***Documentation for Life Expectancy (LE) by CZ Preparation (1990-2019)***

**Updated 8/3/2022 by Isabel De Ramos**

*Data*

* Data source for 1990-2019 mortality data
  + Requested through NAPHSIS/NCHS. Cleaned by Ran Li & Isabel de Ramos, saved as mortXXX.rdatas in ‘NCHS Mortality Data/Clean/Raw’
* Data source for 2000-2019 population data
  + used bridged\_pop\_county\_asrhLT.csv which has county (FIPS) level data by age group, (0,1,5,10…) sex, race (H, NHB, NHW, NHAIAN, NHAPI) , and year (1990-2019)

*Code*

* The code to prepare data for LE by CZ is CZ\_LE\_prep.R
* Prepare mortality data
  + clean\_nchs() finds total deaths by age group, race, year, sex, fips
  + clean\_nchs1() fixes problematic FIPS (see ‘/Data/missing pop denoms/missingpops\_solutions.docx’ for greater detail)
    - removes non-continental FIPS
      * Alaska, America Samoa, Guam, Northern Mariana Islands, Puerto Rico, Virgin Islands
    - removes foreign residents
      * fips = “00000” and “00999” and missing fips (manually checked RAW data, any missing resident fips belonged to US territories e.g. PR, VI, GU, etc.)
    - fixes FIPS that were renamed over time
      * Dade County 12025 renamed as Miami-Dade County 12086
      * Clifton Forge city 51560 absorbed into Alleghany County 51005
      * Yellowstone National Park Territory 30113 and Park 30067 grouped and combined into Gallantin 30031
  + Final NCHS mortality object should have following variables of interest:

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Original variable name in .csv** | **Description** |
| **year** | death\_year | four digit year |
| **fips** | res\_fips\_effective18 | 5 digit fips (state+county) |
| **age\_5yr\_group** | age\_5yr\_group | Age group capped at 85. 0: 0-4, 5: 5-9, 10: 10-14, etc. |
| **sex** | sex | 0=Male, 1=Female |
| **race** | hispanic\_bridged | H, NHB, NHW, NHAIAN, NHAPI |
| **count** |  | total deaths |

* Prepare population data
  + clean\_popdenoms() selects variables of interest - age group, race, year, sex, fips
  + clean\_popdenoms1() fixed problematic FIPS (see ‘/Data/missing pop denoms/missingpops\_solutions.docx’ for greater detail)
    - fixed FIPS that were renamed over time
      * Shannon County 46113 renamed as Oglala Lakota County 46102
      * Clifton Forge city 51560 absorbed into Alleghany County 51005
      * Yellowstone National Park Territory 30113 and Park 30067 grouped and combined into Gallantin 30031
  + Final population denominators object should have following variables of interest:

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Original variable name in .csv** | **Description** |
| **year** | year | four digit year |
| **fips** | fips | 5 digit fips (state+county) |
| **age\_5yr\_group** | age\_5yr\_group | Age group capped at 85. 0: 0-4, 5: 5-9, 10: 10-14, etc. |
| **sex** | sex | 0=Male, 1=Female |
| **race** | hispanic\_bridged | H, NHB, NHW, NHAIAN, NHAPI |
| **pop\_county** | pop\_county | total population |

* Final preparation for LE by CZ
  + crosswalk\_cbsa\_cz() has three main functions
    - merges mortality and population data
    - delineate April 2018 CBSA classification
      * used fips\_msa\_xwalk.csv (see ‘/Crosswalks/US/Clean/fips\_msa\_xwalk.csv')
    - adds commuting zone delineation
      * used county\_FIPS\_CZ\_crosswalk.csv (see ‘/Crosswalks/US/Clean/county\_FIPS\_CZ\_crosswalk.csv')

*Final Products*

* Results from crosswalk\_cbsa\_cz() stored into an object and exported as one .rdata (see ./Clean/1990\_2019\_cbsa\_mortality.rdata’)
* The following variables are present the object:

|  |  |
| --- | --- |
| **Variable name** | **Description** |
| **fips** | 5 digit fips (state+county) |
| **year** | four digit year (2000, …., 2019) |
| **age\_5yr\_group** | Age group capped at 85. 0: 0-4, 5: 5-9, 10: 10-14, etc. |
| **sex** | 0=Male, 1=Female |
| **race** | H, NHB, NHW, NHAIAN, NHAPI |
| **count** | total deaths |
| **pop\_denom** | total population |
| **cbsa0418** | five digit cbsa code based on April 2018 delineation |
| **cz** | commuting zone code |