

Data Cleaning for Preliminary Models

Sebastian Rojas Cabal

7/3/2022

How was the data cleaned?

- 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.