

FinalReport

2025-11-08

```
wealth_data <- read.csv(file.choose())

education_data <- read.csv(file.choose())

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.3.3

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##       filter, lag

## The following objects are masked from 'package:base':
##       intersect, setdiff, setequal, union

education_clean <- education_data %>%
  select(
    Country = Economy,
    Education_Attainment = Educational.attainment..at.least.completed.primary..population.25..years..to.
  ) %>%
  mutate(
    Education_Attainment = as.numeric(Education_Attainment)
  ) %>%
  filter(!is.na(Education_Attainment))

wealth_clean <- wealth_data %>%
  select(
    Country = 2,
    Gini_Coefficient = 3
  )

combined_data <- education_clean %>%
  inner_join(wealth_clean, by = "Country") %>%
  arrange(Country)
cat("Structure of combined dataset:\n")

## Structure of combined dataset:
```

```

str(combined_data)

## 'data.frame':   144 obs. of  3 variables:
##   $ Country           : chr  "Albania" "Algeria" "Angola" "Armenia" ...
##   $ Education_Attainment: num  99.3 64 43.4 99.7 99.5 ...
##   $ Gini_Coefficient    : num  29.4 27.6 51.3 27.2 34.3 30.9 26.6 33.4 24.4 26.4 ...

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3

data <- read.csv("combined_data_simple.csv")

cor_value <- cor(data$Education_Attainment, data$Gini_Coefficient)

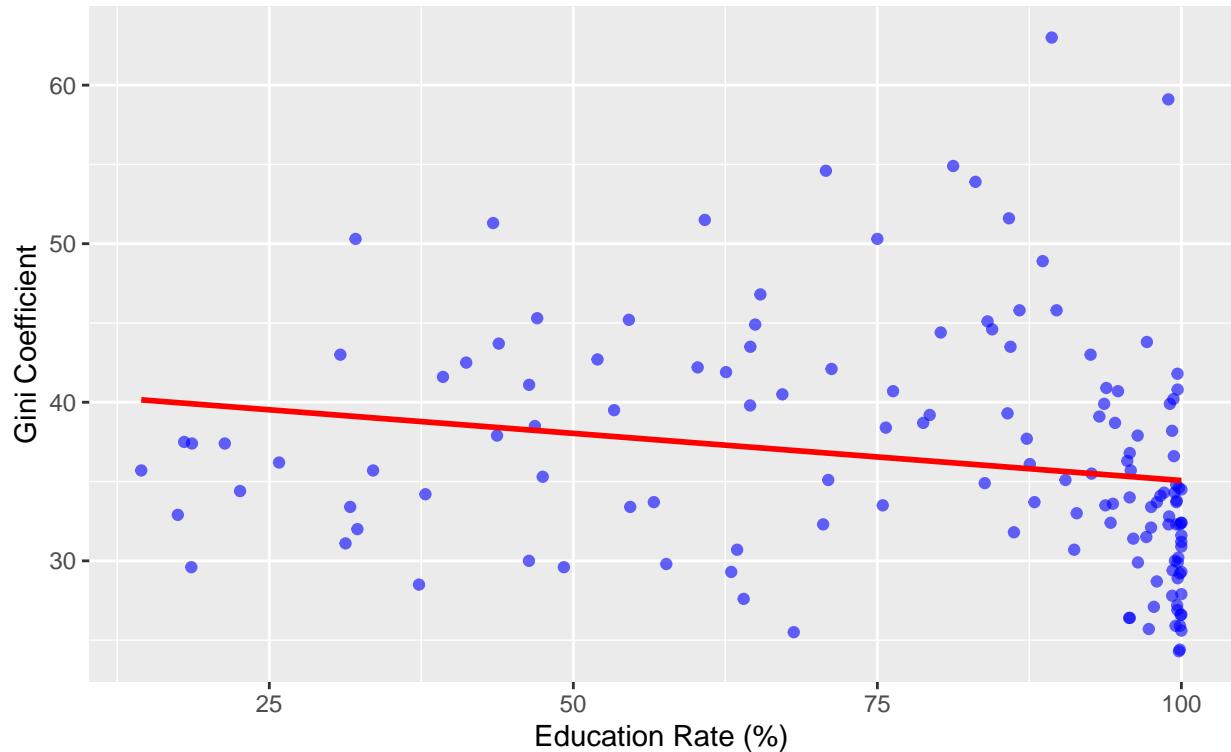
ggplot(data, aes(x = Education_Attainment, y = Gini_Coefficient)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Education vs Wealth Inequality",
       subtitle = paste("Correlation:", round(cor_value, 3)),
       x = "Education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

```

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

Education vs Wealth Inequality

Correlation: -0.201



```

high <- data[order(-data$Gini_Coefficient), ][1:10, ]
low <- data[order(data$Gini_Coefficient), ][1:10, ]

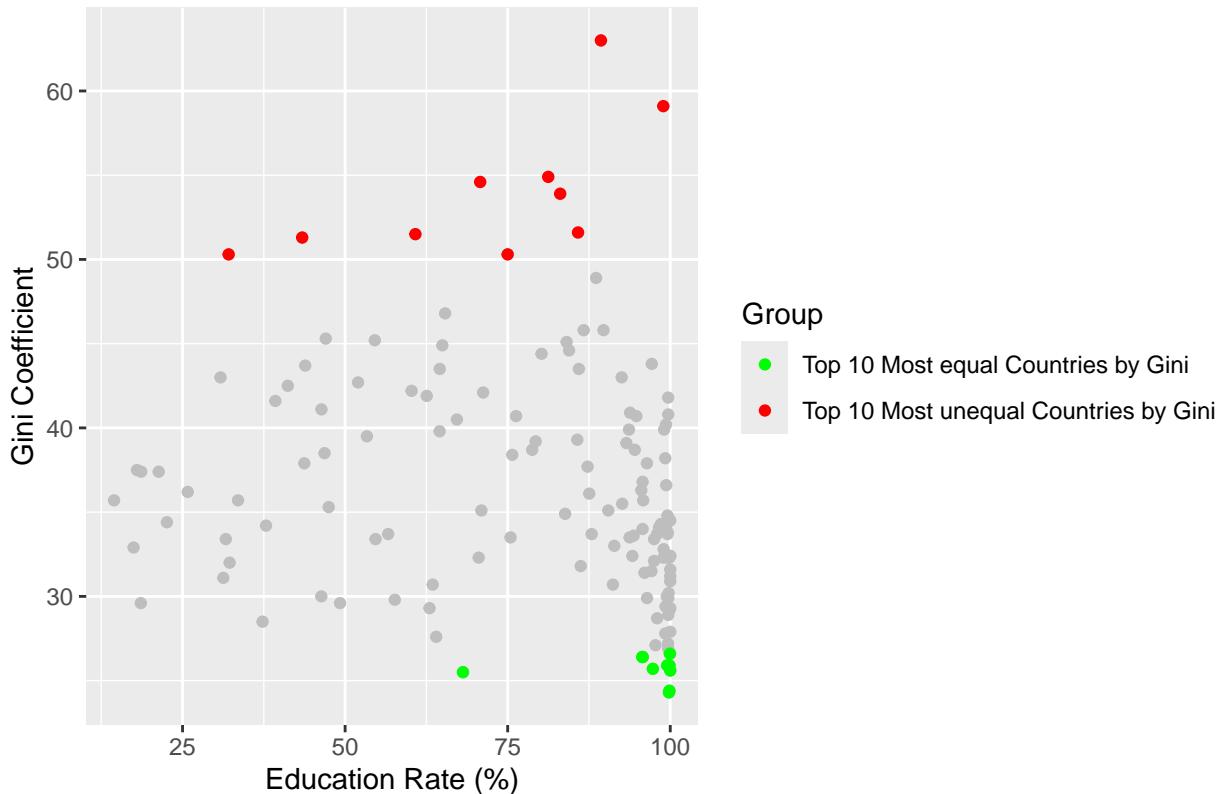
high$Group <- "Top 10 Most unequal Countries by Gini"
low$Group <- "Top 10 Most equal Countries by Gini"

extreme <- rbind(high, low)

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 by Gini" = "red", "Top 10 Most equal Countries by Gini" = "green")) +
  labs(title = "Extreme Inequality Countries",
       x = "Education Rate (%)",
       y = "Gini Coefficient")

```

Extreme Inequality Countries



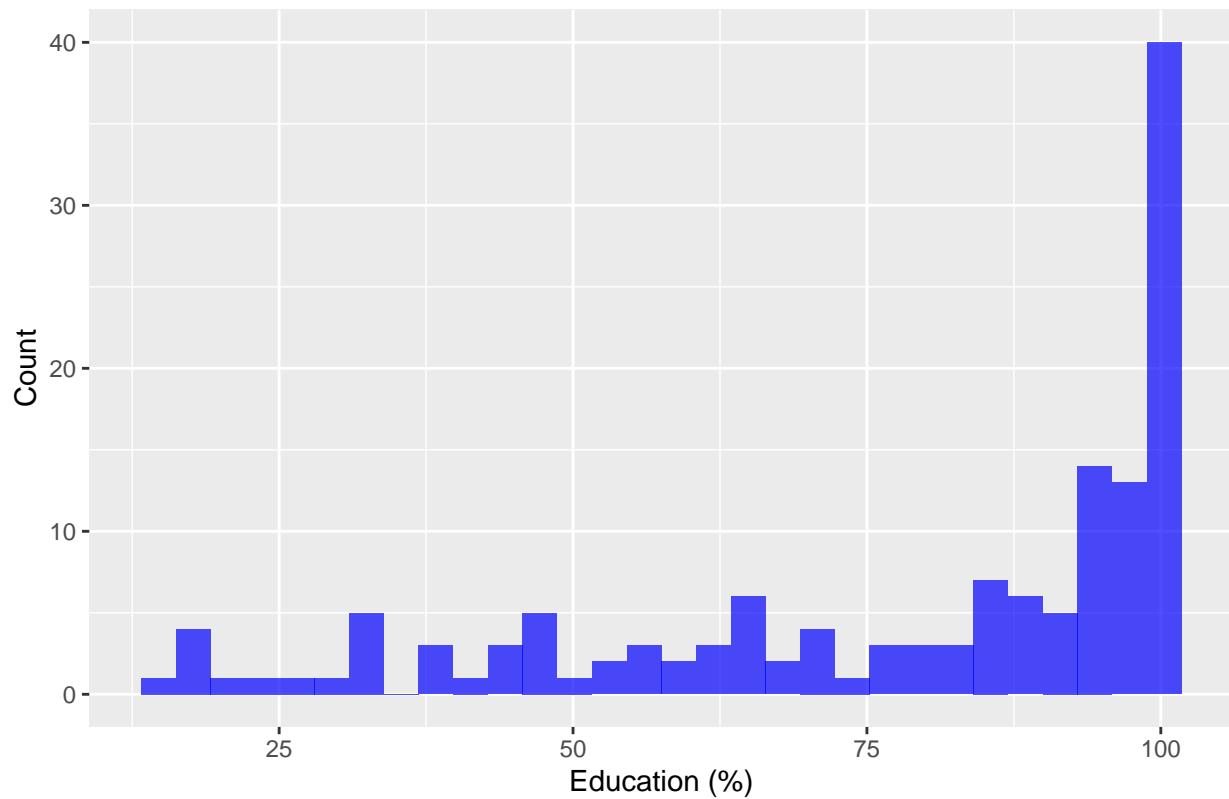
```

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

```

'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

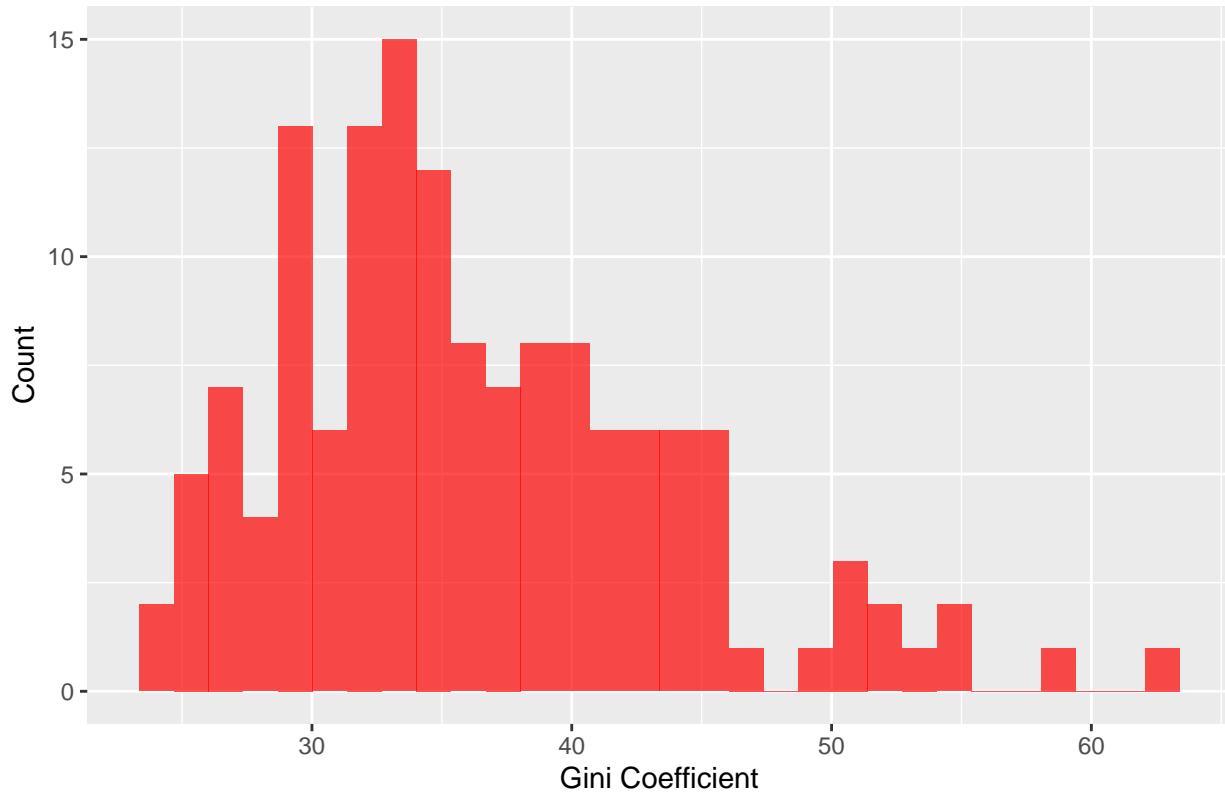
Education Rates



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

'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

Wealth Inequality



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.