Data Manipulation - Dplyr

Importing the libraries

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
## filter, lag
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

Importing the data-sets

```
df = read.csv('Allometry.csv')
head(df)
     species diameter height
                              leafarea branchmass
##
## 1
       PSME
               54.61 27.04 338.485622 410.24638
## 2
       PSME
               34.80 27.42 122.157864
                                         83.65030
## 3
       PSME
               24.89 21.23
                              3.958274
                                          3.51270
               28.70 24.96 86.350653
## 4
       PSME
                                         73.13027
## 5
       PSME
               34.80 29.99 63.350906
                                         62.39044
       PSME
               37.85 28.07 61.372765
## 6
                                         53.86594
```

1) Select the first three columns

```
ans1 <- data.frame(df %>% select(c(1:3)))
head(ans1)
     species diameter height
## 1
                54.61 27.04
        PSME
## 2
        PSME
                34.80 27.42
## 3
        PSME
                24.89 21.23
        PSME
                28.70 24.96
                34.80 29.99
## 5
        PSME
        PSME
                37.85 28.07
```

2) Select all columns except the species column

```
ans2 <- data.frame(df %>% select(-species))
head(ans2)
```

diameter height leafarea branchmass

```
## 1
        54.61 27.04 338.485622
                                  410.24638
## 2
        34.80
               27.42 122.157864
                                   83.65030
## 3
        24.89
               21.23
                        3.958274
                                    3.51270
               24.96
## 4
        28.70
                      86.350653
                                   73.13027
## 5
        34.80
               29.99
                      63.350906
                                   62.39044
## 6
        37.85
               28.07
                      61.372765
                                   53.86594
```

3) Select all the rows where both diameter and height are greater than 20

```
ans3 <- (df %>% filter(diameter > 20 & height > 20))
head(ans3)
```

```
species diameter height
##
                               leafarea branchmass
                54.61 27.04 338.485622
## 1
        PSME
                                          410.24638
## 2
                34.80 27.42 122.157864
        PSME
                                           83.65030
## 3
        PSME
                24.89
                       21.23
                                3.958274
                                            3.51270
## 4
        PSME
                28.70 24.96
                              86.350653
                                           73.13027
## 5
        PSME
                34.80 29.99
                              63.350906
                                           62.39044
## 6
        PSME
                37.85 28.07
                              61.372765
                                           53.86594
```

4) Arrange the rows by height

```
ans4 = df %>% arrange(height)
head(ans4)
```

```
##
     species diameter height
                                leafarea branchmass
## 1
        PIP0
                  4.83
                         3.57
                                2.636336
                                             1.77810
                                             2.67065
## 2
        PSME
                  5.33
                         4.95
                                4.851567
## 3
                  8.38
        PIMO
                         4.95
                                6.551044
                                             4.36969
## 4
        PSME
                  6.10
                         5.26
                               7.595163
                                             3.57145
## 5
        PIMO
                  6.48
                         5.42
                               7.650902
                                             3.50621
## 6
        PSME
                  7.37
                         5.58 11.502851
                                             4.58916
```

5) Arrange the rows by Species and then by height

```
ans5 = df %>% arrange(species, height)
head(ans5)
```

```
##
                                                species diameter height
                                                                          leafarea
## 1 \t \t \t
                      \t
                             \t \t
                                       \t
                                          \t
                                                   \t \t
                                                               NA
                                                                      NA
                                                   PIMO
                                                             8.38
                                                                    4.95
                                                                         6.551044
## 3
                                                   PIMO
                                                             6.48
                                                                    5.42
                                                                          7.650902
## 4
                                                   PIMO
                                                            18.29
                                                                  12.98 82.093031
## 5
                                                   PIMO
                                                            12.95
                                                                   13.44 18.855867
## 6
                                                   PIMO
                                                            18.54 21.29 44.934469
##
     branchmass
## 1
## 2
        4.36969
## 3
        3.50621
## 4
       28.04785
## 5
        8.71068
## 6
       16.54457
```

6) Create a new column by dividing the leafarea with branchmass

```
ans6 = df %>% mutate(leftarea_divide_branchmass = leafarea / branchmass)
head(ans6)
```

```
leafarea branchmass leftarea_divide_branchmass
     species diameter height
## 1
       PSME
                54.61 27.04 338.485622 410.24638
                                                                    0.8250789
## 2
       PSME
                34.80 27.42 122.157864
                                                                    1.4603398
                                          83.65030
## 3
       PSME
                24.89 21.23
                               3.958274
                                           3.51270
                                                                    1.1268466
                28.70 24.96 86.350653
## 4
       PSME
                                          73.13027
                                                                    1.1807785
## 5
       PSME
                34.80 29.99
                              63.350906
                                          62.39044
                                                                    1.0153944
       PSME
                37.85 28.07 61.372765
## 6
                                          53.86594
                                                                    1.1393613
```

7) Find the average diameter for each species

```
ans7 = df %>% group_by(species) %>% summarize(avg_diameter = mean(diameter))
head(ans7)
```

```
## # A tibble: 4 x 2
##
     species
                                                           avg_diameter
     <chr>
                                                                  <dbl>
## 1 "\t \t \t \t
                                                   \t \t"
                       \t
                             \t \t
                                        \t \t
                                                                   NA
## 2 "PIMO"
                                                                   37.3
## 3 "PIPO"
                                                                   36.6
## 4 "PSME"
                                                                   33.1
```

8) Create a new data frame containing 10th to 35th rows of the data

```
ans8 = df %>% slice(10:35)
head(ans8)
```

```
species diameter height leafarea branchmass
## 1
               26.16 25.85 45.02004
       PSME
                                        45.33095
## 2
       PSME
               43.69 31.83 145.80980
                                        96.32311
## 3
       PSME
               69.85 31.35 349.05701 543.97314
       PSME
               44.45 32.61 176.02921
                                      110.12476
## 5
       PSME
               56.64 31.70 319.50711
                                       260.48752
       PSME
               54.61 30.18 234.36878 281.12312
```