**PDAT 610G: Module 2a Assignment**

**Introduction**

UC-Irvine has a famous repository of interesting data sets that are useful for data science and machine learning.  
Consider this dataset on over four thousand Tasmanian Abalone, a yummy sea snail, from the UCI machine learning repository. <https://archive.ics.uci.edu/ml/datasets/Abalone>

Apparently, they are quite delicious, and both the sex and age of the snail influences its taste (and price). You can tell how old an abalone is by opening it and counting the rings, but then it’s no good to eat, so researchers were interested in estimating the age from its size. I have no idea how you tell the sex of an abalone, and apparently, neither could the abalone scientists, since many are marked as I, which means “infant” or “inconclusive.”  
<https://en.wikipedia.org/wiki/Abalone>

Abalone Dataset: *n* = 4177, variables = 10 (1 categorical, 1 integer, 8 numeric)  
 Someone in Tasmania collected 4,000 delicious measurements of abalone sizes   
Categorical: Sex (M, F, I), (I=Indeterminate sex or infant, so the person couldn’t tell)  
Integer: rings (in general, an abalone gets a new ring each year, sort of like a tree)  
Num: Length, Diameter, Height, whole.weight, shucked.weight, viscera.weight, shell.weight  
Units for Numerical Variables: mm and g, as appropriate

**Submission Method**

* Create an R Markdown document containing both your code and any written responses or explanations.
* Make sure to number your answers so that I can easily see which questions you were answering
* The code for any question part should be contained in a single code block.
* When you’re done, knit your code to a PDF file, and submit that file through Blackboard.

**Assignment**

Using abalone.data we discussed on the slide, answer the following questions.

1. What is the class of the data abalone.data? (Use class() function)
2. What is the datatype of variable diameter? (Use typeof() function)
3. Use the function summary() to find basic description of diameter and age (rings) of abalone.
4. Use the function mean() to find the mean of diameter of female abalone.
5. Assume Y: diameter and X: rings. What is? (Hit: Some matrix concepts are used here. means X transpose and means inverse of . You want to find the R functions for transpose and inverse of the matrix. I intentionally left this out so you can practice to find appropriate functions using google.)
6. Can you think of an interesting question about Abalone? (For example, do female abalone weigh more than male abalone?) Make your own question and answer it. You can use mean and standard deviation (function sd()) to answer the question. If you wish, you can also use some graphs and even hypothesis test (we have not discussed about these topics yet thus it is not mandatary for this assignment. We will discuss all of these in later modules. Try to figure out as much as you can and as much as you want).