Using R intervalaverage pkg to constructing 7-year average exposure windows for PM2.5, NO2 and O3 for the MAP cohort

1. Background info & decisions points for air pollution (AP) exposure window construction:
   * AP exposure measure availability: every-two-week average concentration measures available for each participant:
     + NO2 & O3: 01 January 1990 through 31 December 2019
     + PM2.5: 01 January 1999 through 30 June 2017
   * Date of death:
     + Due to data anonymity requirements, date of death was only able to be estimated to within a year of actual date of death.
     + For consistency of exposure window construction across AP measures, and given a) ambiguity of date of death, and b) PM2.5 available only through 30 June 2017:
       - All participants were assigned 01 July as day and month of death, with seven-year exposure windows spanning from 01 July seven years prior to year of death through 30 June of participant year of death
   * R intervalaverage pkg: <https://cran.r-project.org/web/packages/intervalaverage/vignettes/intervalaverage-intro.html>
2. Need: AP datasets with start and end dates for individual pollutant measures (each row is 2- week AP measure for a given participant ID)
   * ap\_exp\_wins\_no2\_o3\_pm25.sas - includes these steps:
     + 1) subset RADC data to a) MAP cohort only, b) with age of death (autopsied) \*/
     + 2) merge MAP data with AP data (2 sets of data: one with no2/o3 exps, one with pm2.5 ) \*/
     + 3) code the following dates: \*/
     + --> a) date of death: death\_yr-MM-DD \*/
     + --> b) start date of 7 year exp window: date of death - 7 yrs (i.e. 07-01-xxxx) \*/
     + --> c) end date of 7 year exp window: 1 day prior to DoD (i.e. 06-30-xxxx) \*/
     + 4) subset datasets to just those inclusive of the 7 yr exp window
   * PM2.5 measures dealt with separately from NO2 and O3 given different span of measures available (NO2 and O3 subsetted together since measure span same dates)
   * subset so only rows inclusive of 7-year exp window prior to death are included
   * data includes only rows inside of each participant’s 7-year window prior to death
     + no2/o3: 01 Jul 1990 earliest start date, latest date: 30 Jun 2019
       - Note that due to rounding of date of death, 01 jul 1991 is the earliest year of death we have
     + pm25: 1999-07-01 earliest start date, 2017-06-30
   * export as csv over to R (both no2 and o3 in the same dataset)
3. in excel (or R): create datasets with periods we want *intervalaverage* to generate averages for
   * make sure start and end date columns are same name here as in the exposure file(s)
   * format dates as YYYY-MM-DD
   * no2/o3:  a row for each 7 year period (a separate file for no2, and one for o3):
     + no2\_start         no2\_end
     + 1991-07-01    1998-06-30 --> because of rounding, our earliest start is jul 1991
     + 1992-07-01    1999-06-30
     + .........
     + 2012-07-01    2019-06-30
   * pm25
     + pm25\_start    pm25\_end
     + 1999-07-01    2006-06-30
     + 2000-07-01    2007-06-30
     + .........
     + 2010-07-01    2017-06-30
4. Construct average exposure windows using intervalaverage in R
   * R code file: AP\_expwins\_no2o3pm25.R
   * Includes the following steps:
     + 1) import & prep exposure window data (using datasets created in #1 above)
     + 2) import and prep table of averaging periods (created in #2 above)
     + 3) run intervalaverage by group/id
     + 4) subset the intervalaverage results to non-NA for each of the AP measures (as there’s only one valid 7-year window calc for each person)
     + 5) boxplot the 7 yr avg exposures by year