The University of British Columbia
Irving K. Barber Faculty of Science
DATA 101

Assignment 2

Please submit your assignment as an R script file named with your last name, student number, assignment number and with the suffix R. For example, if Joe Smith, student number 87654321 hands in Assignment 2, he would name the file `Smith87654321A2.R`.

Within your answer file, include answers with your R code preceded by the `#` sign. For example, to answer the 5th question on an assignment which is "Perform the calculation $2 + 2$", you would type

```
# Question 5
2 + 2 #coding
# 4 (your answer here )
```

Due Date: October 9, 2020

In each question below, type the required lines of R code, together with the answer to the question.

1. (7 marks) Consider the `chickwts` data frame that is built into R.

   (a) Display the 14th row of this data frame.

   (b) Using one R command, display the 7th, 14th and 37th elements of `weight` column.

   (c) Extract the observations from this data frame that correspond to `casein`, and assign the resulting subset to `chickwtsCasein`.

   (d) Find the average of the weights in `chickwtsCasein`.

   (e) Create a new column which contains a factor called `Feed` which has levels `Casein`, `Horsebean`, and so on. That is, `Feed` is the same as `feed` but all entries are capitalized.

   (f) Extract the subset from `chickwts` that contains all observations where weight is less than 240. Call the resulting data frame `chick240`.

   (g) Use the `table()` function to create a vector called `chick240Table` as in

   ```
   chick240Table <- table(chick240$feed)
   ```

   Create a bar plot of the counts in `chick240Table`.

2. (8 marks) If it is not already, installed, install the R package *DAAG* either by using the menu system in RStudio, or by typing the following into an R session:

   ```
   install.packages("DAAG")
   ```

   Load the package into an R session, for example, by typing

   ```
   library(DAAG)
   ```

   (a) How many observations are in the `cuckoos` data frame?

   (b) What is the length measurement for the 27th observation.

   (c) List the elements in the 40th row of the data frame.

(d) Use the `levels` function to identify the types of birds that make up the `species` factor of the `cuckoos` data frame.

(e) Again, using the `levels()` function, create a new factor called `m.pipitFactor` which has levels `meadow.pipit` and `other`, and where `other` is the value assigned to those elements which correspond to any bird species other than `meadow.pipit`.

(f) Create a logical vector column in the data frame called `m.pipit` which contains `TRUE` in the rows corresponding to `meadow.pipit` and `FALSE` in the other rows.

(g) Create two new data frames: one called `cuckoosMPipit`, which consists only of the rows in `cuckoos` corresponding to `meadow.pipit` and one called `cuckoosOther`, which contains all other rows.

(h) Create another data frame called `cuckoosLongLength` which contains all observations in the `cuckoos` data frame where the length exceeds 23. Find the average of the breadth observations in `cuckoosLongLength`.