

Rebecca Higbee, Programmer/Analyst 1 Application

SQL Interview Questions

Comments, denoted by two hyphens, are written often throughout the code for transparency and clarity. SQL best practices for formatting and styling have been adhered to. Code is written with SQLite in DBeaver.

```
<DBeaver Sample Database (SQLite)> SQL Psuedo Code-Higbee, Rebecca X
--#1: How many students are registered each term in Spring 2023 and Fall 2023?

--Naming the CTE "small registration dataframe"
WITH reg_small_df AS

--Each row is a unique combination of anumber, term, and reg_status from the Registration table
(SELECT DISTINCT r.anumber, r.term, r.reg_status
FROM Registration_tbl r
--Filtered by students under the registered status
WHERE r.reg_status = "Reg"
--Filtered by students not in Summer 2023, leaving Spring 2023 and Fall 2023
AND r.term != "Summer 2023")

--From the registration small dataframe, count by term
SELECT *, COUNT(term)
FROM reg_small_df
--Group the counts by term
GROUP BY term;

--#2: Provide a list of students with their email addresses who are registered for Spring 2023.
--Exclude any students who have a FERPA Confidentiality Block.

--Naming the CTE "small student dataframe"
WITH student_small_df AS

--Selecting all necessary columns from the Student and Registration tables
(SELECT s.anumber, s.last_name, s.first_name, s.email, s.ferpa_confid_block,
r.anumber, r.term, r.reg_status
--Left join Registration onto Student using student Anumbers
FROM Student_tbl s
LEFT JOIN Registration_tbl r ON s.anumber = r.anumber
--Filter by Spring 2023 registration term and no FERPA block (denoted by the !=)
WHERE r.term = "Spring 2023" AND ferpa_confid_block != "Y")

--From the student small dataframe, display student full names and respective emails
SELECT last_name, first_name, email
FROM student_small_df;
```

```
<DBBeaver Sample Database (SQLite)> SQL Psuedo Code-Higbee, Rebecca ×

--#3: Provide an email list of students who are registered for an English (ENGL) class.

--Naming the CTE "small course dataframe"
WITH course_small_df AS

--Selecting all necessary columns from the Course, Registration, and Student tables
(SELECT c.crn, c.subject, r.anumber, r.crn, r.reg_status, s.anumber, s.email
--Left join Registration onto Course using crn identification
FROM Course_tbl c
LEFT JOIN Registration_tbl r ON c.crn = r.crn
--Left join Student onto Registration using student anumbers
LEFT JOIN Student_tbl s ON r.anumber = s.anumber
--Filter by registered students in an English course
WHERE c.subject = "ENGL" AND r.reg_status = "Reg")

--From course small dataframe, display only emails in the final output
SELECT email
FROM course_small_df;

--#4: Provide a list of students who are registered Fall 2023 and indicate if the students
--have applied to graduate Fall 2023.

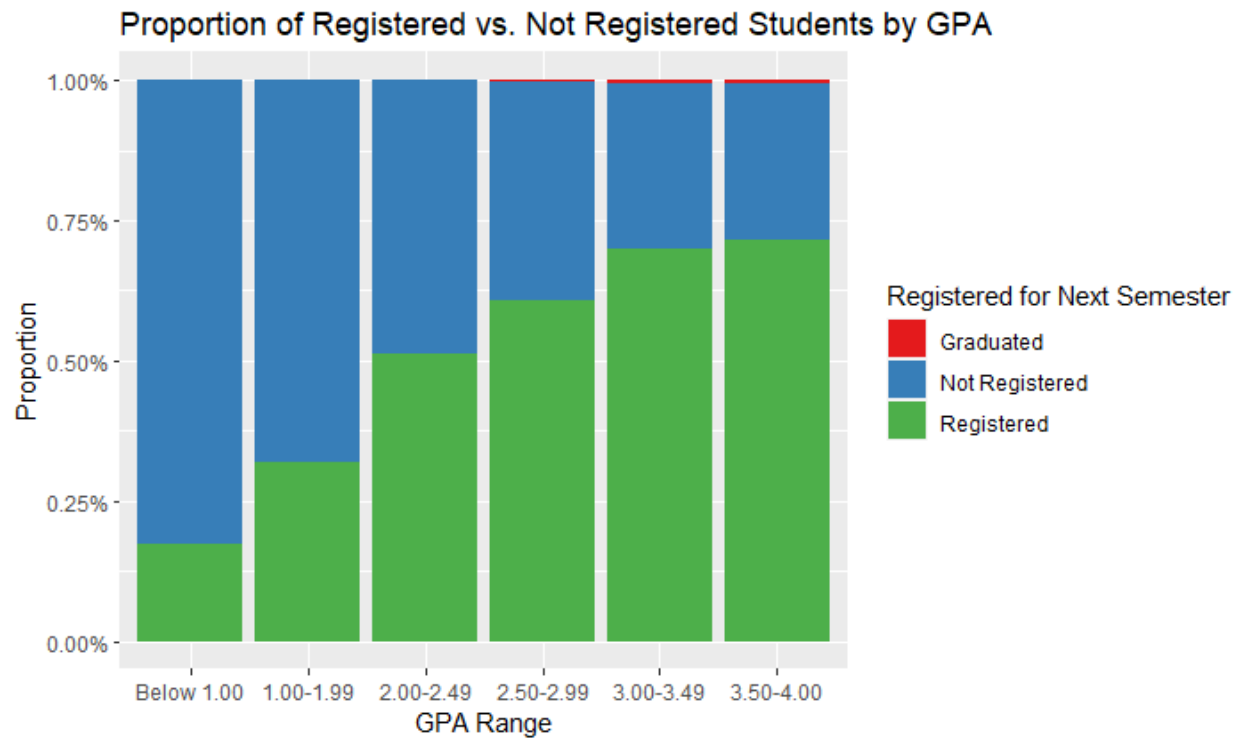
--Naming the CTE "small graduation application dataframe"
WITH grad_small_df AS

--Selecting all necessary columns from the Grad, Registration, and Student tables
(SELECT g.anumber, g.term, r.anumber, r.term, r.reg_status, s.anumber, s.last_name, s.first_name
--Left join Registration onto Grad using student anumbers
FROM Grad_app_tbl g
LEFT JOIN Registration_tbl r ON g.anumber = r.anumber
--Left join Student onto Registration using student anumbers
LEFT JOIN Student_tbl s ON r.anumber = s.anumber
--Filter by registered Fall 2023 students with graduation applications for Fall 2023
WHERE g.term = "Fall 2023" AND r.term = "Fall 2023" AND r.reg_status = "Reg")

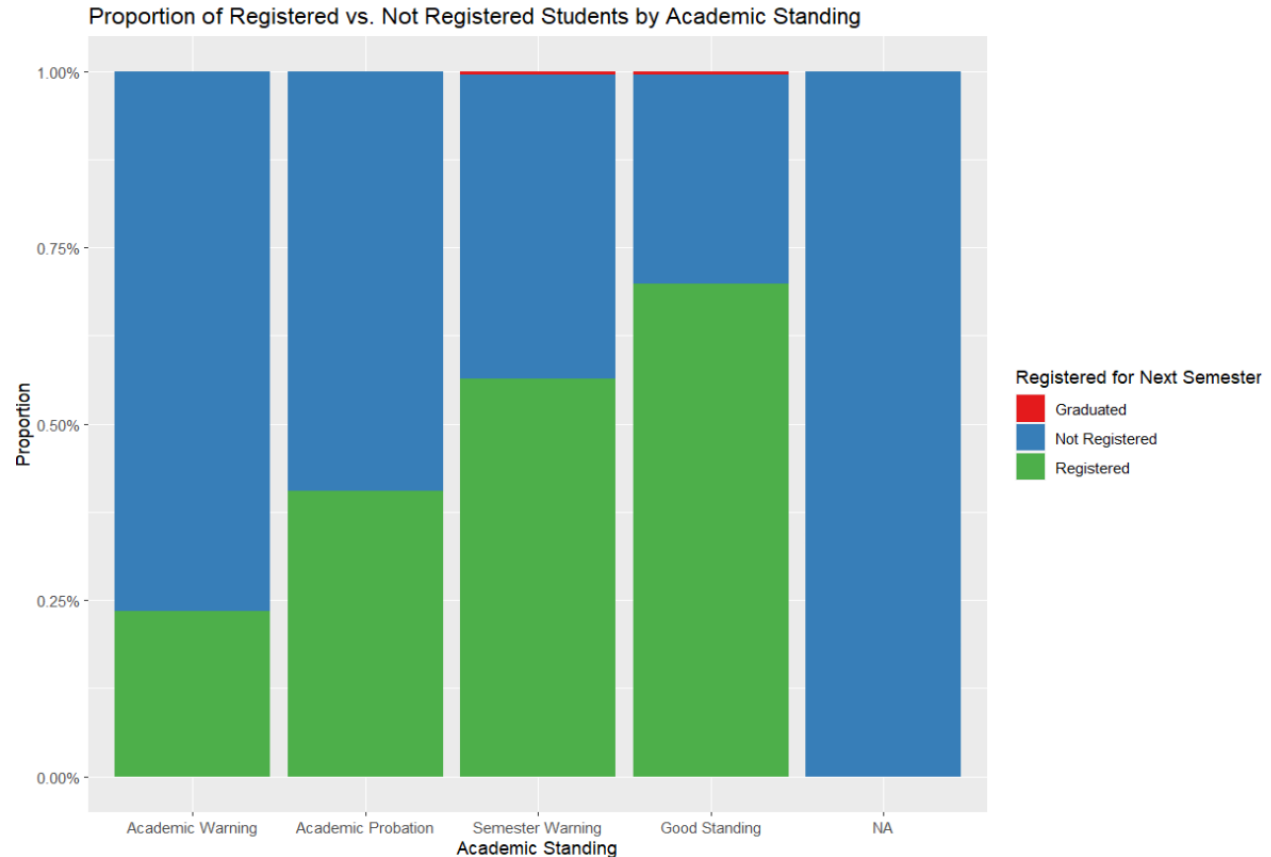
--From the small graduation application dataframe, display only student full names
SELECT last_name, first_name
FROM grad_small_df;
```

Example Retention Dataset:

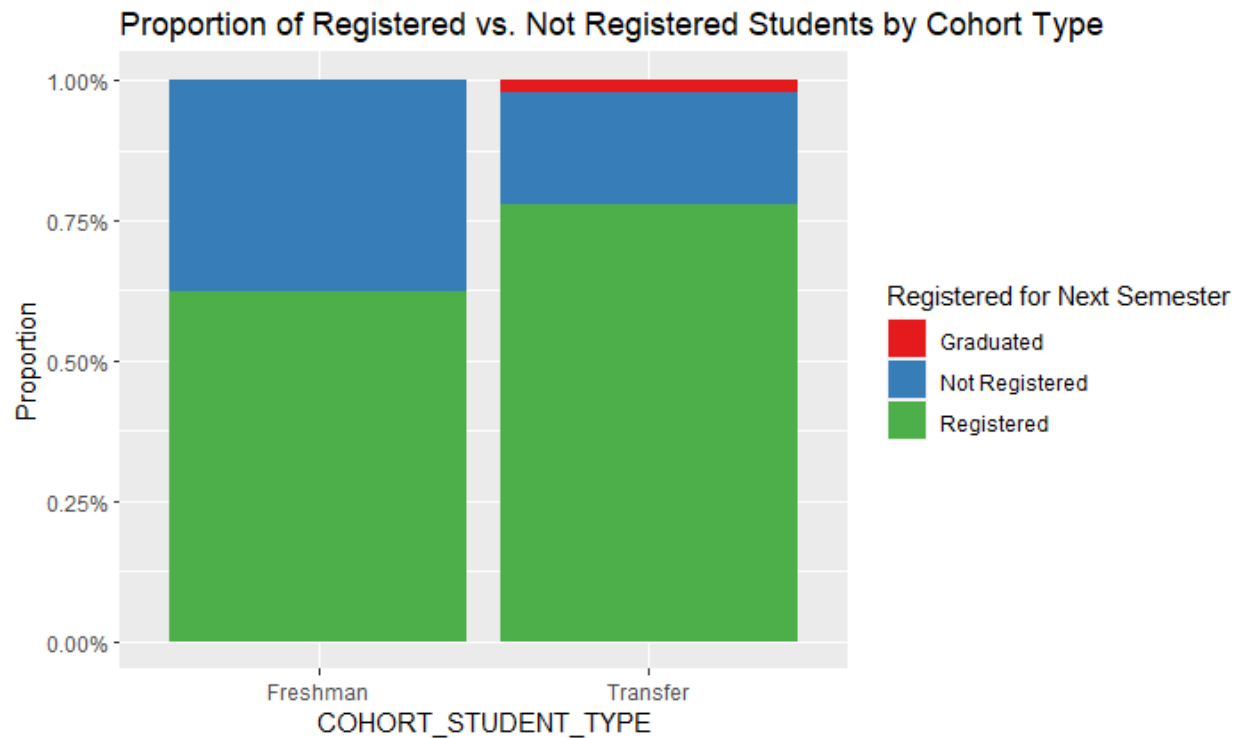
Using R for exploratory data analysis, I have created visualizations that shows the factors that contribute to student retention with insights below each graph. R best practices for formatting and styling have been adhered to, including the use of the lintr and styler packages. Furthermore, color blind appropriate color palettes have been used to increase user readability. To be concise, though many graphs could be shown, only those displaying an effect on student retention will be presented here. Code is included as a separate document in this email for reference, readability, and accountability.



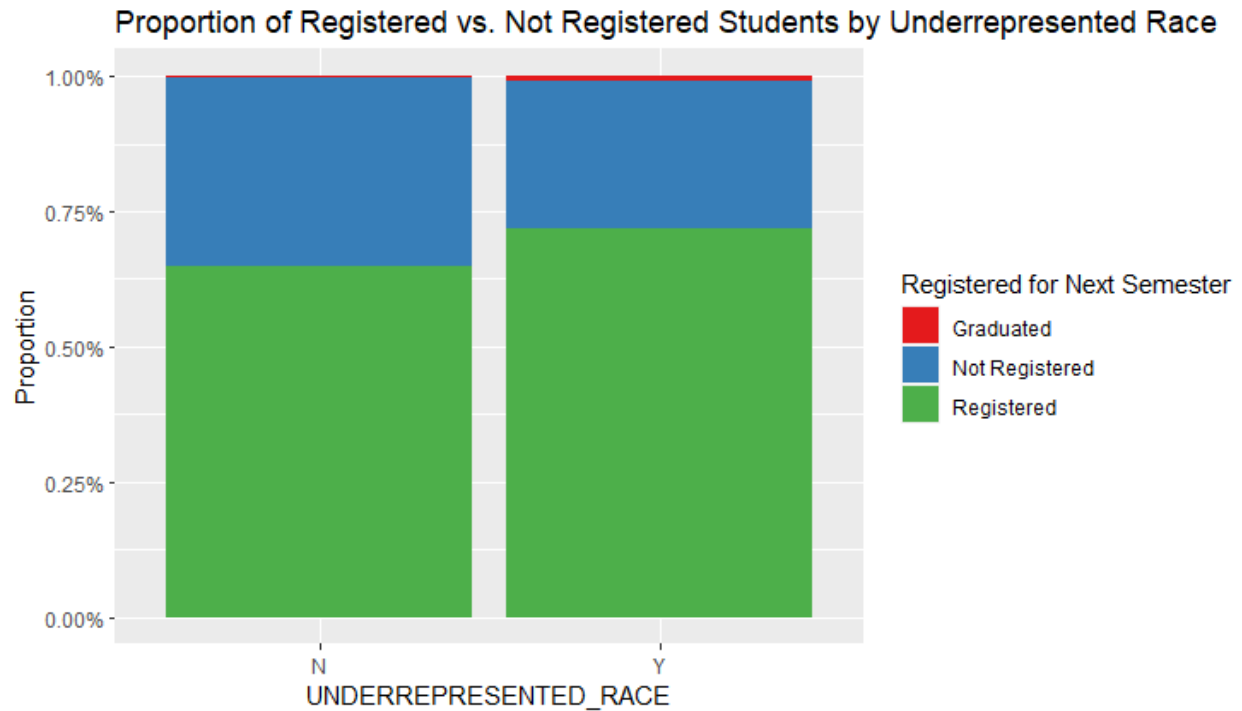
It appears that as a student's GPA range increases, the proportion of students registered for the next semester increases as well. Students who graduated, shown in red, all graduated with a 2.50 GPA or greater. It appears that GPA is a factor of student retention, as a majority of those with lower GPAs tend to not register for the next semester and a majority of those with higher GPAs do register. More data exploration should be done on the factor of GPA.



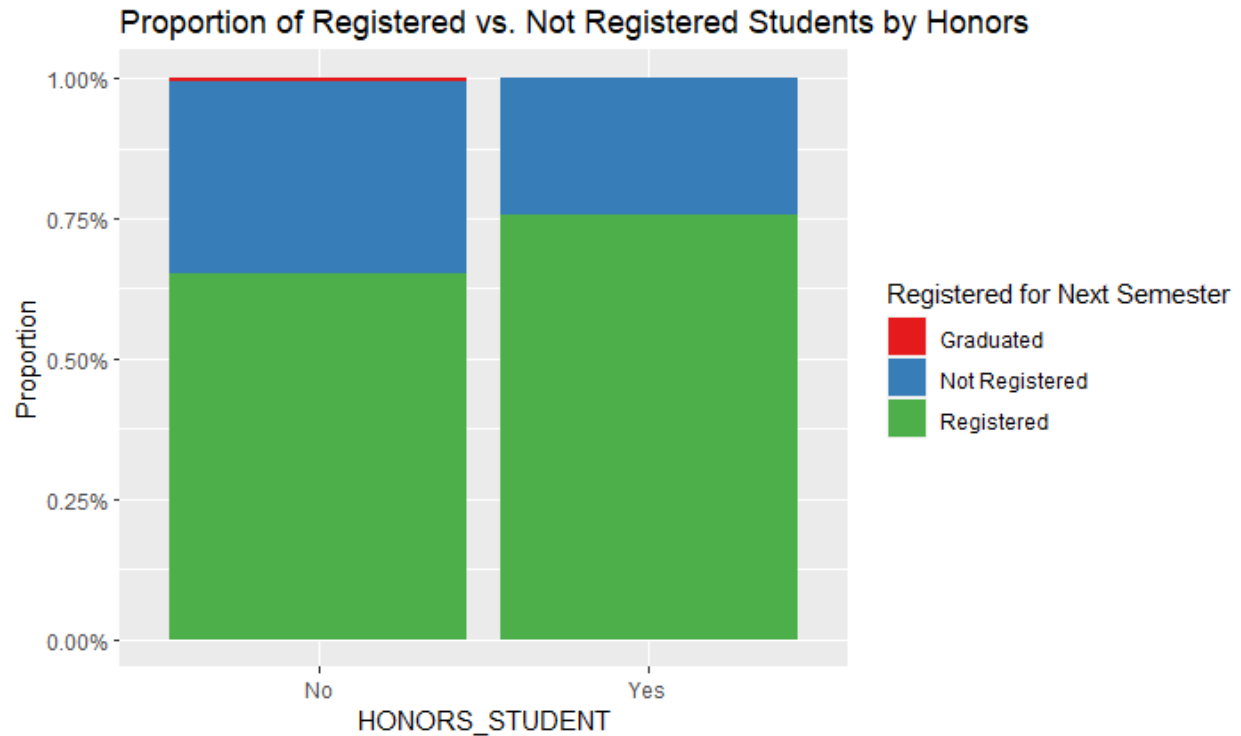
The proportion of registered vs not registered students has been ordered from least number registered to most registered with NA, or those who fall under another category, as the last bar. The greatest proportion of those not registered for the upcoming semester have an academic warning with the second largest proportion in the academic probation category. It will be important to also look into those whose status says semester warning since the number of registered students is only a bit above 50%. Above 70% of students with good academic standing are registered for the next semester. Those who have graduated were either in the semester warning or good standing categories. Further analysis should be conducted on the academic standing factor.



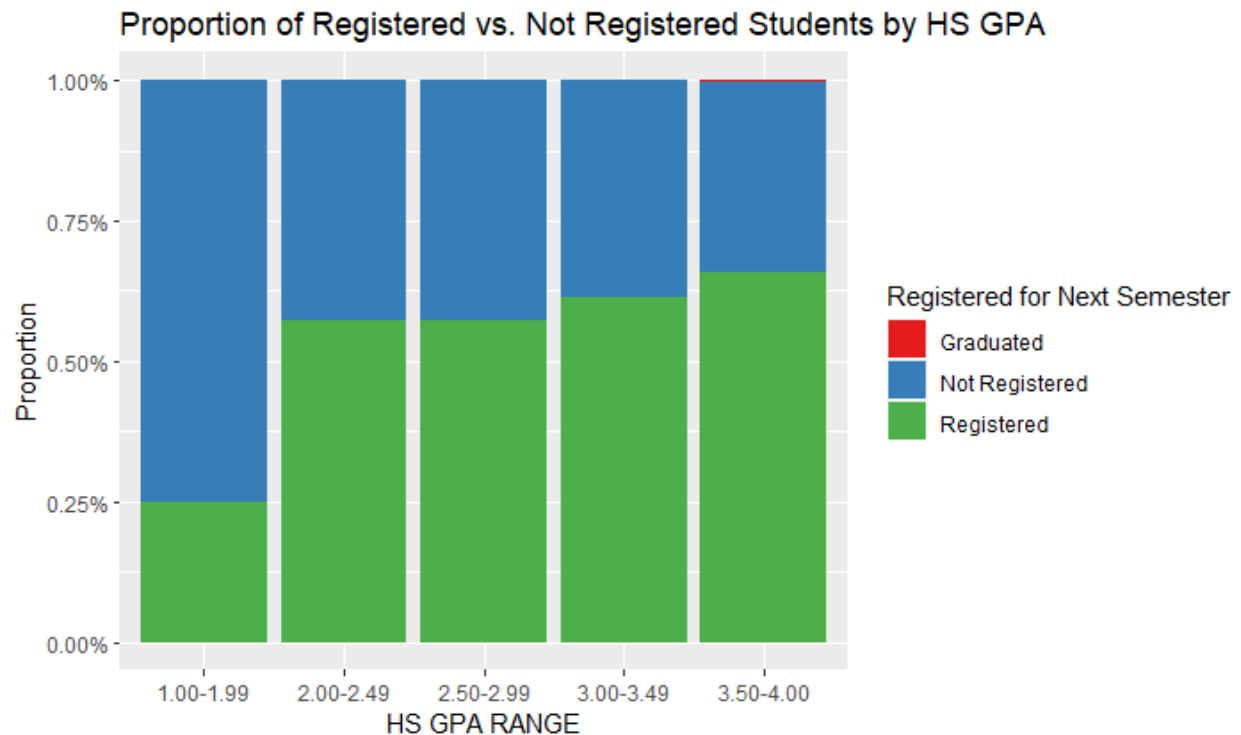
Around 68% of freshmen are registered for the next semester while over 75% of transfer students are registered. This should be explored further, but it makes sense that transfer students would be more motivated or able to continue their schooling since they took the initiative and effort to transfer while freshmen may still be deciding their future plans. All those who graduated fell under the transfer category, which makes sense since the only two categories in the dataset was freshman and transfer and for most students it takes more than one year of college to graduate. More analysis on the relationship between cohort type and registration status would be fascinating.



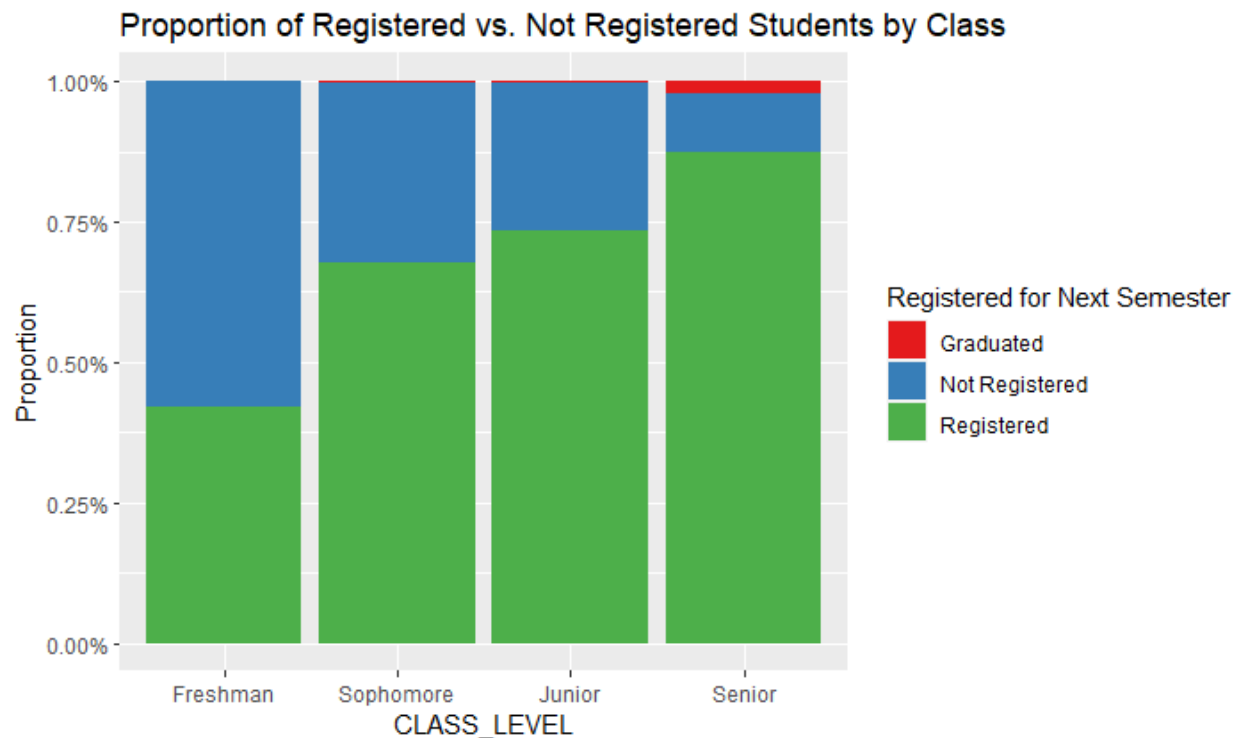
While there may not appear to be a large difference between students registered in both underrepresented and represented races, I think we need to dig deeper here. The counts for those in an underrepresented race are much less than those in a represented race, but that also means that differences we see may be highly significant. Here, the percentages are very close, and in fact underrepresented races have a slightly higher registration rate, but it is important to watch this factor over time to make sure USU supports all races, but especially those denoted by USHE guidelines as underrepresented, so that they have the support system they need here to succeed and stay.



Roughly 10% more honors students are registered for the next semester than those who are not honors students. This may indicate that honors students have more drive or resources to continue their college education than others, and this should be looked into. We also see from our dataset none of those who graduated were honors students.



As high school GPA range increases, the proportion of students registered for the next semester also increases. Those with very low high school GPAs, such as those in the 1.00-1.99 GPA range may indicate that they need help and support in their college experience here at USU.



This bar graph is particularly interesting because it shows that the proportion of those not registered for the next semester decreases as class in school increases. We need to focus on freshman especially for retention rates. It is also interesting to note that there are some students who graduated as sophomores and juniors. I would love to do a deeper analysis on this.

Next steps in the data analysis process would be to consider other factors not listed in the dataset, such as monetary support or family support. I would also love to see the data of these trends over time to see how student retention rates have changed. Code is included in a separate document in the email.