U.S. School Closure & Distance Learning Database

Overview and Codebook

Updated: June 9, 2021

Zachary Parolin
Bocconi University
Columbia University
Zachary.parolin@columbia.edu

Emma Lee
Columbia University

Publication Link: Nature Human Behaviour

Description: This document provides the codebook for the **U.S School Closure and Distance Learning Database**, introduced in Parolin and Lee (2020). For a full overview of data collection, please see the original paper. Here, we provide a brief recap of the data and methods. To measure school closures, we use aggregated, anonymized mobile phone data from SafeGraph. We measure foot traffic to more than 100,000 schools across nearly every county in the U.S. to evaluate how the number of visits to each school in a given month in 2020 (say, September 2020) compares to the number of visits 12 months prior (September 2019). For each month, we track year-over-year change in the number of visitors to each individual school in each month. Negligible year-over-year change in visitor counts for a given school implies that the school is operating normally and is not engaged in large-scale distance learning; in contrast, a large year-over-year decrease in in-person visits implies that the school is closed or engaged in distance learning.

Note: Estimates for 2021 are compared to visit counts in the same month in 2019 (not 2020).

In addition to measuring in-person visits to each school in each month, we identify the name of the school, its geographic location (including state, county, census tract, and census block group), and the grade levels offered at the school. In this public dataset, however, we aggregate data up to the state and county levels. Our primary indicator of interest is **the year-over-year change in total visits to a school in a given month**. For example, say that in September 2019, an average of 100 persons per day visited a given elementary school, but in September 2020, that average fell to 20 persons per day. The year-over-year change for the given school is a decline of 80 percent. Using this indicator, we classify schools experiencing a year-over-year decline of at least 50 percent as being "closed" or "mostly closed." We use the word "closure" in its *de facto* rather than *de jure* form; a school may not officially shut its doors or mandate distance learning, but if more than half the families appear to be engaged in distance learning, the school fits our definition of "school closure," or large-scale distance learning that reduces in-person attendance by at least 50 percent compared to 12 months prior. We also provide the mean year-over-year change in in-person visits, as well as estimates of the share of schools with at least a 25 percent or 75 percent year-over-year decline in attendance as an alternative cutoff point.

Suggested Citation: If using this database, please be sure to include a citation as follows:

Parolin, Z., Lee, E.K. (2021) Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nature Human Behav*iour. https://doi.org/10.1038/s41562-021-01087-8

Overview of Database: There are four datasets, depending on the level of aggregation desired: state, county, census tract, or school district for each month throughout 2020 and 2021. The codebook below describes the variables available in each file.

Codebook:

		Exists in file type:			
Variable	**Description**	**State**	**County**	**Tract**	**District**
statefips	FIPS code for each state	X	X	X	X
state_abb	State abbreviation	X	X	X	X
countyfips	FIPS code for each county (three-digit string without state prefix)		X	X	
countyfips2	FIPS code for each county (1-3 digit integer without state prefix)		X	X	
countyfips3	FIPS code for each county (4-5 digit integer with state prefix)		X	X	
county_name	Name of county		X	X	
census_tract	FIPS code for census tract (6 digit integer without state/county prefix)			X	
leaid	Local education agency identification number (NCES)				X
lea_name	District name				X
year	Year	X	X	X	X
month	Month	X	X	X	X
total_students	Total number of students in the schools in specified geographic area	X	X	X	X
number_schools	Total number of schools identified the specified geographic area	X	X	X	X
share_all_closed_[pc]	For pc=[25, 50, 75], share of **all** schools in geographic area with at least a **[pc]%** year-over-year decline in visitors in given month	X	X	X	X
mean_all_change	Mean year-over-year change in in-person visits for all schools	X	X	X	X
mean_[type]_change	For type=[elem, middlehigh], mean year-over-year change in in-person visits for school type	X	X	X	
share_[type]_closed_[pc]	For pc=[25, 50, 75] and type=[elem, middlehigh], share of school type in geographic area with at least a **[pc]%** year-over-year decline in visitors in given month	X	X	X	