12-14-2017

A Statistical Study of Student Success in the BGSU Honors College

Sarah Hercules
sarherc@bgsu.edu

A Statistical Study of Student Success in the BGSU Honors College

Sarah Hercules

Bowling Green State University

James Albert, Advisor

Department of Mathematics & Statistics

Meredith Gilbertson, Advisor

Department of Sociology

Kacee Snyder, Advisor

Honors College

Abstract

Higher education has long tried to find the best measures to predict student success. Different colleges often have different guidelines, requiring different criteria to be evaluated. The BGSU Honors College has struggled with retention and recruitment of underrepresented students with their current admission criteria. This analysis studies different measures of student success such as BGSU GPA and number of completed Honors credits for high-achieving BGSU students who enrolled from Fall 2013 through Fall 2016 to find the best predictors of student success through regression analysis. Throughout this paper, the impact of ethnicity, gender, the college of a student's program, high school activities, and ACT scores on BGSU GPA is discussed. This study reveals that high school GPA is still the best predictor of BGSU GPA, but other variables improve predictions. Somewhat surprisingly, this analysis finds that the college of student's academic program is a significant factor for predicting BGSU GPA and how many semesters a student is enrolled to complete his or her degree. These findings suggest that the Honors College needs to understand how its curriculum helps or hurts students in different colleges to improve student success and retention, as well as its appeal to prospective Honors students. The best model produced used high school GPA, English ACT scores, athletic activities, ethnicity, and a student's college to predict BGSU GPA. This model can be applied to Honors applicants to help estimate how they will perform at BGSU as an Honors student.

Table of Contents

I. Literature Review

Diversity and access have become the forefront of many higher education discussions. Due to pressure from the government and society, institutions are trying to provide equal access to higher education for all students. One issue with increasing diversity and access of minorities and low-income students is that this action can negatively impact factors for national rankings, such as average test scores (U.S. Department of Education [USDOE], 2015). Universities included in these ranking systems may struggle because they are assessed by their performance in these ranking systems, creating conflict. Since states typically examine these average test scores to determine funding levels, lower average test scores then lead to lower funds for these institutions (Boeckenstedt, 2014). Furthermore, the way to approach racial and economic inequalities should depend upon the specific goals and strengths of an individual institution (USDOE, 2015).

The mission of our Honors College as stated on the BGSU website is to challenge members to "identify their values and broaden their worldview." Promoting and welcoming diversity as an avenue to meaningful discussions aligns with this particular goal. Additionally, the Honors College provides many resources and opportunities that could benefit students of all backgrounds and encourage success. Some of these benefits include academic mentorship and advising, priority registration, and discussion-based classes that challenge students to truly apply and understand the material. One must be cautious, however, to provide multicultural inclusion instead of pushing these students to adopt new beliefs and norms for a dominant culture (Rendón, Jalomo, & Nora, 2000). If students feel pressured to adopt the dominant culture, the benefits of diversity are lost.

Before one can consider inclusion within an institution, one must ask how institutions address racial, gender, and socioeconomic diversity at the admissions stage. As Eric Hoover highlights in several articles about the Coalition for Access, Affordability, and Success, some

schools are embracing a new application platform to try to capture minority students; however, challenges are encountered in developing a fair and equal application (Hoover 2015; Hoover 2016). Standard criteria may not find those minority and low-income students who are likely to succeed in higher education, or in Honors.

Missing potential successful students happens even before the evaluation of applications. After talking with various admissions counselors in the BGSU Office of Admissions, a few patterns regarding recruitment became apparent. First, when meeting with students individually or at school visits, most of the counselors look at the student, or school, profiles to determine whether a student meets Honors criteria, meaning 3.5 GPA and 27 ACT (A. Tracy, D. Perkins, D. Small, & B. Kimbrough, personal communication, December 2, 2016). While historically many of our Honors students have met these criteria, they are not concrete cutoffs and could turn potential students away who do not quite meet those scores. Furthermore, relying on those scores or school profiles to advertise Honors leads to a cycle of attracting the same mix of students, limiting diversity. Drew Small and Daryl Perkins also highlighted that many underrepresented students who meet the standard scores are being recruited by other, more selective, schools to improve the diversity at those institutions as well (personal communication, December 2, 2016). In an essay reflecting on the work of Jerry Herron, Carnicom (2013) argues for the use of regression that includes more factors than just GPA and test scores to provide a more holistic glimpse into a perspective student and improve inclusion in honors.

To improve recruitment at the admissions stage, education is key. Some counselors admitted that they are intimidated by the Honors College, so to present that option to a multicultural student who already faces additional hurdles seems like too much (B. Kimbrough, personal communication, December 2, 2016). If we can highlight the experiences and testimonies

of current, and past, students to better shed light on just what the Honors College is, underrepresented students may have a better chance to embrace this opportunity (A. Tracy, D. Perkins, personal communication, December 2, 2016). Are the counselors right? Do test scores and GPA best predict a student's success?

Several studies have been conducted to determine the usefulness of test scores in predicting student success. One study found that the SAT can indirectly predict graduation by first predicting college GPAs (Mattern, Patterson, & Wyatt, 2013). Additionally, a different study found that ACT scores used with high school GPA can accurately classify whether a student will succeed (Radunzel & Noble, 2013). The limitation with this study, however, is that it fails to account for systemically lower test scores and issues with graduation rates for minority and low-income students. To try to better address social and personality differences, Hannon (2014) conducted a study to use not only test scores and GPA as predictors, but also factors like test anxiety and performance-avoidance goals.

Historically, minority and low-income students have earned lower test scores, often limiting access. A study conducted by Geiser and Santelices (2007) supports the argument that high school GPA is a better predictor of student success – both graduation and cumulative GPA – than test scores. They found that GPA was less correlated with socioeconomic factors than were test scores, though Geiser and Santelices still applied weights for socioeconomic factors in their predictive models. Even subjective factors such as participation in extracurricular activities may limit minority students (Selingo, 2016). Often, these students must help at home or work to provide additional income for their families or themselves, limiting time to participate in organizations and service projects.

While some deem higher education an opportunity for upward mobility, others see

perpetuation of poverty and barriers due to student debt. As William Bowen suggests, the issue with student debt is those students who begin college but do not complete their degree, and thus do not reap the full return on their investment (Bowen & McPherson, 2016). So how do institutions improve completion rates? Bowen (2009) argues that some students do not continue because they are "under-matched", lacking the challenge and engagement needed to encourage them to continue. This argument has led me to wonder whether Honors curriculum could provide these students the needed push to increase graduation rates.

II. Introduction

This project has been inspired first by my passion for the power of data visualization and statistics, but also my knowledge of their misuse. When presented an argument, I think critically about the effectiveness of graphs and presentation of statistics. What do the numbers tell and what information about the numbers is missing? Through my work in admissions as a Campus Tour Guide and Scholar Recruitment Student Coordinator, I regularly encounter applicant data and overhear discussions about higher education. These data and discussions inspired the thought to study admissions data, particularly for the Honors College. To narrow my interests, I pulled on the knowledge I gained through my Principles of Sociology and Exploring Cultural Diversity Through Performance courses. My curiosity was sparked by our discussions about systemic barriers due to race, ethnicity, gender, and socioeconomic status. I recognize my privilege and want to make a change to promote more equal opportunities and allow for more meaningful discussions.

The goal of my project is to study Honors application and admissions data, honors retention and completion data, and BGSU retention and graduation data to find ways to improve access for low-income and minority students and to increase diversity within the Honors College. To achieve this goal, I analyze the best factors for predicting student success. (I will further discuss the

specifics of this process in the methodology section.) I evaluate which response variables are most highly correlated with gender and ethnicity, limiting access for minority students. Through plots and statistical analysis, I describe the impact of these correlations to minimize systemic trends of the predictor variables that limit access for minority students. Instead of simply finding students who are most likely to succeed, I compare Honors and Non-Honors students to determine whether a student will benefit from being in an Honors setting.

III. Research Questions

A lack of diversity and access for minority and low-income students has long been an issue in higher education. There are several areas to be analyzed including marketing information, admissions procedures and criteria, student retention, and graduation rates. I asked the following questions to guide my research on the impact of these areas on diversity in the Honors College:

- What are the best measures of student success?
- Which variables are the best predictors of student success?
- What variables specifically limit access of minority and low-income students?
- What correlations exist between cognitive and non-cognitive predictors?
- How does participation in Honors influence student success?
- Is there a point in time that a student becomes more likely to graduate with Honors if they have made it past that point?
- What traits do Honors graduates share?
- How can the application best be evaluated to find the students who are most likely to graduate with Honors?
- How can the Honors application be improved to promote equal access for all students?

IV. Data

My data came from two main sources: the BGSU Data Warehouse and BGSU Admissions. Though the data does not contain all variables I originally desired to study, these secondary sources previously recorded a large amount of the variables I wanted to evaluate, saving time and hassle to record and extract the data myself.

A. Data Warehouse

The Data Warehouse is a database where various data is stored in tables. Its purpose is to allow efficiency for reporting across campus for various projects and analyses. The data stored comes from the BGSU PeopleSoft system, and data is updated nightly. Theresa Sherwood, a Data Warehouse architect, helped organize files for me. All files I received were sent as .csv files, allowing me to easily load the files in R Studio.

Theresa created ten total files, five for students with the Honors College flag, and five for students who are near Honors criteria but do not have the Honors College flag. The Non-Honors students for my study have a high school GPA of at least 3.4 and/or a 25 composite ACT score. Both of these thresholds are slightly below the 3.5 GPA and 27 ACT typically advertised by Honors, but I believe including these students will allow for a more complete analysis of what predicts student success in Honors since those numbers are not strict cutoffs. Each of the five files for Honors and Non-Honors contain the same variables and information. Records include students in the Fall 2013-2017 cohorts. Records are organized by ascending BGSU ID numbers, and the most recent information for a student is listed first in files that include multiple entries per student.

1. Demographics

The first set of files I received are demographic files containing one row per student. The original Honors file includes 1,319 observations for nine variables. There was one outlier I found

with a very low GPA and ACT, which I removed. The variables I kept from these files include BGSU ID, cohort, gender, ethnicity, high school GPA, ACT composite, SAT, a check for duplicates, and BGSU GPA. Additionally, the Fall 2017 cohort observations had to be removed because there is a 0.000 BGSU GPA listed since no GPA has been earned. The Non-Honors demographic file includes 2,495 observations of the same variables. There were two outliers with high school GPA above 5.0 that were removed. These GPAs were likely measured on a different scale and were not properly converted to a 4.0 scale. Finally, I added a flag to identify Honors students when the files are merged.

2. Courses

The second set of files I received are course files. There is one row per student per class in which the student enrolled. The original Honors file includes 31,804 observations of 15 variables. I kept BGSU ID, the Honors Course Flag, Course GPA, and Units Taken Enrolled to compute new variables that generalize the information. These new variables track how many credits (all, Honors, and Non-Honors courses) students have taken as well as the respective GPA for the category of credits. There were 123 subjects included in the file, so trying to track classes of different subject areas was too specific. Additional opportunities like Education Abroad and internships often do not include a Course GPA, so those observations are not included in my tallies. The Non-Honors file includes 62,133 observations of the same variables.

To reduce this file, I tallied the number of credit hours each student has taken. I excluded credits for which no GPA was recorded. Using the Honors Course flag, I separated these tallies into overall credits, Honors credits, and Non-Honors credits. Additionally, I multiplied the GPA for each course by the number of credit hours for the course. I summed these new values for each student, and I divided the GPA value by the credit hours taken by the student, still excluding NA

values for Course GPA from these computations. Using these calculations and the Honors Course Flag, I found the overall GPA for the student, the Honors GPA, and the GPA in Non-Honors courses.

3. Graduates

The third set of files I received are graduate files. There is one row per student per degree earned. The original Honors file includes 198 observations of nine variables. The variables I kept include BGSU ID, terms enrolled, degree code, and degree description. There are 313 observations of the same nine variables for Non-Honors. There were three double majors in Honors and one double major in Non-Honors. I kept only the first observation of these students, as the first observation is what appeared as the primary major in other files. All primary degrees for the double majors were Bachelor of Arts. I also created a flag to distinguish Honors and Non-Honors graduates. My analysis of Graduates is limited because of the Honors graduates, 177 entered with the Fall 2013 cohort and 21 with the Fall 2014 cohort. Of Non-Honors graduates, 278 entered Fall 2013, 30 Fall 2014, and 5 Fall 2015. These numbers mean there are few observations for students who graduate early, and there is not information for students who do not graduate on time.

4. Areas of Study

The fourth set of files I received track students' majors, minors, and specializations. There is one row per student per major, minor, or specialization per semester. The original Honors file includes 9,697 observations of eight variables. The variables I originally kept include BGSU ID, Academic Term, college, and program description. There are too many majors, minors, and specializations to study individually, so I used the college and a tally of how many majors, minors, and specializations a student had in the most recent semester. This most recent entry is the most recent primary major listed for the student. The Non-Honors file has 18,991 observations of the

same eight variables.

To reduce this file, I created a new variable in which I tallied how many majors, minors, and specializations each student had for each semester and called my new variable "focuses." Students have various reasons for majoring versus minoring, so I lumped all majors, minors, and specializations into one tally. Finally, I merged the most recent semester tally with the reduced file and removed the Academic Term, since that variable is no longer needed. Honors and Non-Honors files were then merged.

5. Semester Tracking

The final set of files I received from the Data Warehouse track students each semester. There is one row per student per semester. The original Honors file includes 6,205 observations of 15 variables. The variables I initially kept include BGSU ID, the number of semesters enrolled, Honors flag, and HLC flag. The Non-Honors file contains 12,852 observations of the same 15 variables. To reduce this file, I tallied the number of semesters each BGSU ID had an Honors flag, HLC flag, as well as lived in Founders Hall. I also kept only the most recent value of how many semesters the student has been enrolled, and the county in which they live.

B. Admissions

The information I received from admissions came as one .csv file with one row per student. My supervisor, Erin Heilmeier helped me get the information from Senior Application Specialist, Robert Theis. The information is collected from the BGSU Online Application, for which much is self-reported information. This file contains all students, both undergraduate and graduate, from the Fall 2013-2017 cohorts. There are multiple records for several IDs if they filled out multiple applications for any reason, including if they applied to the Graduate College. Records are organized by ascending BGSU ID numbers, and the oldest information for a student is listed first

in the file. The file originally had 112,709 observations for 147 variables. I initially kept variables for BGSU ID, address, ACT by subject area, high and final high school GPAs, date admitted, housing (commuter or on campus), various activities, and scholarships.

To clean and reduce this file, I first had to adjust the formatting of the BGSU ID number to match the formatting in the other files. I then removed graduate student applications, and removed any duplicate applications, keeping the most recent only per BGSU ID. This work reduced the file to 93,593 observations. Initially, 32 variables were kept from the original file until further analysis was done to reduce them.

C. Master

Since BGSU ID numbers are unique and are included in all files I used BGSU ID numbers to merge my files into one data frame. I began by merging my Honors and Non-Honors demographic files, creating one data frame with 3,814 observations for the nine variables in those files plus the new variable I created to identify Honors students. The SAT variable is missing too many observations to consider in my analysis. One student appeared in both Honors and Non-Honors, so the Honors observation was removed from the master data frame. I then merged my data frame of the course credit tallies and GPAs for Honors and Non-Honors students. There are now 16 variables in the master data frame. Graduate information was added next, adding four new variables. Information about majors for all students was then merged, adding three new variables to master. Next, the reduced admissions file was merged with master by BGSU ID, bringing the number of variables to 54. My final merge added the tallies for Honors, HLC, Founders, terms enrolled, and county, for a total of 59 variables.

To clean the master file, all names of variables were reassigned as simpler, one-word names. Since there are too many categories of activities for easy analysis, band, choir, and jazz

activities were grouped together to be “music”; communication, debate, news, and theatre activities were grouped as “communication activities”; and honors and leadership activities were grouped together. Students in Firelands as well as outliers for high and low high school GPAs were removed. After these adjustments, the master data frame contains 3579 observations for the 62 variables described in table found in the Appendix. Master serves as a base from which to pick variables to include in various analyses. I reduced my observations further to include just cohorts prior to Fall 2017 since Fall 2017 has very incomplete records. To decide which variables to include in my models, I did some exploratory analysis.

V. Exploratory Analysis

A. Honors versus Non-Honors GPA

Since my background research showed that high school GPA tends to be a good predictor, I decided to start there. Figure 1 shows the distribution of high school GPAs for Honors students versus high school GPAs for Non-Honors students. It is easy to see that Honors Students on average do have a higher high school GPA, but this is only to be expected since I requested my Non-Honors students to have only a 3.4 GPA or higher while Honors historically has looked for a 3.5 GPA. For Non-Honors students, 50% of my observations have a high school GPA below 3.8, but only 25% of Honors students have a high school GPA below 3.8. The average high school GPA for Honors students is 3.99, but the average for Non-Honors students is only 3.81. However, Honors GPAs include plenty of low GPA outliers, even below a 3.0, and Non-Honors GPAs have plenty high GPA outliers up to 4.895.

Although Honors students still have a higher average BGSU GPA, the distribution of BGSU GPAs for both Honors and Non-Honors students is heavily skewed left, with many outliers at the low end of the scale as shown in Figure 2. The biggest difference in these distributions is

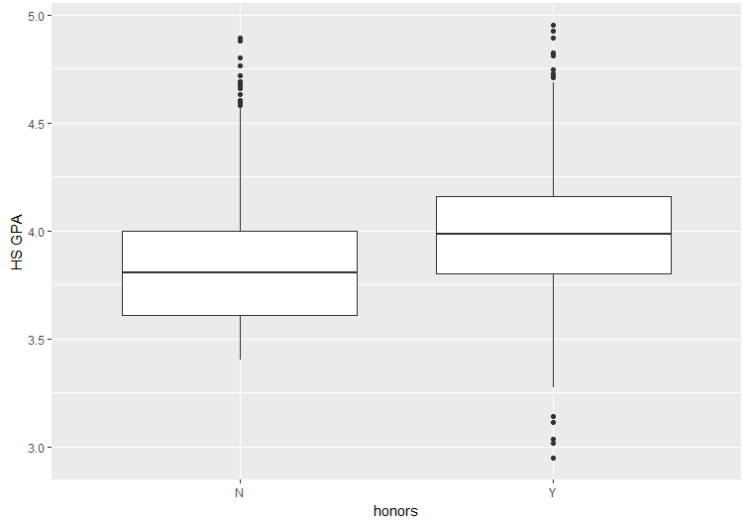

Figure 1: *The distribution of high school GPAs for Honors and Non-Honors BGSU students is graphed in the boxplot above. The Non-Honors GPAs are slightly right skewed, and the middle 50% of scores are between 3.61 and 4.00. The middle 50% of Honors GPAs are between 3.80 and 4.16.*

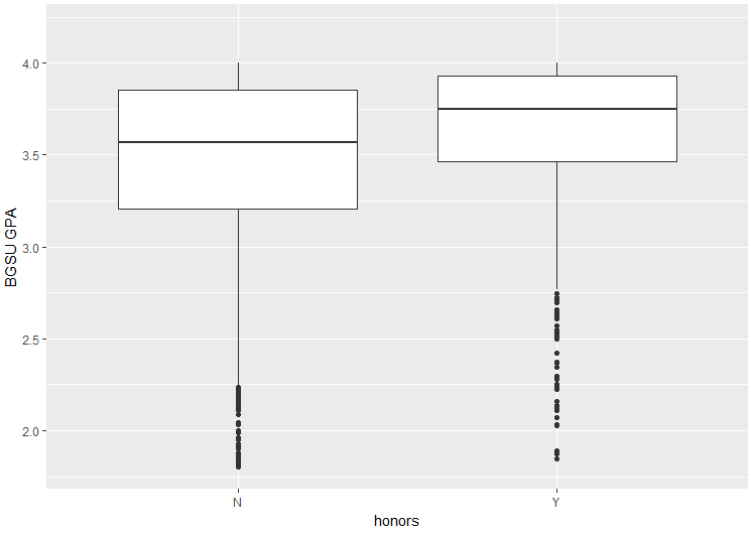

Figure 2: *The distribution of BGSU GPAs for Honors and Non-Honors BGSU students is graphed in the boxplot above. Both GPA distributions are skewed left; however, Honors has a smaller inter-quartile range. The middle 50% of Non-Honors GPAs range 3.15 to 3.84 while the middle 50% of Honors GPAs span only 3.43 to 3.93. Only 50% of Honors students have a GPA below 3.74.*

that Honors students have a smaller range for both the upper 50% and middle 50% of GPAs. While there are low Honors GPAs, the small ranges for the upper quartiles show that more Honors students maintain higher BGSU GPAs than Non-Honors. These findings make sense because Honors students must maintain a 3.5 cumulative GPA to remain in the Honors College.

1. Gender

In addition to differences in Honors versus Non-Honors GPAs, I looked at differences between genders. The distribution of high school GPAs by gender is shown in Figure 3. Females consistently have higher GPAs than males, though Non-Honors students show a greater difference between genders. Non-Honors GPAs are still skewed right. The median high school GPA for Non-Honors males is 3.72 and 3.85 for females. The median GPA for Honors males is 3.93 and 4.00 for females.

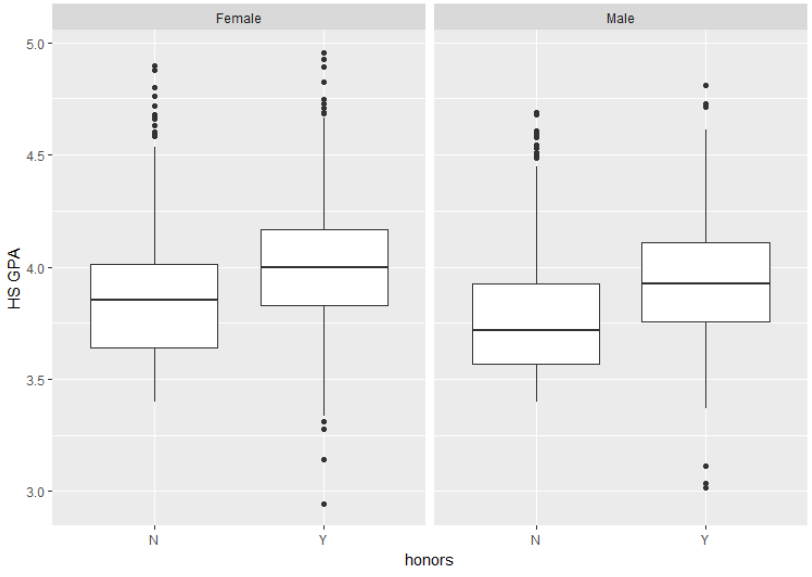

Figure 3: *The distribution of high school GPA separated by gender is shown in the boxplot above. The Non-Honors female median GPA is 0.13 higher than the male median, and the female Honors median is 0.07 higher than the male median.*

Regarding BGSU GPAs, females still tend to have higher GPAs than males. The distribution of BGSU GPAs by gender can be seen in Figure 4. Supporting the discussion in the general Honors versus Non-Honors section, BGSU GPAs for all are still skewed left, and Honors students have a smaller range for the middle 50% of observations. Additionally, the inter-quartile range for females is even smaller than the range for males; however, females have far more low outliers. The median GPA of Non-Honors females is 3.63 while the median for males is only 3.38. The median GPA of females in Honors is 3.77 while the median for males is only 3.64. If GPA is used to measure student success, there is a clear difference between the performance of males and females.

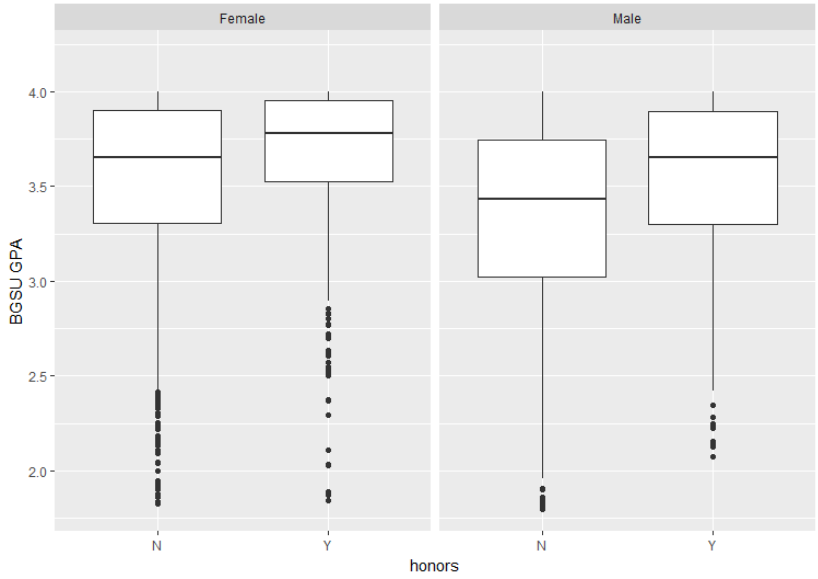

Figure 4: *The boxplot above shows the distribution of BGSU GPAs by gender. Females have both a higher average GPA and smaller inter-quartile range. The median Non-Honors GPA for females is 0.25 higher than the male median and the female Honors median GPA is 0.13 higher than the male median. Both difference are greater than the high school GPA differences.*

2. Cohort

Furthermore, a comparison of the different cohorts suggest that each cohort performs slightly differently. Figure 5 shows the distribution of high school GPAs by cohort. Overall, there is not a great difference between the cohorts, suggesting the students in each cohort began roughly the same regarding high school GPA. Although the high school GPAs have roughly the same distribution, there are larger differences between BGSU GPAs in the different cohorts as shown in Figure 5.

Honors students consistently have a smaller inter-quartile range than Non-Honors students, but the inter-quartile range for the Honors cohorts gets smaller each year as displayed in Figure 6. The shrinking middle range could be due to a variety of factors such as improvements in the Honors College that help students perform better or the range of courses students have taken. That is, the Fall 2013 cohort has likely graduated or at least completed many of their classes for their major while Fall 2014 and 2015 have completed some major classes, and the Fall 2016 cohort has likely mostly taken general courses.

Additionally, though Honors students have a higher median GPA than Non-Honors, certain cohorts have a greater difference. Fall 2014 and Fall 2015 cohorts have a greater difference between Honors and Non-Honors performance, perhaps due to the level of classes being taken. Fall 2013 and Fall 2016 have the smallest differences. Fall 2013 was the first cohort for which Honors was a college, so I am not surprised that there is not as big of a difference in performance, and this cohort has had a full four years at BGSU. Furthermore, Fall 2016 has likely mostly taken general courses, not allowing a chance to differentiate Honors versus Non-Honors performance. These differences suggest that one cohort should not be used to predict other cohorts, and a sample that includes all cohorts will best represent the variety of student stories.

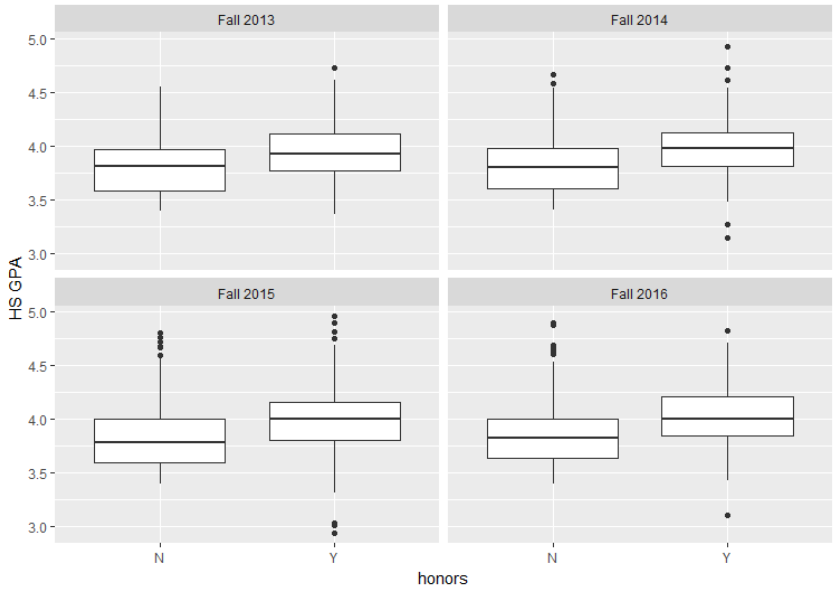

Figure 5: *The distribution of high school GPAs by cohort is displayed in the boxplot above. There do not appear to be major differences in the high school performances of the cohorts.*

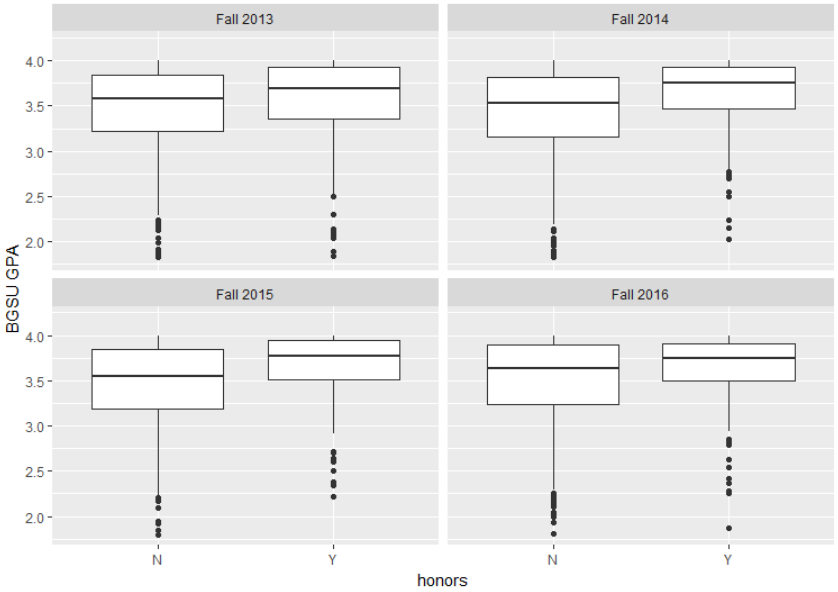

Figure 6: *The boxplot above shows the distribution of BGSU GPAs by cohort. The inter-quartile range of Honors GPAs continues to shrink over time. Honors students consistently have a higher average GPA than Non-Honors, but the difference is greater for Fall 2014 and Fall 2015.*

3. Ethnicity

Finally, I looked at GPA differences by ethnicity as shown in Figure 7. There are no American Indian or Alaskan ethnicities in the Honors group. Most median Non-Honors GPAs for the ethnicities were roughly 3.8, except African American had the highest median at 3.85. Two or more races, Nonresident Alien, and Asian had medians closer to 3.7. Honors students had a different story. Most median Honors GPAs for the ethnicities were roughly 4.0, with Nonresident Alien the highest, only slightly above the others. The Asian median was slightly low at 3.88, but African American had the lowest median with 3.72, ranking below even the Non-Honors median for that ethnicity. Asian Honors students have a slightly bigger inter-quartile range than the others. Additionally, the Asian and African American Honors distributions are skewed right. The Honors distribution for Two or more races is slightly skewed left.

BGSU GPAs do not show the same stories for ethnicities that high school GPAs show. These distributions can be seen in Figure 8. Honors students have a smaller range of middle values than do Non-Honors students for all ethnicities except African American. The Non-Honors Asian distribution is also more left skewed than most of the other Non-Honors distributions. Non-resident aliens have extremely small inter-quartile ranges, but it is difficult to make conclusions due to extremely small sample size. The median GPAs differ quite a bit more than high school GPAs did by ethnicity. The highest Non-Honors median GPA is 3.75 for Nonresident Alien and the lowest median is 3.04 for Asian. Although African American Honors students had lower high school GPAs than did African American Non-Honors students, African American Honors students have a higher BGSU GPA than Non-Honors students. For Honors students, the highest median is 3.83 for Nonresident Alien and the lowest is 3.46 for African American. These differences in average GPA suggest that if high school GPA is used to predict success, different ethnicities treated differently.

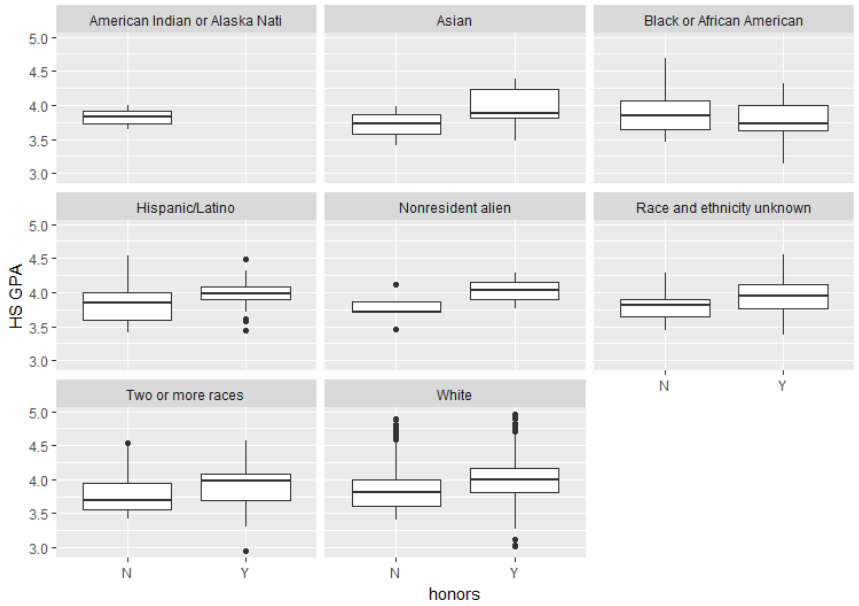

Figure 7: *The boxplot above conveys the differences in high school GPA distribution for different ethnicities. Honors students show higher scores than Non-Honors for all ethnicities except Black or African American.*

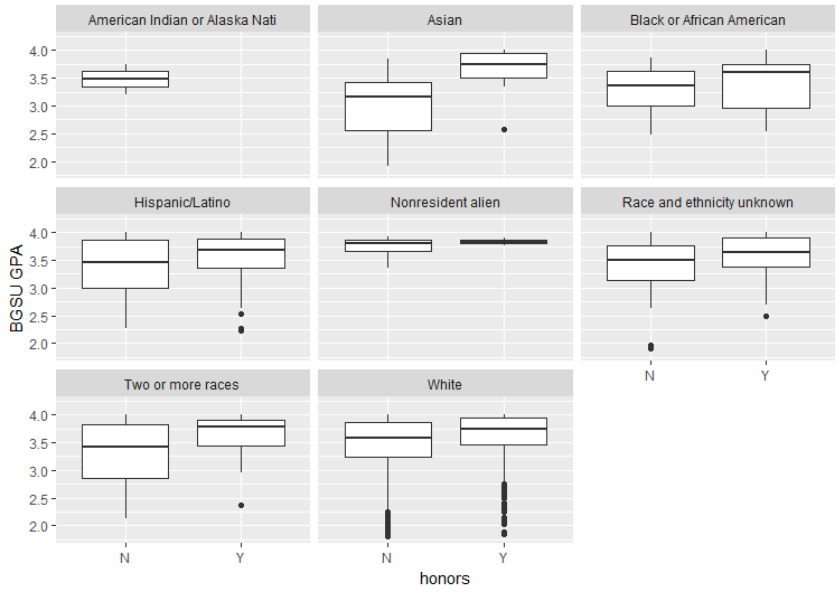

Figure 8: *The boxplot above shows differences in BGSU GPA distribution for different ethnicities. BGSU GPA distributions are slightly more consistent by ethnicity than high school GPAs, but Asian and African American distributions still do not fit the overall pattern.*

4. ACT

To help understand correlations of other variables, I looked at the correlation of ACT test scores, high school GPAs, and BGSU GPAs. Though all scores showed a positive correlation with BGSU GPA, some were more highly correlated than others. The final high school GPA was most highly correlated with BGSU GPA, even above the highest high school GPA. ACT scores had a lower correlation with BGSU GPA than did high school GPA, which was expected. However, I was surprised to find that the English ACT score had the highest positive correlation, even above the composite ACT score. Correlations are summarized in Table 1. Positive correlations show that as high school GPA and English ACT increase, BGSU GPA also increases. Low correlations suggest there are other variables that contribute to movement of BGSU GPA. From these findings, I decided to include gpa_final and act_eng in my full model.

	Final High School GPA	Highest High School GPA	English ACT	Composite ACT
BGSU GPA	0.392	0.366	0.143	0.113

Table 1: *The table above shows correlations of various scores with BGSU GPA. Final high school GPA has the strongest positive correlation and should be selected above highest high school GPA. English ACT has a low positive correlation, but it is higher than Composite.*

5. Honors Credits and Areas of Study

To gauge how academic focus influences BGSU GPA, I considered the number of Honors credits taken and the number of academic focuses. The number of Honors credits is a measure of how involved the student is in the Honors College. Likewise, the number of academic focuses measures how many academic programs the student is involved with at BGSU. Both are positively

correlated with BGSU GPA as can be seen in Table 2. The positive correlations suggest the more academically involved a student is, the better the student will perform. From these observations I have included both focuses and hunits in my full models.

	Honors Credits	Focuses
BGSU GPA	0.197	0.131

Table 2: *The table above shows correlations of Honors credits and academic focuses with BGSU GPA. Positive correlations suggest the more Honors credits and focuses a student has, the better that student performs at BGSU, though the low values show they are not perfectly correlated.*

B. Semesters Enrolled

One question raised during my background research was whether Honors students take more or fewer semesters to graduate. Figure 9 conveys the differences between Honors and Non-Honors students. The number of students was scaled to balance the number of observations. Lower percentages of Honors students graduate in more than eight semesters than Non-Honors students. Additionally, a greater percentage of Honors students graduate in six semesters, or three years, than Non-Honors students. Since my data only includes four full years, any observations above eight semesters mean that students enrolled in summer semesters.

I also wanted to examine the distribution in semesters necessary for graduation in different colleges. These distributions are shown in Figure 10. It is easy to see that the College of Technology has a greater percentage of students who take extra semesters to graduate. More research would have to be done to explain why this is, but I hypothesize that internship requirements for the college play a key role. The College of Arts and Sciences shows a roughly normal distribution, which makes sense due to the wide range of programs. The College of Health

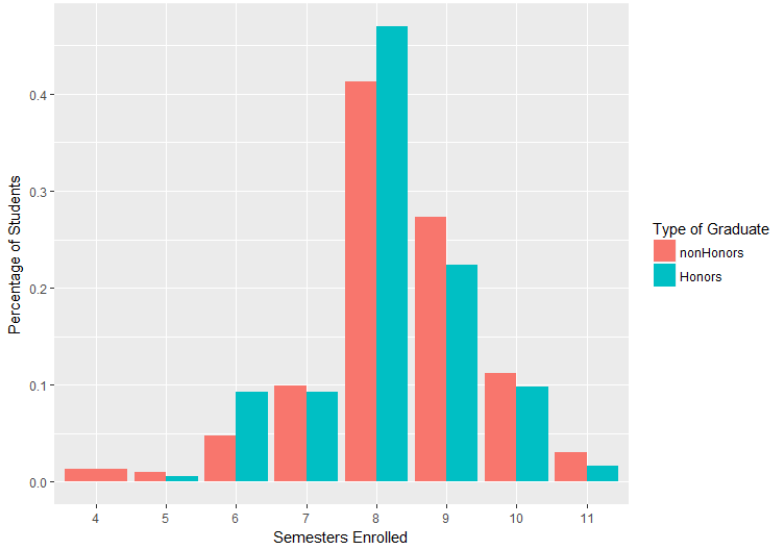

Figure 9: *The percentages of Honors and Non-Honors students who graduate in different numbers of semesters is shown in the boxplot above. Greater percentages of Non-Honors students take more than eight semesters to graduate. More Honors students graduate in six and eight semesters – three or four years.*

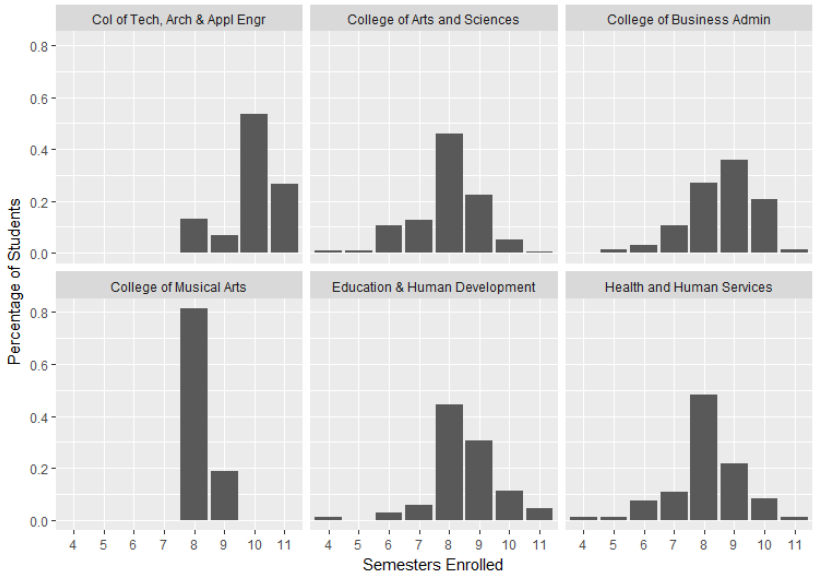

Figure 10: *The distribution of how many semesters a student is enrolled before graduating in each college is shown in the plots above. There is a large variation in the distribution by college. College of Technology has far more students enrolled more than eight semesters than the other colleges.*

and Human Services is also roughly normal. The College of Business is left skewed, with many students taking 8-10 semesters to graduate. The College of Musical Arts has the largest percentage of students graduating on time. The College of Education and Human Development is also skewed with many students taking longer than eight semesters to graduate. Further research needs done to understand why students take different amounts to time to graduate in various colleges.

VI. Methodology

Using multiple regression, I created various models to predict success of students. Multiple regression allows me to consider several variables and how they fit together. It also allows the use of categorical data, like what college the student is in, to predict success. My study uses BGSU undergraduate students from the Fall 2013 through Fall 2016 cohorts. I decided to use two main response variables for the different models as different measures of student success: BGSU GPA and the number of Honors credits. Based on my exploratory analysis, Honors and Non-Honors students were separated to build the BGSU GPA model to see how the model changes for each group. Creating different Honors and Non-Honors models provides an opportunity to explore whether a student performs better as an Honors or Non-Honors student. I considered slightly different collections of inputs for inclusion in the models for each response variable. My models for predicting BGSU GPA considered 11 variables, and my model for predicting the number of honors credits considered 15 variables for inclusion in the model. The response variables and all considered predictor variables are summarized in Tables 3. Full descriptions of the variables can be found in the table in the Appendix. Using these subsets of variables, I created additive models using the forward stepwise method. This method allows me to start with a null model and add one variable at a time that most reduces the error of the model. I utilized R Studio and its packages to

clean, manipulate, summarize, model, visualize, and predict data.

Response Variable	Variables Considered for Inclusion
BG_GPA	gender, ethnicity, college, focuses, act_eng, gpa_final, housing, athlete, music, commActivity, hon_lead
hunits	gender, ethnicity, BG_GPA, hGPA, college, focuses, act_comp, gpa_final, housing, athlete, HLCSem, termsEnroll, music, commActivity, hon_lead

Table 3: *The table above shows each response variable used for the different models and a list of all variables considered for inclusion as predictor variables in the respective models.*

I faced several limitations and challenges in my methodology. The first major issue is the access to relevant data. I am limited to only about four or five years of data, creating very few complete records. Furthermore, there are some factors, like essays and extracurricular participation at BGSU, that I cannot currently quantify to consider in my regression analysis. Additionally, there is no standard for high school GPA determination, so there are limits to the understanding of that number. Furthermore, I do not have any data for AP, College Credit Plus, or Post-Secondary credits and courses. From my conversations with various Honors students and faculty, I believe these credits are extremely important in students' decisions to continue Honors, so not having them in my model limits my analysis. I had to summarize much of my data before I began, so some of the information was lost. My assumptions to measure Honors credits, Honors GPA, and Academic Focuses and to group activities may overgeneralize information, losing key differences in subject areas and double majors or specializations. Regardless of these limitations, I was still able to find some helpful information from my models. My methods for developing each model are

summarized in the following subsections.

A. BGSU GPA

I built two models to predict BGSU GPA, one for Honors students and one for Non-Honors students. The overall process was extremely similar for building each model, so the process is generalized below. Any differences are noted in the respective section.

1. What data was used?

I created a data frame that includes all students who enrolled before Fall 2017 with the variables mentioned in Table 3. I then removed any observations with missing values. Next, a random sample of half of the data was taken for use as a training set with which to build models. The other half was stored as a test set to be used for model validation. To separate Honors and Non-Honors students, the training set was filtered by the honors flag to create an Honors training set and a Non-Honors training set. The test set was filtered in the same way. My Honors training set has 456 observations, and the test set has 478. The training set for Non-Honors has 917 observations and the test set has 896.

2. What inputs were used?

The following variables, described in the table in the Appendix, were considered for inputs: gender, ethnicity, college, focuses, act_eng, gpa_final, housing, athlete, music, commActivity, and hon_lead. My Honors model includes gpa_final, act_eng, athlete, ethnicity, and college as significant predictors. The Non-Honors model includes gpa_final, college, focuses, gender, and ethnicity as the most significant predictors. The models are similar, but the different variables selected show some variables are better at predicting Honors student success and others are more important for predicting Non-Honors student success.

3. What was the process of choosing the model?

To select my model, I used the stepAIC() function in the MASS package in R Studio. Using the "forward" direction, this function begins with a model that includes no predictor variables. In the first step, the function creates one-variable models with each different predictor and keeps the variable for which the model fit is most improved. Model fit in this case is measured by AIC, and the best model will have the smallest AIC. Additional steps add one variable at a time to the model by the same process as long as the model fit is improved. I limited my steps to four so that my model was not too complex to understand.

4. How were models validated?

I used the predicted values from the model and the actual values of the response in my training set to compute the Mean Square Error (MSE). I used the model to make predictions for the test set, and I computed the MSE for the test set. I compared the two MSEs to check the fit of the model.

B. Honors Credits

1. What data was used?

I created a data frame that includes all students who graduated from BGSU who enrolled Fall 2013 or after with the variables mentioned in Table 3. I then removed any observations with missing values. Next, a random sample of half of the data was taken for use as a training set with which to build models. The other half was stored as a test set to be used for model validation. I filtered the training set for observations with at least one for both hSem and hunits to examine just students who at one point were Honors students. I applied the same filter to the test set. My filtered training set includes 86 observations, and the test set includes 89 observations. This sample is significantly smaller than the samples used for my BGSU GPA models.

2. What inputs were used?

The following variables, described in the table in the Appendix, were considered for inputs: gender, ethnicity, BG_GPA, hGPA, college, focuses, act_eng, gpa_final, housing, athlete, HLCSem, termsEnroll, music, commActivity, and hon_lead. My model kept hGPA, focuses, and act_comp as the significant predictors.

3. What was the process of choosing the model?

The process for selecting my model was the same as was used for the BGSU GPA models. Details of the process can be found in section 3 of the BGSU GPA section of methodology. Gender and termsEnroll were originally included in the stepwise model, but they were not significant, so I removed them.

4. How were models validated?

I used the predicted values from the model and the actual values of the response in my training set to compute the Mean Square Error (MSE). I used the model to make predictions for the test set, and I computed the MSE for the test set. I compared the two MSEs to check the fit of the model.

VII. Findings

A. BGSU GPA

1. Honors

My Honors model selected gpa_final, act_eng, athlete, ethnicity, and college as inputs. Since ethnicity and college are categorical variables, dummy variables were added that are stored as 1 if the student is that ethnicity or 0 if not. The null entries are Asian for ethnicity and Academic Enhancement (Undecided) for college. The following is my equation for the model:

BG GPA = -0.46 + 0.73*(high school GPA) + 0.02*(English ACT) + 0.03*(# of athletic activities)
+ 0.00(if Asian) – 0.06(if African American) – 0.06(if Hispanic)
+ 0.67(if Nonresident alien) + 0.31(if Unknown) + 0.35*(if 2 or more) + 0.36(if White)
+ 0.00(if Undecided) + 0.09(if Technology) + 0.19(if Arts & Science) + 0.27(if Business)
+ 0.08(if Musical Arts) + 0.30(if Education) + 0.36(if Health & Human Services)

The model has a R^2 statistic of 0.2785, meaning the model explains about 28% of the variance in the response. While this is a relatively low statistic, it is somewhere to start since there are so many variables that play into student success. The mean square error (MSE) for the training set is 0.189 and the MSE for the test set is 0.332. For a model with just gpa_final as an input, the training MSE is 0.208 and the test MSE is 0.328. These show that the model with additional variables improves predictions for the training set, but the test set error is slightly worse for the bigger model. It may not be worth adding complexity of new variables, but until more tests are performed with new data, we will continue to consider the complex model. These errors are extremely low, but that is likely due to the small range of GPA values. The increase in the test MSE is expected, but it is not enough to be concerned about overfitting.

The first value in the equation is just the intercept to start. Since high school GPAs do not drastically vary, the coefficient can better be understood that for each 0.1-point change in high school GPA, the student's BG GPA is predicted to change 0.073 points in the same direction when the remaining inputs are held constant. Since ACT scores are much higher than GPA values, the coefficient for ACT is only 0.02. That means the ACT score is multiplied by 0.02 and added to the other variables to predict BGSU GPA. Figure 11 conveys the 0.02 increase in BGSU GPA if a student has a one-point higher ACT score while all else is held constant. Similar to English ACT, athletic activities do not have a large influence. For each athletic activity a student participated in

in high school, 0.03 is added to the predicted BGSU GPA. The shift for an increase in athletic activities when other inputs are fixed is displayed in Figure 12.

To understand ethnicities, the base prediction assumes Asian. The coefficients for the other ethnicities convey how much higher or lower to expect a BGSU GPA compared to an Asian student for whom all other predictor variables are fixed. The trend lines for the different ethnicities are graphed in Figure 13. Nonresident Alien is clearly predicted to perform better than other ethnicities; however, there is a very small number of Nonresident Aliens in the sample. Unknown, Two or more races, and White all are predicted roughly the same. There is a group projected lower that includes African American, Asian, and Hispanic.

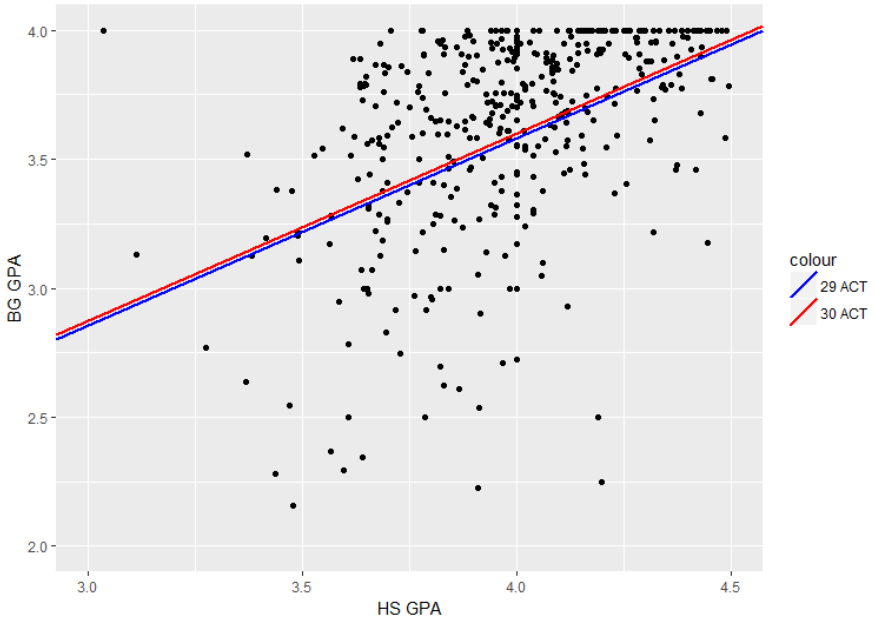

Figure 11: *The graph above shows the impact of English ACT on BGSU GPA when other inputs are held constant. The slope of the line is the impact of high school GPA, and the red line shows the increase in intercept for a one-point increase in English ACT score when all else is constant.*

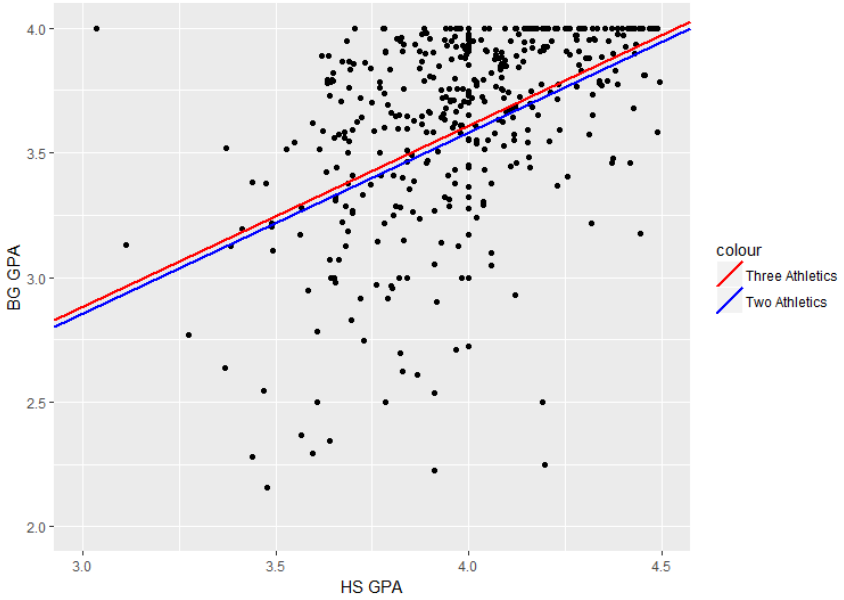

Figure 12: *The graph above shows the impact high school athletics on BGSU GPA when other inputs are held constant. The slope of the line is the impact of high school GPA, and the red line shows the increase in intercept for one additional athletic activity when all else is constant.*

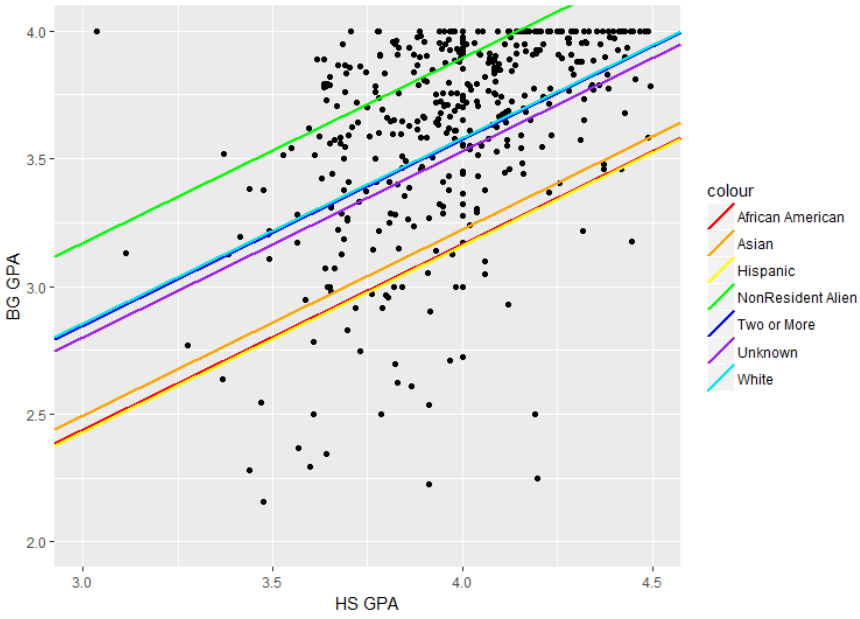

Figure 13: *The graph above shows the impact of ethnicity on BGSU GPA when other inputs are held constant. The slope of the line is the impact of high school GPA, and the colors show shift in intercept for different ethnicities. Three groups of ethnicities emerge.*

Understanding the college coefficients is similar to understanding those for ethnicity. The base value for college is Undecided. The coefficients for each college show how much higher to expect a BGSU GPA for a student in that college versus an Undecided student for whom all other inputs are constant. Differences in college are shown in Figure 14. Unlike ethnicities, no clear groups emerge. Undecided students are predicted to have lower BGSU GPAs, and the College of Health and Human Services is predicted to have the highest BGSU GPAs. The College of Musical Arts and College of Technology have similar influence as do the College of Business and College of Education and Human Development; however, each college appears to have a unique influence.

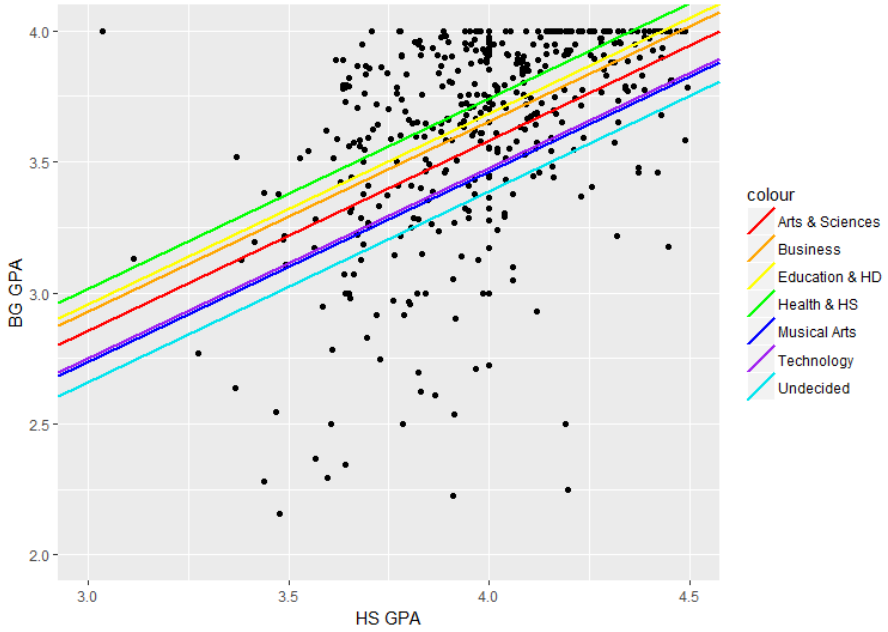

Figure 14: *The graph above shows the impact of college on BGSU GPA when other inputs are held constant. The slope of the line is the impact of high school GPA, and the colors show shift in intercept for different colleges. No groups emerge. Each college is unique.*

The Honors college sample does have lower percentage of students in Business and Education compared to the Non-Honors sample; however, there are a far higher percentage of Arts and Science students in the Honors sample than in the Non-Honors sample. This suggests analysis needs done to understand why the Honors College appeals to certain colleges more than others.

2. Non-Honors

My Non-Honors model selected gpa_final, college, focuses, gender, and ethnicity as significant inputs. The equation for this model is the following:

BG GPA = -.65 + 0.73*(high school GPA) + 0.00(if Undecidedd) + 0.27(if Technology)
+ 0.04(if Arts & Sciences) + 0.22(if Business) + 0.30(if Musical Arts) + 0.40(if Education)
+ 0.34(if Health & Human Services) + 0.16(# of focuses) – 0.16(if Male) + 0.00(if Asian)
+ 0.82(if African American) + 0.67(if Hispanic) + 1.00(if Nonresident alien)
+ 0.90(if Unknown) + 0.59(if 2 or more) + 0.85(if White)

The model has a R^2 statistic of 0.2048, meaning the model explains about 20% of the variance in the response, worse than the Honors only model. The mean square error (MSE) for the training set is 0.370 and the MSE for the test set is 0.375. For a model using just gpa_final as an input, the test MSE is 0.408. The larger error suggests the model with additional variables better predicts BGSU GPA than the simple model. These errors are still low, and there is even less of an increase in the test MSE than in the Honors model, so I am not concerned about overfitting.

Both the Honors and Non-Honors models have similar weights for the high school GPA; however, the distribution of weights for ethnicities and colleges change. The college of Technology and College of Musical Arts have low coefficients for the Honors model, but they are more in the middle range for Non-Honors. Additionally, College of Arts and Sciences has a far lower weight in the Non-Honors model than the Honors model. Perhaps programs in the various

colleges align better with the Honors curriculum to help students succeed. For ethnicities, the biggest differences can be seen for African American and Hispanic categories. In the Honors model, these ethnicities had a negative coefficient, suggesting a lower expected BGSU GPA than the null category, or Asian for a student for whom all else is constant; however, African American and Hispanic both have high, positive coefficients in the Non-Honors model. This disparity suggests that African American and Hispanic Non-Honors students perform better than Asian Non-Honors students with other inputs being fixed; however, Asian Honors students tend to perform better than African American and Hispanic Honors students. The difference in model predictions for a Non-Honors student versus an Honors student with fixed inputs is showed in Figure 15. If the student is an Honors student, they are predicted to have better success when success is measured as BGSU GPA.

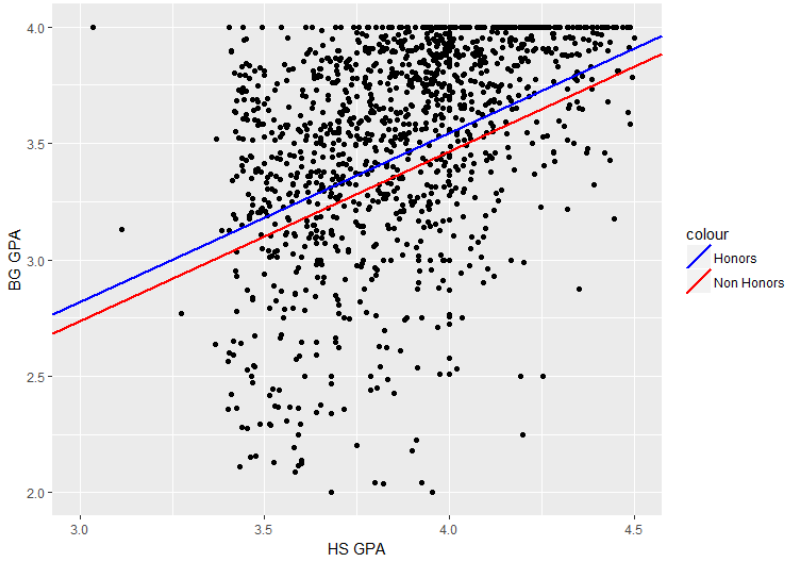

Figure 15: *The graph above shows the difference in predictions from the Honors and Non-Honors model when inputs are held constant. The slope of the line is the impact of high school GPA, and the intercepts are based on an average student. The blue line shows the increase in intercept for BGSU GPA predicted by the Honors model when all else is constant.*

B. Honors Credits

My model to predict the number of completed Honors credits kept hGPA, focuses, and act_comp as inputs. The MSE for the training set is 39.70 and 55.62 for the test set. These errors are very high, suggesting the model is not a good fit. Furthermore, the test MSE for a model that just uses hGPA as the input is 54.40. While this value is close to the other test MSE, the smaller value suggests the simple model may be sufficient for prediction. The R^2 statistic is 0.2419, so it is roughly in the same range as my prior models. Figure 16 shows the actual versus predicted values for the number of Honors credits. The equation for this model is the following:

Hcredits = -19.67 + 3.83*(Honors GPA) + 2.63*(# of focuses) + 0.7117*(Composite ACT score).

This model is difficult to interpret due to the low intercept. Regardless, it is apparent that an increase in Honors GPA, focuses, or Composite ACT will increase BGSU GPA. If all else is held constant, a 0.1 increase in Honors GPA will increase the expected Honors credits by 0.383. Furthermore, an increase of one for the number of focuses will increase expected Honors credits by 2.63 when other inputs are fixed. Finally, a one-point increase in Composite ACT score increased the predicted Honors credits by 0.7117.

After studying the various models, I created, I believe the Honors model to predict BGSU GPA is the best model to predict student success in Honors. Although there is still much unexplained variation, this model uses a select few variables to help explain about 28% of the variation. Additionally, the error was relatively small compared to other models.

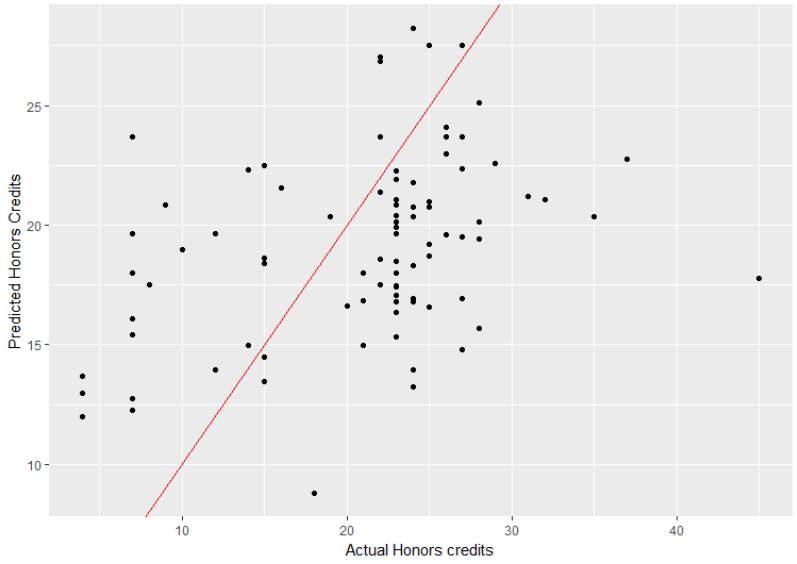

Figure 16: *The graph above shows the actual number of Honors credits versus the predicted number of credits from the model. The red line has a slope of 1 to visualize agreement. The points are evenly split, but there is a large amount of error.*

VIII. Conclusion

My research shows that there is a strong relationship between high school GPA and BGSU GPA. This knowledge can help predict how a student will perform in college; however, I also found that high school GPA tends to vary by gender and ethnicity. These variations suggest that high school GPA should not be studied alone because there are systemic trends in how students perform. My research suggests further studies should be done to determine why females and certain ethnicities have higher GPAs than others. Furthermore, what college a student is enrolled in also influences predicted success, regardless of whether the student is in Honors or Non-Honors. In general, Honors students do perform better than Non-Honors students.

Overall, there is still a lot of work to be done. My analysis cleaned and organized data, paving the way for future studies. As more data is collected, better models can be built; however,

there will always be variation that cannot be explained by the model. There is no perfect way to predict whether a student will succeed in Honors, but by continuing to study data to understand the reasons why students succeed or not, the Honors College can improve marketing to recruit students who will be retained and succeed in Honors.

A. Limitations and Suggestions for the Future

I ran into several issues with my data and my analysis. First, with only four to five years of data, it is difficult to get the big picture of what is happening. As I showed earlier, each cohort perform slightly differently, so you cannot rely on information from one cohort to explain another. With having so few records for graduates since only one cohort has been enrolled for a full four years, I could not predict graduation with Honors like I originally intended. Additionally, there is minimal information for students who do not graduate on time since the extra year for my first cohort would be in process now. The distribution of when students take Honors credits is typically skewed, where far more credits are taken in early semesters. With only having a few completed records, it is nearly impossible to analyze when students drop Honors or predict whether they will complete the curriculum. As more data is collected, studies can be done to track trends for when students lose the Honors flag.

Additionally, I have no information about AP credits or courses, College Credit Plus, or Post-Secondary records. From conversations with Honors students and faculty, I believe these credits play a large role in whether a student decides to join the Honors College and stick with it. For many students, they decide it is too difficult to earn Honors credits since they are past general courses and drop Honors. Information about these credits should be collected to be analyzed with the rest of the information I gathered. Furthermore, I have no additional information from the Honors application. One of the biggest components is the essay, and it plays a large role in the

decision to admit students into Honors. If a measurement is created to score essays, perhaps a 0-10 scale, the essay could be considered in future models to make them more complete.

My models found the student activities in high school played a role in student success, but I do not have data for involvement at BGSU. Other than the Honors Learning Community, I did not have any information to gauge how engaged a student is in the community at BGSU. Do activities increase retention? Do activities lower GPA due to conflicting interests? I believe these questions are worth exploring if data can be collected to track involvement at BGSU. My models also found that demographic information like gender and ethnicity play a role in success, but I do not have any information for socioeconomic status. Furthermore, I do not have scholarship and financial aid information to see how those motivate a student. The only scholarship information I had was determined by GPA and test scores, so it was already indirectly considered in my models.

The data itself is also difficult to use to predict success for a prospective student. Much of the information I have is information collected once a student is enrolled at BGSU. Overlap between Honors and Non-Honors students also makes it difficult to distinguish whether a student would perform better inside or outside of the Honors College. My data was pulled from Fall cohorts, so I do not have information for students who enrolled in the spring Semester or transferred, also limiting my analysis.

After finding how large of a role the college plays in predicting student success, I believe the Honors needs to have a conversation with the various Colleges as well as students from different Colleges to understand what is going on. Until research is done to understand how Honors curriculum works with, or against, curriculum for different colleges, it is difficult to explain the differences to improve student success.

IX. References

Boeckenstedt, J. (2014, December 17). Why the Admissions Office May Be Part of the Problem of College Access. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/Why-the-Admissions-Office-May/150883

Bowen, W. G., & Rudenstine, N.L. (2003, February 7). Race-Sensitive Admissions: Back to Basics. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/Race-Sensitive-Admissions-/22962?cid=cp59

Bowen, W. G., Chignos, M.M., & McPherson, M.S. (2009, September 8). Helping Students Finish the 4-Year Run. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/Helping-Students-Finish-the/48329?cid=cp59

Bowen, W. G., & McPherson, M.S. (2016, July 22). The Student Loan Debt Crisis Is Overblown. The Real Problem Is College Completion Rates. *Vox*. Retrieved from http://www.vox.com/2016/7/22/12254046/myths-higher-education-crisis-debt-loans-free-tuition

Carnicom, S. (2013). Predicting Student Success, Ameliorating Risk, and Guarding against Homogeneity in Honors. *Journal Of The National Collegiate Honors Council*, 14(2), 35-39.

Geiser, S., Santelices, M. V., & University of California, B. E. (2007). Validity of High-School Grades in Predicting Student Success beyond the Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes. Research & Occasional Paper Series: CSHE.6.07. *Center For Studies In Higher Education.*

Hannon, B. (2014). Predicting College Success: The Relative Contributions of Five Social/Personality Factors, Five Cognitive/Learning Factors and SAT Scores. *Journal Of Education And Training Studies*, 2(4), 46-58.

Hoover, E., (2015, October 2). New College-Application Site Aims to Capture Traits of Success - Like Grit and Engagement. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/New-College-Application-Site/233567

Hoover, E., (2016, August 3). 'Coalition' Leader Wants New Application to Serve the Underdog in Admissions. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/Coalition-Leader-Wants/237341?cid=cp50

Mattern, K. D., Patterson, B. F., & Wyatt, J. N. (2013). How Useful Are Traditional Admission Measures in Predicting Graduation within Four Years? Research Report 2013-1. *College Board*.

Neter, J., Kutner, M. H., & Nachtsheim, C. J. (2004). *Applied linear regression* (4th ed.). New York, NY: McGraw-Hill/Irwin.

Radunzel, J., & Noble, J. (2013). Differential Effects on Student Demographic Groups of Using ACT® College Readiness Assessment Composite Score, Act Benchmarks, and High School Grade Point Average for Predicting Long-Term College Success through Degree Completion. ACT Research Report Series, 2013 (5). *ACT, Inc*.

Rendón, L. I., Jalomo, R. E., & Nora, A. (2000). Part II: New Theoretical Directions: Theoretical Considerations in the Study of Minority Student Retention in Higher Education. In , *Reworking the Student Departure Puzzle* (pp. 125-156). Vanderbilt University Press.

Selingo, J. J., (2016, October 22). The Lasting Impact of William Bowen. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/The-Legacy-of-William-Bowen/238155

U.S. Department of Education. (2015, September). *Better Information for Better College Choice & Institutional Performance* [PDF File]. Retrieved from https://collegescorecard.ed.gov/assets/BetterInformationForBetterCollegeChoiceAndInstitutionalPerformance.pdf

U.S. Department of Education. (n.d.). Overview of IPEDS Data. *Institute of Education Sciences, National Center for Education Statistics*. Retrieved from http://nces.ed.gov/ipeds/InsidePages/UseTheData?pageid=58

X. Appendix

Explanation of Variables

	Variable Name	Description
1	ID	Unique 10-digit BGSU ID number for merging as character
2	Cohort	Numeric representation of when student first enrolled
3	Gender	Male or Female as character
4	Ethnicity	Official ethnicity in University system; 8 ethnicities
5	HS_GPA	Numeric high school GPA
6	ACT_comp	Numeric composite score
7	SAT	Numeric total score, but many NAs
8	Duplicate	Numeric tracker used to check duplicates to create files
9	BG_GPA	Numeric cumulative BGSU ID
10	Honors	Character flag of "Y" or "N" to identify Honors students
11	Tunits	Numeric total credits for which student has earned GPA points
12	Hunits	Numeric total Honors credits for which student has earned GPA
13	Nhunits	Numeric total Non-Honors credits for which student has earned GPA
14	tGPA	Cumulative GPA for all credits in which there was a GPA earned
15	hGPA	Cumulative GPA for all Honors credits
16	nhGPA	Cumulative GPA for all Non-Honors credits
17	semEnroll	Numeric semesters graduates enrolled at BGSU
18	Degree	Degree earned by graduates
19	degreeDes	Written out degree earned by graduates
20	Grad	Numeric flag to identify grads: 0 = unfinished, 1 = Non-Honors, 2 = Honors

21	College	Character; 8 different Colleges included
22	Major	Description of major
23	Focuses	Numeric tally of majors, minors, and specializations
24	Address1	Character street address
25	Address2	Character apartment numbers or other address information
26	City	Character city of permanent address
27	State	Character state of permanent address
28	Zip	Character zip code
29	Country	Character country of permanent residence
30	Act_eng	Numeric English ACT score from Admissions
31	Act_math	Numeric Math ACT score from Admissions
32	Act_read	Numeric Reading ACT score from Admissions
33	Act_sci	Numeric Science ACT score from Admissions
34	Act_comp	Numeric Composite ACT score from Admissions
35	Gpa_high	Numeric Highest high school GPA recorded by Admissions
36	Gpa_final	Numeric Final high school GPA recorded by Admissions
37	Admit_date	Character date student was admitted
38	Academic_program	Character college student enrolled into by Admissions
39	Housing	Character “C” or “H” for commuter or on campus Housing
40	Athlete	Numeric number of athletic activities in high school
41	Band	Numeric number of band activities in high school
42	Choir	Numeric number of choir activities in high school
43	Communication	Numeric number of communication activities in high school

44	Debate	Numeric number of debate activities in high school
45	Honors_activity	Numeric number of honors activities in high school
46	Jazz	Numeric number of jazz activities in high school
47	Leadership	Numeric number of leadership activities in high school
48	News	Numeric number of news activities in high school
49	Other	Number of other activities in high school
50	Rotc	Number of ROTC activities in high school
51	Theatre	Number of theatre activities in high school
52	numEthnicity	Number of ethnicities listed on application
53	Scholarship	Character scholarship names awarded by university
54	Scholarship_outstate	Character out of state scholarship offer by university
55	hSem	Numeric number of semesters as Honors student
56	HLCSem	Numeric number of semesters in Honors Learning Community
57	FoundSem	Numeric number of semesters living in Founders
58	termsEnroll	Numeric number of semesters student has been enrolled at BGSU
59	County	Character county of permanent residence
60	Music	Numeric sum of band, choir, and jazz activities
61	commActivity	Numeric sum of communication, debate, news, and theatre activities
62	Hon_lead	Numeric sum of honors and leadership activities

The table above lists all variables originally included in the master data frame. The description explains the structure of the variable and what it conveys. Subsets of these variables were studied and used in predictive models and exploratory analysis.