Aggregate & Longitudinal Analysis Report

Data description

In this analysis, two distinct datasets were utilized to investigate educational disparities within a school district and across states in the United States. The first dataset is used for aggregate analysis, examining student demographics within a particular school district. The second dataset is employed for longitudinal analysis, exploring the trends in the number of students with disabilities receiving special education and related services according to an individualized education program or service plan across states from 2012 to 2017.

Data Source

The data for aggregate analysis originates from two PDFs extracted from chapter 9 of the book "Data Science in Education Using R." The first PDF contains information on student demographics, specifically categorizing students by race (Native American, African American, Asian, Hispanic, White), school name and school group. The second PDF focuses on FRPL data, a metric often used as a proxy for poverty. Students from a household with an income up to 185 percent of the poverty threshold are eligible for free or reduced price lunch. The PDF categorizes students by school name and their FRPL eligibility.

The second dataset for longitudinal analysis is a combination of six separate .csv datasets, each representing child counts for the years 2012 to 2017. These datasets were extracted from chapter 10 of the same book. The information is categorized based on age, gender, race, state, type of disabilities, and education environment.

Data cleaning

For the first dataset, the PDF matrix output was transformed into tibbles, and a unified data frame was created by joining the two datasets based on the school name. This resulted in a comprehensive dataset encompassing the number and percentage of students in each racial group, along with the number of students eligible for free and reduced-price lunch, and its corresponding percentage (frpl_pct). Below are the column names of the processed data frame.

```
[1] "school_name"
 [2] "na_num"
 [3] "na_pct"
 [4] "aa_num"
 [5] "aa_pct"
 [6] "as_num"
 [7] "as_pct"
 [8] "hi_num"
 [9] "hi_pct"
[10] "wh_num"
[11] "wh_pct"
[12] "tot"
[13] "not_eligible_num"
[14] "reduce_num"
[15] "free_num"
[16] "frpl_num"
[17] "frpl_pct"
```

To prepare the second dataset, the columns of the six data files were standardized for consistency, focusing on the number of students by race. The resulting data frame includes the number of disabled children categorized by state, race (American Indian or Alaska Native, Asian, Hispanic, Native Hawaiians and Pacific Islanders, Mixed, White), and age group. The column names for the result data frame are as follows.

```
 [1] "Year"                                               "State Name"
 [3] "SEA Education Environment"                          "SEA Disability Category"
 [5] "American Indian or Alaska Native Age 3 to 5"        "Black or African American Age 3-5"
 [7] "American Indian or Alaska Native Age 6 to21"        "Black or African American Age 6 to21"
 [9] "Asian Age 3-5"                                      "Asian Age 6 to21"
[11] "Hispanic/Latino Age 3-5"                            "Hispanic/Latino Age 6 to21"
[13] "Native Hawaiian or Other Pacific Islander Age 3-5"  "Native Hawaiian or Other Pacific Islander Age 6 to21"
[15] "Two or More Races Age 3-5"                          "Two or more races Age 6 to21"
[17] "White Age 3-5"                                      "White Age 6 to21"
```

Research Questions

Aggregate Analysis Question: How are racial subgroups distributed in low-poverty schools?

Analyzing racial distribution in low-poverty schools provides insights into whether potential disparities persist in more affluent educational environments. This approach might complement studies that focus on high-poverty schools,contributing to a more comprehensive understanding of educational equity across the socioeconomic spectrum.

Longitudinal Analysis Question: What is the trend of the ethnic composition of children with disabilities in the USA from 2012 to 2017?

By answering this question, we aim to identify patterns and changes in the ethnic composition of children with disabilities. Understanding these trends would help us know whether there are disparities in the representation of disabled children across different ethnic groups. This information is crucial for ensuring that educational resources and support are distributed equitably.

Methods

Aggregate Analysis Question

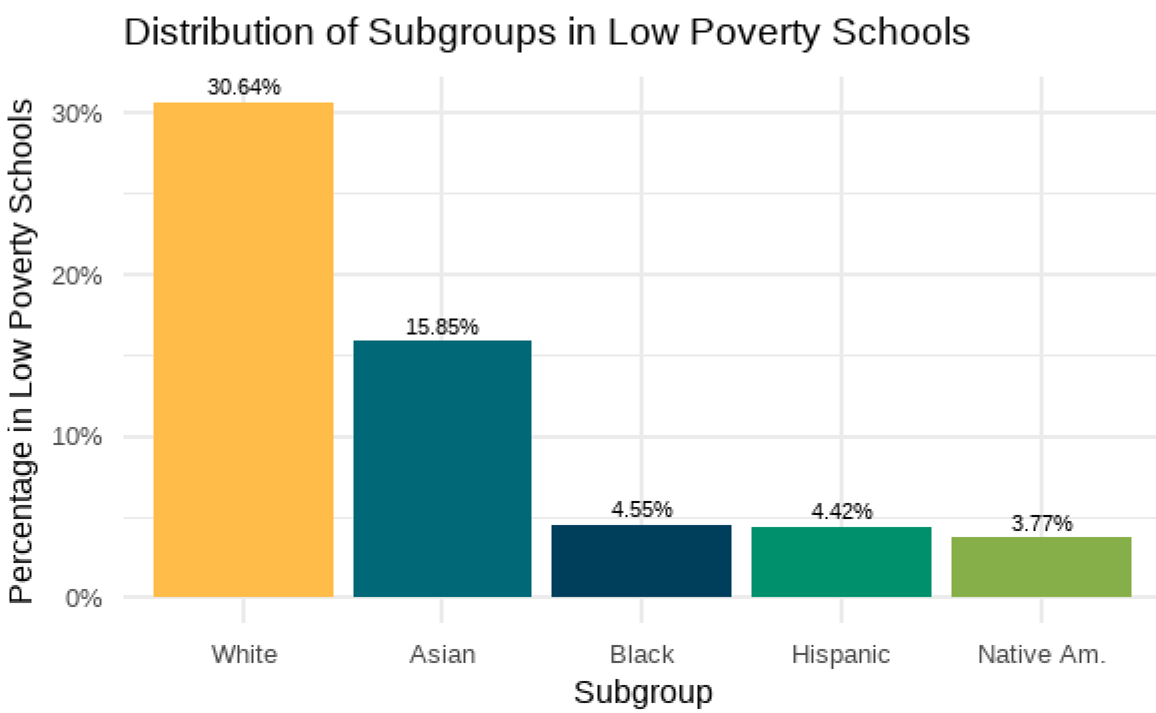
The analysis aimed to calculate the number of students in 'low poverty' schools for each racial subgroup, with 'low poverty' schools defined by NCES as those with a Free/Reduced Price Lunch (FRPL) percentage less than 25%. The process involves identifying schools with FRPL percentages below this threshold and counting the number of students for each racial subgroup in such schools. Then, to calculate the percentage of students by race who are in low poverty schools, the number of students in each race within low poverty schools is divided by the total number of students in that race.

Longitudinal Analysis Question

To conduct the longitudinal analysis, we focused on aggregating the dataset to capture the nationwide trends in the number of disabled children across different ethnic subgroups. Irrelevant variables, such as age groups and state information, were excluded to streamline the analysis. The dataset was grouped by year and race to enable the counting of the number of disabled children in each ethnic subgroup for every year across America. Then, the percentage of each specific ethnic subgroup in each year is calculated.

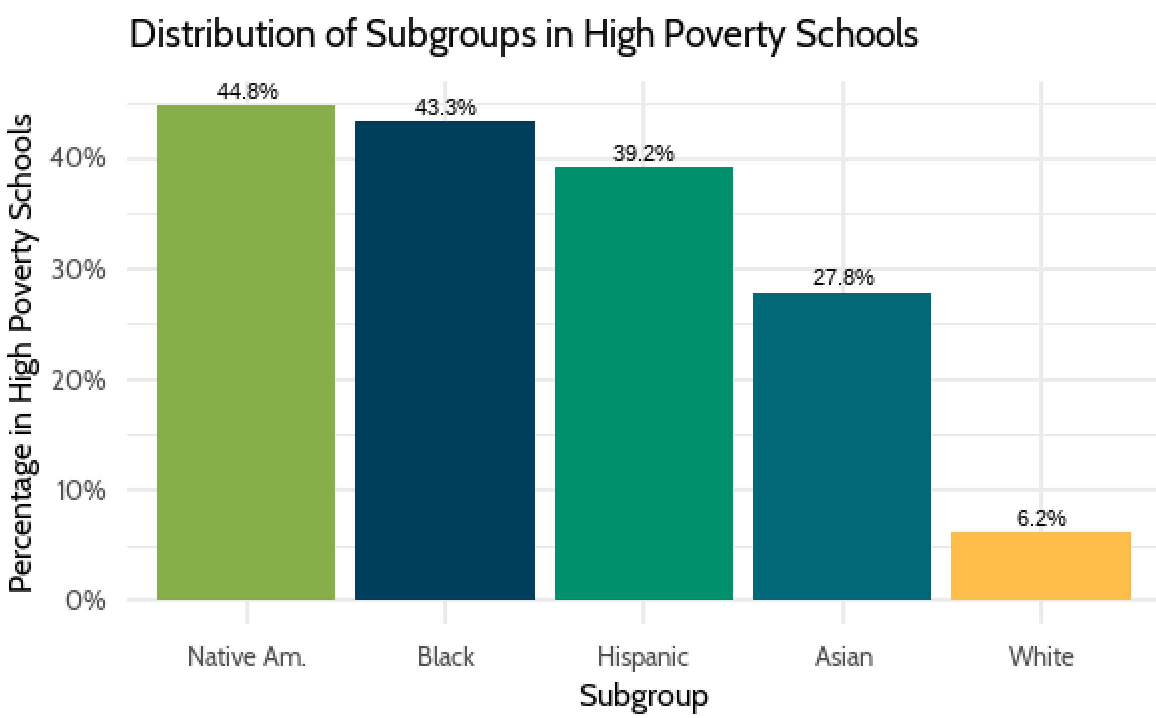
Results

Aggregate Analysis

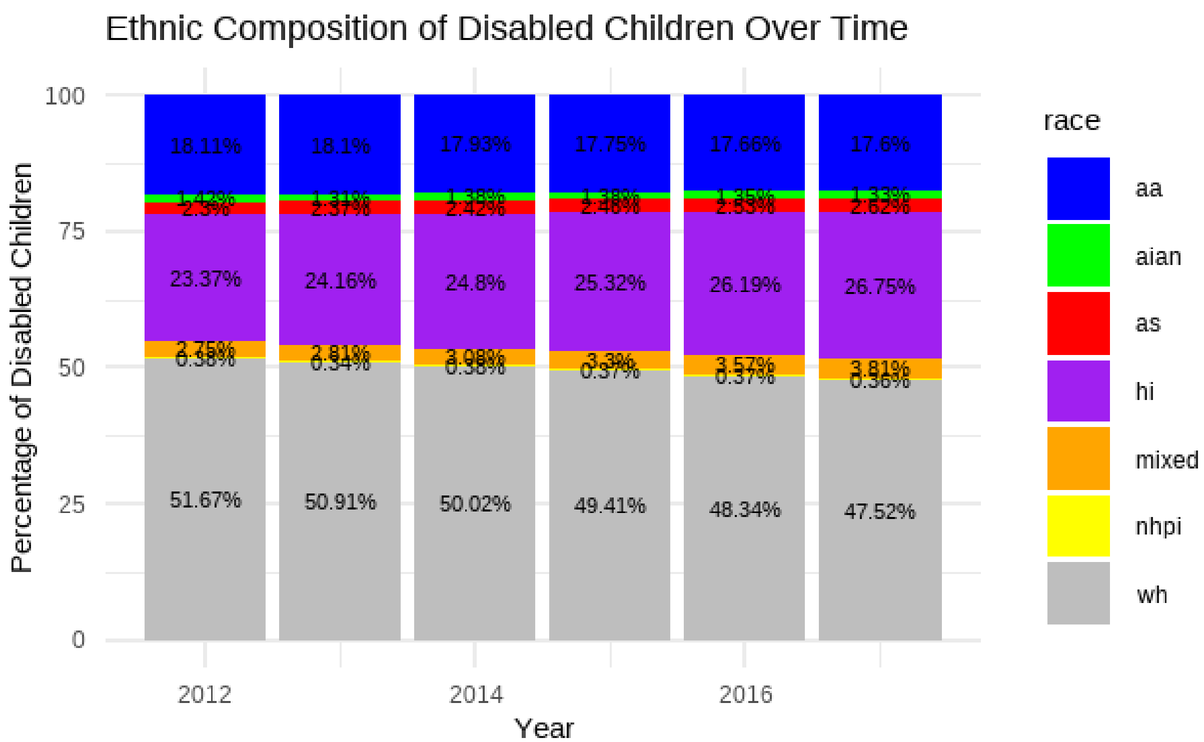


From this visualization, a clear pattern emerges: white students are significantly more prevalent in low poverty schools compared to other ethnic subgroups. Specifically, 30.64% of white students attend low poverty schools, a stark contrast to the representation of 4.55% for black students, 4.42% for Hispanic students, and 3.77% for Native American students. While the percentage for Asian students attending low poverty schools (15.85%) is not as substantial as that for white students, it far surpasses the figures for the other three subgroups.

For comparison, we can also examine the distribution of ethnic subgroups in high poverty schools. There is a notably high percentage of black, Hispanic, and Native American students attending high poverty schools, in stark contrast to the significantly smaller representation of these racial groups in low poverty schools. Particularly noteworthy is the observation that there are five times more white children attending low poverty schools than their counterparts in high poverty schools. While there is also a higher percentage of Asian children in low poverty schools compared to high poverty schools, the distinction is not as pronounced as in the case of white students. From these findings, we can see the notable disparities in the representation of ethnic subgroups across different economic contexts within the district.



Longitudinal Analysis



Examining the ethnic composition of disabled children receiving special education in the United States, a notable pattern can be observed. White disabled students are consistently over-represented across the country, constituting approximately 50% of disabled students each year from 2012 to 2017, although there appears to be a gradual decline in this percentage over time. Equally significant is the observation that the percentage of Hispanic students with disabilities averages around 25%, demonstrating a steady increase from 23.37% to 26.75% over the years. Meanwhile, the percentage for black disabled students hovers around 17-18%, exhibiting a subtle tendency to decline. Although percentages for other racial categories (Asian, American Indians and Alaska Natives, mixed race, and Native Hawaiians and Pacific Islanders) are relatively small, the most minimal representation is observed for Native Hawaiians and Pacific Islanders disabled children, with the figure consistently remaining below 0.4%.

Sources

The NCES Fast Facts Tool provides quick answers to many education questions (National Center for Education Statistics). National Center for Education Statistics (NCES). https://nces.ed.gov/fastfacts/display.asp?id=898

Ryan A. Estrellado, E. A. B. *Data science in education using R*. 9 Walkthrough 3: Using School-Level Aggregate Data to Illuminate Educational Inequities. https://datascienceineducation.com/c09

Ryan A. Estrellado, E. A. B. *Data science in education using R*. 10 Walkthrough 4: Longitudinal Analysis With Federal Students With Disabilities Data. https://datascienceineducation.com/c10