

Exploring AUA's Student Performance

IESM315 Design Analysis and Experiment - Final Project

Instructor: Tadamasa Sawada

Anna Movsisyan

Ani Aloyan

Elina Ohanjanyan

Lusine Aghinyan

10 December, 2024

Contents

Abstract	3
Data Collection	4
A Glance at the Data	5
Literature Review	7
Factors Influencing College Persistence for First-Time Students	7
Collegeboard Analysis	8
Ethnic and gender stereotypes on college students' academic performance	9
Exploratory Data Analysis (EDA)	10
Conducting Analysis	18
Cautions and Results	36
Acknowledgements	37
References	38

Abstract

This report aims to analyze the American University of Armenia (AUA) student GPA statistics provided by AUA's Office of Institutional Research and Assessment (OIRA). The study identifies patterns in academic performance across metrics such as colleges, majors, genders, financial aid status, etc. This analysis highlights trends and potential factors that affect the final cumulative GPA of students.

Data Collection

The data for this study was obtained from AUA's OIRA. Key variables included:

- **Student ID:** Numeric code assigned to students.
- **FirstEnrolled_MajorCode:** Initial major of enrollment.
- **College:** College associated with each major.
- **FincialAid_Received_AtLeast_Once:** Binary variable indicating whether financial aid was received.
- **FirstEnrolled_Year:** Year of initial enrollment.
- **Gender:** Gender information of students.
- **School_GPA:** Standardized GPA (0-5 scale).
- **RoA (Republic of Armenia):** Binary variable indicating citizenship.
- **FirstYear_CGPA:** GPA for the first year.
- **All_CGPA:** Cumulative GPA.

Data privacy concerns led to the exclusion of rows with a count of five or fewer to ensure student anonymity.

A Glance at the Data

```
## # A tibble: 2,706 x 10
```

```
##   Studen~1 First~2 College Finci~3 First~4 Gender Schoo~5 RoA First~6 All_C~7
##   <dbl> <chr>   <chr>   <dbl>   <dbl> <chr>   <dbl> <dbl>   <dbl>   <dbl>
## 1  6.27e 9 BAB    CBE      0    2014 Male    3.81    1  0.189  0.189
## 2  4.18e 9 BAB    CBE      0    2018 Male    4.12    1  NA     NA
## 3  4.69e 9 BSCS   CSE      0    2013 Male    4      1  0.856  0.856
## 4  4.22e 9 BAB    CBE      0    2015 Male    3.56    1  2.25   2.75
## 5  1.50e10 BAB    CBE      0    2015 Male    3.9     1  2.3    1.52
## 6  8.18e 9 BSCS   CSE      0    2015 Male    4.00    1  NA     NA
## 7  6.52e 9 BSCS   CSE      0    2019 Male    4.18    1  2.15   NA
## 8  1.56e10 BAB    CBE      1    2016 Male    4.18    1  NA     NA
## 9  6.08e 9 BAB    CBE      1    2017 Male    4      1  3.87   NA
## 10 1.27e10 BAB    CBE      0    2020 Male    4.44    0  3.13   NA
## # ... with 2,696 more rows, and abbreviated variable names 1: `Student ID`,
## # 2: FirstEnrolled_MajorCode, 3: FincialAid_Received_AtLeast_Once,
## # 4: FirstEnrolled_Year, 5: School_GPA, 6: FirstYear_CGPA, 7: All_CGPA
```

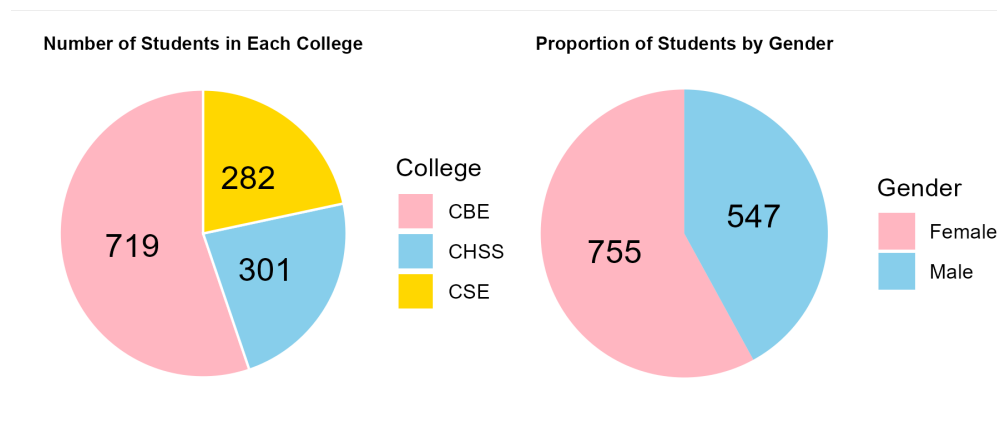


Figure 1: Distribution of AUA Students by College and Gender

The piecharts show the distribution of students by 1) College and 2) Gender.

As it can be seen from Fig. 1, the College of CSE has the most number of enrolled students, CHSS and CSE are relatively closer to each other in numbers and more than twice as less than CBE. Looking at the distribution by gender, it is seen that the number of female students exceeds the number of male students by ~200.

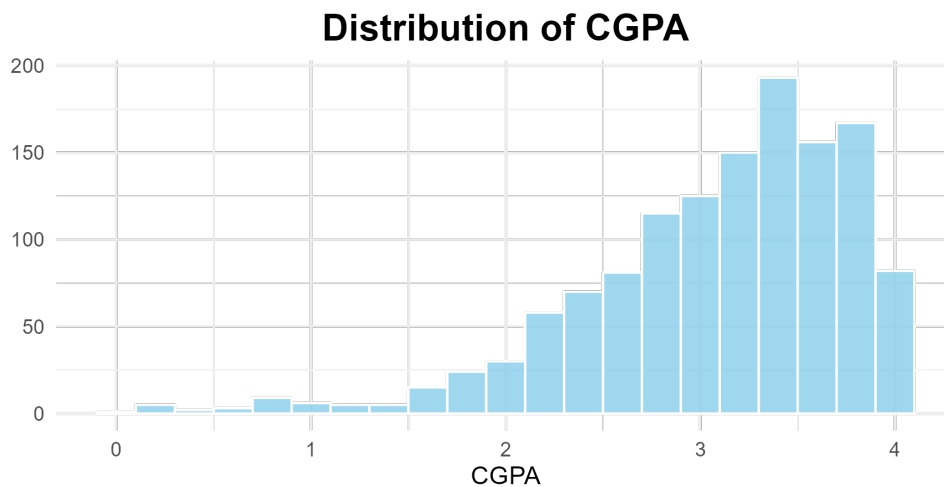


Figure 2: Histogram of CGPA

Now a glance at the distribution of All Cumulative GPA from Fig. 2. From the plot, it can be seen that the distribution is skewed and has a concentration around 3.3-3.7.

Literature Review

To better understand the effects of some of our factors on students' performances, we conducted a Literature Review and studied some articles that tested some similar factors.

Factors Influencing College Persistence for First-Time Students

The first article chosen to study the influence of the factors present in our dataset was article "Factors Influencing College Persistence for First-Time Students" by Sheilynda Stewart, Doo Hun Lim, and JoHyun Kim. It examines various factors that impact whether students persist in their college education (which, admittedly, is different from studying the influence of these factors on the students' GPA). Some of the key points we want to highlight are:

- **Need-Based Grants:** The authors highlight the positive role of financial aid, particularly need-based grants (which is similar to the Financial Aid granted to the students by AUA), in supporting college persistence. These grants reduce financial stress and help students from lower socioeconomic backgrounds stay enrolled, demonstrating their significant influence on retention rates.
- **Gender:** Findings on gender differences in persistence are mixed. Some studies cited in the article reveal that women persist and graduate at higher rates compared to men, while others show no significant gender differences in dropout behavior. The results of the study itself did not indicate that gender is significant when it comes to college persistence. This indicates that gender may interact with other factors, such as major or institutional characteristics, to influence outcomes.
- **School GPA:** According to this study, academic performance, as reflected in a student's High School and First Year College GPA, is a strong predictor of persistence. Students with higher GPAs are more likely to stay enrolled and complete their degrees. A strong academic

foundation during the first year appears critical in sustaining momentum toward graduation. However, it is important to take into consideration the fact that the high school GPAs of the students used in this study were self-reported, so the findings may have some inaccuracies.

Collegeboard Analysis

The second source chosen for understanding our factor is the was an analysis conducted by Collegeboard (the organization responsible for conducting the SAT tests) called “Using SAT Scores to Inform Academic Major-Related Decisions and Planning on Campus” by Paul A. Westrick, Jessica P. Marini, and Emily J. Shaw. While the overall goal of the study was not the analysis of college performance, it did include some interesting findings that may be useful to us in the future.

The most interesting finding for our study was the following plot:

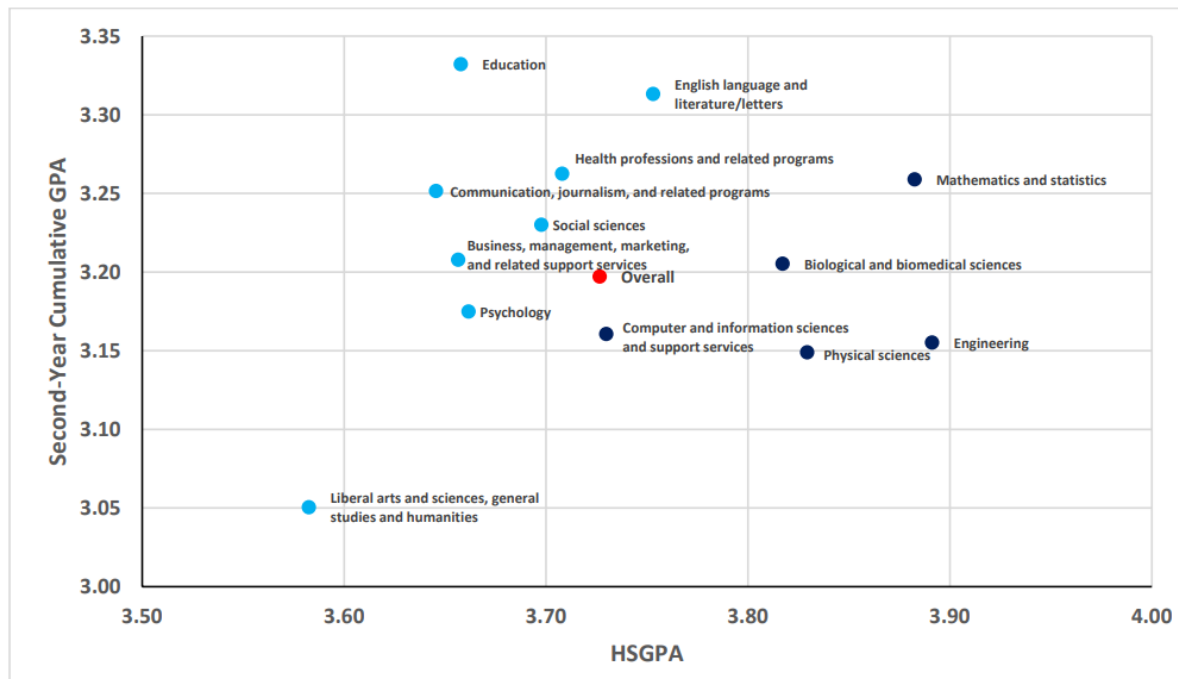


Figure 3: Relationship Between High School and College GPA by Major

As can be seen in Fig. 3, the reason why we wanted to focus on this study is because it shows some interesting trends about how the major of the students can correlate with their High School and College GPA. Some of the interesting points that can be derived from this plot are: - **Science Majors:**

One of the more interesting things on this plot is the fact that while Science majors (Math, Engineering, etc) have the highest GPAs in their school years, their College GPAs seem to be somewhat lower than that of the other majors in college. - **Business Majors:** Another interesting statistic is that Business majors are around the middle in both high school and college GPA. - **Communications Majors:** The last finding that can be somewhat related to our data is the fact Communications majors (which can correspond to the students in AUA's English and Communications program), despite having a relatively lower High School GPA, have one of the highest GPAs in College.

However, it is important to note that the study was conducted on the data of US students and might not reflect some trends present in the Armenian education system.

Ethnic and gender stereotypes on college students' academic performance

The last article we will discuss is "Ethnic and gender stereotypes on college students' academic performance" by Ya-Wen (Melissa) Liang, Don Jones and Rebecca A. Robles-Pina. The reason why this article can be somewhat relevant is because it was conducted at a small university in Texas. While this does not directly correlate to Armenia, it is still important to note that Texas, being one of the more conservative states in the U.S., can show somewhat similar societal factors to Armenia, a very conservative country.

The main finding of this study was that gender had a significant effect on the students' GPA, and female students had a higher GPA than males.

Exploratory Data Analysis (EDA)

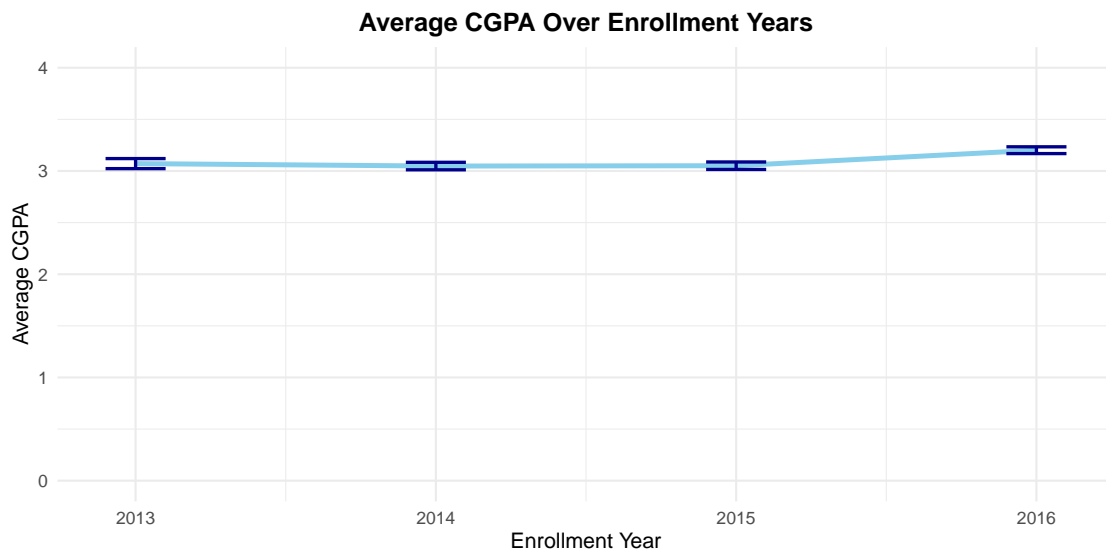


Figure 4: Average CGPA by Cohort

Fig. 4 shows the average CGPA (with error bar by standard error) over the enrollment years. As it can be observed, there is almost no difference between the cohorts of different years and their average is slightly higher than 3.0.

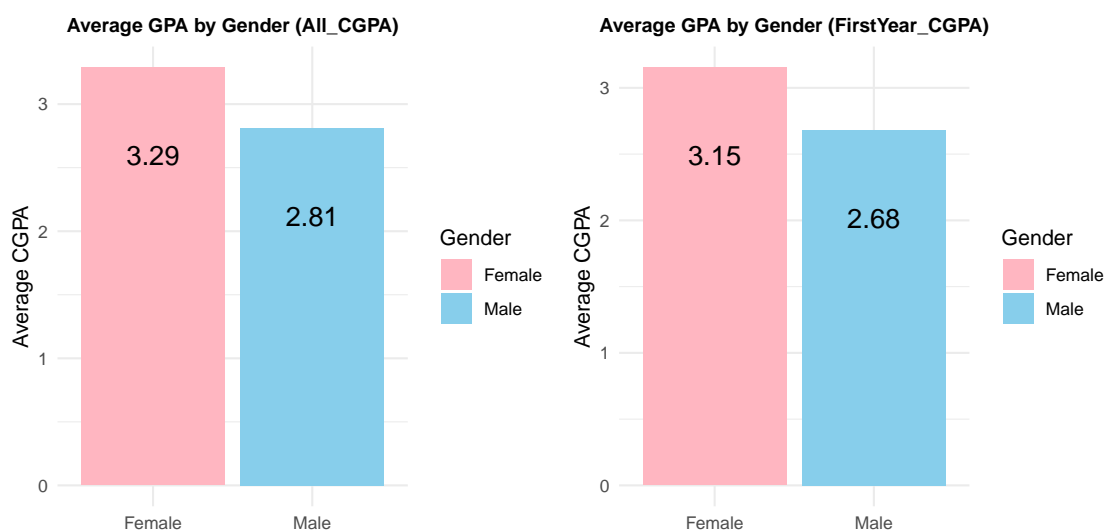


Figure 5: GPA Distribution by Gender

Fig. 5 compares the average CGPA of students by gender, displayed separately for two categories:

All_CGPA and FirstYear_CGPA. Female students demonstrate higher average GPAs in both categories, with an average of 3.29 for All_CGPA and 3.15 for FirstYear_CGPA. In contrast, male students achieve averages of 2.81 for All_CGPA and 2.68 for FirstYear_CGPA.

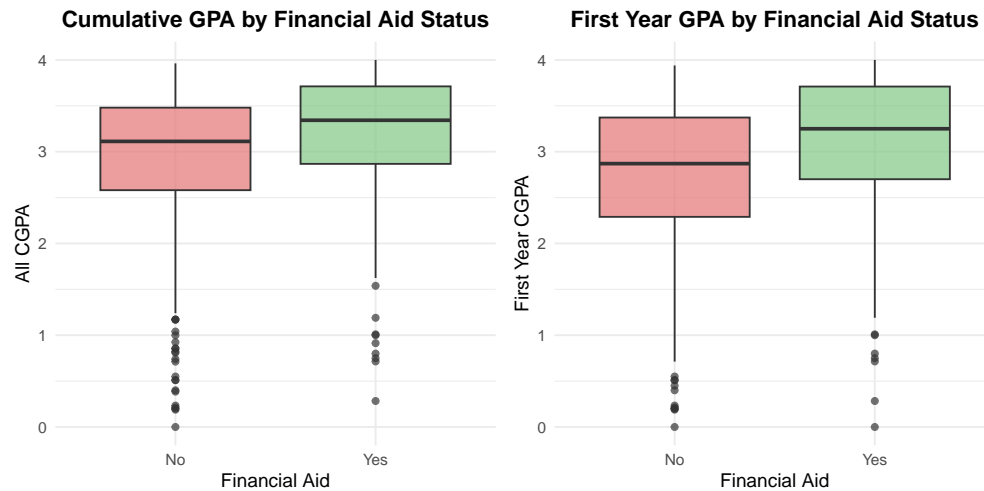


Figure 6: GPA Distribution by Financial Aid Status

This figure displays the distribution of 1st Year/All Cumulative GPAs based on whether students received financial aid. The boxplots of both categories indicates that the median GPA is slightly higher for students who received financial aid (green) compared to those who did not (red). The interquartile range (IQR) is similar for both groups, but there are more outliers in the group without financial aid, particularly on the lower end of the GPA scale. This suggests that financial aid may provide a stabilizing effect on academic performance, reducing extreme variations and potentially supporting better outcomes.

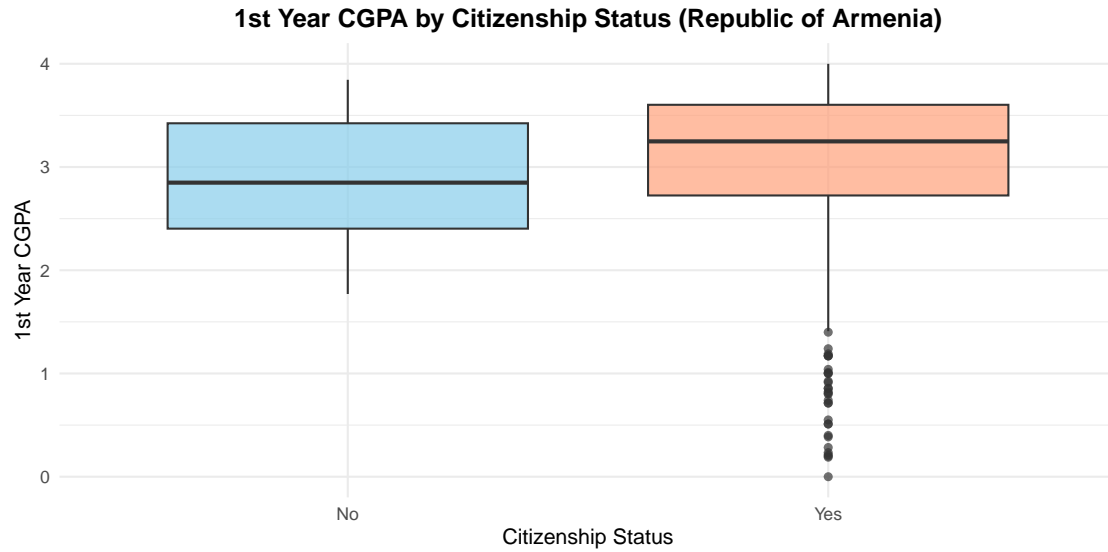


Figure 7: First Year GPA Distribution by RoA Citizenship

This figure shows boxplots comparing the All CGPA distribution of students categorized by citizenship status, where “Yes” indicates students with Armenian citizenship, and “No” represents non-Armenian citizens. Students with Armenian citizenship exhibit slightly higher median GPAs compared to non-Armenian students. The interquartile range (IQR) for Armenian citizens is broader, while non-Armenian students have a more compact distribution with fewer outliers.

However, this comparison is inadmissible due to the stark imbalance in sample sizes: only 26 non-Armenian citizens are included, compared to 1,273 Armenian citizens. Such a significant disparity makes the comparison unreliable. It would be more appropriate to exclude non-Armenian citizens from future analyses to avoid skewed interpretations.

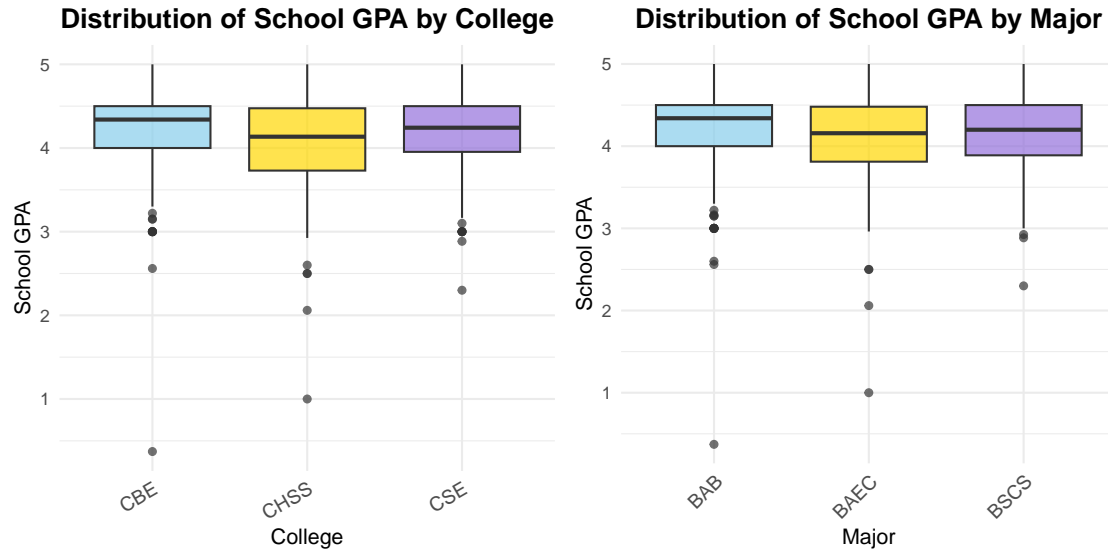


Figure 8: School GPA Distribution by College and Major

This figure displays the distribution of School_GPA across colleges (left) and majors (right). The data is restricted to students enrolled between 2013 and 2016, as these cohorts had completed their full cumulative GPA. During this period, there were only three bachelor programs available: BAB under College CBE, BAEC under CHSS, and BSCS under CSE.

The boxplots illustrate that the distributions are consistent across colleges and their corresponding majors, with comparable medians and interquartile ranges. This indicates limited variability within colleges but highlights the focus of the dataset on a specific time frame and program set. The analysis ensures that comparisons remain valid within the historical context of program availability.

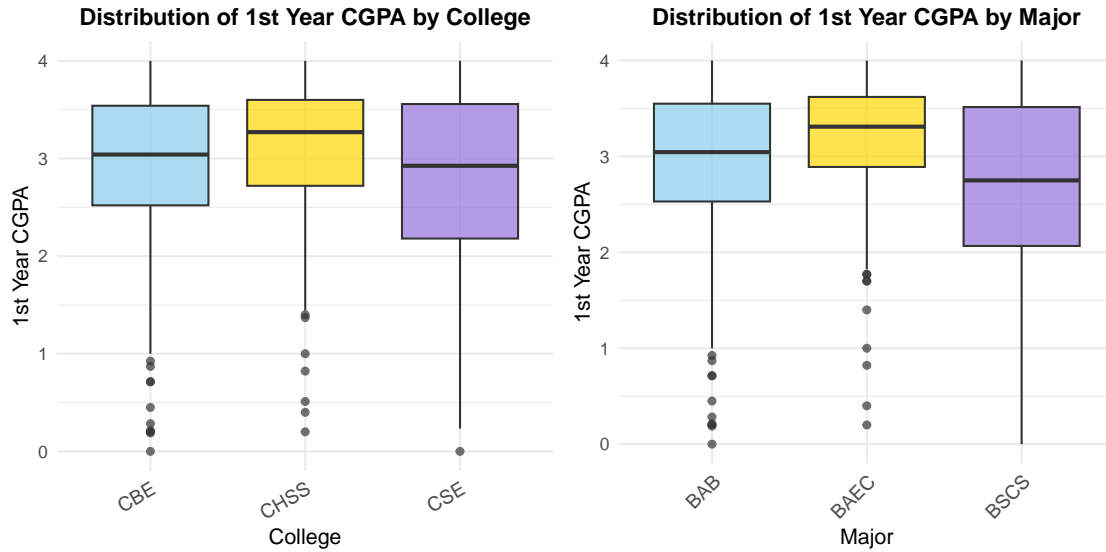


Figure 9: First Year GPA Distribution by College and Major

A similar figure compares the distribution of 1st Year CGPA across colleges (left) and majors (right) for students. The data is restricted to students from the 2013–2016 cohorts, ensuring that the analysis is aligned with available cumulative GPA data. During this period, three main undergraduate programs were available: BAB under CBE, BAEC under CHSS, and BSCS under CSE.

The distributions across colleges and majors show similar patterns, with median GPAs being close across all categories. Notably, the College of Humanities and Social Sciences (CHSS) and its corresponding program (BAEC) demonstrate slightly higher median GPAs compared to other colleges and majors. Outliers are present in all categories, particularly in CSE and BSCS, indicating a broader variation in student performance for these groups. This comparison highlights the relative consistency in GPA performance across colleges and majors while acknowledging differences in variability.

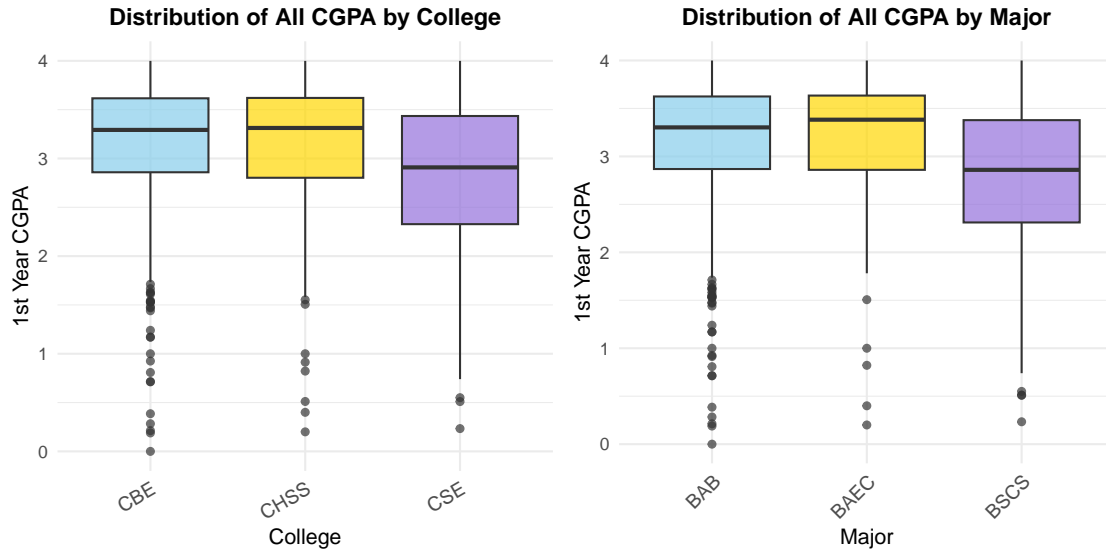


Figure 10: CGPA Distribution by College and Major

This final figure presents the distribution of All CGPA for students across colleges (left) and majors (right). The distributions demonstrate similar patterns to the 1st Year CGPA, with slightly higher median values for CHSS and BAEC compared to the other colleges and majors.

It is important to note that 1st Year CGPA contributes directly to the calculation of All CGPA, making this similarity in distribution an expected behavior. Outliers are present in all categories, reflecting variations in student performance. This consistency reinforces the role of early academic performance in shaping cumulative outcomes over the course of a student's studies.

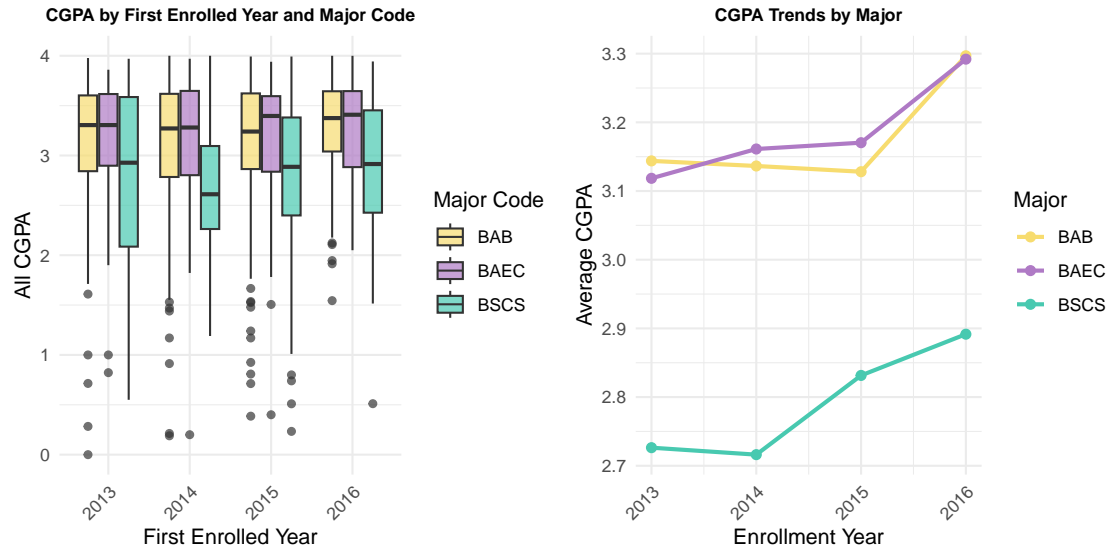


Figure 11: GPA by Enrollment Year and Major

The left panel shows the distribution of Cumulative GPAs (All CGPA) for students grouped by their first enrolled year (FirstEnrolled_Year), with each boxplot colored by their major code (FirstEnrolled_MajorCode). The right panel displays the average cumulative GPA trends for each major across enrollment years.

The boxplots highlight that BAEC majors generally have higher median GPAs, while BSCS majors show greater variability and more outliers. The trend lines reveal that BAEC demonstrates a steady increase in average GPA, peaking in 2016, whereas BAB remains stable across the years. BSCS, although initially lower, shows a consistent upward trend, suggesting improvement over time.

The takeaway is the difference in distributions by Colleges and Enrollment Year. The latter, however, is relatively niche.

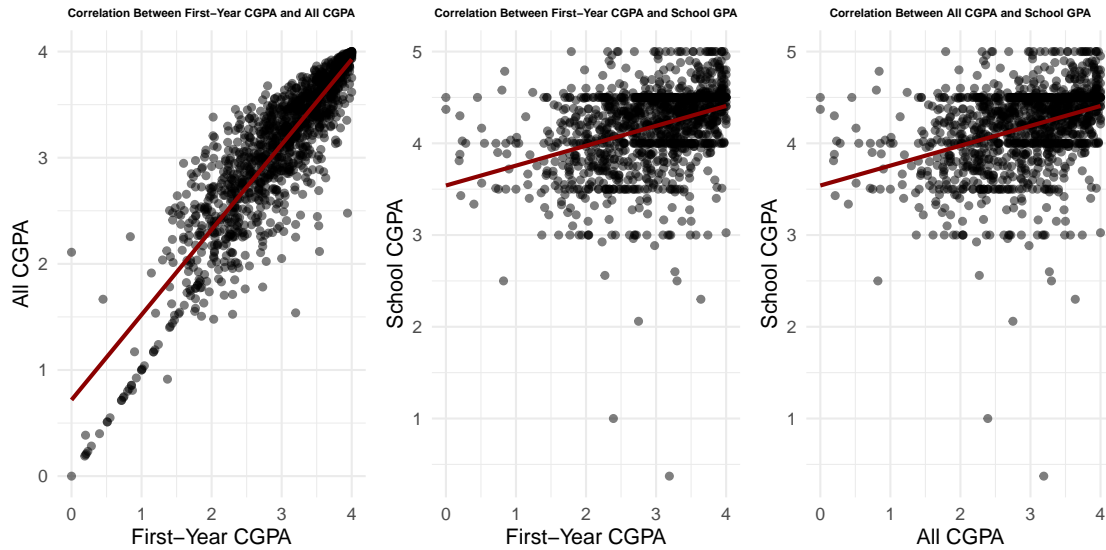


Figure 12: Correlations of GPAs

The figure displays the correlations between different GPA metrics. The left panel shows a strong positive correlation between 1st Year CGPA and All CGPA, which is expected since 1st Year CGPA contributes directly to the cumulative calculation. The middle panel shows a weaker but noticeable positive correlation between 1st Year CGPA and School GPA, highlighting a small association. This is important to consider if both metrics are included in the analysis, as it may introduce collinearity. The right panel illustrates a moderate positive correlation between All CGPA and School GPA, reflecting their cumulative relationship.

Visually, trends are evident, but further numerical analysis is needed to determine whether the weaker correlation (e.g., between 1st Year CGPA and School GPA) justifies including both factors in the final analysis. This step ensures that correlated factors are appropriately managed.

Conducting Analysis

```
# Correlation between 1st Year GPA and School GPA

cor_test <- cor.test(gpa_data$FirstYear_CGPA, gpa_data$School_GPA,
                     method = "pearson")

print(cor_test)

##

## Pearson's product-moment correlation

##

## data:  gpa_data$FirstYear_CGPA and gpa_data$School_GPA
## t = 14.236, df = 1292, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3201474 0.4143975
## sample estimates:
##      cor
## 0.368218

# Correlation between 1st Year GPA and Overall GPA

cor_test_uni_gpa <- cor.test(gpa_data$FirstYear_CGPA, gpa_data$All_CGPA,
                             method = "pearson")

print(cor_test_uni_gpa)

##

## Pearson's product-moment correlation

##
```

```
## data:  gpa_data$FirstYear_CGPA and gpa_data$All_CGPA
## t = 68.054, df = 1300, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8711371 0.8950061
## sample estimates:
##          cor
## 0.8836445
```

The results reveal a strong positive correlation (0.884) between 1st Year GPA and Overall GPA, which is expected as the former contributes directly to the calculation of the latter. On the other hand, the correlation between 1st Year GPA and School GPA is weaker (0.368) but still statistically significant.

We previously discussed the visual relationship between 1st Year GPA and School GPA in the scatterplots, where a weak correlation was observed. Although the correlation coefficient is below 0.5, indicating a modest association, we decided to retain both metrics in the analysis. This decision was made because the correlation is not strong enough to raise serious concerns about collinearity, but the potential unique contributions of both factors justify their inclusion for a comprehensive analysis.

```
# Understanding the impact of College on All CGPA with ANOVA
oneway_anova <- aov(All_CGPA ~ College, data = gpa_data)
summary(oneway_anova)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## College         2    22.4   11.209    24.42 3.87e-11 ***
## Residuals    1299   596.2    0.459
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA test confirms that College significantly impacts All CGPA, as evidenced by the highly significant p-value. This aligns with our earlier visual observations, where the distribution of GPAs differed across colleges.

This analysis further reinforces the role of College as a meaningful factor influencing academic performance. Having established its importance, we will later proceed to examine the effect of College in more detail to deepen our understanding of its influence on student outcomes.

```
# Conduct Tukey's test for College
tukey_test <- TukeyHSD(oneway_anova)
print(tukey_test)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = All_CGPA ~ College, data = gpa_data)
##
## $College
##
```

	diff	lwr	upr	p adj
CHSS-CBE	-0.03637804	-0.1455042	0.07274816	0.7140568
CSE-CBE	-0.32728750	-0.4389750	-0.21559998	0.0000000
CSE-CHSS	-0.29090946	-0.4226450	-0.15917388	0.0000008

The Tukey test reveals clear differences in performance between colleges. While there is no significant difference in GPA between CHSS and CBE, the results indicate that students from CSE perform significantly worse compared to both CHSS and CBE.

This finding aligns with earlier observations of CSE having lower median GPAs and greater variability. It highlights the challenges faced by students in the Computer Science department, where

performance consistently lags behind the other colleges. This analysis reinforces the need for targeted support or further investigation into the factors contributing to this disparity.

```
# T-test for Financial Aid impact on All CGPA  
t_test_finaid <- t.test(All_CGPA ~ FincialAid_Received_AtLeast_Once, data = gpa_data)  
print(t_test_finaid)
```

```
##  
## Welch Two Sample t-test  
##  
## data: All_CGPA by FincialAid_Received_AtLeast_Once  
## t = -7.0574, df = 1187.1, p-value = 2.878e-12  
## alternative hypothesis: true difference in means between group 0 and group 1 is not e  
## 95 percent confidence interval:  
## -0.3430348 -0.1937956  
## sample estimates:  
## mean in group 0 mean in group 1  
## 2.946075 3.214490
```

The t-test results indicate a highly significant difference in cumulative GPA (All CGPA) between students who received financial aid and those who did not. Specifically, students receiving financial aid have a significantly higher mean CGPA (3.214) compared to those who did not (2.946).

The negative confidence interval and t-value further confirm that the GPA for students without financial aid is consistently lower. This finding highlights the potential positive impact of financial aid on student performance, making it a crucial factor in academic success. This result aligns with previous observations from visualizations and reinforces the importance of considering financial aid status in the analysis.

```

# T-test for gender differences in CGPA
t_test_gender <- t.test(All_CGPA ~ Gender, data = gpa_data)
print(t_test_gender)

##
##  Welch Two Sample t-test
##
## data:  All_CGPA by Gender
## t = 12.241, df = 889.26, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group Female and group Male
## 95 percent confidence interval:
##  0.3980936 0.5501203
## sample estimates:
## mean in group Female    mean in group Male
##           3.288330           2.814223

```

The t-test reveals a significant difference in CGPA based on gender ($p < 2.2e-16$). Female students have a higher mean CGPA (3.288) than male students (2.814). This significant gender disparity highlights the potential influence of gender-specific factors on academic performance, warranting further investigation.

```

gpa_data <- gpa_data %>%
  mutate(FirstEnrolled_Year = as.factor(FirstEnrolled_Year))

# One-way ANOVA based on Enrollment Year
anova_year <- aov(All_CGPA ~ FirstEnrolled_Year, data = gpa_data)
summary(anova_year)

```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## FirstEnrolled_Year    3    5.1   1.7030   3.603  0.013 *
## Residuals          1298  613.5   0.4726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA test investigates the influence of enrollment year on cumulative GPA. The p-value (0.013) indicates a statistically significant effect of the enrollment year on cumulative GPA. This suggests slight variations in academic performance based on the year students enrolled. However, the mean differences across years seem relatively small, as seen in the exploratory data analysis.

```
# Tukey test to find differences between years
tukey_posthoc <- TukeyHSD(anova_year)
print(tukey_posthoc)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = All_CGPA ~ FirstEnrolled_Year, data = gpa_data)
##
## $FirstEnrolled_Year
##              diff              lwr              upr              p adj
## 2014-2013 -0.023791341 -0.17004899  0.1224663  0.9753353
## 2015-2013 -0.020520559 -0.16526487  0.1242237  0.9834147
## 2016-2013  0.129845549 -0.02221868  0.2819098  0.1247007
## 2015-2014  0.003270782 -0.12557998  0.1321215  0.9999001
## 2016-2014  0.153636890  0.01661451  0.2906593  0.0207766
## 2016-2015  0.150366108  0.01496024  0.2857720  0.0225587
```

The Tukey post-hoc test further evaluates pairwise differences between enrollment years. Significant difference is observed between 2016 and earlier years, with the largest positive difference in GPA for the 2016 cohort (+0.15 compared to 2013) and differences between other pairs (e.g., 2014-2013, 2015-2014) are not statistically significant as it could be seen from the graph previously.

```
# T-test for Financial Aid (0 = No, 1 = Yes)
t_test_finaid <- t.test(All_CGPA ~ FincialAid_Received_AtLeast_Once, data = gpa_data)
print(t_test_finaid)

##
## Welch Two Sample t-test
##
## data: All_CGPA by FincialAid_Received_AtLeast_Once
## t = -7.0574, df = 1187.1, p-value = 2.878e-12
## alternative hypothesis: true difference in means between group 0 and group 1 is not e
## 95 percent confidence interval:
## -0.3430348 -0.1937956
## sample estimates:
## mean in group 0 mean in group 1
## 2.946075 3.214490
```

The Welch Two-Sample T-Test evaluates the relationship between receiving financial aid and cumulative GPA (CGPA). The results reveal a significant difference in CGPA between students who received financial aid and those who did not, with a p-value of less than 2.878e-12. The mean CGPA for students who received financial aid was 3.214, while for those who did not, it was 2.946. This finding indicates that financial aid might contribute positively to academic performance, potentially by alleviating financial pressures or providing access to necessary resources. However, additional research is needed to establish the exact mechanisms driving this relationship.


```

gpa_data_clean <- gpa_data %>%
  filter(!is.na(FirstYear_CGPA), !is.na(College), !is.na(FirstEnrolled_MajorCode))

anova_major_gpa_college <- aov(All_CGPA ~ FirstEnrolled_MajorCode * FirstYear_CGPA *
  College, data = gpa_data_clean)
summary(anova_major_gpa_college)

```

```

##                                     Df Sum Sq Mean Sq F value
## FirstEnrolled_MajorCode            2   33.0    16.5  171.737
## FirstYear_CGPA                     1  457.8   457.8 4763.936
## College                            2    2.5     1.2   12.759
## FirstEnrolled_MajorCode:FirstYear_CGPA  2    0.6     0.3    2.993
## FirstEnrolled_MajorCode:College        3    0.4     0.1    1.394
## FirstYear_CGPA:College                2    0.5     0.2    2.467
## FirstEnrolled_MajorCode:FirstYear_CGPA:College  3    0.2     0.1    0.779
## Residuals                          1286  123.6     0.1
##                                     Pr(>F)
## FirstEnrolled_MajorCode             < 2e-16 ***
## FirstYear_CGPA                     < 2e-16 ***
## College                            3.26e-06 ***
## FirstEnrolled_MajorCode:FirstYear_CGPA  0.0505 .
## FirstEnrolled_MajorCode:College        0.2429
## FirstYear_CGPA:College                0.0853 .
## FirstEnrolled_MajorCode:FirstYear_CGPA:College  0.5057
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The model evaluates how major codes, first-year CGPA, and colleges collectively influence the cumulative GPA (All_CGPA). The model reveals statistically significant effects for the variables FirstEnrolled_MajorCode ($p < 2e-16$) and FirstYear_CGPA ($p < 2e-16$), indicating that both the choice of major and initial academic performance strongly correlate with final cumulative GPA. However, the interaction effects between FirstEnrolled_MajorCode, FirstYear_CGPA, and College were not statistically significant, suggesting that their combined influence does not add further explanatory power beyond their individual contributions. This indicates that future analyses could focus on the direct effects of these variables rather than their interactions.

Switching

Switching refers to students transitioning from their initially enrolled major to a different program. This factor may influence cumulative GPA, as initial struggles in the original major could impact academic performance. Although the number of such students is relatively small, their impact on GPA trends is worth exploring.

```
## # A tibble: 6 x 3
##   College Switched Count
##   <chr>         <dbl> <int>
## 1 CBE           0     697
## 2 CBE           1      22
## 3 CHSS          0     275
## 4 CHSS          1      26
## 5 CSE           0     273
## 6 CSE           1       9
```

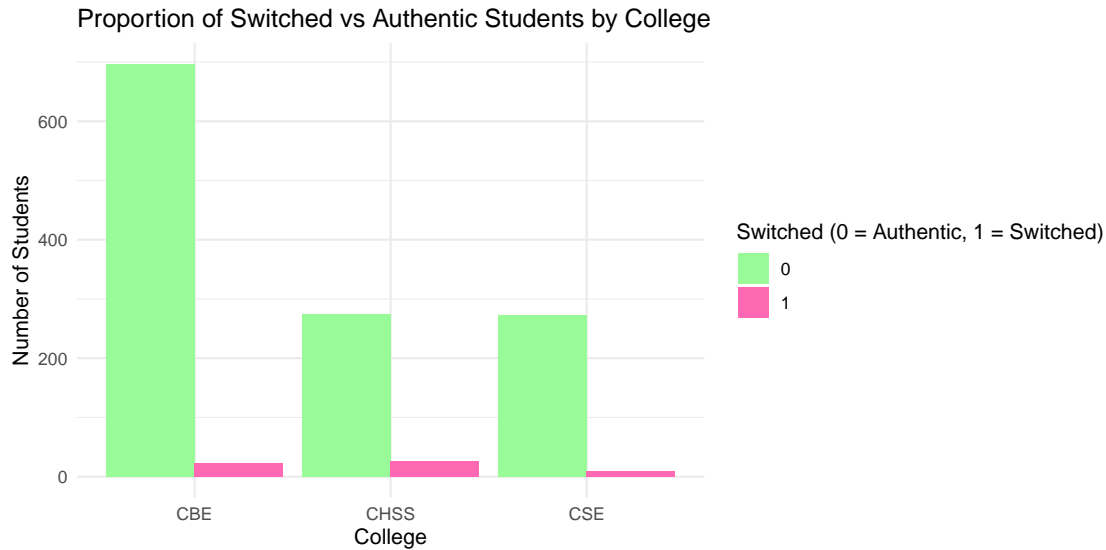


Figure 13: Switched Students by College

```
# T-test for CSE
t_test_cse <- t.test(All_CGPA ~ Switched, data = combined_data %>%
  filter(College == "CSE"))
print(t_test_cse)
```

```
##
## Welch Two Sample t-test
##
## data: All_CGPA by Switched
## t = -3.2026, df = 9.2295, p-value = 0.01045
## alternative hypothesis: true difference in means between group 0 and group 1 is not e
## 95 percent confidence interval:
## -0.9849688 -0.1713170
## sample estimates:
## mean in group 0 mean in group 1
## 2.822705 3.400848
```

For CSE students, a Welch Two Sample t-test revealed that those who switched programs had significantly higher GPAs compared to those who stayed, indicating that switching might have helped alleviate challenges.

```
# T-test for CBE
```

```
t_test_cbe <- t.test(All_CGPA ~ Switched, data = combined_data %>%  
  filter(College == "CBE"))  
print(t_test_cbe)
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: All_CGPA by Switched
```

```
## t = 4.2961, df = 23.624, p-value = 0.0002561
```

```
## alternative hypothesis: true difference in means between group 0 and group 1 is not e
```

```
## 95 percent confidence interval:
```

```
## 0.2256250 0.6435214
```

```
## sample estimates:
```

```
## mean in group 0 mean in group 1
```

```
## 3.181741 2.747168
```

For CBE students, the t-test also showed significantly higher GPAs for those who switched, suggesting that transitioning to another program was beneficial for their academic performance.

```
# T-test for CHSS
```

```
t_test_chss <- t.test(All_CGPA ~ Switched, data = combined_data %>%  
  filter(College == "CHSS"))  
print(t_test_chss)
```

```
##
## Welch Two Sample t-test
##
## data: All_CGPA by Switched
## t = 4.1908, df = 27.57, p-value = 0.0002578
## alternative hypothesis: true difference in means between group 0 and group 1 is not e
## 95 percent confidence interval:
## 0.3602288 1.0500357
## sample estimates:
## mean in group 0 mean in group 1
## 3.192974 2.487842
```

Similarly, CHSS students who switched programs had significantly higher GPAs than those who remained, further supporting the idea that switching can positively impact academic outcomes.

```
anova_all_star_switched <- Anova(lm(All_CGPA ~ FirstYear_CGPA * Switched * College,
  data = combined_data))
print(anova_all_star_switched)
```

```
## Anova Table (Type II tests)
##
## Response: All_CGPA
```

	Sum Sq	Df	F value	Pr(>F)
## FirstYear_CGPA	451.58	1	4701.3154	< 2.2e-16 ***
## Switched	0.60	1	6.2782	0.012345 *
## College	8.94	2	46.5579	< 2.2e-16 ***
## FirstYear_CGPA:Switched	0.08	1	0.8290	0.362721
## FirstYear_CGPA:College	1.06	2	5.5414	0.004015 **

```
## Switched:College                0.82    2    4.2720  0.014152 *
## FirstYear_CGPA:Switched:College  0.83    2    4.3153  0.013555 *
## Residuals                      123.91 1290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the initial ANOVA test, the analysis evaluates the influence of first-year CGPA, switching status, and college on cumulative GPA, including the interactions between these factors. First-year CGPA shows a highly significant effect on cumulative GPA, which is expected since it directly contributes to the calculation of the cumulative GPA. The switching status variable displays marginal significance, suggesting that changing majors might have a potential, albeit weak, influence on cumulative GPA. College remains strongly significant, highlighting notable differences in cumulative GPA based on the college. The interaction between first-year CGPA and college is significant, indicating that the impact of first-year performance on cumulative GPA differs across colleges. However, the interaction between first-year CGPA and switching is not significant, implying that switching majors does not notably modify the relationship between first-year GPA and cumulative GPA.

```
anova_all_star_switched <- Anova(lm(All_CGPA ~ FirstYear_CGPA * Switched * College *
  FirstEnrolled_MajorCode, data = combined_data))
print(anova_all_star_switched)
```

```
## Anova Table (Type II tests)
```

```
##
```

```
## Response: All_CGPA
```

##	Sum Sq	Df	F value
## FirstYear_CGPA	451.38	1	4696.6706
## Switched		0	

## College	0.15	2	0.7683
## FirstEnrolled_MajorCode	0.17	2	0.8809
## FirstYear_CGPA:Switched		0	
## FirstYear_CGPA:College	0.55	2	2.8402
## Switched:College		0	
## FirstYear_CGPA:FirstEnrolled_MajorCode	0.15	2	0.7757
## Switched:FirstEnrolled_MajorCode		0	
## College:FirstEnrolled_MajorCode		0	
## FirstYear_CGPA:Switched:College		0	
## FirstYear_CGPA:Switched:FirstEnrolled_MajorCode		0	
## FirstYear_CGPA:College:FirstEnrolled_MajorCode		0	
## Switched:College:FirstEnrolled_MajorCode		0	
## FirstYear_CGPA:Switched:College:FirstEnrolled_MajorCode		0	
## Residuals	123.59	1286	
##	Pr(>F)		
## FirstYear_CGPA	< 2e-16	***	
## Switched			
## College	0.46402		
## FirstEnrolled_MajorCode	0.41467		
## FirstYear_CGPA:Switched			
## FirstYear_CGPA:College	0.05878	.	
## Switched:College			
## FirstYear_CGPA:FirstEnrolled_MajorCode	0.46058		
## Switched:FirstEnrolled_MajorCode			
## College:FirstEnrolled_MajorCode			
## FirstYear_CGPA:Switched:College			
## FirstYear_CGPA:Switched:FirstEnrolled_MajorCode			

```
## FirstYear_CGPA:College:FirstEnrolled_MajorCode
## Switched:College:FirstEnrolled_MajorCode
## FirstYear_CGPA:Switched:College:FirstEnrolled_MajorCode
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the extended ANOVA test, the effect of the first enrolled major is included, along with its interactions with the other factors. The addition of this variable helps refine the analysis. While the first-year CGPA remains significant, switching status loses its marginal significance when additional variables are considered. College continues to demonstrate a strong significance, reaffirming its impact on cumulative GPA. However, the interaction between switching status, college, and first-year CGPA, which was marginally significant in the previous test, becomes non-significant after the inclusion of the enrolled major. This indicates that the major accounts for variability that switching and college interactions previously explained. Other significant effects persist, but the diminishing significance of the interactions highlights how adding new variables can provide clarity on previously observed effects.

Final Model

```
lm_all_star_switched <- Anova(lm(All_CGPA ~ FirstYear_CGPA * College * FirstEnrolled_MajorCode *
    Switched + Gender + School_GPA + FincialAid_Received_AtLeast_Once, data = combined_data))
print(lm_all_star_switched)
```

```
## Anova Table (Type II tests)
```

```
##
```

```
## Response: All_CGPA
```

```
##
```

	Sum Sq	Df	F value
--	--------	----	---------

## FirstYear_CGPA	328.51	1	3504.1987
## College	0.08	2	0.4394
## FirstEnrolled_MajorCode	0.13	2	0.6710
## Switched		0	
## Gender	2.43	1	25.9547
## School_GPA	0.30	1	3.1540
## FincialAid_Received_AtLeast_Once	0.04	1	0.4380
## FirstYear_CGPA:College	0.52	2	2.7838
## FirstYear_CGPA:FirstEnrolled_MajorCode	0.01	2	0.0418
## College:FirstEnrolled_MajorCode		0	
## FirstYear_CGPA:Switched		0	
## College:Switched		0	
## FirstEnrolled_MajorCode:Switched		0	
## FirstYear_CGPA:College:FirstEnrolled_MajorCode		0	
## FirstYear_CGPA:College:Switched		0	
## FirstYear_CGPA:FirstEnrolled_MajorCode:Switched		0	
## College:FirstEnrolled_MajorCode:Switched		0	
## FirstYear_CGPA:College:FirstEnrolled_MajorCode:Switched		0	
## Residuals	119.53	1275	
##	Pr(>F)		
## FirstYear_CGPA	< 2.2e-16	***	
## College	0.64452		
## FirstEnrolled_MajorCode	0.51140		
## Switched			
## Gender	4.022e-07	***	
## School_GPA	0.07598	.	
## FincialAid_Received_AtLeast_Once	0.50821		

```

## FirstYear_CGPA:College                                0.06218 .
## FirstYear_CGPA:FirstEnrolled_MajorCode                0.95905
## College:FirstEnrolled_MajorCode
## FirstYear_CGPA:Switched
## College:Switched
## FirstEnrolled_MajorCode:Switched
## FirstYear_CGPA:College:FirstEnrolled_MajorCode
## FirstYear_CGPA:College:Switched
## FirstYear_CGPA:FirstEnrolled_MajorCode:Switched
## College:FirstEnrolled_MajorCode:Switched
## FirstYear_CGPA:College:FirstEnrolled_MajorCode:Switched
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

In the final model, we combined all hypothesized predictors of cumulative GPA, including first-year GPA, college, first enrolled major code, switching behavior, gender, school GPA, and financial aid status. Interaction terms were added to explore their combined effects on cumulative GPA. The results show that first-year GPA remains the most significant factor, as expected, given its direct influence on cumulative GPA. Gender also demonstrates significance, indicating differences in performance trends between genders. School GPA is significant, aligning with our earlier decision to include it despite its correlation with first-year GPA. However, financial aid status, while relevant in earlier analyses, does not show significance in this comprehensive model, suggesting it may be overshadowed by stronger predictors. Switching behavior and its interactions with college and major codes provide nuanced insights. These results support the hypothesis that students who switched majors might have experienced challenges that impacted their overall performance. Interactions such as first-year GPA with college and college with first enrolled major code and switching behavior reveal complex relationships that merit further investigation. This final model captures

the multifaceted nature of factors influencing cumulative GPA.

Cautions and Results

The analysis faced certain limitations due to restrictions in accessing complete data provided by OIRA, primarily to safeguard the privacy of students. Such constraints are not uncommon in real-world research, where data quality and completeness often pose significant challenges. It is crucial to fully acknowledge these limitations and ensure that any interpretations or conclusions drawn from the study remain considerate of these factors.

In real-life scenarios, dealing with imperfect or incomplete data is a routine aspect of research. However, what matters is how these challenges are addressed and accounted for in the analysis. The decision to limit the scope of testing to a specific set of analyses was made to maintain focus and clarity, even though other possible tests could have been conducted to explore additional insights. It is important to emphasize that this study's results should be examined with an understanding of the decisions made regarding scope, as well as the inherent data limitations, to prevent misinterpretation.

Acknowledgements

Special thank you to AUA's OIRA for providing the data and our instructor, Tadamasa Sawada, for guidance throughout this project.

References

- Data provided by AUA's Office of Institutional Research and Assessment (OIRA), 2023.
- Liang, Y.-W. (M.), Jones, D., & Robles-Pina, R. A. (2019). Ethnic and gender stereotypes on college students' academic performance. *Educational Research Quarterly*. Retrieved from <https://www.aabri.com/manuscripts/182858.pdf>
- Marini, J. P., Westrick, P. A., & Shaw, E. J. (2021). Using SAT® scores to inform academic major-related decisions and planning on campus. *College Board Research*. Retrieved from https://archive.org/details/ERIC_ED613434
- Stewart, S., Lim, D. H., & Kim, J. (2018). Factors influencing college persistence for first-time students. *Journal of College Student Retention: Research, Theory & Practice*. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1092649.pdf>