

# The democratization of US higher education (1900-1940)

Collin J Wardius

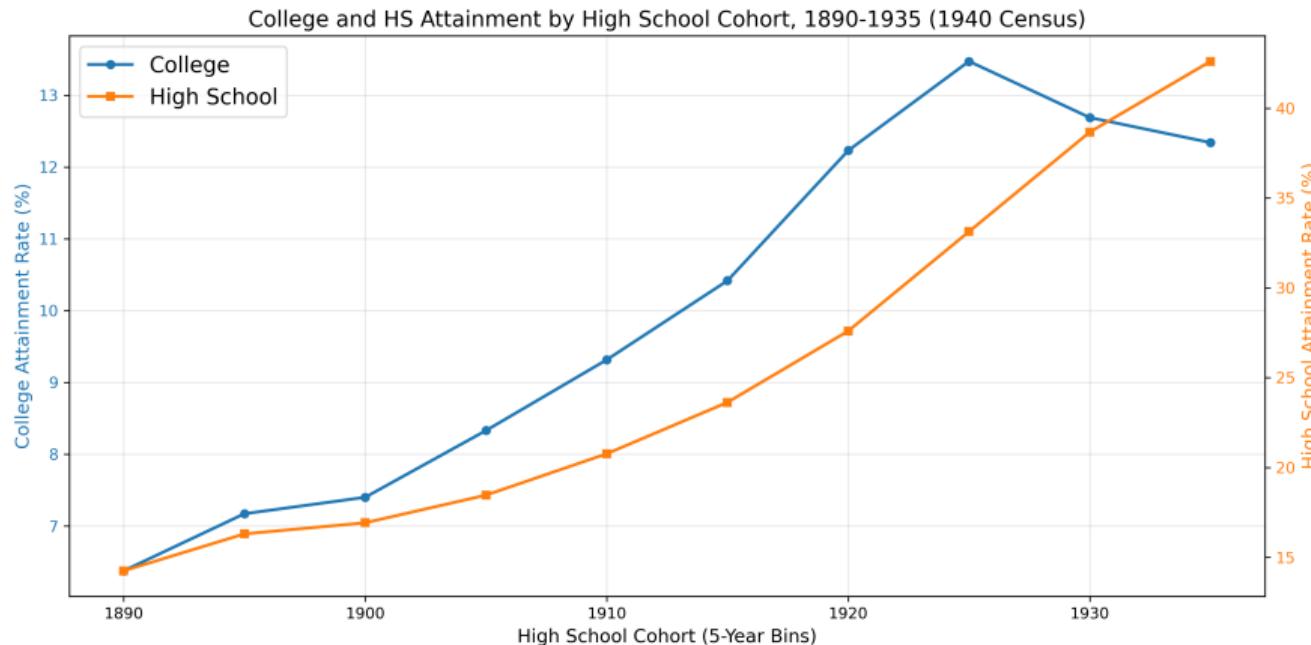
Department of Economics, UC San Diego  
Read by Fabian E, Gaurav, and Julian

🎃 October 30, 2025 🎃

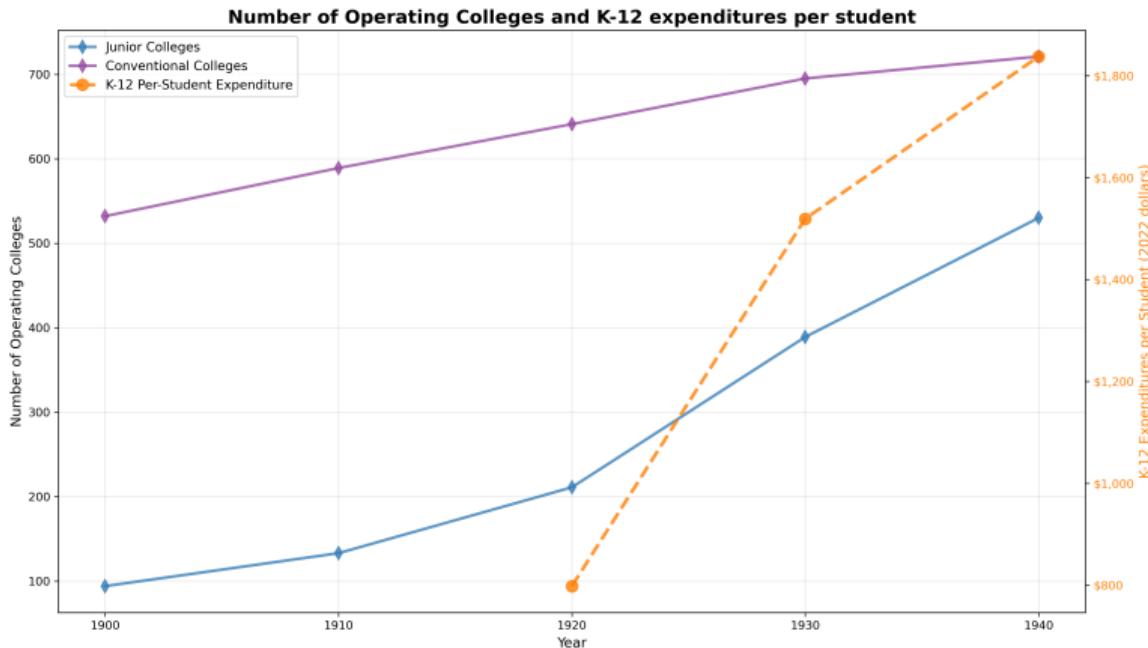
## Education in the US experienced a major transformation in the early 1900s

- Many more students completed high school and college
- Massive increase in capacity and spending at all levels of education

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



# The great expansion of educational resources



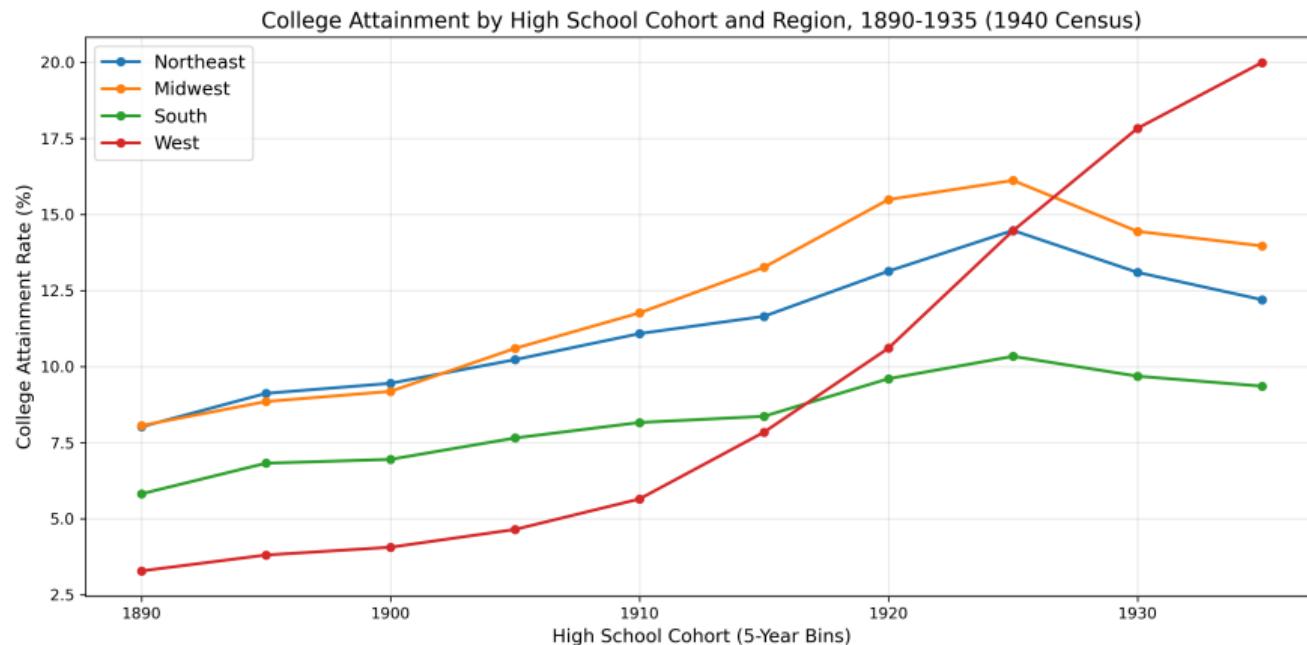
## Educational gains were concentrated in the public-college-oriented West

- People born in the American West went from being the least to the most college educated among people born in any region
- Uniquely, the West's higher education model was centered on public provision of higher education

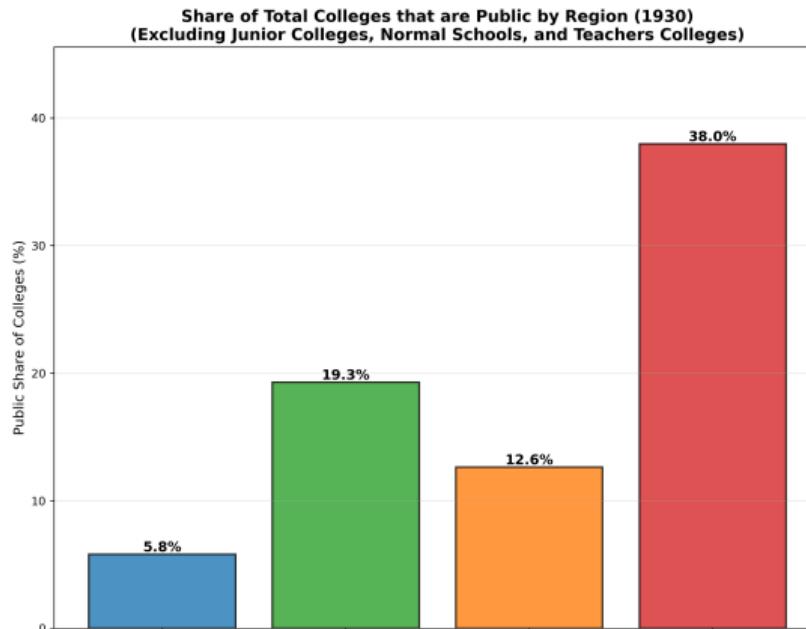
**“Democracy must, if it is to survive and prosper, develop ‘an aristocracy of its own begetting, after its own heart, and dedicated to its own service’; and to that end must provide somewhere the best facilities for the highest education, open freely to all who have the brains and the industry to make use of them.”**

— Robert Gordon Sproul  
President, University of California  
Inaugural Address, October 22, 1930

# The rise of college attainment in the West



# The West's focus on public higher education

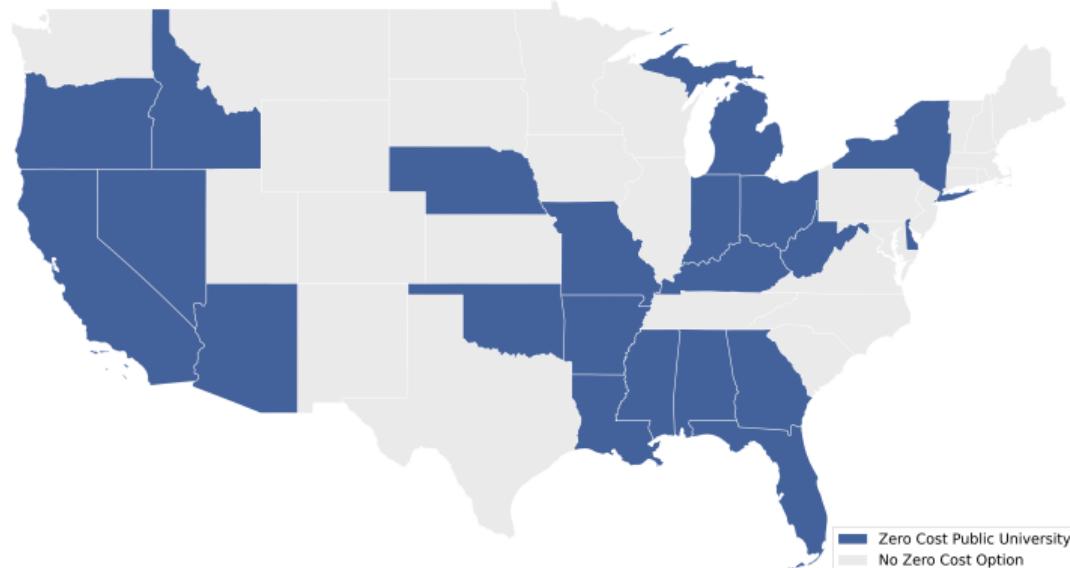


Regional k-12 trends

Regional junior colleges

# The West's focus on public higher education

**States with Zero Cost Public Universities**



## The broad research agenda of this project

1. How did the public college expansion from 1900 to 1940 affect access to college?
2. Quantify the contribution of the public college expansion to the development of the American West

## Current results

- The broad questions of the project presuppose that college expansions have an effect on local college attainment. Do they?
- **Answer: Yes**
- Individuals below age 18 at the time of a college founding are **5-10% more likely to attend college** compared to those above age 18

## The plan for today

- Construct a dataset of county-level, college founding “experiments”
- Use a cohort difference-in-differences approach to estimate the causal effect of a college founding on college attainment

# Literature and contribution

- **History of US higher education (1900-1940)**
  - *Contribution:* Estimate the causal effect of college expansion on education access
    - Goldin (1998), Goldin and Katz (1998), Goldin (2001)
- **Effects of school building in non-US countries**
  - *Contribution:* US college foundings and variation in public vs private control
    - Duflo (2001), Nimier-David (2023)
- **How proximity to college affects attainment and earnings**
  - *Contribution:* Examine extensive margin of college access via new college foundings
    - Card (1993), Acton et al. (2025)
- **Historical US census analysis to answer current questions in economics**
  - *Contribution:* Create a dataset of college expansions and link them to the census data
    - Abramitzky, Boustan, and Eriksson (2014), Derenoncourt (2022), Bleemer and Quincy (2025)
- **Educational investments in general equilibrium**
  - *Contribution:* ...
    - Khanna (2023), Hsiao (2024), Eckert and Kleineberg (2024)

## 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 determine exposure to college foundings
- **1947 College Blue Book (CBB)** 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 Reports on US Education:** college-level data on enrollment, finances, faculty, and programs (novel data in the process of being digitized by me)

# Digitizing the College Blue Book

Digitize using Claude Sonnet 4.5 (Anthropic (2025)) with the prompting approach of Backer-Peral, Meursault, and Severen (2025)

ALABAMA ARKANSAS EDUCATIONAL ATLAS							
COLLEGE OR UNIVERSITY	LOCATION		YEAR FOUNDED	STUDENT CAPACITY	ENROLLMENT		RECOGNIZED BY
	CITY OR TOWN	MAP			M	W	
<b>ALABAMA</b>							
1 Alabama College	W Montevallo	C-4	1 1896 State	900	671	ASDOU	
2 Alabama Poly. Inst. <sup>S</sup>	C Auburn	E-5	4 1872 State	834	595	229	ASDO
3 Alabama University	C University	B-2	1831 State	5500	3577	2267	ASDGUY
4 Athens College	C Athens	C-1	4 1843 ME So.	500	182	277	DO
5 Birmingham So. Coll.	C Birmingham	D-3	267 1856 ME So.	950	550	400	ASDOUY
6 Howard College	C Birmingham	D-3	267 1842 Baptist	1350	678	350	ASDOUY
7 Huntingdon College	W Montgomery	D-5	78 1854 ME So.	650			SDOUY
8 Judson College	W Marion	D-5	2 1838 Baptist	317		317	SDORY
9 Spring Hill College	M Spring Hill	B-9	b 1830 R.C.	350	270		ASDOY
10 State Teachers Coll. <sup>S</sup>	C Florence	B-1	15 1872 State	600	515	600	STD
11 State Teachers Coll. <sup>S</sup>	C Jacksonville	E-3	2 1888 State	1000	414	590	STD
12 State Teachers Coll. <sup>S</sup>	C Livingston	B-5	1 1888 State	350	55	249	*STD
13 State Teachers Coll. <sup>S</sup>	C Troy	E-6	7 1887 State				STD
Colleges Especially for Negroes							
14 Miles College	C Birmingham	C-3	267 1907 C.M.E.	450	94	252	D6
15 Moundville St. T. C.	G Montgomery	D-5	66 1894 State				
16 Oakwood College	G Huntsville	D-3	13 1889 State Ad.				JSD
17 Selma University	G Selma	C-5	18 1878 Baptist	795	300	500	C
18 St. Agri. & Mech. Inst. <sup>S</sup>	G Normal	D-1	b 1875 Public	395	180	210	SD
19 Stillman Institute	C Tuscaloosa	B-4	26 1875 Presby.				
20 Talladega College	C Talladega	D-4	9 1867 A.M.A.	340			SDX
21 Tuskegee Institute <sup>S</sup>	C Tuskegee	E-5	4 1881 Non-B.	2591	1747	844	SDX
Junior Colleges							
22 Marion Institute	M Marion	C-5	2 1842 Private	250	238		DUSJ
23 Sacred Heart Jr. Coll.	W Culver	b 1906 R.C.	175		175		DUJ
24 Snead Junior College	C Boaz	D-2	2 1898 M.E.	177	80	157	JSU
25 Southern Union College	C Wadley	E-4	b 1923 Private	70	23	50	OJD
26 St. Bernard College	M St. Bernard	D-2	b 1892 R.C.	300	200		SO
27 Walker Junior College	C Jasper	C-3	6 1938 Private	62	20	40	DUJ
<b>ALASKA</b>							
1 University of Alaska	C College	b	1917 Terr.	250	180	70	WDO
<b>ARIZONA</b>							
1 Arizona State College <sup>S</sup>	C Flagstaff	D-4	5 1899 State		357	215	NTDO
2 Arizona State College <sup>S</sup>	W Tempe	D-6	3 1885 State	1500	750	696	NTDO
3 Arizona University	C Tucson	E-8	36 1885 State	4000	1033	2478	ANDOUY
Colleges Especially for Negroes							
NONE							
Junior Colleges							
4 Gila Junior College	C Thatcher	F-7	1 1891 Dist.	400	153	150	DUOJ
5 Phoenix Junior College	C Phoenix	C-6	65 1926 Dist.	1450	700	750	DUNJ

# Digitizing the College Blue Book: Example Prompt

## Prompt to Claude Sonnet 4.5

"Use your OCR vision for this task. Please extract the tabular data here to a CSV.

- Extra large headers are **state names** – apply to all colleges until the next state
- Bold headers without numbers classify the **type of college** – apply to succeeding colleges until the next bold header or state name
- No rows should be blank except for the first column
- Verify text outputs to the correct column (watch for nested headers)"

# Biennial Surveys / Commissioner's Reports: Income

418

BIENNIAL SURVEY OF EDUCATION, 1928-1930

TABLE 9.—Receipts—Individual

Institution	From student fees			From produc-tive funds
	For tui- tion and other edu- ca-tional ser-vices	For board and room rent	For other nonedu- ca-tional ser-vices	
1	2	3	4	5
1 Alabama College.....	\$86,043	\$209,115		\$34,963
2 Alabama Polytechnic Institute.....	145,028	66,288	\$63,451	1,20,280
3 University of Alabama.....	399,060	101,842		130,167
4 Alaska Agricultural College and School of Mines.....	961	6,500		
5 University of Arizona.....	138,894	102,516	66,798	1,42,144
6 Arkansas Polytechnic College.....	4,800	31,700		
7 Arkansas State College.....	6,785	33,796	337	
8 State Agricultural and Mechanical College, Jonesboro.....	24,065	24,158	2,331	
9 State Agricultural and Mechanical College, Magnolia.....	2,700	43,104		
10 State Agricultural and Mechanical College, Monteith.....	6,000			
11 University of Arkansas.....	139,525	6,095		1,6,335
12 California Polytechnic School.....				
13 University of California.....	1,747,966	98,429		1,695,640
14 Colorado Agricultural College.....	94,217		42,661	2,28,338
15 Colorado School of Mines.....	56,287			
16 Fort Lewis School of Colorado Agricultural College.....	1,400	14,991	11,568	
17 Colorado School of Grand Junction, Colo.....	7,705			
18 University of Colorado.....	499,200			
19 Connecticut Agricultural College.....	91,862	107,701	21,814	1,282
20 United States Coast Guard Academy.....				
21 University of Delaware.....	96,289	107,389	5,839	1,27,255
22 Gallaudet College.....				
23 Florida State College for Women.....	84,594	296,266		4,100
24 University of Florida.....	190,682	119,584	250,740	1,10,079
25 Georgia School of Technology.....	304,250	85,147		6,500
26 Georgia State College of Milledgeville.....	11,499			
27 Georgia State College for Women.....	47,753	233,843		
28 Medical College, University of Georgia.....	16,200			1,395
29 North Georgia College.....	4,000			
30 South Georgia State College.....	2,720			
31 University of Georgia.....	238,911	128,969		1,19,118
32 University of Hawaii.....	73,633	27,382	2,131	
33 University of Idaho.....	109,956	11,120	2,117,123	
34 University of Illinois.....	87,088	152,099	2,32,111	
35 Indiana University.....	528,149	139,580		4,805
36 Purdue University.....	406,234	314,707	214,004	17,000
37 Iowa State College of Agriculture and Mechanic Arts.....	456,437	110,140		33,034
38 State University of Iowa.....	734,830	289,412		12,082
39 Kansas State Agricultural College.....	292,035	90,781	115,638	3,32,816
40 Municipal University of Wichita (Kans.).....	117,462	18,302	32,403	4,886
41 University of Kansas.....	204,246	93,169		7,14,969
42 Kentucky Polytechnic.....	289,263	95,105		10,334
43 University of Louisville.....	7,328	27,239		
44 Louisiana Polytechnic Institute.....	41,206	60,253	33,734	
45 Louisiana State University and Agricultural and Mechanic Colleges.....	104,473	80,736	14,555	

on 2025-09-18 16:47 GMT  
single-digitized

STATISTICS OF UNIVERSITIES AND COLLEGES

419

publicly controlled institutions,<sup>1</sup> 1929-30

From State or city	From United States Govern- ment	From private benefactions			From all other sources	Total re- ceipts ex- clusive of addi- tions to endow- ment		
		For in- crease of plant	For cur- rent expenses	For in- crease of plant	For endow- ment			
6	7	8	9	10	11	12	13	14
\$106,005	\$284,154	\$3,491	\$1,300			\$46,758	\$771,829	\$771,829
212,500	\$194,261	381,551				631,088	2,470,193	2,470,193
62,500	611,062	8,269	100	\$6,164		12,967	1,332,131	1,325,967
				60,000		1,819	1,170,740	1,169,840
55,000	\$82,745	196,021	15,984	3,317		110,563	1,619,216	1,615,999
						8,000	44,500	44,500
						828	129,000	129,000
						12,124	158,156	158,156
						46,670	191,520	191,520
						12,000	123,200	123,200
						35,100	1,575,167	1,575,167
						80,000	80,000	80,000
						10,386	112,516	112,516
						135,829	15,516,122	15,516,122
						35,039	374,100	374,100
						435	106,159	106,159
						5,080	12,791	12,791
						2,000	2,230,782	2,230,782
						700	280,890	1,249,490
							32,750	32,750
							1,370,613	1,315,643
							120,000	120,000
							1,150,711	1,150,711
							1,284,254	1,284,254
							95,067	2,616,024
							4,990	869,750
							15,611	101,385
							493,606	493,606
							40,011	159,862
							56,000	56,000
							39,408	39,408
							1,750,568	1,750,568
							31,081	31,081
							61,577	61,577
							108,411	108,411
							1,284,254	1,284,254
							7,362,112	7,362,112
							518,753	518,753
							49,786	49,786
							7,362,650	7,362,650
							13,678	13,678
							3,637,908	3,637,908
							32,745	32,745
							1,045,392	1,045,392
							554,810	554,810
							4,069,131	4,069,131
							115,033	115,033
							1,855,077	1,855,077
							5,628,184	5,628,184
							2,790	2,790
							308,591	308,591
							2,885,428	2,885,428
							500	500
							3,783	3,783
							434,610	434,610
							2,742,344	2,742,344
							3,742,344	3,742,344
							1,716,065	1,716,065
							1,716,065	1,716,065

17/32

## Digitizing Biennial Surveys / Commissioner's Reports

- Claude-based digitization is not reliable
- Feed PDFs into Amazon Textract, which returns a CSV file
- Post-processing is very time-intensive:
  - Checking back to the original documents to create variable names
  - Harmonizing variable names across CSVs
  - Correcting for formatting variations (tables split across multiple pages, tables without headers, ...)

## Why do this work?

- Data on the intensive margin of funding (e.g., California ties funding to enrollments in 1911)
- Data on available programs
- Data on tuition

## Preview of identification approach

- **Identifying variation:** College founding
- Some people are born in counties just late enough to access a new college
- Some people are born in counties too early to access a new college

# What data do we need to perform this analysis?

1. Counties where a college opens
2. Where people were before they turned 18

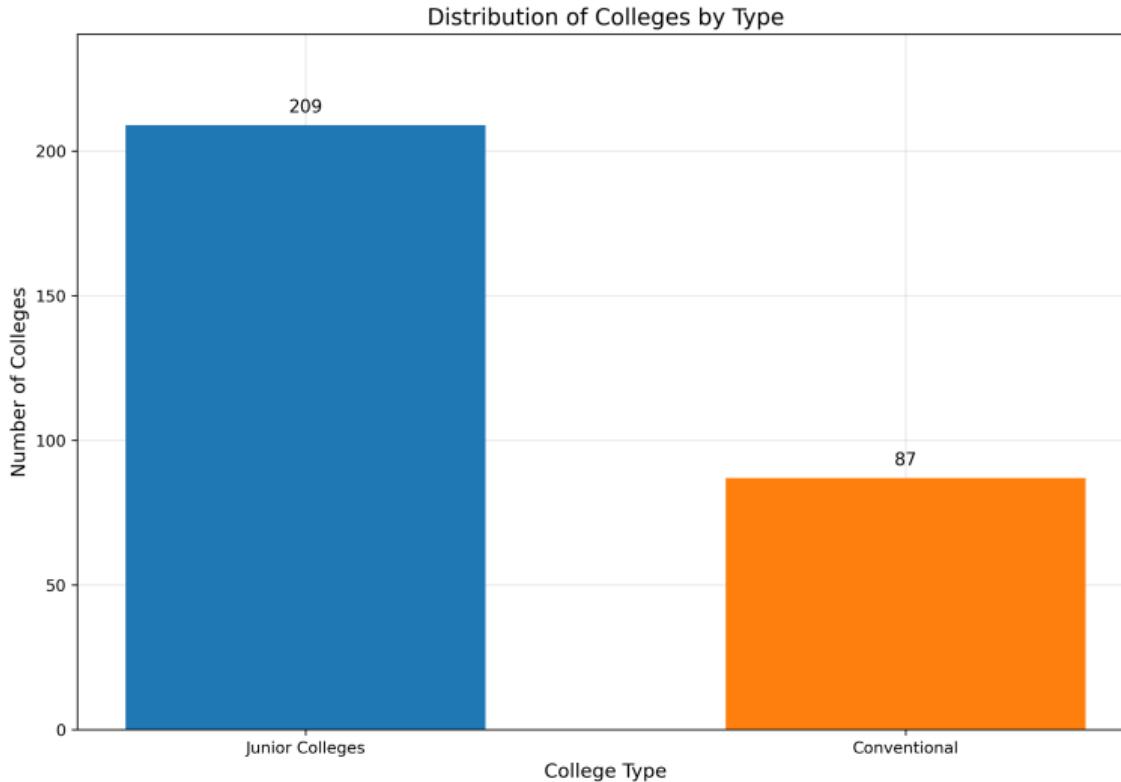
## Identifying counties that experienced a college expansion

- Harmonize county boundaries across census years
- Intersect county boundaries with geocoded college coordinates from the CBB
- Restrict to counties that **gained exactly one college between 1900 and 1940 and had no colleges prior**
- **The final sample includes 296 county-level “experiments”**

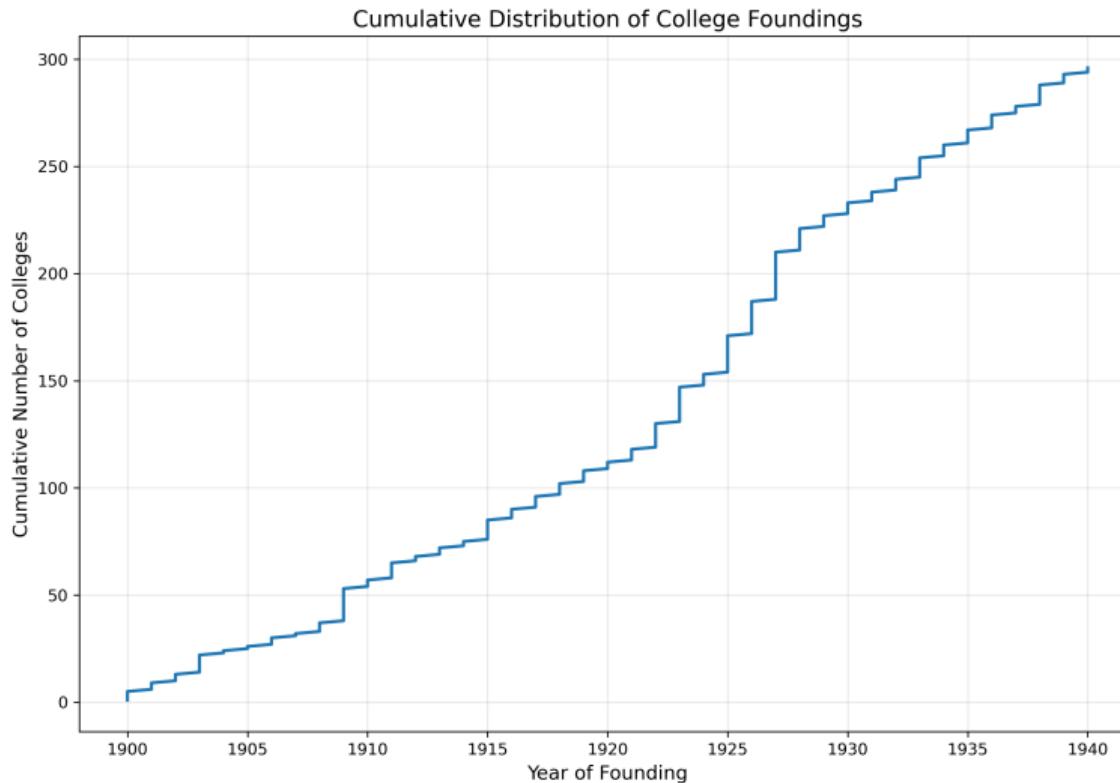
County spatial stability

Creating county crosswalk

# Heterogeneity in college type



# New college founding dates



## Determining whether an individual lived near a college founding

1. Identify adults (age 25-70) in the 1940 census, observe time-invariant demographics and educational attainment
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 their 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 aged between 25 and 70 who were successfully linked to pre-age 18 observations versus those who were not linked.

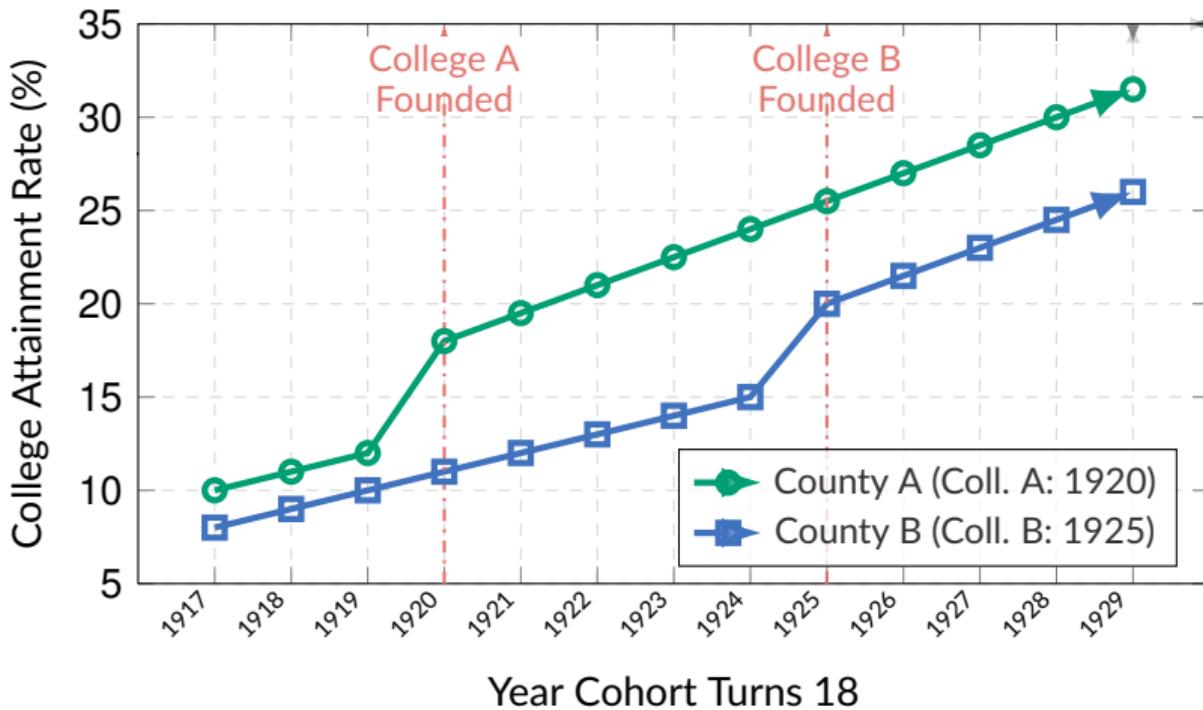
## Estimating the effect of a college founding on college attainment: cohort DD approach

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

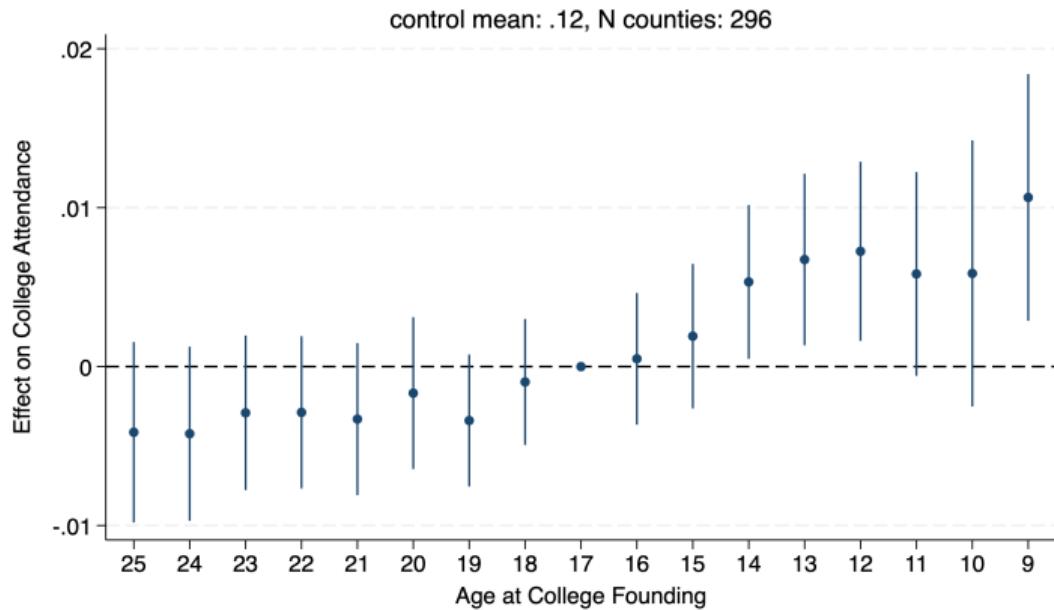
$$y_{ick} = \alpha_c + \lambda_k + \sum_{j \neq 17} \beta_j \mathbb{1}\{\text{Cohort } k \text{ age } j \text{ at time of college founding in } c\} + \gamma \mathbf{X}_{ick} + \epsilon_{ick} \quad (1)$$

- $i$ : individual,  $c$ : pre-18 county,  $k$ : birth cohort
- $j < 17$ : Cohorts young enough to benefit from the new college (treatment effects)
- $j > 17$ : Cohorts too old to benefit (test for pre-trends)
- **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



# Effect of college founding on college attendance



**Figure:** Standard errors clustered at the county level. All specifications include birth year, county, nativity, race, mother's birth place, father's birth place, and sex fixed effects along with controls for pre-18 residential moves.

## Robustness exercises

- TWFE variation in treatment timing bias [here](#)
- Controlling for county-by-cohort linear time trends [here](#)
- Restricting the sample to only counties that have minimum 30 observations in each age bin [here](#)
- Placebo test comparing attainment of individuals born before versus far before the college founding [here](#)

# Next steps

## 1. Heterogeneity analysis

- 1.1 On the person-side, estimate effects by race, sex, and parent's income
- 1.2 On the college-side, estimate effects by junior versus conventional college and state versus non-state college
- 1.3 Complementarity in investment between HS education spending and college founding

## 2. Other identification exercises on the effect of college expansion on attainment

- 2.1 Incorporate demographic-specific college expansions [here](#)
- 2.2 Incorporate distance from college founding

## 3. Data

- 3.1 Fully digitize the Biennial Surveys / Commissioner's Reports on Higher Education

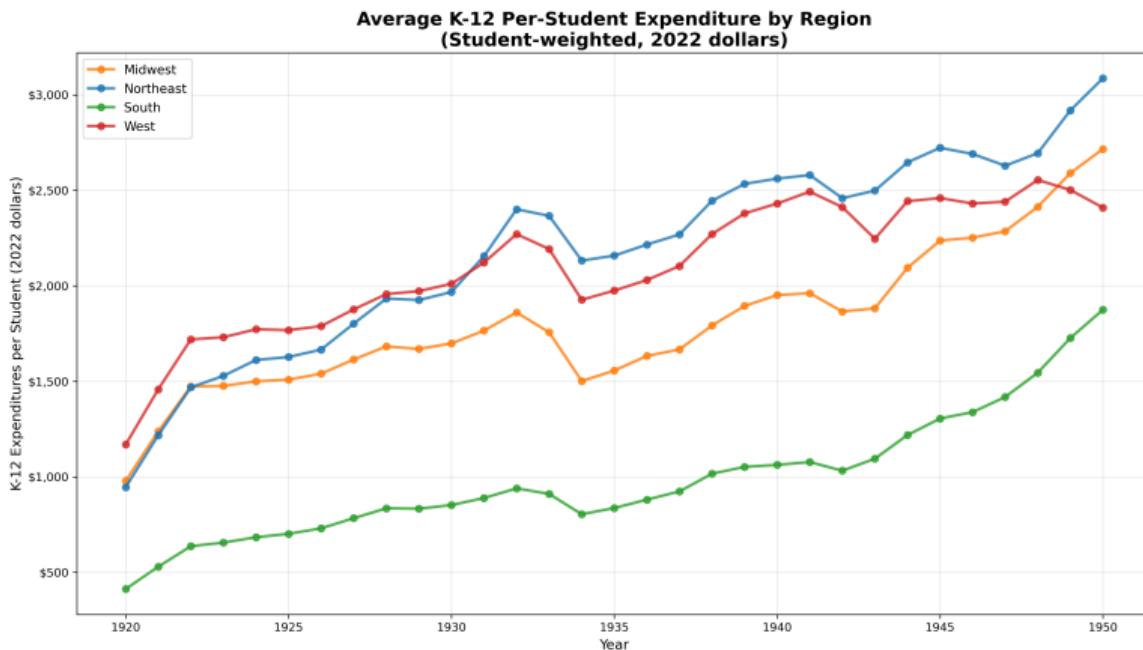
## 4. Model

- 4.1 Write down a two-stage model where individuals (1) invest in education then (2) migrate for work. College expansions decrease the cost of education. [rough sketch](#)

cwardius@ucsd.edu

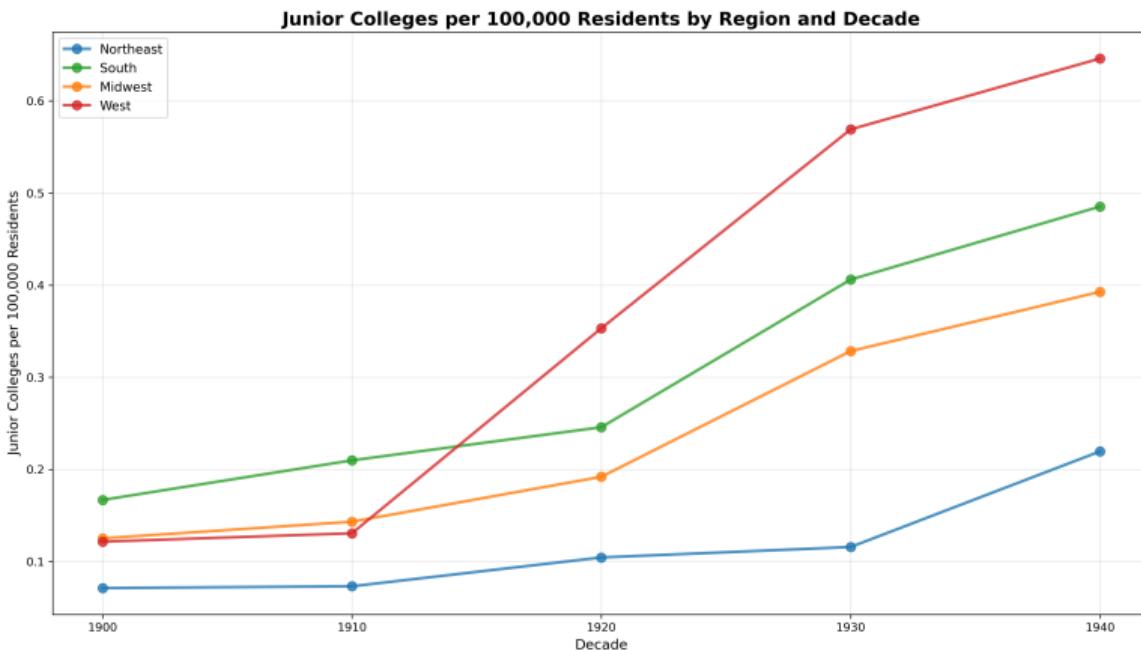


# Regional trends in K-12 spending per capita



Back to slide

# Regional trends in K-12 spending per capita



[Back to slide](#)

## 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.

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

# Creating a county crosswalk

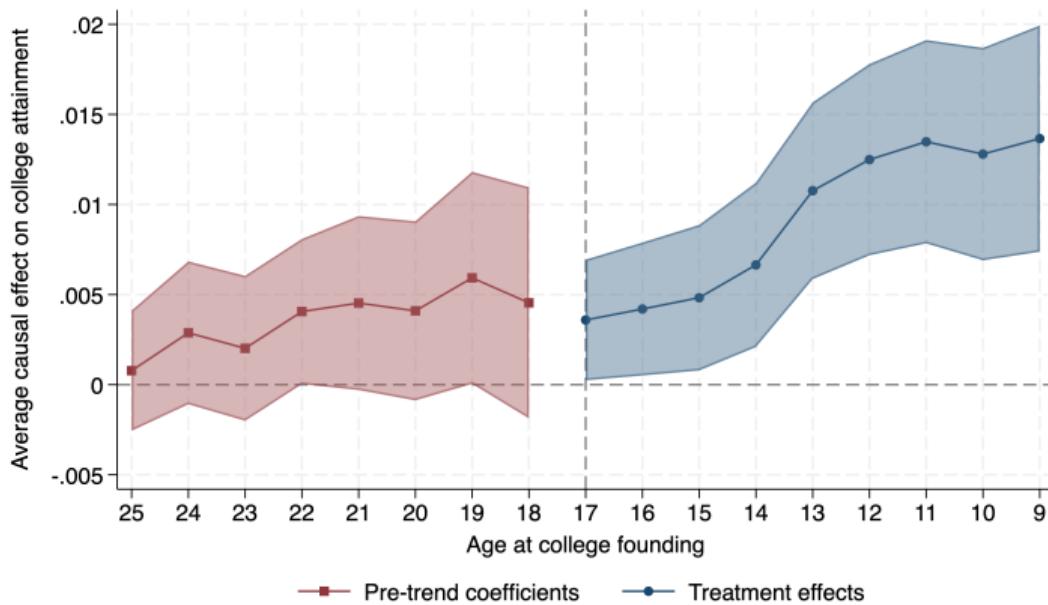
We need consistent county boundaries to accurately assign which people experience a college creation versus which do not.

## 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

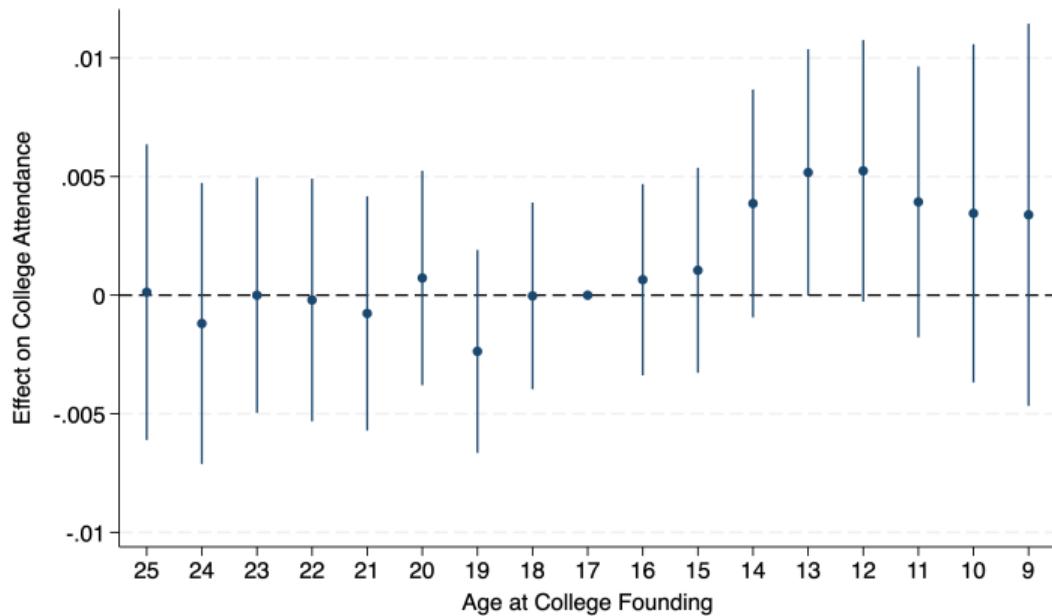
[Back to Isolating treated and control counties](#)

# Effects adjusting for treatment effect heterogeneity: Borusyak, Jaravel, and Spiess (2024)



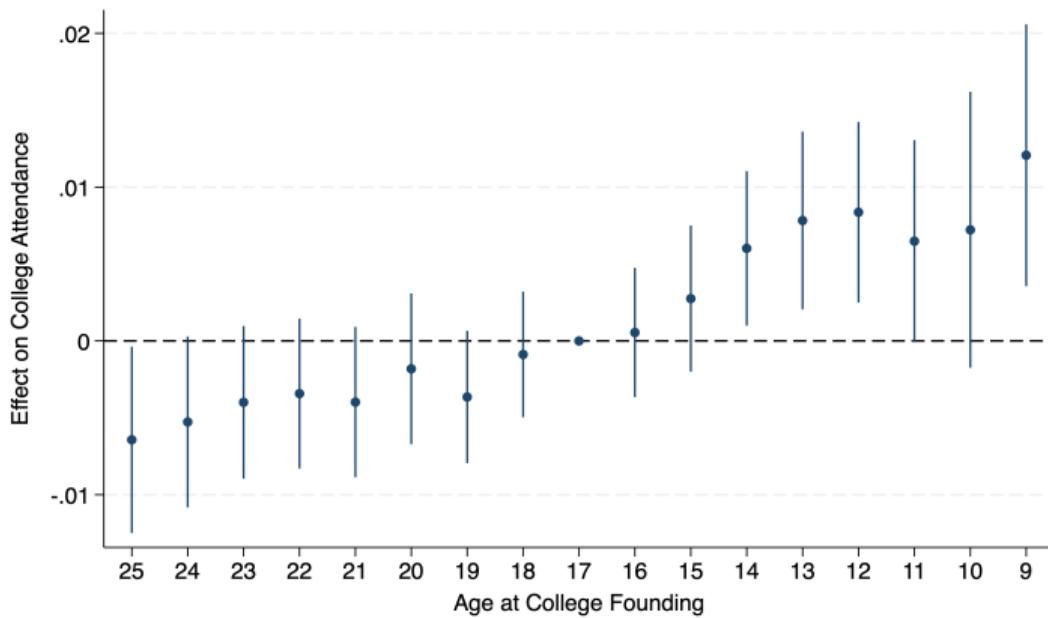
Back to robustness

## Effects controlling for county-by-cohort linear trends



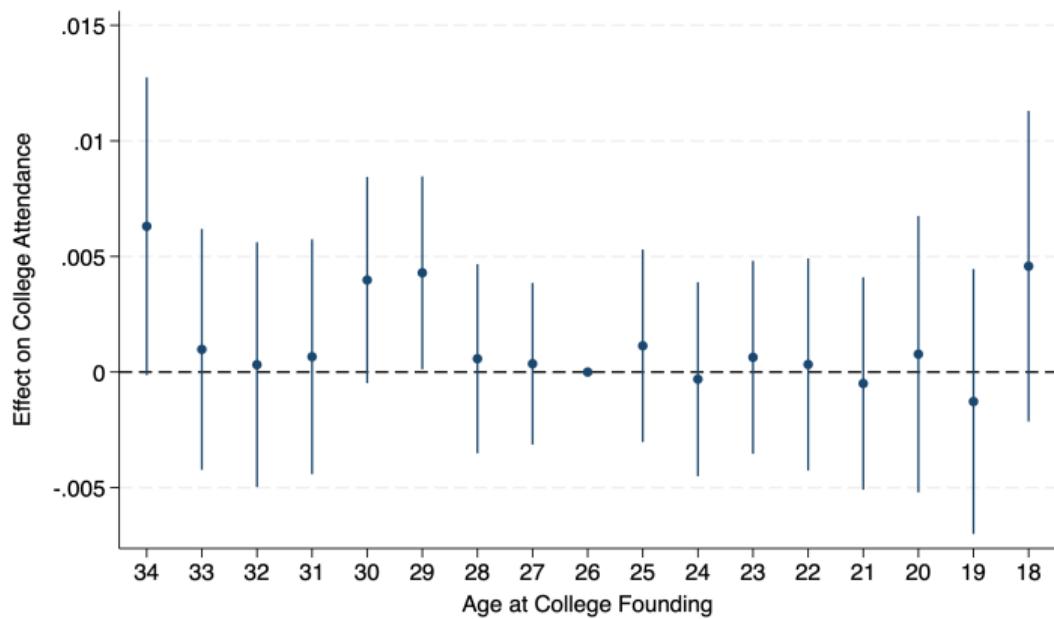
[Back to robustness](#)

## Effects restricting the sample to counties with more than 30 observations in each age bin



[Back to robustness](#)

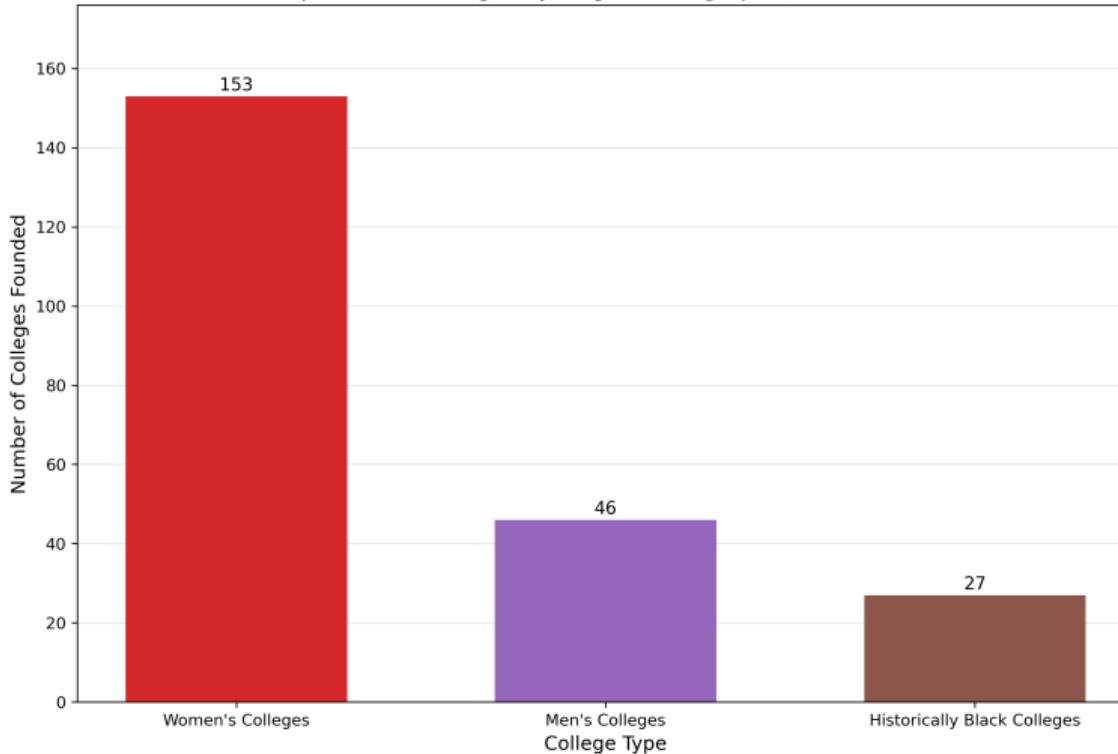
# Placebo test comparing attainment of individuals born before versus far before the college founding



Back to robustness

# Demographic-specific college foundings

Specialized Colleges by Target Demographic, 1900-1940



## Model-implied estimating equations

Hsiao (2024) model implied estimating equation,  $j$ : origin,  $k$  birth cohort,  $l$  destination

$$\log \bar{w}_{jkI} - \log \bar{e}_{jk} = \log \frac{\tilde{\epsilon}}{\epsilon} - \log a_I + \log \tau_{jk}^e + \log \tau_{jkl}^m \quad (2)$$

From which  $\tau_{jk}^e$ , the cost of education, can be estimated. Use the Cohort DD from earlier as a source of quasi-experimental variation in education costs:

$$\log \tau_{jk}^e = \alpha_j + \alpha_k + \beta S_j T_k + \epsilon_{jk} \quad (3)$$

where  $S_j = 1$  if  $j$  received a new college and  $S_k = 1$  if  $k$  was “exposed” to the college founding.

[back to next steps](#)