Data Wrangling Assignment

In this assignment, we are interested in exploring the effects of the exposome on pediatric asthma. Towards that we will assess and wrangle the provide data by answering the following questions. You will use the codebook, covariates, exposome and phenotype files. 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. Using the covariates dataset, group the data by sex "e3_sex_None". Which gender has the highest weight "hs_c_weight_None '' and what is its corresponding mean value? (20 Points)

2. Create a dataframe called "phenotype" using the phenotype dataset and answer the following questions:

   a. Count the number of missing values per column in the above dataframe. (20 Points)

   b. Extract and Print all individuals (IDs) from the dataframe what have a Body mass index categories at 6-11 years old - WHO reference of <u>Overweight</u> or <u>Obese</u>. Refer the Codebook file to find the field and values. (20 Points)

   c. Identify any records (IDs) that are duplicative in the records you extracted in Step b. (20 Points)