Campos and Kearns
"The Impact of Public School Choice: Evidence from Los Angeles' Zones of Choice"
Replication README

Overview
The code in this replication package constructs the analysis files from a total of 212 files provided by LAUSD. The replication process proceeds in three steps:

1. The main build – creates a panel of LAUSD student data covering 2002-2021
2. The analysis build – using the panel created in step 1, this step creates analysis data that are used to create the main figures and tables in the paper, as well as appendix material
3. The analysis – uses the data created in step 2, to create the main figures and tables in the paper

The code in the replication uses a combination of R and Stata code. The code takes about 3 hours to run in its entirety.

Data Availability and Provenance Statements

The data used for this paper can be accessed by submitting a formal request to the Los Angeles Unified School District (LAUSD) through their standard portal. Any interested researcher can obtain a Data Use Agreement (DUA) with the Los Angeles Unified School District. They must first submit a formal request in their standard portal. It can take some months to negotiate data use agreements and gain access to the data. The author will assist with any reasonable replication attempts for two years following publication.

Details about each data source:

Data Group	Number of files	Example name:	Provided
Demographics and enrollment	20	demo2002.csv	No
CST Scores	14	CST 2003.csv	No
SBAC Scores	6	sbac2015.csv	No
Suspensions	1	suspensions.csv	No
GPA	20	GPA 2002.csv	No
SAT	11	SAT 2007.txt	No
UC/CSU Indicators	15	Uc2004.csv	No
Grades for each course	20	grades2002.csv	No

Attendance	20	Secondary Attendance 2002.csv	No
Census Blocks	1	Addresses_geocoded2001_2019.dta	No
Census Blocks with Zone Information	1	Blocks_with_zones.csv	No
ZOC School list	1	Zoc_high_schools_list.csv	No
Rank-ordered lists (ZOC)	1	zocchoices.csv	No
ZOC Assignments	1	zocassignments.csv	No
Crosswalk between LAUSD and California Department of Education (CDE) identifiers	1	Plocn_cde_xwalk.dta	No
CDE School Directory (includes school grid coordinates)	1	Pubschls.xlsx	No
National Student Clearinghouse	2	National_Student_Clearinghouse_2009_2013.csv	No
HS Attendance Zones for Census Blocks (pre-period)	1	Census_block_id_hs_attendance_zones.csv	No
Census Tract Income	1	La_06037_census_tract_income.csv	No
Attendance Zone Boundary Shape File	1	zone_boundaries_zoc_included.shp	No

Computational Requirements

Stata (code was last run with version 18): the program "`0_installpackages.do`" will install all dependencies locally, and should be run once.

R (code was last run with version 4.3.1): any R scripts that are called within the analysis do-files install any relevant packages

The maximum RAM used across the replication is roughly 70GB and the replication takes roughly 3 hours to run.

Descriptions of programs/code

- The script master.do calls every other script and reproduces the analysis.
- The script in code/0_installpackages.do installs all the relevant Stata packages.
- The script in code/0_installpackages.R installs all the relevant R packages.
- The programs in code/0_buildLAUSD.do read in the raw data and construct a student-level panel with a variety of outcomes and student demographic information.

    - 34 minutes to run
- The programs in code/01_buildZOC.do construct a variety of analysis datasets.
    - Takes 1-2 hours to run
- The programs in code/2_ANALYSIS_ZOC.do creates all the relevant tables and figure content in the paper.
- The script in code/makeZOCMap.R creates one figure in the paper and is called within code/2_ANALYSIS.do
    - Note that this script pings Census API. We removed our API key from the script so a replicator has to add their API key in order to ping the Cenus API.
- The script in code/zoc_va_qte.R creates data used to create on figure in the paper and is called within code/2_ANALYSIS.do

Instructions to replicators

Submit a proposal to LAUSD to and enter in a data use agreement with the district. They can then share the relevant data files and you can replicate the analysis. With all the relevant files a replicator should run the master.do script after updating file paths in all the relevant scripts.

List of tables and programs

Figure/Table #	Program	Note
Figure 1	2_ANALYSIS_ZOC.do – zocMap	Depends on tidycensus package in R to obtain Census tract attributes. Replicators should get an API key.
Figure 2	2_ANALYSIS_ZOC.do – choseNonNeighborhoodDescriptive	
Figure 3	2_ANALYSIS_ZOC.do – Event_study College_event_study Event_study_itt College_event_study_itt	Each panel of Figure 3 is created by a separate program within 2_ANALYSIS_ZOC.do
Figure 4	2_ANALYSIS_ZOC.do – vaQTE	Calls an R program
Table 1	2_ANALYSIS_ZOC.do – descriptive_stats	
Table 2	2_ANALYSIS_ZOC.do – deltaRegressions	
Table 3	2_ANALYSIS_ZOC.do – ovgRobustness	
Table 4	2_ANALYSIS_ZOC.do – studentBehaviorES	