

Data Cleaning for Preliminary Models

Sebastian Rojas Cabal

7/3/2022

How was the data cleaned?

Control Data

GDP by State

- State GDP corresponds to the Real GDP by state with NAICS classification. Table number: SAGDP9N.
 - Source: <https://apps.bea.gov/regional/downloadzip.cfm>
 - For info on NAICS classification: <https://www.bea.gov/cautionary-note-about-annual-gdp-state-discontinuity>
- Quantities are in chained 2012 millions of US Dollars
 - “Chained-dollar values of GDP by state are derived by applying national chain-type price indexes to the current dollar values of GDP by state for the 21 North American Industry Classification System -based industry sectors. The chain-type index formula that is used in the national accounts is then used to calculate the values of total real GDP by state and real GDP by state at more aggregated industry levels. Real GDP by state may reflect a substantial volume of output that is sold to other states and countries. To the extent that a state’s output is produced and sold in national markets at relatively uniform prices (or sold locally at national prices), real GDP by state captures the differences across states that reflect the relative differences in the mix of goods and services that the states produce. However, real GDP by state does not capture geographic differences in the prices of goods and services that are produced and sold locally.” From: <https://www.bea.gov/news/2021/gross-domestic-product-state-4th-quarter-2020-and-annual-2020-preliminary>
- Quantities were converted to hundreds of thousands of US Dollars.

Nonprofit Data

- Clean nonprofit data stored as **anti_sample_220703.csv** and **nonanti_sample_220703.csv**
- Cases where for Foreign Expenses, Foreign Services or Total Expenses were negative (lower than zero) were dropped from the sample.
- Cases where Total Expenses were missing (NA) were dropped from the sample.
- Cases where Foreign Expenses, Total Expenses, or Foreign Services were missing (NA), they were converted to zero (0).
- Cases where the Proportion of Foreign Expenses relative to the Total Expenses was larger than 1.02 (102%) were dropped from the sample.
- In cases of duplicated filings, only the most recent tax filing was kept. The slice_max function was used.
- Dollar amounts were converted into 100,000 USD.
- Dollar amounts were converted into constant or real 2013 dollars based on the deflation method used by the BLS.
 - Info from the method and the relevant CPIs
 - <https://www.bls.gov/cpi/tables/supplemental-files/historical-cpi-u-202203.pdf>

- <https://www.bls.gov/cpi/factsheets/cpi-math-calculations.pdf>)
- A variable was added for the year marriage equality (not domestic partnerships or civil unions) was passed in each state.
 - Source: https://www.lgbtmap.org/equality-maps/marriage_relationship_laws
- A variable was added that is equal to zero (0) for every year in which same-sex marriage was not legal in the nonprofit's state. It is equal to one (1) for every year that it was legal.
- Data goes from 2013-2020