

The effect of college expansion on college attainment: evidence from historical US censuses

Collin J Wardius

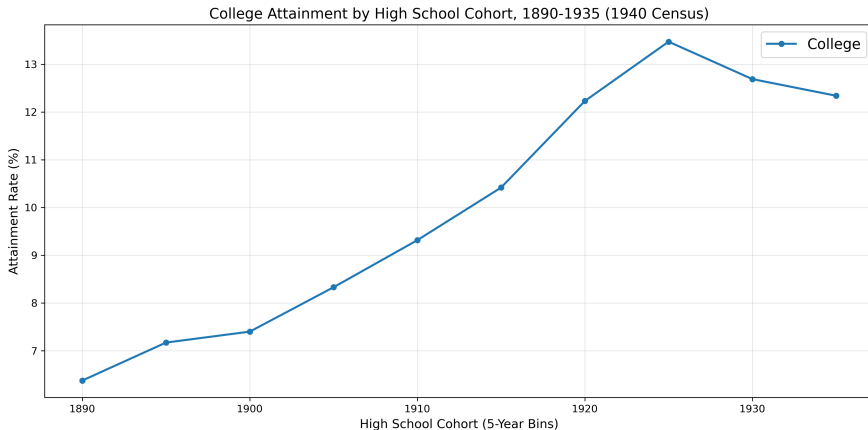
Department of Economics, UC San Diego
Approved by

October 30, 2025

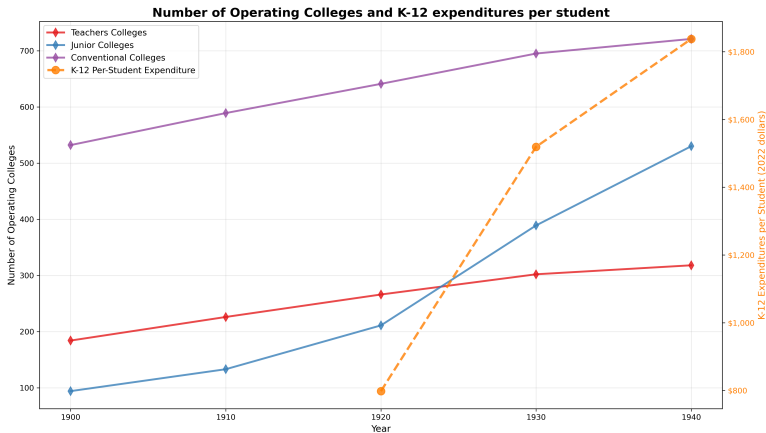
Education in the US experienced a major transformation in the early 1900s

- Many more students completed HS and college
- Massive increase in capacity and spending at all levels of education
- **My focus: dramatic expansion in college openings and enrollment**

By 1940, younger Americans were much more college educated than their parents



The great expansion of educational resources



Research Question

- Do supply-side expansions of colleges drive increased educational attainment?
- Specifically: How did new college openings affect local college attainment?

Literature

- **History of US higher education (1900-1940)**

- *My contribution:* Quantify the causal effect of university expansion on education access
 - Goldin (1998), Goldin and Katz (1998), Goldin (2001)

- **Effects of school building in non-US countries**

- *My contribution:* US university foundings and variation in public vs private control
 - Duflo (2001), Nimier-David (2023)

- **How proximity to college affects attainment and earnings**

- *My contribution:* Examine extensive margin of college access via new university foundings
 - Card (1993), Acton et al. (2025)

- **Historical census analysis to answer current questions in economics**

- *My contribution:* Create a dataset of university expansions and link them to the census data
 - Abramitzky, Boustan, and Eriksson (2014), Derenoncourt (2022), Bleemer and Quincy (2025)

Data

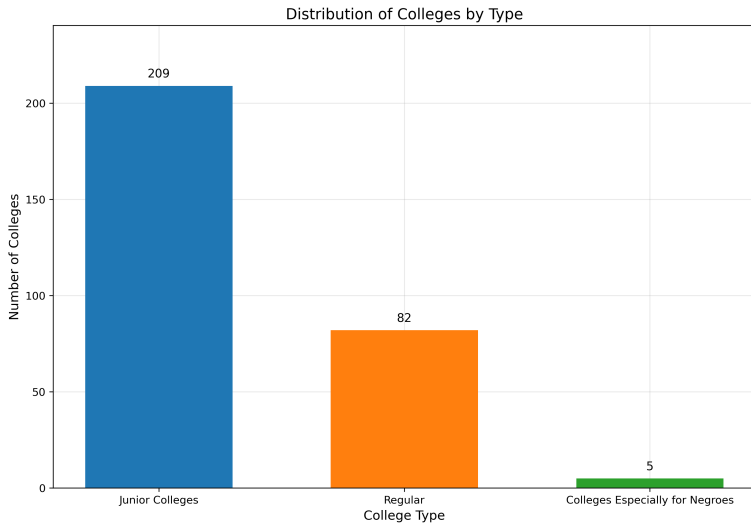
- **1900-1940 Decennial, Linked Full-Count US Censuses** Ruggles et al. (2025): Adult outcomes measured in 1940 (occupation, income, education, location); childhood location (pre-18) linked from earlier censuses to assign treatment
- **1947 College Blue Book** H.W. Hurt, H.J. Hurt, and Burckel (1947): college founding year, enrollment, student capacity, state or private control, location
- **Biennial Surveys of Education and Commissioner's on US Education**: college-level data on enrollment, finances, faculty, and programs (novel data in the process of being digitized by me)

Preview of identification approach

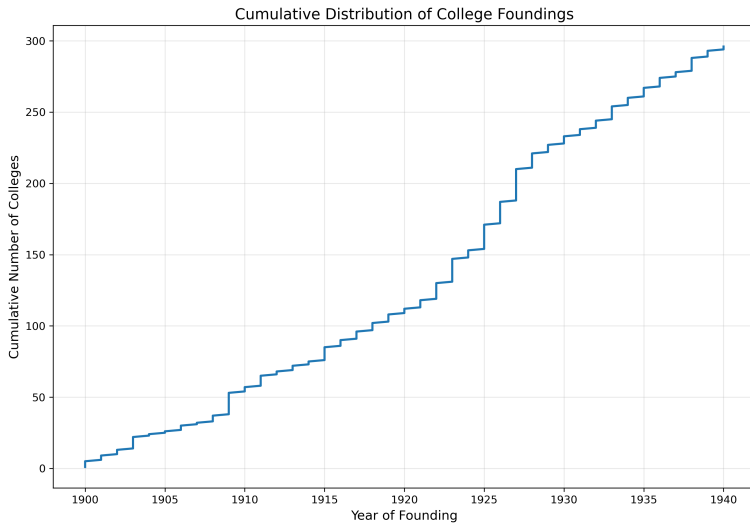
- **Identifying variation:** quasi-random founding date of a university
- Some people are born just late enough to access a new university
- Some people are born too early to access a new university

What data do we need to

New college types



New college founding dates



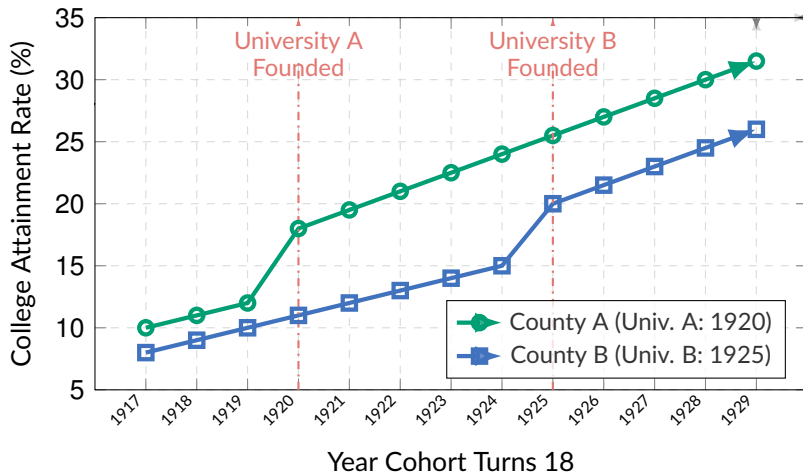
Estimating the effect of a university founding on college attainment: cohort DD approach

Cross sectional regression, identifying variation is at the age cohort-by-county level.

$$y_{ick} = \alpha_c + \lambda_k + \sum_{j \neq -1} \beta_j \mathbb{1}\{\text{Cohort } k \text{ born } j \text{ years relative to college founding in } c\} + \xi \mathbf{X}_{ick} + \epsilon_{ick} \quad (1)$$

- i : individual, c : pre-18 county, k : birth cohort
- $j < 0$: Cohorts born *before* college founding (test for pre-trends)
- $j \geq 0$: Cohorts born *after* college founding (treatment effects)
- **Identifying assumption:** Conditional on controls, counties that gained a college would have experienced parallel trends in attainment across cohorts absent the new college

Visualization of the identification assumption



Determining which individuals experiences a college expansion

We only observe education and location in 1940, after individuals either received or did not receive a college education

1. Identify adults (age 25+) in the 1940 census
2. Link back to the censuses for which they are below the age of 18 using Ruggles et al. (2025) longitudinal linkage
3. If an individual is observed twice before 18, take the latest observation
4. Assign the individual that county of residence for the purposes of treatment assignment

Comparing linked versus unlinked individuals in the census

Table: Comparison of 1940 Characteristics: Linked vs Unlinked Individuals

	Linked Mean	Unlinked Mean	Difference
Female (%)	23.3	63.3	-40.0
Age	37.2	49.7	-12.5
College (%)	13.4	7.9	5.5
Married (%)	70.5	83.7	-13.1
White (%)	95.1	93.1	2.0
N	18,521,950	26,557,936	
% of Total	41.1%	58.9%	

Note: This table compares mean characteristics in 1940 for individuals age between 25 and 70 who were successfully linked to pre-age 18 observations versus those who were not linked.

Effects on college and BA attainment

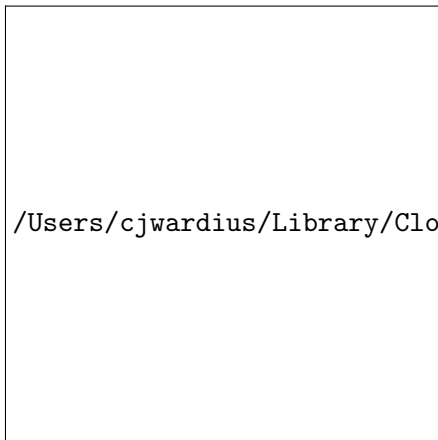
Table: Effect of College Opening on College Attainment

	(1) Any College	(2) BA Degree
Exposed to new college	0.006** (0.002)	0.001 (0.001)
Observations	1838538	1838538
N counties	296	296
Mean of Dep. Var.	0.134	0.062

All regressions control for county and age cohort FE. SEs clustered at the county level.

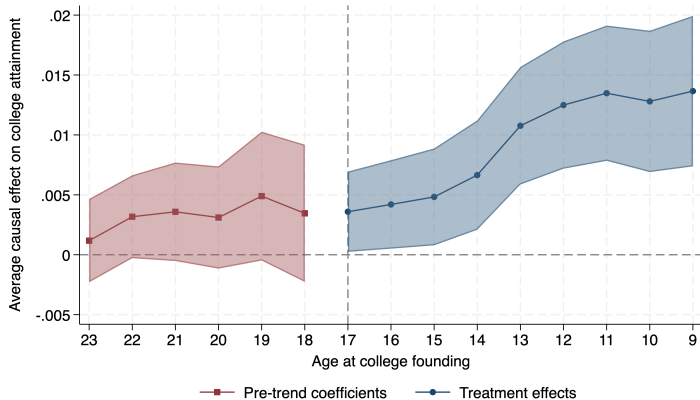
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Dynamic effects on college attendance



/Users/cjwardius/Library/CloudStorage/OneDrive-UCSanD

Dynamic effects adjusting for treatment heterogeneity



County spatial stability over 1900-1940

Table: County Boundary Stability Between 1900 and 1940

Overlap Threshold	Reference Period	
	1940 Counties	1900 Counties
Total Counties	3108	2848
99% or more overlap	2852 (91.8%)	2538 (89.1%)
95% or more overlap	2941 (94.6%)	2616 (91.9%)
90% or more overlap	2976 (95.8%)	2647 (92.9%)
80% or more overlap	3005 (96.7%)	2681 (94.1%)

Notes: The 1940 Counties column shows the percentage of 1940 counties that overlap with a single 1900 county at the specified threshold. The 1900 Counties column shows the percentage of 1900 counties that overlap with a single 1940 county.

Review of estimates of school construction on attainment

Table: Effects of school building on educational outcomes

Study	Country	Intervention	ID Strategy	Years of Schooling (SE)	Other Outcomes (SE)
<i>Panel A: Indonesia INPRES Program</i>					
Duflo (2001)	Indonesia	INPRES primary schools	DiD	0.12-0.19 (0.025-0.029) per school/1,000; Wage earners: 0.20-0.26 (0.042-0.050)	Log wages: 0.015-0.027 (0.007-0.009); Returns (IV): 0.068-0.106 (0.028-0.022) per year
Hsiao (2024)	Indonesia	Same PRES	IN- DiD + spatial eq.	Full: 0.085 (0.036); Wage: 0.099 (0.039) per school/1,000	Log wages: 0.026 (0.009); Mkt access doubles education effects
Akresh et al. (2023)	Indonesia	Same PRES	IN- DiD	M: +0.27; F: +0.23 per school/1,000	Formal work (M): +1.1 pp; Migration: +0.7-0.8 pp; Expenditure: +2.1-3.2%; Child educ: +0.10-0.17 yrs

Notes:

Creating a county crosswalk

We need consistent county boundaries to accurately assign which people experience a college creation versus which do not.

Approach:

1. Use 1940 as the reference year
2. Spatially intersect 1900, 1910, 1920, and 1930 boundaries with 1940 boundaries
3. Match counties where the intersection exceeds 70% overlap
4. Retain only counties that appear consistently across all census years

[Back to Isolating treated and control counties](#)

My reference on how the spatial join is performed

Consider A and B from 1900 (the base year) and 1940 (the target year) respectively. Then get the area of $A \cap B$. We then compare this to the area of the target to calculate

$$\frac{A \cap B}{B} \quad (2)$$

and we map B to A if this is above some threshold. I am using 70% as of right now but this could be modified. In practice, almost 90% of the counties have close to 100% overlap.