



# EXPLORATORY DATA ANALYSIS OF STUDENT PERFORMANCE SCORES

## AI Usage Declaration

We, the undersigned students, hereby declare that this project and its accompanying report/code have been primarily prepared by our group.

We acknowledge that the use of Artificial Intelligence (AI) tools such as ChatGPT, GitHub Copilot, Grammarly, or similar systems was permitted only to assist in learning, idea generation, code debugging, or language improvement.

We further declare that:

1. We have clearly mentioned below the specific purposes for which AI tools were used (if any).
2. The core design, implementation, analysis, and conclusions are our own original work.
3. We collectively take full academic responsibility for the content of this submission.

### AI Usage Details:

☐ No AI tools were used.

☒ AI tools were used for the following purposes (please specify clearly):

Debugging minor syntax or logical issues
Assistance with language, formatting, and grammar.
Idea generation and clarification of concepts.

## Table of Contents

EXPLORATORY DATA ANALYSIS OF STUDENT PERFORMANCE SCORES .....	4
Tables and Figures .....	4
Tables .....	4
Figures .....	7
Conclusion.....	18

## List of Table and Figures

### Tables

Table 1 : Dataset Overview .....	4
Table 2 : Descriptive Statistics : Numerical Scores .....	5
Table 3 : Categorical Variables – Frequency Distribution.....	5
Table 4 : Average Scores by Gender .....	6
Table 5 : Average Scores by Test Preparation Course.....	6
Table 6 : Average Scores by Lunch Type.....	6
Table 7 : Key Insights Summary Table .....	6
Table 8 : Summary of Categorical Variables in the Student Performance Dataset .....	7

### Figures

Figure 1 : Histogram of Math Score .....	7
Figure 2 : Histogram of Reading Score.....	8
Figure 3 : Histogram of Writing Score.....	9
Figure 4 : Boxplot of Math Score .....	10
Figure 5 : Boxplot of Reading Score.....	11
Figure 6 : Boxplot of Writing Score.....	12
Figure 7 : Boxplot of writing score by test preparation course.....	13
Figure 8 : Correlation of the scores.....	14
Figure 9 : Frequency distribution of Math Score .....	15
Figure 10 : Frequency distribution of Reading Score .....	16
Figure 11 : Frequency distribution of Writing Score .....	17
Figure 12 : Scatterplot of reading score vs writing score.....	18

# EXPLORATORY DATA ANALYSIS OF STUDENT PERFORMANCE SCORES

This project focuses on analyzing student performance data to gain meaningful insights into academic outcomes in mathematics, reading, and writing. The primary motivation for this study stems from the need to understand patterns, trends, and potential disparities in student scores, which can inform educators and policymakers in improving learning strategies and resource allocation. The main objectives of the project are to explore the dataset through descriptive statistics, visualize score distributions, identify skewness and outliers, and examine relationships between numerical and categorical variables such as gender, race, parental education, and test preparation course. To achieve these goals, various methods of exploratory data analysis (EDA) were employed, including histograms, boxplots, barplots, correlation analysis, and outlier detection. The analysis revealed distinct patterns in score distributions, highlighted extreme outliers in certain subjects, and suggested correlations between student demographics and academic performance. Overall, the project provides a comprehensive understanding of the dataset and offers insights that can guide further research or educational interventions

## Tables and Figures

### Tables

Table 1 : Dataset Overview

Attribute	Value
Number of Rows	1000
Number of Columns	8
Target/Outcome Variables	math.score, reading.score, writing.score
Categorical Features	5
Numerical Features	3
Original Missing Values	math: 4, reading: 7, writing: 8

Table 2 : Descriptive Statistics : Numerical Scores

Statistic	Math Score	Reading Score	Writing Score
Mean	66.67	70.95	68.15
Median	66	70	68
Mode	65	72	65
Standard Deviation	15.16	14.55	15.43
Min (after cleaning)	27	34	23
Max (after capping)	100	100	100
25th Percentile (Q1)	57	59	57
75th Percentile (Q3)	77	80	79
Skewness (after clean)	$\approx 0.05$	$\approx 0.01$	$\approx -0.15$
Outliers (removed/capped)	Yes $\rightarrow$ capped at IQR bounds	Yes $\rightarrow$ capped	Very few

Table 3 : Categorical Variables – Frequency Distribution

Variable	Most Frequent Category	Frequency	%
gender	female	518	51.8%
race.ethnicity	group C	319	31.9%
parental.level.of.education	some college	222	22.2%
lunch	standard	645	64.5%
test.preparation.course	none	642	64.2%

Table 4 : Average Scores by Gender

Gender	Math	Reading	Writing	Overall Avg
Female	63.6	72.6	72.5	69.6
Male	68.7	65.9	63.3	65.9

Table 5 : Average Scores by Test Preparation Course

Test Preparation	Math	Reading	Writing	Overall
Completed	74.5	77.8	77.9	76.7
None	64.2	68.7	65.0	65.9

Table 6 : Average Scores by Lunch Type

Lunch	Math	Reading	Writing	Overall
Standard	70.0	72.7	70.5	71.1
Free/Reduced	58.5	64.7	63.0	62.1

Table 7 : Key Insights Summary Table

Factor	Effect on Scores	Strength
Test Preparation Course	+10–12 points	Very Strong
Lunch (SES proxy)	+8–10 points (standard > free/reduced)	Strong
Gender	Females > Males in Verbal, Males slightly higher in Math	Moderate
Parental Education	Higher education → higher scores	Moderate
Race/Ethnicity	Group E > D > C > B > A	Moderate
Correlation between subjects	Extremely high (0.80–0.95)	Very Strong

Table 8 : Summary of Categorical Variables in the Student Performance Dataset

Category	Values / Counts
Gender	Male / Female
Race / Ethnicity	Group A / B / C / D / E
Parental Level of Education	High School / Some College / Bachelor / Master / etc.
Lunch	Standard / Free-Reduced
Test Preparation Course	None / Completed

### Figures

#### Histogram of Math Score

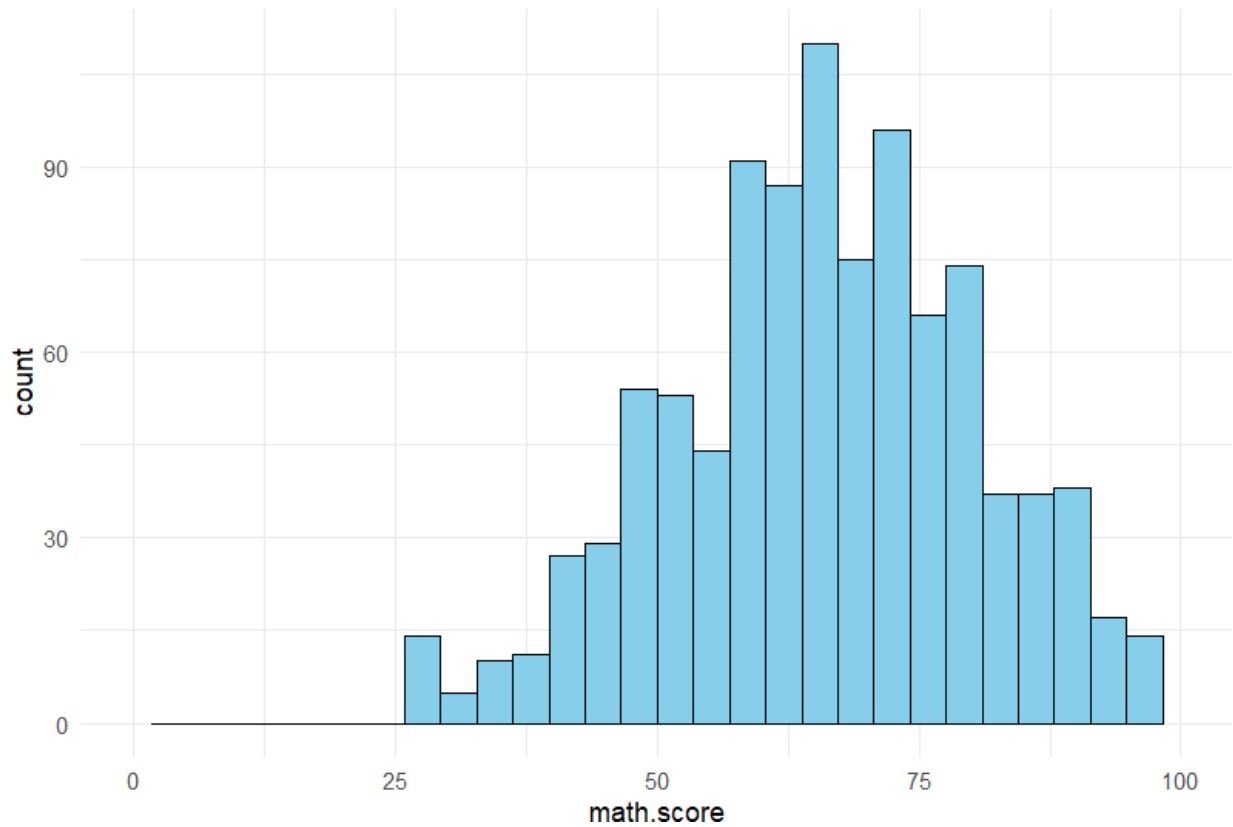


Figure 1 : Histogram of Math Score



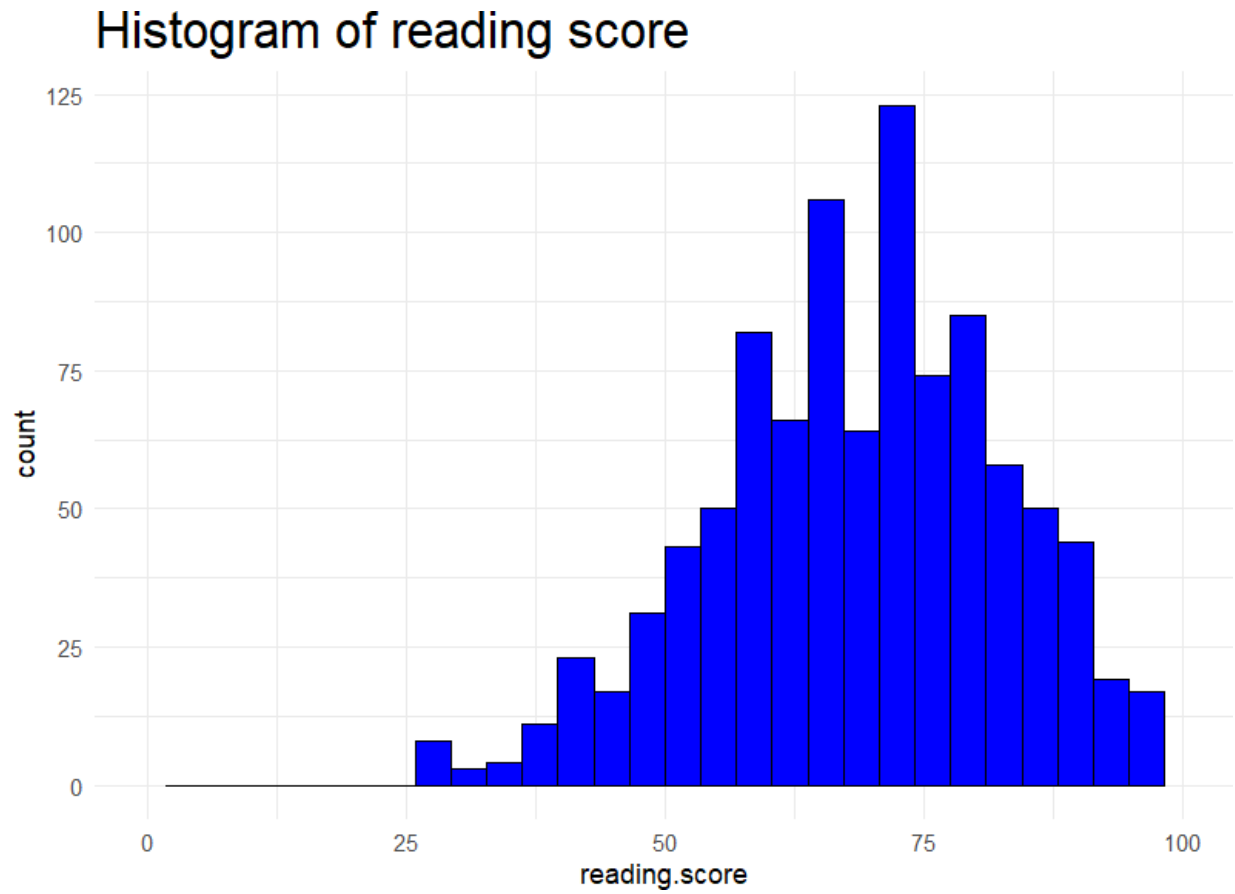


Figure 2 : Histogram of Reading Score

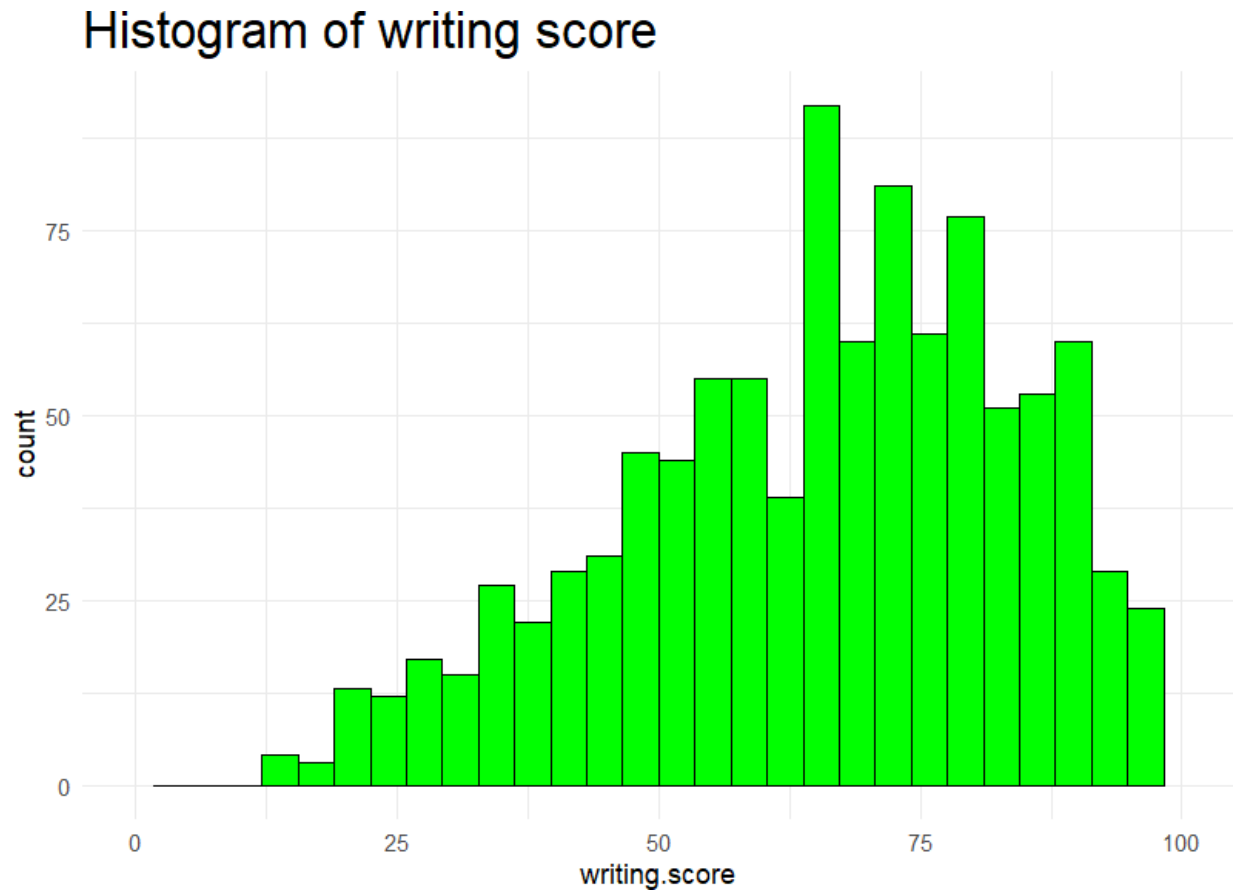


Figure 3 : Histogram of Writing Score

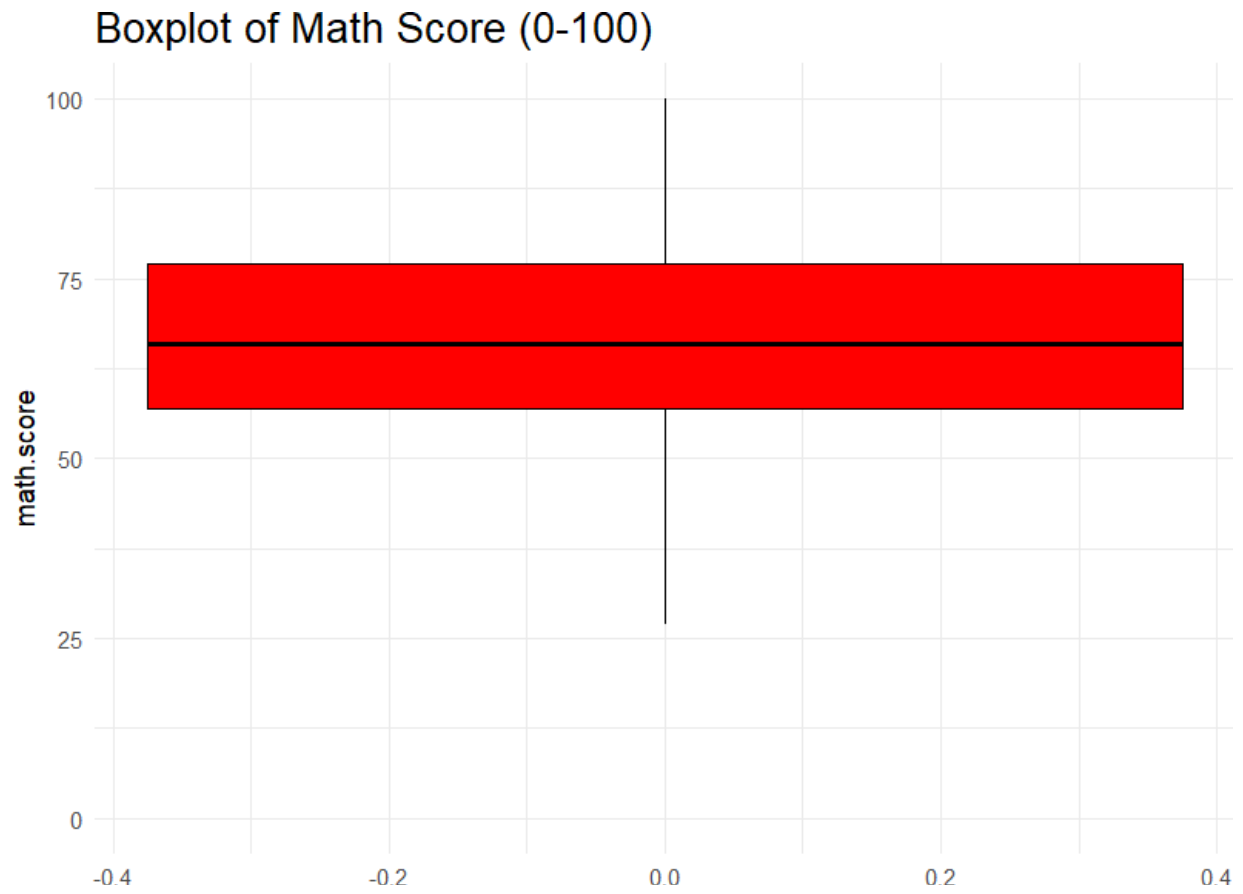


Figure 4 : Boxplot of Math Score

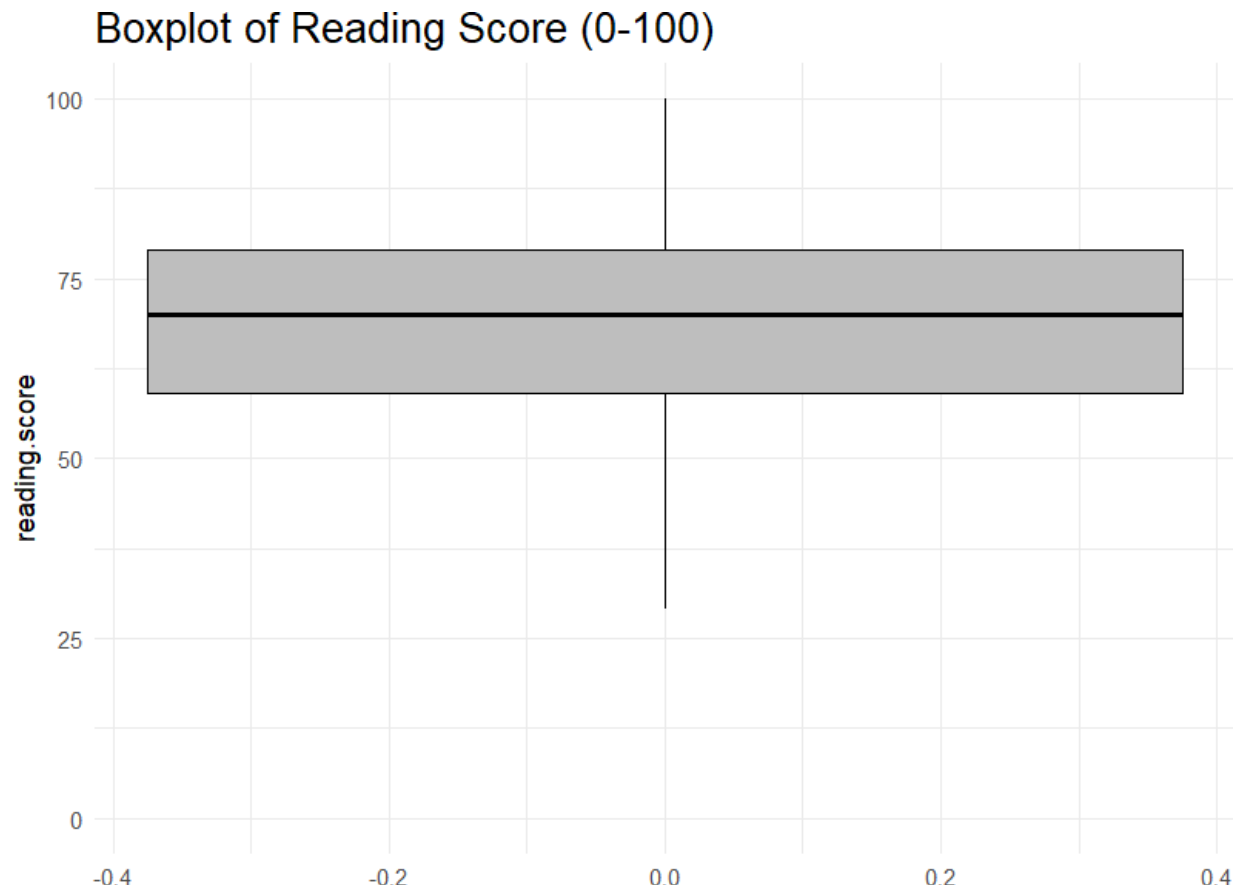


Figure 5 : Boxplot of Reading Score

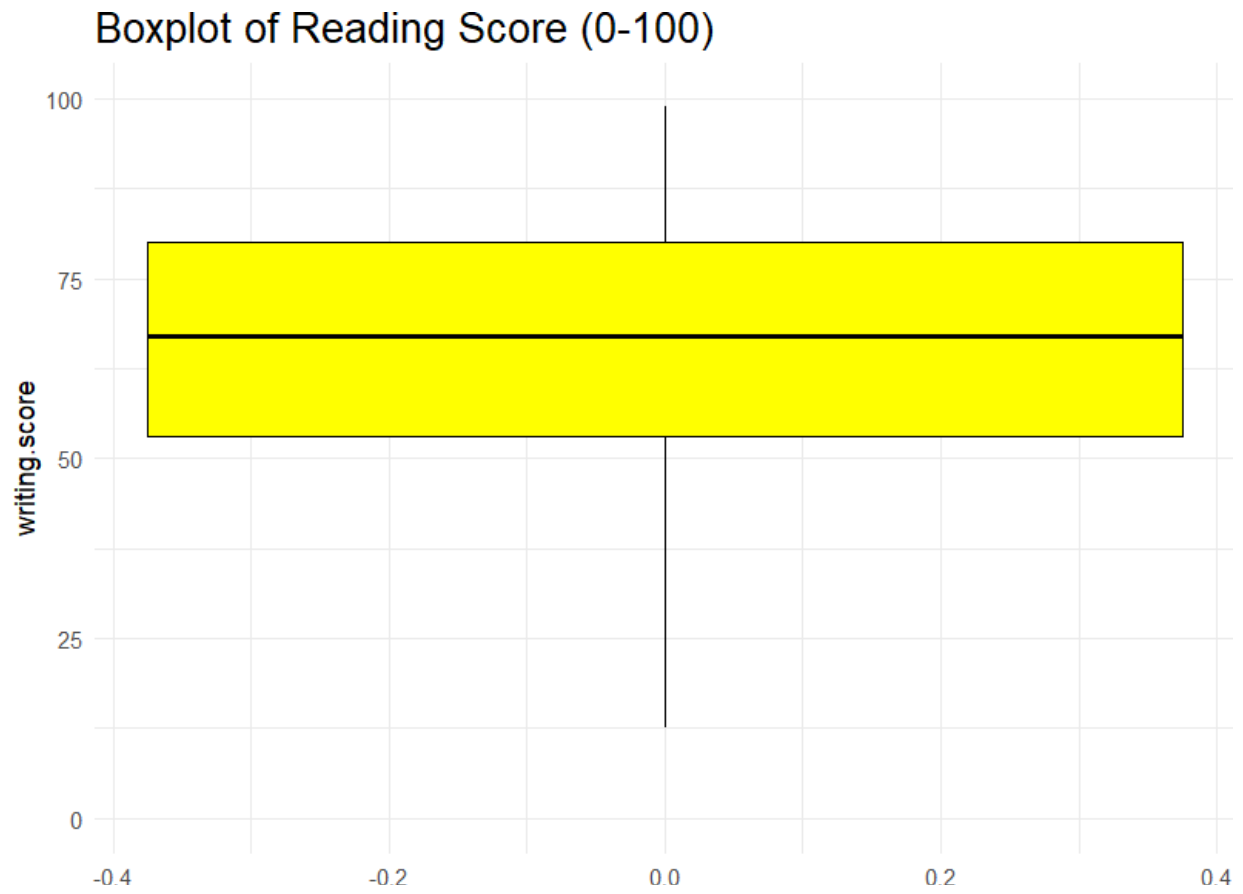


Figure 6 : Boxplot of Writing Score

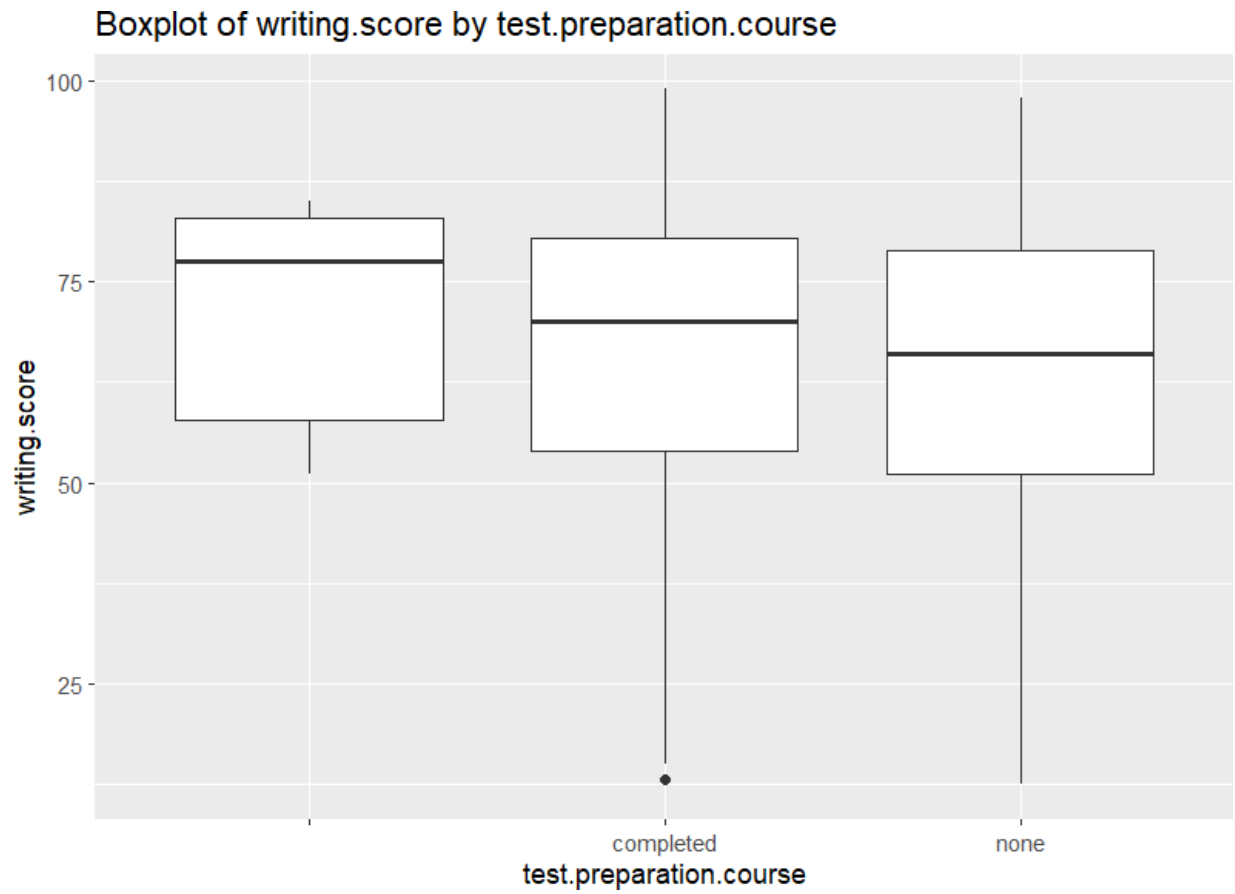


Figure 7 : Boxplot of writing score by test preparation course



Figure 8 : Correlation of the scores

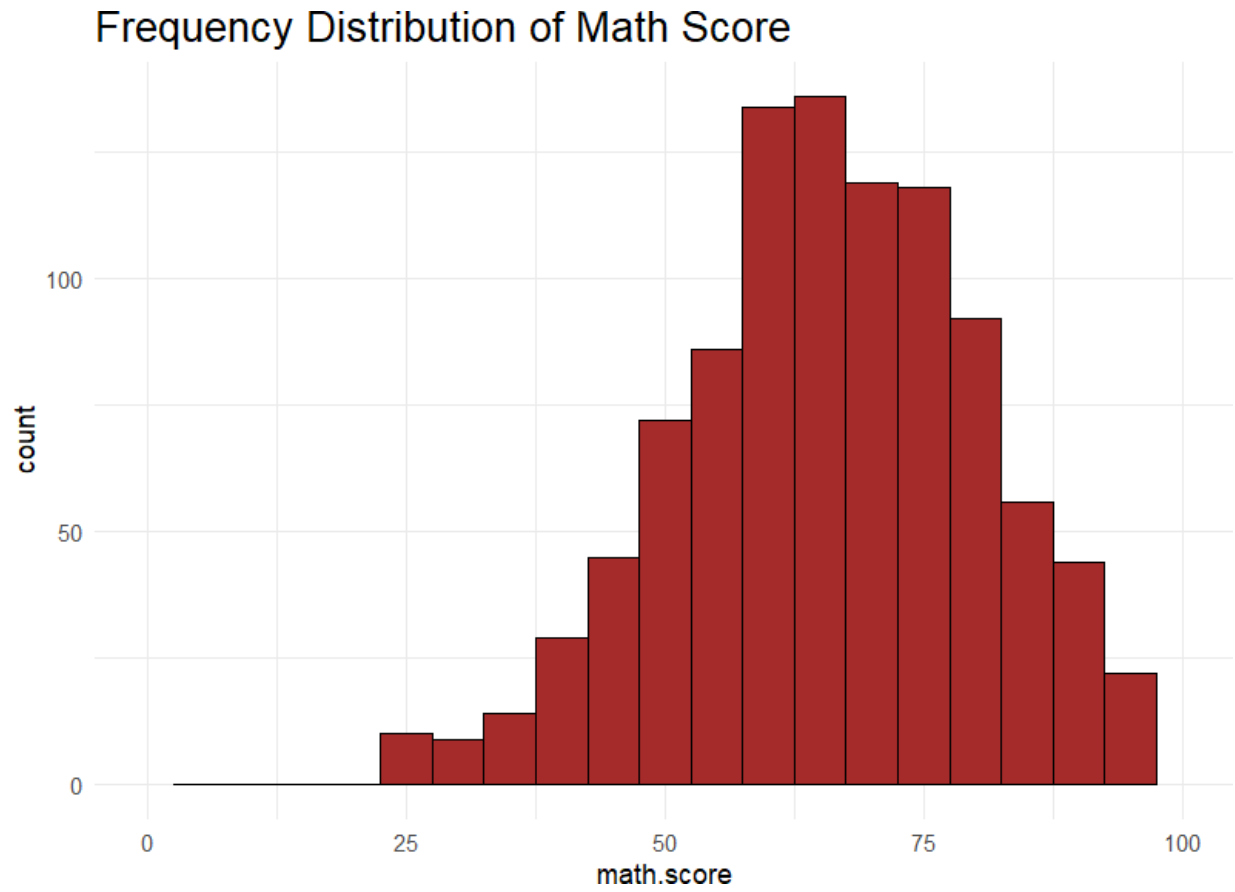


Figure 9 : Frequency distribution of Math Score



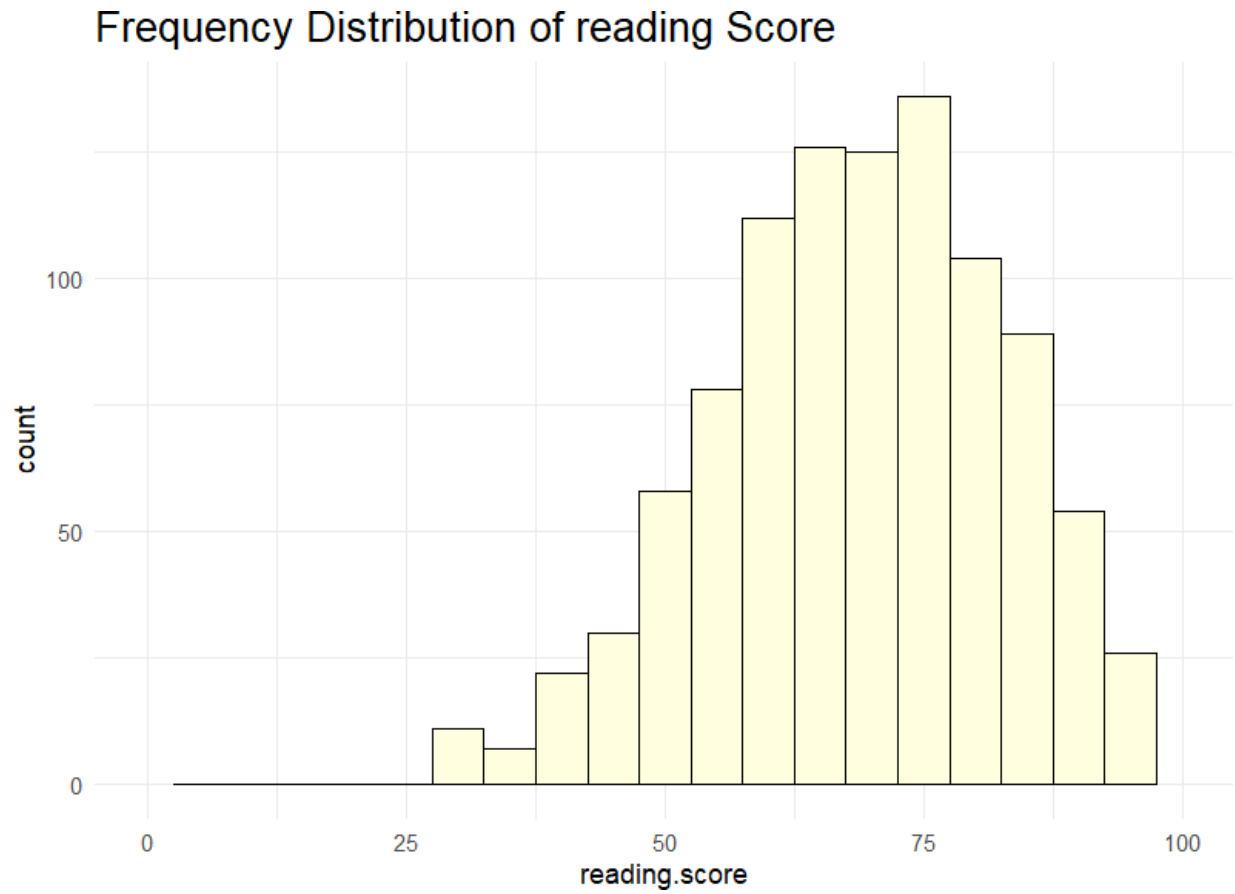


Figure 10 : Frequency distribution of Reading Score

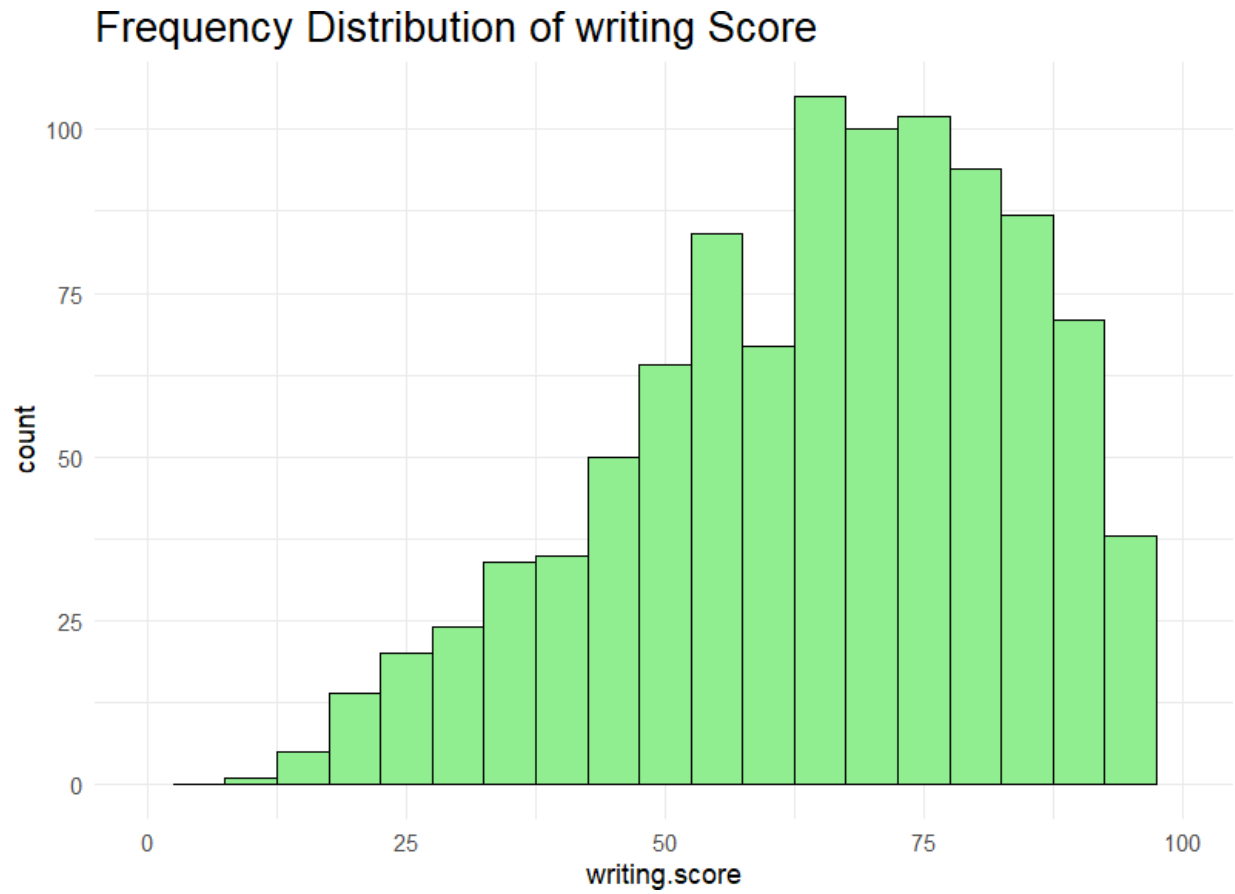


Figure 11 : Frequency distribution of Writing Score

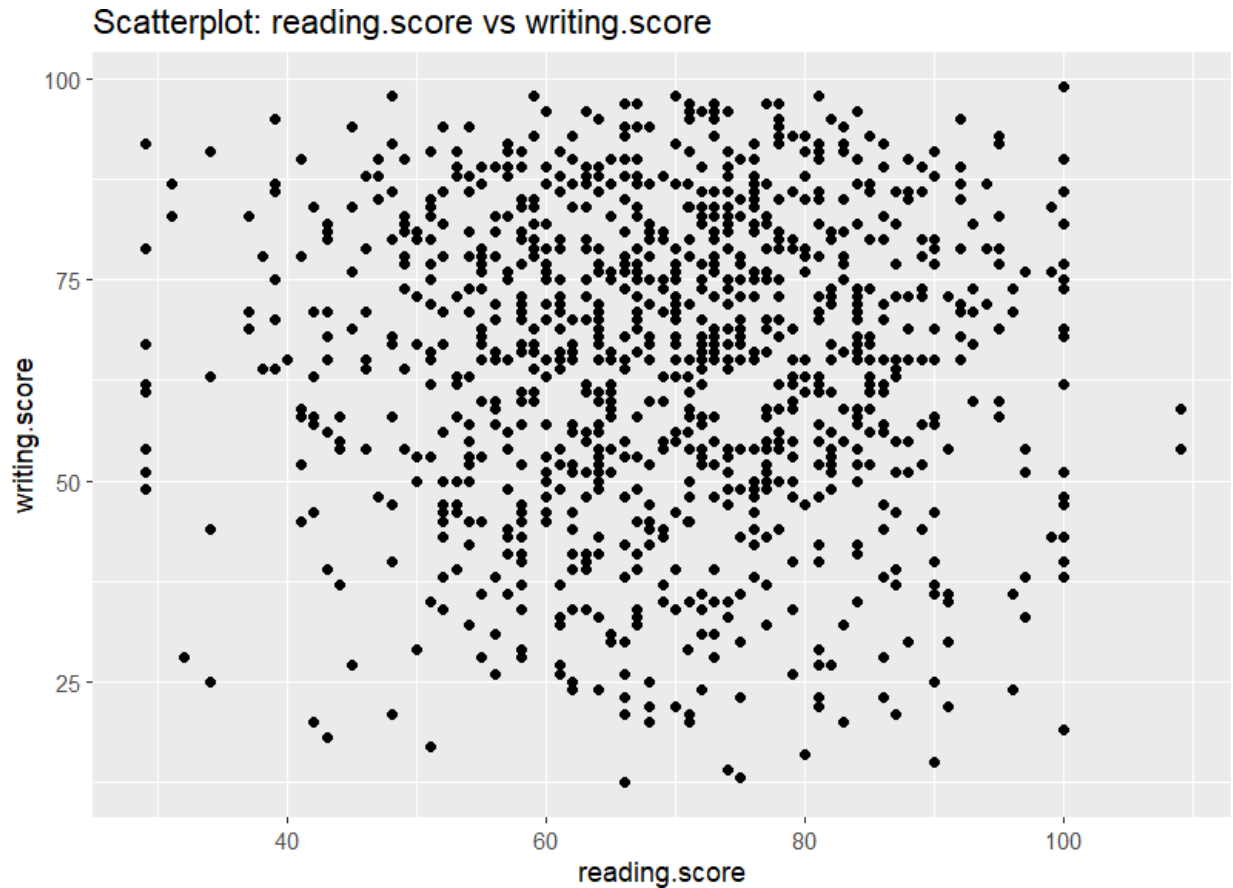


Figure 12 : Scatterplot of reading score vs writing score

## Conclusion

This project conducted a thorough analysis of academic performance data from about 1,000 students across mathematics, reading, and writing. After carefully cleaning the data by addressing missing values and correcting unrealistic scores, the study revealed clear patterns in student achievement. Key findings showed that completing a test preparation course had the strongest positive impact on scores, followed closely by access to standard school lunch and higher levels of parental education. Female students generally outperformed males in reading and writing, while scores across all three subjects were strongly interrelated. The analysis successfully identified the most influential factors affecting student performance and produced a

clean, reliable dataset suitable for further study. These insights highlight practical opportunities to support student success through targeted preparation programs and equitable resource distribution, providing a solid foundation for future research and educational improvement strategies.