1 Data read in system

The first challenge part of the project was to create a data read in system in order to allow data analysis. The data files come in the form on .csv files which has the data structure seen in Figure 1.


```
1    Stationsnamn;Klimatnummer;Mäthöjd (meter över marken)
2    Falsterbo;52230;2.0
3
4    Parameternamn;Beskrivning;Enhet
5    Lufttemperatur;momentanvärde, 1 gång/tim;degree celsius
6
7    Tidsperiod (fr.o.m);Tidsperiod (t.o.m);Höjd (meter över
8    1951-01-01 00:00:00;2015-11-30 23:59:59;5.0;55.3836;12.8
9
10   Datum;Tid (UTC);Lufttemperatur;Kvalitet;;Tidsutsnitt:
11   1951-01-01;06:00:00;-2.4;Y;;Kvalitetskontrollerade histo
12   1951-01-01;12:00:00;-0.8;Y;;Tidsperiod (fr.o.m.) = 1951-
13   1951-01-01;18:00:00;0.0;Y;;Tidsperiod (t.o.m.) = 2015-09
14   1951-01-02;06:00:00;1.0;Y;;Samplingstid = Ej angivet
15   1951-01-02;12:00:00;1.4;Y;;
16   1951-01-02;18:00:00;1.4;Y;;Kvalitetskoderna:
17   1951-01-03;06:00:00;2.0;Y;;Grön (G) = Kontrollerade och
18   1951-01-03;12:00:00;1.4;Y;;Gul (Y) = Misstänkta eller ac
19   1951-01-03;18:00:00;1.2;Y;;
20   1951-01-04;06:00:00;2.0;Y;;Orsaker till saknade data:
21   1951-01-04;12:00:00;2.0;Y;;   stationen eller givaren har
22   1951-01-04;18:00:00;1.8;Y;;   kvalitetskontrollerna har t
23   1951-01-05;06:00:00;2.0;Y
```

Figure 1: Data structure of the files used for analysis. Note that for this file line 1-10 contains no data points.

The first couple of lines consists of information which is irrelevant for reading in, which is why a for loop was used in order to skip the first lines until it reaches the first data point, which can be identified by its "year-month-day;hour:minute:second;temperature" format.

In order to obtain the information in each data point it was first split into three separate strings at each ";". In the next step the "year-month-day" string was split further at each "-" and stored into separate vectors. The same was done to the "hour:minute:second" string for each ":" and the hour was stored in a vector (the data points contains no minutes or seconds and are thus not stored in vectors). Last step was to convert the temperature to a float and then store it in a vector.

The next step was to create a date vector which calculates the time in terms of year (in decimal) so that one can display all the data points in the same plot without having for example data point "1951-06-11 6:00:00" and data point "1951-06-11 12:00:00" having the same x-value. The code itself converts hours, days and months into equivalent amount of one year and adds it all together with the year and then stores the result in a vector. The code also checks if the current year is a leap year and then takes into consideration the extra day when converting things to years.