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.

   ```
   chickwts[14,]

   ##    weight    feed
   ## 14    141 linseed
   ```

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

   ```
   chickwts$weight[c(7,14,37)]

   ## [1] 108 141 423
   ```

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

   ```
   chickwtsCasein<-subset(chickwts, feed=="casein")
   chickwtsCasein

   ##    weight   feed
   ## 60    368 casein
   ## 61    390 casein
   ## 62    379 casein
   ## 63    260 casein
   ## 64    404 casein
   ## 65    318 casein
   ## 66    352 casein
   ## 67    359 casein
   ## 68    216 casein
   ## 69    222 casein
   ## 70    283 casein
   ## 71    332 casein
   ```

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

```
mean(chickwtsCasein$weight)

## [1] 323.5833
```

(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.

```
chickwts$Feed <- chickwts$feed
levels(chickwts$Feed)<- c("Casein", "Horsebean", "Linseed", "Meatmeal","Soybean"
chickwts$Feed

##  [1] Horsebean Horsebean Horsebean Horsebean Horsebean Horsebean Horsebean
##  [8] Horsebean Horsebean Horsebean Linseed   Linseed   Linseed   Linseed
## [15] Linseed   Linseed   Linseed   Linseed   Linseed   Linseed   Linseed
## [22] Linseed   Soybean   Soybean   Soybean   Soybean   Soybean   Soybean
## [29] Soybean   Soybean   Soybean   Soybean   Soybean   Soybean   Soybean
## [36] Soybean   Sunflower Sunflower Sunflower Sunflower Sunflower Sunflower
## [43] Sunflower Sunflower Sunflower Sunflower Sunflower Sunflower Meatmeal
## [50] Meatmeal  Meatmeal  Meatmeal  Meatmeal  Meatmeal  Meatmeal  Meatmeal
## [57] Meatmeal  Meatmeal  Meatmeal  Casein    Casein    Casein    Casein
## [64] Casein    Casein    Casein    Casein    Casein    Casein    Casein
## [71] Casein
## Levels: Casein Horsebean Linseed Meatmeal Soybean Sunflower
```

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

```
chick240<-subset(chickwts, weight<240)
chick240

##    weight      feed      Feed
## 1     179 horsebean Horsebean
## 2     160 horsebean Horsebean
## 3     136 horsebean Horsebean
## 4     227 horsebean Horsebean
## 5     217 horsebean Horsebean
## 6     168 horsebean Horsebean
## 7     108 horsebean Horsebean
## 8     124 horsebean Horsebean
## 9     143 horsebean Horsebean
## 10    140 horsebean Horsebean
## 12    229   linseed   Linseed
## 13    181   linseed   Linseed
## 14    141   linseed   Linseed
## 16    203   linseed   Linseed
## 17    148   linseed   Linseed
## 18    169   linseed   Linseed
## 19    213   linseed   Linseed
```

```
## 24     230    soybean    Soybean
## 29     193    soybean    Soybean
## 33     199    soybean    Soybean
## 34     171    soybean    Soybean
## 35     158    soybean    Soybean
## 42     226 sunflower Sunflower
## 54     153  meatmeal   Meatmeal
## 57     206  meatmeal   Meatmeal
## 68     216    casein     Casein
## 69     222    casein     Casein
```
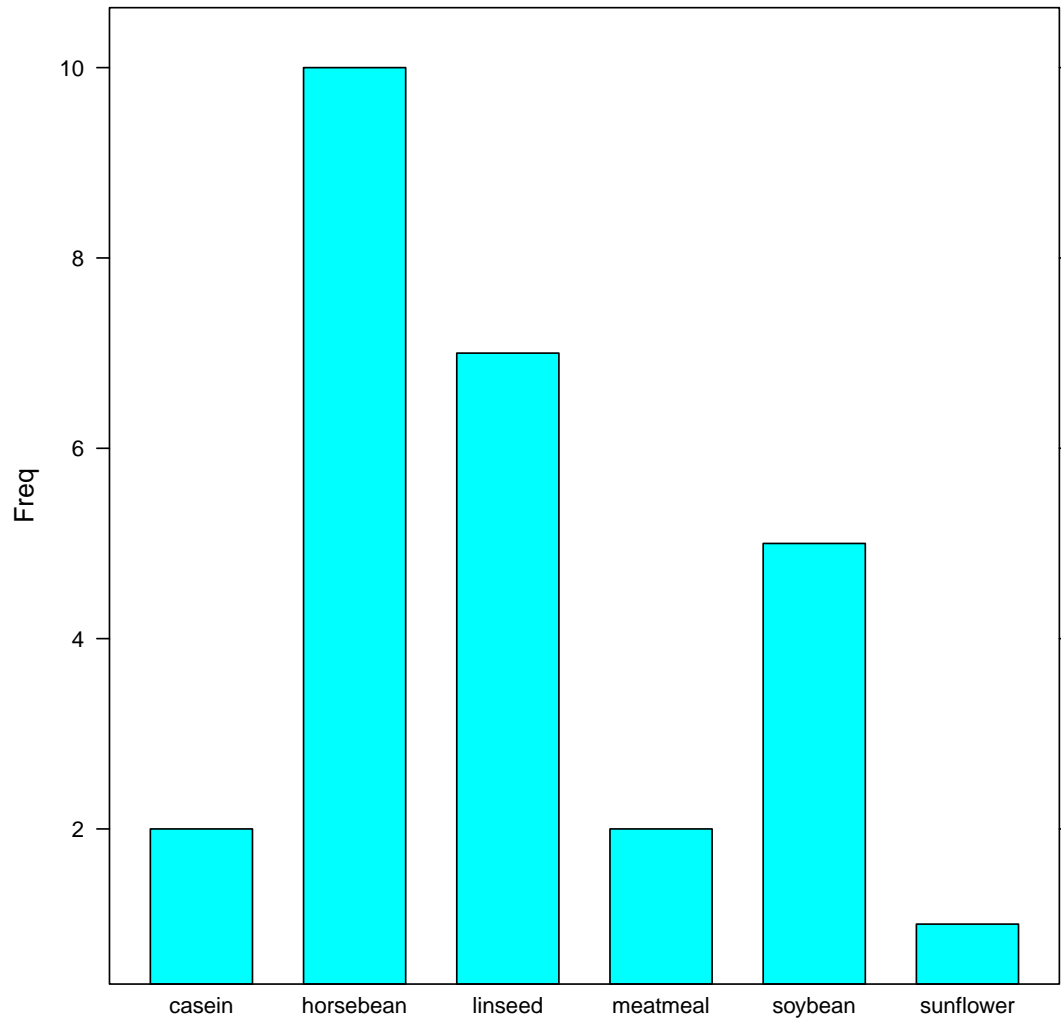
(g) Use the `table()` function to create a vector called `chick240Table` as in

```
chick240Table <-as.data.frame( table(chick240$feed))
#chick240Table <- table(chick240£feed)
chick240Table

##          Var1 Freq
## 1    casein    2
## 2 horsebean   10
## 3   linseed    7
## 4  meatmeal    2
## 5   soybean    5
## 6 sunflower    1
```

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

```
barchart(Freq ~ Var1, data=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:

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

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

```r
library(DAAG)
```

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

```r
dim(cuckoos)[1]
## [1] 120
```

or

```
nrow(cuckoos)

## [1] 120
```

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

```
cuckoos$length[27]

## [1] 22
```

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

```
cuckoos[40,]

##    length breadth      species id
## 40  22.9    17.2 meadow.pipit 60
```

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

```
levels(cuckoos$species)

## [1] "hedge.sparrow" "meadow.pipit"  "pied.wagtail"  "robin"
## [5] "tree.pipit"    "wren"
```

(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`.

```
cuckoos$m.pipitFactor <- cuckoos$species
levels(cuckoos$m.pipitFactor) <- c("other", "meadow.pipit", rep("other",4))
```

(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.

```
cuckoos$m.pipit<- cuckoos$species=="meadow.pipit"
```

(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.

```
cuckoosMPipit <- subset(cuckoos, m.pipitFactor=="meadow.pipit")
cuckoosOther <- subset(cuckoos, m.pipitFactor=="other")
```

(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`.

```
cuckoosLongLength <- subset(cuckoos, length>23)
mean(cuckoosLongLength$breadth)

## [1] 16.78824
```