

## Technical Notes

---

### Overview

The data in this report come from surveys conducted by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF), other federal agencies, and nonfederal organizations. Users should take great care when comparing survey data from these different sources. Differences in definitions, survey procedures, and phrasing of questions, among other things, make these data less than strictly comparable. Efforts have been made to maintain consistency throughout these tables, but it has sometimes been necessary, for accuracy, to use distinct terminology that does not match that used in other tables.

### Reporting categories

#### Racial and ethnic information

Achieving consistency over time is difficult when measuring racial and ethnic characteristics of the population due to the methods of collecting and reporting race and ethnicity. First, both the naming of population subgroups and their definitions have changed over time. Second, many of the groups of particular interest are quite small, so it is difficult to measure them accurately without larger samples or surveys of the entire population of interest. In some instances, sample surveys may not have had sufficient sample size to permit the calculation of reliable racial or ethnic population estimates for all groups; consequently, data are not always shown for some groups. For example, the Bureau of Labor Statistics' Current Population Survey (CPS) does not provide data on unemployment among American Indians or Alaska Natives. Third, data on race and ethnicity are often based on self-identification, which can change based on the range of answer options and the purpose of the survey. Fourth, it is easy to overlook or minimize heterogeneity within racial or ethnic subgroups, such as Hispanics or Latinos, when only a single statistic is estimated for their entire population.

#### ***Office of Management and Budget's categories and guidelines***

In October 1997, the Office of Management and Budget (OMB) announced new government-wide standards for the collection of data on race and ethnicity ([https://obamawhitehouse.archives.gov/omb/fedreg\\_1997standards](https://obamawhitehouse.archives.gov/omb/fedreg_1997standards)) that became effective 1 January 2003. OMB specified the following categories and definitions of racial and ethnic groups:

- Black or African American: A person having origins in any of the Black racial groups of Africa.
- American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.
- Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent; for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
- Native Hawaiian or Other Pacific Islander: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific islands.
- White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
- Hispanic or Latino: A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.

Respondents can also select one or more racial designations, and those who do are classified under "more than one race."

The Department of Education published final guidance in the Federal Register on 19 October 2007 (72 Fed. Reg. 59267) to transition to the new OMB standards for reporting race and ethnicity. Previously, the Department of Education's National Center for Education Statistics (NCES) had identified mutually exclusive racial and ethnic groups as White, Black, Hispanic, Asian or Pacific Islander, and American Indian or Alaska Native. In 2008, NCES changed race and ethnicity reporting for degree completion data and for enrollment data. For the degree completion data, reporting in the new categories became mandatory for the 2011–12 data collection (i.e., 2011 data). For the fall enrollment data, reporting in the new categories became mandatory for the 2010–11 data collection (i.e., 2010 data). However, institutions were not required to update the race and ethnicity data of individuals who were already in their systems. In this report, the racial and ethnic groups detailed in Integrated Postsecondary Education Data System (IPEDS) tables, which run through 2018, incorporate OMB's new race and ethnicity reporting standards for all years for which data are provided. For more information, see <https://www2.ed.gov/policy/rschstat/guid/raceethnicity/index.html>.

In this report, NCES IPEDS data show racial and ethnic information only for U.S. citizens and permanent residents. For all other data sources, racial and ethnic information is only for U.S. citizens and permanent residents, unless the table specifies that temporary visa holders are also included.

### ***Minority-serving institutions***

High-Hispanic-enrollment institutions (HHEs) are nonprofit public and private institutions of higher education whose full-time equivalent (FTE) enrollment of undergraduate students is at least 25% Hispanic. The FTE enrollment of Hispanic students is determined by enrollment data that institutions reported to the fall 2018 IPEDS Enrollment Survey conducted by NCES. NCES determined FTE enrollment by estimating that approximately three part-time students are equivalent to one full-time student. Because IPEDS does not collect part-time credit hour information, the FTE numbers are only an approximation. The list includes only nonprofit public and private institutions of higher education.

Historically Black colleges and universities (HBCUs) are academic institutions listed by the White House Initiative on Historically Black Colleges and Universities. The Higher Education Act of 1965, as amended, defines an HBCU as "any historically Black college or university that was established prior to 1964, whose principal mission was, and is, the education of Black Americans, and that is accredited by a nationally recognized accrediting agency or association determined by the Secretary [of Education] to be a reliable authority as to the quality of training offered or is, according to such an agency or association, making reasonable progress toward accreditation." See <https://sites.ed.gov/whhbcu/one-hundred-and-five-historically-black-colleges-and-universities/>.

Tribal colleges are 32 fully accredited academic institutions on a list maintained by the White House Initiative on Tribal Colleges and Universities. See <https://sites.ed.gov/whiaiane/tribes-tcus/tribal-colleges-and-universities/>.

### **Information about people with disabilities**

For several reasons, data on people with disabilities have limitations. First, the operational definitions of disability may vary across a wide range of physical and mental impairments, and these definitions may not be comparable. The Americans with Disabilities Act of 1990 (ADA) encouraged progress toward standard definitions. Under ADA, an individual is considered to have a disability if he or she has a physical or mental impairment that substantially limits one or more of his or her major life activities, has a record of such impairment, or is regarded as having such impairment. ADA also contains definitions of specific disabilities.

Second, data on disabilities frequently are not included in comprehensive institutional records (e.g., in registrars' records in institutions of higher education). If included at all, such data may be kept only in confidential files at an office responsible for providing special services to students. Institutions of higher education are unlikely to have information regarding students with disabilities who have not requested that they be provided with special services related to their

disabilities. Whereas in elementary or secondary school programs that receive funds to provide special education, statistics on all students identified as having special needs are centrally available.

Third, information about people with disabilities that is gathered from surveys is often obtained from self-reported responses. Typically, respondents are asked to state whether they have any specified physical, mental, or sensory impairment or limitation in order to classify them as having a disability. Resulting data therefore reflect individual perceptions of their functioning, rather than more objective measures of functioning that use standardized criteria, such as those used in clinical studies of disability.

Fourth, some surveys have recently changed the wording of their questions on disability, lowered thresholds for defining disability, or both. This has resulted in an increase in the number of individuals with disability *as measured*.

Sources of data on people with disabilities cited in this report include the National Survey of College Graduates (NSCG), Survey of Doctorate Recipients (SDR), and Survey of Earned Doctorates (SED), all conducted by NCSES; and the American Community Survey (ACS), conducted by the Census Bureau. These sources are described in more detail later in this appendix; the following is a brief description of how each source treats the issue of disability.

The ACS (2019) asks a series of yes or no questions about each individual at a sampled address as to whether the person has serious difficulty hearing; seeing; concentrating, remembering, or making decisions; walking or climbing the stairs; dressing or bathing; or doing errands alone. The Census Bureau categorizes respondents with one or more “yes” responses as having a disability.

The SED (2019), SDR (2019), and NSCG (2019) all have the same series of disability questions. The respondent is asked, “What is the USUAL degree of difficulty you have with” seeing; hearing; walking; lifting 10 pounds; and concentrating, remembering, or making decisions. The response choices are none, slight, moderate, severe, and unable to do. Respondents who report “moderate,” “severe,” or “unable to do” for any activity were classified as having a disability.

## Primary data sources

This section provides summary descriptions of primary data sources and links to more detailed survey information.

### Primary NCSES sources

The following sources from NCSES were used for data tables in this publication. Published data tables from these surveys can be accessed on the NCSES website at <https://nces.gov/>. In addition, researchers may access data directly from the NCSES Interactive Data Tools (<https://ncesdata.nsf.gov/home>) or the Scientists and Engineers Statistical Data System (SESTAT) Data Tool (<https://ncesdata.nsf.gov/sestat/sestat.html>).

### Survey of Earned Doctorates

The Survey of Earned Doctorates (SED) is an annual census of individuals who earned a research doctorate from an accredited U.S. academic institution. The most common research doctorate degree is the doctor of philosophy (PhD). Recipients of professional degrees, such as the juris doctor (JD) and doctor of medicine (MD), are not included in the SED. Data are collected directly from individual doctorate recipients contacted through their university. Responses were gathered primarily through a Web-based questionnaire with a small number of responses from paper questionnaires and computer assisted telephone interviews. The recipients are asked to provide information about their field of doctoral study, educational history, postgraduate plans for work and further study, and demographic characteristics. Since the survey’s inception in 1957, more than 90% of the annual cohort of doctorate recipients has typically responded to the questionnaire each year.

For individuals who do not respond to the SED, data that are available from public sources (e.g., field of doctorate) are added to the file. No adjustments are made for nonresponse, and no imputation is used for missing items among

respondents. The data for a given year include all doctorates awarded in the 12-month period ending on 30 June of that year.

The SED is sponsored by NCSES and three other federal agencies: the National Institutes of Health, the Department of Education, and the National Endowment for the Humanities. Further information about the SED can be found at <https://nsf.gov/statistics/srvydoctorates/>.

### ***Survey of Graduate Students and Postdoctorates in Science and Engineering***

The Survey of Graduate Students and Postdoctorates in Science and Engineering, more commonly referred to as the Graduate Students Survey (GSS), is an annual census of all U.S. academic institutions granting research-based master's degrees or doctorates in science, engineering, and selected health fields as of fall of the survey year. The survey, sponsored by NSF and the National Institutes of Health, collects the total number of graduate students, postdoctoral appointees (postdocs), and doctorate-level nonfaculty researchers by demographic and other characteristics, such as source of financial support. Results are used to assess shifts in graduate enrollment and postdoctoral appointments and trends in financial support.

The survey collects data from institutions' branch campuses, affiliated research centers, medical schools, schools of nursing, and schools of public health. The 2018 survey covered 715 academic institutions. Data are collected separately for each eligible organizational unit (academic department or program, research center, or health facility).

Approximately 99% of institutions and affiliated units respond to the survey. Missing data for nonresponding units are imputed by using prior years' data, when available or by using data provided from similar units at a peer institution.

In 2017 the GSS taxonomy was revised to align with the NCSES Taxonomy of Disciplines. As a result, two fields became partially ineligible, three fields became ineligible, and several broad fields were reorganized.

In 2016, the survey included a pilot data collection designed to assess the feasibility of (1) reporting master's and doctoral student data separately, (2) using Classification of Instructional Programs codes issued by the NCES for reporting GSS data, and (3) expanding the use of file uploads for data submission. Data provided by pilot coordinators are included in the 2016 data products.

In 2014, the survey frame was updated following a comprehensive frame evaluation study that identified potentially eligible but not previously surveyed U.S. academic institutions. A total of 151 newly eligible institutions were added, and two private for-profit institutions offering mostly practitioner-based graduate degrees were removed as no longer eligible. For more information, see <https://www.nsf.gov/statistics/srvygradpostdoc/>.

Due to the 2014 methodological changes and other changes that have occurred in recent cycles, care should be used when assessing trends within GSS data.

Reporting of race and ethnicity since 2008 is likely to have been affected by changes in reporting in IPEDS. Starting in 2008, IPEDS respondents were asked to use a new race and ethnicity classification that included a category for persons who are not Hispanic, a category for persons who identify with more than one race, and a category for Native Hawaiians and Other Pacific Islanders, separate from Asians. The new classification was optional in 2008 and 2009 IPEDS but mandatory in 2010, and it may have contributed to a significant increase in GSS reporting of "Not Hispanic or Latino, more than one race" within GSS data.

Further information about the GSS can be found at <https://www.nsf.gov/statistics/srvygradpostdoc/>.

### ***National Survey of College Graduates***

The National Survey of College Graduates (NSCG) is a repeated cross-sectional survey conducted biennially since the 1990s that provides data on the nation's population of college graduates, with a particular focus on those in the science and engineering (S&E) workforce. The survey samples individuals who are living in the United States during the survey

reference week, have at least a bachelor's degree, and are under the age of 76. This survey is a unique source for examining various characteristics of college-educated individuals, including occupation, work activities, salary, the relationship of degree field and occupation, and demographic information.

The 2019 NSCG includes over 92,500 respondents, and the overall response rate is 68%, representing a population of about 65 million college graduates living in the United States. Of these college graduates, an estimated 36 million are classified as scientists and engineers. These are individuals with a bachelor's or higher-level degree educated or employed in a S&E or S&E-related field. Individuals not included in the survey frame for the 2019 NSCG are U.S. educated scientists and engineers earning degrees after 31 December 2017 and foreign-educated scientists and engineers who came to the United States after 31 December 2017.

The NSCG classifies the following broad categories as S&E occupations: computer and mathematical scientists, life and related scientists, physical and related scientists, social and related scientists, and engineers. Postsecondary teachers are included within each of these groups. The following are considered S&E-related occupations: health and related occupations; S&E managers; S&E precollege teachers; S&E technicians and technologists, including computer programmers; and other S&E-related occupations, such as architects and actuaries. All other occupations are non-S&E occupations. Among the largest are non-S&E managers, non-S&E teachers, social services and related occupations, and sales and marketing occupations. Further information on the NSCG can be found at <https://www.nsf.gov/statistics/srvygrads/>.

### ***Survey of Doctorate Recipients***

The Survey of Doctorate Recipients (SDR) is a repeated cross-sectional survey conducted biennially since 1973 that provides demographic and career history information about individuals with a research doctoral degree in a science, engineering, or health (SEH) field from a U.S. academic institution. The survey, sponsored by NCSSES and the National Institutes of Health, follows a sample of SEH doctorate holders throughout their careers from the year of their degree award until age 76. The panel is refreshed each survey cycle with a sample of new SEH doctoral degree earners. Results are used to make decisions related to the educational and occupational achievements and career movement of the nation's doctoral scientists and engineers.

For the 2019 SDR, all 2015 and 2017 sampled members who remained age eligible were retained for the 2019 cycle. The 2015 sampled members who did not respond in both the 2015 and 2017 surveys were dropped from the 2019 sample. As with prior survey cycles, a sample of 10,000 new graduates who had earned their degrees from 1 July 2015 to 30 June 2017 were added. An additional sample of 14,564 SEH doctoral degree holders eligible for the 2015 sample but not previously selected were also added to support an overall sample of 120,000 cases. The 2019 sample response rate was 69%, and the sample represented approximately 1,148,800 U.S.-trained research doctorate recipients under 76 years of age. Beginning with the 2017 SDR, field of study reporting was revised and updated to better align with the NCSSES Taxonomy of Disciplines, which more closely aligns with the Classification of Instructional Programs issued by the NCES.

Further information on the SDR is available at <https://www.nsf.gov/statistics/srvydoctoratework/>.

### ***Early Career Doctorates Survey***

The Early Career Doctorates Survey (ECDS) is sponsored by NCSSES and the National Institutes of Health. The survey gathers in-depth information about individuals who earned their first doctoral degree (PhD, MD, or equivalent) in the past 10 years in order to better understand the labor market and research and employment opportunities for early career doctorate holders. These individuals are employed in federally funded research and development centers (FFRDCs) and U.S. master's- and doctorate-granting academic institutions, excluding affiliated medical schools and centers. The ECDS collects details about demographics; professional activities and achievements; professional and personal life balance; mentoring, training, and research opportunities; and career paths and plans. The survey includes individuals with doctoral

degrees earned in any field and any country, and it covers all types of positions (e.g., postdocs, junior faculty, nonfaculty researchers, and other staff). The sample size is 256 institutions and 15,465 individuals.

Further information on the ECDS is available at <https://www.nsf.gov/statistics/srvyecd/>.

## Primary non-NCSES sources

The following non-NCSES sources were used for data tables in this report.

### ***The Integrated Postsecondary Education Data System: Fall Enrollment, Completions, and Institutional Characteristics Survey Components***

National Center for Education Statistics, Department of Education

The Integrated Postsecondary Education Data System (IPEDS) is a collection of survey programs that surveys all U.S. postsecondary institutions, including universities and colleges and the institutions that offer technical and vocational education. Starting in 1992, the completion of all IPEDS surveys is mandatory for all institutions that participate in or are applicants for participation in any federal financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended. IPEDS comprises several integrated survey components. These surveys obtain information about types of institutions where postsecondary education is available, student participants, fall enrollments, programs offered and completed, graduation rates, and the human and financial resources involved in the delivery of postsecondary education. In this report, data are primarily drawn from the IPEDS Fall Enrollment Survey and the IPEDS Completions Survey, which is administered to all institutions offering degrees at the bachelor's degree level and above, 2-year institutions, and less-than-2-year institutions.

NCES changed degree-level categories in the IPEDS Completions Survey in fall 2008, but reporting in the new categories was optional for 2008 and 2009 data. Reporting in the new degree-level categories was mandatory for the 2010–11 IPEDS Completions collection. Before 2008, the post-baccalaureate degree categories were master's, first professional, and doctor's. With the 2008 changes, the category first professional degree is no longer used. Programs and awards in that category (e.g., medicine, law, pharmacy, and theology) have been reclassified as either master's degrees or as one of three types of doctor's degrees: doctor's-research/scholarship, doctor's-professional practice, or doctor's-other. Numbers reported here for 2008 and 2009 doctoral degrees combine doctor's degrees reported by institutions using the pre-2008 reporting categories and doctor's-research/scholarship degrees reported by institutions using the 2008 reporting categories. Data for 2010 include only doctorates reported as doctor's-research/scholarship.

Data from the IPEDS Completions Survey includes Title IV institutions in the 50 states, the District of Columbia, and other U.S. jurisdictions. The IPEDS Completions data cover all awards granted between 1 July and 30 June. The year of the data indicates the end of the academic year in which the degrees were awarded. This report also uses data from the Fall Enrollments Survey component, which provides a snapshot of the enrollment at an institution for a specific time in the fall. For the IPEDS Fall Enrollment Survey, institutions with traditional academic year calendar systems (semester, quarter, trimester, or 4-1-4) report their enrollment as of 15 October or the official fall reporting date of the institution. Institutions with calendar systems that differ by program or allow continuous enrollment report students that are enrolled at any time between 1 August and 31 October; these institutions are typically for-profit institutions. Enrollment numbers reported are as of that date in the indicated year.

Further information on the IPEDS is available at <https://nces.ed.gov/ipeds>.

### ***Current Population Survey***

Bureau of Labor Statistics, Department of Labor

The Current Population Survey (CPS) is a monthly household survey conducted by the Census Bureau for the Bureau of Labor Statistics. It provides data on employment and unemployment by age, sex, race, and a variety of other characteristics, and it is the source of the monthly official U.S. unemployment rate. Estimates calculated from the CPS reflect the civilian noninstitutional population ages 16 and older. CPS gathers information from approximately 60,000 households monthly through personal and telephone interviews. Basic labor-force data are gathered monthly; data on special topics are gathered in periodic supplements. Consecutive monthly estimates are often averaged to produce quarterly or annual average estimates.

Further information on the CPS is available at <https://www.bls.gov/cps/>.

### ***Enterprise Human Resources Integration-Statistical Data Mart***

Office of Personnel Management

The Office of Personnel Management (OPM) provides estimates of federally employed scientists and engineers through its Enterprise Human Resources Integration Statistical Data Mart (EHRI-SDM). The data cover most executive branch agencies and some legislative and judicial branch agencies. Coverage is limited to federal employees with at least a bachelor's degree and is subject to change over time. For example, the State Department stopped providing data on Foreign Service personnel in 2006 and stopped providing all data in 2015. In 2017, the question on disability was changed, and the new wording resulted in an increase in the number of federal employees with disabilities as *measured*.

More information on OPM's estimates of federally employed scientists and engineers is available at <https://www.opm.gov/policy-data-oversight/data-analysis-documentation/>. Information about EHRI-SDM is available at [https://www.fedscope.opm.gov/datadefn/ae\\_hri\\_sdm.asp](https://www.fedscope.opm.gov/datadefn/ae_hri_sdm.asp).

### ***American Community Survey***

Census Bureau

The American Community Survey (ACS) is an ongoing survey that produces annual estimates of the U.S. population as well as various demographic, labor force, and housing characteristics. The survey was designed to allow for estimates of small geographical areas that previously were only possible using decennial census data.

Further information on the ACS is available at <https://www.census.gov/programs-surveys/acs/>.

## **Sampling and nonsampling errors**

The data from all the sources used for this report are subject to error. For non-census survey programs (NSCG, SDR, ECDS, CPS, ACS), accuracy is determined by the joint effects of sampling and nonsampling errors. Sampling errors arise because estimates based on a sample differ from figures that would have been obtained if a complete population had been surveyed. The sample selected for any survey is only one of a large number of possible samples of the same size and design that could have been selected. Even if all other aspects of the survey remained fixed, such as the questionnaire and instructions, the estimates from each sample would differ. This variability, termed sampling error, occurs by chance and is measured by the standard error associated with a particular estimate.

The standard error of a sample survey estimate measures the precision with which an estimate from one sample approximates the true population value, and it can be used to construct a confidence interval for a survey parameter to assess the accuracy of the estimate. For further information on sampling error sources and its impact on the survey estimates, see each survey's website.

Nonsampling errors can arise from design, reporting, and processing errors, as well as from errors due to nonresponses or faulty responses. These errors can occur in data from sample surveys, from census surveys (SED, GSS, IPEDS), and from administrative data (OPM). Nonsampling errors include respondent-based events, such as some respondents

interpreting questions differently from other respondents, respondents making estimates rather than giving actual data, and respondents being unable or unwilling to provide complete, correct information. Errors can also arise during the processing of responses, such as during recording and keying. Nonsampling errors are difficult to measure, and estimates of nonsampling errors are not available for data in this report.