

Data Manipulation - Dplyr

Importing the libraries

```
library(dplyr)

##
## 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
## 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
```

1) Select the first three columns

```
ans1 <- data.frame(df %>% select(c(1:3)))
head(ans1)
```

```
##   species diameter height
## 1    PSME    54.61  27.04
## 2    PSME    34.80  27.42
## 3    PSME    24.89  21.23
## 4    PSME    28.70  24.96
## 5    PSME    34.80  29.99
## 6    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
```


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

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

##	species	diameter	height	leafarea	branchmass	leftarea_divide_branchmass
## 1	PSME	54.61	27.04	338.485622	410.24638	0.8250789
## 2	PSME	34.80	27.42	122.157864	83.65030	1.4603398
## 3	PSME	24.89	21.23	3.958274	3.51270	1.1268466
## 4	PSME	28.70	24.96	86.350653	73.13027	1.1807785
## 5	PSME	34.80	29.99	63.350906	62.39044	1.0153944
## 6	PSME	37.85	28.07	61.372765	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"      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	PSME	26.16	25.85	45.02004	45.33095
## 2	PSME	43.69	31.83	145.80980	96.32311
## 3	PSME	69.85	31.35	349.05701	543.97314
## 4	PSME	44.45	32.61	176.02921	110.12476
## 5	PSME	56.64	31.70	319.50711	260.48752
## 6	PSME	54.61	30.18	234.36878	281.12312