Scientific Programming Assignment 2 *DRAFT*

MPhil in Computational Biology

October 28, 2016

If there are errors found, I will update the assignment on the web at
`http://github.com/sje30/rpc2016`

Due date: to be determined in class.

Please submit your report to the Moodle as a single .Rnw file. Name your file `spa2_XXX.Rnw`, where XXX is your CRSid. (For example, I would save my file as `spa2_sje30.Rnw`.). You must ensure that the file can be compiled into a PDF by someone else using:

```
require(knitr); knit2pdf('spa2_sje30.Rnw')
```

Your .Rnw file should dynamically compute and report answers, rather than you writing the code and then typing your answers manually into the latex part of the document. In particular, it should run on subliminal, using only the system packages installed. If in doubt about whether you can use a given package, ask me in advance. (You may use .Rmd, rather than .Rnw as long as you indicate clearly how to convert it to a pdf.)

Your report must be a maximum of ten pages, excluding the appendix. This course work will consist of 20% towards your overall mark for this module.

1 Cambridge weather data set

The data files for this assignment will be available on subliminal at the following location:

```
data.dir = "/local/data/public/sje30/weather/2016/daily-text"
```

These data files come from `http://www.cl.cam.ac.uk/research/dtg/weather/`. Read that page to find out the contents of the data (and any errors). You will not need to copy these files into your directory if you are working on subliminal, as you are expected not to edit them.

Your task is to answer the following questions.

1. Look through the data files and decide how many files need to be excluded from your analysis. These may be because the data are incomplete, or inaccurate. (You should analyse only the files that have complete information.) Report which files you have removed and why. (You may need to complete the entire assignment and then revisit this question as you analyse the data.) Assuming the database started on 1995-07-01 and finished on 2016-10-25, what percentage of days do we have complete data for? [4 marks]

2. Plot the variation in temperature for one day in the database, 2012-12-25 [1 mark]

3. For each day, compute the mean temperature over the day, and plot this mean temperature as a function of the date. What was the hottest day on record, and what was the coldest day on record? Mark those extremes on the plot. [4 marks]

4. For each day of the year, plot the 'mean of the mean temperatures' as a function of the day of the year. Plot the mean temperature for the year which is most similar to this average profile. Repeat for the year that is least similar. State briefly how you measured similarity. [3 marks]

5. Draw a scatterplot of the mean temperature and the total amount of rain in each day. Report the Pearson correlation between the two variables. [2 marks]

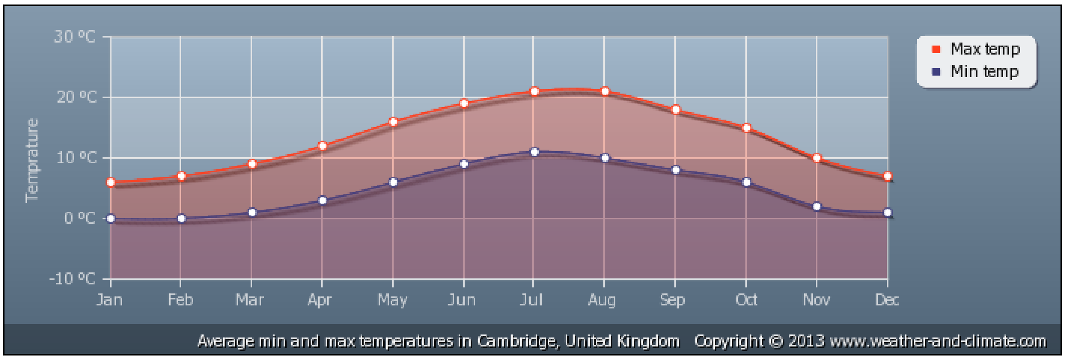

Figure 1: Example min/max mean temperature by month. Graph copied from `http://www.weather-and-climate.com/average-monthly-min-max-Temperature,cambridge,United-Kingdom`.

6. Draw a scatterplot of the total number of hours of sunshine against the mean daily temperature. What is the Pearson correlation between these two variables? [2 marks]

7. Draw a graph similar to Figure 1, using our data for Cambridge. How do the graphs compare? [2 marks]

8. Define a "wet day" as one where the total rainfall in a day exceeds a certain threshold that you should clearly define. When was the longest consecutive number of days that it rained according to your method? [2 marks]

Hints: R can work with date/time objects quite nicely.

```
as.Date("2012_10_29", "%Y_%m_%d") + 3

## [1] "2012-11-01"

strptime("20/2/06 11:16:16.683", "%d/%m/%y %H:%M:%OS")

## [1] "2006-02-20 11:16:16 GMT"
```