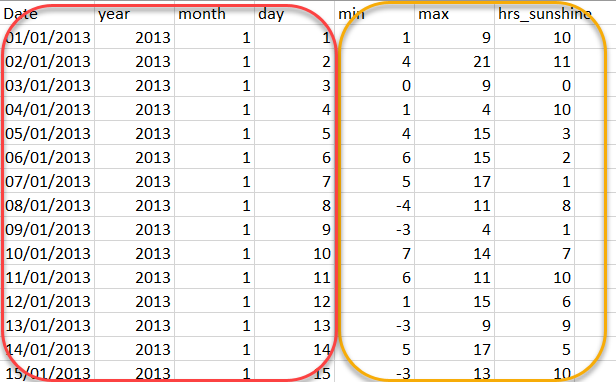
# Scenario

The data\_folder folder contains a set of files with recorded sunshine information in them.

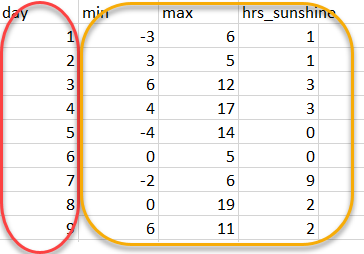
There are some files which contain a full years worth of daily records. The year is indicated in file name. The file ‘y2013all.csv’ looks like this;



The first four columns have date information in them and the last three columns have the sunshine information. The file contains one row for each day of the year.

A second set of files contain the same sunshine information, but each file has only 1 month of data in each. The filename indicates the year and the month involved.

The file ‘y2016\_mar.csv’ looks something like this.



The sunshine information is in the same format, but there is only a single column with the day number within the month, the rest of the date being in the filename.

We would like to create a single file containing all of the available data in the format of the ‘yXXXXall.csv’ files.

Tasks

1. Combined all of the month files into respective years, adding in the needed date columns
2. Combine all of the year files into a single file

Problems

1. Need to either combine the month files in the correct chronological order or sort the results afterwards
2. Need to either combine the year files in the correct chronological order or sort the results afterwards

## Approach

Regardless of whether we use Python or Powershell, the approach will be the same.

### Sorting

Although there is limited data in the folder currently we would like to make the solution as general as possible, so although we could probably afford to sort at the end, it will generally be more efficient to put the year files in order before we concatenate them.

For the month files however, as we know that there is a maximum of 366 records for any given year, this is small enough a number for us to sort after we have collected them together. This means that we don’t need to concern ourselves with the problem of sorting alphabetic month names in date order.

### Assumptions:

1. Year files are complete
2. There is a complete set of month files for years except the most recent year which may be incomplete
3. All of the files have a columns heading record which is consistent with the file type. I.e. as outlined above.

Outline :

1. Extract all of the month files
2. For each year (within the month files) create a list of the filenames involved
3. Create year file from the month files for all relevant years. (the latest year may not be complete)
   1. create a file, with the correct year name and insert a year file header record
   2. concatentate the month files
   3. sort the records into date order
4. Concatenate all of the year files into a single file.