Using Big Data To Solve Economic and Social Problems

Professor Raj Chetty

Head Section Leader Rebecca Toseland

Photo Credit: Florida Atlantic University

Missing Applicants to Elite Colleges

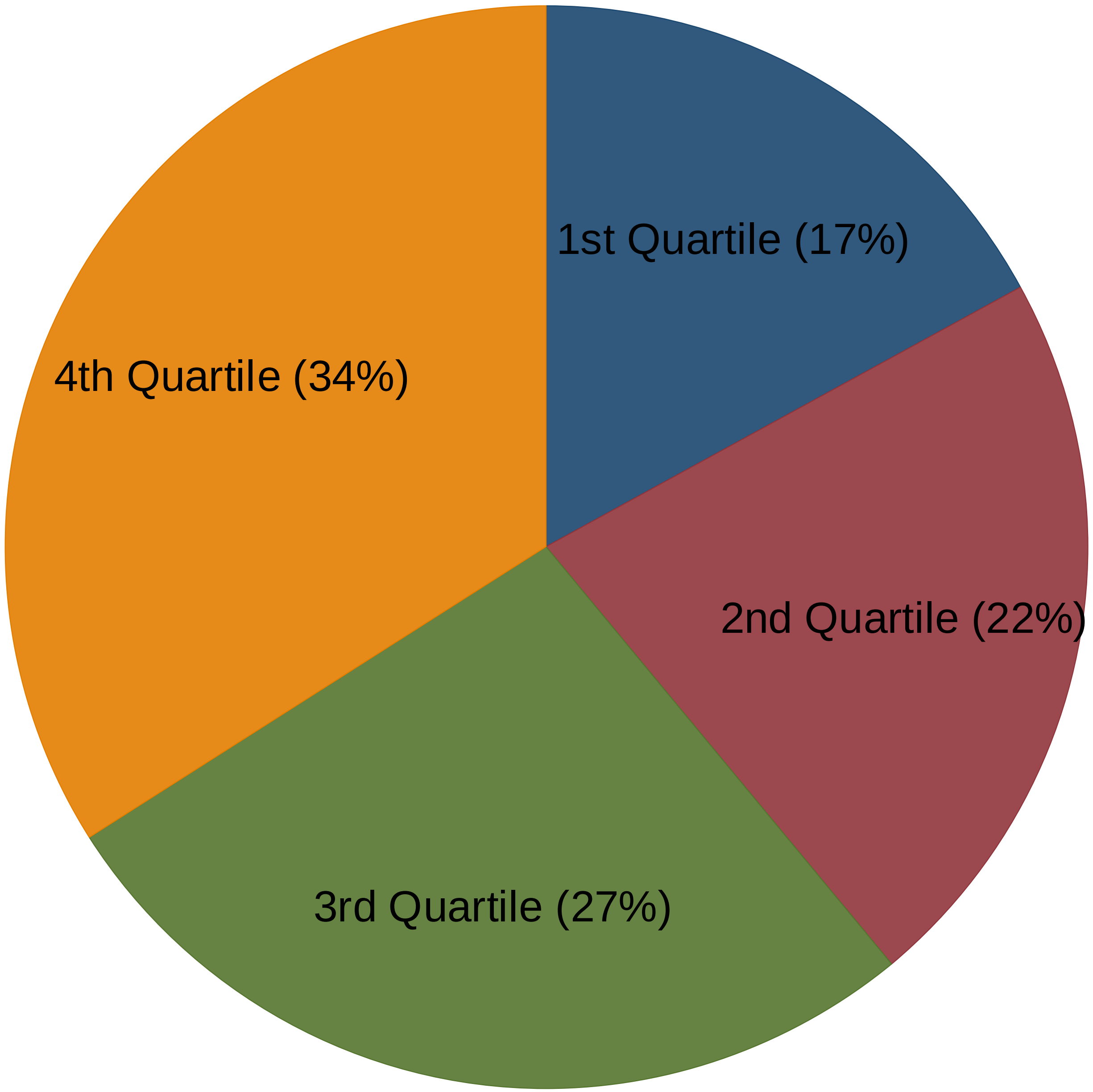
- What can we do to increase the number of low-income students who attend highly selective colleges?

- Hoxby and Avery (2013) show that a key factor is that many low-income, high achieving students do not *apply* to top colleges

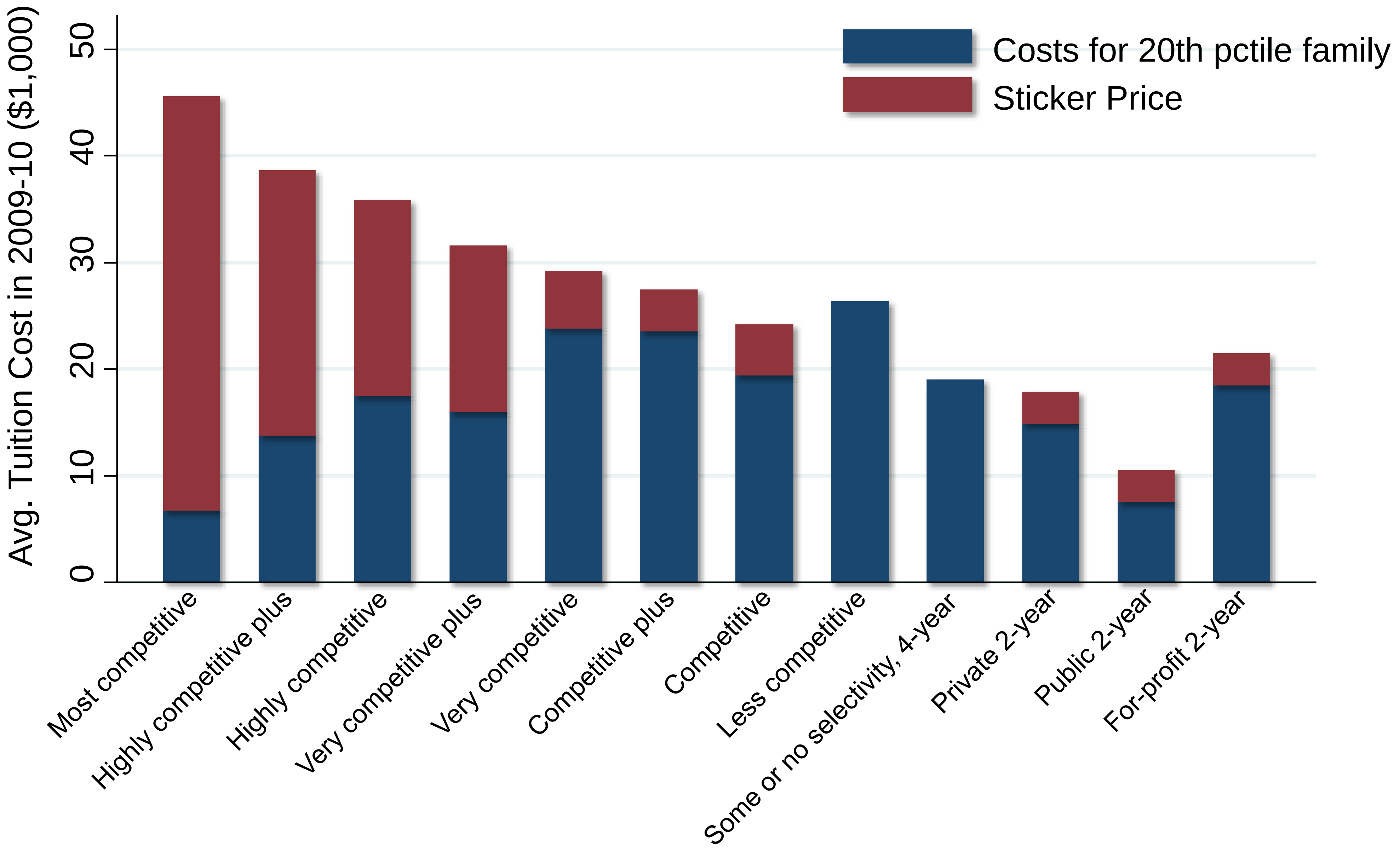
Missing Applicants to Elite Colleges

- Data: College Board and ACT data on test scores and GPAs of all graduating high school seniors in 2008
  - Also know where students sent their SAT/ACT scores, which is a good proxy for where they applied
- Focus on “high-achieving” students: those who score in the top 10% on SAT/ACT and have A- or better GPA

Share of High-Achieving Students by Parent Income Quartile



Costs of Attending Colleges by Selectivity Tier for Low-Income Students



Missing Applicants to Elite Colleges

- Next, examine where low-income (bottom quartile) and high-income (top quartile) students apply

- Focus on difference between college's median SAT/ACT percentile and student's SAT/ACT percentile

  – How good of a match is the college for the student's achievement level, as judged by peers' test scores?

Figure 8. Distribution of High-Achieving, High-Income Students' College Applications, by Student-College Match[a]

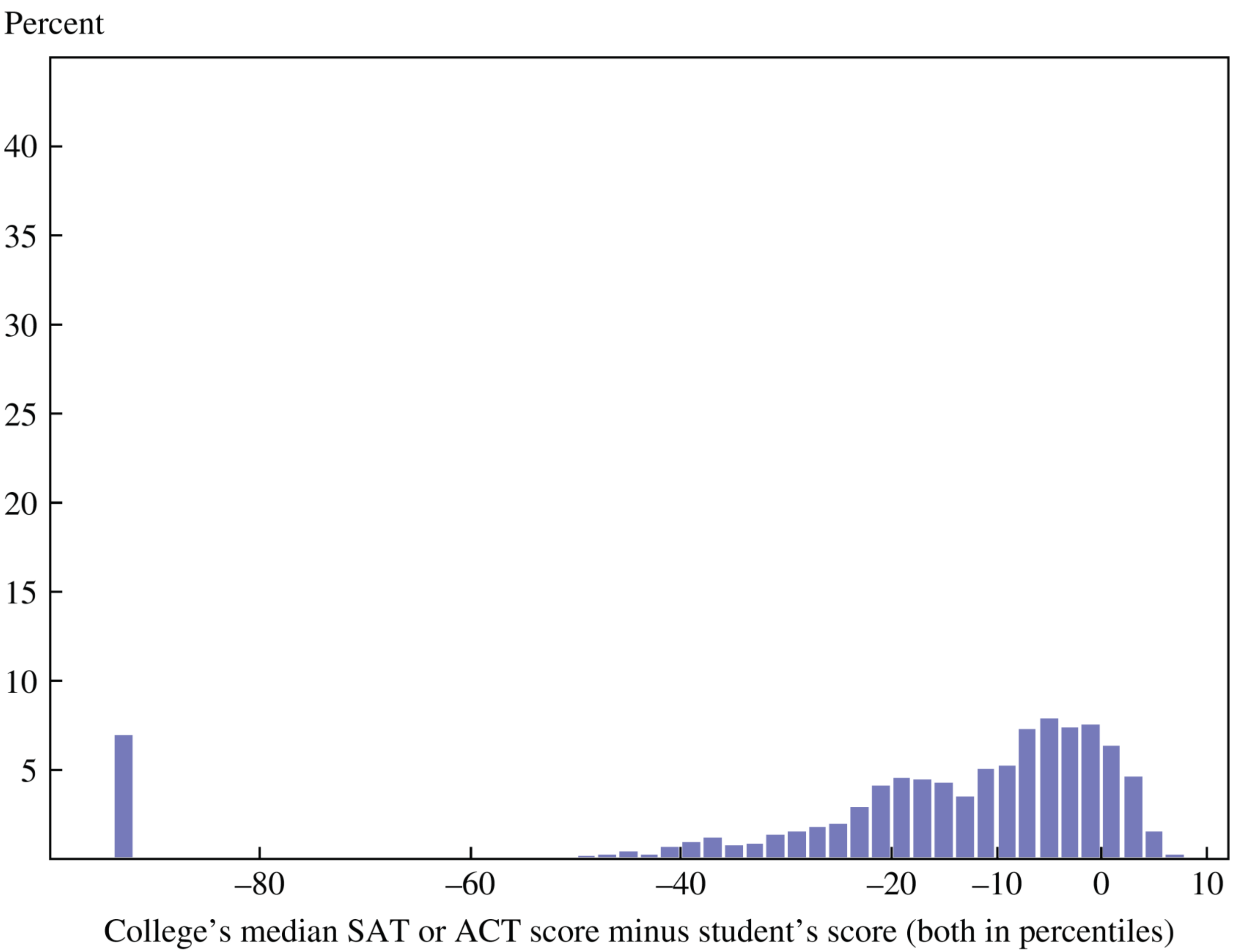
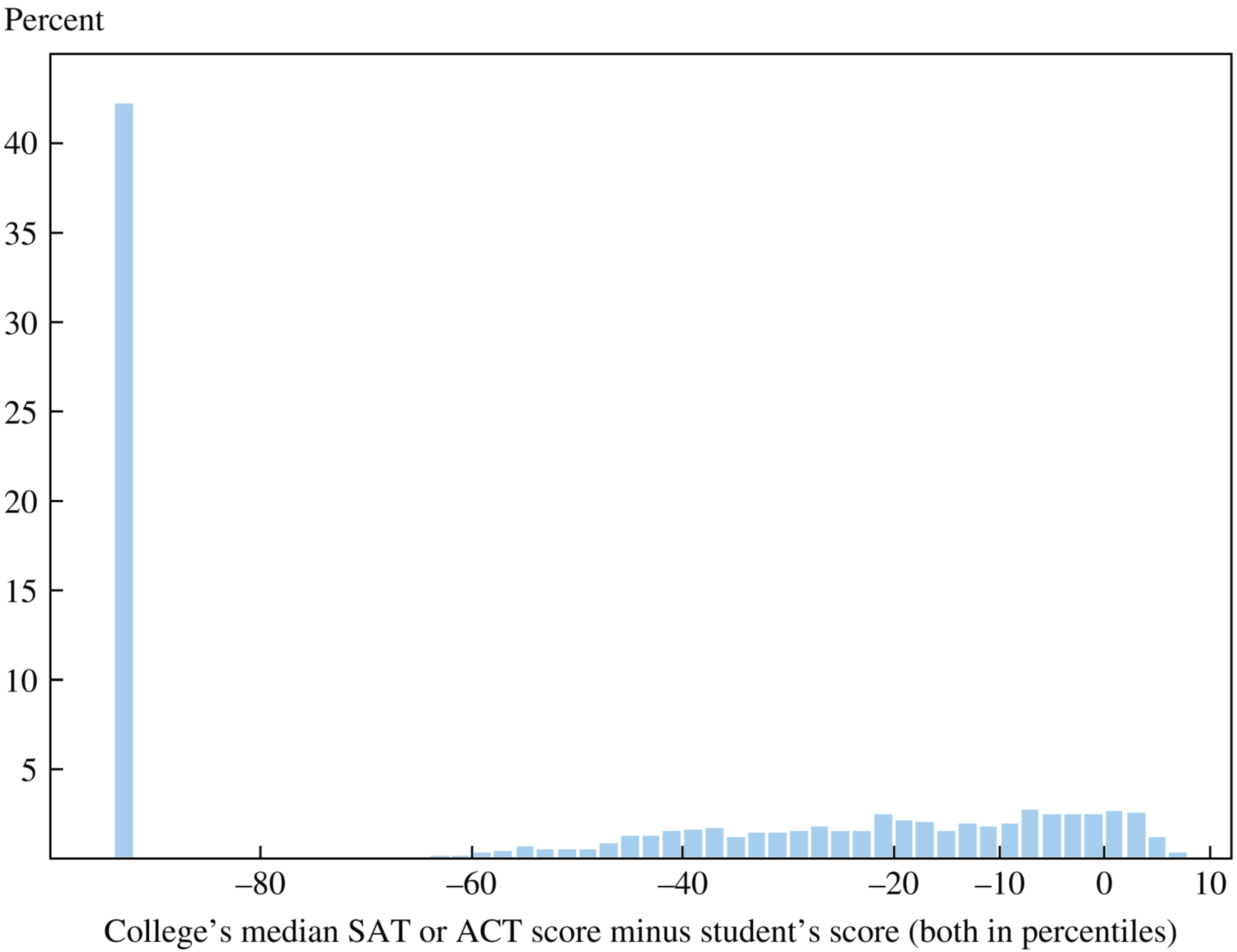
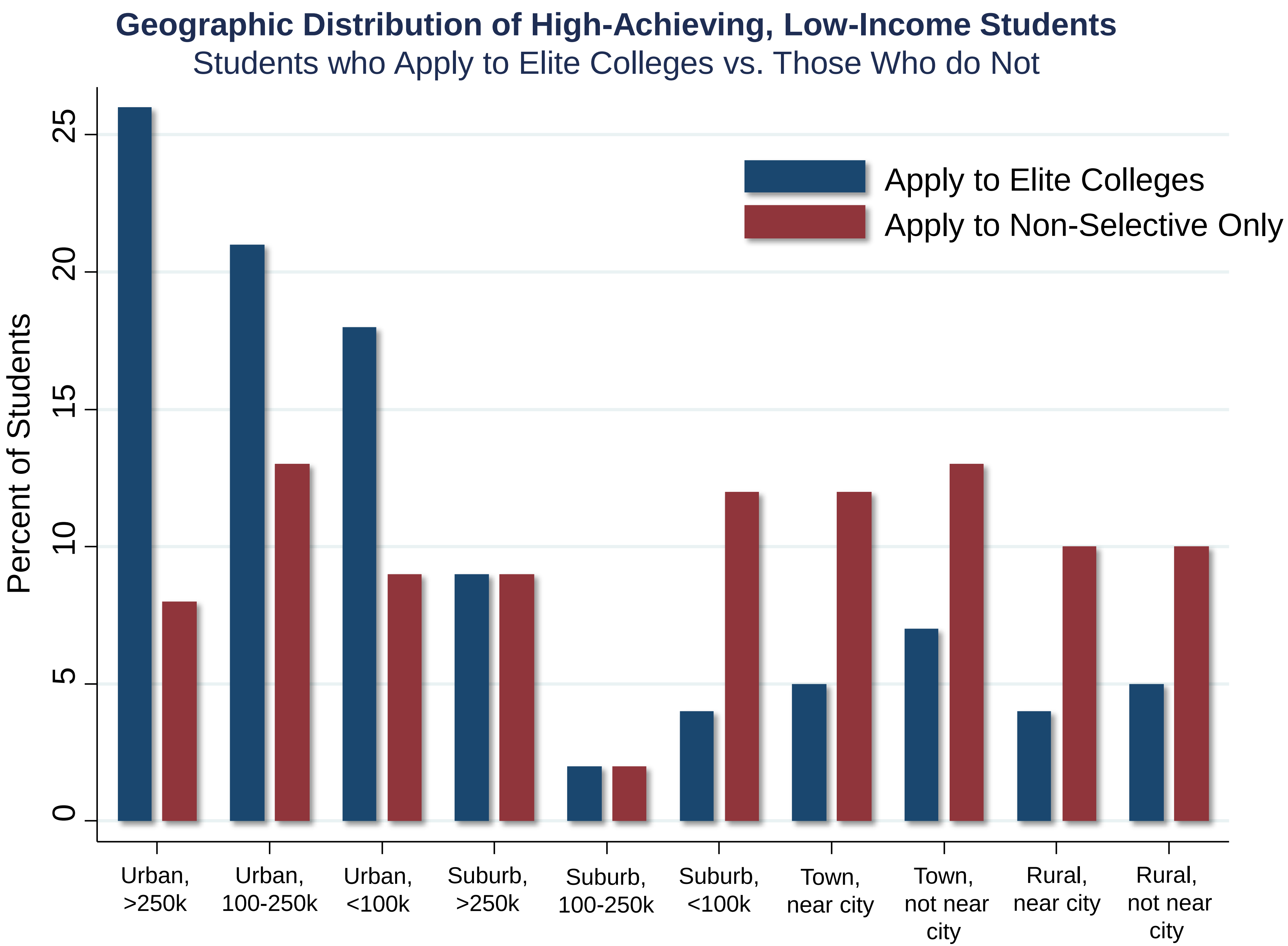


Figure 9. Distribution of High-Achieving, Low-Income Students' College Applications, by Student-College Match[a]



Why Do Many Smart Low-Income Kids Not Apply to Elite Colleges?

- One plausible explanation: lack of information
- Children from high-income families have guidance counselors, relatives, and peers who provide advice
- Lower-income students may not have such resources
- Test this hypothesis by exploring which types of high-achieving low-income students apply to elite colleges
  - Compare 8% of students who apply to elite colleges vs. 50% who apply only to non-selective colleges


Geographic Distribution of High-Achieving, Low-Income Students
Students who Apply to Elite Colleges vs. Those Who do Not
Percent of Students
0
5
10
15
20
25
Apply to Elite Colleges
Apply to Non-Selective Only
Urban, >250k
Urban, 100-250k
Urban, <100k
Suburb, >250k
Suburb, 100-250k
Suburb, <100k
Town, near city
Town, not near city
Rural, near city
Rural, not near city

Why Do Many Smart Low-Income Kids Not Apply to Elite Colleges?

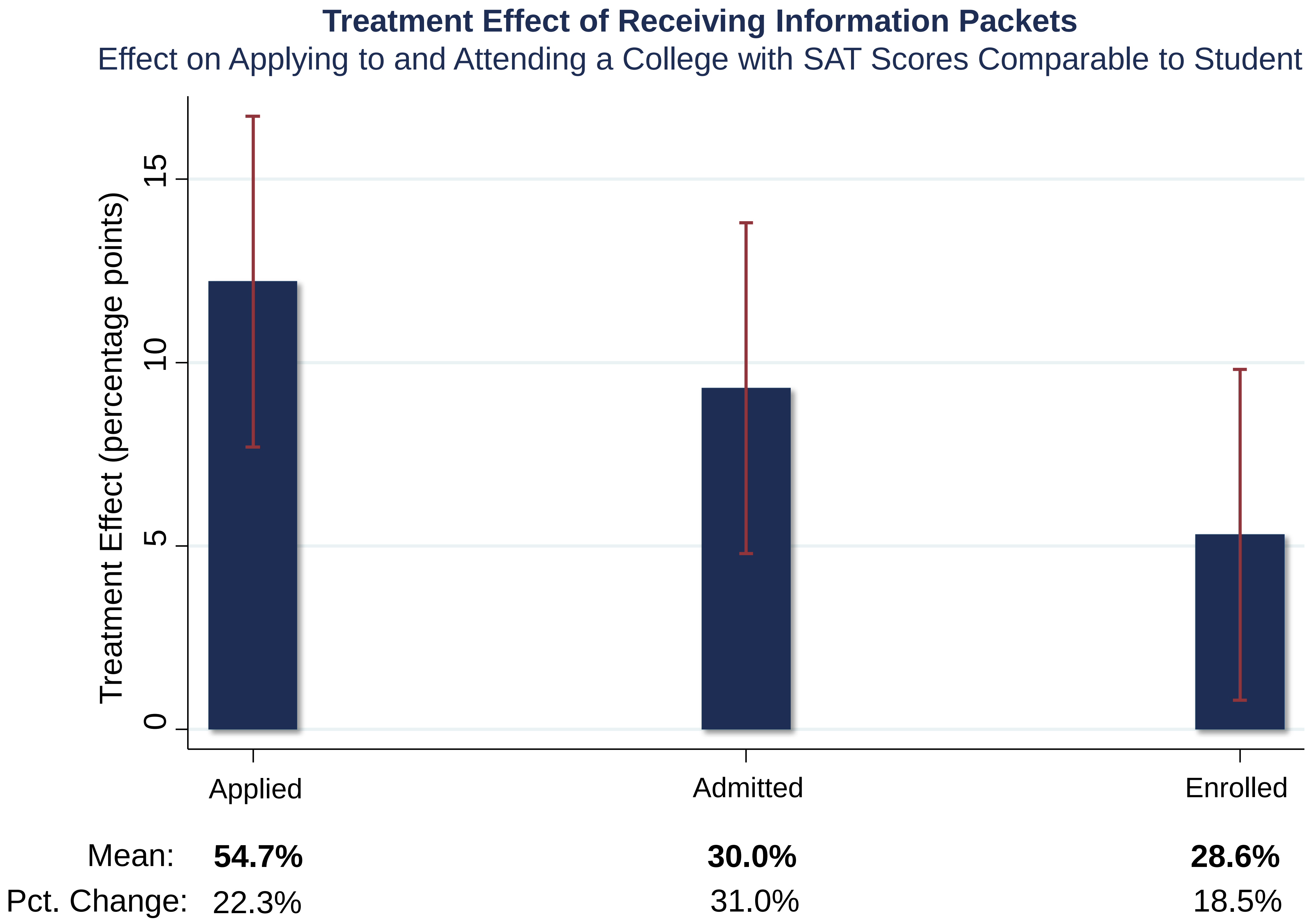
- Further suggestive evidence for information hypothesis: those who apply to elite colleges tend to:
  - Live in Census blocks with more college graduates
  - Attend schools with many other high achievers who apply to elite colleges (e.g., magnet schools)

Informational Mailings to Low-Income High Achievers

- Hoxby and Turner (2013) directly test effects of sending students information on college using a randomized experiment
  - Idea: traditional methods of college outreach (visits by admissions officials) hard to scale in rural areas to reach “missing one-offs”
  - Therefore use mailings that provide customized information:
    - Net costs of local vs. selective colleges
    - Application advice (rec letters, which schools to apply to)
    - Application fee waivers

Informational Mailings to Low-Income High Achievers

- Expanding College Opportunities experimental design:
  - 12,000 from low-income students who graduated high school in 2012 with SAT/ACT scores in top decile
  - Half assigned to treatment group (received mailing)
  - Half assigned to control (no mailing)
  - Cost of each mailing: $6
  - Tracked students application and college enrollment decisions using surveys and National Student Clearinghouse data


Treatment Effect of Receiving Information Packets
Effect on Applying to and Attending a College with SAT Scores Comparable to Student
Treatment Effect (percentage points)
0
5
10
15
Applied
Admitted
Enrolled
Mean:
54.7%
30.0%
28.6%
Pct. Change:
22.3%
31.0%
18.5%

Missing Applicants to Elite Colleges: Lessons

1. Part of the reason there are so few low-income students at elite colleges like Stanford is that smart, low-income kids don't apply

2. This phenomenon is partly driven by a lack of exposure, consistent with other evidence on neighborhood effects

3. Low-cost interventions like informational mailings can close part of the application gap

   – But kids from low-income families remain less likely to attend elite colleges

Directions for Future Work on Higher Education Using Big Data

1. How can we further increase access to elite colleges to provide more pathways to upper-tail outcomes?

   – Identify more highly qualified low-income children who are not currently being admitted and/or not applying using outcome data

   – Can we reach such students using social networks?

2. How can we expand access to colleges that may be “engines of upward mobility”?

   – Estimate value-added of high-mobility-rate colleges using experiments/quasi-experiments and study their recipe for success

K-12 Education

K-12 Education: Background

- U.S. spends nearly $1 trillion per year on K-12 education

- Decentralized system with substantial variation across schools

  – Public schools funded by local property taxes → sharp differences in funding across areas

  – Private schools and growing presence of charter schools

K-12 Education: Overview

- Main question: how can we maximize the effectiveness of this system to produce the best outcomes for students?

  – Traditional approach to study this question: qualitative work in schools

  – More recent approach: analyzing big data to evaluate impacts

- References:

Chetty, Friedman, Hilger, Saez, Schanzenbach, Yagan. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR” *QJE* 2011.

Reardon, Kalogrides, Fahle, Shores. “The Geography of Racial/Ethnic Test Score Gaps.” Stanford CEPA Working Paper 2016

Fredriksson, Ockert, Oosterbeek. “Long-Term Effects of Class Size.” *QJE* 2012

Chetty, Friedman, Rockoff. “Measuring the Impacts of Teachers I and II” *AER 2014*

Using Test Score Data to Study K-12 Education

- Primary source of big data on education: standardized test scores obtained from school districts
  - Quantitative outcome recorded in existing administrative databases for virtually all students
  - Observed much more quickly than long-term outcomes like college attendance and earnings

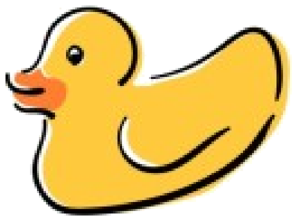
Using Test Score Data to Evaluate Primary Education

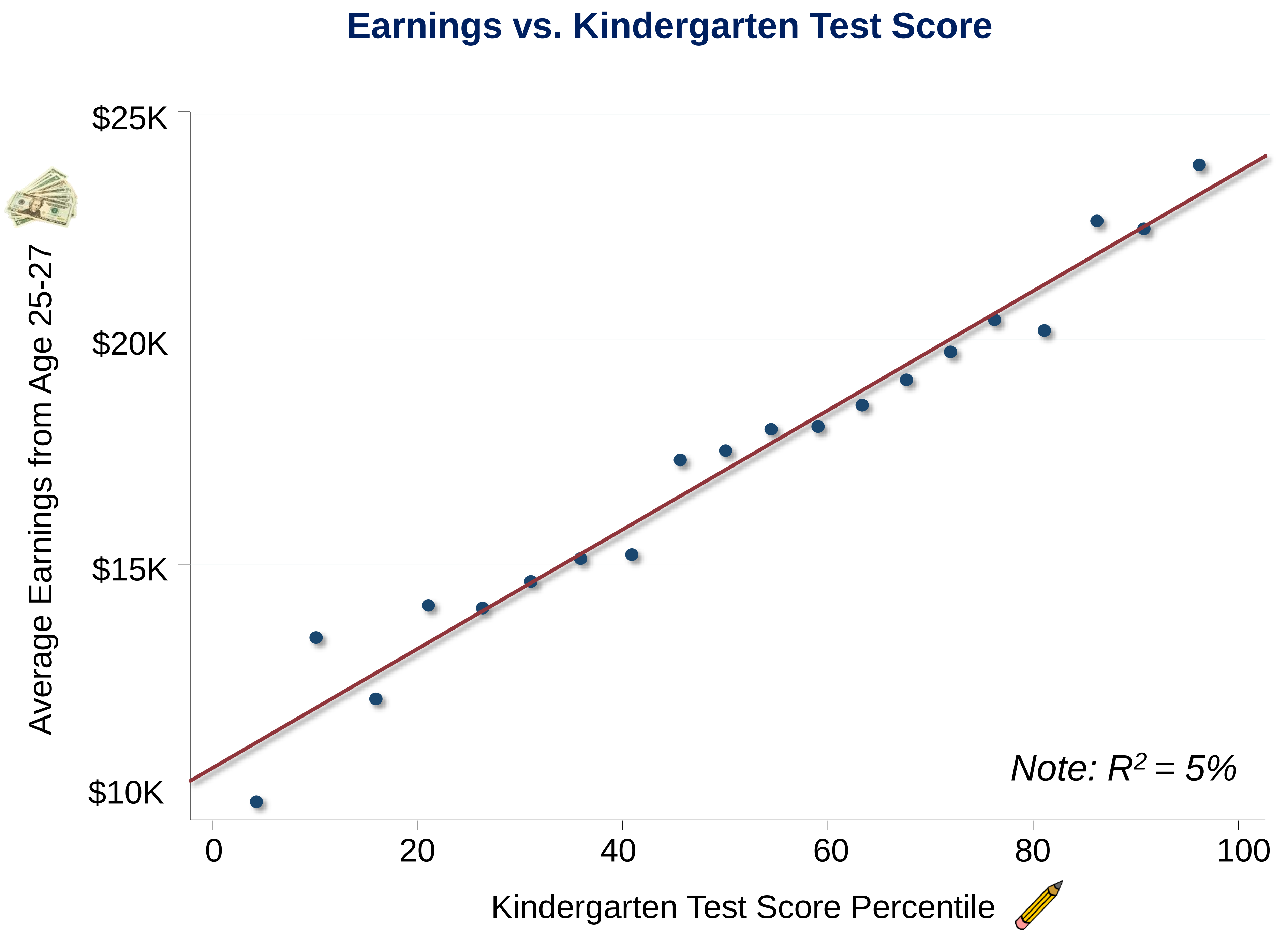
- Common concern: are test scores a good measure of learning?
  - Do improvements in test scores reflect better test-taking ability or acquisition of skills that have value later in life?
- Chetty et al. (2011) examine this issue using data on 12,000 children who were in Kindergarten in Tennessee in 1985
  - Link school district and test score data to tax records
  - Ask whether KG test score performance predicts later outcomes

A Kindergarten Test

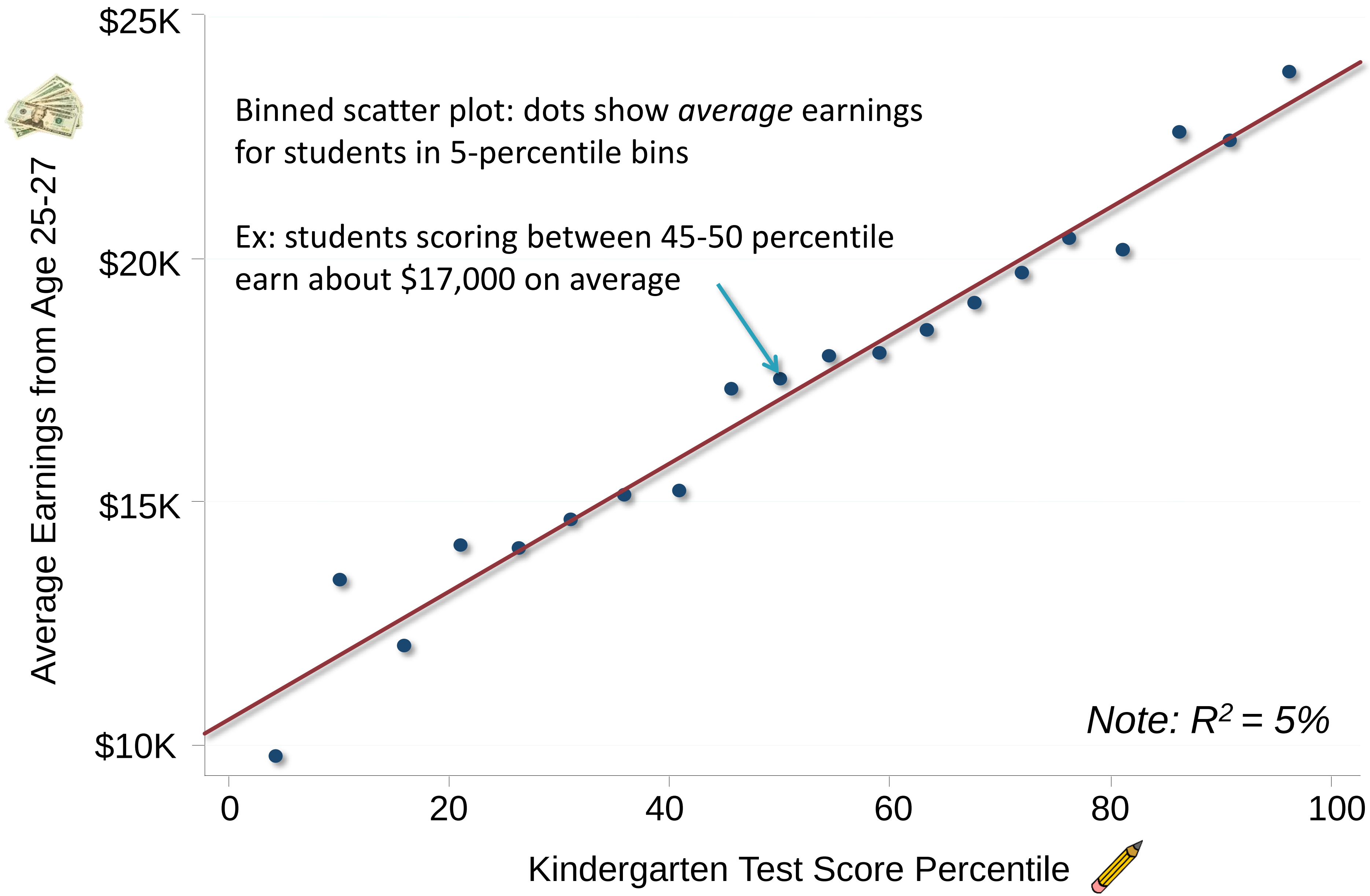
- I'll say a word to you. Listen for the *ending* sound.
- You circle the picture that *starts* with the same sound

"cup"

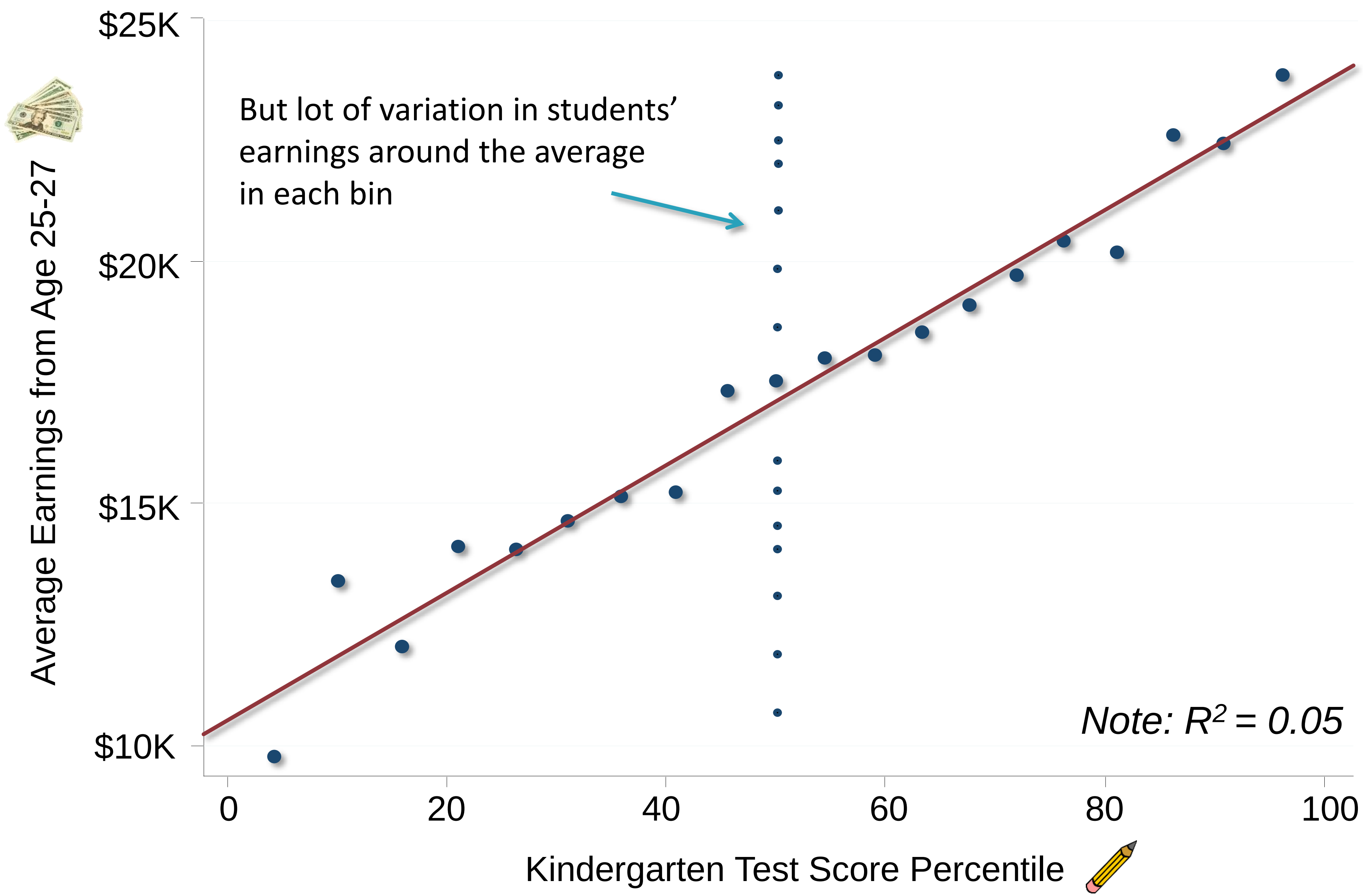



Earnings vs. Kindergarten Test Score
$25K
$20K
$15K
$10K
Average Earnings from Age 25-27
0
20
40
60
80
100
Kindergarten Test Score Percentile
Note: R^2 = 5%

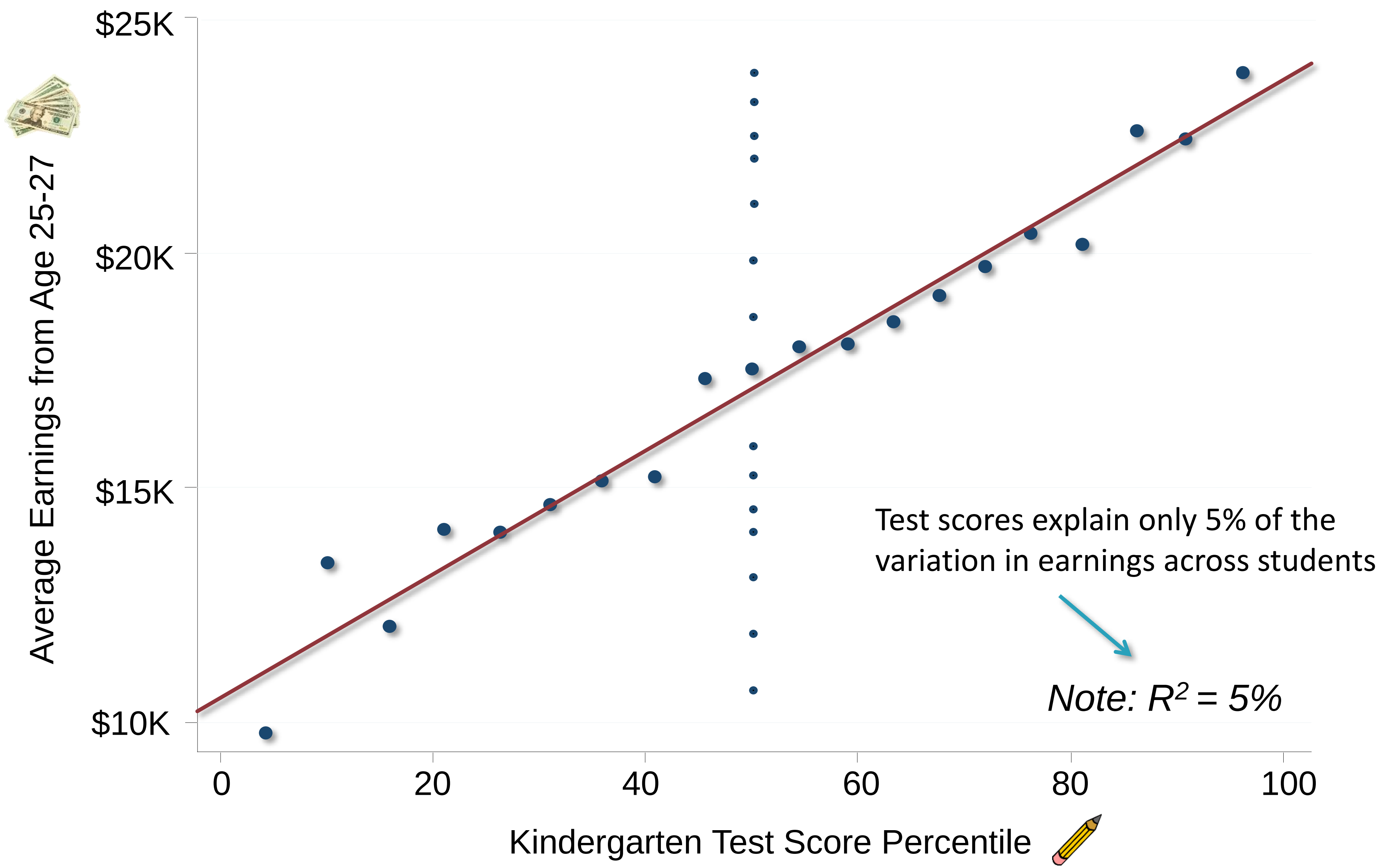
Earnings vs. Kindergarten Test Score



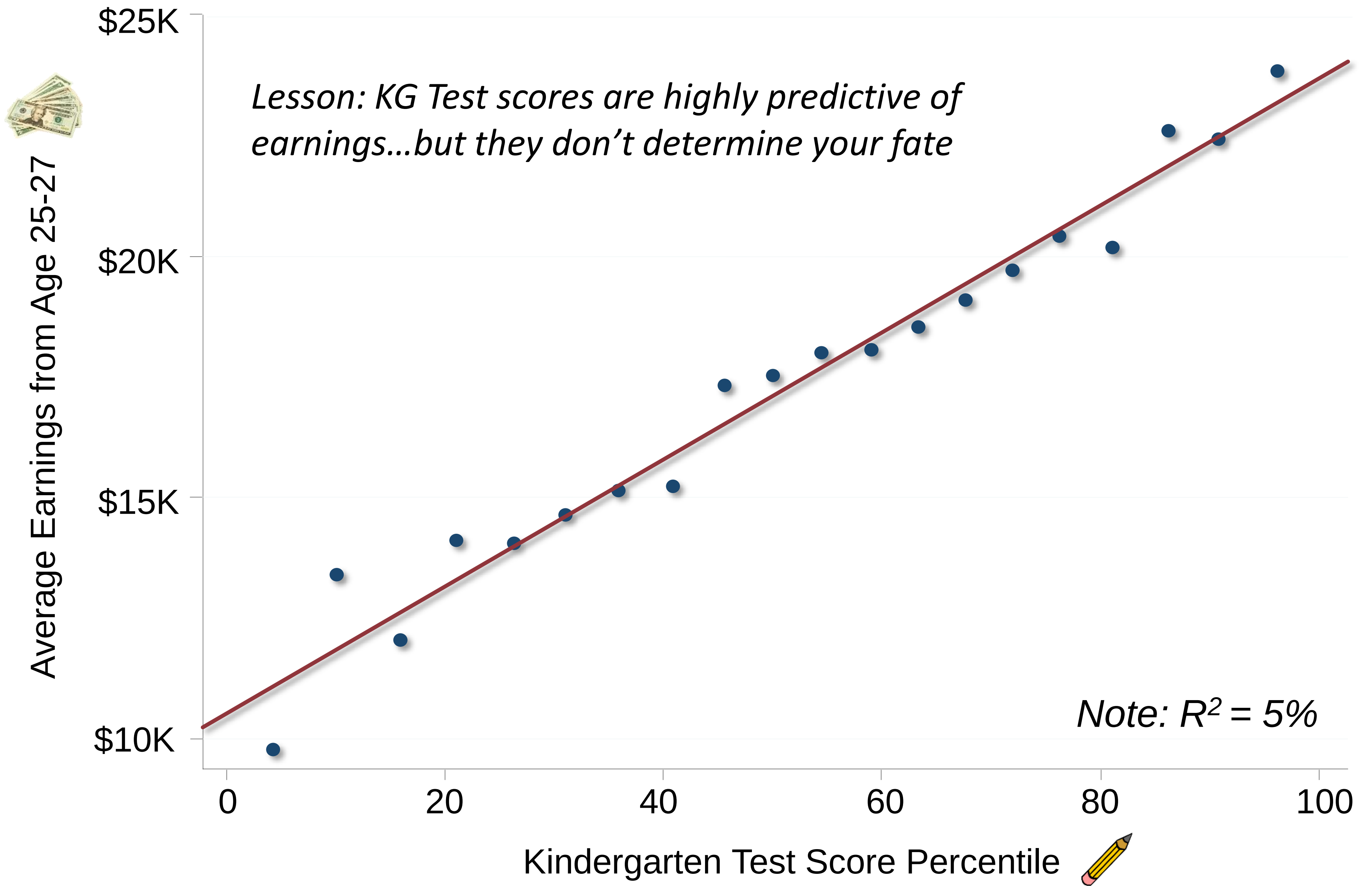
Earnings vs. Kindergarten Test Score



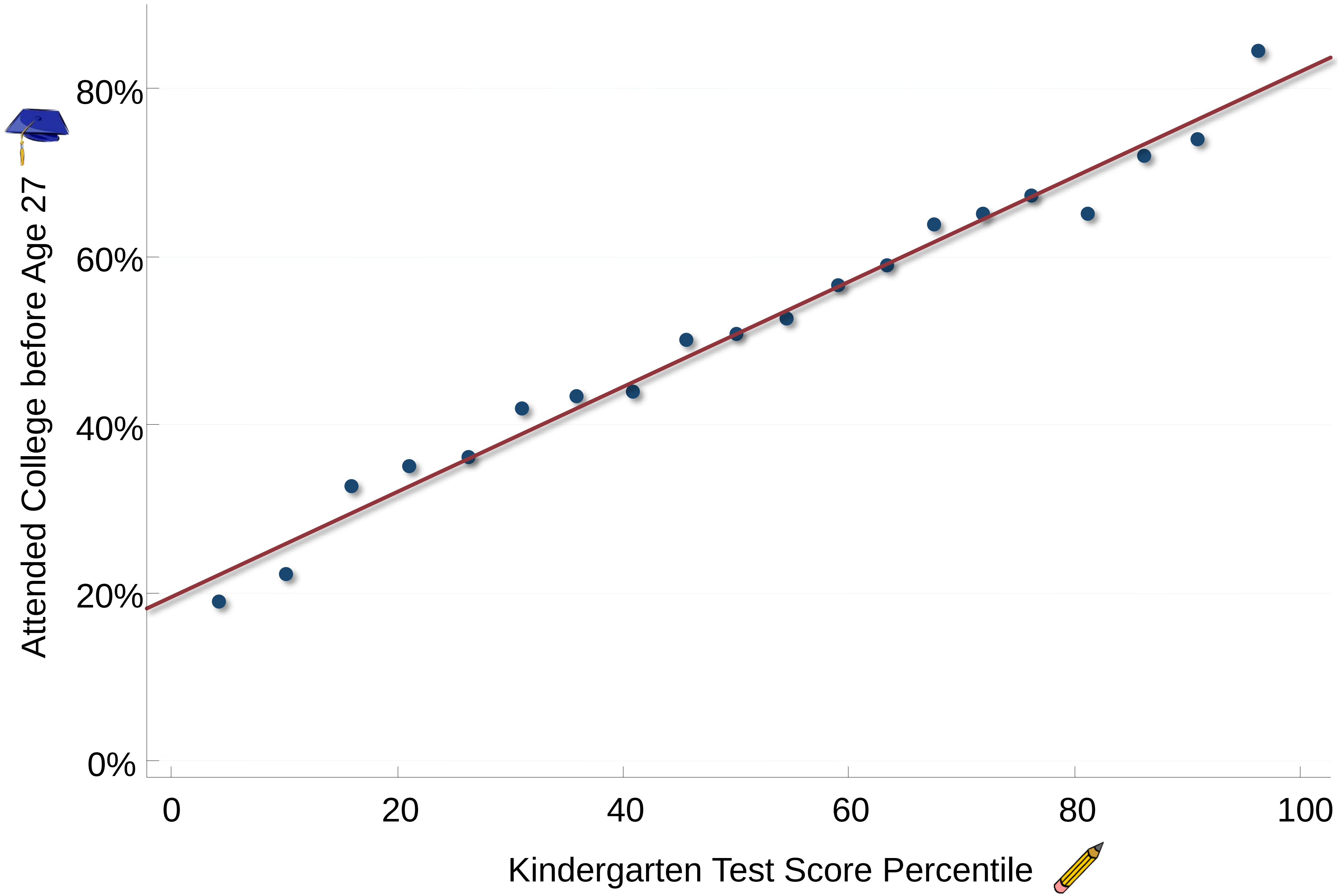
Earnings vs. Kindergarten Test Score

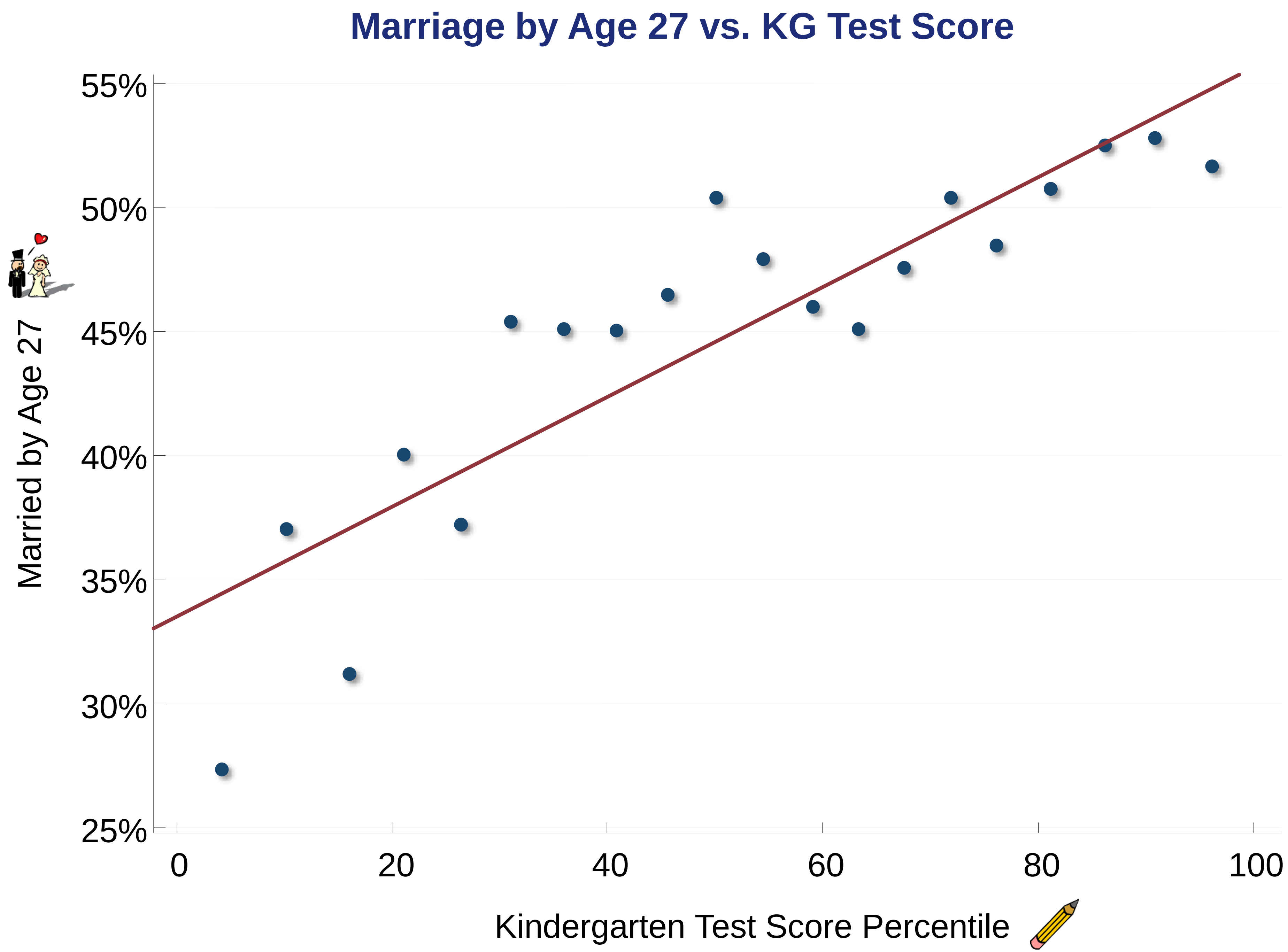


Earnings vs. Kindergarten Test Score



College Attendance Rates vs. KG Test Score




Marriage by Age 27 vs. KG Test Score
55%
50%
45%
40%
35%
30%
25%
Married by Age 27
0
20
40
60
80
100
Kindergarten Test Score Percentile

Studying Differences in Test Score Outcomes

- Test scores can provide a powerful data source to compare performance across schools and subgroups (e.g., poor vs. rich)

- Problem: tests are not the same across school districts and grades → makes comparisons very difficult

- Reardon et al. (2016) solve this problem and create a standardized measure of test score performance for all schools in America
  - Use 215 million test scores for students from 11,000 school districts across the U.S. from 2009-13 in grades 3-8

Making Test Score Scales Comparable Across the U.S.

- Convert test scores to a single national scale in three steps:

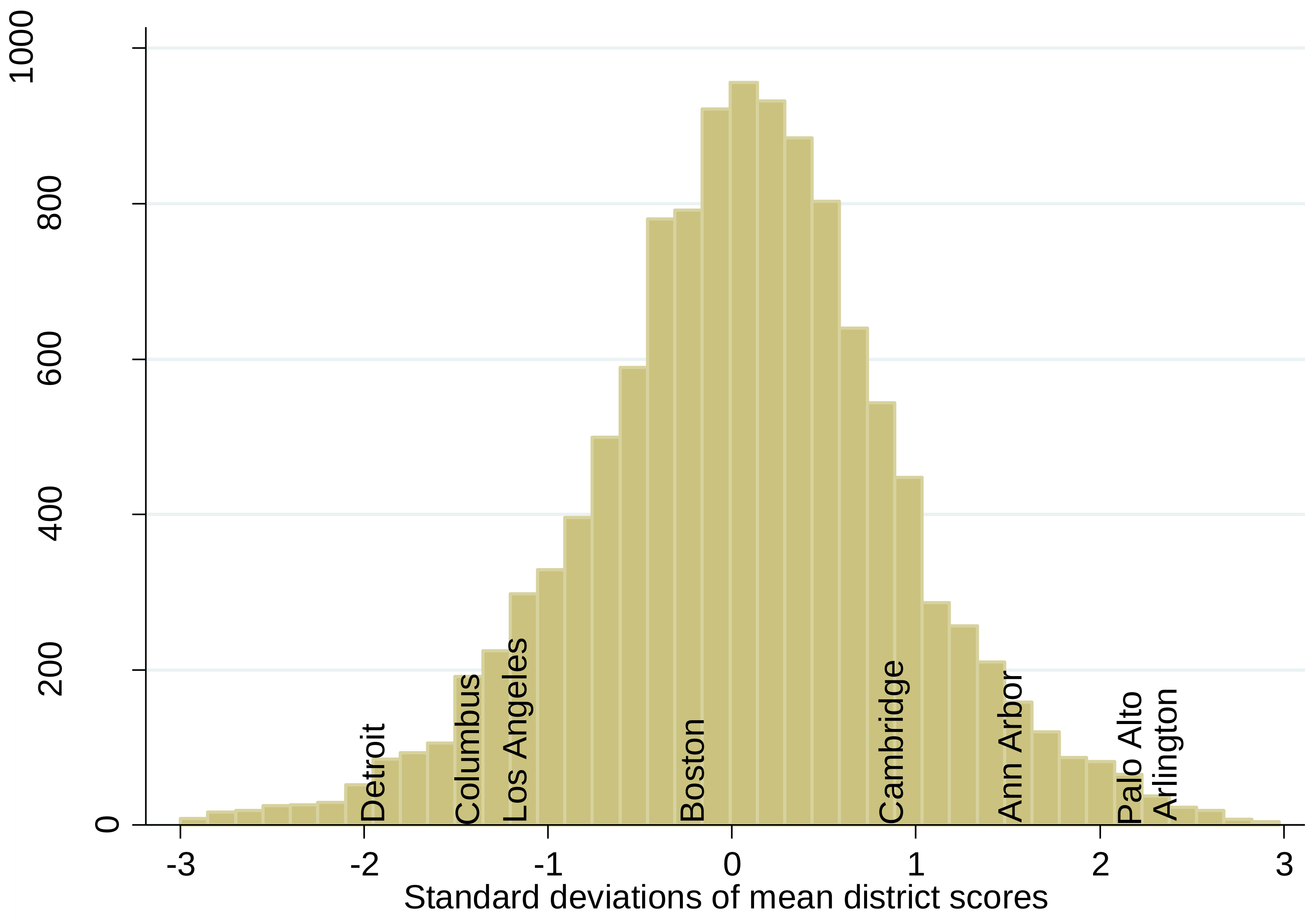
1. Rank each school district's average scores in the statewide distribution (for a given grade-year-subject)

2. Use data from a national test administered to a sample of students by Dept. of Education to convert state-specific rankings to national scale

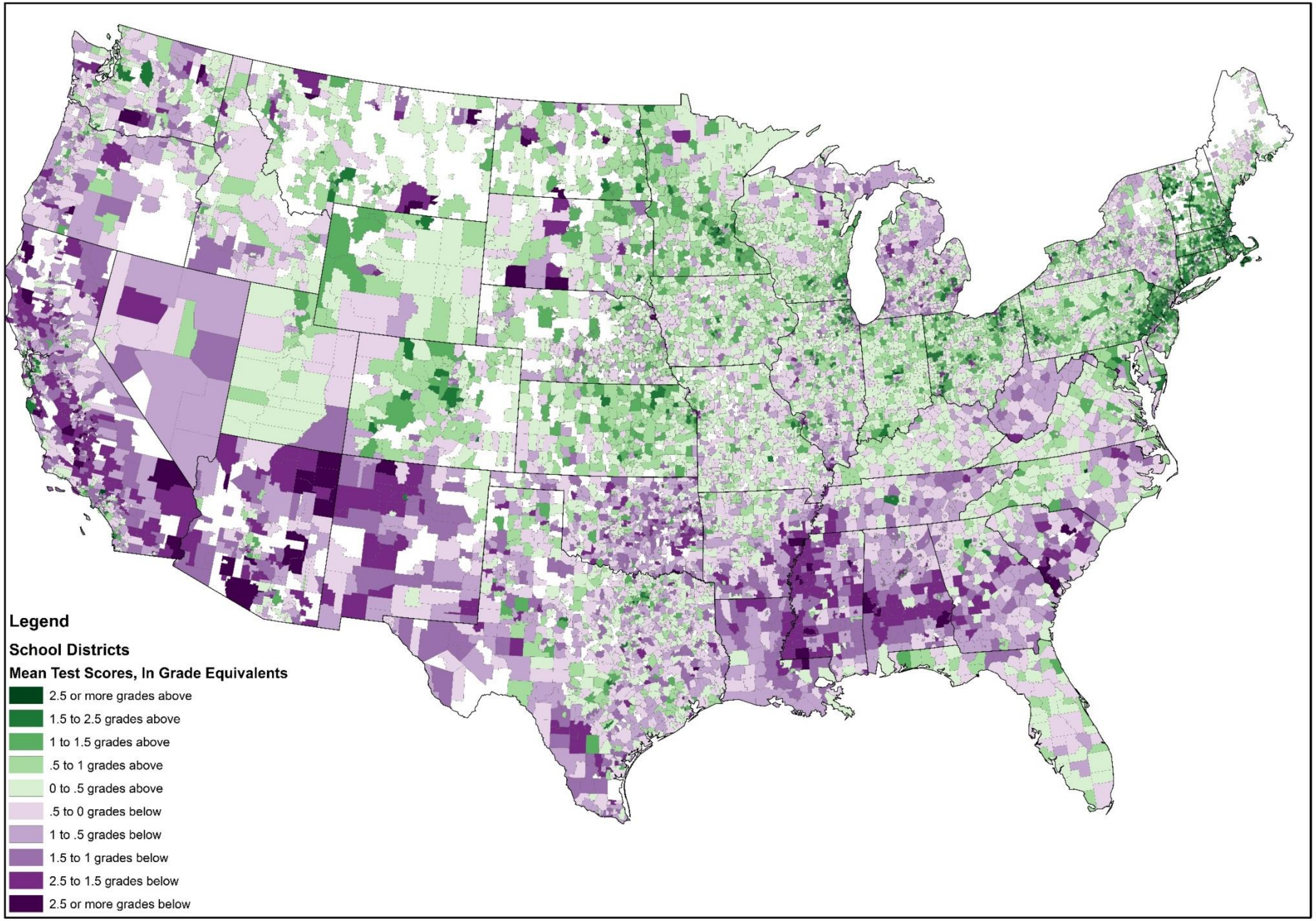
    - Ex: suppose CA students score 5 percentiles below national average
    - Then a CA school whose mean score is 10 percentiles below CA mean is 15 percentiles below national mean

3. Convert mean test scores to "grade level" equivalents

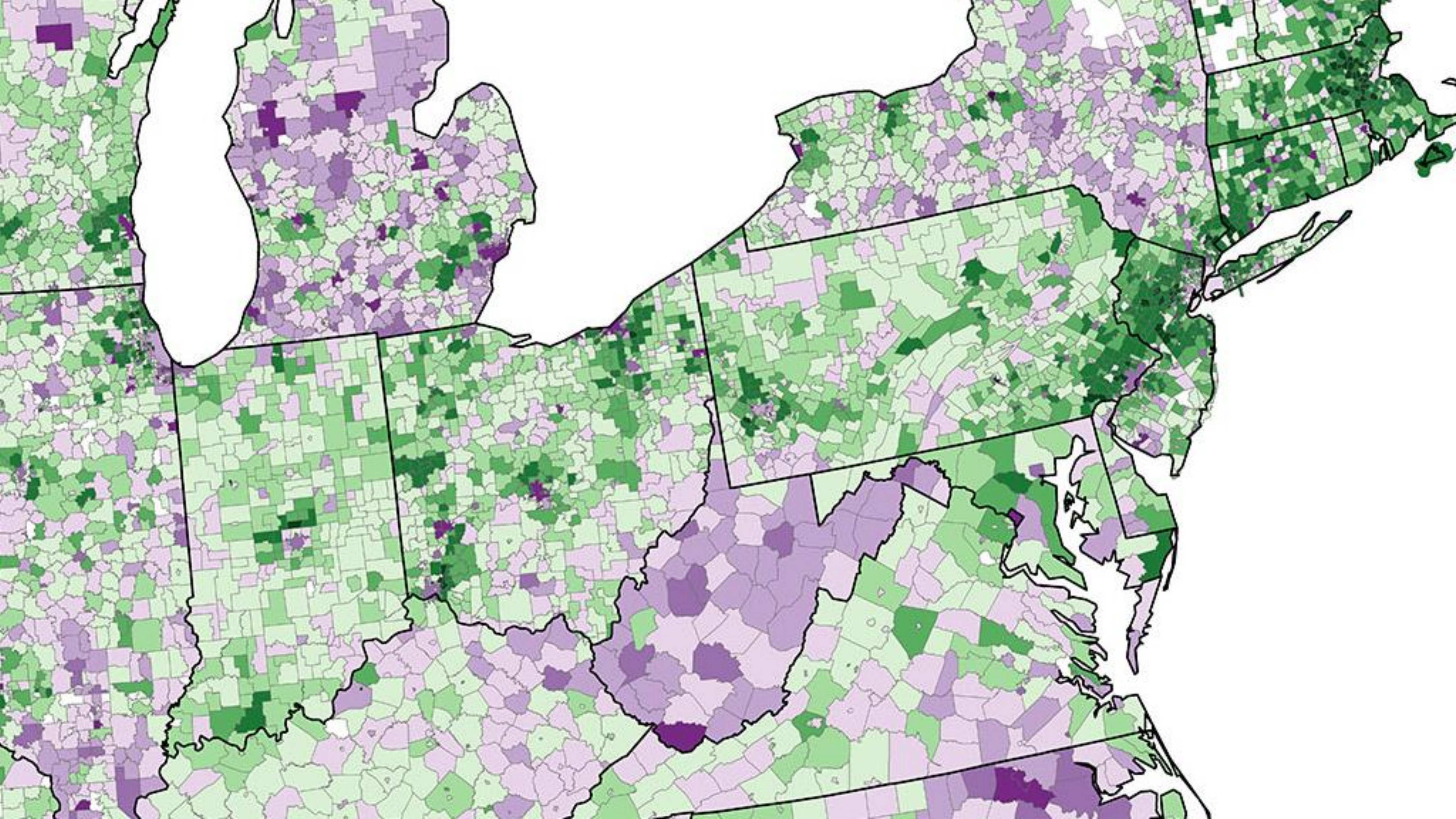
Nationwide District Achievement Variation, 2009-2013



Average Test Scores, by School District, Grades 3-8, 2009-2013



© (2016) sean f. reardon, Demetra Kalogrides, Erin Fahle, Kenneth Shores, and Benjamin Shear. Stanford Education Data Archive: seda.stanford.edu



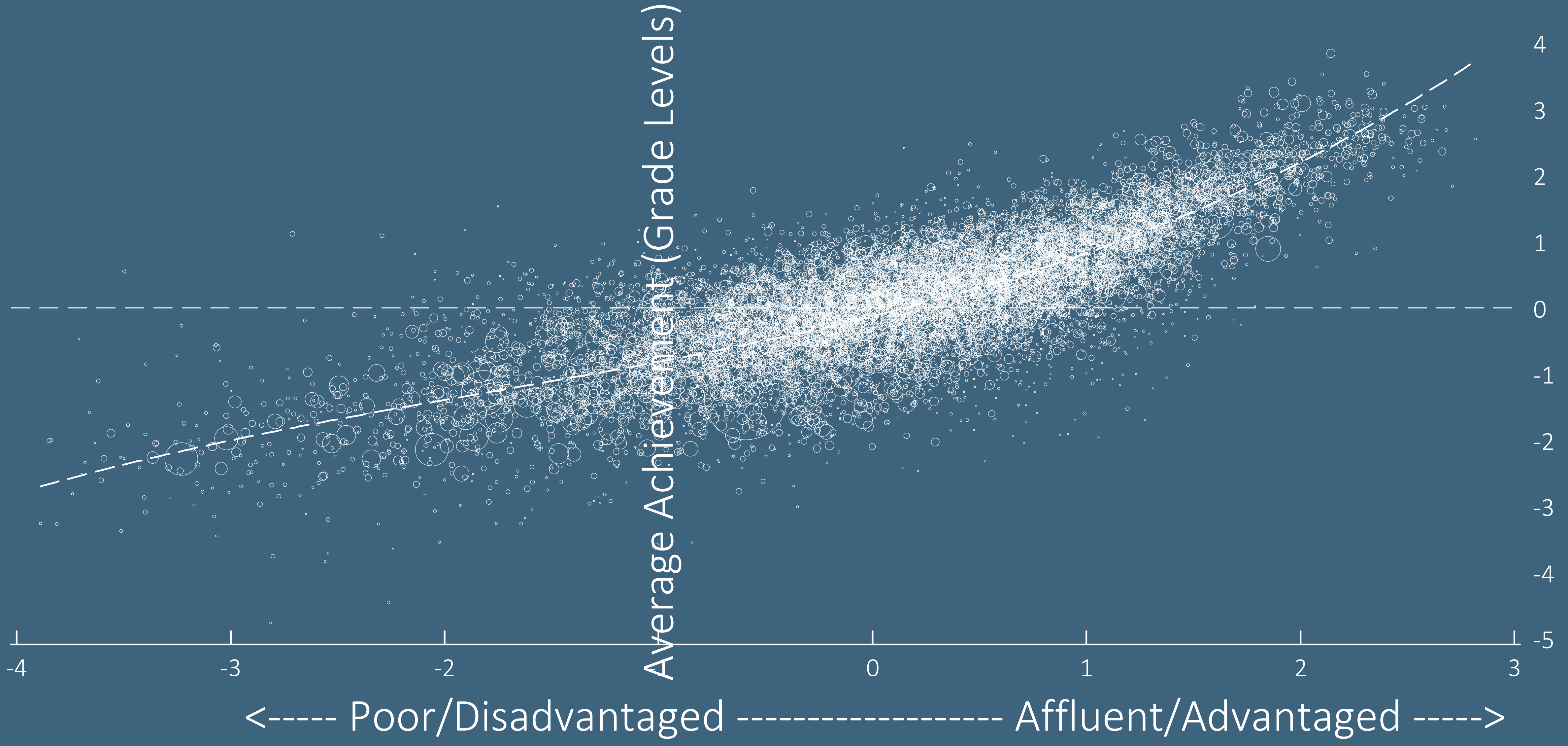
Achievement Gaps in Test Scores by Socioeconomic Status

- Next, use these data to examine how test scores vary across socioeconomic groups

- Define an index of socioeconomic status (SES) using Census data on income, fraction of college graduates, single parent rates, etc.

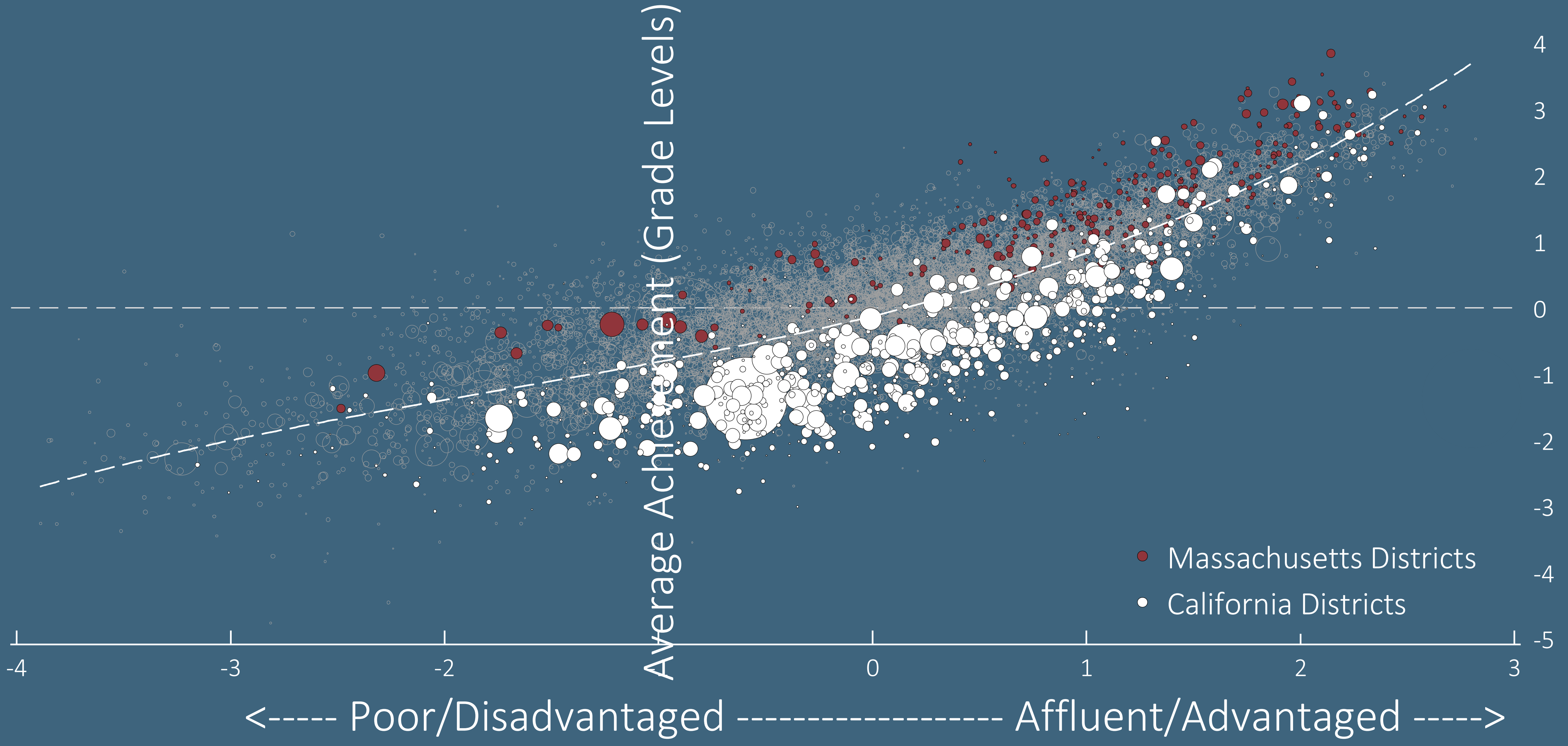
Academic Achievement and Socioeconomic Status

US School Districts, 2009-2013



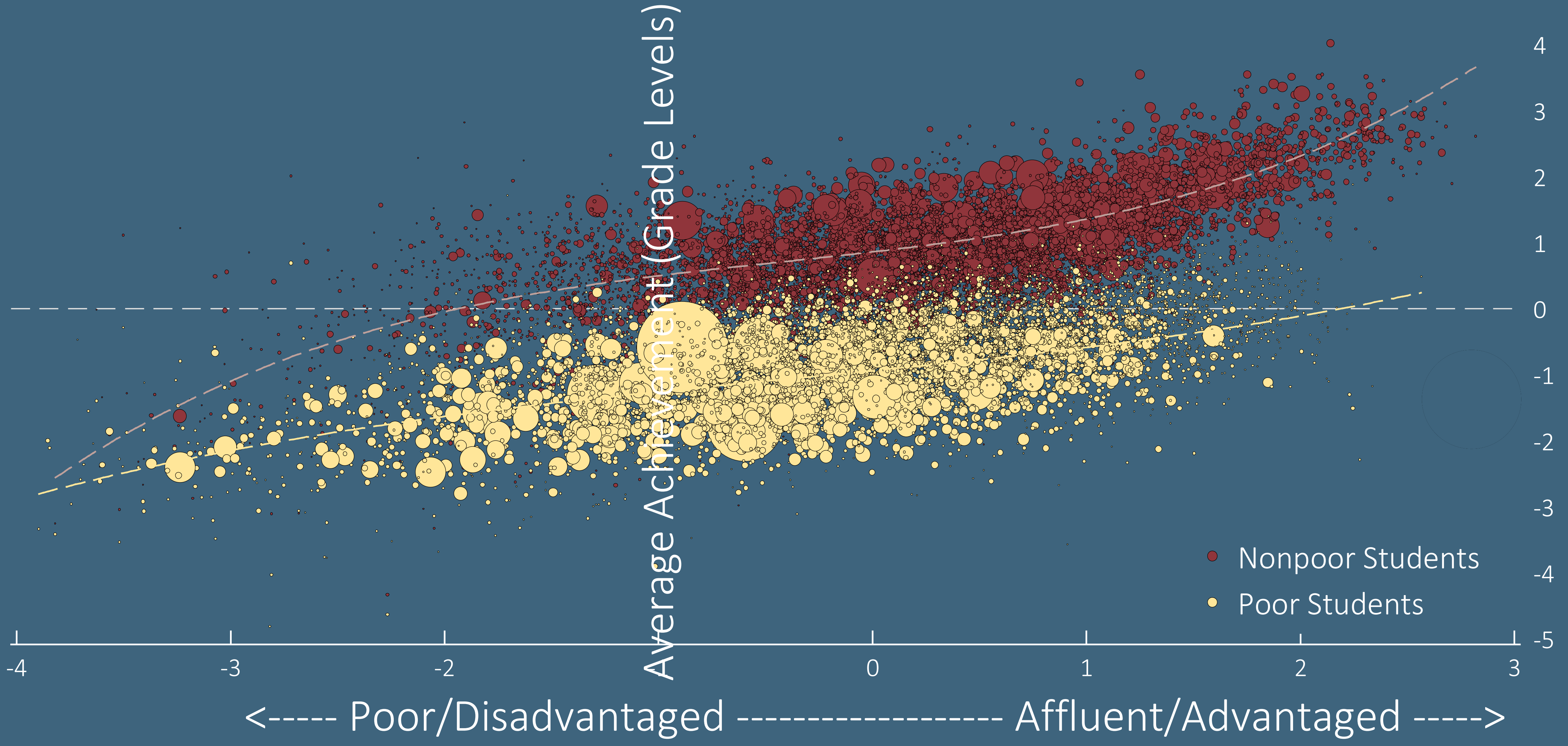
Academic Achievement and Socioeconomic Status

California and Massachusetts School Districts, 2009-2013



Academic Achievement and Socioeconomic Status, by Poverty Status

US School Districts With 20+ Students of a Given Economic Status, 2009-2013

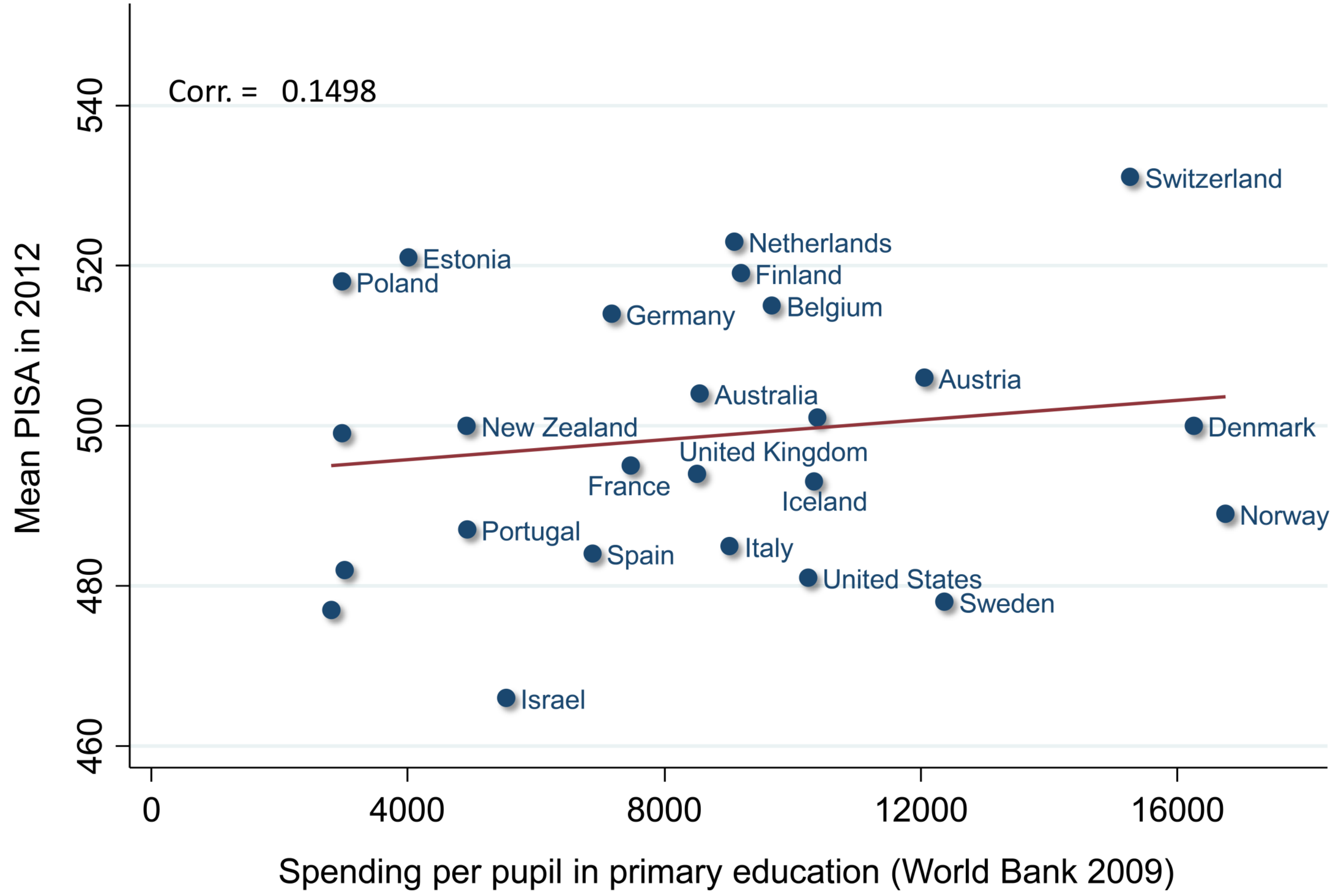


How Can We Improve Poorly Performing Schools?

- There are many school districts in America where students are two grade levels behind national average, controlling for SES

- How can we improve performance in these schools?
  - Simply spending more money on schools is not necessarily the solution…

Test Scores vs. Expenditures on Primary Education Across Countries



Two Policy Paradigms to Improve Schools

- Two distinct policy paradigms to improve schools

  1. Government-based solutions: improve public schools by reducing class size, increasing teacher quality, etc.

  2. Market-based solutions: charter schools or vouchers for private schools

- Contentious policy debate between these two approaches

  – We will consider each approach in turn

Government-Based Solutions: Improving Schools

Improving Schools: The Education Production Function

- Improving public schools requires understanding the education production function

- How should we change schools to produce better outcomes?

Better Teachers?

Smaller Classes?

Better Technology?

Effects of Class Size

- Begin by analyzing effects of class size
- Cannot simply compare outcomes across students who are in small vs. large classes
  - Students in schools with small classes will generally be from higher-income backgrounds and have other advantages
  - Therefore simply comparison in observational data will yield overstate causal effect of class size
- Need to use experimental/quasi-experimental methods instead

Effects of Class Size: Tennessee STAR Experiment

- Student/Teacher Achievement Ratio (STAR) experiment
  - Conducted from 1985 to 1989 in Tennessee
  - About 12,000 children in grades K-3 at 79 schools

- Students and teachers randomized into classrooms within schools
  - Class size differs: small (~15 students) or large (~22 students)
  - Classes also differ in teachers and peers

Effects of Class Size: Tennessee STAR Experiment

- Evaluate impacts of STAR experiment by comparing mean outcomes of students in small vs. large classes

- Report impacts using regressions of outcomes on an indicator (0-1 variable) for being in a small class [Krueger 1999, Chetty et al. 2011]

STAR Experiment: Impacts of Class Size

Dep Var:	Test Score	College Attendance	Earnings
Outcome	(1)	(2)	(3)
Small Class	4.81	2.02%	-$4
	(1.05)	(1.10%)	($327)
Observations	9,939	10,992	10,992
Mean of Dep. Var.	48.67	26.4%	$15,912

STAR Experiment: Impacts of Class Size

Dep Var:	Test Score	College Attendance	Earnings
	(1)	(2)	(3)
Small Class	4.81	2.02%	-$4
Estimated Impact	(1.05)	(1.10%)	($327)
Observations	9,939	10,992	10,992
Mean of Dep. Var.	48.67	26.4%	$15,912

Estimated impact of being in a small KG class:
4.81 percentile gain in end-of-KG test score

STAR Experiment: Impacts of Class Size

	Dep Var:	Test Score	College Attendance	Earnings
		(1)	(2)	(3)
Small Class		4.81	2.02%	-$4
		(1.05)	(1.10%)	($327)
	Standard Error			
Observations		9,939	10,992	10,992
Mean of Dep. Var.		48.67	26.4%	$15,912

95% chance that estimate lies within +/-2 times standard error
→ test score impact between 2.71 and 6.91 percentiles

Repeat experiment 100 times → 95 of the 100 estimates will lie between 2.71 and 6.91 percentiles

STAR Experiment: Impacts of Class Size

Dep Var:	Test Score	College Attendance	Earnings
	(1)	(2)	(3)
Small Class	4.81	2.02%	-$4
	(1.05)	(1.10%)	($327)
Observations	9,939	10,992	10,992
Mean of Dep. Var.	48.67	26.4%	$15,912

Mean Value of Outcome

STAR Experiment: Impacts of Class Size

Dep Var:	Test Score	College Attendance	Earnings
	(1)	(2)	(3)
Small Class	4.81 (1.05)	2.02% (1.10%)	-$4 ($327)
Observations	9,939	10,992	10,992
Mean of Dep. Var.	48.67	26.4%	$15,912

STAR Experiment: Impacts of Class Size

Dep Var:	Test Score	College Attendance	Earnings
	(1)	(2)	(3)
Small Class	4.81	2.02%	-$4
	(1.05)	(1.10%)	($327)
Observations	9,939	10,992	10,992
Mean of Dep. Var.	48.67	26.4%	$15,912

95% chance that estimate lies within +/-2 times standard error
→ Earnings impact could be as large as ***$650*** *(4% increase)*

Effects of Class Size: Quasi-Experimental Evidence

- Limitation of STAR experiment: insufficient data to estimate impacts of class size on earnings precisely

- Fredriksson et al. (2013) use administrative data from Sweden to obtain more precise estimates
  - No experiment here; instead use a quasi-experimental method: regression discontinuity