Tutorial 3

John O'Sullivan

1. Using the commands **library(boot)** and **data(melanoma)**, load the in-built data set called melanoma into R. These data are from measurements made on patients with malignant melanoma. They are stored as a data frame with records for 205 patients. Use the help file to familiarise yourself with the data set.

2. Use the **str()** function to see a summary of the variables in the data.

3. Give 3 different ways of accessing the age of the person stored in the 95th record (row).

4. Create a new variable which gives the number of years since the patients had their operation. Attach this column to the dataframe, giving it a sensible name.

5. Create a new data frame which contains only the 'time', 'age' and 'thickness' columns. Use **sapply()** to get the mean, standard deviation, and interquartile range (function **IQR()**) of these variables.

6. Using **subset()**, create two dataframes - one containing the male patients and one containing the female patients. How many female patients are there? Find the mean age at the time of operation for these two groups. Is there a difference between them?

7. The **findwords()** function is at the end of Lecture 3 Code. Look at this function carefully, and see how it works.

Find a new block of text from an article from today's Irish Times on www.irishtimes.com. Run it through the **findwords()** function.