

Duplicate Records and Cross-Year Identifier Inconsistency in Census Year Service Industry Data

Lucas Threinen

Macalester College

CES Technical Notes

CES-TN-2019-01

In the process of our data analysis, we uncovered facts about Census data files that are likely to be of interest to the Census Bureau and other researchers. We also developed code to correct data inconsistencies and errors, particularly in the 2002 Census of Services (CSR). These facts are described below, as are our methods for dealing with the associated data problems.

Relevant Datasets: CSR, SSL

Required Dataset Permissions: CSR 2002, SSL 2002

CES Technical Notes may contain confidential data and, thereby, disclosure is prohibited. The opinions, conclusions and suggestions expressed in these notes are those of the authors and do not necessarily represent those of the CES or the U.S. Census Bureau. Researchers (CES staff, SSEs and other Census employees) who wish to access the full text of a technical note or submit their own technical note should contact CES.Technical.Notes.List@census.gov.