Data Processing Guide

Preparing the datasets:

1. Download .zip Mortality Multiple Cause Files From: <https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple>
2. Download .PDF User's Guide Files From: <https://www.cdc.gov/nchs/nvss/mortality_public_use_data.htm>
3. Unzip Mortality Multiple Cause Files To My OneDrive Folder:

Parsing the datasets:

* Prerequisites:
  + Anaconda: <https://www.anaconda.com/distribution/>
  + Python files:
    - CJ\_parser.py
      * This file converts the CDC multiple cause files from space delimited to comma separated. Comma separated values(CSV) are easier to work with.
    - Concatenate.py
      * Merges the 20 sequelae variables
      * Removes the spaces in the column names
      * Creates CSV files for a merged MCOD
      * Creates CSV files for drug overdose death ICD10 codes.

1. Run CJ\_parser.py, replacing the file names with the appropriate names from the extracted data.
2. Run concatenate.py, replacing the file names with the appropriate names from the CSV files that CJ\_parser.py created. Steps in data handling
   1. Create PANDAS dataframe using CSV output from parser.
   2. Strip column names from parser
   3. Verify column names
   4. Concatenate RA1:RA20 into a combined ICD-10 multiple cause of death column
   5. Export “Cleaned” dataframe to new CSV file with the structure : “year\_all\_mortality\_concat.csv”
      1. I.e. 2012 deaths would look like “2012\_all\_mortality\_concat.csv”.

Find only overdose deaths using CDC case definition

Overdose Deaths for Each Year, starting in 2012

Concatenated .CSV files; contain additional column called “all\_MCOD” which merges RA1:RA20 fields into one giant text string with hypes separating it:

Code:

Concatenate

Parsed .CSV Files

Names: VS12MORT.csv, where x is the last digit of the year.

Parse Files to CSV

Unzip files

CDC-Raw-Space-Separated-Values

Names: VS1xMORT.DUSMCPUB, where x in the last digit of the year.

CDC-Mortality Multiple Cause Files in .ZIP format

Names: mort201xus.zip, where x is the last digit of the year.