Technical Documentation

2017 Nonemployer Statistics by Demographics

Overview

The Nonemployer Statistics by Demographics series (NES-D) provides information on the demographic characteristics of nonemployer businesses. The NES-D is the result of a research project by the Census Bureau to complete the picture of U.S. business ownership by demographics for the United States. Historically, the quinquennial Survey of Business Owners (SBO) provided the only comprehensive source of information on both employer and nonemployer businesses by demographic characteristics of the business owners. In 2017, the SBO was replaced by the Annual Business Survey (ABS). The ABS is an annual survey that collects demographic characteristics from employer businesses. However, the ABS excludes the collection of demographic data from nonemployer businesses. The NES-D was developed to produce similar estimates as ABS on owner demographics for nonemployer businesses. The NES-D is not a survey; rather, it leverages existing individual-level administrative records to assign demographic characteristics to the universe of nonemployer businesses. Demographic characteristics including sex, ethnicity, race, veteran status, owner age, place of birth, and U.S. citizenship are assigned to nonemployer business owners.

Together, the NES-D and the ABS will continue to provide the only source of detailed and comprehensive statistics on the scope, nature and activities of all U.S. businesses by the demographic characteristics of the business owners. NES-D data will be available annually by detailed geography and industry levels, receipt-size class, and legal form of organization (LFO).

Connect with us
@uscensusbureau

For more information:
2020CENSUS.GOV

APPROVAL ID: CBDRB-FY20-390

Shape
your future
START HERE >

United States®
Census
2020

Data Sources and the Nonemployer Universe

NES-D is created from a variety of administrative records (AR) and Census Bureau data sources that include the Business Register (BR), Internal Revenue Service (IRS) tax Form 1040 data, tax Schedule K-1 data, Decennial Census and American Community Survey (ACS) data, Social Security Administration Numident data, and AR from the Department of Veterans Affairs (VA). The Census Bureau identifies and extracts the universe of nonemployer businesses from the BR. The nonemployer universe is comprised of businesses with no paid employment or payroll, annual receipts of $1,000 or more ($1 or more in the construction industries), and filing IRS tax forms for sole proprietorships (Form 1040, Schedule C), partnerships (Form 1065), or corporations (the Form 1120 series). The BR also provides the LFO of the business as well as its receipts, industry classification, and geography classification. For more information on how nonemployer businesses are identified and defined, visit the Nonemployer Statistics technical documentation page: https://www.census.gov/programs-surveys/nonemployer-statistics/technical-documentation/methodology.html

The primary source of data for race and Hispanic origin information is Decennial Census and ACS data, with the Census Numident serving as a secondary source. To assign race and Hispanic origin responses, priority is given to the most recent data from Decennial and ACS data; that is, first, to post-2011 ACS data, then the 2010 or most recent Census, followed by 2001-2010 ACS data, and finally Census 2000. Whenever an owner cannot be assigned a race or Hispanic origin by Decennial or ACS data, then the Numident is used to assign a race or Hispanic origin to that owner (for additional details on this topic, see https://www2.census.gov/ces/wp/2019/CES-WP-19-34.pdf).

In adherence to the revisions to Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting issued by the Office of Management and Budget, NES-D (as well as other business demographics surveys, including the ABS) does not include a "Some Other Race" or SOR race category. For those individuals with an SOR race in Decennial or ACS data, NES-D first tries to find a non-SOR race value in older decennial, ACS data or the Numident. If a non-SOR race value is not found, the race category is imputed.

Shape
your future
START HERE >

United States®
Census
2020

Following the legacy SBO and the ABS, NES-D does not include a "multiple race" category for individuals indicating they are of multiple races. Instead, for owners who report multiple races in decennial or ACS data, and are tabulated as "multiple race," NES-D uses the detailed Census or ACS race information to assign the owner to each of the corresponding racial categories. For example, an owner who reports as white and American Indian and Alaska Native (AIAN) will be assigned and tabulated to both the white race category and the AIAN race category. For this reason, summed totals in NES-D tables for owner race and firm race will be greater than the summed totals for binary demographic categories such as Hispanic origin.

The Census Numident is the primary source for the age, sex, place of birth, and U.S. citizenship status of the business owner, with Decennial and ACS data used as a secondary source. Finally, the Department of Veteran Affairs (VA) USVETS data provide AR on veteran status. Title 38 of the U.S. Code of Federal Regulations gives VA the authority to determine veterans' status. Luque et al. (2019) (https://www2.census.gov/ces/wp/2019/CES-WP-19-01.pdf) provides a discussion of VA's data, how the concept of a veteran captured by the SBO/ABS questions is broader than VA's veteran (official) definition, and how Department of Defense (DOD) data could potentially be used in NES-D in the future as an additional source that may be able to complement VA's data to bring the AR-based definition closer to the survey-based veteran concept.

Business Owner Identification and Demographics Linkage

Anonymized unique individual identifiers are used to identify owners and attach demographic characteristics from the data sources above to nonemployer business owners. The anonymized unique identifiers are assigned to individuals in AR and census data sources upon receipt of data at the Census Bureau. These identifiers are known as Protected Identification Keys or PIKs and are used as linking keys across data sources to obtain demographic information and attach those demographic characteristics to owners of nonemployer businesses.

Depending on the LFO of the business, two IRS forms are used to obtain PIKs: IRS Form 1040 for sole proprietors, and Schedule K-1 for owners of partnerships and S-corps. For C-corporations, there is no tax form or business registry that clearly and unequivocally identifies all owners of this type of business. For this reason, C-corporations are not included in the 2017 NES-D. Research is currently underway to explore whether demographics can be reliably imputed for C-corporations. C-corporations constitute only about 2 percent of the nonemployer universe and approximately 4 percent of receipts.

Connect with us
@uscensusbureau

For more information:
2020CENSUS.GOV

Shape
your future
START HERE >

United States®
Census
2020

APPROVAL ID: CBDRB-FY20-390

The administrative and census records data sources mentioned above provide demographic characteristics coverage for the vast majority of identified nonemployer business owners (not including owners of C-corporations). Sex, age, place of birth, and U.S. citizenship are available for approximately 98 to 99 percent of identified owners. Data for Hispanic origin are available for about 95 percent of all records and about 90 percent for race.

Whenever missing, demographic characteristics are imputed using the same imputation method as the Annual Business Survey (see https://www.census.gov/programs-surveys/abs/technical-documentation/methodology.html for treatment of nonresponse). For more details on NES-D administrative records coverage and related issues, see https://www2.census.gov/ces/wp/2019/CES-WP-19-34.pdf.

Firm-level Demographic Assignment

Assigning demographic characteristics to owners of sole proprietorships, and by extension to the firms themselves, is straightforward. Only individuals can own sole proprietorships, and each sole proprietorship has only one owner. Hence if the PIK of the sole proprietor can be linked to a given demographic data source, then the sole proprietor's firm will be assigned that demographic characteristic.

For partnerships and S-corps, the assignment of demographic characteristics to the firm as a whole is more complicated since these types of firms can have more than one owner, and not all owners are necessarily individuals. Following the ABS and legacy SBO, NES-D assigns firms to demographic groups by determining the total share of firm ownership held by individual members of each (demographic) group. A firm is assigned to a given group if owners of that group collectively own a majority stake (more than 50 percent) in the firm. NES-D uses ownership share information in Schedule K-1 data to determine what demographic group holds a majority stake in the firm (see https://www2.census.gov/ces/wp/2019/CES-WP-19-34.pdf for additional details). Those characteristics that have only two categories at the individual level (e.g., sex, Hispanic origin or veteran status) also have a third category at the firm level: equally-owned. For characteristics that have more than two individual-level categories, such as race, it is possible that no one group will collectively own a majority of the firm. In such cases, the firm is not assigned to any race category.

Shape
your future
START HERE >

United States®
Census
2020

NES-D also provides minority-owned, nonminority-owned, and equal minority-nonminority-owned categories based on the race and Hispanic origin of the owners. Specifically, individuals who are non-Hispanic white are considered to be part of the nonminority group. Also see the Tabulation section of the ABS methodology: https://www.census.gov/programs-surveys/abs/technical-documentation/methodology.html.

Not all firms are eligible for demographic classification; these firms are labeled as "unclassifiable". Following the methodology of the ABS (and legacy SBO), i) only firms where the person with the largest ownership share owns at least 10 percent of the firm are eligible for demographic assignment, ii) up to 4 owners with the largest ownership shares in the firm are considered in the assignment, iii) only person owners are used in the estimation, and hence iv) only firms with person owners are used in the calculation.

To make nonemployer owner-level and firm-level demographics estimates consistent, and also consistent with (ABS) employer estimates, only the top 4 owners of classifiable firms are included in the calculations.

Industry Classification of Firms

The industry classifications of firms are based on the 2017 North American Industry Classification System (NAICS) (https://www.census.gov/eos/www/naics/). The source of industry codes is primarily from Internal Revenue Service filings and are self-classified by tax filers. In scope are all NAICS sectors except those classified in the following NAICS industries:

Crop and Animal Production (NAICS 111 and 112)

Rail Transportation (NAICS 482)

Postal Service (NAICS 491)

Monetary Authorities-Central Bank (NAICS 521)

Funds, Trusts, and Other Financial Vehicles (NAICS 525)

Private Households (NAICS 814)

Public Administration (NAICS 92)

The 2017 NES-D provides data at the 2-digit and 3-digit NAICS level. The Census Bureau is researching if additional detail can be published in future years. For more information on the classification of industries, refer to the nonemployer statistics methodology: https://www.census.gov/programs-surveys/nonemployer-statistics/technical-documentation/methodology.html#par_textimage_36648475.

Connect with us
@uscensusbureau

For more information:
2020CENSUS.GOV

Shape
your future
START HERE >

United States®
Census
2020

Geographic Classification of Firms

The 2017 NES-D provides data for the U.S., States, and Metropolitan/Micropolitan Statistical Areas (MSA). Most geography codes are derived from the business owner`s mailing address identified from administrative records. Because the owner's mailing address may not be the same as the physical location of the business, the resulting geography codes do not always represent where business is actually conducted, but this represents the best information available regarding the location of the business. The Census Bureau is researching if additional detail can be published in future years.

Disclosure

In accordance with U.S. Code, Title 13, Section 9, no data are published that would disclose the operations of an individual business. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied. (Approval ID: CBDRB-FY20-390.)

Disclosure avoidance is the process used to protect the confidentiality of data provided by an individual or firm. Using disclosure avoidance procedures, the Census Bureau modifies or removes the characteristics which may disclose confidential information. Although it may appear that a table shows information about a specific individual or business, the Census Bureau has taken steps to mask or suppress the original data while making sure the results are still useful.

NES-D uses noise infusion as the primary method of disclosure avoidance for receipts. Noise infusion perturbs data values prior to tabulation by applying a random noise multiplier to receipts data. Disclosure protection is accomplished in a manner that causes the vast majority of cell values to be perturbed by, at most, a few percentage points. Each published cell value has an associated noise flag indicating the relative amount of distortion in the cell value resulting from the perturbation of the data for the contributors to the cell. In certain circumstances, some individual cells may be suppressed for additional disclosure avoidance. Suppressed data are replaced by one of the following symbols:

N - Not available or not comparable

S - Withheld because estimates did not meet publication standards

X - Not applicable

Connect with us
@uscensusbureau

For more information:
2020CENSUS.GOV

Shape
your future
START HERE >

United States®
Census
2020

APPROVAL ID: CBDRB-FY20-390

The level of noise applied to the receipts are identified by the following symbols:

G - Low noise: the cell value was changed by less than 2 percent by the application of noise

H - Moderate noise: the cell value was changed by 2 percent or more but less than 5 percent by the application of noise

J - High noise: the cell value was changed by 5 percent or more by the application of noise

The following rounding rules are applied to the number of nonemployer firms:

Round count cells that are less than 15 to "<15"

If N is between 15 and 99, round to the nearest 10

If N is between 100-999, round to the nearest 50

If N is between 1,000-9,999, round to the nearest 100

If N is between 10,000-99,999, round to the nearest 500

If N is between 100,000-999,999, round to the nearest 1,000

If N is 1,000,000 or more, round to four significant digits

Reliability of Data

The data sources used to produce these estimates are of the highest quality, and well-grounded in a body of proven administrative records research that shows the quality and suitability of those data sources to directly replace demographic information in business, as well as household, surveys. The NES-D are derived from AR data and are not subject to sampling error. Therefore, there is no relative standard error or standard error due to sampling.

The data compiled for NES-D are subject to non-sampling errors, which can be attributed to many sources. For instance, administrative records data may contain measurement error because of issues such as coverage problems (e.g., the data source may not cover certain populations as well as others); linking or matching issues which may lead to bias problems; conceptual and timing misalignments; reporting errors; definition and classification difficulties; errors in recording or coding the data obtained; and other errors of coverage, processing, and estimation for missing or misreported data. In the case of NES-D, coverage and bias problems are not as pronounced because nonemployer business owners are well represented in tax and other administrative and census records data. The accuracy of tabulated data is determined by the joint effects of the various non-sampling errors. Precautionary steps were taken in all phases of the processing to minimize the effects of non-sampling errors. For a detailed discussion of this topic, see the Limitations and Challenges section of the NES-D working paper: https://www2.census.gov/ces/wp/2019/CES-WP-19-34.pdf.

Connect with us
@uscensusbureau

For more information:
2020CENSUS.GOV

Shape
your future
START HERE >

United States®
Census
2020

APPROVAL ID: CBDRB-FY20-390

Tabulation

Tabulations were produced for the final data and are linked to the ABS website (https://www.census.gov/programs-surveys/abs/data/tables.html). Data are tabulated by the sex, ethnicity, race, and veteran status of the firm owners. Business ownership is defined as having 51 percent or more of the stock or equity in the business and is categorized by firms classifiable by sex, ethnicity, race, and veteran status and firms unclassifiable by sex, ethnicity, race, and veteran status. Firms classifiable by sex, ethnicity, race, and veteran status are categorized by the following:

Sex

Female

Male

Equally male/female

Ethnicity

Hispanic

Non-Hispanic

Equally Hispanic/non-Hispanic

Race

White

Black or African American

American Indian and Alaska Native

Asian

Native Hawaiian and Other Pacific Islander

Minority (classified as any race and ethnicity combination other than non-Hispanic and white)

Nonminority (classified as non-Hispanic and white)

Equally minority/nonminority

Veteran status

Veteran

Nonveteran

Equally veteran/nonveteran

Connect with us
@uscensusbureau

For more information:
2020CENSUS.GOV

Shape
your future
START HERE >

United States®
Census
2020

APPROVAL ID: CBDRB-FY20-390

Comparability to the Survey of Business Owners

NES-D and SBO estimates are not directly comparable due to differences in survey and administrative records responses, non-sampling error or other issues such as definitional differences between the survey and AR data, and allowable survey responses for sole proprietorships that do not have a parallel in tax data. The highlights of these differences are described below. For a detailed discussion of this topic, see the 'Comparison to SBO' section of the following working papers:

https://www2.census.gov/ces/wp/2019/CES-WP-19-34.pdf

https://www2.census.gov/ces/wp/2019/CES-WP-19-01.pdf

Regarding race, i) the SBO included a "Some-Other-Race" category (which is no longer allowed in business statistics or surveys), and ii) AR research finds that agreement rates for race between AR and survey responses are very high, but tend to be lower for small population groups (e.g., American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander) relative to other race groups.

Regarding firm ownership by sex, the SBO response allowed for sole proprietorships to be equally owned by a man and a woman (usually married couples) while tax records can only consider the sex of the person that appears as the owner of the sole proprietorship on the 1040 tax Form. Consequently, the AR-based firm equally-owned category is expected to be, and is, lower than the SBO estimate. Note though that for a large share of nonemployer sole proprietors, the 2012 SBO already used AR for direct substitution of core demographics including sex. This resulted in lower 2012 SBO equally-owned estimates for nonemployer businesses than in the 2007 SBO.

Regarding firm ownership by veteran status, as mentioned earlier, the concept of veteran captured by the SBO/ABS is broader than VA's (official) definition of a veteran. Specifically, VA's veteran definition does not include some military personnel such as individuals who are currently on active military duty and individuals serving in the National Guard/Reserve Component who never served on active duty in the past. In addition, also as mentioned earlier, some older and healthier veterans are less well represented in VA's data. For these reasons, AR-based estimates are expected to be, and are, lower than SBO estimates.

Contact Information

For assistance with these data, please email the program area at
adep.annual.business.survey@census.gov or call 301.763.3316 between 7 a.m. and 5 p.m. (EST), Monday
through Friday.