

## shrub volume data

### Exercise 1

Create a new Rmd file called “shrub\_volume.Rmd”. Write a short paragraph describing the data set before starting your analysis. Include a figure.

The dataset is based on a study of the factors that control the size and carbon storage of shrubs. An experiment in which the effects of three various treatments on shrub volume located in four different areas is seen. The site describes the area in which the scrubs are seen. The experiment describes thich type of treatment is implemented on the shrubs. The length, width, and height, constitute measurement factprs that are utilized to determine the volume of each of the shrubs.

```
knitr::include_graphics("../fig/shrubs-image.jpeg")
```



```
library(magrittr)
```

```
## Warning: package 'magrittr' was built under R version 4.0.5
```

```
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
```

Import the data set using `read.csv()`. Assign it to an object called `shrubs`.

```
shrubs <- read.csv(file = "../data-raw/shrub-volume-data.csv")
```

Select the data from the “length” column and print it out (using `select()`).

```
select(shrubs, length) %>% print()
```

```
##   length
## 1    2.2
## 2    2.1
## 3    2.7
## 4    3.0
## 5    3.1
## 6    2.5
## 7    1.9
## 8    1.1
## 9    3.5
## 10   2.9
## 11   4.5
## 12   1.2
```

Select the data from the `site` and `experiment` columns and print it out (using `select()`).

```
select(shrubs, site, experiment) %>% print()
```

```
##      site experiment
## 1      1           1
## 2      1           2
## 3      1           3
## 4      2           1
## 5      2           2
## 6      2           3
## 7      3           1
## 8      3           2
## 9      3           3
## 10     4           1
## 11     4           2
## 12     4           3
```

Add a new column named “area” containing the area of the shrub, which is the length times the width (using `mutate()`).

```
mutate(shrubs, area = length*width)
```

```
##      site experiment length width height  area
## 1      1           1    2.2   1.3    9.6  2.86
## 2      1           2    2.1   2.2    7.6  4.62
## 3      1           3    2.7   1.5    2.2  4.05
## 4      2           1    3.0   4.5    1.5 13.50
## 5      2           2    3.1   3.1    4.0  9.61
## 6      2           3    2.5   2.8    3.0  7.00
## 7      3           1    1.9   1.8    4.5  3.42
## 8      3           2    1.1   0.5    2.3  0.55
## 9      3           3    3.5   2.0    7.5  7.00
## 10     4           1    2.9   2.7    3.2  7.83
## 11     4           2    4.5   4.8    6.5 21.60
## 12     4           3    1.2   1.8    2.7  2.16
```

Sort the data by length (using `arrange()`).

```
arrange(shrubs, length)
```

```
##      site experiment length width height
## 1      3           2    1.1   0.5    2.3
## 2      4           3    1.2   1.8    2.7
## 3      3           1    1.9   1.8    4.5
## 4      1           2    2.1   2.2    7.6
## 5      1           1    2.2   1.3    9.6
## 6      2           3    2.5   2.8    3.0
## 7      1           3    2.7   1.5    2.2
## 8      4           1    2.9   2.7    3.2
## 9      2           1    3.0   4.5    1.5
## 10     2           2    3.1   3.1    4.0
## 11     3           3    3.5   2.0    7.5
## 12     4           2    4.5   4.8    6.5
```

Filter the data to include only plants with heights greater than 5 (using `filter()`).

```
filter(shrubs, height > 5)
```

```
##   site experiment length width height
## 1    1           1   2.2   1.3   9.6
## 2    1           2   2.1   2.2   7.6
## 3    3           3   3.5   2.0   7.5
## 4    4           2   4.5   4.8   6.5
```

Filter the data to include only plants with heights greater than 4 and widths greater than 2 (using `,` or `&` to include two conditions).

```
filter(shrubs, height > 4 & width > 2)
```

```
##   site experiment length width height
## 1    1           2   2.1   2.2   7.6
## 2    4           2   4.5   4.8   6.5
```

Filter the data to include only plants from Experiment 1 or Experiment 3 (using `|` for “or”).

```
filter(shrubs, experiment == 1 | experiment == 3)
```

```
##   site experiment length width height
## 1    1           1   2.2   1.3   9.6
## 2    1           3   2.7   1.5   2.2
## 3    2           1   3.0   4.5   1.5
## 4    2           3   2.5   2.8   3.0
## 5    3           1   1.9   1.8   4.5
## 6    3           3   3.5   2.0   7.5
## 7    4           1   2.9   2.7   3.2
## 8    4           3   1.2   1.8   2.7
```

Filter the data to remove rows with null values in the height column (using `!is.na()`)

```
filter(shrubs, !is.na(height))
```

```
##   site experiment length width height
## 1    1           1   2.2   1.3   9.6
## 2    1           2   2.1   2.2   7.6
## 3    1           3   2.7   1.5   2.2
## 4    2           1   3.0   4.5   1.5
```

```
## 5      2      2      3.1  3.1    4.0
## 6      2      3      2.5  2.8    3.0
## 7      3      1      1.9  1.8    4.5
## 8      3      2      1.1  0.5    2.3
## 9      3      3      3.5  2.0    7.5
## 10     4      1      2.9  2.7    3.2
## 11     4      2      4.5  4.8    6.5
## 12     4      3      1.2  1.8    2.7
```

Create a new data frame called `shrub_volumes` that includes all of the original data and a new column containing the volumes ( $\text{length} * \text{width} * \text{height}$ ), and display it.

```
mutate(shrubs, volume = length*width*height) %>%
  select(site, experiment, length, width, height, volume)
```

```
##      site experiment length width height  volume
## 1      1          1      2.2  1.3    9.6  27.456
## 2      1          2      2.1  2.2    7.6  35.112
## 3      1          3      2.7  1.5    2.2   8.910
## 4      2          1      3.0  4.5    1.5  20.250
## 5      2          2      3.1  3.1    4.0  38.440
## 6      2          3      2.5  2.8    3.0  21.000
## 7      3          1      1.9  1.8    4.5  15.390
## 8      3          2      1.1  0.5    2.3   1.265
## 9      3          3      3.5  2.0    7.5  52.500
## 10     4          1      2.9  2.7    3.2  25.056
## 11     4          2      4.5  4.8    6.5 140.400
## 12     4          3      1.2  1.8    2.7   5.832
```

## Exercise 2

Modify the code to calculate and print the average height of a plant in each experiment.

```
shrub_dims <- read.csv('../data-raw/shrub-volume-data.csv')
by_experiment <- group_by(shrub_dims, experiment)
avg_height_by_experiment <- summarize(by_experiment, avg_height = mean(height))
avg_height_by_experiment
```

```
## # A tibble: 3 x 2
##   experiment avg_height
##       <int>      <dbl>
## 1         1         4.7
## 2         2         5.1
## 3         3         3.85
```

Use `max()` to determine the maximum height of a plant at each site.

```
shrub_dims <- read.csv('../data-raw/shrub-volume-data.csv')
by_site <- group_by(shrub_dims, site)
max_height <- summarize(by_site, max_height = max(height))
max_height
```

```
## # A tibble: 4 x 2
##   site max_height
##   <int>      <dbl>
## 1     1         9.6
## 2     2         4
## 3     3         7.5
## 4     4         6.5
```

Write the same code but as a pipeline (using the pipe `|>` or `>`) to get the same result.

```
shrub_dims <- read.csv('../data-raw/shrub-volume-data.csv')
avg_height <- shrub_dims %>%
  group_by(site) %>%
  summarize(avg_height = mean(height))
```

```
shrub_dims <- read.csv('../data-raw/shrub-volume-data.csv')
avg_height_by_experiment <- shrub_dims %>%
  group_by(experiment) %>%
  summarize(avg_height = mean(height))
avg_height_by_experiment
```

```
## # A tibble: 3 x 2
##   experiment avg_height
##       <int>      <dbl>
## 1         1         4.7
## 2         2         5.1
## 3         3         3.85
```

```
shrub_dims <- read.csv('../data-raw/shrub-volume-data.csv')
max_height <- shrub_dims %>%
  group_by(site) %>%
  summarize(max_height = max(height))
max_height
```

```
## # A tibble: 4 x 2
##   site max_height
##   <int>      <dbl>
## 1     1         9.6
## 2     2         4
## 3     3         7.5
## 4     4         6.5
```

## Exercise 3

Fix the errors in the code so that it does what it's supposed to.

Add a comment to the end of each line of code explaining what it does.

When importing the data we need to ensure that we show the location of where it is, so our R studio can know where to find the file or dataset.

```
read.csv("../data-raw/shrub-volume-data.csv")
```

	site	experiment	length	width	height
## 1	1	1	2.2	1.3	9.6
## 2	1	2	2.1	2.2	7.6
## 3	1	3	2.7	1.5	2.2
## 4	2	1	3.0	4.5	1.5
## 5	2	2	3.1	3.1	4.0
## 6	2	3	2.5	2.8	3.0
## 7	3	1	1.9	1.8	4.5
## 8	3	2	1.1	0.5	2.3
## 9	3	3	3.5	2.0	7.5
## 10	4	1	2.9	2.7	3.2
## 11	4	2	4.5	4.8	6.5
## 12	4	3	1.2	1.8	2.7

We need to create a new object labeled as `shrub_data` using the data. After importing the data using the `read.csv` function, it is important to assign it to an object; in this case, `shrub_data`.

```
shrub_data <- read.csv("../data-raw/shrub-volume-data.csv")
```

After assigning the dataset to an object we can use the object and pipeline to start it off

```
shrub_data <- read.csv("../data-raw/shrub-volume-data.csv")
svc <- shrub_data %>%
  mutate(volume = length * width * height) %>%
  group_by(site) %>%
  summarize(mean_volume = mean(volume))
```

When using `summarize`, instead of the `mean` volume taking the average of the volumes, it took the maximum of the volumes, which is incorrect. To fix this, the `max` function was changed to the `mean` function.

```
shrub_data <- read.csv("../data-raw/shrub-volume-data.csv")
sve <- shrub_data %>%
  mutate(volume = length * width * height) %>%
  group_by(experiment) %>%
  summarize(mean_volume = mean(volume))
```

This shows the correct code to import the shrub volume data and calculate the average shrub volume for each site and, separately, for each experiment.

Homework: Shrub volume data set - part 2



## Exercise 4

Import the experiments data and then use `inner_join()` to combine it with the shrub dimensions data to add a “manipulation” column to the shrub data.

```
shrub_data <- read.csv("../data-row/shrub-volume-data.csv")
shrub_volume_experiments <- read.csv("../data-row/shrub-volume-experiments.csv")
shrub_volume_sites <- read.csv("../data-row/shrub-volume-sites.csv")

combined_sd <- inner_join(shrub_data, shrub_volume_experiments, by = "experiment")
combined_full_sd <- inner_join(combined_sd, shrub_volume_sites, by = "site")

combined_full_sd
```

```
##      site experiment length width height manipulation latitude longitude
## 1      1           1   2.2   1.3   9.6      control    29.65    -82.32
## 2      1           2   2.1   2.2   7.6        burn     29.65    -82.32
## 3      1           3   2.7   1.5   2.2      rainout    29.65    -82.32
## 4      2           1   3.0   4.5   1.5      control    29.26    -82.42
## 5      2           2   3.1   3.1   4.0        burn     29.26    -82.42
## 6      2           3   2.5   2.8   3.0      rainout    29.26    -82.42
## 7      3           1   1.9   1.8   4.5      control    29.80    -82.15
## 8      3           2   1.1   0.5   2.3        burn     29.80    -82.15
## 9      3           3   3.5   2.0   7.5      rainout    29.80    -82.15
## 10     4           1   2.9   2.7   3.2      control    29.99    -82.62
## 11     4           2   4.5   4.8   6.5        burn     29.99    -82.62
## 12     4           3   1.2   1.8   2.7      rainout    29.99    -82.62
##      elevation
## 1             54
## 2             54
## 3             54
## 4             50
## 5             50
## 6             50
## 7             57
## 8             57
## 9             57
## 10            62
## 11            62
## 12            62
```

## Exercise 5

The smallest value of length, width and height.

```
length <- c(2.2, 2.1, 2.7, 3.0, 3.1, 2.5, 1.9, 1.1, 3.5, 2.9)
width <- c(1.3, 2.2, 1.5, 4.5, 3.1, NA, 1.8, 0.5, 2.0, 2.7)
height <- c(9.6, 7.6, 2.2, 1.5, 4.0, 3.0, 4.5, 2.3, 7.5, 3.2)
```



The smallest value of: length: 1.1 width: 0.5 height: 1.5

The largest value of: length: 3.5 width: 4.5 height: 9.6

The sum value of: length: 25 width: 19.6 height: 45.4

The average value of: length: 2.5 width: 2.178 height: 4.540

```
min(length, na.rm = TRUE)
```

```
## [1] 1.1
```

```
min(width, na.rm = TRUE)
```

```
## [1] 0.5
```

```
min(height, na.rm = TRUE)
```

```
## [1] 1.5
```

```
max(length, na.rm = TRUE)
```

```
## [1] 3.5
```

```
max(width, na.rm = TRUE)
```

```
## [1] 4.5
```

```
max(height, na.rm = TRUE)
```

```
## [1] 9.6
```

```
sum(length, na.rm = TRUE)
```

```
## [1] 25
```

```
sum(width, na.rm = TRUE)
```

```
## [1] 19.6
```

```
sum(height, na.rm = TRUE)
```

```
## [1] 45.4
```

```
mean(length, na.rm = TRUE)
```

```
## [1] 2.5
```

```
mean(width, na.rm = TRUE)
```

```
## [1] 2.177778
```

```
mean(height, na.rm = TRUE)
```

```
## [1] 4.54
```

The volume of each shrub ( $\text{length} \times \text{width} \times \text{height}$ ). Storing this as an object or variable will make some of the next problems easier.

```
volume <- length*width*height  
volume
```

```
## [1] 27.456 35.112 8.910 20.250 38.440 NA 15.390 1.265 52.500 25.056
```

The sum of the volume of all of the shrubs. 224.379

```
sum(volume, na.rm = TRUE)
```

```
## [1] 224.379
```

A vector of the height of shrubs with lengths  $> 2.5$ .

```
height_2.5 <- subset(height, height > 2.5)  
height_2.5
```

```
## [1] 9.6 7.6 4.0 3.0 4.5 7.5 3.2
```

A vector of the height of shrubs with heights  $> 5$ .

```
height_5 <- subset(height, height > 5)  
height_5
```

```
## [1] 9.6 7.6 7.5
```

A vector of the heights of the first 5 shrubs (using []).

```
height_first_5 <- height[1:5]  
height_first_5
```

```
## [1] 9.6 7.6 2.2 1.5 4.0
```

A vector of the volumes of the first 3 shrubs (using []).

```
volume_first3 <- volume[1:3]
volume_first3
```

```
## [1] 27.456 35.112 8.910
```

A vector of the volumes of the last 5 shrubs with the code written so that it will return the last 5 values regardless of the length of the vector (i.e., it will give the last 5 values if there are 10, 20, or 50 individuals).

```
volume_last5 <- tail(volume, n = 5)
volume_last5
```

```
## [1] NA 15.390 1.265 52.500 25.056
```

## Exercise 6

A vector of shrub lengths

```
shrub_dimensions_labeled <- read.csv("../data-raw/shrub-dimensions-labeled.csv")
shrub_length <- shrub_dimensions_labeled$length
```

A vector of the volume of each of the shrubs

```
shrub_height <- shrub_dimensions_labeled$height
shrub_width <- shrub_dimensions_labeled$width
volume_dimensions <- shrub_length*shrub_height*shrub_width
```

A data frame with just the shrubID and height columns

```
df_id_height <- select(shrub_dimensions_labeled, shrubID, height)
```

A data frame with the second row of the full data frame

```
df_second_row <- shrub_dimensions_labeled[2,]
```

A data frame with the first 5 rows of the full data frame

```
df_first_five_rows <- shrub_dimensions_labeled[1:5,]
```