CIND 123 Summer 2019 - Assignment #1

Tasdeed Aziz

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

Use RStudio for this assignment. Edit the file `A1-S19-Q.Rmd` and insert your R code where wherever you see the string "#INSERT YOUR ANSWER HERE"

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

Sample Question and Solution

Use `seq()` to create the vector $(1, 2, 3, \ldots, 10)$.

```
seq(1,10)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

Question 1

a) Use the `seq()` function to create the vector $(1, 7, 13, \ldots, 61)$. Note that each term in this sequence is of the form $1 + 6n$ where $n = 0, \ldots, 10$.

```
seq(1,61,by=6)
```

```
##  [1]  1  7 13 19 25 31 37 43 49 55 61
```

b) Use `seq()` and `c()` to create the vector $(1, 2, 3, \ldots, 10, 9, 8, \ldots, 3, 2, 1)$.

```
c(1:10,seq(9,1,by=-1))
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10  9  8  7  6  5  4  3  2  1
```

c) Use 'seq()' function to create a vector with starting value 1.725 and end value 5.725. Each value in the sequence should be increased by 0.5.

```
seq(1.725,5.725,by=0.5)
```

```
## [1] 1.725 2.225 2.725 3.225 3.725 4.225 4.725 5.225 5.725
```

d) Execute the following commands: seq_len(17), seq_along(1:17), seq(1,17). Write your observations.

```r
## The commands computes output of same length
seq_len(17)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
```

```r
seq_along(1:17)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
```

```r
seq(1:17)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
```

Question 2

a) Compute:

$$\sum_{n=1}^{100} n$$

```r
sum(seq(1,100))
```

```
## [1] 5050
```

b) Compute:

$$\sum_{n=10}^{100} (n-3)^5$$

```r
sum((seq(10,100)-3)^5)
```

```
## [1] 143159211832
```

c) Create a sequence of odd numbers between 1 to 100 in a vector. Find the sum of those odd numbers.

```r
a<-seq(1,100,by=2)
a
```

```
##  [1]  1  3  5  7  9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
## [26] 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99
```

```r
sum(a)
```

```
## [1] 2500
```

d) What are the outputs for the following cases: (i) sum(1:5, NA), and (ii) sum(1:5, NA, na.rm = TRUE). Give your intuition or reasonging behind each.

```r
#(i)Does not output the sum as result of 'NA' (missing values)
sum(1:5,NA)
```

```
## [1] NA
```

```r
#(ii) Computes the sum as 'NA' (missing values) are replace
sum(1:5,NA,na.rm=t)
```

```
## [1] 15
```

Question 3

a) Create an empty list `mylist`.

```r
mylist <- list()
```

b) Add a component named `firstAttr` whose value is 42.

```r
mylist$firstAttr <-42
```

c) Add a component named `secondAttr` a 4x3 matrix whose elements are $(1, 2, \ldots, 12)$ in row-wise order.

```r
mylist$secondAttr <- matrix(1:12,nrow=4)
```

d) Add a component named `thirdAttr` a 4x3x2 array whose elements are $(1, 2, \ldots, 24)$.

```r
mylist$thirdAttr <- array(1:24,c(4,3,2))
mylist
```

```
## $firstAttr
## [1] 42
##
## $secondAttr
##      [,1] [,2] [,3]
## [1,]    1    5    9
## [2,]    2    6   10
## [3,]    3    7   11
## [4,]    4    8   12
##
## $thirdAttr
## , , 1
##
##      [,1] [,2] [,3]
## [1,]    1    5    9
## [2,]    2    6   10
## [3,]    3    7   11
## [4,]    4    8   12
##
```

```
## , , 2
##
##      [,1] [,2] [,3]
## [1,]   13   17   21
## [2,]   14   18   22
## [3,]   15   19   23
## [4,]   16   20   24
```

Question 4

`iris` data set gives the measurements in centimeters of the variables sepal length, sepal width, petal length and petal width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica.

Install the 'iris' data set on your computer using the command `install.packages("datasets")`. Then load the `datasets` package into your session using the following command.

```
library(datasets)
```

a) Display the first 6 rows of the `iris` data set

```
head(iris,n=6)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

b) Compute the average of the first four variables (Sepal.Length, Sepal.Width, Petal.Length and Petal.Width) using `sapply()` function.

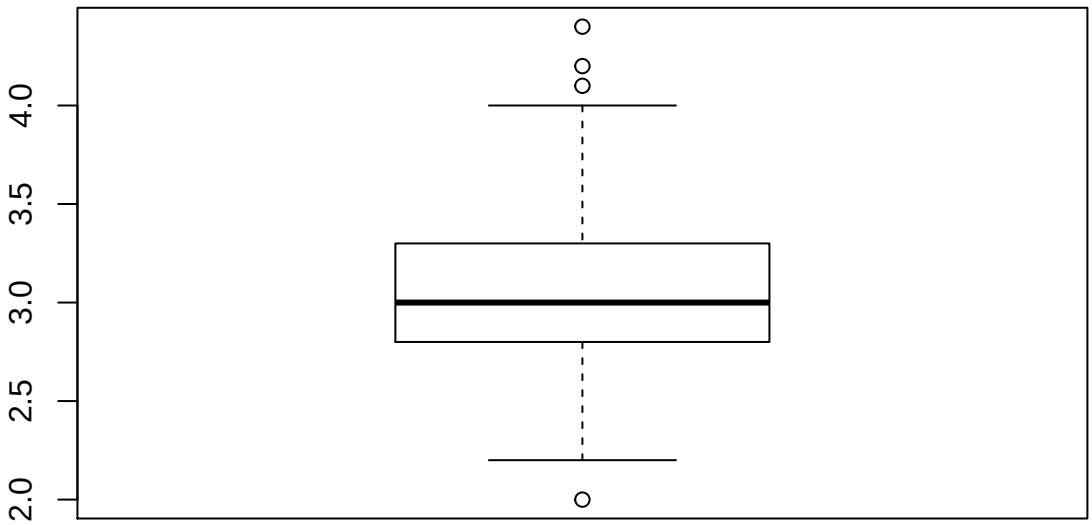
Hint: You might need to consider removing the `NA` values, otherwise the average will not be computed.

```
sapply(iris[1:4],mean, use.names=T, na.rm=T)
```

```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##     5.843333     3.057333     3.758000     1.199333
```

c) Construct a boxplot for `Sepal.Width` variable, then display all the outliers.

```
boxplot(iris[2:2])
```



d) Compute the lower and the upper quartiles of `Sepal.Width` variable

```
a <- iris$Sepal.Width
quantile(a,0.25)
```

```
## 25%
## 2.8
```

```
quantile(a,0.75)
```

```
## 75%
## 3.3
```

##Question 5

Install the 'mtcars' data set on your computer using the command `install.packages("datasets")`. Then load the `datasets` package into your session using the following command.

```
library(datasets)
```

a) Attache `mtcars` dataset into your session, then assign the dataset to a new variable called `newmtcarsData`.

Hint: You can use 'attach()' function to call the dataset in the current R session.

```
attach(mtcars)
newmtcarsData <- mtcars
detach(mtcars)
```

b) Display the structure of `newmtcarsData` using `str()` function, then change the data type of the `am` variable from `numeric` to `factor` with the following labels: `manual` for `1` and `automatic` for `0`.

```r
str(newmtcarsData)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```r
newmtcarsData$am <- replace(newmtcarsData$am,newmtcarsData$am==0,'automatic')
newmtcarsData$am <- replace(newmtcarsData$am,newmtcarsData$am==1,'manual')
newmtcarsData$am <- as.factor (newmtcarsData$am)
str(newmtcarsData)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : Factor w/ 2 levels "automatic","manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

c) Obtain the dimension of the `newmtcarsData` using `dim()` function, then write a command to display only the number of columns of `newmtcarsData` variable.

```r
dim(newmtcarsData)
```

```
## [1] 32 11
```

```r
ncol(newmtcarsData)
```

```
## [1] 11
```

d) Get the five point summary for "new.data" using `summary()` function, then write a command to dispaly only the maximum value of each attribute.

```r
summary(newmtcarsData)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##          am          gear            carb      
##  automatic:19   Min.   :3.000   Min.   :1.000  
##  manual   :13   1st Qu.:3.000   1st Qu.:2.000  
##                 Median :4.000   Median :2.000  
##                 Mean   :3.688   Mean   :2.812  
##                 3rd Qu.:4.000   3rd Qu.:4.000  
##                 Max.   :5.000   Max.   :8.000
```

```r
#Position of variable 'am' has been change as max cannot be calculated on factorial
new <- newmtcarsData[c(1,2,3,4,5,6,7,8,10,11,9)]
sapply(new[1:10],max,na.rm=t)
```

```
##     mpg     cyl    disp      hp    drat      wt    qsec      vs    gear    carb 
##  33.900   8.000 472.000 335.000   4.930   5.424  22.900   1.000   5.000   8.000
```

END of Assignment #1.