597PR. A1– Intro data analysis using fundamental data structures

Download the two data files from this Assignment. They are called “co2_hawaii.txt” and “co2_alaska.txt”. These are from the U.S. National Oceanic and Atmospheric Administration (NOAA). They contain over 40 years of measurements of carbon dioxide levels (CO_2) in the atmosphere, taken from instruments at Barrow, Alaska and Mauna Loa, Hawaii.

View and familiarize yourself with their contents and format, but make sure you do not modify the input files while viewing them. There are detailed comments in the data files explaining the meaning of each column in the data.

- Time yourself as you work on this program. When finished, you’ll be asked how long you worked on it.
- If after reading the instructions you think you can complete it with ONLY the standard Python library (NOT using Pandas or Numpy for example), then go ahead! The “CSV” and “Collections” packages might be helpful. The whole list is at https://docs.python.org/3/library/
- OR if you’re not that comfortable with Python yet, then use any general-purpose programming language of your choice (whatever you’re strongest in). Do not use a spreadsheet application, SQL, SAS or SPSS – those are not general-purpose languages. But C++, Java, PHP, Perl, Ruby, Go, Lisp, JavaScript, etc. are examples that are all fine. If using R, try it without DataFrames (which are similar to Pandas).

Write a program to automatically perform the following data processing steps:

1. Read the two unmodified files and load all the data values into some kind of appropriate array, list, dictionaries, or similar data structure(s).
2. Parse the Quality Control Flags to know which rows have invalid data that must be ignored in calculations. Typically, you’ll see “-999.99” in place of the co2 value, but there may be other kinds of invalid rows, so determine validity from the “qcflag” column only.
3. Calculate the following items PER YEAR, and output a new text file called “co2_report.txt” containing a table with year down the side and each column heading indicated, for both locations:
    a. “MAX_LEVEL”: The highest CO_2 daily level recorded.
    b. “MEAN_LEVEL”: The mean (average) of daily CO_2 levels recorded.
    c. “%CHANGE”: Percent change of “Mean level” compared to previous year. Calculating percent change is always simple subtraction and division. Example: if one row has value 50 and the next value is 52, then the change is +4%. (52 – 50) / 50 = 0.04 = 4%
4. If you can, *try* to arrange the output data like the following and round the mean CO2 values to 2 decimal places. If you can’t figure out how to list the Alaska and Hawaii columns side-by-side, then print them separately:

```
     ALASKA                        HAWAII
YEAR MAX_LEVEL MEAN_LEVEL %CHANGE MAX_LEVEL MEAN_LEVEL %CHANGE
1973
1974
1975
...
2015
2016
```

5. Lastly, have the program calculate the MEAN of the annual “%CHANGE” column for each location and add those results below the table.

Once your program is finished, submit your program source code file(s) and the data output file to Canvas. You can use a tool in Canvas to spot check your calculations.