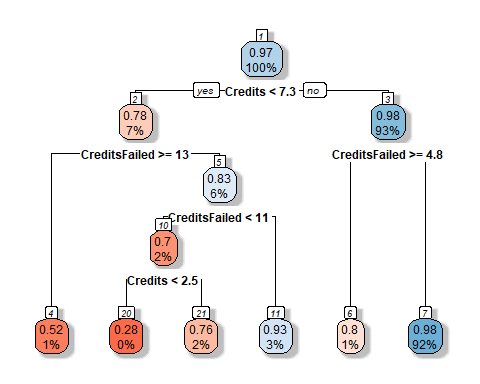
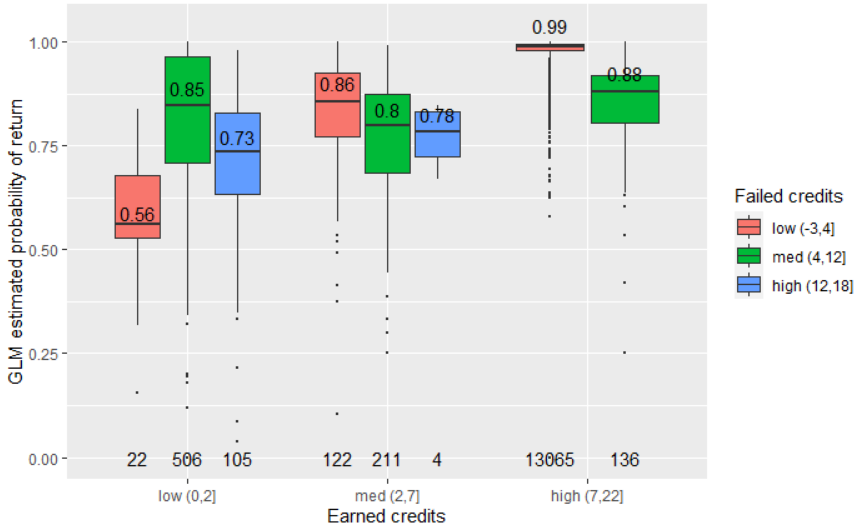
## Story 1: The most important factor for predicting a student’s retention is their performance



**Figure 1 Main splits of a decision tree classification model are based on credits earned and failed.**

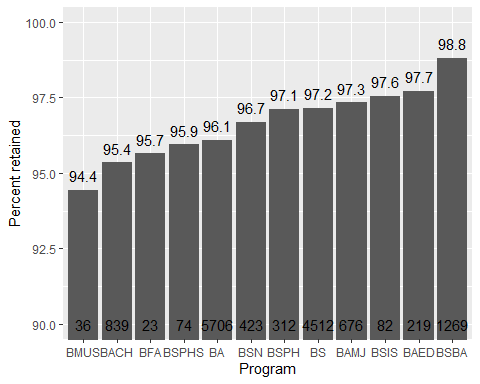
We fitted a decision tree model with retention rate being the dependent variable and class, residency, double major, enrollment, credits, credits failed, and program as the independent variables. Decision tree is a classification algorithm and uses gini impurity criteria to select splits, which will divide dataset into most homogenous sets. The fact that this tree splits mainly on credits and credits failed, indicating that these are the two of the most important variables for classification. In addition, we discovered the splits used for classification, which is Credits <2.5 and <7.3; splits for CreditsFailed are >=4.8, >=13. The highest retention group are students who earn >=7.3 credits and fail <4.8 credits (retention rate = 0.98, 92% of students are in this group). The lowest retention group (reddest) are students who earn <2.5 credits and fail 11~13 credits (retention rate = 0.28).



**Figure 2 The interaction effect between earned and failed credits on predicted retention rate.**

Since we identified a significant interaction effect between earned and failed credits on retention rate in both a GLM model and a decision tree model, now we will investigate further into this interaction effect. The categories for earned and failed credits were defined by the decision tree split shown above, in this way, we can split the data into the most homogeneous groups while enabling easier interpretation. When earn less than 2 credits, retention rate is low. Within this group, if failed credit is <=4, the median return rate is the lowest, around 56%, probably because so few credits were attempted in total. Interestingly, when failed 4~12 credits, the return rate increases to 85%, then when more credits are failed, the return rate drops again to around 73%. This implies a combination effect of how many credits are attempted in total versus how many are failed. When earns 2-7 credits, the return rate is similar regardless of failed credits, with a slight decreasing trend with increased failed credits. When earn more than 7 credits, there is no student fail more than 12 credits. The lower failed credit results in highest return rate ~99%, which is also the category with the most number of students. This trend is consistent with the real retention rate by earned and failed credits, providing support for the GLM model predictions.

## Story 2: How does program affect retention rate?

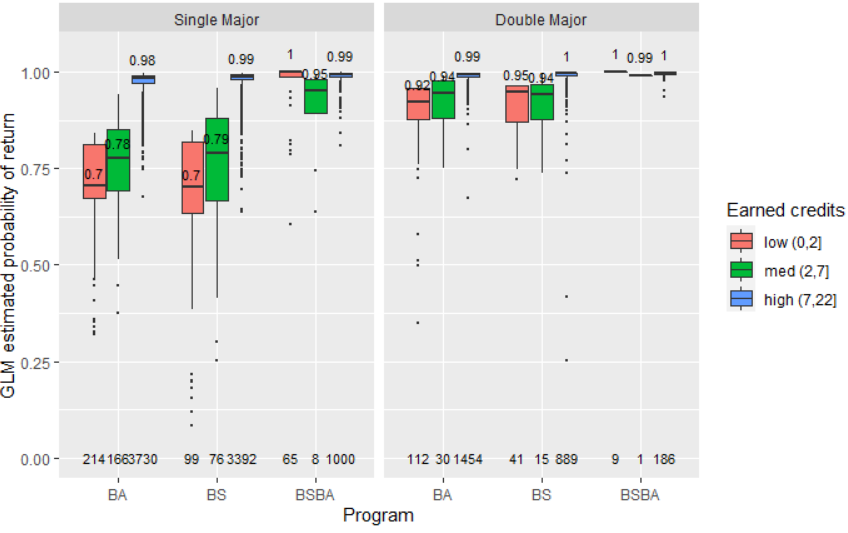


**Figure 3 Percent of students retained in each program.**

The programs with lowest retention rate is BMUS (Music Performance, 36 students), BACH (College of Arts and Sciences, 848 students), BFA (Studio Arts, 23 students), BSPHS (Pharmaceutical Sciences, 74 students), but all with smaller sample sizes. The programs with highest retention rate is BSBA (Business admin, 1270 students), BAED (School of education, 219 students), and BSIS (Information Science, 84 students).

Then within each program, how does earned credits and double major affects retention rate?

I picked 3 programs with the highest number of students as an example: science, art, and business, because the GLM model would more likely to give accurate prediction when the sample size is large.



**Figure 4 Earned credits interact with program and double major to affect predicted retention rate.**

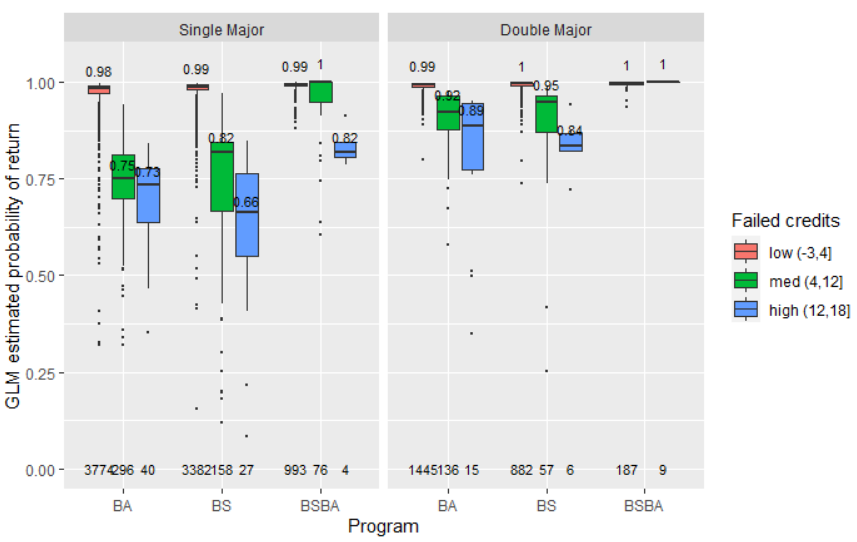
Next, we are interested in how would earned credits and double major affects students' return rate in different majors. We picked the three representative programs with the highest number of students: science, art, and business. Y axis indicates the probability of a student to return estimated by a full GLM model. Several interesting take-homes are:

1. There is a high return rate for business major, regardless of double major status and earned credits. This explains why there is such a high return rate (98.8%) in general.

2. For BA and BS, if a student earns high credits, there is a very high probability (98%) that they will return.

3. However, if the student earns low to median credits, double major students are more likely to return (~94%) than single major students (70~80%). In addition, single major students showed a gradient effect of increased credits will result in higher probability of return, whereas double major students remained a high probability of return for both low and median level of earned credits.

Thus, the single major students are more sensitive to how many credits they earned, except for students from certain programs such as business.



**Figure 5 Failed credits interact with program and double major to affect predicted retention rate, but differently from earned credits.**

Unlike earned credits, when we look at failed credits, it seems like all students are demotivated (have lower probability of return) by increased failed credits, as shown by the gradient decrease in probability of return as failed credits increase. Consistent with previous findings, double major students are more resilient to failed credits in two ways. One is that they have a generally higher probability of return at each failed credit level, which is 10~20% higher than single major students. Second is that they are less sensitive to increased failed credits (less steep slope) compared to single major students. In addition, business students are still the most resilient and have higher probability of return. However we now see that even single major business students are demotivated by a high level of failed credits (>12 credits failed).