Perceptions of Corruption in Nigeria*

Rural Nigerians and Less Educated Groups Report Higher Corruption

Fatimah Yunusa

November 30, 2024

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

Table of contents

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

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....

2 Data

2.1 Measurement

For this analysis, i use the Afrobarometer Round 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 survey dataset contains responses from 1,556 people and shows 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.

Our primary measurement task is to capture public opinion towards economic, political and socioeconomic issues and translate it into actionable lintel. 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.

The data for this study was downloaded, cleaned, analyzed, modeled, and visualized using R (R Core Team 2023), a statistical programming language. The following packages were used:

- **`modelsummary`** (Arel-Bundock 2022): Used to create well-formatted regression tables and summaries.
- **`rstanarm`** (Team 2021): Facilitated Bayesian regression modeling using the Stan backend.
- **`dplyr`** (Wickham et al. 2023): Streamlined data manipulation, transformation, and summarization processes.
- **`here`** (Müller 2020): Simplified the organization of file paths, improving project reproducibility.
- **`arrow`** (Richardson et al. 2024): Enabled efficient handling of large datasets with fast reading and writing capabilities.
- **`bayesplot`** (Gelman, Gabry, et al. 2021): Assisted in producing diagnostic checks and posterior predictive visualizations for Bayesian models.
- **`ggplot2`** (Wickham 2016): Provided extensive tools for creating clear and impactful visualizations.
- **`patchwork`** (Pedersen 2024): Combined multiple plots into cohesive, professional visuals.
- **`knitr`** (Xie 2021): Enabled dynamic report generation, integrating code, output, and text seamlessly.
- **`kableExtra`** (Zhu 2024): Enhanced table styling and formatting for clean, professional presentations.
- **`ggalt`** (Rudis, Bolker, and Schulz 2017): Allowed the creation of specialized plots, enhancing the depth of data visualizations.
- **`haven`** (Wickham, Miller, and Smith 2023): Facilitated the import of `.sav` survey data from the Afrobarometer.

- `broom.mixed` (Bolker and Robinson 2024): Streamlined the extraction and tidying of results from mixed models for easy analysis and visualization.
- `afrobarometer` (Afrobarometer 2023): Served as the primary data source, providing valuable survey data for analysis.
- ***Telling Stories with Data*** (Alexander 2023): This book was referenced for its code and methodologies.

2.2 Cleaned Data Overview

Table 1: Preview of the Cleaned Dataset

corruption_presidency	education	gender	age_group	urban_rural	region
Some Corruption	Some university	Female	25-34	Urban Residents	634
Some Corruption	Primary school completed	Male	45-54	Urban Residents	634
High Corruption	Some university	Male	18-24	Urban Residents	634
Some Corruption	Some secondary schooling	Male	35-44	Urban Residents	634
Some Corruption	University completed	Male	25-34	Urban Residents	634
Some Corruption	Post-graduate	Female	35-44	Urban Residents	634

Table 1 represents the first six cleaned rows of the dataset, showing the important variables used in the analysis. The variables were constructed from specific survey questions in the Afrobarometer Round 9 Nigeria Survey (Afrobarometer 2023), which captures several socioeconomic sentiments within the country.The column corruption_presidency captures perceptions of corruption within the office of the presidency, derived from responses to the question, "How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: the president and officials in his office?" The variable education reflects respondents' highest level of education completed, gathered from the question, "What is the highest level of education you have completed?" The gender variable indicates the respondent's gender, collected through direct demographic inquiry. The column age_group categorizes respondents into age brackets based on the question, "How old are you?" The urban_rural variable distinguishes between urban and rural respondents, as recorded by interviewers based on the primary sampling unit (PSU) during data collection (Afrobarometer 2022b). Lastly, region specifies the geographical region of each respondent, identified using a unique code for Nigerian states. This table not only demonstrates the diversity of the dataset but also highlights the breadth of demographic and governance-related variables included for analysis.

Table 2: Detailed Summary Statistics of the Cleaned Dataset (Transposed)

Statistic	Value
Most Frequent Age Group	25-34
Urban Percentage	43.38%
Male Percentage	49.68%
Corruption Counts	None: 61 Some: 574 High: 593 All: 328
Education Counts	Primary: 203 Secondary: 173 University: 102 Postgraduate:

The cleaned dataset used in this study is summarised in Table 2, which provides key statistics to offer an overview of the dataset's structure and characteristics. The most frequent age group among respondents is identified, representing the predominant demographic in terms of age. The table also highlights the urban and male percentages, which indicate the proportion of urban residents (43.38%) and male respondents (49.68%) in the sample.

Further, the table breaks down the perceptions of corruption into four categories—"None," "Some," "High," and "All"—with counts for each, offering insights into respondents' views on corruption in the presidency. Additionally, the table details the distribution of respondents across education levels, including categories such as "Primary," "Secondary," "University," and "Postgraduate." Together, these summary statistics provide a comprehensive snapshot of the dataset, enabling a clear understanding of its demographic and categorical variables.

2.3 Variables of Interest

Figure 1 illustrates the distribution of survey respondents based on their residential location, categorized as urban or rural. From this, we see that the sample has more rural residents compared to urban residents. Specifically, the Afrobarometer survey for Nigeria, conducted in 2022, reports that approximately 53% of Nigeria's population resides in rural areas, this is is consistent with national demographic trends (Bank 2022). This rural-urban distribution is important for helping us understand the perspectives seen in the survey, since rural populations may have different challenges in governance and corruption compared to urban populations. The imbalance in representation shows that looking at geographical disparities when interpreting results from nationwide surveys is important.

Figure 2 presents the distribution of respondents' perceptions of corruption within the Nigerian presidency. The variable "Perceived Corruption in the Presidency" captures respondents' perceptions of the extent of corruption within the Nigerian presidency and its officials. This variable originates from Question 38A of the Afrobarometer survey, which asked: *"How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: the president and officials in his office?"* (Afrobarometer 2022a) Respondents could choose from the following values:

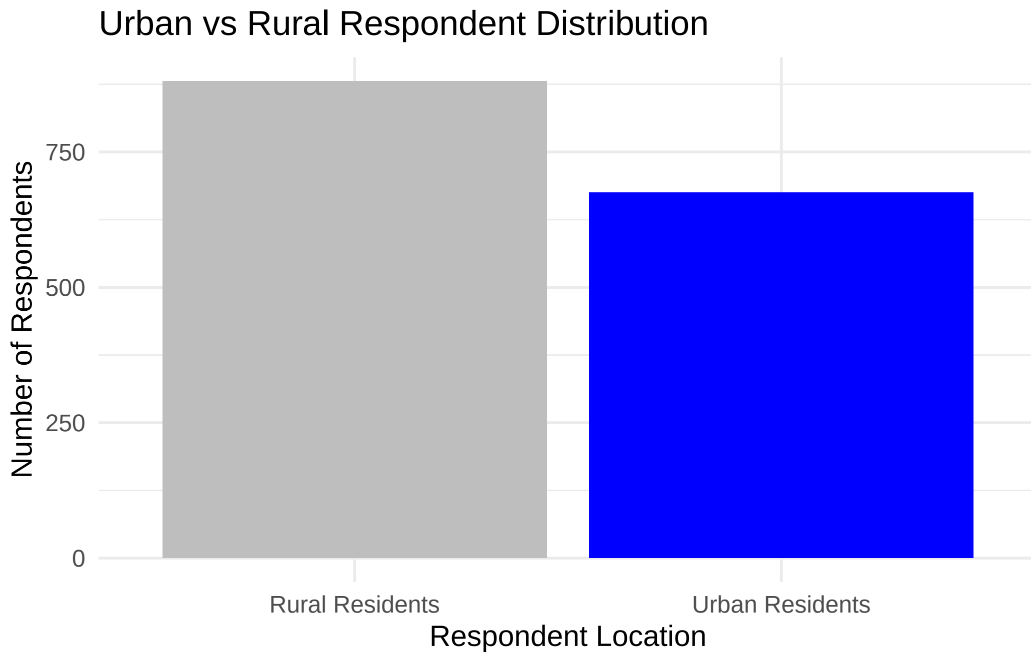

Figure 1: Urban vs Rural Respondent Distribution. This bar chart compares the number of respondents residing in urban versus rural areas in Nigeria.

- **0 (None)**: Indicates that the respondent perceives no corruption in the office of the presidency.
- **1 (Some of them)**: Suggests that the respondent believes a few individuals in the presidency are involved in corruption.
- **2 (Most of them)**: Reflects the perception that corruption is prevalent among the majority of the presidency's officials.
- **3 (All of them)**: Suggests the respondent perceives widespread corruption, implicating all officials in the presidency.
- **8 (Refused)**: Indicates the respondent declined to answer the question.
- **9 (Don't know/Haven't heard enough)**: Suggests the respondent was unsure or lacked sufficient knowledge to provide an answer.
- **-1 (Missing)**: Represents missing or unavailable responses.

For the analysis, only the substantive responses (0-3) were considered, in order to help us look into perceptions of corruption within the presidency. The high frequency of responses in the "High Corruption" and "Some Corruption" categories reflects persistent concerns about the integrity of key officials in Nigeria, aligning with broader findings on corruption in the region. Figure 2 presents the distribution of respondents' perceptions of corruption within the Nigerian presidency. Among these, the category "High corruption" is the most frequently reported, followed by "Some corruption". "All corrupt" ranks third, with "No Corruption" being the

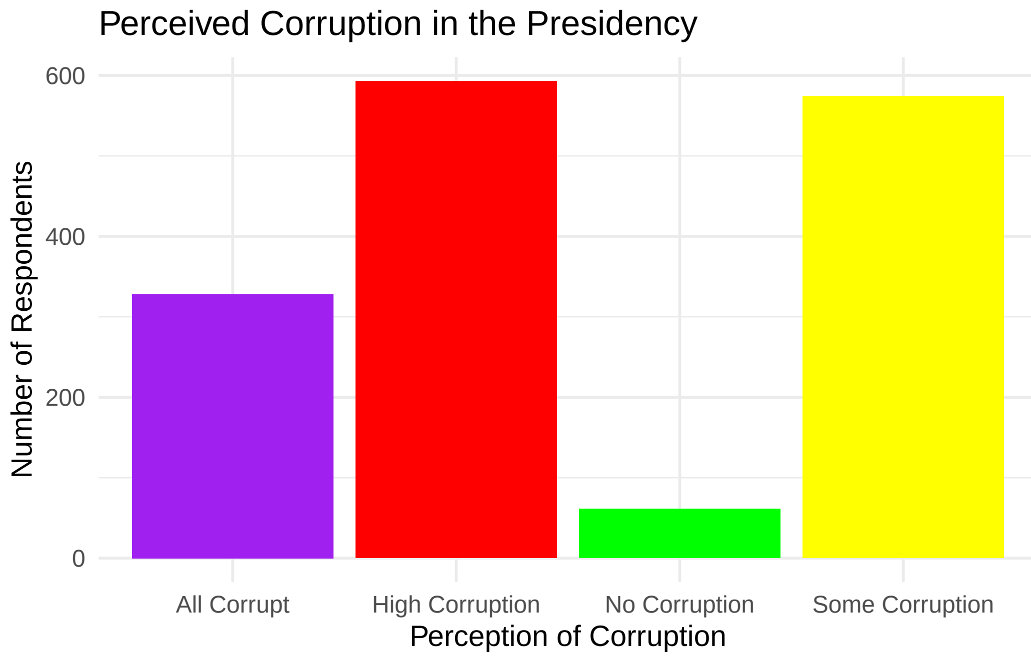

Figure 2: Perceived Corruption in the Presidency. This bar chart shows the distribution of respondents' perceptions of corruption in the presidency in Nigeria.

least chosen category which is no surprise given our data is in the context of Nigeria. These proportions show significant public concern about corruption within the presidency, consistent with broader trends in public concern abut corruption corruption within the presidency in Nigeria.

Figure 3 visually represents the educational attainment levels of respondents in the Afrobarometer survey for Nigeria. The data reflects responses to the question: *What is your highest level of education?*. The chart indicates that the most frequent reported level of education is "secondary school completed," followed by "primary school completed" and "no formal schooling". Fewer respondents reported having attained "post secondary qualifications," "university completed," or "Post-graduate" education, shows the disparity in advanced educational attainment across the population. This distribution shows that basic education levels are more prevalent among respondents, which may influence their perceptions and responses in the broader survey. Including educational attainment is important in our analysis because it is an important socioeconomic factor.

Figure 4 illustrates the distribution of survey respondents across five distinct age groups: 18-24, 25-34, 35-44, 45-54, and 55+. The data reveals that the 25-34 age group is the most represented, with over 500 respondents, followed by the 35-44 age group. The youngest cohort, aged 18-24, comprises approximately 300 respondents, while the older age groups (45-54 and 55+) are less represented in the dataset, with fewer than 200 respondents each. This distribution

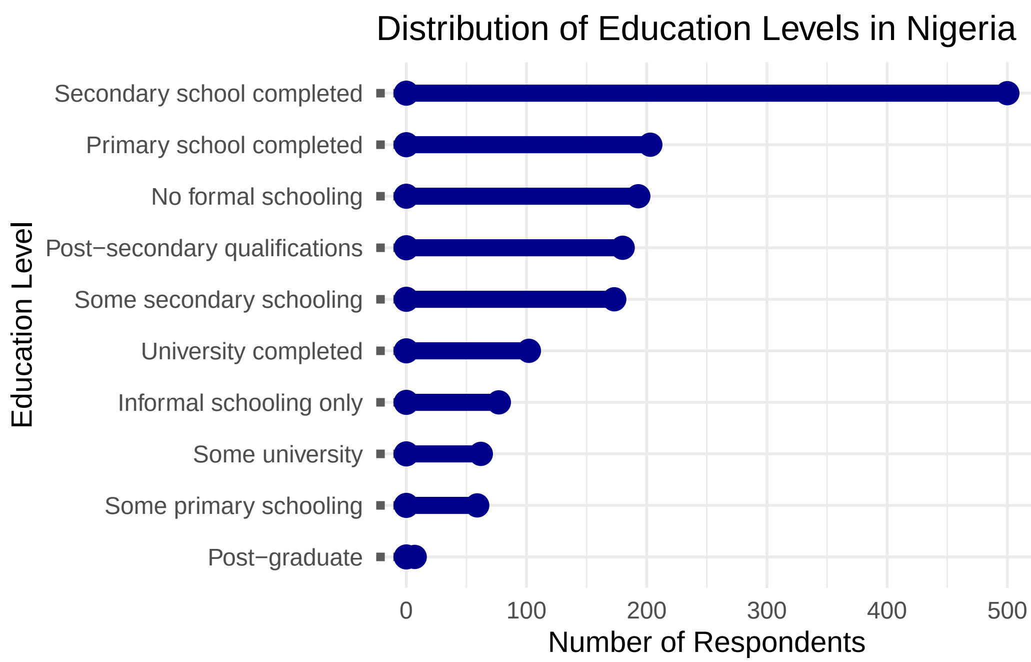

Figure 3: Distribution of Education Levels in Nigeria

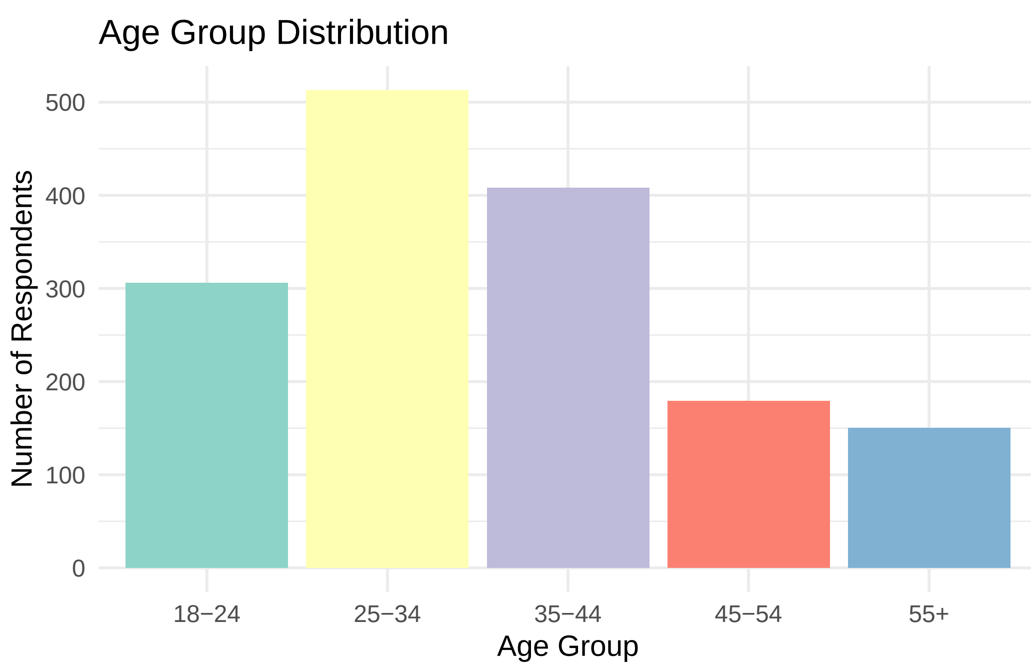

Figure 4: Age Group Distribution. This bar chart displays the number of respondents in each age group in Nigeria.

highlights a concentration of responses from younger demographics, particularly those aged 25-34, which may reflect broader societal trends such as population age structure or survey reach and accessibility.

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:

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

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

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.
  - 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 the Appendix.

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 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

Table 3: 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

The analysis revealed distinct trends in Corruption perceptions in different demographics. Our results are summarized in **?@tbl-modelresults.This** presents the estimated coefficients and standard errors for each predictor included in the model. The dependent variable is the perception of corruption in the presidency, categorized into ordered levels. The table includes predictors such as education level, gender, urban or rural location, and agre group, which are modeled to explain variations in corruption perceptions. Each predictor's coefficient represents its effect on the log odds of perceiving corruption in the presidency at a specific level relative to lower levels, holding other variables constant.

Higher education levels(e.g., university or post-graduate qualifications) arre associated with positive effects on corruption perceptions, indicating a greater likelihood of perceived higher levels of corruption. Begin female shows a small positive association with higher corruption perceptions. Residing in rural areas is negatively associated with corruption perceptions, suggesting lower perceptions of corruption compared to urban counterparts. Agegroups show minimal variation in their association with corruption perceptions, with no significant deviations observed among different age brackets.

4.2 Predicted Probabilities

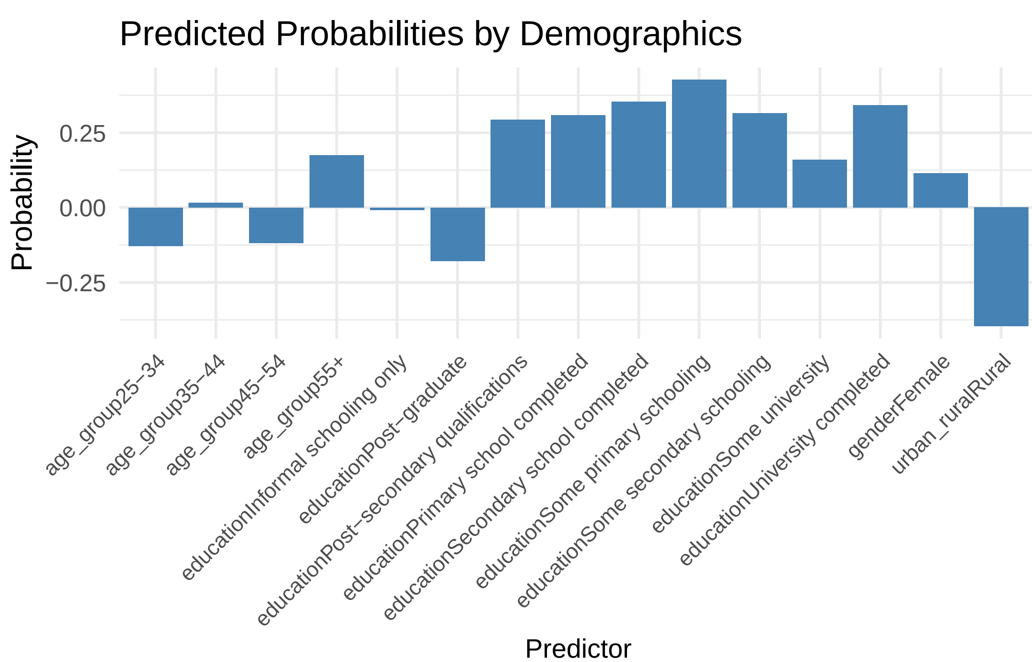

Figure 5: Predicted probabilities of high corruption perceptions by key predictors.

Figure 5 illustrates the predicted probabilities derived from the Bayesian ordinal logistic regression model for different demographic predictors influencing corruption perceptions in Nigeria.

Each bar represents the estimated impact of a specific predictor on the likelihood of perceiving higher corruption levels in the presidency.

Higher education levels consistently show positive probabilities, indicating that individuals with more advanced education levels are more likely to perceive higher corruption levels. Living in rural areas has a significantly negative probability, suggesting that rural residents are less likely to perceive high corruption compared to urban residents. Gender shows a moderate positive probability, with females slightly more likely to perceive higher corruption levels than males. There seems to be minimal variation observed across age groups, suggesting that age does not substantially influence perceptions of corruption.

The results show that higher levels of education, particularly university completion and post-secondary qualifications, are associated with greater perceptions of corruption in the presidency. Urban residents demonstrate significantly higher perceived corruption levels compared to rural residents. Gender differences reveal that female respondents show slightly higher corruption perceptions, while age groups show minimal variation. These findings are shown in the regression estimates and predicted probabilities, as visualized in Figure 5 and summarized in Table 3.

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

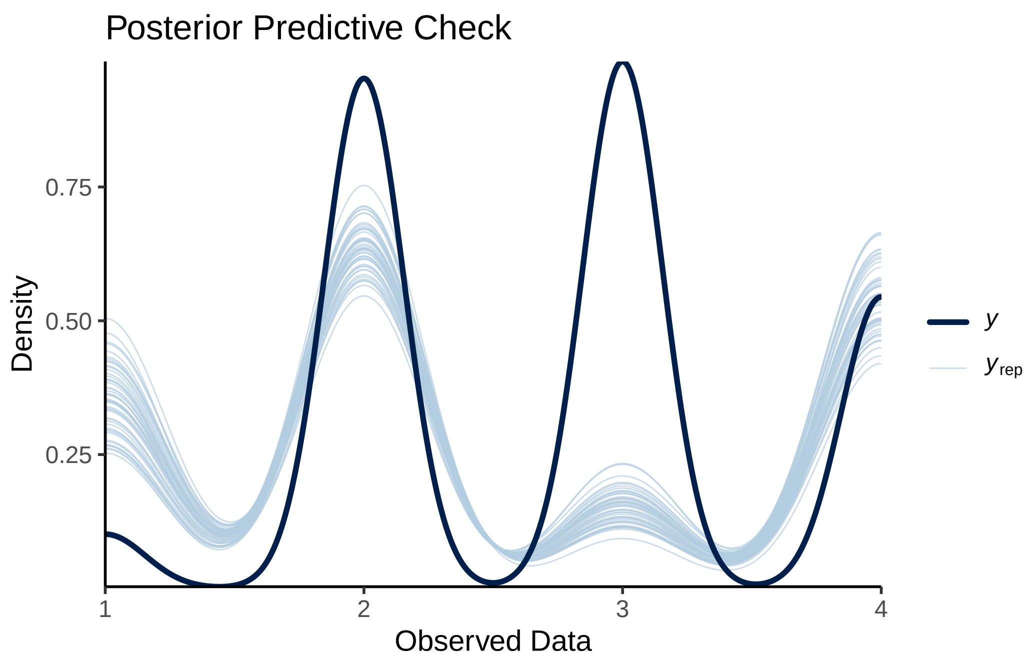

Figure 6: Posterior Predictive Check for the Bayesian ordinal logistic regression model.

B.2 Diagnostics

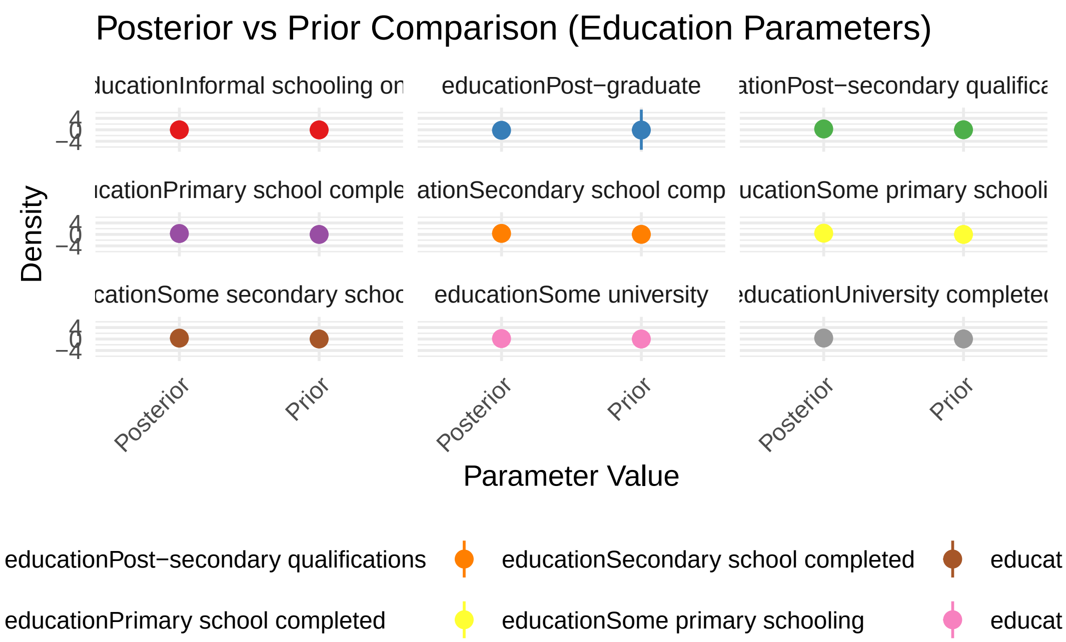

Figure 7: Posterior vs Prior Comparison for Education Parameters.

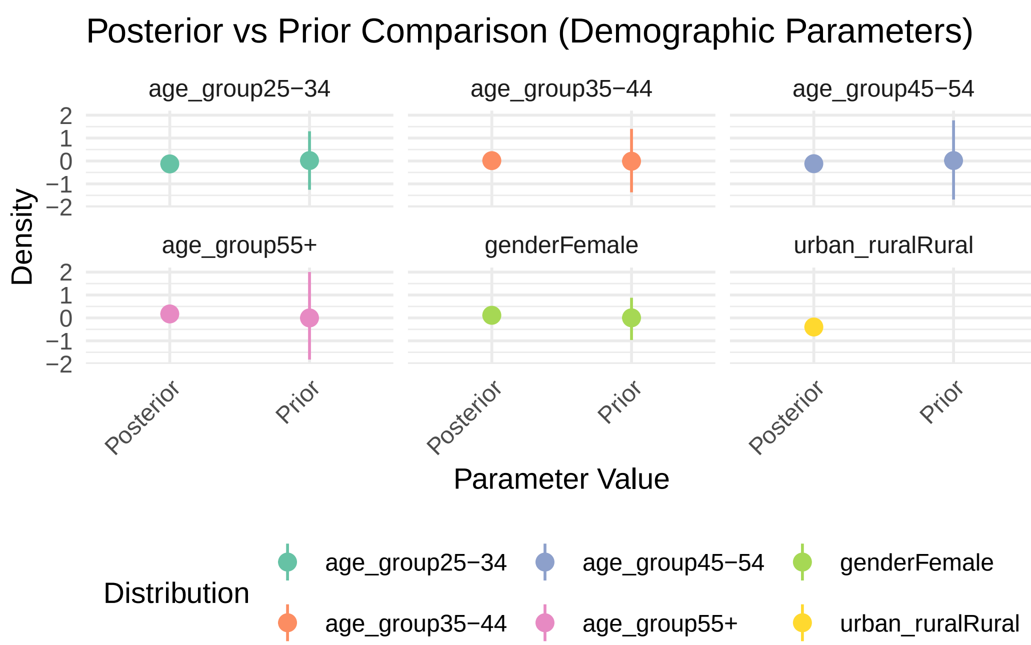

Figure 8: Posterior vs Prior Comparison for Demographic Parameters.

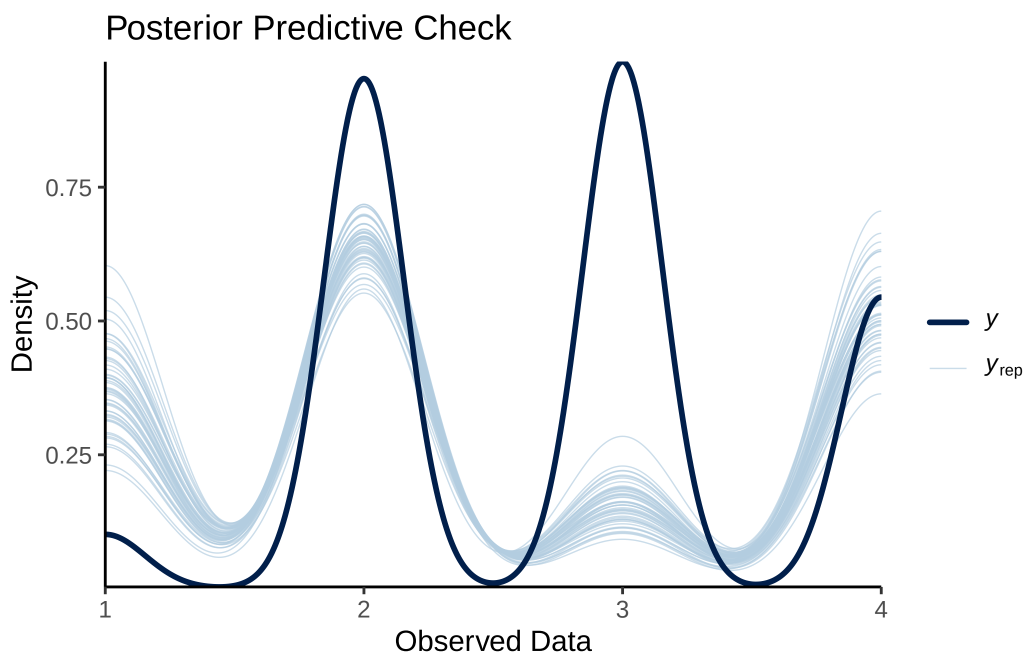

Figure 9: Posterior Predictive Check for the Bayesian ordinal logistic regression model.

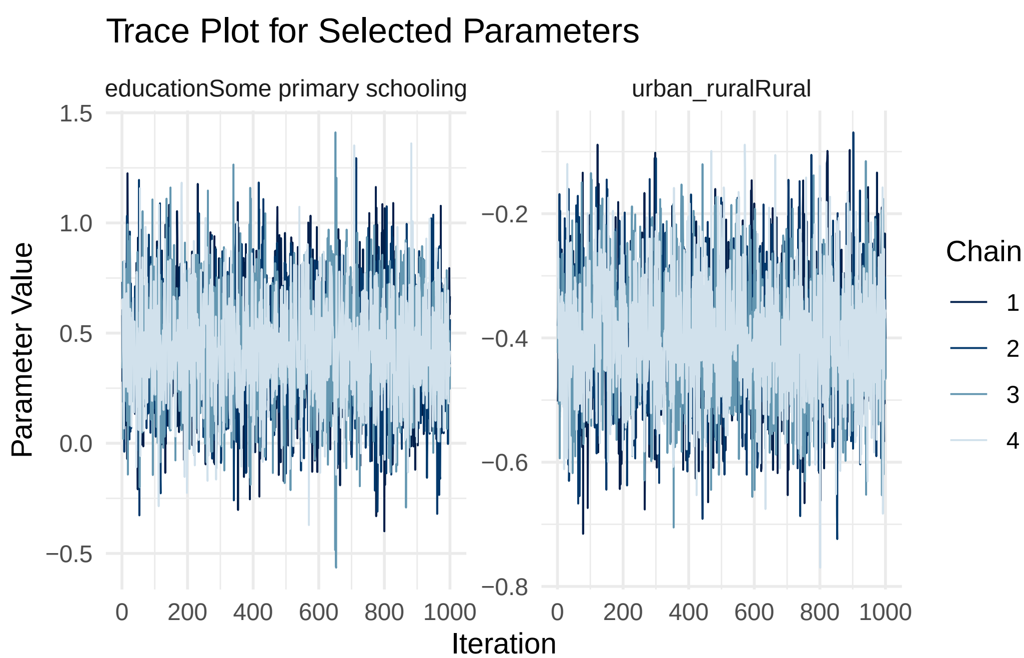

Figure 10: Trace plot for selected parameters in the Bayesian ordinal logistic regression model.

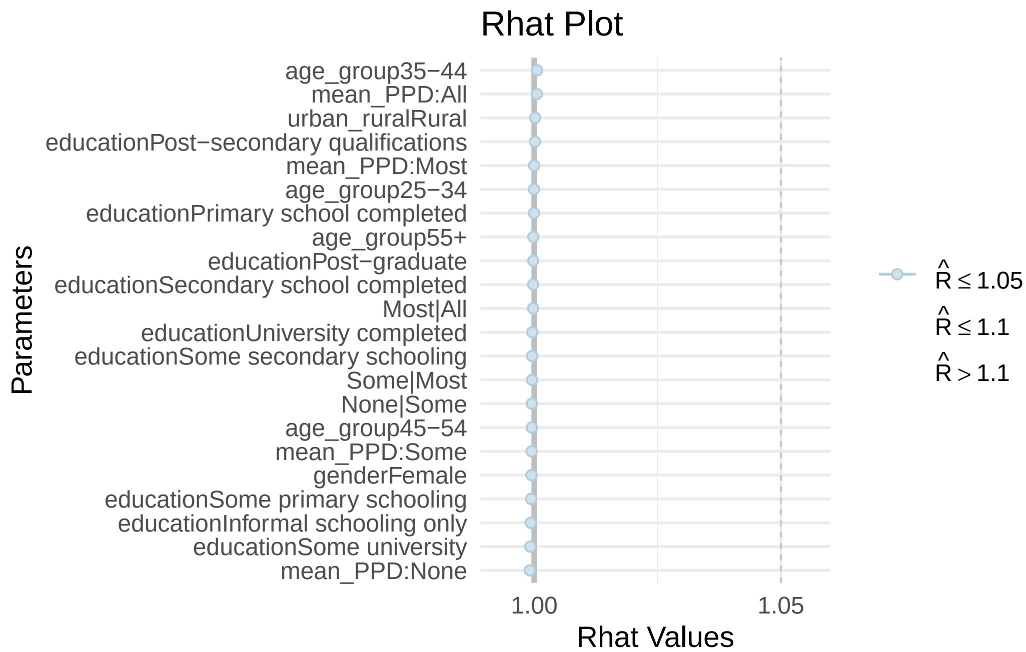

Figure 11: Rhat values for model convergence in the Bayesian ordinal logistic regression model.

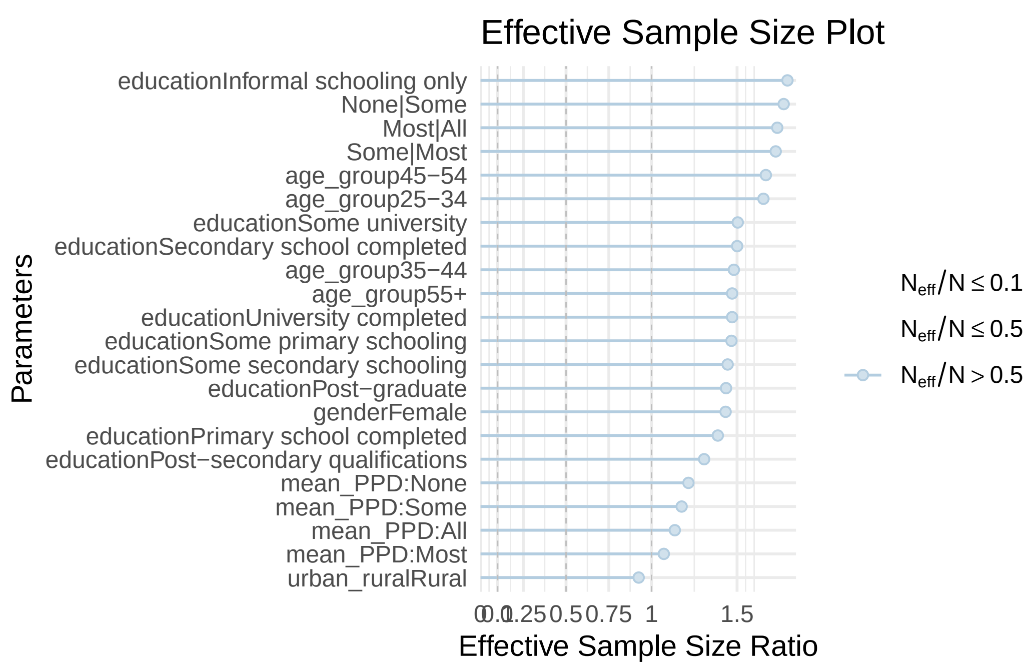

Figure 12: Effective sample size plot for parameters in the Bayesian ordinal logistic regression model.

References

Afrobarometer. 2022a. *Afrobarometer Nigeria Round 9 Codebook, 2022*. https://www.afrobarometer.org/survey-resource/nigeria-round-9-codebook-2022/.

———. 2022b. *Afrobarometer Round 9: Survey Manual*. https://www.afrobarometer.org/wp-content/uploads/2022/07/AB_R9.-Survey-Manual_eng_FINAL_20jul22.pdf.

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

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bank, World. 2022. "Rural Population (." https://data.worldbank.org/indicator/SP.RUR.TOTL?locations=NG.

Bolker, Ben, and David Robinson. 2024. *Broom.mixed: Tidying Methods for Mixed Models*. https://CRAN.R-project.org/package=broom.mixed.

Gelman, Gabriel, Jonah Gabry, et al. 2021. *Bayesplot: Plotting for Bayesian Models*. https://mc-stan.org/bayesplot.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Pedersen, Thomas Lin. 2024. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

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

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Rudis, Bob, Ben Bolker, and Jan Schulz. 2017. *Ggalt: Extra Coordinate Systems, 'Geoms', Statistical Transformations, Scales and Fonts for 'Ggplot2'*. https://CRAN.R-project.org/package=ggalt.

Team, Stan Development. 2021. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Evan Miller, and Danny Smith. 2023. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https:

//CRAN.R-project.org/package=kableExtra.