Preliminaries:

Before you begin, ensure you have done the following (instructions are contained within the slides):

- Create a directory (folder) in your Documents called r_course/R_Course (note, folder names are case sensitive).
- Set the working directory as the r_course folder.
- From the course website, download the data set **colon_cancer_data_set.txt** and transfer to the r_course directory.
- To download file, press control and click on the link – select the "Download Linked File As" option. Make sure to save the file in the r_course folder.
- Using any suitable variable name, read in the colon_cancer_data_set.txt file into RStudio (`read.table()`).

The colon_cancer_data_set.txt files contains matched gene expression data for tumour and normal cells for a cohort of patients. The gene names are given accession numbers.

Q1. Using the functions, `names()`, `dim()`, `nrow()`, `ncol()`, `row.names()`, find out the following (Note, it is often useful to just look at the first few rows and columns of the data set, e.g., `df[1:5, 1:10]`):

1.1 How many tumor and normal samples are in the data set?

1.2 What data is contained in the last column of the data set?

1.3 For how many genes has gene expression been measured? (Note, the last column of the data set is not a gene.)

1.4 What is the range of columns that contain gene expression data?

Q2. From the original dataframe, create two dataframes, one called `affected` containing only gene expression for the tumour samples

(**Hint:** Status=A) and another called `unaffected` containing only gene expression for the normal samples (**Hint:** Status=U).

2.1 How many tumour samples are in the data set?

2.2 How many normal samples are in the data set?

Q3. For the affected and unaffected data separately:

3.1 Use the `apply()` function to compute the mean gene expression for each gene.

3.2 Combine the vector of values into a single matrix remembering to give it a variable name (**Hint:** `rbind()`).

3.3 Write this combined matrix of gene names to a file (**Hint:** `write.table()`).

Q4. Once completed:

4.1 Save your script as **problems_1.R**

4.2 Save your session as **problems_1.RData**

Note! Please ensure to save your work as we will use this data set continuously throughout the course.