Data Guide
The Evolution of Segregation and Education
Prepared for EC970, Harvard University Department of Economics
Spring 2024

Danielle Graves

The goal of this resource is to orient you with the publicly available datasets to researchers. Often, these datasets are linked together.

Education researchers also often work with student-level administrative data for individual states. This data has to be applied for and worked with in a secure environment to ensure confidentiality of students. Permission is granted at the discretion of state departments of education.

★ sources appear frequently in the literature. We'll call them the "As Seen on TV" sources. Blue indicates a clickable link.

School Spending and Enrollment

- ★**Non-Fiscal Common Core of Data (CCD Non-Fiscal)**
    - Outcomes: "school types, enrollment, demographics, and more"
    - Coverage: "all public elementary and secondary schools and school districts"
    - Frequency: Annual
    - Years available: 1986-2023
- ★**Fiscal Common Core of Data (CCD Fiscal)**
    - Outcomes: "revenues and expenditures"
    - Coverage: "public elementary and secondary schools, school districts, and state departments of education"
    - Frequency: Annual
    - Years available: 1986-2023
- ★**Integrated Postsecondary Education Data System (IPEDS)**
    - Outcomes: "institutional characteristics; institutional prices; admissions; enrollment; student financial aid; degrees and certificates conferred; student persistence and success; and institutional resources including human resources, finances, and academic libraries"
    - Coverage: post secondary institutions (community colleges and 4 year colleges)
    - Frequency: Annual
    - Years available: 1999-2023
- **County-level school enrollment and resources in ten segregated Southern states**
    - Outcomes: enrollment, attendance, number of schools, number of teachers, revenue by source, expenditures, salaries
    - Level: County (often equivalent to district in the south). Disaggregated by race because schools were segregated in this time and place
    - Available years: 1910-1940
    - As seen in: Carruthers, Celeste K., and Marianne H. Wanamaker. "Returns to School Resources in the Jim Crow South." Explorations in Economic History 64 (April 2017): 104–10. https://doi.org/10.1016/j.eeh.2017.02.004

Student Achievement

- ★**Nation's Report Card (NAEP)**
    - Outcomes: major subject areas (math, reading, writing, science)
    - Level: National and state. Some urban districts since 2002

- Available years: 1990-2023
- Ages: Varies by subject area
- **High School Transcript Study (HSTS)**
  - Outcomes: course-taking and GPA
  - Level: National
  - Available years: 1990, 2000, 2005, 2009, 2019
  - Ages: high school seniors who graduated from high school
- **Long Term Trends**
  - Outcomes: math and reading assessments
  - Level: National
  - Available years: 1970-2023
  - Ages: 9, 13, 17
- ★ **Stanford Education Data Archive**
  - **Documentation**
  - Outcomes: test scores. Includes covariates, including demographic information
  - Level: school-district and county
  - Available years: 2009-2018 ("Educational Opportunity"), 2019-2023 ("Educational Recovery")
  - This data pulls from and combines NAEP and EDGE (see below)

School District and County Demographics

These data sources can be used as controls or to study heterogeneous effects across populations.

- **Education Demographic and Geographic Estimates (EDGE)**
  - Outcomes: characteristics of parents and children from the ACS (2005-2017) and community level characteristics from the decennial census (1990-2020)
- ★ **IPUMS NHGIS**
  - Data from the Census and ACS, compiled at the geographic level of your choice
  - *Note: data is not available at the school district level. For this, use EDGE. If you need historic data, there are school district x county crosswalks available.*

Segregation

- ★**Office of Civil Rights Data (CRDC)**
  - Outcomes: course types, staff, teachers, bullying, student discipline, class size, resources.
  - Level: school and district. Often "disaggregated by race/ethnicity, sex, disability and whether students are English Learners."
  - Available years: 1968-2020. Historical data (1968-1998) may be a different format than modern data.
- **Groeger, Waldman and Eads. "Miseducation: Is There Racial Inequality at Your School?" *Pro Publica*. 2018**
  - *Great interactive data tool for understanding the different manifestations of school segregation*
- **Mapping Inequality**
  - *Interactive data tool*
  - "In the 1930s the federal government created redlining maps for almost every major American city. Mapping Inequality lets you explore these maps and the history of racial and ethnic discrimination in housing policy."
- **Segregation Contribution Index**

- *Interactive data tool*
- Allows researchers to explore how much individual schools "contribute to segregation" by comparing school district demographics and school demographics'

Court Orders

Court orders are not the simple exogenous variation we hope them to be. First, court *cases* do not always lead to a court *order*. Second, the date of a court *order* does not necessarily coincide with the date of *implementation* of that order. Third, sources do not always agree about the timing of these events. Many scholars end up cross referencing many sources to determine the timing of the events of interest in their areas of study. These sources are listed below.

Comprehensive datasets

These list all court *cases*. Cleaning is needed to determine (1) the relevant court case (often districts have multiple court cases levied against them) and (2) which court cases led to orders.

- **Pro Publica**
- **American Communities Project (ACP)**
- **United States Commission on Civil Rights**
  - *This is a report that contains helpful context on each court case*

Researcher compiled datasets

"Tried and true" ... mostly

These are the safest datasets to use out-of-the-box. However, they are often limited in coverage. Make sure to check the data section of the paper to understand the limitations of that data. Often, researchers will catch small errors in other researchers' data. Keep this in mind!

- **Anstreicher, Fletcher, and Thompson (2022)**
- **Reber, S. "Court Ordered Desegregation: Successes and Failures Integrating American Schools since Brown versus Board of Education."** Journal of Human Resources, 2005, vol. 40, issue 3.
  - *See Table A1*
- **Reardon, Sean F., Elena Grewal, Demetra Kalogrides and Erica Greenberg. (2012). "Brown Fades: The End of Court Ordered School Desegregation and the Resegregation of American Public Schools."** Journal of Policy Analysis and Management, 31(4): 876-904.

Exceptional Replication Packages

A great way to get started using education data sources is by looking at replication packages of papers you like. These replication packages are generally available on ICPSR and occasionally available on an author's website.

- **Thompson, O. "School Desegregation and Black Teacher Employment."** (2022) Review of Economics and Statistics, 104(5), 962-980.
- **Cascio, Elizabeth U, and Sarah Reber. "The Poverty Gap in School Spending Following the Introduction of Title I."** American Economic Review 103, no. 3 (May 2013): 423–27.
- **Johnson, Rucker C., and C. Kirabo Jackson. "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending."** American Economic Journal: Economic Policy 11, no. 4 (November 1, 2019): 310–49.

Bonus Resources

The **Pro Publica Education Datasets** contain links to a wide variety of free, publicly available data. Instead of listing them all here, I encourage you to explore the data store for yourself. You might just get a research paper idea out of it!