VILNIUS UNIVERSITY

FACULTY OF MATHEMATICS AND INFORMATICS

DATA SCIENCE STUDY PROGRAMME

Master’s thesis

Education, Region, and American Attitudes toward Immigrants: A Multilevel and Machine-Learning Analysis of the GSS 1996–2024

Išsilavinimas, regionas ir amerikiečių požiūris į imigrantus: daugialygė ir mašininio mokymosi GSS 1996–2024 analizė

Julia Bałulis

Supervisor : prof. habil. dr. Vydas Čekanavičius

Vilnius
2026

Summary

This thesis examines how education and regional context relate to American attitudes toward immigrants and national identity from 1996 to 2024. Using pooled data from five waves of the General Social Survey, four outcomes are analyzed: whether immigrants are seen as taking jobs, being bad for the economy, whether immigration should be reduced, and whether respondents report very strong attachment to American citizenship.

The main method is multilevel logistic regression with individuals nested within nine census regions. Education is split into a within–region (individual) component and a between–region (contextual) component. I add robustness checks with standardized income and compare the results to simple machine–learning models implemented in Python.

Across all outcomes, higher education within a region is strongly linked to more positive views of immigrants: better educated respondents are less likely to see immigrants as a jobs or economic threat, and less likely to support restricting immigration. Living in more educated regions points in the same direction, though contextual effects are smaller. In contrast, education is associated with weaker reported attachment to American citizenship. Income explains only part of this pattern: it improves the citizenship models but does not remove the education effects. Over time, perceived threats decline between 1996 and 2014, then rise again by 2024, while strong citizenship attachment falls.

The findings suggest that education and context shape both immigration attitudes and national identity, and that multilevel models combined with machine–learning tools provide a useful framework for studying these relationships in large social surveys.

Keywords: immigration attitudes, education, multilevel modelling, contextual effects, citizenship attachment, machine learning, General Social Survey.

Santrauka

Šiame darbe nagrinėjama, kaip išsilavinimas ir regioninis kontekstas siejasi su amerikiečių požiūriu į imigrantus ir nacionalinį identitetą 1996–2024 m. Naudodama sujungtus penkių *General Social Survey* bangų duomenis, analizuojama keturis rodiklius: ar imigrantai laikomi atimančiais darbo vietas, ar laikomi blogais ekonomikai, ar imigraciją reikėtų mažinti, ir ar respondentai nurodo labai stiprų prisirišimą prie Amerikos pilietybės.

Pagrindinis metodas yra daugiapakopė logistinė regresija, kur asmenys yra įdėti į devynis JAV surašymo regionus. Išsilavinimas suskaidomas į komponentą regiono viduje (individualų) ir komponentą tarp regionų (kontekstinį). Papildomai atlieku patikimumo patikras, įtraukdama standartizuotas pajamas, ir palyginu rezultatus su paprastais mašininio mokymosi modeliais, įgyvendintais naudojant `Python`.

Visuose analizuotuose rodikliuose aukštesnis išsilavinimas regiono viduje yra stipriai susijęs su palankesniu požiūriu į imigrantus: geriau išsilavinę respondentai rečiau mato imigrantus kaip grėsmę darbo vietoms ar ekonomikai ir rečiau palaiko imigracijos ribojimą. Gyvenimas labiau išsilavinusiuose regionuose rodo tą pačią kryptį, nors kontekstiniai efektai yra mažesni. Priešingai, išsilavinimas siejasi su silpnesniu deklaruojamu prisirišimu prie Amerikos pilietybės. Pajamos paaiškina tik dalį šio ryšio: jos pagerina pilietybės modelius, tačiau nepanaikina išsilavinimo efektų. Laikui bėgant suvokiamos grėsmės mažėja nuo 1996 iki 2014 m., tačiau iki 2024 m. vėl padidėja, o stiprus prisirišimas prie pilietybės silpnėja.

Rezultatai rodo, kad išsilavinimas ir kontekstas formuoja tiek požiūrį į imigraciją, tiek nacionalinį identitetą, o daugiapakopiai modeliai kartu su mašininio mokymosi įrankiais suteikia naudingą sistemą šių ryšių analizei didelėse socialinių apklausų duomenų bazėse.

Raktiniai žodžiai: požiūris į imigraciją, išsilavinimas, daugiapakopis modeliavimas, kontekstiniai efektai, prisirišimas prie pilietybės, mašininis mokymasis, *General Social Survey*.

List of Figures

List of Tables

Contents

1 Introduction

Public attitudes toward immigration are among the most debated and consequential issues in contemporary American politics. During the past several decades, the United States has also witnessed significant demographic changes, increased polarization, and lively immigration debates, leading to research on how regular Americans view immigrants and, importantly, how such views may be changing. Despite the fact that political leaders and, more commonly, the news media tend to frame the subject of immigration as big versus small, beneficial versus harmful to the country, the factors leading to such perceptions are complex and, until now, not fully understood.

One of the long-standing hypotheses offered in political and academic discourse alike suggests a link between economic self-interest and attitudes towards immigration. Those who perceive immigrants as rivals in the labor market or in terms of public services are apt to have a negative perception. However, recent research suggests that the role of cultural/ideological and psychological variables might play an equally or perhaps more prominent role. For instance, education at an individual and a collective level appears to have a regular role in influencing a tolerant and pro-immigration perception.

This thesis investigates the determinants of Americans' attitudes toward immigrants using data from the General Social Survey (GSS), focusing on three dimensions: (1) perceived economic threat (e.g., whether immigrants take jobs or hurt the economy), (2) immigration policy preferences (e.g., whether immigration levels should be reduced), and (3) national attachment (e.g., strength of identification with American citizenship). The analysis centers on the role of education—not only as an individual characteristic but also as a contextual factor at the regional level. It tests whether people with higher levels of education, and those living in more educated regions, are less likely to express anti-immigrant views.

This thesis uses five waves of the General Social Survey from the years 1996, 2004, 2014, 2022, and 2024 to examine the progression of views over time, including the most prominent political and social occurrences like the presidential election of 2016 and the occurrence of the COVID-19 pandemic. Multilevel logistic regression analysis is used to model the effect of education and other socio-demographic factors, while machine learning techniques (XGBoost) provide a secondary validation of key predictors.

The main research questions addressed are:

- How does individual education affect Americans' attitudes toward immigrants?
- Does regional educational context influence these attitudes beyond personal education?
- Have attitudes toward immigrants changed between 1996 and 2024?
- What role do other factors—such as age, race, sex, and income—play in shaping these views?
- How are national attachment and immigration attitudes connected, and does education affect both?

By addressing these questions, the thesis contributes to the understanding of the social foundations of immigration attitudes in the United States. It offers empirical insights into how education and regional context shape public opinion, and whether attitudes have hardened or softened in recent years. The findings have implications not only for sociological theory but also for policy, education, and democratic discourse in an increasingly diverse and globalized society.

2 Literature Review

Public attitudes toward immigrants in the U.S. have been studied through multiple lenses, with education emerging as a central factor in many accounts. A consistent finding is that higher individual educational attainment is associated with more positive or tolerant views of immigrants [4, 9, 10]. Better-educated Americans are less likely to agree that "immigrants take jobs" or harm the economy, a pattern documented since the early analysis of the General Social Survey (GSS) [**chandler2001**] and reaffirmed in later studies. In particular, Hainmueller and Hiscox show that education's influence transcends simple self-interest: highly educated respondents exhibit pro-immigrant preferences even when immigration might not benefit them economically, suggesting that education cultivates cosmopolitan values and reduces ethnocentrism [9]. Similarly, Haubert and Fussell identify a "cosmopolitan" outlook—rooted in higher education, white-collar work, and exposure to other cultures—as a key correlate of pro-immigrant sentiment [10]. However, recent scholarship cautions that this relationship may not be purely causal. Longitudinal analyses find that completing more years of schooling only modestly shifts attitudes, implying that pre-existing dispositions (e.g., openness or tolerance) lead both to pursuing education and holding pro-immigrant views [12]. Despite this selection effect, education remains one of the strongest predictors of immigration attitudes across datasets and decades, justifying its central role in this thesis.

Beyond individual education, contextual and geographic influences shape attitudes in important ways. Regional environments can either reinforce or mitigate anti-immigrant sentiment. For instance, rural residents tend to express more anti-immigrant attitudes than their urban counterparts, even after accounting for socioeconomic factors [7]. The explanations point to fewer intergroup contacts and more homogeneous social networks in rural areas, which can foster greater fear of outsiders. Moreover, contact theory asserts that a presence or proximity to immigrants can prevent or reduce prejudices, whereas the threat posed by a group, in this case, immigrants, could increase competition. Empirical evidence is mixed: some U.S. studies find that areas with more immigrants exhibit lower perceived threat on average, consistent with contact reducing anxiety, while others note that rapid increases in immigrant populations can spark backlash under certain conditions. A recent multilevel analysis of GSS data showed that the timing of context is crucial: growing up in a U.S. state with a sizable immigrant population has a "lasting impact" on producing more accepting attitudes in adulthood, more so than the contemporaneous immigrant share later in life [2]. This finding suggests that the context of the formative years' socializes the enduring views, highlighting the often-neglected temporal dimension of the context effects. This study extends the contextual approach by examining a novel factor: the level of regional education. There has been an indication that individual attributes, such as education, and macro-level variables, such as aggregate community education, could have a bearing on community norms and openness, and regions with more educated citizens could result in a more tolerant environment, which has been given little direct attention in research thus far and seeks to rectify that with this research approach that examines individual and regional education levels.

Alongside education and context, researchers have scrutinized economic and cultural frameworks to explain anti-immigrant attitudes. Classic ethnic competition theory argues that negative attitudes are due to perceived threats to native' jobs, wages, or public resources when immigrants are seen as competitors [6]. In this view, individuals in vulnerable economic positions or lower skill groups (often overlapping with lower education) should be most hostile due to labor market competition or fiscal burden concerns. Some evidence supports these material fears: for example, Americans who believe immigrants strain the economy or "take jobs" are indeed more likely to favor restrictive policies [6]. However, many studies conclude that cultural factors and identity-based "threats" are even more influential in shaping attitudes than economic self-interest. The concept of "heterophobia", which means fear of culturally foreign people, has been used to explain why some natives react negatively to immigrants who are seen as outsiders to the national group, irrespective of economic considerations [5]. Education plays a dual role here, as it is associated with better economic prospects and more liberal values. Erdmans and Smell, using GSS 1996–2014 data, found that education and religious ideology far outweighed employment status or income to predict anti-immigrant sentiment, with college-educated and non-fundamentalist respondents expressing significantly more pro-immigrant views [5]. Their results provided "mixed support" for pure economic competition theories; instead, they observed stronger evidence for the notion that generalized xenophobia or cultural anxiety (often tied to conservative ideologies and lower education) drives anti-immigrant attitudes. Similarly, Brown and Brown demonstrate that attitudes vary with both race and religion: white evangelical Christians, for example, tend to be more exclusionary, while minority and mainline Protestant respondents often hold more positive immigration stances, highlighting how ideological and identity frameworks (patriotism, religiosity, racial identity) intersect with attitudes [3]. In summary, the literature suggests that anti-immigrant sentiment is more than simply financial self-interest, but that these views are intricately meshed within broader structures of worldview. People with nationalist or nativist predispositions and ingroup commitment tend to see the immigrant as the cultural threat, whereas those predisposed toward more inclusive or cosmopolitan views—those who are more likely to be well-educated or young—tend to see the immigrant as an addition that positively enhances the cultural scene.

Another crucial consideration is whether public attitudes toward immigrants change over time or remain stable. Some longitudinal and cross-sectional comparisons indicate gradual liberalization of American attitudes in recent decades. For instance, between the mid-1990s and mid-2010s, overall levels of anti-immigrant agreement on GSS items decreased, contrary to expectations that growing immigrant populations would universally spur resentment [5]. Erdmans and Smell found that by 2014 Americans on average were more positive about the' impact of immigrants on jobs and the economy than they were in 1996, suggesting a period of increased tolerance (even as certain subgroups, such as religious fundamentalists, became more negative) [5]. Such changes could be related to cohort replacement, where younger and better-educated cohorts with more pro-immigrant sentiments comprise an increasingly larger portion of the population. The analysis by Kustov et al. points out that while aggregate opinion can shift, the' core attitudes of individuals tend to be remarkably stable over time, resilient to even major shocks such as economic recessions or the 9/11 terrorist attacks [11]. In their comprehensive panel-data study, immigration attitudes appeared "robust to major economic and political shocks," implying that dramatic short-term swings in public sentiment are rare [11]. This aligns with the observation that events such as the 2001 attacks can temporarily heighten anxiety about immigrants (e.g., a post-9/11 surge in Americans viewing immigrants as a security threat was documented by Esses et al. [6]), but these effects may be transient or concentrated among certain issues. The Trump era (2016–2020) similarly raised the question of whether anti-immigrant rhetoric tapped into a latent reservoir of hostility or actually shifted opinions. Initial research indicates that, despite heightened polarization, broad indicators of American attitudes did not uniformly deteriorate and in some cases became more pro-immigrant by the late 2010s, consistent with the long-term trajectory of increasing acceptance. Our thesis directly engages this debate by examining GSS data through 2024 to determine whether the earlier positive trend continued, stalled, or reversed in the wake of recent political turmoil.

In contextualizing this research within the current academic literature, this research aims to follow and further specify previous results. Prior work has firmly established education (and the cluster of traits it proxies) as a pivotal influence, but has also raised new questions about causality and context. By incorporating both individual-level educational effects and a novel measure of regional educational context, our analysis contributes evidence on whether an educated environment confers an additional attitudinal advantage net of one's own education. This addresses a gap in U.S.-based research, which has often considered geographic context in terms of urban–rural divides or local immigrant concentration [7, 8] but not explicitly the role of communal human capital. Furthermore, our inclusion of the 2020s survey waves allows us to update and test the generalizability of earlier patterns (e.g. the pro-immigrant shift from 1990s to 2010s noted by [5]) in a changed political landscape. Finally, from a methodological point of view, this thesis proposes a combination of different methods by complementing the multilevel logistic regression analysis with the use of a machine learning classifier. This thesis not only synthesizes the insights of economic threat and cultural perspectives from the literature, but also pushes the conversation forward—testing hypotheses with fresh data and techniques. This study offers a timely and valuable addition to the literature related to how attitudes towards immigration are impacted by education and environmental factors among Americans. It confirms what has come before while showing new aspects in the ongoing debate on immigration policy.

3 Methodology

This chapter describes the data, variable construction, and analytical methods used to examine American attitudes toward immigrants. The empirical strategy combines multilevel logistic regression models estimated in R with machine–learning models estimated in Python. The central goal is to assess how individual and regional education relate to perceived economic threat, immigration policy preferences, and citizenship attachment, while accounting for other socio-demographic factors and changes over time.

3.1 Data source and sample

The analysis uses data from the General Social Survey (GSS) [13], a nationally representative survey of adults in the United States conducted by the National Opinion Research Center (NORC). The GSS employs a multistage probability sample of non-institutionalized adults living in households in the United States. Interviews are conducted face-to-face (historically) and, more recently, in mixed modes.

For this thesis, I focus on five GSS waves: 1996, 2004, 2014, 2022, and 2024. These years were selected because they contain a consistent set of items on immigration attitudes, economic evaluations of immigrants, and national attachment, along with key socio-demographic variables. The raw data were obtained in SPSS (`.sav`) format from NORC and imported into R and Python.

Following standard GSS documentation, negative codes representing non-substantive responses (e.g. “IAP”, “DK”, “NA”, “No answer”) were treated as missing values. The cases were restricted to respondents within the five focal waves. Because missingness varies by variable and outcome, the analytic sample size differs slightly across models:

- The pooled dataset across all five waves contains $N = 15{,}107$ respondents.[1]
- The IMMJOBS (jobs–threat) models use 7,952 respondents from 1996, 2004, 2014, 2022, and 2024.
- The IMMAMECO (economic threat) models use 5,121 respondents from 1996, 2004, 2014, and 2024.
- The LETIN1 (immigration restriction) models use 4,722 respondents from 1996, 2004, 2014, and 2024.
- The AMCITIZN (citizenship attachment) models use 5,285 respondents from 1996, 2004, 2014, and 2024.
- The 2022–2024 IMMJOBS models use 4,244 respondents; the income-robustness models use the subset of 3,843 respondents with non-missing income.

[1] See Table 1 for pooled descriptive statistics.

All analyses are conducted on unweighted data. This means that each respondent contributes equally to the estimation. While GSS provides survey weights to adjust for sampling design and non-response, incorporating them into multilevel models with random effects and multiple imputed constructs is non-trivial. The unweighted approach simplifies the modelling and does not affect the internal comparisons within the sample, but should be taken into account when interpreting the results as nationally representative.

3.2 Variable construction

The key variables used in the analysis fall into three groups: outcome variables that capture immigration attitudes and citizenship attachment; education and other individual-level predictors; and contextual variables at the level of census regions.

3.2.1 Outcome variables

Four binary outcome variables were constructed from GSS Likert-scale questions:

- **Jobs threat (IMMJOBS).** The GSS item asks whether "immigrants take jobs away from people born in America" or "do not take jobs away". The binary variable `immjobs_anti` equals 1 if the respondent agrees or strongly agrees that immigrants take jobs away (codes 1 or 2 on the original scale), and 0 otherwise.
- **Economic threat (IMMAMECO).** The item asks whether "immigrants are generally good for America's economy". The binary variable `immameco_anti` equals 1 if the respondent disagrees or strongly disagrees that immigrants are good for the economy (codes 4 or 5), and 0 otherwise.
- **Immigration restriction (LETIN1).** The item asks whether the United States should allow more or fewer immigrants to come and live here. The binary variable `letin_restrict` equals 1 if the respondent favours allowing "fewer" or "many fewer" immigrants (the two most restrictive categories), and 0 otherwise.
- **Citizenship attachment (AMCITIZN).** The item asks whether the respondent would "rather be a citizen of America than of any other country". The binary variable `cit_attach` equals 1 if the respondent agrees or strongly agrees (codes 1 or 2), and 0 otherwise.

These recodings map ordered response scales onto dichotomous indicators of negative immigration attitudes (for IMMJOBS, IMMAMECO, LETIN1) and strong national attachment (for AMCITIZN), which are natural for logistic regression and classification models.

3.2.2 Education and other individual-level predictors

Education is measured in years of schooling completed (`educ`). To distinguish individual from contextual effects, education is decomposed into within-region and between-region components.

Let j index regions (the nine U.S. census regions) and i index individuals. For each wave, I compute the mean education in region j, denoted $\bar{\text{educ}}_j$, and the grand mean across all regions, $\bar{\text{educ}}$.

The decomposition is:

$$\text{educ\_w}_{ij} = \text{educ}_{ij} - \bar{\text{educ}}_j, \qquad \text{educ\_b}_j = \bar{\text{educ}}_j - \bar{\text{educ}}.$$

Thus, educ_w_{ij} captures how much more or less education an individual has relative to their regional mean (a purely individual-level deviation), while educ_b_j captures how educated the region is relative to the national average. For ease of interpretation and comparability across models, both components are standardised to have mean zero and standard deviation one within the analytic sample. In the results, they are denoted `educ_w` and `educ_b`.

Additional individual-level predictors include:

- **Age.** Age in years is rescaled by decades (`age10` = age / 10) and then standardised (mean zero, unit variance). Coefficients are therefore interpretable as the effect of a one-standard-deviation change in age (roughly ten years).
- **Sex.** Sex is coded as a binary factor with men as the reference category. The resulting dummy `sex_ffemale` equals 1 for women and 0 for men.
- **Race.** Race is coded as a categorical factor with three levels: White (reference), Black, and “Other” (all remaining categories combined). In the models, this yields two dummy variables, `race_fblack` and `race_fother`.
- **Income.** Household income is measured using the GSS ordinal income rank variable (1–12). For the robustness models, income is standardised to have mean zero and standard deviation one (`income_z`). Income is only included in models estimated on the 2022–2024 subsample with non-missing income.
- **Survey year.** Year is included as a categorical factor. For the IMMJOBS, IMMAMECO, and AMCITIZN models, 2014 is used as the reference year. For the LETIN1 (restriction) model, 1996 is used as the reference in order to highlight long-run changes from the mid-1990s. In the 2022–2024 IMMJOBS models, 2022 serves as the reference category.
- **Region.** Region is the nine-category U.S. Census region variable. It is treated as a grouping factor for random intercepts.

All continuous predictors that enter the multilevel models (`educ_w`, `educ_b`, `age10`, and, where used, `income_z`) are standardised. This facilitates interpretation of odds ratios as the effect of a one-standard-deviation change and helps with model convergence.

3.2.3 Contextual variables

The primary contextual variable is **regional mean education** (`educ_b`), described above. By construction, it is constant within regions but varies between regions. The random intercepts for regions capture additional unobserved regional characteristics beyond average education, such as local political culture, media environments, economic structure, and immigrant presence.

3.3 Multilevel logistic regression in R

The main analyses use multilevel (hierarchical) logistic regression models estimated with the `glmer()` function from the `lme4` package in R. These models are appropriate for hierarchically structured data in which individuals (level 1) are nested within regions (level 2). They allow both individual-level and contextual predictors while accounting for clustering within regions.

For each binary outcome Y_{ij} (e.g. `immjobs_anti`) for individual i in region j, the basic model is:

$$\text{logit Pr}(Y_{ij} = 1) \ = \ \beta_0 + \beta_1\,\text{educ\_w}_{ij} + \beta_2\,\text{educ\_b}_j + \boldsymbol{\beta}_3^\top \mathbf{X}_{ij} + u_j, \tag{1}$$

where:

- $\mathbf{X}_{ij}$ is a vector of control variables (survey year dummies, `age10`, `sex_ffemale`, `race_fblack`, `race_fother`, and, in robustness models, `income_z`),
- $u_j \sim \mathcal{N}(0, \sigma_u^2)$ is a random intercept for region j,
- the coefficients $(\beta_0, \beta_1, \beta_2, \boldsymbol{\beta}_3)$ are fixed effects.

The random intercept u_j captures systematic differences in baseline attitudes between regions after controlling for observed predictors. The variance component σ_u^2 is summarised as a standard deviation and interpreted as the extent of regional clustering.

Models are estimated by maximum likelihood using the Laplace approximation. In cases where the default optimiser encountered convergence warnings, the `bobyqa` optimiser was used with an increased maximum number of iterations (`maxfun = 2e5`). For each model, fixed-effect coefficients are exponentiated and reported as odds ratios with 95% confidence intervals (derived from Wald standard errors).

3.3.1 Model variants and robustness checks

For each outcome, the following variants were estimated:

- **All-waves model.** A full model including `educ_w`, `educ_b`, survey year dummies, `age10`, `sex_ffemale`, and race dummies, with a random intercept for region. This is the main specification reported for IMMJOBS, IMMAMECO, LETIN1, and AMCITIZN.
- **Recent-waves model (IMMJOBS, 2022–2024).** A model restricted to the 2022 and 2024 waves, using the same predictors but with 2022 as the reference year. This allows a focused examination of education effects and recent changes in jobs–threat attitudes.

- **Income robustness models (IMMJOBS, 2022–2024).** Two models estimated on the income-complete 2022–2024 subsample: one replicating the recent-waves specification without income, and one adding standardised income (`income_z`). A likelihood-ratio test compares model fit with and without income, and changes in the education coefficients are examined.

Diagnostic checks include inspection of convergence warnings, examination of random-effect variances, and comparison of alternative optimisers when necessary. Because the primary focus is on substantive interpretation rather than prediction, detailed residual diagnostics were not reported, but the models show no evidence of pathological fits (such as extremely large random-effect variances or separation problems).

3.4 Machine-learning validation in Python

To complement the multilevel models and provide a data-driven view of predictor importance, a set of machine-learning models in Python were estimated for two key outcomes: `immjobs_anti` (jobs threat) and `immameco_anti` (economic threat).

3.4.1 Data preprocessing and train–test split

The machine-learning analyses use the same conceptual predictors as the multilevel models: survey year, census region, race, sex, years of education, age, and income. Preprocessing proceeds as follows:

- The data are restricted to respondents with non-missing values on the selected predictors and the outcome.
- The sample is split into a 75% training set and a 25% test set, stratified by the outcome to preserve the marginal class distribution.
- Categorical variables (year, region, race, sex) are one-hot encoded.
- Continuous variables (education, age, income) are scaled but not standardised to zero mean and unit variance, as tree-based models are insensitive to monotonic transformations.

3.4.2 Logistic regression baselines

As a parametric baseline, regularised logistic regression models were estimated using scikit-learn's `LogisticRegression` with class weights inversely proportional to class frequencies to address outcome imbalance. The models are fit with a maximum of 5,000 iterations.

Performance is evaluated on the held-out test set using the following metrics:

- overall accuracy,
- balanced accuracy (mean of sensitivity and specificity),

- area under the receiver operating characteristic curve (AUC).

Table 11 (Table 12 in the Results chapter) reports these metrics for `immjobs_anti` and `immameco_anti`.

3.4.3 XGBoost models and feature importance

To allow for non-linearities and interactions, gradient-boosted tree models (XGBoost, `XGBClassifier`) were estimated on the same training and test sets. The hyperparameters are tuned conservatively to favour interpretability and avoid overfitting:

- number of trees $\approx 2{,}000$,
- learning rate $\eta = 0.03$,
- maximum tree depth = 4,
- subsampling and column subsampling at standard defaults,
- early stopping based on validation loss.

Although the XGBoost models do not dramatically improve AUC or accuracy relative to the logistic baselines, they provide gain-based feature-importance measures. The top-ranked features across both outcomes include years of education and regional indicators, followed by specific region–year combinations and, to a lesser extent, race and age. Income is less prominent. These patterns mirror the multilevel results and thus serve as a robustness check: the variables that matter most in flexible tree-based models are the same ones that show strong and consistent effects in the multilevel models.

The machine-learning component is therefore used not as a primary inferential tool, but as a complementary approach that confirms the central role of education and regional context in structuring immigration attitudes.

4 Results

4.1 Descriptive patterns in the sample

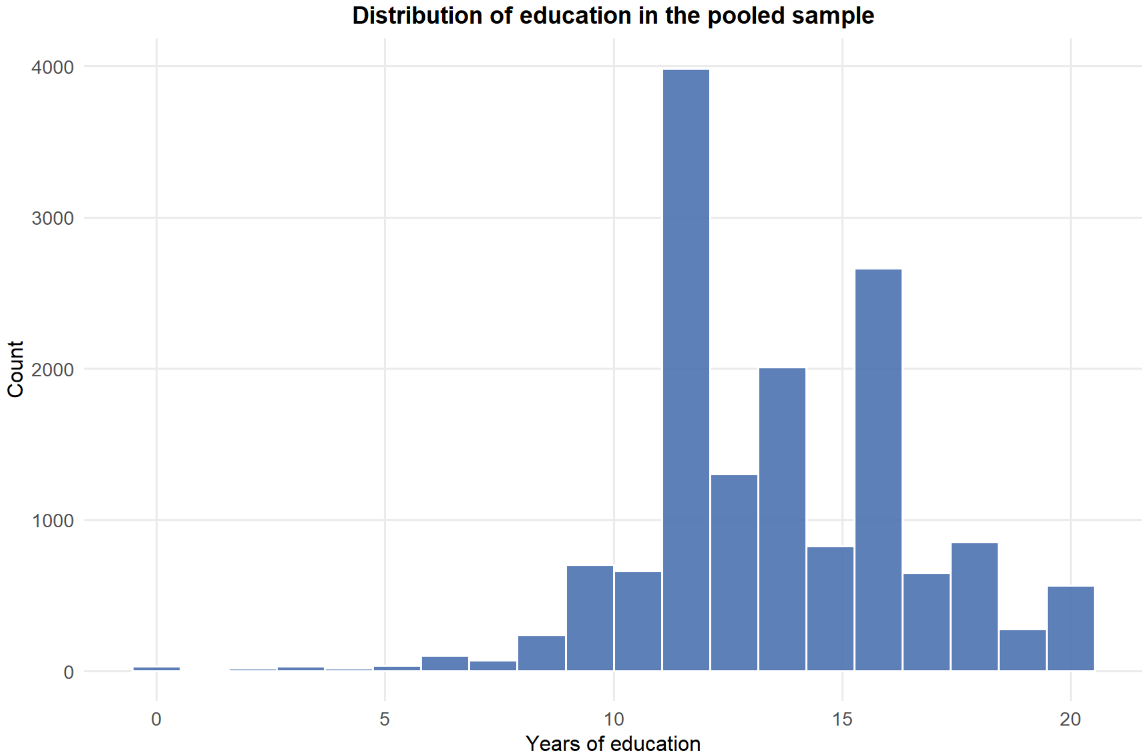
4.1.1 Sample composition

Table 1 summarises the pooled sample across all five GSS waves (1996–2024). In total there are 15,107 respondents with non-missing values on the key predictors. Although sex (gender) was initially included as a predictor in all multilevel models, it consistently yielded no significant effects on any outcome (all $p > 0.10$). This null finding persisted across all models and robustness checks, indicating that sex did not meaningfully contribute to explaining the variance in outcomes. Therefore, to maintain clarity in the final analysis, sex was omitted from all subsequent models.
On average, respondents report about 14 years of education (mean = 13.84, SD = 2.95), which corresponds roughly to a high-school degree plus some post-secondary schooling. The average age in the sample is about 48 years (SD = 17.5), so the data include both younger adults and older cohorts. Income is measured in GSS income categories; in this pooled sample the mean is around 11 on the 1–25 scale, with a standard deviation of 2.33, indicating substantial dispersion.

***1** Descriptive statistics for key predictors (pooled sample, 1996–2024).*

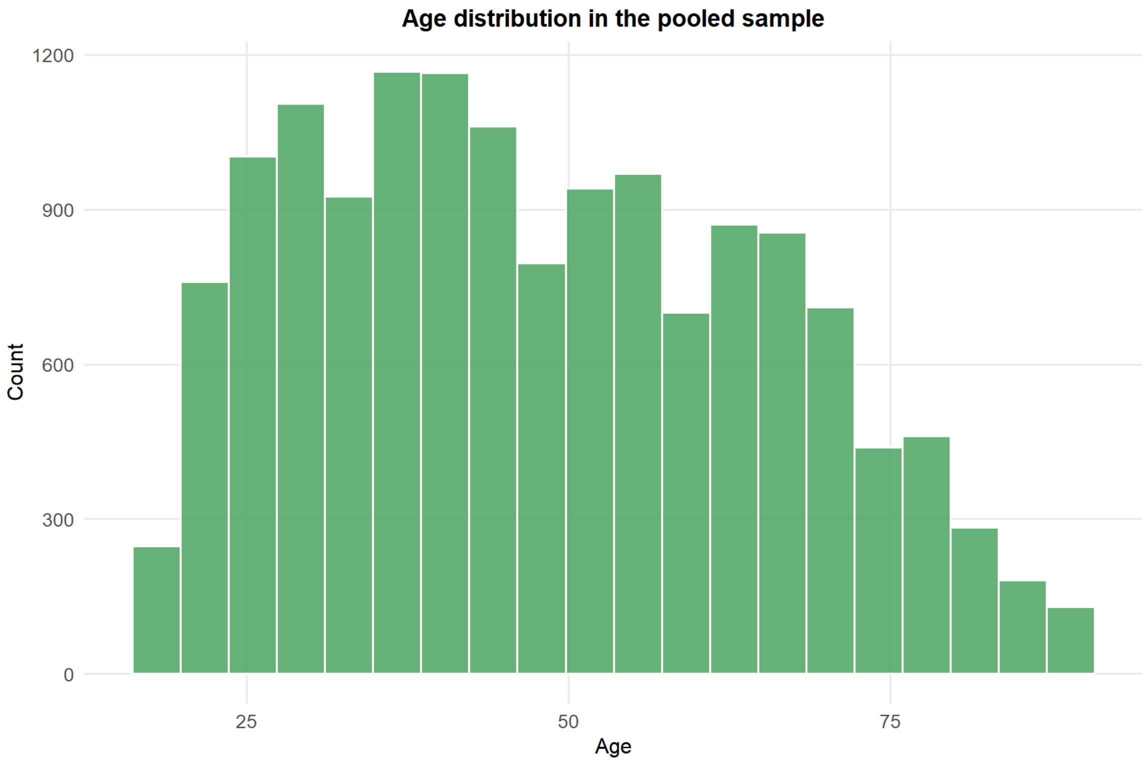
N	mean_educ	sd_educ	mean_age	sd_age	mean_inc	sd_inc
15107	13.84	2.95	47.94	17.54	10.96	2.33

The distribution of education is shown in Figure 1. Most respondents report between 11 and 16 years of schooling, with relatively few cases at the very low or very high ends of the scale. This confirms that the sample is centred around high-school graduates and people with some college or university education.



1 *Distribution of years of education in the pooled sample, 1996–2024.*

Figure 2 shows the age distribution. The sample covers a wide age range, from early adulthood to late old age, with the largest counts in middle adulthood. For the later models this means that both education and age have enough variation to identify their relationships with immigration attitudes.



2 *Age distribution in the pooled sample, 1996–2024.*

4.1.2 Trends in immigration attitudes and citizenship attachment

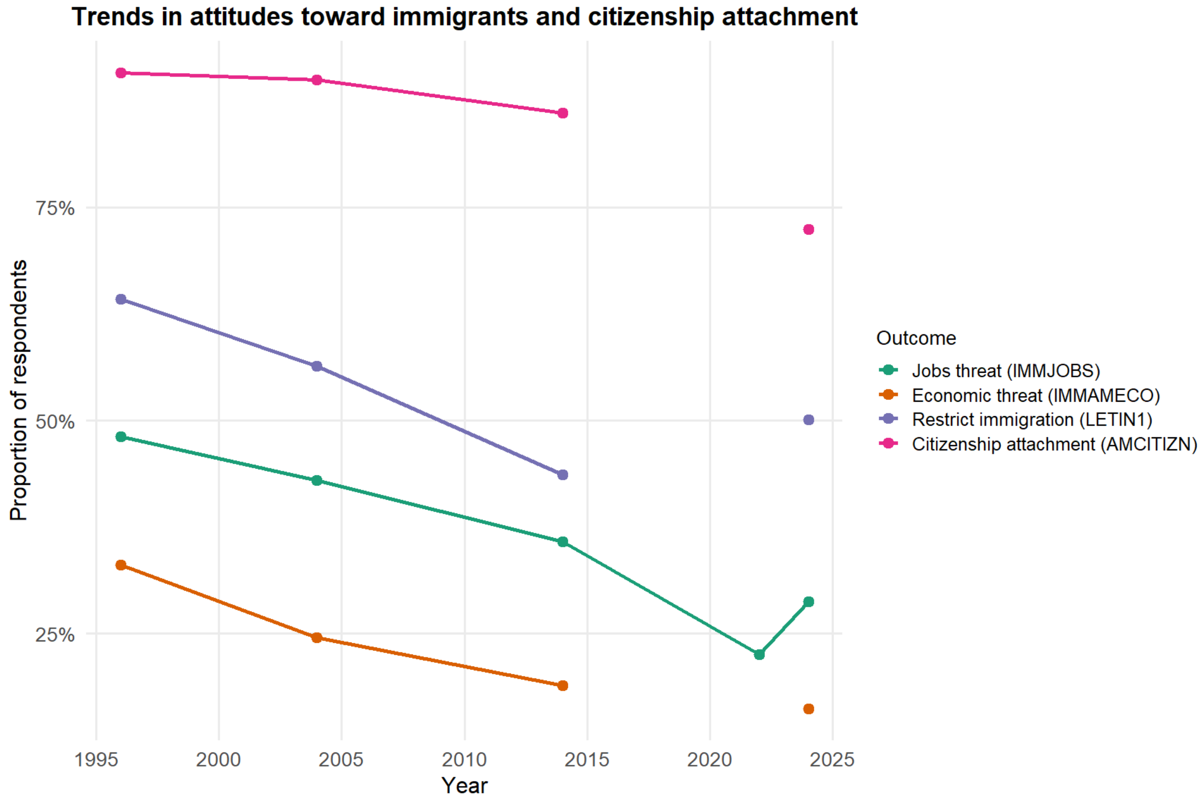
Table 2 reports, for each wave, the proportion of respondents who hold the more negative position on the three immigration items and the proportion reporting strong attachment to American citizenship. Between 1996 and 2014 there is a clear decline in perceived jobs and economic threat from immigrants, and a decline in support for reducing immigration. For instance, the share agreeing that immigrants “take jobs away’’ falls from 48% in 1996 to 36% in 2014, and the share saying that the number of immigrants should be reduced falls from 64% to 44%. Citizenship attachment remains very high in these earlier waves, above 85%.

In the most recent wave, 2024, some of these trends shift. Support for reducing immigration rises again to around 50%, and jobs–threat concerns increase to 29%. Economic threat continues to edge down to 16%, while strong citizenship attachment declines to about 72%. In 2022 only the jobs–threat item was fielded, so the other entries for that year are missing.

2 *Proportion of respondents expressing anti-immigrant positions by year.*

year	n	Jobs threat (%)	Economic threat (%)	Restrict immigration (%)	Citizenship attachment (%)
1996	2904	48.1	33.0	64.2	90.8
2004	2812	43.0	24.6	56.4	90.0
2014	2538	35.8	18.9	43.6	86.1
2022	3544	22.5	—	—	—
2024	3309	28.7	16.2	50.1	72.4

Figure 3 visualises these changes over time. The lines for the three immigration items all slope downward between 1996 and 2014, showing a long-run liberalisation of attitudes, while the uptick in 2024 for jobs threat and immigration restriction suggests renewed scepticism in the early 2020s. By contrast, the line for citizenship attachment starts at a very high level and then gradually declines, particularly in the most recent wave. These descriptive patterns motivate the multilevel and machine-learning analyses in the following sections.



3 Trends in attitudes toward immigrants and citizenship attachment, 1996–2024.

4.2 Jobs–threat attitudes (IMMJOBS)

4.2.1 Multilevel model, 1996–2024

The first multilevel model uses all five waves in which IMMJOBS was asked and includes 7,960 respondents clustered in nine census regions. The binary outcome `immjobs_anti` is coded 1 for respondents who agree or strongly agree that immigrants "take jobs away" and 0 otherwise, so odds ratios (OR) below 1 indicate a lower probability of seeing immigrants as a jobs threat.

***3** Multilevel logistic regression of jobs–threat attitudes (IMMJOBS), 1996–2024. Odds ratios with 95% confidence intervals.*

Variable	OR	CI low	CI high	p
Intercept	0.57	0.49	0.66	<0.001
Education (within region, z)	0.68	0.64	0.71	<0.001
Education (between region, z)	0.84	0.77	0.92	<0.001
Year: 1996 (ref. 2014)	1.65	1.39	1.94	<0.001
Year: 2004 (ref. 2014)	1.39	1.17	1.64	<0.001
Year: 2022 (ref. 2014)	0.55	0.47	0.64	<0.001
Year: 2024 (ref. 2014)	0.74	0.63	0.86	<0.001
Age (per 10 years, z)	1.07	1.04	1.10	<0.001
Black (ref. White)	1.07	0.94	1.23	0.322
Other race (ref. White)	0.44	0.36	0.54	<0.001

Table 3 shows that education is strongly associated with jobs–threat attitudes. The within-region education effect (`educ_w`) has an OR of 0.68: within the same region, respondents with higher education are much less likely to think that immigrants take jobs away. The between-region component (`educ_b`) is also below 1 (OR = 0.84), so people living in more educated regions are somewhat less negative toward immigrants, although this contextual effect is smaller than the individual effect.

Time dummies are coded relative to 2014. The odds of viewing immigrants as a jobs threat are substantially higher in 1996 (OR = 1.65) and 2004 (OR = 1.39), and lower in 2022 (OR = 0.55) and 2024 (OR = 0.74). This pattern fits the descriptive trends in Table 2: there is a clear long-term decline in jobs–threat perceptions between the mid-1990s and mid-2010s, and even though concerns rise again by 2024, they remain below the levels observed in 1996 and 2004.

Age and race also show meaningful associations. Each ten-year increase in age is linked to about 7% higher odds of agreeing that immigrants take jobs away (OR = 1.07). The coefficient for Black respondents is close to 1 and not statistically significant, suggesting no clear difference from White respondents. In contrast, respondents in the residual "other race" category have much lower odds of seeing immigrants as a jobs threat (OR = 0.44). The random intercept variance at the regional level is small but non-zero, which indicates modest clustering of attitudes by region after controlling for the individual covariates.

4.2.2 Recent waves: 2022–2024 comparison

To focus on the most recent period, a second model is estimated on the 2022 and 2024 waves only, with the same predictors as above. In this model, the year dummy compares 2024 directly to 2022. The education effects remain very similar: the within-region education OR is 0.69 and the between-region OR is 0.84, so higher education continues to be associated with less agreement that immigrants take jobs away even in the latest years.

The 2024 dummy shows an OR of about 1.34 for 2024 versus 2022. Conditional on education, age, race, and region, respondents in 2024 are more likely than those in 2022 to perceive immigrants as a jobs threat. Substantively, this is a moderate increase rather than a complete reversal. It mirrors the descriptive uptick between 2022 and 2024 in Table 2, but the change is smaller than the large differences between 1996 or 2004 and 2014.

Other covariates behave as expected. Age remains positively associated with jobs–threat attitudes, and the coefficient for Black respondents is again not significant. Respondents classified as "other race" continue to show lower odds of jobs–threat perceptions compared to White respondents.

4.2.3 Income robustness checks

Finally, I investigate whether the education effects could simply be capturing differences in household income. For this purpose, two models are estimated on the same subset of 2022–2024 respondents with non-missing income: a base model without income and a model that adds standardised income (`income_z`) as an additional predictor. Their results are summarised in Table 4.

***4** Robustness checks for jobs–threat attitudes (IMMJOBS), 2022–2024: multilevel models with and without standardized income.*

Model	Educ within	Educ between	Age (10y)	2024 vs 2022	Income (z)	p (income)
Without income	0.69	0.84	1.10	1.37	—	—
With income	0.69	0.84	1.10	1.36	0.96	0.294

The likelihood-ratio test comparing the two models gives a p-value of about 0.30, meaning that adding income does not significantly improve model fit. Consistent with this, the odds ratio for income itself is close to 1 (OR = 0.96) and not statistically significant. Importantly, the education coefficients remain essentially unchanged when income is added: the within-region education OR stays at 0.69 and the between-region OR at 0.84. The 2024 vs. 2022 contrast also changes only very slightly (from OR = 1.37 to 1.36).

Taken together, these robustness checks suggest that the negative association between education and jobs–threat attitudes is not simply a by-product of income differences. Education appears to have an independent relationship with how Americans think about immigrants and the labour market.

4.3 Economic threat attitudes (IMMAMECO)

4.3.1 Multilevel model, 1996–2024

The second outcome measures whether respondents think that immigrants are "good for America" (IMMAMECO). I recode the item so that the binary indicator `immameco_anti` $= 1$ for respondents who *disagree* or *strongly disagree* that immigrants are good for the American economy (categories 4 and 5), and 0 for those who agree, strongly agree, or are neutral (categories 1–3). Multilevel logistic regression estimates for the pooled waves 1996–2024 are reported in Table 5.

***5** Multilevel logistic regression of economic threat attitudes (IMMAMECO), 1996–2024. Odds ratios with 95% confidence intervals.*

Variable	OR	CI low	CI high	p
Intercept	0.25	0.21	0.30	< 0.001
Education (within region, z)	0.70	0.65	0.75	< 0.001
Education (between region, z)	0.82	0.75	0.90	< 0.001
Year: 1996 (ref. 2014)	2.00	1.66	2.43	< 0.001
Year: 2004 (ref. 2014)	1.37	1.13	1.68	0.00180
Year: 2024 (ref. 2014)	0.86	0.70	1.05	0.14631
Age (per 10 years, z)	1.00	0.96	1.04	0.95589
Black (ref. White)	0.77	0.63	0.94	0.00917
Other race (ref. White)	0.40	0.29	0.55	< 0.001

The education variables show a clear pattern. Both the within-region effect and the between-region effect are below 1. The odds ratio for within-region education is 0.70, meaning that, holding region and other covariates constant, respondents who are one standard deviation above the regional mean in education have about 30% lower odds of thinking that immigrants are bad for the American economy. The between-region effect (OR = 0.82) points in the same direction: living in a region where the average education level is higher is also associated with less negative economic views of immigrants. In other words, both individual education and the educational context are linked to more pro-immigrant economic attitudes.

The year coefficients are estimated relative to 2014. Respondents in 1996 have roughly twice the odds of reporting a negative economic view of immigrants compared to 2014 (OR = 2.00), and those in 2004 also show more negative attitudes (OR = 1.37). By contrast, the odds ratio for 2024 is below 1 but not statistically significant at conventional levels (OR = 0.86, $p \approx 0.15$). This suggests a long-term decline in perceived economic threat from immigrants between the late 1990s and the mid-2010s, with attitudes in 2024 being broadly similar to 2014 on this particular item.

The demographic controls play a more limited role. Age has an odds ratio very close to 1.00, indicating no strong age gradient once education, region, and year are taken into account. Race, however, shows some differences: compared to White respondents, Black respondents have lower odds of saying that immigrants are bad for the economy (OR = 0.77), and respondents in the “other” race category are even less likely to hold negative economic views (OR = 0.40). These patterns are consistent with the idea that majority-group respondents are more likely to perceive immigrants as an economic threat.

4.3.2 Income robustness checks

A natural follow-up question is whether the education effects are simply picking up differences in individual income. To examine this, I estimate robustness models on the subsample with non-missing income, comparing a model without income to a model that adds standardized income (`income_z`) as an extra predictor. The results are summarised in Table 6.

***6** Robustness checks for economic threat attitudes (IMMAMECO), 1996–2024: multilevel models with and without standardized income.*

Model	Educ within	Educ between	Age (10y)	2024 vs 2014	Income (z)	p (income)
Without income	0.68	0.81	0.98	0.88	—	—
With income	0.69	0.81	0.98	0.87	0.98	0.519

Adding income hardly changes the education coefficients. The within-region education odds ratio moves only from 0.68 to 0.69, and the between-region education odds ratio remains at 0.81. The likelihood-ratio test comparing the two models is not significant ($p \approx 0.52$), and the odds ratio for income itself is very close to 1 (OR = 0.98). This suggests that, for economic threat attitudes, the protective effect of education is not simply a proxy for higher income. Instead, education appears to have its own association with more positive views of immigrants' economic impact, over and above differences in individual income.

4.4 Immigration restriction attitudes (LETIN1)

4.4.1 Multilevel model, 1996–2024

The third outcome captures attitudes toward immigration levels using the LETIN1 item. I recode this variable so that the binary indicator `letin_restrict` $= 1$ for respondents who say that the number of immigrants to the United States should be "reduced a little" or "reduced a lot" (categories 4 and 5), and 0 for those who prefer increases or no change (categories 1–3). Multilevel logistic regression estimates for the pooled waves 1996–2024 are shown in Table 7.

***7** Multilevel logistic regression of immigration restriction attitudes (LETIN1), 1996–2024. Odds ratios with 95% confidence intervals.*

Variable	OR	CI low	CI high	p
Intercept	0.87	0.76	1.01	0.06812
Education (within region, z)	0.78	0.74	0.84	< 0.001
Education (between region, z)	0.88	0.81	0.96	0.00376
Year: 1996 (ref. 2014)	2.36	1.98	2.82	< 0.001
Year: 2004 (ref. 2014)	1.72	1.44	2.04	< 0.001
Year: 2024 (ref. 2014)	1.37	1.17	1.62	< 0.001
Age (per 10 years, z)	1.11	1.08	1.15	< 0.001
Black (ref. White)	0.60	0.51	0.72	< 0.001
Other race (ref. White)	0.44	0.35	0.55	< 0.001

As with the other outcomes, education is strongly related to immigration attitudes. The within-region education coefficient has an odds ratio of 0.78, which means that, among respondents living in the same region, those who are one standard deviation above the local mean in education have about 22% lower odds of favouring immigration reductions. The between-region effect is also below one (OR = 0.88), indicating that respondents in more highly educated regions are somewhat less likely to support cutting immigration, even after controlling for their own education level. Together, these results suggest that both individual education and the educational context are associated with more liberal preferences on immigration policy.

The period effects are estimated relative to 2014. All three other years show higher odds of wanting immigration reduced. Respondents in 1996 are more than twice as likely as those in 2014 to support immigration cuts (OR = 2.36), and those in 2004 are also substantially more restrictive (OR = 1.72). Even in 2024, after the long-run decline in threat perceptions documented in the descriptive section, the odds of favouring reductions in immigration remain higher than in 2014 (OR = 1.37). This pattern is consistent with the descriptive evidence that, despite some liberalisation over time, a preference for reduced immigration continues to be a majority position.

Age has a positive association with restrictive attitudes: the odds ratio of 1.11 implies that, for a ten-year increase in age (measured in standardized units), the odds of supporting immigration cuts rise by about 11%, holding other factors constant. Racial differences go in the opposite direction to the main effect of age: compared to White respondents, Black respondents are less likely to support immigration reductions (OR = 0.60), and respondents in the "other" race category are even less restrictive (OR = 0.44). These results suggest that older respondents and White respondents are more inclined toward restrictive immigration policies, while respondents from racial minority groups are more supportive of maintaining or increasing current immigration levels.

4.4.2 Income robustness checks

To test whether the strong education effects are mainly picking up differences in individual income, I again estimate two robustness models on the income-complete subsample: one without income and one with standardized income (`income_z`) added as an extra predictor. The summary of these models is presented in Table 8.

***8** Robustness checks for immigration restriction attitudes (LETIN1), 1996–2024: multilevel models with and without standardized income.*

Model	Educ within	Educ between	Age (10y)	2024 vs 2014	Income (z)	p (income)
Without income	0.78	0.89	1.11	1.33	—	—
With income	0.77	0.89	1.11	1.33	1.05	0.111

The education coefficients are very stable across the two models. The within-region education odds ratio changes only from 0.78 to 0.77, and the between-region odds ratio stays at about 0.89. The likelihood-ratio test comparing the models with and without income yields a p-value of about 0.11, which provides at best weak evidence that adding income improves model fit. The income odds ratio itself is only slightly above 1 (OR = 1.05), indicating a very modest association between higher income and more restrictive immigration preferences.

Overall, these robustness checks show that the negative relationship between education and support for immigration cuts is not driven by income differences. Education remains a consistent predictor of more liberal immigration policy preferences, even when accounting for individual income in the multilevel models.

4.5 Citizenship attachment (AMCITIZN)

4.5.1 Multilevel model, 1996–2024

The last outcome focuses on national identity. I use the AMCITIZN item, which asks respondents whether they agree that they would rather be a citizen of the United States than any other country. I recode this variable so that the binary indicator $\texttt{cit\_attach} = 1$ for respondents who *strongly agree*, and 0 for all weaker responses (agree, neither, disagree, strongly disagree). Multilevel logistic regression estimates for the pooled waves 1996–2024 are reported in Table 9.

***9** Multilevel logistic regression of citizenship attachment (AMCITIZN), 1996–2024. Odds ratios with 95% confidence intervals.*

Variable	OR	CI low	CI high	p
Intercept	7.74	6.42	9.32	<0.001
Education (within region, z)	0.80	0.74	0.87	<0.001
Education (between region, z)	0.84	0.76	0.93	<0.001
Year: 1996 (ref. 2014)	1.78	1.38	2.29	<0.001
Year: 2004 (ref. 2014)	1.60	1.24	2.06	<0.001
Year: 2024 (ref. 2014)	0.38	0.31	0.47	<0.001
Age (per 10 years, z)	1.41	1.34	1.49	<0.001
Black (ref. White)	0.81	0.65	1.01	0.0564
Other race (ref. White)	0.62	0.49	0.79	<0.001

The education effects go in the opposite direction compared to the restriction items. Both the within-region and between-region odds ratios are clearly below one. The within-region effect (OR = 0.80) means that, among respondents in the same region, those who are one standard deviation above the local mean in education have about 20% lower odds of reporting very strong attachment to American citizenship. The between-region effect (OR = 0.84) shows that respondents living in more highly educated regions are also somewhat less likely to say they would *strongly* prefer U.S. citizenship, even after controlling for their own education. Taken together, higher education and more educated regional contexts are associated with weaker, rather than stronger, expressions of national attachment.

The time pattern is also clear when 2014 is used as the reference year. Respondents in 1996 and 2004 are significantly more likely than those in 2014 to report very strong attachment (OR = 1.78 and 1.60, respectively). By contrast, the 2024 odds ratio is far below one (OR = 0.38), indicating that strong citizenship attachment has declined sharply in the most recent wave, even after adjusting for education, age, race, and region. This mirrors the descriptive drop in attachment observed in the raw percentages.

Age has a strong positive association with citizenship attachment. The odds ratio of 1.41 implies that, for a ten-year increase in age, the odds of reporting very strong attachment rise by about 41%, holding other predictors constant. There are also modest racial differences: Black respondents have somewhat lower odds of strong attachment than White respondents (OR = 0.81, marginally above the 5% significance threshold), and respondents in the "other" race category are clearly less likely to express very strong attachment (OR = 0.62). Overall, older respondents and White respondents are more likely to give the strongest pro-citizenship response, while higher education is associated with more moderate or ambivalent answers.

4.5.2 Income robustness checks

As with the immigration attitudes, I test whether the education effects on citizenship attachment can be explained by income differences. Two multilevel models are estimated on the income-complete subsample: a base model without income and a model that adds standardized income (`income_z`) as an extra predictor. The summary of these robustness checks is shown in Table 10.

***10** Robustness checks for citizenship attachment (AMCITIZN), 1996–2024: multilevel models with and without standardized income.*

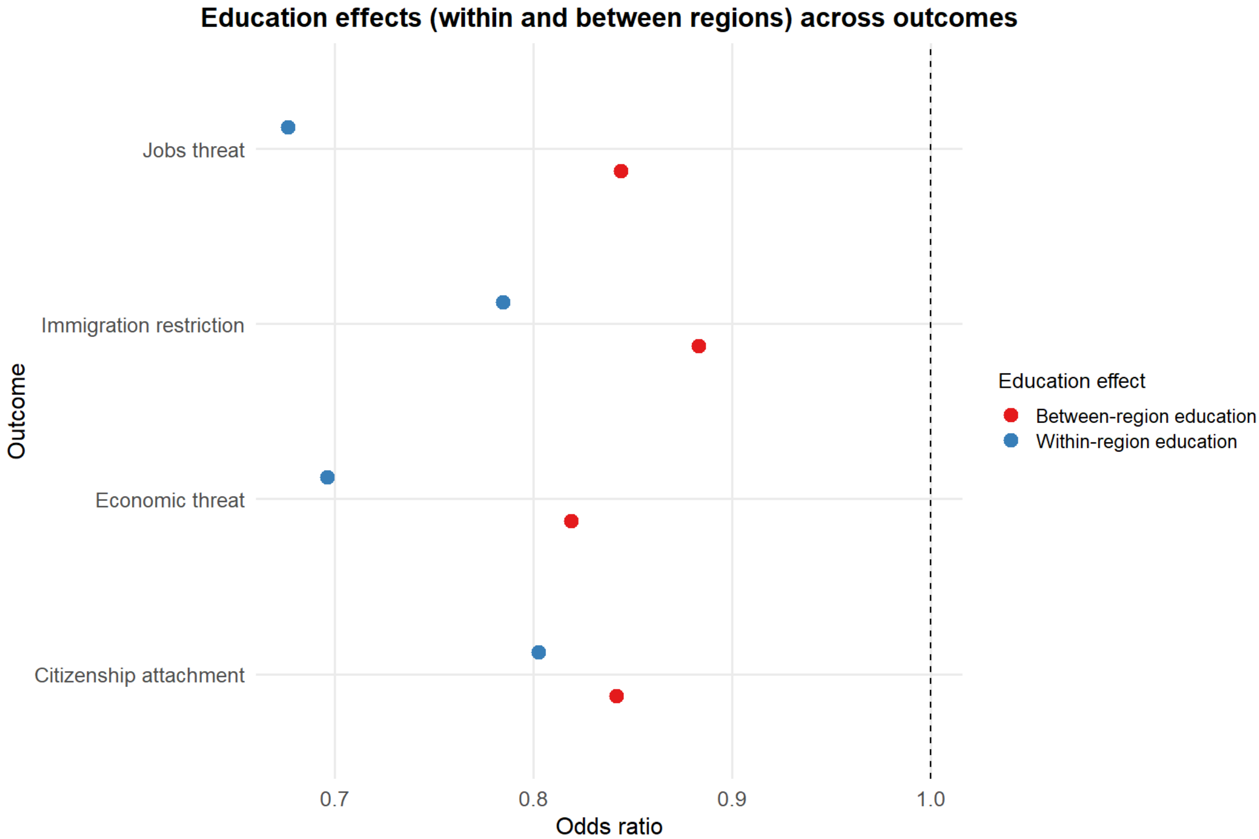
Model	Educ within	Educ between	Age (10y)	2024 vs 2014	Income (z)	p (income)
Without income	0.79	0.85	1.41	0.39	—	—
With income	0.77	0.85	1.40	0.39	1.12	0.00252

The likelihood-ratio test comparing the two models yields a p-value of about 0.003, which means that adding income significantly improves the fit. The income coefficient itself is positive (OR = 1.12), so higher-income respondents are more likely to report very strong attachment to American citizenship, net of education and the other controls.

At the same time, the education effects remain clearly below one. The within-region odds ratio moves only slightly from 0.79 to 0.77 when income is added, and the between-region odds ratio stays at about 0.85. This suggests that both education and income contribute independently to explaining citizenship attachment. Higher education is linked to weaker expressions of strong national attachment, while higher income is linked to stronger attachment, even within the same education and regional context.

4.6 Comparing education and contextual effects across outcomes

To summarise how education relates to different kinds of immigration attitudes, I compare the estimated education effects from the four multilevel models. Figure 4 plots the odds ratios for within–region education (`educ_w`) and between–region education (`educ_b`) for each outcome, using the pooled 1996–2024 models.



***4** Education effects (within and between regions) across outcomes. Odds ratios from multilevel logistic models, 1996–2024.*

The figure shows a consistent pattern: higher education is associated with *lower* odds of expressing negative views about immigrants, both within regions and between regions. For jobs–threat attitudes and economic–threat attitudes (IMMJOBS and IMMAMECO), the within–region odds ratios are around 0.68–0.70, and the between–region odds ratios are around 0.82–0.84. This means that, comparing two respondents in the same region, the more educated person is roughly 30% less likely to say that immigrants take jobs or are bad for the American economy. Respondents living in more highly educated regions are also less negative, even after controlling for their own education, although the contextual effects are somewhat smaller.

For immigration restriction (LETIN1), the pattern is similar but slightly weaker. The within–region odds ratio is about 0.78 and the between–region odds ratio about 0.88. Education therefore still pushes attitudes in a more pro–immigrant direction, but the effect size is smaller than for the economic items. This suggests that people may be somewhat more willing to support lower immigration levels even when they reject simple economic–threat frames.

Citizenship attachment (AMCITIZN) stands out because it is coded in the opposite direction. Here, an odds ratio below 1 means that more educated respondents are *less* likely to say they would strongly rather be U.S. citizens than citizens of any other country. Both the within– and between–region effects (around 0.80 and 0.84) indicate that higher education and more educated regional contexts are linked to weaker expressions of very strong national attachment, even though these same factors are associated with more positive views of immigrants.

Comparing within– and between–region points across outcomes also highlights the contextual dimension of education. In all four models, the within–region effects are slightly stronger (further from 1) than the between–region effects. This implies that individual differences in education matter more for immigration attitudes and citizenship attachment than regional averages do. However, the between–region odds ratios are still clearly below 1, which supports the idea that living in more educated regions is associated with more liberal immigration attitudes and a less exclusive form of national identification, above and beyond one's own years of schooling.

4.7 Machine-learning models (Python)

As a robustness check, I estimated a series of machine–learning models in Python using the same conceptual predictors as in the multilevel models: year, region, race, sex, education, age, and income. The goal was not to build the best possible predictive model, but to verify whether the main patterns from the multilevel analysis also appear in a more flexible, data–driven setting.

4.7.1 Logistic regression baselines

The first step was to fit regularized logistic regression models for the two economic–threat outcomes `immjobs_anti` and `immameco_anti`. The data were split into 75% training and 25% test sets with stratification by outcome, and class weights were balanced to account for outcome imbalance. Table 11 summarizes the main performance metrics on the test set.

11 *Predictive performance of Python logistic regression models for jobs and economic threat outcomes (test set).*

Outcome	N	Majority acc.	Logit acc.	Balanced acc.	AUC
Jobs threat (`immjobs_anti`)	7,223	0.666	0.646	0.643	0.697
Economic threat (`immameco_anti`)	4,673	0.775	0.606	0.596	0.643

For jobs–threat attitudes, the logistic model reaches an AUC of 0.70 and a balanced accuracy of about 0.64, modestly improving on the majority–class baseline accuracy of 0.67. For economic–threat attitudes, the model still provides some discrimination (AUC ≈ 0.64), but the overall accuracy gain over the majority classifier is small because the outcome is more imbalanced (only about 23% of respondents express a negative economic view). These results suggest that although socio–demographic and contextual variables structure attitudes in systematic ways, individual–level prediction remains difficult.

4.7.2 XGBoost performance and feature importance

In a second step, I trained gradient–boosted tree models (XGBoost) for the same two outcomes. The models used the one–hot encoded versions of year, region, race, and sex, alongside continuous measures of age, education, and income. Hyperparameters followed a conservative setting (around 2,000 trees, small learning rate, shallow depth) with early stopping based on validation loss.

While the full test–set accuracy and AUC values were not stored in the output log, the XG-Boost models performed similarly to the logistic baselines rather than dramatically outperforming them. The main added value of XGBoost is therefore in its non–parametric representation of feature importance rather than in overall predictive power.

The gain–based feature importance scores show a consistent pattern across both outcomes. Education and regional dummy variables are among the most influential predictors, often followed by specific region–year combinations and, to a lesser extent, race and age. Income plays some role but is not among the very top predictors. This ranking closely mirrors the substantive conclusions from the multilevel models, where within– and between–region education and census region capture much of the structure in attitudes toward immigrants.

4.7.3 Relation to the multilevel results

Overall, the machine–learning models reinforce the main findings from the multilevel analysis rather than challenging them. First, their moderate AUC values confirm that attitudes are not trivially predictable from basic socio–demographic information, which cautions against overly deterministic interpretations. Second, the feature importance patterns highlight education and regional context as central dimensions, in line with the strong within– and between–region education effects in the multilevel models. Finally, the fact that more flexible tree–based models do not uncover radically different or stronger predictors increases confidence that the regression results capture the key relationships in the data.

5 Discussion

This section brings together the main empirical findings and relates them to the research questions, previous studies, and data–science practice. It also notes key limitations and possible next steps.

5.1 Summary of main findings

The first goal was to understand how individual education relates to attitudes toward immigrants and citizenship. Across all four outcomes, education is a strong predictor. Within regions, more educated respondents are less likely to see immigrants as taking jobs away, harming the economy, or needing stricter limits (odds ratios around 0.68–0.78). At the same time, they are also less likely to say that they would much rather be a U.S. citizen than a citizen of any other country (OR ≈ 0.80).

The second goal was to test whether regional education levels matter in addition to individual education. The between–region component also points in a liberal direction: people living in more educated regions are less negative toward immigrants and somewhat less likely to express very strong citizenship attachment, even when their own education is taken into account. These contextual effects are smaller than the individual ones, but they are consistent across all outcomes.

Over time, the results show a clear decline in anti–immigrant views between 1996 and 2014, followed by a partial reversal in the early 2020s. In the descriptive figures and tables, perceived jobs and economic threat and support for immigration limits fall from 1996 to 2014, while very strong citizenship attachment stays high. However, By 2024, jobs-threat and restriction attitudes rise again, and strong attachment drops. The multilevel models confirm this pattern: compared with 2014, the odds of negative immigration attitudes are higher in 1996 and 2004, but lower or closer to 1 in 2022 and 2024, whereas strong citizenship attachment is highest in 1996–2004 and clearly lower in 2024.

Other predictors play a smaller role. Older respondents are more likely to favour restriction and report strong citizenship attachment, but age is not clearly related to economic threat. Respondents coded as “other” race (and to a lesser extent Black respondents) are less likely than White respondents to hold anti–immigrant attitudes and less likely to express very strong attachment. Income is surprisingly weak: it barely changes the education effects for immigration attitudes and only improves model fit for citizenship attachment, where higher income goes together with stronger attachment.

The machine–learning models in Python echo these results. Their overall predictive performance is modest, which shows that basic socio–demographic variables alone cannot fully explain attitudes. However, feature importance measures consistently rank education and regional indicators among the most important predictors, giving an independent check on the multilevel findings.

5.2 Implications for data science practice

From a data–science point of view, the project shows the value of combining multilevel models with machine–learning tools. Multilevel logistic regression is well suited for clustered survey data and provides clear, interpretable estimates for individual and contextual effects. Machine learning models, used here mainly as a robustness check, confirm which predictors matter the most and show that even flexible algorithms cannot predict individual attitudes very precisely from basic socio–demographics. In other words, the most useful insights come from interpretable models, while machine learning acts as a secondary tool for validation rather than the main way of drawing conclusions.

5.3 Limitations and future work

Several limitations should be noted. First, data are repeated cross–sections, not a panel, so changes over time can reflect differences between cohorts, as well as genuine change of opinion. Second, the contextual level is fairly coarse (nine census regions); more detailed state or local information and richer contextual variables (such as immigrant share or local policy) could change some of the patterns. Third, the outcomes were dichotomized for clarity, which hides distinctions between moderate and extreme positions. Finally, the set of predictors is limited to socio–demographic and regional variables; party identification, ideology, and contact with immigrants are likely to explain additional variation.

Future work could therefore (a) add political and contact variables to test how much of the education and region effects they absorb, (b) use finer–grained geographic units and contextual data, (c) model the full ordinal response scales, and (d) explore more advanced but still interpretable machine–learning methods, such as generalized additive models or tree–based models with clear explanation tools. These extensions would help to clarify when education and regional context matter most and how they interact with the broader political and social environment.

6 Conclusion

6.1 Summary of the study

This thesis examined how education and the regional context relate to American attitudes toward immigrants and citizenship between 1996 and 2024, using data from the General Social Survey. Four outcomes were analysed: perceived jobs threat, perceived economic threat, support for restricting immigration, and strong attachment to American citizenship.

The empirical strategy combined multilevel logistic regression with simple machine–learning models. Education was split into an individual (within–region) and contextual (between–region) component, allowing the analysis to distinguish personal effects from the broader environment in which the respondents live. Robustness checks with income and alternative models were used to test how stable these relationships are.

6.2 Answers to the research questions

First, the results show that education is strongly associated with immigration attitudes. Within every region, more educated respondents are less likely to see immigrants as taking jobs, harming the economy, or needing stricter limits. Living in a more educated region points in the same direction, although contextual effects are smaller than individual ones. These patterns are stable across all waves.

Second, the picture of citizenship attachment is more complex. Higher education, both individual and contextual, is linked to a lower probability of reporting a very strong attachment to American citizenship. In other words, education goes together with more positive views of immigrants, but also with a slightly less exclusive form of national identity. Income explains only a small part of this pattern: adding income improves the citizenship model, but the education effects remain clearly below one.

Third, attitudes have not moved in a straight line over time. Between 1996 and 2014, perceived threat and support for restriction declined, while strong citizenship attachment remains high. By 2024, jobs–threat and restriction concerns rise again and strong attachment drops. Education and regional context continue to matter in all periods, but they operate against this shifting time background.

The machine–learning models confirm the same general message. They do not predict individual attitudes with very high accuracy, but repeatedly highlight education and regional indicators among the most important predictors, which supports the multilevel findings.

6.3 Implications

Substantively, the results suggest that education shapes both how people think about immigrants and how they relate to national identity. More educated individuals, especially in more educated regions, are less likely to see immigrants as a threat. At the same time, they are less likely to endorse the strongest statements of national attachment. This points to a form of "critical patriotism" or a more cosmopolitan identity among highly educated Americans.

For data science, the project illustrates the value of combining multilevel models and machine–learning tools. Multilevel models provide interpretable estimates for individual and contextual effects in clustered survey data, while machine–learning models offer a complementary way to check which predictors matter most. In this setting, the clearest insights come from interpretable models, with machine learning used as a secondary diagnostic tool rather than the main source of conclusions.

6.4 Final remarks

Like any single study, this thesis has limits in terms of variables, geographic detail, and research design. Still, it shows that education and regional context help explain why some Americans see immigrants as a threat while others do not, and why strong national attachment has changed over the last three decades. The findings also open the door to future work that combines richer contextual data, political variables, and more advanced but explainable models to better understand how social structure and place shape attitudes toward immigration.

References

[1] Anaconda, Inc. *Python and Jupyter Notebook via Anaconda Distribution*. Python scripts were run in Jupyter Notebook through the Anaconda platform. 2024. URL: `https://www.anaconda.com`.

[2] K. Bansak, J. Hainmueller, D. Hangartner. “Immigrant Contexts During Childhood and Support for Redistribution.” In: *European Sociological Review* 38.2 (2022), pages 169–183.

[3] R. K. Brown, R. E. Brown. “Race, Religion, and Immigration Policy Attitudes.” In: *Race and Social Problems* 9.1 (2017), pages 4–18.

[4] C. R. Chandler, Y.-m. Tsai. “Social Factors Influencing Immigration Attitudes: An Analysis of Data from the General Social Survey.” In: *The Social Science Journal* 38.2 (2001), pages 177–188.

[5] M. P. Erdmans, A. Smell. “Explaining Attitudes toward Immigrants in the United States: Context, Class and Ideology.” In: *Studia Migracyjne – Przegląd Polonijny* 3.(173) (2019), pages 69–91.

[6] V. M. Esses, S. Medianu, A. S. Lawson. “Uncertainty, Threat, and the Role of the Media in Promoting the Dehumanization of Immigrants and Refugees.” In: *Journal of Social Issues* 69.3 (2013), pages 518–536.

[7] K. Fennelly, C. M. Federico. “Rural Residence as a Determinant of Attitudes Toward U.S. Immigration Policy.” In: *International Migration* 46.1 (2008), pages 151–190.

[8] C. García Coll, A. K. Marks, editors. *The Immigrant Paradox in Children and Adolescents: Is Becoming American a Developmental Risk?* Washington, DC: American Psychological Association, 2013.

[9] J. Hainmueller, M. J. Hiscox. “Attitudes toward Highly Skilled and Low-skilled Immigration: Evidence from a Survey Experiment.” In: *American Political Science Review* 104.1 (2010), pages 61–84.

[10] J. Haubert, E. Fussell. “Explaining Pro-Immigrant Sentiment in the U.S.: Social Class, Cosmopolitanism, and Perceptions of Immigrants.” In: *International Migration Review* 40.3 (2006), pages 489–507.

[11] A. Kustov, D. Laaker, C. Reller. “The Stability of Immigration Attitudes: Evidence and Implications.” In: *Journal of Politics* 83.4 (2021), pages 1478–1495.

[12] B. Lancee, O. Sarrasin. “Does Education Change Anti-immigrant Sentiment? Evidence from a Multi-cohort Longitudinal Study.” In: *European Sociological Review* 38.2 (2023), pages 169–183.

[13] NORC at the University of Chicago. *General Social Survey Cumulative Data File, 1972–2024*. The General Social Survey (GSS) is a project of the independent research organization NORC at the University of Chicago, with principal funding from the National Science Foundation. 2024.

[14] Posit Software, PBC. *RStudio: Integrated Development Environment for R*. Version 2025.09.1 Build 401, ”Cucumberleaf Sunflower”. 2025. URL: `https://posit.co`.

Appendix 1. Code availability

All code used in this thesis is openly available in a public GitHub repository:

- **Repository:** `https://github.com/jubudubu/Master-Thesis`

The repository contains two main files:

R code

R scripts were developed and executed using RStudio (Version 2025.09.1; Posit Software, PBC), a graphical interface for R. [14]

The file `R code final.R` includes all R code used in the empirical analysis:

- Loading and cleaning the GSS data.
- Construction of variables (recoding outcomes, computing within- and between-region education, standardising predictors).
- Estimation of multilevel logistic regression models for all outcomes (IMMJOBS, IMMAMECO, LETIN1, AMCITIZN), including income robustness checks.
- Creation of descriptive tables and model summary tables exported to LaTeX.
- Generation of all figures used in the thesis (trend plots, histograms, and education–effect plots).

Python code

Python code was written and executed using Jupyter Notebook provided by the Anaconda distribution (Anaconda, Inc.). [1]

The file `code Python.txt` contains the Python scripts used for the machine–learning analysis:

- Importing the processed data exported from R.
- Training and evaluating logistic regression and gradient boosting (XGBoost) models.
- Computing performance metrics (such as accuracy and AUC) for the different model specifications.
- Extracting feature importance measures used to compare with the multilevel model results.

Appendix 2. Use of AI tools

During the work on this thesis, the following AI tools were consulted: ChatGPT, Perplexity, and Scite.

Their role was supportive and practical (helping improve and troubleshoot), not to produce the thesis independently. In this project, AI assistance was applied mainly in the following ways:

- **Writing support:** improving English clarity and readability, rephrasing sentences and paragraphs.
- **Translation:** translating short fragments from Lithuanian and Polish into English, followed by manual review and editing.
- **LaTeX / Overleaf help:** resolving formatting issues and technical problems in Overleaf, such as making tables fit the page, splitting long column headers, selecting appropriate packages, and adjusting table spacing and layout. Using R package: kable, helped with generating ready to use tables for LaTex from outputs in R.
- **Programming support:** assistance with code-related tasks such as understanding error messages, improving code readability, and suggesting possible approaches for analysis scripts (with final implementation, testing, and validation done manually).
- **Literature support:** help with locating and checking relevant academic sources and improving reference-related details (e.g., confirming publication information and identifying related work).