Overview

--------

Save the attached file to your workspace and unpack the files. You should now have a folder, coding-data-exam, with three sub-folders:

wx\_data/ Weather data files

yld\_data/ Yield data files

answers/ Empty folder where you will place your answers to the questions on this exam

src/ Empty folder where you will place your source code that generated the answers

You can use whatever software tools you would like to answer the exam problems. Each problem will require some development and analysis from you. Read the instructions carefully-- you will be graded for your ability to provide the precise answer to these problems.

Weather Data Description

------------------------

The wx\_data folder has files containing weather data records from 1-Jan-1985 to 31-Dec-2014. Each file corresponds to a particular weather station from Nebraska, Iowa, Illinois, Indiana, or Ohio.

Each line in the file contains 4 records separated by tabs:

1. The date (YYYYMMDD format)

2. The maximum temperature for that day (in tenths of a degree Celsius)

3. The minimum temperature for that day (in tenths of a degree Celsius)

4. The amount of precipitation for that day (in tenths of a millimeter)

Missing values are indicated by the value -9999.

Yield Data Description

----------------------

The yld\_data folder has a single file, US\_corn\_grain\_yield.txt, containing a table of the total harvested corn grain yield in the United States measured in 1000s of megatons for the years 1985 - 2014.

Problem 1

---------

Write a software program to calculate for each weather data file the number of days in which the maximum temperature and minimum temperature data are present but the precipitation data is missing. The program should create one line of output for each of the weather data files. Each line of output should have the filename (e.g. USC00339312.txt) and the number of days you have identified. Separate the columns by a tab. Write the output in ascending order by filename and store it in the file MissingPrcpData.out in the answers folder.

Problem 2

---------

Write a software program to calculate the following values for every year for every station, ignoring missing data:

Average maximum temperature (in degrees Celsius)

Average minimum temperature (in degrees Celsius)

Total accumulated precipitation (in centimeters)

The program should create one line of output for each of the weather stations. Each line of output should have the following information separated by tabs: the filename, the year, the average maximum temperature to 2 decimal places, the average minimum temperature to 2 decimal places, and the total accumulated precipitation to 2 decimal places. If a value cannot be calculated assign a value of -9999.00 when writing your output. Store your results in the file YearlyAverages.out in the answers folder, sorted in ascending order by filename.

Problem 3

---------

Write a software program that tabulates how often each year from 1985-2014 had the highest average maximum temperature, highest average minimum temperature, and highest total accumulated precipitation from the set of weather stations. Use the results of Problem 2 as input to your program. The resulting output should indicate how frequently a given year had the highest values for the 167 weather stations present in wx\_data.

Output should be written to the file YearHistogram.out in the answers folder. The file should contain 4 columns: the year, the count of how many weather stations had that year as the highest annual average maximum temperature, a count of how many weather stations had that year as the highest annual average minimum temperature, and a count of how many weather stations had that year as the highest total annual precipitation. Separate each column by tabs and sort the output in ascending order by year. In addition, create a histogram of the output and save the result as YearHistogram.png.

Problem 4

---------

Use the annual average maximum temperature, annual average minimum temperature, and total annual precipitation results from Problem 2 and calculate the Pearson correlation between these variables and the grain yield data stored in US\_corn\_grain\_yield.txt. Output the correlations for every station. Write your output to the file Correlations.out in the answers folder. The columns of the file should be separated by tabs and will contain the filename of the weather station (e.g. USC00339312.txt) and the three correlation values to 2 decimal places. Sort the output in ascending order by filename.

Conclusion

----------

When you have completed the exam create a compressed archive of the answers/ and src/ folders and email it to the hiring manager or your contracting representative.