Analysis of the Impact of Advising on UVU Students' Performance – 2018-19 School Year

Andre Rossi de Oliveira

Jacob Sorenson

Samuel Elzinga

Gregory Brooks

1. Introduction

In this report, we investigate the impact of academic advising on student performance using data from the 2018-19 academic year. This investigation is limited in scope by design, both in terms of the data used and the objectives pursued. We have complete data on student characteristics and advising since 2012, but we elected to work only with the 2018-19 data initially for two main reasons. The first reason is that the size of the entire data set is so large that running the code we wrote for the statistical analysis is a very time-consuming experience. We will have to secure more powerful computer resources before we can use the full data set at once. The second is that, by focusing on a smaller data set, we can hone our data mining techniques and learn lessons that can be applied to the rest of the data.

Our main objective in this report is to understand the relationship between student visits to advisors and student performance as measured by their average course grade in a semester. There are many other measures of student success in our data set that we could have and will eventually use, like retention and graduation rates, but we decided to limit the analysis to course grades for now in order to assess the suitability of our statistical methods and generate a relatively small batch of initial results to kick off the discussion with Academic Advising.

The report is organized as follows. The next section gives a brief description of the two data sets we worked with and some of the main issues we encountered. Section 3 contains a collection of plots that show how our data set is distributed according to student characteristics and visits to advisors. Section 4 also provides data visualizations tools, but it focuses on the outcome variable, namely semester grade point average. Data visualization is complemented by tests of measures of central tendency. In section 5, we employ regression analysis in a first attempt to quantify the impact of advising on student performance. Section 6 contains our concluding remarks.

2. Data issues

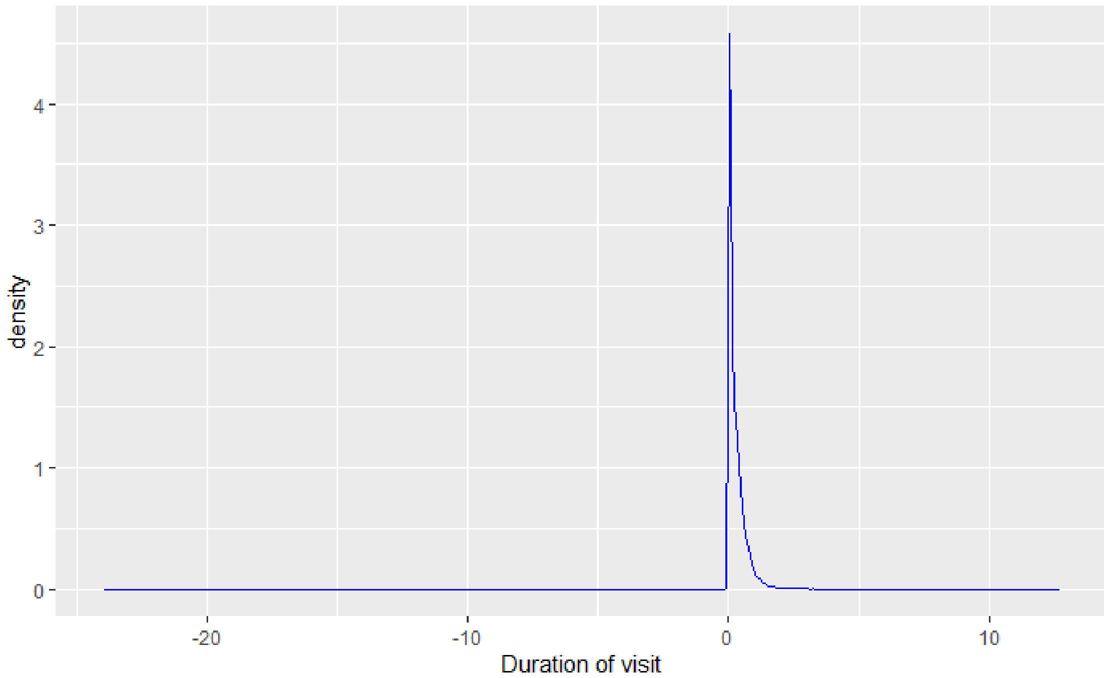
The data for the 2018-2019 school year that we used in this study comes from two data sets. The first, provided by UVU's Institutional Research (IR), contains data on student characteristics such as gender, age and ethnicity, as well as academic measures like courses taken, GPA, course grade and major. The second data set comes from UVU's Academic Advising (AA), and covers data on student visits to advisors, including dates and times, reason for visit and advising center.

The IR data set for the 2018-19 academic year consists of 887,601 observations, but it is important to stress that there are several entries per student. Each combination of course taken and change of major corresponds to a different entry. There are many data issues that we had to deal with, including clerical errors and outliers. Some issues that remain and must be addressed in the future are the following:

- Identify data, if any, that refers to concurrent enrollment students.
- Some students have GPA equal to 0.
- The definitions of some variables are not entirely clear, including 'Major Changes', 'Major Change Term', 'QL Complete' and 'CC Complete'.

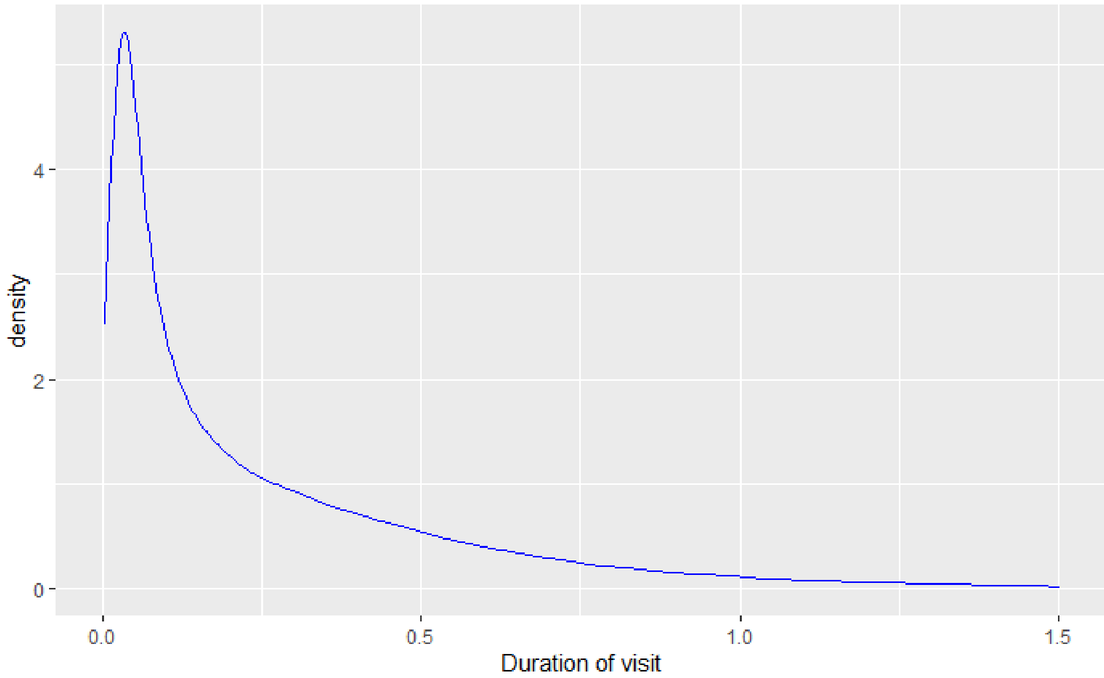
The AA data set (2018-19) consists of 170,084 observations, but it is important to keep in mind that many students have more than one visit to an advising office in a given semester. There are 165,152 observations that correspond to visits shorter than 1.5 hours (and positive), 2,835 to visits that did not include an end time, 1,874 to visits that lasted between 1.5 and 4 hours, 211 to visits longer than 4 hours (up to 12.6), and 10 to visits with negative durations. Clearly, some of these records are in error. The distribution of the visits by their duration can be seen in the figure below (missing values have been eliminated).

Figure 1: Distribution of visits by duration – Full data



The distribution of the data after potential errors have been purged (165,152 data points, corresponding to 97.1% of the data) is shown in the next figure.

Figure 2: Distribution of visits by duration – Adjusted data



Reasons for errors in the recording of visit duration may include but are not limited to, switching the start and end times, not recording the end time of the visit, and switching AM and PM on the recorded times. In particular, it is important to investigate if the students whose visits has no recorded time out are those who didn't show up for the meeting.

3. Data visualization

3.1. Visualization of student characteristics

Before we set out to investigate how student performance is affected by advising, we take a closer look at our data set. In this section, we provide plots to visualize the distribution of the following student characteristics: gender, ethnicity, age, residency, class standing, and generation (first generation or not).

As can be seen in the first figure below, almost 45% of the students in the 2018-19 academic year were female and a little over 55% were male.

Figure 3: Distribution by gender

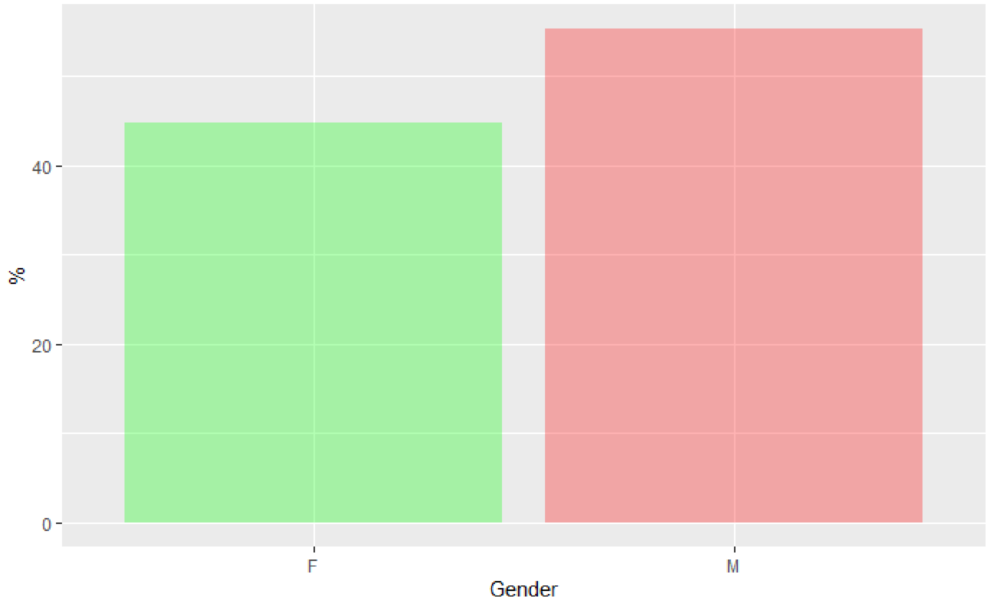
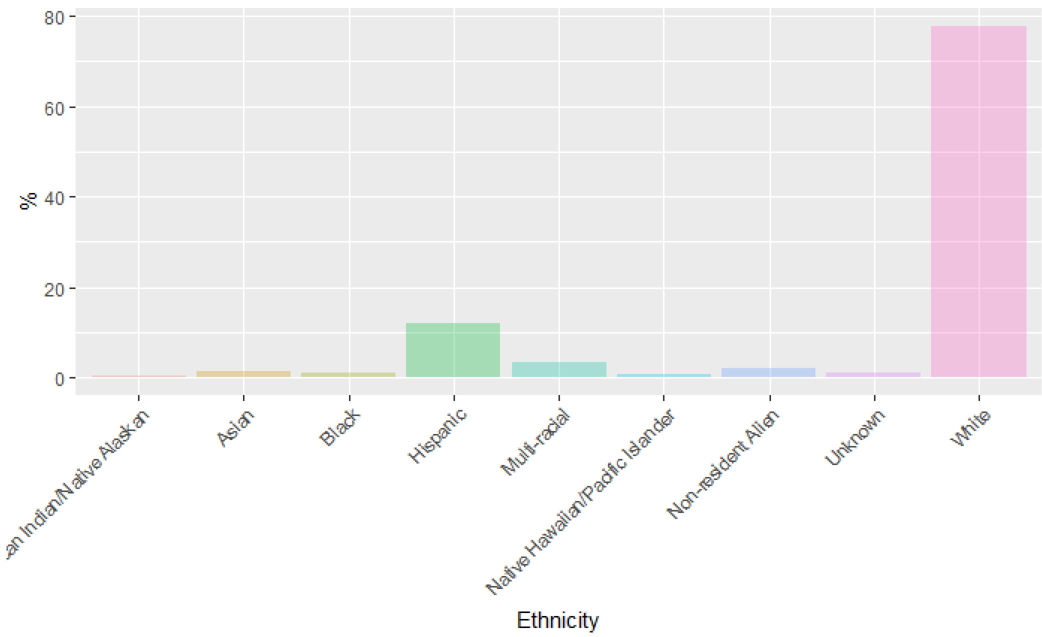


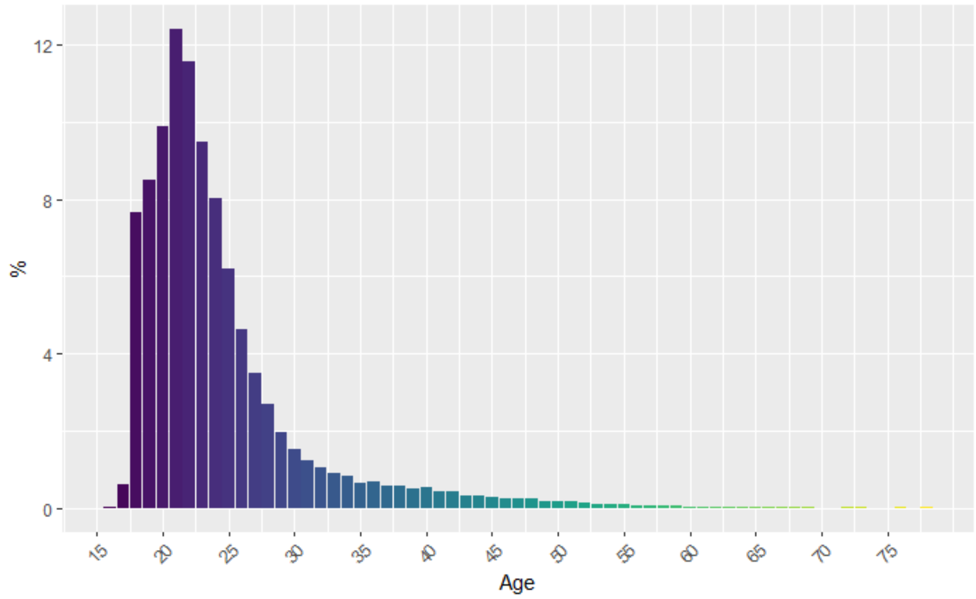
Figure 5 shows the distribution by ethnicity. Notice that the overwhelming majority (almost 78%) of students in our data set are white, followed by Hispanics (almost 12%).

Figure 4: Distribution by ethnicity



The distribution of UVU's student population by age in the 2018-19 academic year can be seen in the picture below. The only age groups that have a share higher than 4% are those comprised of students 18 to 26 years old.

Figure 5: Distribution by age



The next few plots show how students in our data set are distributed by residency, class standing, and generation (first or not).

Figure 6: Distribution by residency

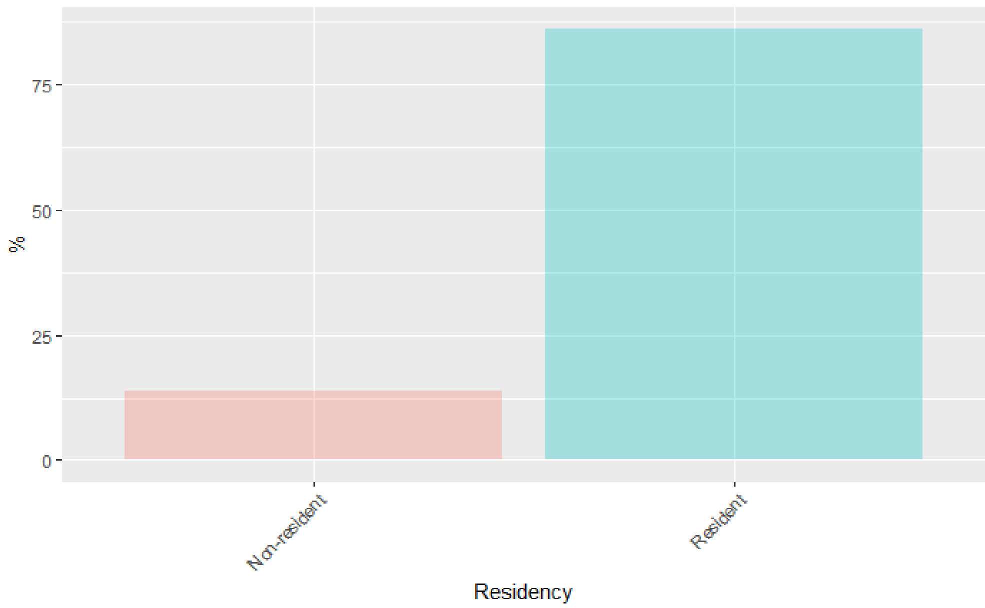


Figure 7: Distribution by class standing

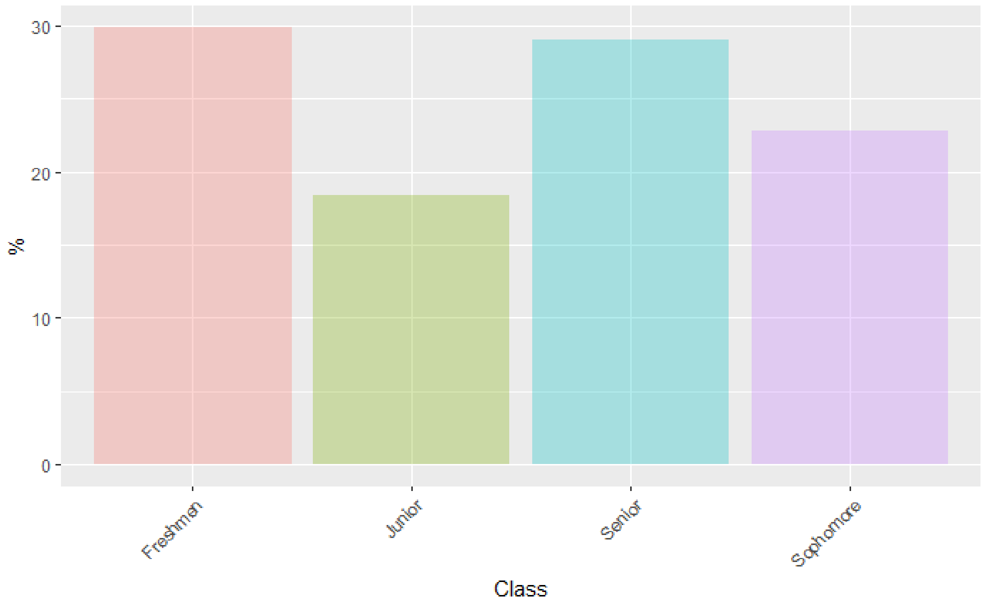
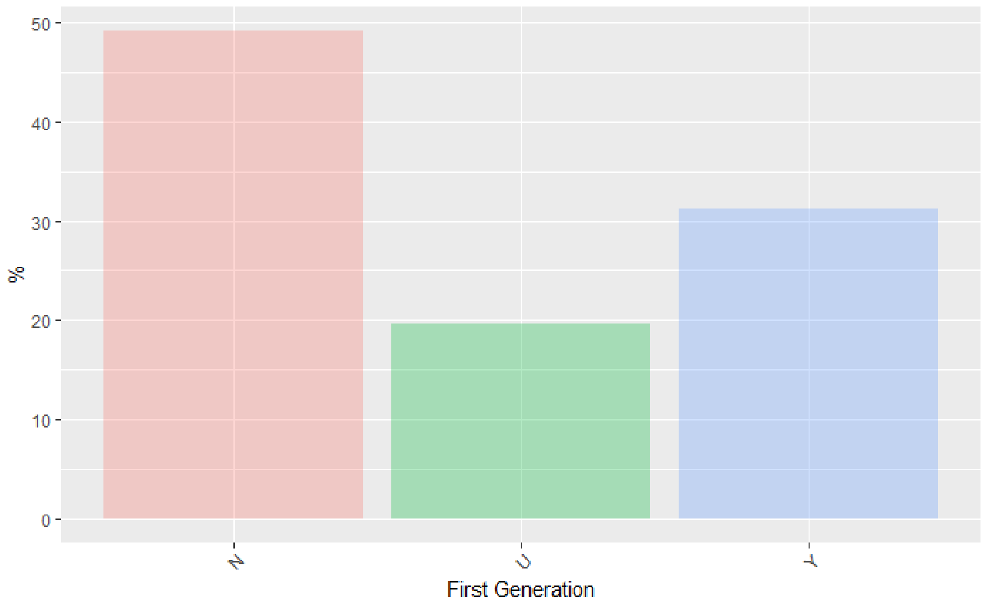


Figure 8: Distribution by generation



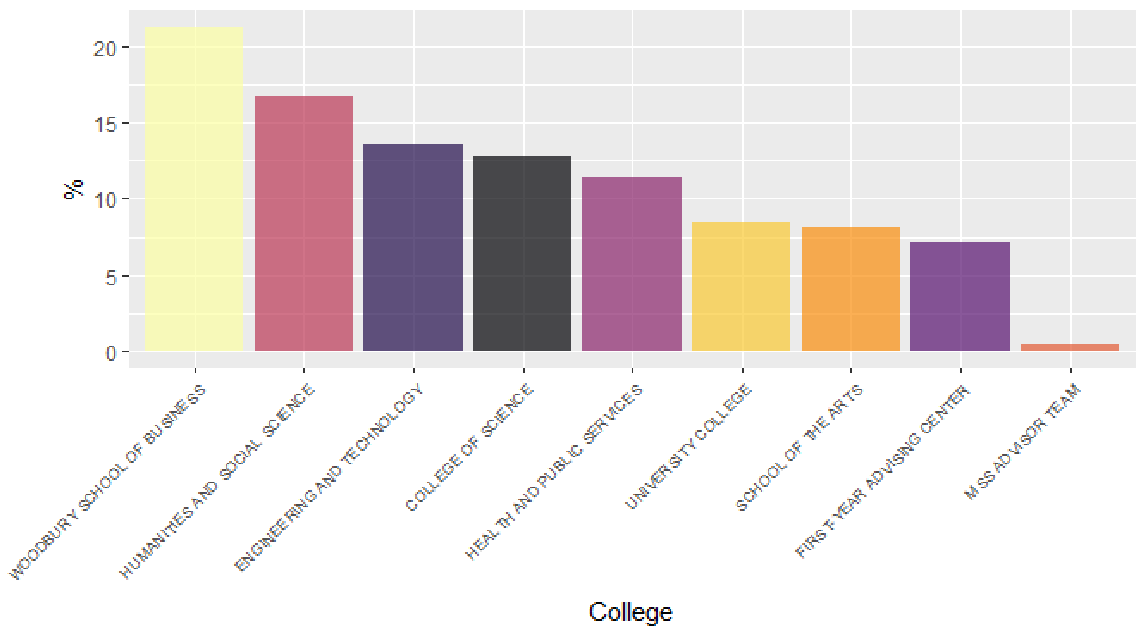
The large majority of UVU students, over 86%, are Utah residents. Given that UVU is a regional institution, that does not come as a surprise. Most students in the 2018-19 academic year were freshmen, followed by seniors, sophomores, and juniors. Finally, a large share of our data set (over 31%) is comprised of first-generation students, but the majority are non-first-generation students (over 49%)[1].

3.2. *Visualization of visits to advisors*

In this section, we highlight some aspects of the interaction between UVU advisors and students that may provide valuable input to the Advising Office, especially with regard to optimizing the use of their resources.

We first order colleges/schools and advising centers by their shares of total visits to advisors in the 2018-19 academic year. As can be seen below, the Woodbury School of Business had the largest share, with over 20% of visits. University College and School of the Arts had the smallest shares among colleges/schools, with around an 8% share each. Specialized centers, such as the First Year Advising Center and the Multicultural Student Services (MSS) Advisor Team had markedly smaller percentages.

Figure 9: Distribution of advising visits by college/school



[1] The label U stands for unknown.

The next seven figures show the five majors with the largest shares of visits in each college/school.

Figure 10: Distribution of visits to advisors by majors – Woodbury School of Business

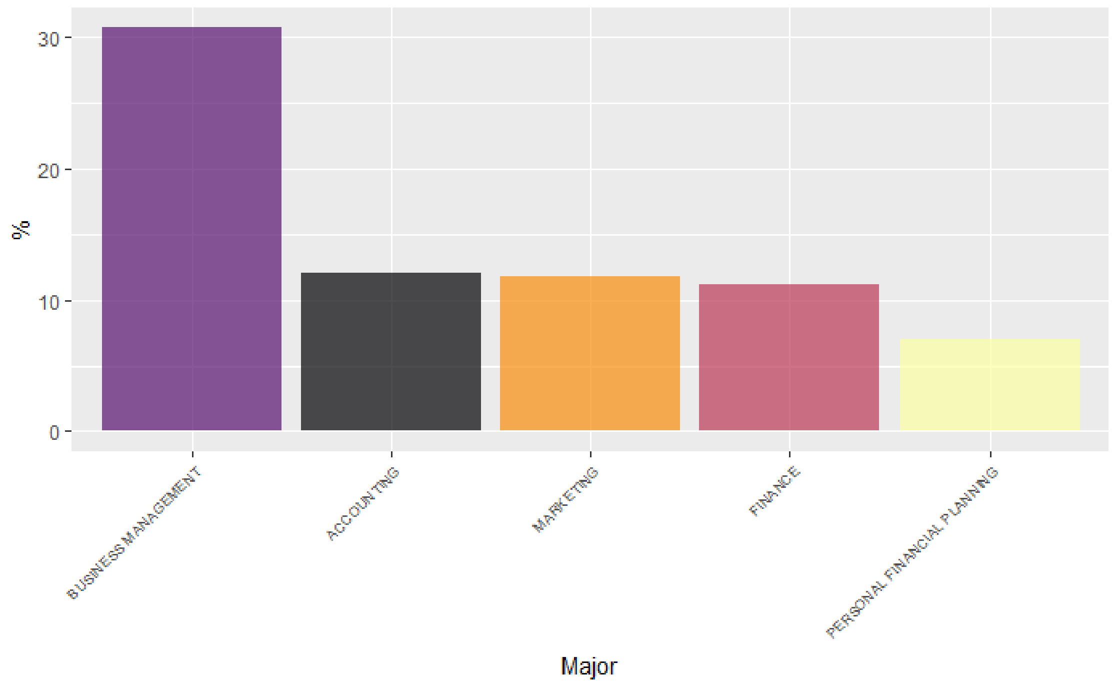


Figure 11: Distribution of visits to advisors by majors – Humanities and Social Science

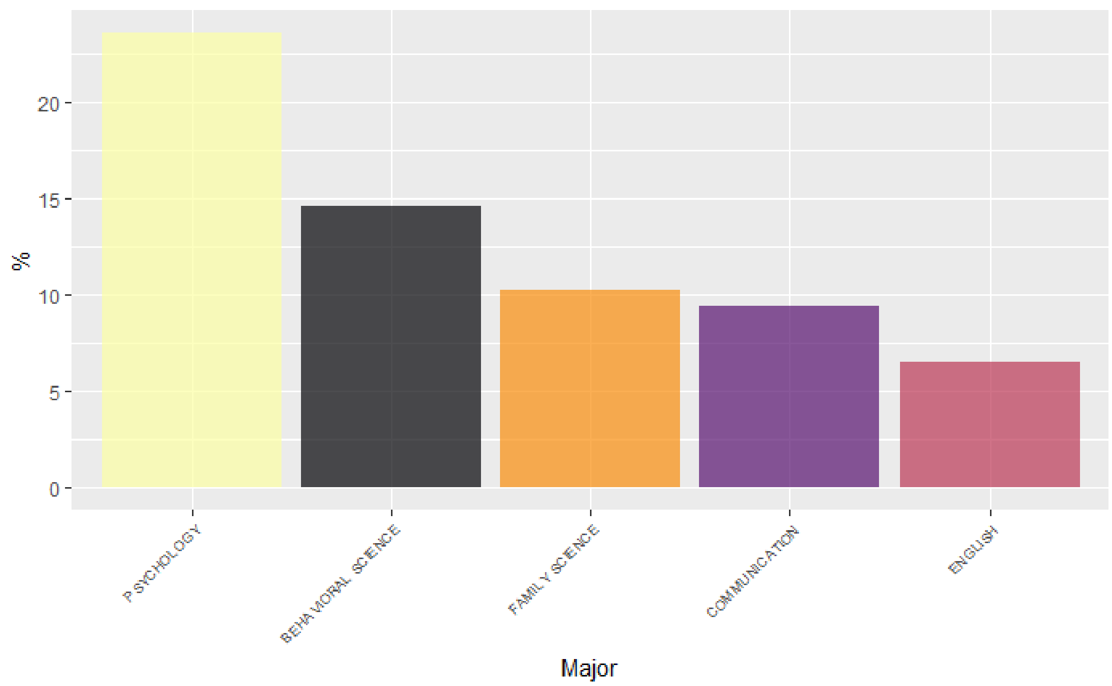


Figure 12: Distribution of visits to advisors by majors – Engineering and Technology

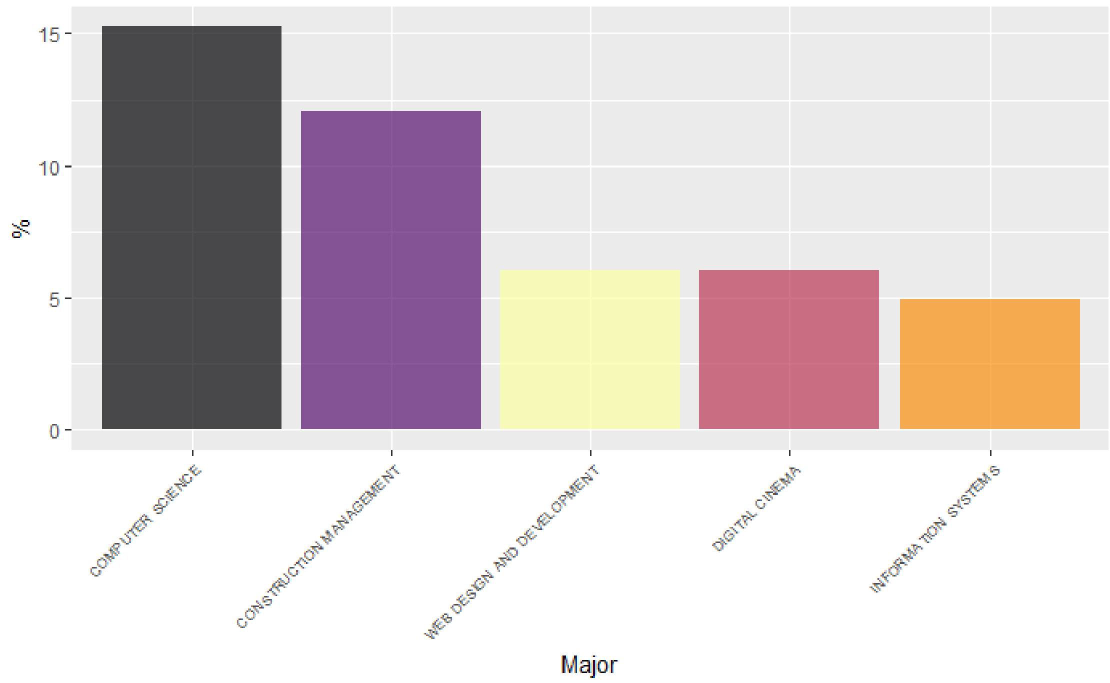


Figure 13: Distribution of visits to advisors by majors – College of Science

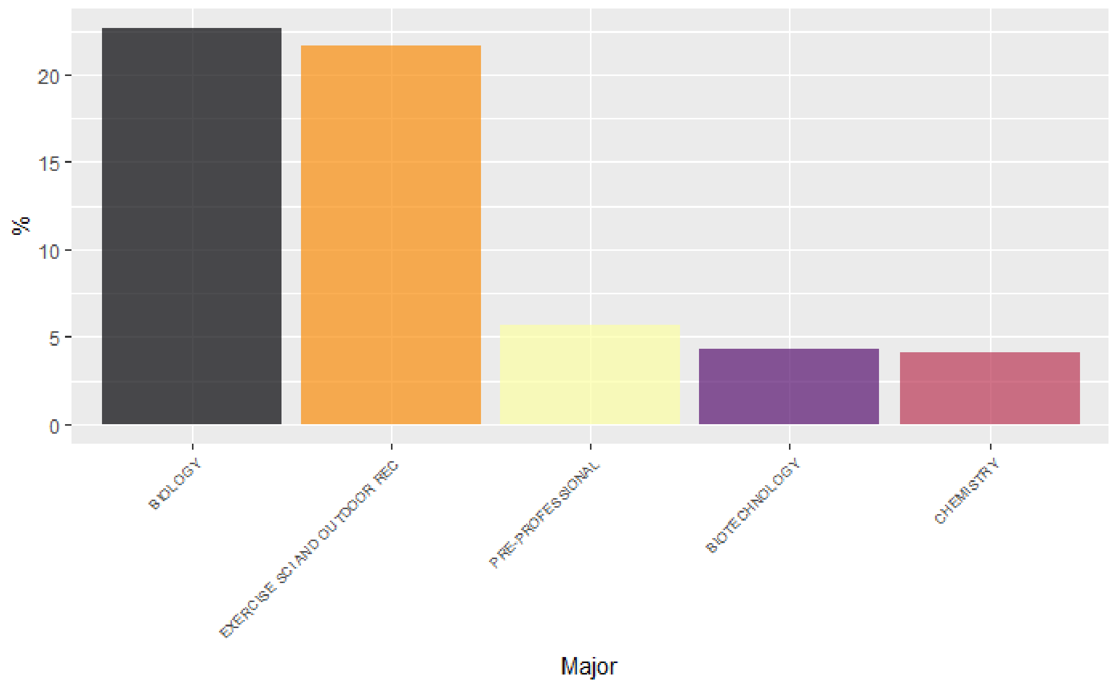


Figure 14: Distribution of visits to advisors by majors – Health and Public Services

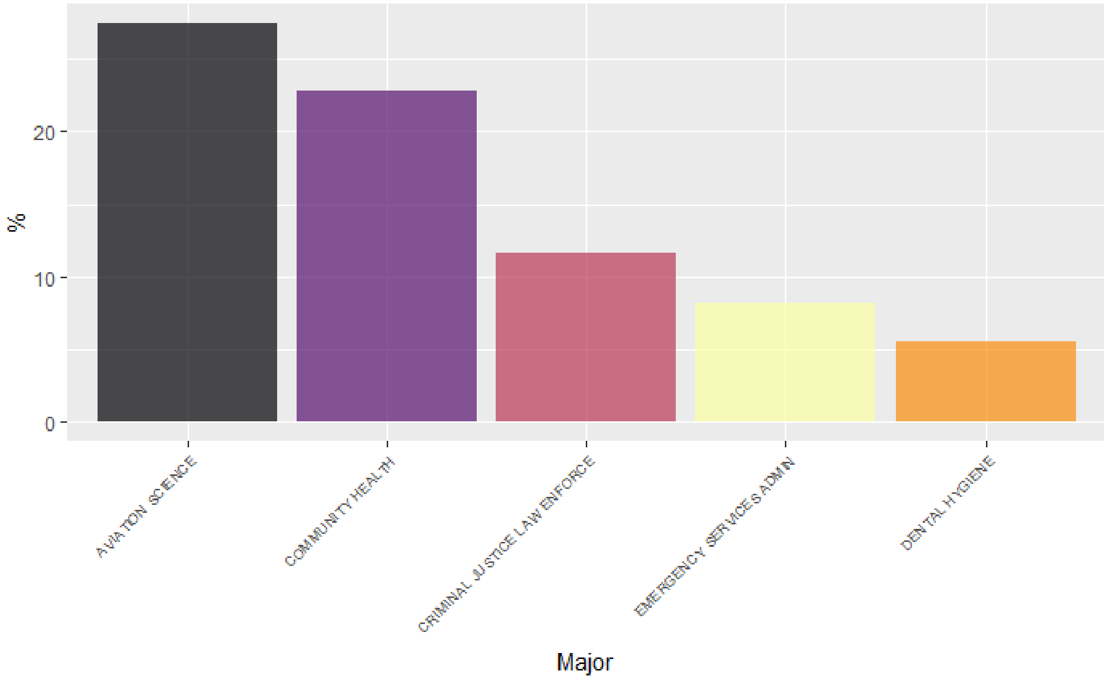


Figure 15: Distribution of visits to advisors by majors – University College

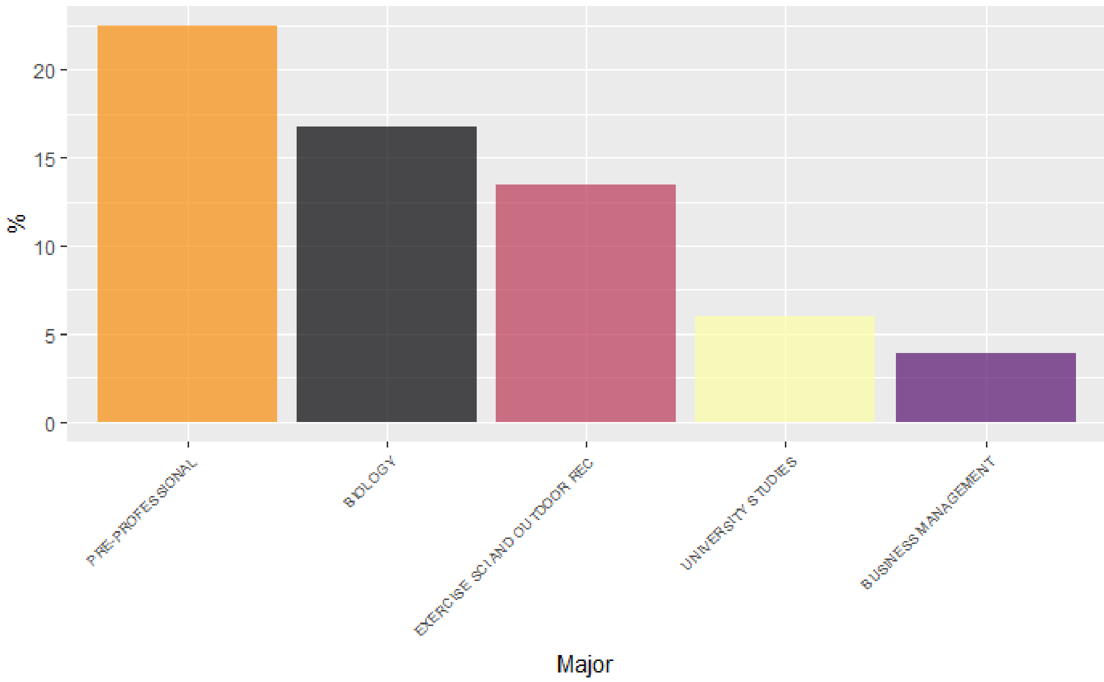
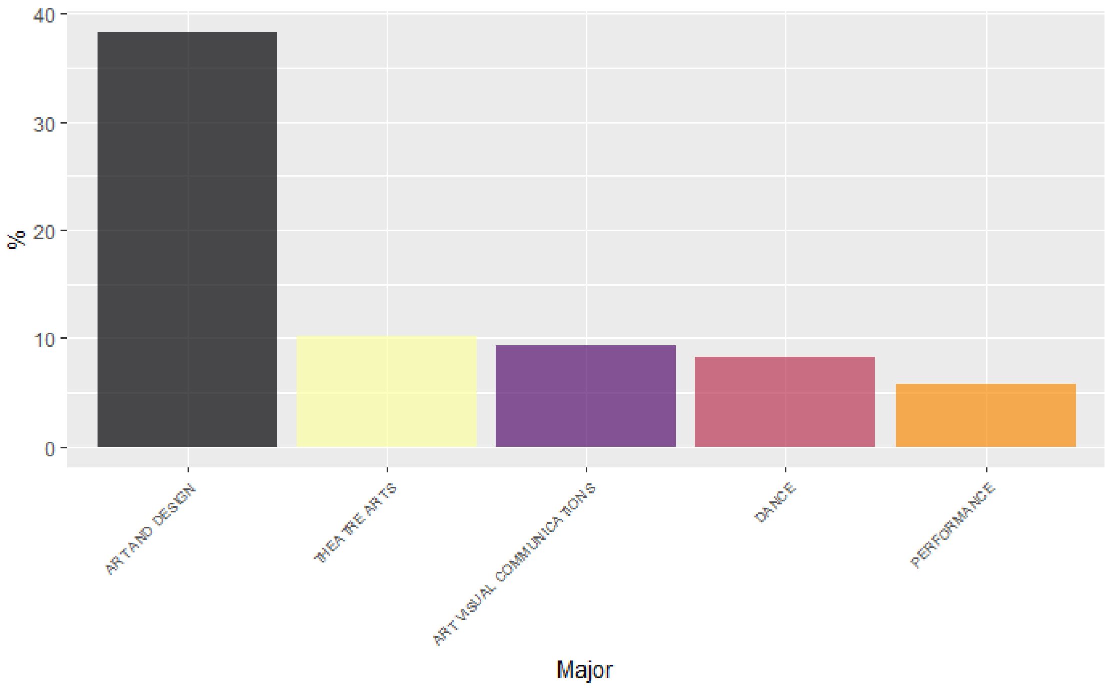
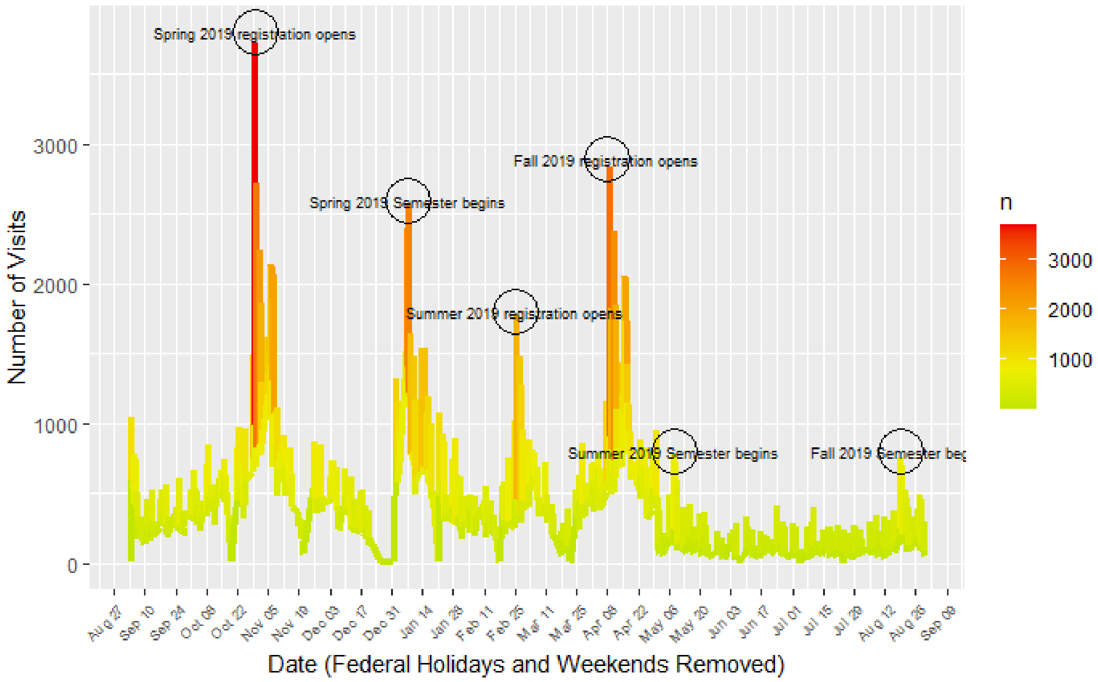


Figure 16: Distribution of visits to advisors by majors – School of the Arts



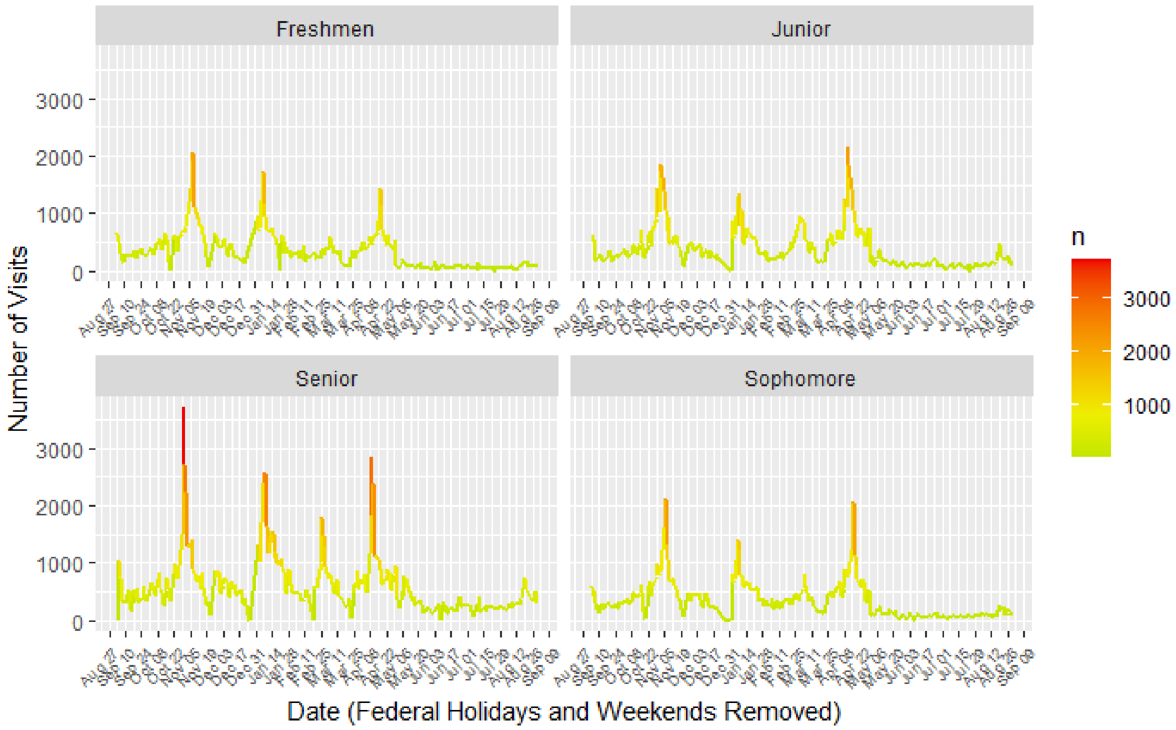
We also conducted a time series analysis to determine when advisors were busiest during the academic year.

Figure 17: Time series of visits to advisors



For this analysis, weekends and federal holidays were removed, as advisors do not regularly meet with students during those times. Notice that students are more likely to visit advisors around registration times and the beginning of new semesters. This pattern is shown by students of all class levels, as can be seen below, even though the magnitudes may differ.

Figure 18: Time series faceted by class level



4. Visualization and analysis of outcome variable by visits to advisor

In this section, we begin to investigate the potential impact of advising on student performance, measured here by semester grade point average. This outcome is a numeric variable that was constructed by converting letter grades to numerical grades as follows:

Table 1: Letter grade to points conversion table

Letter Grade	Points
A	4
A-	3.7
B+	3.4
B	3
B-	2.7
C+	2.4

C	2
C-	1.7
D+	1.4
D	1
D-	0.7
E (Fail)	0
UW (unofficial withdrawal)	0

As can be seen in the next two figures, students who paid a visit to an advisor had higher average and median scores both in Fall 2018 and Spring 2019. Moreover, the entire distribution of average grades for those who visited with an advisor is higher than that of students who didn't in both semesters.

Figure 19: Boxplot of semester average grade by visit to advisor – Fall 2018

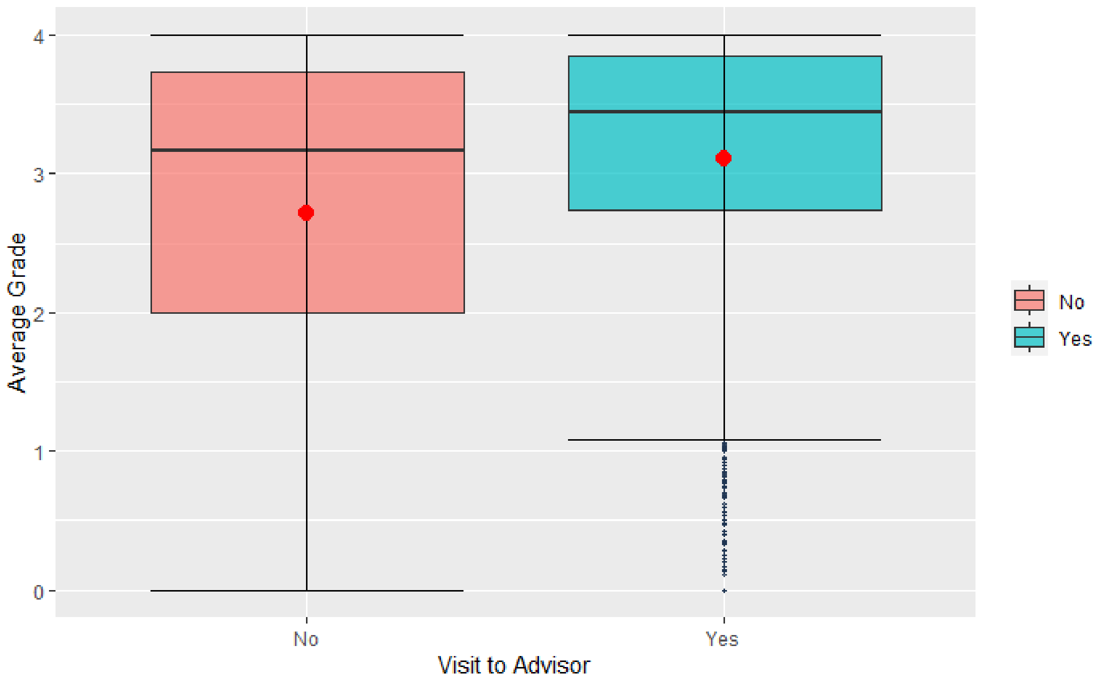
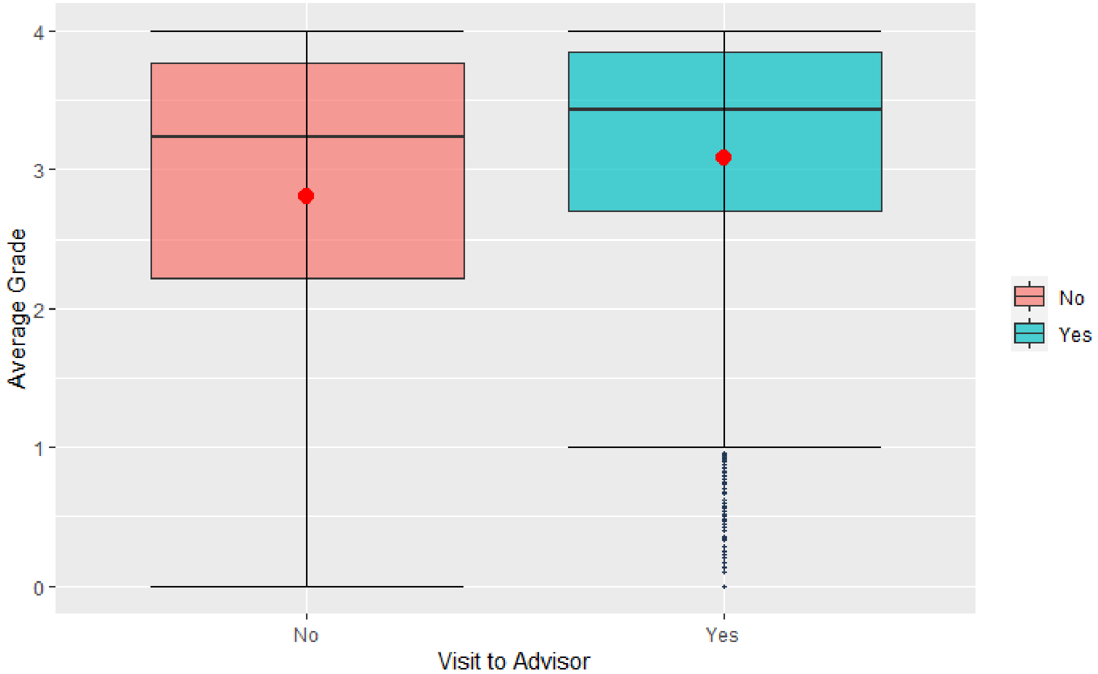


Figure 20: Boxplot of semester average grade by visit to advisor – Spring 2019



In order to check if these differences are statistically significant, we ran a t-test using data for the entire 2018-19 academic year, whose results are found in the table below:

Table 2: t-test semester average grade

Value of t statistic	-29.854
Degrees of freedom	42,470
P-value	0.000000
Mean visited	3.094068
Mean did not visit	2.810079

Since the p-value is lower than 0.05, we reject the null hypothesis that the difference in means is equal to zero. In other words, the two means are statistically different.

Let's now run Mood's Median test to check if the medians are statistically different.

Table 3: Mood's Median test semester average grade

Value of z statistic	-19.26
P-value	0.000000

The p-value is practically zero, so we reject the null hypothesis. This is evidence that there is a statistically significant difference in the medians of the two groups.

We also investigated the relationship between timing of the visit to an advisor and semester average grade. We created dummy variables based on whether students met with their advisors before or after registration for the following semester. For instance, the dummy variable in the Spring 19 semester was based on whether they met with their advisor before or after registration for Fall 19.

Whether students visited their advisors before or after registration did not have a significant impact on their semester grade averages, as can be seen in the next two boxplots. The medians and averages are very close, and the distributions are similar overall.

Figure 21: Boxplot of semester average grade by timing of visit to advisor – Fall 2018

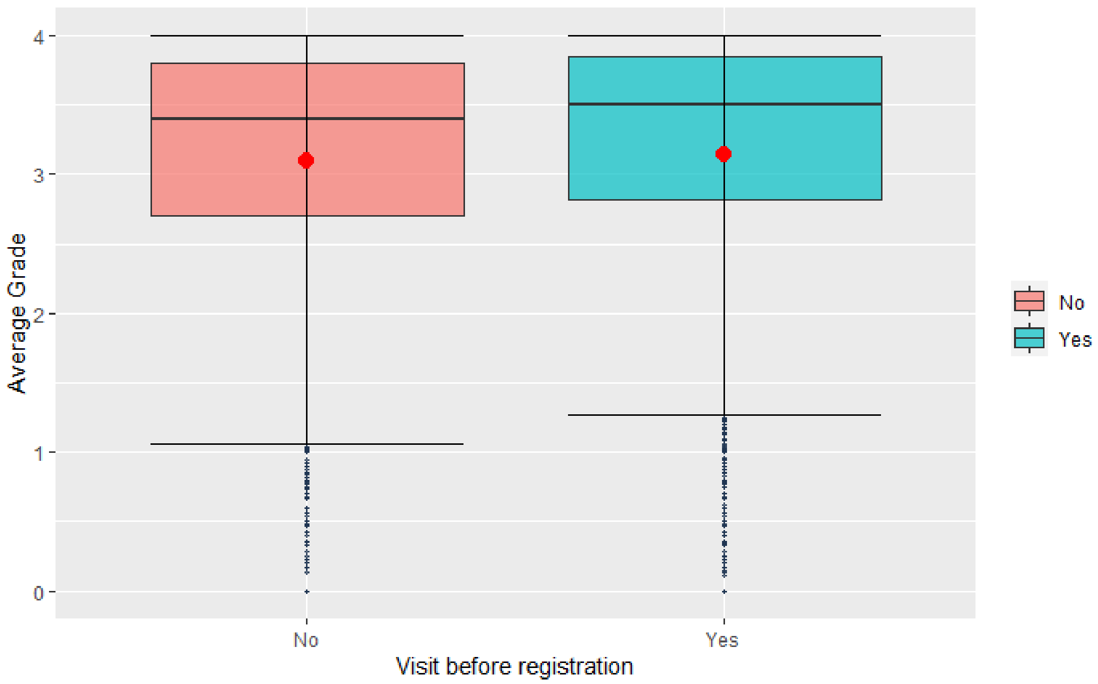
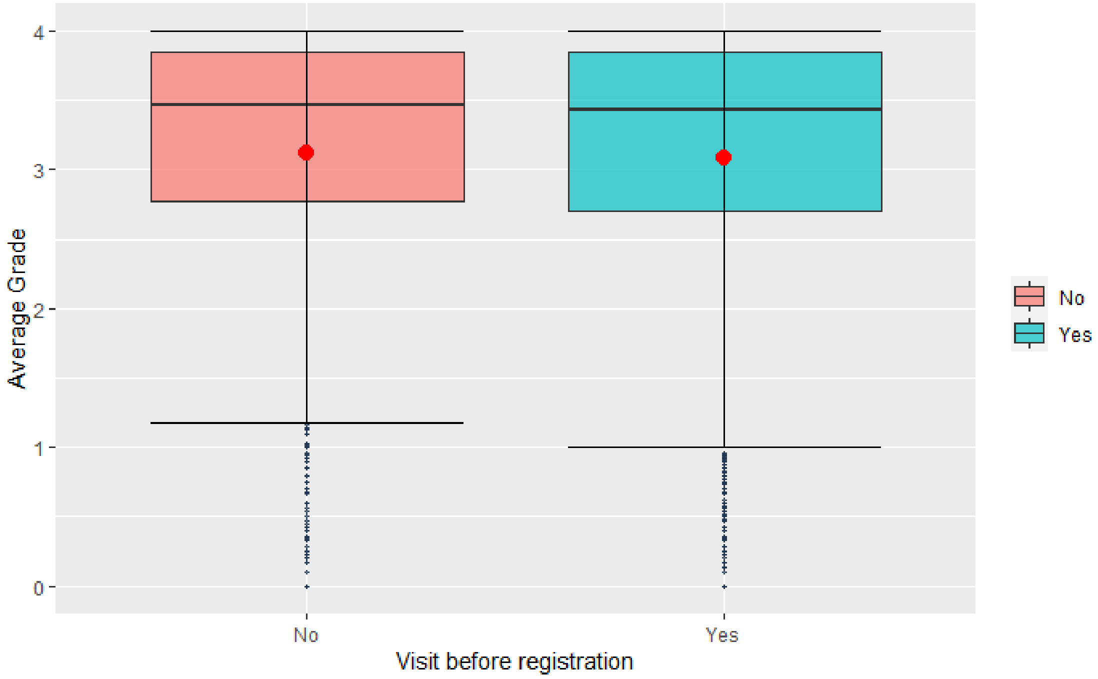


Figure 22: Boxplot of semester average grade by timing of visit to advisor – Spring 2019



5. Estimation of treatment effects

In this section, we employ regression analysis to evaluate the potential impact of visits to advisors on student performance. The dependent variable is 'grade point average', and the explanatory variable we are mainly interested in is 'visit', a dummy equal to 1 when the student visited with an advisor and 0 otherwise. The other explanatory variables are GPA, age, and dummies for gender, ethnicity, residency, class standing and generation (first or not).

The results of the linear regression can be found in the table below.

Table 4: Linear regression results

Variable	Coefficient	St. error	T statistic	P-value
Intercept	0.161071	0.059268	2.718	0.00658
Visit	0.20046	0.008095	24.765	< 2e-16
GPA	0.805789	0.005654	142.513	< 2e-16
Gender	-0.09733	0.007927	-12.278	< 2e-16
Age	0.002953	0.000675	4.374	1.22E-05
Resident	-0.00942	0.013358	-0.705	0.48051

Asian	0.102988	0.062554	1.646	0.09969
Black	0.069416	0.066279	1.047	0.29496
Hispanic	0.040486	0.053903	0.751	0.4526
Multi-racial	0.087435	0.057073	1.532	0.12553
Native Hawaiian/Pacific Islander	-0.03882	0.069417	-0.559	0.57599
Non-resident Alien	0.032883	0.060349	0.545	0.58583
Ethnicity Unknown	0.109641	0.064142	1.709	0.08739
White	0.138634	0.052899	2.621	0.00878
Junior	0.11531	0.012374	9.319	< 2e-16
Senior	0.166392	0.012408	13.41	< 2e-16
Sophomore	0.049523	0.011719	4.226	2.39E-05
Unknown Generation	-0.10911	0.011604	-9.402	< 2e-16
First Generation	-0.04828	0.009058	-5.33	9.84E-08

Notice that the 'visit' variable is significant (p-value less than 0.05) and positive. This means there is statistical evidence that visiting an advisor in a given semester has a positive impact on students' grades in that semester.

Among the other explanatory variables, GPA, gender, age, and all the class standing and generation dummies are statistically significant. The residency dummy and all the ethnicity dummies, with the exception of 'white', are not significant.

6. Concluding remarks

In this report, we have shown there is evidence that visiting an advisor in a given semester has a positive and significant impact on the student's grade point average in that semester. We reached this conclusion based on data visualization, *t* and median tests, and multiple regression analysis applied to data from the 2018-19 academic year.

However, this study should be seen only as a preliminary investigation of the relationship between academic advising and student success. There are many other aspects of this relationship that need to be studied, which can be accomplished by including other measures of student success and advising experience and using data from other periods. In addition, the results we have obtained so far may be spurious, for the analysis did not account for possible bias in the estimation of the impact of advising on student performance. The next step will be to apply matching and possibly other techniques to the data to try and minimize that bias, given that a randomized trial, the ideal experiment in this setting, is not a viable option.

We hope the results of this study are useful to UVU's Academic Advising and look forward to receiving their comments and suggestions.