Assignment 3 (Maximum: 80 marks)
Due date: 22-04-2021

Temperature readings of several years are provided in the file 'tdata.csv' for several thousand stations (each station is identified by a latitude and a longitude). The assignment is to find (1) year-wise minimum temperature across all the stations for each year and (2) global minimum across all stations and all years.

You must use 1 process to read the entire data from a file named 'tdata.csv' using sequential I/O calls (for e.g., open/fopen, read/fread, close/fclose etc.). You must distribute the data to all the processes, you're free to select any data distribution strategy. Every process may find its local minimum. This depends on the way data has been distributed. You may select any strategy to parallelize, keep performance in mind. You must read the file (any one process may read the file) immediately after MPI_Init. You must time your entire code (start timer after reading the file and stop timer before MPI_Finalize(), report the maximum time across all processes). Show the scaling of your code on 1 and 2 nodes with 1, 2 and 4 cores per node (i.e. a total of 6 process counts).

The entire code must be written using C+MPI. Name your source code as 'src.c'. The output should contain three lines (no extra lines to be printed). The first line should contain the yearly minimum values in a csv format (number of columns = number of years), the second line should contain the global minimum (1 value) and the third line should contain the time (stop – start, maximum across all processes). The output file name must be 'output.txt'. You must include a 'report.pdf' detailing the data distribution strategy you used and why, you must also describe your code in the report, as well as your observations regarding the scaling and speedup. Please include the plot in report.pdf.

Input file description: The file contains a header. The first two columns in this file are latitudes and longitudes. The next 41 columns show the temperature readings for 41 years (1960 – 2000). You must not assume anything about the number of rows and columns in the file. Rows may be in the range of a few hundred thousand lines.

The provided file is just an example, we'll test with other data files.

Assignment3 directory should necessarily contain
- src.c (should take any only the filename as an input argument)
  -- we should be able to run your code using "make; mpicc –np X ./code tdata.csv"
- report.pdf
- run.sh (should run on 6 process counts as specified above)
- Makefile

You will be graded out of 5 for progress on git (start early!) and the remaining 75 for coding (understandable and documented), report (readable and clear) and plots (complete and clear).