Statistics 133 Final Paper: Cal Answers Data Exploration

Introduction

Cal Answers is a tool created by UC Berkeley that tracks student and faculty data. Data is stored in "Dashboards" that are specific to a topic of interest. In a Dashboard, users can select variables and see tables, charts, and graphs based on the varaibles that they selected. The amount of data on Cal Answers is vast, but our research is focused on a few specific Dashboards. We investigated the ethnic makeup of applicants and students, GPA averages by department and semester, and time taken for students to graduate.

With the data from Cal Answers, we can see various combinations of demographic data, education data, and class data over a long period of time. By examining these comibinations, we are able to analyze what changes in the student body are occurring over time. We can also investigate what effects economic, social, and policy changes have on the Berkeley campus. Understanding these effects provides insight into what causes changes in student behavior. This would be important information for the administration to help them create good policies and provide the best experience for Berkeley students. It is also interesting just to see what trends occur in the student body and what causes these trends.

Grade Point Averages of Departments

Among faculty and students, there is a competitive spirit in the conversation of who has the toughest department or classes. There is great honor knowing that one is the taking courses in toughest department at UC Berkeley. Typically, there is not much agreement on which department is the hardest to take courses in because of bias, so this paper will try to find the true answer.

First, we took 18 semesters of grades starting in undergraduate courses from the 2010 school year to the present (2010 summer semester to 2016 spring semester). Our GPA calculation deviates from the normal Berkeley GPA calculation–instead, we counted A+ as 4.3 grade points. Our reasoning is that A+ is the hardest grade to get; in fact, out of 1189583 letter grades only 61124 people got A+'s, which only is five percent of that population! In addition, a course that is consistently giving out A+'s to students is definitely easier than a course than gives out as many A's. So, we found the average GPA by department, and the twenty departments with the highest GPA's came out as one might predict. 70% of this category is made up by language departments like Filipino, Tamil, Turkish, Korean, Vietnamese, etc. However, does this near-uniformity in subject hold for the twenty lowest GPA's?

Figure 1 shows that Biology, Statistics, Mathematics, and Physics are within the five departments with the lowest GPA's. It may be because they all are very popular departments and their lower division courses are 'weeder' courses. Courses like Biology 1A, Math 1B, Physics 7A, and Statistics 20 are required courses for many majors outside of the courses' respective departments.

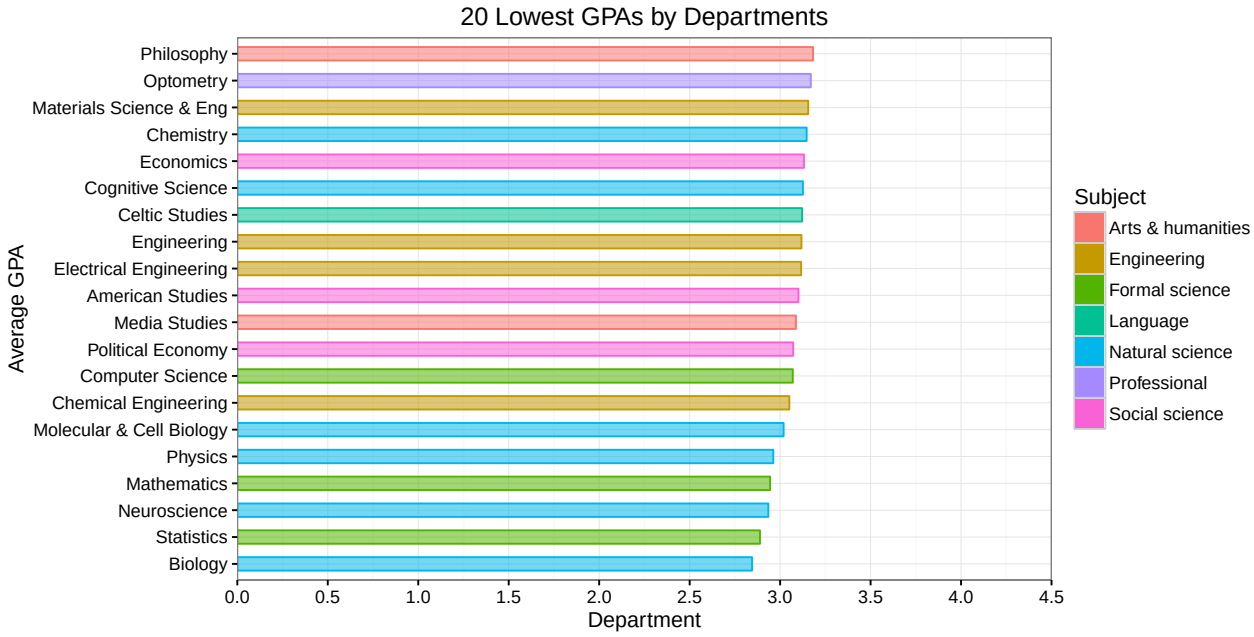

Figure 1: Each subject has a near-equal representation in the lowest 20

Another topic we wanted to study was the the differences in difficulty for regular-semester courses (spring and fall semesters) and summer courses. So, we found the average GPA for courses in the summer and in the regular school year over the period 2010 to 2015. Then, we computed the absolute difference of courses offered in the regular school year and in the summer. We found that there was an average of 0.21 GPA difference and that courses in the school year had a higher GPA by 0.04. Next, we were interested in the department that might be the most beneficial to take courses in the school year. Demography had the highest difference, a 0.72 difference, almost one letter grade! However, there is only one class in that department that is both offered in the summer and the regular school year. If we look for departments that offer more than five courses both in the summer and the regular school year, we find that Architecture has nine courses offered in both periods of time, which affects. In Architecture, the regular school year had a higher GPA by 0.56, which is half a letter grade, and this affect 2413 students. It would definitely be advisable to take Architecture courses during the school year then!

Finally, we wanted to analyze grade trends with respect to changes in departmental policy. The most well-known policy change is the GPA cap for Computer Science majors. The program has had both GPA thresholds and numerical limits on how many majors Berkeley could accommodate, but the departmented transitioned away from student limits in favor of the more transparent GPA threshold. In a conversation with Chris Hunn, the Computer Science major advisor, we learned that the GPA requirement has varied from 2.0 to 3.5+, with the late 80's and late 90's being the most competitive. Set more than 20 years ago, the historical record was ~380 majors, which was still the highest capacity up to 2012. After the dotcom bubble burst, numbers waned and the GPA was adjusted accordingly until it was removed altogether in 2007. UC Berkeley has 1,100+ majors now and still haven't reached the highest GPA requirement set in the past. From 2010 to the

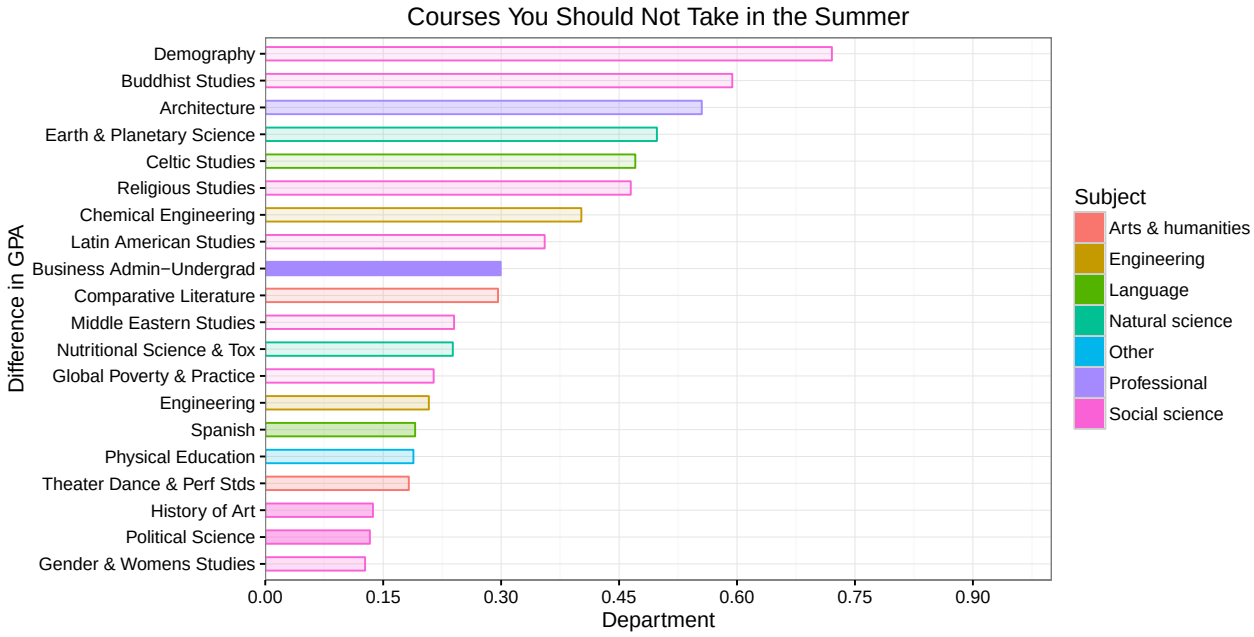

Figure 2: 20 Highest Differences between Summer/Regular Semester

present, Computer Science has been the fourth most popular department for students, and in spring 2014 to fall 2015 the 3.0 GPA cap was implemented. We decided to study this GPA cap rather than the most recent bump to 3.3 GPA in spring 2016 due to availability of data.

We looked at only the upper division CS courses, since there are many people in the lower division courses who do not get into the CS major, as well as people who take the classes for fun. In the whole dataset, there are more people who take CS lower division courses than CS upper division courses and the GPA is higher by 0.11 for upper division courses, indicating that the smarter students were accepted into the major or that the upper division courses are easier (which is probably not the case from my experience). From figure 3, there is an increase in CS course enrollment. Comparing fall 2012 and spring 2016, the enrollment count doubled and the gpa stayed relatively the same. In addition the GPA average took a drop, and it is climbing up and looks to increase even more. It does seem that the CS department is dealing with the high increase of students very well; the GPA's are consistent from before the heavy increase of students. One may think that GPA would be lower since there are a finite amount of resources for teaching students and faculty size hasn't really changed much in the CS department to accomodate for more students.In fact, a professor in the last offering of CS61A (the introduction course in CS, which has the highest enrollment of all CS courses) advised students to study indepedently by watching webcasts and to not come to lecture or lab physically sections since there were not enough resources for each individual student. Although this might hinder student performance, the GPA, from the past school year, for the most important CS upper division courses like 162 and 170 has never been higher. All in all, the data supports the idea that the CS department is handling the depletion of resources well.

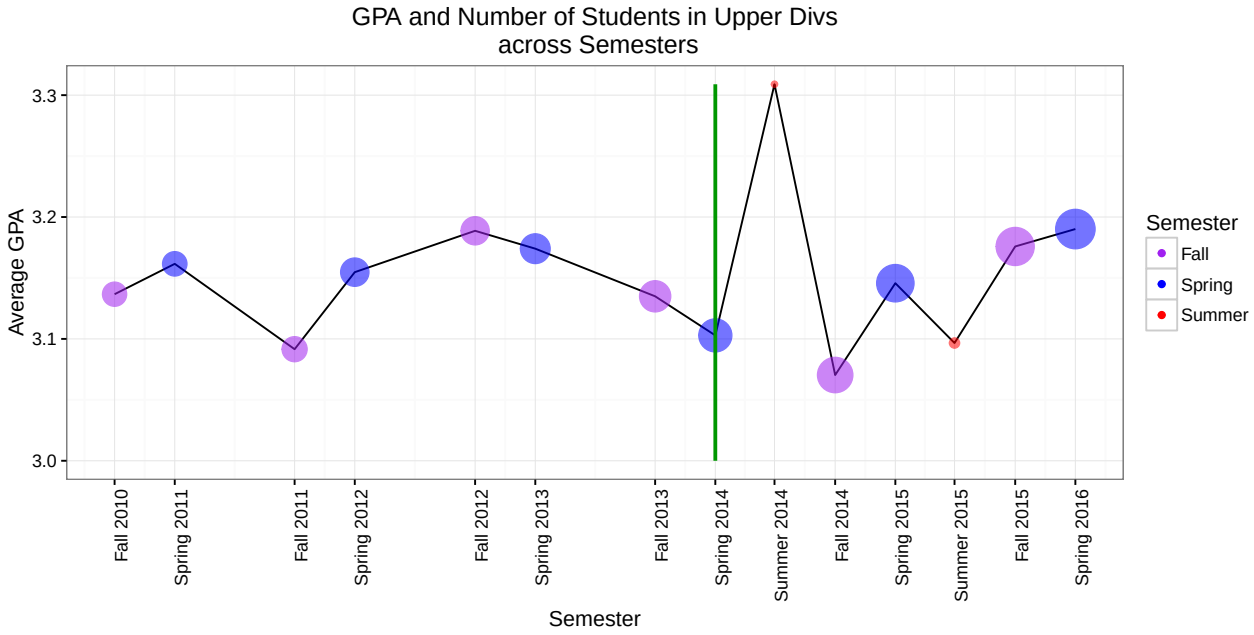

Figure 3: Green line represents when GPA cap was implemented, size of the points represents number of CS students

Major and Ethnic Breakdown

Here at Berkeley, there is a high value placed on promoting a connected community of different people. Many of the humanities majors aim to spread education of a certain group of people, through teaching of the language or promoting a cultural history. A good way to spread cultural understanding is having different groups of people undertaking intensive study of specific ethnicities or cultures. An examination of the ethnic composition of the applicants of these majors could reveal patterns of cross-cultural connectednes on the Berkeley campus.

To begin our analysis, we selected list of majors, such as African American studies and other majors that encompass the study of a subset of the community and its culture. To simplify visualizations, we aggregated some groups under one larger umbrella group. Although a more granular analysis would be possible, aggregating the data still allows for an analysis of the diversity of the applicants of a major. Specifically for the case of African American Studies and majors pertaining to Asian and Asian American culture, we aggregated related majors together as it would better capture the scope of the study of these two cultures. International students and students who declined to state their ethnicity were not included in this analysis because it is unclear which ethnic categories would be appropriate for these applicants. Finally, we calculated the number of people in each ethnic group that applied to the major and the respective percentage out of the total number of applicants.

For applicants of the major of American Studies, there is a noticable increase in the diversity of the applicants over the years. Examining figure 4, in the year 2000, the applicants are predominantly white and although this continues over the course of the 15 years, there is a general trend of an increasing percentage of applicants from the other

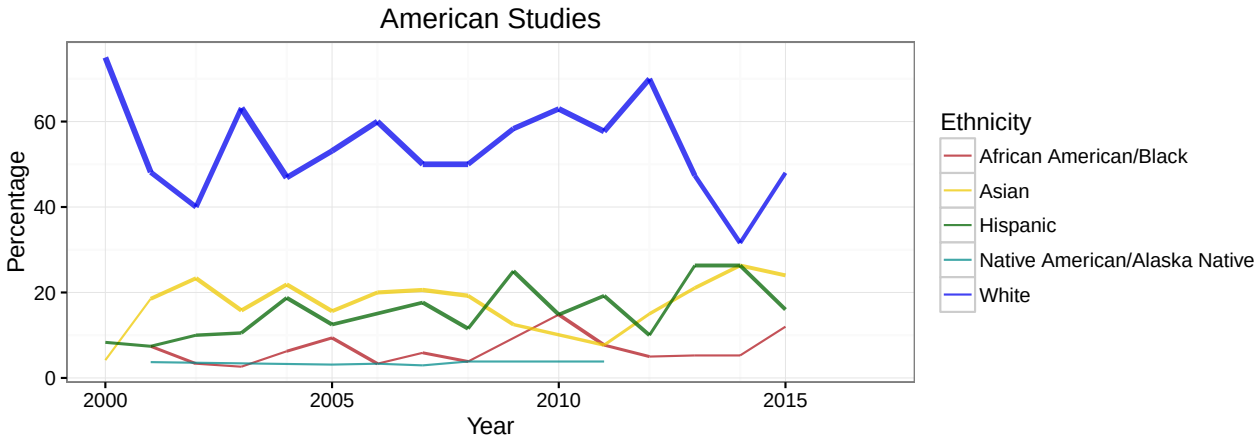

Figure 4: Ethnic Breakdown of American Studies Applicants

groups.

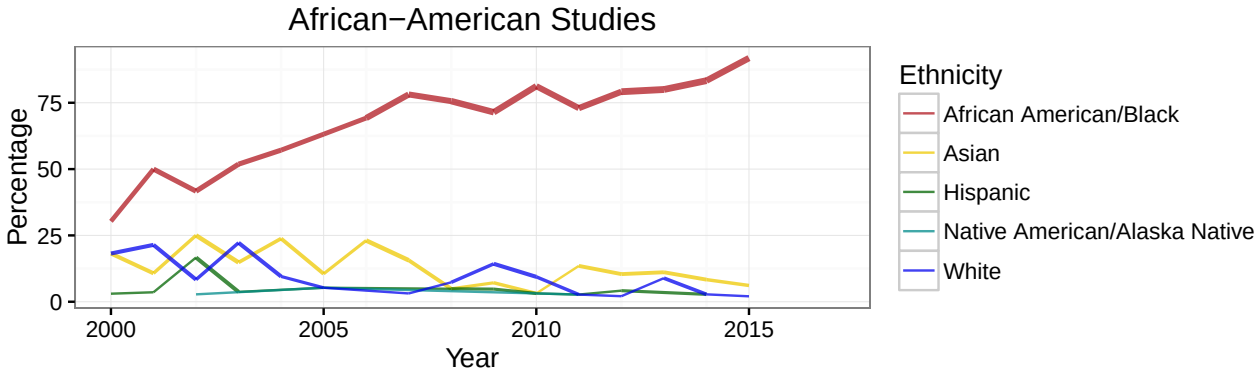

Figure 5: Ethnic Breakdown of African American Studies Applicants

For African American studies and related majors, as shown in figure 5, the trend is the opposite of American studies. Over fifteen years, the African American group's prescense in the majors increases dramatically, while the ethnic composition started off quite even. It is important to note that there has been a consistent increase in the number of total applicants, but also a decreasing number of other applicants. More African Americans apply to these majors as there are less applicants from other ethnicities, so there is an overall decrease in the diversity.

Looking at both African American studies and Asian studies and related majors, as well as other majors such as majors studying Asian culture and language, there appears to be a dominant ethnic group, and that group is the one that is associated with what the major aims to teach. For African American studies, that group would be African Americans and for Asian studies that group would be Asians. In these two groups in particular, the largest group makes up the vast majority of the applicants. This is would be consistent with the intuition that many poeple would choose to study what is familiar to them, at least when they are first choosing their major. They would generally be the most knowledgable of their own culture and in some cases would be the most interested

in learning more about their cultural history. The explicit focus of these majors would compound the pattern further.

On the other hand, an interesting analysis emerges when looking at applicant data for the Ethnic Studies and American Studies majors. These major titles are comparatively less ethnically specific and focus on a broader range of groups. American Studies, for instance, looks into American history and society. Of course, the United States has a long history and a strong legacy of immigration. Although it can be difficult to define one as an American or not, this distinction would be more encompassing and would draw more associations than a distinction such as African American.

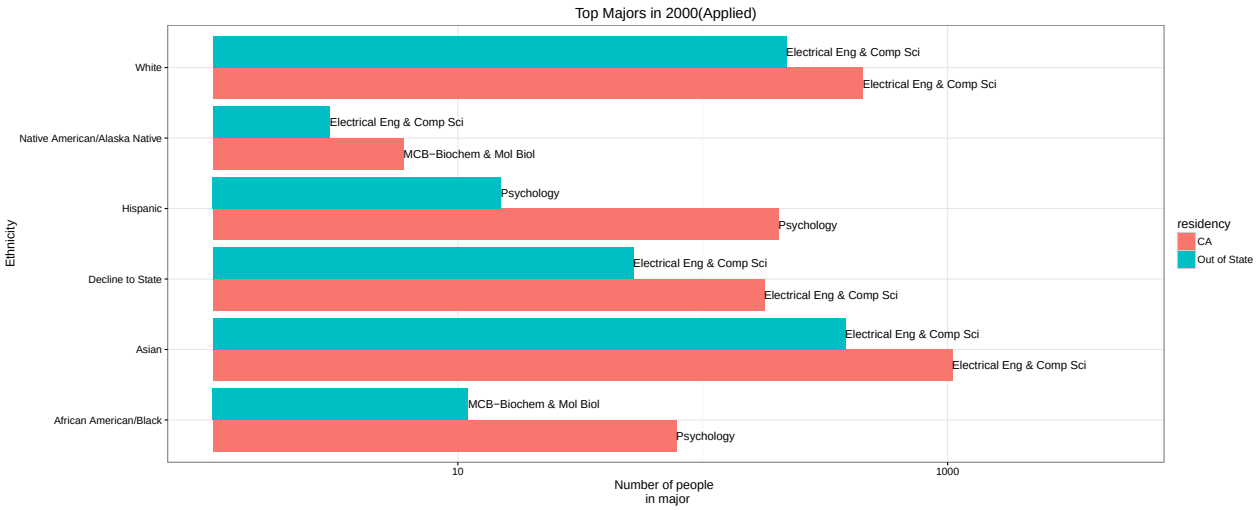

Figure 6: Top Majors 2000 by Ethnicity

This idea plays out even more so for Ethnic Studies. The percentages of applicants from respective ethnic groups indicates that for many years, particularly around 2006 and 2007, each ethnic group was well represented with each group making up at least near 20% of applicants. Although this diverse composition disappears past 2010, the makeup of the applicants of the ethnic studies major further shows how diversity of students within a major is strongly affected by who the major aims to study. As such, there are certainly limitations to promoting cross-cultural connectedness and understading through a focused and thorough education path.

As opposed to a more ethnic specific major, a nonhumanities major would be of broad appeal across groups. Comparing the second-most popular majors to the various ethnic groups shows interesting similarities, especially for students from other states. We can see from figure 6 and 7 that for most groups that come from other states, electrical engineering and computer science are the most popular majors, and there are less groups that have non-science majors as the top choice of applicants. The major choice of the decline to state group further shows the widespread popularity of science and technology major, particularly electrical engineering, if it is assumed that this group is a collection of various ethnic groups. Of course, this would be difficult to prove but it is a reasonable assumption.

A potential driver of this pattern is the fact that out of state students likely have a

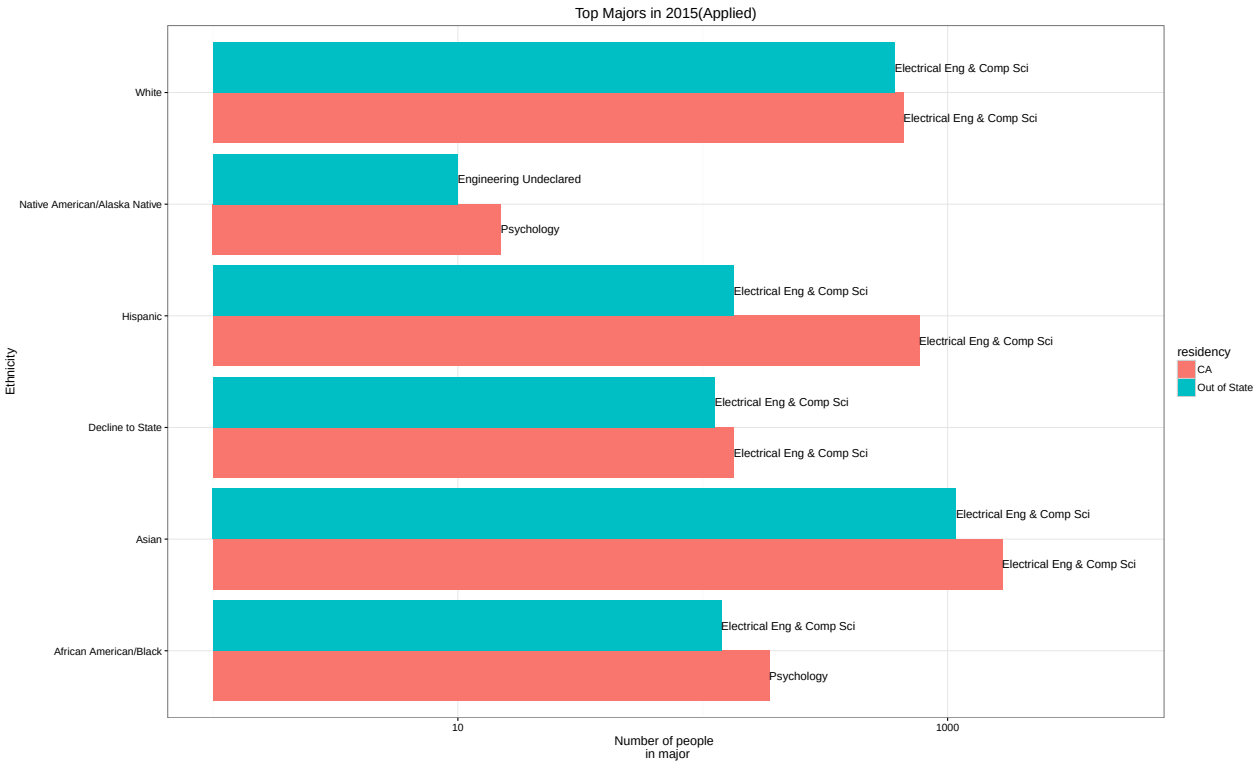

Figure 7: Top Majors 2015 by Ethnicity

higher cost of education than in state students. These students would choose to study majors that are widely perceived to lead to higher incomes, such as engineering. A lower financial burden possibly influences the major choices of domestic applicants, although engineering and related STEM majors are also widely popular but to a lesser extent. The major choices could also be attributed to the fact that there are substantially fewer out-of-state applicants than there are in state students, thus there are more limited and less diverse interest.

Affirmative Action and the Ethnic Makeup of UC Berkeley Students

Affirmative action is a hotly debated issue. Proponents argue that it gives a boost to minorities who badly need the increased educational opportunities. Opponents argue that the college admissions should be as close to a meritocracy as possible and that race-based admissions are unfair to the top students. Whether affirmative action has had a positive or negative on the school and its students is a difficult question to address, but the direct effect of affirmative action policies on the composition of the student body is clear. UC Berkeley had an affirmative action policy in place up until 1996, when California made it illegal for public universities to discriminate on the basis of race. By observing the racial composition of graduating students over time, we can see how the 1996 ban changed the ethnic makeup of the student body.

Figure 8 shows the ethnic breakdown of graduating students from 1983 to 2008. From 1983 to 1996, the percentage of white students dropped drastically. At the same time, the

percentage of Asian, Latino, and international students grew significantly. Interestingly, despite being considered a primary target for affirmative action policies, the population of black students from 1983 to 1996 barely fluctuated, maintaining a steady average of 4.879% of the student population. So, despite strong affirmative action policies, black students were still underrepresented at UC Berkeley.

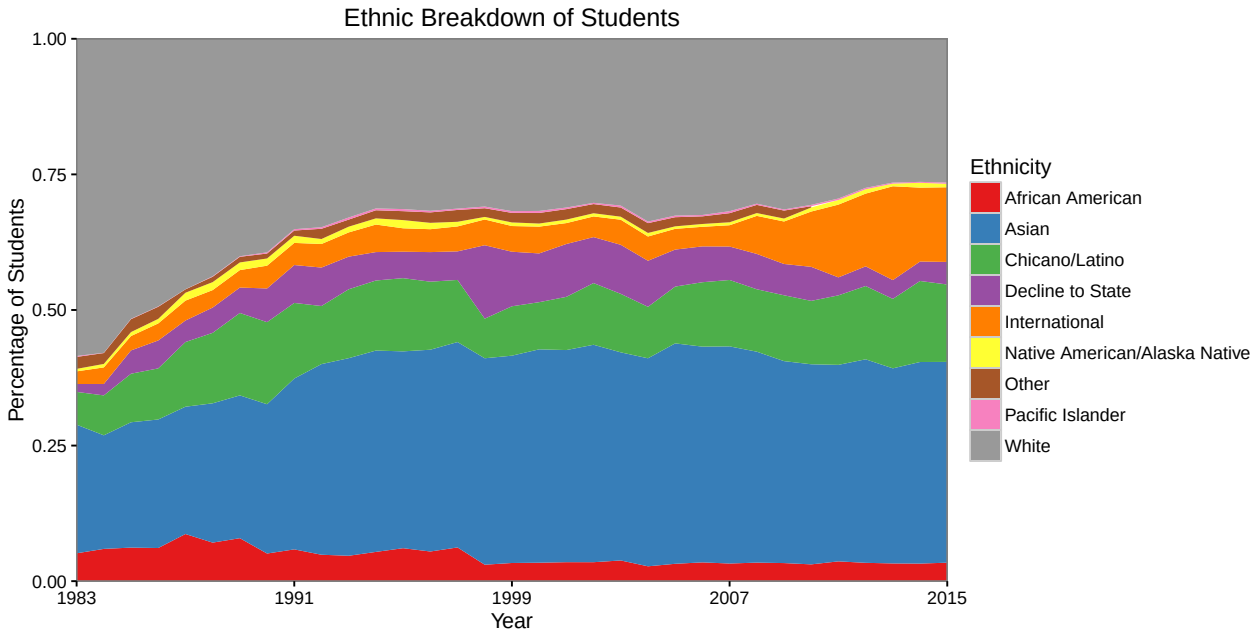

Figure 8: Ethnic breakdown of Berkeley over time, by percentage of the student population.

After 1996, the ethnic composition of Berkeley began to stabilize. The average percentage of black students between 1996 and 2015 was 4.382%, so it slightly declined during this time. However, besides this small decline, no group made any large fluctuations in percentage of the population until 2007. In 2007, the population of white and Asian students began do decline while the population of international students began to rise. This change can most likely be explained by economic reasons. Berkeley was likely feeling the effects of the sinking economy on its budget. By admitting more international students, a larger percentage of the student population would be paying out-of-state tuition. This could help Berkeley maintain its budget without having to cut services to students. However, this strategy decreased the in-state student acceptance rate, a side effect that did not make California tax-payers happy.

It is also interesting to note which populations did not see any effect from affirmative action. The percentage of the student body that was Native American and Pacific Islander remained small and relatively unchanged over the thirty two year period. These groups are clearly underrepresented minorities, but affirmative action policies did not appear to have any effect on how many were admitted.

How Long do Students Take to Graduate

The time it takes to graduate can be influenced by a variety of factors. The difficulty of the major, the individual academic strength of the student, and the cost of staying in school all affect how long it takes for a student to earn his or her degree. Over the past thirty years, the cost of tuition has skyrocketed. We can observe how this trend affects how long students stay in school, and we can further separate the student body by other metrics to see if there are other strong determiners of graduation time.

According to the data, students have been graduating faster over the past thirty years. The average number of years spent in school by a student graduating in 1983 was 4.620. This number decreased almost every single year for the next thirty two years, with the average reaching 4.049 in 2008. We predicted this trend of time spent in school decreasing, as we expected rising tuition costs to push people to graduate faster. Still, the full extent of the trend was surprising. Students in 1983 spent on average more than an extra semester in school compared to students in 2008.

Further trends can be discovered by dividing the students up into different categories. Females graduate faster than males, on average. Out-of-state students graduate slightly faster than in-state students. The ethnic breakdown of graduation time is shown in figure 9. The residency gap can be explained by the increased tuition costs for out-of-state students. This would also explain why international students graduate faster compared to other groups of students. The rest of the ethnicity gap may be explained cultural and socioeconomic factors that affect school performance, but further research would be required to draw any definitive conclusions.

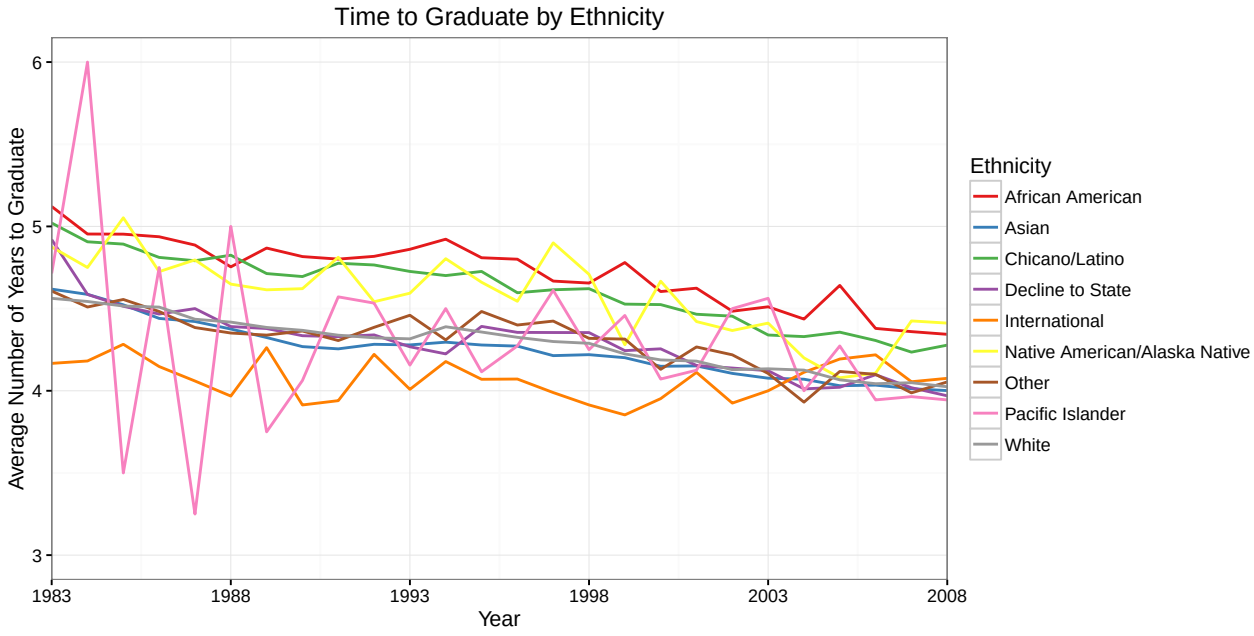

Figure 9: Graduation time by ethnicity. Only includes data from students who entered Berkeley as a freshman.

An interesting metric to analyze is the graduation time by college. Figure 10 shows this breakdown. Haas is not included because incoming freshman are not admitted into

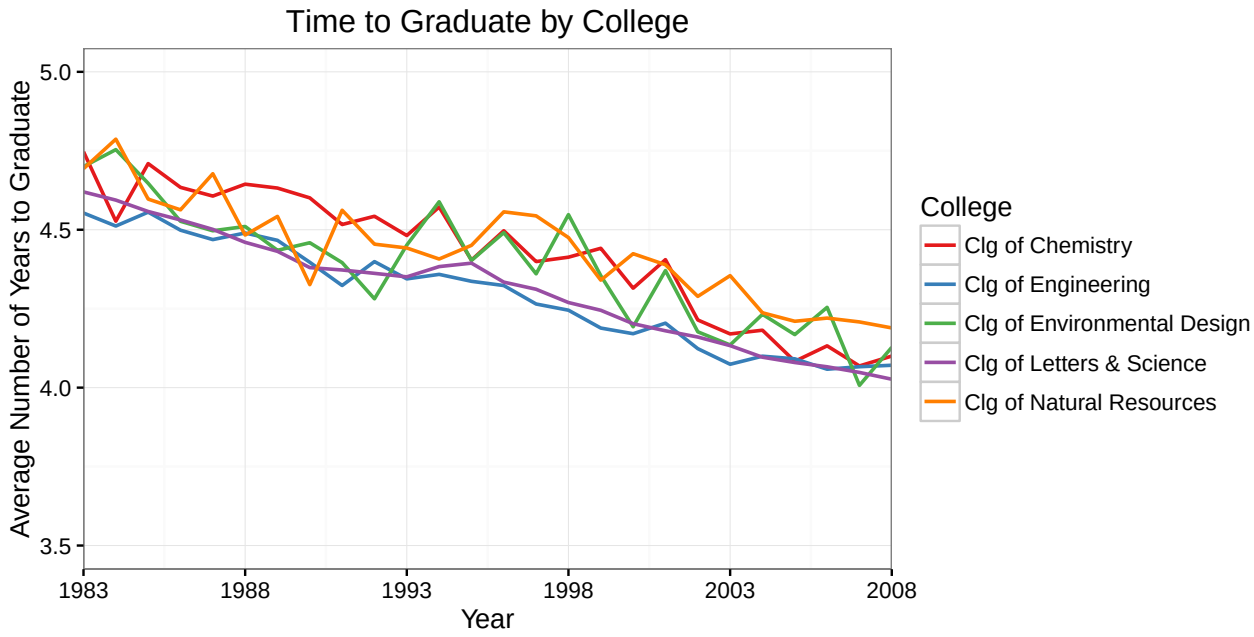

Figure 10: Gradation time by College. Only includes data from students who entered Berkeley as a freshman.

Haas, so the time taken to graduate by students in Haas is skewed compared to students in other colleges. Students in Chemistry, Environmental Design, and Natural Resources take longer to graduate than their peers in Letters & Science and Engineering. The majors offered by the Engineering department are generally considered to be difficult, but students actually complete these majors relatively quickly compared to majors offered by other colleges. The longer graduation times from the Chemistry, Environmental Design, and Natural Resources colleges may be explained by the career paths for students in these departments. Perhaps it is common in these fields to take an extra semester in order to learn as much as possible before beginning research or industry jobs. Perhaps it is common to take a co-op job, which often delays graduation by a semester or two. Whatever the reason may be, it appears that the perceived difficulty of the majors offered by each department do not correlate with how long it takes to graduate.

Conclusion

The data and analysis revealed many interesting results. Trends in the economy can have an impact on which students are admitted, what majors are chosen, and how fast people graduate. Policy changes made by the state of California regarding affirmative action have greatly affected the ethnic makeup of Berkeley students. The composition of inteded majors can greatly differ in terms of ethnic makeup. Policy changes are an appropriate method for dealing with the limited resources and growing student interest in computer science.

Although we had a wealth of data to work with, our research did have some limitations. While some Dashboards go as far back as 1983, some only contain a few years worth of data. In some cases, the data is split up over many tables with no varaibles to properly join the tables together. So, it is not always possible to directly compare two variables with each other. Ideas that could be incorporated into future research are getting access to the restricted Dashboards, gathering data from outside sources, and finding other ways to categorize and measure the student body.

There are a number of further topics that could be explored. What is the perceived difficulty of classes according students? How do these perceptions match up with the difficulty according to GPA? What percentile you need to achieve a particular letter grade for a given class? What is the average GPA of students by ethnicity, gender, and residency status? What factors influence the majors that students pick? All of these questions would be excellent springboards for future investigations of UC Berkeley student data.