ST662 Topics in Data Analytics
2018-19 Semester 2
Assignment Sheet 2

Due at 2pm Monday 4th March 2019

A grassland biodiversity experiment was conducted at many sites across Europe and one in Canada. The data from this experiment was published in the journal called *Ecology*. Information on the experiment is available at

Abstract: `http://onlinelibrary.wiley.com/doi/10.1890/14-0170.1/abstract`.
Datasets for download: `http://www.esapubs.org/archive/ecol/E095/232/`.
Datasets' descriptions: `http://www.esapubs.org/archive/ecol/E095/232/metadata.php`.

Write a SAS programme to do the following data manipulation exercises.

1. (a) Download the biomass.csv dataset and read it into SAS.

   (b) Restrict the dataset to only sites 13, 14, 23, 25, 33 and 52, to only the first year of experimental data, and to only treatment 1.

   (c) Create a new dataset that provides the annual yield for each plot at each site.

   (d) Create a new dataset that provides the average annual yield for each site (i.e. averaged across all plots).

2. (a) Download the climate.csv dataset and read it into SAS.

   (b) Restrict the dataset to only sites 13, 14, 23, 25, 33 and 52.

   (c) Create a new dataset that provides the average 'air_mean' for each site and each year.

3. (a) Merge the biomass dataset created in Qu 1d with the relevant year of the climate dataset created in Qu 2c.

   (b) Create a scatter plot of average annual yield versus average annual temperature. Ensure the quality of the scatterplot is suitable for including in a presentation or report (e.g. put a title on it, check the font sizes of labels, perhaps label points within the graph etc).

Details on what you have to submit for this assignment

Submission of this assignment is in two parts:

1. Submit on Moodle the SAS programme (code only, no output) you created to address Qu's 1-3 above. This must be done before the start of class, i.e. BEFORE 2PM. Do not leave this until the last minute as Moodle submission will close at this time.

2. Submit a printed hard copy of the dataset generated in Qu 3a and of the scatter plot created in Qu 3b. These should both fit on one page. Also submit a hard copy of your programme (code only, no output). This printout will be submitted at the beginning of the lecture at 2pm.