# STAT 4410/8416 Homework 1

*lastName firstName*

*Due on Sep 11, 2018*

1. Based on your reading assignments answer the following questions

   a) What is data science?
   b) Explain with an example what you mean by data product.
   c) Carefully read the Cleveland's paper shown in lecture 2 and discuss what he suggested about the field of statistics and data science.
   d) Explain in a short paragraph how data science is different from computer science.
   e) What is data literacy? Is it important to be data literate in this modern world? Explain why or why not.
   f) In his article, Donoho discussed about common task framework. Explain what it is and why he mentioned it.
   g) According to Donoho, what are the activities of greater data science?

2. What are the very first few steps one should do once data is loaded onto **R**? Demonstrate that by loading tips data from http://www.ggobi.org/book/data/tips.csv

3. In our **R** class we learnt about recursive functions that produce a sequence of numbers upto a given number say $n$ as demonstrated in the following codes.

```r
foo <- function(x){
  print(x)
  if(x>1) foo(x-1)
}

moo <- function(x){
  if(x>1) moo(x-1)
  print(x)
}

foo(3)
```

```
## [1] 3
## [1] 2
## [1] 1
```

```r
moo(3)
```

```
## [1] 1
## [1] 2
## [1] 3
```

Explain why function `moo()` prints 1 through 3 while function `foo()` prints from 3 through 1.

4. The function `sqrt()` provides the square root of any number. However, it can't provide any square root of negative number. We want to create our own function to provide a message for negative number.
   a) Create a new **R** function `getRoot()` that will provide square root of any number. If the number is negative it should return 'not possible'. Demonstrate your function such that it produces the following outputs. `getRoot(4) = 2`, `getRoot(-4) = 'not possible'`
   b) Does your function produce expected results for vector input as well? For example does it give the following result? Explain why or why not. `getRoot(c(4,-4, 9, -16))=2 'not possible' 3 'not possible'`
   c) If your function does not work as expected, how can you make it work properly?

5. Write a program that will do the following. Include your codes and necessary outputs to demonstrate your work.
   a) Generate 500000 random numbers from an exponential distribution with rate=0.2 and store these numbers in a vector called myVector. **Report** a histogram of the numbers you just generated.
   b) Convert myVector into a matrix of 500 columns and assign it to an object called myMatrix. **Report** the dimension of myMatrix.
   c) Compute the column means of myMatrix. **Report** a histogram of those column means.
   d) Explain why the two histograms you have created in questions (4a) and (4c) are different in shapes.
6. This problem will give you some practice with creating and manipulating vectors.
   a) Using `seq()`, create a vector consisting of an arithmetic sequence of integers from 0 to 12 with a common difference of 3 stored in a variable called mySeq. Report mySeq.
   b) Describe how the different arguments in each of the three following commands changes the output of `rep()`: `rep(mySeq,3)`, `rep(mySeq,each=3)`, and `rep(mySeq,mySeq)`.
   c) Concatenate the sequence 1:6 to the end of the vector described by `rep(mySeq,mySeq)` and store the resulting vector in the same mySeq variable. Report the length of mySeq.
   d) Create a square matrix populated row-wise from your mySeq vector and store it in a variable called sqMtrx. Report the element located in the 4th row and 3rd column of sqMtrx.