

Evaluating the Effect of Racial Minority on Minority Graduation

Lab 2 - Team: Galactic Mavericks

Introduction:

In recent decades, collegiate education has experienced a significant surge in enrollment, with undergraduate enrollment numbers increasing by almost 30% in the past two decades (1). As educational opportunities have expanded and become more accessible to the general public, it is crucial to examine the factors that contribute to student success, particularly among those from underrepresented minority groups. The composition of the student body has emerged as a crucial factor affecting student outcomes. As colleges and universities have become more racially diverse, scholars have begun to investigate the impact of this diversity on student achievement.

Therefore, our study will focus on the academic performance of ethnic minority students within a diverse student body, with the ultimate goal of providing insight into the factors that contribute to their academic success. By examining this relationship, our analysis aims to provide practical implications for educational institutions and policymakers seeking to improve the educational outcomes of underrepresented minority students. Specifically, the research question that we will strive to answer is, Does having a higher percentage of racial minority composition within a student body lead to higher graduation rates for minorities?

Data and Methodology:

The data used in this research is being derived from the National Center for Education Statistics', Integrated Postsecondary Education System (IPEDS) and the Voluntary System of Accountability's Student Success and Progress rate. The data was compiled and uploaded to kaggle by Jonathan Ortiz (2). In our analysis, we included data on five racial categories: black, Asian, Hispanic, American Indian, and white. It is important to note that we considered the first four racial groups to be minority groups according to the US Census Bureau. This designation is based on the bureau's classification system, which defines a minority group as any group that is not non-Hispanic white alone.

Data Manipulation:

Our dataset consisted of one main table, `state_sector_grad`, which contains the graduation rates by state, institution type and race, and 2 secondary tables, `state_sector_details` and `institution_details` that had information about state appropriations and student aid respectively. In order to successfully prep the data for our analysis, we had to perform the following data transformations within R. We initially performed all exploration and analysis work on a 30% subsample of the data. We then used the remaining 70% as the confirmation set to generate the statistics in this report.

The first step taken was to calculate the graduation rate of each minority group. We first filtered the data to show only data from the total race to get the total student body count ("`total_cohort`") grouped by the following fields: `state`, `control`, `level`, `year`, and `cohort`. This data frame was called `data_totals`. We then created a separate dataframe, "`df`", that was grouped by the following fields: "`state`", "`control`", "`level`", "`year`", "`cohort`", "`gender`" and we then dropped all columns that we deemed irrelevant for our analysis. We joined the initial dataset back to the totals dataframe based on the following fields: "`state`", "`control`", "`level`". In order to gather the endowment and aid information, we joined the `institution_details` and the `state_data` to the main fact table on fields: "`state`", "`control`", "`level`". In regards to the metric calculations such as `grad_150_rate`, we decided to create our own graduation rate.

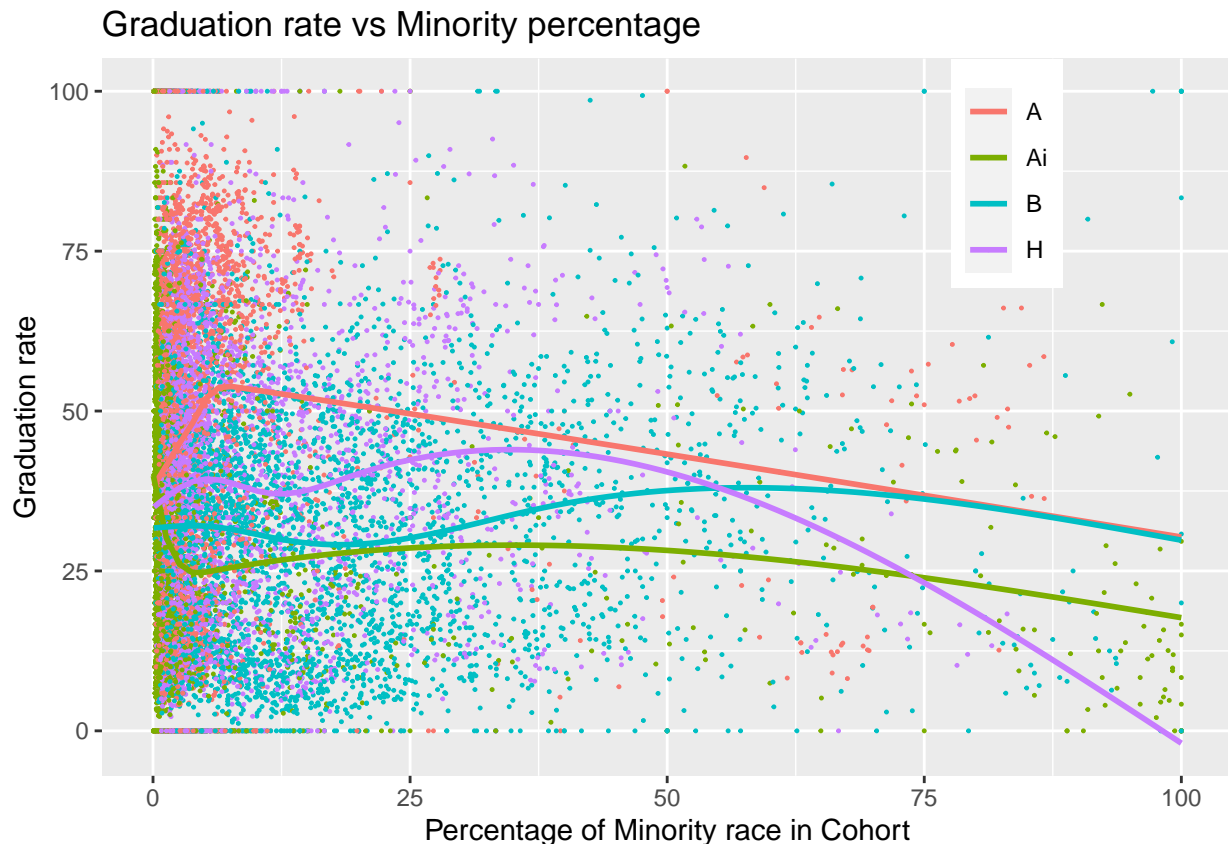
The reason we did this is that this will make it easier to summarize the data when we are grouping. This was made possible by the `total_cohort` field from the `totals` df.

In addition to the data transformations mentioned earlier, we then omitted the variables ‘year’ and ‘cohort’ from our analysis because the graduation rates remained constant throughout the years in the dataset and the cohort type (whether it’s a 4-year bachelor program or another 4-year program) didn’t have any noticeable effect. Omitting these variables allowed us to simplify our analysis and focus on other relevant variables. However, it is important to note that omitting a relevant variable can lead to omitted variable bias. In order to avoid this bias, we conducted a sensitivity analysis where we included the ‘year’ and ‘cohort’ variable in our analysis and assessed its impact on our findings. Our sensitivity analysis indicated that the inclusion of the variables did not significantly affect our results, as the variables were not associated with our outcome variable of interest, the graduation rate.

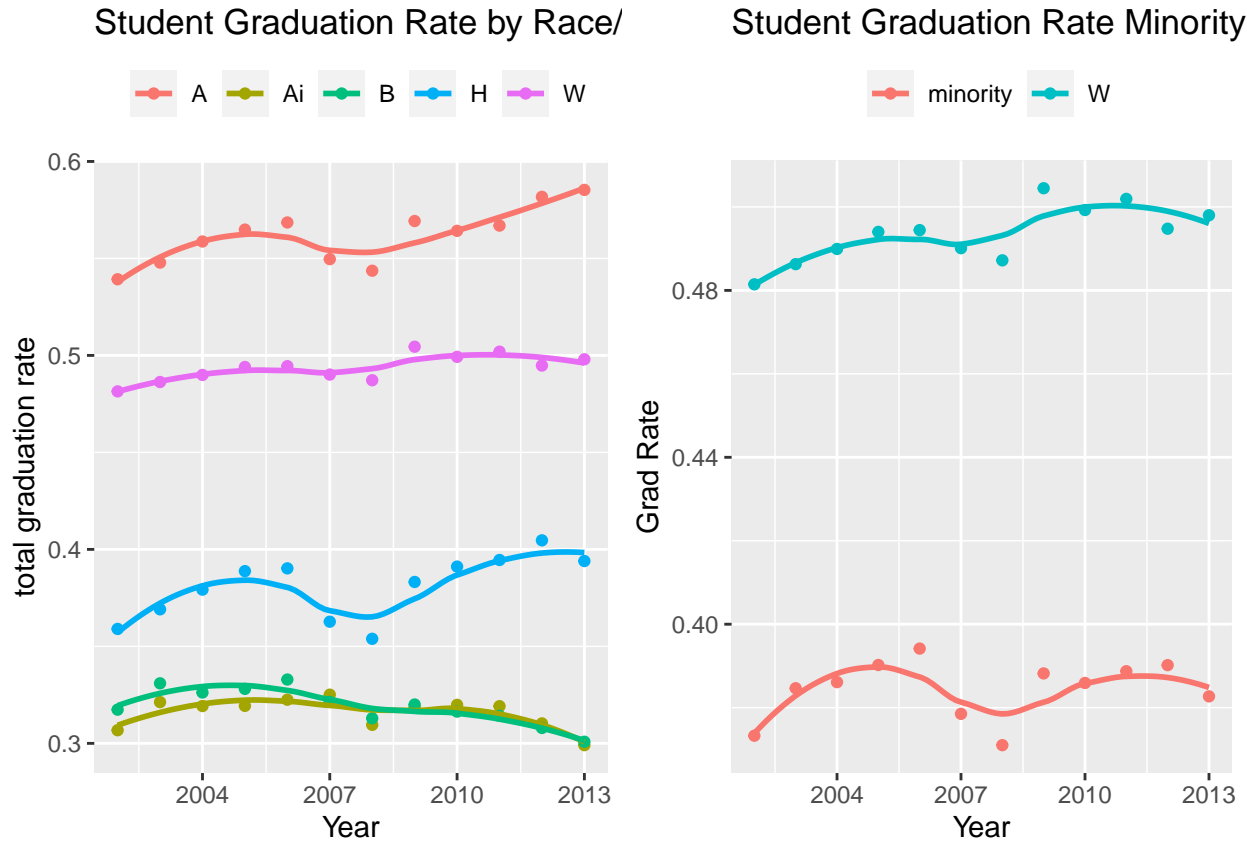
Our team then omitted all values with the race “white” in order to target minorities within our analysis. We also dropped the state “United States” as it included the national averages. Our final dataset consisted of the following 5 base metrics: `perc_minority` (percent minority), `grad_100_rate` (those that graduate on time), `grad_150_rate` (those that graduate on time and late), `aid_value` (aid going to student recipients), and `state_appr_value` (State appropriations to higher education). For the purpose of our analysis we decided to focus on the variable `grad_150_rate` instead of the `grad_100_rate` as this more closely aligns with our research question.

Initial Observations/Exploratory Analysis:

In our first analysis we plotted the percentage minority against the graduation rate (Figure 1). This gave us an initial discovery that we decided to pursue. Our analysis led us to believe that there is an inverse relationship between graduation rate and minority percentage as the less minorities are within a cohort, the higher the graduation rate for those minorities are.



We then evaluated the graduation rate among all races (including the omitted race “white”) to gain an understanding of our data. Although Whites had by far the largest graduation numbers, Asians had the highest graduation rate (Figure 2). Given that Asians are considered a minority within our analysis, it was interesting to see that despite this, the minority graduation rate is still well below that of Whites (Figure 3).



Our regression Models:

The goal of our regression is to try and identify and quantify the relationship between our dependent variable, graduation rate, and the main independent variable, Minority percentage. We end up modeling the following three regressions:

$$\widehat{\text{graduation rate}} = \beta_0 + \beta_1 \cdot (\text{Minority percentage}) + \epsilon$$

$$\widehat{\text{graduation rate}} = \beta_0 + \beta_1 \cdot (\text{Minority percentage}) + \beta_2 \cdot (\text{Financial Aid}) + \epsilon$$

$$\widehat{\text{graduation rate}} = \beta_0 + \beta_1 \cdot (\text{Minority percentage}) + \beta_2 \cdot (\text{Financial Aid}) + \beta_3 \cdot (\text{State Appropriation}) + \epsilon$$

Results:

Starting off with the first model, or our base model, we tested the percent minority against the graduation rate. In the second model we tested the percent minority and financial aid value. The third model we tested percent minority, aid value and the state appropriate value. Despite our initial assumptions that there may have been a negative correlation between percent minority and graduation rate, our results were returned to be inconclusive. Per the stargazer table depicted below, the perc_minority field is not statistically significant. We thus cannot reject the null hypothesis of the percent minority composition having an impact on the graduation rate for minorities.

However, during our analysis our team did discover an interesting finding. Student Aid returned a highly significant P-value. Per the stargazer table, for every \$500 increase in student aid will result in an increase of one percent in student graduation rate.

```
##
## Estimated Regressions
## =====
##                               Output Variable: Graduation Rate
## -----
##                               (1)                (2)                (3)
## -----
## Percent Minority             -0.053             0.002             0.021
##                               (0.046)            (0.049)            (0.049)
##
## Student Aid                  0.002***          0.002***
##                               (0.0001)          (0.0001)
##
## State appropriations                -0.011
##                               (0.007)
##
## Constant                    39.626***          27.508***          30.484***
##                               (0.705)            (2.158)            (2.158)
## -----
## Observations                 1,037             1,037             1,026
## R2                          0.001             0.135             0.129
## Adjusted R2                 0.0002            0.133             0.126
## Residual Std. Error   19.375 (df = 1035)   18.039 (df = 1034)   18.006 (df = 1022)
## F Statistic           1.187 (df = 1; 1035) 80.708*** (df = 2; 1034) 50.418*** (df = 3; 1022)
## =====
## Note:                               *p<0.1; **p<0.05; ***p<0.01
```

There are limitations to our analysis though. For regression estimates to be reliable, it is necessary to assume that the observations are independent and identically distributed (iid). There are some factors that we see might create dependencies between different samples. First of all, based on figure (1) we can safely say that the distribution of graduation rate varies significantly between races as we see clusters. Secondly, our dataset considers only 5 racial minorities. This can be problematic because it does not account for the diversity within each of these categories. Moreover, racial categories are social constructs and can vary across different cultures and societies. Therefore, a limited number of categories may not accurately capture the diversity of racial identities and experiences. Additionally, some people may not feel that any of the available categories accurately represent their racial identity. This can lead to people feeling forced to choose a category that they do not fully identify with or feeling excluded from the survey altogether. For example, Middle eastern or North African Americans don't perceive themselves as white (4). Despite our hopes of discovering conclusive results, our team decided to be careful to not fall into the trap of "fishing for results" and we opted to publish our findings.

Conclusion:

The Galactic Mavericks team sought to answer the following research question : Does having a higher percentage of racial minority composition within a student body lead to higher graduation rates for minorities? Despite initial optimism from our exploratory analysis, our results came back as inconclusive. As such we cannot determine if having a higher percentage of minority students will lead to higher graduation rates for those students. Despite this, this cannot be viewed as a complete failure. Our team did an honest investigation and did not attempt to search for results by modifying our analysis to find significant results.

However if time was not a constraint, our team would have done additional research and analysis to evaluate the effect on graduation rates for minorities. Currently we're implicitly looking at graduation rate from the lens of race. In reality there are multiple causal factors at play in determining one's likelihood of graduation. In order to capture all those factors, we would have to look at this problem from various angles simultaneously. For instance, during our analysis, our team discovered that there is a strong correlation between the amount of student financial aid received and graduation rates. If we had more time for our research, our team would have been interested in investigating several other factors, such as the impact of sexual orientation, socioeconomics, and epigenetics to name a few.

Citations:

1. Hanson, Melanie. "College Enrollment & Student Demographic Statistics" EducationData.org, July 26, 2022, <https://educationdata.org/college-enrollment-statistics>
2. "College Completion and Efficiency Measures for US." Kaggle, <https://www.kaggle.com/datasets/thedevastator/college-completion-and-efficiency-measures-for-u>.
3. "About the Topic of Race" United States Census Bureau, <https://www.census.gov/topics/population/race/about.html>
4. "Middle Eastern and North African Americans may not be perceived, nor perceive themselves, to be White" PNAS, <https://www.pnas.org/doi/10.1073/pnas.2117940119>