

Week_2_project

November 26, 2017

1 Project 2: Holiday weather

by Rob Griffiths, 11 September 2015, updated 11 April 2017

This is the project notebook for Week 2 of The Open University's [Learn to code for Data Analysis](#) course.

There is nothing I like better than taking a holiday. In the winter I like to have a two week break in a country where I can be guaranteed sunny dry days. In the summer I like to have two weeks off relaxing in my garden in London. However I'm often disappointed because I pick a fortnight when the weather is dull and it rains. So in this project I am going to use the historic weather data from the Weather Underground for London to try to predict two good weather weeks to take off as holiday next summer. Of course the weather in the summer of 2016 may be very different to 2014 but it should give me some indication of when would be a good time to take a summer break.

1.1 Getting the data

Weather Underground keeps historical weather data collected in many airports around the world. Right-click on the following URL and choose 'Open Link in New Window' (or similar, depending on your browser):

<http://www.wunderground.com/history>

When the new page opens start typing 'London' in the 'Location' input box and when the pop up menu comes up with the option 'London, United Kingdom' select it and then click on 'Submit'.

When the next page opens with London Heathrow data, click on the 'Custom' tab and select the time period From: 1 January 2014 to: 31 December 2014 and then click on 'Get History'. The data for that year should then be displayed further down the page.

The data can be obtained month by month in CSV format. For example, the January 2014 data is available at this URL:

<https://www.wunderground.com/history/airport/LHR/2014/1/1/MonthlyHistory.html?format=1>

where LHR is the airport code. Changing the first 1 to 2, 3, etc. will show instead the data for February, March, etc. You can also change the 2014 in the URL to another year.

The date column heading will be the timezone, e.g. 'GMT' (Greenwich Mean Time) for London, or the time offset, e.g. '-02'. It may even change throughout the year. For example, Delhi (airport code DEL) is '+0430' from March to August and '+0330' in the other months.

You can copy each month's data directly from the browser to a text editor like Notepad or TextEdit, to obtain a single file with as many months as you wish.

Weather Underground has changed in the past the way it provides data and may do so again in the future. I have therefore collated the whole 2014 data in the provided 'London_2014.csv' file.

Now load the CSV file into a dataframe making sure that any extra spaces are skipped:

```
In [1]: import warnings
        warnings.simplefilter('ignore', FutureWarning)

        from pandas import *
        london = read_csv('London_2014.csv', skipinitialspace=True)
```

1.2 Cleaning the data

First we need to clean up the data. I'm not going to make use of 'WindDirDegrees' in my analysis, but you might in yours so we'll rename 'WindDirDegrees< br />' to 'WindDirDegrees'.

```
In [2]: london = london.rename(columns={'WindDirDegrees<br />' : 'WindDirDegrees'})
```

remove the < br /> html line breaks from the values in the 'WindDirDegrees' column.

```
In [3]: london['WindDirDegrees'] = london['WindDirDegrees'].str.rstrip('<br />')
```

and change the values in the 'WindDirDegrees' column to float64:

```
In [4]: london['WindDirDegrees'] = london['WindDirDegrees'].astype('float64')
```

We definitely need to change the values in the 'GMT' column into values of the datetime64 date type.

```
In [5]: london['GMT'] = to_datetime(london['GMT'])
```

We also need to change the index from the default to the datetime64 values in the 'GMT' column so that it is easier to pull out rows between particular dates and display more meaningful graphs:

```
In [6]: london.index = london['GMT']
```

1.3 Finding a summer break

According to meteorologists, summer extends for the whole months of June, July, and August in the northern hemisphere and the whole months of December, January, and February in the southern hemisphere. So as I'm in the northern hemisphere I'm going to create a dataframe that holds just those months using the datetime index, like this:

```
In [13]: summer = london.loc[datetime(2014,6,1) : datetime(2014,8,31)]
```

I now look for the days with warm temperatures.

```
In [8]: summer[summer['Mean TemperatureC'] >= 25]
```

Summer 2014 was rather cool in London: there are no days with temperatures of 25 Celsius or higher. Best to see a graph of the temperature and look for the warmest period.

So next we tell Jupyter to display any graph created inside this notebook:

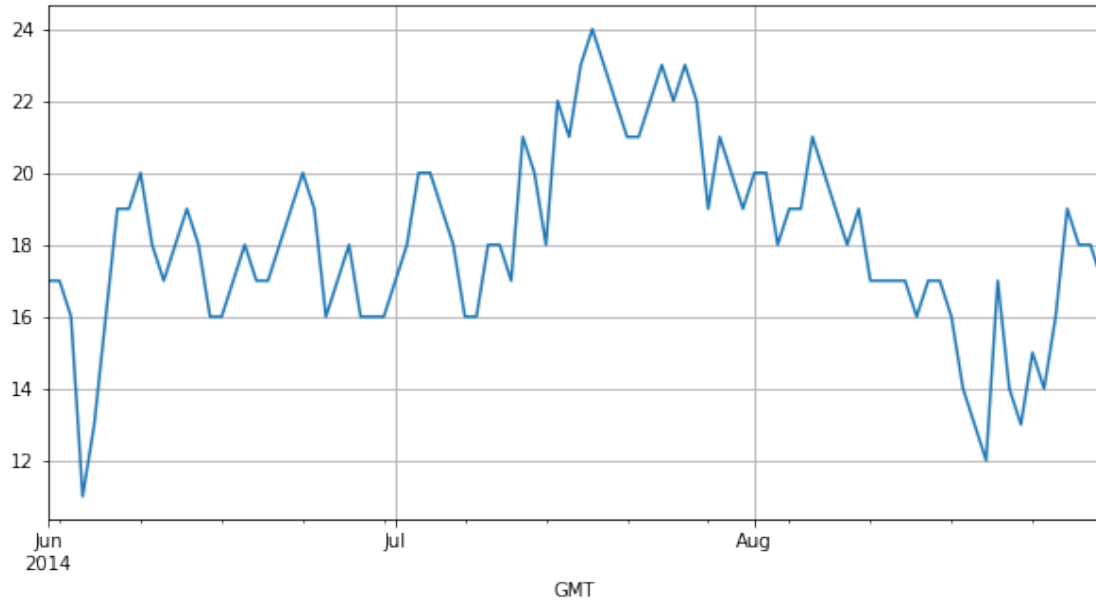
```
In [9]: %matplotlib inline
```

Now let's plot the 'Mean TemperatureC' for the summer:

```
In [10]: summer['Mean TemperatureC'].plot(grid=True, figsize=(10,5))
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7ff284e8c390>
```

```
Out[10]:
```

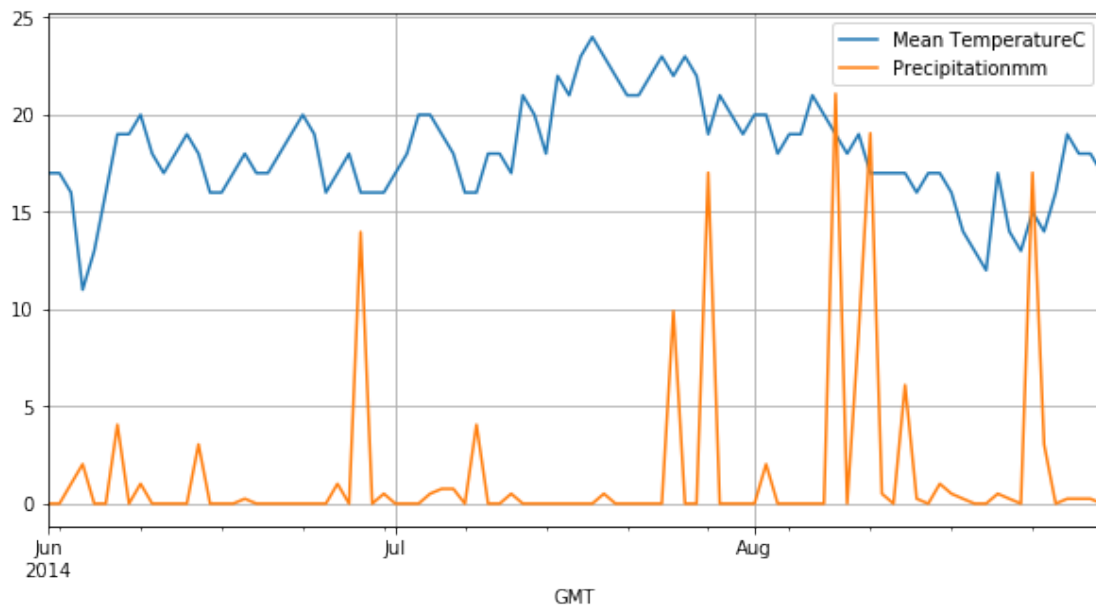


Well looking at the graph the second half of July looks good for mean temperatures over 20 degrees C so let's also put precipitation on the graph too:

```
In [11]: summer[['Mean TemperatureC', 'Precipitationmm']].plot(grid=True, figsize=(10,5))
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7ff284da8630>
```

```
Out[11]:
```

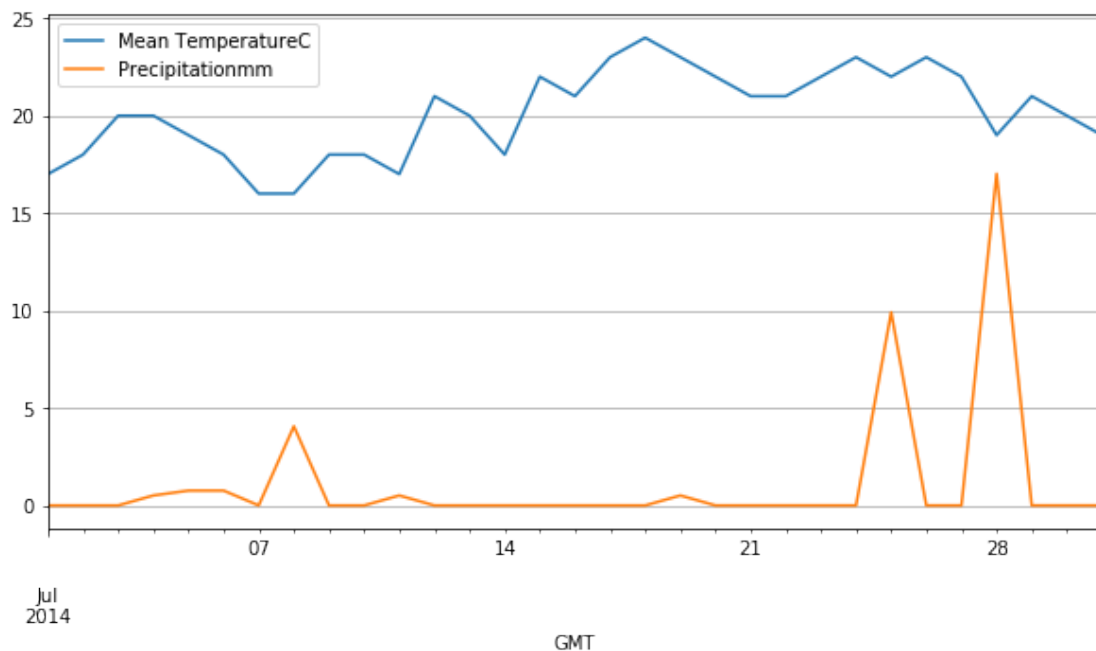


The second half of July is still looking good, with just a couple of peaks showing heavy rain. Let's have a closer look by just plotting mean temperature and precipitation for July.

```
In [14]: july = summer.loc[datetime(2014,7,1) : datetime(2014,7,31)]
         july[['Mean TemperatureC', 'Precipitationmm']].plot(grid=True, figsize=(10,5))
```

Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x7ff284e3e278>

Out[14]:



Yes, second half of July looks pretty good, just two days that have significant rain, the 25th and the 28th and just one day when the mean temperature drops below 20 degrees, also the 28th.

1.4 Conclusions

The graphs have shown the volatility of a British summer, but a couple of weeks were found when the weather wasn't too bad in 2014. Of course this is no guarantee that the weather pattern will repeat itself in future years. To make a sensible prediction we would need to analyse the summers for many more years. By the time you have finished this course you should be able to do that.