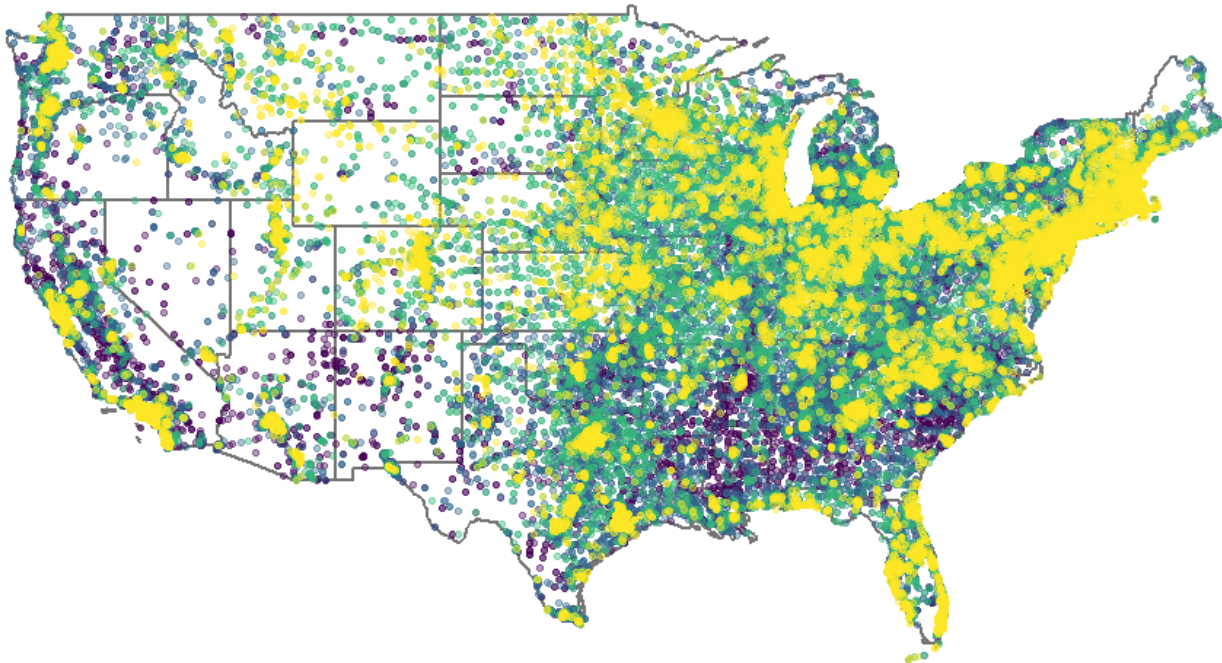# PREDICTING EDUCATIONAL OPPORTUNITY:
## A METHODS COMPARISON USING POPULATION DATA

CLAIRE ALLEN-PLATT
STAT 571 DATA SCIENCE LIVE
APRIL 2021

# Measuring educational opportunity

Starting in 2002, all schools in the United States were mandated to implement annual standardized tests of reading and math in Grades 3-8. Each state administers its own test, and tests sometimes change from year to year, troubling researchers' ability to make national comparisons of student achievement.

# Measuring educational opportunity

In 2016, Reardon et al. linked every state's test data to a common national scale and published a public-use dataset, the Stanford Education Data Archive (SEDA), which is regularly updated.

SEDA reports an average reading and average math score for over 80,000 elementary and middle schools in every public school district in the United States. Each school has a single pooled estimate of reading and math achievement, pooled across grades 3-8 and the 10-year period SY 2008-09 to 2017-18. The dataset combines approximately 450 million standardized math and reading tests and represents about 57 million unique students.

Sources: The Educational Opportunity Project at Stanford University, 2021; Reardon et al., 2016
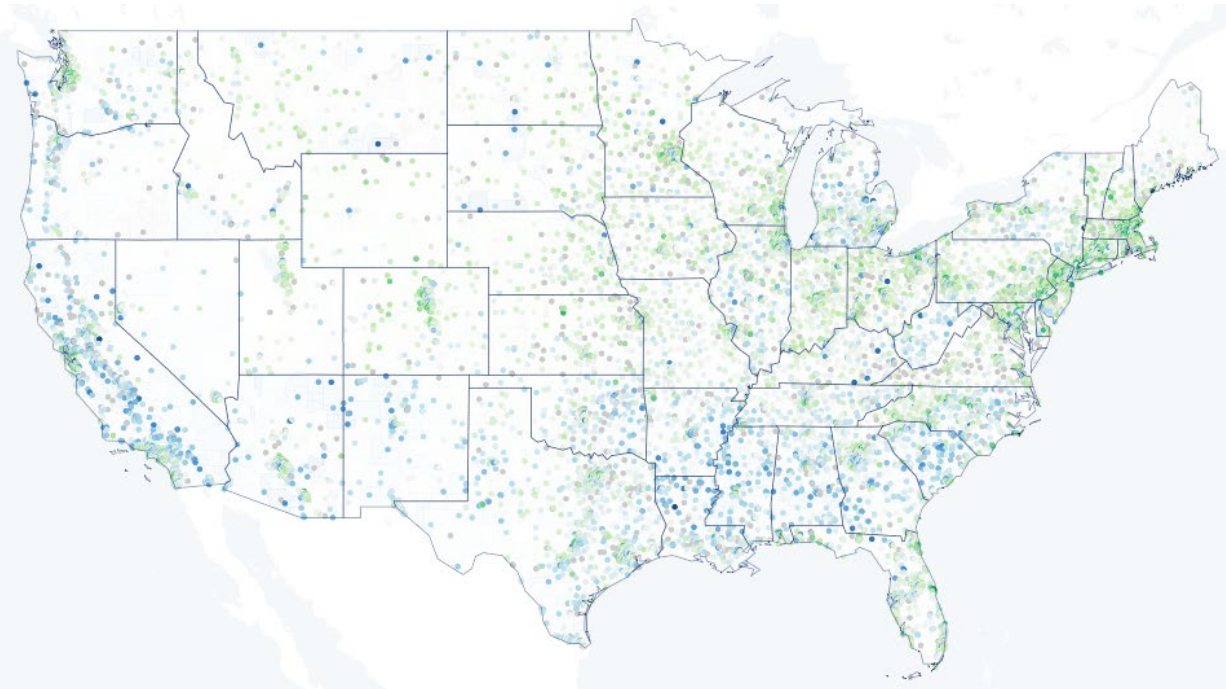
Average Test Scores By School
in grade levels, relative to the U.S. average

-5       U.S. average       5

# Measuring educational opportunity

Most standardized achievement tests assign each test taker a numeric score on a continuous scale. States then coarsen the numeric scores into an ordinal category designation, such as "below basic," "basic," "proficient," and "advanced," based on score thresholds designated by a committee of teachers and other experts.



Sources: Shear & Reardon, 2020
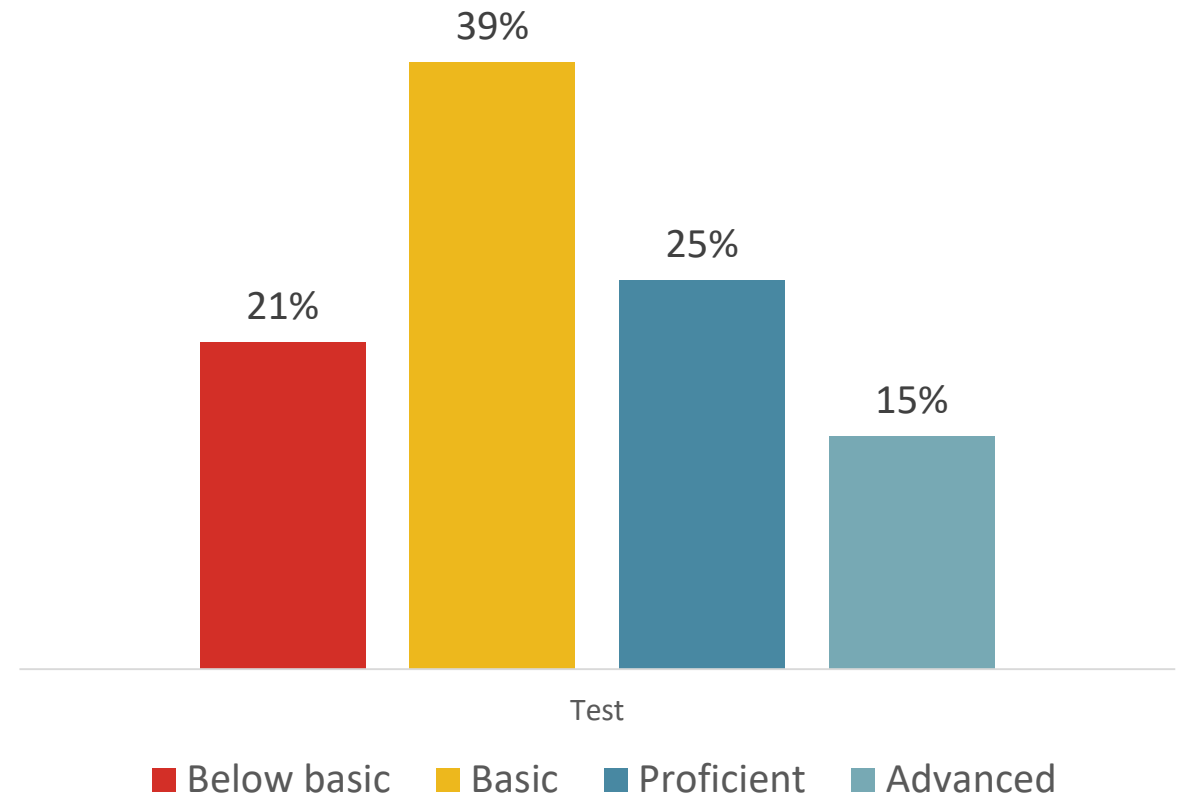
# Measuring educational opportunity

Reardon (2019) argues that the average standardized test score in a school or district "can be thought of as reflecting the average cumulative set of educational opportunities children in a community have had up to the time when they take a test" (p. 41).
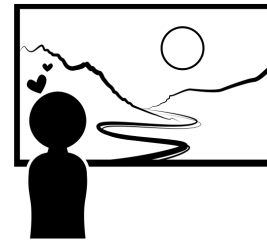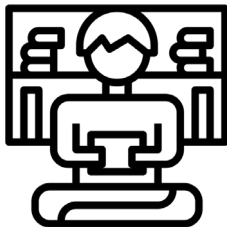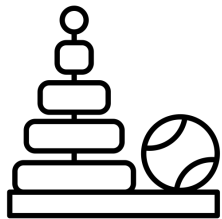
"Educational opportunities" can be thought of as *inputs* to the education process, such as opportunities for informal learning or enrichment at home or in the community; the quality and availability of early childhood care or preschool; housing, financial, and food security; and the availability of resources like libraries or reliable Internet.
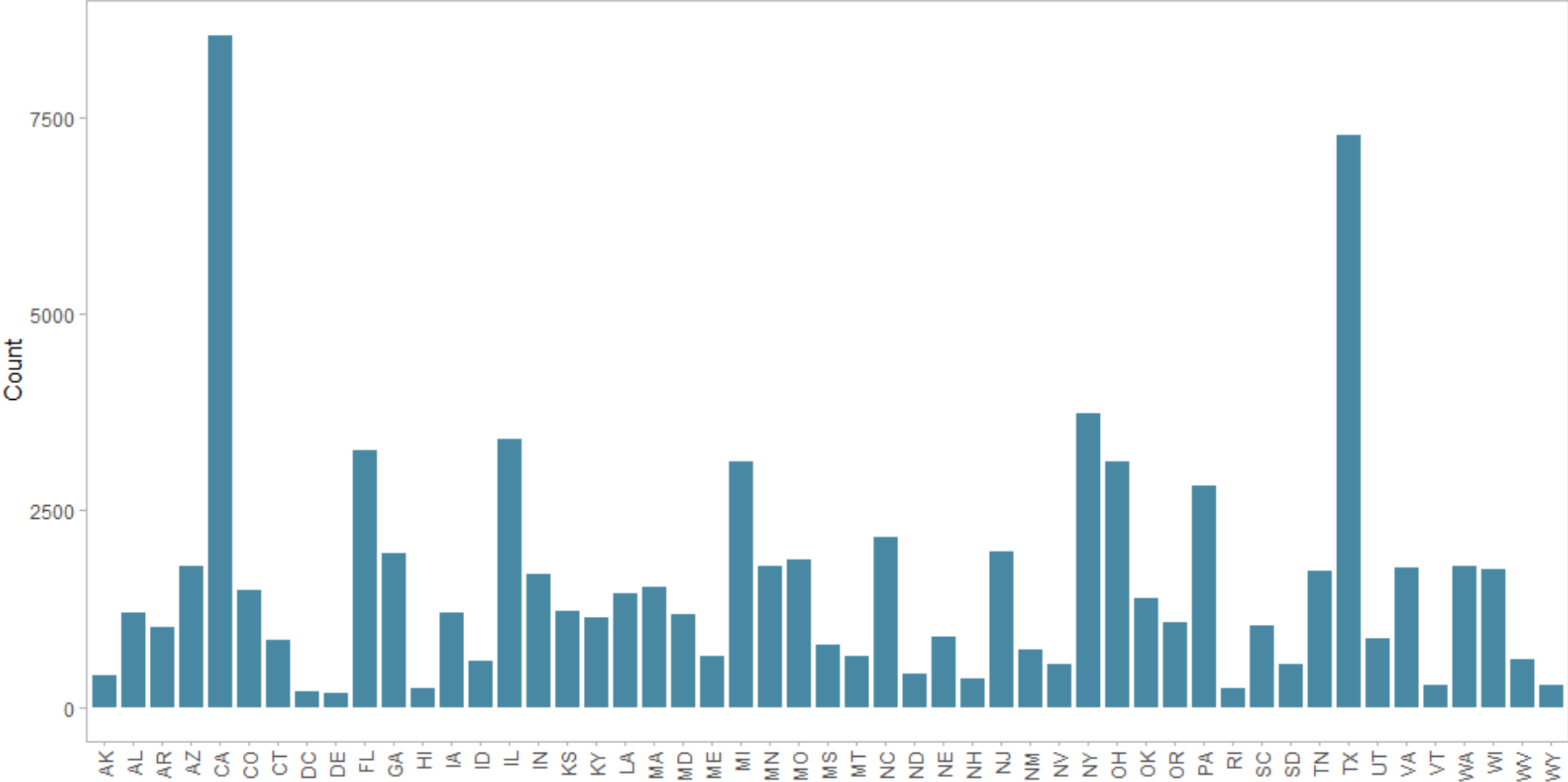
# Goals of study

Methodological and substantive:

1. How well can we investigate hypotheses about educational opportunity with public-use data?
2. What modeling techniques are well-suited to investigating hypotheses about educational opportunity with public-use data?
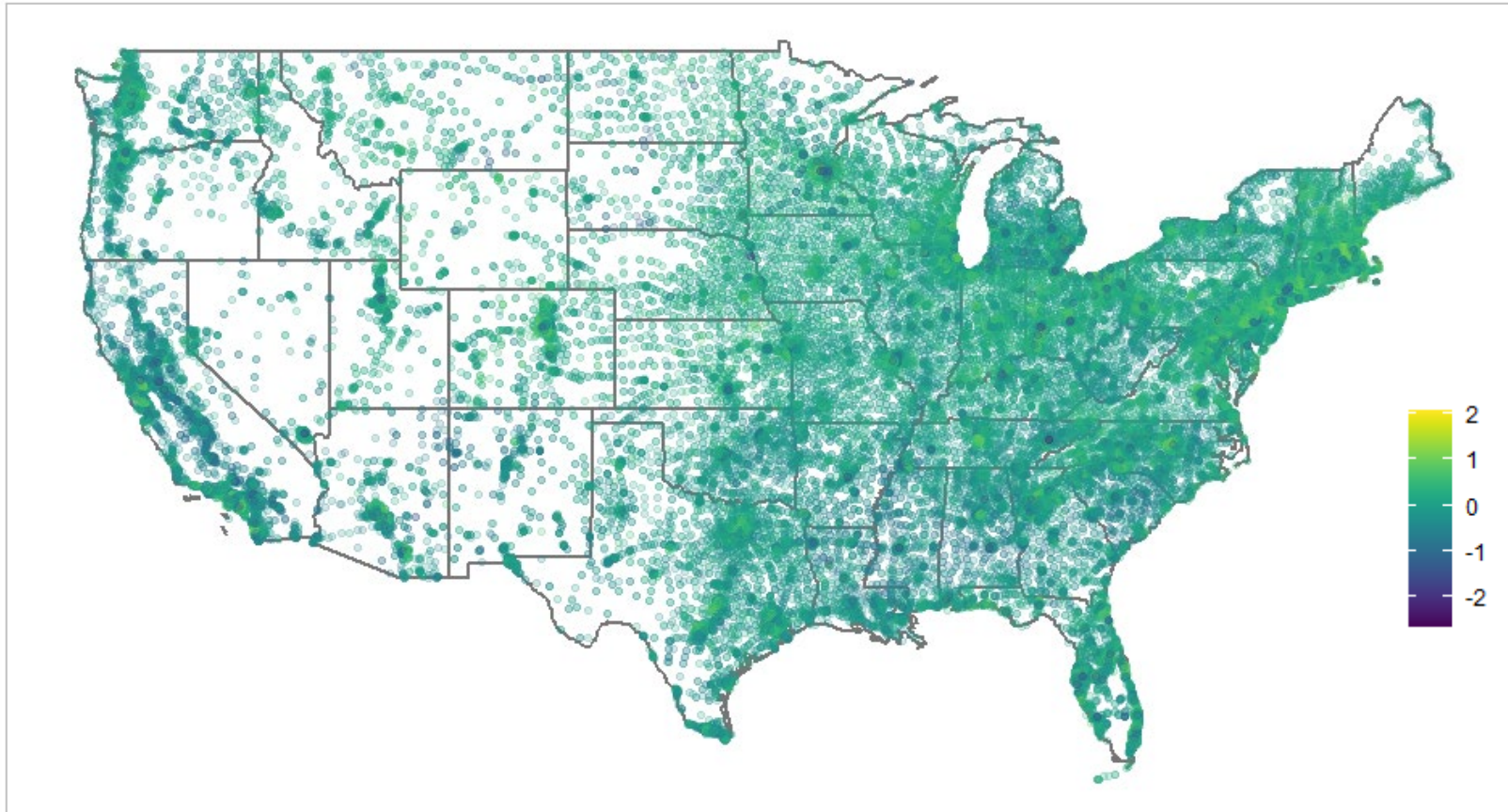
# Data



**Count of elementary and middle schools per state in SEDA dataset**
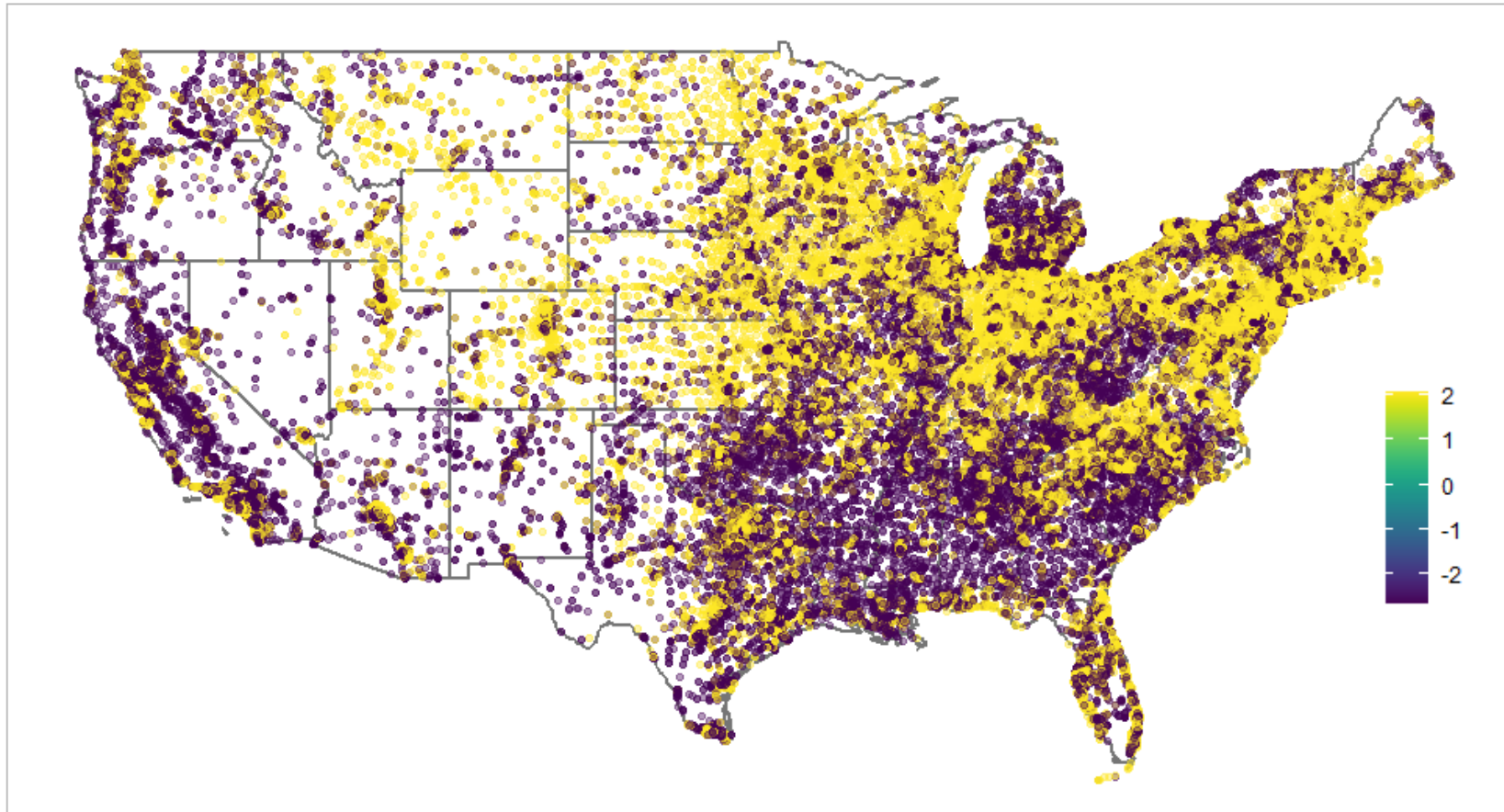Data source: SEDA

# Outcome variable: continuous



**Average achievement of every elementary and middle school in the United States**
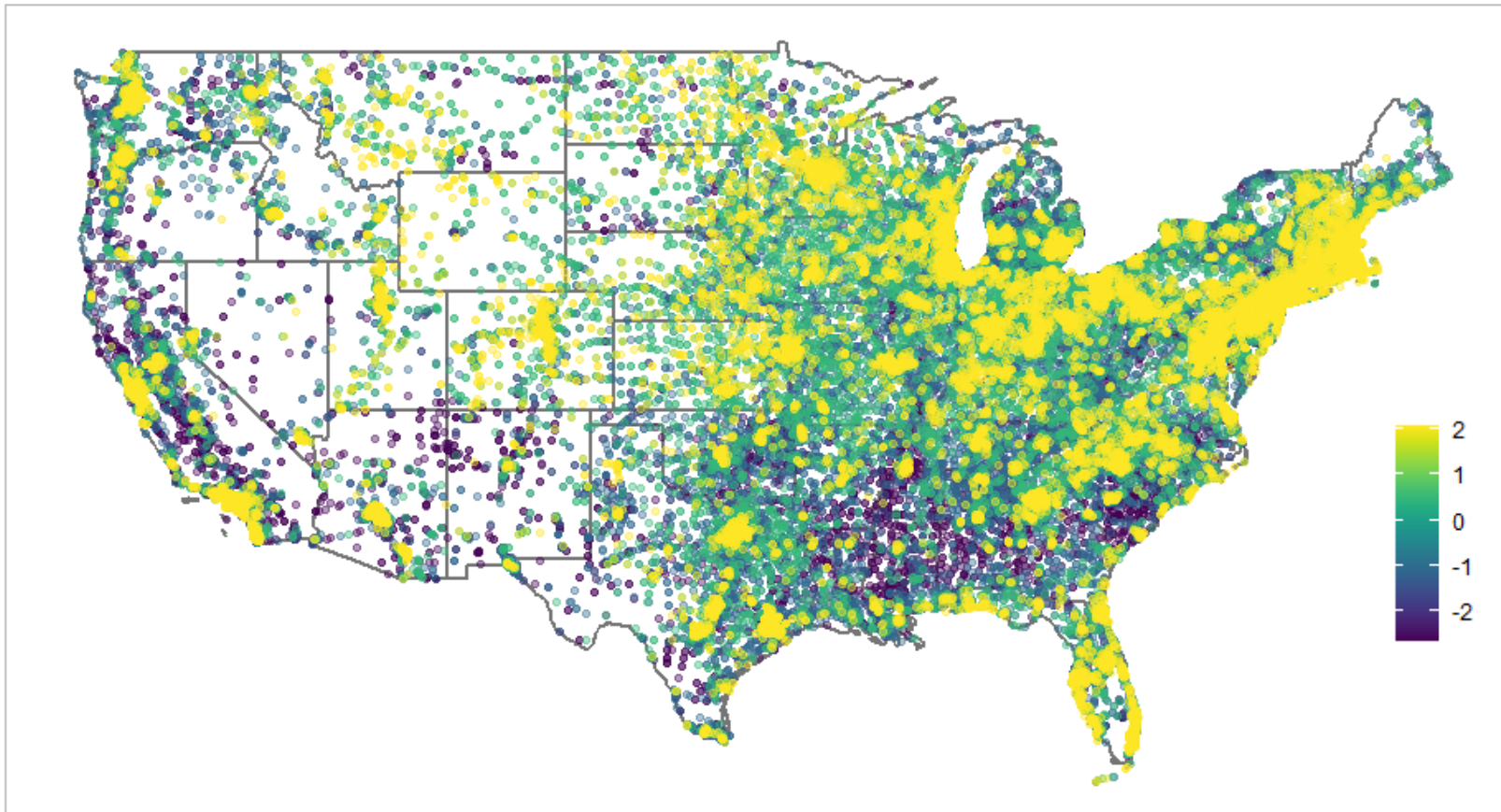
Data source: SEDA

# Outcome variable: binary



**Average achievement of every elementary and middle school in the United States**
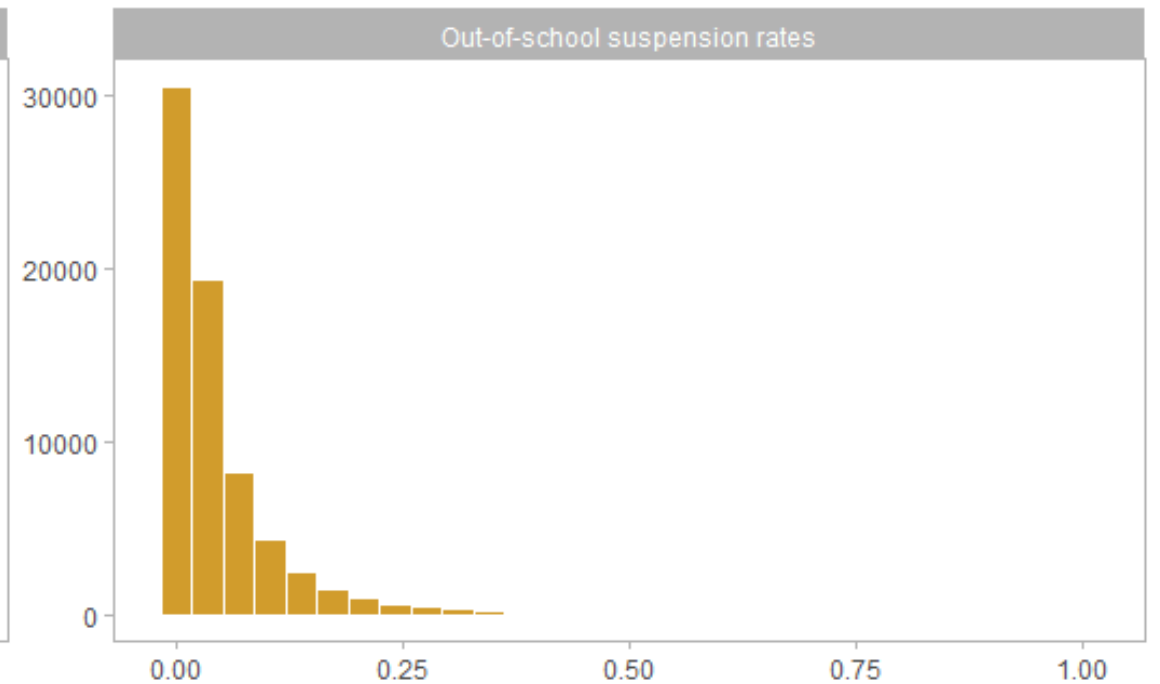
Data source: SEDA
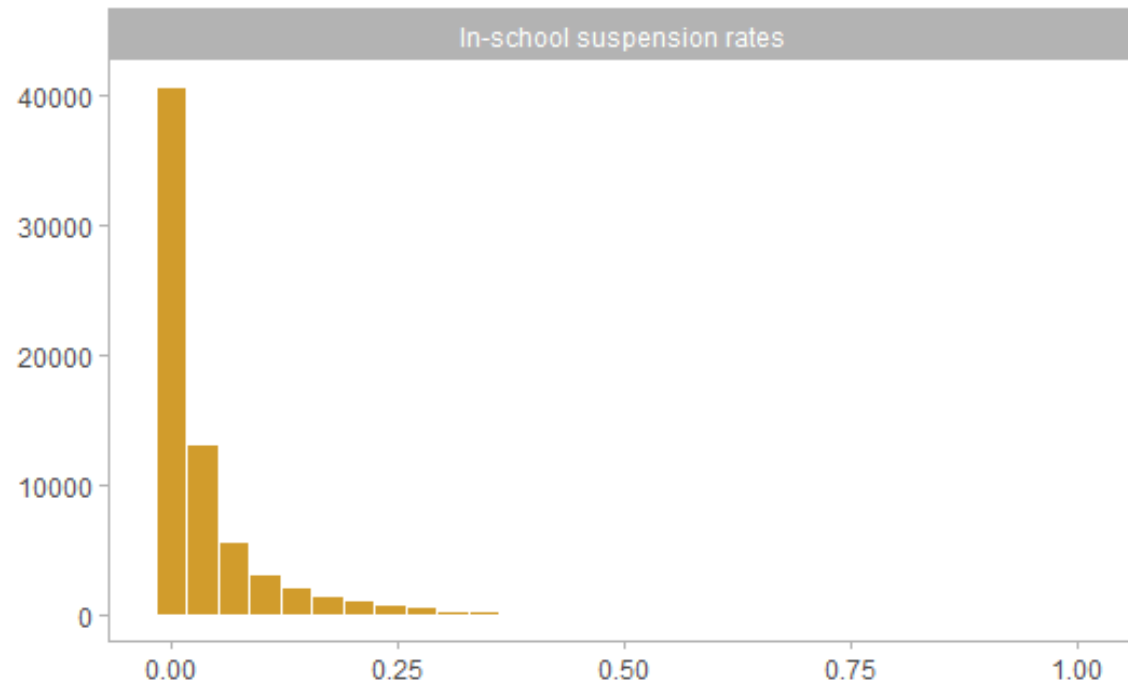
# Outcome variable: quartiles



**Average achievement of every elementary and middle school in the United States**

Data source: SEDA

# Features of opportunity in public-use data

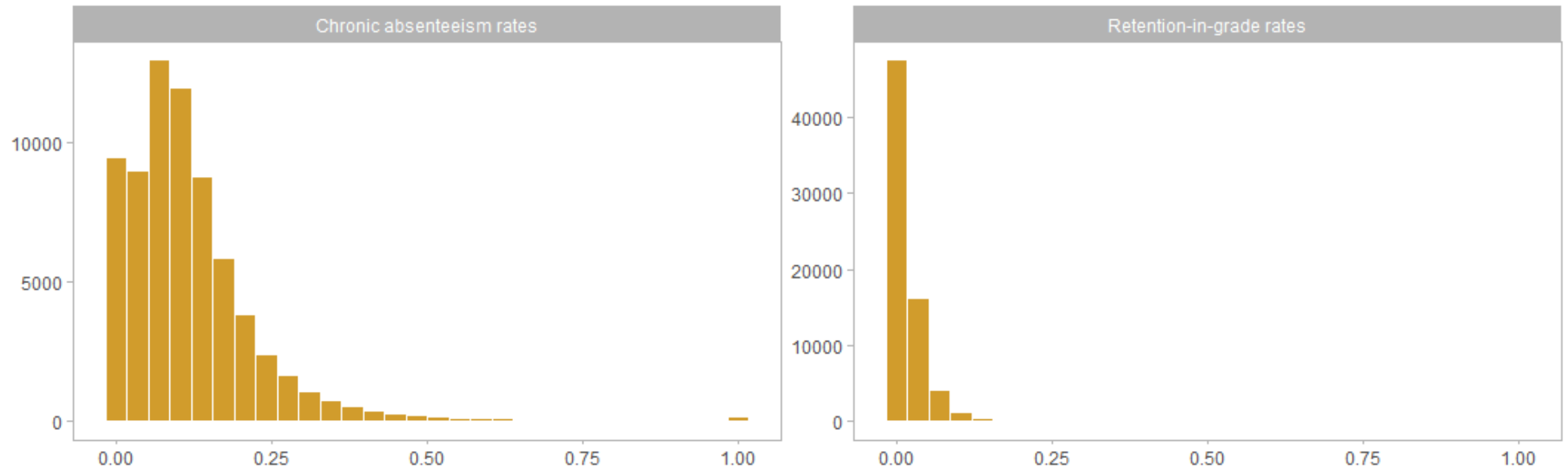**School equity indicators: punitive discipline**
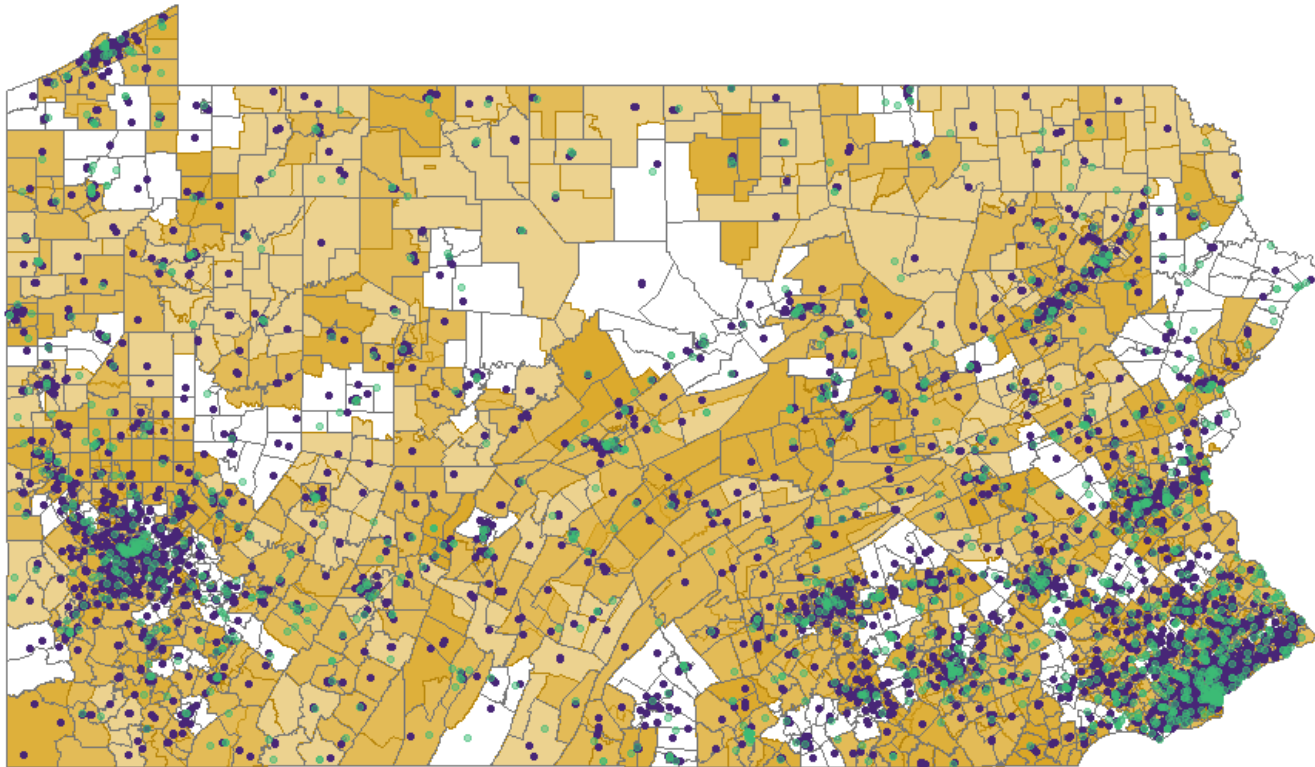
Data source: Civil Rights Data Collection

# Features of opportunity in public-use data

**School equity indicators: absenteeism and grade repetition**

Data source: Civil Rights Data Collection
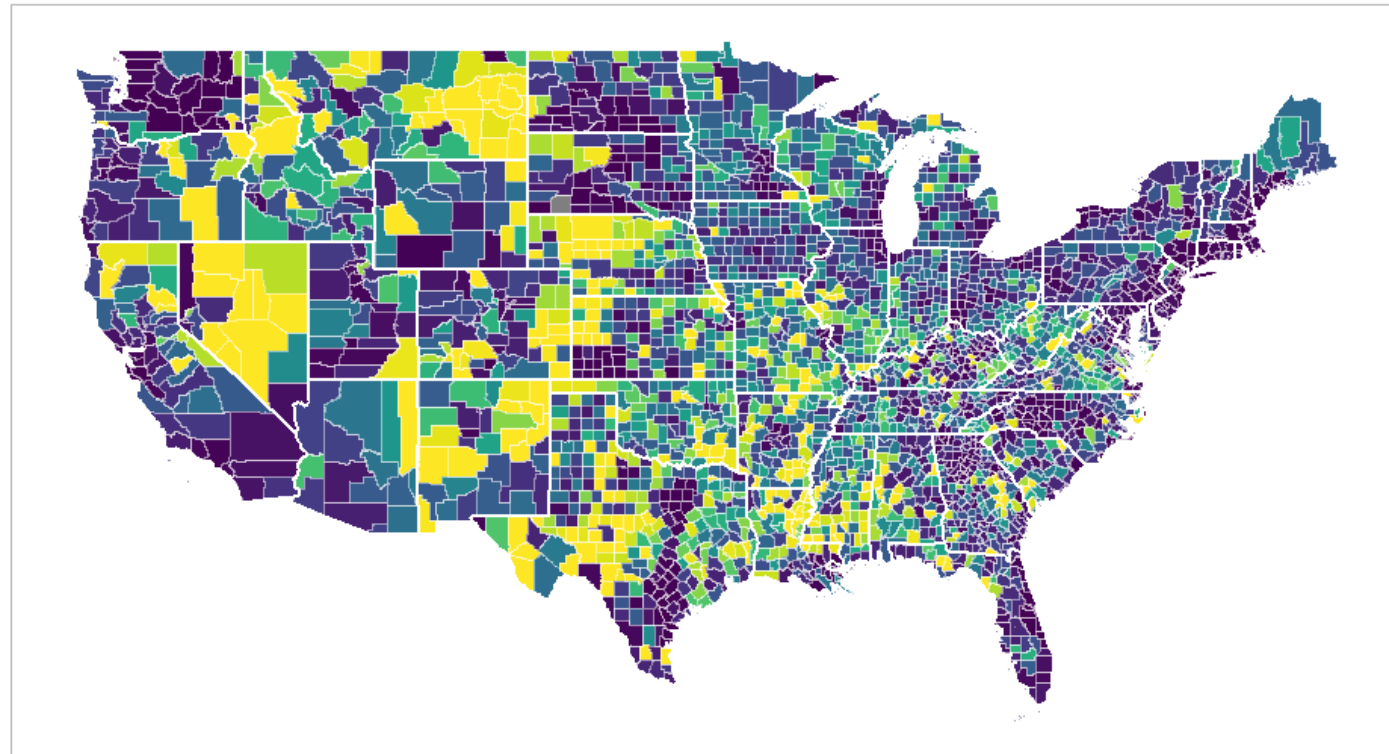
# Features of opportunity in public-use data



**Cultural institutions (green)**
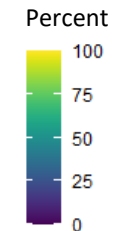
**Schools (purple)**

**School attendance boundaries**

Data source: National Center for Education Statistics; School Attendance Boundary Survey; Institute of Museum and Library Services

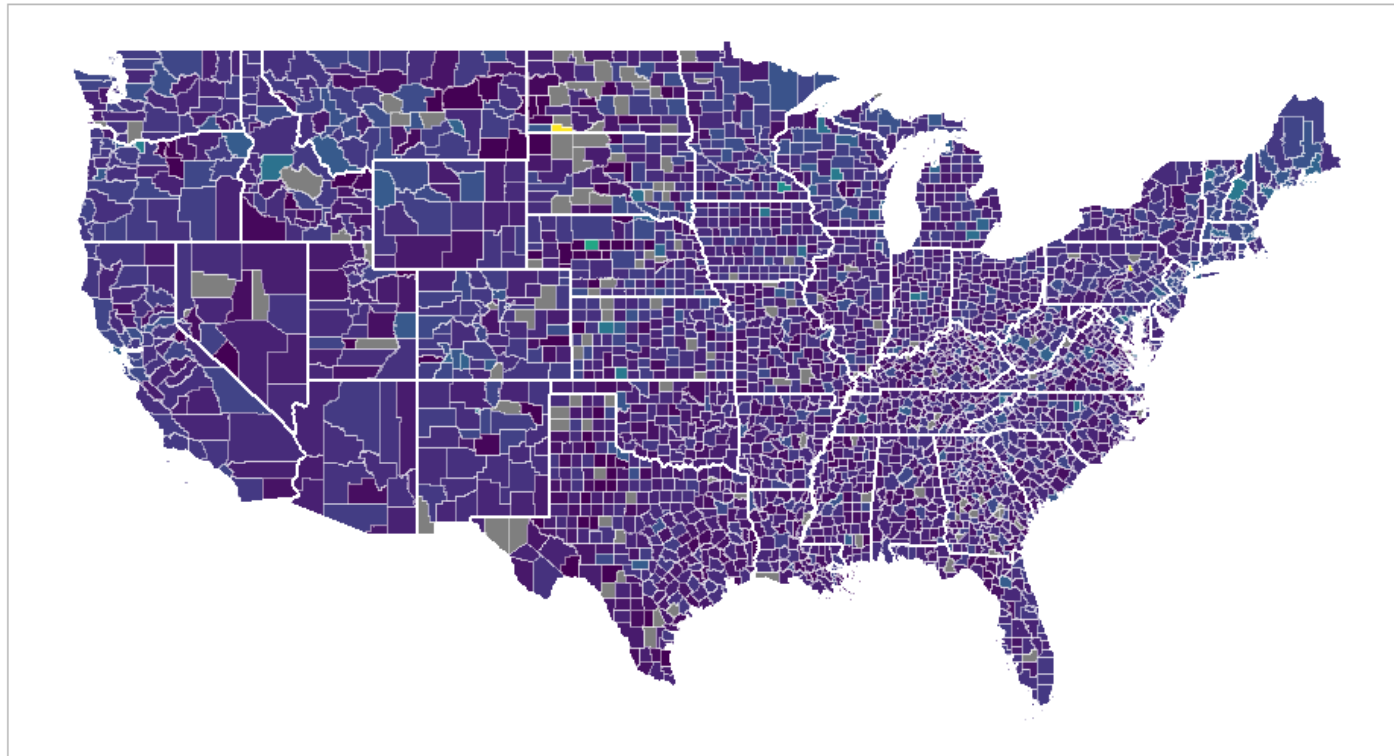# Features of opportunity in public-use data



**Percent of county population without access to Broadband internet**

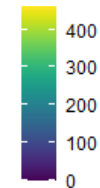Data source: Federal Communications Commission

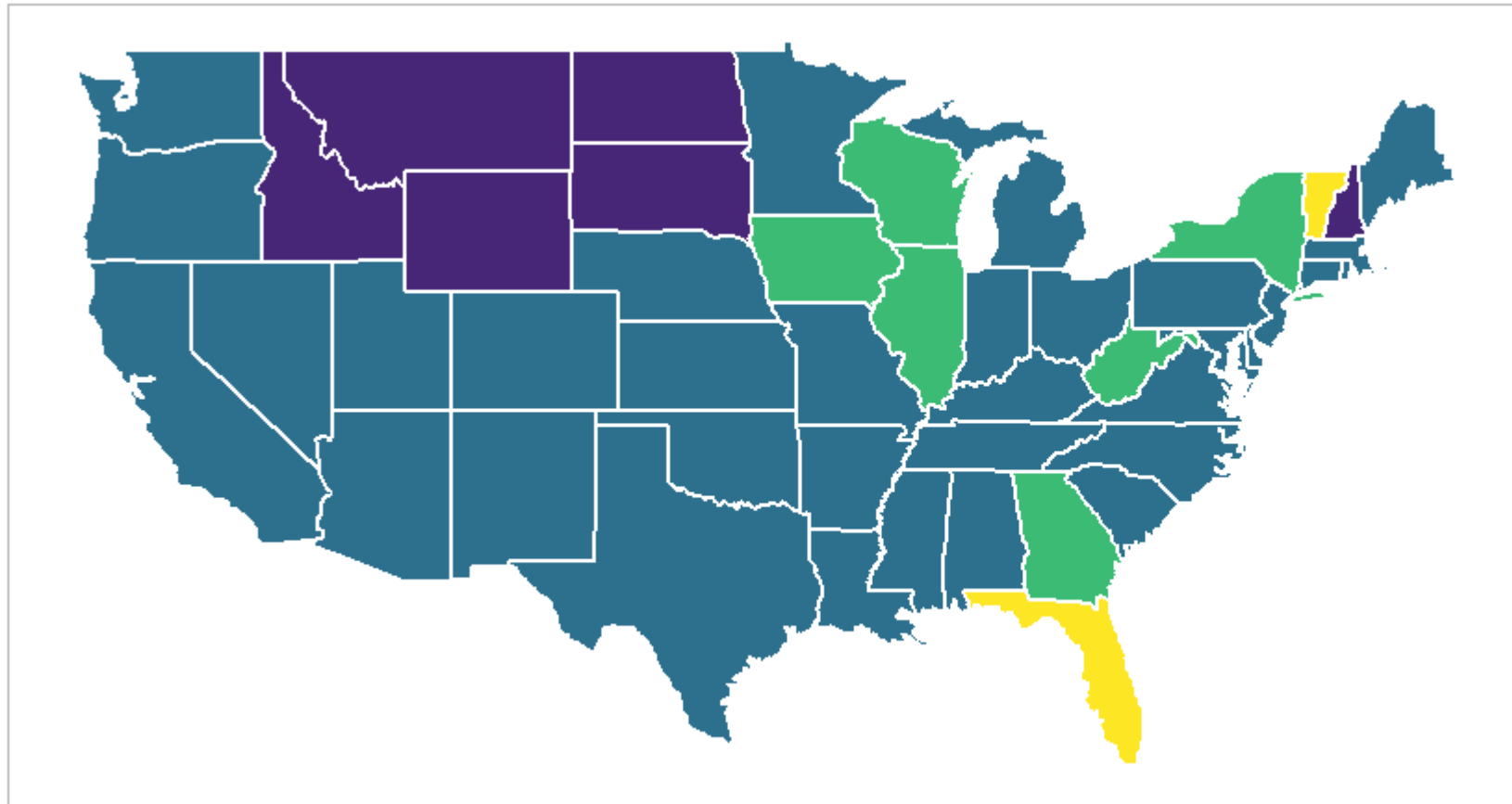# Features of opportunity in public-use data



**Primary Care Providers per 100,000 county residents**
Data source: Federal Communications Commission

# Features of opportunity in public-use data



**State preschool provision**
Data source: Education Commission of the States
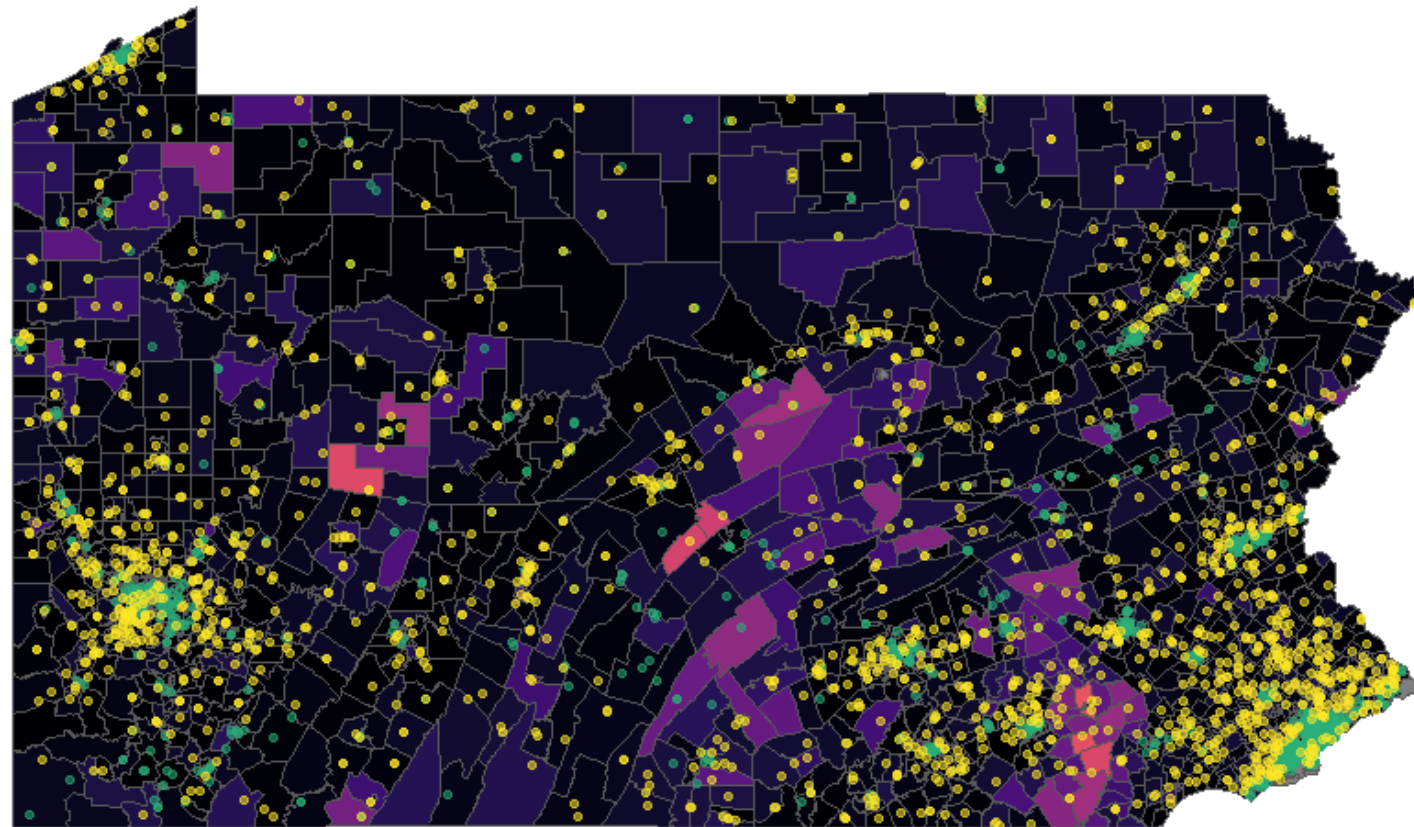
How state funds pre-k
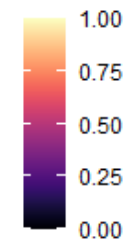- Fully universal Pre-K
- Mostly universal Pre-K
- Non-universal, state-funded Pre-K
- No state-funded Pre-K

# Features of opportunity in public-use data



Proportion of children with no health insurance

1.00
0.75
0.50
0.25
0.00

**School achievement and child health insurance coverage**
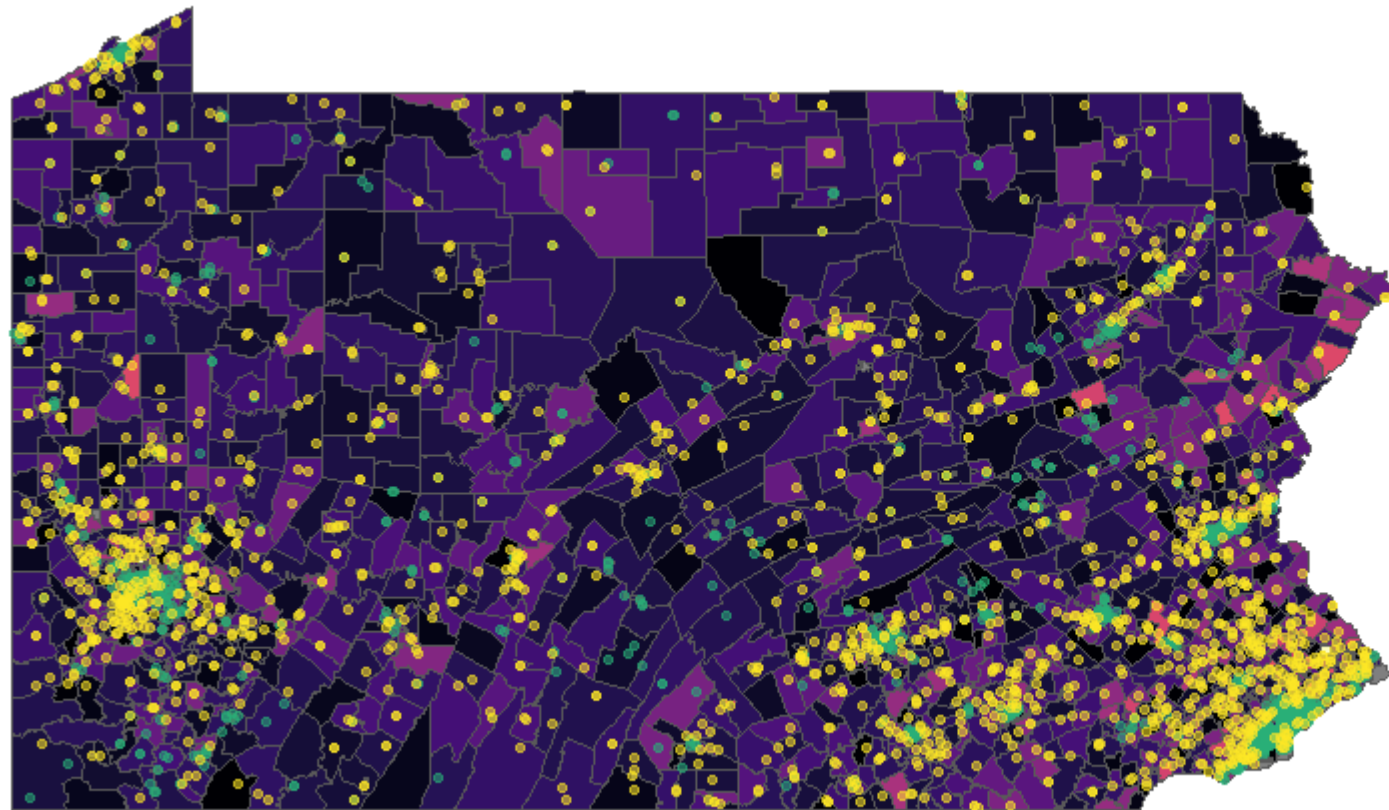Data source: American Community Survey; SEDA

Above average achievement?

- 0
- 1

# Features of opportunity in public-use data



**School achievement and gross rent as a proportion of income**

Data source: American Community Survey; SEDA

Proportion of households paying 50% or more of income toward rent

1.00
0.75
0.50
0.25
0.00

Above average achievement?

● 0
● 1

# Analysis plan

If the average standardized test score in a school measures opportunity, then identifying the factors associated with opportunity gaps could have implications for policymakers who wish to allocate focused resources to high-needs children.

Compare methods for predicting average achievement:

- **LASSO with logistic regression**. A linear function of features is used to model the log odds of the event of interest (above average achievement = 1, below average achievement = 0). Constrain the solution via regularization to find important predictors from among 100+ variables
- **LASSO with ordinal logistic regression.** Here, the event of interest is a multi-category outcome, quartiles of achievement relative to the national average.
- **LASSO with multiple regression.** Here, the standard score units of the outcome (average achievement) are retained.

# References

The Educational Opportunity Project at Stanford University. (2021). Introduction to the Stanford Education Data Archive (SEDA). Workshop.

No Child Left Behind Act, H.R.1, 107th Congress. (2002). Retrieved from https://www.congress.gov/bill/107th-congress/house-bill/1/text

Reardon, S. (2019). Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age. *Russell Sage Foundation, 5*(2), 40-68.

Shear, B., & Reardon, S. (2020). Using pooled heteroskedastic ordered probit models to improve small-sample estimates of latent test score distributions. *Journal of Educational and Behavioral Statistics, 46*(1).

Van Matre, J.C., Shores, K., Kalogrides, D., & Reardon, S. (2016). Stanford Education Data Archive Technical Documentation.
https://cepa.stanford.edu/sites/default/files/SEDA%20Technical%20Documentation%20Version1_1.pdf