STAT40730
Data Programming with R (Online).
Lab 3: lists and data frames.

1. Using the command `data(esoph)` load the inbuilt data set called `esoph` into `R`. These data are from a case-control study of oesophageal cancer in France. They are stored as a data frame with records for 88 age/alcohol/tobacco combinations. Use the help file to familiarise yourself with the data set.

2. Use the `colnames` function to give your new data frame some neater column names.

3. Give 3 different ways of accessing the number of cases in the 15th record.

4. Create a new variable which gives the number of observations in each record i.e. the combined total of cases and controls.

5. Create a new data frame which contains only the number of cases and the number of controls columns. Use `sapply` to get the mean, standard deviation, and interquartile range (function `IQR`) of these variables.

6. Using `subset`, find the mean number of cases in the set of records which have low alcohol intake (i.e. 0-39g/day) and the mean number of cases in the set of records which have high alcohol intake (i.e. 120+). Is there a difference between the groups?

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

8. Re-create the graph from Lecture 3 slide 28 but this time using mother's weight (variable `lwt`) instead of age. Is it a positive or negative relationship?