# College Access and Notability in the United States*

Samuel Stemper†

August 11, 2024

**Abstract**

I use comprehensive geolocated data on prominent, famous, and influential Americans across various fields born throughout the 19th and 20th centuries to characterize the distribution and determinants of notability rates across the U.S. I combine this data with information on college site selection experiments—historical instances in which multiple candidate locations were considered as the sites of new colleges—to estimate the effect of college access on notability. Comparing notability rates in counties selected for a college to those in runner-up counties indicates that college placement generates a large, immediate, and persistent increase in notability rates. Analysis of biographical texts suggests that 20 to 40 percent of these effects are driven by college attendance.

**JEL Codes**: I23; I26; J24; N31

**Keywords**: higher education; access to education; returns to education; economic history; fame

*First version: March 21, 2024. This version: August 11, 2024. I am grateful to Michael Andrews, Ezra Karger, and Étienne Wasmer for helpful comments and suggestions. All remaining errors are mine.

†University of Auckland. Email: sam.stemper@auckland.ac.nz

Across various fields, the likelihood of achieving prominence varies significantly based on place of birth. For example, over 16% of U.S.-born CEOs of S&P 500 firms were born in New York State, nearly 3 times its share of the U.S. population (Bernile et al., 2017).[1] 5 of every 1,000 children from San Jose, California go on to become inventors, 2.5 times the national average (Bell et al., 2019). In sports and entertainment, nearly 6 percent of National Hockey League players were born in Minnesota,[2] and over 10 percent of Academy Award-winning actors and directors were born in New York City.[3]

What drives variation in rates of notability across different areas? This paper explores the extent to which these differences are driven by access to higher education. I use geolocated historical data on notable individuals born in America between 1820 and 1980, derived from various editions of Wikipedia and Wikidata, allowing for a comprehensive analysis of the distribution and determinants of notability rates.

In the first part of the paper, I document three novel historical facts about the variation in notability rates among these American birth cohorts. First, notability rates vary substantially by place of birth. Notability rates for cohorts born in Suffolk County, Massachusetts are nearly three times the rates for cohorts born in Erie County, New York. Second, notability rates in the 19th and 20th centuries are highly correlated with 20th century socioeconomic characteristics. A county's level of education, skill content of the workforce, and access to higher education are all highly predictive of notability rates. Of these features, levels of higher education are most predictive; the share of adults with bachelor's degrees in 1940 explains over 12 percent of the across-county variation in subsequent notability rates. Third, spatial variation in notability rates is highly persistent over time; the rank-rank correlation between across-county notability rates among individuals born in the 1820s/1830s and individuals born in the 1960s/1870s is roughly 0.2.

Motivated by these facts, I investigate one potential driver of these differences in notability: college access. I use data on college site selection experiments—instances in which multiple can-

[1] Data in Bernile et al. (2017) includes of 1,508 U.S.- born CEOs of firms in the S&P 500 from 1992 to 2012, that have non-missing county of birth data (from various internet sources) and non-missing firm-level data in Execucomp. 251 of these 1,508 CEOs were born in New York. As of the 2020 U.S. Census, the population of New York was 20,202,320 and the U.S. population was 331,464,948. See U.S. Census Bureau, Quick Facts: New York and U.S. Census Bureau, Quick Facts: United States.

[2] "North Star Rising: Minnesota Rivals Quebec, Alberta for Second-Most NHL Skaters Behind Ontario," The Hockey News, December 4, 2023.

[3] Appendix Table B.1 provides the relevant calculations alongside links to IMDb.

didate locations were considered for a new college but only one of these locations was selected—from Andrews (2023). This setting lends itself to a difference-in-differences design, in which I compare counties that were selected for college sites to runner-up counties that were not selected.

Rates of notability are highly sensitive to these changes in college access; notability rates increased substantially among exposed birth cohorts: individuals born in winning counties who were younger than 18 at the time of college establishment. These increases are large—more than 50 percent of the pre-experiment mean—and set in quickly—among cohorts that were children at the time of college establishment. Effects are broad-based, but largest among individuals whose adult occupations were in culture (e.g. actors, singers, writers) or sports (e.g. football players, baseball players, basketball players). Using data from Wikipedia biographical texts, I estimate that 20 to 40 percent of the increase in notability rates is accounted for by individuals whose Wikipedia pages mention the name of the college. These facts suggest that college attendance plays a substantial role in explaining these results.

There are numerous accounts of individuals from modest origins who altered their life trajectories through the opportunities afforded by the establishment of these new colleges. I describe three such anecdotes below.

Robert M. La Follette was born on a farm in Dane County, Wisconsin in 1855. This timing was auspicious, as the University of Wisconsin would open in Dane County 11 years later. Ambitious but poor, he would later write in his autobiography,

> “My single term at the university [of Wisconsin] law school had been rendered possible only through the consideration of the faculty in [...] permitting me to enter without paying the usual matriculation fee. I had no money—but as fine an assortment of obligations and ambitions as any young man ever had."

At university, he was recognized for his oration skills; La Follette was selected to represent the University in the state collegiate oratorical contest. After winning both state and interstate competitions, La Follette was given a reception at the Wisconsin State House. La Follette noted that this experience ultimately helped launch his first political campaign: his successful run for District Attorney of Dane County in 1880. La Follette later won election to the U.S. House of Representatives and served as Governor of Wisconsin from 1901 to 1906 (La Follette, 1913).

George Washington Pierce was born to a ranching family in Travis County, Texas in 1872; 9 years later, the University of Texas was established in Travis County. As a child, Pierce's biography described him as "an avid reader, encouraged in this perhaps by the discovery that he could avoid some farm chores by pleading preoccupation with homework." Pierce attended the University of Texas and fellow students later "remember[ed] him as a brilliant and advanced student," whose "work in physics and mathematics was sufficiently outstanding to gain him exemption from many final examinations and to earn him the B.Sc. degree in three years." During his senior year, Pierce's physics professor hired him as a research assistant and their joint work led to Pierce's first publication in *Physical Review*. Pierce's study of physics would lead him to Harvard, where he received his PhD in 1900 and gained the rank of Professor in 1917. As a physicist and inventor, Pierce advanced telecommunications by improving long-distance call quality and pioneered the development of crystal oscillators essential for precise timekeeping in electronic devices (Saunders and Hunt, 1959).

Finally, brothers George and John McCutcheon were born in Tippecanoe County, Indiana, in 1866 and 1870, respectively. In 1869, Purdue University was established in the same county. Both brothers attended Purdue, where their father had assumed a managerial position in the commissariat in 1876. George left the university after one year to become a newspaper reporter.[4] John completed his degree, crediting Purdue with nurturing his interest in drawing and sketching. He later recalled, "My childhood interest in drawing and sketching had a chance to root at Purdue and my drawings improved... or at least did not become any worse ... and I found creative outlets for my cartoon ability." John contributed cartoons and illustrations to the first Purdue University yearbook in 1889, as well as to the local newspaper, *The Purdue*. During their time at Purdue, the brothers formed a lifelong friendship and working relationship with George Ade, a fellow student who would become a well-known playwright and humorist, earning the nickname "Aesop of Indiana."[5] Both brothers would ultimately become prominent for their contributions to journalism and illustration, highlighted by John's Pulitzer Prize award for his 1931 editorial cartoon "A Wise Economist Asks a Question."[6]

[4] George Barr McCutcheon, Britannica.

[5] "John T. McCutcheon (1870-1949)," Purdue University Retirees Association. "George Ade, Everybody's Friend," Indiana State Library.

[6] "A wise economist asks a question," U.S. Library of Congress.

What is important about these anecdotes is the pivotal role of these newly established colleges in identifying, cultivating, and publicizing skills that may have otherwise gone unnoticed. Absent the establishment of a local college, it might have been considerably more challenging for these individuals to access the education and opportunities that shaped their rise to prominence. All four individuals mentioned above—La Follette, Pierce, and the McCutcheon brothers—are represented in my data, and my research design allows me to construct their counterfactual outcomes, their expected probability of notability absent the establishment of a college, using data from cohorts in runner-up counties.

This work relates to the economics of talent broadly and to the economics of access to higher education specifically.

Regarding the prior, numerous studies have found significant variation in adult outcomes across places of birth (e.g. Chetty et al., 2014, Bosquet and Overman, 2019). A smaller set of literature examines the drivers of specific exceptional outcomes such as becoming an inventor (Bell et al., 2019), earning in the top 1% of income (Chetty et al., 2018), or working at a prestigious firm (Chetty et al., 2023).[7] One commonality across these studies is the role of education, both individually and in aggregate. For example, data used in Bell et al. (2019) indicates that inventors were 38 percentage points more likely to have attend college at age 20 than non-inventors.[8] Data from Chetty et al. (2018) indicates the county-level correlation between the share of children born whose earnings as adults place them in the top 1% and the proportion of residents with college degrees is 0.51.[9] Finally, the closest paper to mine is Doxey et al. (2022), which finds that historical expansions in high school access in the U.S. increased the likelihood that nearby residents would grow up to be federal judges, congresspeople, or notable scientists, businesspeople, or artists.

With respect to access to higher education, this work is closely related to numerous studies that examine the effect of college openings. The specific set of college site selection experiments I use in this study has been used to study local invention (Andrews, 2023) and local educational

[7] A separate literature, reviewed in Rosenthal and Strange (2004), considers the productivity effects of agglomeration. Some strands of this literature focus on productivity within specific, high-prominence positions, including the arts (Andersson et al., 2014; Borowiecki, 2013; Borowiecki and Dahl, 2021), entrepreneurship (Delgado et al., 2010; Glaeser et al., 2010), and invention (Moretti, 2021).)

[8] See Bell et al. (2019) Table 1. 86.0 percent of inventors attended college at age 20, versus 47.7 percent of non-inventors.

[9] These calculations do not appear directly in Chetty et al. (2018) but are based on public data available on the authors' accompanying website. Data on income is based on U.S. cohorts born between 1978 and 1983. The college share of the population is measured between 2012 and 2016.

attainment (Russell et al., 2022). Howard et al. (2022) use a similar site selection design to study the effect of local universities on economic resilience. A much larger set of papers studies the long-term effects of U.S. Land Grant Universities, most of which were established in 1862 (e.g. Moretti, 2004; Shapiro, 2006; Iranzo and Peri, 2009; Liu, 2015).[10]

The rest of the paper proceeds as follows. Sections 1 and 2 describe and summarize the data used in this paper. Section 3 describes the methodology used to identify causal effects of college access using college site selection experiments. Section 4 presents my results and section 5 concludes.

# 1 Data Sources

## 1.1 Notable People Data

My source of data on notable people comes from Laouenan et al. (2022), who compile a comprehensive compilation of notable figures based on data from multiple editions of Wikipedia and Wikidata.[11] Laouenan et al. (2022) cross-verify the database by including entries "only when the content could be cross-verified between different language editions and Wikidata." The authors verify algorithms with 5,000 manual checks; error rates are less than 1 percent. Laouenan et al. (2022) capture numerous biographical details for individuals represented in the data; these details include their place and date of birth and their primary occupation category. Occupation categories include Culture, Discovery/Science, Leadership, Sports/Games, or Missing/Other. Notable people data also contain detailed occupations, which include over 1,000 unique occupations.[12] Overall, the dataset contains information on approximately 2.29 million individuals born between 3500BC and 2018AD.

To this list, I make several sample restrictions for the purposes of this study. First, I limit my sample to individuals with U.S. citizenship with non-missing birthplaces. Second, to reduce the possibility that non-notable individuals are included in my sample, I exclude entries with fewer

[10]This list includes some studies that use the location of Land Grant Universities as an instrument for local levels of education.

[11]The authors have made this data publicly available for download from the project website.

[12]Appendix Figure B.1 displays the most common detailed occupations for each occupation category for notable individuals in my sample. These figures are calculated after implementing the sample restrictions described in this section.

than 1,000 Wikipedia visits between 2015 and 2018.[13] Next, I restrict my focus to individuals born in the 160 years between 1820 and 1980, inclusive.[14] Using this data, I map each individual to a modern U.S. county based on the latitude and longitude of their birthplace.[15] I exclude any individuals whose birthplaces do not map to a modern U.S. county and, due to difficulty assigning births in New York City, New York, I exclude the five counties in New York City.

After these sample restrictions, I am left with a set of 184,921 individuals who were born in the U.S. between 1820 and 1980. Appendix Figure B.3 shows the individuals whose Wikipedia pages were most read, separately for each 2-decade period in my sample. In early periods, these lists are comprised primarily of individuals who are notable for leadership in business or government: John D. Rockefeller, Theodore Roosevelt, and Henry Ford, for example. Over time, the representation of individuals associated with entertainment, art, or cultural contributions grows: Harry Houdini, Ernest Hemingway, and Walt Disney. The changing composition of notable individuals over time is shown more explicitly in Appendix Figure B.2.

To summarize the presence of notable people across geographies and time, I collapse this data to the county-by-birth year level, counting the number of notable individuals in each county and birth year overall and separately by occupation category.

### 1.2 County Birth Rates

I calculate annual county-level birth rates using decennial Census data from IPUMS USA and NHGIS (Ruggles et al., 2024; Manson et al., 2023). I provide details regarding this procedure in Appendix A. Broadly, when available, I use full-count IPUMS decennial Census data to calculate the number of children aged 0 in each county. When county-level data from IPUMS is unavailable, I use data from NHGIS. In some cases, NHGIS data reports population by age in the form of ranges (e.g. 0-4 years old), in which case I divide these population counts by the number of years they represent.

Over this historical period, the borders that delineate U.S. counties changed periodically. To account for these changes, I convert historical counties to their 1990 county equivalents using the

[13] In robustness tests, I show that I obtain nearly identical results using cutoffs of 0 or 100 Wikipedia visits.

[14] I end this data in 1980 due to the very small number of (currently) notable people born after this year. Appendix Figure B.2 shows the number of notable individuals by year of birth and primary occupation, demonstrating a large decline in total notable individuals after 1990 and a large decline in total non-sports-related notable individuals after 1980.

[15] Throughout this paper, I use U.S. county borders as of 1990.

method described in and data provided by Eckert et al. (2020). Eckert et al. (2020) calculate the share of the area of each "reporting" unit (e.g., 1900 county) that overlaps with a "reference" unit (e.g., 1990 county), which allows me to re-aggregate data from historical counties to 1990 counties for each decennial observation.

To estimate annual birthrates, I linearly interpolate between decennial birthrates separately for each county. I link this annual birth rate data to county-by-birth year counts of notable people and calculate notability rates: the number of (subsequently) notable people as a share of births in a county. (For ease of interpretation, I report notability rates per 10,000 births throughout this paper.) Notability rates exhibit relatively less long-run over-time variation compared to counts of notable individuals. To illustrate this point, Appendix Figure B.4 shows annual notability rates by year of birth and primary occupation. Compared to the over-time patterns in total notable individuals (shown in Appendix Figure B.2), the long-run variation in rates of notability over time is relatively smaller.

### 1.3 College Site Selection Experiments Data

I use data on college site selection experiments from Andrews (2023). These experiments arose during the expansion of American higher education over the 19th and 20th centuries. To collect this data, Andrews (2023) consults extensive historical records, identifying instances in which multiple candidate locations were considered for a new college but only one of these candidate locations was selected. The usefulness of this setting is that finalist sites that did not win ("losing sites" or "runner-up sites") provide a reasonable counterfactual for the sites that did win ("winning sites") in a difference-in-differences design.

The methods used to choose between candidate sites differed substantially across institutions. For instance, the locations of the University of North Dakota and North Dakota State University were determined through a lottery draw, while Macon was selected as Georgia Tech's location after a state committee vote, beating Atlanta by a single vote. From a list of 181 colleges for which candidate locations were found, Andrews (2023) makes two additional restrictions. First, Andrews (2023) restricts his attention to "high-quality" experiments: experiments where, "conditional on being a finalist, the site selection decision is as good as random." Second, he excludes colleges established before 1836. These restrictions produce a set of 63 site selection experiments.

I further restrict the set of experiments to those for which I have notability rates for the county of the winning site and at least one county of a losing site for cohorts that were at least at 30 at the time of the college establishment. Including sufficiently old cohorts (i.e. individuals who were adults at the time of college establishment) ensures that my data includes cohorts that were older than typical college-age students at the time the college was established. Practically, because my data starts in 1820, this restriction excludes any colleges established before 1850.

Altogether, these restrictions generate a list of 46 college experiments. Appendix Figure B.5 shows the locations of the winning and losing counties. The experiments included in my study, alongside the winning and losing counties, are listed in Appendix Table B.2.

# 2 Summarizing Notability Rates

## 2.1 Geographic Distribution of Notability

Figure 1 displays the geographic distribution of notability rates in my sample. Labeled counties in Figure 1 correspond to the 10 counties with the highest number of births during my sample. Overall, the average county-year observation produced roughly 4.3 notable people per 10,000 births.[16] However, there is substantial heterogeneity in notability rates across counties. These rates tend to be highest in urban areas and in the northeastern states, as well as in Florida and California generally. Among the 10 counties with the most births over this period, Suffolk County, Massachusetts has the highest notability rate: 16 of every 10,000 children born in this county would ultimately become notable, nearly three times the rate in Erie County, New York, and twice that of neighboring Middlesex County, Massachusetts.

## 2.2 Correlates of Notability

To assess the degree to which notability covaries with county characteristics, I first collapse my county-by-birth year panel data to the county level, calculating each county's average notability rate, weighted by the number of births in each year. I denote this notability rate $\overline{Notability}_i$, and estimate regressions of the form below.

$$\overline{Notability}_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \tag{1}$$

[16] This and other summary statistics are shown in Appendix Table B.3.

where $X_i$ represents some characteristic of county $i$. These regressions are meant to be descriptive, so these coefficients should be interpreted as associative rather than causal.

I estimate Equation 1 using four county characteristics as of 1940: the share of adults over age 25 with a high school education, the share of adults over age 25 with a college education, the share of employment in high-skill work,[17] and the ratio of college students aged 19 to 21 in each county in 1940 to the size of the county's birth cohorts 19 to 21 years prior to 1940. This latter measure is meant to measure access to college; counties with large college enrollments will tend to have higher rates of enrollment relative to cohort sizes. I refer to this measure as the "College Access Index."

I use data from 1940 because this was the first U.S. Census that asked respondents for their level of education. As with county birth rate data, I convert 1940 counties to 1990 counties using the method and data in Eckert et al. (2020). For ease of interpretation, I convert each independent variable into percentiles.

My results are displayed in Table 1. I focus on notability rates between 1941 to 1980, as they represent notability rates for cohorts born after the 1940 Census data was collected. I restrict my sample to counties that have nonmissing data throughout the range.

Columns 1 to 4 display bivariate relationships between county characteristics in 1940 and subsequent notability rates. Across the four variables I consider, all have strong, positive associations with notability rates; counties with more highly-educated adults, with more high-skill employment, and higher levels of college access tend to have higher rates of notability. The largest of these bivariate relationships is with respect to the college-educated share; a 100 percentile increase in the 1940 share of adults with a college education is associated with an increase in notability rates over the next 40 years by 4.7 per 10,000 births, an increase of roughly 1.2 standard deviations. College access also plays a large role: moving from the bottom to the top of the distribution of college access is associated with an increase in subsequent notability rates of 4.0 per 10,000 births.

Column 5 displays the results of a "horse race" regression, which includes all four characteristics simultaneously. Here again, the college-educated share and college access index exhibit large, positive effects on future notability rates.

[17] I categorize the following occupation groups as high-skill: "professional," "semi-professional," and "proprietors, managers, and officials (except farm)."

Overall, the findings suggest that one possible environmental factor influencing notability is the availability and take-up of higher education. However, I note that these connections are merely observational. Later, I more precisely evaluate the causal impact of access to higher education using college site selection experiments.

### 2.3 Persistence of Notability

I next assess the persistence of notability rates over time. Specifically, I calculate average notability rates in each 2-decade birth cohort in my data, $\overline{Notability}_{it}$[18] and estimate rank-rank correlations between all combinations of birth cohorts.

I highlight two general patterns in my results, which are shown in Figure 2. First, rank-rank correlations are largest for estimates with closer year ranges; for adjacent cohorts, correlation coefficients are between 0.3 and 0.5. Second, even among the most distant relationship I estimate—the relationship between notability rates in the 1820s to notability rates in 1960s—the correlation coefficient remains quite high: 0.21.

While these persistence patterns are striking, they are not particularity informative about the degree to which changes in environmental factors—such as college access—lead to higher or lower rates of notability over these periods. In the section below, I describe my methodology to assess the role of a specific change within many U.S. counties: the opening of a college.

## 3 Methodology

To assess the effect of college access directly, I estimate the short and long-run effects of college openings by comparing outcomes in counties where a college was established to outcomes in runner-up counties. I estimate effects using a stacked difference-in-differences approach.[19]

To construct my stacked data, I collect, for each experiment, data for both winning and losing counties and stack these event-specific datasets. With this stacked data, I estimate the regression below.

$$Notability_{iet} = \delta College_{ie} \times PostCollege_{iet} + \gamma_{ie} + \lambda_{et} + \varepsilon_{iet}, \quad (2)$$

where $i$ indexes counties, $e$ indexes experiments, and $t$ indexes birth years. $College_{ie}$ is a binary

[18]For each period, I calculate each county's average notability rate weighted by the number of births in each year.

[19]This approach was introduced in Cengiz et al. (2019) has since become a popular approach to avoiding potential bias associated with two-way fixed effects estimators (Goodman-Bacon, 2021).

variable equal to one for the winning county, $PostCollege_{iet}$ is a binary variable equal to one for cohorts that were 18 years old or younger at the time of college establishment.[20] $\gamma_{ie}$ is a county-by-experiment fixed effect.[21] $\lambda_{et}$ is a experiment-by-birth year fixed effect. For all regressions, I cluster standard errors at the county level.

In my main specification, I include experiment-by-county and experiment-by-birth year fixed effects, $\gamma_{ie}$ and $\lambda_{et}$. Doing so allows over-time trends in notability across birth years to vary across experiments. However, I note that Andrews (2023) does not include these terms. In Appendix B, I show that using the specification from Andrews (2023) produces qualitatively similar results.

In addition to my static stacked difference-in-difference results, I also estimate event studies that estimate the dynamic effects of college selection, replacing $PostCollege_{iet}$ in Equation 2 with indicators for birth years relative to their age at college establishment. For precision, I group birth years into groups of five (e.g. 0 to 4, 5 to 9, etc.). I exclude fixed effects for cohorts that were age 33 to 24, so event study coefficients represent differences between winning and losing counties, relative to the differences in those cohorts.

In this setting, identification of causal effects relies on a parallel trends assumption: absent the opening of a college, notability patterns in winning and losing counties would have moved in parallel. While I cannot evaluate this assumption directly, I test for differences in trends and levels of notability prior to college establishment. Specifically, I restrict my data to cohorts that were older than 18 at the time of college establishment, and estimate the two regressions below.

$$Notability_{iet} = \theta College_{ie} \times Birthyear_t + \gamma_{ie} + \lambda_{et} + \varepsilon_{iet} \tag{3}$$

$$Notability_{iet} = \rho College_{ie} + \lambda_{et} + \varepsilon_{iet}, \tag{4}$$

where $Birthyear_t$ denotes the birth year for cohort $t$, $\gamma_{ie}$ is a county-by-experiment fixed effect, $\lambda_{et}$ is a experiment-by-birth year fixed effect. Intuitively, Equation 3 tests whether, within a cohort and set of candidate counties, and conditional on county-by-experiment fixed-effects, pre-college trends differed between the winning and losing counties. Equation 4 tests whether, within a cohort

[20] I choose 18 because cohorts older than age 18 are less likely to attend the opening college. Later, I describe event study analyses that flexibly estimate dynamic effects across cohorts.

[21] A small number of counties appear in more than one experiment. For example, Lenoir County in North Carolina was a runner-up for both North Carolina State University in 1886 and East Carolina University in 1907.

and set of candidate counties, levels of pre-college notability differed between the winning and losing counties.

Appendix Table B.4 displays my estimates of Equations 3 and 4. In Column 1, I test for differences in trends. Overall, I find little evidence that trends between treated and control counties are different prior to college opening. In Column 2, I test for differences in levels. Here, I similarly find little evidence that levels of notability in the years prior to college opening are were different. These facts bolster a causal interpretation of my difference-in-difference results.

## 4 Results

Table 2 shows my main difference in differences estimates. In Panel A, I display results with respect to county notability rates per 10,000 births. Column 1 shows the effects on notability rates across all occupations. The results indicate that, following college establishment, notability rates in winning counties increased by 2.4 persons per 10,000 births relative to runner-up counties. This is a large effect, equal to more than 50% of the pre-experiment mean in winning or losing counties. Columns 2 through 5 show effects on occupation-specific notability rates. Point estimates are positive across all categories but statistically significant only for culture and sports occupations.

In Panel B, I show the effects on the number of notable persons per birth year. Overall, these results are qualitatively similar in direction and magnitude, but are less precise; 95% confidence intervals include 0 for all estimates.[22]

Figure 3 examines the effect over time and provides evidence that these changes appear among the earliest college-eligible cohorts. Panel A displays how the notability rate per 10,000 births evolved across cohorts, separately for winning and losing counties. For cohorts that turned 18 prior to college establishment, losing counties exhibit slightly higher notability rates. However, this pattern flips for cohorts that were college-eligible; in winning counties, cohorts that were adolescents or younger at the time of college establishment exhibit large and persistent increases in notability rates. Panel B displays the corresponding event study results, which estimate differences in notability rates between winning and losing counties across birth cohorts. The largest effects appear among individuals who were children or adolescents (e.g. -1 to 18 years old) at the

[22] In Appendix Table B.6 (described in more detail below), I show that I obtain much more precise difference-in-differences estimates for the number of notable persons per birth year when I include controls for cohort size.

time of college establishment.[23]

In Panel C, I show that the same over-time patterns appear for data on the raw number of notable persons per birth year. Among cohorts that were above the age of 18 at the time of college establishment, differences between winning and losing counties were small; in both groups, the average number of notable people per birth cohort was roughly 0.2. Following college establishment, these trends diverge: in winning counties, the number of notable people per birth cohort increases much more rapidly than in losing counties. Panel D shows the corresponding event study results.

To what degree are these effects driven by college access (either directly or among peers) versus college-induced changes in population size or composition? Overall, the over-time patterns are more consistent with the college access channel. That the positive effects of college establishment are sudden—showing up among individuals who were children or adolescents at the time of college establishment—and initial magnitudes are consistent with difference-in-differences estimates based on long-run data suggest that the effects of colleges on college access may be more relevant in this context than the long-run effects of county population or composition.

As a more direct test of the role of college enrollment in driving these effects, I capture text data from the individual Wikipedia pages in this sample. With this text, I identify whether the name of the college is mentioned in the text, and calculate what I refer to as "college-specific" notability rates. For a given opening-specific county in my data, college-specific notability measures the frequency of notable individuals whose Wikipedia pages mention that college's name.[24] Taking the opening of Pennsylvania State University as an example, college-specific notability for the winning and runner-up counties (Centre County and Blair County, respectively) measures the frequency of notable individuals born in these counties who (a) had Wikipedia pages and (b) these pages included the text "Pennsylvania State University."

As a rough approximation of the role of college attendance in driving my main results, I reproduce my difference-in-differences and event study estimates using measures of college-specific notability. These results are shown in Table 3 and Figure 4. Column 1 in Panel A of Table 3 indicates that, following college establishment, college-specific notability rates in winning counties

[23] This cohort corresponds to the "0 to 19" group in Panel B of Figure 3.

[24] I include additional details in Appendix A. Because Princeton University was called The College of New Jersey before 1892, I exclude The College of New Jersey from this analysis.

increased by 0.8 persons per 10,000 births relative to runner-up counties. This is roughly one-third of my main effect in Table 2. The equivalent estimate in Panel B suggests a slightly smaller ratio, suggesting roughly one-fifth of the increase in overall notability is accounted for by increases in college-specific notability. Importantly, college-specific notability follows similar over-time patterns as overall notability. In Figure 4, gaps in college-specific notability appear among individuals who were children or adolescents at the time of college establishment.

I subject these results to several robustness checks.

First, I conduct a permutation test to verify that my results fall outside the bounds of estimates driven by typical over-time variation in notability. Specifically, I reproduce my difference-in-differences estimates 1,000 times after (a) removing the treated county and (b) randomly selecting a control county for "placebo" treatment. By its nature, this procedure excludes experiments in which there was only one runner-up county; there are 18 such experiments in my data, leaving a set of 28 experiments included in each permutation test.

Figure B.6 shows the distribution of t-statistics among my 1,000 permutation estimates. Panel A shows the distribution for difference-in-difference estimates for notability rates. This distribution is centered around 0 and exhibits very limited mass beyond -2 and 2. I compare this distribution to two estimates: my main difference-in-difference estimate (shown in Table 2) and an equivalent estimate using only the 28 experiments included in each permutation test. Both estimates fall beyond the 99th percentile of permutation estimates. Panel B shows the equivalent distribution for estimates with respect to the number of notable people per year. These estimates are less statistically abnormal, but actual estimates both fall beyond the 90th percentile of permutation estimates.

Second, I show that my results are not driven by any one college site selection experiment. In Appendix Figure B.7, I show that my main difference-in-differences estimate is stable to the exclusion of any one site selection event.

I also show that the choice to exclude individuals with fewer than 1,000 Wikipedia visits between 2015 and 2018 has no effect on my results. In Tables B.7 and B.8, I reproduce Table 2 using a cutoff of 0 visits and 100 visits, respectively, and obtain nearly identical results.

Next, I show that I obtain nearly identical results regardless of the set of relative age cohorts I use to estimate my difference-in-difference estimates. Appendix Figure B.8 displays my difference-in-difference coefficient under 192 combinations of restrictions on cohorts—defined based on their age in the year of college establishment—included in my sample. The coefficient estimates all fall between 1.5 and 2.5, and are all statistically significant at 10% and all but 3 (of 192) estimates significant at 5%.

In addition, I note that the set of fixed effects used in Andrews (2023) differs slightly from those used in my main estimates. Specifically, I fully interact birth year and county fixed effects with experiment fixed effects. In Appendix Table B.5, I show that I obtain nearly identical coefficients when I instead use the set of fixed effects used in Andrews (2023).

Finally, Andrews (2023) points out that the establishment of a college leads to population growth in winning counties. In Appendix Table B.6, I test whether my effects are robust to controls for county population, which I measure using the number of births per birth cohort in each county. The regressions in Appendix Table B.6 include linear and squared terms for the number of births. My results in Panel A, which estimate effects on notability rates, are nearly identical to my main results in Table 2. In Panel B, I estimate effects on the number of notable persons per birth year. Here, the addition of controls for cohort size reduces the magnitude of my estimates by roughly 60 percent. This is consistent with Andrews (2023): after a college is established, large increases in population may increase the number of notable persons per birth year simply due to population size. Including these controls will limit the role of this *scale* effect. That the statistical significance of these effects increase when cohort size controls are included suggests that population growth is not the exclusive driver of the treatment effects I estimate.[25]

## 5 Discussion

In this paper, I study variation in notability rates in the United States. Using geolocated data on notable individuals, I estimate the historical likelihood of achieving notability among cohorts over time and across counties. I combine this data with information on college site selection experiments and document increases in notability rates following the establishment of a local college. These effects are large—50 percent of the pre-college mean—and appear among cohorts that were adolescents at the time of college establishment.

[25]Of course, these controls do not account for changes in the *composition* of county populations. For example, the establishment of a college may have disproportionately increased county population among subgroups that are more likely to become notable.

I note several limitations and opportunities for future research.

First, while I framed my results in the context of college access, I note that college establishment has a myriad of other effects on an area: the establishment of a college likely drives population growth via migration, changes in the demographic composition of the population, and increased economic activity. One limitation of this work is that my setting does not permit a separate examination of the individual contributions of these potential factors beyond the information contained in Wikipedia biographies.

Additionally, on a more granular level, the specific roles of skill development versus networking in mediating these observed effects remain ambiguous. Recent evidence from Michelman et al. (2022) suggests that exposure to high-status peers played an important role in achieving elite status among Harvard College students in early 20th century America. These mechanisms are likely relevant in the context of 19th and 20th century U.S. college openings as well.

Finally, future work may seek to identify other drivers of differences in notability across places and over time. While labor economists often estimate causal effects on reasonably “ordinary" outcomes such as college attendance, employment, or occupation choice, notability data represents a rare window into *exceptional* outcomes. These outcomes are rare but important, and offer insights into the dynamics of achievement and recognition.

## References


**Andersson, Åke E, David Emanuel Andersson, Zara Daghbashyan, and Björn Hårsman**, "Location and spatial clustering of artists," *Regional Science and Urban Economics*, 2014, *47*, 128–137.

**Andrews, Michael J**, "How do institutions of higher education affect local invention? Evidence from the establishment of US colleges," *American Economic Journal: Economic Policy*, 2023, *15* (2), 1–41.

**Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen**, "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 2019, *134* (2), 647–713.

**Bernile, Gennaro, Vineet Bhagwat, and P Raghavendra Rau**, "What doesn't kill you will only make you more risk-loving: Early-life disasters and CEO behavior," *The Journal of Finance*, 2017, *72* (1), 167–206.

**Borowiecki, Karol Jan**, "Geographic clustering and productivity: An instrumental variable approach for classical composers," *Journal of Urban Economics*, 2013, *73* (1), 94–110.

_ **and Christian Møller Dahl**, "What makes an artist? The evolution and clustering of creative activity in the US since 1850," *Regional Science and Urban Economics*, 2021, *86*, 103614.

**Bosquet, Clément and Henry G Overman**, "Why does birthplace matter so much?," *Journal of Urban Economics*, 2019, *110*, 26–34.

**Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer**, "The effect of minimum wages on low-wage jobs," *The Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

**Chetty, Raj, David J Deming, and John N Friedman**, "Diversifying society's leaders? The causal effects of admission to highly selective private colleges," Technical Report, National Bureau of Economic Research 2023.

_ **, John N Friedman, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter**, "The opportunity atlas: Mapping the childhood roots of social mobility," Technical Report, National Bureau of Economic Research 2018.

_ **, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez**, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

**Delgado, Mercedes, Michael E Porter, and Scott Stern**, "Clusters and entrepreneurship," *Journal of economic geography*, 2010, *10* (4), 495–518.

**Doxey, Alison, Ezra Karger, and Peter Nencka**, "The Democratization of Opportunity: The Effects of the US High School Movement," 2022.

**Eckert, Fabian, Andrés Gvirtz, Jack Liang, and Michael Peters**, "A method to construct geographical crosswalks with an application to us counties since 1790," Technical Report, National Bureau of Economic Research 2020.

**Follette, Robert M La**, *A Personal Narrative of Political Experiences*, Robert M. La Follette, 1913.

**Glaeser, Edward L, William R Kerr, and Giacomo AM Ponzetto**, "Clusters of entrepreneurship," *Journal of urban economics*, 2010, *67* (1), 150–168.

**Goodman-Bacon, Andrew**, "Difference-in-differences with variation in treatment timing," *Journal of econometrics*, 2021, *225* (2), 254–277.

**Howard, Greg, Russell Weinstein, and Yuhao Yang**, "Do universities improve local economic resilience?," *Review of Economics and Statistics*, 2022, pp. 1–51.

**Iranzo, Susana and Giovanni Peri**, "Schooling externalities, technology, and productivity: Theory and evidence from US states," *The Review of Economics and Statistics*, 2009, *91* (2), 420–431.

**Laouenan, Morgane, Palaash Bhargava, Jean-Benoît Eyméoud, Olivier Gergaud, Guillaume Plique, and Etienne Wasmer**, "A cross-verified database of notable people, 3500BC-2018AD," *Scientific Data*, 2022, *9* (1), 290.

**Liu, Shimeng**, "Spillovers from universities: Evidence from the land-grant program," *Journal of Urban Economics*, 2015, *87*, 25–41.

**Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles**, "IPUMS National Historical Geographic Information System: Version 18.0 [dataset]," https://doi.org/10.18128/D050.V18.0 2023.

**Michelman, Valerie, Joseph Price, and Seth D Zimmerman**, "Old boys' clubs and upward mobility among the educational elite," *The Quarterly Journal of Economics*, 2022, *137* (2), 845–909.

**Moretti, Enrico**, "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data," *Journal of econometrics*, 2004, *121* (1-2), 175–212.

__ , "The effect of high-tech clusters on the productivity of top inventors," *American Economic Review*, 2021, *111* (10), 3328–3375.

**Rosenthal, Stuart S and William C Strange**, "Evidence on the nature and sources of agglomeration economies," in "Handbook of regional and urban economics," Vol. 4, Elsevier, 2004, pp. 2119–2171.

**Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler**, "IPUMS USA: Version 15.0 [dataset]," https://doi.org/10.18128/D010.V15.0 2024.

**Russell, Lauren C, Lei Yu, and Michael J Andrews**, "Higher education and local educational attainment: evidence from the establishment of US colleges," *Review of Economics and Statistics*, 2022, pp. 1–32.

**Saunders, Frederick A and Frederick V Hunt**, "George Washington Pierce," *Biographical Memoirs. National Academy of Sciences*, 1959, *33*, 351–80.

**Shapiro, Jesse M**, "Smart cities: quality of life, productivity, and the growth effects of human capital," *The review of economics and statistics*, 2006, *88* (2), 324–335.

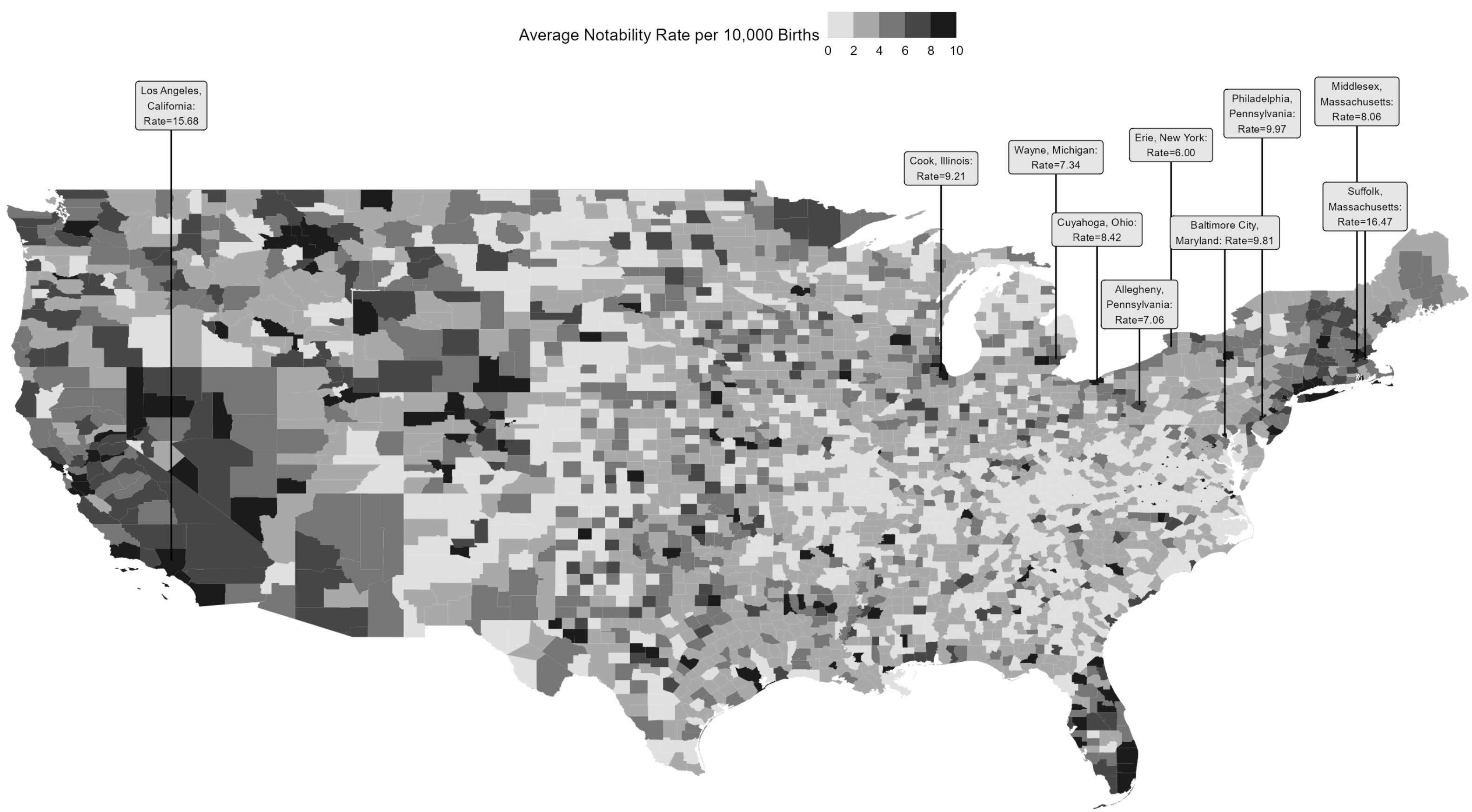


**Figure 1:** Average Notability Rate in US Counties: 1820 to 1980

**Note:** Figure displays the geographic distribution of notability rates in US counties from 1820 to 1980. Notability rates over this period are calculated as the average notability rate in each year, weighted by the number of births in that year. Labeled counties are the 20 counties with the highest number of births. Due to difficulty assigning births in New York City, New York, the five counties in New York City are excluded.

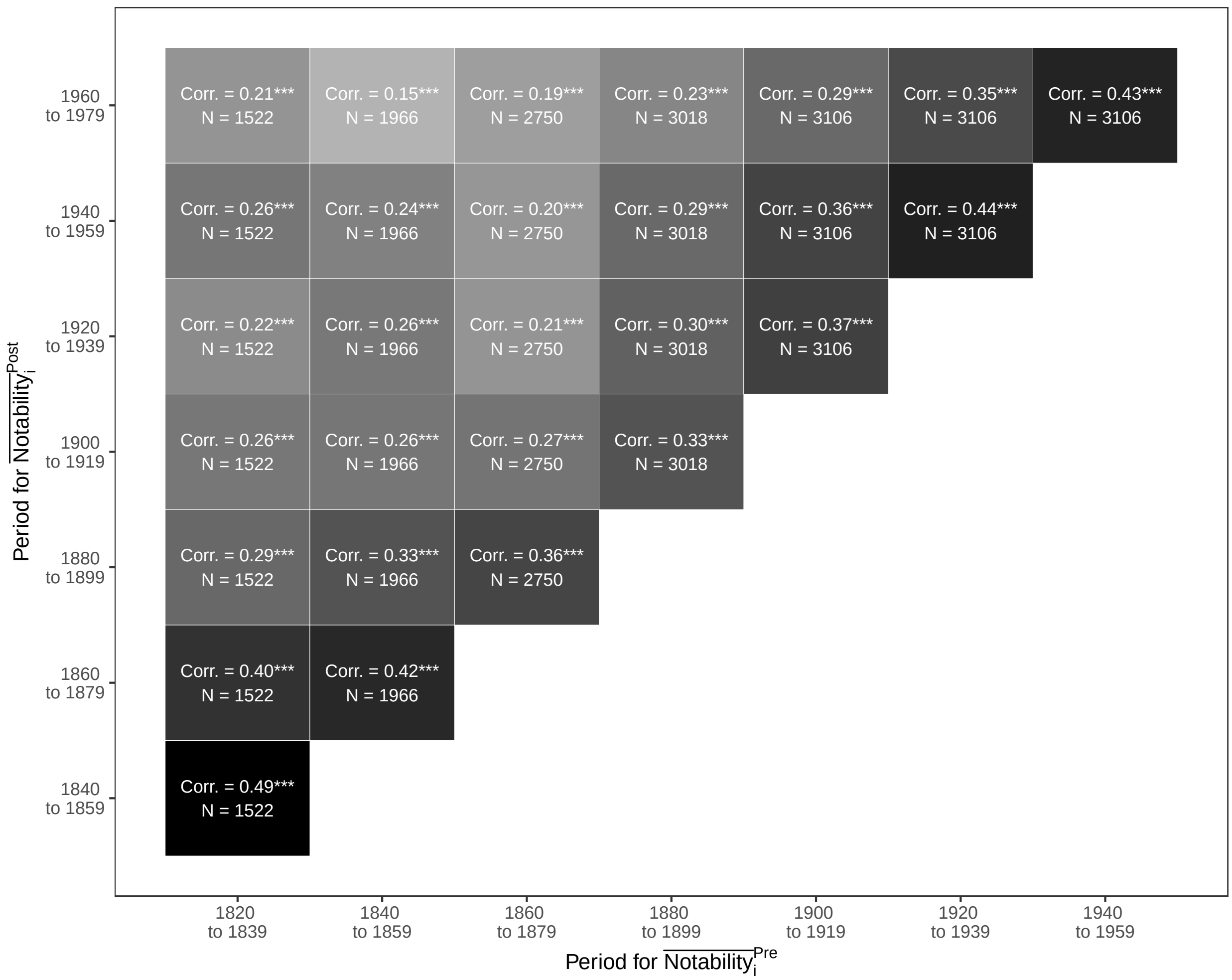


**Figure 2:** Rank-Rank Correlations of Notability Rates Across Historical Birth Cohorts

**Note:** Figure displays rank-rank correlations in notability rates among 2-decade birth cohorts between 1820 to 1980. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

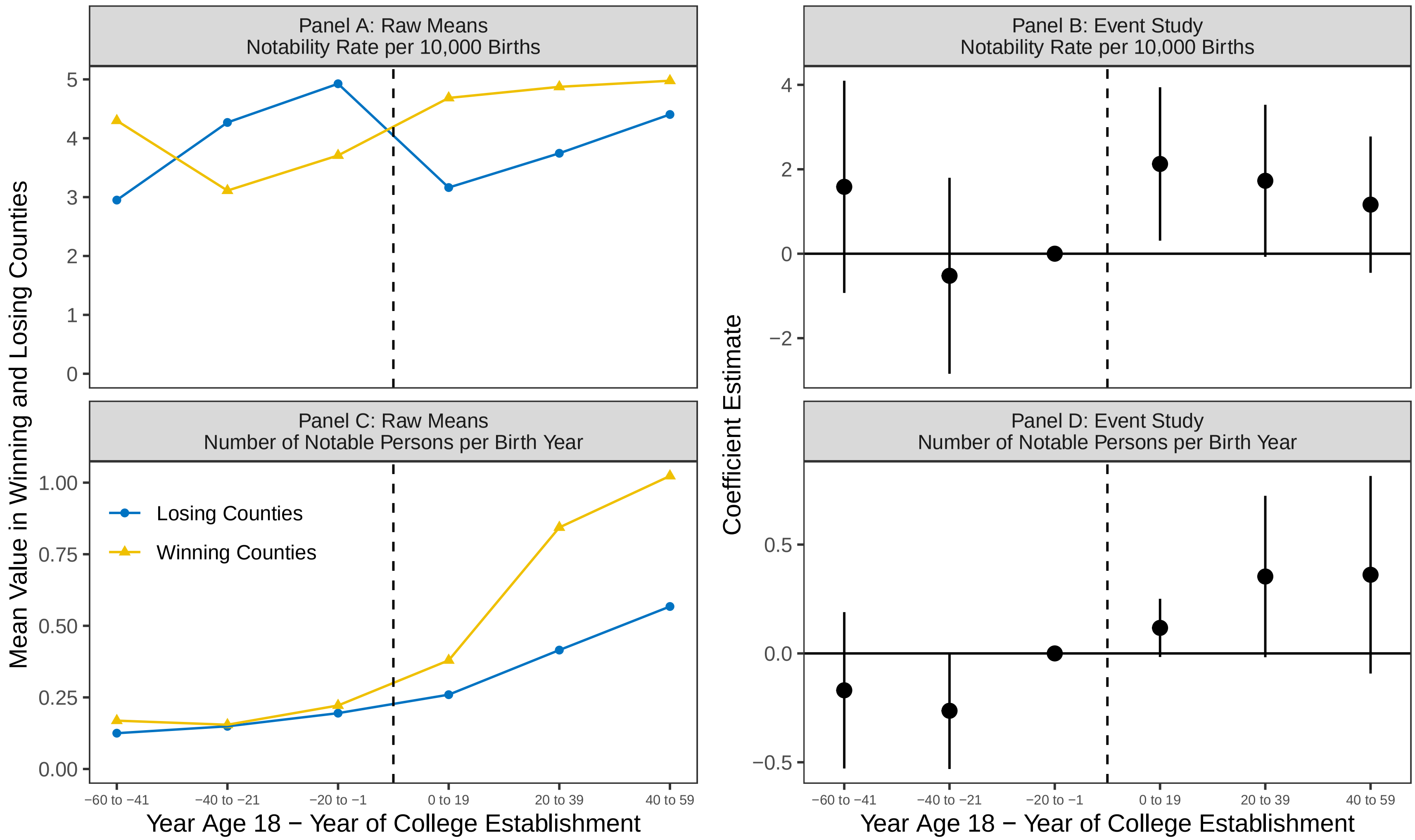


**Figure 3:** Dynamic Impact of College Establishment on Notability

**Note:** Figure displays trends in notability (in Panels A and C) and event study results (in Panels B and D) showing the dynamic effect of college establishment on notability. Panel A displays averages of notability rates per 10,000 births among winning and losing counties in 5-year birth cohorts, relative to the year of college establishment. Panel B displays event study results showing the dynamic effect of college establishment on notability rates. Panel C displays averages of the number of notable persons per birth year among winning and losing counties in 20-year birth cohorts, relative to the year of college establishment. Panel D displays event study results showing the dynamic effect of college establishment on the number of notable persons per birth year. Error bars in Panels B and D are 95% confidence intervals. All panels trim the data to include cohorts that were 18 years old 60 years prior to and 60 years after college establishment; Appendix Figures B.9 and B.10 display results for all time periods.

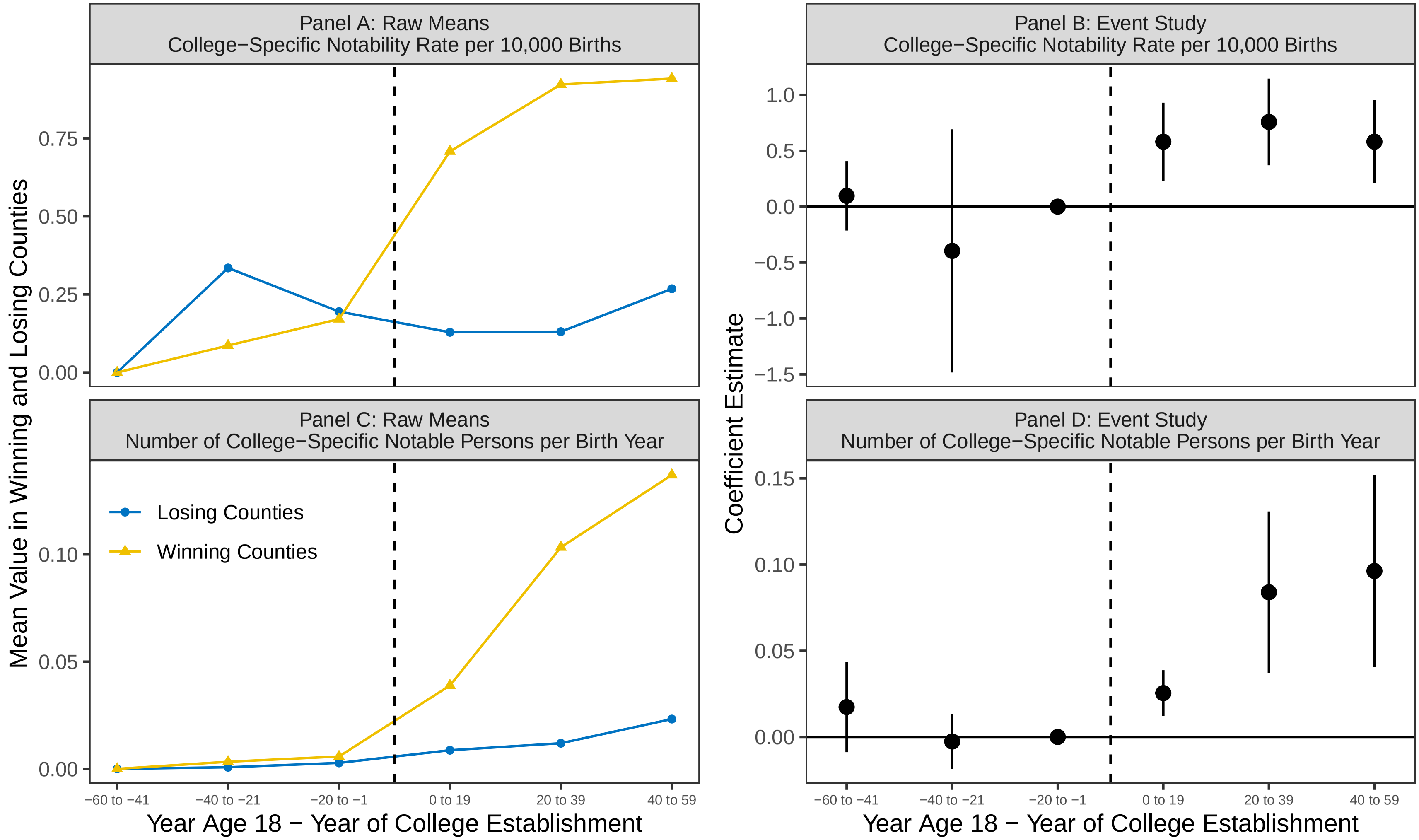


**Figure 4:** Dynamic Impact of College Establishment on College-Specific Notability

**Note:** Figure displays trends in notability (in Panels A and C) and event study results (in Panels B and D) showing the dynamic effect of college establishment on notability. For a given opening-specific county in my data, college-specific notability measures the frequency of notable individuals whose Wikipedia pages mention the name of the college. Because Princeton University was called The College of New Jersey prior to 1892, I exclude The College of New Jersey from this analysis. Panel A displays averages of college-specific notability rates per 10,000 births among winning and losing counties in 20-year birth cohorts, relative to the year of college establishment. Panel B displays event study results showing the dynamic effect of college establishment on college-specific notability rates. Panel C displays averages of the number of college-specific notable persons per birth year among winning and losing counties in 5-year birth cohorts, relative to the year of college establishment. Panel D displays event study results showing the dynamic effect of college establishment on the number of college-specific notable persons per birth year. Error bars in Panels B and D are 95% confidence intervals. All panels trim the data to include cohorts that were 18 years old 60 years prior to and 60 years after college establishment.

**Table 1:** Correlates of Notability Rates

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dep. Var.: Notability Rate 1941 to 1980 | | | | | |
| Pct. Rank: HS Share (1940) | 3.490*** | | | | -2.323*** |
| | (0.231) | | | | (0.513) |
| Pct. Rank: BA Share (1940) | | 4.659*** | | | 5.273*** |
| | | (0.224) | | | (0.447) |
| Pct. Rank: High-Skill Empl. Share (1940) | | | 3.018*** | | 0.076 |
| | | | (0.233) | | (0.374) |
| Pct. Rank: College Access Index (1940) | | | | 4.014*** | 1.676*** |
| | | | | (0.228) | (0.379) |
| Dep. Var. Mean | 3.38 | 3.38 | 3.38 | 3.38 | 3.38 |
| Dep. Var. SD | 3.85 | 3.85 | 3.85 | 3.85 | 3.85 |
| Raw Indep. Var. Mean | 0.20 | 0.03 | 0.13 | 0.05 | - |
| Raw Indep. Var. SD | 0.08 | 0.02 | 0.04 | 0.07 | - |
| Num.Obs. | 3105 | 3105 | 3105 | 3105 | 3105 |
| R2 | 0.069 | 0.122 | 0.051 | 0.091 | 0.132 |

**Note:** Table displays OLS estimates of Equation 1, which summarize the relationship between county notability rates and county characteristics in 1940. All variables are standardized such that they have mean 0 and standard deviation 1. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table 2:** Difference-in-Differences Estimates

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Category | All | Culture | Disc/ Science | Leader- ship | Sports |
| Panel A: Dep. Var. is Notability Rate per 10,000 Births | | | | | |
| Post × College | 2.387** | 0.465* | 0.150 | 1.221+ | 0.475*** |
| | (0.823) | (0.212) | (0.148) | (0.672) | (0.125) |
| Runner-Up Mean: Pre-Experiment | 4.54 | 0.50 | 0.46 | 3.17 | 0.28 |
| Runner-Up Mean: Post-Experiment | 4.09 | 1.15 | 0.37 | 1.12 | 1.39 |
| Winning Mean: Pre-Experiment | 3.57 | 0.65 | 0.39 | 2.25 | 0.19 |
| Winning Mean: Post-Experiment | 5.55 | 1.68 | 0.55 | 1.44 | 1.76 |
| DD Coef./Winning Pre-Treat. Mean | 0.67 | 0.71 | 0.39 | 0.54 | 2.51 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.330 | 0.343 | 0.286 | 0.312 | 0.376 |
| Panel B: Dep. Var. is Number of Notable Persons per Birth Year | | | | | |
| Post × College | 0.471+ | 0.200 | 0.025 | 0.079 | 0.164+ |
| | (0.280) | (0.122) | (0.016) | (0.051) | (0.095) |
| Runner-Up Mean: Pre-Experiment | 0.18 | 0.03 | 0.02 | 0.10 | 0.03 |
| Runner-Up Mean: Post-Experiment | 0.80 | 0.27 | 0.06 | 0.17 | 0.28 |
| Winning Mean: Pre-Experiment | 0.19 | 0.04 | 0.02 | 0.10 | 0.02 |
| Winning Mean: Post-Experiment | 1.13 | 0.41 | 0.09 | 0.23 | 0.39 |
| DD Coef./Winning Pre-Treat. Mean | 2.52 | 4.68 | 1.04 | 0.82 | 9.82 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.712 | 0.655 | 0.450 | 0.526 | 0.648 |

**Note:** Table displays difference-in-differences results estimating the effects of college establishment on notability rates (in Panel A) and the number of notable persons per birth year (in Panel B). All regressions include experiment-by-county and experiment-by-birth year fixed effects. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table 3:** College-Specific Difference-in-Differences Estimates

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Category | All | Culture | Disc/ Science | Leader- ship | Sports |
| Panel A: Dep. Var. is College-Specific Notability Rate per 10,000 Births | | | | | |
| Post × College | 0.832*** | 0.172** | 0.095* | 0.389+ | 0.175*** |
| | (0.224) | (0.059) | (0.044) | (0.205) | (0.035) |
| Runner-Up Mean: Pre-Experiment | 0.19 | 0.05 | 0.00 | 0.13 | 0.00 |
| Runner-Up Mean: Post-Experiment | 0.23 | 0.03 | 0.02 | 0.10 | 0.08 |
| Winning Mean: Pre-Experiment | 0.11 | 0.01 | 0.01 | 0.09 | 0.00 |
| Winning Mean: Post-Experiment | 1.04 | 0.21 | 0.17 | 0.36 | 0.28 |
| DD Coef./Winning Pre-Treat. Mean | 7.91 | 31.32 | 7.63 | 4.45 | Inf |
| Num.Obs. | 21047 | 21047 | 21047 | 21047 | 21047 |
| R2 | 0.432 | 0.252 | 0.346 | 0.476 | 0.365 |
| Panel B: Dep. Var. is Number of College-Specific Notable Persons per Birth Year | | | | | |
| Post × College | 0.105*** | 0.025*** | 0.013*** | 0.029*** | 0.038*** |
| | (0.024) | (0.006) | (0.004) | (0.007) | (0.010) |
| Runner-Up Mean: Pre-Experiment | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Runner-Up Mean: Post-Experiment | 0.03 | 0.00 | 0.00 | 0.01 | 0.01 |
| Winning Mean: Pre-Experiment | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Winning Mean: Post-Experiment | 0.16 | 0.04 | 0.02 | 0.05 | 0.05 |
| DD Coef./Winning Pre-Treat. Mean | 28.41 | 47.51 | 25.04 | 10.83 | Inf |
| Num.Obs. | 21047 | 21047 | 21047 | 21047 | 21047 |
| R2 | 0.518 | 0.433 | 0.406 | 0.458 | 0.433 |

**Note:** Table displays difference-in-differences results estimating the effects of college establishment on college-specific notability rates (in Panel A) and the number of college-specific notable persons per birth year (in Panel B). For a given opening-specific county in my data, college-specific notability measures the frequency of notable individuals whose Wikipedia pages mention the name of the college. Because Princeton University was called The College of New Jersey prior to 1892, I exclude The College of New Jersey from this analysis. All regressions include experiment-by-county and experiment-by-birth year fixed effects. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

# A Data Construction

This appendix describes the process used to calculate annual county-level notability rates in the U.S. between 1820 and 1980.

## A.1 Notable People Data

From the raw data from Laouenan et al. (2022), I restrict my sample to individuals with U.S. citizenship with non-missing birthplaces, with 1,000 or more Wikipedia visits between 2015 and 2018, and with birth years between 1820 and 1980, inclusive. I map each individual to a 1990 U.S. county based on the latitude and longitude of their birthplace, as coded by Laouenan et al. (2022). A small number of notable individuals have birthplaces listed as either a U.S. state, "United States of America," or "NO Place of Birth." I exclude these individuals from my sample Finally, I exclude any individuals who do not map to a modern U.S. county and, due to difficulty assigning births in New York City, New York, I exclude the five counties in New York City.

In my analysis of college openings, measurement notability rates are very sensitive to areas with low birth rates. For counties-years in my analysis with less than 10 births per year, I manually confirm the birth locations of relevant individuals. This process led to one correction related to Louis Hermann Pammel, whose birthplace was erroneously listed in Laouenan et al. (2022) data as La Crosse, Washington, rather than the correct location: La Crosse, Wisconsin.[26]

In addition, I augment college openings data by searching individual Wikipedia pages for the names of each college in my analysis. To do so, I capture the entire text of each individual's Wikipedia page. From this text, I exclude the sections beyond any of the sections titled "See also," "Notes," "References," "Bibliography," "Further reading," or "External links." This restriction limits the possibility of false positives related to sources such as the "Louisiana State University Press" or the "Penn State University Press."

With this text, I search for the names of each college in my analysis. I make a small number of changes and additions to college names relative to the names listed in Appendix Table B.2. These changes and additions are listed below:

- University of Mississippi: Additionally search for "Ole Miss"
- University of California, Berkeley: Additionally search for "University of California" and "UC Berkeley"
- Lincoln College (IL): Additionally search for "Lincoln College"
- Missouri University of Science and Technology: Additionally search for Missouri S&T
- Texas A and M University: Additionally search for "Texas A and M" and "Texas A&M"
- Virginia Polytechnic Institute: Additionally search for "Virginia Tech"

[26] "Louis H Pammel," Iowa State University University Library Online Exhibits.

- Georgia Institute of Technology: Additionally search for "Georgia Tech"
- Alabama Agricultural and Mechanical University: Additionally search for "Alabama Agricultural and Mechanical" and "Alabama A&M"
- North Carolina A and T University: Additionally search for "North Carolina A and T" and "North Carolina A&T"
- University of California, Davis: Additionally search for "UC Davis"
- US Merchant Marine Academy: Additionally search for "Merchant Marine Academy"
- US Air Force Academy: Additionally search for "Air Force Academy"

As noted in the main text, I exclude The College of New Jersey from analyses involving this data. The College of New Jersey has undergone multiple name changes over its history. Further, before 1896, Princeton University was named The College of New Jersey. Both of these facts limit the usefulness of this exercise concerning The College of New Jersey.

### A.2 County Birth Rates

I calculate annual county-level birth rates using decennial Census data from IPUMS USA and NHGIS (Ruggles et al., 2024; Manson et al., 2023). For each year, my source for county births is given below.

- **1820**: NHGIS 1820 Decennial Census, $1/10\times$ Free White Persons Under 10 years of age + $1/14\times$ Colored Population Under 14 years of age
- **1830**: NHGIS 1830 Decennial Census, $1/5\times$ Free White Persons Under 5 years of age + $1/10\times$ Colored Population Under 10 years of age
- **1840**: NHGIS 1840 Decennial Census, $1/5\times$ Free White Persons Under 5 years of age + $1/10\times$ Colored Population Under 10 years of age
- **1850**: IPUMS Full-Count 1850 Decennial Census, Persons Age 0
- **1860**: IPUMS Full-Count 1860 Decennial Census, Persons Age 0
- **1870**: IPUMS Full-Count 1870 Decennial Census, Persons Age 0
- **1880**: IPUMS Full-Count 1880 Decennial Census, Persons Age 0
- **1890**: None
- **1900**: IPUMS Full-Count 1910 Decennial Census, Persons Age 10
- **1910**: IPUMS Full-Count 1910 Decennial Census, Persons Age 0
- **1920**: IPUMS Full-Count 1920 Decennial Census, Persons Age 0

- **1930**: IPUMS Full-Count 1930 Decennial Census, Persons Age 0
- **1940**: NHGIS, $1/5\times$ Persons Under 5 years of age
- **1950**: NHGIS, $1/5\times$ Persons Under 5 years of age
- **1960**: NHGIS 1960 Decennial Census, Persons Under 1 year of age
- **1970**: NHGIS 1970 Decennial Census, Persons Under 1 year of age
- **1980**: NHGIS 1980 Decennial Census, Persons Under 1 year of age

Over this historical period, the borders that delineate U.S. counties changed periodically. To account for these changes, I convert historical counties to their 1990 county equivalents using the method described in and data provided by Eckert et al. (2020). Eckert et al. (2020) calculate the share of the area of each “reporting" unit (e.g., 1900 county) that overlaps with a “reference" unit (e.g., 1990 county), which allows me to re-aggregate data from historical counties to 1990 counties for each decennial observation. I remove any county-year observation for which less than 90% of the county’s 1990 area is accounted for in Eckert et al. (2020) data.

To estimate annual birthrates, I linearly interpolate between decennial birthrates separately for each county. I link this annual birth rate data to county-by-birth year counts of notable people and calculate notability rates: the number of (subsequently) notable people as a share of births in a county. For ease of exposition, I report notability rates per 10,000 births throughout this paper.

# B Additional Figures and Tables

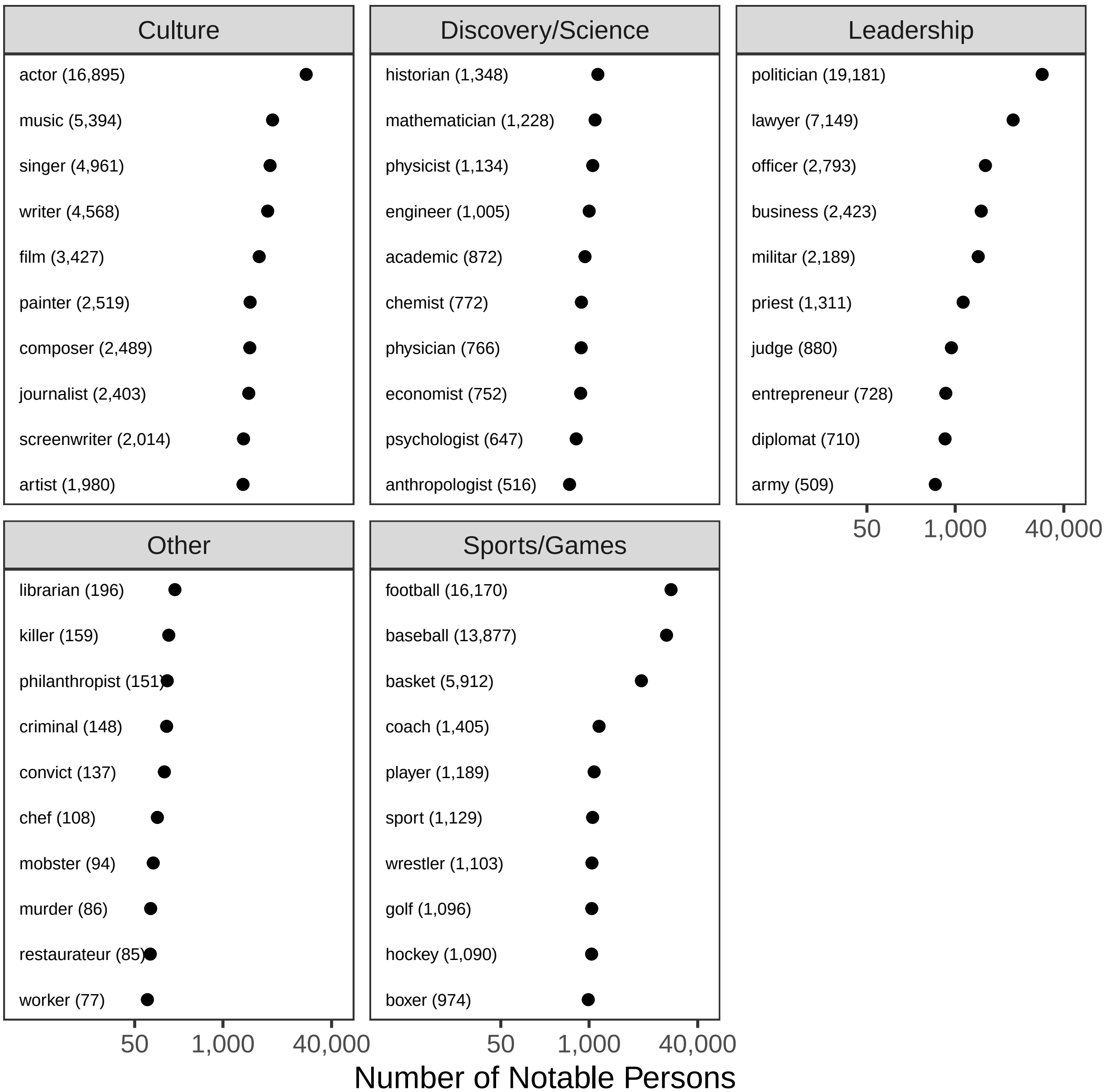


**Figure B.1:** Most Common Detailed Occupations within Each Occupation Category

**Note:** Figure displays the 10 most common detailed occupations within each occupation category. The horizontal axis is displayed in log-scale.

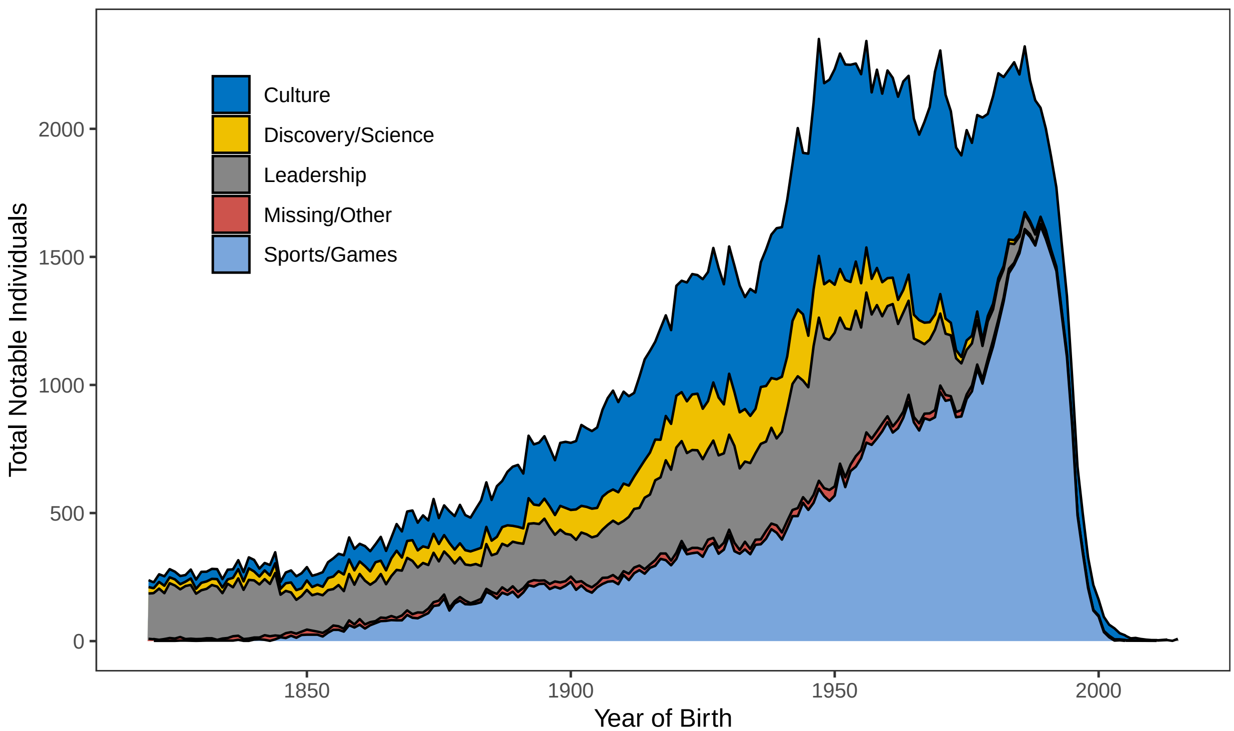


**Figure B.2:** Notable Individuals by Birth Year and Main Occupation Category

**Note:** Figure displays the number of US-born notable individuals by birth year and main occupation category.

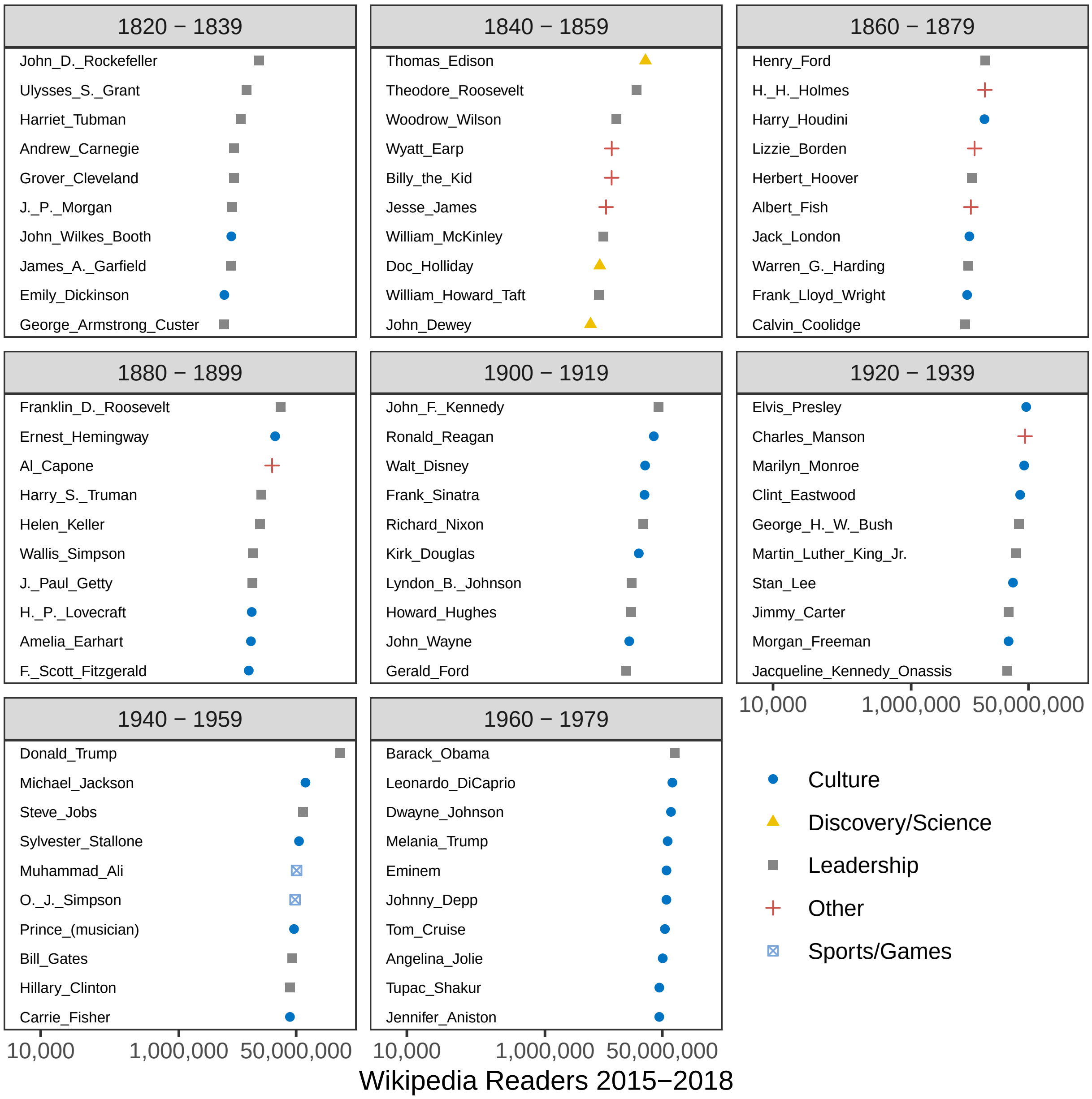


**Figure B.3:** Most-Read Wikipedia Entries by Historical Period

**Note:** Figure displays the 10 individuals whose Wikipedia pages were most-read, separately for each 2-decade period in my sample. The horizontal axis is displayed in log-scale.

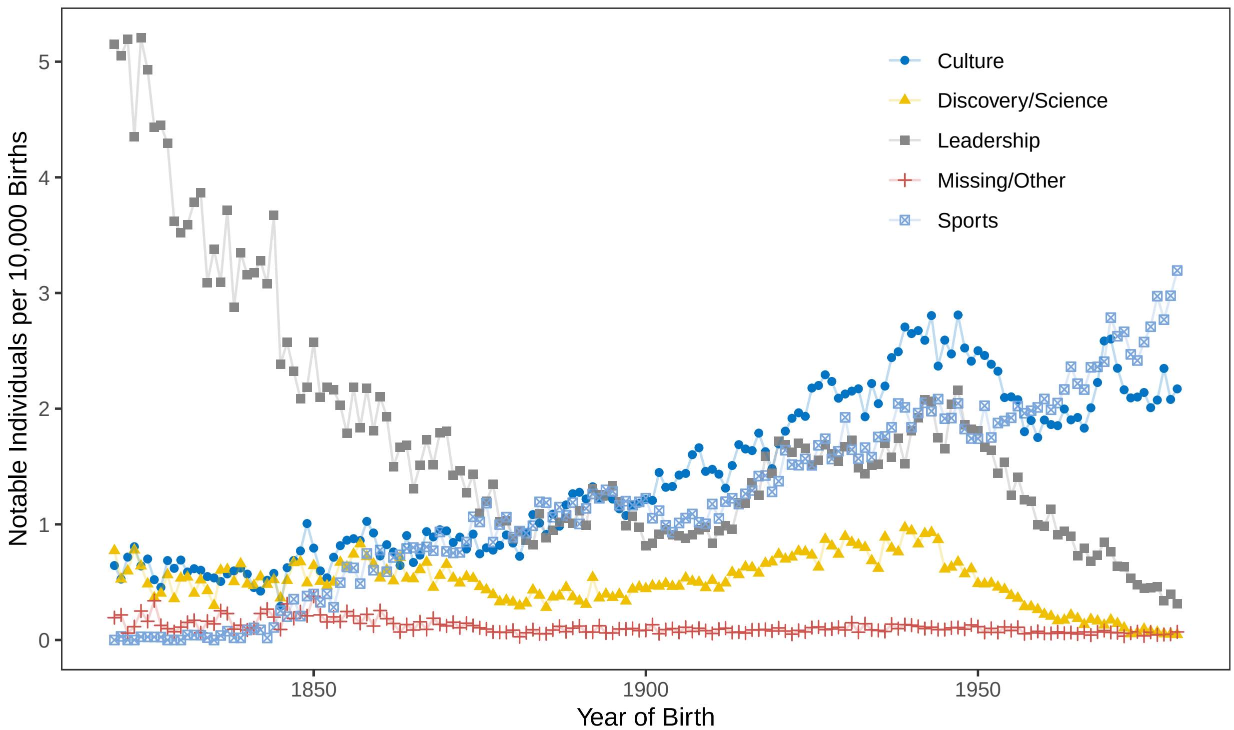


**Figure B.4:** Notable Individuals per 10,000 Births by Birth Year and Main Occupation Category

**Note:** Figure displays the number of notable individuals per 10,000 births among US-born individuals by birth year and main occupation category.

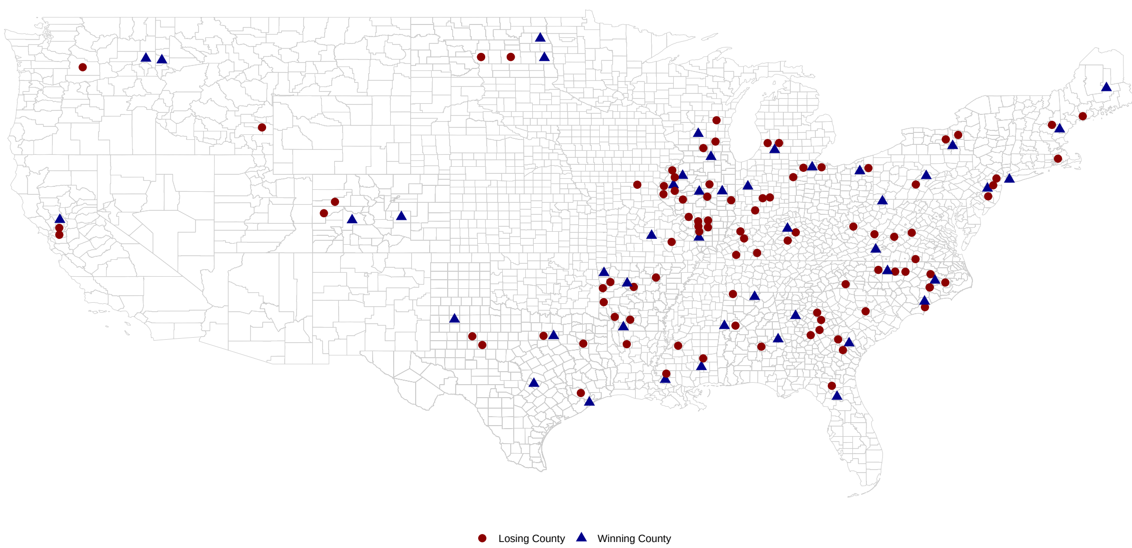


**Figure B.5:** Location of College Site Selection Experiments

**Note:** Figure displays the locations of winning and losing counties in my college site selection experiments.

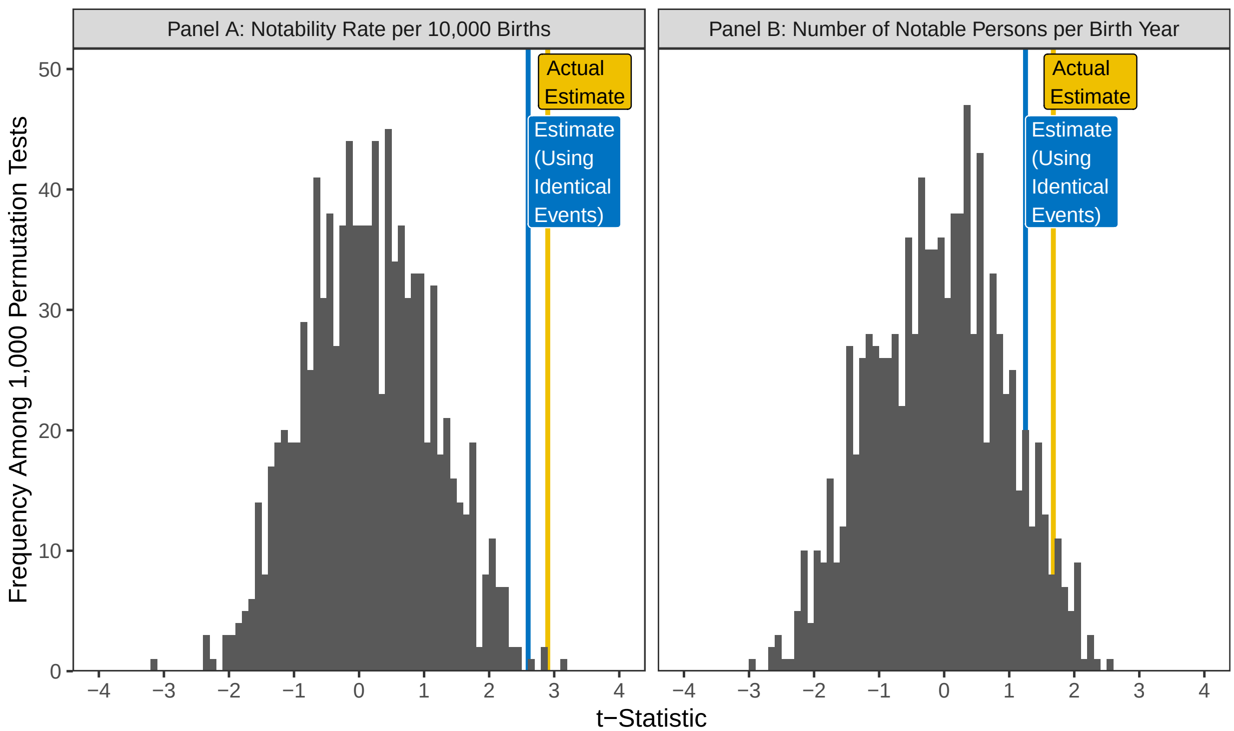


**Figure B.6:** Permutation Tests versus Actual Estimates

**Note:** Figure displays the distribution t-statistics of 1,000 permutation estimates as well as those associated with actual estimates.

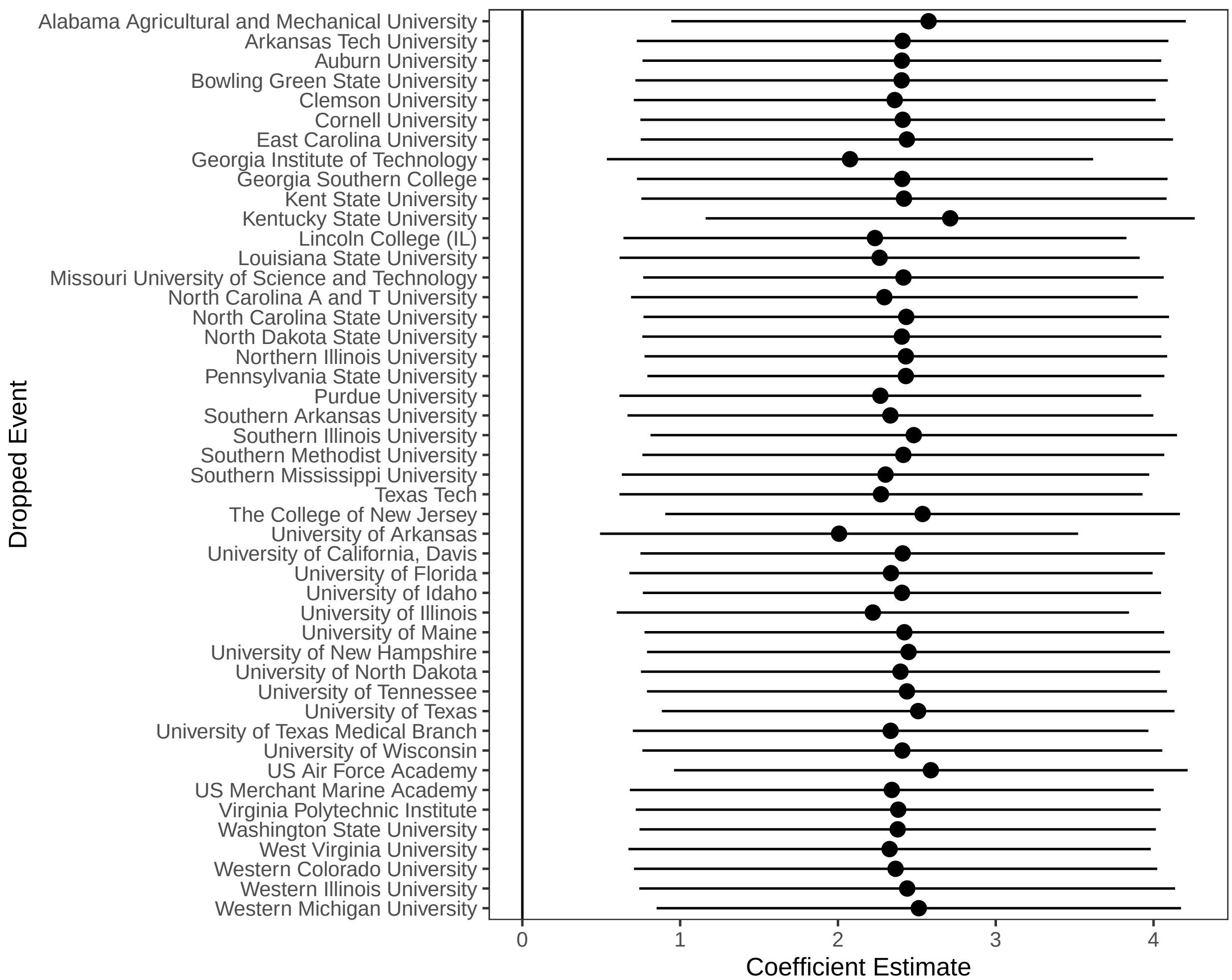


**Figure B.7:** Difference-in-Differences Estimates: Sensitivity to Omission of Individual Site Selection Experiments

**Note:** Figure displays difference-in-differences results estimating the effects of college establishment on notability rates, estimated 46 times; each time after dropping one site selection event from the data. Error bars are 95% confidence intervals.

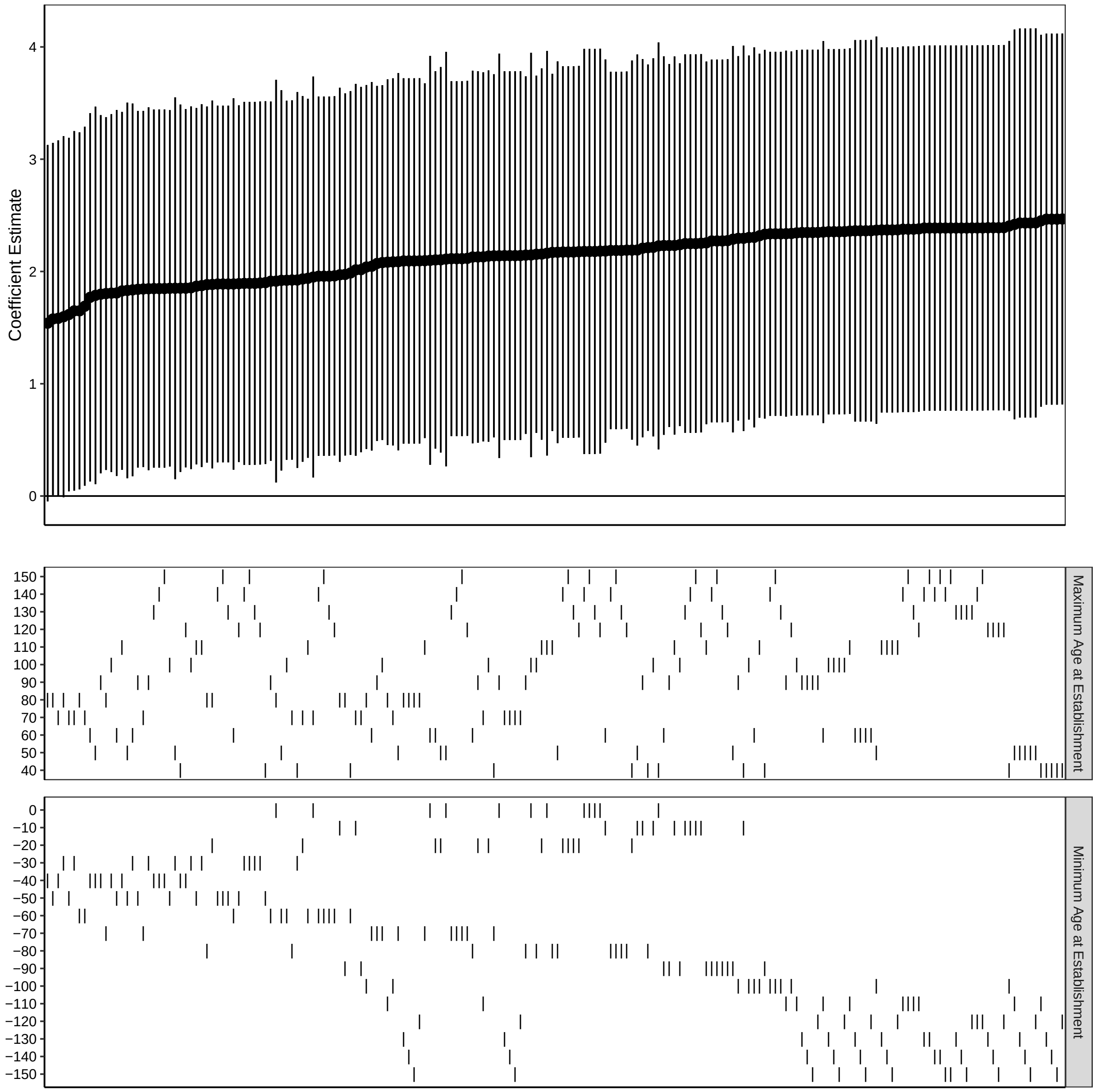


**Figure B.8:** Difference-in-Differences Estimates: Sensitivity to Different Cohort Ranges

**Note:** Figure displays difference-in-differences results estimating the effects of college establishment on notability rates, estimated for different ranges of birth cohorts, defined by their age at college establishment. Each point represents a difference-in-differences estimate, estimated after including a different set of cohorts (shown in bottom panels) in the data. Error bars are 95% confidence intervals.

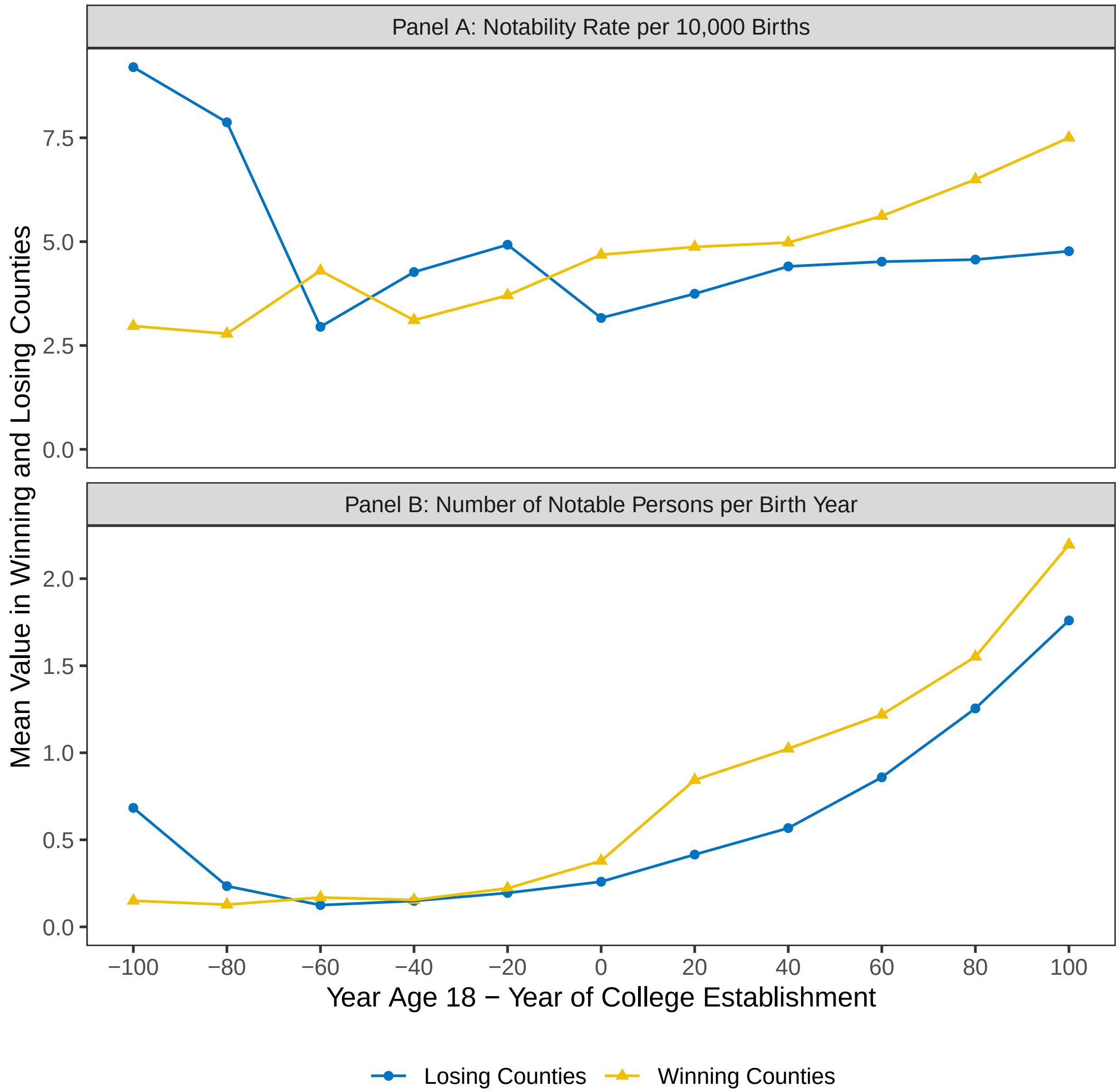


**Figure B.9:** Trends in Notability Among Winning Counties and Losing Counties (Full Data)

**Note:** Figure displays trends in notability among winning and losing counties in 5-year birth cohorts, relative to the year of college establishment. Panel A displays averages of notability rates per 10,000 births. Panel B displays averages of the number of notable persons per birth year.

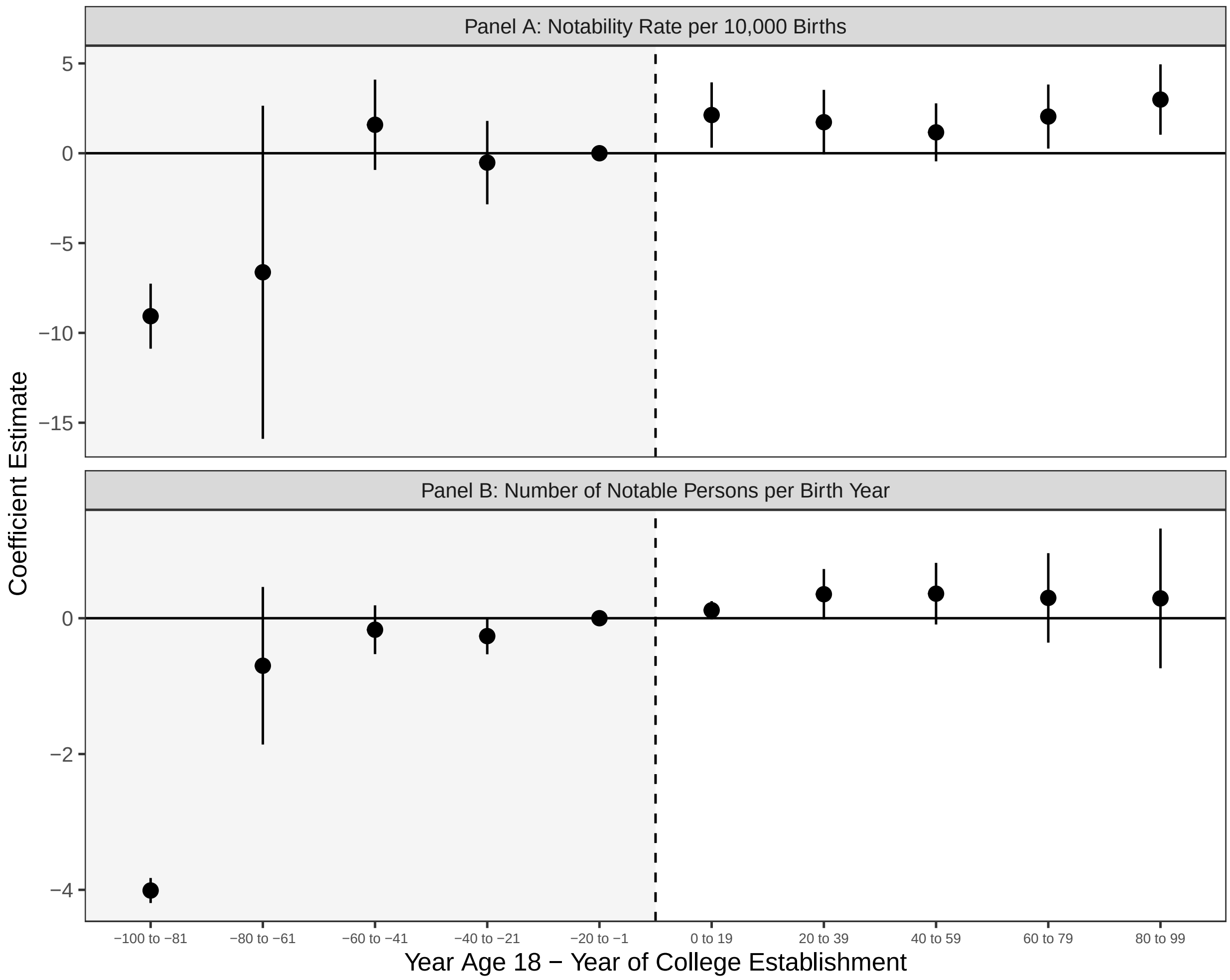


**Figure B.10:** Dynamic Impact of College Establishment on Notability

**Note:** Figure displays event study results showing the dynamic effect of college establishment on notability rates (in Panel A) and the number of notable persons per birth year (in Panel B). Error bars are 95% confidence intervals.

**Table B.1:** Calculating the Share of Academy Award-Winning Actors and Directors Born in New York City

| | Total | Total Born in NYC | Share Born in NYC |
|---|---|---|---|
| **Award Counts** | | | |
| Best Actor Winners | 86<br>Link | 8<br>Link | 9.30% |
| Best Actress Winners | 79<br>Link | 7<br>Link | 8.86% |
| Best Supporting Actor Winners | 79<br>Link | 14<br>Link | 17.72% |
| Best Supporting Actress Winners | 86<br>Link | 13<br>Link | 15.12% |
| Best Director Winners | 75<br>Link | 6<br>Link | 8.00% |
| **Multiple-Award Winners** | | | |
| Actor/Supporting Actor | 6<br>Link | 2<br>Link | |
| Actress/Supporting Actress | 7<br>Link | 0<br>Link | |
| Actor/Director | 0<br>Link | 0<br>Link | |
| Actress/Director | 0<br>Link | 0<br>Link | |
| Supporting Actor/Director | 0<br>Link | 0<br>Link | |
| Supporting Actress/Director | 0<br>Link | 0<br>Link | |
| **Total less Multiple-Award Winners** | **392** | **46** | **11.73%** |

**Note:** Table displays calculations and documentation necessary to calculate the share of Academy Award-winning actors and directors who were born in New York City.

**Table B.2:** List of College Site Selection Experiments

| College | Experiment Year | State | Winning County | Runner-Up Counties |
|---|---|---|---|---|
| Pennsylvania State University | 1855 | Pennsylvania | Centre | Blair |
| The College of New Jersey | 1855 | New Jersey | Mercer | Middlesex, Burlington, Essex |
| Lincoln College (IL) | 1864 | Illinois | Logan | Edgar, Warrick, Macon |
| Cornell University | 1865 | New York | Tompkins | Schuyler, Seneca, Onondaga |
| University of Maine | 1866 | Maine | Penobscot | Sagadahoc |
| University of Wisconsin | 1866 | Wisconsin | Dane | Fond Du Lac |
| University of Illinois | 1867 | Illinois | Champaign | Mclean, Morgan |
| West Virginia University | 1867 | West Virginia | Monongalia | Greenbrier, Kanawha |
| Purdue University | 1869 | Indiana | Tippecanoe | Monroe, Marion, Hancock |
| Southern Illinois University | 1869 | Illinois | Jackson | Perry, Clinton, Marion, Washington, Jefferson |
| University of Tennessee | 1869 | Illinois | Knox | Rutherford |
| Louisiana State University | 1870 | Louisiana | East Baton Rouge | Bienville, East Feliciana |
| Missouri University of Science and Technology | 1870 | Missouri | Phelps | Iron |
| University of Arkansas | 1871 | Arkansas | Washington | Independence |
| Auburn University | 1872 | Alabama | Lee | Tuscaloosa, Lauderdale |
| Virginia Polytechnic Institute | 1872 | Virginia | Montgomery | Albemarle, Rockbridge |
| University of Texas | 1881 | Texas | Travis | Smith |
| University of Texas Medical Branch | 1881 | Texas | Galveston | Harris |
| North Dakota State University | 1883 | North Dakota | Cass | Burleigh, Stutsman |
| University of North Dakota | 1883 | North Dakota | Grand Forks | Burleigh, Stutsman |
| Georgia Institute of Technology | 1886 | Georgia | Fulton | Clarke, Greene, Baldwin, Bibb |
| Kentucky State University | 1886 | Kentucky | Franklin | Boyle, Warren, Daviess, Christian, Fayette |
| North Carolina State University | 1886 | North Carolina | Pender | Lenoir, Mecklenburg |
| Clemson University | 1889 | Alabama | Pickens | Richland |
| University of Idaho | 1889 | Idaho | Latah | Bonneville |
| Alabama Agricultural and Mechanical University | 1891 | Alabama | Madison | Montgomery |
| University of New Hampshire | 1891 | New Hampshire | Strafford | Belknap |
| Washington State University | 1891 | Washington | Whitman | Yakima |
| North Carolina A and T University | 1892 | North Carolina | Guilford | Durham, New Hanover, Alamance, Forsyth |
| Northern Illinois University | 1895 | Illinois | De Kalb | Winnebago |
| Western Illinois University | 1899 | Illinois | Mcdonough | Adams, Hancock, Warren, Schuyler, Mercer |
| Western Michigan University | 1903 | Michigan | Kalamazoo | Barry, Allegan |
| University of Florida | 1905 | Florida | Alachua | Columbia |
| Georgia Southern College | 1906 | Georgia | Bulloch | Tattnall, Emanuel |
| University of California, Davis | 1906 | California | Yolo | Solano, Contra Costa |
| East Carolina University | 1907 | North Carolina | Pitt | Lenoir, Beaufort, Edgecombe |
| Western Colorado University | 1909 | Colorado | Gunnison | Garfield, Mesa |
| Arkansas Tech University | 1910 | Arkansas | Pope | Sebastian, Conway, Franklin |
| Bowling Green State University | 1910 | Ohio | Wood | Henry, Van Wert, Sandusky |
| Kent State University | 1910 | Ohio | Portage | Trumbull |
| Southern Arkansas University | 1910 | Arkansas | Columbia | Hempstead, Ouachita, Polk |
| Southern Mississippi University | 1910 | Mississippi | Forrest | Jones, Hinds |
| Southern Methodist University | 1911 | Texas | Dallas | Tarrant |
| Texas Tech | 1923 | Texas | Lubbock | Scurry, Nolan |
| US Merchant Marine Academy | 1941 | New York | Nassau | Bristol |
| US Air Force Academy | 1954 | Colorado | El Paso | Walworth, Madison |

**Note:** Table displays the set of college site selection experiments included in my sample.

**Table B.3:** Summary Statistics

| Statistic | N | Mean | St. Dev. |
|---|---|---|---|
| Panel A: Full Sample; 3,136 Unique Counties | | | |
| Notability Rate per 10,000 Births: All | 446,683 | 4.386 | 96.378 |
| Notability Rate per 10,000 Births: Culture | 446,683 | 0.958 | 33.937 |
| Notability Rate per 10,000 Births: Discovery/Science | 446,683 | 0.497 | 47.593 |
| Notability Rate per 10,000 Births: Leadership | 446,683 | 1.816 | 70.501 |
| Notability Rate per 10,000 Births: Sports | 446,683 | 0.992 | 19.488 |
| Notability Rate per 10,000 Births: Missing/Other | 446,683 | 0.121 | 15.840 |
| Number of Notable Persons: All | 446,696 | 0.342 | 2.284 |
| Number of Notable Persons: Culture | 446,696 | 0.115 | 1.147 |
| Number of Notable Persons: Discovery/Science | 446,696 | 0.029 | 0.259 |
| Number of Notable Persons: Leadership | 446,696 | 0.086 | 0.452 |
| Number of Notable Persons: Sports | 446,696 | 0.106 | 0.754 |
| Number of Notable Persons: Missing/Other | 446,696 | 0.006 | 0.082 |
| Total Births | 446,696 | 650.157 | 2,142.600 |
| Birthyear | 446,696 | 1,907.284 | 43.820 |
| Panel B: Event Sample; 138 Unique Counties | | | |
| Notability Rate per 10,000 Births: All | 21,691 | 4.473 | 13.011 |
| Notability Rate per 10,000 Births: Culture | 21,691 | 1.102 | 4.388 |
| Notability Rate per 10,000 Births: Discovery/Science | 21,691 | 0.430 | 3.326 |
| Notability Rate per 10,000 Births: Leadership | 21,691 | 1.699 | 10.525 |
| Notability Rate per 10,000 Births: Sports | 21,691 | 1.151 | 4.323 |
| Notability Rate per 10,000 Births: Missing/Other | 21,691 | 0.091 | 1.687 |
| Number of Notable Persons: All | 21,691 | 0.698 | 2.095 |
| Number of Notable Persons: Culture | 21,691 | 0.237 | 0.939 |
| Number of Notable Persons: Discovery/Science | 21,691 | 0.056 | 0.278 |
| Number of Notable Persons: Leadership | 21,691 | 0.163 | 0.516 |
| Number of Notable Persons: Sports | 21,691 | 0.231 | 0.890 |
| Number of Notable Persons: Missing/Other | 21,691 | 0.011 | 0.107 |
| Total Births | 21,691 | 1,142.219 | 2,350.622 |
| Birthyear | 21,691 | 1,903.050 | 45.239 |
| Experiment Year | 21,691 | 1,888.502 | 21.011 |

**Note:** Table displays summary statistics for all counties (in Panel A) and for counties in the event sample (in Panel B). + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table B.4:** Testing for Differences in Pre-College Trends and Levels

| | (1) | (2) |
|---|---|---|
| Testing for Differences in: | Trends | Levels |
| Panel A: Dep. Var. is Notability Rate per 10,000 Births | | |
| Birthyear × College | 0.053 | |
| | (0.039) | |
| College | | -0.847 |
| | | (0.734) |
| Num.Obs. | 6183 | 6183 |
| R2 | 0.328 | 0.284 |
| Panel B: Dep. Var. is Number of Notable Persons per Birth Year | | |
| Birthyear × College | 0.002 | |
| | (0.002) | |
| College | | -0.021 |
| | | (0.033) |
| Num.Obs. | 6183 | 6183 |
| R2 | 0.540 | 0.482 |

**Note:** Table displays estimates of Equations 3 (in Column 1) and Equations 4 (in Column 2), which test for differences in notability trends prior to college establishment. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table B.5:** Difference-in-Differences Estimates (Simple Fixed-Effects)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Category | All | Culture | Disc/ Science | Leader-ship | Sports |
| Panel A: Dep. Var. is Notability Rate per 10,000 Births | | | | | |
| Post × College | 2.577** | 0.507* | 0.218 | 1.233+ | 0.525* |
| | (0.895) | (0.231) | (0.162) | (0.624) | (0.201) |
| Runner-Up Mean: Pre-Experiment | 4.54 | 0.50 | 0.46 | 3.17 | 0.28 |
| Runner-Up Mean: Post-Experiment | 4.09 | 1.15 | 0.37 | 1.12 | 1.39 |
| Winning Mean: Pre-Experiment | 3.57 | 0.65 | 0.39 | 2.25 | 0.19 |
| Winning Mean: Post-Experiment | 5.55 | 1.68 | 0.55 | 1.44 | 1.76 |
| DD Coef./Winning Pre-Treat. Mean | 0.72 | 0.78 | 0.56 | 0.55 | 2.77 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.051 | 0.060 | 0.028 | 0.033 | 0.059 |
| Panel B: Dep. Var. is Number of Notable Persons per Birth Year | | | | | |
| Post × College | 0.613+ | 0.259 | 0.033+ | 0.101 | 0.215+ |
| | (0.367) | (0.164) | (0.019) | (0.063) | (0.127) |
| Runner-Up Mean: Pre-Experiment | 0.18 | 0.03 | 0.02 | 0.10 | 0.03 |
| Runner-Up Mean: Post-Experiment | 0.80 | 0.27 | 0.06 | 0.17 | 0.28 |
| Winning Mean: Pre-Experiment | 0.19 | 0.04 | 0.02 | 0.10 | 0.02 |
| Winning Mean: Post-Experiment | 1.13 | 0.41 | 0.09 | 0.23 | 0.39 |
| DD Coef./Winning Pre-Treat. Mean | 3.28 | 6.05 | 1.38 | 1.04 | 12.86 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.457 | 0.383 | 0.167 | 0.242 | 0.326 |

**Note:** Table displays difference-in-differences results estimating the effects of college establishment on notability rates (in Panel A) and the number of notable persons per birth year (in Panel B). All regressions include county and birth year fixed effects. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table B.6:** Difference-in-Differences Estimates (Adding Cohort Size Controls)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Category | All | Culture | Disc/ Science | Leader- ship | Sports |
| Panel A: Dep. Var. is Notability Rate per 10,000 Births | | | | | |
| Post × College | 2.359** | 0.432* | 0.152 | 1.250+ | 0.447*** |
| | (0.804) | (0.195) | (0.146) | (0.673) | (0.121) |
| Total Births (1000s) | 0.209 | 0.150 | -0.004 | -0.092 | 0.174** |
| | (0.274) | (0.094) | (0.053) | (0.140) | (0.066) |
| Total Births (1000s) Squared | -0.009 | -0.004+ | 0.000 | 0.001 | -0.006*** |
| | (0.007) | (0.002) | (0.001) | (0.004) | (0.002) |
| Runner-Up Mean: Pre-Experiment | 4.54 | 0.50 | 0.46 | 3.17 | 0.28 |
| Runner-Up Mean: Post-Experiment | 4.09 | 1.15 | 0.37 | 1.12 | 1.39 |
| Winning Mean: Pre-Experiment | 3.57 | 0.65 | 0.39 | 2.25 | 0.19 |
| Winning Mean: Post-Experiment | 5.55 | 1.68 | 0.55 | 1.44 | 1.76 |
| DD Coef./Winning Pre-Treat. Mean | 0.66 | 0.66 | 0.39 | 0.56 | 2.35 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.330 | 0.344 | 0.286 | 0.312 | 0.376 |
| Panel B: Dep. Var. is Number of Notable Persons per Birth Year | | | | | |
| Post × College | 0.188* | 0.087* | 0.008 | 0.033 | 0.061* |
| | (0.086) | (0.042) | (0.011) | (0.024) | (0.028) |
| Total Births (1000s) | 0.857*** | 0.331*** | 0.071*** | 0.178*** | 0.271*** |
| | (0.133) | (0.064) | (0.019) | (0.024) | (0.048) |
| Total Births (1000s) Squared | -0.004 | -0.001 | -0.002*** | -0.003*** | 0.001 |
| | (0.003) | (0.001) | (0.000) | (0.001) | (0.001) |
| Runner-Up Mean: Pre-Experiment | 0.18 | 0.03 | 0.02 | 0.10 | 0.03 |
| Runner-Up Mean: Post-Experiment | 0.80 | 0.27 | 0.06 | 0.17 | 0.28 |
| Winning Mean: Pre-Experiment | 0.19 | 0.04 | 0.02 | 0.10 | 0.02 |
| Winning Mean: Post-Experiment | 1.13 | 0.41 | 0.09 | 0.23 | 0.39 |
| DD Coef./Winning Pre-Treat. Mean | 1.01 | 2.03 | 0.34 | 0.34 | 3.66 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.862 | 0.782 | 0.471 | 0.574 | 0.784 |

**Note:** Table displays difference-in-differences results estimating the effects of college establishment on notability rates (in Panel A) and the number of notable persons per birth year (in Panel B). All regressions include experiment-by-county and experiment-by-birth year fixed effects. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table B.7:** Difference-in-Differences Estimates (Visit Cutoff = 0)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Category | All | Culture | Disc/ Science | Leader- ship | Sports |
| Panel A: Dep. Var. is Notability Rate per 10,000 Births | | | | | |
| Post × College | 2.523** | 0.517* | 0.164 | 1.240+ | 0.500*** |
| | (0.856) | (0.222) | (0.186) | (0.688) | (0.121) |
| Runner-Up Mean: Pre-Experiment | 5.22 | 0.62 | 0.58 | 3.53 | 0.34 |
| Runner-Up Mean: Post-Experiment | 4.54 | 1.25 | 0.44 | 1.26 | 1.53 |
| Winning Mean: Pre-Experiment | 4.25 | 0.71 | 0.51 | 2.68 | 0.25 |
| Winning Mean: Post-Experiment | 6.11 | 1.81 | 0.65 | 1.61 | 1.92 |
| DD Coef./Winning Pre-Treat. Mean | 0.59 | 0.73 | 0.32 | 0.46 | 2.01 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.337 | 0.337 | 0.289 | 0.317 | 0.377 |
| Panel B: Dep. Var. is Number of Notable Persons per Birth Year | | | | | |
| Post × College | 0.496+ | 0.202 | 0.030 | 0.085 | 0.176+ |
| | (0.297) | (0.127) | (0.019) | (0.053) | (0.102) |
| Runner-Up Mean: Pre-Experiment | 0.21 | 0.03 | 0.03 | 0.11 | 0.03 |
| Runner-Up Mean: Post-Experiment | 0.86 | 0.29 | 0.07 | 0.19 | 0.30 |
| Winning Mean: Pre-Experiment | 0.22 | 0.05 | 0.03 | 0.11 | 0.02 |
| Winning Mean: Post-Experiment | 1.21 | 0.43 | 0.10 | 0.24 | 0.42 |
| DD Coef./Winning Pre-Treat. Mean | 2.28 | 4.20 | 1.09 | 0.75 | 8.22 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.715 | 0.658 | 0.454 | 0.531 | 0.648 |

**Note:** Table displays difference-in-differences results estimating the effects of college establishment on notability rates (in Panel A) and the number of notable persons per birth year (in Panel B). Sample of notable individuals includes individuals with at least 1 Wikipedia visit between 2015 and 2018. All regressions include experiment-by-county and experiment-by-birth year fixed effects. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table B.8:** Difference-in-Differences Estimates (Visit Cutoff = 100)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Category | All | Culture | Disc/ Science | Leader- ship | Sports |
| Panel A: Dep. Var. is Notability Rate per 10,000 Births | | | | | |
| Post × College | 2.548** | 0.516* | 0.164 | 1.255+ | 0.512*** |
| | (0.854) | (0.222) | (0.186) | (0.686) | (0.121) |
| Runner-Up Mean: Pre-Experiment | 5.21 | 0.62 | 0.58 | 3.53 | 0.34 |
| Runner-Up Mean: Post-Experiment | 4.53 | 1.25 | 0.44 | 1.25 | 1.52 |
| Winning Mean: Pre-Experiment | 4.23 | 0.71 | 0.51 | 2.66 | 0.25 |
| Winning Mean: Post-Experiment | 6.10 | 1.80 | 0.65 | 1.61 | 1.92 |
| DD Coef./Winning Pre-Treat. Mean | 0.60 | 0.72 | 0.32 | 0.47 | 2.06 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.336 | 0.336 | 0.289 | 0.317 | 0.377 |
| Panel B: Dep. Var. is Number of Notable Persons per Birth Year | | | | | |
| Post × College | 0.497+ | 0.202 | 0.030 | 0.085 | 0.177+ |
| | (0.297) | (0.126) | (0.019) | (0.053) | (0.102) |
| Runner-Up Mean: Pre-Experiment | 0.21 | 0.03 | 0.03 | 0.11 | 0.03 |
| Runner-Up Mean: Post-Experiment | 0.86 | 0.29 | 0.07 | 0.19 | 0.30 |
| Winning Mean: Pre-Experiment | 0.22 | 0.05 | 0.03 | 0.11 | 0.02 |
| Winning Mean: Post-Experiment | 1.20 | 0.43 | 0.10 | 0.24 | 0.42 |
| DD Coef./Winning Pre-Treat. Mean | 2.29 | 4.20 | 1.10 | 0.75 | 8.26 |
| Num.Obs. | 21691 | 21691 | 21691 | 21691 | 21691 |
| R2 | 0.715 | 0.658 | 0.454 | 0.531 | 0.648 |

**Note:** Table displays difference-in-differences results estimating the effects of college establishment on notability rates (in Panel A) and the number of notable persons per birth year (in Panel B). Sample of notable individuals includes individuals with at least 100 Wikipedia visits between 2015 and 2018. All regressions include experiment-by-county and experiment-by-birth year fixed effects. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.