

# Data Manipulation - Dplyr

Prashanth.S (19MID0020)

## Questions

Perform the following operations on the attached file using methods available in the dplyr library

- 1) Select the first three columns
- 2) Select all columns except the species column
- 3) Select all the rows where both diameter and height are greater than 20
- 4) Arrange the rows by height
- 5) Arrange the rows by Species and then by height
- 6) Create a new column by dividing the leafarea with branchmass
- 7) Find the average diameter for each species
- 8) Create a new data frame containing 10th to 35th rows of the data

## Code

```
1 library(utils)
2 library(dplyr)
3
4 df=read.csv("Allometry.csv")
5
6 ## 1st 3 columns
7 ans1 <- data.frame(df %>% select(c(1:3)))
8 #head(ans1)
9
10 ## exclude the species column
11 ans2 <- data.frame(df %>% select(-species))
12 #head(ans2)
13
14 ## take out the rows where the (diameter and height) > 20
15 ans3 <- (df %>% filter(diameter > 20 & height > 20))
16 #ans3
17
18 ## arrange the rows by height
19 ans4 = df %>% arrange(height)
20 #ans4
21
22 ## arrange the rows by species and then by height
23 ans5 = df %>% select(c(1:5)) %>%
24   arrange(species, height) ## give only arrange(species, height)
25 #ans5
26
27 ## leafarea_divide_branchmass = (leafarea) / (branchmass)
28 ans6 = df %>% mutate(leafarea_divide_branchmass = leafarea / branchmass)
29 #ans6
30
31 ## average diameter for each species
32 ans7 = df %>% group_by(species) %>% summarize(avg_diameter = mean(diameter))
33 #ans7
34
35 ## 10 to 35 row of the data
36 ans8 = df %>% slice(10:35)
37 #ans8
```

★ dplyr\_library.R Allometry.csv

PRASHANTH S 19MID0020

Description

Perform the following operations on the attached file using methods available in the dplyr library

- 1) Select the first three columns
- 2) Select all columns except the species column
- 3) Select all the rows where both diameter and height are greater than 20
- 4) Arrange the rows by height
- 5) Arrange the rows by Species and then by height
- 6) Create a new column by dividing the leafarea with branchmass
- 7) Find the average diameter for each species
- 8) Create a new data frame containing 10th to 35th rows of the data

<https://moovit.vit.ac.in/>

## Data-Set

★ dplyr_library.R [3] Allometry.csv [3]		PRASHANTH S 19MID0020
		Description
<pre>1 species,diameter,height,leafarea,branchmass 2 PSME,54.61,27.03999951,338.485622,410.24638 3 PSME,34.8,27.4199988,122.157864,83.6503 4 PSME,24.89,21.23000017,3.958274,3.5127 5 PSME,28.7,24.95999968,86.350653,73.13027 6 PSME,34.8,29.98999968,63.350906,62.39044 7 PSME,37.85,28.06999906,61.372765,53.86594 8 PSME,22.61,23.42999913,32.077794,22.22474 9 PSME,39.37,32.44999809,147.270523,119.38869 10 PSME,39.88,30.10000028,141.787332,95.58503 11 PSME,26.16,25.84999877,45.020041,45.33095 12 PSME,43.69,31.82999998,145.809802,96.32311 13 PSME,69.85,31.35000017,349.05701,543.97314 14 PSME,44.45,32.61000019,176.029213,110.12476 15 PSME,56.64,31.69999766,319.507115,260.48752 16 PSME,54.61,30.17999806,234.368784,281.12312 17 PSME,5.33,4.949999782,4.851567,2.67065 18 PSME,6.1,5.26000011,7.595163,3.57145 19 PSME,7.37,5.580000195,11.502851,4.58916 20 PSME,8.33,6.470000131,25.380647,8.63615 21 PSME,13.51,8.910000365,65.698749,34.93701 22 PSME,51.31,33.30000015,160.839174,165.05529 23 PSME,22.35,19.38999953,31.780702,35.64921 24 PIPO,69.6,39.36999864,189.733007,452.42455 25 PIPO,58.42,35.8099995,253.308265,595.64015 26 PIPO,33.27,20.80000143,91.538428,160.44416 27 PIPO,44.2,29.11000066,90.453658,149.72883 28 PIPO,30.48,22.39999889,99.73679,44.13532 29 PIPO,27.43,27.69000074,34.464685,22.9836 30 PIPO,43.18,35.58000047,68.150309,106.4041 31 PIPO,38.86,33.12000106,46.32606,58.24071 32 PIPO,52.58,41.16000258,160.993131,214.34109 33 PIPO,20.83,23.34000042,9.806496,8.25614 34 PIPO,24.13,25.94000095,20.74328,22.60111 35 PIPO,24.89,25.11000022,21.649603,16.77015 36 PIPO,45.97,30.3899993,66.633675,87.36908 37 PIPO,35.05,28.39999855,54.267994,51.09006 38 PIPO,23.88,23.38000056,19.84468,13.98343</pre>		Perform the following operation on the attached file using method available in the dplyr library  1) Select the first three column  2) Select all column except the species column  3) Select all the rows where both diameter and height

Mam, I implemented the code in moodle as-well as in RStudio also.

In-order to avoid confusion with continuous output, I am including my implementation in RStudio also.

