

README file for replication code and data accompanying Koppa, Vijetha, and Jeremy West. "School reopenings, COVID-19, and employment." *Economics Letters*. March 2022. <https://doi.org/10.1016/j.econlet.2022.110310>

## **Overview**

With the exception of school policy data from MCH Strategic Data, all data used in this analysis is provided by public sources. Because the MCH Strategic Data are proprietary, we do not provide raw source data. However, we do provide the compiled study data along with all R-code programs (scripts) used to reproduce the study data preparation and the exhibits in the paper, including reproduction of all figures and tables in the paper and accompanying appendix.

## **Data sources**

- 1) County COVID-19 cases and deaths by date from the New York Times at <https://github.com/nytimes/covid-19-data>
- 2) County employment measurements by month from the Bureau of Labor Statistics at <https://www.bls.gov/lau/lausad.htm>
- 3) County community mobility indexes by date from Google Community Mobility Reports at <https://www.google.com/covid19/mobility/>
- 4) County boundaries shape file from the Census Bureau at <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2019.html>  
The school district and county boundaries shape files are used to merge districts to counties  
These data are also used to determine county coordinates (latitude and longitude)
- 5) School district boundaries shape file from National Center for Education Statistics at <https://nces.ed.gov/programs/edge/Geographic/DistrictBoundaries>
- 6) County and state COVID-19 policy ordinances from Department of Health and Human Services at <https://healthdata.gov/dataset/COVID-19-State-and-County-Policy-Orders/gyqz-9u7n>  
  
We also use a crosswalk of state abbreviations to county FIPS codes from Kieran Healy at <https://github.com/kjhealy/fips-codes>
- 7) County population estimates from the Census Bureau at <https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html>
- 8) County Gross Domestic Product for 2019 from the Bureau of Economic Analysis at

<https://apps.bea.gov/regional/downloadzip.cfm>

- 9) County Small Area Income and Poverty Estimates (SAIPE) for 2019  
<https://www.census.gov/data/datasets/2019/demo/saipe/2019-state-and-county.html>
- 10) School district policies for the 2020-2021 school year from MCH Strategic Data at  
<https://www.mchdata.com/covid19/schoolclosings> (account required)
- 11) County COVID-19 test volume by date from the Carnegie Mellon University Delphi Group at  
<https://delphi.cmu.edu/covidcast/export/>
- 12) County demographic labor force participation from the American Community Survey 2019 5-year from the Census Bureau at  
<https://data.census.gov/>  
Using selected labor force participation data from table S2301
- 13) County demographic characteristics from the American Community Survey 2019 5-year from the Census Bureau at  
<https://data.census.gov/>  
Using selected demographic data from table DP05
- 14) County 2019 health care characteristics from the Centers for Medicare and Medicaid Services at  
<https://data.cms.gov/mapping-medicare-disparities>  
Using 2019 county influenza vaccine rates
- 15) County 2019 total and ICU hospital bed counts from ArcGIS (ESRI) at  
<https://nation.maps.arcgis.com/home/item.html?id=58c01cc29cad449baa356049f4d53388>

## **Description of Data Preparation and Analysis code**

Below we provide a description of the R-code program files used to prepare the data and used in the analysis. We describe what each code file does and provide a list of output files saved by each R program. Each code file has a descriptive title that includes a number. The data preparation code (numbered 1) is provided only for reference, because some of the raw source data is proprietary as discussed above. The resulting prepared data is included in the provided data archive. The remainder of the R-code files (numbered 2 and 3) prepare and export the full set of figures and tables, respectively, included in the paper and appendix. These files can be run in any numerical order. The specific exhibits generated by each code file are listed below.

## **Computational requirements**

Hardware: The authors executed the R-code on a 64-bit desktop computer running a Linux operating system, but any current desktop or laptop computer with at least 4GB of RAM should be adequate to run the code files, using either Windows, MacOS, or Linux.

Software: R version 4.1 was used to conduct this analysis, although other versions of the software might also function properly with the provided code files.

In addition, the code files make use of the following user-added libraries/packages:

- R package augsynth version 0.2.0
- R package broom version 0.7.12
- R package cowplot version 1.1.1
- R package data.table version 1.14.2
- R package dplyr version 1.0.7
- R package fixest version 0.10.1
- R package ggplot2 version 3.3.5
- R package gridExtra version 2.3
- R package haven version 2.4.3
- R package huxtable version 5.4.0
- R package lwgeom version 0.2-8
- R package readxl version 1.3.1
- R package sf version 1.0-6
- R package usmap version 0.5.2
- R package zoo version 1.8-9

Run-time: approximately 1 hour from start to finish for all code used in the study, although the runtime will vary greatly with CPU performance of the hardware used to execute the code files.

## **Code for data preparation**

This file is provided only for reference, because some of the raw source data is proprietary as discussed above. Users who are only interested in reproducing one or more of the figures and tables do not need to run this data preparation file, as the output “county\_panel.csv” file is included in the provided data archive.

### ***1\_Compile\_Data.R***

Description: downloads (if necessary) most of the raw data input files from public sources, and compiles all input data files into the panel of data for the study analysis. The input file data sources are listed above and also described in the R-code file.

Input files:

- [multiple] see code description section and/or data source list above.

Output files:

- county\_panel.csv
- county\_panel.dta (Stata version 13)

## **Code for reproducing all figures and tables**

These files reproduce all figures and tables included in the paper, and may be run in any order.

### ***2\_Figures.R***

Description: generates all figures for the paper and appendix.

Input file:

- county\_panel.csv

Output files:

- Figure 1:** Figure01-counties-map-school-policy.pdf (also saves a version as .png)
- Figure 2A:** Figure02A-event-study-daily-covid-deaths-per-million-campus.pdf
- Figure 2B:** Figure02B-event-study-daily-covid-deaths-per-million-hybrid.pdf"
- Figure 3:** Figure03-line-graph-employment-school-policy.pdf
- Figure B1:** FigureA01-line-graph-school-started-school-policy.pdf
- Figure B2:** FigureA02-line-graph-mobility-school-policy.pdf
- Figure B3A:** FigureA03A-line-graph-covid-cases-school-policy.pdf
- Figure B3B:** FigureA03B-line-graph-covid-deaths-school-policy.pdf

### ***3\_Tables.R***

Description: generates all tables for the paper and appendix. Also conducts robustness checks using synthetic control estimation, though no files are saved for this.

Input file:

- county\_panel.csv

Output files:

- Table 1:** Summary\_Statistics.tex
- Table 2:** Regressions\_Employment.tex
- Table B1:** Regressions\_Covid.tex
- Table B2:** Regressions\_Covid\_States\_Any\_Campus.tex
- Table B3:** Regressions\_Covid\_Monthly.tex
- Table B4:** Regressions\_Employment\_States\_Any\_Campus.tex