

Data Paper

Absenteeism in Calaveras County, CA: Pre- and Post-COVID-19 Pandemic Dataset

Authors

Amanda Heit

Keywords

Absenteeism; Calaveras County, California; COVID-19 Pandemic

Specifications Table

| | |
|-----------------------------------|--|
| Subject Area | Education, Social Science, Psychology, Demographics |
| More Specific Subject area | Educational demographics |
| Type of data | Text file |
| How data was acquired | Extracted from source text and reformatted |
| Data format | Comma-separated values text file (.csv) |
| Data source location | Calaveras County, California, United States of America; California Department of Education |
| Data accessibility | https://github.com/AmHeit/DataSet.git |

Abstract

This dataset showcases reasons for absenteeism in Calaveras County, California before and after the COVID-19 Pandemic. The years the dataset covers are the 2017-2018 school year and the 2022-2023 school year. The dataset is composed of a spreadsheet in .csv formatting. Both unexcused and excused absences are listed, and demographic categories are used to distinguish between groups of students in the county. The dataset was initially created in separate parts by the state of California to keep track of absenteeism in the state. A small section of it was

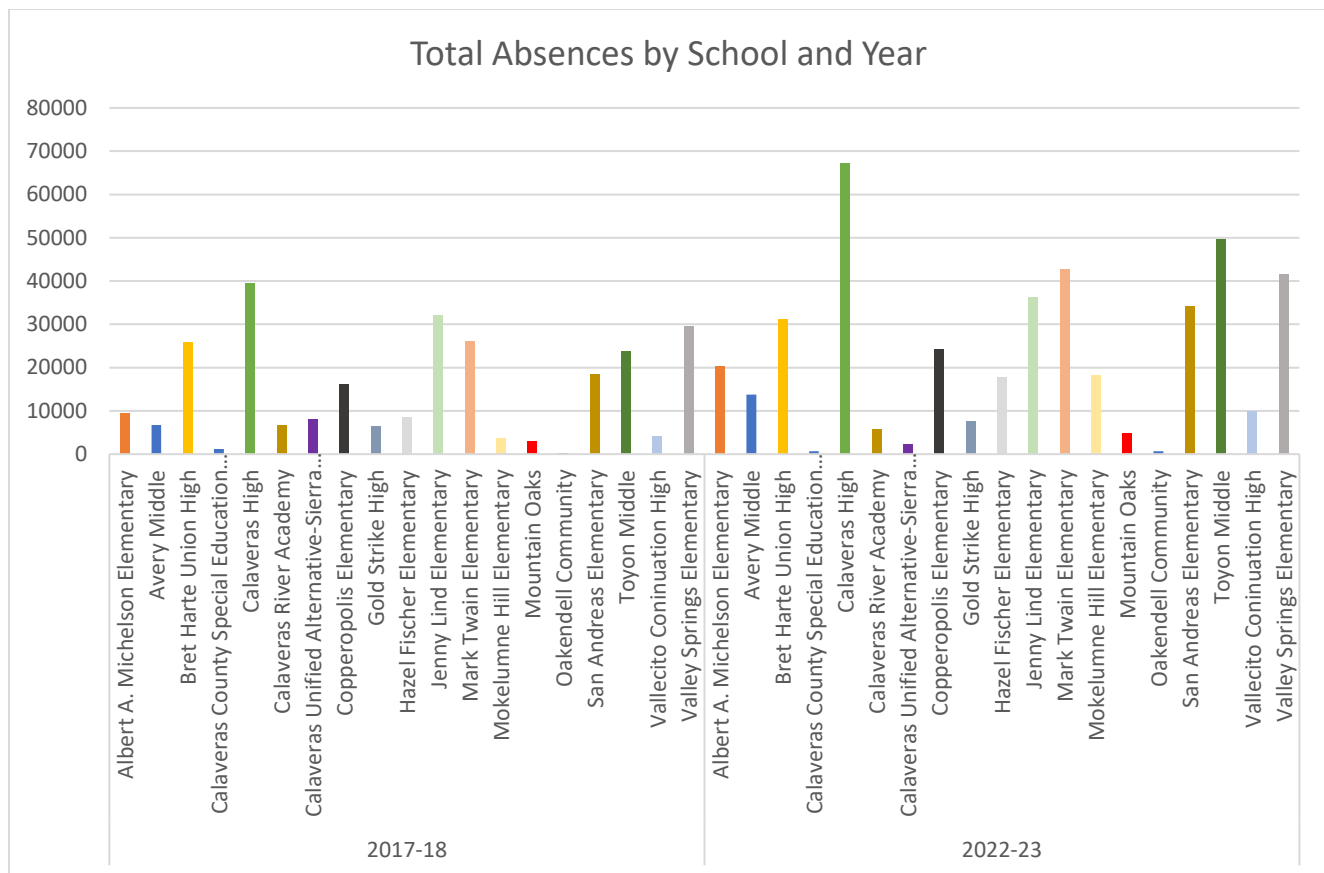
combined for comparison purposes by the author.

Overview

Chronic Absenteeism is defined in the State of California as any student that is absent for 10 percent or more of enrolled instructional days (California Department of Education, n.d.). This count does not include students who are enrolled less than 30 days. The pandemic saw an increase across the United States in absenteeism reaching up to 40 percent (nearly doubling) in some school districts (Smith, 2022). Being chronically absent creates significant barriers to success and educational equity for students making it necessary for schools to utilize data to determine individual societal needs in their area.

Data can help point out population groups that need further support in recovering. These groups experience lower grades and performance on standardized tests, higher rates of dropouts, diminished health, poor job market prospects, and increased criminal involvement (The United States Government, 2023). Since schools gain funding based on attendance and test scores, absenteeism has been calculated to have a cost. A student generates \$65.34 a day in the United States when they attend school, and with 164 million lost school days in 2022 across the country, \$10.7 billion in funding was lost (Box Pure Air, 2022). In the state of California, 30% of students (nearly 2 million) were chronically absent in 2022, while the average was 10% pre-pandemic (Jones, 2023). Reasons for missing school range from transportation to bullying, to childcare issues, to illness.

The dataset in this paper was collected to help educators and council members make informed decisions on the needs of their students in Calaveras County, California. Data comes from the California Department of Education's public yearly reports and includes data before and after the pandemic to be used as a comparison tool for individual schools in the county. Not every school in this dataset experienced an increase in absenteeism after the pandemic, but most saw changes, and Calaveras High School saw an increase of nearly 60% in absenteeism. The largest reporting categories for total absences were white, socioeconomically disadvantaged, and male for both years studied. After the pandemic, those who were chronically absent rose nearly 40% among all the schools in the county.



Description

The dataset is a single spreadsheet file (.csv) that contains a list of individual schools in Calaveras County, CA from the 2017-2018 and 2022-2023 school years. This data is taken from the original text files for the corresponding years which included data from all counties in the state. Each school is described by its type of school and includes if it is a public or charter school. Cumulative enrollment is included along with suspensions and independent study students. The absences reporting category names were rewritten focusing on clarity and inclusion, and include the following demographic groups: Not Chronically Absent, Chronically Absent, Female, Male, Grades 1-3, Kindergarten, Asian, African American, Did Not Report, Filipino, Hispanic or Latino, American Indian or Alaska Native, Pacific Islander, Two or More Races, White, Students with Disabilities, Socioeconomically Disadvantaged, and Total.

Object Name

Calaveras-absenteeism-2018-2023.csv

Language

English

Format Names and Versions

Version 1.0 Calavaras-absenteeism

Creation Date

February, 8, 2024

License

CC0-1.0 -Creative Commons public domain zero

Repository

Github: <https://github.com/AmHeit/DataSet.git>

Date of publication

March 10, 2024

Methods

Data from the mentioned school years was extracted from the California Department of Education Absenteeism by reason data (California Department of Education, 2023). This data was presented in TXT format, and so it was converted into a Microsoft Excel spreadsheet. Each school year had over 370,000 rows of data covering the whole state. Using filtering and search features, the data from Calaveras County was selected and extracted into two .csv files; one for each year. The data was then merged into a single file and cleaned with OpenRefine.

Blank and extra rows, along with inconsistent categories, were removed. Categories were inconsistent if they did not appear in the same school in both years. Often these categories included sections for grade levels not offered at the school. For quality control purposes, the following schools were removed from the list as they did not have data listed in both years of the project: John Verra High, Learners, Empowered Academic Progress, Rail Road Flat Elementary, and Vallecito Independent Study.

Schools included in this dataset include; Albert A. Michelson Elementary, Avery Middle, Bret Harte Union High, Calaveras County Special Education (SELPA), Calaveras High, Calaveras River Academy, Calaveras Unified Alternative-Sierra Hills Education Center, Copperopolis Elementary, Gold Strike High, Hazel Fischer Elementary, Jenny Lind Elementary, Mark Twain Elementary, Mokelumne Hill Elementary, Mountain Oaks, Oakendell Community, San Andreas Elementary, Toyon Middle, Vallecito Continuation High, and Valley Springs

Elementary.

Since reporting categories were hard to understand with codes like “RI,” the data was enhanced with full words, and grade levels were made consistent. For example, Kindergarten was used instead of GRKN and GRK.

The final data was saved as a .csv file named calaveras-absenteeism.csv and sorted by individual school names so that a school could locate data from both years easily. Quality control was used to ensure the same categories and identifying reasons were used for each school listed.

Reuse Potential

This dataset is useful for school administrators and researchers in comparing the impact of COVID-19 on absences at school. Schools can use the data to form programs and measures to help groups of students or grade levels. Data evaluation can help inform where funds need to be allocated along with where funds are facing a loss due to missed school days.

Data may point out if certain subgroups are experiencing more bullying than other groups to help schools direct efforts to diversity and inclusion. While working with subgroups, the data may point out problems relating to health, childcare, transportation, or mental well-being.

Researchers can reuse the data in later years to measure strategies taken and their impacts on students attending school.

The absence of certain subgroups in the data can also highlight that “what is measured matters,” and cause evaluators to seek out and find lost categories of students that could be slipping through the system.

Data could also be used in correlation with standardized test scores to determine how absences impact educational retention.

Limitations

The dataset includes information for two years before and after the pandemic. While this ensures that the same students will be sampled in that time frame, it doesn’t account for students who have graduated from high school, dropped out, or moved locations during these years.

There is also the potential of oversampling from the data if one is not careful. Some categories might include overlap with other categories that focus on race or economic status as the exact reporting of specific categories is unclear. Overlap categories include Grades 1-3, Kindergarten, Male, Female, Socioeconomically Disadvantaged, and Students with Disabilities.

Ethics Statement

Since this data involves human participants, it is important to note that no identifying information was included in the dataset. Categories that contained ten or fewer students were marked with an asterisk (*) in the dataset to protect student identity and preserve privacy.

Declaration of Competing Interests

There were no competing interests in the creation of this dataset.

Acknowledgments

The California Department of Education collected the initial TXT data for both years included in this dataset (California Department of Education, 2023). This dataset is a subset only of Calaveras County, California information.

References

- Box Pure Air. (2022, January 4). *Who pays the price when students Miss School?*. BOX PURE AIR. <https://www.boxpureair.com/blog/who-pays-the-price-when-students-miss-school>
- California Department of Education. (n.d.). *Chronic absenteeism indicator*. Chronic Absenteeism Indicator - California School Dashboard and System of Support (CA Dept of Education). <https://www.cde.ca.gov/ta/ac/cm/dashboardchronic.asp>
- California Department of Education. (2023). *Absenteeism by reason data*. Absenteeism by Reason Data - Accessing Educational Data (CA Dept of Education). <https://www.cde.ca.gov/ds/ad/filesabr.asp>
- Jones, C. (2023, September 7). *Soaring chronic absenteeism in California schools is at “pivotal moment.”* CalMatters. <https://calmatters.org/education/k-12-education/2023/09/soaring-chronic-absenteeism-in-california-schools-is-at-pivotal-moment/>
- Smith, C. (2022, July 5). *Chronic absenteeism is a huge school problem. can data help?*. Governing. <https://www.governing.com/now/chronic-absenteeism-is-a-huge-school-problem-can-data-help>
- The United States Government: Council of Economic Advisors. (2023, September 15). *Chronic absenteeism and disrupted learning require an all-hands-on-deck approach*. The White House. <https://www.whitehouse.gov/cea/written-materials/2023/09/13/chronic-absenteeism-and-disrupted-learning-require-an-all-hands-on-deck-approach/>