Metadata

```
Course:   DS 5100
Module:   11 R Programming 2
Topic:    HW on Tidyverse
Author:   R.C. Alvarado (adapted)
Date:     6 July 2022
```

Student Metadata

```
Name:  Rachel Grace
UVA ID: rg5xm
```

Instructions

Perform the tasks below to write the necessary code and include all solutions.

Read about the Abalone dataset here.

Grab the `abalone.data` dataset from this URL:

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

Hine 1: You can pass the URL directly to `read.csv()`.

Hint 2: there is no header row.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

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

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

Questions

Q1

(1 POINT)

Print the number of rows in the dataset.

```
# CODE HERE
nrow(dataset)
```

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

Q2

(1 POINT)

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

```
# CODE HERE
max(dataset$X15)
```

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

Q3

(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
records <- dataset %>%
  filter(M =="I")
nrow(records)
```

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

Q4

(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
records2 <- dataset %>%
  filter(M == "I" | M == "M")
nrow(records2)
```

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

Q5

(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(dataset$M)
```

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

Q6

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

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

Q7

(1 POINT)

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

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

```
# CODE HERE
dataset %>%
  filter(M == "M") %>%
  summarize(median=median(X0.455))
```

```
##   median
## 1   0.58
```

Submission

Save as PDF and upload to Gradescope.