What Attributes Lead to Success in Secondary Education?

Nanor Asadourian

2024-08-07

Data description

This dataset is an extraction of a survey conducted by the Univeristy of Calfornia Irvine and is part of their Machine Learning Repository as of 2014. Information on the survey and sampling methods can be found here.

Analysis

Questions to Answer:

This dataset was surveyed to best predict student performance, and I want to better understand what are the traits/attributes in a student's life that most heavily influence their academic success. The students surveyed attended two different schools in Portugal and the area of academic success this data measures is only in Mathematics. This dataset is representing a small portion of students attending secondary education, and different countries/schools/cultures will have differing variables and results.

Here are some questions that will be guiding this analysis:

- Does the location a student lives or the school they attend affect their learning?
- Does a divorced family that lives apart affect a student's performance in school?
- Does a parent's education level affect their children's want to succeed and try for higher education?
  - If so, how much does this impact?
- Does a parent's occupation affect their children's want to succeed and try for higher education?
  - If so, how much does this impact?
- Does studying for more hours a day pay off? Is there an ideal study time? Is there a perfect balance of studying vs. free-time?
- Does participating in extracurricular activities help students in their academics?
- Does the number of absences affect classroom participation and a student's ability to pass?
  - Is there a threshold as to how many absences will cause negative impact?
- Does drinking affect classroom participation and a student's ability to pass?
  - Is there a threshold as to how many drinks will cause negative impact?
- Is a social student, one who values going out with friends and/or is in a relationship, more likely to academically succeed?
- Do family relationships impact success?
- What is the percentage of additional support (either from family or school) that a student needs to succeed?

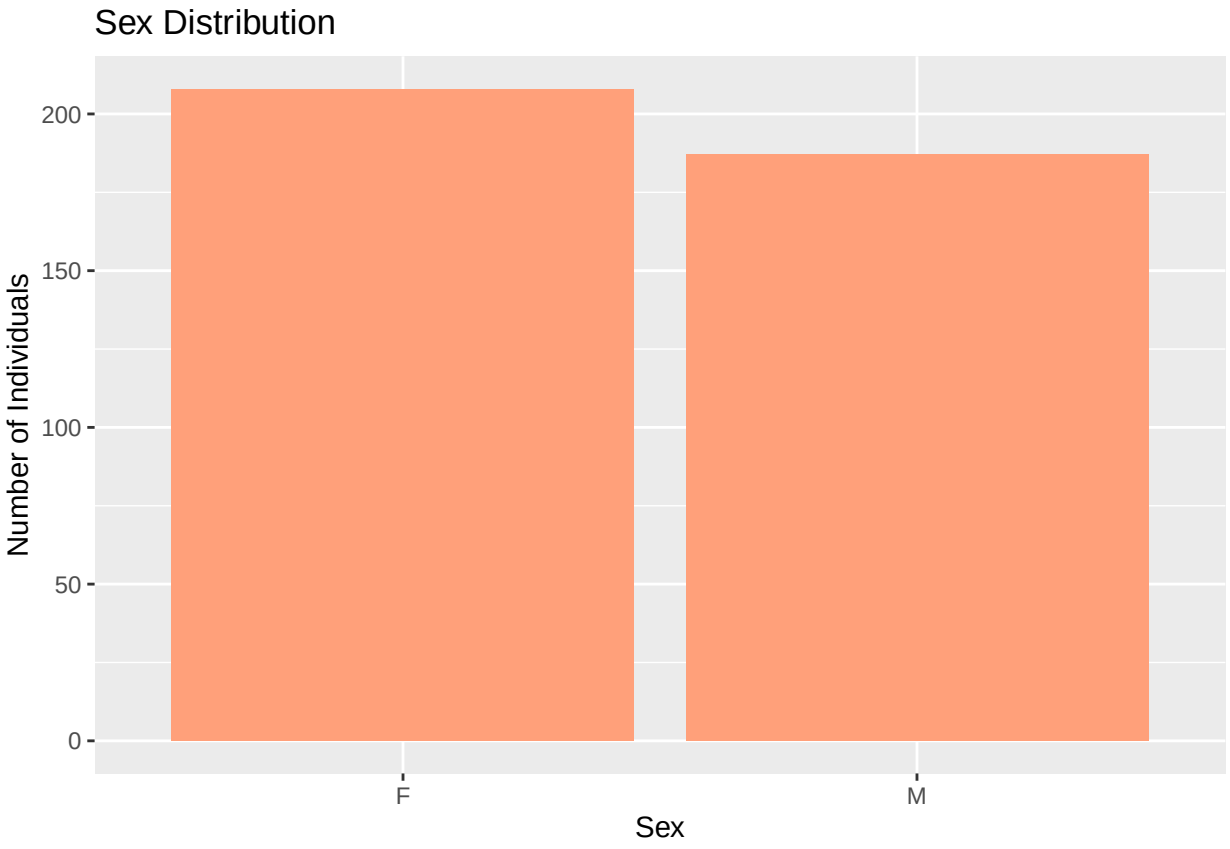
Population and sampling

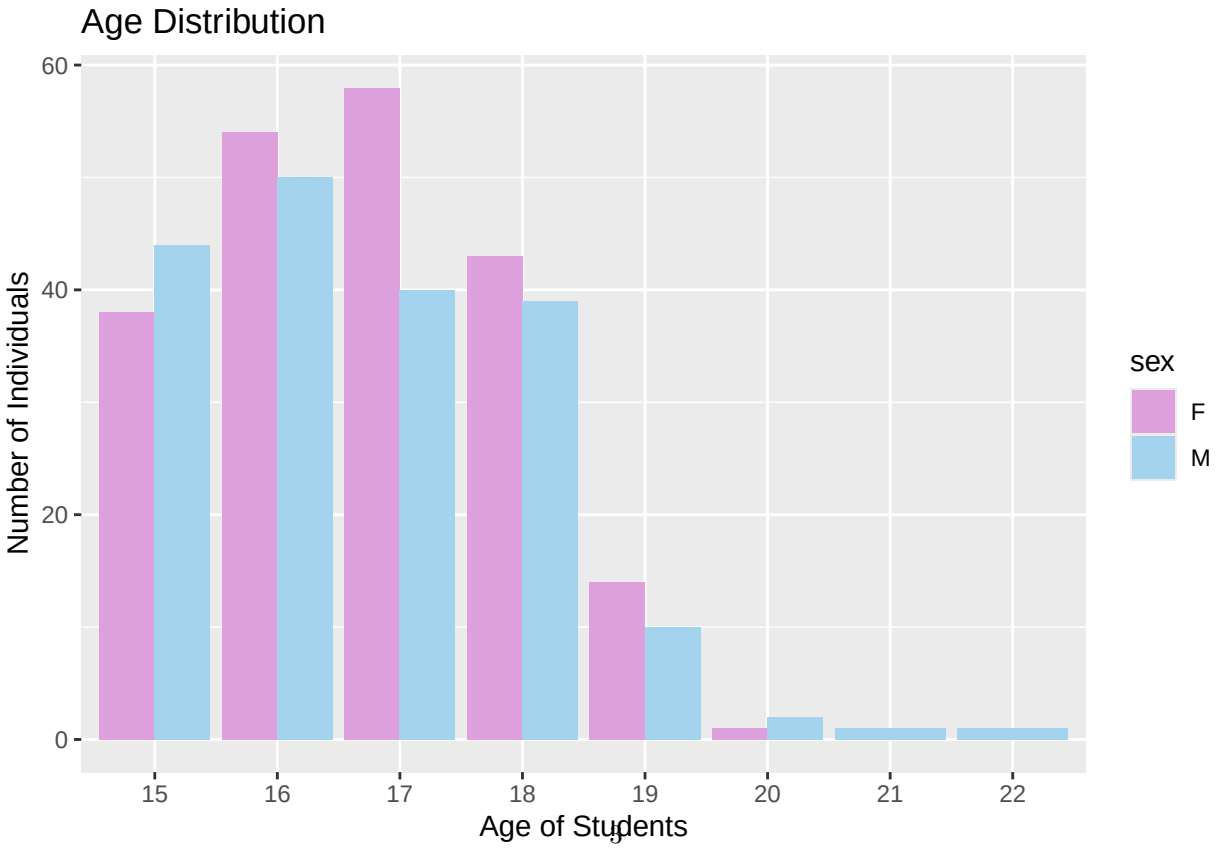
Population: The population for this survey is students in secondary education of two Portuguese schools.

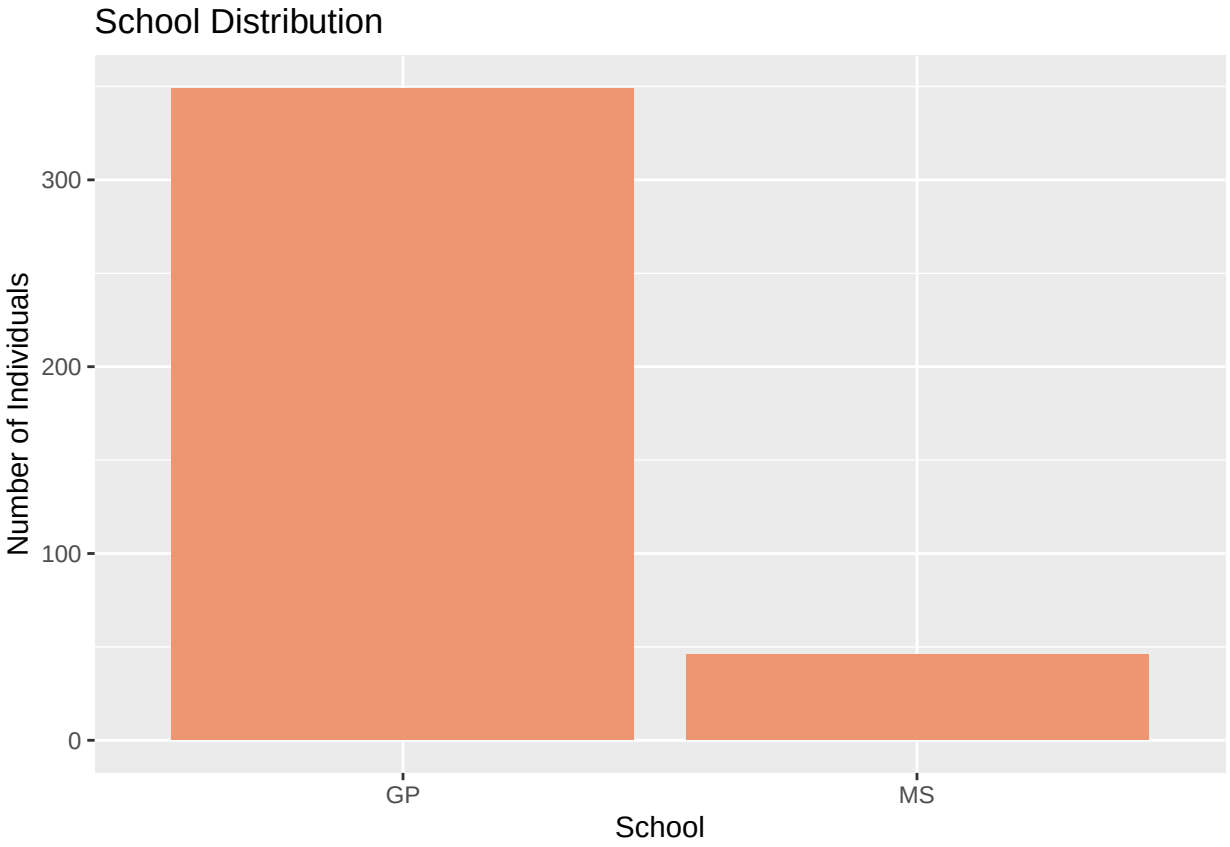
Sample: individuals from the population

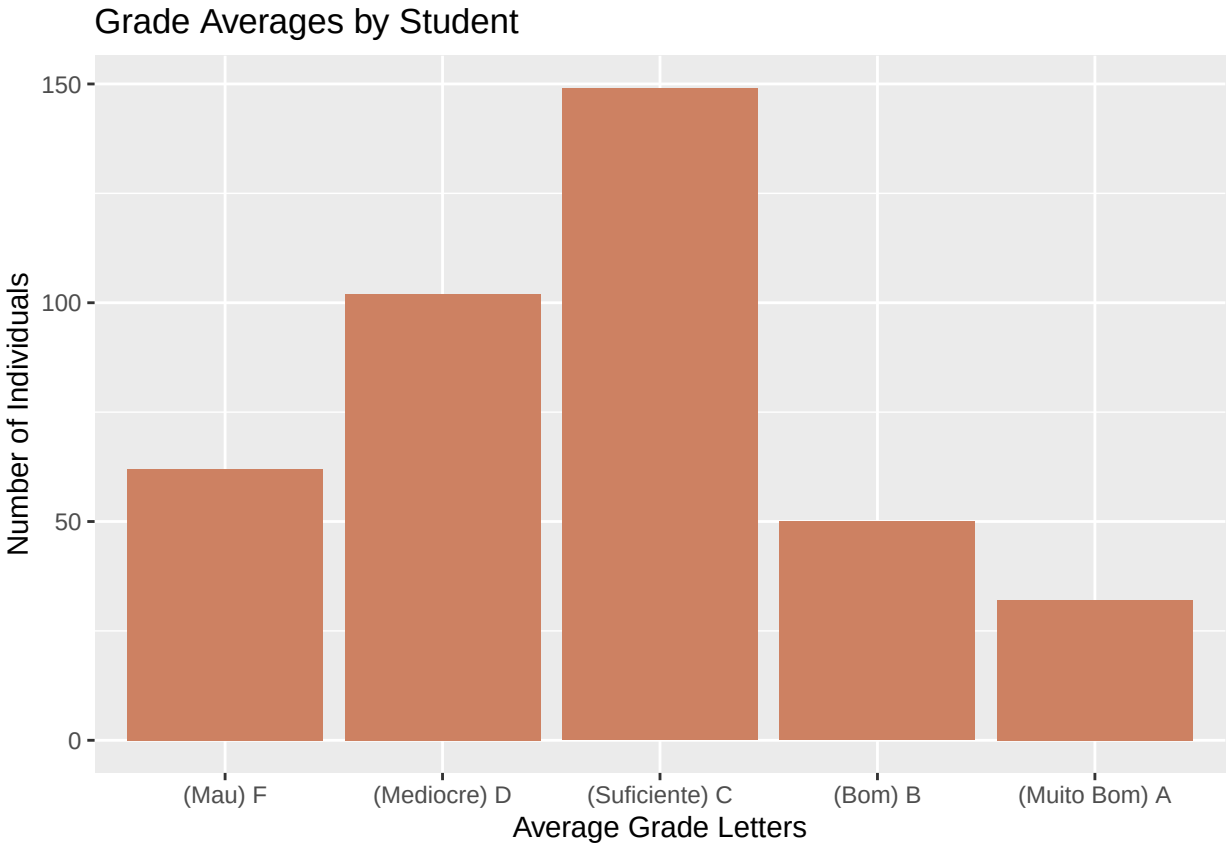
Biases: This survey has a location bias as it is only targeting students in two Portuguese schools in Mathematics for it's subject.

Demographics



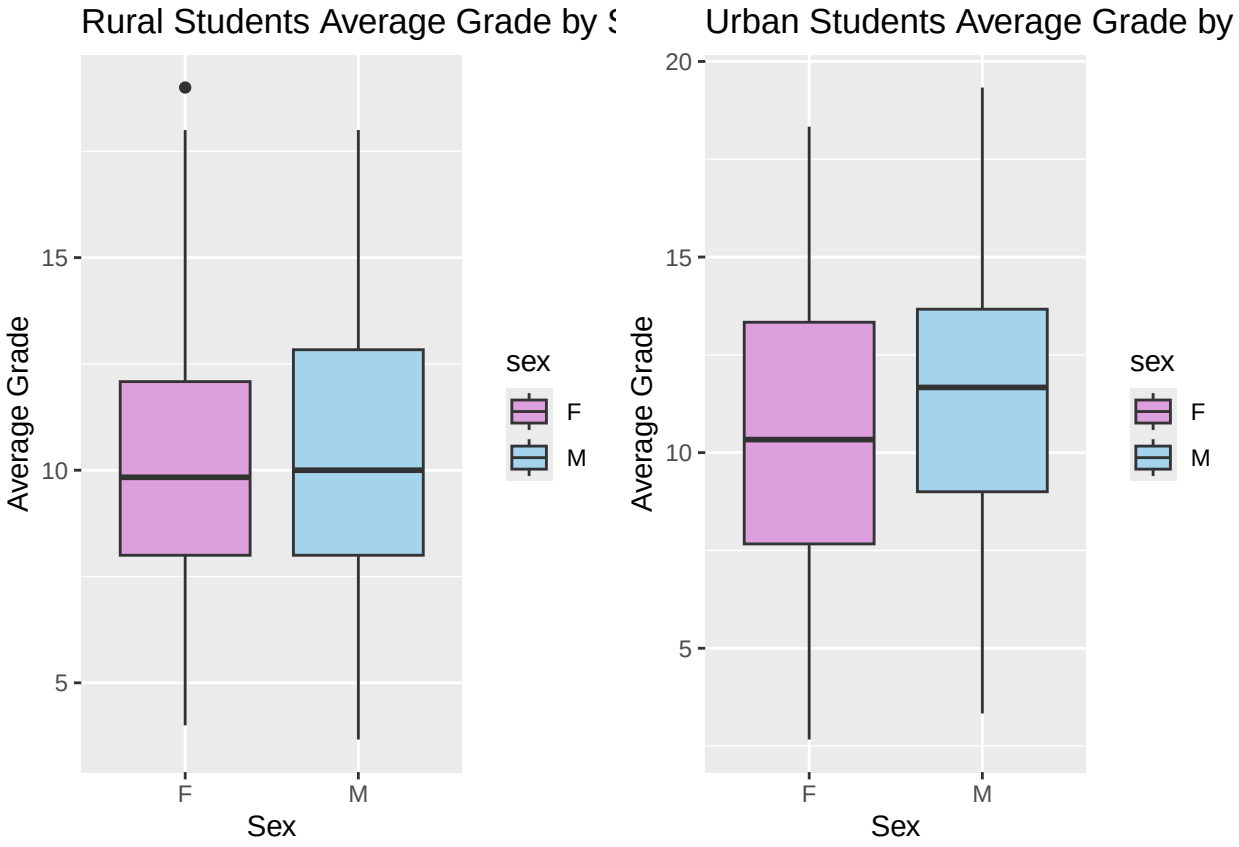



School Distribution
Number of Individuals
300
200
100
0
GP
MS
School

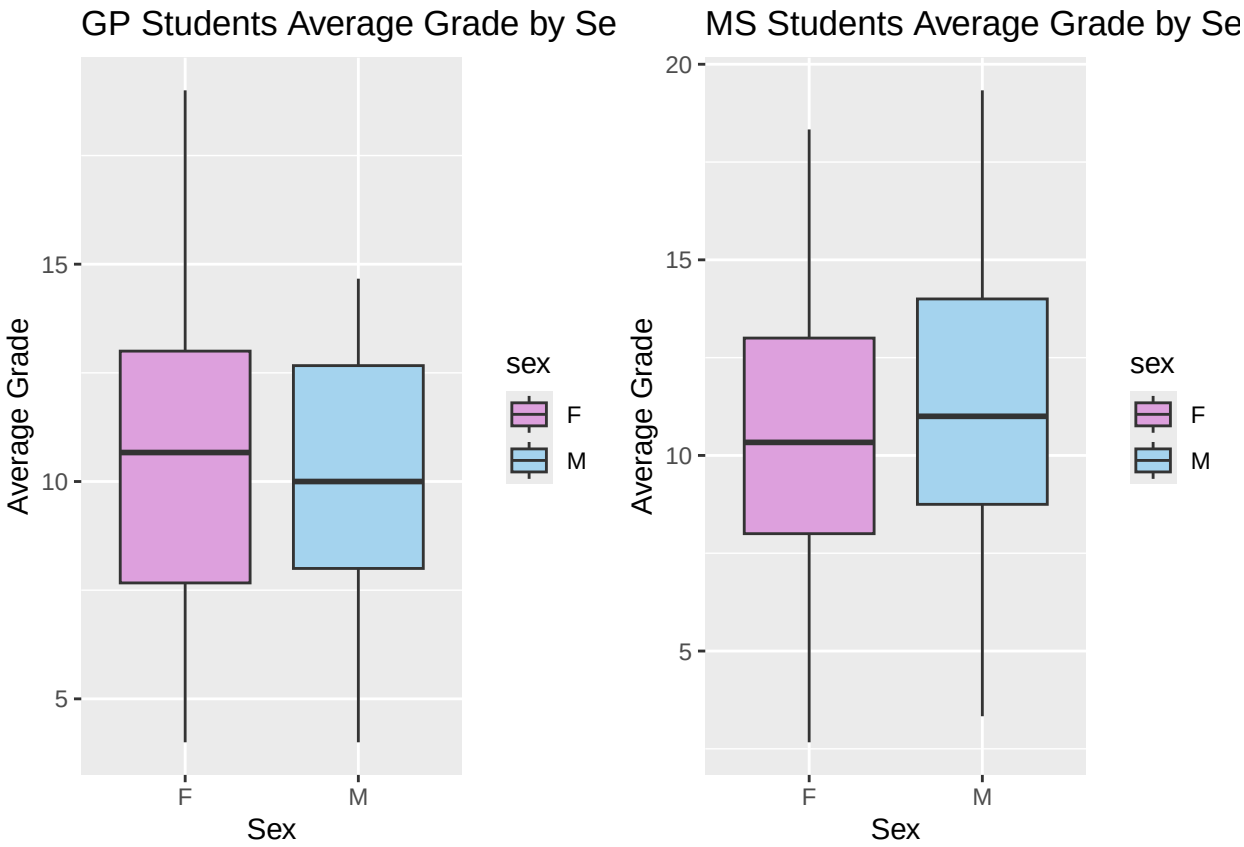


The *Sex* plot shows that the participants were nearly matched in numbers, with slightly more females attending the school surveys. According to the plot, over 200 of the respondents were female whereas only approximately 180 were male. The *Age* plot shows that most of the respondents in the surveys were minors, with only 100 participants out of 395 being the age over 18. The *School* plot shows that most of the respondents in the surveys attended the Gabriel Pereira School, and less than 50 attended Mousinho da Silveira.

Location

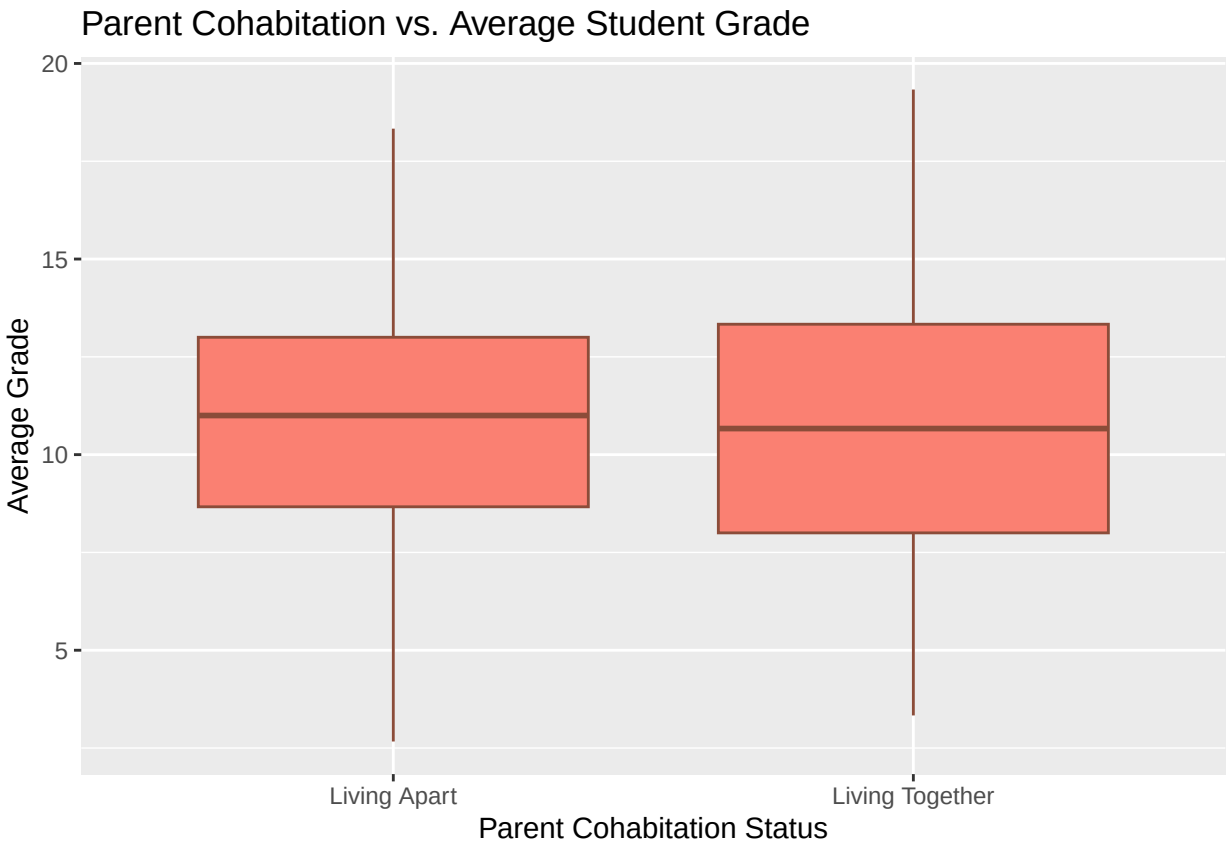


Because *Sex* didn't seem to make a big difference, I wanted to take it a step further if student's address and in turn their school affected this. These plots seem to show this. Students who live in Rural communities seem to have a similar average and distribution. Male students living in Urban communities have a higher average and a larger spread, whereas female students have a slightly lower average and more centered values.



Taking this a step further, while comparing schools, the averages differ more between genders. This could be caused by the small number of students surveyed who attend Mousinho da Silveirs as there are only 50 students.

Family Influence




```
# Living Apart
summary(student_data$Gmean[student_data$Pstatus == "A"])
```

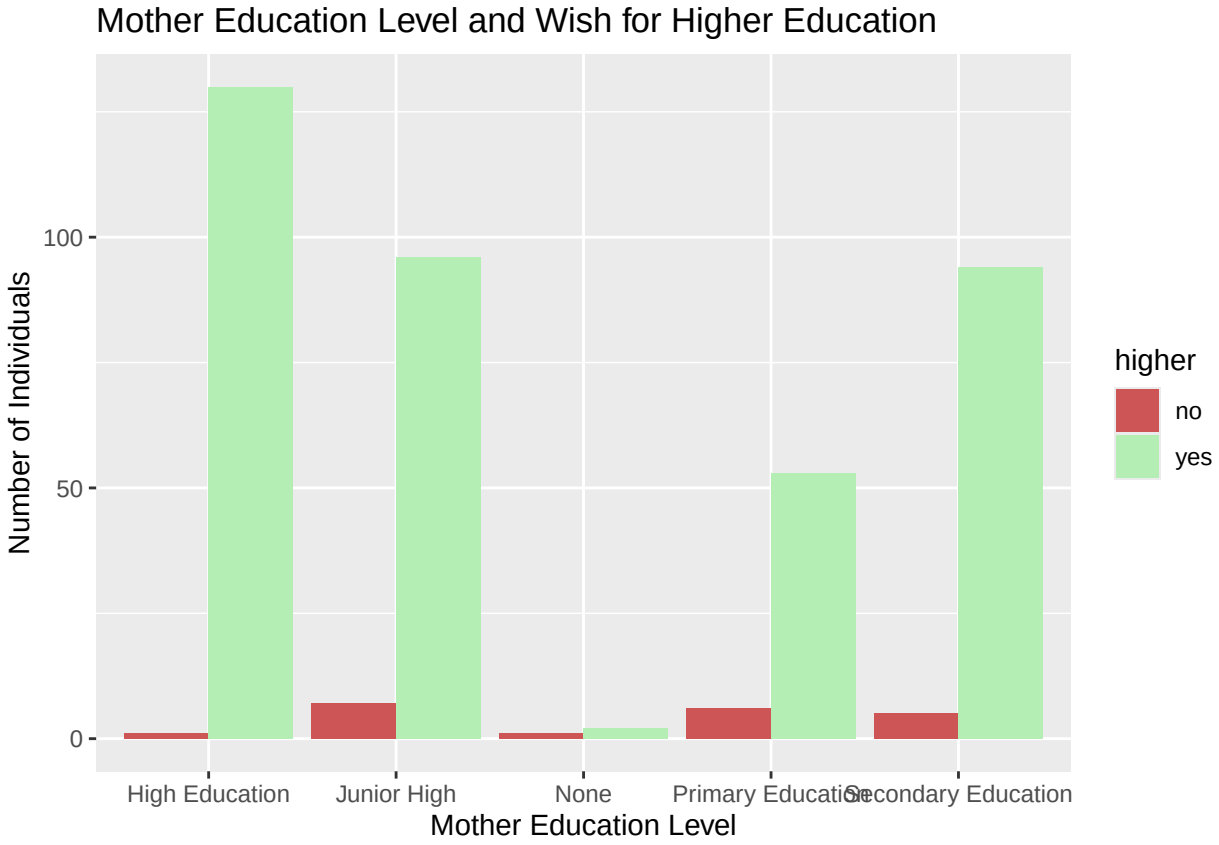
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.667   8.667  11.000  11.114  13.000  18.333
```

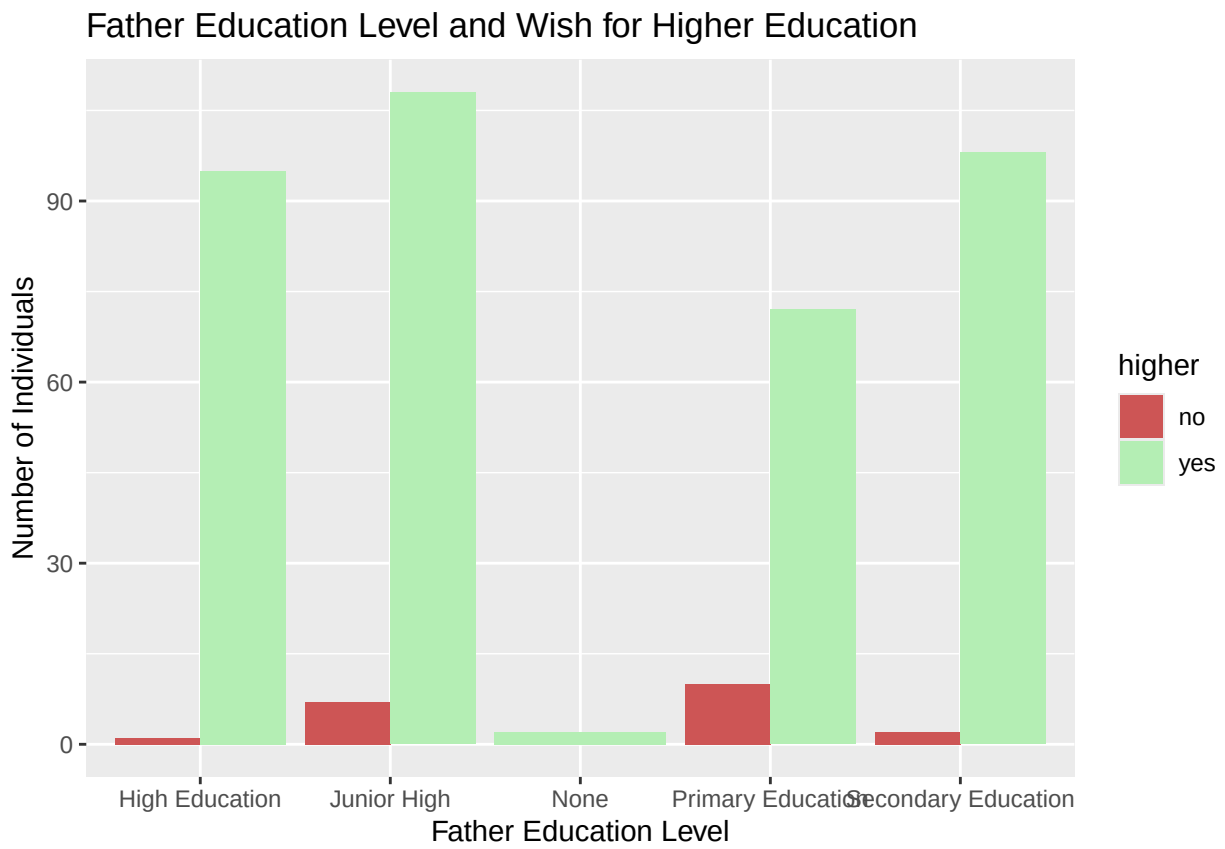
```
# Living Together
summary(student_data$Gmean[student_data$Pstatus == "T"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.333   8.000  10.667  10.702  13.333  19.333
```

Here we see that there is not a big difference in the average grades of students based on their parent's cohabitation status. The difference comes in where the data is centered. If a students' parents are living together, the average minimum and average maximum of the student's grades are 1 point higher.

Family Education





According to these two plots, there is not a significant difference on a student's choice to continue with higher education, but there are differences on their choice to not continue.

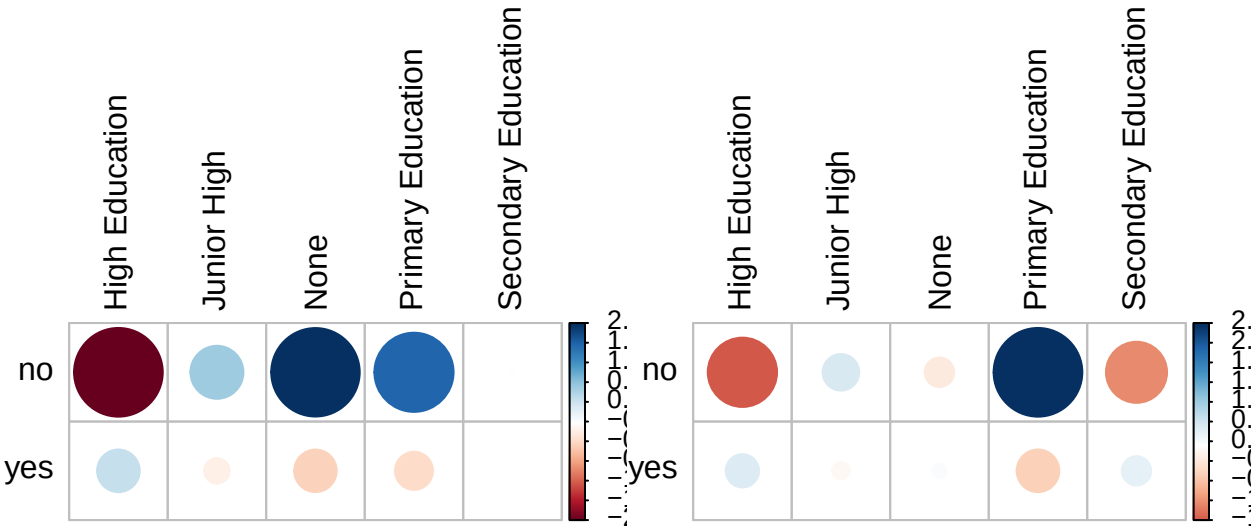
A Chi-squared test was conducted to test if the difference between these two variables are statistically significant.

The null hypothesis is that the influence a parent's education level has on their child's decision is not significant. The alternate hypothesis is that the influence a parent's education level has on their child's decision is statistically significant.

The tests produced a p-value of 0.007721 and 0.006636 for the mother's and father's education level, respectively. Both these values are significant with an alpha of 1%. We reject the null hypothesis and can conclude there are significant influences of a parent's education level.

```
##
##  Pearson's Chi-squared test
##
## data:  tab2.1
## X-squared = 13.87, df = 4, p-value = 0.007721
```

```
##
##  Pearson's Chi-squared test
##
## data:  tab2.2
## X-squared = 14.216, df = 4, p-value = 0.006636
```

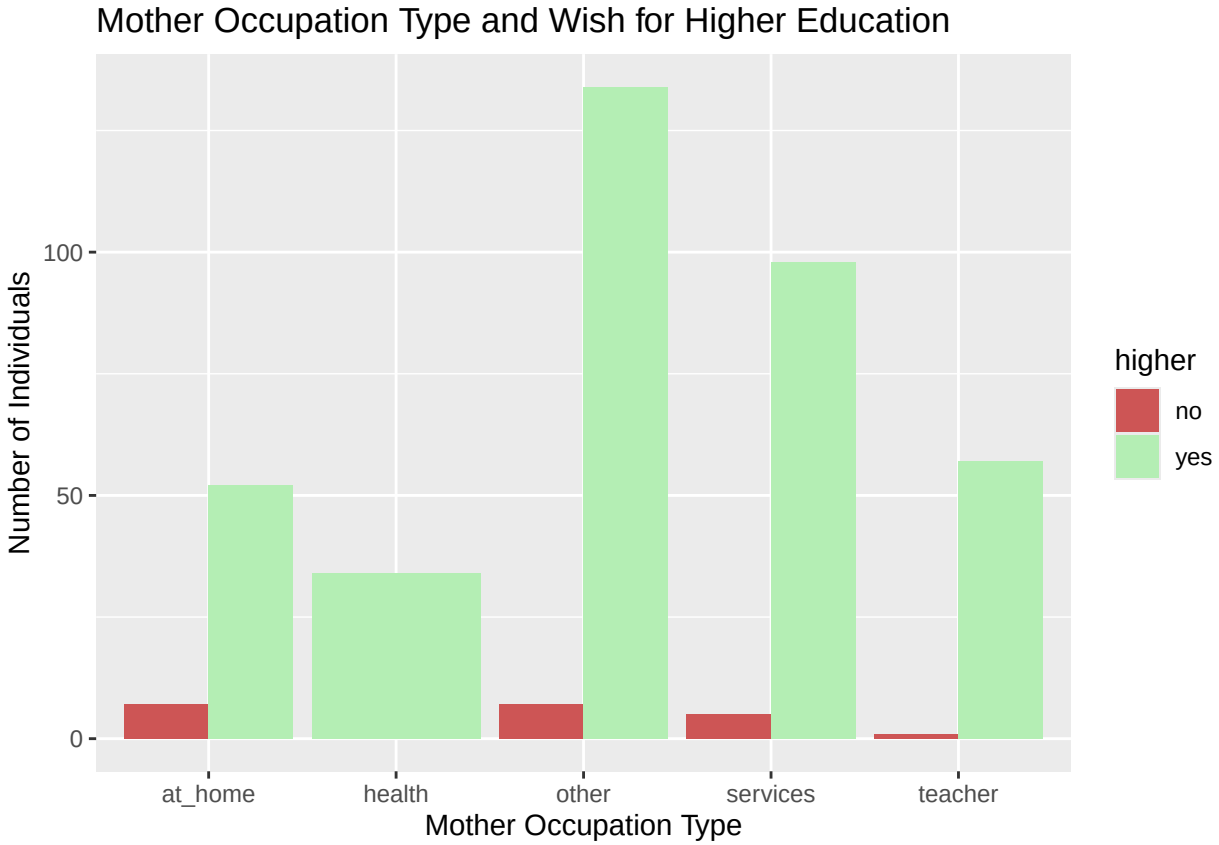


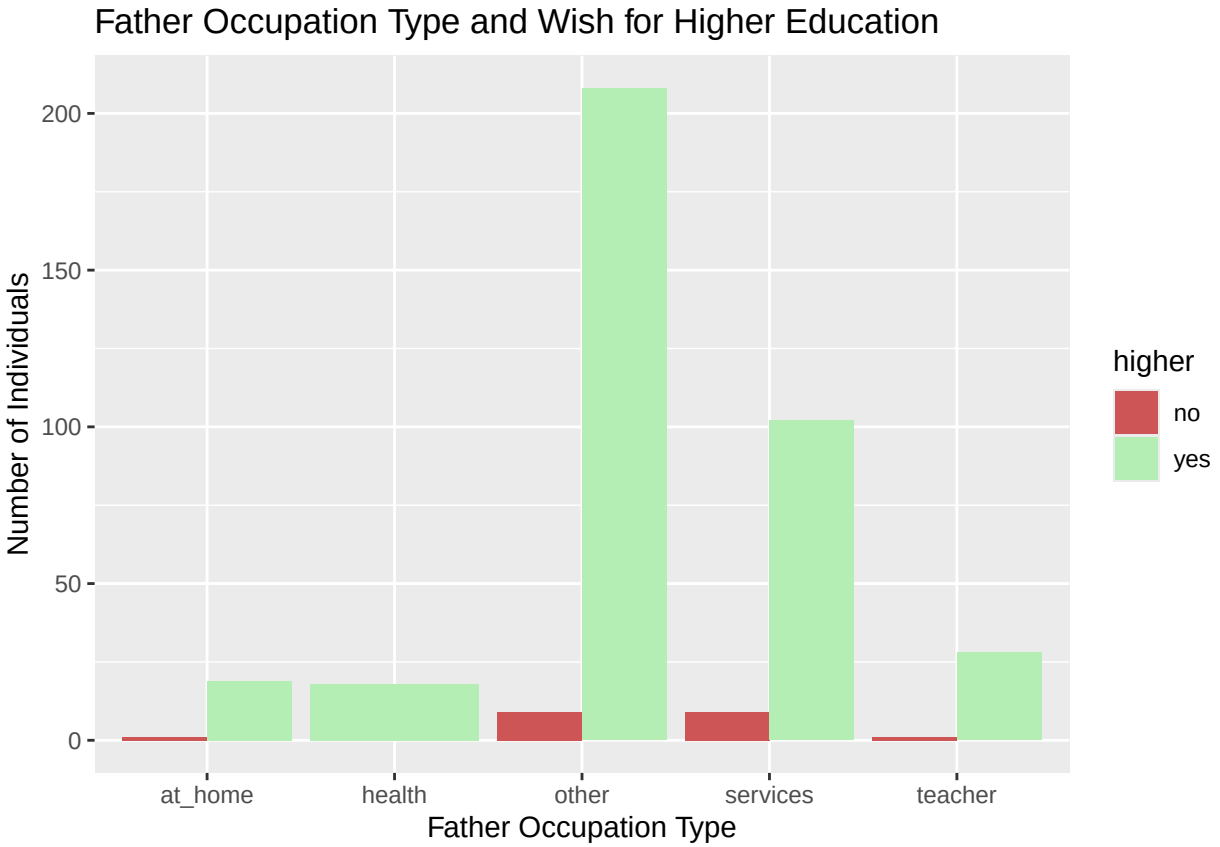
Due to the 4 degrees of freedom, the residuals would provide a better insight into exactly what is causing this. Looking at both graphs, there is a general stronger correlation with the mother's education level than the father's.

As seen above, there is a ***strong negative correlation*** on a student's decision to not attend higher education if their parent attended higher education. This means a student is less likely to say no if either parent went on to higher education.

There is also a ***strong positive correlation*** between a student's decision to not attend higher education and their parent having no education or up to a primary education. This means a student is more likely to say no if their parent had little to no education.

Family Ocupation



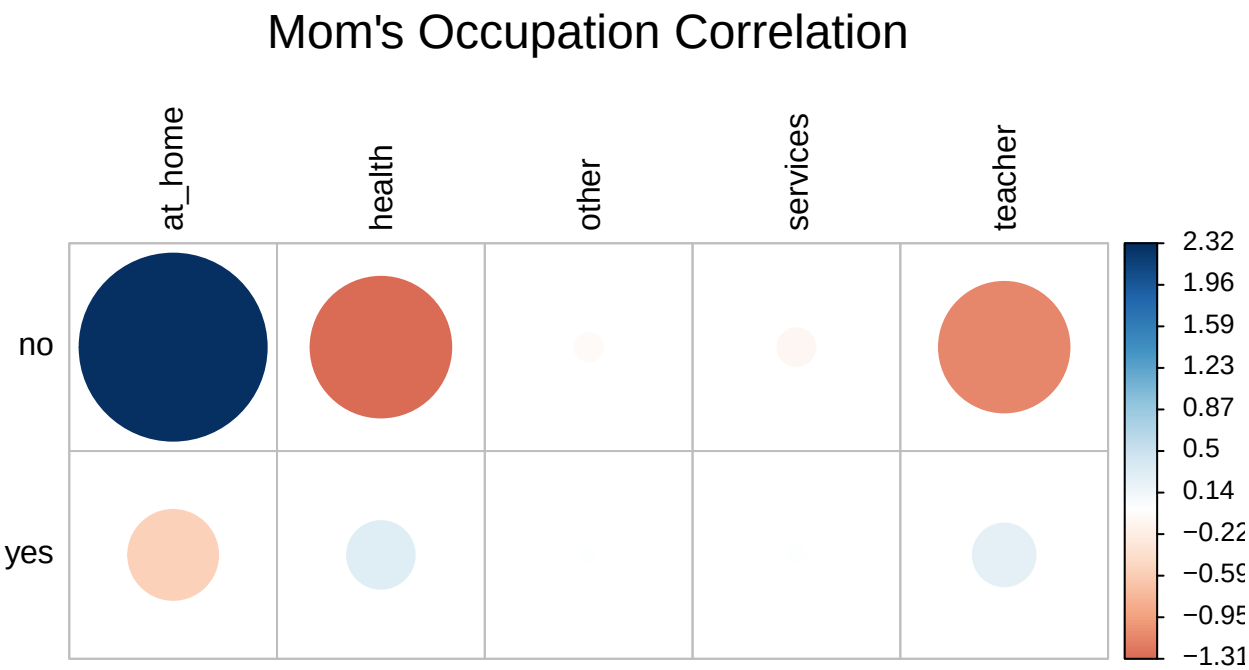


According to these plots, ***student's whose parents are in the health industry all wish to attend higher education***.

A Chi-squared test was conducted to test if the difference between these two variables are statistically significant.

The null hypothesis is that the influence a parent's occupation type has on their child's decision is not significant. The alternate hypothesis is that the influence a parent's occupation type has on their child's decision is statistically significant.

The tests produced a p-value of 0.06501 and 0.4574 for the mother's and father's occupation types, respectively. Only the mother's occupation is considered significant with an alpha of 1%. We reject the null hypothesis and can conclude there are influences of a mother's occupation type.

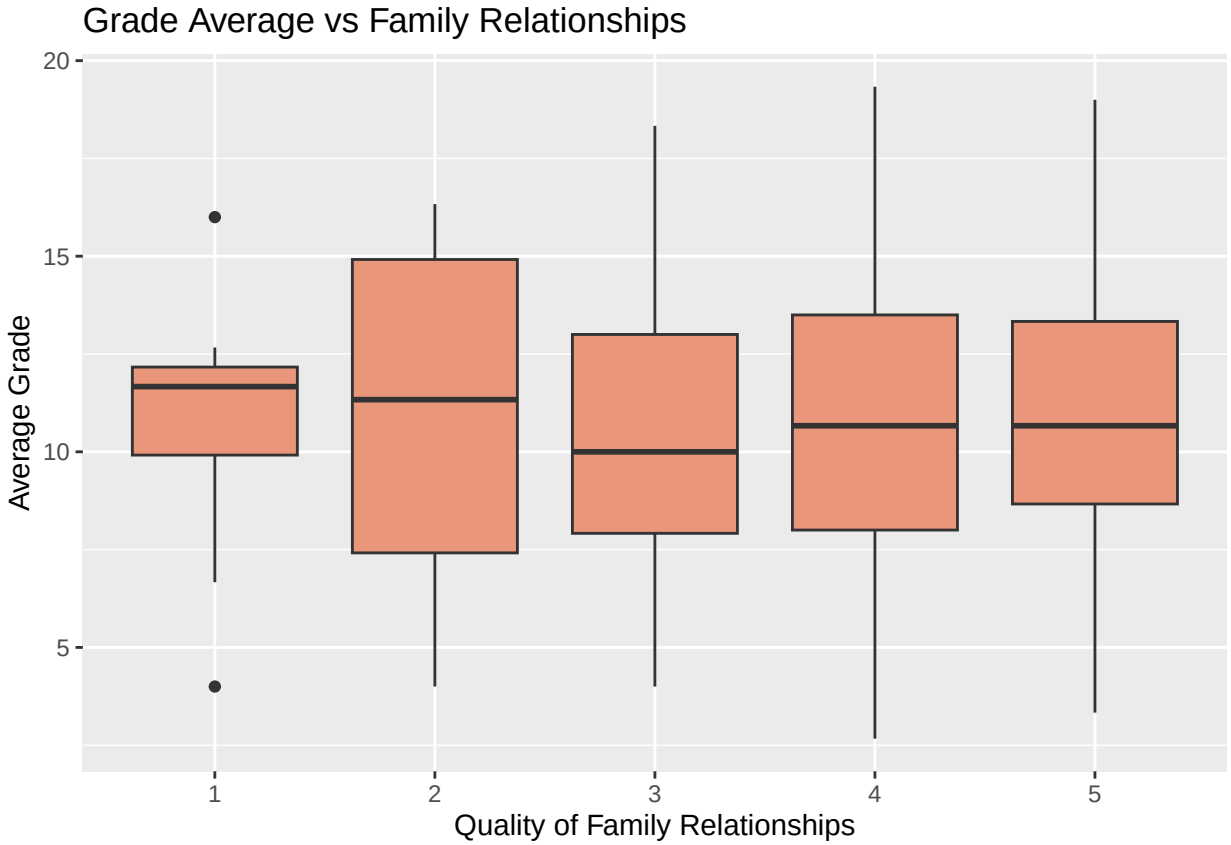


Due to the 4 degrees of freedom, the residuals would provide a better insight into exactly what is causing this.

As seen above, there is a ***strong positive correlation*** on a student's decision to not attend higher education if their mother is a stay-at-home parent. This means a student is more likely to say no if they had a stay-at-home mother.

There is also a ***strong negative correlation*** between a student's decision to not attend higher education and their mother being in the health or education industry. This means a student is less likely to say no if their mother was a teacher or healthcare professional.

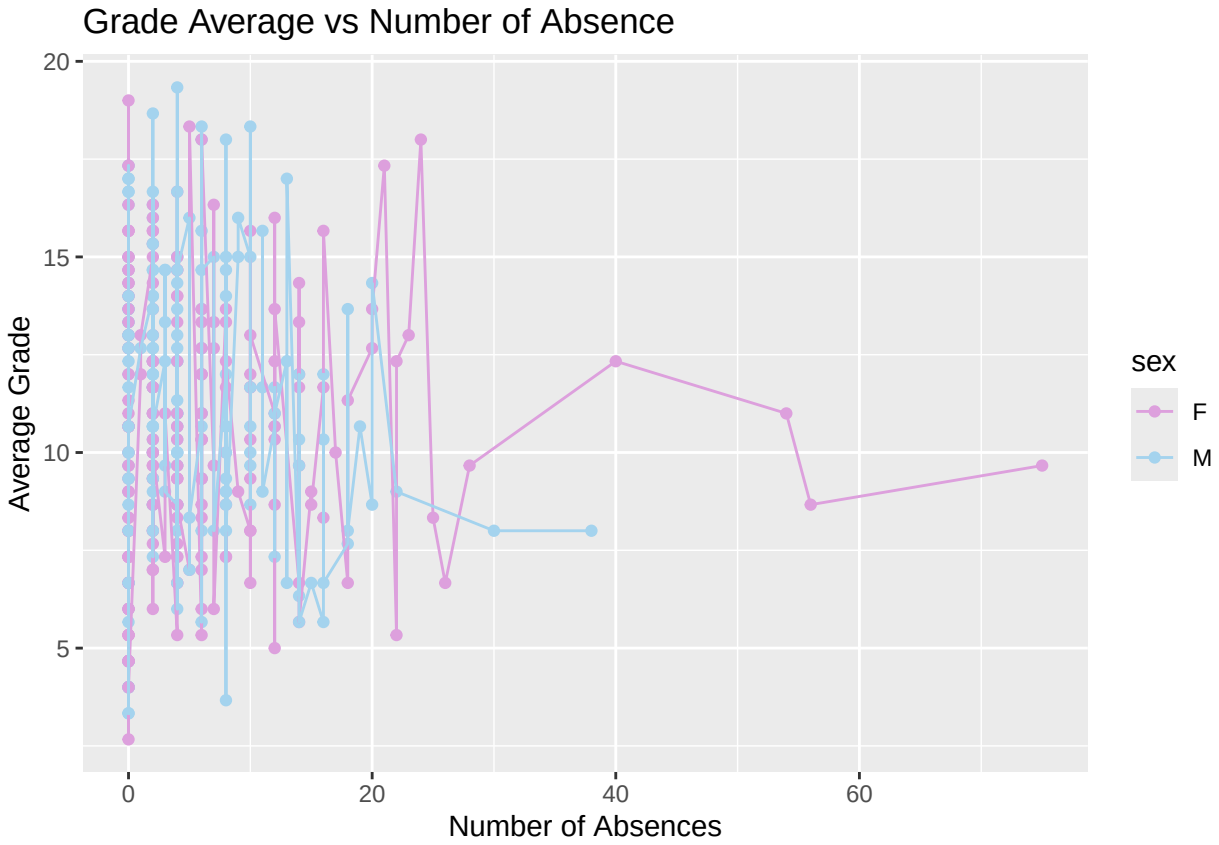
Family Relationships



Even though the averages are all nearly the same, between 10.73 and 10.94, there is a much bigger spread towards when students have a better quality of family relationships. There are only 26 students who have a 1 or 2 quality of family relationship, which is less than 7% of all students surveyed.

Academic Ability

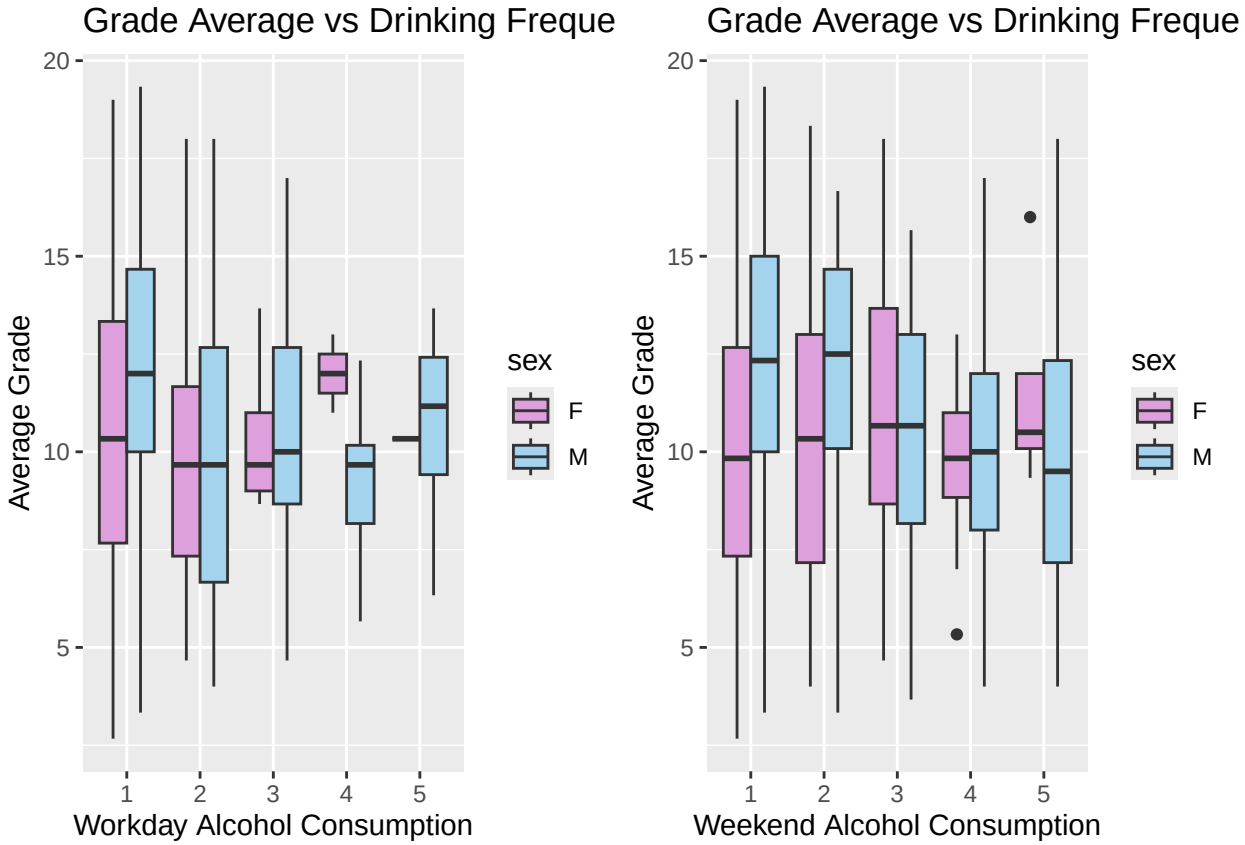
Student Absences



From this plot, we notice that the average student absence is around 5, and that students less than or around the average absences receive the full spectrum of grades. As students gain absences, their grades seem to converge to the average. This means students with more absences can lead to an average grade. There are only a handful of students who have more than 20 absences, but it is a majority fo female students who are in that upper range of absences.

About 10 absences appears to be the threshold of when students begin to score closer to the average grade

Student Drinking



Looking at these plots, we have some takeaways. There looks to be a normal distribution when comparing the frequency of alcohol consumption during a *workday* and the student's average grade. When a student drinks more than moderately (3 and above), their grades seem to converge to the average. This means students with a higher alcohol consumption during the workday will likely result in an average grade.

This does not seem to be the case when comparing grade averages with the frequency of alcohol consumption during a *weekend*. Students consuming alcohol high amounts of alcohol on weekends are still able to score high grade averages. This means weekend alcohol consumption doesn't affect student's average grades.

```
# Percentage of Male Students who Drink Moderately or More on Workdays
nrow(MDrink) / length(student_data$sex[student_data$sex == "M"])
```

```
## [1] 0.1925134
```

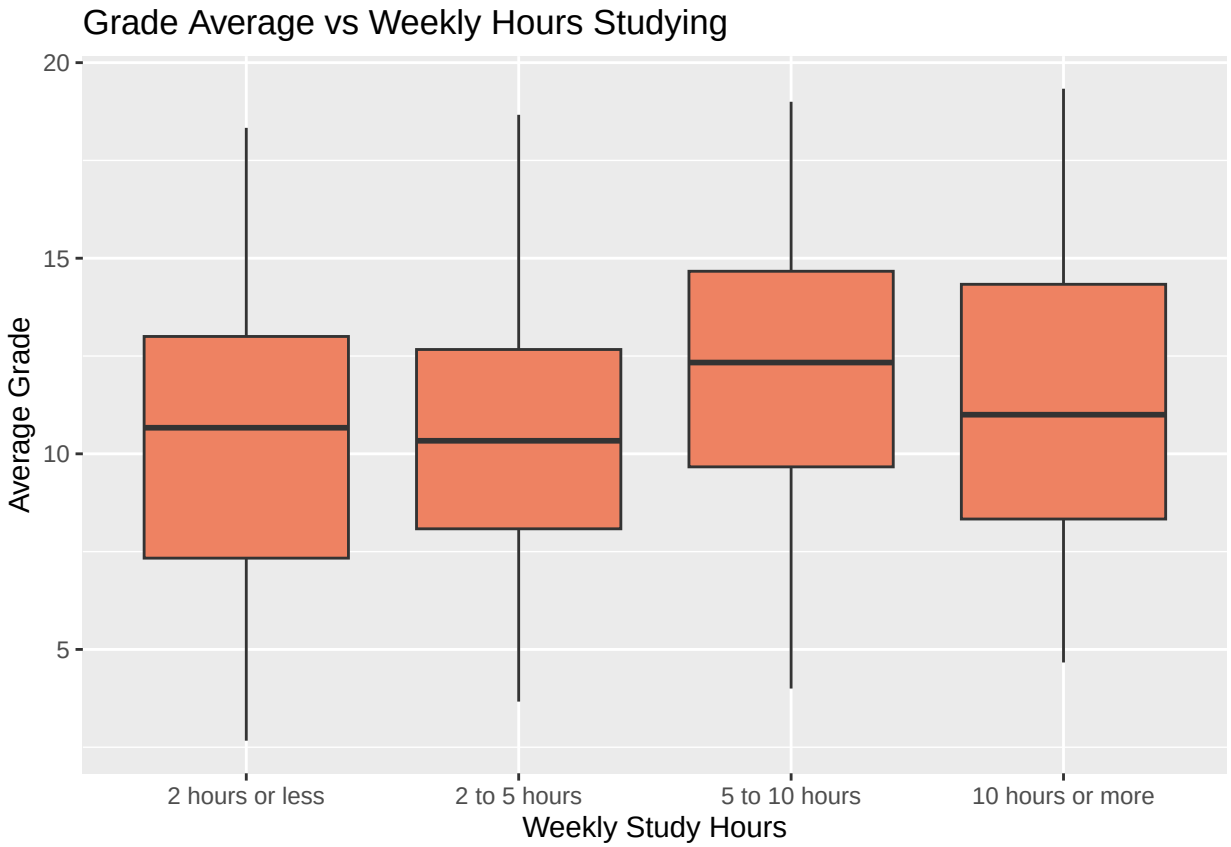
```
# Percentage of Female Students who Drink Moderately or More on Workdays
nrow(FDrink) / length(student_data$sex[student_data$sex == "F"])
```

```
## [1] 0.03846154
```

There is a larger percentage of male students drinking on the higher end than female students even though there are more female students reported in this survey. Roughly 19.25% of male students drink more than moderately on workdays, where only 3.8% of female students drink more than moderately on workdays.

A score of a three, or moderate drinking, appears to be the threshold of when students begin to score closer to the average grade.

Study Habits



Looking at our plot, we notice a slight positive trend between time spent studying and the average grade. Spending at least 5 hours a week studying will lead to a much higher grade average than not but upon studying greater than 10 hours a week, the average falls a little bit. Less than a 25% of the students surveyed study for 5+ hours a week, but there seems to be an advantage to doing so.

```
# Percentage of Students who study more than 5 Hours Weekly
length(student_data$studytime[student_data$studytime >=3])/nrow(student_data)
```

```
## [1] 0.2329114
```

```
# Students who study for 5 to 10 hours weekly
summary(student_data$Gmean[student_data$studytime == 3])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.000   9.667  12.333  11.831  14.667  19.000
```

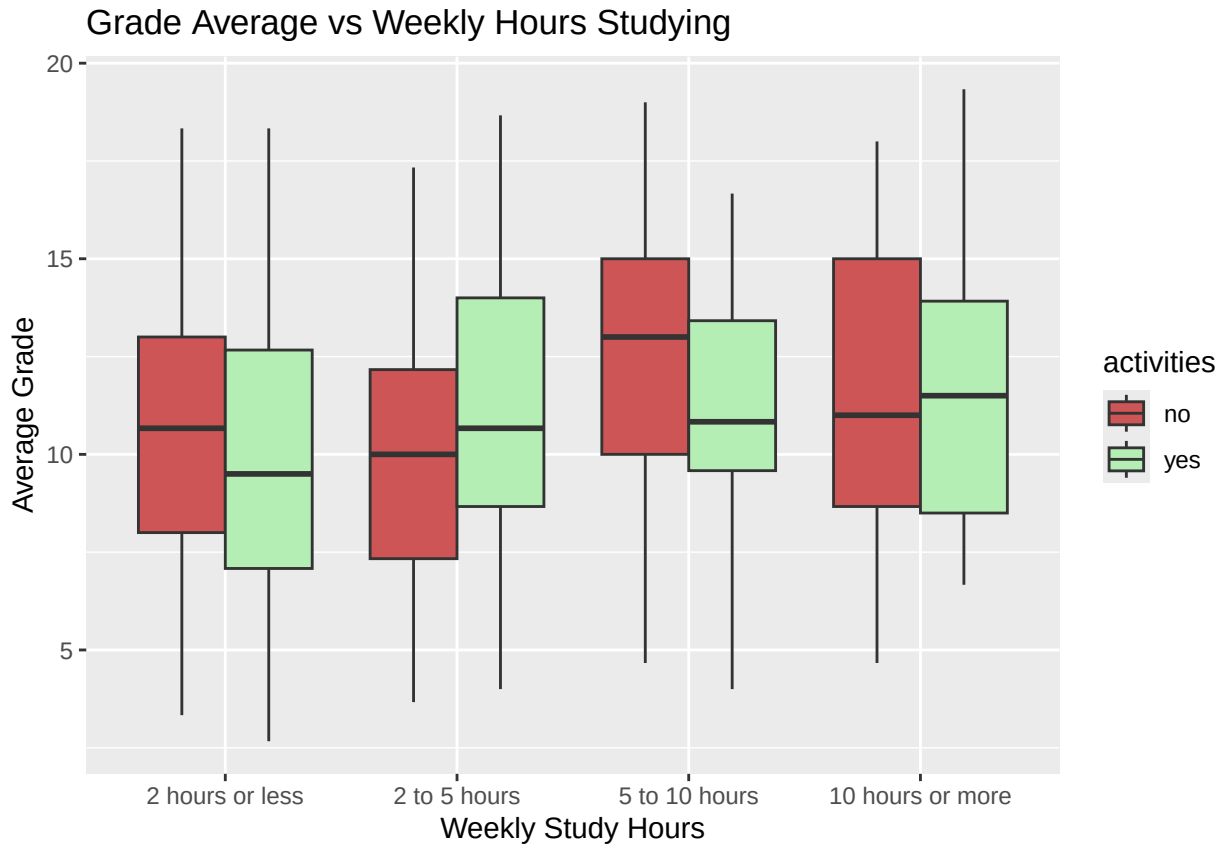
```
# Students who study for 10+ hours weekly
summary(student_data$Gmean[student_data$studytime == 4])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.667   8.333  11.000  11.679  14.333  19.333
```

When looking at the summary between studying for 5-10 hours a week (Option 3) and studying for 10+ hours a week (Option 4), we can identify some key differences. The 1st Quartiles and Median of Option 3

are at least 1 point higher, while the Mean and 3rd quartiles has a 1/3 point lead. Option 4 only yields a higher maximum and minimum score.

Studying for 5-10 hours weekly appears to be the best amoutnof time to yeild the highest average grade.



This plot breaks down a student's studying habits further and includes their participation in extracurricular activities and how both can affect their grades. It appears that when student's study the least, defined as less than two hours, and when they study the most effectively, defined as when their study hours yield the best grade, extracurricular activities will lead to a worse grade average. One reason for this could be that the EC activity is the cause for a low study time of less than 2 hours or a distraction for those in the higher weekly study times.

```
# Students who study for 5 to 10 hours weekly with an EC Activity
summary(student_data$Gmean[student_data$studytime == 3 & student_data$activities == "yes"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.000   9.583  10.833  11.157  13.417  16.667
```

```
# Students who study for 5 to 10 hours weekly without an EC Activity
summary(student_data$Gmean[student_data$studytime == 3 & student_data$activities == "no"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.667  10.000  13.000  12.667  15.000  19.000
```

For those students who study moderately, between 2 to 5 hours weekly, and excessively, more then 10 hours weekly, extracurricular activities will help the student's average grades. One reason for this could be that the EC activity is driving the student to perform better.

Extra-Curricular activities could be a driving force to study or a distraction from studying based on the students.

Social Relationships

Social Life

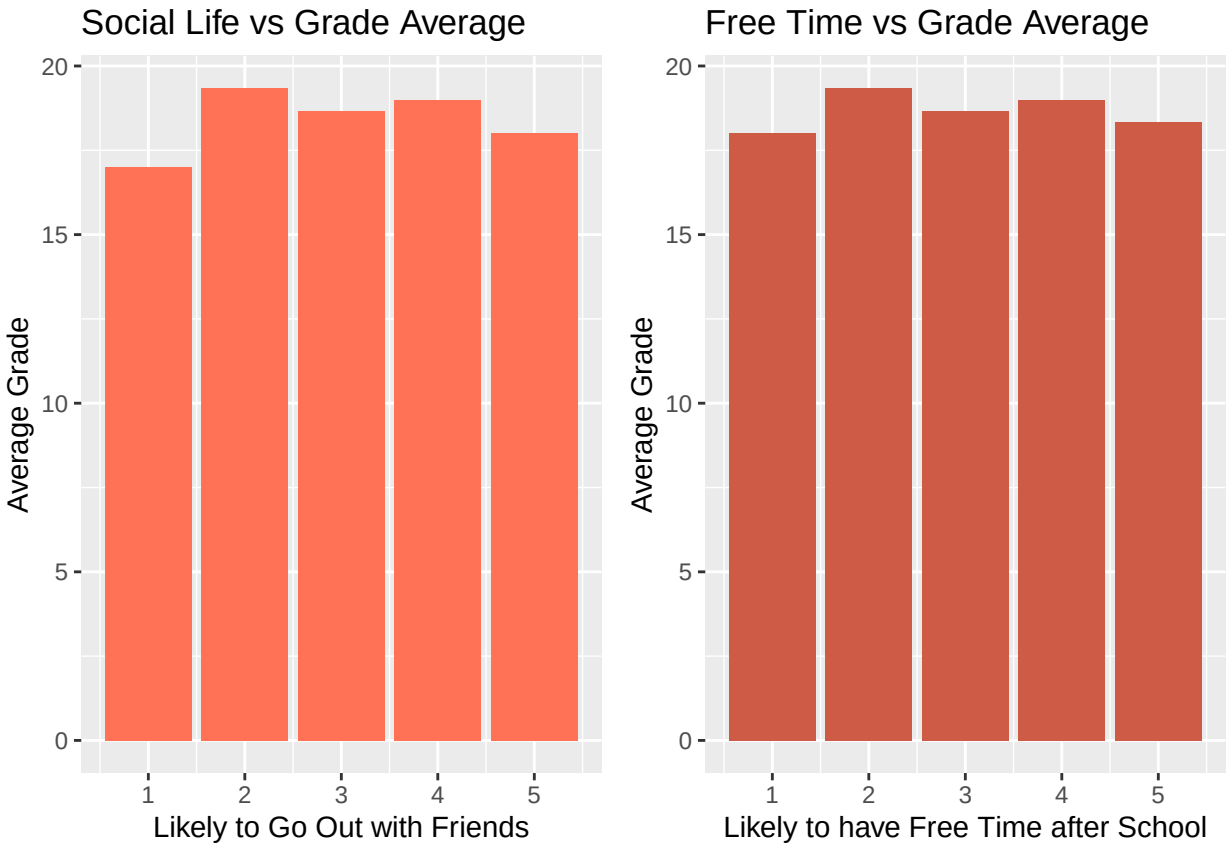
```
# Students who go out moderately (3 or above)
length(student_data$goout[student_data$goout >=3])/nrow(student_data)
```

```
## [1] 0.6810127
```

```
# Students who have a moderate amount of freetime (3 or above)
length(student_data$freetime[student_data$freetime >=3])/nrow(student_data)
```

```
## [1] 0.7898734
```

A majority of students, 68% and 78%, have a healthy social life with spending time with friends and having ample free-time after school.



Between the different levels of social lives the student hold, there is little difference. When looking at the levels of free time students have, we also see little difference. Overall, students who have free time and/or go out with friends have a higher average grade compared to those that don't.

Romantic Relationships

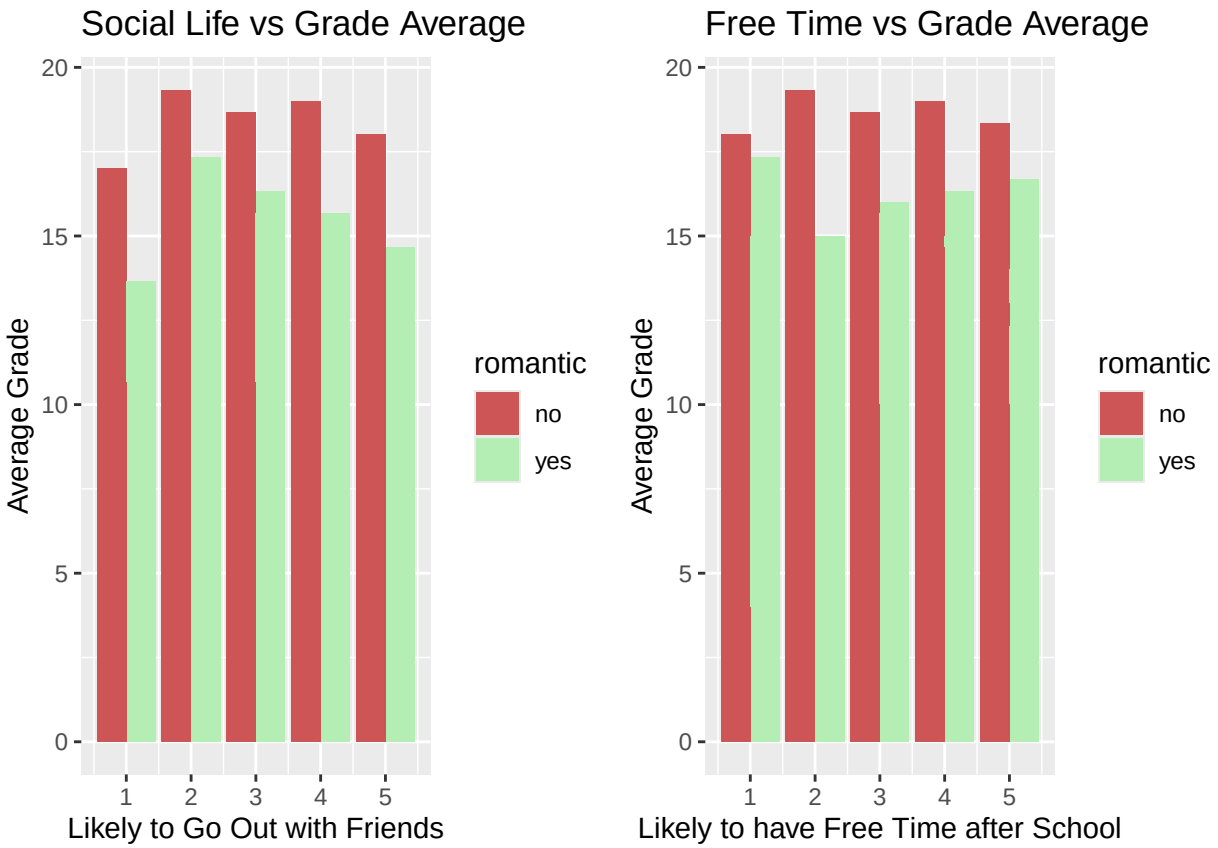
```
# Students who are in a relationship
summary(student_data$romantic)
```

```
##  no yes
## 263 132
```

```
length(student_data$romantic[student_data$romantic == "yes"])/nrow(student_data)
```

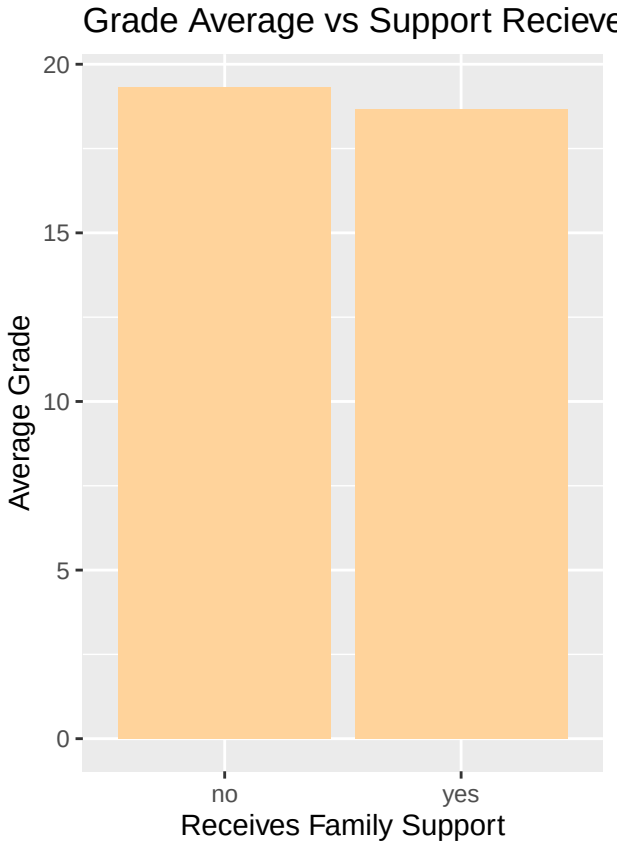
```
## [1] 0.3341772
```

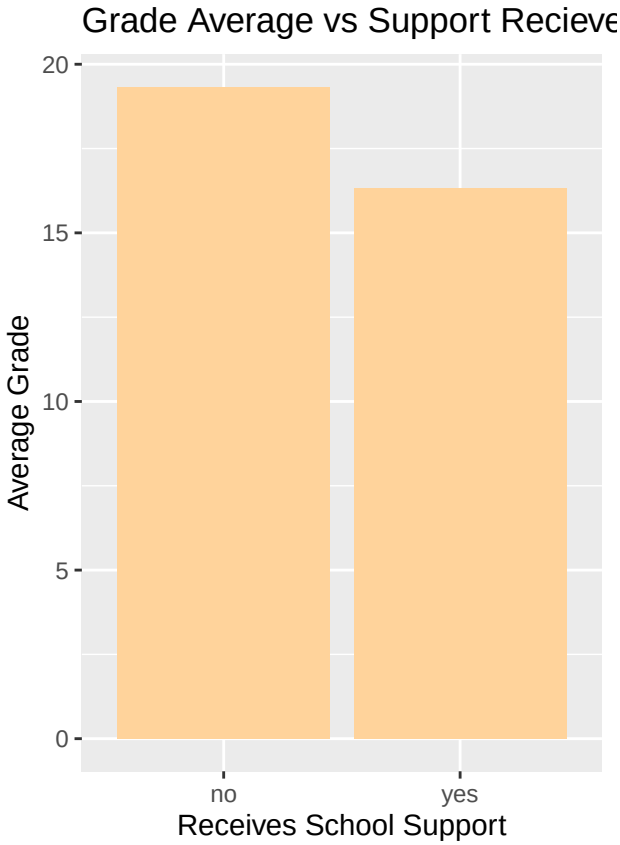
Looking at these percentages, about one third of the students surveyed are in a romantic relationship. We do not have additional information on if both parties in the relationship were surveyed.

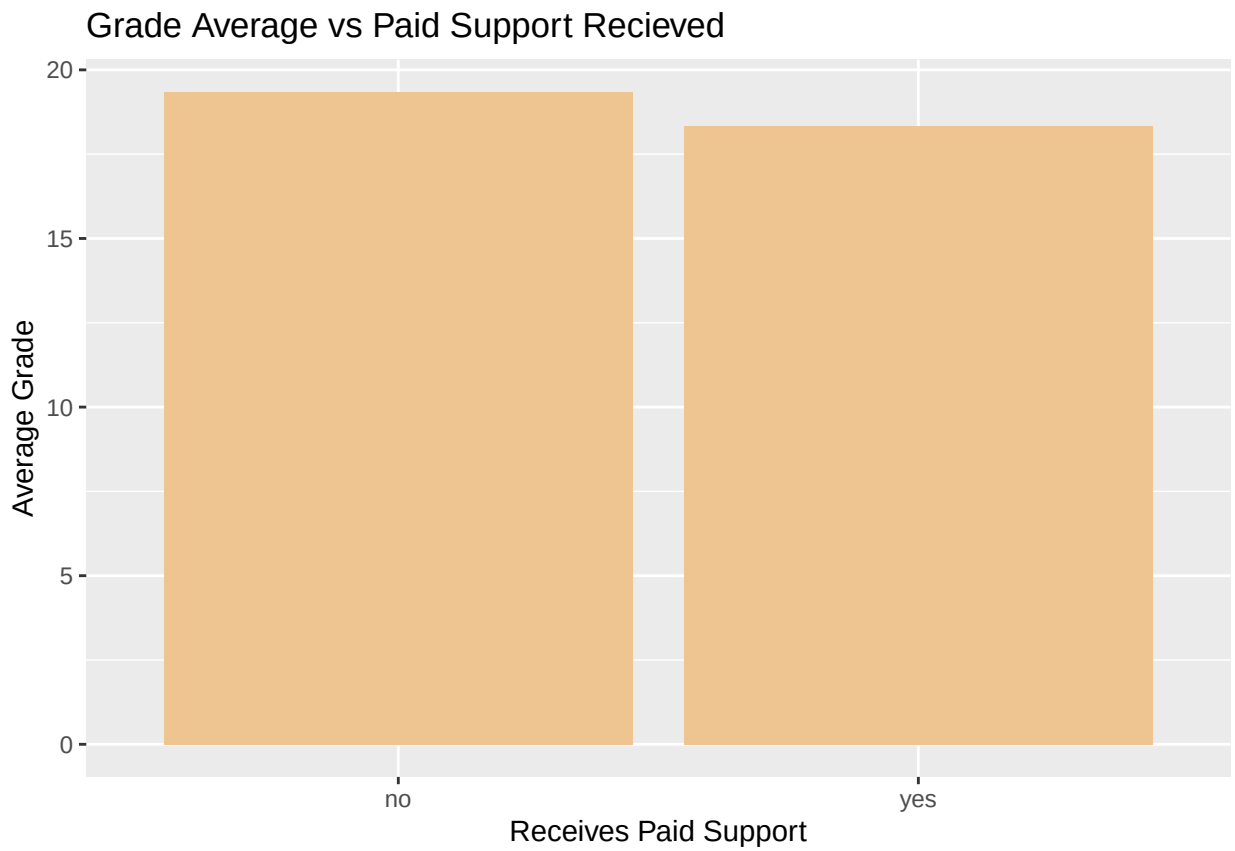


When looking a bit further in social relationships when compared to romantic ones, we can see across the board those who are involved with a romantic relationship have a slightly lower average than those who don't.

Additional Support







Looking at the few plots above, it seems those who are receiving support from either the school or their families are performing slightly worse than those who do not. The average grade for those enrolled in extra paid classes within the course subject is slightly lower than those not enrolled in one. These results could be that the student's lower grades warranted the need for additional support, either from families, the school or through paid classes.

Conclusion

From this analysis, we see that there are a few factors heavily contributing to better student performance in Mathematics as well as certain traits that can hinder a student's ability.

It appears the **highest performing students** are from an *urban neighborhood* and *well educated family*, *who study 5-10 hours weekly* and have *less than 10 absences* in their school.

Drinking during the workday and *maintaining a romantic relationship* can **impair** a students academic success.

Students who receive *additional support* and *participate in Extra-Curricular activities* are **no more ahead** in their academics.