The Relationship of Student Attendance and Student Proficiency in English, Language, and Arts

Tram Anh Hoang, Haitham Anbar, Maria Coronado, Erick Njue

2022-12-05

Abstract

Introduction

We use data from the Oregon Department of Education (ODE) to examine the following research questions:

- Does attendance rate affect student proficiency in English, Language, and Arts? and;
- Does such a relationship vary for different racial and ethnic groups?

We use data of student attendance rate and student proficiency rate from the ODE assessment group reports for all the schools in the state. We will base our assessment on the latest academic year available for both data sets, which is 2021 – 2022.

Methods

In this research, we run a regression model on attendance data and proficiency data of schools in Oregon. We obtain two different data sets from the ODE: attendance rate by school and student group, and proficiency in English, Language, and Arts by student group.

Data description and preparation

We use (R Core Team 2021) package (Müller 2020) to read in the two data sets from Oregon Department of Education (ODE)into the (R Core Team 2021) environment. We use (Firke 2021) to clean the names. After preparing the data sets, we display them as tables using (Wickham et al. 2019).

The first data set is student performance in English Language and Arts (ELA) from all schools in the state. The data set provides the number and percentage of students that are proficient in each of the four levels of ELA proficiency, stratified by school, race and ethnicity, and grade level. This data set has 20046 rows and 20 columns.

The second data set is student attendance data from all the schools in Oregon. This data set provides the number and percentage of regular attenders as well as those of chronically absent students, stratified by school and either race and ethnicity or grade level. This data set has 42295 rows and 11 columns.

We clean the two data sets by removing rows with no data and rows with suppressed data, marked as "- -" and "*" respectively. We use {dplyr} package functions in (Wickham et al. 2019) to do so.

After cleaning, we merge the two data sets by matching schools' IDs and student race and ethnicity. Rows with attendance data only or proficiency data only will be removed. We use the *left_join* in the {dplyr} and *drop_na* in{tidyr}, all contained in (Wickham et al. 2019)

Methods

Results

We use the function *pivot_wider* by student group in (Wickham et al. 2019) package {tidyr} to have columns by race. This is followed by *pivot_longer* and another *pivot_wider* to make the table presentable. We use the *kable* function found in (Xie 2021) to present the results.

Table 1: Student Proficiency Rate by Race and Institution types

Institution type	Asian	Black African American	Hispanic/Latino	Multi racial	Pacific Islander	White
High School	63.16	16.27	33.39	52.18	16.95	51.50
School	64.33	19.90	25.01	51.45	15.77	50.88

To interpret our results, we further present them in a scatter plots using {ggplot2} of (Wickham et al. 2019). In Plot 1, we plot the percentage of proficient students based on the rate of regular attendance. In Plot 2 below, we use *facet_wrap* function of {ggplot2} plot the rate of regular attendance by percentage of proficient students based on the race.

Plot 1

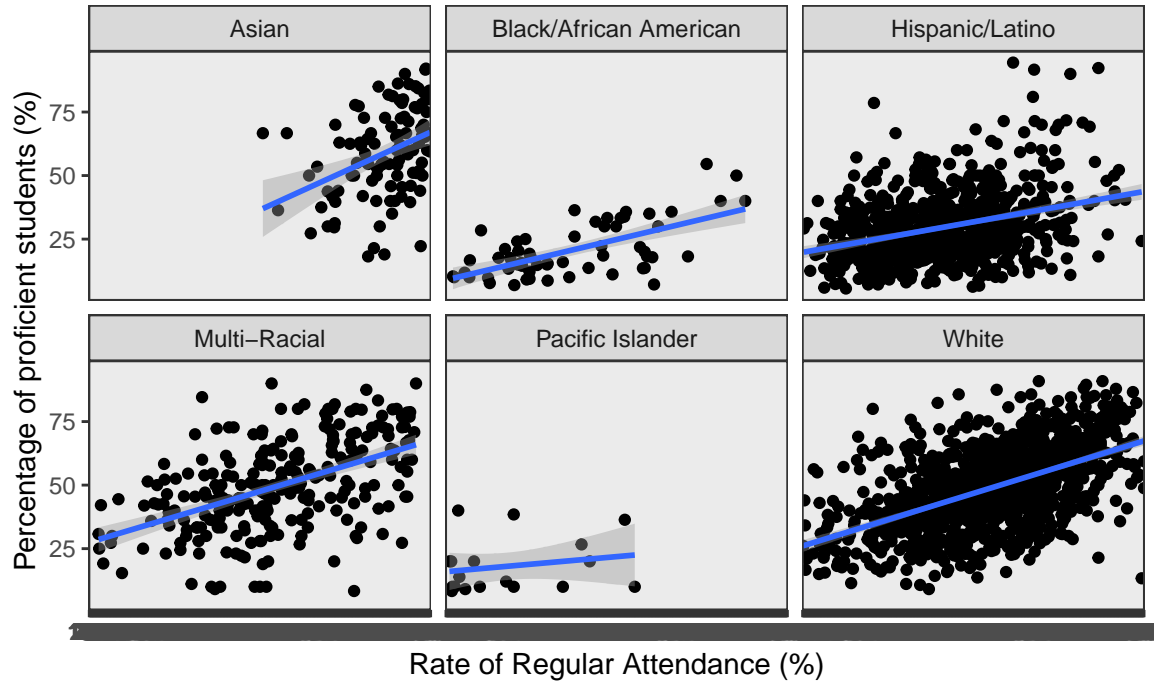Student Attendance and Proficiency Rate in English, Language, and Arts
Data from Oregon Education Department
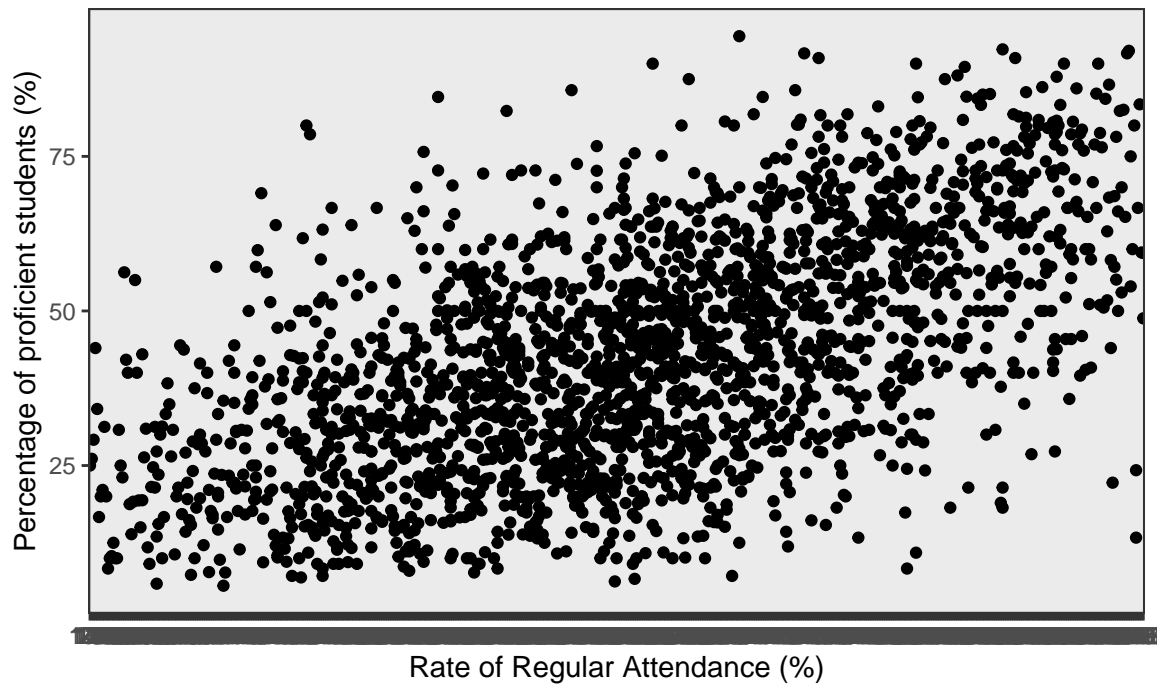
Plot 2

Student Attendance and Proficiency Rate
in English, Language, and Arts by Different Races

Data from Oregon Education Department

The estimated y-intercept is 46.57 which tells us the predicted value of proficiency when school attendance is zero.

Discussions

References

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2021. "Knitr: A General-Purpose Package for Dynamic Report Generation in R." https://yihui.org/knitr/.