Integrated Report: Immigration Policy Attitudes

Ann Jiang

2025-03-17

# Introduction

This report analyzes anti-immigration attitudes using the NSL2018 dataset. The document describes the data preparation steps, presents descriptive analyses (including frequency tables and group comparisons), reports multiple regression models, and displays several visualizations. The goal is to examine how nativity and immigrant generation are associated with anti-immigration attitudes.

# Data Preparation

In this section, we load the required packages, set the working directory, read the dataset, and recode key variables.

# Load required packages  
library(dplyr)  
library(ggplot2)  
library(scales)  
library(broom) # for tidying regression output  
  
# Set working directory and read dataset (update the path accordingly)  
setwd("/Users/azure/Desktop/Coursework/PhD Coursework/1.2 Winter 2025/1.2.5 Survey Analysis--Anti-Immigration Immigrant/Pew-Research-Center\_2018-National-Survey-of-Latinos-Dataset 2")  
nsl\_data <- read.csv("NSL2018\_data.csv", stringsAsFactors = FALSE)

## Recoding Variables

### Nativity

We recode qn4 so that respondents are flagged as foreign-born (1) if their answer is “another country” (case-insensitive), and U.S.-born (0) otherwise.

nsl\_data <- nsl\_data %>%  
 mutate(foreign\_born = ifelse(tolower(trimws(qn4)) == "another country", 1, 0))

### Generation Status

Based on the immgen variable, we recode generation status as follows:  
- “First Gen FB w/o PR” → “First Gen”  
- “Second Gen NB” → “Second Gen”  
- “Third or more” → “Third+ Gen”  
Other responses become NA.

nsl\_data <- nsl\_data %>%  
 mutate(generation = case\_when(  
 tolower(trimws(immgen)) == "first gen fb w/o pr" ~ "First Gen",  
 tolower(trimws(immgen)) == "second gen nb" ~ "Second Gen",  
 tolower(trimws(immgen)) == "third or more" ~ "Third+ Gen",  
 TRUE ~ NA\_character\_  
 ))  
nsl\_data$generation <- factor(nsl\_data$generation, levels = c("First Gen", "Second Gen", "Third+ Gen"))

### Immigration Policy Indicators

We recode three items: - **qn14a (Trump Support):** 1 if “approve”, 0 if “disapprove”. - **qn29 (Border Wall Favorability):** 1 if “favor”, 0 if “oppose”. - **qn31 (Opinion on Immigrant Population):** 1 if “too many”, 0 if “too few” or “right amount”.

nsl\_data <- nsl\_data %>%  
 mutate(trump\_support\_bin = case\_when(  
 tolower(trimws(qn14a)) == "approve" ~ 1,  
 tolower(trimws(qn14a)) == "disapprove" ~ 0,  
 TRUE ~ NA\_real\_  
 ),  
 favor\_wall = case\_when(  
 tolower(trimws(qn29)) == "favor" ~ 1,  
 tolower(trimws(qn29)) == "oppose" ~ 0,  
 TRUE ~ NA\_real\_  
 ),  
 too\_many\_imm = case\_when(  
 tolower(trimws(qn31)) == "too many" ~ 1,  
 tolower(trimws(qn31)) %in% c("too few", "right amount") ~ 0,  
 TRUE ~ NA\_real\_  
 ))

## Composite Variable

We create the composite immigration policy score by summing the three binary indicators. The result is then categorized as a factor with levels 0–3.

nsl\_data <- nsl\_data %>%  
 mutate(immigration\_policy = trump\_support\_bin + favor\_wall + too\_many\_imm,  
 immigration\_policy\_cat = factor(immigration\_policy, levels = 0:3, ordered = TRUE))  
# Check the composite variable summary  
summary(nsl\_data$immigration\_policy)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.0000 0.0000 0.0000 0.7738 1.0000 3.0000 356

# Descriptive Analyses

In this section, we output summaries, frequency tables, group comparisons, and statistical tests.

## Independent Variables

### Nativity

cat("Nativity (foreign\_born):\n")

## Nativity (foreign\_born):

print(table(nsl\_data$foreign\_born, useNA = "ifany"))

##   
## 0 1   
## 745 756

print(summary(nsl\_data$foreign\_born))

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 1.0000 0.5037 1.0000 1.0000

### Generation Status

cat("Generation Status (generation):\n")

## Generation Status (generation):

print(table(nsl\_data$generation, useNA = "ifany"))

##   
## First Gen Second Gen Third+ Gen <NA>   
## 759 318 405 19

print(summary(nsl\_data$generation))

## First Gen Second Gen Third+ Gen NA's   
## 759 318 405 19

## Dependent Variable

### Composite Immigration Policy

cat("Composite Immigration Policy (immigration\_policy):\n")

## Composite Immigration Policy (immigration\_policy):

print(summary(nsl\_data$immigration\_policy))

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.0000 0.0000 0.0000 0.7738 1.0000 3.0000 356

cat("\nFrequency Table for immigration\_policy\_cat:\n")

##   
## Frequency Table for immigration\_policy\_cat:

print(table(nsl\_data$immigration\_policy\_cat, useNA = "ifany"))

##   
## 0 1 2 3 <NA>   
## 611 278 160 96 356

cat("\nExplanation: NA values in 'immigration\_policy\_cat' indicate that one or more of the component indicators is missing.\n")

##   
## Explanation: NA values in 'immigration\_policy\_cat' indicate that one or more of the component indicators is missing.

## Group Comparisons

### Mean & SD by Nativity

mean\_by\_nativity <- nsl\_data %>%  
 group\_by(foreign\_born) %>%  
 summarise(mean\_composite = mean(immigration\_policy, na.rm = TRUE),  
 sd\_composite = sd(immigration\_policy, na.rm = TRUE),  
 n = n())  
cat("Mean and SD of immigration\_policy by foreign\_born:\n")

## Mean and SD of immigration\_policy by foreign\_born:

print(mean\_by\_nativity)

## # A tibble: 2 × 4  
## foreign\_born mean\_composite sd\_composite n  
## <dbl> <dbl> <dbl> <int>  
## 1 0 0.896 1.09 745  
## 2 1 0.641 0.820 756

### T-Test: Foreign-born vs. U.S.-born

t\_test\_result <- t.test(immigration\_policy ~ foreign\_born, data = nsl\_data)  
cat("T-test for immigration\_policy by foreign\_born:\n")

## T-test for immigration\_policy by foreign\_born:

print(t\_test\_result)

##   
## Welch Two Sample t-test  
##   
## data: immigration\_policy by foreign\_born  
## t = 4.4868, df = 1099.1, p-value = 7.989e-06  
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0  
## 95 percent confidence interval:  
## 0.1433785 0.3662363  
## sample estimates:  
## mean in group 0 mean in group 1   
## 0.8959732 0.6411658

### Cross-Tabulations and Chi-Square Tests

cat("Cross-tabulation: Foreign-born vs. immigration\_policy\_cat:\n")

## Cross-tabulation: Foreign-born vs. immigration\_policy\_cat:

cross\_tab <- table(nsl\_data$foreign\_born, nsl\_data$immigration\_policy\_cat)  
print(cross\_tab)

##   
## 0 1 2 3  
## 0 312 111 96 77  
## 1 299 167 64 19

cat("\nChi-square Test (Foreign-born vs. immigration\_policy\_cat):\n")

##   
## Chi-square Test (Foreign-born vs. immigration\_policy\_cat):

chi\_sq\_result <- chisq.test(cross\_tab)  
print(chi\_sq\_result)

##   
## Pearson's Chi-squared test  
##   
## data: cross\_tab  
## X-squared = 51.156, df = 3, p-value = 4.532e-11

cat("\nCross-tabulation: Generation vs. immigration\_policy\_cat:\n")

##   
## Cross-tabulation: Generation vs. immigration\_policy\_cat:

cross\_tab\_gen <- table(nsl\_data$generation, nsl\_data$immigration\_policy\_cat)  
print(cross\_tab\_gen)

##   
## 0 1 2 3  
## First Gen 301 167 64 20  
## Second Gen 145 56 39 22  
## Third+ Gen 159 54 54 52

cat("\nChi-square Test (Generation vs. immigration\_policy\_cat):\n")

##   
## Chi-square Test (Generation vs. immigration\_policy\_cat):

chi\_sq\_gen <- chisq.test(cross\_tab\_gen)  
print(chi\_sq\_gen)

##   
## Pearson's Chi-squared test  
##   
## data: cross\_tab\_gen  
## X-squared = 60.654, df = 6, p-value = 3.314e-11

# Regression Analyses

We run three models: one with nativity only, one with generation among U.S.-born respondents, and one using a combined categorical variable.

## Separate Models

### Model 1: Immigration Policy ~ Nativity

model\_nat <- lm(immigration\_policy ~ foreign\_born, data = nsl\_data)  
cat("Regression Model: Immigration Policy ~ Nativity\n")

## Regression Model: Immigration Policy ~ Nativity

print(summary(model\_nat))

##   
## Call:  
## lm(formula = immigration\_policy ~ foreign\_born, data = nsl\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.8960 -0.8960 -0.6412 0.3588 2.3588   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.89597 0.03978 22.525 < 2e-16 \*\*\*  
## foreign\_born -0.25481 0.05744 -4.436 1.01e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9711 on 1143 degrees of freedom  
## (356 observations deleted due to missingness)  
## Multiple R-squared: 0.01692, Adjusted R-squared: 0.01606   
## F-statistic: 19.68 on 1 and 1143 DF, p-value: 1.006e-05

### Model 2: Immigration Policy ~ Generation (U.S.-born only)

nsl\_usborn <- subset(nsl\_data, foreign\_born == 0)  
model\_gen <- lm(immigration\_policy ~ generation, data = nsl\_usborn)  
cat("Regression Model: Immigration Policy ~ Generation (U.S.-born only)\n")

## Regression Model: Immigration Policy ~ Generation (U.S.-born only)

print(summary(model\_gen))

##   
## Call:  
## lm(formula = immigration\_policy ~ generation, data = nsl\_usborn)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.0000 -0.9969 -0.7634 1.0031 2.2366   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1.000000 0.626589 1.596 0.111  
## generationSecond Gen -0.236641 0.630166 -0.376 0.707  
## generationThird+ Gen -0.003135 0.629528 -0.005 0.996  
##   
## Residual standard error: 1.085 on 581 degrees of freedom  
## (161 observations deleted due to missingness)  
## Multiple R-squared: 0.01138, Adjusted R-squared: 0.007979   
## F-statistic: 3.345 on 2 and 581 DF, p-value: 0.03596

## Combined Categorical Variable

We create a combined variable that assigns “Foreign-born (First Gen)” if foreign\_born == 1, and uses the generation value for U.S.-born respondents.

nsl\_data <- nsl\_data %>%  
 mutate(nativity\_generation = ifelse(foreign\_born == 1, "Foreign-born (First Gen)", as.character(generation)))  
nsl\_data$nativity\_generation <- factor(nsl\_data$nativity\_generation, levels = c("Foreign-born (First Gen)", "Second Gen", "Third+ Gen"))  
cat("Combined Nativity/Generation Variable Distribution:\n")

## Combined Nativity/Generation Variable Distribution:

print(table(nsl\_data$nativity\_generation, useNA = "ifany"))

##   
## Foreign-born (First Gen) Second Gen Third+ Gen   
## 756 318 405   
## <NA>   
## 22

model\_combined <- lm(immigration\_policy ~ nativity\_generation, data = nsl\_data)  
cat("Regression Model: Immigration Policy ~ Combined Nativity/Generation\n")

## Regression Model: Immigration Policy ~ Combined Nativity/Generation

print(summary(model\_combined))

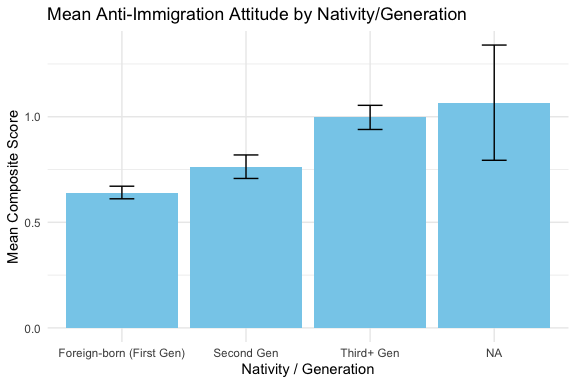
##   
## Call:  
## lm(formula = immigration\_policy ~ nativity\_generation, data = nsl\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.9969 -0.7634 -0.6412 0.3588 2.3588   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.64117 0.04113 15.589 < 2e-16 \*\*\*  
## nativity\_generationSecond Gen 0.12219 0.07236 1.689 0.0916 .   
## nativity\_generationThird+ Gen 0.35570 0.06784 5.243 1.89e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9637 on 1127 degrees of freedom  
## (371 observations deleted due to missingness)  
## Multiple R-squared: 0.02382, Adjusted R-squared: 0.02209   
## F-statistic: 13.75 on 2 and 1127 DF, p-value: 1.257e-06

# Visualizations

Below are two types of visualizations: a bar chart of mean composite scores and a stacked bar chart of the distribution of composite scores.

## Bar Chart: Mean Composite Score by Nativity/Generation

mean\_data <- nsl\_data %>%  
 group\_by(nativity\_generation) %>%  
 summarise(mean\_policy = mean(immigration\_policy, na.rm = TRUE),  
 se\_policy = sd(immigration\_policy, na.rm = TRUE) / sqrt(n()),  
 n = n())  
  
ggplot(mean\_data, aes(x = nativity\_generation, y = mean\_policy)) +  
 geom\_bar(stat = "identity", fill = "skyblue") +  
 geom\_errorbar(aes(ymin = mean\_policy - se\_policy, ymax = mean\_policy + se\_policy), width = 0.2) +  
 labs(title = "Mean Anti-Immigration Attitude by Nativity/Generation",  
 x = "Nativity / Generation",  
 y = "Mean Composite Score") +  
 theme\_minimal()



## Stacked Bar Chart: Distribution of Composite Scores

We calculate percentages in the natural order (0 to 3) and then plot a stacked bar chart.

dist\_data <- nsl\_data %>%  
 filter(!is.na(nativity\_generation), !is.na(immigration\_policy\_cat)) %>%  
 group\_by(nativity\_generation, immigration\_policy\_cat) %>%  
 summarise(count = n(), .groups = "drop") %>%  
 group\_by(nativity\_generation) %>%  
 mutate(  
 percentage = count / sum(count) \* 100,  
 label\_pos = cumsum(percentage) - 0.5 \* percentage,  
 sample\_size = sum(count)  
 ) %>%  
 arrange(nativity\_generation, immigration\_policy\_cat)  
  
my\_palette <- c(  
 "0" = "#F0E442", # Yellow  
 "1" = "#009E73", # Green  
 "2" = "#56B4E9", # Blue  
 "3" = "#E69F00" # Orange  
)  
  
ggplot(dist\_data, aes(x = nativity\_generation, y = percentage, fill = immigration\_policy\_cat)) +  
 geom\_bar(stat = "identity", width = 0.7, position = "stack") +  
 geom\_text(aes(y = label\_pos, label = sprintf("%.1f%%", percentage)),  
 color = "black", size = 3.5,  
 data = subset(dist\_data, percentage > 7)) +  
 scale\_fill\_manual(  
 values = my\_palette,  
 name = "Anti-Immigration\nAttitude Score",  
 labels = c("0 (Low)", "1", "2", "3 (High)")  
 ) +  
 labs(  
 title = "Distribution of Anti-Immigration Attitudes by Nativity/Generation",  
 subtitle = "Score=0 is at the bottom, Score=3 is at the top",  
 x = "",  
 y = "Percentage of Respondents",  
 caption = "Groups: Foreign-born (First Gen), Second Gen, Third+ Gen"  
 ) +  
 scale\_y\_continuous(limits = c(0, 100), labels = percent\_format(scale = 1)) +  
 theme\_minimal() +  
 theme(  
 legend.position = "right",  
 plot.title = element\_text(face = "bold", size = 14),  
 plot.subtitle = element\_text(size = 11, color = "#636363"),  
 plot.caption = element\_text(size = 9, color = "#636363", hjust = 0),  
 axis.text = element\_text(size = 10),  
 axis.title.y = element\_text(size = 11),  
 panel.grid.minor = element\_blank()  
 ) +  
 geom\_text(aes(y = -5, label = paste0("n=", sample\_size)),  
 size = 3, color = "#636363")

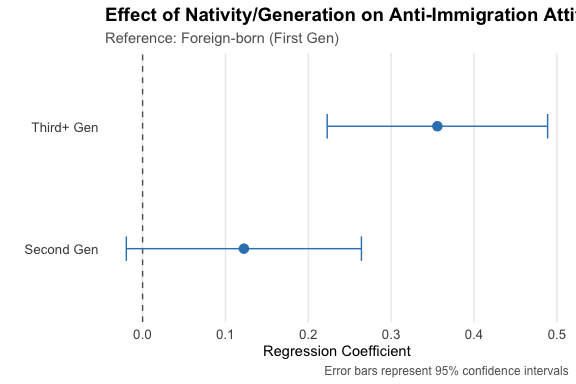
A chart with different colored squares

Description automatically generated

## Regression Coefficient Plot

We visualize the coefficients from the combined categorical model.

coef\_data <- tidy(model\_combined)  
# Exclude the intercept for clarity  
coef\_data <- coef\_data[coef\_data$term != "(Intercept)", ]  
# Clean term labels  
coef\_data$term <- gsub("nativity\_generation", "", coef\_data$term)  
  
ggplot(coef\_data, aes(x = estimate, y = reorder(term, estimate))) +  
 geom\_vline(xintercept = 0, linetype = "dashed", color = "#636363") +  
 geom\_point(size = 3, color = "#3182bd") +  
 geom\_errorbarh(aes(xmin = estimate - 1.96 \* std.error, xmax = estimate + 1.96 \* std.error),  
 height = 0.2, color = "#3182bd") +  
 labs(  
 title = "Effect of Nativity/Generation on Anti-Immigration Attitudes",  
 subtitle = "Reference: Foreign-born (First Gen)",  
 x = "Regression Coefficient",  
 y = "",  
 caption = "Error bars represent 95% confidence intervals"  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_text(face = "bold", size = 14),  
 plot.subtitle = element\_text(size = 11, color = "#636363"),  
 plot.caption = element\_text(size = 9, color = "#636363"),  
 axis.title.x = element\_text(size = 11),  
 axis.text = element\_text(size = 10),  
 panel.grid.minor = element\_blank(),  
 panel.grid.major.y = element\_blank()  
 )



# Conclusion

This report has provided a comprehensive overview of the analysis of anti-immigration attitudes: - **Descriptive statistics** were generated for key independent and dependent variables. - **Group comparisons** (means, t-tests, cross-tabulations, and chi-square tests) were performed. - **Regression models** were run using separate and combined approaches to examine the influence of nativity and generation. - **Visualizations** including bar charts, stacked bar charts, and a regression coefficient plot are provided for clear interpretation of the results.

==========================================================================

# End of Report

```