

## Data Wrangling Assignment Part 2

Submit your responses in a PDF file including your code and results. You can use Python or R. Do ask for help and if you have any questions.

1. Create a dataframe called 'exposome' using the exposome dataset and answer the following questions. Refer the Codebook file to find the field and values. You will use the datasets and dataframe from Part 1 of this assignment.
  - a. Add a new column to the dataframe exposome called 'cotinine\_combination' which combines all possible values of cotinine in mother (hs\_cotinine\_mcat\_None) and cotinine in child (hs\_cotinine\_cdich\_None). Create values of your choice in this new field and provide a count for each of these combined values. (15 Points)
  - b. Using the exposome and phenotype dataframes, plot 'cotinine\_combination' and the patients having doctor diagnosed asthma. 15 Points)
2. Please use the file 'NO2\_2020\_California' to answer the following questions. You can use Python, R, or other programming languages that you are familiar with:
  - a. How many unique county names are in this file? After checking how many unique counties in this file, please choose a reasonable plot to show the distribution of counties. You can use either matplotlib or seaborn. (15 Points)
3. Please use the file 'NO2\_2020\_California' to answer the following questions:
  - a. Find the geo-coordinates for 'Martin Luther King High School' with address '9301 Wood Rd, Riverside, CA 92508'. (5 Points)
  - b. Find the nearest monitor by distance in the file 'NO2\_2020\_California' to the above school. (10 Points)
  - c. Next, find the monthly average, monthly minimum and monthly maximum for NO2 for this monitor for the year 2020. (15 Points)
  - d. Plot the monthly average, monthly minimum and monthly maximum NO2 on a graph using matplotlib or seaborn. (15 Points)