NITYA NIMMAGADDA  
FINAL PROJECT  
12/13/23   
  
1)  
**Research Scenario**

Scenario: Analyzing Factors Influencing Academic Performance in High School Mathematics

The focus on personalized education is growing in today's academic world, particularly in understanding what affects students' success. Our study centers on the performance of high school students in math, a subject many find difficult. We aim to explore different factors that could impact their learning outcomes. By analyzing data from a large urban high school, which includes everything from demographic details to study habits and the type of support students receive at school, we hope to gain insights that can help improve teaching methods, shape educational policies, and enhance support for students. This extensive dataset offers a chance to really understand what drives academic success and pinpoint the crucial elements that affect how students learn and perform.

**Research Questions**

Primary Question: What is the impact of study time on students' final grades in high school mathematics?  
  
This question seeks to investigate how directly the time students spend studying math correlates with their achievement in the subject.

**Secondary Questions:**

How do demographic factors, such as gender influence mathematics performance?

Does receiving school or family support correlate with better academic outcomes in mathematics?  
  
These questions let us have a comprehensive look at student achievement, taking into account personal factors such as the time spent studying and external factors like support received. The results of this study could be very impactful for teachers, education authorities, and students themselves, providing important information that could help improve teaching approaches and the support systems available in high school math education.  
  
2) <https://www.kaggle.com/datasets/dillonmyrick/high-school-student-performance-and-demographics>  
  
The dataset student\_math\_clean.csv contains detailed information about students, their demographic backgrounds, and academic performance in mathematics.

Here is a brief overview of some of the key variables in the dataset:

student\_id: Unique identifier for each student.

school: School attended by the student.

sex: Gender of the student (F/M).

age: Age of the student.

address\_type: Type of address (Urban/Rural).

family\_size: Family size (Greater than 3/Less than or equal to 3).

parent\_status: Parental cohabitation status.

mother\_education, father\_education: Education level of mother and father.

mother\_job, father\_job: Occupation of mother and father.

travel\_time: Time taken to travel from home to school.

study\_time: Weekly study time.

failures: Number of past class failures.

school\_support, family\_support: Whether the student receives educational support from school or family.

paid\_classes: Whether the student has paid classes in the subject.

activities: Participation in extra-curricular activities.

nursery: Whether the student attended nursery school.

higher: Desire to pursue higher education.

internet: Access to the Internet at home.

romantic: Involvement in a romantic relationship.

family\_relationship, free\_time, social: Measures of family relationships, free time, and social life.

weekday\_alcohol, weekend\_alcohol: Alcohol consumption on weekdays and weekends.

health: Current health status.

absences: Number of school absences.

grade\_1, grade\_2, final\_grade: Academic grades.  
  
3)  
  
**Statistical Methods for Research**

To investigate the research questions concerning the impact of various factors on high school students' performance in mathematics, the following statistical methods will be employed:

**Linear Regression Analysis:**

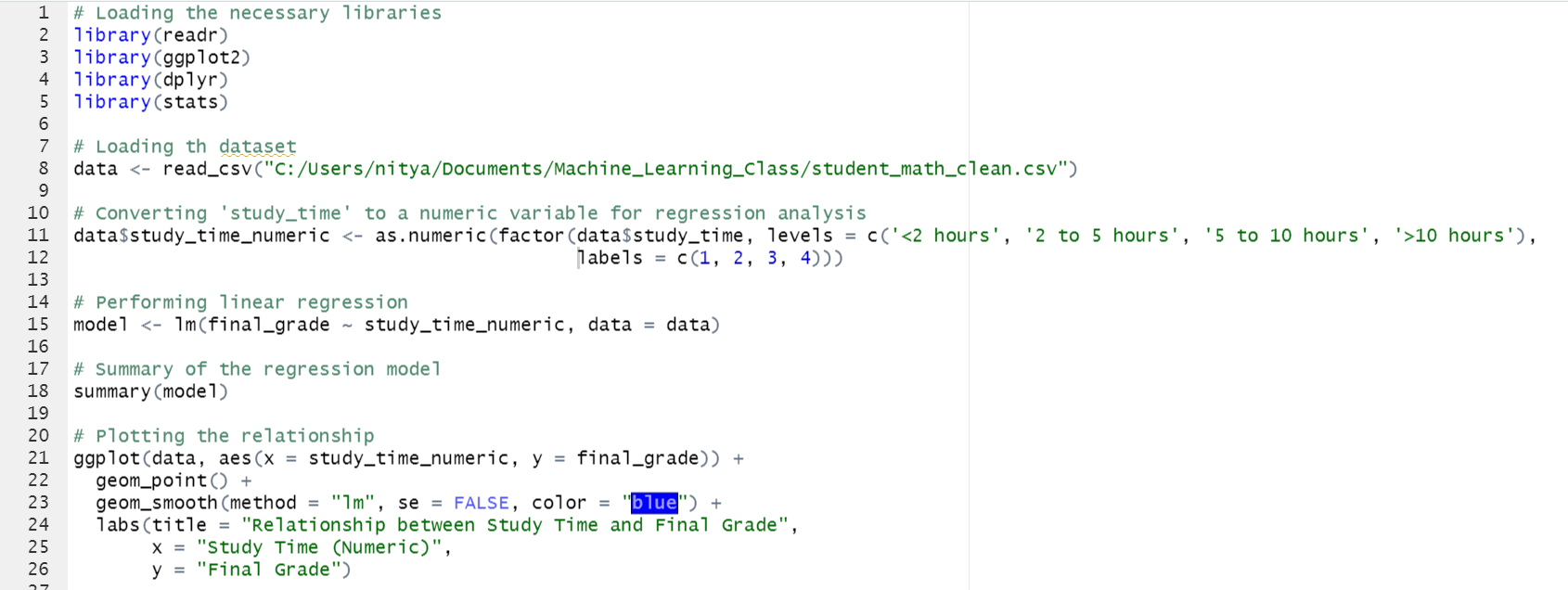
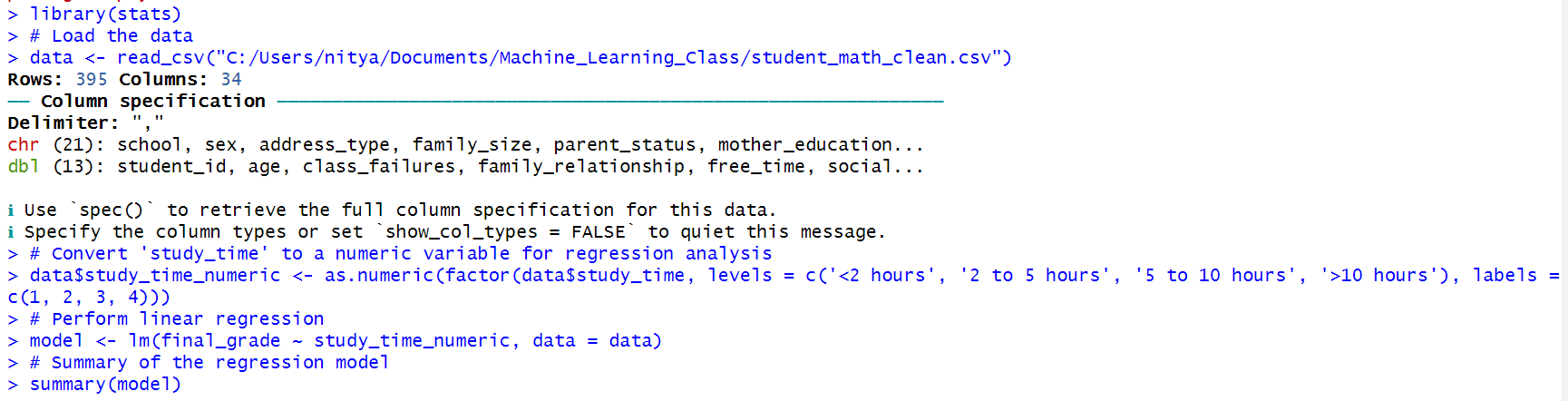
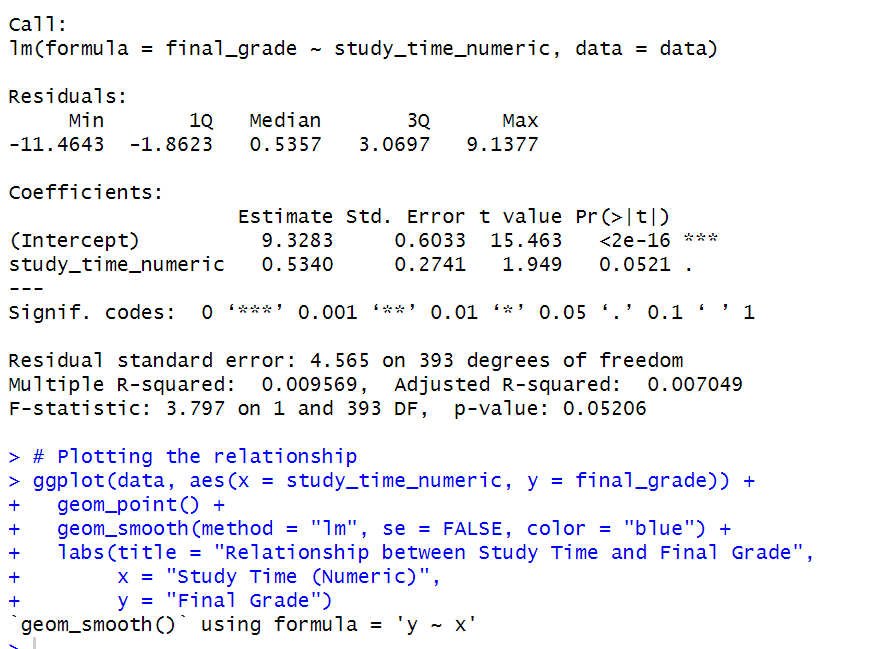
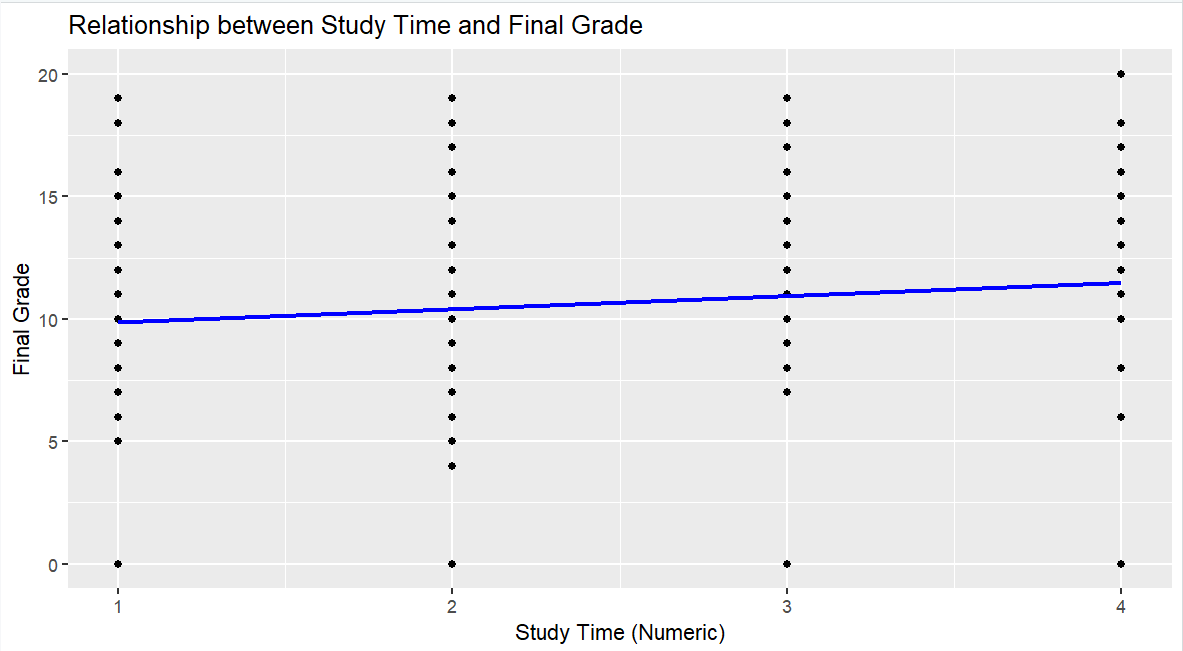
Purpose: To assess the relationship between continuous predictors (like study time) and the continuous outcome variable (final grade in mathematics).

Application: This method will be primarily used to answer the primary research question regarding the impact of study time on final grades. It will help quantify the strength and direction of this relationship.

**Multiple Regression Analysis:**

Purpose: To evaluate the combined effect of multiple predictors (including both continuous and categorical variables like age, gender, school support, and family support) on the final grade.

Application: This approach will address the secondary research questions, allowing for the assessment of how demographic factors and support mechanisms influence academic performance, while controlling for other variables.

4) **Results/ Conclusions/ Limitations**  
  
Analysis (Simple Linear Regression):  
  
  
  
  
  
  
  
**Results**

Regression Coefficients:

Intercept (Constant): 9.3283

This suggests that when the study time is zero (or at the base level), the expected final grade is approximately 9.33.

Study Time Numeric: 0.5340

This coefficient indicates that for each unit increase in the study time category, the final grade is expected to increase by 0.534 points on average.

**Statistical Significance:**

The p-value for the study time numeric variable is 0.0521, which is a little above the conventional significance level of 0.05. This suggests that the relationship between study time and final grade is borderline statistically significant.

**Model Fit:**

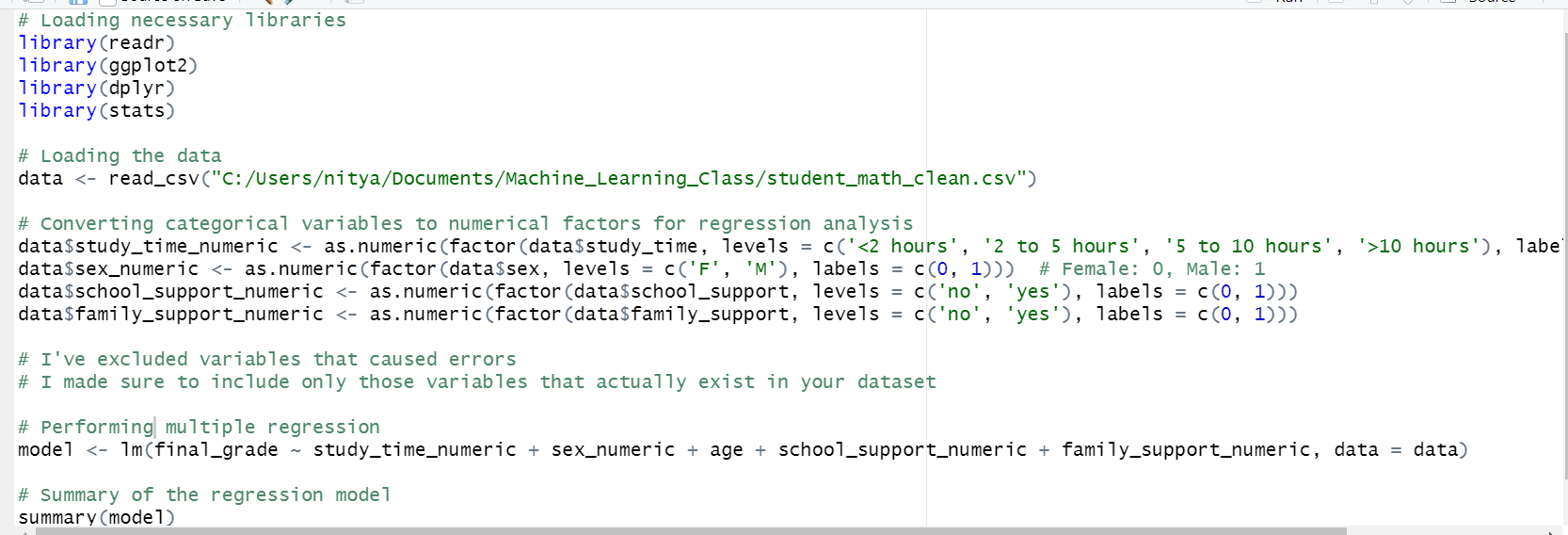
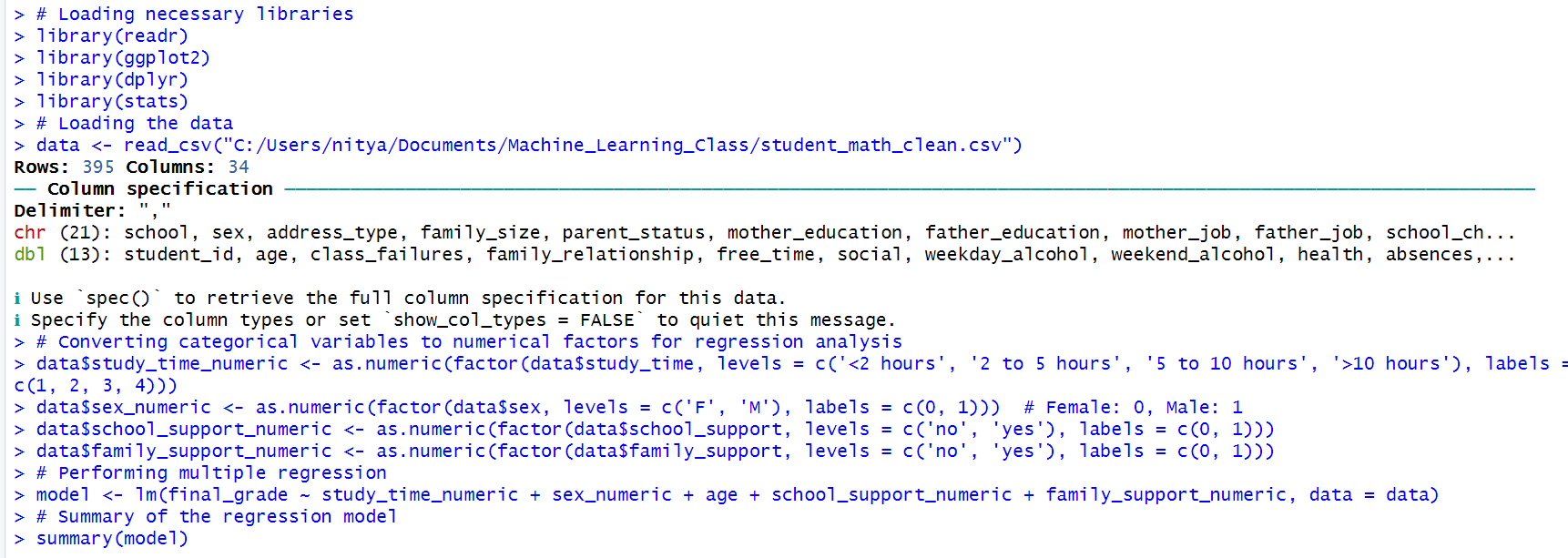
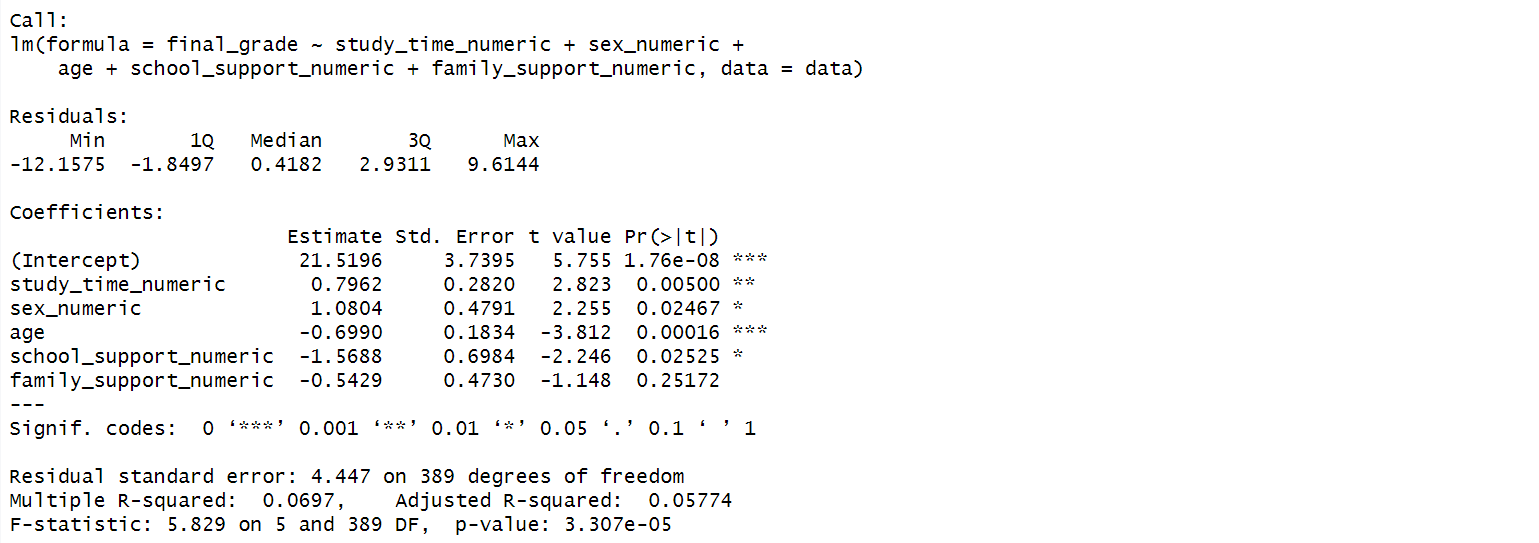
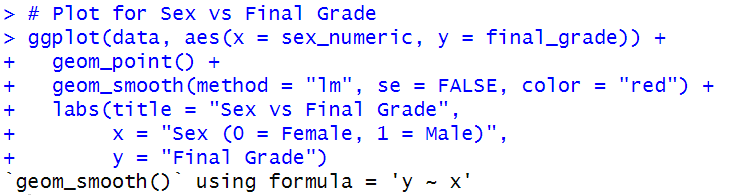
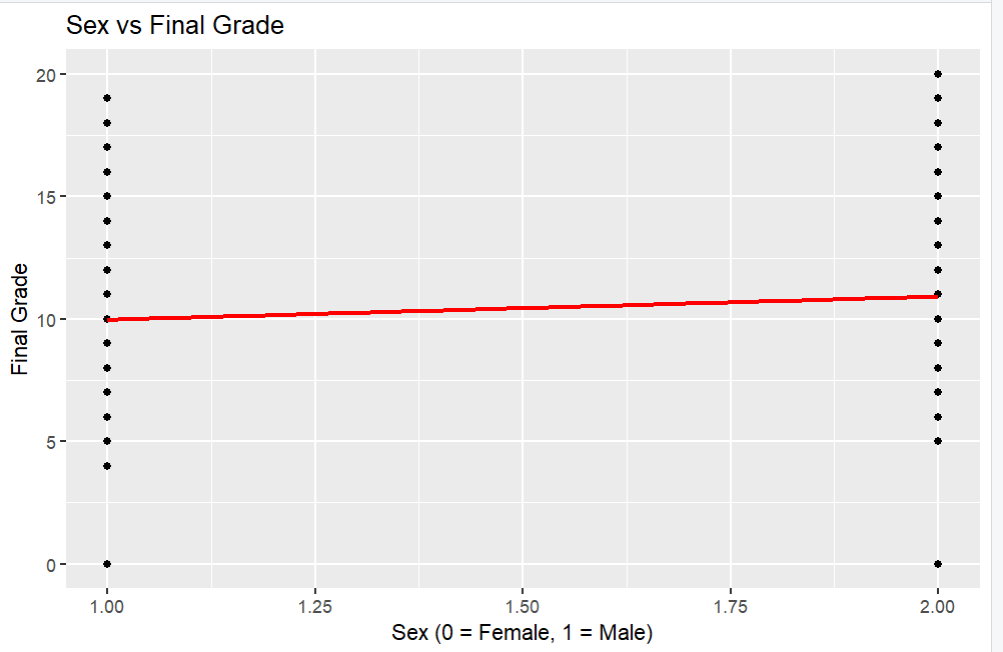
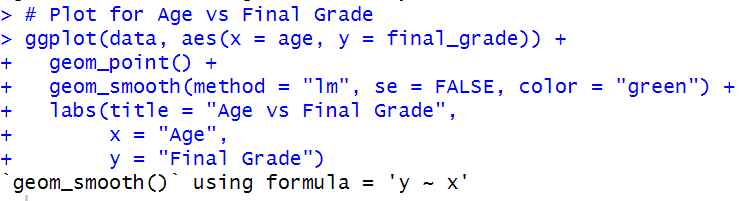
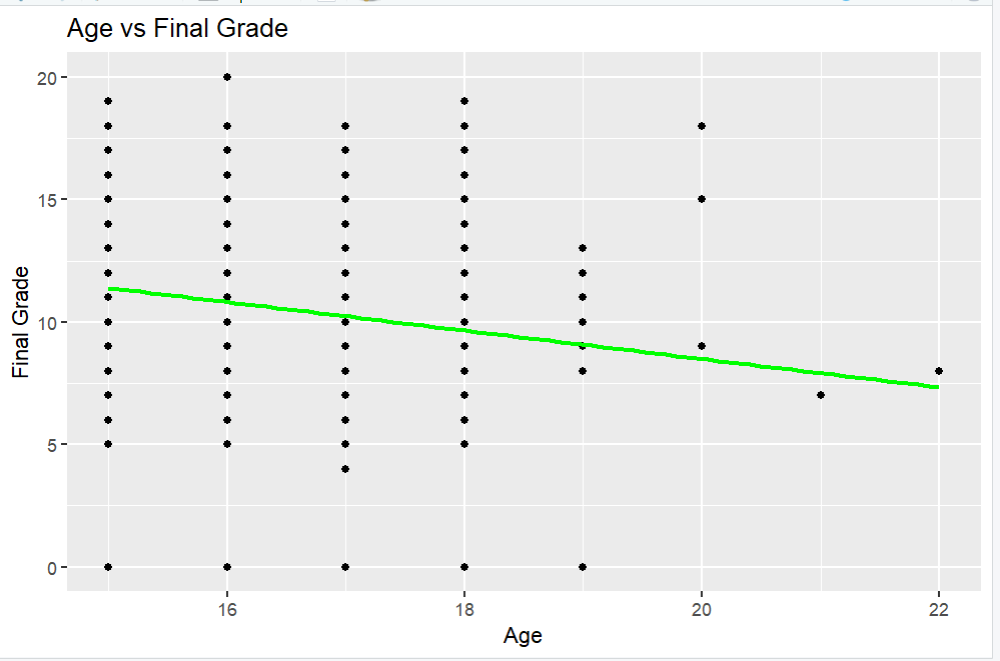
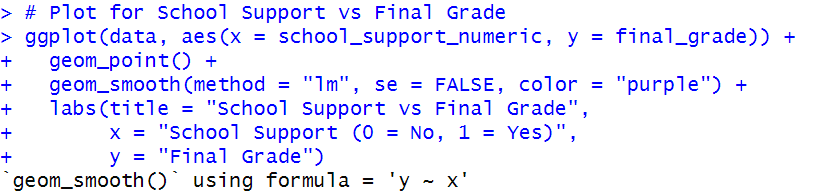
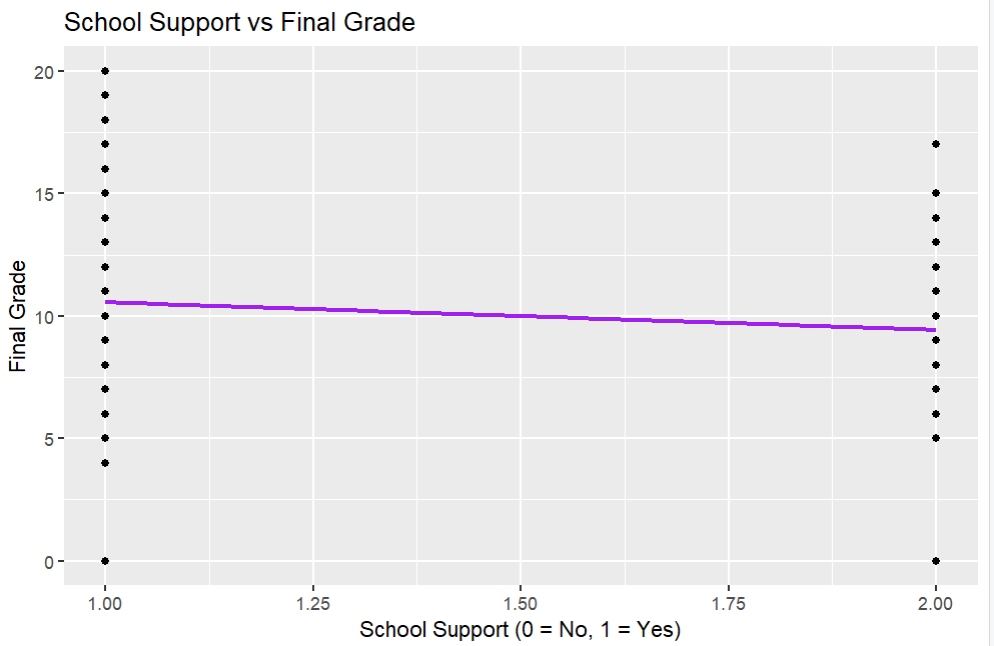
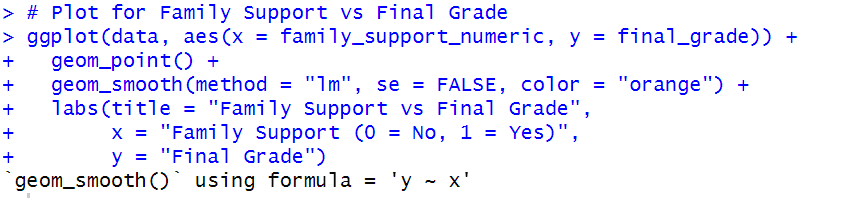
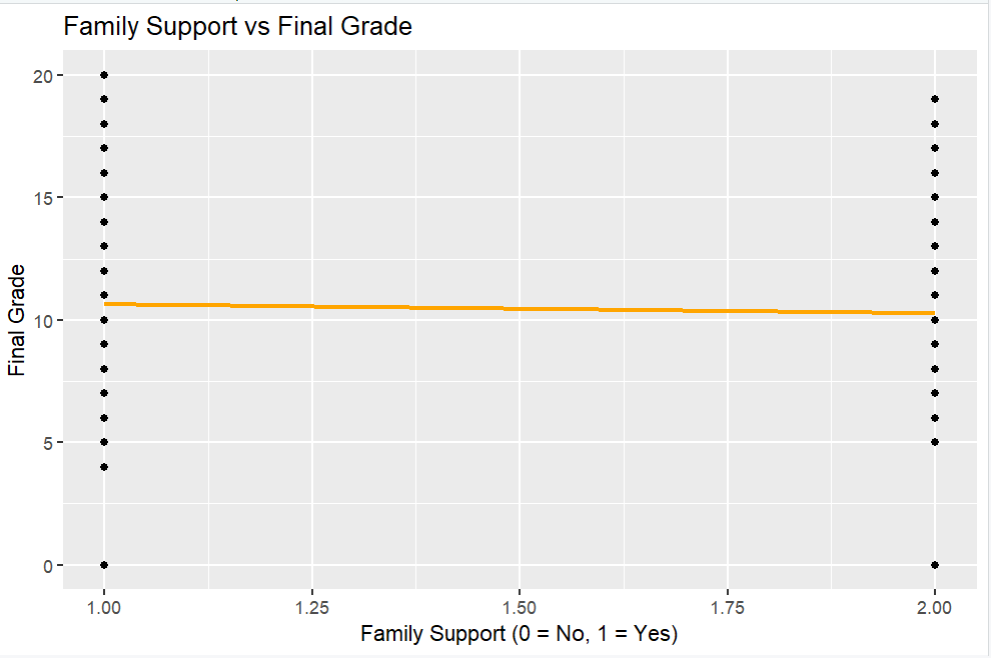
R-squared: 0.009569 (Approximately 0.96%)

This indicates that only about 0.96% of the variability in the final grades can be explained by the study time. This suggests a very weak explanatory power of the model.

**Conclusions**

The results show a slight but not strong link between the time spent studying and math grades. Although it seems that more studying could lead to somewhat better grades, this connection isn't particularly strong or statistically significant. The study time's minimal significance means there's a possible trend, but it's not solid enough to be definitive at the 5% significance level. This may be because of various factors not included in the model.

The small R-squared value also tells us that study time isn't a major predictor of final grades on its own. This underlines the complexities in academic achievement and points to other potential influences like teaching quality, student drive, family engagement, and individual situations playing important parts in shaping student performance. **Limitations**

The study's limitations include its focus on just one predictor, study time, whereas student performance is likely shaped by various factors. Also, the data pertains to a specific student group, meaning the findings might not apply to different student demographics. Furthermore, dividing study time into only four categories could lead to an oversimplified view of students' diverse study habits.  
  
Hence, now we will perform a multiple linear regression analysis:  
  
Multiple Linear Regression  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
**Results:**

Regression Coefficients:

Intercept (Constant): 21.5196

This suggests that the base level final grade (with all predictors at zero) is approximately 21.52.

Study Time Numeric: 0.7962

For each unit increase in study time category, the final grade increases by 0.7962 points on average.

Sex Numeric: 1.0804

Male students (coded as 1) have a 1.0804 point higher final grade than female students (coded as 0).

Age: -0.6990

With each additional year in age there is an average decrease of 0.6990 points in the final grade.

School Support Numeric: -1.5688

Students receiving school support (coded as 1) have a 1.5688 point lower final grade than those not receiving school support (coded as 0).

Family Support Numeric: -0.5429

The coefficient for family support is negative, but it is not statistically significant (p-value > 0.05).

**Statistical Significance:**

Study time, sex, age, and school support show statistically significant effects on final grades at the 5% significance level.

Family support is not statistically significant in this model.

**Model Fit:**

R-squared: 0.0697

Approximately 6.97% of the variability in final grades is explained by the model. This is relatively low, indicating that other factors not included in the model also play a significant role in determining final grades.

**Conclusions**

Study Time Effect: The study found a positive and meaningful link between study time and math grades, indicating that more time spent studying is likely to result in better grades. This result is in line with the general expectation that increased study effort can improve academic performance.

Gender Disparities: According to the model, boys generally achieve higher grades than girls, a finding that needs further exploration to understand what factors contribute to this difference.

Age Factor: The model shows a negative relationship between age and grades, suggesting that older students might have lower scores. This could be due to a range of social and educational factors that affect students as they grow older.

School Support Impact: The analysis shows us that students receiving school support tend to have lower grades, which may imply that those requiring additional help might already be struggling academically.

Family Support Influence: The analysis shows a negative but statistically insignificant relationship between family support and grades, shows that there's no clear link between these two factors in this particular data set.  
  
**Limitations**  
The model only accounts for a limited amount of the variation in final grades, suggesting that other significant factors influencing student performance were not included in this analysis. It's important to note that this study doesn't establish cause-and-effect relationships; it only indicates possible correlations. Additionally, the model is based on the assumption of linear relationships, which may not fully reflect the complex ways in which these factors affect academic performance.

**Suggestions for Further Research**

Including additional variables such as teaching quality, student motivation, extracurricular activities, and socioeconomic status might provide a more comprehensive understanding.

Exploring non-linear models or interaction effects could also give uss more insights into the dynamics affecting student performance.  
  
**References:**

<https://www.kaggle.com/datasets/dillonmyrick/high-school-student-performance-and-demographics>  
  
Creswell, J. W. (2014). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage Publications.

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics. Sage Publications.

Cohen, L., Manion, L., & Morrison, K. (2018). Research Methods in Education. Routledge.