

Technical Notes

Definition of small business

A single firm may encompass multiple establishments if it operates at multiple locations. Small businesses are generally defined in these profiles as firms with fewer than 500 employees across all establishments. However, in the data used for the business dynamics feature on the first page of each state profile, the association of establishments and employment with small businesses is based on the number of employees linked with an Employer Identification Number (EIN). 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.

Data

All data used in these profiles are publicly available and published online. The data and associated technical documentation can be accessed using the hyperlinks 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.

Software

The profiles were created using R with formatting through HTML and CSS. Data visualizations were primarily created using the ggplot2 package in R. Tables were primarily created using the knitr and gt packages. Elements were combined into an HTML document using the knitr package. A PDF document was created from the HTML document using Prince XML.

Rounding

Numbers in these profiles have been rounded if they include a decimal point, with the number of digits after the decimal point indicating the precision preserved under rounding. Numbers were rounded using the round_half_up command from the janitor package in R, which rounds up numbers ending in 5. For example, 1.05 would be rounded to 1.1, while 1.04 would be rounded to 1.0.

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 largely predate the COVID-19 pandemic. 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 effects of the pandemic on small businesses can be found at advocacy.sba.gov/tag/covid-19.

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 the first page of each state profile showing small business share of employment by county were based primarily on 2018 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 in R.

The business ownership shares by demographic group feature on the third page of each state profile may include 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 time series figure on the fourth page of each state profile showing total employment by business size category may include imputed values. Where a value for a size category was missing or reported as zero, the value was imputed by carrying forward the latest valid value for that category from a preceding year.

The time series figure in each territory profile showing total employment by establishment size category may also include imputed values. Where a value for a size category was missing, that value was sometimes imputed by subtracting the values reported for the other categories from the total for all categories. When that was not possible, if a value for a preceding year was available, the missing value was imputed by carrying forward the latest preceding value. When that also was not possible, the missing value was imputed by carrying backward the earliest value.

Contact

Any additional information that may be helpful for interpreting or reproducing the analysis is available by request. The SBA Office of Advocacy can be contacted at advocacy@sba.gov.