shrub_volume.Rmd

Marcos Padilla-Ruiz

2022-10-18

```
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
```

Exercise 1: data wrangling basics ## shrub data

The shrub volume dataset consists of a table made up of five columns labeled site, experiment, length, width, and height with a total of 12 rows

read data

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

selecting data

```
select(shrubs, length)
```

```
##    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(shrubs, site, experiment)
```

```
##    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
```

adding clomun using mutate function

```
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 data by length with arrange function

```
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 data

```
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
```

```
# shrubs with height greater than 5

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
```

```
# shrubs with height greater than 4 and width greater than 2

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
```

```
# shrubs from experiment 1 or 3

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
```

```
# filter to remove null values from height column
```

new shrub data frame with added column

```
shrub_volumes <- mutate(shrubs, volumes = length*width*height)

shrub_volumes
```

```
##    site experiment length width height volumes
## 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: data aggregation

rewrite code as pipeline for maximum plant height

```
avg_height <- shrubs |>
  group_by(site) |>
  summarize(avg_ht = max(height))


avg_height
```

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

Exercise 3: fix the code

```r
shrubs <- read.csv(file = "../raw_data/shrub-volume-data.csv") # they did not specify the location of t


shrub_data_site <- shrubs |> # 'shrub_data_site' is the name of the new data frame we're creating and '
  mutate(volume = length * width * height) |> # this line creates a new column labeled 'volume' and is
  group_by(site) |> # this function allows us to group the rows by the values in the site column
  summarize(mean_volume = max(volume)) # this summarizes the maximum volumes in each site into one new

shrub_data_site
```

```
## # A tibble: 4 x 2
##    site mean_volume
##   <int>       <dbl>
## 1     1        35.1
## 2     2        38.4
## 3     3        52.5
## 4     4       140.
```

```r
shrub_data_exp <- shrubs |> # 'shrub_data_exp' is the new data frame and we are once again working with
  mutate(volume = length * width * height) |> # creates new 'volume' column
  group_by(experiment) |> # groups rows by the values in the experiment column
  summarize(mean_volume = mean(volume)) # summarizes the average volumes within each experiment

shrub_data_exp
```

```
## # A tibble: 3 x 2
##   experiment mean_volume
##        <int>       <dbl>
## 1          1        22.0
## 2          2        53.8
## 3          3        22.1
```