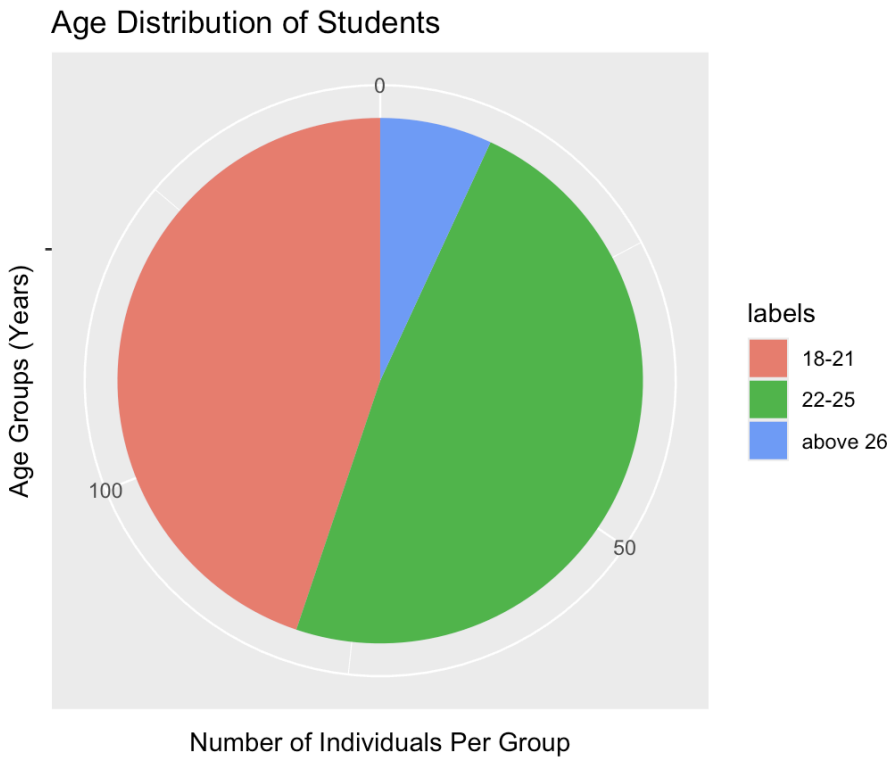
Capstone Project 1 Report
Abhi Wagh

Demographics Data

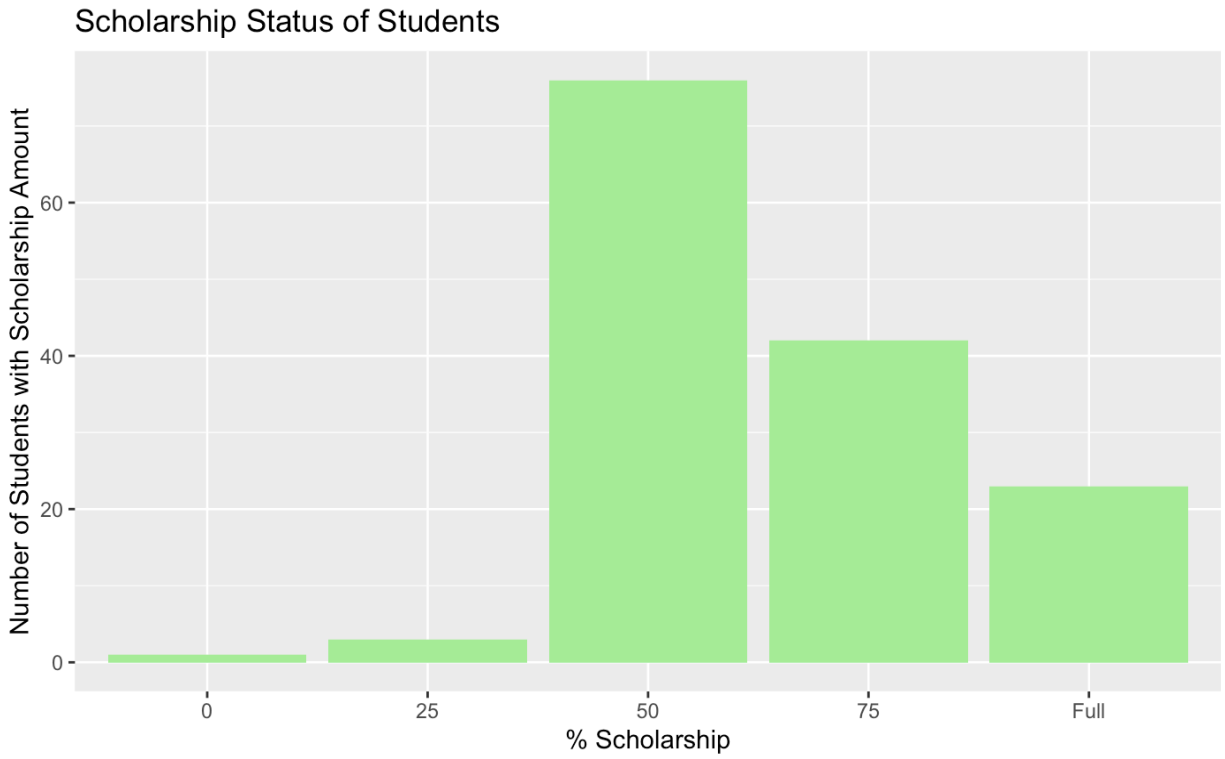
The dataset I chose to examine was established in 2019, likely surveying Engineering and Educational Sciences majors at a university in Cyprus (not clear whether it's referring to Cyprus the country or a city/town). Before delving into my main research questions, I wanted to get a sense of the demographics of the student population that was surveyed. The data shows more men than women were surveyed in this data (87 men vs. 58 women). I hypothesized the ratio of men to women would be closer to equal, as this would yield more accurate insights upon analysis.

Additionally, the majority of students fall into Group 2 age-wise and are between 22-25 years old. Only 10 students or about 6.90% of the students surveyed are above 26 years old. This data suggests that the university where the data was collected mostly has undergraduate programs. It could serve a smaller population of graduate students or there are just some older undergraduate students, which could account for students above 26.



Furthermore, I analyzed whether students had a partner or not, their parental status, and what scholarship they were on (if they were on one). The ratio of not partnered to partnered students was 84:61. Parental status data showed that the majority of students' parents were alive and married. The ratio of students with married parents to divorced parents to deceased

parent(s) was 127:11:7. Lastly, scholarship data suggested that 52.41% of all students or 76 students surveyed had a 50% scholarship.



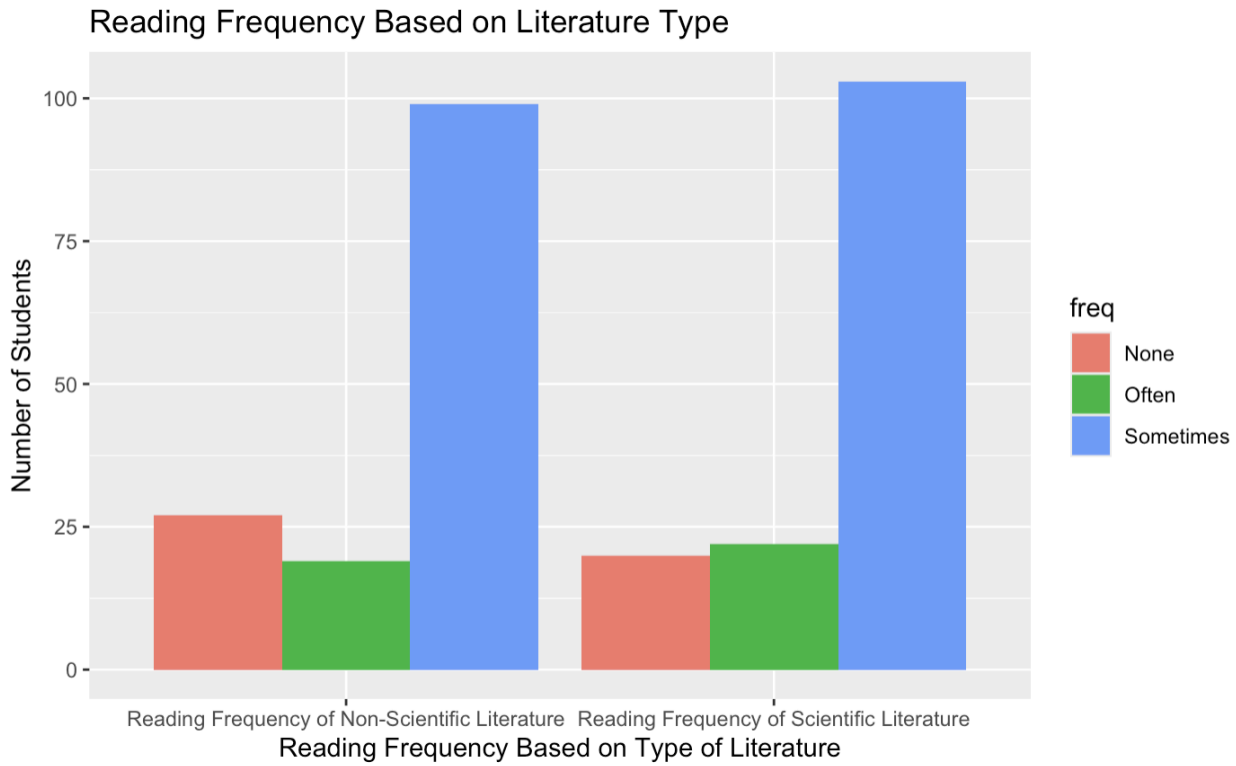
Research Question 1: What trends exist in studying behaviors among university students? How do those relate to cumulative GPA?

My main research question focused on how cumulative GPA was impacted by assorted studying behaviors. The studying behaviors I chose were reading scientific literature or not, studying alone or with friends, and the number of hours students studied per week. I also analyzed which types of high school most university students came from and how that might tie in with cumulative GPA. My hypothesis was that behaviors such as reading scientific literature, studying several hours a week (>20), and studying with friends predict an overall higher GPA.

Research Question 1a: Do students in this dataset read scientific literature or non-scientific literature more often? How is this related to cumulative GPA?

Context-wise, the student population that was surveyed in this dataset is mostly STEM majors (engineering and Educational Sciences). I hypothesized that students read scientific literature more often, based on the background information of the dataset. Overall, there were more students that read scientific literature than non-scientific literature. Out of the students who read scientific literature, 71.03% of them or 103 students read it sometimes. Out of those that read non-scientific literature, 68.28% of students or 99 students read it sometimes. Because of these values, 103 and 99, it's likely that some students read both scientific and non-scientific literature. (It would've been more accurate to get the overlap value of students that read both types of literature and their frequency of reading them.)

Lastly, I chose to analyze cumulative GPA trends only for scientific literature reading frequency. (Because the results were inconclusive for this analysis, I didn't analyze the same trends for non-scientific literature reading frequency.) I found that most data points fell in the 'sometimes' category. The highest mean cumulative GPA, according to this analysis, was 3.25 for students who never read scientific literature. It could be that they are also part of the group that reads non-scientific literature, or other confounding variables are muddying the data. Data analysis on this aspect of the data proved to be inconclusive.



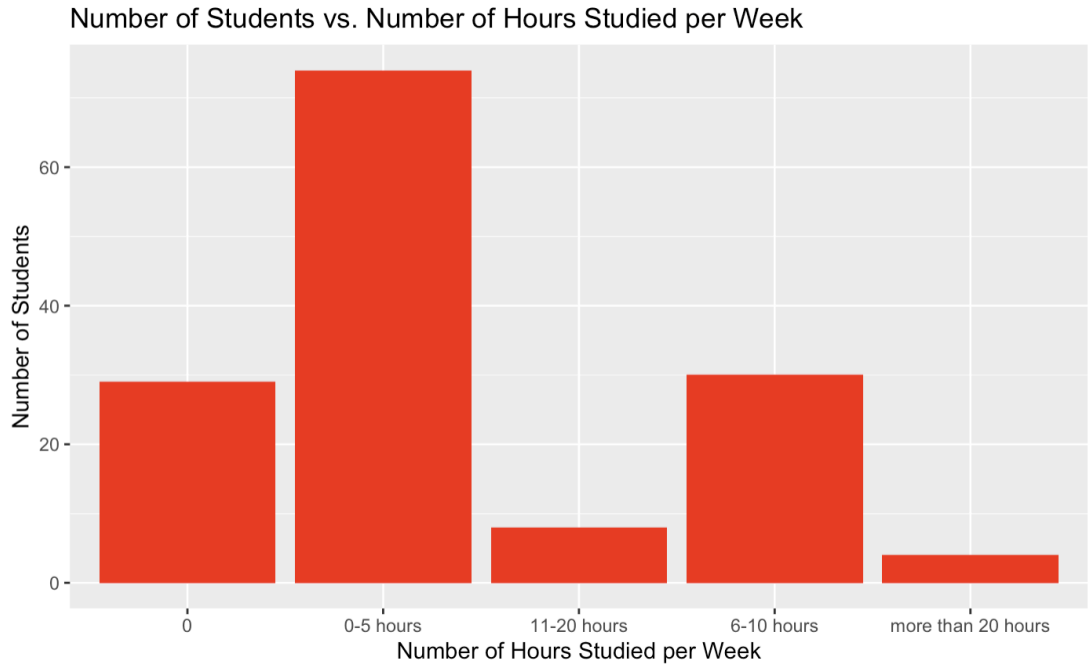
Research Question 1b: Is studying alone versus studying with friends for midterms associated with higher cumulative GPA?

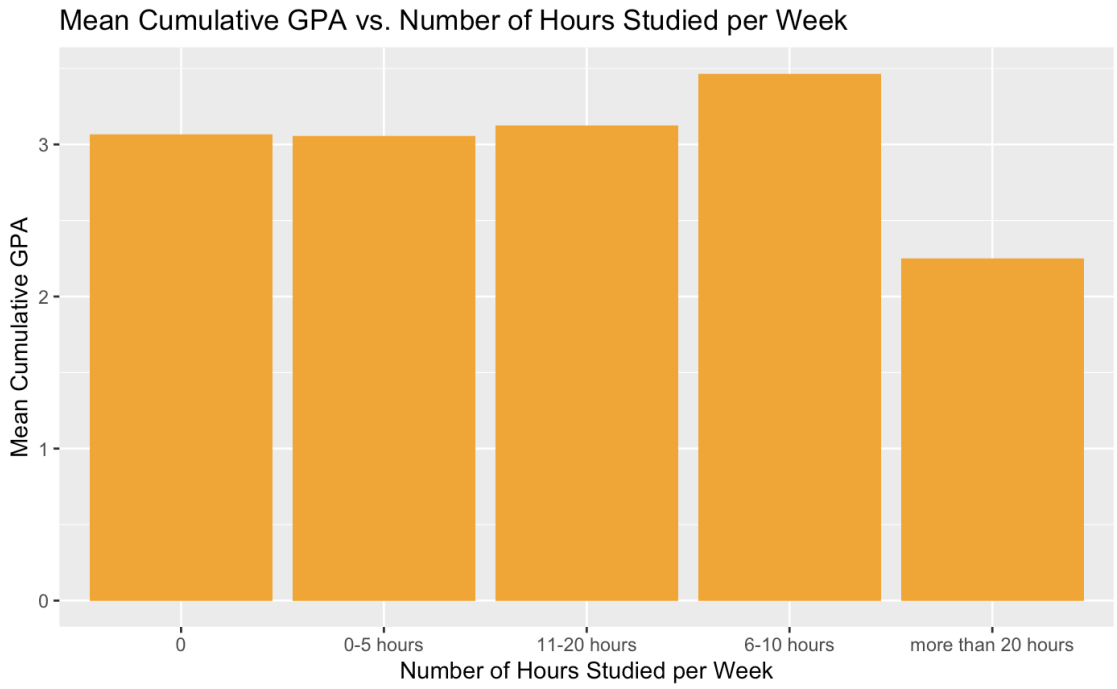
Upon analyzing how students study for midterm exams, I found that 107 students study alone, 27 students study with friends, and 11 students don't study. There are significantly more data points for students who study alone than those who study with friends. Again, the relationship between these parameters and cumulative GPA is likely inaccurate due to the skewed data.

The students who didn't study for midterms had the lowest calculated GPA at 2.82. Mean GPA is almost 0.2 higher for studying alone than with friends, but data is likely skewed. It also doesn't account for people who might do both study behaviors.

Research Question 1c: What is the distribution of the number of hours students study per week? How does it connect to cumulative mean GPA?

74 students or about 51.03% of students studied less than 5 hours a week. The next most popular amounts of time to study weekly were 6-10 hours and 0 hours. As for GPA data, students who study 6-10 hours per week have the highest mean cumulative GPA (3.47). Only 4 students study more than 20 hours a week, and their average cumulative GPA is 2.25. This could again be an inaccurate statistic, but it's possible that studying for more than 20 hours a week doesn't actually help boost GPA. Studying 6-10 hours a week could be the ideal amount of time to study to boost or maintain a high GPA, as the GPA decreases the more subsequent hours one studies after that.

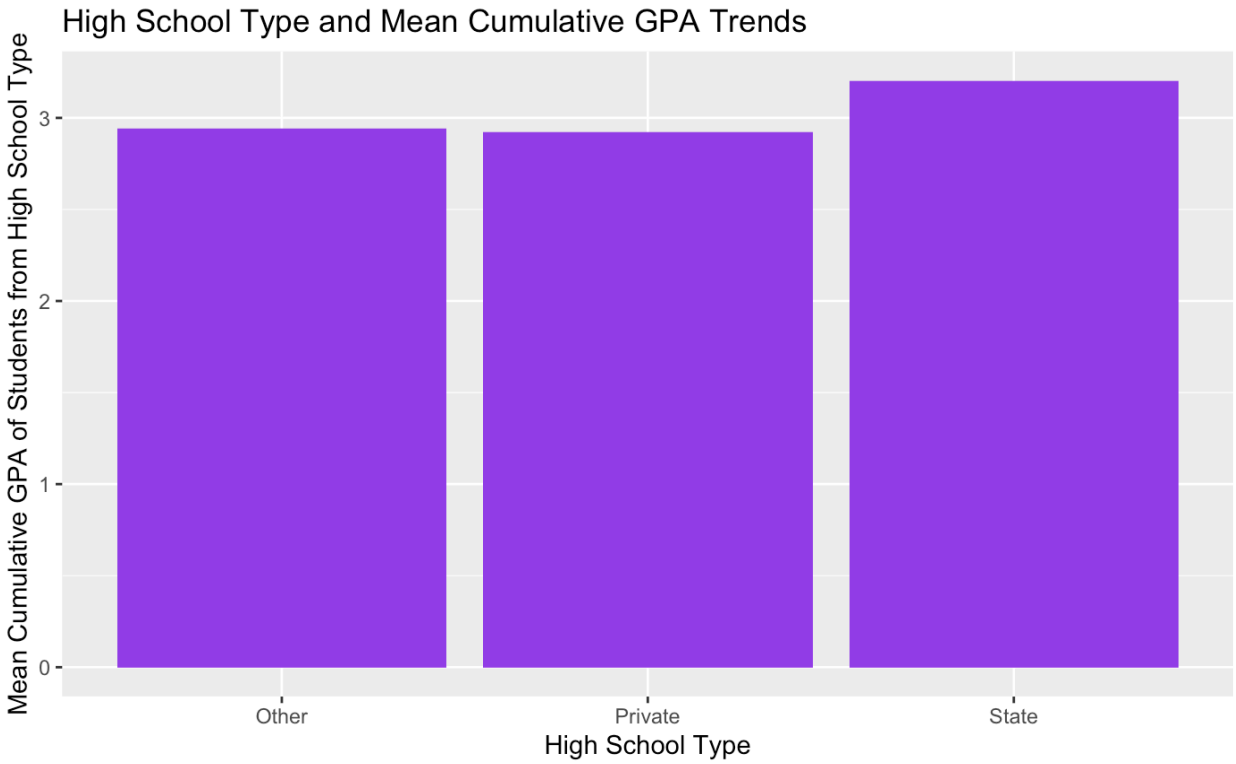




Research Question 1d: What is the relationship between high school type and cumulative GPA?

After looking into study behaviors, I was curious about how the type of high school university students graduated from were related to mean cumulative GPA. There has been a raging debate that continues today about whether private or public schools offer a more quality education. The breakdown for university students by the type of high school they graduated from was 103 students from state high schools, 25 students from private high schools, and 17 from other types of high schools. (Although it was not specified what 'other' meant, I assume it could represent high schools with specialized programs such as arts schools.)

71.03% of students surveyed graduated from state high schools, and it appears they had the highest cumulative GPA (3.20) out of all the groups studied. Again, this could be a skewed statistic based on most of the data points being from the state high school group.



Research Question 2: Which learning methods are preferred by students most in university? (taking notes, listening, discussion)

As this student population primarily consists of STEM majors, it's likely that they have to learn terminology-heavy content. This made me curious as to what their preferred learning methods are. Students also tend to differ in their learning preferences, so this analysis could provide instructors at the university with useful data as to what majority of their students prefer. It could allow them to also diversify the way they teach to accommodate students who don't learn in the same way as most of their peers.

For students who always use a specific method of learning or find it always useful, note-taking is the preferred method of learning for 84 students and discussion in class is a close second for 66 students. For those who only sometimes use a particular method, listening in class was the winner and discussion again a close second. Overall, listening in class was the least preferred method of learning.

Conclusion

Although this dataset was flawed, several insights were extracted from this dataset. The majority of the student population surveyed was under age 26. This suggested the university where the data was collected from was likely mostly serving undergraduate students with a smaller graduate student population. Further information would be needed though to fully verify this assumption. Additionally, more students surveyed were men than women. Although this can result in a slightly imbalanced dataset, the reasoning behind this could be that the university or the two departments of engineering and educational sciences are male-dominated.

There were also slightly more unpartnered students than partnered students. This could be due to the stress of college and the lack of free time to develop relationships. Further information would again be needed to verify this. Most students also had parents who were both alive and married, which could be relevant if they are footing the bill for the students' tuition and fees. Dual-income households might have a higher rate of paying for their children's college fees. Lastly, about half the student population surveyed were at university on a 50% scholarship.

This might suggest that the university has a robust scholarship-providing program, especially for the specific majors of engineering and educational sciences.

Analysis for my main research question showed that the data was skewed, so a large part of it is likely inaccurate. However, some clear insights did come out of it. Students tend to read more scientific literature than non-scientific literature, which makes sense given their educational backgrounds. Most students prefer to study alone instead of with friends for midterm exams. It would be interesting to get more data on if studying habits change for smaller unit exams or larger exams like finals. About half of students also tended to study between 0-5 hours per week, and the next most frequent amount of studying was between 6-10 hours. Most students, around 70%, graduated from state high schools. Lastly, I analyzed preferred learning methods for students in this dataset. Overall, it seems that note-taking was the most preferred method of learning, while listening in class was the least preferred method.