Notes on Data and Methods

The 2021 Small Business Profiles for the States provide a description of the status and contributions of small businesses in the United States, in each individual state, and in the District of Columbia. The 2021 edition has been revised and updated to include new data sources, new statistics, and new data visualizations. The following sections provide technical information that may be helpful in interpreting the statistics or reproducing the analyses.

Definition of small businesses

Small businesses are generally defined in the profiles as firms with fewer than 500 employees. However, in the data used for the business dynamics feature on page 1 of each profile, the association of establishments and employment with small businesses is based on the number of employees linked with an Employer Identification Number (EIN).[1] Since a single firm may use multiple EINs, some of the establishments and employment associated with small businesses in that feature would not be associated with small businesses in other features.

Software

The profiles were created using R and LaTeX. Data visualizations were created primarily using the ggplot2 package, and the final document was generated using knitr. Numbers in the text were rounded using the round_half_up command from the janitor package. Numbers ending in 5 were rounded up, and final zeros after the decimal indicate the precision preserved under rounding. For example, 1.05 would have been rounded to 1.1, while 1.04 would have been rounded to 1.0.

Sources of data

All data used in the 2021 edition of the state small business profiles are publicly available and published online. Hyperlinks to sources are provided in the source notes for each feature. Years indicated in the source notes are the years the data describe, which may not match the naming convention used by the data publisher. Where no year is listed, data from multiple years were used.

The demographics features on page 3 debut in these profiles data from the Annual Business Survey and from Nonemployer Statistics by Demographics, two new publications from the Census Bureau. The Annual Business Survey collects demographic information about employer businesses.

---

[1] US Bureau of Labor Statistics, "Business Employment Dynamics Size Class Data: Questions and Answers," last modified December 15, 2005, https://www.bls.gov/bdm/sizeclassqanda.htm.

Nonemployer Statistics by Demographics provides estimates of the number of nonemployer businesses by demographic group derived by the Census Bureau using administrative data.[2]

Imputation

Values were imputed for some features when published data were missing, obviously erroneous, or otherwise problematic. A source note that begins "Source:" indicates that statistics in a feature are as reported by the source, with no imputation. A source note that begins "Source of data:" indicates that the feature includes imputed values for at least one profile.

Maps on page 1 showing small business share of employment by county were based primarily on 2017 Statistics of US Businesses data. Where data for a county were insufficient to calculate the share, if the missing data were available for that county for any of the four preceding years, data from the most recent year were carried forward. If data were not available for a recent year, shares were imputed by averaging the shares of neighboring counties. If shares for some neighboring counties were also missing, the share for the state as a whole was used as a substitute for those counties in the calculation. Imputed values were calculated using the spdep package.

The business ownership shares by demographic group feature on page 3 also includes imputed values. The counts for some categories were often missing, but the ownership shares could still be calculated using the data available. For example, a count may have been published for all businesses classifiable by owner sex and for businesses with majority ownership by men and for businesses with majority ownership by women, but not for businesses owned equally by men and women. The number of businesses for which women were half or majority owners would have been calculated by subtracting the number of businesses with majority ownership by men from the number of businesses with ownership classifiable by sex.

In the same feature, the share of businesses owned by racial minorities was approximated by subtracting the number of businesses with majority ownership by Whites from the total number of businesses with ownership classifiable by race. This method was used instead of adding the counts for each separate racial minority group because the racial categories are not mutually exclusive and because the counts for some racial minority groups were often missing. The share reported is therefore an estimate of the share of businesses with half or majority ownership by racial minorities and without simultaneous majority ownership by Whites.

The last feature involving imputation is the time series figure on page 4 showing total employment by business size. Where the value for a year was missing or reported as zero, the value was imputed by carrying forward the latest valid value from a preceding year.

---

[2] US Census Bureau, "Technical Documentation: 2017 Nonemployer Statistics by Demographics," accessed on August 6, 2020, https://www2.census.gov/programs-surveys/abs/data/2017/2017%20NESD%20Technical%20 Documentation_final.pdf.

Effects of the COVID-19 pandemic

Data on small businesses are generally only available with a substantial lag relative to the period they describe. The data analyzed in these profiles predate the COVID-19 pandemic. Early research indicates that the effects of the pandemic on small businesses have been substantial and have varied by industry, geography, and demographic group. Research from the Office of Advocacy describing the impacts of the pandemic on small businesses can be found at https://advocacy.sba.gov/tag/covid-19/.

Contact

For additional information, contact the SBA Office of Advocacy at advocacy@sba.gov.