

# STAT107\_Project: Education Attainment vs Income Inequality

Alyssia Zhan, Liam Renehan, Sanjana Rayala

2025-12-05

---

## Abstract

This project investigates the relationship between national education attainment and wealth inequality across countries. Using publicly available datasets containing primary education completion rates and Gini coefficients, we assess whether higher education levels are associated with more equitable economic outcomes. After cleaning and merging the datasets to create a unified set of 144 countries, we used statistical modeling and data visualization to characterize the association between education and inequality. Our analysis reveals a statistically significant negative relationship, indicating that countries with higher primary education attainment tend to exhibit lower levels of income inequality. Although the explanatory power of education alone is limited, the observed trend suggests that expanding educational access may play a meaningful role in reducing global wealth disparities. These findings highlight the importance of education as a potential policy lever for promoting economic equity.

## Introduction

The central goal of this project is to examine whether higher levels of educational attainment are associated with reductions in national wealth inequality. Inequality, commonly measured using the Gini coefficient, remains a major global concern, and understanding its structural determinants is essential for developing effective policy. Education is often considered an important contributor to economic opportunity because it enhances workforce skills, increases earning potential, and broadens access to employment. However, the extent to which education correlates with national income inequality across countries remains an empirical question. By analyzing primary education completion rates alongside Gini coefficients from a wide range of nations, this project seeks to quantify the relationship between these two variables and identify broader trends. The analysis also highlights countries that represent extreme cases of equality or inequality, helping contextualize the broader global landscape.

## Data

Two datasets were used in this study: one containing national primary education completion rates for adults age 25 and older, and another reporting each country's Gini coefficient as a measure of wealth inequality. After importing both datasets, several preprocessing steps were performed to ensure consistency and accuracy. These steps included selecting the relevant variables, converting the education metrics to numeric format, removing missing or incomplete observations, and merging the datasets by country name. The resulting combined dataset consists of 144 countries and includes three key variables: country name, primary education attainment, and the Gini coefficient. This unified dataset provided the foundation for visualization, correlation analysis, and regression modeling, enabling a comprehensive examination of the relationship between education and economic inequality across nations.

## Visualization

### Scatterplot: Education vs Wealth Inequality

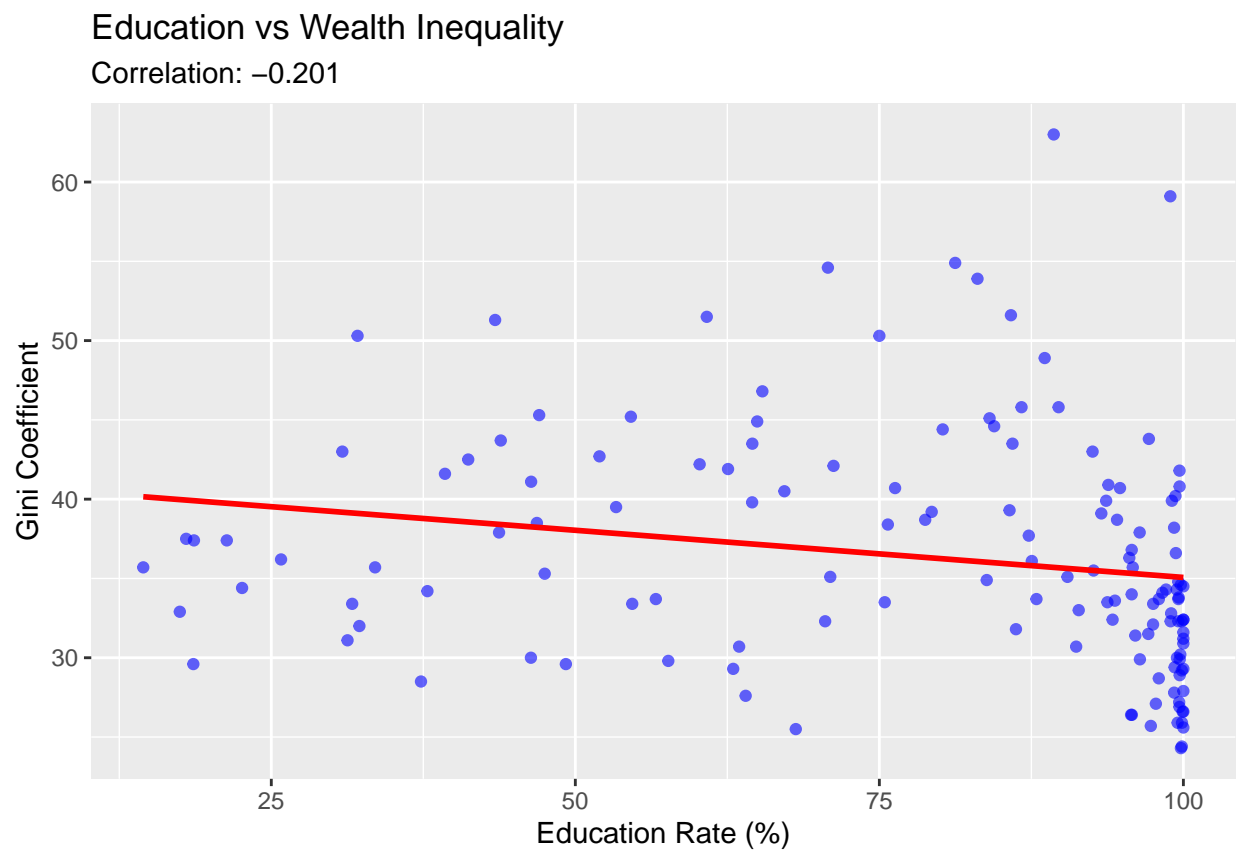
```
library(ggplot2)

data <- combined_data

# Calculate correlation
cor_value <- cor(data$Education_Attainment, data$Gini_Coefficient)

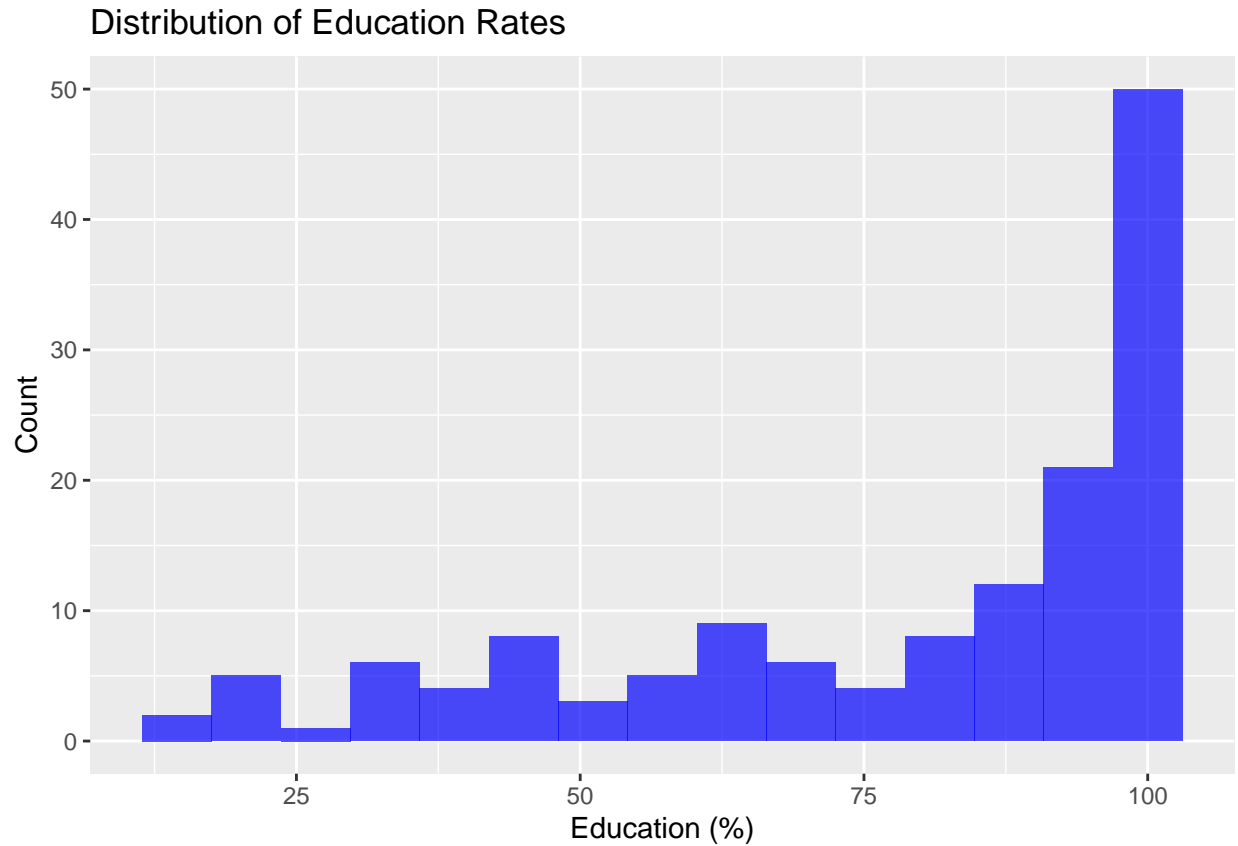
# Scatterplot with regression line
ggplot(data, aes(x = Education_Attainment, y = Gini_Coefficient)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Education vs Wealth Inequality",
       subtitle = paste("Correlation:", round(cor_value, 3)),
       x = "Education Rate (%)",
       y = "Gini Coefficient")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



### Histogram: Education Rates

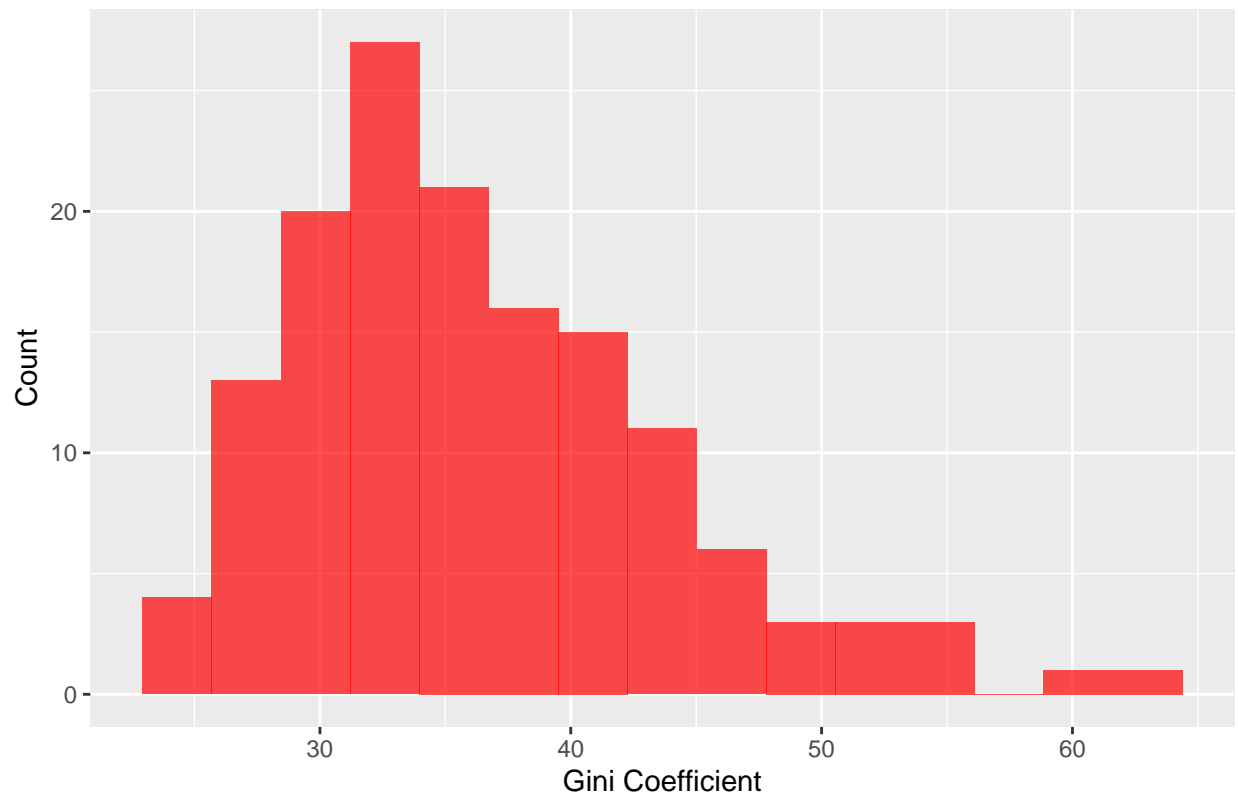
```
ggplot(data, aes(x = Education_Attainment)) +  
  geom_histogram(fill = "blue", alpha = 0.7, bins = 15) +  
  labs(title = "Distribution of Education Rates", x = "Education (%)", y = "Count")
```



### Histogram: Wealth Inequality

```
ggplot(data, aes(x = Gini_Coefficient)) +  
  geom_histogram(fill = "red", alpha = 0.7, bins = 15) +  
  labs(title = "Distribution of Wealth Inequality", x = "Gini Coefficient", y = "Count")
```

## Distribution of Wealth Inequality



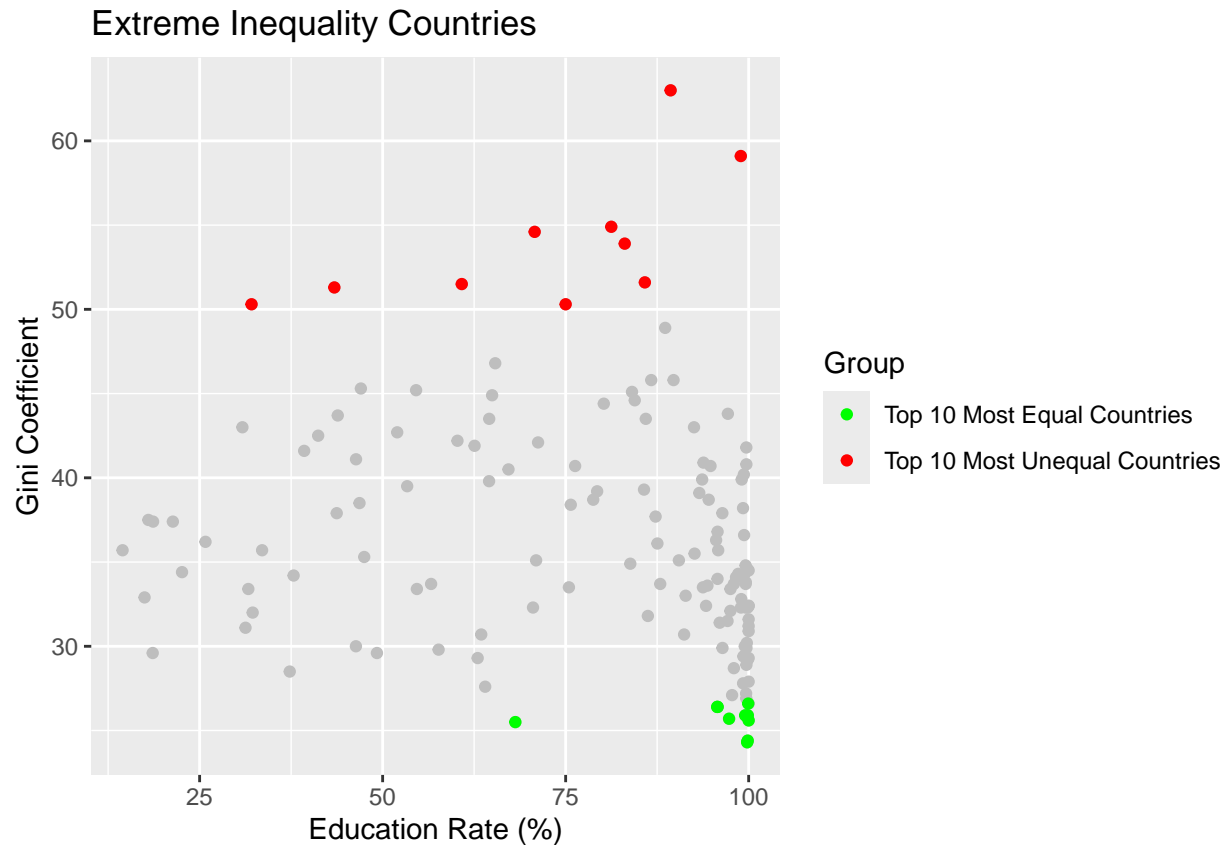
## Extreme Inequality Countries

```
# Identify top 10 most unequal and top 10 most equal countries
high <- data[order(-data$Gini_Coefficient), ][1:10, ]
low <- data[order(data$Gini_Coefficient), ][1:10, ]

high$Group <- "Top 10 Most Unequal Countries"
low$Group <- "Top 10 Most Equal Countries"

extreme <- rbind(high, low)

# Plot extreme inequality countries
ggplot() +
  geom_point(data = data, aes(x = Education_Attainment, y = Gini_Coefficient), color = "gray") +
  geom_point(data = extreme, aes(x = Education_Attainment, y = Gini_Coefficient, color = Group)) +
  scale_color_manual(values = c("Top 10 Most Unequal Countries" = "red", "Top 10 Most Equal Countries" = "gray")) +
  labs(title = "Extreme Inequality Countries", x = "Education Rate (%)", y = "Gini Coefficient")
```



## Analysis

We began by visually exploring the relationship between education attainment and wealth inequality using scatterplots, histograms, and summary statistics. The scatterplot, paired with a fitted regression line, suggests a modest negative association between primary education completion and the Gini coefficient. A correlation analysis confirms this pattern, revealing a coefficient of approximately  $-0.20$ , which indicates a weak but meaningful inverse relationship. To further quantify this association, we fitted a linear regression model with Gini coefficient as the outcome and education attainment as the predictor. The slope of the model is statistically significant and negative, implying that increases in education attainment are associated with decreases in inequality. However, the model explains only a small portion of the overall variance, as reflected by an R-squared value of about 0.04. Despite this, the pattern is strengthened by examining extreme countries: those with the highest inequality tend to have lower educational attainment, while countries with the lowest inequality generally exhibit higher completion rates. These results reinforce the overall trend, suggesting that education is one of several factors shaping national inequality levels.

```
lm_model <- lm(Gini_Coefficient ~ Education_Attainment, data = combined_data)
summary(lm_model)
```

```
##
## Call:
## lm(formula = Gini_Coefficient ~ Education_Attainment, data = combined_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -11.460  -5.117  -1.436   4.284  27.302
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    41.01273    1.99250  20.584  <2e-16 ***
## Education_Attainment -0.05949    0.02427  -2.451   0.0155 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.32 on 142 degrees of freedom
## Multiple R-squared:  0.0406, Adjusted R-squared:  0.03384
## F-statistic: 6.008 on 1 and 142 DF,  p-value: 0.01545
```

## Conclusions

This analysis demonstrates a statistically significant negative relationship between primary education attainment and national wealth inequality. Countries with higher rates of primary education completion tend to achieve more equitable income distributions, while those with lower attainment levels experience greater disparities. Although education alone explains only a small fraction of the variation in inequality, its consistent association with lower Gini coefficients underscores its importance as part of broader economic and social policy strategies. Expanding educational access may therefore contribute to more equitable economic outcomes, especially when combined with other structural reforms. Future research could extend this work by incorporating secondary and tertiary education data, examining additional socioeconomic predictors, or using causal modeling approaches to better understand the mechanisms linking education and inequality.