

A1.

1)

There are 1500 students in the dataset

2)

The age range for students is 15-18

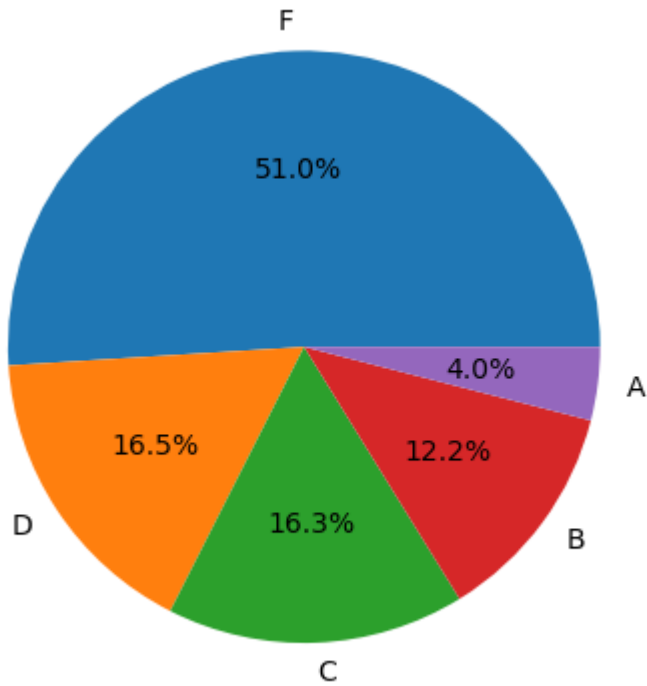
3)

Column Name	Data Type
Student ID	Numerical, Discrete
Age	Numerical, Discrete
Parental Education	Categorical, Ordinal
Weekly Study Time	Numerical, Continuous
Absences	Numerical, Discrete
Tutoring	Binary
Parental Support	Numerical, Discrete, Ordinal
Extracurricular	Binary
Sports	Binary
Music	Binary
Volunteering	Binary
GPA	Numerical, Continuous
Grade Class	Categorical, Ordinal

4)

Grade	Proportion
F	51%
D	16.53%
C	16.27%
B	12.2%
A	4%

5)



The pie chart shows that the grade data is positively skewed, with much of the class cohort failing the class (over 50%!). Fewer students are able to reach higher grades, as seen by the decreasing proportion of students as the grade letter increases.

## A2.

1)

77 Parents have an education level of "Higher"

154 Parents have no education

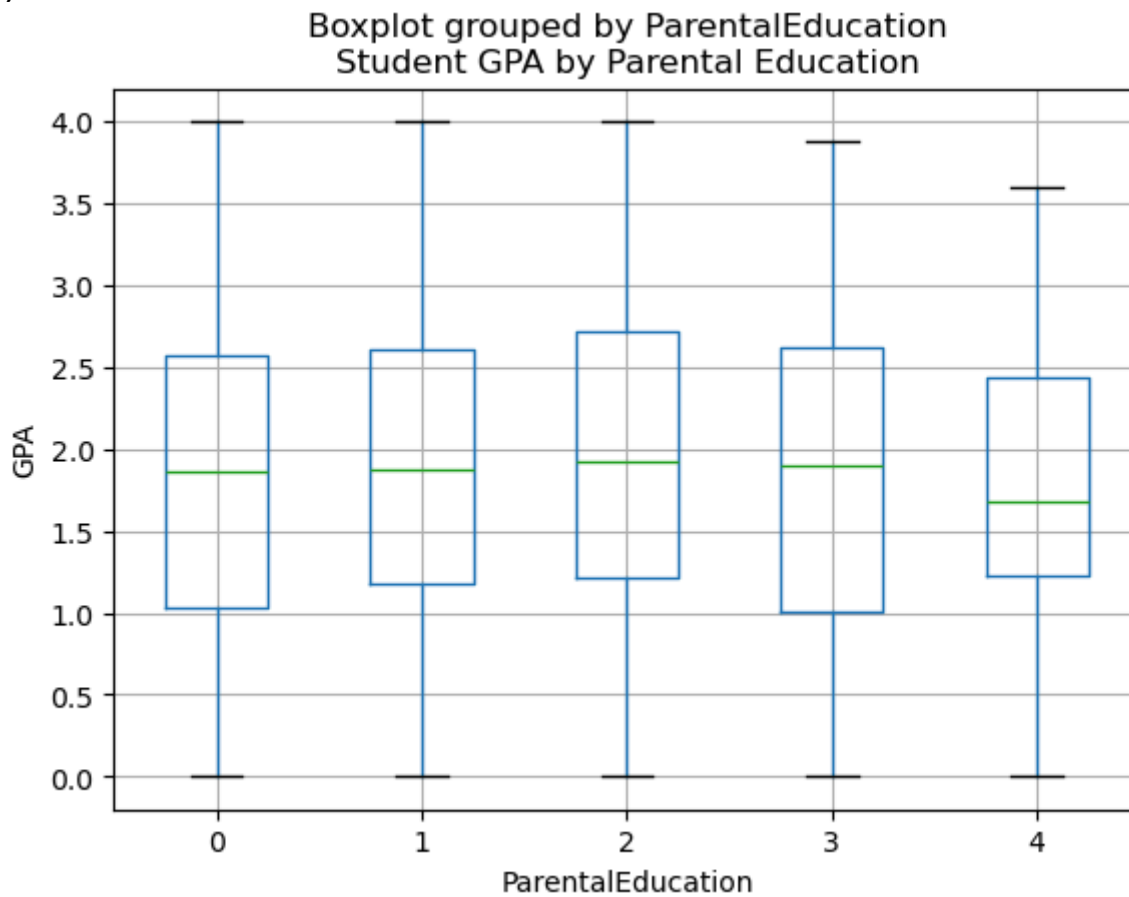
The most common level of parental education is "High School"

2)

line 16 in jupyter notebook:

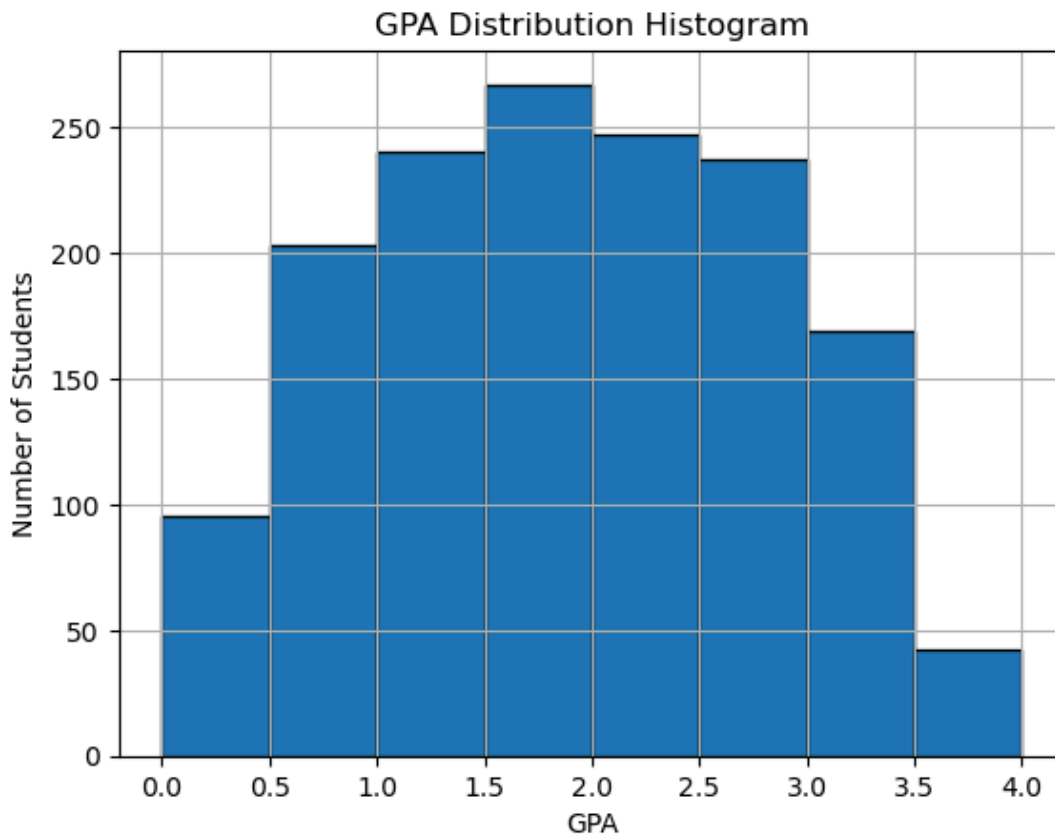
```
[16]: data["ParentalEducation"] = data["ParentalEducation"].replace(  
      {"No Education": 0, "High School": 1, "Some College": 2, "Bachelor's": 3, "Higher": 4})
```

3)



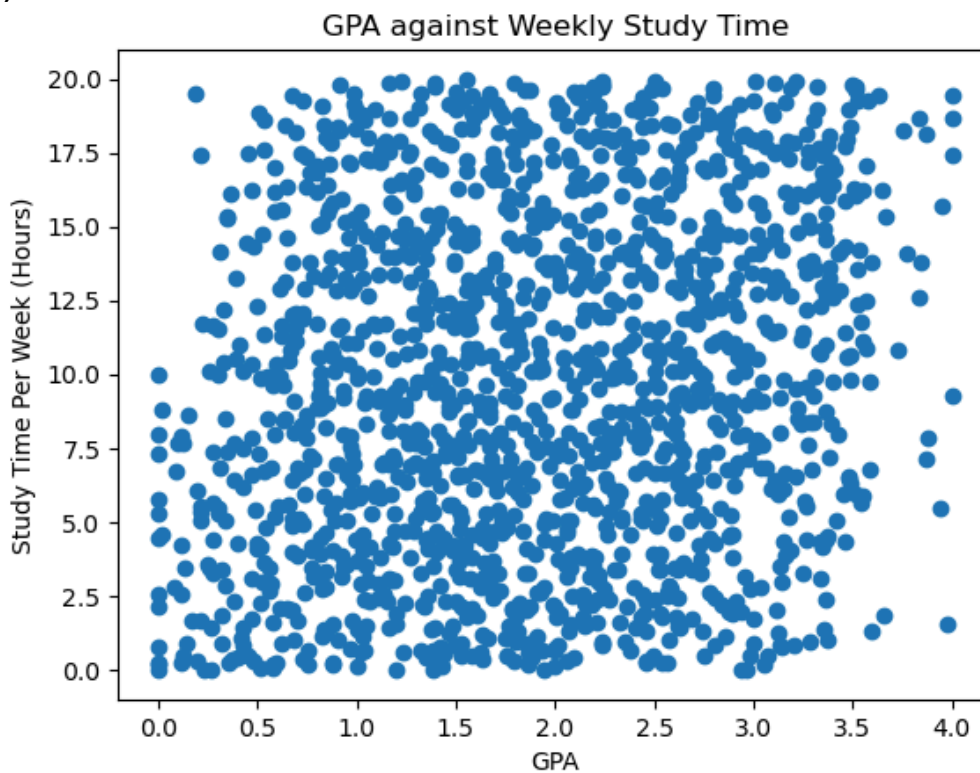
Looking at the boxplot, I can say that there is not a meaningful relationship between parental education and student GPA. This is because the median values are all very similar for all parental education levels, and while children of parents with a “Higher” level of education seem to have a lower maximum and tighter quartiles, I believe this is simply because there are fewer parents with this level of education, leading to a lack of data.

A3.  
1)



We can see that most students score around the middle of the GPA range, from 1.0 - 3.0 GPA. On the extremes, there are more students who are scoring very low (0 - 0.5 GPA) than there are scoring very high (3.5 - 4.0 GPA).

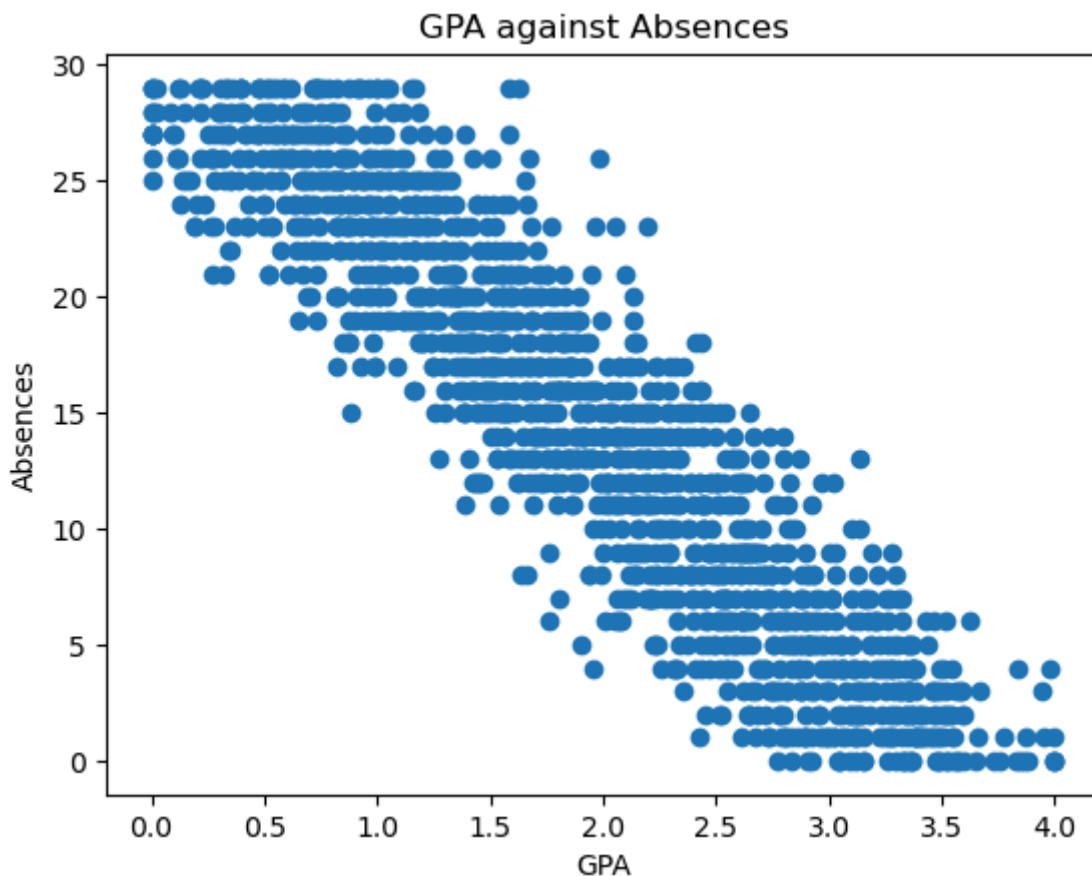
2)



Correlation Coefficient == 0.1904931303711251

There is a weak positive relationship between weekly study time and GPA

3)



Correlation Coefficient == -0.9194876943290947

There is a strong negative relationship between absences and GPA. This means that having a higher number of absences has a big impact on GPA

**A4.**

1)

There are 5 students who participate in all of the listed activities

2)

There are 432 Students who participate in none of the listed activities

3)

Group A mean GPA == 2.45

Group B mean GPA == 1.73

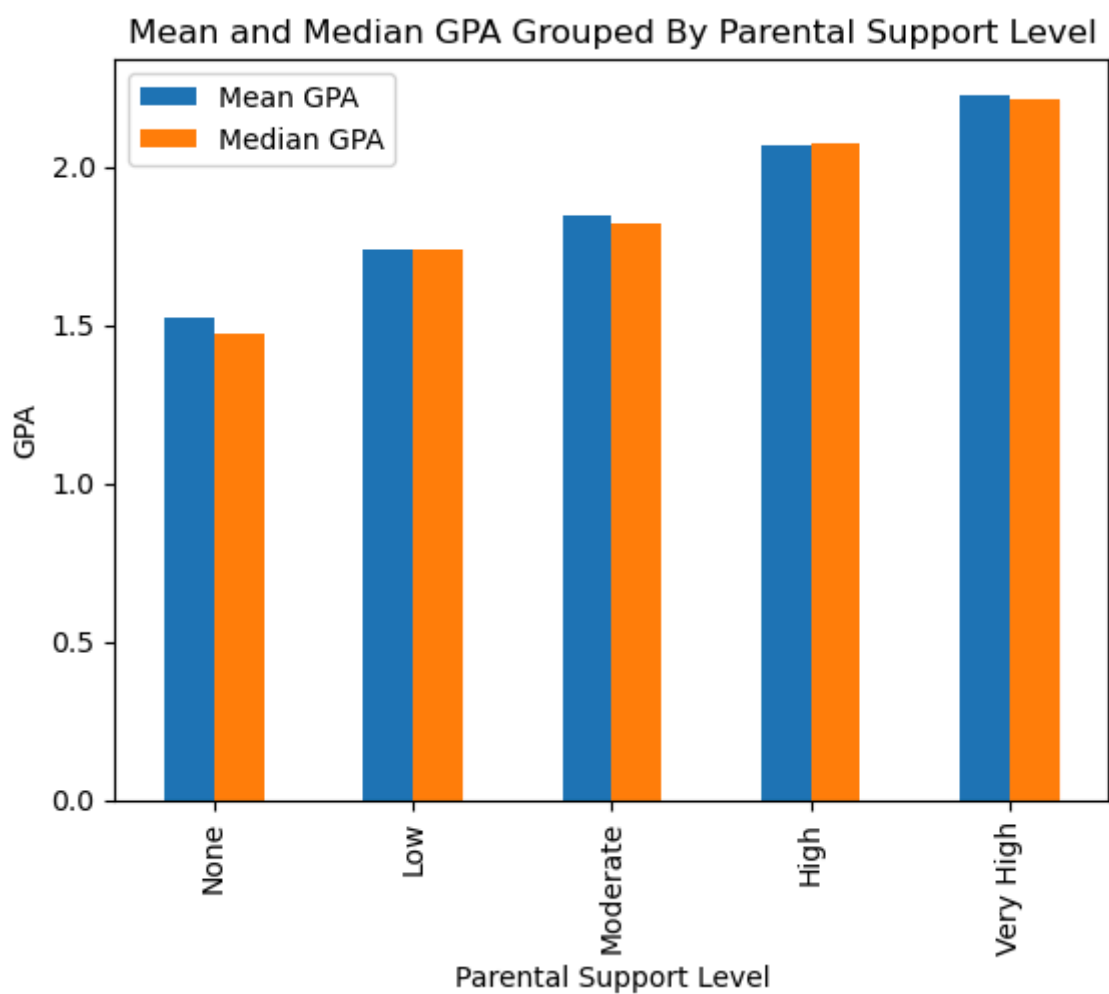
Students who took part in all possible non study activities had a higher mean GPA than students who took part in none. However, there are only 5 students who took part in all activities available, so I would be hesitant to use this information to make any meaningful conclusions, as the sample size is too low

A5.

1)

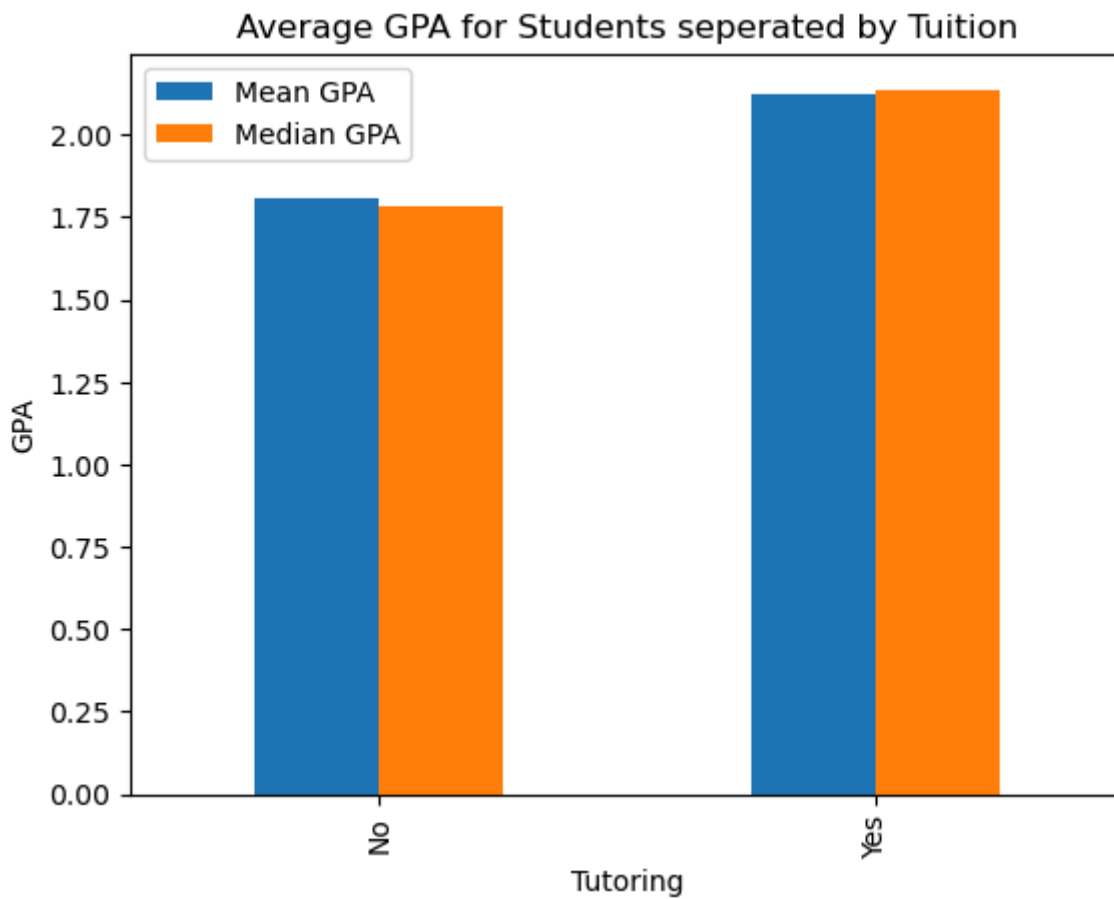
Parental Support	Mean GPA	Median GPA	No. 18 Year Olds
0	1.521602	1.471672	33
1	1.735855	1.740455	80
2	1.845914	1.817007	98
3	2.068174	2.070669	116
4	2.227639	2.215516	35

2)



With increasing levels of parental support, we can see that there are increases in both mean and median GPA. This means that more parental support helps almost all students perform better at school.

3)



This bar chart shows that students who have tutoring perform, on average, better than those who do not have tutoring. This can be seen by the higher levels of both mean and median GPA scores for students in tutoring against lower values for students who do not participate in tutoring.