hw11

Tom Lever

11/11/22

Metadata

Name: hw11

URL: https://github.com/tslever/DS5100-2022-08-tsl2b/blob/main/lessons/M11/hw11.Rmd

Course: DS 5100

Term: Fall 2022 Online

Module: M11: R Programming

Topic: Tidyverse

Author: Tom Lever

Net ID: tsl2b

Created: 11 November 2022

Updated: 11 November 2022

Overview

In this homework, you will work on the Abalone datset from the UCI Machine Learning Repository.

To get started, download and import the `abalone.data` dataset from this URL:

- https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data

You can pass the URL directly to `read.csv`. There is no header row.

Note: The instruction to print in the questions below can be accomplished either through the `print` function or by displaying a value directly.

Total Points: 7

Tasks

Task 0

(0 points)

Get the dataset.

```
data_set <- read.csv(
   'https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data',
   header = FALSE
```

```
)
head(data_set, n = 3)
```

```
##   V1    V2    V3    V4     V5     V6     V7   V8 V9
## 1  M 0.455 0.365 0.095 0.5140 0.2245 0.1010 0.15 15
## 2  M 0.350 0.265 0.090 0.2255 0.0995 0.0485 0.07  7
## 3  F 0.530 0.420 0.135 0.6770 0.2565 0.1415 0.21  9
```

Task 1

(1 point)

Print the number of rows in the dataset.

```
print(nrow(data_set))
```

```
## [1] 4177
```

Task 2

(1 point)

The rightmost column is the number of rings. Print the maximum number of rings.

```
library(dplyr)
print(data_set %>% select(V9) %>% max())
```

```
## [1] 29
```

Task 3

(1 point)

The leftmost column is the gender with values `M` (Male), `F` (Female), and `I` (Infant).

Apply the `filter` function from `tidyverse` to select only rows where gender is infant, and print the number of records.

```
print(data_set %>% filter(V1 == 'I') %>% nrow())
```

```
## [1] 1342
```

Task 4

(1 point)

Apply the `filter` function from `tidyverse` to select only rows where gender is infant or male, and print the number of records.

```
print(data_set %>% filter (V1 == 'I' | V1 == 'M') %>% nrow())
```

```
## [1] 2870
```

Task 5

(1 point)

Call the `table` function on the abalone genders to find out how many of each gender are present.

```
data_set %>% select(V1) %>% table()
```

```
## V1
##    F    I    M
## 1307 1342 1528
```

Task 6

(1 point)

Compute the mean value of column 2 (V2) grouped by gender.

V2 is the longest shell measurement.

Requirements: Use the `%>%` operator to chain commands, and the `group_by` and `summarize` functions.

```
data_set %>% group_by(V1) %>% summarize(average = mean(V2))
```

```
## # A tibble: 3 x 2
##   V1    average
##   <chr>   <dbl>
## 1 F       0.579
## 2 I       0.428
## 3 M       0.561
```

Task 7

(1 point)

Compute the median value of longest shell measurement for only the males.

Requirements: Use the `%>%` operator to chain commands.

```
data_set %>% filter(V1 == 'M') %>% summarize(the_median = median(V2))
```

```
##   the_median
## 1       0.58
```

```
data_set %>% group_by(V1) %>% summarize(the_median = median(V2)) %>% filter(V1 == 'M')
```

```
## # A tibble: 1 x 2
##   V1    the_median
##   <chr>      <dbl>
## 1 M           0.58
```