Univariate Analysis - Student Performance Dataset

Raju Ahmed

2025-12-27

1. Introduction

In this document, I present my univariate analysis for the Student Performance dataset. My analysis covers the following variables:

Binary/Nominal Variables: sex, paid, activities, higher, internet

Ordinal Variables: Medu (Mother's Education), famrel (Family Relationships)

Numeric Discrete Variable: age

Additionally, I include one bivariate analysis: **Medu vs G3** (Mother's Education vs Final Grade).

2. Data Loading

```
# Load the selected dataset
data <- read_csv("student-mat-selected.csv")

# Display structure
glimpse(data)
```

```
## Rows: 395
## Columns: 13
## $ sex        <chr> "F", "F", "F", "F", "F", "M", "M", "F", "M", "M", "F", "F",~
## $ age        <dbl> 18, 17, 15, 15, 16, 16, 16, 17, 15, 15, 15, 15, 15, 15, 15,~
## $ Medu       <dbl> 4, 1, 1, 4, 3, 4, 2, 4, 3, 3, 4, 2, 4, 4, 2, 4, 4, 3, 3, 4,~
## $ traveltime <dbl> 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 3, 1, 2, 1, 1, 1, 3, 1, 1,~
## $ studytime  <dbl> 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 3, 1, 2, 3, 1, 3, 2, 1, 1,~
## $ failures   <dbl> 0, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0,~
## $ paid       <chr> "no", "no", "yes", "yes", "yes", "yes", "no", "no", "yes", ~
## $ activities <chr> "no", "no", "no", "yes", "no", "yes", "no", "no", "no", "ye~
## $ higher     <chr> "yes", "yes", "yes", "yes", "yes", "yes", "yes", "yes", "ye~
## $ internet   <chr> "no", "yes", "yes", "yes", "no", "yes", "yes", "no", "yes",~
## $ famrel     <dbl> 4, 5, 4, 3, 4, 5, 4, 4, 4, 5, 3, 5, 4, 5, 4, 4, 3, 5, 5, 3,~
## $ absences   <dbl> 6, 4, 10, 2, 4, 10, 0, 6, 0, 0, 0, 4, 2, 2, 0, 4, 6, 4, 16,~
## $ G3         <dbl> 6, 6, 10, 15, 10, 15, 11, 6, 19, 15, 9, 12, 14, 11, 16, 14,~
```

```
## Dataset Dimensions: 395 rows x 13 columns
```

---

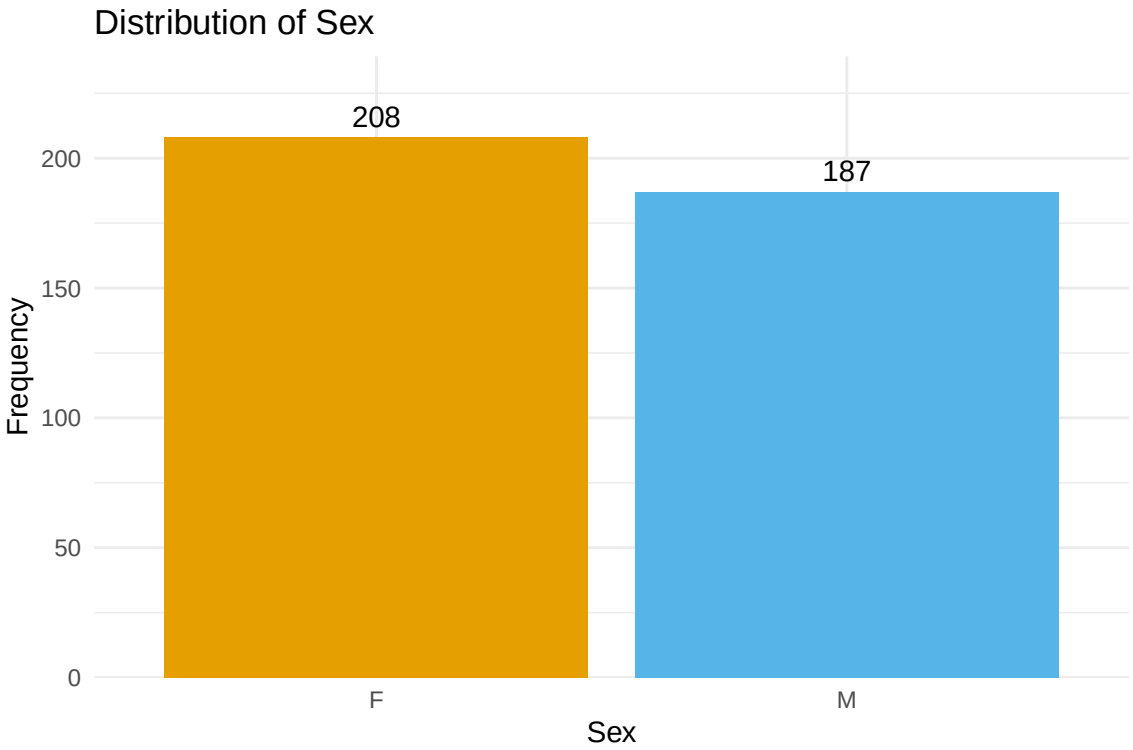
3. Univariate Analysis - Binary Variables

3.1 Sex (Student's Sex)

Table 1: Frequency Distribution of Sex

Category	Absolute_Freq	Relative_Freq
F	208	0.527
M	187	0.473

```
##
## Mode: F
```



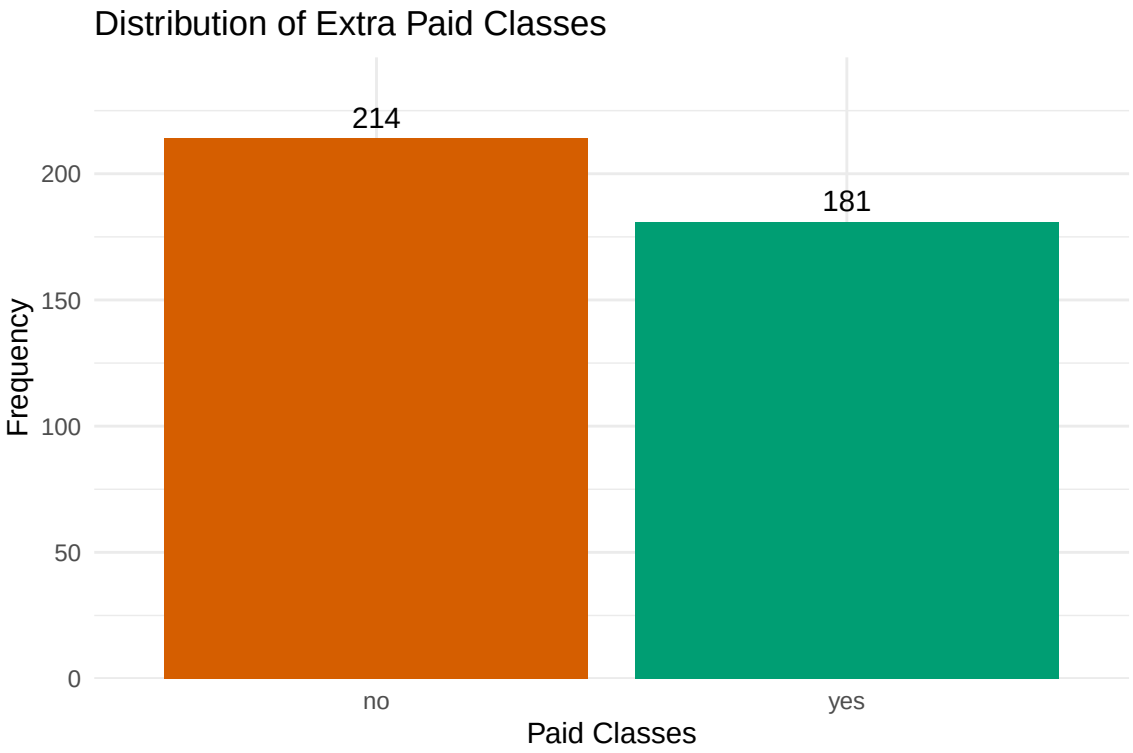
Interpretation: The dataset contains 208 female and 187 male students. Females represent 52.7% of the sample.

3.2 Paid (Extra Paid Classes)

Table 2: Frequency Distribution of Paid Classes

Category	Absolute_Freq	Relative_Freq
no	214	0.542
yes	181	0.458

```
##
## Mode: no
```



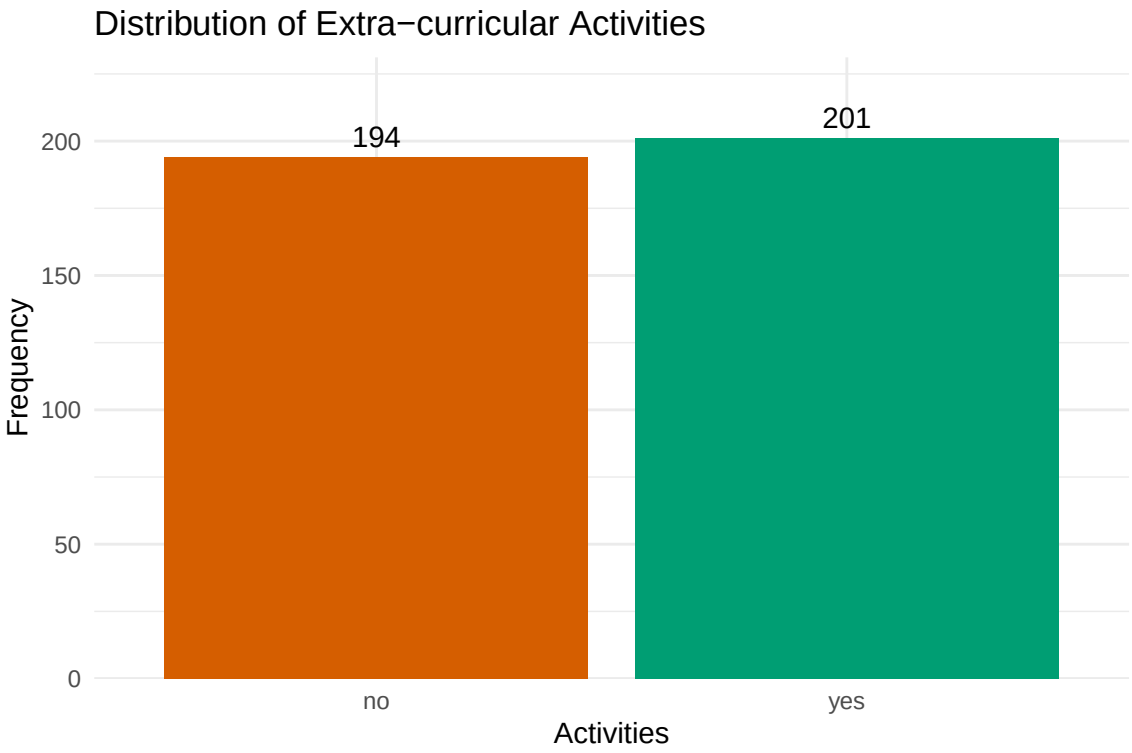
Interpretation: 45.8% of students take extra paid classes within the Math course.

3.3 Activities (Extra-curricular Activities)

Table 3: Frequency Distribution of Extra-curricular Activities

Category	Absolute_Freq	Relative_Freq
no	194	0.491
yes	201	0.509

```
##
## Mode: yes
```



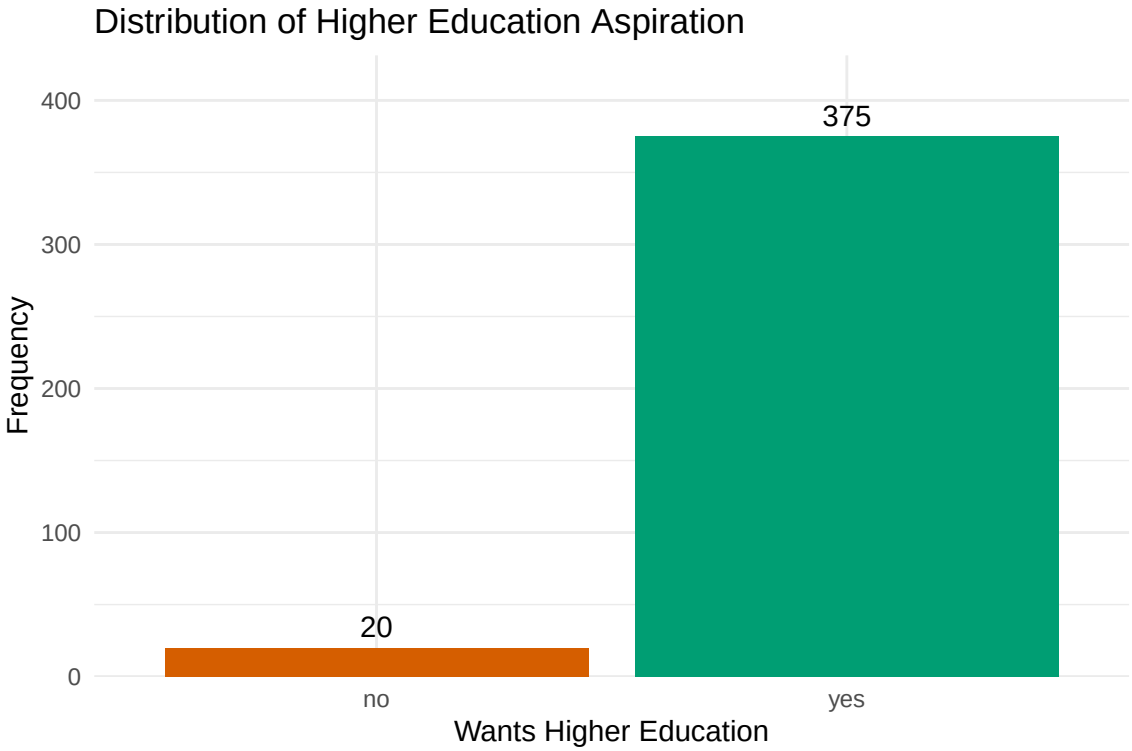
Interpretation: 50.9% of students participate in extra-curricular activities.

3.4 Higher (Wants Higher Education)

Table 4: Frequency Distribution of Higher Education Aspiration

Category	Absolute_Freq	Relative_Freq
no	20	0.051
yes	375	0.949

```
##
## Mode: yes
```



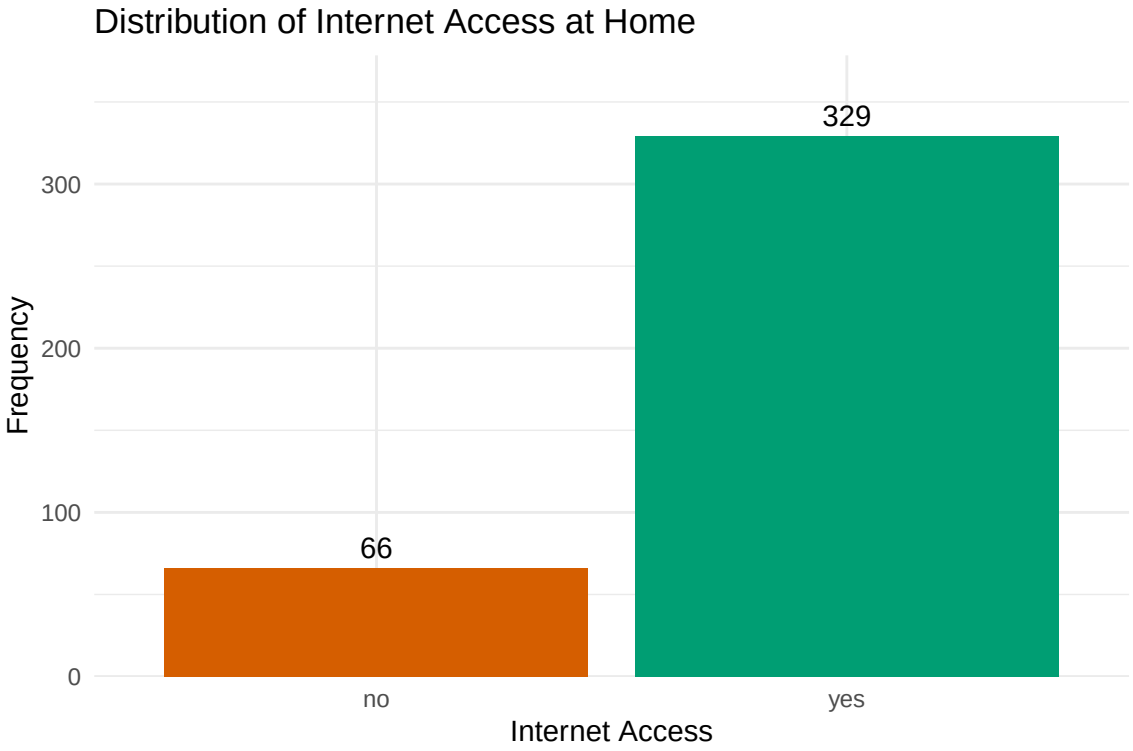
Interpretation: An overwhelming 94.9% of students want to pursue higher education.

3.5 Internet (Internet Access at Home)

Table 5: Frequency Distribution of Internet Access

Category	Absolute_Freq	Relative_Freq
no	66	0.167
yes	329	0.833

```
##
## Mode: yes
```



Interpretation: 83.3% of students have internet access at home.

---

4. Univariate Analysis - Ordinal Variables

4.1 Medu (Mother's Education Level)

Mother's education is coded as:

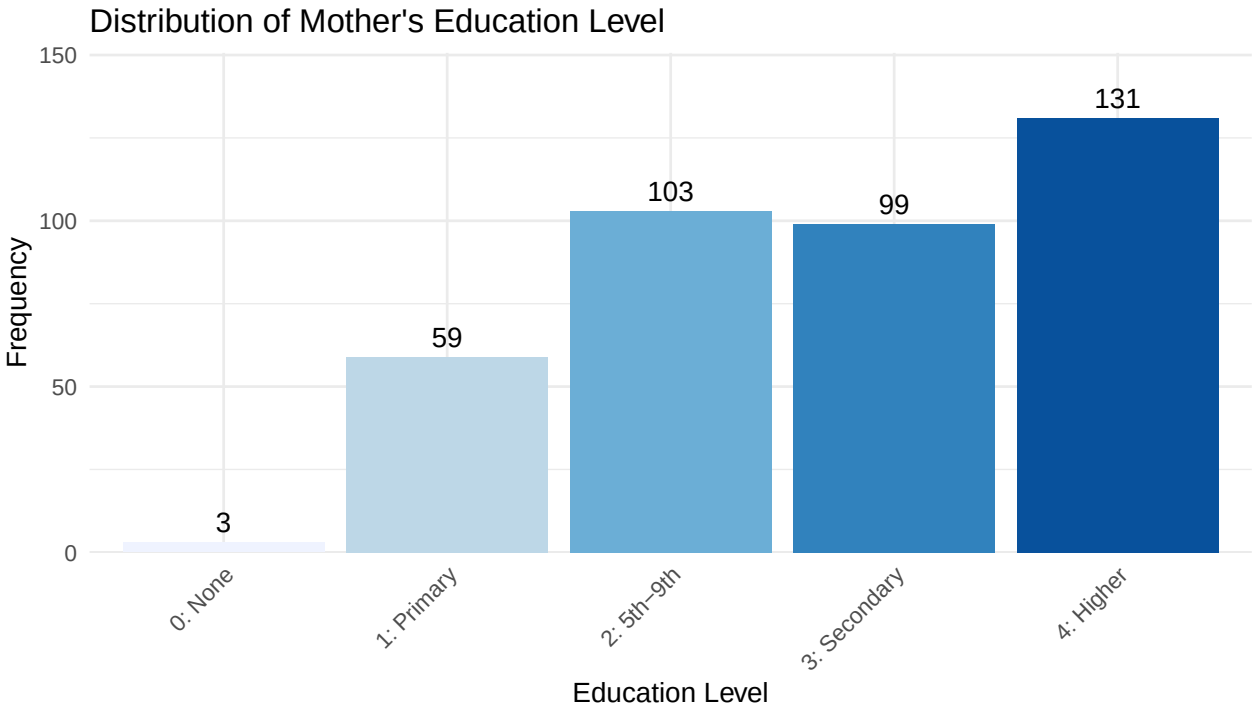
- 0: None
- 1: Primary education (4th grade)
- 2: 5th to 9th grade
- 3: Secondary education
- 4: Higher education

Table 6: Frequency Distribution of Mother's Education

Level	Absolute_Freq	Relative_Freq	Cumulative_Freq
0	3	0.008	0.008
1	59	0.149	0.157
2	103	0.261	0.418
3	99	0.251	0.668
4	131	0.332	1.000

```
##
## Mode: 4
```

```
##
## Median: 3
```



Interpretation: The most common mother's education level is 4 (mode), with a median of 3. This indicates that most mothers have at least secondary education or higher.

4.2 Famrel (Quality of Family Relationships)

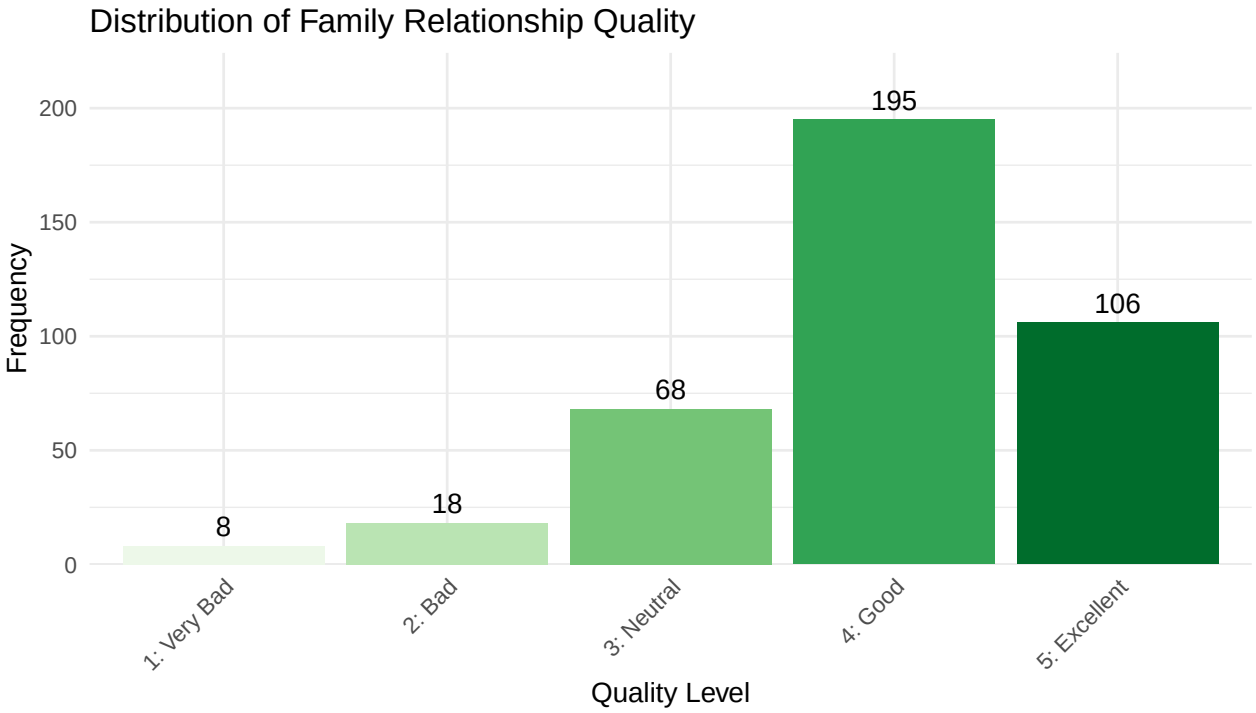
Family relationship quality is coded on a scale from 1 (very bad) to 5 (excellent).

Table 7: Frequency Distribution of Family Relationships

Level	Absolute_Freq	Relative_Freq	Cumulative_Freq
1	8	0.020	0.020
2	18	0.046	0.066
3	68	0.172	0.238
4	195	0.494	0.732
5	106	0.268	1.000

```
##
## Mode: 4
```

```
##
## Median: 4
```



Interpretation: Most students report good to excellent family relationships (mode = 4, median = 4). Only a small minority report poor family relationships.

---

5. Univariate Analysis - Numeric Variable

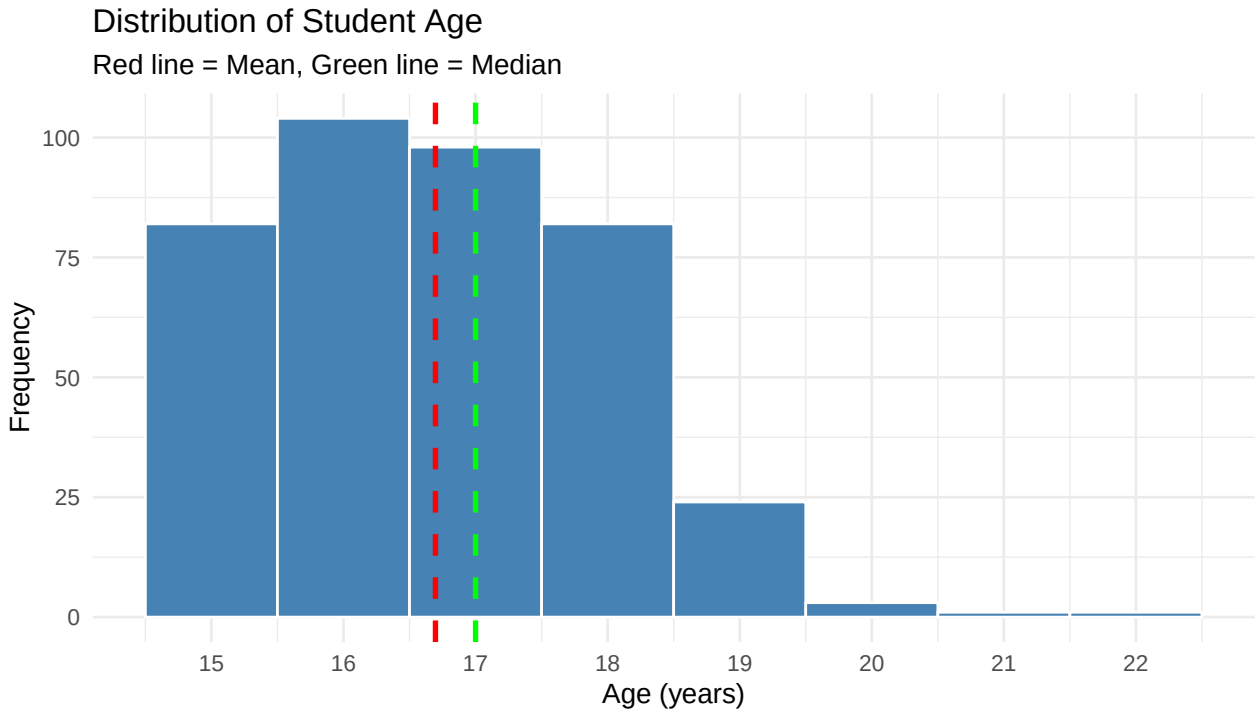
5.1 Age (Student's Age)

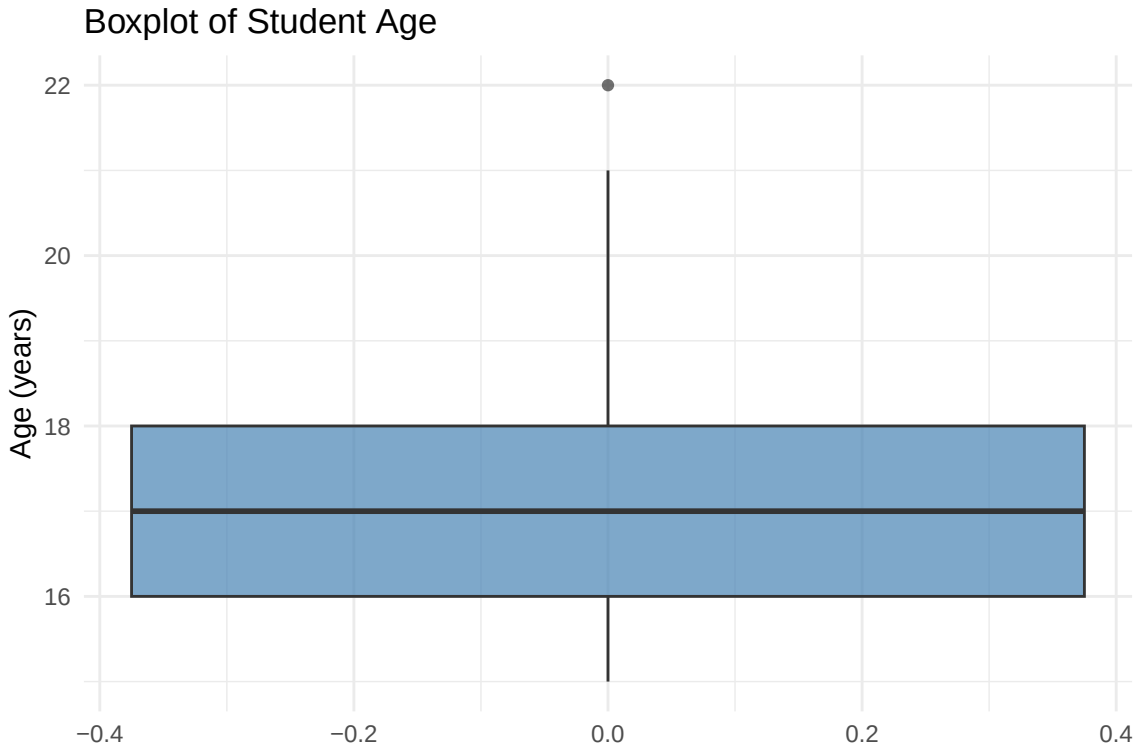
```
## === Central Tendency ===
## Mean: 16.7
## Median: 17
## Mode: 16
##
## === Dispersion ===
## Variance: 1.63
## Standard Deviation: 1.28
## Range: 15 - 22
## IQR: 2
## Coefficient of Variation: 7.64 %
##
## === Five-Number Summary ===
## [1] 15 16 17 18 22
##
## === Quartiles ===
```

```
##   0%  25%  50%  75% 100%
##   15   16   17   18   22
```

Table 8: Frequency Distribution of Age

Age	Absolute_Freq	Relative_Freq	Cumulative_Freq
15	82	0.208	0.208
16	104	0.263	0.471
17	98	0.248	0.719
18	82	0.208	0.927
19	24	0.061	0.987
20	3	0.008	0.995
21	1	0.003	0.997
22	1	0.003	1.000





Interpretation:

- Students' ages range from 15 to 22 years
- Mean age is 16.7 years (SD = 1.28)
- The distribution is slightly right-skewed (mean > median), indicating some older students
- Most students (IQR) are between 16 and 18 years old
- The mode is 16 years, the most common age

---

6. Summary Table - All Variables

Table 9: Summary Statistics for Raju's Variables

Variable	Type	n	Mode	Median	Mean	SD
sex	Binary	395	F	-	-	-
age	Numeric	395	16	17	16.7	1.28
Medu	Ordinal	395	4	3	-	-
paid	Binary	395	no	-	-	-
activities	Binary	395	yes	-	-	-
higher	Binary	395	yes	-	-	-
internet	Binary	395	yes	-	-	-
famrel	Ordinal	395	4	4	-	-

---

7. Bivariate Analysis: Mother's Education vs Final Grade

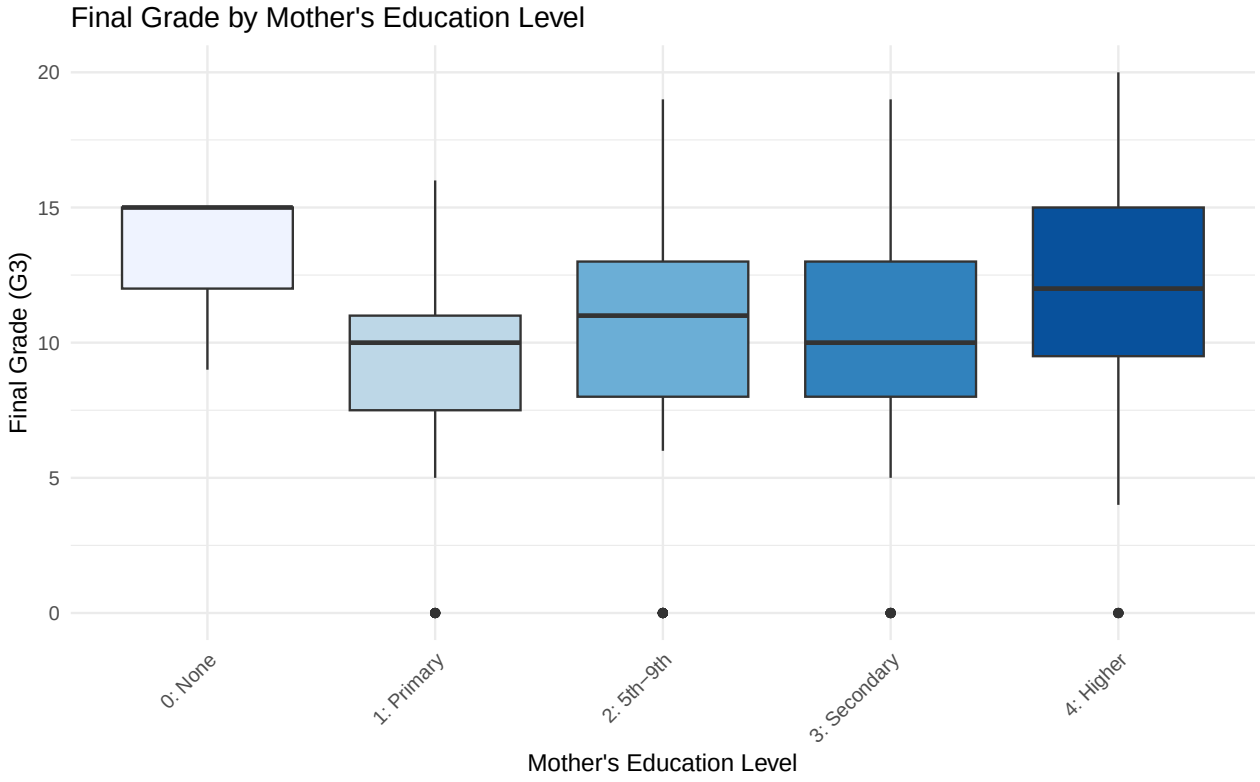
This section examines the relationship between mother's education level (Medu) and student's final grade (G3).

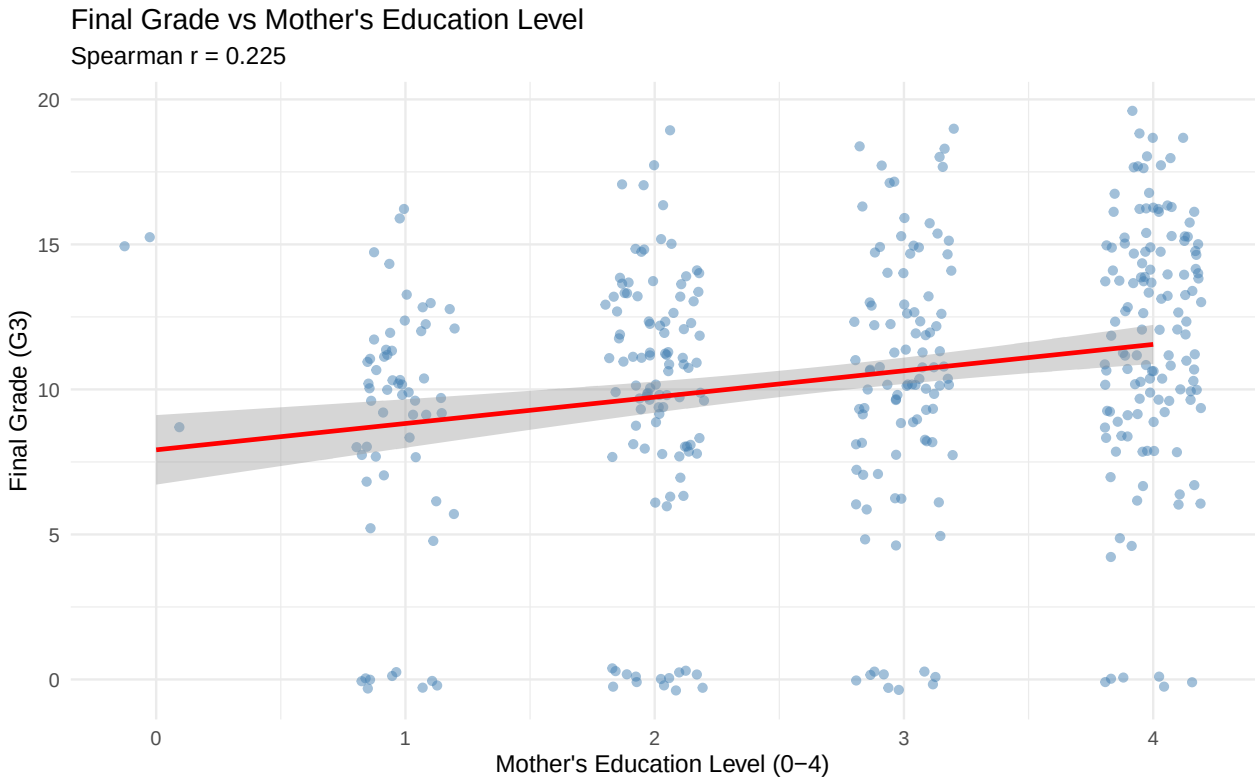
Table 10: Final Grade (G3) Statistics by Mother's Education Level

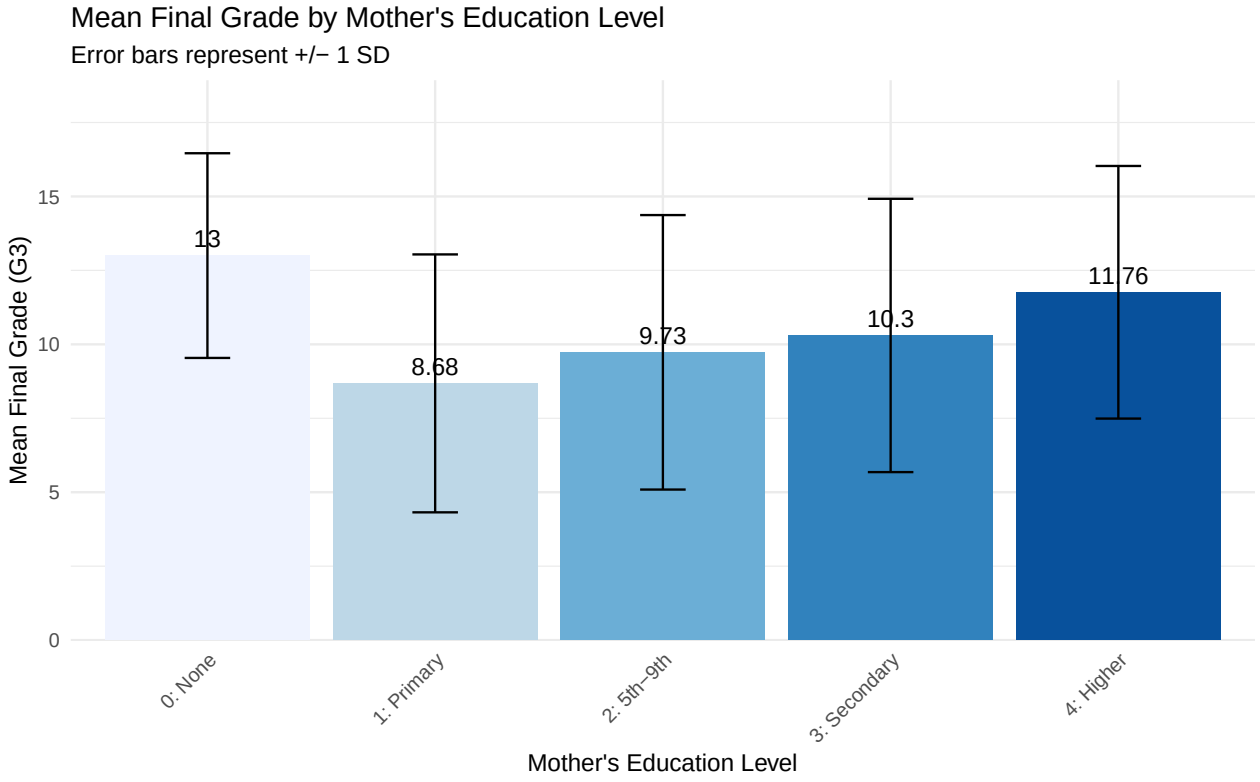
Medu	n	Mean_G3	SD_G3	Median_G3	Min_G3	Max_G3
0	3	13.00	3.46	15	9	15
1	59	8.68	4.36	10	0	16
2	103	9.73	4.64	11	0	19
3	99	10.30	4.62	10	0	19
4	131	11.76	4.27	12	0	20

```
## Spearman Correlation (Medu vs G3): 0.225
```

```
## Pearson Correlation (Medu vs G3): 0.217
```







Interpretation:

- There is a **positive correlation** between mother's education and student's final grade (Spearman r = 0.225)
- **Important caveat:** Medu = 0 shows a high mean (13.0), but this is a **small sample artifact** with

only n = 3 students (grades: 9, 15, 15). This group should be excluded from trend interpretation.

- **Excluding Medu = 0**, there is a clear positive trend: as mother's education increases from level 1 to 4, mean grades increase ($8.68 \rightarrow 9.73 \rightarrow 10.30 \rightarrow 11.76$)
- Students with mothers who have higher education (level 4) achieve the highest mean grade (11.76)
- Students with mothers having only primary education (level 1) have the lowest mean grade (8.68)
- The relationship suggests that **parental education is a meaningful predictor** of student academic performance