Retention Summary Analysis

April 18, 2019

This analysis shows breakdowns of retention data by year, major, and gender. Only minimal statistical analyses are performed, namely, linear regressions. No causal hypotheses are proposed. By looking at major and gender breakdowns, including retention rates as well as raw student counts, we may begin to focus our questions on how to retain more specific kinds of students (e.g., females in department X), and thereby improve the overall retention rate.

All data were obtained via the collection of Power BI reports known as “Retention,” and include retention data from Fall 2009 to Fall 2017 for University-wide data, and Fall 2012 to Fall 2017 for major-specific data.

University-wide Retention

The following plot shows the overall retention rate over the years, plus a linear regression.

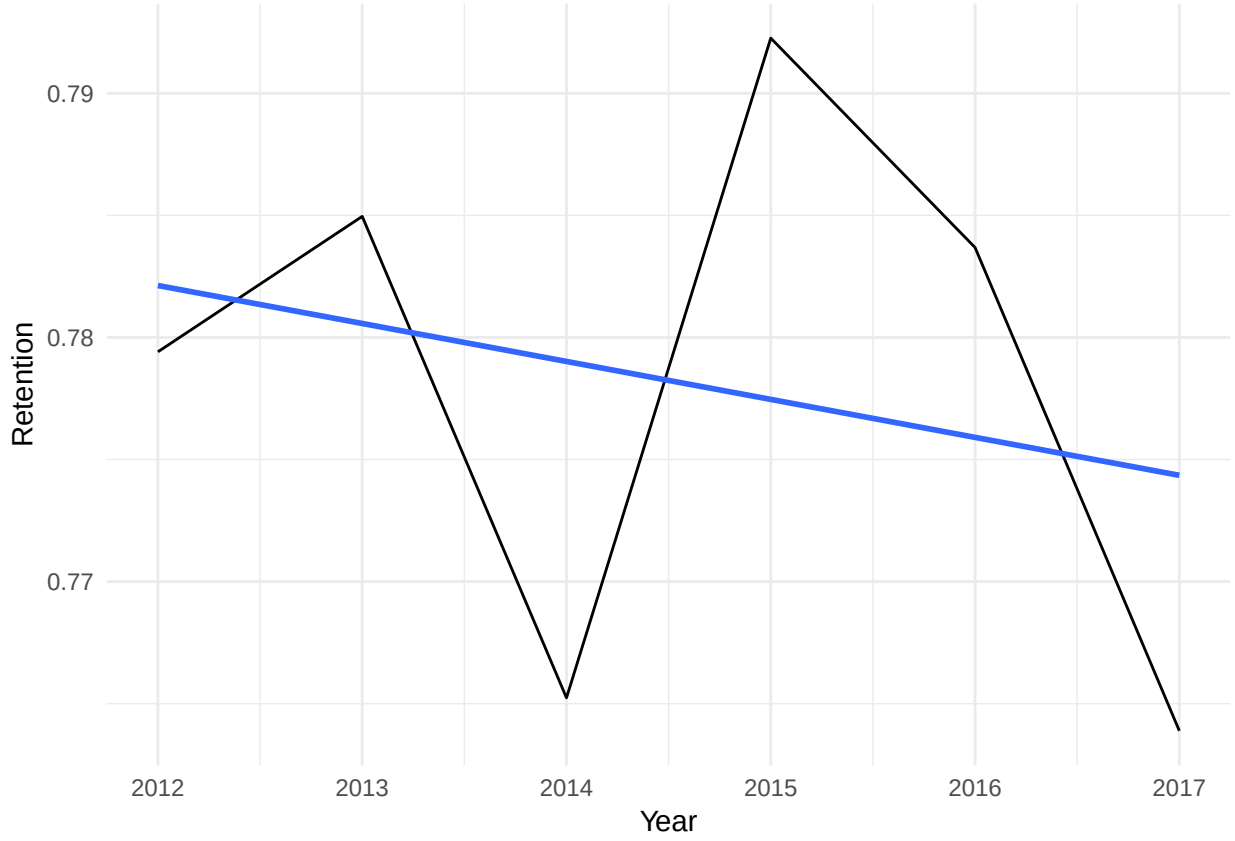

Figure 1: Retention University-wide with Linear Regression, Per Year 2012-2017

It might be important to note that the linear regression slope differs depending on which range of years are included. The plot below shows regressions from different ranges: 2009-2017, 2010-2017, etc. In only one case (2009-2017) is the regression slope positive. In most cases, the goodness of fit is quite low.

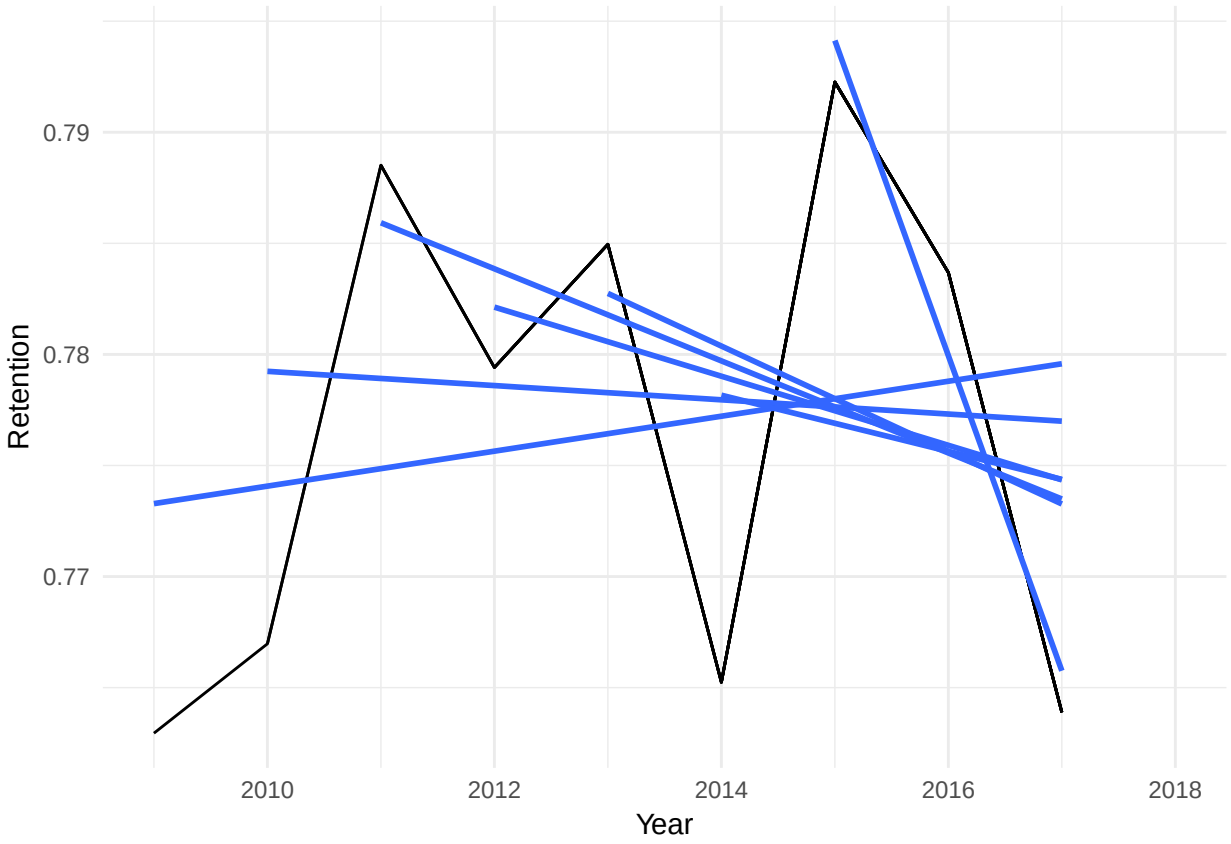

Figure 2: Retention University-wide with Linear Regression, Varying Starting Year

The next plot shows the raw counts of students who retained (white) and students who did not (red).

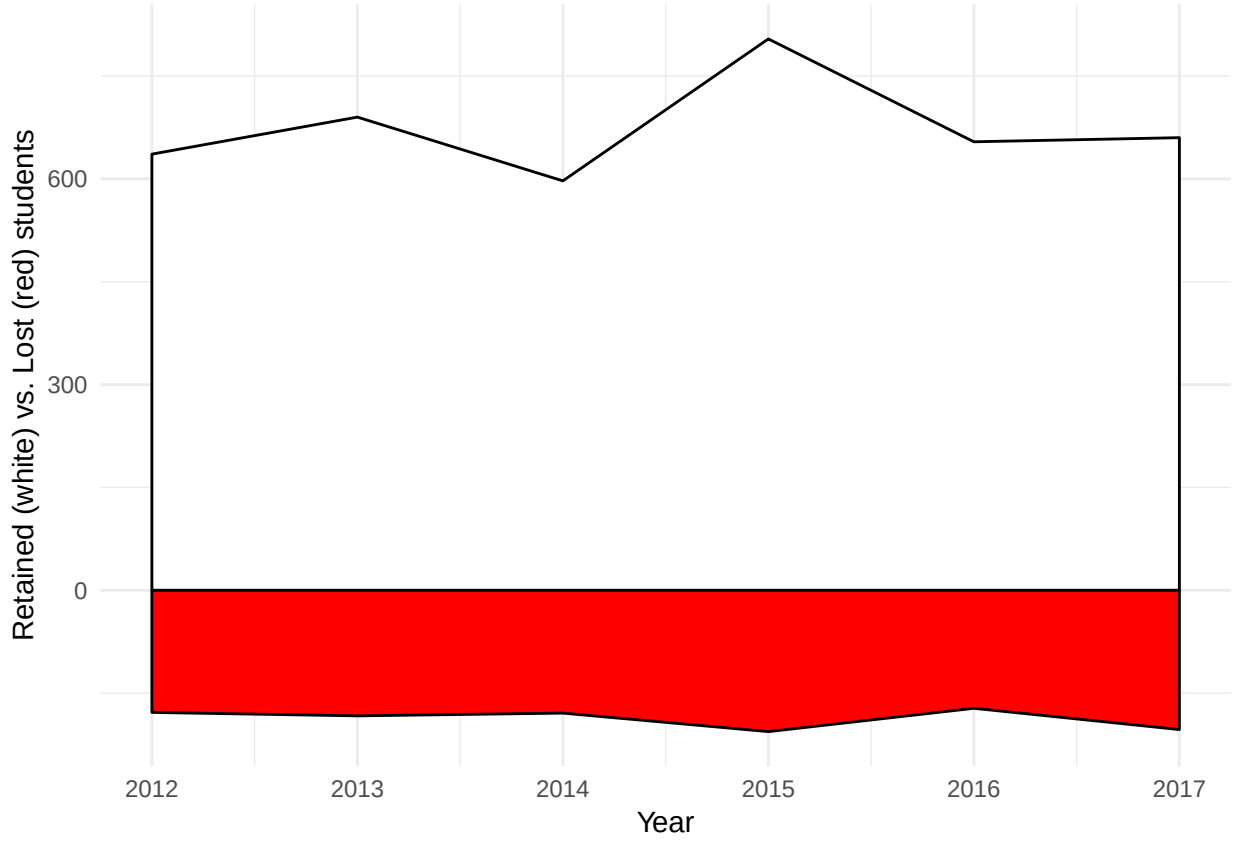

Figure 3: Total Retained and Lost Student Counts, 2012-2017

Major-specific Retention

Next, we break out retention by major. We limit the data to majors with at least 10 students in all years 2012 to 2017, and for which retention data is available for all years. These filters eliminate majors that do not individually have much contribution to overall University-wide retention.

In the plot below, we see linear regression models for majors in which the slope of the model gives an increase or decrease of 1% retention or more per year. Thus, the plot shows majors whose retention rates have dramatically changed over the years.

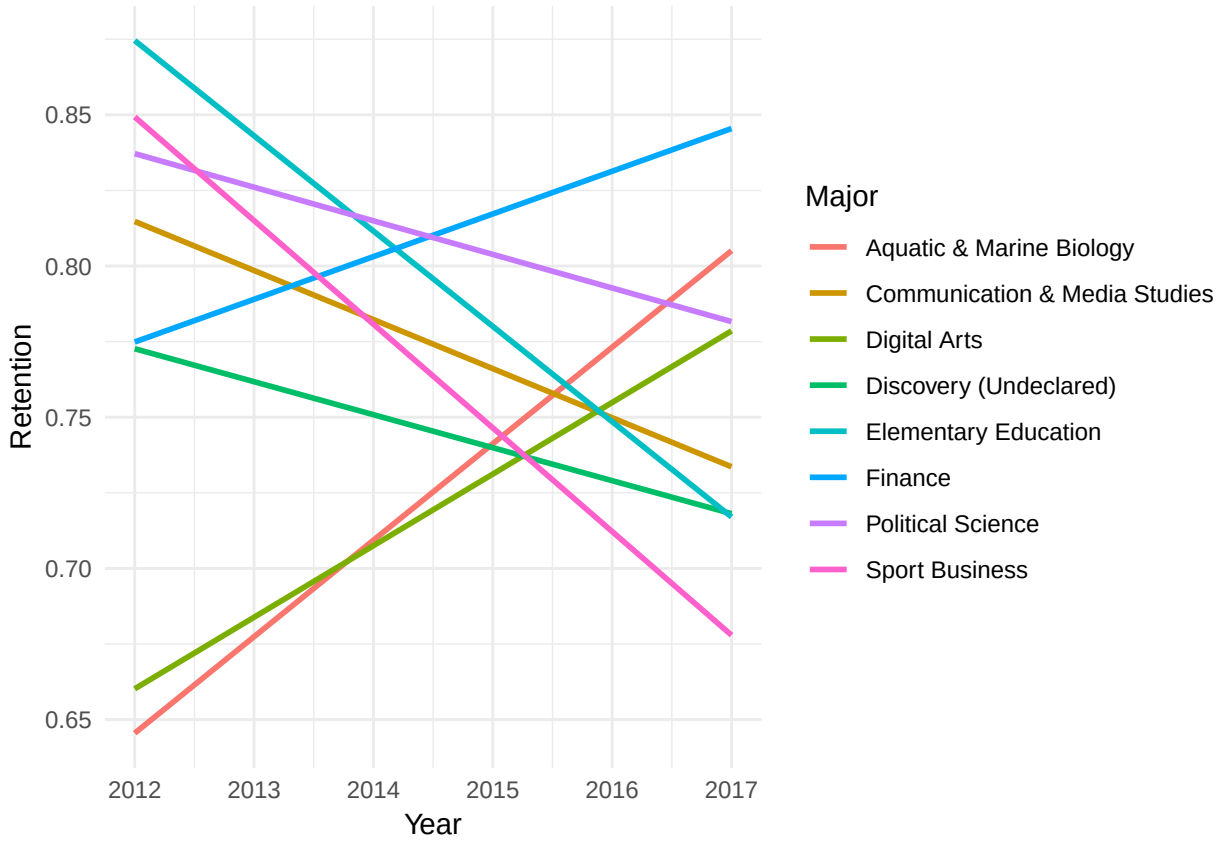

Figure 4: Linear Regression for Retention, just majors with dramatic changes (increase or decrease of 1% retention or more per year)

Next, we show the raw student counts for those who retained (white) and those who did not (red), per major. The counts are summed across all years (2012-2017).

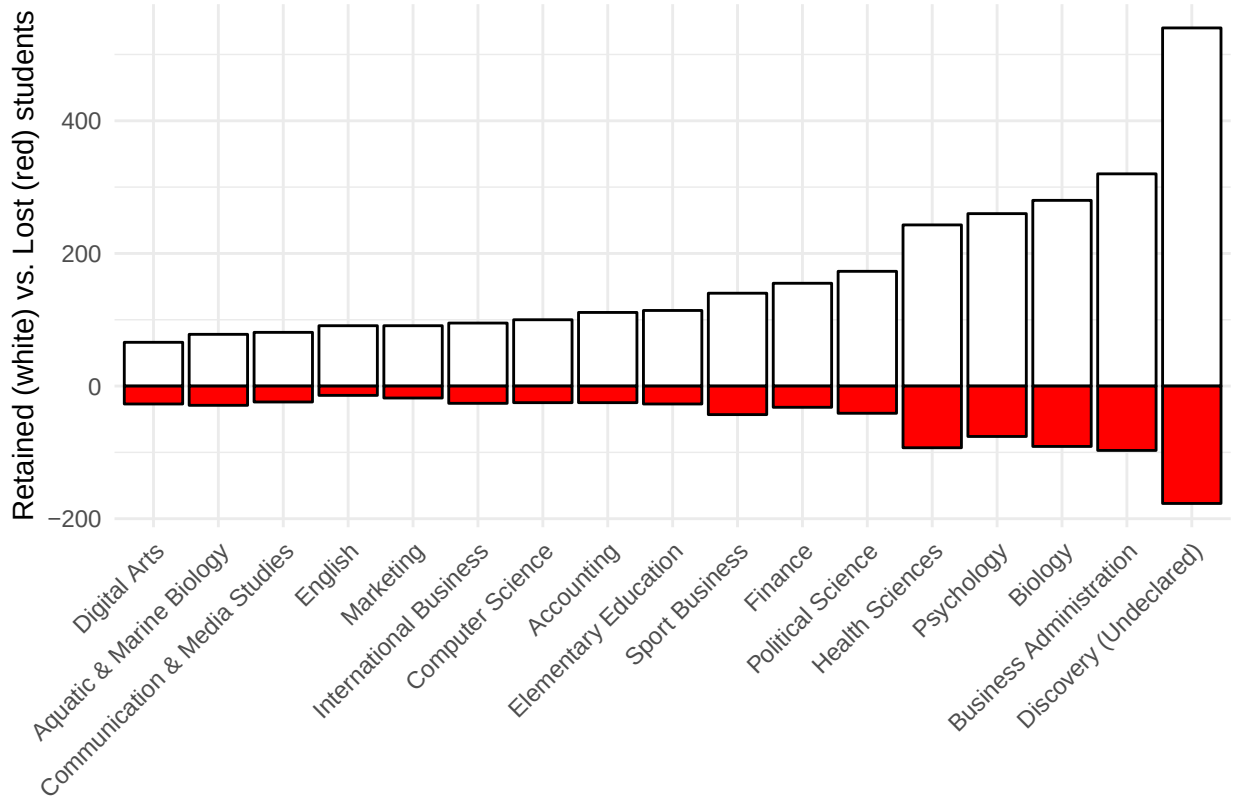

Figure 5: Total Retained and Lost Student Counts, 2012-2017

Rather than sum across all years, we can also show a yearly breakdown of raw student counts, per major, as shown in the next figure.

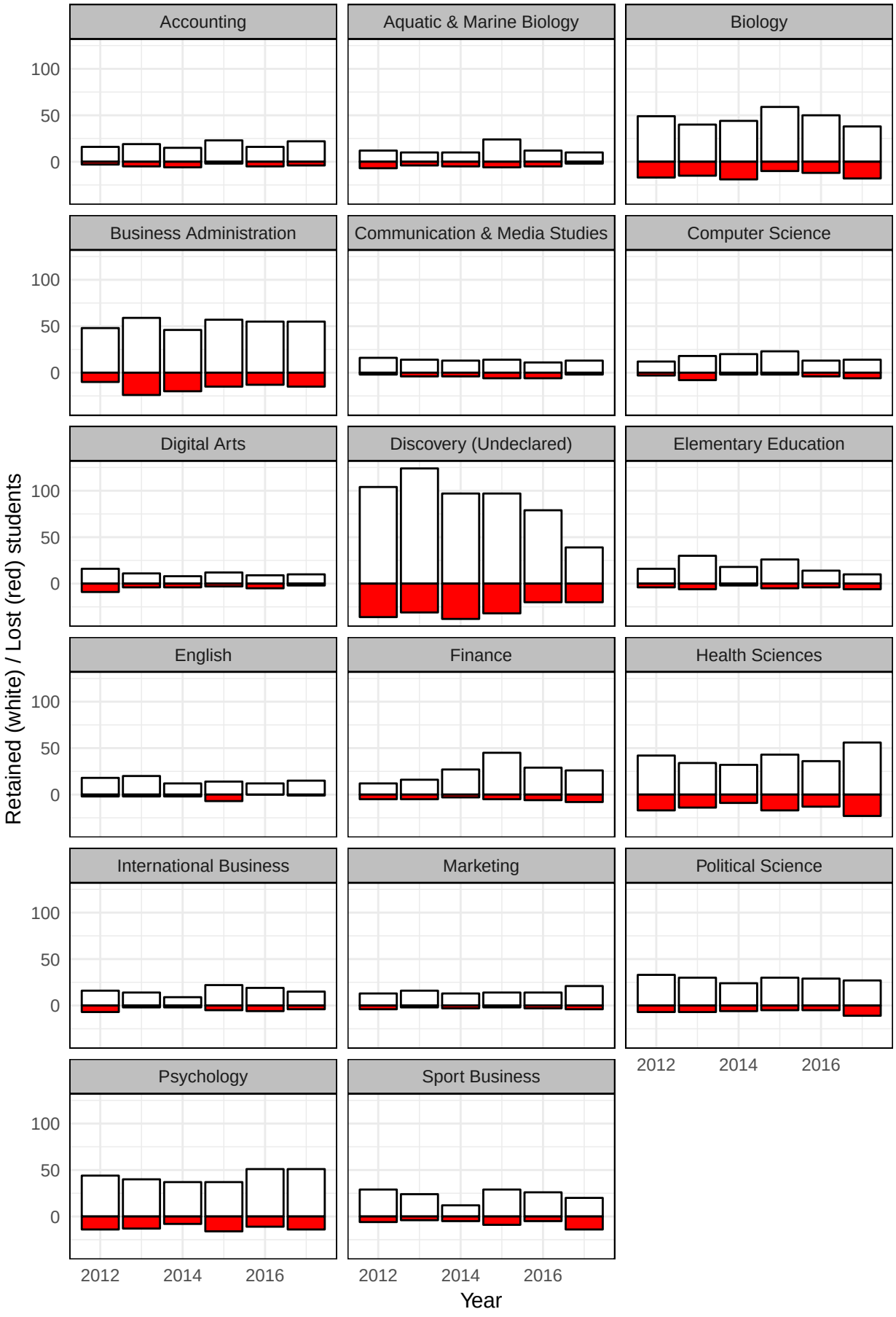

Figure 6: Retained and Lost Student Counts, Per Year

Major- and Gender-specific Retention

In order to drill down further, we can distinguish retention data by student gender. First, it is important to understand the overall gender distribution per major (summed across all years), as shown in the figure below.

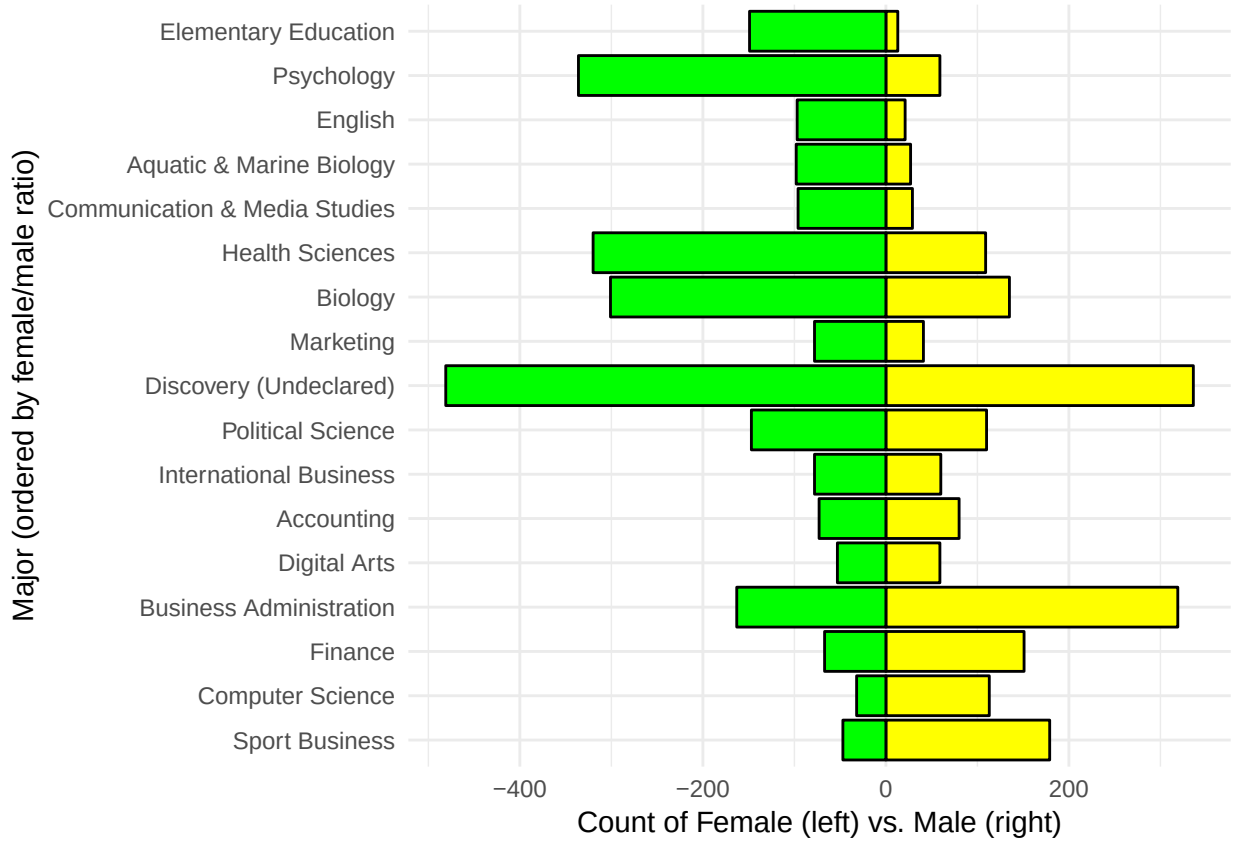

Figure 7: Count of Students by Gender, 2012-2017

Next, we can visualize the difference in retention rate per gender, per major (averaged across all years). We compute male retention rate minus female retention rate. So a bar left of 0 means female retention was higher than male retention in that major, and a bar right of 0 means male retention was higher.

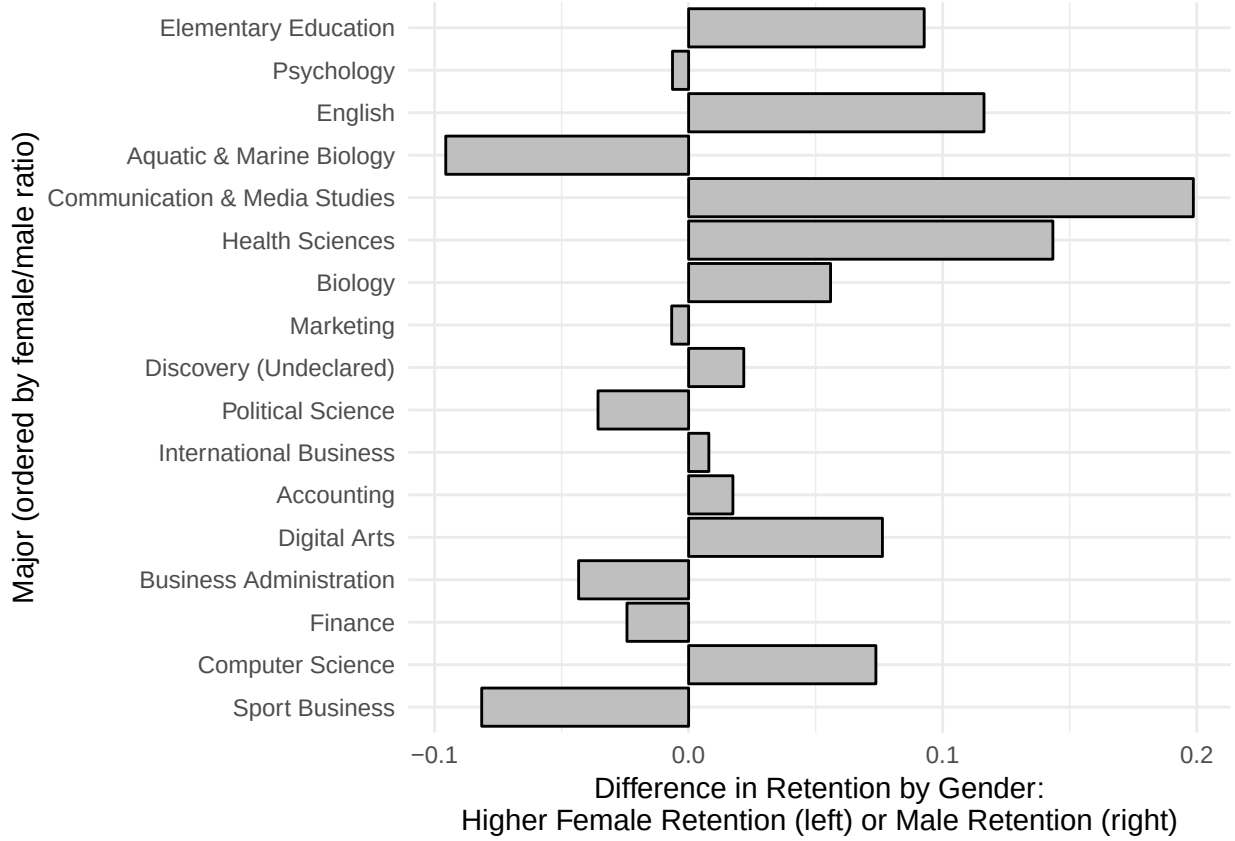

Figure 8: Difference in Female vs. Male Retention, 2012-2017 average

We can also show these data on a per-year basis, as shown in the next two plots.

Figure 9: Count of Students by Gender, Per Year

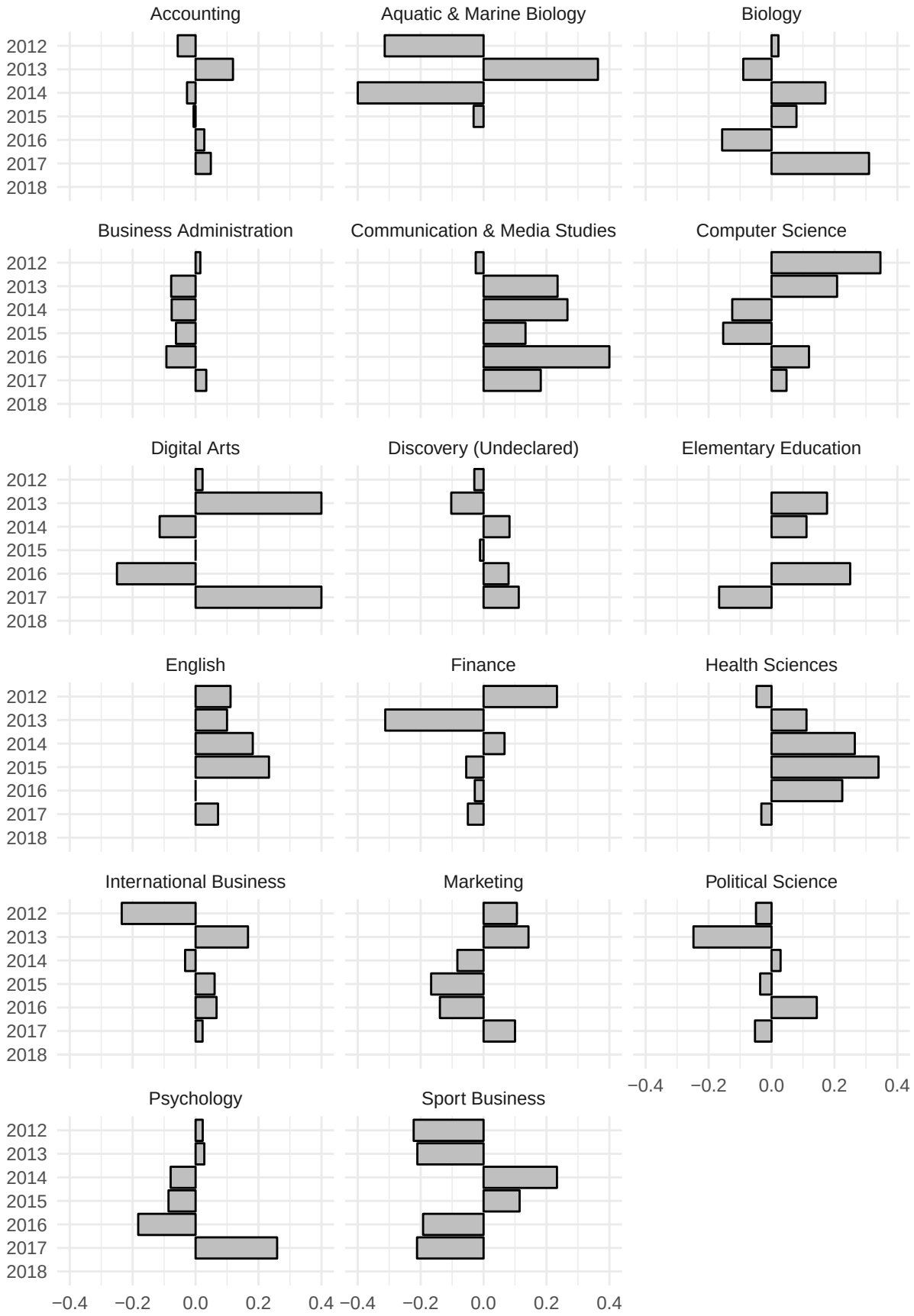

Figure 10: Difference in Female vs. Male Retention, Per Year

Finally, we can show the raw counts of students retained (white) and not retained (red), per major, per gender, and per year. This helps us isolate which majors and years witnessed retention issues for certain genders.

Figure 11: Retained and Lost Female Student Counts, Per Year

Figure 12: Retained and Lost Male Student Counts, Per Year