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:  Courtney Hodge
Net ID: yss2zv
File GitHub URL: https://github.com/courtneyhodge/DS5100-2024-06-R/blob/main/hw/M11/M11-HW.pdf
```

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 GitHib.

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.

```
# CODE HERE
df <- read.csv("https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data")

df
```

M <chr>	X0.455 <dbl>	X0.365 <dbl>	X0.095 <dbl>	X0.514 <dbl>	X0.2245 <dbl>	X0.101 <dbl>	X0.15 <dbl>	X15 <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.

```
print(nrow(df))
```

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

Task 2

(1 point)

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

```
# CODE HERE
print(max(df$X15))
```

```
## [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.

```
# CODE HERE
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
task3 <- df |>
  filter(M == "I")

print(nrow(task3)) #print the number of records
```

```
## [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.

```
# CODE HERE
task4 <- df |>
  filter(M == "M" | M == "I")

print(nrow(task4))
```

```
## [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.

```
# CODE HERE
table(df$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.

```
# CODE HERE
task6 <- df |>
  group_by(M) |> #this is column 2 of our abalone df
  summarize(mean(X0.455))

print(task6)
```

```
## # A tibble: 3 × 2
##   M     `mean(X0.455)`
##   <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.

```
# CODE HERE
task7 <- df |>
  filter(M == "M") |>
  summarize(median(X0.455))

print(task7)
```

```
##   median(X0.455)
## 1           0.58
```