1. GEOID
   1. GEOID should be a character variable with str\_length=11.
   2. Current data set uses variable name, census\_tract\_fips, for GEOID.
   3. leading "0" in GEOID is lost if a csv file is opened using excel.
2. Before data merge, process data set from each data source separately,
   1. Adding mode="wb" for windows downloading in function download.file(). E.g., download.file(url, destfile=destfile, mode = "wb").
   2. Data at different levels (tracts, single county, county subdivision) have to be consisted to the census tract level
      1. For data at single county level, map to tract level by merging county level data using state+county FIPS with a mapping file containing state+county FIPS and census tract FIPS as variables (MUA and food insecurity data).
      2. For data at county subdivision level, since there is no nested relationship between county subdivision and census tracts, the mapping requires tract-level spatial overlays for the whole country.
   3. Consist across data sources the year of census tract fips used. Data from current data sources are consisted to the 2010 census tracts before merging (mrfei data are mapped from 2000 to 2010 census tracts).
   4. Files read from local folder: Food insecurity data from feedingamerica.org. Data file downloading has to be requested.
   5. SVI data set: -999 and -999.0 are used as NA.
   6. Medically underserved areas (MUA):
      1. Limit to “Medically Underserved Areas”, excluding “Medically Underserved Populations.”
      2. Keep only MUA/P Status designated or proposed for withdrawal.
3. During data merge,
   1. Left\_join data sets from different sources to 2010 tigris::tracts, excluding the following areas: United States Virgin Islands, Commonwealth of the Northern Mariana Islands, Guam, American Samoa, and Puerto Rico.
4. After data merge,
   1. Save the final data set in both rds and csv format.
   2. Map the final data set from 2010 to 2020 census tracts based on census tract crosswalk. All variables need to be recalculated for 2020 census tracts using weights.
   3. Keep numeric representations of dichotomous variables. E.g.,

census\_tract\_fips\_2020 census\_tract\_fips\_2010 AREALAND\_PART AREALAND\_TRACT\_20 weight\_inverse mua

1 06085504424 06085504417 8591682 25455452 0.338 1

2 06085504424 06085504420 16847398 25455452 0.662 0

> subset(data2020, census\_tract\_fips =='06085504424')$mua\_num

[1] 0.3375183

* 1. Handle NA: usda\_low\_food\_access\_flag and mua were set to 0 if NA in the final data set.

1. Census tract crosswalk:
   1. Census tract crosswalks describe how U.S. census tracts from one census year correspond to census tract from another year. The tracts could split or merge. Use the calculated weight inverse to calculate the new data values for the new tracts.
   2. Weight inverse calculation: weight\_inverse = AREALAND\_PART / AREALAND\_TRACT\_20

|  |  |
| --- | --- |
| AREALAND\_TRACT\_20 | Total Land area of 2020 tract in square meters |
| AREALAND\_PART | Calculated Land area of the overlapping part in square meters |

Source: 2020 Census Tract to 2010 Census Tract Relationship File Layout (https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/2020-comp-record-layout.html)

* 1. New data value calculation based on weight inverse:
     1. Split: remain the same for the new census tracts.
     2. Merge: a weighted sum of the merged tracts.
     3. Minor boundary corrections / refinement (weight inverse < 0.05): not considering the refinement in new data value calculation by filtering out changes with weight inverse less than 0.05. After filtering, adjust weight inverses, so that the weights within each 2020 tract totaling 1.