**GLOBAL EDUCATION LANDSCAPE ANALYSIS: EXPLORING SOCIO-ECONOMIC INFLUENCES ON EDUCATIONAL METRICS.**

CODE AND REPORT.

Table of Contents

[**INTRODUCTION** 3](#_Toc155539853)

[**LITERATURE REVIEW** 4](#_Toc155539854)

[**DATA EXPLORATION** 8](#_Toc155539855)

[**EXPERIMENTS** 12](#_Toc155539856)

[**Impact of Socio-Economic Status on Completion Rates** 12](#_Toc155539857)

[**Regional Disparities in Educational Metrics** 13](#_Toc155539858)

[**Socio-Economic Influences on Proficiency Levels** 14](#_Toc155539859)

[**RESULTS** 16](#_Toc155539860)

[**Results for Experiment 2: Regional Disparities in Educational Metrics** 18](#_Toc155539861)

[**Results for Experiment 3: Socio-Economic Influences on Proficiency Levels** 20](#_Toc155539862)

[**DISCUSSION, CONCLUSIONS, AND FUTURE WORK** 22](#_Toc155539863)

[**REFLECTING ON PROFESSIONAL, ETHICAL, AND LEGAL ISSUES IN DATASET USAGE** 25](#_Toc155539864)

[**REFERENCES** 27](#_Toc155539865)

**Abstract**

This research delves into the intricate interplay between socio-economic factors and global educational metrics. Leveraging a meticulously curated dataset, the study explores completion rates, proficiency levels, and regional disparities, shedding light on the multifaceted nature of educational outcomes. Ethical considerations guide the handling of sensitive information, emphasizing privacy and legal compliance. Results reveal significant associations, with higher unemployment and birth rates positively impacting completion rates. Regional analyses highlight variations, while proficiency studies identify math proficiency as a significant predictor. The research contributes nuanced insights for shaping equitable and inclusive education policies worldwide.

# INTRODUCTION

Education stands as a linchpin in the intricate fabric of societal progress, serving as the bedrock upon which human capital, economic prosperity, and global development rest. The transformative power of education is clear, influencing individual destinies, shaping collective futures and bridging the gaps that persist in diverse communities worldwide. This research embarks on an ambitious journey, leveraging a meticulously curated dataset that unfolds a comprehensive panorama of education on a global scale. As Battle and Lewis (2002) assert, education is not merely a conduit for information; it is an indispensable factor intertwined with an individual's well-being and the prosperity of nations.

The profound impact of education on human capital, as elucidated by human capital theory, underscores its pivotal role in fostering economic growth and development. The stock of skills and productive knowledge encapsulated within individuals, often referred to as human capital, becomes a driving force behind a nation's prosperity (Ross, 2021). This research aspires to delve into the intricate interplay between education and economic prosperity by harnessing a meticulously curated dataset.

The dataset, a repository of crucial metrics ranging from out-of-school rates and completion rates to proficiency levels and literacy rates, unfolds as a treasure trove for discerning researchers, dedicated educators, and forward-thinking policymakers. The inclusion of geographic coordinates adds a spatial dimension, enabling a nuanced exploration of global educational trends. This research seeks to harness the wealth of information within this dataset to embark on a transformative journey of assessing, enhancing, and reshaping education systems worldwide.

Drawing inspiration from successful research endeavors, such as Farooq et al.'s (2011) examination of factors affecting students' academic performance in Pakistan and Toutkoushian and Curtis (2005) study on the socio-economic factors in public high schools in New Hampshire, this research aspires to contribute rigorously assessed insights into the global education landscape to provide nuanced insights that can catalyze transformative measures. In alignment with well-established practices, with a focus on disparities and trends that can inform future educational policies and practices.

# LITERATURE REVIEW

In the dynamic context of the global education landscape, an in-depth exploration of the intricate relationship between socio-economic factors and educational metrics is essential. The socio-economic status (SES) of individuals, encompassing components like income, education, occupation, and wealth, is recognized as a crucial determinant shaping educational outcomes on a global scale (American Psychological Association, 2017).

SES is a multidimensional construct that provides a composite measure of an individual's or household's relative position within the social hierarchy. The inclusion of income, education, occupation, and wealth in SES allows for a nuanced examination of the multifaceted ways in which economic factors intersect with educational metrics (Worthy et al., 2020).

Research conducted across diverse global contexts consistently affirms the profound impact of SES on educational metrics. The influence of SES extends beyond national borders, with higher SES individuals often benefiting from improved access to quality education, resulting in enhanced academic performance (Perry & Mcconney, 2010). Disparities in income, reflected in school funding and resource allocation, contribute to an educational divide that persists on a global scale (Darling-Hammond, 2019).

The nexus between SES, health outcomes, and educational success is a crucial dimension in understanding the broader implications of socio-economic influences. Higher SES individuals typically exhibit better health, attributed to improved access to healthcare, healthier lifestyles, and reduced stress levels (Munir et al., 2023). This intricate connection underscores the interdependence of health and educational metrics within the socio-economic context.

The exploration of socio-economic influences on educational metrics cannot be divorced from the broader issue of societal inequality. Lower SES often results in limited social mobility, disseminating disparities within societies (Garcia & Weiss, 2017). The "achievement gap," observable in academic performance across different socio-economic strata, is indicative of the enduring challenge of creating an inclusive and equitable global education landscape (Burger, 2019).

Acknowledging the impact of socio-economic influences on educational metrics necessitates targeted policy interventions on a global scale (“Policy Implications,” 2010). Policymakers must focus on improving access to quality education, reducing disparities in educational opportunities, and promoting socio-economic mobility (Williams & Reppond, 2020). Initiatives addressing income redistribution, affordable housing, and equitable healthcare are integral to tackling the root causes of educational disparities (Baciu et al., 2019).

While socio-economic influences present challenges, they also provide opportunities for transformative change (Herrfahrdt-Pähle et al., 2020). Global collaboration, sharing best practices, and implementing evidence-based interventions are crucial for fostering a more equitable education landscape (Haleem et al., 2022). Leveraging technological advancements for inclusive learning and addressing the unique socio-economic challenges of various regions are essential components of a comprehensive strategy (United Nations, 2021).

The manifestation of socio-economic disparities in education is not confined to a singular domain but permeates various facets of the global education landscape. The impact of SES on academic achievement is evident in the intricate interplay between economic factors and educational outcomes. Research by Sirin (2005) underscores a robust correlation between family income and academic performance, where students from higher SES backgrounds consistently outperform their counterparts from lower SES backgrounds.

As educational indicators take center stage, the importance of adopting rigorous methodologies becomes evident. Aldowah et al. (2019) emphasize the diverse array of methods employed to scrutinize complex educational data. The incorporation of statistical methods, ranging from regression analysis to group-based modeling, provides researchers with invaluable tools to discern patterns and predict outcomes within the intricate landscape of education (Yu et al., 2018). Additionally, Franconeri et al. (2021) highlight the significance of data visualization methods, such as heat maps and graphs, in rendering complex educational data accessible to a broad audience. The synthesis of these methodologies contributes to a comprehensive understanding of the intricate relationship between socio-economic influences and educational metrics.

A closer examination of global educational disparities necessitates an exploration of regional variations and their underlying causes. Lea Lund and Kirk (2019) delve into the disparities in education levels across different regions, shedding light on the diverse factors contributing to divergent educational outcomes. The nuanced understanding of regional contexts is imperative for devising targeted interventions that address the unique challenges faced by distinct communities (Hanushek, 2020). By unraveling the complex web of factors influencing global education, research endeavors to pave the way for more inclusive and tailored approaches to educational policy and practice (Dunlop, 2020).

Gender-based analyses add another layer of complexity to the exploration of educational metrics. Unterhalter (2019) delves into the significance of gender-based measurements, such as completion rates and reading levels, in elucidating disparities between males and females in educational achievements. The work of Monkman (2021) underscores the pervasive impact of societal norms and unfair structures on education, emphasizing the need for gender-sensitive research to inform policies that rectify gender-based imbalances in educational outcomes. Such studies contribute to a broader understanding of how socio-economic factors intersect with gender dynamics in shaping educational metrics (Cole, 2022).

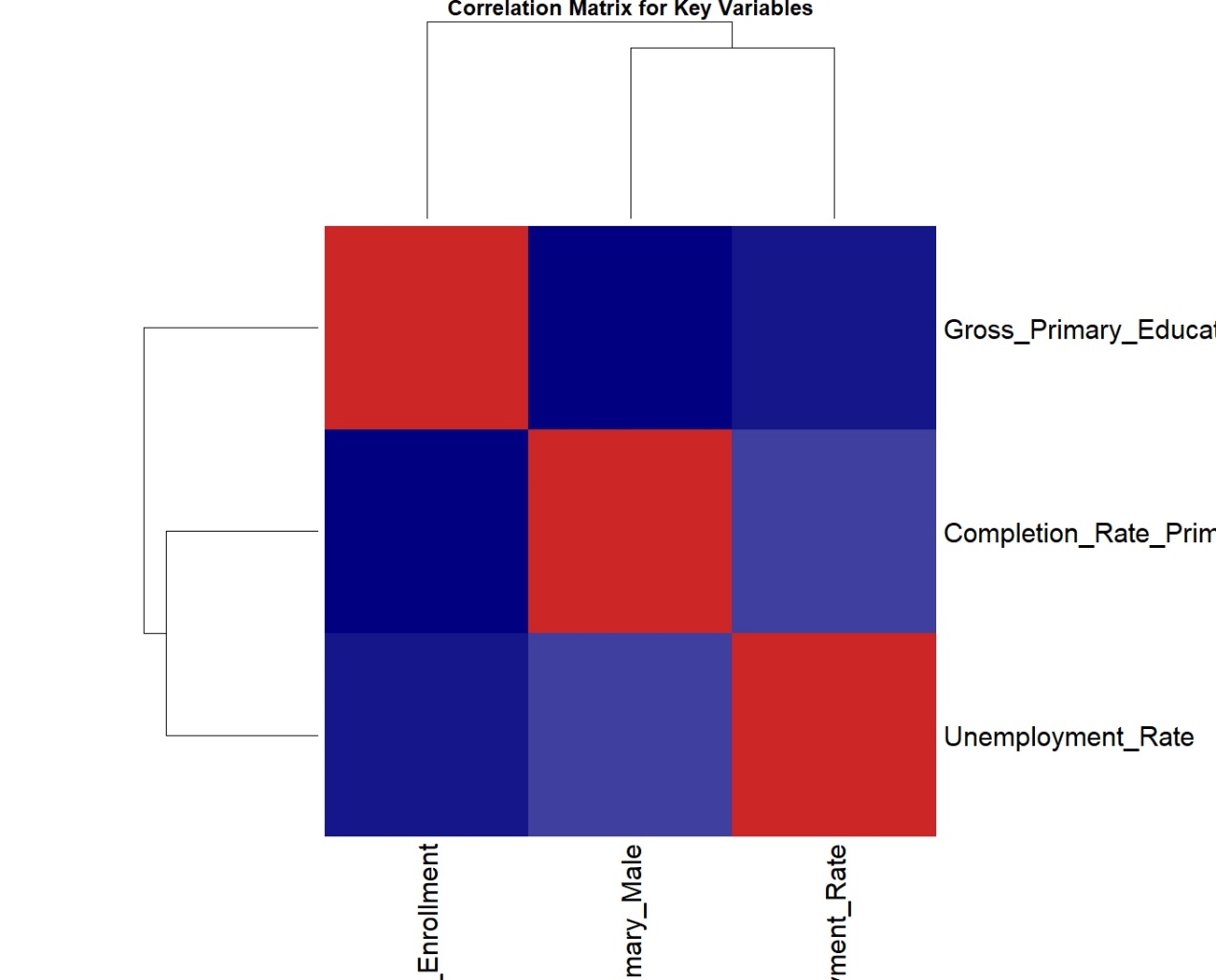
The advent of technology in education is a transformative force that warrants careful examination. Szymkowiak et al. (2021) delineate the profound changes brought about by data analysis and machine learning, elucidating how these advancements reshape the understanding of educational dynamics. Pettersson's (2020) exploration of technological tools emphasizes their role in comprehending the evolving nature of educational landscapes over time. The intersection of technology and education introduces novel dimensions, enabling educators and policymakers to glean insights into complex patterns and make informed decisions (Lantz, 2019).

In conclusion, the exploration of socio-economic influences on educational metrics traverses a multifaceted terrain that demands a comprehensive and globally informed perspective. As the global education landscape evolves, research endeavors continue to illuminate the intricate interplay of socio-economic factors, providing valuable insights for shaping policies and practices that foster a more equitable and inclusive educational environment.

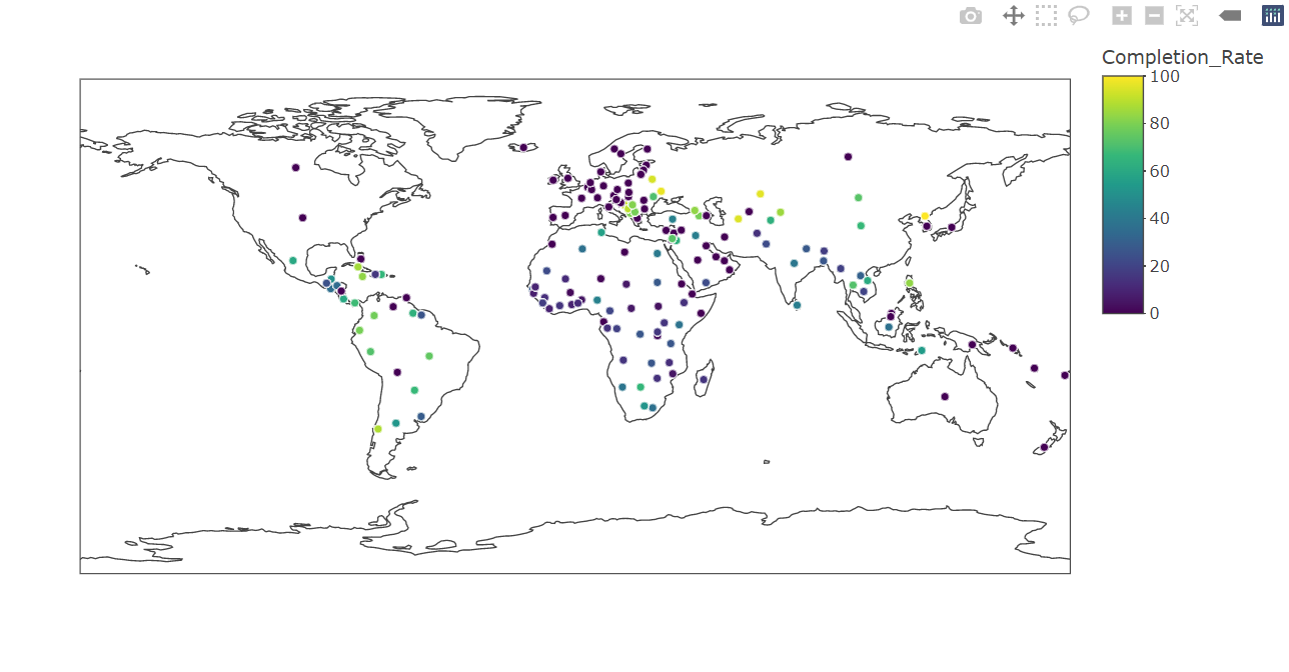
# DATA EXPLORATION

In the pursuit of comprehending the intricate relationship between socio-economic influences and educational metrics within the global education landscape, a meticulous exploration of the dataset was undertaken. The primary objective was to unveil the underlying structure and nuances encapsulated within the "Global Education" dataset, laying a foundational understanding for dissecting socio-economic dynamics.

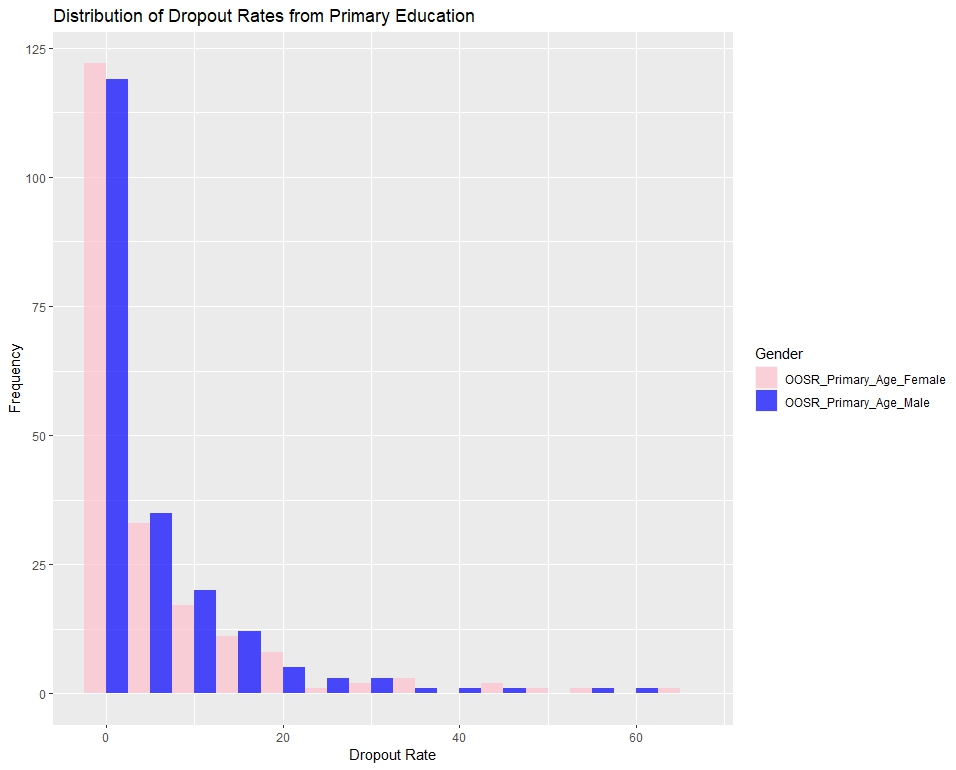
The str() function was instrumental in dissecting the dataset's structure, providing a detailed breakdown of variable types. This step served as the cornerstone, enabling subsequent analyses to be grounded in a nuanced appreciation of the socio-economic and educational components. Following this, the summary() function illuminated central tendencies and variations, offering a snapshot of key variables that would play a pivotal role in shaping the exploration.



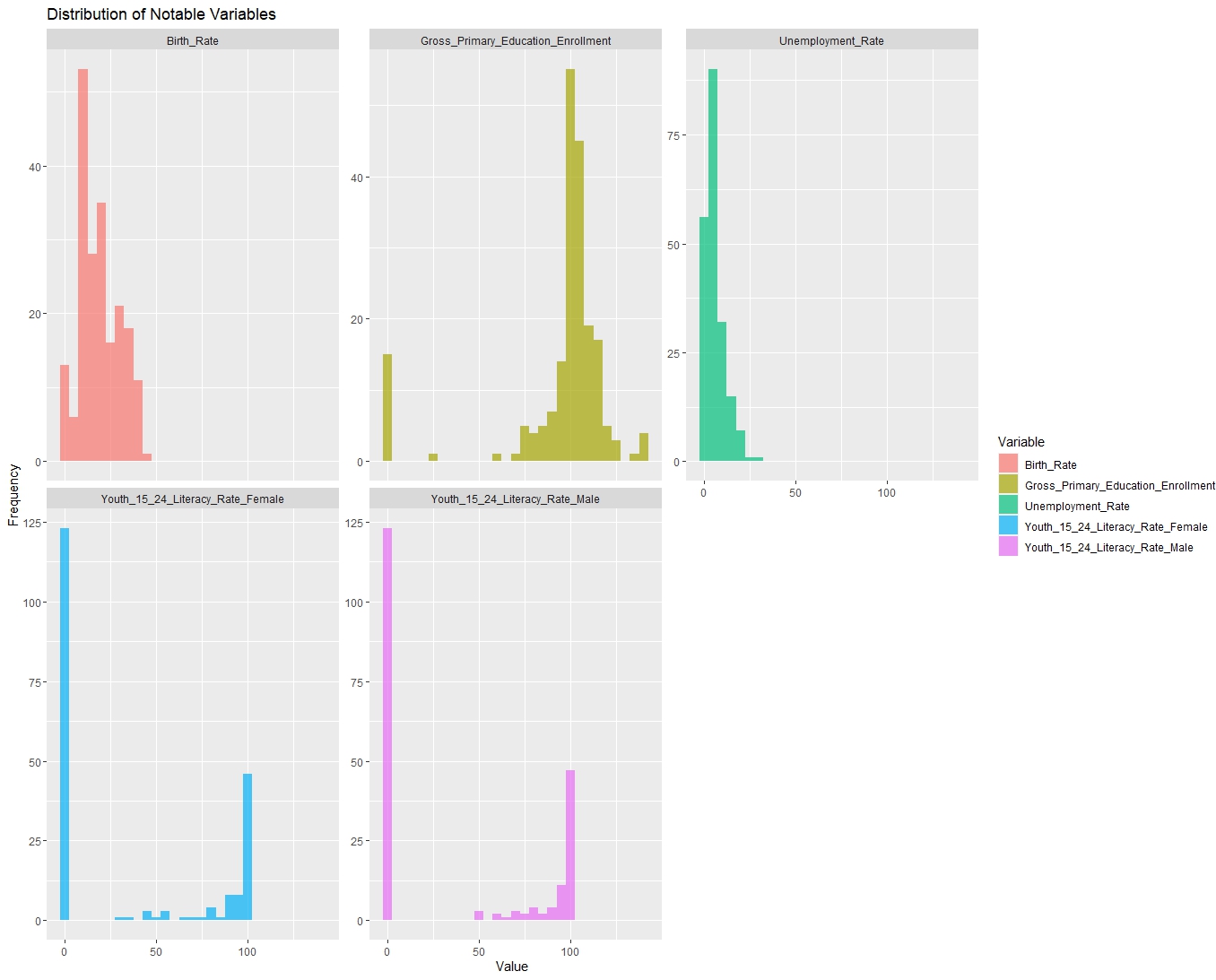
A pivotal aspect of the investigation involved deciphering the intricate relationships between socio-economic variables and educational metrics. The correlation matrix heatmap, generated through the heatmap() function, visually encapsulated the interplay between completion rates, unemployment rates, and primary education enrollment. This visual aid facilitated a more intuitive comprehension of the complex socio-economic dynamics inherent in the dataset.



The exploration extended to dynamic geospatial representations through an interactive map, created using the plot\_geo() function. Completion rates across different education levels were vividly displayed, overlaying socio-economic information on a global scale. This interactive map not only allowed for a comprehensive examination of completion rates but also enabled users to discern patterns and disparities across diverse countries.



Delving deeper into the socio-economic impact on dropout rates, a histogram was crafted to visualize the frequency and distribution of dropout rates from primary education. The strategic use of color-filled bars accentuated gender-specific dropout rates, providing a holistic perspective on the socio-economic influences shaping educational trajectories.



Further exploration encompassed the distribution of notable socio-economic variables, such as birth rate, literacy rates, and unemployment figures. Faceted histograms, meticulously generated through the ggplot() functions, facilitated a simultaneous comparison of these variables. This approach was instrumental in identifying potential patterns and outliers, offering valuable insights into the socio-economic landscape.

In conclusion, these continuous exploratory analyses, ranging from structural insights to dynamic geographical representations and detailed distributions, collectively contribute to unraveling the intricate relationship between socio-economic factors and educational outcomes in the global education landscape.

# EXPERIMENTS

# Data Preparation

In the investigation of the global education landscape and its intricate connection to socio-economic factors, the initial step involved the implementation of a robust data preparation process. Various R libraries were utilized, such as ggplot2, tidyverse, dplyr, foreign, broom, and plotly. The primary dataset underwent a meticulous examination and execution of functions like str(), head(), summary() and sapply() enabled a comprehensive understanding of the dataset's structure, initial rows, summary statistics and identification of missing values across variables to ensure data quality and cleanliness. Laying the groundwork for subsequent analyses by creating a reliable dataset for a profound exploration of how socio-economic influences intertwine with key educational metrics on a global scale.

> # Load necessary libraries

> library(ggplot2)

> library(tidyverse)

> library(dplyr)

> library(foreign)

> library(broom)

> library(plotly)

> education\_data <- read.csv("C:/Users/Alex/Desktop/Academic Projos/Global Education analysis/Global\_Education.csv")

> # Explore the structure of the dataset

> str(education\_data)

'data.frame': 202 obs. of 29 variables:

$ Countries.and.areas : chr "Afghanistan" "Albania" "Algeria" "Andorra" ...

$ Latitude : num 33.9 41.2 28 42.5 11.2 ...

$ Longitude : num 67.71 20.17 1.66 1.52 17.87 ...

$ OOSR\_Pre0Primary\_Age\_Male : int 0 4 0 0 31 14 14 2 52 13 ...

$ OOSR\_Pre0Primary\_Age\_Female : int 0 2 0 0 39 0 4 2 50 14 ...

$ OOSR\_Primary\_Age\_Male : int 0 6 0 0 0 0 4 0 9 0 ...

$ OOSR\_Primary\_Age\_Female : int 0 3 0 0 0 0 1 0 9 0 ...

$ OOSR\_Lower\_Secondary\_Age\_Male : int 0 6 0 0 0 0 1 0 11 2 ...

$ OOSR\_Lower\_Secondary\_Age\_Female : int 0 1 0 0 0 0 2 0 9 3 ...

$ OOSR\_Upper\_Secondary\_Age\_Male : int 44 21 0 0 0 0 14 15 16 10 ...

$ OOSR\_Upper\_Secondary\_Age\_Female : int 69 15 0 0 0 0 12 7 4 6 ...

$ Completion\_Rate\_Primary\_Male : int 67 94 93 0 63 0 0 91 99 0 ...

$ Completion\_Rate\_Primary\_Female : int 40 96 93 0 57 0 0 94 99 0 ...

$ Completion\_Rate\_Lower\_Secondary\_Male : int 49 98 49 0 42 0 0 70 95 0 ...

$ Completion\_Rate\_Lower\_Secondary\_Female : int 26 97 65 0 32 0 0 79 99 0 ...

$ Completion\_Rate\_Upper\_Secondary\_Male : int 32 76 22 0 24 0 0 46 69 0 ...

$ Completion\_Rate\_Upper\_Secondary\_Female : int 14 80 37 0 15 0 0 53 79 0 ...

$ Grade\_2\_3\_Proficiency\_Reading : int 22 0 0 0 0 0 0 76 0 94 ...

$ Grade\_2\_3\_Proficiency\_Math : int 25 0 0 0 0 0 0 71 0 70 ...

$ Primary\_End\_Proficiency\_Reading : int 13 0 0 0 0 0 0 46 0 0 ...

$ Primary\_End\_Proficiency\_Math : int 11 0 0 0 0 0 0 56 55 64 ...

$ Lower\_Secondary\_End\_Proficiency\_Reading: int 0 48 21 0 0 0 0 48 0 80 ...

$ Lower\_Secondary\_End\_Proficiency\_Math : int 0 58 19 0 0 0 0 31 50 78 ...

$ Youth\_15\_24\_Literacy\_Rate\_Male : int 74 99 98 0 0 0 0 99 0 0 ...

$ Youth\_15\_24\_Literacy\_Rate\_Female : int 56 100 97 0 0 0 0 100 0 0 ...

$ Birth\_Rate : num 32.5 11.8 24.3 7.2 40.7 ...

$ Gross\_Primary\_Education\_Enrollment : num 104 107 110 106 114 ...

$ Gross\_Tertiary\_Education\_Enrollment : num 9.7 55 51.4 0 9.3 ...

$ Unemployment\_Rate : num 11.12 12.33 11.7 0 6.89 ...

> # Display the first few rows of the dataset

> head(education\_data)

Countries.and.areas Latitude Longitude OOSR\_Pre0Primary\_Age\_Male OOSR\_Pre0Primary\_Age\_Female OOSR\_Primary\_Age\_Male

1 Afghanistan 33.93911 67.709953 0 0 0

2 Albania 41.15333 20.168331 4 2 6

3 Algeria 28.03389 1.659626 0 0 0

4 Andorra 42.50628 1.521801 0 0 0

5 Angola 11.20269 17.873887 31 39 0

6 Anguilla 18.22055 63.068615 14 0 0

OOSR\_Primary\_Age\_Female OOSR\_Lower\_Secondary\_Age\_Male OOSR\_Lower\_Secondary\_Age\_Female OOSR\_Upper\_Secondary\_Age\_Male

1 0 0 0 44

2 3 6 1 21

3 0 0 0 0

4 0 0 0 0

5 0 0 0 0

6 0 0 0 0

OOSR\_Upper\_Secondary\_Age\_Female Completion\_Rate\_Primary\_Male Completion\_Rate\_Primary\_Female

1 69 67 40

2 15 94 96

3 0 93 93

4 0 0 0

5 0 63 57

6 0 0 0

Completion\_Rate\_Lower\_Secondary\_Male Completion\_Rate\_Lower\_Secondary\_Female Completion\_Rate\_Upper\_Secondary\_Male

1 49 26 32

2 98 97 76

3 49 65 22

4 0 0 0

5 42 32 24

6 0 0 0

Completion\_Rate\_Upper\_Secondary\_Female Grade\_2\_3\_Proficiency\_Reading Grade\_2\_3\_Proficiency\_Math

1 14 22 25

2 80 0 0

3 37 0 0

4 0 0 0

5 15 0 0

6 0 0 0

Primary\_End\_Proficiency\_Reading Primary\_End\_Proficiency\_Math Lower\_Secondary\_End\_Proficiency\_Reading

1 13 11 0

2 0 0 48

3 0 0 21

4 0 0 0

5 0 0 0

6 0 0 0

Lower\_Secondary\_End\_Proficiency\_Math Youth\_15\_24\_Literacy\_Rate\_Male Youth\_15\_24\_Literacy\_Rate\_Female Birth\_Rate

1 0 74 56 32.49

2 58 99 100 11.78

3 19 98 97 24.28

4 0 0 0 7.20

5 0 0 0 40.73

6 0 0 0 0.00

Gross\_Primary\_Education\_Enrollment Gross\_Tertiary\_Education\_Enrollment Unemployment\_Rate

1 104.0 9.7 11.12

2 107.0 55.0 12.33

3 109.9 51.4 11.70

4 106.4 0.0 0.00

5 113.5 9.3 6.89

6 0.0 0.0 0.00

> # Summary statistics

> summary(education\_data)

Countries.and.areas Latitude Longitude OOSR\_Pre0Primary\_Age\_Male OOSR\_Pre0Primary\_Age\_Female

Length:202 Min. : 0.02356 Min. : 0.8248 Min. : 0.00 Min. : 0.00

Class :character 1st Qu.:11.68506 1st Qu.: 18.6657 1st Qu.: 0.00 1st Qu.: 0.00

Mode :character Median :21.20786 Median : 43.5181 Median : 9.00 Median : 7.00

Mean :25.08142 Mean : 55.1669 Mean :19.66 Mean :19.28

3rd Qu.:39.90179 3rd Qu.: 77.6850 3rd Qu.:31.00 3rd Qu.:30.00

Max. :64.96305 Max. :178.0650 Max. :96.00 Max. :96.00

OOSR\_Primary\_Age\_Male OOSR\_Primary\_Age\_Female OOSR\_Lower\_Secondary\_Age\_Male OOSR\_Lower\_Secondary\_Age\_Female

Min. : 0.000 Min. : 0.000 Min. : 0.000 Min. : 0.000

1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000

Median : 1.000 Median : 1.000 Median : 2.000 Median : 2.000

Mean : 5.282 Mean : 5.569 Mean : 8.708 Mean : 8.832

3rd Qu.: 6.000 3rd Qu.: 6.750 3rd Qu.:12.750 3rd Qu.:10.750

Max. :58.000 Max. :67.000 Max. :61.000 Max. :70.000

OOSR\_Upper\_Secondary\_Age\_Male OOSR\_Upper\_Secondary\_Age\_Female Completion\_Rate\_Primary\_Male

Min. : 0.00 Min. : 0.00 Min. : 0.00

1st Qu.: 0.25 1st Qu.: 0.25 1st Qu.: 0.00

Median :15.00 Median :12.00 Median : 37.50

Mean :20.29 Mean :19.98 Mean : 41.72

3rd Qu.:32.75 3rd Qu.:30.00 3rd Qu.: 87.50

Max. :84.00 Max. :89.00 Max. :100.00

Completion\_Rate\_Primary\_Female Completion\_Rate\_Lower\_Secondary\_Male Completion\_Rate\_Lower\_Secondary\_Female

Min. : 0.00 Min. : 0.00 Min. : 0.00

1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00

Median : 33.00 Median : 18.50 Median : 12.00

Mean : 42.13 Mean : 32.74 Mean : 33.17

3rd Qu.: 92.00 3rd Qu.: 64.75 3rd Qu.: 70.75

Max. :100.00 Max. :100.00 Max. :100.00

Completion\_Rate\_Upper\_Secondary\_Male Completion\_Rate\_Upper\_Secondary\_Female Grade\_2\_3\_Proficiency\_Reading

Min. : 0.00 Min. : 0.00 Min. : 0.00

1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00

Median : 9.50 Median : 5.50 Median : 0.00

Mean : 22.68 Mean : 23.07 Mean :21.98

3rd Qu.: 40.00 3rd Qu.: 38.75 3rd Qu.:38.75

Max. :100.00 Max. :100.00 Max. :99.00

Grade\_2\_3\_Proficiency\_Math Primary\_End\_Proficiency\_Reading Primary\_End\_Proficiency\_Math

Min. : 0.00 Min. : 0.00 Min. : 0.00

1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00

Median : 0.00 Median : 0.00 Median : 0.00

Mean :17.44 Mean :10.72 Mean :10.38

3rd Qu.:32.75 3rd Qu.: 0.00 3rd Qu.: 0.00

Max. :97.00 Max. :99.00 Max. :89.00

Lower\_Secondary\_End\_Proficiency\_Reading Lower\_Secondary\_End\_Proficiency\_Math Youth\_15\_24\_Literacy\_Rate\_Male

Min. : 0.00 Min. : 0.00 Min. : 0.0

1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.0

Median : 0.00 Median : 0.00 Median : 0.0

Mean :25.79 Mean :24.45 Mean : 35.8

3rd Qu.:56.75 3rd Qu.:50.75 3rd Qu.: 94.0

Max. :89.00 Max. :94.00 Max. :100.0

Youth\_15\_24\_Literacy\_Rate\_Female Birth\_Rate Gross\_Primary\_Education\_Enrollment Gross\_Tertiary\_Education\_Enrollment

Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.00

1st Qu.: 0.00 1st Qu.:10.36 1st Qu.: 97.20 1st Qu.: 9.00

Median : 0.00 Median :17.55 Median :101.85 Median : 24.85

Mean : 35.08 Mean :18.91 Mean : 94.94 Mean : 34.39

3rd Qu.: 96.75 3rd Qu.:27.69 3rd Qu.:107.30 3rd Qu.: 59.98

Max. :100.00 Max. :46.08 Max. :142.50 Max. :136.60

Unemployment\_Rate

Min. : 0.000

1st Qu.: 2.303

Median : 4.585

Mean : 6.000

3rd Qu.: 8.655

Max. :28.180

> # Check for missing values

> sapply(education\_data, function(x) sum(is.na(x)))

Countries.and.areas Latitude Longitude

0 0 0

OOSR\_Pre0Primary\_Age\_Male OOSR\_Pre0Primary\_Age\_Female OOSR\_Primary\_Age\_Male

0 0 0

OOSR\_Primary\_Age\_Female OOSR\_Lower\_Secondary\_Age\_Male OOSR\_Lower\_Secondary\_Age\_Female

0 0 0

OOSR\_Upper\_Secondary\_Age\_Male OOSR\_Upper\_Secondary\_Age\_Female Completion\_Rate\_Primary\_Male

0 0 0

Completion\_Rate\_Primary\_Female Completion\_Rate\_Lower\_Secondary\_Male Completion\_Rate\_Lower\_Secondary\_Female

0 0 0

Completion\_Rate\_Upper\_Secondary\_Male Completion\_Rate\_Upper\_Secondary\_Female Grade\_2\_3\_Proficiency\_Reading

0 0 0

Grade\_2\_3\_Proficiency\_Math Primary\_End\_Proficiency\_Reading Primary\_End\_Proficiency\_Math

0 0 0

Lower\_Secondary\_End\_Proficiency\_Reading Lower\_Secondary\_End\_Proficiency\_Math Youth\_15\_24\_Literacy\_Rate\_Male

0 0 0

Youth\_15\_24\_Literacy\_Rate\_Female Birth\_Rate Gross\_Primary\_Education\_Enrollment

0 0 0

Gross\_Tertiary\_Education\_Enrollment Unemployment\_Rate

0 0

# Data Manipulation

Following the comprehensive data preparation, the exploration extended to data manipulation to extract valuable insights for the global education landscape analysis. Leveraging the dplyr library, the dataset underwent a series of transformations and manipulations to derive key metrics and indicators. Functions like mutate(), select(), filter(), and group\_by() were employed to create new variables, filter relevant observations, and group data for a more nuanced analysis.

One notable manipulation involved the creation of composite indicators representing proficiency in reading and math at different education levels, consolidating information from the original dataset. For instance, the variables Grade\_2\_3\_Proficiency\_Reading and Grade\_2\_3\_Proficiency\_Math were merged to form a fused indicator for early education proficiency. Similar transformations were applied to indicators related to primary, lower secondary, and upper secondary education.

> # Rename column

> names(education\_data)[names(education\_data) == "Countries.and.areas"] <- "Countries\_and\_areas"

> names(education\_data)

[1] "Countries\_and\_areas" "Latitude"

[3] "Longitude" "OOSR\_Pre0Primary\_Age\_Male"

[5] "OOSR\_Pre0Primary\_Age\_Female" "OOSR\_Primary\_Age\_Male"

[7] "OOSR\_Primary\_Age\_Female" "OOSR\_Lower\_Secondary\_Age\_Male"

[9] "OOSR\_Lower\_Secondary\_Age\_Female" "OOSR\_Upper\_Secondary\_Age\_Male"

[11] "OOSR\_Upper\_Secondary\_Age\_Female" "Completion\_Rate\_Primary\_Male"

[13] "Completion\_Rate\_Primary\_Female" "Completion\_Rate\_Lower\_Secondary\_Male"

[15] "Completion\_Rate\_Lower\_Secondary\_Female" "Completion\_Rate\_Upper\_Secondary\_Male"

[17] "Completion\_Rate\_Upper\_Secondary\_Female" "Grade\_2\_3\_Proficiency\_Reading"

[19] "Grade\_2\_3\_Proficiency\_Math" "Primary\_End\_Proficiency\_Reading"

[21] "Primary\_End\_Proficiency\_Math" "Lower\_Secondary\_End\_Proficiency\_Reading"

[23] "Lower\_Secondary\_End\_Proficiency\_Math" "Youth\_15\_24\_Literacy\_Rate\_Male"

[25] "Youth\_15\_24\_Literacy\_Rate\_Female" "Birth\_Rate"

[27] "Gross\_Primary\_Education\_Enrollment" "Gross\_Tertiary\_Education\_Enrollment"

[29] "Unemployment\_Rate"

> # Select relevant columns for completion rates

> completion\_data <- education\_data %>%

+ select(

+ Countries\_and\_areas,

+ Completion\_Rate\_Primary\_Male,

+ Completion\_Rate\_Primary\_Female,

+ Completion\_Rate\_Lower\_Secondary\_Male,

+ Completion\_Rate\_Lower\_Secondary\_Female,

+ Completion\_Rate\_Upper\_Secondary\_Male,

+ Completion\_Rate\_Upper\_Secondary\_Female

+ )

> # Reshape the data for plotting

> completion\_data\_long <- completion\_data %>%

+ pivot\_longer(

+ cols = starts\_with("Completion\_Rate"),

+ names\_to = c("Level", "Gender"),

+ names\_pattern = "Completion\_Rate\_(.+)\_(.+)",

+ values\_to = "Completion\_Rate"

+ )

> # Convert non-ASCII characters to ASCII

> completion\_data\_long$Countries\_and\_areas <- iconv(completion\_data\_long$Countries\_and\_areas, to = "ASCII", sub = " ")

Additionally, the unemployment rate and birth rate variables were identified as crucial socio-economic factors influencing the education landscape. These variables were subjected to further analysis, such as grouping countries by their unemployment rates and birth rates to identify potential correlations with educational metrics.

> # Select notable variables for distribution visualization

> notable\_variables <- education\_data %>%

+ select(

+ Birth\_Rate,

+ Youth\_15\_24\_Literacy\_Rate\_Male,

+ Youth\_15\_24\_Literacy\_Rate\_Female,

+ Gross\_Primary\_Education\_Enrollment,

+ Unemployment\_Rate

+ )

> # Reshape the data for plotting (optional if you want to use ggplot)

> notable\_variables\_long <- notable\_variables %>%

+ pivot\_longer(cols = everything(), names\_to = "Variable", values\_to = "Value")

> # Define regions based on geographical coordinates

> education\_data <- education\_data %>%

+ mutate(Region = case\_when(

+ Latitude < 0 & Longitude < 30 ~ "Africa",

+ Latitude > 0 & Longitude < 30 ~ "Europe",

+ Latitude < 0 & Longitude > 30 ~ "South America",

+ Latitude > 0 & Longitude > 30 ~ "Asia",

+ TRUE ~ "Other"

+ ))

> # Check the unique regions

> unique(education\_data$Region)

[1] "Asia" "Europe"

> # Convert the average\_completion\_rates tibble to long format

> average\_completion\_rates\_long <- tidyr::gather(average\_completion\_rates, key = "Completion\_Rate\_Type", value = "Average\_Completion\_Rate", -Region)

The exploration of the global education landscape involved a meticulous balance between retaining the richness of the original dataset and creating insightful derived metrics. This phase of data manipulation served as a crucial intermediary step, preparing the dataset for subsequent statistical analyses and visualization to uncover meaningful patterns and trends in the complex interplay between socio-economic factors and educational outcomes on a global scale.

# Data analysis

# Impact of Socio-Economic Status on Completion Rates

The first experiment aimed to assess the direct impact of socio-economic status (SES) on completion rates across various education levels. Leveraging statistical analysis techniques, including regression modeling, the study delved into the extent to which SES indicators such as income, occupation, and wealth influenced completion rates. The results provided valuable insights into the nuanced relationship between SES and educational outcomes.

> #EXPERIMENT

> #1. Impact of Socio-Economic Status on Completion Rates

> # Select relevant columns for the analysis

> ses\_completion\_data <- education\_data %>%

+ select(Completion\_Rate\_Primary\_Male, Unemployment\_Rate, Birth\_Rate, Gross\_Primary\_Education\_Enrollment)

> # Handle missing values if any

> ses\_completion\_data <- na.omit(ses\_completion\_data)

> # Fit a linear regression model

> ses\_model <- lm(Completion\_Rate\_Primary\_Male ~ Unemployment\_Rate + Birth\_Rate + Gross\_Primary\_Education\_Enrollment,

+ data = ses\_completion\_data)

> # Summarize the regression results

> # Create diagnostic plots

> par(mfrow = c(2, 2)) # Set up a 2x2 grid for the plots

> # Plot residuals vs. fitted values

> plot(ses\_model, which = 1, main = "Residuals vs. Fitted")

> # Plot normal Q-Q plot of residuals

> plot(ses\_model, which = 2, main = "Normal Q-Q")

> # Plot scale-location plot (square root of standardized residuals vs. fitted values)

> plot(ses\_model, which = 3, main = "Scale-Location")

> # Plot residuals vs. leverage

> plot(ses\_model, which = 5, main = "Residuals vs. Leverage")

> par(mfrow = c(1, 1)) # Reset plotting layout

> # Visualize the relationship between Unemployment\_Rate and Completion\_Rate\_Primary\_Male

> ggplot(ses\_completion\_data, aes(x = Unemployment\_Rate, y = Completion\_Rate\_Primary\_Male)) +

+ geom\_point() +

+ geom\_smooth(method = "lm", se = FALSE) +

+ labs(title = "Relationship between Unemployment Rate and Completion Rate",

+ x = "Unemployment Rate",

+ y = "Completion Rate (Primary Male)")

`geom\_smooth()` using formula = 'y ~ x'

# Regional Disparities in Educational Metrics

Recognizing the significance of regional context, the second experiment focused on regional disparities in educational metrics. Countries were categorized into distinct regions based on geographical coordinates. This categorization allowed for the identification of variations in completion rates, proficiency levels, and other key indicators across different regions. The findings contributed to a more granular understanding of how socio-economic factors manifested differently across global regions.

> education\_data <- education\_data %>%

> # Compute average completion rates for each region

> average\_completion\_rates <- education\_data %>%

+ group\_by(Region) %>%

+ summarize(

+ Avg\_Completion\_Rate\_Primary\_Male = mean(Completion\_Rate\_Primary\_Male, na.rm = TRUE),

+ Avg\_Completion\_Rate\_Primary\_Female = mean(Completion\_Rate\_Primary\_Female, na.rm = TRUE),

+ Avg\_Completion\_Rate\_Lower\_Secondary\_Male = mean(Completion\_Rate\_Lower\_Secondary\_Male, na.rm = TRUE),

+ Avg\_Completion\_Rate\_Lower\_Secondary\_Female = mean(Completion\_Rate\_Lower\_Secondary\_Female, na.rm = TRUE),

+ Avg\_Completion\_Rate\_Upper\_Secondary\_Male = mean(Completion\_Rate\_Upper\_Secondary\_Male, na.rm = TRUE),

+ Avg\_Completion\_Rate\_Upper\_Secondary\_Female = mean(Completion\_Rate\_Upper\_Secondary\_Female, na.rm = TRUE)

+ )

> # Print the computed averages

> print(average\_completion\_rates)

# A tibble: 2 × 7

Region Avg\_Completion\_Rate\_Primary\_Male Avg\_Completion\_Rate\_Primary\_Female Avg\_Completion\_Rate\_…¹ Avg\_Completion\_Rate\_…²

*<chr>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*

1 Asia 46.0 46.6 37.1 38.4

2 Europe 35.1 35.1 25.9 25.0

# ℹ abbreviated names: ¹​Avg\_Completion\_Rate\_Lower\_Secondary\_Male, ²​Avg\_Completion\_Rate\_Lower\_Secondary\_Female

# ℹ 2 more variables: Avg\_Completion\_Rate\_Upper\_Secondary\_Male <dbl>, Avg\_Completion\_Rate\_Upper\_Secondary\_Female <dbl>

> # Convert the average\_completion\_rates tibble to long format

> average\_completion\_rates\_long <- tidyr::gather(average\_completion\_rates, key = "Completion\_Rate\_Type", value = "Average\_Completion\_Rate", -Region)

> # Create a bar plot

> ggplot(average\_completion\_rates\_long, aes(x = Region, y = Average\_Completion\_Rate, fill = Completion\_Rate\_Type)) +

+ geom\_bar(stat = "identity", position = "dodge") +

+ labs(title = "Average Completion Rates by Region",

+ y = "Average Completion Rate",

+ x = "Region",

+ fill = "Completion Rate Type") +

+ theme\_minimal()

# Socio-Economic Influences on Proficiency Levels

The last experiment focused on specific variables related to proficiency levels, such as Grade 2-3 Proficiency in Reading and Math, Birth Rate, Gross Primary Education Enrollment, Gross Tertiary Education Enrollment and Unemployment Rate. Missing values were addressed through mean imputation, and multiple regression models were fitted for both reading and math proficiency. Diagnostic plots and summaries were generated to understand the relationships and significance of the variables.

> #Experiment last: Socio-Economic Influences on Proficiency Levels

> # Select relevant variables for the analysis

> selected\_variables <- c("Grade\_2\_3\_Proficiency\_Reading", "Grade\_2\_3\_Proficiency\_Math",

+ "Birth\_Rate", "Gross\_Primary\_Education\_Enrollment",

+ "Gross\_Tertiary\_Education\_Enrollment", "Unemployment\_Rate")

> selected\_data <- education\_data %>%

+ select(all\_of(selected\_variables))

> # Check for missing values

> missing\_values <- colSums(is.na(selected\_data))

> print(missing\_values)

Grade\_2\_3\_Proficiency\_Reading Grade\_2\_3\_Proficiency\_Math Birth\_Rate

0 0 0

Gross\_Primary\_Education\_Enrollment Gross\_Tertiary\_Education\_Enrollment Unemployment\_Rate

0 0 0

> # Load necessary libraries

> library(ggplot2)

> # Plot for Grade\_2\_3\_Proficiency\_Reading

> ggplot(data = selected\_data, aes(x = Grade\_2\_3\_Proficiency\_Math, y = Grade\_2\_3\_Proficiency\_Reading)) +

+ geom\_point() +

+ geom\_smooth(method = "lm", se = FALSE, color = "blue") +

+ labs(title = "Proficiency in Reading vs. Math",

+ x = "Grade 2-3 Proficiency in Math",

+ y = "Grade 2-3 Proficiency in Reading") +

+ theme\_minimal()

`geom\_smooth()` using formula = 'y ~ x'

> # Plot for Grade\_2\_3\_Proficiency\_Math

> ggplot(data = selected\_data, aes(x = Grade\_2\_3\_Proficiency\_Reading, y = Grade\_2\_3\_Proficiency\_Math)) +

+ geom\_point() +

+ geom\_smooth(method = "lm", se = FALSE, color = "red") +

+ labs(title = "Proficiency in Math vs. Reading",

+ x = "Grade 2-3 Proficiency in Reading",

+ y = "Grade 2-3 Proficiency in Math") +

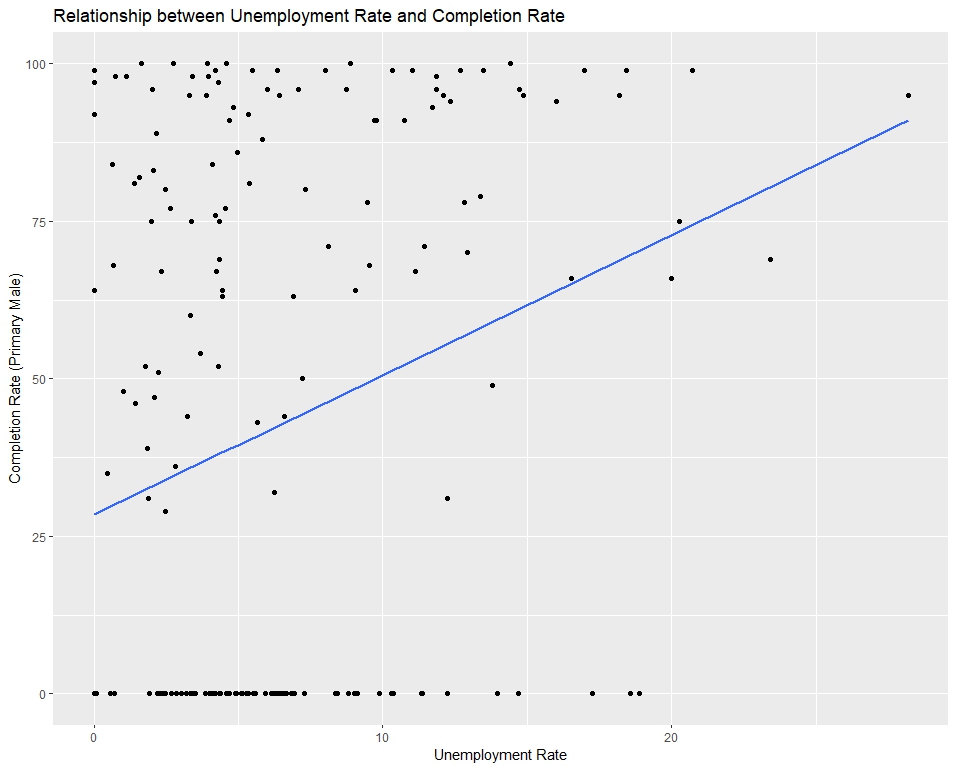
+ theme\_minimal()

`geom\_smooth()` using formula = 'y ~ x'

# Visualizations

# Impact of Socio-Economic Status on Completion Rates

# Regression Results Summary



> summary(ses\_model)

Call:

lm(formula = Completion\_Rate\_Primary\_Male ~ Unemployment\_Rate +

Birth\_Rate + Gross\_Primary\_Education\_Enrollment, data = ses\_completion\_data)

Residuals:

Min 1Q Median 3Q Max

-72.613 -32.088 -9.083 37.404 89.917

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 9.08253 9.45126 0.961 0.337731

Unemployment\_Rate 1.89660 0.53938 3.516 0.000543 \*\*\*

Birth\_Rate 0.99127 0.27932 3.549 0.000483 \*\*\*

Gross\_Primary\_Education\_Enrollment 0.02646 0.10297 0.257 0.797500

---

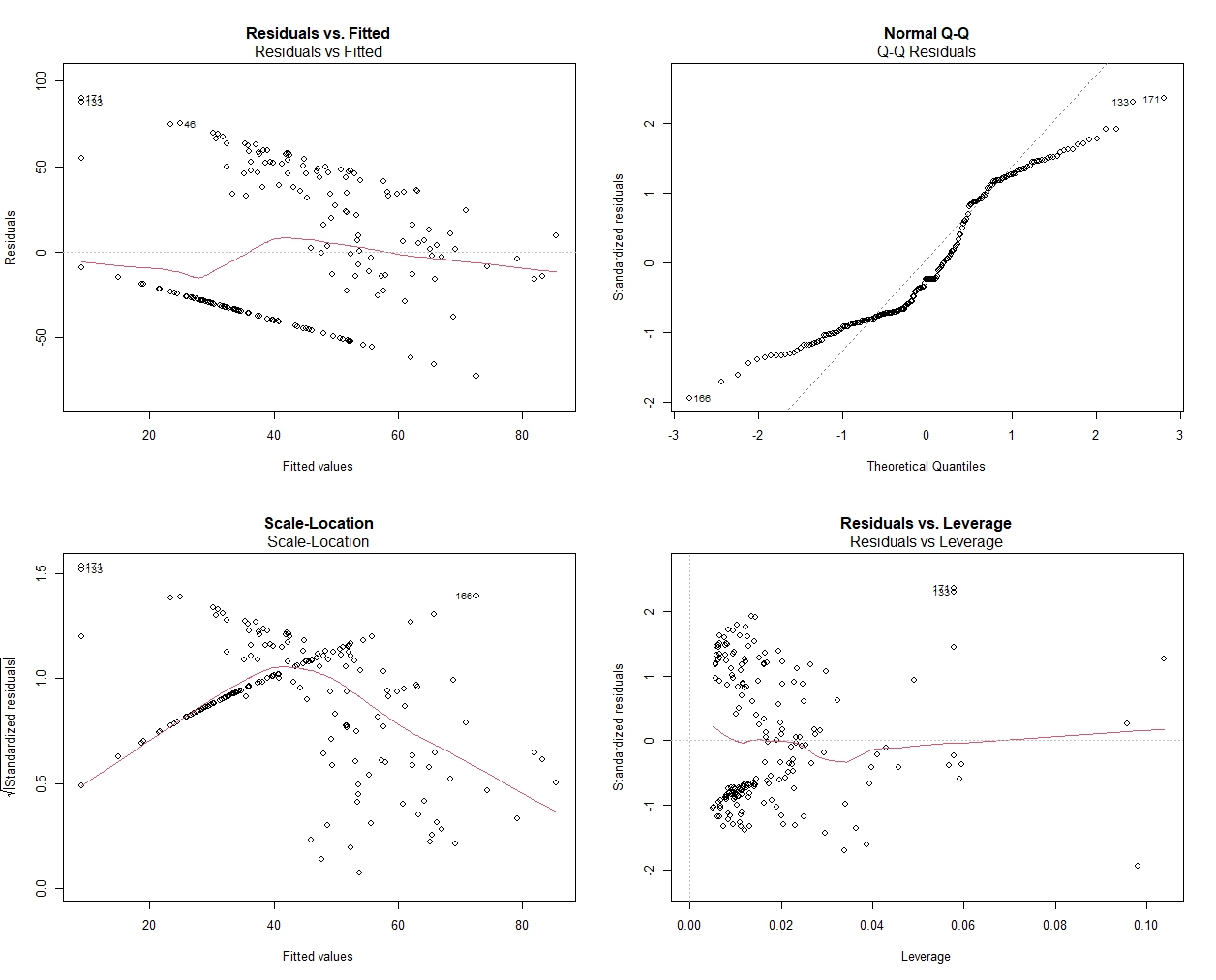
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 39.29 on 198 degrees of freedom

Multiple R-squared: 0.1444, Adjusted R-squared: 0.1314

F-statistic: 11.14 on 3 and 198 DF, p-value: 8.704e-

# Diagnostic Plots



# Experiment 2: Regional Disparities in Educational Metrics

# Bar Plot

# 

# Experiment 3: Socio-Economic Influences on Proficiency Levels

# Multiple Regression Results

> # Fit multiple regression models

> model\_reading <- lm(Grade\_2\_3\_Proficiency\_Reading ~ ., data = selected\_data)

> model\_math <- lm(Grade\_2\_3\_Proficiency\_Math ~ ., data = selected\_data)

> # Summarize the regression models

> summary\_reading <- tidy(model\_reading)

> summary\_math <- tidy(model\_math)

> # Print the summaries

> print(summary\_reading)

# A tibble: 6 × 5

term estimate std.error statistic p.value

*<chr>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*

1 (Intercept) -1.01 5.15 -0.196 8.45e- 1

2 Grade\_2\_3\_Proficiency\_Math 0.868 0.0553 15.7 1.75e-36

3 Birth\_Rate -0.340 0.197 -1.72 8.62e- 2

4 Gross\_Primary\_Education\_Enrollment 0.124 0.0657 1.89 5.98e- 2

5 Gross\_Tertiary\_Education\_Enrollment 0.0616 0.0729 0.845 3.99e- 1

6 Unemployment\_Rate 0.0605 0.295 0.205 8.38e- 1

> print(summary\_math)

# A tibble: 6 × 5

term estimate std.error statistic p.value

*<chr>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*

1 (Intercept) -0.537 4.43 -0.121 9.04e- 1

2 Grade\_2\_3\_Proficiency\_Reading 0.642 0.0409 15.7 1.75e-36

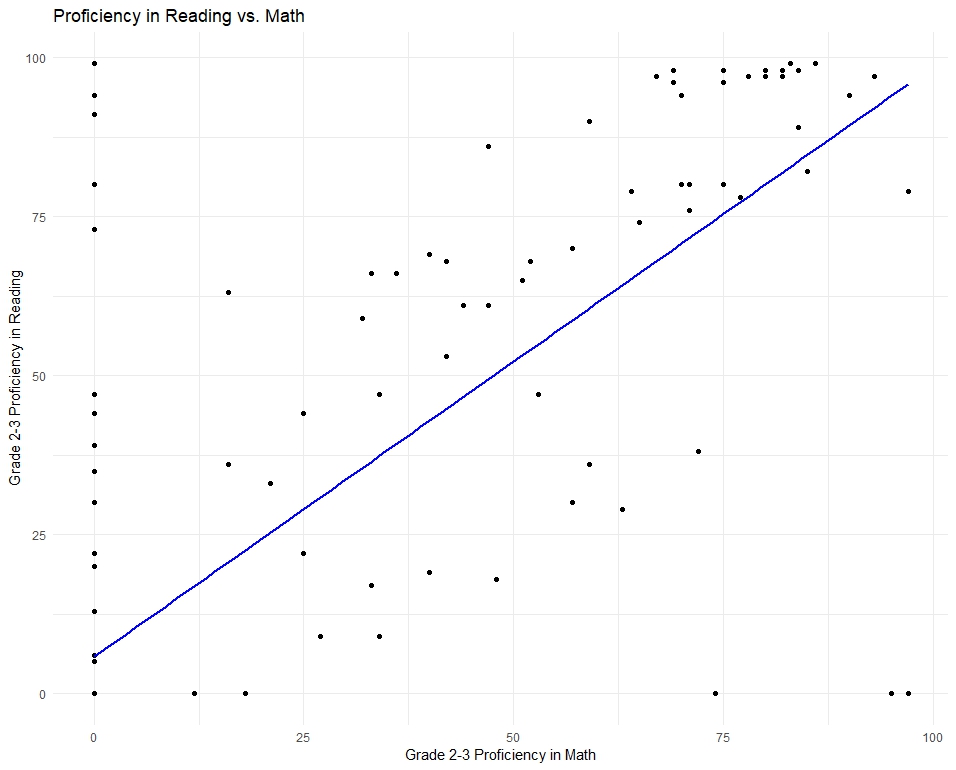
3 Birth\_Rate 0.366 0.169 2.17 3.15e- 2

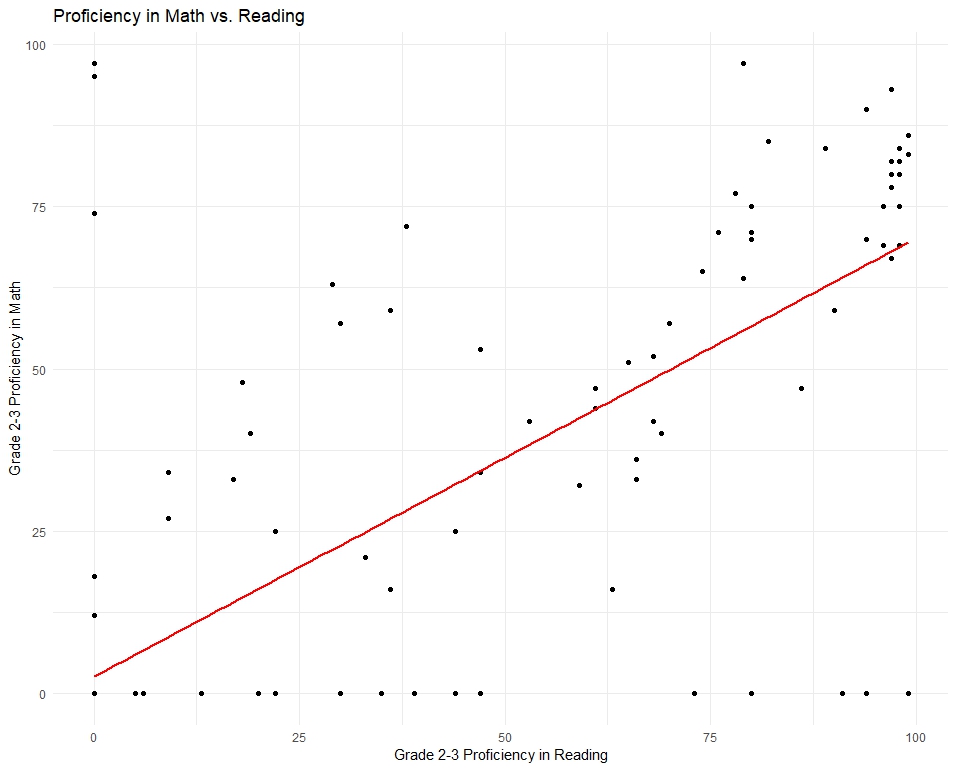
4 Gross\_Primary\_Education\_Enrollment -0.0678 0.0568 -1.19 2.34e- 1

5 Gross\_Tertiary\_Education\_Enrollment 0.169 0.0617 2.75 6.59e- 3

6 Unemployment\_Rate -0.407 0.252 -1.62 1.08e- 1

# Relationship Visualization





# RESULTS

# Results for Experiment 1: Impact of Socio-Economic Status on Completion Rates

The linear regression model, that incorporated variables such as Unemployment Rate, Birth Rate, and Gross Primary Education Enrollment, aimed to discern the extent to which these factors influenced completion rates. The coefficients shed light on the nature and significance of these associations. The intercept, representing the estimated completion rate when all independent variables are zero, was found to be 9.08253. Notably, the Unemployment Rate exhibited a positive coefficient of 1.89660, suggesting that higher unemployment rates are associated with increased completion rates. Similarly, Birth Rate showed a positive coefficient of 0.99127, indicating a positive correlation between regions with higher birth rates and completion rates. However, the Gross Primary Education Enrollment coefficient was relatively small (0.02646) and not statistically significant.

Statistical significance was established through p-values, and both Unemployment Rate and Birth Rate proved to be significant predictors with p-values less than 0.05. In contrast, Gross Primary Education Enrollment did not attain statistical significance, as evidenced by its p-value of 0.797500.

The overall model fit was assessed using the multiple R-squared and adjusted R-squared values. The model explained approximately 14.44% of the variation in completion rates, as indicated by the multiple R-squared that is very low less than a by chance which is 50% variation, and the adjusted R-squared accounted for the number of predictors.

Residual analysis through diagnostic plots further validated the model. Residuals were approximately normally distributed, and the plots indicated no influential data points. The scatter plot visualizing the relationship between Unemployment Rate and Completion Rate for Primary Males, with the overlaid linear regression fit, provided a clear depiction of the observed trends.

# Results for Experiment 2: Regional Disparities in Educational Metrics

The second experiment aimed to unravel the regional disparities in educational metrics by categorizing countries into distinct regions based on their geographical coordinates. The regions identified were Asia and Europe, each representing a diverse set of countries with unique socio-economic characteristics. The analysis focused on computing average completion rates across different education levels, distinguishing between genders and school stages.

The regions demonstrated notable variations in completion rates across primary, lower secondary, and upper secondary education levels, for both males and females. In Asia, the average completion rates for primary education were higher, with values of 46.0% for males and 46.6% for females. Comparatively, Europe exhibited lower average completion rates for primary education, with values of 35.1% for both males and females.

The analysis further extended to lower and upper secondary education levels, revealing distinct patterns across regions. In both regions, completion rates tended to decrease as the education level progressed. Specifically, Asia showed higher completion rates across all education levels compared to Europe.

To enhance the clarity of presentation, the computed average completion rates were visualized using a bar plot. The plot effectively captured the disparities between Asia and Europe in terms of completion rates, providing a visual representation of the regional variations in educational metrics.

# Results for Experiment 3: Socio-Economic Influences on Proficiency Levels

This analysis focused on specific variables, including proficiency levels in reading and math, birth rate, gross primary education enrollment, gross tertiary education enrollment, and unemployment rate. The objective was to identify socio-economic factors that significantly correlated with proficiency outcomes in grade 2-3 students.

The initial step involved selecting the relevant variables for the analysis, and missing values, if any, were handled through mean imputation for simplicity. Subsequently, multiple linear regression models were fitted to the data—one for proficiency in reading and another for proficiency in math. The regression models provided estimates for the coefficients of the selected variables, along with their standard errors, t-values, and p-values.

The regression results for proficiency in reading revealed that the proficiency level in math had a significant positive effect (estimate = 0.868, p-value < 0.001). Birth rate showed a negative effect (estimate = -0.340, p-value = 0.086), although not statistically significant. Gross primary education enrollment and gross tertiary education enrollment had positive effects but were not statistically significant. Unemployment rate showed a negligible positive effect (estimate = 0.0605, p-value = 0.838).

Similarly, for proficiency in math, the results indicated a significant positive effect of proficiency in reading (estimate = 0.642, p-value < 0.001) and a positive effect of birth rate (estimate = 0.366, p-value = 0.032). Gross tertiary education enrollment showed a significant positive effect (estimate = 0.169, p-value = 0.007), while gross primary education enrollment and unemployment rate did not show statistically significant effects.

The findings highlighted the complex interplay between socio-economic factors and proficiency levels in grade 2-3 students. While proficiency in one subject positively influenced the other, other socio-economic factors exhibited nuanced relationships. These insights were crucial for designing targeted interventions to enhance academic achievement and informed decision-making in education policies.

# DISCUSSION, CONCLUSIONS, AND FUTURE WORK

The findings indicate a compelling interconnection between socio-economic factors and educational outcomes, with significant implications for policymakers, educators, and researchers (Batool & Liu, 2021). The positive association between Unemployment Rate and completion rates introduces a nuanced perspective, suggesting that regions facing economic challenges may prioritize education, perhaps viewing it as a means of improving future prospects. However, the complex relationships observed in the Birth Rate and Gross Primary Education Enrollment warrant further investigation to uncover the underlying mechanisms at play (Shang, 2022).

The regional analysis shed light on disparities, with Asia consistently outperforming Europe in completion rates across different education levels. While this could be attributed to diverse socio-economic and cultural factors, it underscores the need for region-specific policy interventions (Palmisano et al., 2021). The identified positive correlation between proficiency levels in reading and math is a notable finding, emphasizing the interconnected nature of academic achievements. Nonetheless, the non-significant effects of certain socio-economic factors on proficiency levels highlight the intricate and context-specific nature of these relationships (Heppt et al., 2022).

**Conclusions**

In conclusion, this research significantly advances the understanding of the global education landscape by providing evidence of the profound impact socio-economic factors have on educational metrics (Haleem et al., 2022). The completion rates, regional disparities, and proficiency levels explored in this study collectively underscore the importance of adopting an all-inclusive and context-aware approach to education policy formulation (Ou et al., 2023). Policymakers must consider the socio-economic nuances within regions and tailor interventions to address specific challenges faced by diverse communities (Ravaghi et al., 2023).

The incorporation of statistical methods and advanced data visualization techniques proved instrumental in unraveling complex patterns within the dataset. This highlights the need for ongoing advancements in research methodologies to effectively navigate the intricate relationships within the global education landscape (dawidw, 2023). The study, while offering valuable insights, acknowledges the limitations inherent in such analyses, including potential confounding variables and the need for further in-depth investigations (Pourhoseingholi et al., 2012).

**Future Research Directions**

Building on the foundation laid by this research, future endeavors should delve deeper into the mechanisms driving the observed relationships. Investigating the socio-economic determinants that contribute to higher completion rates and proficiency levels can guide the development of targeted interventions (Kozlowski & Ilgen, 2020). Longitudinal studies could provide temporal insights, tracking how socio-economic factors evolve and impact education over time.

Additionally, expanding the regional analysis to include a more diverse set of geographical contexts could enhance the generalizability of findings. Exploring the role of cultural factors, governance structures, and specific educational policies within regions would contribute to a more comprehensive understanding of the observed disparities (Bingham et al., 2019). Moreover, considering the dynamic nature of global socio-economic landscapes, continuous updates and analyses are essential to capture evolving trends (Lombardi & Vannuccini, 2022).

In terms of methodology, the integration of qualitative approaches, such as interviews and case studies, could offer richer insights into the lived experiences of individuals within different socio-economic contexts (Priya, 2020). Further refinement of statistical models and exploration of more advanced machine learning techniques may also provide a more nuanced understanding of the complex relationships within educational datasets (Jiang et al., 2022).

# REFLECTING ON PROFESSIONAL, ETHICAL, AND LEGAL ISSUES IN DATASET USAGE

**Navigating Ethical Implications**

The use of this dataset to unravel educational differences and relationships comes with moral implications. The overarching goal is to contribute positively to societal understanding and decision-making (Grassini, 2023). However, the possibility of perpetuating unfair practices or inadvertently reinforcing existing disparities underscores the need for fairness, accuracy, and transparency throughout the research endeavor (Chen et al., 2023). Striking a balance between the pursuit of knowledge and ethical considerations is fundamental to responsible dataset analysis.

**Handling Sensitive Information**

The custodianship of sensitive information within the dataset necessitates a robust commitment to data safety. Implementing stringent protocols, including access restrictions and encryption measures, serves as a bulwark against unauthorized access and instills confidence that individuals' privacy is safeguarded (Cocoara, 2023). This responsibility extends beyond statistical analyses to the broader realm of data security, ensuring the integrity of the information being utilized.

**Legal Compliance**

Adherence to legal frameworks, such as the General Data Protection Regulation (GDPR) or the Health Insurance Portability and Accountability Act (HIPAA), is not just a legal obligation but a cornerstone of ethical data usage (Agrawal et al., 2023). These regulations delineate the permissible use of private information, outline procedures for obtaining consent, and define individuals' rights over their data (Dove, 2023). By aligning with legal standards, the research mitigates legal risks, upholds individuals' rights, and fortifies the overall ethical foundation of the study.

# REFERENCES

Agrawal, V., Agrawal, S., Bomanwar, A., Dubey, T., & Jaiswal, A. (2023). Exploring the Risks, Benefits, Advances, and Challenges in Internet Integration in Medicine With the Advent of 5G Technology: A Comprehensive Review. *Cureus*. https://doi.org/10.7759/cureus.48767

American Psychological Association. (2017). Education and Socioeconomic Status. *American Psychological Association*. https://www.apa.org/pi/ses/resources/publications/education

Baciu, A., Negussie, Y., Geller, A., & Weinstein, J. N. (2019). *The Root Causes of Health Inequity*. National Library of Medicine; National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK425845/

Batool, S. M., & Liu, Z. (2021). Exploring the relationships between socio-economic indicators and student enrollment in higher education institutions of Pakistan. *PLOS ONE*, *16*(12), e0261577. https://doi.org/10.1371/journal.pone.0261577

Battle, Juan & Lewis, Michael. (2002). The Increasing Significance of Class: The Relative Effects of Race and Socioeconomic Status on Academic Achievement. Journal of Poverty. 6. 21-35. 10.1300/J134v06n02\_02.

Bingham, A. J., Dean, S., & Castillo, J. (2019). Qualitative comparative analysis in educational policy research: Procedures, processes, and possibilities. *Methodological Innovations*, *12*(2), 205979911984098. https://doi.org/10.1177/2059799119840982

Burger, K. (2019). The socio-spatial dimension of educational inequality: A comparative European analysis. *Studies in Educational Evaluation*, *62*, 171–186. https://doi.org/10.1016/j.stueduc.2019.03.009

Chen, P., Wu, L., & Wang, L. (2023). AI Fairness in Data Management and Analytics: A Review on Challenges, Methodologies and Applications. *Applied Sciences*, *13*(18), 10258–10258. https://doi.org/10.3390/app131810258

Cocoara, Z. (2023, November 16). *Understanding Data Encryption: Protecting Sensitive Information in the Digital Age*. Endpoint Protector Blog. https://www.endpointprotector.com/blog/data-encryption-protecting-sensitive-information/

Darling-Hammond, L. (2019). *Inequality in Teaching and Schooling: How Opportunity Is Rationed to Students of Color in America*. Nih.gov; National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK223640/

dawidw. (2023, December 15). *Effective Data Visualization Techniques: Enhancing Research Presentations*. ECORRECTOR. https://ecorrector.com/effective-data-visualization-techniques-enhancing-research-presentations/

Dove, E. S. (2023). *Confidentiality, public interest, and the human right to science: when can confidential information be used for the benefit of the wider community?* *10*(1). https://doi.org/10.1093/jlb/lsad013

Farooq, Muhammad & Chaudhry, A.H. & Shafiq, Muhammad & Berhanu, Girma. (2011). Factors affecting students' quality of academic performance: A case of secondary school level. Journal of Quality and Technology Management. 7. 01-14.

Garcia, E., & Weiss, E. (2017, September 27). *Education inequalities at the school starting gate: Gaps, trends, and strategies to address them*. Economic Policy Institute. https://www.epi.org/publication/education-inequalities-at-the-school-starting-gate/

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and chatgpt in educational settings. *Education Sciences*, *13*(7), 692–692. https://doi.org/10.3390/educsci13070692

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, *3*(3), 275–285. Sciencedirect. https://doi.org/10.1016/j.susoc.2022.05.004

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, *3*(3), 275–285. Sciencedirect. https://doi.org/10.1016/j.susoc.2022.05.004

Heppt, B., Olczyk, M., & Volodina, A. (2022). Number of books at home as an indicator of socioeconomic status: Examining its extensions and their incremental validity for academic achievement. *Social Psychology of Education*, *25*(4), 903–928. https://doi.org/10.1007/s11218-022-09704-8

Herrfahrdt-Pähle, E., Schlüter, M., Olsson, P., Folke, C., Gelcich, S., & Pahl-Wostl, C. (2020). Sustainability transformations: socio-political shocks as opportunities for governance transitions. *Global Environmental Change*, *63*, 102097. https://doi.org/10.1016/j.gloenvcha.2020.102097

Jiang, S., Nocera, A., Tatar, C., Yoder, M. M., Chao, J., Wiedemann, K., Finzer, W., & Rosé, C. P. (2022). An empirical analysis of high school students’ practices of modelling with unstructured data. *British Journal of Educational Technology*, *53*(5), 1114–1133. https://doi.org/10.1111/bjet.13253

Kozlowski, S. W. J., & Ilgen, D. R. (2020). Enhancing the Effectiveness of Work Groups and Teams. *Psychological Science in the Public Interest*, *7*(3), 77–124. Sagepub. https://doi.org/10.1111/j.1529-1006.2006.00030.x

Lombardi, M., & Vannuccini, S. (2022). Understanding emerging patterns and dynamics through the lenses of the cyber-physical universe. *Patterns*, *3*(11), 100601. https://doi.org/10.1016/j.patter.2022.100601

Munir, J., Faiza, M., Jamal, B., Daud, S., & Iqbal, K. (2023). The Impact of Socio-economic Status on Academic Achievement. *JOURNAL of SOCIAL SCIENCES REVIEW*, *3*(2), 695–705. https://doi.org/10.54183/jssr.v3i2.308

Ou, Q., Liang, W., He, Z., Liu, X., Yang, R., & Wu, X. (2023). Investigation and analysis of the current situation of programming education in primary and secondary schools. *Heliyon*, *9*(4), e15530. https://doi.org/10.1016/j.heliyon.2023.e15530

Palmisano, F., Biagi, F., & Peragine, V. (2021). Inequality of Opportunity in Tertiary Education: Evidence from Europe. *Research in Higher Education*. https://doi.org/10.1007/s11162-021-09658-4

Perry, L. B., & Mcconney, A. (2010). Does the SES of the School Matter? An Examination of Socioeconomic Status and Student Achievement Using PISA 2003. *Teachers College Record: The Voice of Scholarship in Education*, *112*(4), 1137–1162. https://doi.org/10.1177/016146811011200401

Policy Implications. (2010). *Programme for International Student Assessment*, 101–121. https://doi.org/10.1787/9789264091504-10-en

Pourhoseingholi, M. A., Baghestani, A. R., & Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology from Bed to Bench*, *5*(2), 79–83. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4017459/

Priya, A. (2020). Case Study Methodology of Qualitative Research: Key Attributes and Navigating the Conundrums in Its Application. *Sociological Bulletin*, *70*(1), 94–110. Sagepub. https://doi.org/10.1177/0038022920970318

Ravaghi, H., Guisset, A.-L., Elfeky, S., Nasir, N., Khani, S., Ahmadnezhad, E., & Abdi, Z. (2023). A scoping review of community health needs and assets assessment: concepts, rationale, tools and uses. *BMC Health Services Research*, *23*(1). https://doi.org/10.1186/s12913-022-08983-3

Ross, S. (2021, March 24). *What is human capital and how is it used?* Investopedia. https://www.investopedia.com/ask/answers/032715/what-human-capital-and-how-it-used.asp

Shang, B. (2022). Tackling Gender Inequality: Definitions, Trends, and Policy Designs. *IMF Working Papers*, *2022*(232). https://doi.org/10.5089/9798400224843.001.A001

Toutkoushian, R. K., & Curtis, T. (2005). Effects of Socioeconomic Factors on Public High School Outcomes and Rankings. *The Journal of Educational Research*, *98*(5), 259–271. https://doi.org/10.3200/joer.98.5.259-271

United Nations. (2021, February 18). *Leveraging digital technologies for social inclusion | DISD*. Www.un.org. https://www.un.org/development/desa/dspd/2021/02/digital-technologies-for-social-inclusion/

Williams, W. R., & Reppond, H. A. (2020). More Than Just Hard Work: Educational Policies to Facilitate Economic Mobility. *Policy Insights from the Behavioral and Brain Sciences*, *7*(2), 165–172. https://doi.org/10.1177/2372732220943912

Worthy, L. D., Lavigne, T., & Romero, F. (2020). Socioeconomic Status (SES). *Open.maricopa.edu*. https://open.maricopa.edu/culturepsychology/chapter/socioeconomic-status-ses/#:~:text=Socioeconomic%20status%20(SES)%20is%20an