LIFE CYCLE OF

COLLECTION, EXPLORATORY DATA ANALYSIS, AND CLEANING

Obtain School-by-School Spending Data File

Download, scrape, request from state contact, or submit a FOIA request

Post Raw File for Download on NERD$ Website

Exploratory Data Analysis

a. Confirm data meet reporting requirement

b. If missing, add enrollment, state IDs, NCES IDs, and Census ID

c. Reasonableness check

Basic Checks: Cleaning Round 1

a. Convert into NERD$ template and variable names

b. Run R script for data cleaning to highlight high-level anomalies such as empty cells, missing schools, unusual values

c. Add NRD (No Data Reported) designation in empty cells

Improve Usefulness: Cleaning Round 2

a. Add variables from Common Core of Data

- NCES enrollment
- Grade Level
- Charter status
- Locale
- Magnet status
- Title 1 status
- FRPL status

b. Standardize state-provided file and CCD data:

- State ID formats
- District and school names
- Ensure all CCD schools are included

Post Cleaned Data File for Download on NERD$ Website

INTERNAL AND EXTERNAL DATA VALIDATION

Validate Data by School

a. Run R script to detect concerning values:

Spending outside of a range considered typical (too high or too low)

High federal spending relative to state/local

High central share

Enrollment discrepancies between state, NCES, and CCD enrollment counts

b. Review online sources to check for reasonableness

c. Check report card for updated data

d. Generate **flag_NERDS**

Validate Data by District

a. Roll up total reported spending by district

b. Adjust for state-published exclusions

c. Compare to F33 Total Current Expenditures

d. Check district financial files

e. Verify exclusions with the state

f. Generate **flag_F33** for districts with concerning values

g. Review school values in flagged districts, adjust **flag_NERDS** if needed

STANDARDIZATION, DOCUMENTATION, AND PUBLICATION

Create Normed Variable to Allow for Cross-State Comparison

a. Adjust to ensure comparable expenditure categories

b. Adjust for differences in the way states count enrollment

c. Generate normed school value **pp_total_norm_NERDS**

Verify

a. Confirm data file is complete

b. Finalize documentation and code book

Post Data File to NERD$ Database and Website for Quick Download

NERD$ National Education Resource Database on Schools is a joint effort of Georgetown University's Edunomics Lab and Massive Data Institute. This work is funded in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D200032 to Georgetown University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. https://edunomicslabhttps://edunomicslab..org/nerds/org/nerds/