***Documentation for Life Expectancy (LE) NCHS Mortality Data Preparation (2000-2019)***

**Updated 12/13/2021 by Isabel De Ramos**

*Data*

* Data source for 2000-2019 mortality data
  + Requested through NAPHSIS/NCHS.
  + ­2005-2017
    - raw data originally uploaded by Usama (see ‘/Mortality/Documentation/ Bilal NAPHSISProjectReviewForm\_revised.doc’)
    - cleaned by Ran (see ‘/Mortality/Data/mortality\_code.R’) using get\_mortality\_05\_19(), and uploaded as mortXXXX.csv’s in ‘/MS\_thesis\_project/Clean’
  + 2000-2004, 2018-2019
    - raw data originally uploaded by Usama
    - cleaned by Isabel (see ‘/Mortality/Data/mortality\_code.R’ and ‘/Mortality/Data/mortality\_functions\_IDR.R’) using get\_mortality\_05\_19() for 2018-2019, and uploaded as mortXXXX.csv’s in ‘/MS\_thesis\_project /Clean’
* Data source for 2000-2019 population data
  + Originally obtained from Census Bureau, cleaned and coded by Ran (see ‘/MS\_thesis\_project /Data/Population Denom ASHR Protocol.docx’)
  + used bridged\_pop\_county\_asrh.csv which has county (FIPS) level data by age group, sex, race (H, NHB, NHW, NHAIAN, NHAPI) , and year (2000-2019)

*Code*

* The code to prepare data for LE calculations and analysis is LE\_data\_prep.R
* Prepare mortality data
  + clean\_nchs() found total deaths by age group, race, year, sex, fips
  + clean\_nchs1() fixed problematic FIPS (see ‘/Clean/missingpops\_solutions.docx’ for greater detail)
    - removed non-continental FIPS
      * Alaska, America Samoa, Guam, Northern Mariana Islands, Puerto Rico, Virgin Islands
    - removed 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.)
    - fixed 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 ‘/Clean/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 calculations
  + crosswalk\_metro\_regions() has three main functions
    - merge mortality and population data
    - delineate urbanicity classification
      * used fips\_urbanrural\_xwalk.csv (see ‘/MS\_thesis\_project/Crosswalks/US/Clean/FIPS Urban-Rural Documentation.docx’)
    - delineate census region classification
      * used fips\_censusregion\_xwalk.csv (see ‘/MS\_thesis\_project/Crosswalks/US/Clean/FIPS Census Region Documentation.docx’)

*Final Products*

* Results from crosswalk\_metro\_region() stored into an object and exported as one .rdata (see ./Clean/00\_19\_nchs\_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 |
| **metro6** | 1=large central metropolitan, 2=large fringe metropolitan, 3=medium metropolitan, 4=small metropolitan, 5=micropolitan, 6=noncore |
| **metro2** | 1=metropolitan, 0=nonmetropolitan |
| **census\_region** | 1=Northeast, 2=Midwest, 3=South, 4=West |