Perceptions of Corruption in Nigeria*

Rural Nigerians and Less Educated Groups Report Higher Corruption

Fatimah Yunusa

November 28, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

We use the statistical programming language R (R Core Team 2023)....

2 Data

2.1 Measurement

For this analysis, i use the Afrobarometer Rouund 9, Nigeria, 2023 survey. Afrobarometer conducts surveys that are nationally representative across African nations to capture public perceptions on topics like governance, democracy, and economic issues. The Nigerian Round 9 surveey dataset contains responses from 1,556 people and is refective of the diverse demographic statistics across gender, region, and residence type(urban or rural). The data is freely available on the Afrobarometer website(Afrobarometer 2023) and was collected through a multi-stage, stratified sampling method. This is to ensure representativeness and enables analysis of different demographics and socioeconomic indicators.

*Code and data are available at:https://github.com/fatimahsy/Nigeria_Democracy-/tree/main

Our primary measurement task is to capture public opininon towards economic, political and socioeconomic issues and translate it into actionable intell. The Afrobarometer survey translates these real world perceptions into quantifiable data using survey instruments.The process begins with identifying relevant phenomena in the Nigerian socio-political and economic context, such as public attitudes toward democracy, demand for governance accountability, and perceived economic conditions. These phenomena are operationalized into survey questions that are clear, culturally sensitive, and tailored to elicit meaningful responses.

For example:

- Perceptions of Democracy: A phenomenon such as the demand for democracy is operationalized by questions asking respondents whether they prefer democracy over other forms of governance, and whether they believe Nigeria is a democratic nation. These questions are grounded in observable public discourse, historical trends, and policy outcomes.
- Economic Prosperity: Attitudes toward economic conditions are measured by asking individuals to rate the current state of the economy and their own living standards. These subjective evaluations stem from tangible experiences like inflation, unemployment, or changes in public service delivery.

Each survey question acts as a data proxy for these complex real-world dynamics, reducing them to quantifiable variables while maintaining their conceptual integrity.

This dataset captures a unique moment in Nigeria's socio-political landscape, reflecting how citizens perceive democracy and governance in the context of ongoing economic and political challenges. Alternative datasets, such as Nigeria's national statistics or World Bank survey data, could provide economic indicators but lack the granularity of Afrobarometer's public opinion data.

2.2 Variables of Interest

2.3 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

Background details and diagnostics are included in Appendix B.

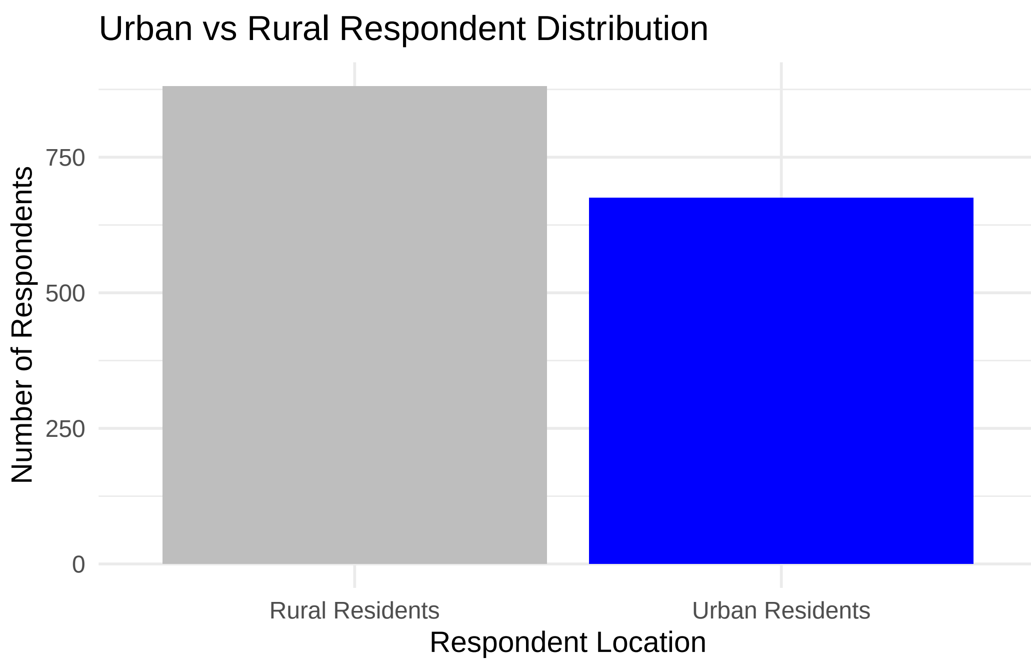

Figure 1: Figure 1: Urban vs Rural Respondent Distribution. This bar chart compares the number of respondents residing in urban versus rural areas.

3 Model

3.1 Model set-up

A Bayesian ordinal logistic regression model is used to estimate the relationship between various demographic factors and perceptions of corruption in the presidency. Ordinal logistic regression is a statistical technique used for ordered categorical outcomes to predict the cumulative probability of being in a particular category or below, given a set of predictor variables.

My model includes four independent demographic variables: `education`, `gender`, `urban_rural`, and `age_group`. The dependent variable is `corruption_presidency`, which has four ordered categories: "None," "Some," "Most," and "All."

The ordinal logistic regression model I used is:

$$\log\left(\frac{P(y_i \leq j)}{P(y_i > j)}\right) = \alpha_j - (\beta_1 \times \text{education} + \beta_2 \times \text{gender} + \beta_3 \times \text{urban rural} + \beta_4 \times \text{age group}) \tag{1}$$

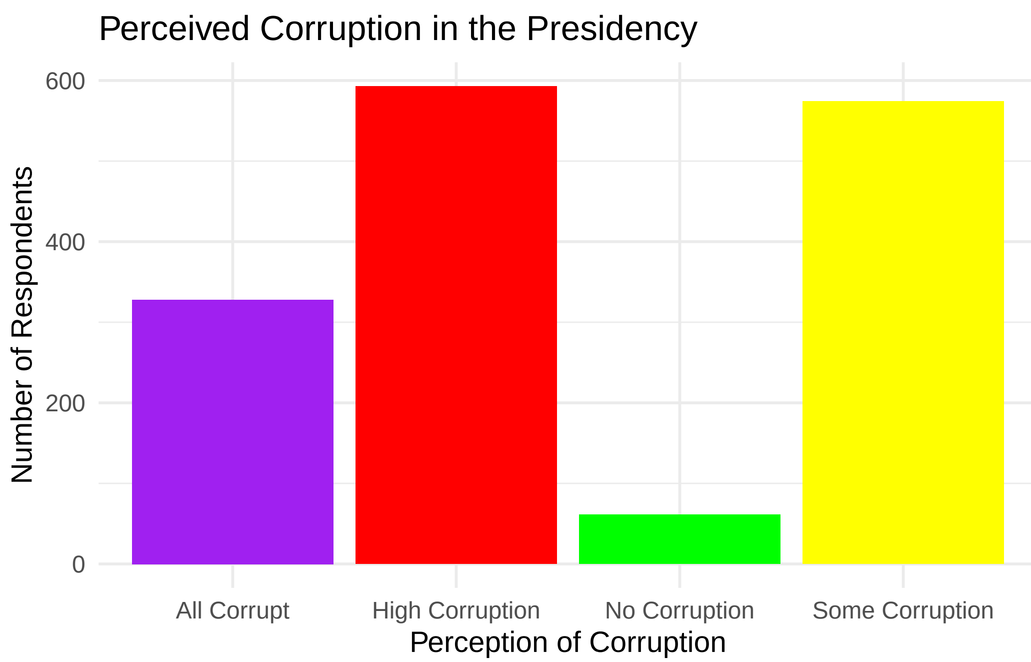

Figure 2: Figure 2: Perceived Corruption in the Presidency. This bar chart shows the distribution of respondents' perceptions of corruption in the presidency.

$$
\begin{aligned}
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\beta_2 &\sim \text{Normal}(0, 2.5) \\
\beta_3 &\sim \text{Normal}(0, 2.5) \\
\beta_4 &\sim \text{Normal}(0, 2.5) \\
\alpha_j &\sim \text{Normal}(0, 2.5) \quad \text{for } j = 1, 2, 3 \\
\pi_{ij} &\sim \text{Dirichlet}(\alpha = 1)
\end{aligned}
$$

where:

- y_i represents the ordinal outcome for respondent i, indicating their perception of corruption in the presidency.
- $P(y_i \leq j)$ is the cumulative probability of respondent i falling into category j or below.
- α_j are the threshold parameters that define the boundaries between the corruption categories.
- $\beta_1, \beta_2, \beta_3, \beta_4$ are the coefficients corresponding to the predictors:
  - β_1: Education level of the respondent.

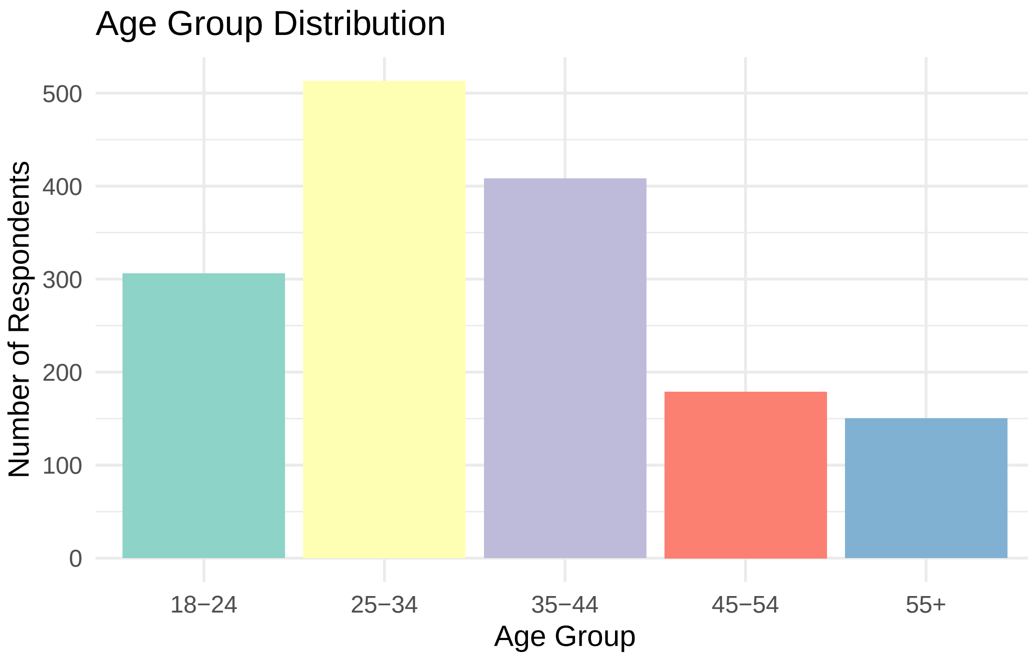

Figure 3: Figure 3: Age Group Distribution. This bar chart displays the number of respondents in each age group.

  - β_2: Gender of the respondent (male or female).
  - β_3: Urban or rural location of the respondent.
  - β_4: Age group of the respondent.

- π_{ij} is the probability that respondent i falls into corruption category j.

In this model, the priors reflect a belief that the predictors explain about 20% of the variability in the outcome, as defined by the $R^2(0.2)$ prior. The Normal priors with a mean of 0 and standard deviation of 2.5 for the coefficients and threshold parameters are relatively weakly informative, ensuring the data have a significant role in determining the posterior distributions.

The threshold parameters (α_j) separate the ordered categories of the outcome variable, while the Dirichlet prior for π_{ij} assumes no strong prior preference for any specific corruption category.

The use of these priors provides flexibility for the model while preventing overfitting. The posterior distributions for the coefficients and thresholds, along with diagnostic checks for model convergence and predictive accuracy, can be found in **?@sec-appendix**.

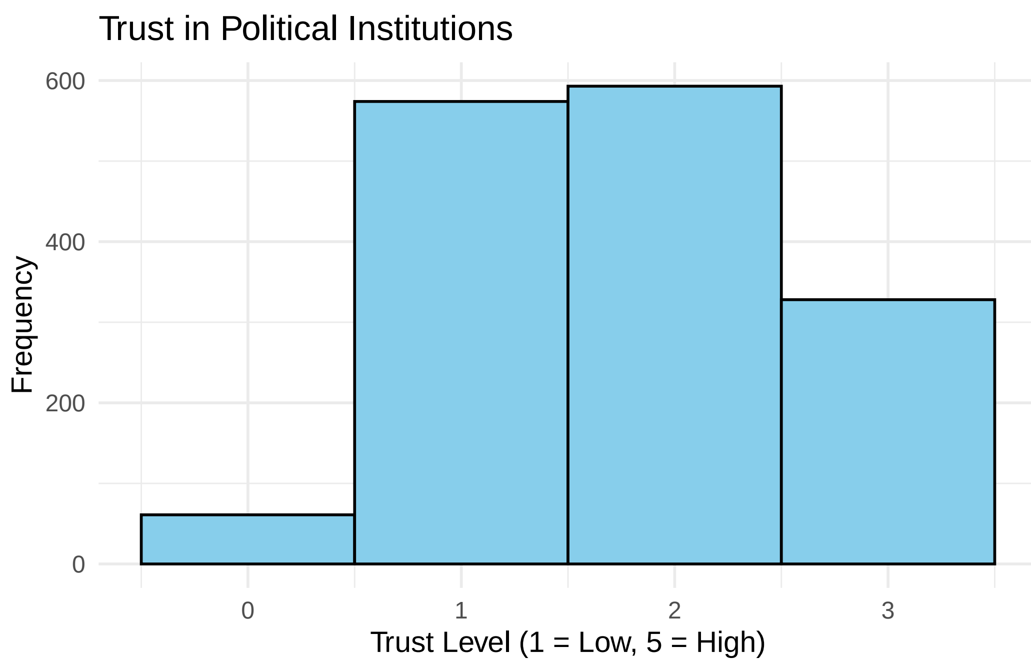

Figure 4: Figure 4: Distribution of Trust in Political Institutions. This histogram shows how respondents rated their trust in political institutions, from low to high trust.

3.1.1 Model Assumptions

The model relies on the following assumptions:

- Proportional Odds: The relationship between the predictors and the log-odds of being in a higher corruption category is constant across thresholds.
- Ordinal Outcomes: The response categories are ordered, but the distances between them are not quantitatively defined.
- Independence of observations: Responses from different individuals are assumed to be independent.

3.2 Model Justification

The Bayesian approach was chosen because of its flexibility in incorporating prior beliefs and its ability to provide real quantification for uncertainty for parameter estimates. An ordinal logistic regression model is appropriate because the outcome variable (perceived corruption) is inherently ordered.

The predictors-education, gender, urban/rural location, and age group- were selected based on their theoretical relevance to perceptions of governmental corruption in Nigeria, as discussed

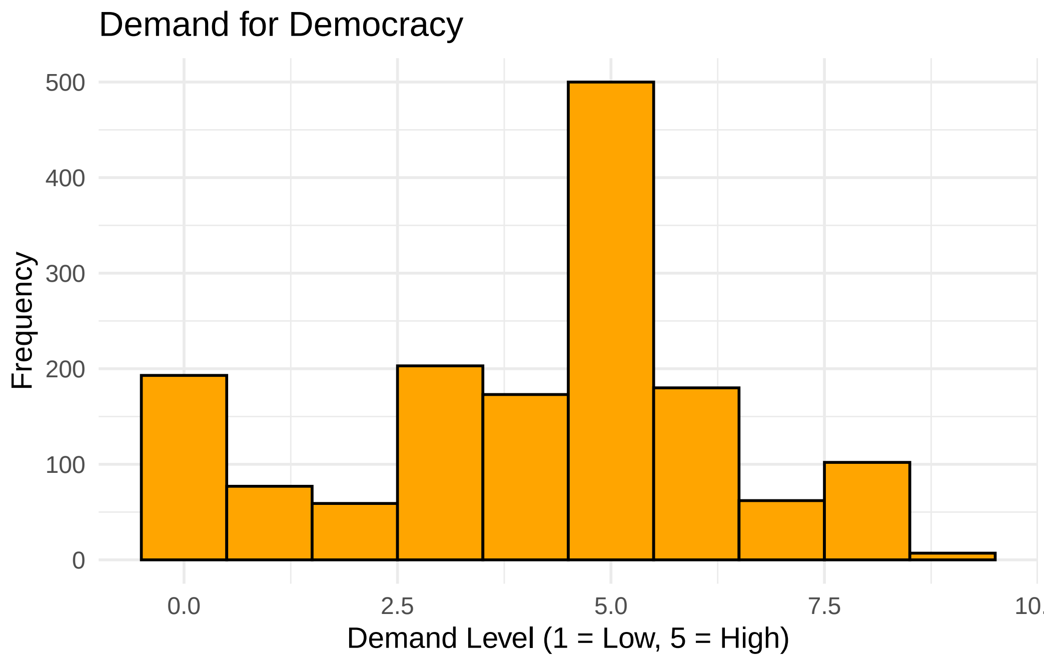

Figure 5: Figure 5: Distribution of Demand for Democracy. This histogram illustrates respondents' levels of demand for democracy, with higher values reflecting stronger democratic preferences.

in the data section. For example, urban and rural individuals may influence exposure to governance systems, while education might shape perceptions of systemic corruption.

The modeling decisions align closely with the structure and characteristics of the dataset to ensure interpretability and relevance. Age was included as a categorical variable reflecting predefined groups, allowing the model to capture non-linear relationships between age and perceptions of corruption. Education was treated as a categorical variable to account for differences across distinct milestones, such as “Primary school completed” and “Post-graduate qualifications.” Gender was modeled as a binary variable (“Male” or “Female”) to reflect societal influences on corruption perceptions. Urban versus rural context was included as a binary variable to capture potential differences in exposure to governance and information access. While the dataset includes regional information, province effects were excluded to maintain focus on demographic factors and simplify the model. Finally, the ordinal logistic regression framework was chosen to respect the ordered nature of the outcome variable (“None,” “Some,” “Most,” “All”) without assuming equal distances between categories. These decisions ensure the model is both consistent with the data and well-suited to addressing the research question.

4 Results

4.1 Corruption Perceptions Model

The analysis revealed distinct trends in Corruption perceptions in different demographics. Our results are summarized in Table 1.

Table 1: Summary of the Bayesian Model

term	estimate	std.error
educationInformal schooling only	-0.01	0.25
educationSome primary schooling	0.43	0.25
educationPrimary school completed	0.31	0.17
educationSome secondary schooling	0.32	0.19
educationSecondary school completed	0.35	0.17
educationPost-secondary qualifications	0.29	0.20
educationSome university	0.16	0.27
educationUniversity completed	0.34	0.23
educationPost-graduate	-0.18	0.74
genderFemale	0.11	0.10
urban_ruralRural	-0.40	0.10
age_group25-34	-0.13	0.13

age_group35-44	0.02	0.14
age_group45-54	-0.12	0.17
age_group55+	0.17	0.19
None|Some	-3.17	0.23
Some|Most	-0.32	0.21
Most|All	1.41	0.21

4.2 Predicted Probabilities

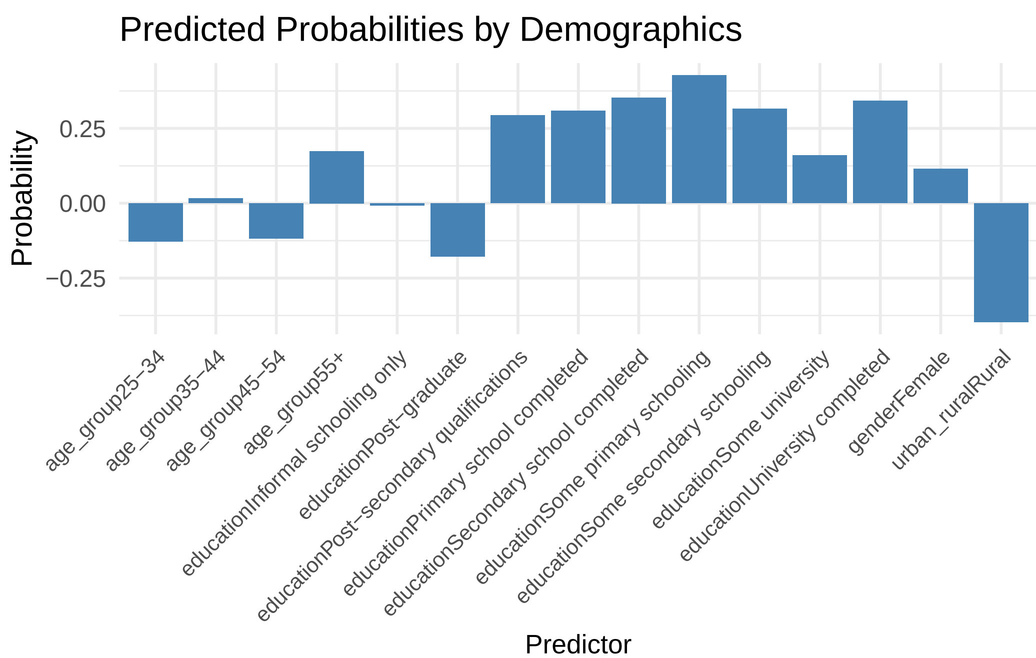

Figure 6: Predicted probabilities of high corruption perceptions by key predictors.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior Predictive Check

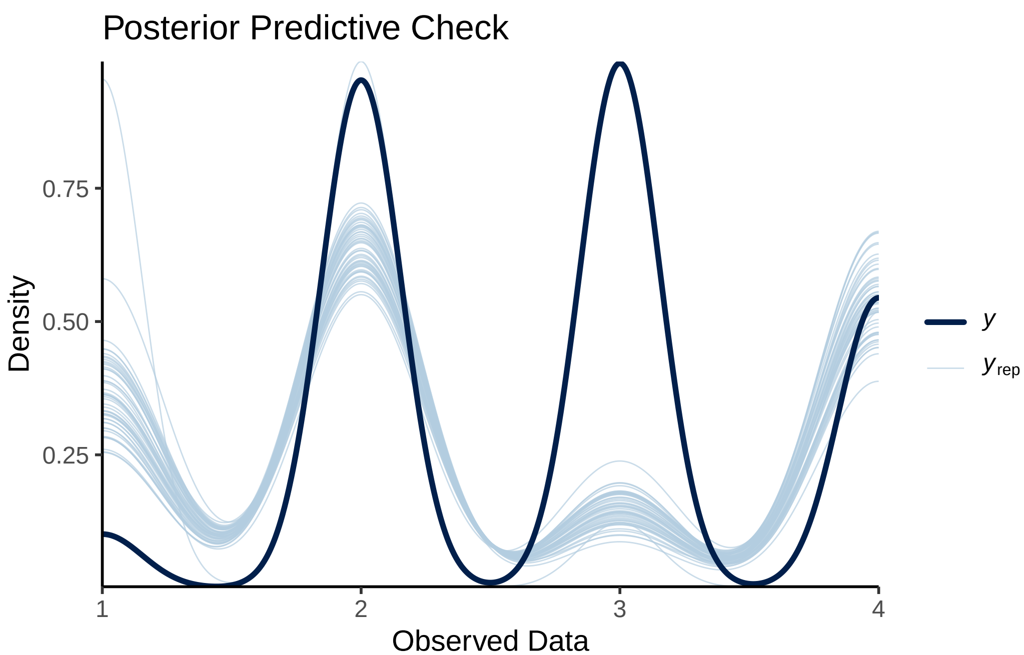

Figure 7: Posterior Predictive Check for the Bayesian ordinal logistic regression model.

B.2 Diagnostics

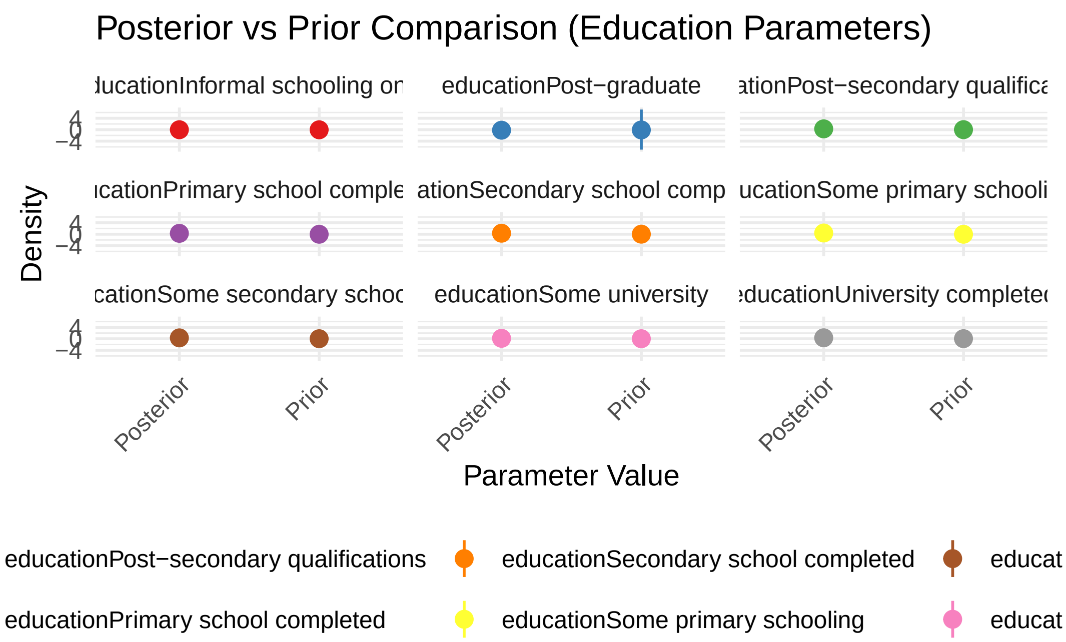

Figure 8: Posterior vs Prior Comparison for Education Parameters.

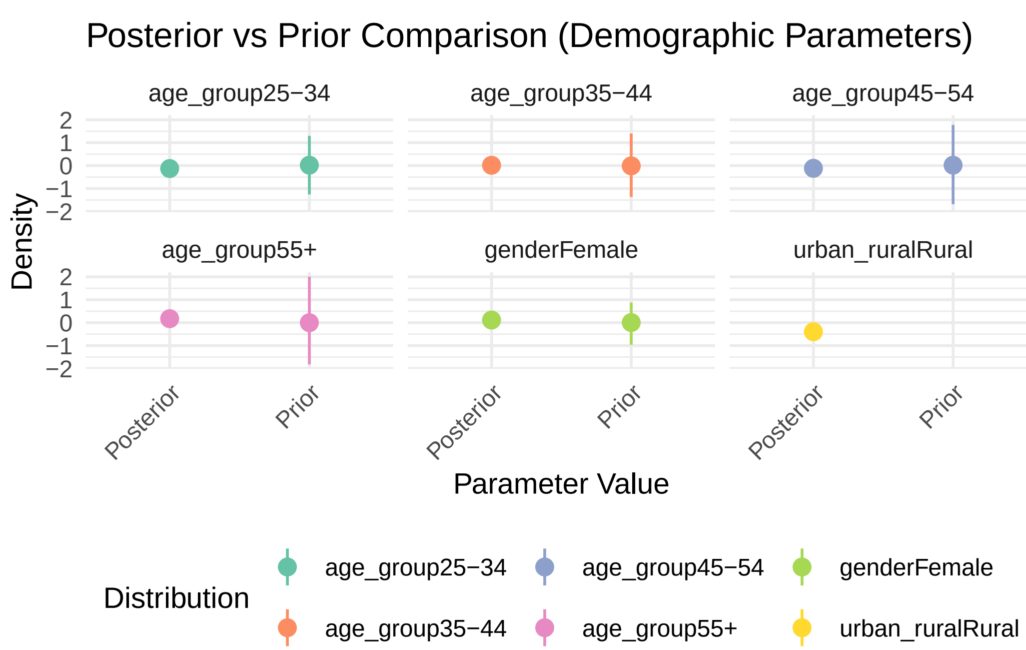

Figure 9: Posterior vs Prior Comparison for Demographic Parameters.

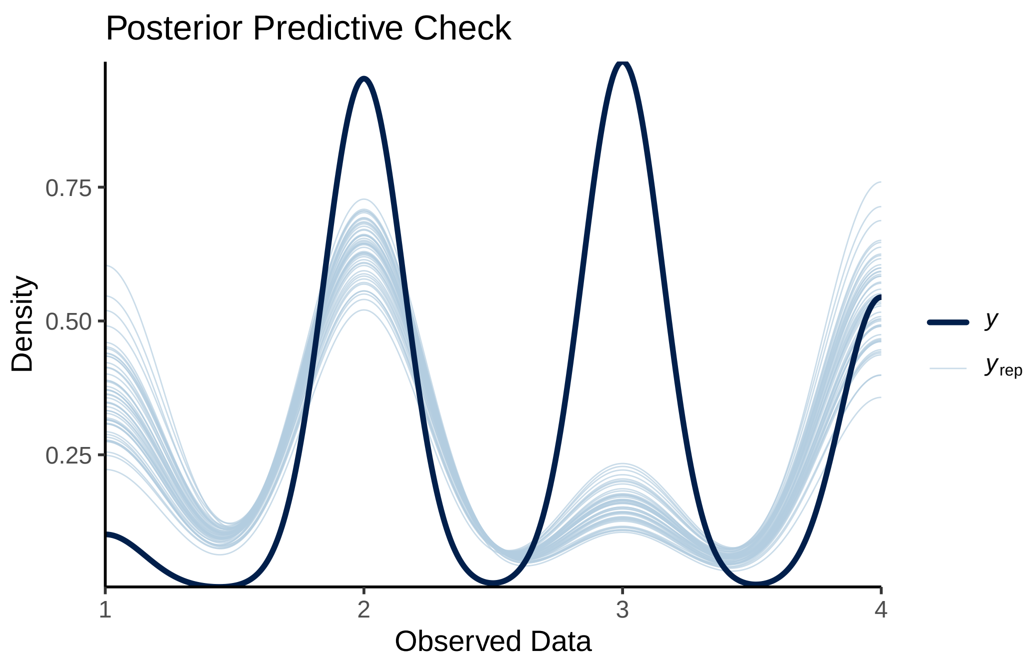

Figure 10: Posterior Predictive Check for the Bayesian ordinal logistic regression model.

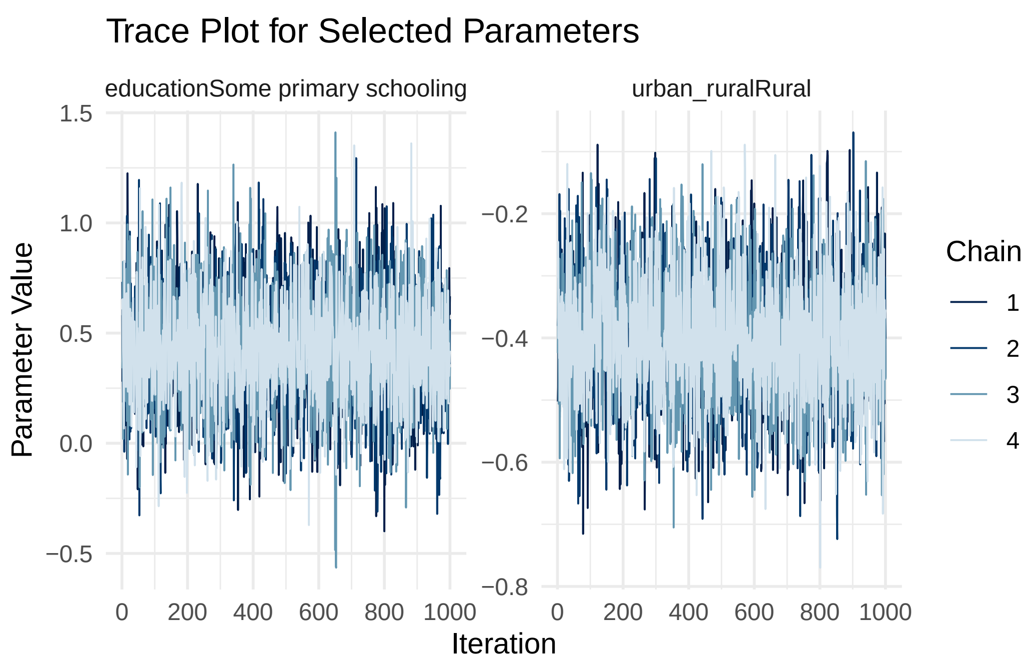

Figure 11: Trace plot for selected parameters in the Bayesian ordinal logistic regression model.

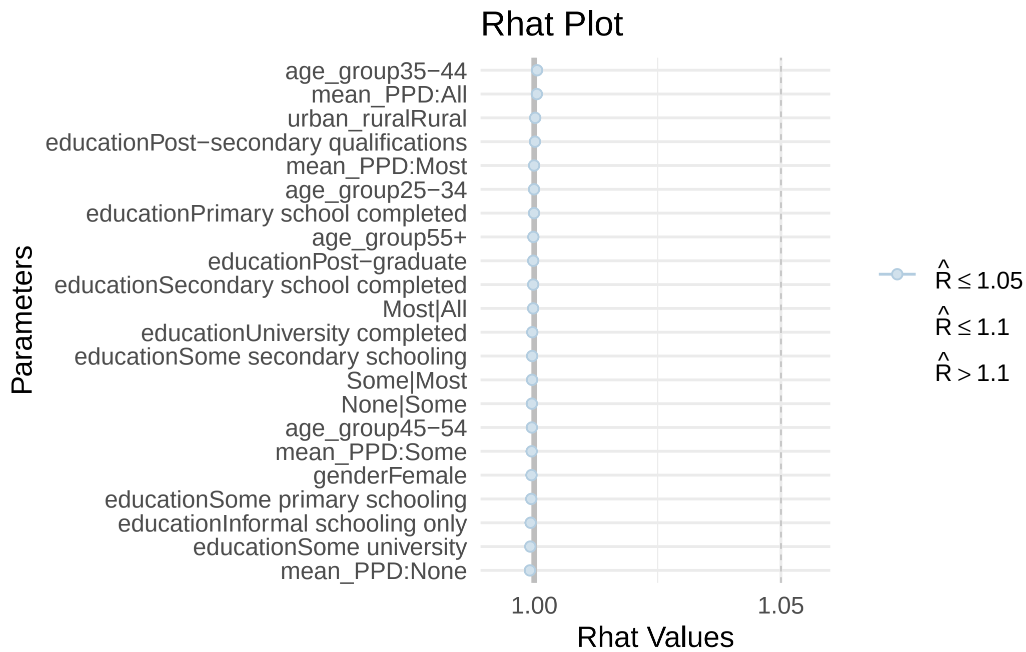

Figure 12: Rhat values for model convergence in the Bayesian ordinal logistic regression model.

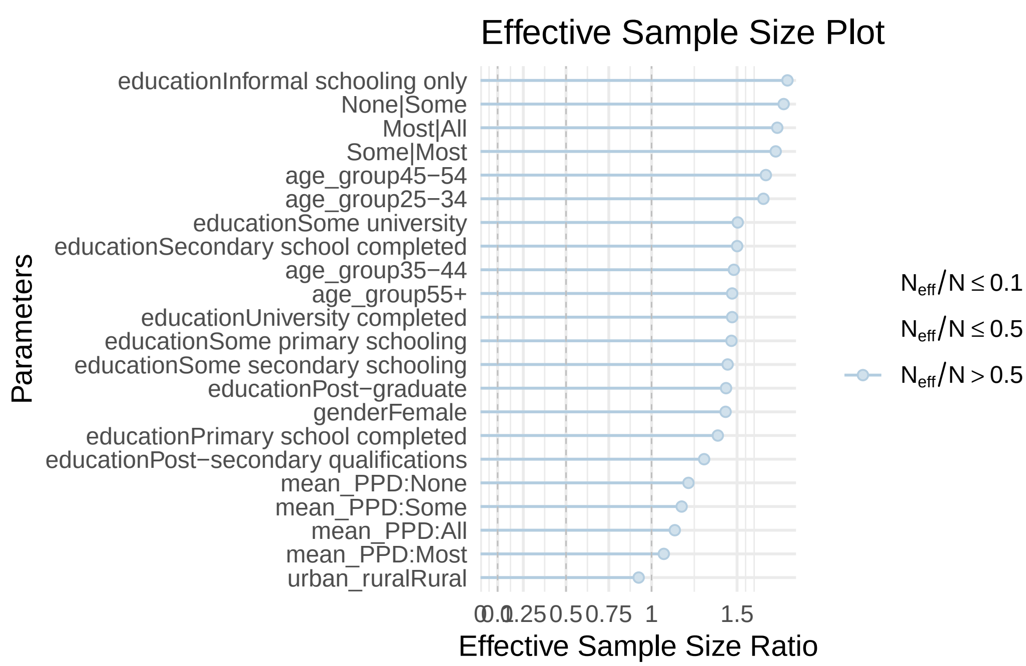

Figure 13: Effective sample size plot for parameters in the Bayesian ordinal logistic regression model.

References

Afrobarometer. 2023. "Afrobarometer Round 9, Nigeria, 2023." Afrobarometer Network. https://www.afrobarometer.org/survey-resource/nigeria-round-9-data-2023/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.