Investigative Questions:

1. With areas (county and city) held constant, is it easier to be admitted to California universities through transfer from a college or direct admission from high school?
2. Is geographical location independent of admission rate for high school? In other words, do all high schools from one location yield a high admission rate, and other high schools from another location yield a low admission rate?
3. Is ethnicity a factor in high school and college transfer admission rates?

Results and Discussion:

1. We find the average admission rates in each area (county, city) for high school and college respectively. Then, we compare the admission rate of high school and college of the same county and city. If the high school admission rate is higher than college admission rates, then we would claim that for this county and city, it is easier to apply from high school than transfer from college. In a broader perspective, we found that over 65% of the areas have higher admission rate when a student is applying from high school.

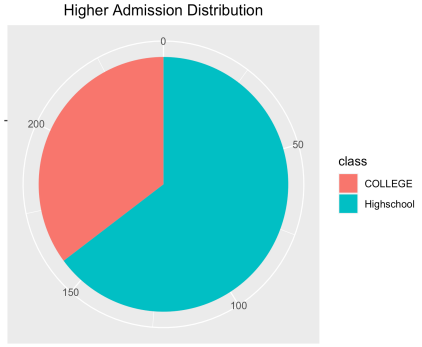

Figure 1: Pie chart on better admission rate

But is this result due to chance? The green box marks the average high school admission rate of the corresponding county and city. For example, in Scottsdale Arizona, the overall admission rate of high school students applying to universities is 62%. Similarly, the red box marks the average admission rate of college applicants from local community colleges to universities. In this case, Scottsdale Arizona has an overall transfer admission rate of 50% (Figure 2).

To investigate if there is a significant difference in the high school admission distribution and the transfer admission distribution. We can conduct a simple t-test to find out.

On the right-hand side, we can see the two plotted distributions to verify that the two distributions are both normal, with similar variances. Simply put, they are valid for t-test.

The overall average of high school admission rate is 58% across all counties and cities, which is significantly higher than the overall college admission rate of 52%. The result of the t-test shows that the difference between the two distributions is significant, not due to chance.

From the above results, we claim that with areas held constant, students have a higher chance to be admitted to universities through direct admission from high school. We can argue from an observational analysis result: over 65% of the areas have higher admission rate as a high school direct admission. Statistically speaking, there is significant

difference in the admission distribution between the two means of application.

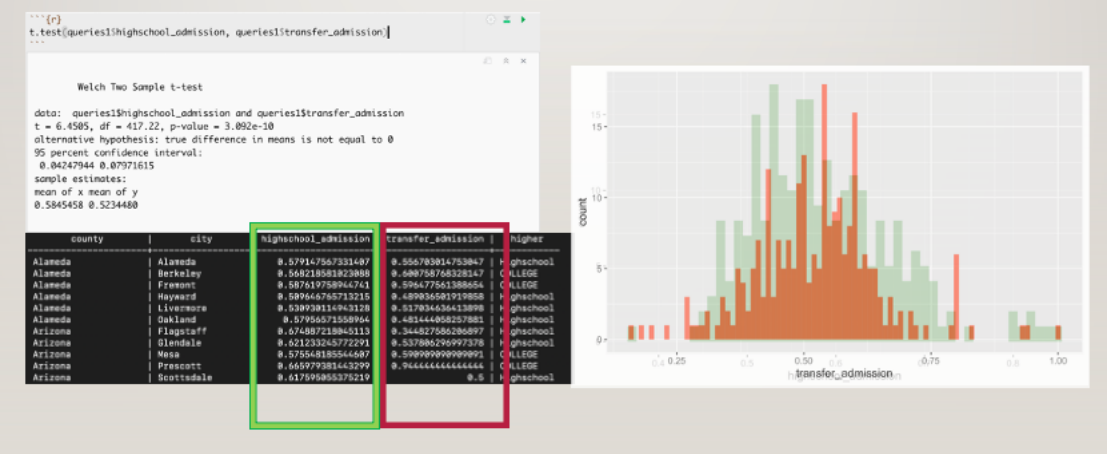

Figure 2: T-test

2. We look at colleges with admission rates below average, and we found that the ethnicity distribution among these colleges, Asian represents most of the ethnicity distribution (Figure 3). This finding will be relevant later with the third investigative question. Note that in the process of coming to this conclusion, we have filter out rows where a college has less that 400 applicants of the same ethnic group. For example, this row of information: the Hispanic group of 32 applicants in Golden West College applying to Santa Barbara in the term 2012 will not be considered (Figure 4). Because we feel like with a small number of applicants, it doesn't represent the overall results of admission rates.

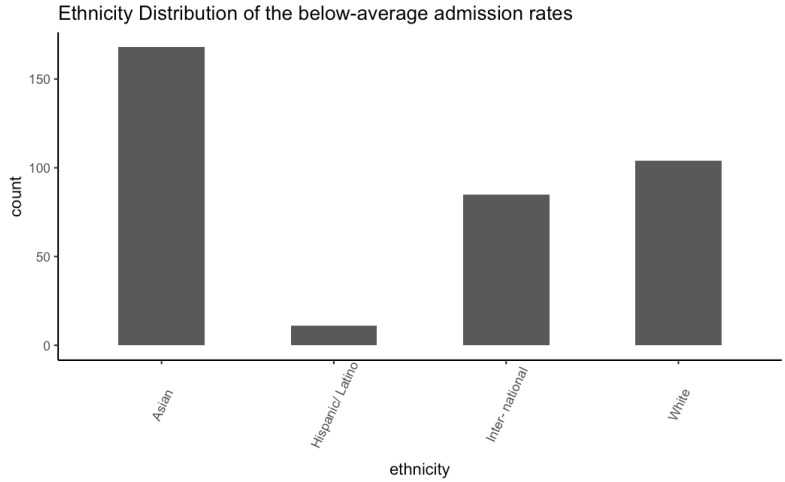

Figure 3 Ethnicity Distribution

Santa Barbara	2012	GOLDEN WEST COLLEGE	Hispanic/ Latino	32	16	8	0.50000000

Figure 4: Example of elimination of data

3. When we look at high school admission rates categorized by ethnicity, we see that they are relatively similar to each other, with Aboriginals being the highest, and Hispanic being the group with the lowest admission rate. We found that there is no ethnicity group with an admission rate that is significantly higher or lower than others.

We can find similar distribution with the data from college admission rates, with African American being the group with the lowest admission rate (Figure 5).

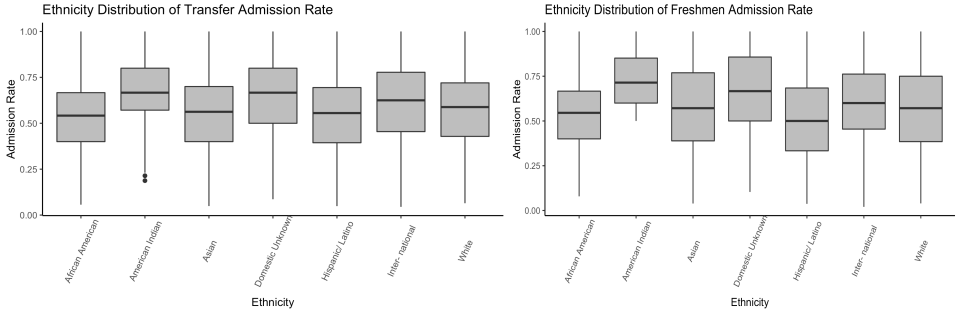

Figure 5 Overall distribution of Transfer and Direct Admission

If we compare the findings between question 2 and question 3. We see contradictory results at first glance. If Asians constitute most to the below-average admission rates among colleges, then on the right-hand side, why is the Asian box chart not the lowest among all ethnic groups?

Similarly, If Hispanic college applicants were the "minority" among the colleges with below average admission rates, we would naturally assume that the Hispanic college applicants all have a higher chance to be admitted to universities. However, why is it scoring the lowest admission rates across groups?

Here, we must consider the distribution within each ethnic group.

Let's zoom in on Asian and Hispanic group of college transfer students. The 2 distributions are very similar in mean, variance and tails (Figure 6). This corresponds to the box plots we have for question 3.

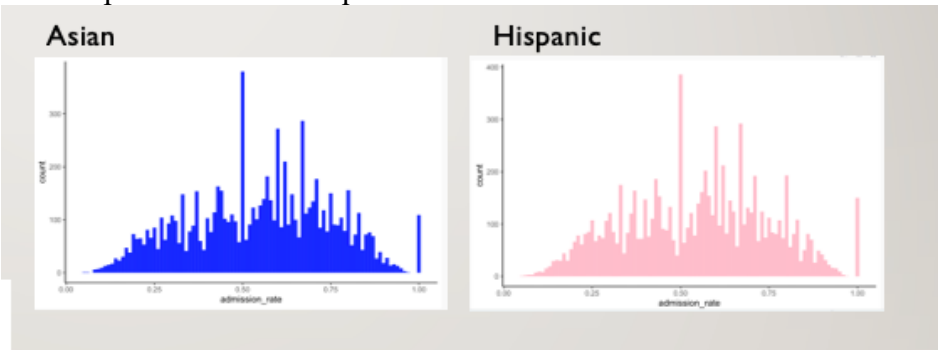

Figure 6 Asian and Hispanic College admission

However, if we use the same constraints as we did for question2, that is, only consider applicants greater than 400, and look at the below average admission rates: then we can see that most of the information we have on the Hispanic group is lost due to the constraint on the number of applicants (Figure 7).

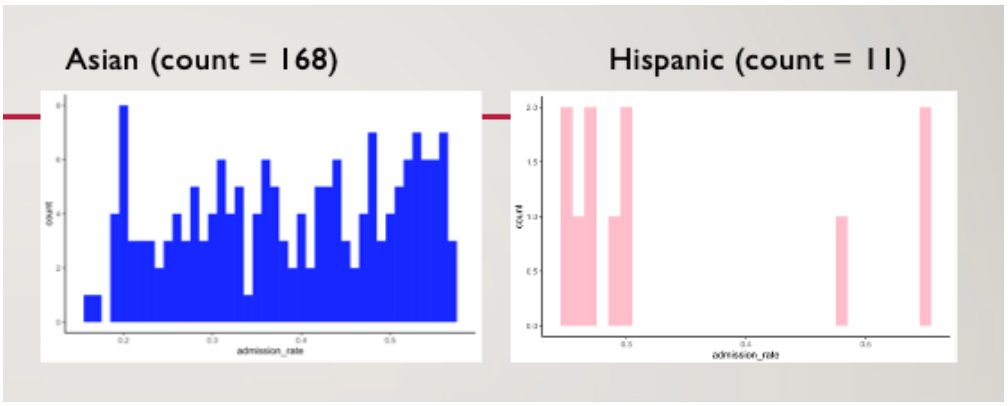

Figure 7 Question 2 Query Result

This explains why we have a big discrepancy between Asian and Hispanic group as the result of the 2nd investigative question. If we eliminate such constraints, then we know that there is no significant difference between the Hispanic and Asian group (Figure 8).

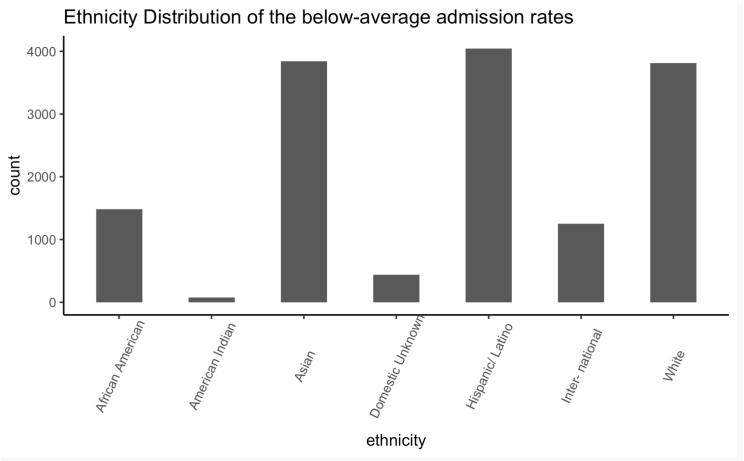

Figure 8: Without the applicant > 400 constraint

Conclusion:

Through the investigative questions, we were able to learn more in-depth about the university admission. Different means of application (college transfer and high school direct admission) were examined, and various ethnic groups were closely analyzed through visual presentation.