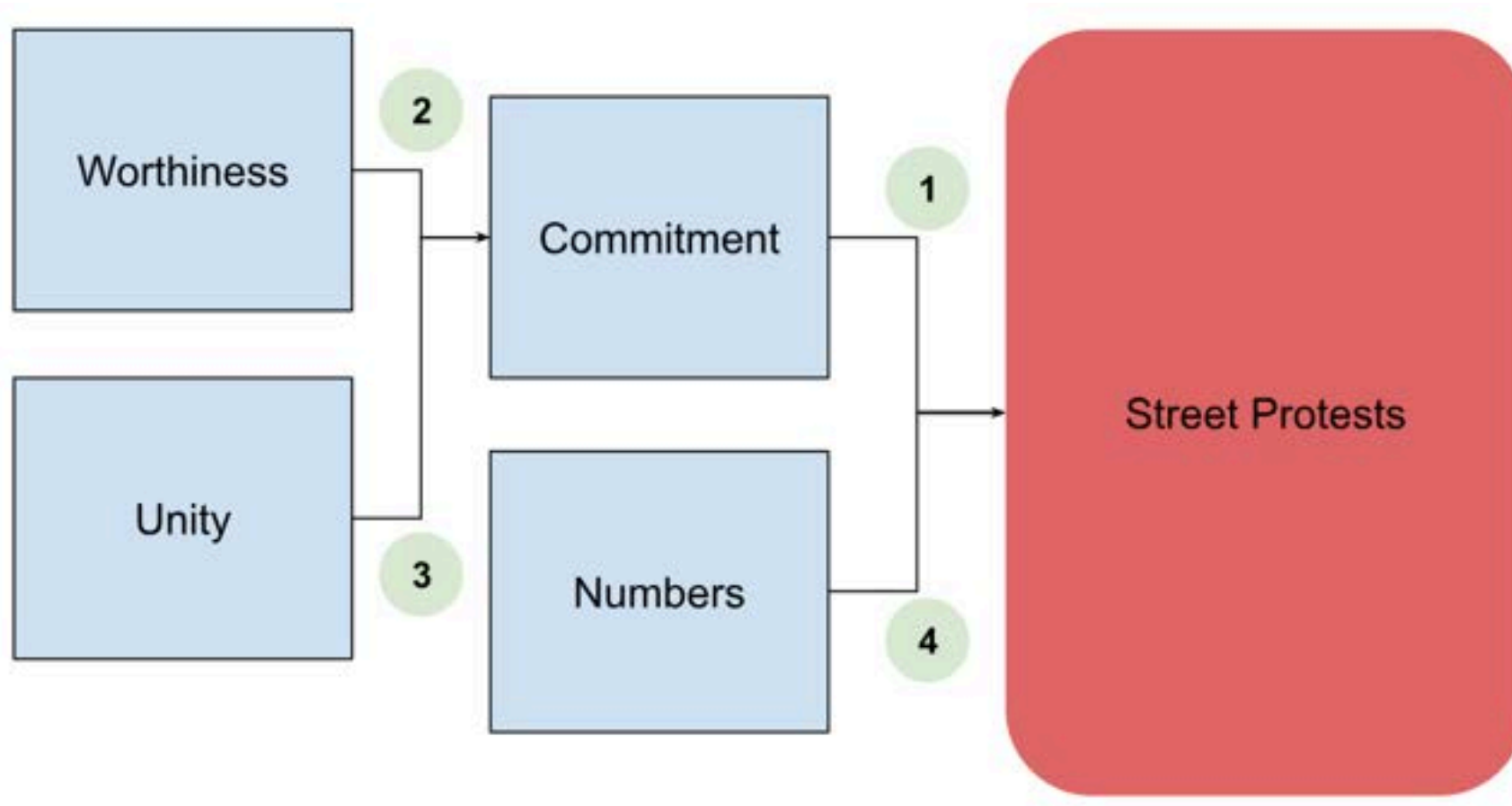## Understanding Social Movements with *WUNC* framework

- **Worthiness**
  - Demonstration of righteous and virtuous cause
  - Convincing power holders and the public of movement's merit
- **Unity**
  - Shared identity, solidarity, and mass mobilization
  - Strengthening the cause and its appeal

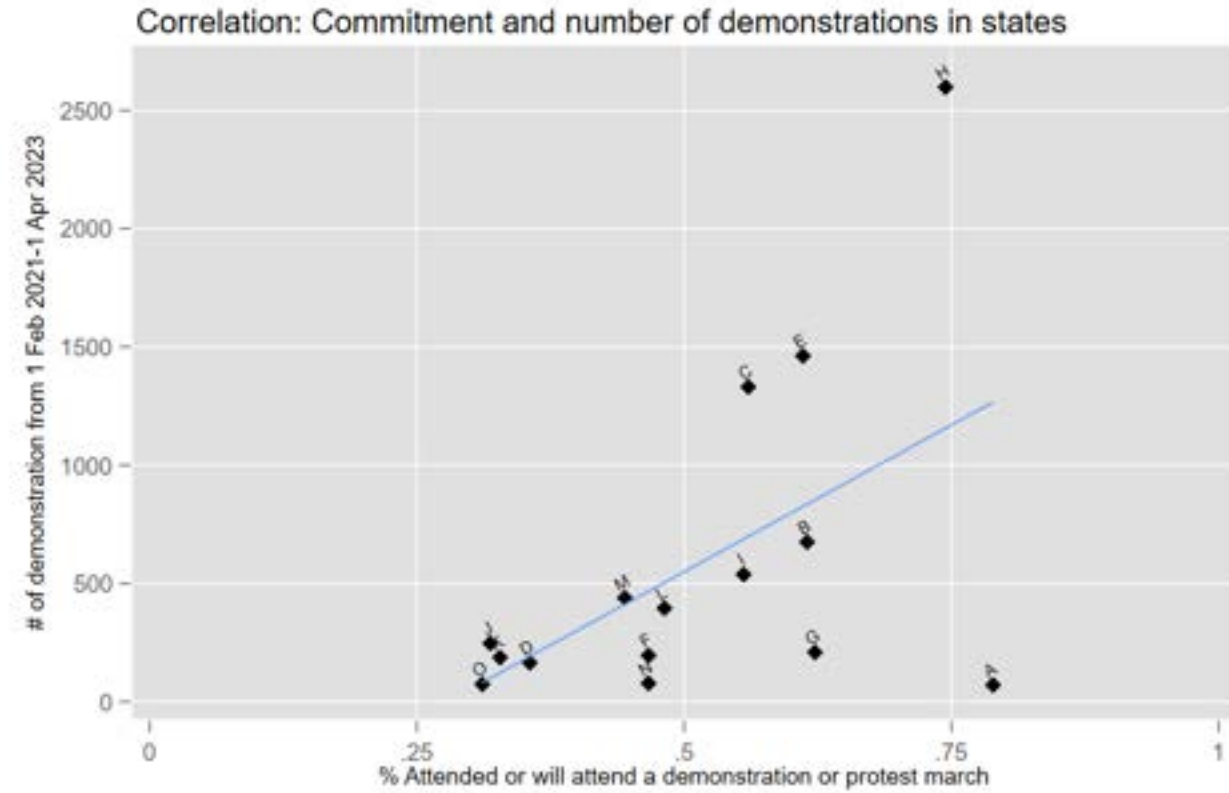# Understanding Social Movements with *WUNC* framework
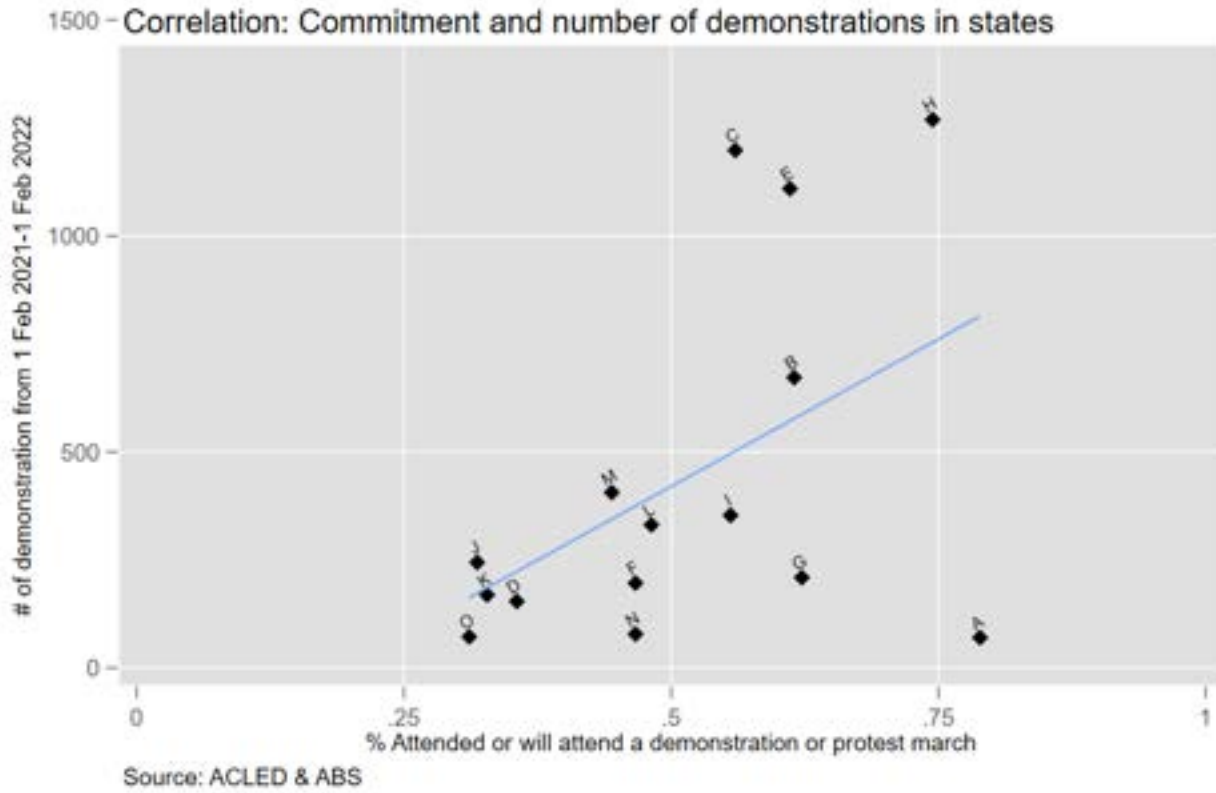
- **Numbers**
  - Power in mass mobilization in reflection of broad support

- **Commitment**
  - Willingness to endure risks and costs
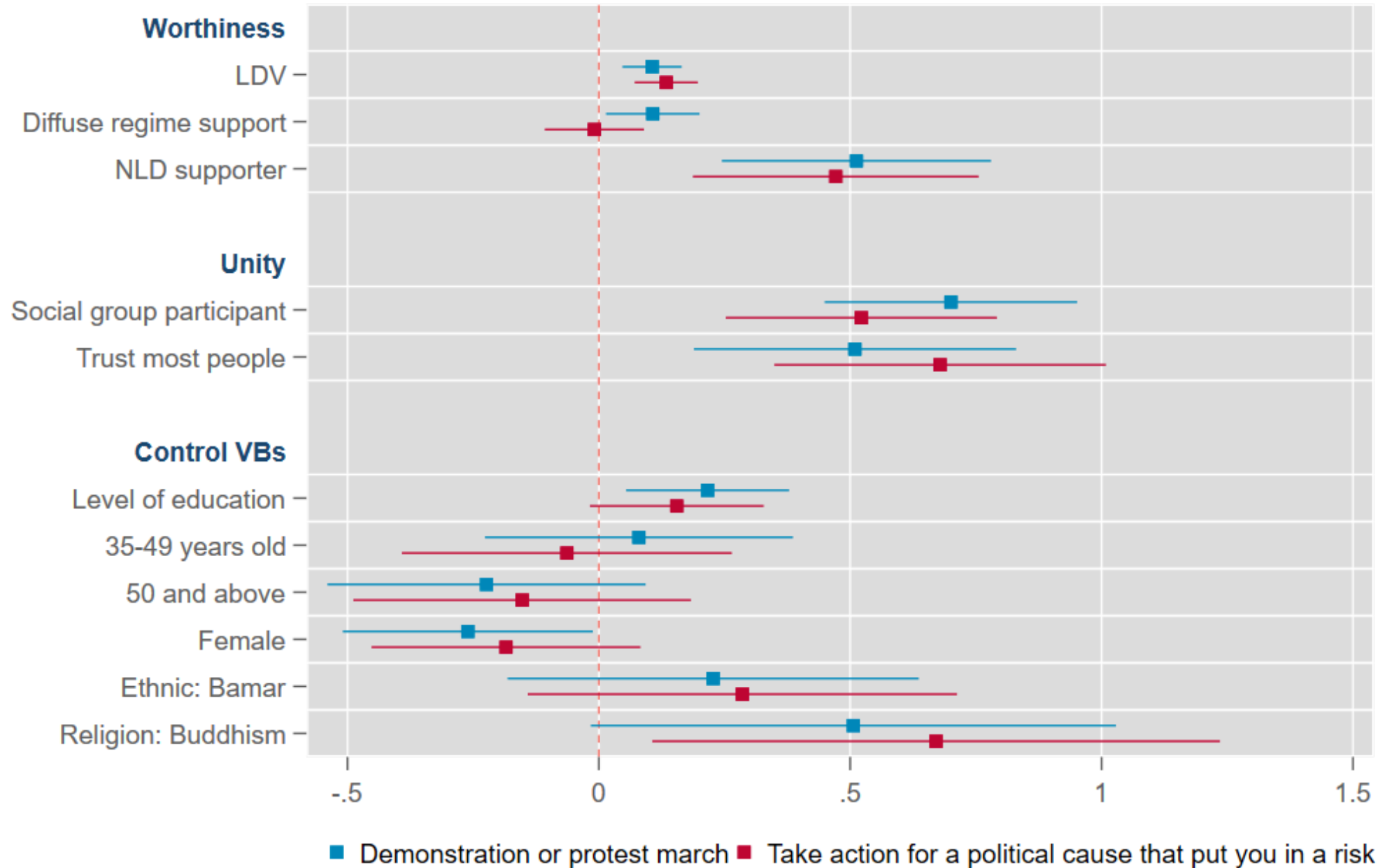 - Persistence in the face of repression or opposition

- "Commitment" as risk-taking mindset: key driver for mobilization
- "Numbers" as a reference of level of mobilization
- "Worthiness" and "unity" as catalysts for mobilization

Findings from correlation and regression

Correlation: Commitment and number of demonstrations in states

Worthiness
- LDV
- Diffuse regime support
- NLD supporter

Unity
- Social group participant
- Trust most people

Control VBs
- Level of education
- 35-49 years old
- 50 and above
- Female
- Ethnic: Bamar
- Religion: Buddhism

Legend:
- ■ Demonstration or protest march
- ■ Take action for a political cause that put you in a risk

Summary

- **Democratic aspiration inflames people's commitment to continuing protest**:
  - Importance of democratic values and NLD support in galvanizing social resistance
  - Empirical support for social capital's role in fostering civic engagement
  - Commitment and Numbers matter in predicting future demonstrations

"Remember, remember..."

- Probability sampling and sampling frame
- PPS (probability proportional to size) sampling for large survey project
- Utilize national survey data to answer political science and public policy research questions
- Values of national representative survey

Thanks for having me!

Contact:

Osbern Huang

osberntw.github.io

osbern@gmail.com