Uninstitutional Political Participation in Taiwan:Hierarchical Logistic Regression Analysis of the 2019 WVS

Jinyi Tzeng (47969938)

Table of contents

1 Abstract

Non-institutional political participation—such as protests, petitions, and boycotts—represents expressive forms of civic engagement that require substantial individual motivation and resources. Building on the typology by Verba, Schlozman, and Brady (1995), this study investigates the drivers of such participation in Taiwan using data from the 2019 Taiwan subsample of the World Values Survey (World Values Survey Association, 2019).

This study applies a logistic regression framework guided by the Civic Voluntarism Model (CVM), incorporating predictors in conceptually structured blocks: socioeconomic status (education, income, subjective social class), demographics (age, gender), psychological engagement (political interest, discussion frequency), media use, and Taiwan-specific variables (party preference, institutional trust, organizational participation, diffuse support).

Findings indicate that income level, age, political interest, discussion frequency, social media use, and organizational participation are significant positive predictors. Males were significantly less likely to participate than females. In contrast to expectations, education level and political trust variables did not show significant effects. These results highlight the combined roles of structural resources and psychological engagement, while also suggesting that institutional trust may not directly influence non-institutional participation in Taiwan's 2019 context.

2 Introduction

Politics encompasses the management and organization of public affairs, yet its meaning varies across contexts. In democratic systems, political participation not only reflects civic engagement but also serves as a mechanism for citizens to influence decision-making and hold leaders accountable. Understanding the drivers of such participation is thus essential for maintaining a responsive democracy.

Scholars like Verba, Schlozman, and Brady (1995) classify political participation into three categories: institutional (e.g., voting), non-institutional (e.g., protests, petitions), and contact-based (e.g., contacting officials). Among these categories , non-institutional participation is particularly expressive and resource-dependent, reflecting citizens' motivations and capacity to act. Earlier studies (e.g., Dalton, 2008) highlight its growing prevalence, especially among youth and digitally active populations, making it a critical area for contemporary research.

In Taiwan, research on political engagement has long focused on institutional behaviors like voting. However, non-institutional participation—such as demonstrations or petitions—has gained traction, particularly among younger generations and civil society groups. Despite this shift, individual-level empirical studies examining its predictors remain scarce. Most existing work is either descriptive or centered on institutional forms, leaving a gap in understanding who engages in alternative modes of participation and why.

To address this gap, we employ the Civic Voluntarism Model (Verba et al., 1995), which identifies resources, psychological engagement, and recruitment as key drivers of participation. Using this framework, we build a theory-driven logistic regression model, beginning with socioeconomic factors (e.g., education, income), then adding demographic variables (e.g., age, gender), and finally incorporating psychological predictors (e.g., political interest). Our study aims to identify the strongest predictors of non-institutional political participation in Taiwan, offering an empirical overview based on nationally representative survey data.

3 Methodology

This study uses data from Wave 7 of the World Values Survey (WVS), a global research initiative that collects data on public attitudes and values across multiple domains, including political behavior, cultural norms, religion, and scientific views. The analysis focuses on the 2019 Taiwan subsample, which includes 1,223 respondents and is designed to be nationally representative through stratified random sampling. Post-stratification weights were not applied in this analysis due to variable-level alignment limitations (World Values Survey Association, 2019).

To investigate the factors associated with non-institutional political participation in Taiwan, a binary participation indicator was constructed based on 11 relevant items from the WVS questionnaire (Q209–Q220, excluding Q217). These items capture a range of non-institutional political actions, such as protesting, petitioning, and boycotting. Each item was recoded as 1 if the respondent reported having participated in the activity, and 0 otherwise. A binary variable, participation_binary, was created to indicate whether a respondent engaged in at least one such action. Respondents with missing values across all 11 items were excluded through listwise deletion.

Table 1: Variables used to construct the political participation index.

Question	Description
Q209	Signing a petition
Q210	Joining in boycotts
Q211	Attending peaceful demonstrations
Q212	Joining unofficial strikes
Q213	Donating to a political campaign
Q214	Contacting a government official
Q215	Encouraging others to take action about political issues
Q216	Encouraging others to vote
Q218	Signing an electronic petition
Q219	Encouraging others to take any form of political action
Q220	Organizing political activities, events, or protests

A hierarchical logistic regression framework is employed to examine the predictors of non-institutional participation. Here, the term "hierarchical" refers to a theory-driven, blockwise modeling strategy in which predictors are introduced sequentially in conceptually organized layers—rather than to multilevel (nested) data structures. The five conceptual layers are as follows:

- **Layer 1: Core Socioeconomic Status (SES)**
  Guided by the Civic Voluntarism Model (Verba, Schlozman, & Brady, 1995), the baseline model includes three resource-related variables: education level, income level, and subjective social class.
- **Layer 2: Demographic Controls**
  Following guidance from Gelman and Hill (2006), demographic characteristics—specifically age and gender—are included to control for population heterogeneity.
- **Layer 3: Psychological Engagement**
  Drawing on Dalton's (2008) typology of political orientations, this layer incorporates political interest and the frequency of political discussion.
- **Layer 4: Media Use**
  Based on Prior's (2007) media choice theory, this layer includes both traditional and digital media use variables capturing respondents' exposure to political content.
- **Layer 5: Taiwan-Specific Contextual Factors**
  The final layer includes variables commonly featured in Taiwanese political research, such as party preference (categorized as Pan-Blue, Pan-Green, or Neutral/Other), political trust in government and parties, diffuse support for democracy, and organizational participation.

This layered modeling approach enables the systematic evaluation of how distinct theoretical domains contribute to non-institutional political participation while maintaining comparability across nested models. All predictors were preprocessed using theory-informed recoding. Ordered categorical variables (e.g., education, political interest, trust) were converted into numeric scales preserving their ordinal structure. Political participation was measured using an 11-item index from the WVS (Q209–Q220, excluding Q217), and a binary variable was created to indicate any engagement. Media use and trust items were standardized into numeric scores, and organizational participation was summarized as a binary indicator. All factor levels were explicitly defined, and a complete-case dataset was used for model estimation.

Multicollinearity was assessed using variance inflation factors (VIF), with all values below 2. Model fit and parsimony were evaluated using likelihood ratio tests (LRT), Akaike information criterion (AIC), and Bayesian information criterion (BIC), with McFadden's pseudo-R^2 reported as a measure of explanatory power. Although sensitive political behaviors such as protesting may be underreported due to social desirability bias, this limitation is acknowledged when interpreting results.

Table 2: Table 2. Five Conceptual Layers in the Hierarchical Modeling Strategy

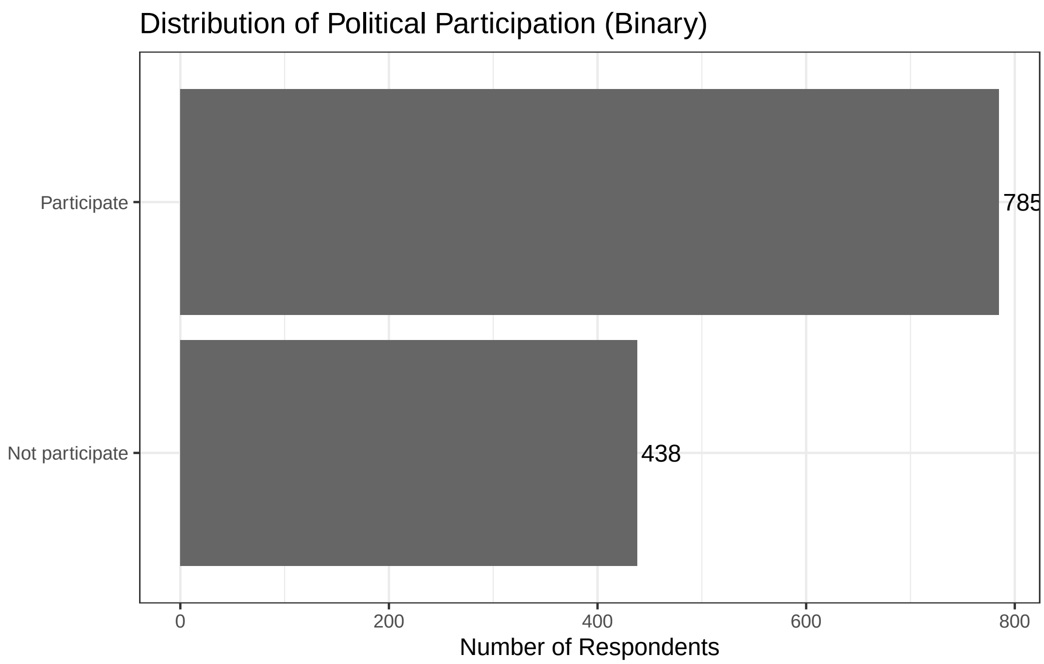
Layer	Label	Description
Layer 1	Core SES	Education, income, subjective social class (Verba et al., 1995)
Layer 2	Demographics	Age and gender (Gelman & Hill, 2006)
Layer 3	Psychological Engagement	Political interest, discussion frequency (Dalton, 2008)
Layer 4	Media Use (4 sub-layers)	Includes traditional news, traditional politics, digital news, digital politics (Prior, 2007)
Layer 5	Taiwan-Specific Context	Party ID, political trust, diffuse support, org. participation

4 Exploratory Data Analysis

4.1 Participation Overview

Among the 1,223 respondents, 785 individuals (64.2%) reported having engaged in at least one type of non-institutional political activity—such as petitioning, protesting, or boycotting—while 438 (35.8%) reported no participation in any of the listed actions.

This indicates that expressive civic engagement was relatively prevalent in Taiwan in 2019, and provides a meaningful base for comparing participation rates across socioeconomic, psychological, and contextual factors in the following sections.



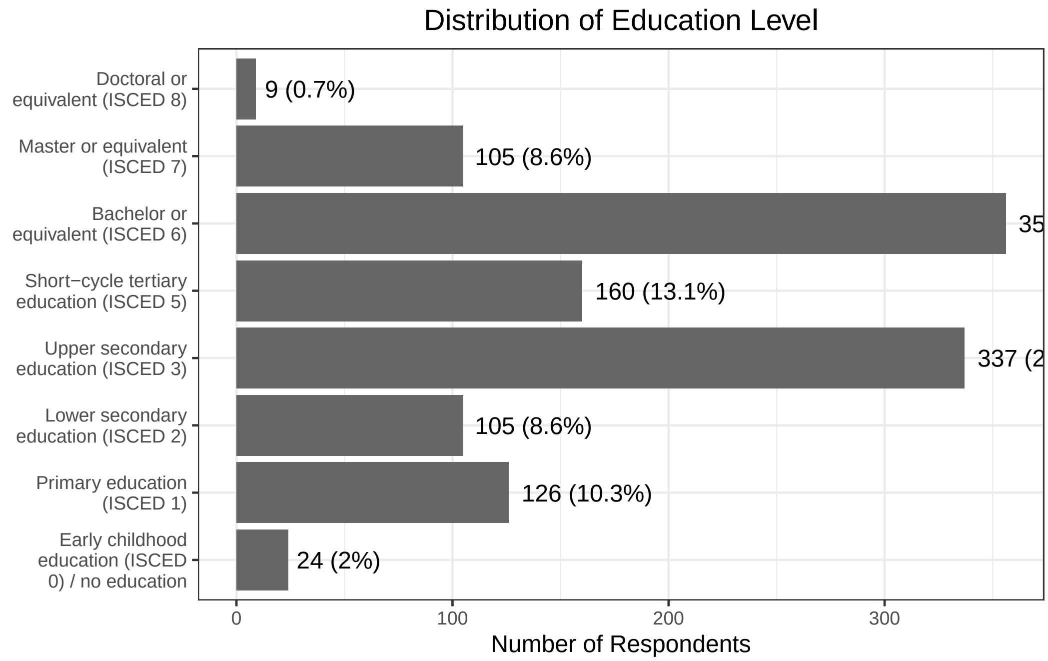
4.1.1 SES layer

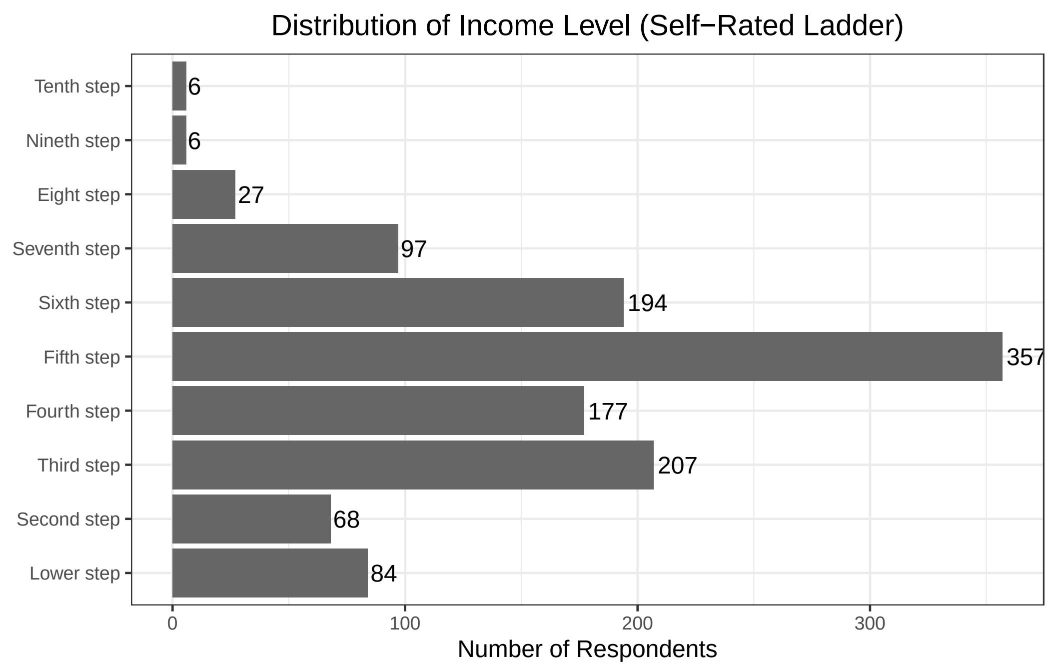
Education Level presents the distribution of respondents' highest level of education.The majority of respondents possess medium to high levels of education. In particular, Bachelor's degree (ISCED 6) and Upper secondary education (ISCED 3) are the most common, accounting for approximately 28.1% and 27.6% of the sample, respectively. This suggests that a substantial portion of the adult population in Taiwan holds considerable educational capital.

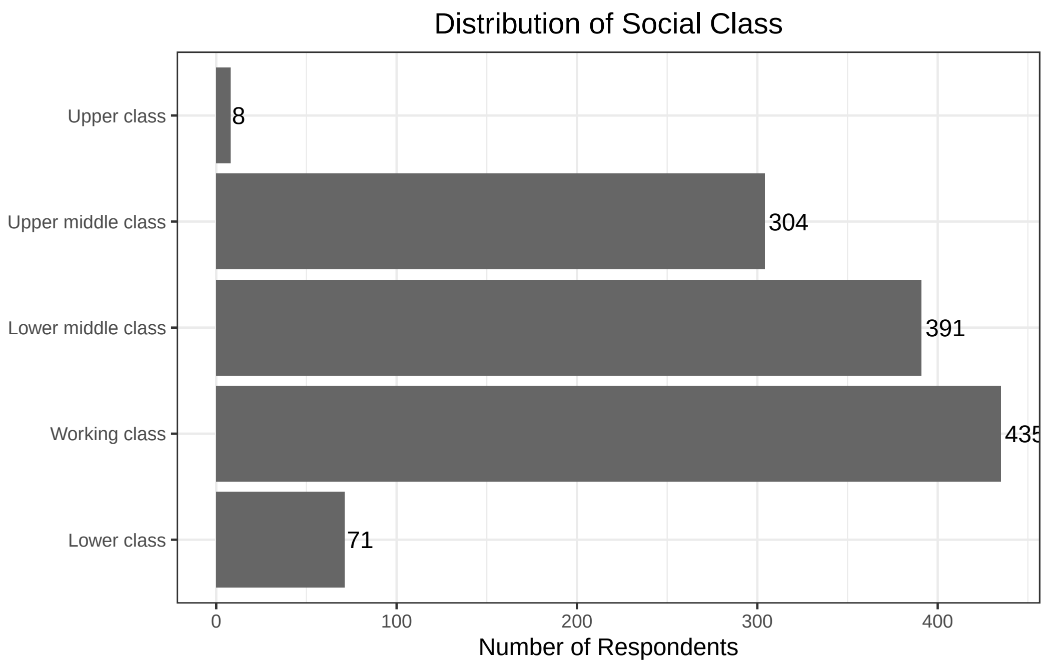
Income Level (Self-Placement on 10-Step Ladder) As shown in figure. Respondents were asked to rate their perceived income status on a 10-step ladder, with 1 representing the lowest and 10 the highest level of income. This subjective measure captures how individuals view their economic position relative to others in society., with Step 5 (30.6%) being the most frequently selected. The distribution is roughly bell-shaped, skewed slightly left, indicating a self-perceived concentration around the middle. This variable captures perceived economic standing, rather than objective income, and helps reflect respondents' relative sense of economic status.

The **Social Class** figure displays the distribution of respondents' self-identified social class. Most individuals identified as working class (33.6%), lower middle class (30.2%), or upper middle class

(23.4%), with only a small minority selecting lower class (5.5%) or upper class (0.6%). As with the income ladder, this is a self-assessed measure, but it offers insight into how individuals view their place in the social hierarchy.



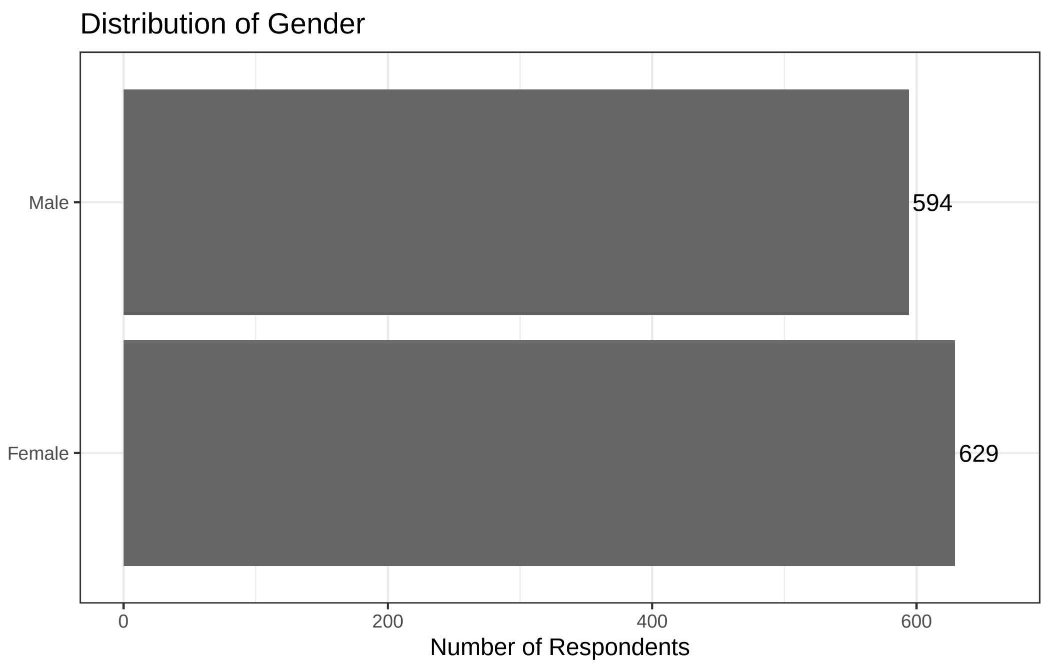


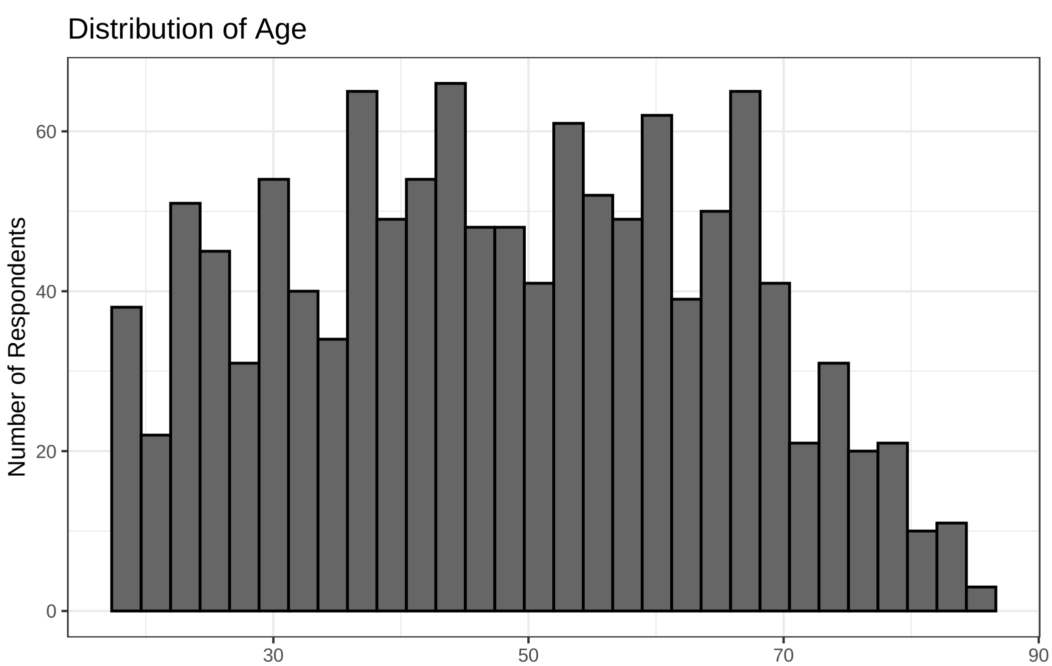


4.1.2 Demographic layer

The **gender** distribution is relatively balanced, but women show slightly higher levels of non-institutional participation. This could reflect differences in civic engagement motivation or opportunity structures by gender.Compared to Western democracies where male citizens tend to participate more in protests and petitions, the higher female engagement observed in Taiwan might reflect stronger mobilization among women through community-based or social justice movements.

The **age** distribution of political participation in Taiwan reveals a pattern in which younger and middle-aged respondents (especially those aged 20–50) are more likely to engage in non-institutional forms of political action. This trend aligns with global findings (Dalton, 2008) that younger generations are more inclined toward expressive and unconventional forms of participation.

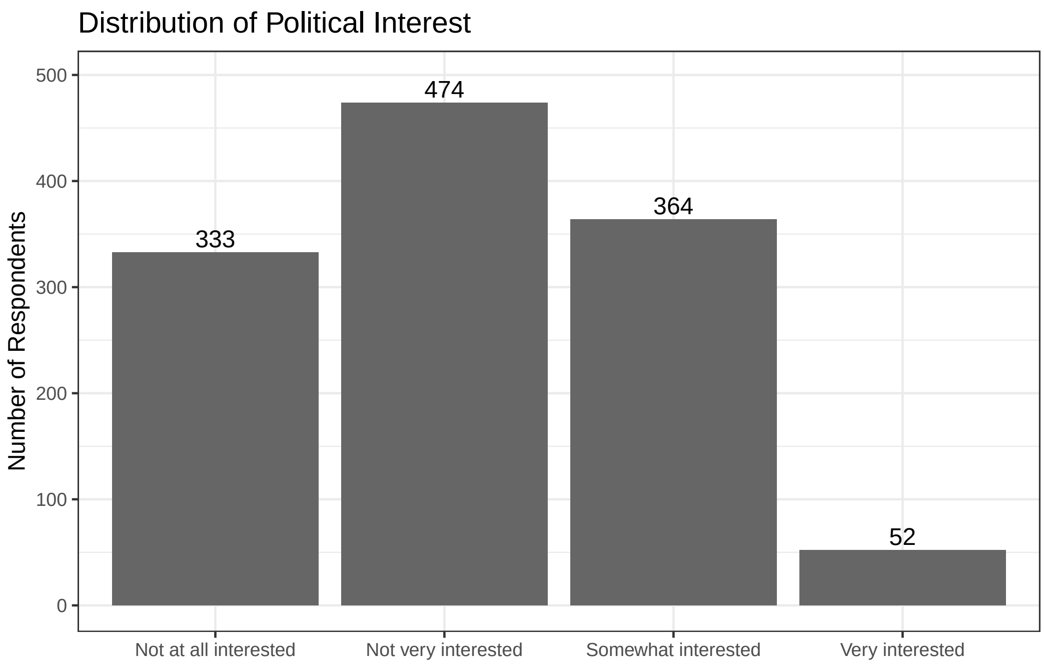


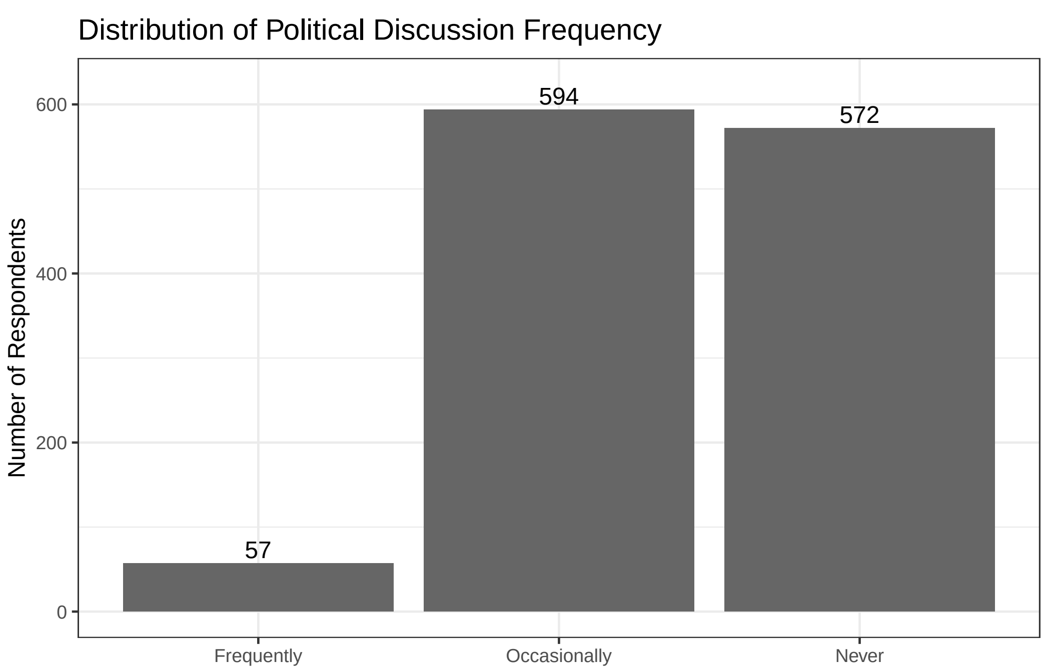


4.1.3 Psychological Engagement layer

The distribution of **political interest** in Taiwan reveals a population that is largely disengaged from politics. The majority of respondents identified as either "Not very interested" or "Not at all interested," with only a small minority reporting high levels of interest. Specifically, just 52 individuals expressed being "Very interested" in politics. This pattern suggests that political interest remains relatively low among the general public, which may limit psychological motivation for non-institutional political participation. According to Dalton (2008), low political interest is a common feature in many democracies and tends to suppress citizen engagement in expressive forms of action such as protests or petitions.

The distribution of **political discussion frequency** indicates that most respondents rarely engage in political conversations. Nearly half reported that they “never” discuss politics, while another substantial portion said they do so only “occasionally.” In contrast, just 57 individuals reported discussing politics “frequently.This limited frequency of political discussion may reflect a broader sense of disengagement or discomfort with political dialogue.





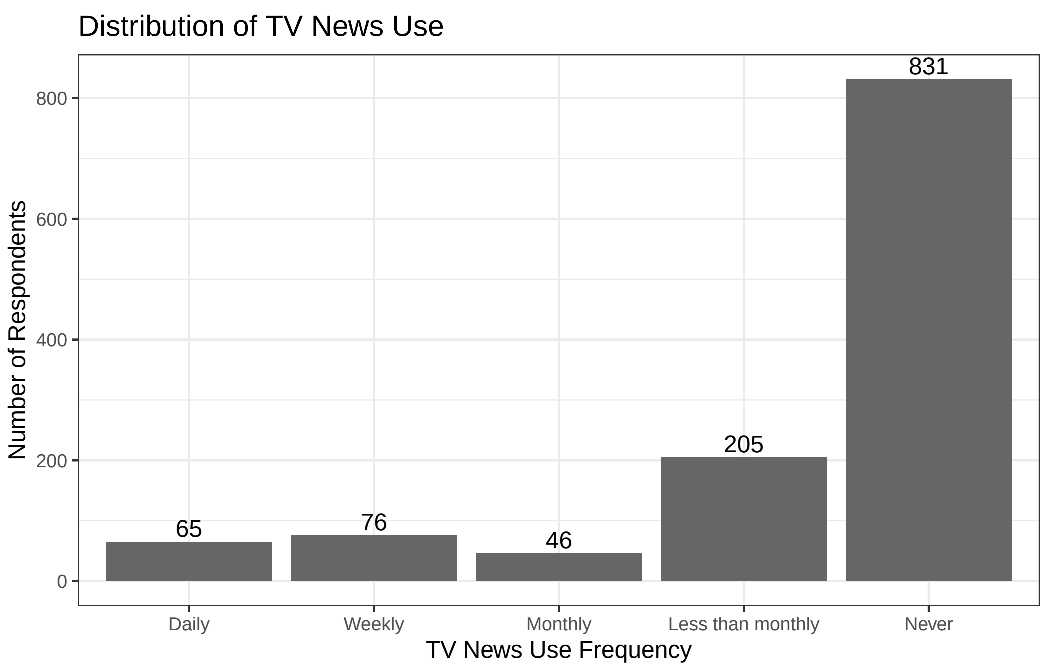
4.1.4 Media Use layer

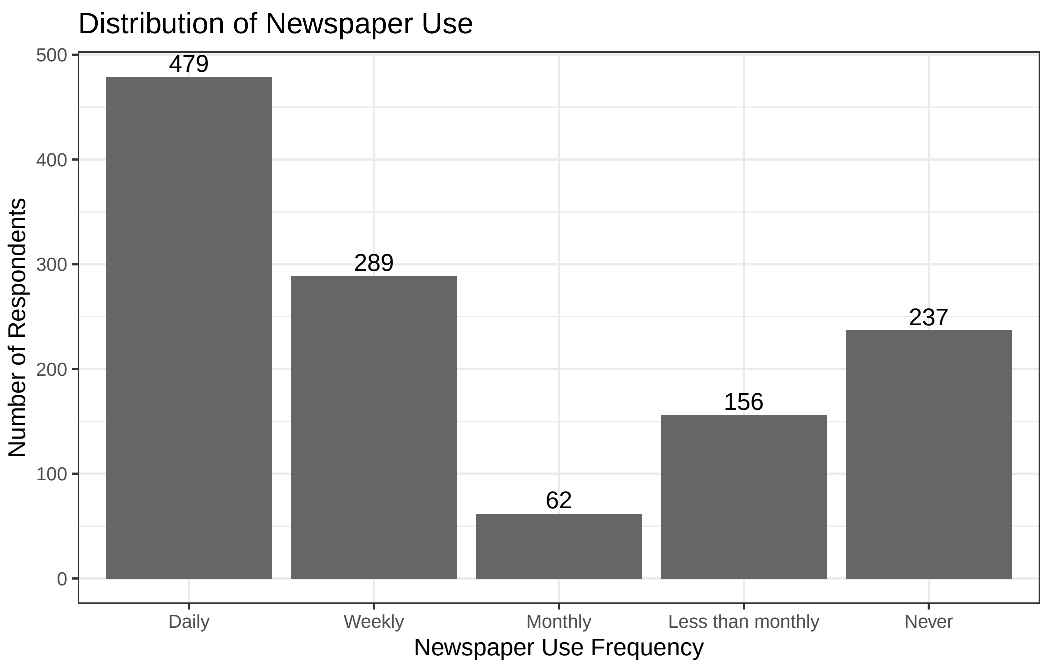
The distribution of **TV news** shows that an overwhelming majority of respondents (831) reported never watching TV news. Daily viewers were the smallest group (65), with the rest falling in between. This pattern suggests that traditional television news consumption is relatively uncommon among the surveyed population. The decline in regular TV viewership may reflect generational media shifts or growing reliance on alternative digital sources.

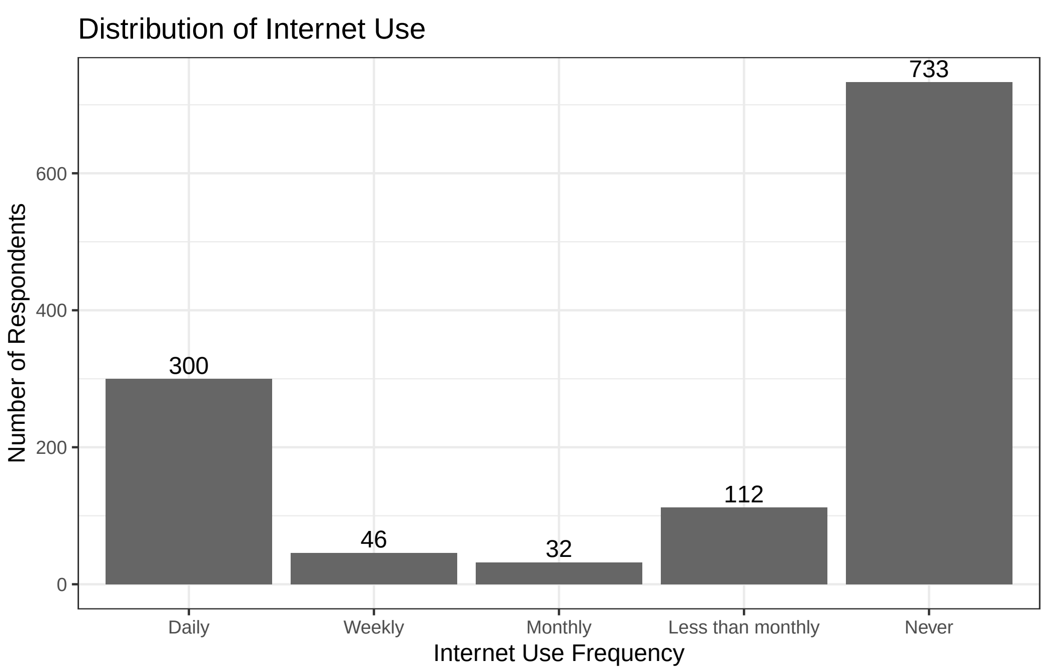
The newspapers shows a substantial number of respondents (479) reported reading newspapers daily, followed by weekly readers (289). Only 237 respondents reported never reading newspapers. This indicates that newspapers may remain a relevant source of political information for certain demographic groups, possibly older or more civically engaged individuals.

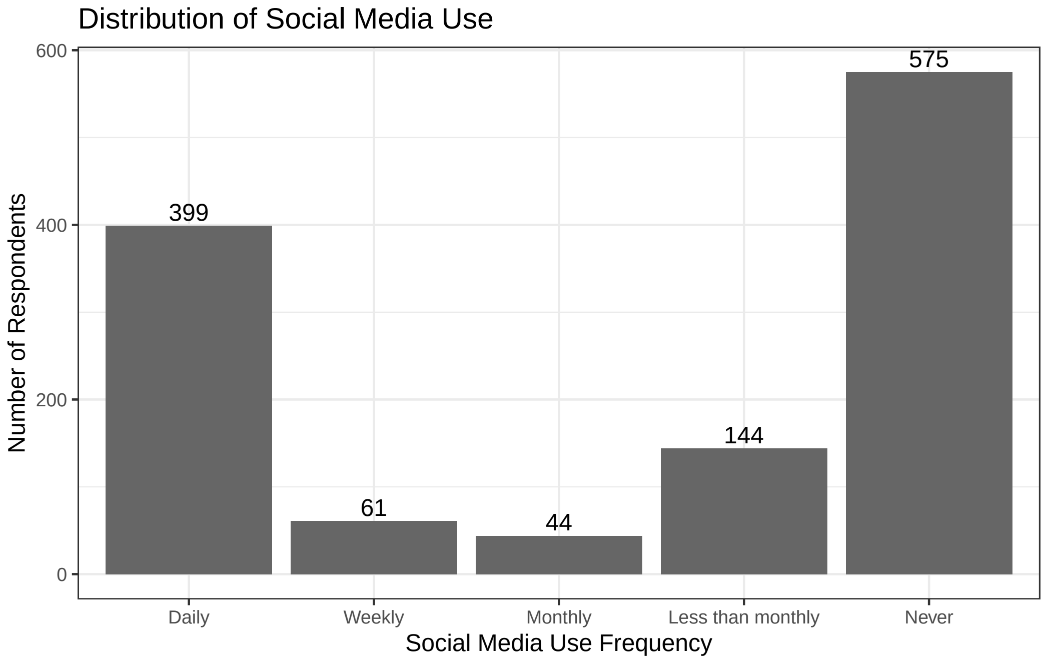
The internet usage distribution reveals a striking contrast: while 300 respondents reported daily use of the internet for news, an even larger group (733) reported never using it. This suggests a significant digital divide within the sample. Although internet access is widespread in Taiwan, its use for political information may still vary by age, education, or interest level.

Social media use was the most common response (399), followed by a notable number of respondents (575) who never used social media for political information. This split suggests that while social media is emerging as a key source of political engagement for some, a large portion of the population still abstains—possibly due to trust, preference, or generational differences.




Distribution of Newspaper Use
Number of Respondents
0
100
200
300
400
500
479
289
62
156
237
Daily
Weekly
Monthly
Less than monthly
Never
Newspaper Use Frequency

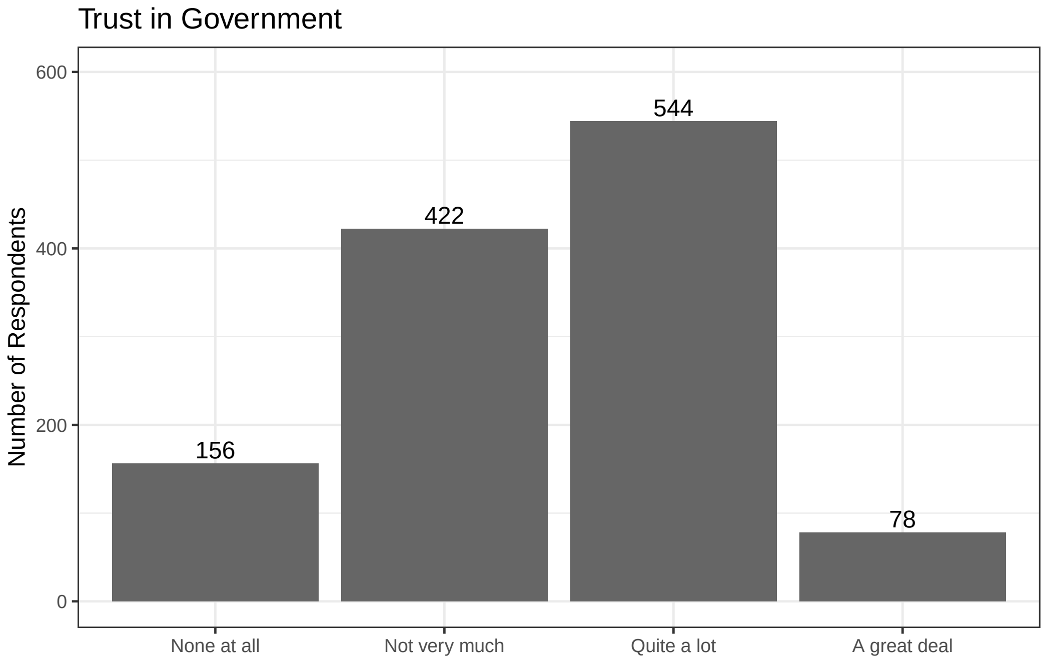

Distribution of Internet Use
Number of Respondents
0
200
400
600
300
46
32
112
733
Daily
Weekly
Monthly
Less than monthly
Never
Internet Use Frequency

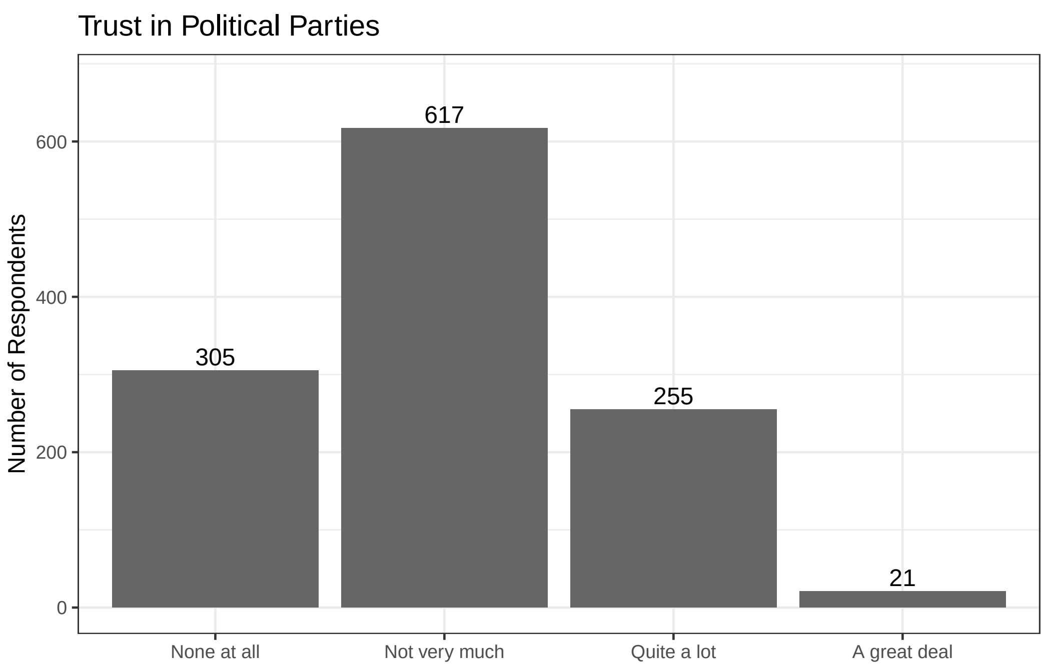


4.1.5 Taiwan-Specific Factors layer (Common in Taiwan Research)

In terms of **political trust**,the majority of respondents reported moderate to high levels of trust in the government, with 544 selecting “Quite a lot” and 422 selecting “Not very much.” Only a small minority expressed extreme trust (“A great deal,” n = 78) or no trust at all (n = 156). This pattern suggests that while full trust is rare, outright distrust is also limited.

For **political parties trust** , respondents generally expressed low levels of trust in political parties. A combined 922 respondents selected “Not very much” (n = 617) or “None at all” (n = 305), while only 276 respondents indicated moderate to high trust (“Quite a lot” or “A great deal”). These results align with the broader trend of institutional skepticism toward parties in Taiwan.

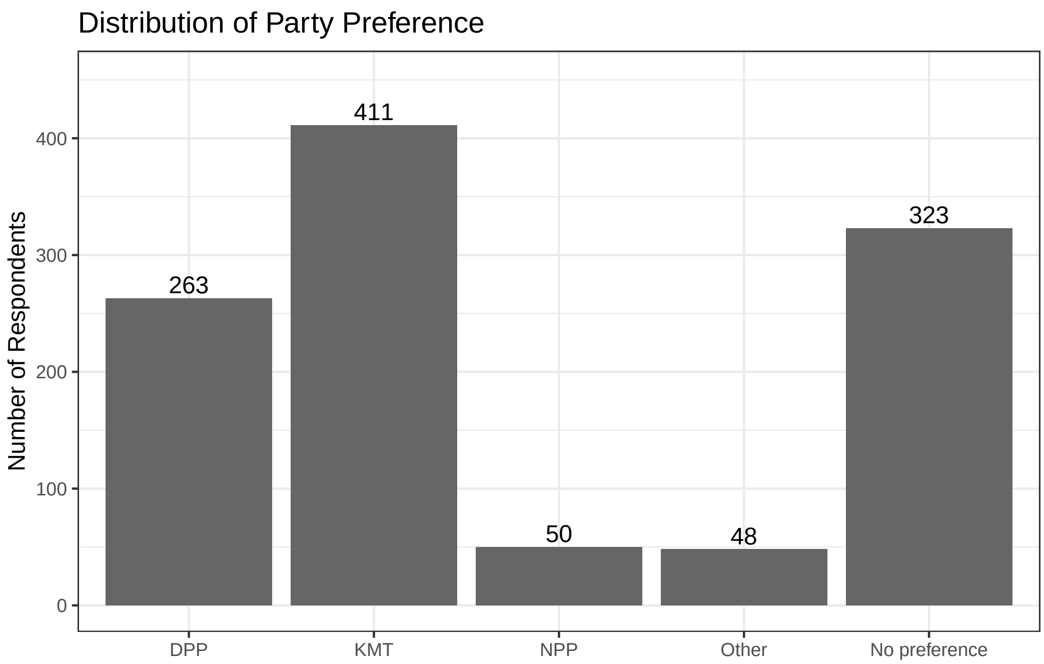


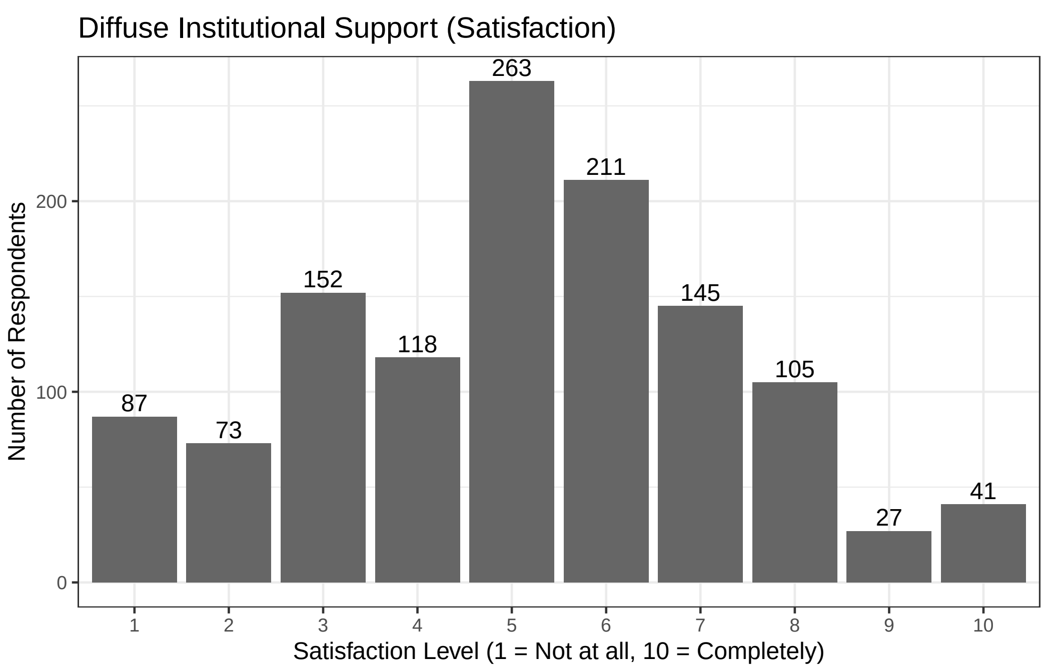


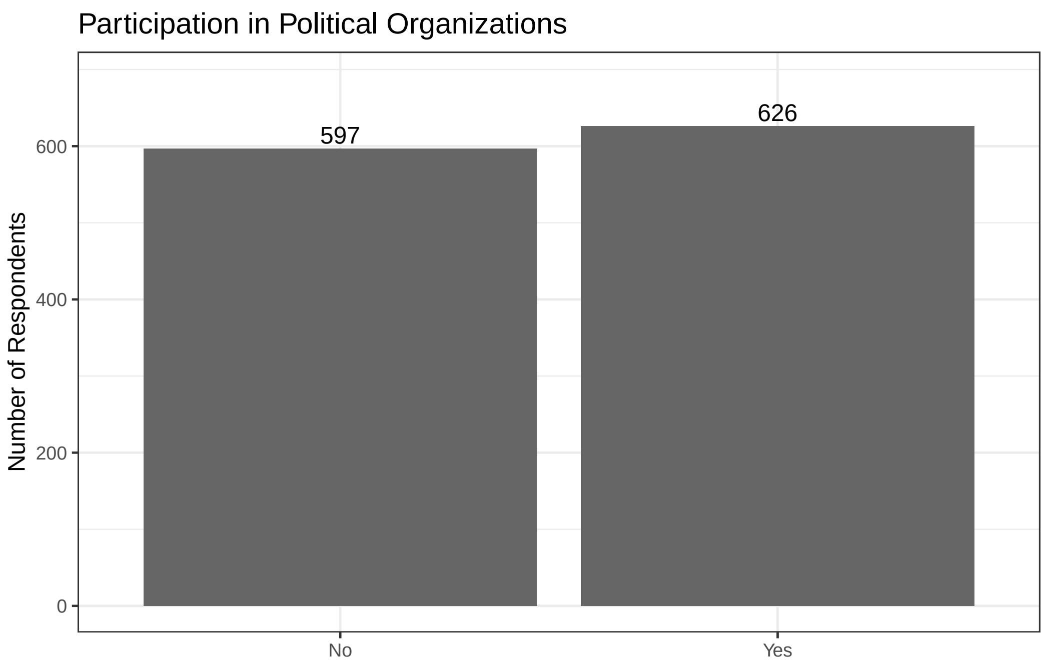
The distribution of **party preference** shows that most respondents identified with either the KMT (411) or the DPP (263), while a substantial number (323) reported having no party preference. Support for smaller parties like the NPP or "Other" was relatively low. This pattern reflects Taiwan's long-standing two-party dominance, but also highlights a notable segment of politically unaffiliated individuals.

Diffuse Institutional Support shows that the Respondents' satisfaction with the political system was generally moderate. Most clustered around the midpoint (5 or 6 on a 10-point scale), with fewer respondents expressing either very low (1–2) or very high (9–10) satisfaction. This suggests a general ambivalence toward institutional performance, possibly reflecting ongoing political polarization or distrust in the system's effectiveness.

Organizational participation in this study captures involvement in a range of formal group activities, including labor unions, political parties, environmental and professional associations, charitable organizations, and women's groups. Based on this composite indicator, a slightly higher number of respondents (626) reported participating in at least one such organization, compared to 597 who did not. This reflects a moderate level of civic engagement through formal institutions in Taiwan, particularly within issue-based or advocacy-oriented groups.






Participation in Political Organizations
Number of Respondents
600
400
200
0
597
626
No
Yes

5 Result

5.1 Model Comparison for variables

To identify the best-fitting model, we constructed a sequence of nested logistic regression models by incrementally adding conceptually grouped blocks of predictors. Model fit was evaluated at each step using Akaike Information Criterion (AIC), pseudo R^2, and likelihood ratio test (LRT) p-values.

The AIC values progressively declined across the first four models, suggesting improved model fit with each added block. The full model—which includes all five conceptual layers—achieved the lowest AIC (1276.487), indicating the best trade-off between fit and parsimony. Although the addition of the Trust & Party block slightly increased the AIC, the full model still outperformed all others overall.

Pseudo R^2 values also increased with each model extension, from 0.024 in the baseline SES-only model to 0.096 in the full model. This steady gain in explained variance suggests that each added predictor block contributed meaningfully to the model's explanatory power.

LRT p-values confirmed these gains, with statistically significant improvements observed at every step except for the Trust & Party block ($p = 0.154$). The final model's improvement over the previous model was significant ($p = 0.002$), further justifying its selection.

Taken together, the AIC, pseudo R^2, and LRT results consistently support the full model as the best-fitting specification.

Table 3: Model Comparison Summary

Model	AIC	LogLik	Pseudo R^2	Deviance Improvement	LRT p-value
SES only	1342.321	-667.160	0.024	NA	NA
+ Demographics	1320.692	-654.346	0.043	25.63	0.000
+ Psychological	1296.296	-640.148	0.064	28.40	0.000
+ Media use	1282.757	-629.379	0.080	21.54	0.000
+ Trust & Party	1285.396	-624.698	0.086	9.36	0.154
+ Org & Support (Full)	**1276.487**	**-618.244**	**0.096**	**12.91**	**0.002**

5.2 Key Predictors of Political Participation

The table below presents the results from the full logistic regression model used to examine predictors of non-institutional political participation.
The **odds ratio (OR)** reflects the multiplicative change in the odds of participation for a one-unit increase in the predictor.An OR greater than 1 indicates a higher likelihood of participation, while an OR less than 1 indicates a lower likelihood. **Confidence intervals** (CIs) that do not include 1 imply statistical significance, and p-values below 0.05 are considered statistically significant.

5.2.1 Socioeconomic Status (SES)

Income Level (numeric coding):
Individuals with higher income levels were more likely to participate in non-institutional political

activities (OR = 1.150, $p < 0.05$).
For instance, moving up three steps in income level corresponds to an approximately **52% increase** in the odds of participation.

5.2.2 Demographic Variables

Gender (reference: Female):
Male respondents were significantly less likely to engage in political participation compared to females (OR = 0.533, $p < 0.001$),
meaning the odds of participation for males were **roughly half** that of females.

Age (numeric coding):
Older individuals were more likely to participate (OR = 1.023, $p < 0.001$).
A 10-year increase in age was associated with about a **26% increase** in the odds of participating.

5.2.3 Psychological Engagement

Political Interest (numeric coding):
Greater political interest was associated with higher odds of participation (OR = 1.285, $p = 0.011$).
A two-level increase in interest (e.g., from "Not very interested" to "Very interested") raised the odds of participation by about **65%**.

Discussion Frequency (numeric coding):
Respondents who discussed politics less frequently were less likely to participate (OR = 0.712, $p = 0.013$).
This suggests that infrequent political discussants had about **29% lower odds** of participation compared to frequent discussants.

5.2.4 Media Use

Social Media Use (numeric coding):
Increased use of social media was positively associated with participation (OR = 1.129, $p = 0.032$).
For example, shifting two levels upward in frequency of use (e.g., from "Monthly" to "Daily") corresponded to roughly a **27% increase** in participation odds.

5.2.5 Taiwan-Specific Variables

Organizational Participation:
Respondents who were involved in at least one organization were significantly more likely to participate (OR = 1.644, $p < 0.001$).
This reflects a **64% increase** in the odds of political participation compared to those not involved in any organizations.

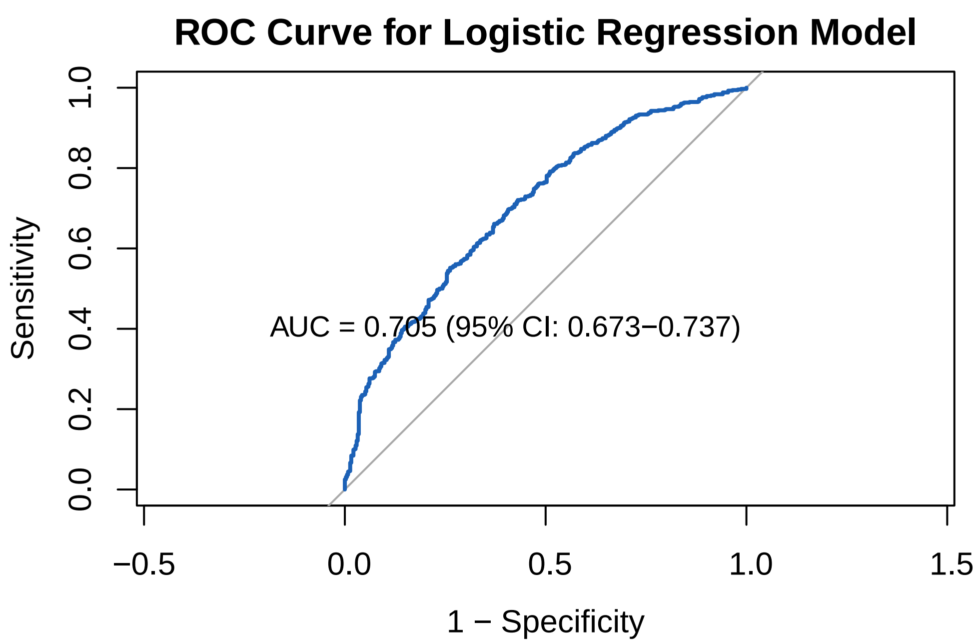
Other predictors—including **education level**, **trust in political institutions**, and **traditional media use**—did **not** reach statistical significance in the final model.

Table 4: Coefficients and Odds Ratios from the Final Logistic Regression Model

Predictor	Coefficient (log-odds)	Odds Ratio	95% CI Lower	95% CI Upper	p-value
Education	0.103	1.109	0.978	1.257	0.107
Social Class	-0.048	0.953	0.794	1.142	0.6
Income Level	0.139	1.150	1.044	1.268	0.005**
Sex (Male)	-0.629	0.533	0.402	0.704	0***
Age	0.023	1.023	1.011	1.035	0***
Political Interest	0.250	1.285	1.063	1.554	0.01**
Discussion Frequency	-0.340	0.712	0.543	0.931	0.013*
Newspaper	-0.063	0.939	0.856	1.029	0.177
TV News	0.040	1.041	0.915	1.183	0.535
Internet	0.112	1.118	0.989	1.265	0.075
Social Media	0.121	1.129	1.010	1.261	0.032*
Trust: Government	-0.121	0.886	0.706	1.110	0.292
Trust: Political Parties	-0.173	0.841	0.661	1.070	0.159
partyKMT	-0.159	0.853	0.585	1.242	0.409
Party: NPP	0.541	1.717	0.839	3.688	0.15
Party: Other	-0.028	0.973	0.485	1.982	0.938
Party: No Preference	-0.096	0.908	0.610	1.351	0.636
Diffuse Support	0.022	1.022	0.948	1.101	0.569
Org Participation	0.497	1.644	1.251	2.163	0***

5.3 Model Predictive Performance

The final logistic regression model demonstrated moderate but limited predictive performance, with an area under the ROC curve (AUC) of 0.705 . This indicates that the model is able to distinguish between political participants and non-participants approximately 71% of the time.



5.4 Model Diagnostics

To evaluate the robustness and validity of the final logistic regression model, several diagnostic checks were performed.

5.4.1 Multicollinearity

Variance Inflation Factors were calculated for all predictors. All GVIF values, as well as their adjusted forms were well below the commonly used threshold of 5. The highest adjusted GVIF observed was approximately 1.59 for *internet_use*, suggesting no evidence of problematic multicollinearity.

5.4.2 Influential Observations

Cook's Distance plots indicated that a small number of observations (e.g., 95, 580, and 915) had relatively high influence, but none exceeded the conventional threshold of 1. Similarly, leverage vs. standardized residuals plots did not reveal any severe outliers, although a few observations exhibited slightly higher leverage.
DFBeta plots for key predictors such as *education*, *income_level*, and *sex* also showed that individual cases had minimal influence on coefficient estimates, indicating overall model stability.

5.4.3 Model Fit and Predictive Accuracy

Pseudo R^2 (McFadden): 0.096, with additional R^2 measures (ML: 0.117, CU: 0.161) suggesting modest explanatory power.

5.4.4 Hosmer–Lemeshow Test

Hosmer–Lemeshow Test yielded a significant result ($p < 0.001$), indicating some lack of fit; however, this test is known to be sensitive to large sample sizes and may overstate lack-of-fit.

5.4.5 Residual Diagnostics (DHARMa)

DHARMa simulation-based residual diagnostics indicated no significant issues:
The Kolmogorov–Smirnov (KS) test and dispersion test both yielded non-significant p-values.
Residual vs. predicted plots showed uniform scatter, with no signs of overdispersion, outliers, or prediction bias.

5.4.6 Linearity Assumption (Box–Tidwell Test)

To assess whether the logit of the outcome was linearly related to continuous predictors, interaction terms with log-transformed variables were tested. All interaction terms—including those for *education*, *income_level*, *age*, and media use variables—were non-significant ($p > 0.05$), supporting the assumption of linearity in the logit for these predictors.

Note: All diagnostic plots (e.g., Cook's Distance, DFBeta, leverage plots, residual checks) are provided in Appendix A for reference.

6 Conclusion

This study finds that income level, social media use, and organizational participation are the most influential predictors of non-institutional political participation in Taiwan. In contrast, traditional factors such as education and institutional trust showed little explanatory power. These results indicate a shift toward more personalized, digitally-driven, and network-based forms of civic engagement, particularly among younger or more connected citizens. This highlights the growing importance of civic networks over conventional political structures in shaping participatory behavior. Future research could further explore generational differences or test interaction effects to better understand the complexity of political participation in evolving digital societies.

7 Discussion

This study integrates theoretical frameworks, empirical findings, and statistical modeling to draw valid inferences about non-institutional political participation in Taiwan in 2019. The selection of variables and model layers was informed by theories of socioeconomic status (SES), psychological engagement, and commonly used predictors in Taiwan-based political studies, in order to minimize selection bias. A sequence of logistic regression models was constructed to identify the best-fitting model, determine significant predictors, and evaluate model assumptions. We also used the tidymodels framework to assess generalization performance and visualized variable contributions via a variable importance plot (VIP).

In relation to the Civic Voluntarism Model (CVM), income level emerged as a consistent and significant predictor, reinforcing the idea that economic resources matter for civic engagement. However, education level and subjective social class showed no significant association, which diverges from CVM's classical expectations. This suggests that in Taiwan's recent context, income might serve as a more meaningful indicator of participatory capacity than formal education or class identification.

Among demographic variables, women were more likely to participate than men, and participation likelihood increased with age. The age effect, while modest (OR = 1.023), indicates a steady rise in engagement across age groups. These patterns may reflect both generational shifts and broader social changes in civic awareness.

In line with Dalton's theory of political engagement, both political interest and political discussion frequency significantly predicted participation. These findings reinforce the role of psychological readiness in enabling citizens to act, especially through non-traditional channels.

Regarding media use, only social media emerged as a significant predictor, while traditional outlets like TV and newspapers did not. This highlights the early but growing role of digital platforms in mobilizing political action in Taiwan, especially among younger and more digitally connected populations.

Lastly, among Taiwan-specific predictors, party preference and political trust showed no statistical association with non-institutional participation. Only organizational involvement stood out as a strong predictor. This implies that while institutional trust may influence formal political behaviors, collective networks—such as unions or civic groups—are more relevant for explaining who takes part in protests or petitions.

Taken together, the findings suggest a shift toward decentralized, digitally enabled, and network-based forms of civic engagement. Traditional indicators such as education and institutional alignment may no longer fully capture the dynamics of political action in Taiwan's evolving participatory landscape.

8 Limitations & Future Research

This study is subject to several limitations. First, it relies on cross-sectional survey data, which limits the ability to make causal inferences about the relationships between predictors and political participation. The use of self-reported measures also raises the possibility of social desirability bias. Second, although the model incorporates multiple theoretically informed layers, it does not explore potential interaction effects or nonlinear patterns that may better reflect the complexity of participatory behavior.

For future research, longitudinal data would enable the tracking of changes in political engagement over time and offer stronger evidence for causal relationships. Expanding the model to include external factors—such as the media environment, civic networks, or political opportunity structures—could further enrich the analysis. Comparative studies across East Asian contexts may also help clarify how institutional and cultural differences influence political participation. Lastly, more in-depth exploration of generational dynamics could shed light on evolving forms of activism, particularly among younger fellows.

9 References

Bennett, W. L., & Segerberg, A. (2013). *The logic of connective action: Digital media and the personalization of contentious politics.* Cambridge University Press.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Verba, S., Schlozman, K. L., & Brady, H. E. (1995). *Voice and equality: Civic voluntarism in American politics.* Harvard University Press.

Brady, H. E., Verba, S., & Schlozman, K. L. (1995). Beyond SES: A resource model of political participation. *American Political Science Review, 89*(2), 271–294. https://doi.org/10.2307/2082427

Chang, Y. T., & Chu, Y. H. (2006). Corruption and trust: Exceptionalism in Asian democracies? *The Journal of Politics, 68*(1), 259–271. https://doi.org/10.1111/j.1468-2508.2006.00404.x

Dalton, R. J. (2008). Citizenship norms and the expansion of political participation. *Political Studies, 56*(1), 76–98. https://doi.org/10.1111/j.1467-9248.2007.00718.x

10 Appendix- feature engineering

10.0.1 Uninstitutional Political Participation(Response)

Table 5: Variables used to construct the political participation index.

Question	Description
Q209	Signing a petition
Q210	Joining in boycotts
Q211	Attending peaceful demonstrations
Q212	Joining unofficial strikes
Q213	Donating to a political campaign
Q214	Contacting a government official
Q215	Encouraging others to take action about political issues
Q216	Encouraging others to vote
Q218	Signing an electronic petition
Q219	Encouraging others to take any form of political action
Q220	Organizing political activities, events, or protests

10.0.2 Organization index variable

Table 6: Selected Items for Organizational Participation Variable

Question Code	Organization Type
Q97	Labor Union
Q98	Political Party
Q99	Environmental Organization
Q100	Professional Association
Q101	Humanitarian or Charitable Organization
Q104	Women's Group

11 Appendix-Diagnostic plots & summary

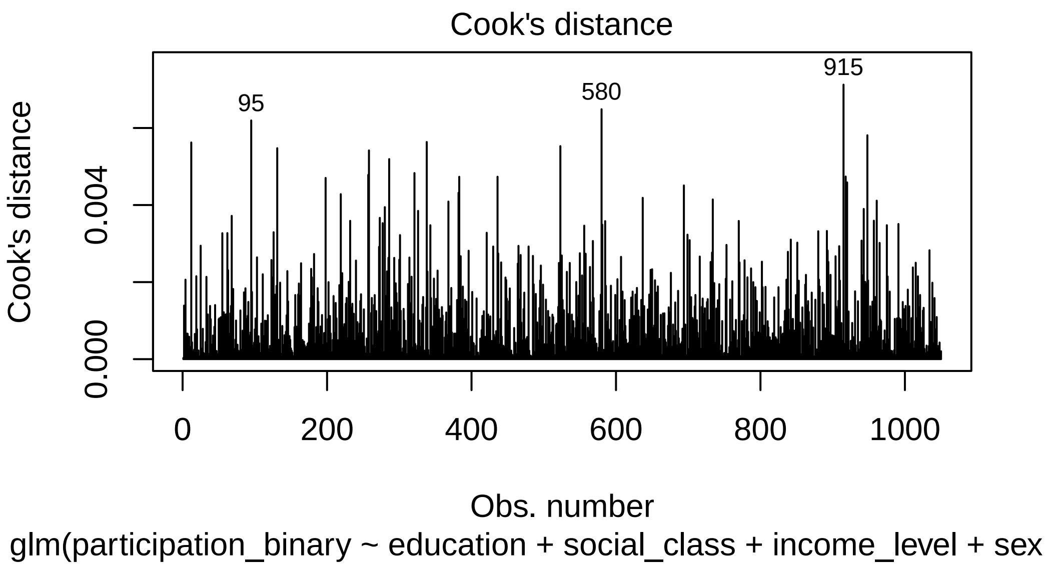
11.1 Diagnostic Plot

11.1.1 VIF

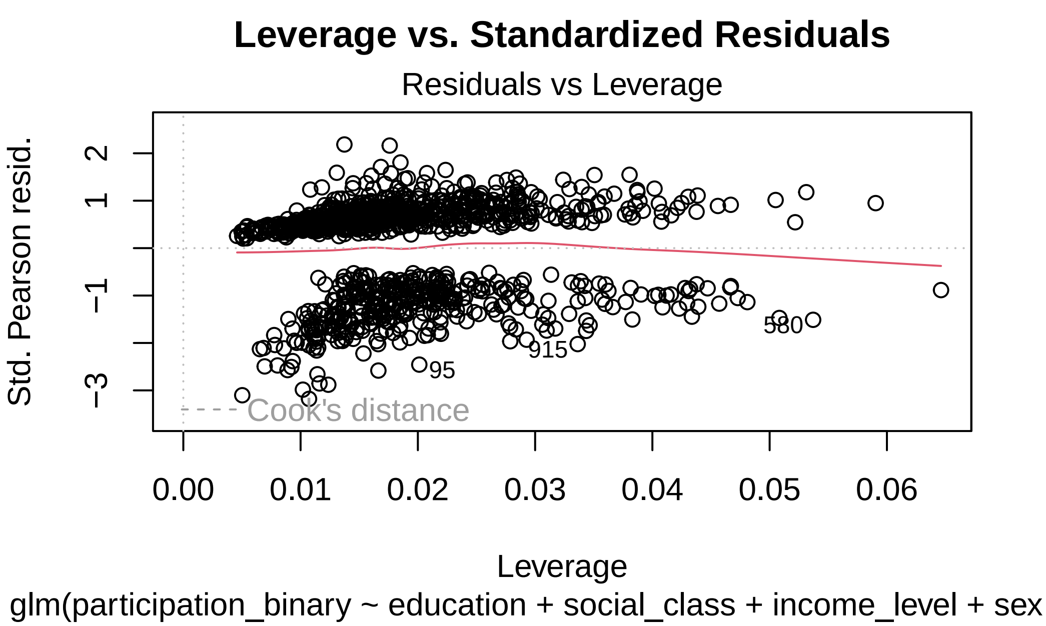
Table 7: Variance Inflation Factors (VIF) for the Final Model

	GVIF	Df	GVIF^(1/(2*Df))
education	2.02	1	1.42
social_class	1.48	1	1.21
income_level	1.53	1	1.24
sex	1.08	1	1.04
age	2.04	1	1.43
political_interest	1.39	1	1.18
discussion_frequency	1.28	1	1.13
newspaper_use	1.16	1	1.08
tvnews_use	1.12	1	1.06
internet_use	2.54	1	1.59
socialmedia_use	2.23	1	1.49
trust_government	1.79	1	1.34
trust_parties	1.72	1	1.31
party	1.38	4	1.04
diffuse_support	1.43	1	1.19
org_participation_binary	1.03	1	1.02

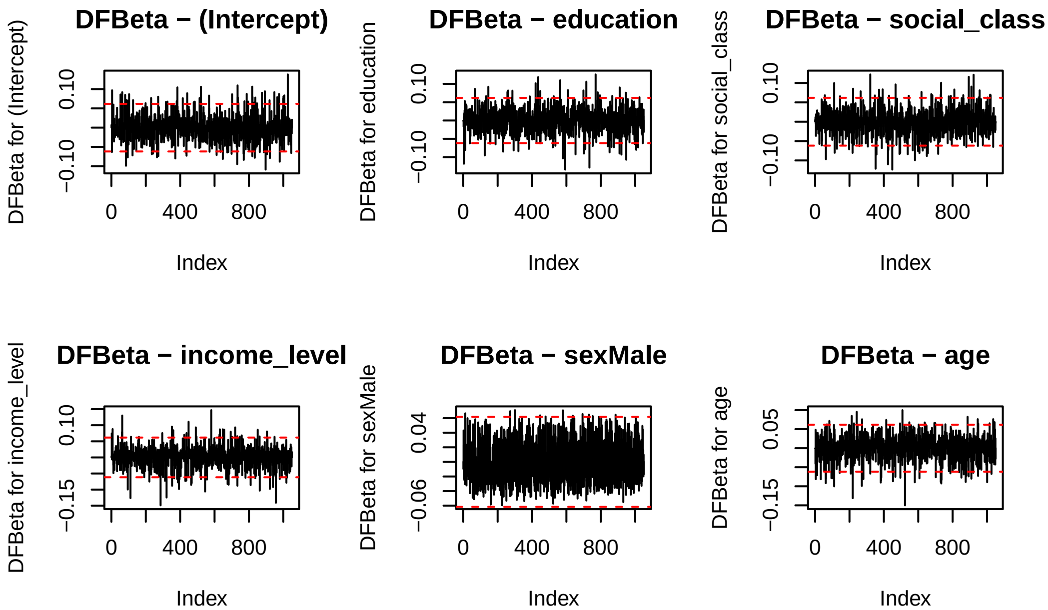
11.1.2 cook distance



11.1.3 leverage



11.1.4 dfbeta



11.1.5 Hosmer-Lemeshow Test

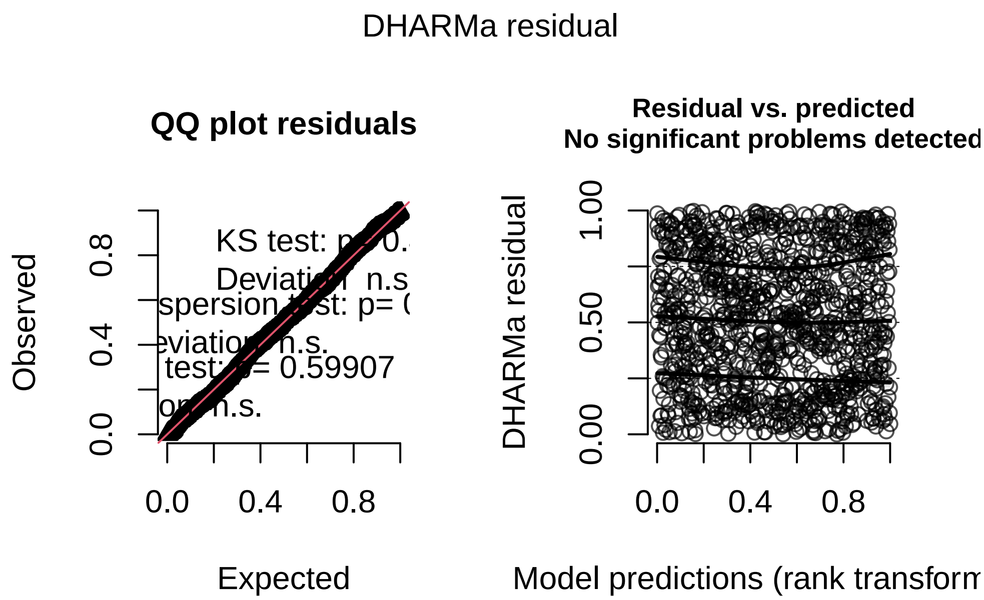
```
fitting null model for pseudo-r2

         llh       llhNull            G2      McFadden          r2ML          r2CU 
-618.2435977 -683.7545277   131.0218599     0.0958106     0.1173113     0.1611157 


	Hosmer and Lemeshow goodness of fit (GOF) test

data:  df_model_df$participation_binary, fitted(model_full)
X-squared = 1050, df = 8, p-value < 2.2e-16
```

11.1.6 DHARMa Residual Diagnostics



11.1.7 Box-Tidwell

```
Call:
glm(formula = participation_binary ~ education + income_level +
    age + newspaper_use + tvnews_use + internet_use + socialmedia_use +
    diffuse_support + I(education * log_education) + I(income_level *
    log_income) + I(age * log_age) + I(newspaper_use * log_newspaper) +
    I(tvnews_use * log_tvnews) + I(internet_use * log_internet) +
    I(socialmedia_use * log_socialmedia) + I(diffuse_support *
    log_diffuse), family = binomial, data = df_box)

Coefficients:
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                          -2.910939   1.827783  -1.593   0.1112
education                            -0.988652   0.578678  -1.708   0.0875 .
income_level                          0.142684   0.328828   0.434   0.6643
age                                   0.131710   0.112677   1.169   0.2424
newspaper_use                        -0.013246   0.483604  -0.027   0.9781
tvnews_use                            1.628165   0.746432   2.181   0.0292 *
internet_use                          0.169069   0.708408   0.239   0.8114
socialmedia_use                      -0.469652   0.627030  -0.749   0.4539
diffuse_support                      -0.304732   0.276793  -1.101   0.2709
I(education * log_education)          0.451313   0.233847   1.930   0.0536 .
I(income_level * log_income)         -0.005061   0.136639  -0.037   0.9705
```

```
I(age * log_age)                    -0.021794   0.023343  -0.934   0.3505
I(newspaper_use * log_newspaper)    -0.010247   0.241690  -0.042   0.9662
I(tvnews_use * log_tvnews)          -0.745905   0.350266  -2.130   0.0332 *
I(internet_use * log_internet)      -0.018882   0.346681  -0.054   0.9566
I(socialmedia_use * log_socialmedia)  0.312304   0.308309   1.013   0.3111
I(diffuse_support * log_diffuse)     0.113333   0.107798   1.051   0.2931
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1367.5  on 1049  degrees of freedom
Residual deviance: 1282.1  on 1033  degrees of freedom
AIC: 1316.1

Number of Fisher Scoring iterations: 4
```

12 Appendix- Source code

12.1 Data preprocessing

```
#load dataset
df_raw <- read_csv(here("data", "Taiwan.csv")) %>%
  clean_names()

#package
library(here)
library(readr)
library(dplyr)
library(janitor)
library(DataExplorer)
library(skimr)
library(conflicted)
library(ggplot2)
library(knitr)
library(naniar)
library(broom)
library(knitr)
library(kableExtra)
library(stringr)
library(tidyr)
library(tidymodels)
library(broom)
library(pscl)
library(dplyr)
library(kableExtra)
library(purrr)
library(forcats)
library(car)

#### Data preprocessing
#Define the 11 variables used for political participation index
participation_vars <- c("q209", "q210", "q211", "q212", "q213",
                        "q214", "q215", "q216", "q218", "q219", "q220")

#Select organization variables
org_vars_selected <- c("q97", "q98", "q99", "q100", "q101", "q104")






# Feature engineering and preprocessing to final dataframe
df_final <- df_raw %>%
```

```r
# 1. Create Political Participation index
mutate(across(all_of(participation_vars), ~ case_when(
  . == "Have done" ~ 1,
  . %in% c("Would never do", "Might do") ~ 0,
  TRUE ~ NA_real_
), .names = "{.col}_bin"),
participation_index = rowSums(across(ends_with("_bin")), na.rm = TRUE),
participation_binary = factor(if_else(participation_index > 0, 1, 0),
                              levels = c(0, 1),
                              labels = c("Not participate", "Participate")),
) %>%

# 2. SES variables
mutate(
  q275 = na_if(trimws(q275), "No answer"),
  q287 = case_when(q287 %in% c("No answer", "Don't know") ~ NA_character_,
                   TRUE ~ trimws(q287)),
  q288 = if_else(trimws(q288) == "Eight  step", "Eight step", trimws(q288)),
  q288 = na_if(q288, "No answer"),

  q275_fct = factor(q275, levels = c(
    "Early childhood education (ISCED 0) / no education",
    "Primary education (ISCED 1)",
    "Lower secondary education (ISCED 2)",
    "Upper secondary education (ISCED 3)",
    "Short-cycle tertiary education (ISCED 5)",
    "Bachelor or equivalent (ISCED 6)",
    "Master or equivalent (ISCED 7)",
    "Doctoral or equivalent (ISCED 8)"
  ), ordered = TRUE),
  q275_num = as.numeric(q275_fct),

  q287_fct = factor(q287, levels = c(
    "Lower class", "Working class", "Lower middle class",
    "Upper middle class", "Upper class"
  ), ordered = TRUE),
  q287_num = as.numeric(q287_fct),

  q288_fct = factor(q288, levels = c(
    "Lower step", "Second step", "Third step", "Fourth step", "Fifth step",
    "Sixth step", "Seventh step", "Eight step", "Nineth step", "Tenth step"
  ), ordered = TRUE),
  q288_num = as.numeric(q288_fct)
) %>%

# 3. Demographic Variables
mutate(
```

```r
    sex_fct = factor(q260),
    sex_num = as.numeric(sex_fct),
    age = as.numeric(q262)
  ) %>%

  # 4. Psychological Variables
  mutate(
    q199_fct = factor(q199, levels = c("Not at all interested",
                                      "Not very interested",
                                      "Somewhat interested",
                                      "Very interested"), ordered = TRUE),
    q199_num = as.numeric(q199_fct),

    q200_fct = factor(q200, levels = c("Frequently",
                                      "Occasionally",
                                      "Never"), ordered = TRUE),
    q200_num = as.numeric(q200_fct)
  ) %>%

  # 5. Media variables
  mutate(
    newspaper_use = as.numeric(factor(q201, levels = c("Never",
                                                      "Less than monthly",
                                                      "Monthly", "Weekly",
                                                      "Daily"),
                                     ordered = TRUE)),
    tvnews_use = as.numeric(factor(q202, levels = c("Never",
                                                   "Less than monthly",
                                                   "Monthly", "Weekly",
                                                   "Daily"), ordered = TRUE)),
    internet_use = as.numeric(factor(q206, levels = c("Never",
                                                     "Less than monthly",
                                                     "Monthly", "Weekly",
                                                     "Daily"),
                                    ordered = TRUE)),
    socialmedia_use = as.numeric(factor(q207, levels = c("Never",
                                                        "Less than monthly",
                                                        "Monthly", "Weekly",
                                                        "Daily"),
                                       ordered = TRUE))
  ) %>%

  # 6. Taiwan research variables
  mutate(
    party = case_when(
      q223 %in% c("TWN: Nationalist Party", "TWN: The pan-Blue coalition",
                  "TWN: People First Party",
```

```r
                  "TWN: Chinese / New Party", "TWN: Minkuotang") ~ "KMT",
    q223 %in% c("TWN: Democratic Progressive Party",
                "TWN: The pan-Green coalition") ~ "DPP",
    q223 == "TWN: New Power Party" ~ "NPP",
    q223 %in% c("TWN: Non-partisan & Non-Partisan Solidarity Union",
                "TWN: Green Party Taiwan",
                "TWN: Social Democratic Party", "Other") ~ "Other",
    q223 %in% c("None", "Don't know", "I would not vote", "No answer") ~
      "No preference",
    TRUE ~ NA_character_
  ),
  party = factor(party, levels = c("DPP", "KMT", "NPP", "Other",
                                  "No preference"))
) %>%

# 7. Political Trust
mutate(
  q71 = ifelse(q71 == "Don't know", NA, q71),
  q72 = ifelse(q72 == "Don't know", NA, q72),
  q71_fct = factor(q71, levels = c("None at all",
                                  "Not very much",
                                  "Quite a lot", "A great deal"),
                   ordered = TRUE),
  q72_fct = factor(q72, levels = c("None at all", "Not very much",
                                  "Quite a lot", "A great deal"),
                   ordered = TRUE),
  trust_government = as.numeric(q71_fct),
  trust_parties = as.numeric(q72_fct),
  political_trust_index = rowMeans(across(c(trust_government,
                                            trust_parties)), na.rm = TRUE)
) %>%

# 8. Organizational participation
mutate(across(all_of(org_vars_selected), ~ case_when(
  . %in% c("Active member", "Inactive member") ~ 1,
  . %in% c("Not a member", "Don’t belong") ~ 0,
  TRUE ~ NA_real_
), .names = "{.col}_org"),
org_participation_binary = if_else(rowSums(across(ends_with("_org")),
                                           na.rm = TRUE) > 0, 1, 0)
) %>%

# 9. Diffuse support
mutate(
  q252_clean = case_when(
    q252 == "Not satisfied at all" ~ "1",
    q252 == "Completely satisfied" ~ "10",
```

```r
      q252 == "Don't know" ~ NA_character_,
      TRUE ~ q252
    ),
    diffuse_support = as.numeric(q252_clean)
  )

#-------------------------------------------------------------------------------
# numeric encoding of ordinal and factor variables
df_final_clean <- df_final %>%
  transmute(
    participation_index,
    participation_binary,
    education = q275_num,
    social_class = q287_num,
    income_level = q288_num,
    sex = sex_fct,
    age = age,
    political_interest = q199_num,
    discussion_frequency = q200_num,
    newspaper_use,
    tvnews_use,
    internet_use,
    socialmedia_use,
    party,
    trust_government,
    trust_parties,
    political_trust_index,
    org_participation_binary,
    diffuse_support
  )

#-------------------------------------------------------------------------------
# df for table + factor levels
df_final_clean_fct <- df_final %>%
  transmute(
    participation_index,
    participation_binary,
    education = q275_fct,
    social_class = q287_fct,
    income_level = q288_fct,
    sex = sex_fct,
    age = age,
    political_interest = q199_fct,
    discussion_frequency = q200_fct,
    newspaper_use,
    tvnews_use,
    internet_use,
```

```r
    socialmedia_use,
    party,
    trust_government = q71_fct,
    trust_parties = q72_fct,
    political_trust_index,
    org_participation_binary = factor(org_participation_binary,
                                      levels = c(0,1), labels = c("No", "Yes")),
    diffuse_support = factor(diffuse_support, levels = 1:10)
  )

df_pred_fct <- df_final_clean_fct %>% drop_na()


#for pred
df_pred_clean <- df_pred_fct %>%
  select(-participation_index, -political_trust_index) %>%
  drop_na()

#------------------------------------------------------------------------------
# Define question sets
participation_vars <- c("q209", "q210", "q211", "q212", "q213",
                        "q214", "q215", "q216", "q218", "q219", "q220")

org_vars_selected <- c("q97", "q98", "q99", "q100", "q101", "q104")

# Build clean prediction-ready dataset
df_model_predict <- df_raw %>%

  # 1. Response: Participation Index + Binary
  mutate(across(all_of(participation_vars), ~ case_when(
    . == "Have done" ~ 1,
    . %in% c("Would never do", "Might do") ~ 0,
    TRUE ~ NA_real_
  ), .names = "{.col}_bin")) %>%
  mutate(
    participation_index = rowSums(across(ends_with("_bin")), na.rm = TRUE),
    participation_binary = factor(
      if_else(participation_index > 0, "Participate", "Not participate"),
      levels = c("Not participate", "Participate")
    )
  ) %>%

  # 2. SES
  mutate(
    q275 = na_if(trimws(q275), "No answer"),
    q287 = na_if(trimws(q287), "Don't know"),
    q288 = str_replace(trimws(q288), "Eight  step", "Eight step"),
```

```
  q288 = na_if(q288, "No answer"),

  education = factor(q275, levels = c(
    "Early childhood education (ISCED 0) / no education",
    "Primary education (ISCED 1)",
    "Lower secondary education (ISCED 2)",
    "Upper secondary education (ISCED 3)",
    "Short-cycle tertiary education (ISCED 5)",
    "Bachelor or equivalent (ISCED 6)",
    "Master or equivalent (ISCED 7)",
    "Doctoral or equivalent (ISCED 8)"
  ), ordered = TRUE),

  social_class = factor(q287, levels = c(
    "Lower class", "Working class", "Lower middle class",
    "Upper middle class", "Upper class"
  ), ordered = TRUE),

  income_level = factor(q288, levels = c(
    "Lower step", "Second step", "Third step", "Fourth step", "Fifth step",
    "Sixth step", "Seventh step", "Eight step", "Nineth step", "Tenth step"
  ), ordered = TRUE)
) %>%

# 3. Demographic + Psychological
mutate(
  sex = factor(q260),
  age = as.numeric(q262),

  political_interest = factor(q199, levels = c(
    "Not at all interested", "Not very interested",
    "Somewhat interested", "Very interested"
  ), ordered = TRUE),

  discussion_frequency = factor(q200, levels = c(
    "Frequently", "Occasionally", "Never"
  ), ordered = TRUE)
) %>%

# 4. Media usage
mutate(
  newspaper_use = as.numeric(factor(q201, levels = c("Never",
                                                     "Less than monthly",
                                                     "Monthly", "Weekly",
                                                     "Daily"),
                                    ordered = TRUE)),
  tvnews_use = as.numeric(factor(q202, levels = c("Never",
```

```
                                                    "Less than monthly",
                                                    "Monthly", "Weekly",
                                                    "Daily"), ordered = TRUE)),
    internet_use = as.numeric(factor(q206, levels = c("Never",
                                                      "Less than monthly",
                                                      "Monthly", "Weekly",
                                                      "Daily"), ordered = TRUE)),
    socialmedia_use = as.numeric(factor(q207, levels = c("Never",
                                                         "Less than monthly",
                                                         "Monthly", "Weekly",
                                                         "Daily"),
                                        ordered = TRUE))
  ) %>%

  # 5. Party
  mutate(
    party = case_when(
      q223 %in% c("TWN: Nationalist Party", "TWN: The pan-Blue coalition",
                  "TWN: People First Party", "TWN: Chinese / New Party",
                  "TWN: Minkuotang") ~ "KMT",
      q223 %in% c("TWN: Democratic Progressive Party",
                  "TWN: The pan-Green coalition") ~ "DPP",
      q223 == "TWN: New Power Party" ~ "NPP",
      q223 %in% c("TWN: Non-partisan & Non-Partisan Solidarity Union",
                  "TWN: Green Party Taiwan", "TWN: Social Democratic Party",
                  "Other") ~ "Other",
      q223 %in% c("None", "Don't know", "I would not vote",
                  "No answer") ~ "No preference",
      TRUE ~ NA_character_
    ),
    party = factor(party, levels = c("DPP", "KMT", "NPP", "Other",
                                     "No preference"))
  ) %>%

  # 6. Trust
  mutate(
    q71 = ifelse(q71 == "Don't know", NA, q71),
    q72 = ifelse(q72 == "Don't know", NA, q72),

    trust_government = factor(q71, levels = c("None at all", "Not very much",
                                               "Quite a lot", "A great deal"),
                              ordered = TRUE),
    trust_parties = factor(q72, levels = c("None at all", "Not very much",
                                            "Quite a lot", "A great deal"),
                           ordered = TRUE)
  ) %>%
```

```r
# 7. Organizational participation
mutate(
  org_participation_binary = if_else(
    rowSums(across(all_of(org_vars_selected), ~ case_when(
      . %in% c("Active member", "Inactive member") ~ 1,
      . %in% c("Not a member", "Don’t belong") ~ 0,
      TRUE ~ NA_real_
    )), na.rm = TRUE) > 0,
    "Yes", "No"
  ),
  org_participation_binary = factor(org_participation_binary,
                                    levels = c("No", "Yes"))
) %>%

# 8. Diffuse support
mutate(
  q252_clean = case_when(
    q252 == "Not satisfied at all" ~ "1",
    q252 == "Completely satisfied" ~ "10",
    q252 == "Don't know" ~ NA_character_,
    TRUE ~ q252
  ),
  diffuse_support = factor(as.integer(q252_clean), levels = 1:10,
                           ordered = TRUE)
) %>%

# 9. Final selection
transmute(
  participation_binary,
  education,
  social_class,
  income_level,
  sex,
  age,
  political_interest,
  discussion_frequency,
  newspaper_use,
  tvnews_use,
  internet_use,
  socialmedia_use,
  party,
  trust_government,
  trust_parties,
  org_participation_binary,
  diffuse_support
)
```

```
df_pred<-drop_na(df_model_predict)
```

12.2 Modeling

```
df_model_df <- df_final_clean %>%
  drop_na()

#ses model
model_ses <- glm(participation_binary ~ education + social_class + income_level,
                     family = binomial,
                     data = df_model_df)




#ses model+demographic
model_ses_dem <- glm(participation_binary ~ education + social_class +
                       income_level+sex+age,family = binomial,
                     data = df_model_df)




#model_ses_dem_phy
model_ses_dem_phy <- glm(participation_binary ~ education + social_class
   + income_level + sex + age + political_interest + discussion_frequency,
                  family = binomial,
                  data = df_model_df)

# model_ses_dem_phy_media
model_ses_dem_phy_media <- glm(participation_binary ~
  education + social_class + income_level +
  sex + age + political_interest + discussion_frequency +
  newspaper_use + tvnews_use + internet_use + socialmedia_use,
  family = binomial,
  data = df_model_df
)

# model_ses_dem_phy_media_trust
model_ses_dem_phy_media_trust <- glm(participation_binary ~
  education + social_class + income_level +
  sex + age + political_interest + discussion_frequency +
  newspaper_use + tvnews_use + internet_use + socialmedia_use +
  trust_government + trust_parties + party,
  family = binomial,
  data = df_model_df
```

```
)

# model full
model_full <- glm(participation_binary ~
  education + social_class + income_level +
  sex + age + political_interest + discussion_frequency +
  newspaper_use + tvnews_use + internet_use + socialmedia_use +
  trust_government + trust_parties + party +
  diffuse_support + org_participation_binary,
  family = binomial,
  data = df_model_df
)
```

12.3 Model results &testing

```
# Model comparison table
model_list <- list(
  model_ses, model_ses_dem, model_ses_dem_phy,
  model_ses_dem_phy_media, model_ses_dem_phy_media_trust, model_full
)

model_table <- tibble(
  Model = c("SES only", "+ Demographics", "+ Psychological",
           "+ Media use", "+ Trust & Party", "+ Org & Support (Full)"),
  AIC = sapply(model_list, AIC),
  LogLik = sapply(model_list, logLik),
  `Pseudo R²` = sapply(model_list, function(m) {
    pscl::pR2(m)["McFadden"] |> as.numeric()
  })
)

# Likelihood ratio tests
lrt_results <- list(
  anova(model_ses, model_ses_dem, test = "Chisq"),
  anova(model_ses_dem, model_ses_dem_phy, test = "Chisq"),
  anova(model_ses_dem_phy, model_ses_dem_phy_media, test = "Chisq"),
  anova(model_ses_dem_phy_media, model_ses_dem_phy_media_trust, test = "Chisq"),
  anova(model_ses_dem_phy_media_trust, model_full, test = "Chisq")
)

# Add test results to table
model_table <- model_table |>
  mutate(
    `Deviance Improvement` = c(NA, sapply(lrt_results, \(x) round(x$Deviance[2], 2))),
    `LRT p-value` = c(NA, sapply(lrt_results, \(x) round(x[["Pr(>Chi)"]][2], 4)))
  )
```