In [ ]:

```
---
editor_options:
  markdown:
    wrap: 72
output:
  pdf_document: default
  html_document:
    df_print: paged
---

# 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:  Kyler Halat-Shafer
    UVA ID: uxt5qb

# Instructions

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

Read about the Abalone dataset
[here](https://archive.ics.uci.edu/ml/datasets/Abalone).

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

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

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

Hint 2: 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**

# Questions

## Q1

(1 POINT)
```

Print the number of rows in the dataset.

```{r}
data <- read.csv('abalone.data', header = F)
nrow(data)

#Solution:4177
```

Q2

(1 POINT)

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

```{r}
#The rightmost column is column 9, which you can call using [], then wrappin
# in the that column
#data[9]
max(data[9])

#Solution: 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.

```{r}
nrow(data%>%
  filter(V1 == 'I'))

#Solution: 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.

```{r}
nrow(data%>%
  filter(V1 == 'I'| V1 == 'M'))
```

```
#Solution: 2870
```

Q5

(1 POINT)

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

Print the result.

```{r}
table(data[1])

#Solution:
# F = 1307
# I = 1342
# M = 1528
```

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.

```{r}
data %>%
  group_by(V1) %>%
  summarize(mean(V2))

#Solution:
#F      0.5790933
#I      0.4277459
#M      0.5613907
```

Q7

(1 POINT)

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

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

```{r}
```

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

#Solution:
#0.58

#I just wanted to do a quick test to see that this was correct since the M c
#data%>%
#  group_by(V1)%>%
#  summarize(median(V2))
```

```
# Submission

Save as PDF and upload to Gradescope.
```