Read me: Detailed data description

1. Data file naming rule:

Every single CSV spreadsheet contains employment in one occupation at the census tract level in the contiguous 48 states and the District of Columbia in the United States. Each table is named as

EST_##-####_XXXXX

In the name, 'EST' means 'estimates;' '##-####' represents a six-digit standard occupational classification (SOC) code; 'XXXXX' represents the occupation title associated with the SOC code. For example, 'EST_11-1011_Chief Executives' means the estimated employment at census tract level for 11-1011 Chief Executives.

2. Columns' definition in the file:

In each file, there are ten columns as explained in the following table.

OCC_CODE	The SOC code of the occupation
OCC_TITLE	The title of the occupation associated with the SOC code
STATE	The state where the census tract is.
GEOID	The Federal Information Processing Standard (FIPS) code of the census tract.
BLS_EMP_TOT	The total employment of the occupation in the corresponding state.
EST_BASE	The estimated employment of the occupation in the corresponding census tract, based on the Base Method.
EST_FULL_STATE	The estimated employment of the occupation in the corresponding census tract, based on the Full Method with the CPS sociodemographic sample at the state level.
EST_FULL_DIVISION	The estimated employment of the occupation in the corresponding census tract, based on the Full Method with the CPS sociodemographic sample at the census division level.
EST_FULL_REGION	The estimated employment of the occupation in the corresponding census tract, based on the Full Method with the CPS sociodemographic sample at the census region level.
EST_FULL_NATION	The estimated employment of the occupation in the corresponding census tract, based on the Full Method with the CPS sociodemographic sample at the national level.

3. Python codes and original data

We provide python codes and original data for calculating the estimated employment at census tract levels. They are:

occupational_downscaling.ipynb	Python Jupyter notebook file
state_M2017_dl.xlsx	2017 Occupational Employment and Wage Statistics (OEWS) data
cps_00023.csv.gz	Original dataset: Current Population Survey (CPS) data
nhgis0001_ds245_20195_tract.tab	Original dataset: 2015-2019 5-Year American Community Survey (ACS) data: Sex by Age by Race
nhgis0001_ds244_20195_tract.tab	Original dataset: 2015-2019 5-Year American Community Survey (ACS) data: Sex by Occupation for the Civilian Employed Population 16 Years and Ove
field_descriptions.xlsx	2017 Occupational Employment and Wage Statistics (OEWS) data description
nhgis0001_ds245_20195_tract_codebook.txt	American Community Survey (ACS) data codebook: Sex by Age by Race
nhgis0001_ds244_20195_tract_codebook.txt	American Community Survey (ACS) data codebook: Sex by Occupation for the Civilian Employed Population 16 Years and Over
state_divison_region.xlsx:	Crosswalk between three geographic units in the US: states, census divisions, and census regions
soc_census_crosswalk.xlsx	crosswalk between SOC codes and Census occupation codes

People can use the Python codes in the Python Jupyter Notebook file and the original datasets to replicate our results.