

Metadata

Course: DS 5100
Module: 11 R Programming 2
Topic: HW on Tidyverse
Author: R.C. Alvarado (adapted)
Date: 7 July 2023 (revised)

Student Info

Name: Hilde Younce
Net ID: ksg8xy
File GitHub URL: <https://github.com/hyounce/DS5100-ksg8xy/blob/main/lessons/M11/M11-HW-2.Rmd>

Instructions

In your **private course repo** use this notebook to write code that performs the tasks below.

Save your notebook in the `M11` directory.

Remember to add and commit these files to your repo.

Then push your commits to your repo on GitHub.

Be sure to fill out the **Student Info** block above.

To submit your homework, save your results as a PDF and upload it to GradeScope.

TOTAL POINTS: 7

Overview

In this homework, you will work with the Abalone dataset (<https://archive.ics.uci.edu/ml/datasets/Abalone>) 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>
(<https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data>)

You can pass the URL directly to `read.csv()` and that 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.

```
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

abalone = read.csv("https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data")
abalone
```

M	X0.455	X0.365	X0.095	X0.514	X0.2245	X0.101	X0.15	X15					
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>					
M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.0700	7					
F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.2100	9					
M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.1550	10					
I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.0550	7					
I	0.425	0.300	0.095	0.3515	0.1410	0.0775	0.1200	8					
F	0.530	0.415	0.150	0.7775	0.2370	0.1415	0.3300	20					
F	0.545	0.425	0.125	0.7680	0.2940	0.1495	0.2600	16					
M	0.475	0.370	0.125	0.5095	0.2165	0.1125	0.1650	9					
F	0.550	0.440	0.150	0.8945	0.3145	0.1510	0.3200	19					
F	0.525	0.380	0.140	0.6065	0.1940	0.1475	0.2100	14					
1-10 of 4,176 rows				Previous	1	2	3	4	5	6	...	418	Next

Task 1

(1 point)

Print the number of rows in the dataset.

```
nrow(abalone)
```

```
## [1] 4176
```

Task 2

(1 point)

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

```
max_rings <- abalone %>%  
  arrange(desc(X15))  
print(max_rings[1,9])
```

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

Task 3

(1 point)

The leftmost column is the gender with these values: M : male, F : female, I : infant.

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

```
abalone %>%  
  filter(M == "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.

```
abalone %>%  
  filter(M == "I" | M == "M") %>%  
  nrow()
```

```
## [1] 2869
```

Task 5

(1 point)

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

Print the result.

```
abalone %>%
  select(M) %>%
  table()
```

```
## M
##   F   I   M
## 1307 1342 1527
```

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.

```
abalone %>%
  group_by(M) %>%
  summarize(mean = mean(X0.455))
```

M	mean
<chr>	<dbl>
F	0.5790933
I	0.4277459
M	0.5614604
3 rows	

Task 7

(1 point)

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

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

```
abalone %>%
  filter(M == "M") %>%
  summarize(median = median(X0.455))
```

		median
		<dbl>
		0.58
1 row		