Data Sources for "Life, Literacy, and the Pursuit of Prosperity: Party Competition and Policy Outcomes in 50 States"

Measures of Health, Education, and Income

Infant Mortality

In 1890, the federal census reported figures on infant mortality for the first time—defined that year, as it has been defined since, as the death rate of persons under one year of age for every thousand births—but it appears that federal officials quickly became convinced that the state reports could not be trusted. Raw numbers of births and infant deaths were again published in 1900, but that year the federal census calculated an infant mortality rate only for nine "registration states." The number of "registration states" fell to seven in the 1910 census (every New England state except Rhode Island, as well as Michigan and Pennsylvania), which must have reflected an even more rigorous set of standards than was used in 1900, and raw numbers that year were reported for no other states. In 1920, the list of "registration states" had risen to 23, and, by 1930, all but 2 states had achieved that status. South Dakota and Texas, the last holdouts, became "registration states" in 1932 and 1933, respectively. Since 1933, infant mortality has been reported regularly for every state in the union.

U.S. Department of the Interior, Census Office. *Report on Vital and Social Statistics in the United States at the Eleventh Census: 1890—Part I: Analysis and Rate Tables.* Washington, D.C.: Government Printing Office, 1896. The data are from pp. 513-657. The heading of the table is "Table 1.—Population, Births, Deaths, and Death Rates at Certain Ages, and Deaths From Certain Causes, with Distinction of Sex, Color, General Nativity, and Parental Nativity." Missing: AK and HI.

U.S. Census Office. *Twelfth Census of the United States, Taken in the Year 1900, Census Reports, Volume III—Vital Statistics, Part I, Analysis and Ratio Tables.* Washington D.C.: United States Census Office, 1902. The data are from pp. 285-555. The heading of the table is "Table 19.—Population, Births, Deaths, and Death Rates at Certain Ages, and Deaths From Certain Causes, by Sex, Color, General Nativity, and Parent Nativity: Census Year 1900." For all data, missing: AK and HI. For "Deaths under [Age] 1 Per 1,000 Births," all states missing except CT, ME, MA, MI, NH, NJ, NY, RI, and VT.

U.S. Department of Commerce and Labor, Bureau of the Census. *Mortality Statistics: 1910.* Bulletin 109. Washington D.C.: Government Printing Office, 1912. The data are from pg. 18. The table is unnamed, but shows for "certain states and for certain cities of 50,000 population or over in 1910, for each of the years 1908 to 1910, the number of living births registered, the number of deaths of infants under 1 year of age, and the infant mortality or ratio of deaths of infants under 1 year to every 1,000 births." For infant mortality rate, all states missing except CT, ME, MA, MI, NH, PA, RI (exclusive of Providence and Pawtucket), and VT. For births, all states missing except CT, ME, MA, NH, PA, RI (exclusive of Providence and Pawtucket), and VT.

U.S. Department of Commerce, Bureau of the Census. *Mortality Statistics: 1910—Eleventh Annual Report.* Washington D.C.: Government Printing Office, 1913. The data are from pp. 304-331. The heading of the table is "Table 5.—Deaths (Exclusive of Stillbirths) in the Registration Area, Its Main Subdivisions, and Decedent: Each Registration State, By Age and By Sex, Color, General Nativity, and Parent Nativity of 1910." Only these states are included: CA, CO, CT, IN, ME, MD, MA, MI, MN, MT, NH, NJ, NY, NC, OH, PA, RI, UT, VT, WA, and WI.

U.S. Department of Commerce, Bureau of the Census. *Fourteenth Census of the United States—Taken in the Year 1920—Volume II—Population, 1920: General Report and Analytical Tables*. Washington D.C.: Government Printing Office, 1922. The data are from pp. 188-285. The heading of the table is "Table 13.—Distribution By Age Periods For the Total Population and By Single Years of Age For Persons Under 25, For Population Classes, Both Sexes, 1920 and 1910, and Each Sex, 1920, By Divisions and States." Missing: AK and HI. For 1910 population data.

U.S. Department of Commerce, Bureau of the Census. *Fourteenth Census of the United States Taken in the Year 1920—Volume III—Population, 1920: Composition and Characteristics of the Population by States*. Washington D.C.: Government Printing Office, 1922. The data are from pp. 26-33. The heading of the table is "Table 9.—Age, For Divisions, 1920 and 1910, and For States, 1920." Missing: AK and HI.

U.S. Department of Commerce, Bureau of the Census. *Mortality Statistics, 1920: Twenty-First Annual Report*. Washington D.C.: Government Printing Office, 1922. The data are from pp. 180-213. The heading of the table is "Deaths (Exclusive of Stillbirths) in the Registration Area (Exclusive of Hawaii), Its Main Subdivisions, and Each Registration State, By Age, Sex, Color, General Nativity, and Parent Nativity of Decedent." Only these states are included: CA, CO, CT, DE, FL, IL, IN, KS, KY, LA, ME, MD, MA, MI, MN, MS, MO, MT, NE, NH, NJ, NY, NC, OH, OR, PA, RI, SC, TN, UT, VT, VA, WA, and WI.

U.S. Department of Commerce. *Statistical Abstract of the United States, 1931, Fifty-Third Number.* Washington, D.C.: Government Printing Office, 1931. The data are from pp. 83-84. The heading of the table is "No. 75.—Births and Excess of Births Over Deaths: By States." Only these states are included: CA, CT, IN, KS, KY, ME, MD, MA, MI, MN, NE, NH, NY, NC, OH, OR, PA, SC, UT, VT, VA, WA, and WI. For 1920 number of births.

U.S. Department of Commerce. *Statistical Abstract of the United States, 1931, Fifty-Third Number.* Washington, D.C.: Government Printing Office, 1931. The data are from pp. 87-88. The heading of the table is "No. 79.—Deaths of Infants Under 1 Year of Age: Number and Rates, By States, For the Birth Registration Area." The data provides the "deaths of infants under 1 year of age per 1,000 births." For 1920, only these states are included: CA, CT, IN, KS, KY, ME, MD, MA, MI, MN, NE, NH, NY, NC, OH, OR, PA, SC, UT, VT, VA, WA, and WI. For 1930, all states are included except AK, SD, and TX. In U.S. Department of Commerce, Bureau of the Census, *Mortality Statistics, 1931 and 1932: Selected Tables* (Washington, D.C.: Government Printing Office, 1934), the

"Letter of Transmittal" reads: "The death registration area for which statistics are here presented included in 1931 all States except Texas and South Dakota, and in 1932 all except Texas. Texas has now been admitted. So the statistics for 1933 will cover for the first time the entire area of continental United States. This will be the consummation of the more than 30 years of persistent organized effort on the part of the Bureau of the Census and the American Public Health Association, beginning when the registration area for deaths covered only 10 States with the District of Columbia." For "1930" figures for South Dakota, we use 1932 data; for Texas, we use 1933 data. See also https://www.cdc.gov/nchs/data/vsushistorical/birthstat_1934.pdf

Robert D. Grove and Alice M. Hetzel, *Vital Statistics Rates in the United States, 1940-1960*. Washington, D.C.: National Center for Health Statistics, Public Health Service, U.S. Department of Health, Education, and Welfare, 1968. The data are from pp. 214-235. The heading of the table is "Table 41.—Infant mortality rates by age and color: United States and each state, 1940-60," with the rates given being "infant deaths per 1,000 live births in specified color-sex group." We drew data from the "total" number (combining "white" and "nonwhite") "under 1 year." For 1940, 1950, and 1960, only AK is missing.

Vital Statistics of the United States, 1970—Vol. II—Mortality, Part A. Rockville, Maryland: National Center for Health Statistics, U.S. Department of Health, Education, and Welfare, 1974. The data are in "Table 2-6. Infant Mortality Rates by Color: United States, Each Division and State, 1966-70," with the rates given being "deaths under 1 year per 1,000 live births." We drew data from the "total" number (combining "white" and "all other"). No missing states.

U.S. Census Bureau, *Statistical Abstract of the United States: 2003 (123rd Edition), Section 2: Vital Statistics.* Washington, D.C., 2003. The data are from pg. 90. The heading of the table is "No. 114. Infant Mortality Rates by Race—States: 1980 to 2001," with the rate representing "deaths per 1,000 live births, by place of residence." The variable is further defined as "deaths of infants under 1 year old, exclusive of fetal deaths" and "excludes deaths of nonresidents of the United States." We drew data from the "total" number (combining "white" and "Black" and including "other races, not shown separately"). No missing states. For 1980, 1990, and 2000.

T.J. Mathews and Marian F. MacDorman, "Infant Mortality Statistics from the 2010 Period Linked Birth/Infant Death Data Set," *National Vital Statistics Reports*, vol. 62, no. 8. National Center for Health Statistics, Centers for Disease Control and Prevention, U.S Department of Health and Human Services, 18 Dec. 2013. The data are from pg. 6. The heading of the table is "Table C. Infant mortality rates, number of infant deaths, and percent change: United States and each state, Puerto Rico, Virgin Islands, and Guam, 2005 and 2010 linked files," with the rates given being "infant mortality rate per 1,000 live births." No missing states.

Life Expectancy

In 1890, life expectancy was reported only for one state (Massachusetts), and, in 1900 and 1910, life expectancy was reported only for Indiana, Massachusetts, Michigan, New Jersey, and New York. Data for half of the states was reported in 1920, then no data at all were reported in 1930. Only in 1933 did "death registration procedures" exist in all states (Pope 1992, 268), but, in 1940, the first year that data were reported for every state, life expectancy was reported only for people classified as "white." A full data series for life expectancy—including all people in every state—has existed only since 1950. Accurate data exist between 1890 and 1940, but only for some states, and only if assumptions are made about the death rate for people in racial categories not reported in census data.

In 1920, life expectancy is given at the state level only for "whites" in states where data are reported, for "Negroes" at the national but not state level (with national data broken down into two categories, composed of states where Blacks made up less than 4% of the population and states where Blacks made up at least 5% of the population), and for "all races combined" in the territory of Hawaii. To estimate life expectancy for the entire population, we drew on the 1920 population census to determine the number of whites, Blacks, and other nonwhites in each state's population. In the absence of better information (which existed nowhere except the Hawaii territory), we grouped nonwhites with Blacks, rather than whites, for the sake of these estimates. Noting whether nonwhites made up less than 4% or greater than 5% of the population, we then used the figures reported by the census to estimate life expectancy for each state's whole population in states where data are reported for whites. (For California, where nonwhites made up 4.7% of the population, we used the life expectancy figure for states whose populations were greater than 5% Black.) In 1940, life expectancy is reported at the state level only for "whites," so the 1940 figures in our analysis are for whites only.

U.S. Department of Commerce, Bureau of the Census. *United States Life Tables: 1890, 1901, 1910, and 1901-1910.* Washington, D.C.: Government Printing Office, 1921. The data are from pp. 126-67. For 1890, all states missing except MA. For 1900, all states missing except IN, MA, MI, NJ, and NY. For 1910, all states missing except IN, MA, MI, NJ, and NY. According to Pope (1992, 268)—Clayne L. Pope, "Adult Mortality in America before 1900: A View from Family Histories," 267-96 in *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel,* eds. Claudia Goldin and Hugh Rockoff (Chicago: University of Chicago Press, 1992)—"Adequate death registration procedures did not exist for all states until 1933. Until 1910, death registration coverage was concentrated in the more industrialized and urbanized states of the Northeast. Consequently, the oft-cited life tables from 1890 to 1920 based on the death registration area cannot be representative of the nation as a whole unless regional variation in mortality was unimportant by the turn of the century. Life tables constructed for periods before the development of a significant death registration area (before 1900) are limited in geographical coverage and do not, in most cases, provide evidence on mortality for long time periods."

U.S. Department of Commerce, Bureau of the Census. *United States Abridged Life Tables, 1919-1920.* Washington, D.C.: Government Printing Office, 1923. Data are provided only for "white" and "Negro," with state-by-state statistics available only for the "white" population. For the "Negro" population, life expectancy was reported for the group of "states with less than 4 per cent Negroes" and "states with more than 5 per cent

Negroes." Lacking any data for other nonwhite groups, we assumed that life expectancy for other nonwhites was better estimated by the "Negro" figure than by the "white" figure, and we calculated statewide figures for whole populations accordingly. For 1920, the following states are missing: AL, AK, AZ, AR, DE, FL, GA, ID, IA, LA, ME, MS, MT, NE, NC, NH, NM, ND, OK, RI, SD, TX, VT, WV, and WY. Note that Hawaii, not yet a state, is included.

U.S. Public Health Service, National Office of Vital Statistics, Federal Security Agency. *State and Regional Life Tables, 1939-41*. For 1940, information was only provided for white Americans, so we only used the data for states that were predominantly white. We defined states as predominantly white if, in the 1950 census, the government reported life expectancy data only for white people in the state.

U.S. Department of Health, Education, and Welfare; Public Health Service; National Office of Vital Statistics. *Life Tables for 1949-51*. Vital Statistics—Special Reports. Vol. 41, Nos. 1-5.

U.S. Department of Health, Education, and Welfare; Public Health Service. *State Life Tables: 1949-51*. Vital Statistics—Special Reports. Vol. 41 Supplement. Washington, D.C.: Government Printing Office, 1956. In 1950, the census for the first time disaggregated whites and nonwhite, but only for states with substantial nonwhite populations.

U.S. Department of Health, Education, and Welfare; Public Health Service. *United States Life Tables: 1959-61*. Public Health Service Publication No. 1252: Vol. 1, No. 1. Washington, D.C., Dec. 1964.

U.S. Department of Health, Education, and Welfare; Public Health Service. *State Life Tables: 1959-61*. Public Health Service Publication No. 1252: Vol. 2, Nos. 1-51.

U.S. Department of Health, Education, and Welfare; Public Health Service; Health Resources Administration; National Center for Health Statistics. *United States Life Tables: 1969-71*. DHEW Publication No. (HRA) 75-1150. Vol. 1, No. 1. Rockville, Md., May 1975.

U.S. Department of Health, Education, and Welfare; Public Health Service; Health Resources Administration; National Center for Health Statistics. *State Life Tables: 1969-71*. DHEW Publication No. (HRA) 75-1151. Vol. 2, Nos. 1-51. Rockville, Md., June 1975.

U.S. Department of Health and Human Services, Public Health Service, National Center for Health Statistics. *U.S. Decennial Life Tables for 1979-81*. Vol. 1, No. 1—*United States Life Tables*. Washington, D.C.: Government Printing Office, Aug. 1985.

U.S. Department of Health and Human Services, Public Health Service, National Center for Health Statistics. *U.S. Decennial Life Tables for 1979-81*. Vol. 2, Nos. 1-51—*State Life Tables*. Washington, D.C.: Government Printing Office, 1985-86.

U.S. Department of Heath and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. *U.S. Decennial Life Tables for 1989-91*. Vol. 1, No. 1—*United States Life Tables*. Washington, D.C.: Government Printing Office, 1997.

U.S. Department of Heath and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. *U.S. Decennial Life Tables for 1989-91.* Vol. 2, Nos. 1-51—*State Life Tables*. Washington, D.C.: Government Printing Office, 1998.

U.S. Department of Heath and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System. *U.S. Decennial Life Tables for 1999-2001—United States Life Tables.* National Vital Statistics Reports, Vol. 57, No. 1. 5 Aug. 2008.

U.S. Department of Heath and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System. *United States Decennial Life Tables, 1999-2001—State Life Tables.* National Vital Statistics Report, Vol. 60, No. 9. 14 Sept. 2012.

Lewis, Kristen, and Sarah Burd-Sharps. *American Human Development Report: The Measure of America, 2013-2014.* Social Science Research Council.

Illiteracy Rate

Rates referred to the population 10 years old and over in 1880 and 1890, to the population 15 years old and over between 1900 and 1930, and to the population 14 years old and over in 1950 and 1960. No rates of illiteracy were reported by the census in 1940. A 1963 report defined illiteracy as "the inability to read and write a simple message either in English or in any other language"—a lower standard than "functional illiteracy," which included not only people who were entirely illiterate but also those whose level of literacy did not allow them to read and carry out written instructions (Bureau of the Census 1963, 1).

Bureau of the Census. "Estimates of Illiteracy, by States: 1960." *Current Population Reports*, Series P-23, No. 8. Washington, D.C., 12 Feb. 1963. Includes estimates of illiteracy for 1900-1960, except 1940. According to this report, "Data on illiteracy for States were last collected in the 1930 Census. National data on illiteracy have been collected in sample surveys taken in 1947, 1952, and 1959; but these surveys did not provide figures for States." No comparable data exist for the period since 1960.

High School Graduation Rate, 1890-1940

Until the mid-20th century, the federal government did not track or report high school graduation rates. These data do not exist for earlier years. So, to extend this data series back in time, Claudia Goldin (1994) proposed a new measure—the number of each year's high school graduates, as a proportion of that state's 18-year-olds. The quality of this measure depends on the accuracy of both the numerator (number of high school graduates) and the denominator (number of 18-year-olds in the state). We find no reason to question the census's population counts, so calculating the number of 18-year-olds is straightforward: either the census reports the number directly or it reports a slightly larger group, such as people between the ages of 15 and

19, which allows us to estimate the number of 18-year-olds with reasonable accuracy. But counting high school graduates in a given year is much more difficult. High schools are local institutions, so statewide measures of high school graduates require that states or federal agencies receive accurate counts from large numbers of public high schools, private high schools, secondary programs at universities, and other institutions providing equivalent educations. As Goldin (1994, 2) points out, high school graduates were almost certainly undercounted, with small schools in rural areas less likely to report their graduates than large urban high schools.

Goldin notes that the *Biennial Survey of Education* reported, without comment, two different sets of statistics for each state for the period 1920-38—one set drawn from a survey of schools, the other drawn from a survey of state officials. Only the school survey contained numbers of graduates (the state surveys reported enrollments by grade, but not graduates each year), but Goldin concluded that the state-reported numbers were more accurate than the school-reported numbers. So she made adjustments to the school-reported numbers to align them more closely with the state numbers and to account for other idiosyncrasies in the data. After an extended and close examination of the primary data, in the interest of extending Goldin's work to other years, we realized that the state-reported numbers were internally inconsistent.[1]

Consequently, for the period 1890-1940, we rely on Goldin's numbers for the three years she estimated (1910, 1928, and 1938), while using the raw school-reported numbers for other years (1890, 1900, 1920). For the latter three years, we add together the uncorrected graduation

---

[1] To take one example, here are enrollment figures for secondary students in California in 1924, which Goldin used as her base year:

Table 7 (p. 347), Total secondary, 308,952 (public) + 11,344 (private)

Table 2/4 (pp. 352/355) Total secondary, 308,534 (public) + 9,238 (private)

Table 29 (p. 386), Total secondary, 151,568 (public)

Here are the same figures for Massachusetts:

Table 7 (p. 347), Total secondary, 128,757 (public) + 15,287 (private)

Table 2/4 (pp. 352/355) Total secondary, 128,478 (public) + 13,819 (private)

Table 29 (p. 386), Total secondary, 111,006 (public)

And the same set of figures for Oregon:

Table 7 (p. 347), Total secondary, 34,737 (public) + 1,843 (private)

Table 2/4 (pp. 352/355) Total secondary, 34,719 (public) + 1,193 (private)

Table 29 (p. 386), Total secondary, 34,719 (public)

Looking at these numbers, we see that all three tables report virtually identical numbers for Oregon—two of them listing 34,719 students in public secondary schools, one of them listing 34,737. Given Goldin's assumption that the state-reported numbers were generally accurate, this consistency between the three tables, all of them state-reported numbers, is reassuring. But that assumption breaks down as we examine the California numbers, where two tables report 309,000 students in public secondary schools and a third table reported 152,000—or for Massachusetts, where similar, though smaller, discrepancies exist between the three tables. We cannot explain these differences, given that all three tables were ostensibly drawing on identical bodies of state-reported data, and, as far as we can tell, neither Goldin nor the primary reports makes any mention of these differences.

data for 1890, 1900, and 1920 with estimated graduates from college preparatory programs in universities and colleges. Here we follow Goldin (1994, 9) in omitting normal schools and in assuming that 16% of students enrolled in these college preparatory programs went on to graduate. We do worry that the undercount of graduates from secondary schools is greater than the undercount of students in these college preparatory programs, so we use the ratios in Goldin's Table A2 (60.8% for 1890, 77.4% for 1900, 85% for 1920) to adjust the secondary school numbers upward before adding the estimated graduates from college preparatory programs.

Goldin, Claudia. 1994. "Appendix to: 'How America Graduated from High School, 1910 to 1960,' Construction of State-Level Secondary School Data." NBER Working Paper Series on Historical Factors in Long Run Growth, Historical Paper No. 57. Cambridge, Mass.: National Bureau of Economic Research.

Report of the Commissioner of Education for the Year 1891-'92. Vol. 2. Washington: Government Printing Office. Data are drawn from p. 686 (Table 1: "Summary of Statistics of Public High Schools for 1891-'92—Schools, Instructors, and Students") and pp. 688-89 (Table 2: Summary of Statistics of Endowed Academies, Seminaries and Other Private Secondary Schools for 1891-'92").

U.S. Department of the Interior, Census Office. *Report on Population of the United States at the Eleventh Census: 1890.* Vol. 1. Washington, D.C.: Government Printing Office, 1894. Data are drawn from Part II, Table 2.

Report of the Commissioner of Education for the Year 1899-1900. Vol. 2. Washington: Government Printing Office, 1901. Data are drawn from p. 2161 (Table 31: "Combined Statistics of Public High Schools and Private High Schools and Academies—Number of Schools, Instructors, and Students in 1899-1900") and p. 2162 (Table 32: "Combined Statistics of Public High Schools and Private High Schools and Academies—College Preparatory Students and Graduates in 1899-1900").

U.S. Census Office. *Twelfth Census of the United States, Taken in the Year 1900, Census Reports, Volume II—Population, Part II.* Washington D.C.: United States Census Office, 1902. Data are drawn from Table 2.

U.S. Department of the Interior, Bureau of Education. *Biennial Survey of Education, 1920-1922.* Vol. 2. Bulletin, 1924, No. 14. Washington: Government Printing Office, 1925. Data are drawn from pp. 67-68 (Table 29: "Enrollment of Pupils by Grades, 1921-22"), pp. 540-41 (Table 7: "White and Colored Pupils Enrolled in Public High Schools, 1921-22"), pp. 559-60 (Table 22: "Graduates from Public High Schools and Number of Graduates Continuing Their Education in 1921-1922"), pp. 608-9 (Table 6: "Graduates in Private High Schools and Academies and Graduates Continuing Their Education, 1921-22"), and p. 610 (Table 7: "Classification of Students Enrolled in Private High Schools and Academies by Years, 1921-22").

U.S. Department of the Interior, Bureau of Education. *Biennial Survey of Education, 1922-1924.* Bulletin, 1926, No. 23. Washington: Government Printing Office, 1927. Data are drawn from p. 841 (Table 6: "Graduates in Private High Schools and Academies and Graduates Continuing Their Education, 1923-1924").

U.S. Department of Commerce, Bureau of the Census. *Fourteenth Census of the United States Taken in the Year 1920—Volume III—Population, 1920: Composition and Characteristics of the Population by States*. Washington D.C.: Government Printing Office, 1922. Data are drawn from Table 9.

High School Graduation Rate, 1940-2009

U.S. Census Bureau. *A Half-Century of Learning: Historical Census Statistics on Educational Attainment in the United States, 1940 to 2000*. April 2006. Data are drawn from Table 5 ("Percent of the Total Population 25 Years and Over with a High School Diploma or Higher by Sex, for the United States, Regions, and States: 1940 to 2000").

U. S. Census Bureau. "Educational Attainment in the United States: 2009—Population Characteristics." Feb. 2012. Data are drawn from Table 2.

Income Per Capita

Klein, Alexander. "Personal Income of U.S. States: Estimates for the Period 1880-1910." Warwick Economic Research Papers, No. 916. Department of Economics, University of Warwick, Sept. 2009.

Lee, Everett S., Ann Ratner Miller, Carol P. Brainerd, and Richard A. Easterlin. *Population Redistribution and Economic Growth: United States, 1870-1950.* Vol. 1. *Methodological Considerations and Reference Tables.* Philadelphia: American Philosophical Society, 1957. Data are drawn from Table Y-1, p. 753.

Bureau of Economic Analysis. SA1 Personal Income Summary: Personal Income, Population, Per Capita Personal Income. Updated 24 Mar. 2016. Accessed at: http://www.bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=5#reqid=70&step=1&isuri=1

Data Sources for Party Competition and all Control Variables

These are all documented extensively in our article itself, in the section entitled "DATA SOURCES." To provide a single resource that describes all of our data, we have included that section in its entirety, in manuscript format, below:

DATA SOURCES

Our analysis proceeds in two stages, testing first the hypothesis that party competition spurs investment in human capital and infrastructure, and next the hypothesis that this investment pays off in improving the health, wellness, and prosperity of state residents. This section introduces our measures and data sources.

Party Competition

In order to analyze our contention that tightly matched parties are better able to assemble coalitions for broad-based spending, we calculate the margin of control over legislative seat shares, averaged (except in post-1935 Nebraska, which has a unicameral legislature) across a state's two legislative chambers, then subtract that margin from 100%. This measure of competitiveness can range from 100% if the two parties are evenly matched to 0% if one party holds every seat in a legislature. To calculate these margins, we use the seat shares reported by Dubin (2007). While we collected these data for every state, at 10-year intervals from 1880 to the present day, in this article we draw only on data through 1980, since we are examining the impact of party competition and spending decisions in one era on income and wellness outcomes in a future era (through 2010).

Figure 1 illustrates the trend of party competition in three states, two of which are close to ideal types. (These states, and others discussed below, are chosen only for purposes of illustration.) Between 1880 and 1980, Alabama was a one-party state, with no serious competition to the dominant Democratic party and almost no change over time. In 1940, for example, Democrats held all but one seat in Alabama's 106-person lower house and every seat in the senate, yielding a party competitiveness score that year of 0.9%. Illinois, in contrast, was a highly competitive state through almost the entirety of this period: in 1970, the two parties each had 29 seats in the senate, and Republicans had a 90–87 edge in the house—for a party competitiveness score of 99.2%. Of the eleven years in our study, Illinois was non-competitive only in 1920. Illinois, like Alabama, has experienced little variation, year-to-year, in its level of party competition.

If the whole world were Alabama and Illinois, we would have little causal leverage for this project. Our theory, of course, is that a two-party state like Illinois spends more on healthcare, education, and transportation—with better outcomes for its citizens in terms of health, literacy, and prosperity—than a one-party state like Alabama. But we have no convincing way to test that specific proposition in these two states because, alongside party competition, there exists a multitude of other factors, some measurable but many not, that distinguish Illinois from Alabama. To understand the impact of party competition, we must hold these other factors constant, and the best way to do that is to examine changes in party

competitiveness over time within a given state. That is the basis of our research design: state fixed effects allow us to hold constant consistent differences across states and focus our analysis instead on the decade-to-decade variation in levels of party competition within states.

Michigan, as Figure 1 suggests, offers a path forward—and Michigan, not Alabama or Illinois, is representative of the great majority of states. In most states, the level of party competition ebbed and flowed between 1880 and 1980. Michigan provides a rich body of data for analysis, with party competitiveness fluctuating over time, between a low in 1920 of 0%—when both chambers in the legislature were entirely Republican—to a high of 97.3% in 1970, when Democrats controlled the house by a narrow 58–52 margin, and the senate was equally split between the two parties. Figure 2 provides illustrations of some of the types represented by the American states. Some states, like Oregon and Pennsylvania, had competitive party systems in the late 19th century and in the second half of the 20th century, but were dominated by one party in the early 20th century. Others, like Kentucky and Maryland, were the mirror image, with competitiveness peaking in 1920. And still others, like Florida and Vermont, were one-party states, one Democratic and the other Republican, until the 1960s, when both became competitive. Over a century's time, there are as many patterns of changes in party competition as there are American states—and it is this very variety, above all the changes from decade to decade within each state, that provides the foundation for the causal claims we investigate.

We focus on party competition in the state legislature because that is where the collective spending and policy decisions that should directly impact state development take place. However, because executive branch officials also exert an important influence over policy, party competition for statewide offices may be an additional relevant measure. Using the dataset generously shared by Ansolabehere and Snyder (2002), we compiled the average margin of victory in the contests for U.S. Senate, all U.S. House races, governor, and any other statewide offices held in a state in a given year. This is the same measure that Besley, Persson and Sturm (2010) use in their study examining the relationship between party competition and economic growth. We include this electoral measure alongside the legislative margin of competition to hold this factor constant and to see if it adds any additional explanatory power in our models.

Control Variables

Our hypotheses about the impact of party dynamics are *ceteris paribus* expectations. Yet because all is never equal in the American states, we include in our models a set of political, economic, and demographic control variables in order to isolate the impact of party competition.

Democratic Party Strength. Spending patterns may be more a function of which party controls a state rather than the level of two-party competition. While the literature probing the effects of state party control on spending has yielded mixed findings based on the era studied or the "party cleavage types" (Dawson and Robinson 1963; Dye 1966; Hofferbert 1966; Winters 1976; Barrilleaux and Miller 1988; Schneider 1988; Brown 1995; Smith 1997), this is still an important alternative explanation for us to consider. To account for this, we construct measures of which party controlled the upper and lower houses in each legislature, based on the data reported by Dubin (2007), to determine whether Democratic control led to higher spending.

Since governors, too, are central in making state spending decisions (Kousser and Phillips 2012; Caughey, Warshaw, and Xu 2017), we record the partisanship of the governor in each state in each decade and include this measure in our regression models.

Income per capita. A primary factor in determining a state's ability to invest in governmental spending—as well as an important determinant of its rate of development, which we will discuss in the next section—is the economic resources of its residents. Indeed, state income has been a main variable in work in political science (Dawson and Robinson 1963; Dye 1966; Hofferbert 1966; Winters 1976; Barrilleaux and Miller 1988; Schneider 1988; Brown 1995; Smith 1997) and in economics (Besley and Case 2003; Michener and McClean 2003; Frank 2009; Besley, Persson and Sturm 2010) that have looked at state policy choices and growth rates. We expect that more affluent states will be able to spend more in all areas, and that economic attainment may also condition future development. Our measure of per capita income comes from Klein (2009) for 1880–1910, from Lee et al. (1957, 753) for 1920, and from Bureau of Economic Analysis (2016) for 1930–2010. To convert this measure into current (2016) dollars, we rely on McCusker (1991, 297–373) for data from 1880–1910 and on Bureau of Labor Statistics (2016) thereafter.

Foreign-Born Percentage of State Population. A consistent finding of the literature on state and local politics in America is that jurisdictions with many immigrants face particular obstacles when it comes to obtaining funding or finding legislative success (Gamm and Kousser 2013; Abrajano and Hajnal 2015). Immigration has been a flashpoint throughout American history, with different national origin groups facing discrimination but the presence of discrimination remaining a constant. To test for this, we use census data to measure the percentage of foreign-born residents in each state. For 1880 through 1990, this is reported in Carter et al. (2006). For 2000 and 2010, we use Grieco et al. (2012).

Black Percentage of State Population. At least as much as immigrants have faced discrimination, Black Americans have long seen significantly lower spending levels on social services in the states in which they are most numerous (Plotnick and Winters 1985; Grogan 1994). To control for this factor, we use the census figures in Carter et al. (2006) before 2000 and in Rastogi et al. (2011) thereafter.

Other Nonwhite Percentage of State Population. Because ethnicity along with race has been a source of discrimination in the United States, we also measure the "other nonwhite" population in the state—a census category that, depending on the year, includes Asian-American, Native American, Latino/a/x, and Hawaiian and Pacific Islander residents. Hajnal and Trounstine (2005) offer evidence of ethnic and racial biases in local spending, and Hero and Tolbert (1996), Hero (1998), Filindra and Pearson-Merkowitz (2013), Butler (2014), and Harden (2015) show how policymakers discriminate against ethnic minorities at the state level. We gather these figures from Carter et al. (2006) before 2000 and in Hoeffel et al. (2012), Hixon et al. (2012), and Norris et al. (2012) afterward.

Urban Percentage of State Population. Finally, states with dense urban populations, compared to those with a larger percentage of rural residents, face different demands for the scope of government and the types of services that the state is called upon to provide. States

with more city dwellers may also have a different capacity to develop quickly. To control for the impact of the urban share of the population, we collect demographic figures from Carter et al. (2006) for all years through 1990, from State Data Center of Iowa (2000) for 2000, and from U.S. Census Bureau (2010) for 2010.

State Spending

How much each state spends—and in what areas—is a pivotal variable in this study. For our first set of models, state spending functions as the dependent variable, as we test the hypothesis that party competition affects the amount and allocation of dollars. And, for the second set of models, state spending becomes the crucial independent variable, as we ask whether expenditures on human capital and transportation infrastructure affect the well-being of people living in the state.

We rely on two main sources of data. Our primary source is a series of reports issued by the U.S. government aggregating local and state expenditures for each state, with figures for total spending as well as figures for spending by subject area. These reports—which cover state and local spending in 1902, 1913, 1932, 1942, 1962, 1972, and 1982—were aggregated by Sylla, Legler, and Wallis (1995) into a single database, which we draw on for our analysis. To extend this time series back to the 19th century, Sylla, Legler, and Wallis (1993) scoured archives and libraries to collect budget records reported at the state level. The resulting data series is less comprehensive than the one created by the federal government, with some obvious gaps in state coverage and no data on local spending. Still, given that spending figures in the two data series are strongly correlated during the period they overlap in the early 20th century and given, too, that no other comprehensive record exists for this era, we rely on this second database for state spending in 1880 and 1890. Most of our analysis, however, relies simply on the 20th-century spending data.

Our theory links expenditures to party competition at the state level and to decisions of state legislators, but local governments have historically played powerful roles in allocating money for public goods, especially in the realm of education but also in such areas as sanitation, transportation, and public safety—and states differ considerably on this dimension (Burns 1994; Teaford 1984, 2002; Bloom 2019). Comparing statewide investments requires, then, that we aggregate, as much as possible, expenditures made at both the state and local levels. That is consistent with the practice in these federal reports. This approach is appropriate because, in the American system, state and local governments are not separate layers of government but intertwined entities, with states exercising authority over local governments through the judicial precedent of Dillon's Rule and with many states heavily funding local governments. While power and spending are devolved in different ways in different states, the federal reports account for this by aggregating total state and local spending.

These reports include money from federal transfers to state and local governments alongside own-source revenues as the funding that supports total expenditures. Since federal transfers are often not segregated in the ways that they are spent, including these dollars is necessary (and, given the nature of state spending, unavoidable) if we are to provide a

comprehensive picture of a state's investment in infrastructure and human capital. However, we need to attend to the concern that these federal transfers might be substantial in scale and correlated with levels of party competition in the state legislature. While we have no theory that would predict that federal transfers are related to a state's partisan competitiveness, if the two variables were to be positively correlated, we could not rule out the alternate hypothesis that federal transfers, rather than two-party competition, are driving investments in education, healthcare, and transportation. The federal government began reporting data on federal transfers in its 1942 report—between 1880 and 1932, before the New Deal, federal transfers to state and local governments were quite small—so we did a special analysis of the data from the 1942, 1962, 1972, and 1982 reports. We find that, even in the most recent decades, states still draw on own-source revenues for the vast majority of their spending[2] and that there is no correlation between the level of federal aid to each state and its level of party competition.[3] This finding alleviates any concern that competitive legislatures are investing more in education, health, and transportation simply because they have more federal money to spend.

As Figure 3 shows, spending on education, healthcare, and transportation have generally risen over time, in constant dollars, across all states. These parallel trends satisfy a foundational requirement of the difference-in-differences technique. Just as important, though, is the substantial variation among the states. We offer three sets of examples in Figure 3. On education spending, Idaho consistently spent more than Delaware from 1900 until 1940—though the lines on the graph are compressed, the differences, such as $173 v. $52 in 1910, were very large—then, in 1940, the lines crossed. Since then, Delaware has been the bigger spender on education, though the two states came close to converging again in 1980 after a large drop in spending in Delaware. On health spending, Figure 3 shows two states, Indiana and Minnesota, with nearly identical levels of spending from 1900 to 1960, which then diverge sharply in 1970 and 1980. And Georgia and Maine, which we illustrate for transportation spending, began at the same level, but diverged early. By 1930, Maine spent more than four times as much as Georgia on transportation, and Maine continued to outpace Georgia in this realm by a considerable amount until 1980, when Maine's spending fell sharply, in constant dollars.

This decade-by-decade variation, replicated across all the states, is crucial for our study. Coupled with variation in levels of party competition, it allows us, first, to perform within-state analysis of the impact of changes in party competitiveness on state spending decisions. And, second, that same variation allows us to engage in within-state analysis of how changes in spending affect the well-being of state residents.

---

[2] Averaging across all states, federal aid accounted for 10.8% of state revenues in 1940, rose to 14.1% in 1960, peaked at 18.0% amidst the Great Society programs going into effect in 1970, and then dipped slightly to 16.7% by 1980. It never rose above 27.1% for any state in any year.

[3] The correlation between the proportion of a state's revenues that come from federal aid and its level of legislative party competition is -0.04, falling far short of statistical significance, over this period. Calculating that correlation separately for each decade also does not yield any significant positive correlations between federal aid and party competition.

Measures of Health, Education, and Income

Identifying measures for wellness outcomes, especially at the state level and over a century's time, was a wide-ranging research challenge. To our knowledge, no comparable dataset has been assembled before. We drew on an array of sources, identified on our APSR Dataverse page, and consulted other scholars to construct datasets that document residents' well-being in these states at 10-year intervals between 1880 and 2010. For this article, we only present data in which we have great confidence: where we could not resolve significant historical problems or doubts, we do not use those data. For each variable (except high school graduation rates, which we explain below), all the data come from a consistent set of sources and, in any given year, the data for every state is drawn from exactly the same source.

Figure 4 summarizes our five primary wellness measures, with each state in each decade represented as a dot. The broad trend lines are clear: on all five wellness measures, the state of the American population has improved over time. Infant mortality and illiteracy have fallen, while life expectancy, graduation rates, and income have all risen over the last century. As was the case with state spending, the patterns apparent in this figure provide support for a key assumption of the difference-in-differences models: the assumption of parallel trends. What Figure 4 conceals, though, is state-by-state variation and year-by-year changes in individual states. Our within-state analysis requires that, just as changes in party competition and spending vary from state to state, the same variation be present in this set of dependent variables. We proceed, then, to present figures on measures of well-being that focus on one measure at a time, allowing us to isolate individual states and to compare changes in one state to those in another.

Life—Infant Mortality. While the federal census reported figures on infant mortality beginning in 1890, not until 1910 did it begin reporting data only for "registration states," where federal officials trusted the state data. The number of registration states increased from seven in 1910, to 23 in 1920, to 46 in 1930, and to all 48 in 1933. Figure 5 illustrates infant mortality rates (where data exist) over the period 1920–2010. Each state is represented by a gray line in the figure, with the lines of two states highlighted. As this figure shows, there was historically great variation in levels of infant mortality and in the rate of improvement. Mississippi and New Mexico, for example, dramatically reversed positions over time. In 1930 the infant mortality rate in Mississippi, 68 deaths for every 1,000 live births, ranked it somewhere in the middle of states. But by 1960, Mississippi's rate, at 42, was the highest in the country, and it has been the highest ever since, even as it has continued to decline. New Mexico, in contrast, which in 1930 had by far the highest rate in the country with 145 infant deaths for every 1,000 live births, now ranks in the middle of the pack. Rates in all states have fallen and converged over the last few decades, but there remain significant differences between states. Mississippi's rate in 2010, 9.6 deaths per thousand, is 70 percent higher than New Mexico's, at 5.6.

Life—Life Expectancy at Birth. The federal census reported life expectancy data for just one state (Massachusetts) in 1890, five states in 1900 and 1910, and half the states in 1920. The 1920 and 1940 state-level data were reported only for whites, and there was no 1930 data at

all.[4] Only in 1950 did the federal census begin systematically reporting life expectancy for all people in each state. Examining Figure 6, the general trend is clear: steady improvements in life expectancy at birth. But differences between states have been substantial. In 1920, life expectancy in the territory of Hawaii was just 47.5 years, the lowest in the country, and in California it was 55.8 years. By 1960, with full data reported for all states, Hawaii's rate, at 71.6, slightly exceeded California's, at 70.8. In subsequent decades, Hawaii—which had once had the lowest life expectancy in the country—has had the highest. California, which lagged behind Hawaii in the second half of the 20th century, nearly closed the gap by 2010. The variation among the states is evident throughout the whole time period, including the period since 1950 when the fullest data are available.

Literacy—Illiteracy Rates. Between 1880 and 1960, when the data series ends, the federal census reported illiteracy rates for each state's population. As Figure 7 suggests, state populations once differed dramatically on this measure. In some states, like Virginia, the rate has plummeted over time—from 40.6% (1880) to 24.3% (1900) to 3.4% (1960). In others, like New York, illiteracy rates were already low, at 5.5% in 1880 and decreased only slightly over time, to 2.9% in 1960—a small decline next to the experience in Virginia, but still a considerable improvement from the perspective of New Yorkers.

Literacy—High School Graduation Rates. To measure high school graduation rates, we draw on two different data series. For the period 1890–1940, our data estimate the proportion of 18-year-olds who graduated in a given year from high school. For the period 1950–2010, our data show the proportion of all adults who are high school graduates. In an age when high school graduation rates steadily increased—which by definition means that the proportion of young adults with high school degrees is greater than the proportion of older adults with high school degrees—the first measure is markedly higher. Both data series, separated by a gap, are illustrated in Figure 8.

Before 1940, the federal government did not collect data on the number of high school graduates in each state's adult population. So we instead report an estimated number of each year's high school graduates, as a proportion of that state's 18-year-olds. (In calculating this estimate, we draw extensively on Goldin 1994 as well as on a number of primary sources. See our APSR Dataverse page for a detailed discussion.) As Figure 8 indicates, these data show great improvement over the period 1890–1940, with all states trending sharply up. But states, on this as on other measures, varied. Ohio, which had a lower graduation rate than New Hampshire at the turn of the 20th century—with Ohio at 10.7% in 1900, and New Hampshire at 17.6%—had risen to 37.9% by 1930, with New Hampshire at 37.8% that year.

Since the middle of the 20th century, data on graduation rates is straightforward to locate: the federal government has regularly surveyed the population and reported the percentage of adults in each state who graduated from high school. These are the data we report for 1950–2010. The story here, again, is of progress across the board, but each state charted its own path. Starting by this measure at the same place in 1950, at 37%, New Hampshire and Ohio followed

[4] As we discuss on our APSR Dataverse page, we are able to draw on other data in the census to estimate life expectancy for the entire population of each state in 1920, including nonwhites, but the 1940 figures represent data only for people classified by the census as white.

different trajectories over the next 60 years. New Hampshire, which at the turn of the 20th century had more teenagers graduating from high school than Ohio, had more high school graduates in its population at the turn of the 21st century. In 2010, 91.3% of New Hampshire's adults and 87.6% of Ohio's adults had graduated from high school.

Prosperity—Income per capita. The final measure of well-being, income per capita, is drawn from the same data sources identified in the previous section. In the first stage of our analysis, we use present income as a variable to predict the level of state spending, given our expectation that states with the capacity to raise more in tax revenue are likely to spend more than other states. But, in the second stage of our analysis, we use future income as a measure of a state's success in stimulating greater prosperity after investments in human capital and infrastructure have time to pay off. Using per capita income as a measure of economic growth follows the approach of other historical works on political economy in the states, since state GDP data are not available before the 1960s (Barro and Sala-i-Martin 2004; Besley, Persson, and Sturm 2010). Figure 9 offers an example of the disparities among states on this measure. Measured in current (2016) dollars, per capita income in Connecticut, already relatively high at $6,300 in 1880, rose to $63,000 by 2010, higher than that of any other state. Montana, in contrast, buoyed by silver and copper mining in the late 19th century, began at a very high level —$10,600 in 1880—but rose to just $38,000 by 2010.