# Diversity

Qi Feng

1. Racial Diversity

Racial segregation, with its root in the practice of slavery and segregation, is still present today in the United States. Rcacial separation impedes the establishment of a multicultural and equal society. In particular, it poses challenges for minority students to improve in educational settings, as discrimination, racial biases and prejudices pervade. On the contrary, racial diversity can benefit students for it fosters collabration, encourages creativity, and ultimately improves academic achievement. Therefore, we would like to provide an evaluation of racial diversity among colleges.

To begin analysis, we first explore **how common each race is in colleges**.
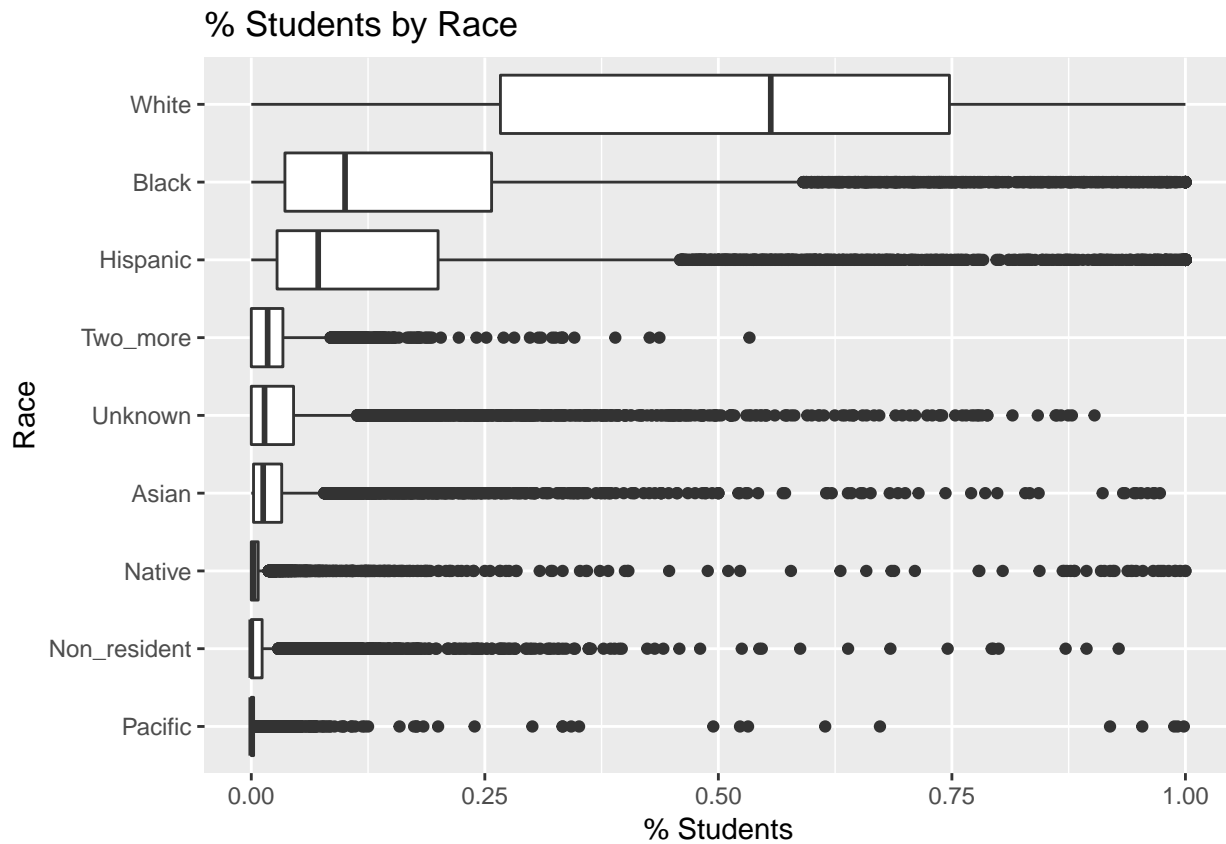
% Students by Race (capped at 0.6)
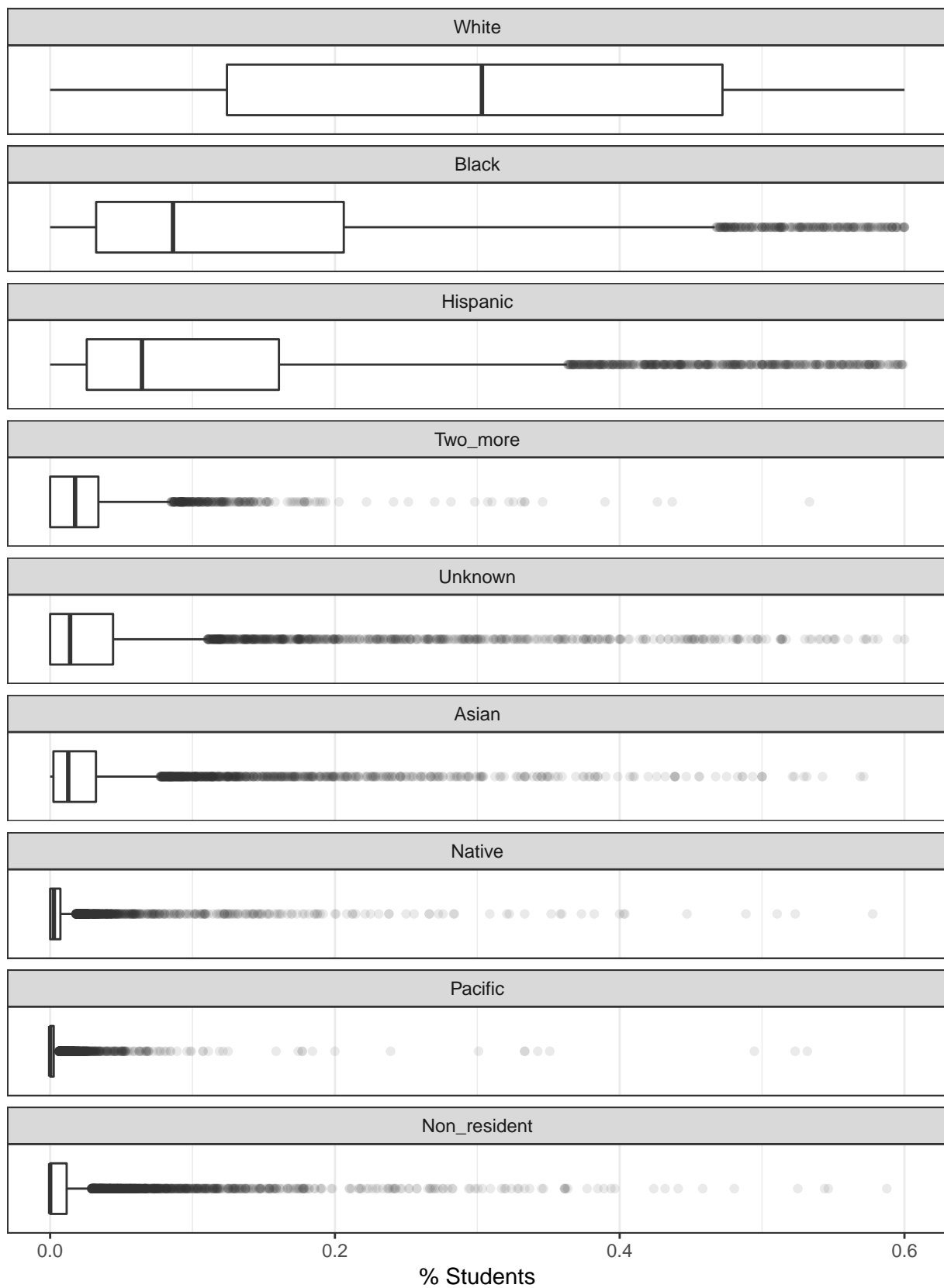
Comparing the medians in the boxplots we find the between-group variation to be significant. Specifically, `White`, `Black` and `Hispanic` seem to be the most common race while `Pacific` and `Non_resident` are the least. Also, in-group variation is larger for `White`, `Black` and `Hispanic` as well, so that the proportions of these racial groups vary more from college to college. Note that there are a great many outliers in the dataset.

Racial Diversity Index (RDI)

Next, to analyze racial diversity, we introduce *Racial Diversity Index (RDI)*, which is defined by the probability that any two students selected at random would have different races. This index was developed by Missouri State University (https://diversity.missouristate.edu/DiversityIndex.htm) and is calculated by

$$RDI = 1 - Pr(\text{same race}) = 1 - \sum_i (\text{race i})^2,$$

where *race i* is the percentage of students of *race i*.
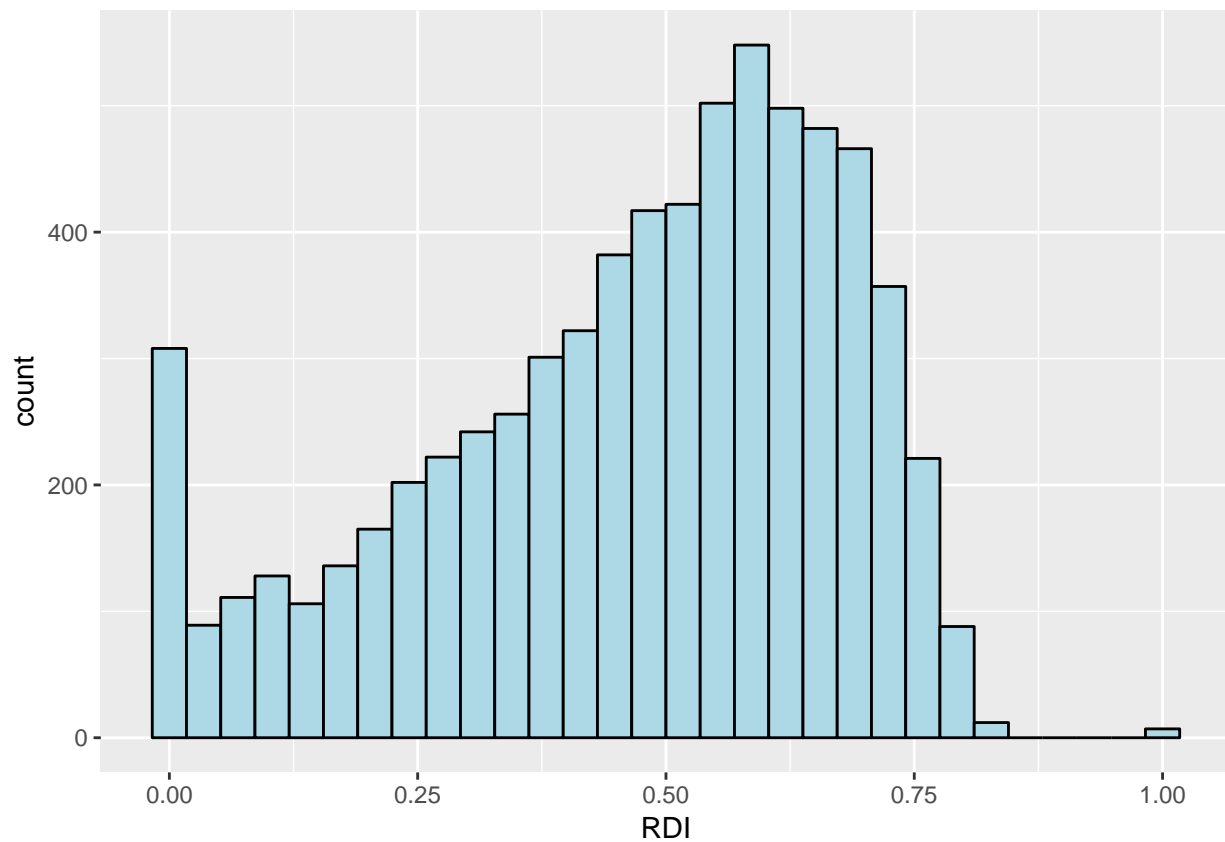
An RDI close to 1 implies high racial diversity while an RDI close to 0 implies low. From the histogram, we observe that the distribution of `RDI` is slightly negatively skewed, so that more colleges have moderate to high `RDI`'s (over 0.50). The mode is approximately 0.6 so that most colleges have moderately high racial diversity.

Note that there are a few colleges scoring 0.00 (*Table 1*), as all of their students are from one exclusive race. For instance, all students of *Thunderbird School of Global Management* are *Hispanic*. At the other tail, note the isolated bar for `RDI = 1.00`, which is impractical as no college would have exactly even distribution of races. A closer look at the data reveals that the `race` features for these colleges are all 0, indicating that their `population` might not have been recorded properly (*Table 2*). Excluding these colleges, we find the most racially diverse colleges listed in *Table 3*.

Finally, taking a look at the make-ups of the three most racially diverse colleges, we can see that proportions of races can still vary a lot within each college. For instance, at *Chaminade University of Honolulu*, almost

Table 1: Colleges with RDI = 0

name	White	Black	Hispanic	Asian	Native	Pacific	Two_more	Non_resident	Unknown	RDI
Thunderbird School of Global Management	0	0	1	0	0	0	0	0	0	0
Velvatex College of Beauty Culture	0	1	0	0	0	0	0	0	0	0
Universal College of Beauty Inc-Los Angeles 1	0	1	0	0	0	0	0	0	0	0
Carver Bible College	0	1	0	0	0	0	0	0	0	0
Mr Leon's School of Hair Design-Moscow	1	0	0	0	0	0	0	0	0	0
Hebrew Theological College	1	0	0	0	0	0	0	0	0	0

Table 2: Colleges with RDI = 1

name	White	Black	Hispanic	Asian	Native	Pacific	Two_more	Non_resident	Unknown	RDI
American Conservatory Theater	0	0	0	0	0	0	0	0	0	1
Prince Institute-Rocky Mountains	0	0	0	0	0	0	0	0	0	1
Lyme Academy College of Fine Arts	0	0	0	0	0	0	0	0	0	1
Professional Business College	0	0	0	0	0	0	0	0	0	1
Spanish-American Institute	0	0	0	0	0	0	0	0	0	1
Taft University System	0	0	0	0	0	0	0	0	0	1

one-third of students are `Asian` but only few are `Native` and `Non-resident`.

% Students by Race
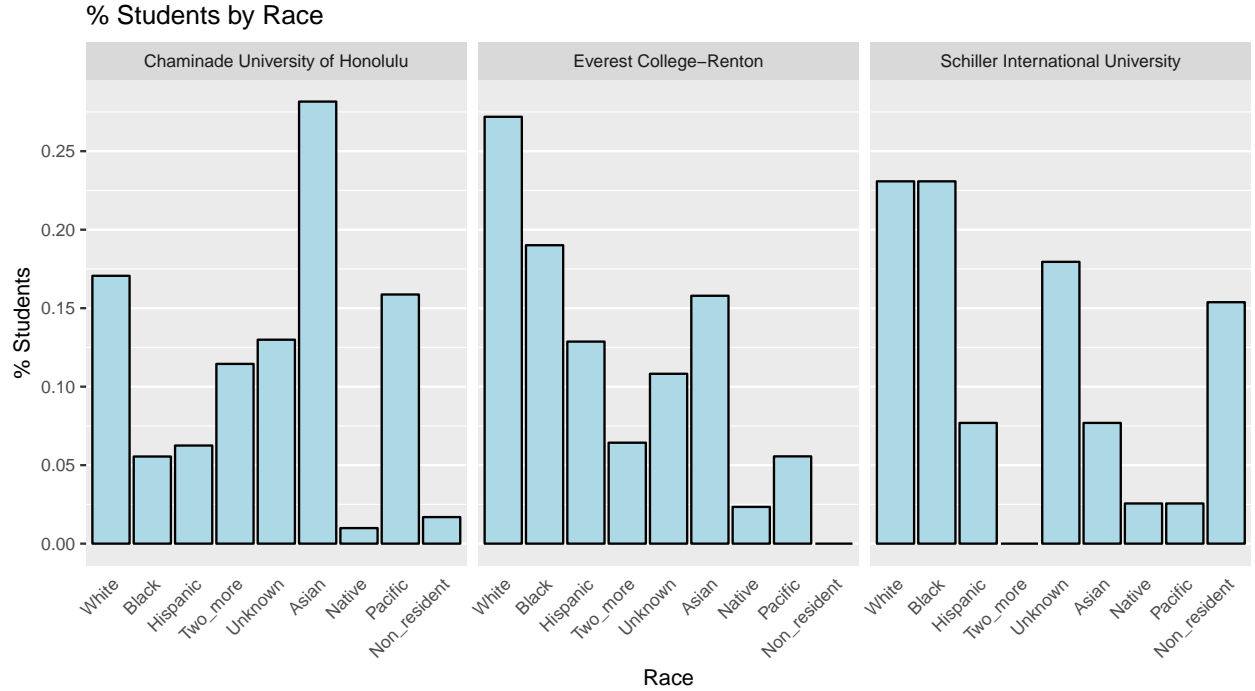
Table 3: Colleges with Highest Racial Diversity

name	RDI
Chaminade University of Honolulu	0.8290570
Everest College-Renton	0.8289556
Schiller International University	0.8244501
Pacific Rim Christian University	0.8215979
Golden Gate University-San Francisco	0.8205131
Charles A Jones Career and Education Center	0.8203344

2. Gender Diversity

Gender equality has been an issue in workplace, where women are known to have struggled for parity in pay and influence. To form awareness of gender equality and reduce challenges in workplace in the future, more and more colleges have been working on promoting gender diversity.

To investigate gender diversity, we first explore **how common each gender is in colleges**. Comparing the medians in the boxplot we see that most colleges seem to have more female students than male.
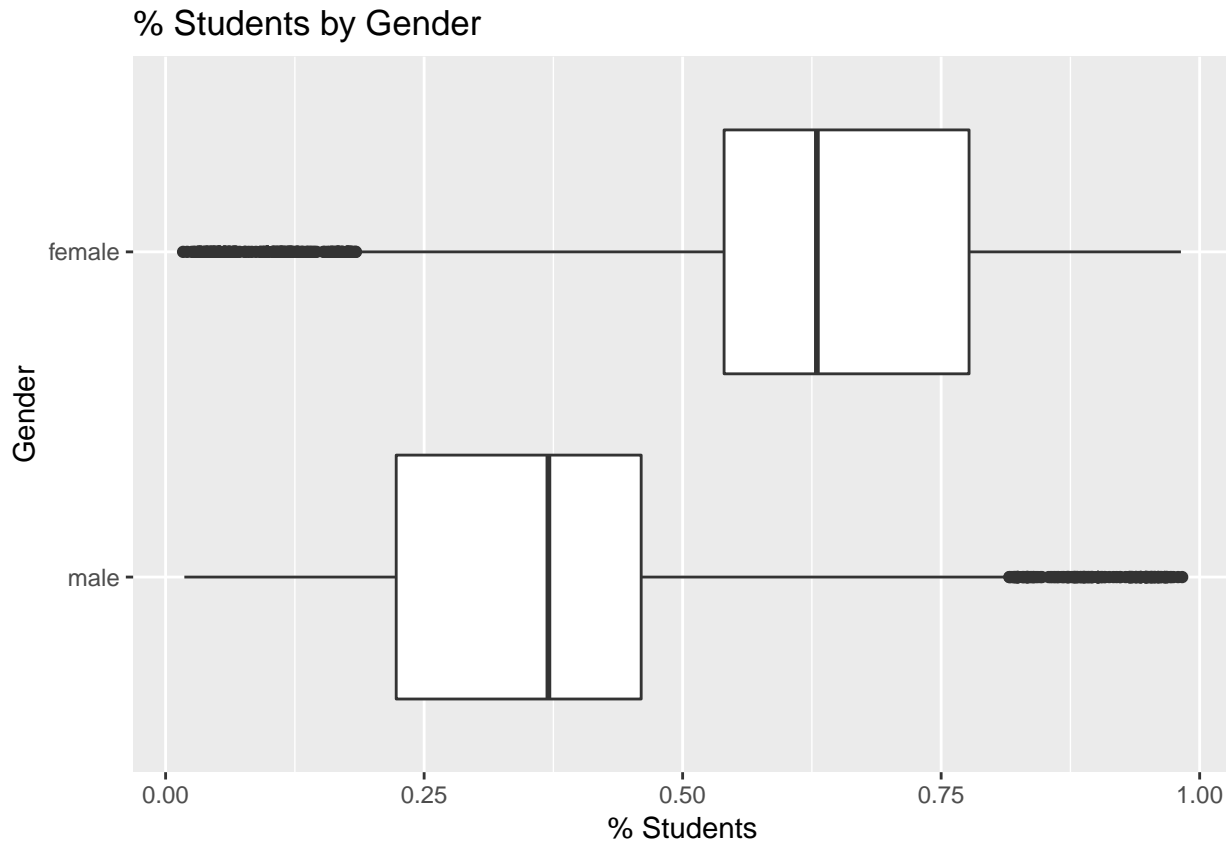
% Students by Gender



Gender Diversity Index (GDI)

To give a substantial measurement of gender diversity, we develop the *Gender Diversity Index (GDI)*. Like **racial diversity**, it is defined by the probability that any two students selected at random would have different genders:

$$GDI = 1 - Pr(\text{same gender}) = 1 - (\text{male})^2 - (\text{female})^2,$$

where *male* and *female* are % of male and female students.

Table 4: Colleges with Highest Gender Diversity

name	female	male	GDI
VanderCook College of Music	0.5000000	0.5000000	0.5
South Carolina State University	0.5000000	0.5000000	0.5
American Institute of Beauty	0.5000000	0.5000000	0.5
SOLEX College	0.5000000	0.5000000	0.5
Texas State Technical College-Harlingen	0.5001274	0.4998726	0.5
University of Idaho	0.5001372	0.4998628	0.5
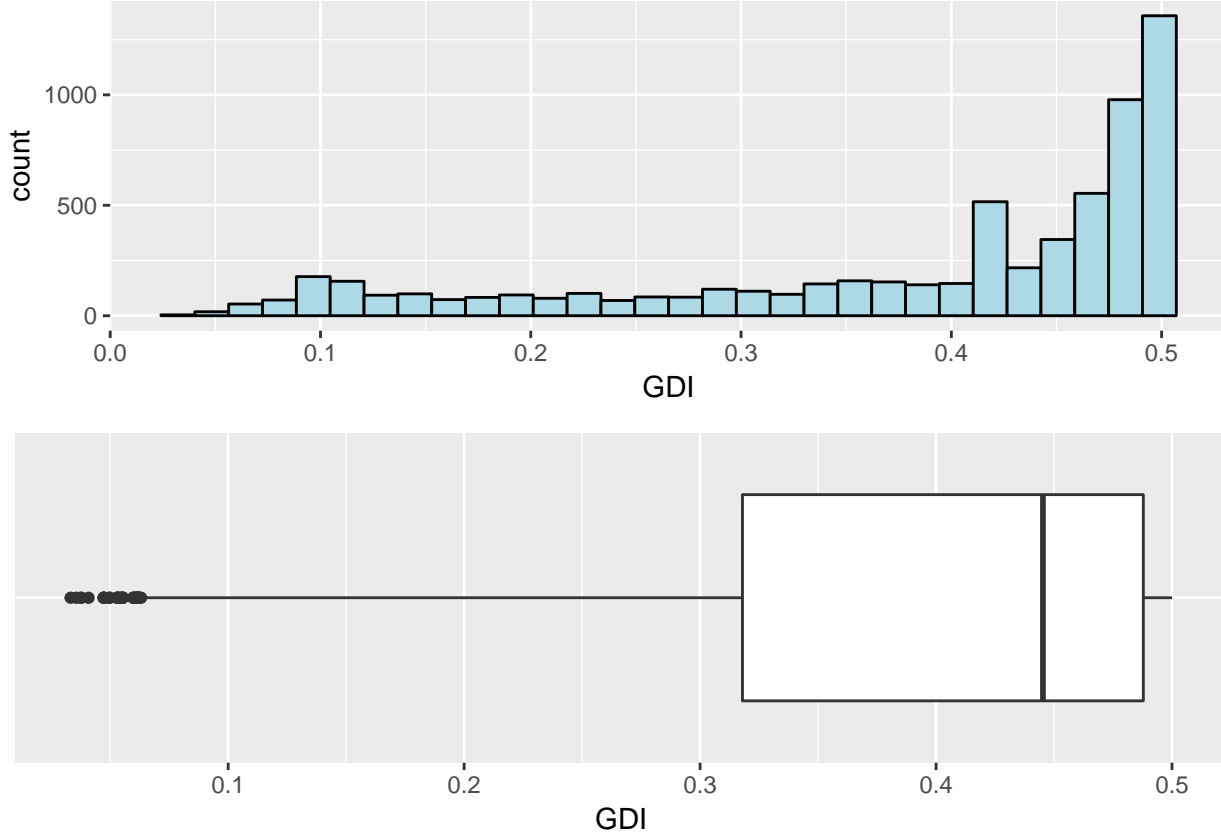
A GDI close to 0.5 implies high racial diversity (i.e., the % of male students are close to % of female students) while a GDI close to 0 implies low racial diversity. From the histogram, we observe that the distribution of GDI is negatively skewed with a long left tail. Along with the boxplot, the spread for the bottom 25% of GDI is much wider than the top 25%, meaning that the colleges having the lowest gender diversity vary a lot in GDI, while the colleges with the highest gender diversity have similar GDI. Note that the mode is 0.5, so that most colleges have equal % of female and male students.

Finally, the colleges with the highest gender diversity are the followings: