Lab Assignment-4

Name: Khushi Prasad
SubGroup: 3CO20
Roll No.: 102183044
Subject: Data Science Fundamentals

Q1. Vector creation Write R code to generate the following vectors, explore the functions seq() and rep() using the help on commands:

- 1.3 1.6 1.9 2.2 2.5 2.8 3.1 3.4 3.7 4.0 4.3 4.6 4.9
- 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
- 14 12 10 8 6 4 2 0
- 5 5 12 12 13 13 20 20

CODE:

```
a<-seq(1.3,4.9,0.3)
print(a)
b<-rep(c(1,2,3,4),5)
print(b)
c<-seq(14,0,-2)
print(c)
d<-rep(c(5,12,13,20),each=2)
print(d)
```

OUTPUT:

```
> print(a)
 [1] 1.3 1.6 1.9 2.2 2.5 2.8 3.1 3.4 3.7 4.0 4.3 4.6 4.9
> print(b)
 [1] 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
> print(c)
[1] 14 12 10  8  6  4  2  0
> print(d)
[1]  5  5 12 12 13 13 20 20
```

Q2. Loading and exploring data structure Load the iris data that R provides internally by typing data(iris)

A. What sort of data type is iris?

B. How many rows (observations) and columns (variables) does the iris dataset have?

C. Which variable of the data frame iris is a factor and how many levels does it have?

CODE:

```
data(iris)
# A
structure(iris)
typeof(iris)
class(iris)
# B
ncol(iris)
nrow(iris)
# C
Filter(is.factor,iris)
levels(iris$Species)
```

OUTPUT:

```
> structure(iris)
    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
> typeof(iris)
[1] "list"
> class(iris)
[1] "data.frame"
> # B
> ncol(iris)
[1] 5
> nrow(iris)
[1] 150
> # C
> Filter(is.factor,iris)
       Species
1       setosa
2       setosa
51  versicolor
52  versicolor
101  virginica
102  virginica
> levels(iris$Species)
[1] "setosa"     "versicolor" "virginica"
```

Q3. Use the “iris” dataset to find
a) The mean and standard deviation of the sepal width and sepal length for each type of species.
b) Create a new dataset called iris.class from the iris dataset. Use a loop and ifelse statement to create a vector in the iris.class dataset called Calyx.Width, which is “short” if Sepal.Length is less than 5, and otherwise is “long.” (The sepals of a flower are collectively known as the calyx.)
CODE:

```
25  # A
26  s1<-subset(iris,Species=="setosa")
27  mean(s1$Sepal.Length)
28  sd(s1$Sepal.Length)
29  s2<-subset(iris,Species=="virginica")
30  mean(s2$Sepal.Length)
31  sd(s2$Sepal.Length)
32  s3<-subset(iris,Species=="versicolor")
33  mean(s3$Sepal.Length)
34  sd(s3$Sepal.Length)
35  mean(s1$Sepal.Width)
36  sd(s1$Sepal.Width)
37  mean(s2$Sepal.Width)
38  sd(s2$Sepal.Width)
39  mean(s3$Sepal.Width)
40  sd(s3$Sepal.Width)
41
42  # B
43  iris.class<-iris
44  Calyx.width<-vector(mode="character",length=150)
45  for(i in 1:150){
46    if(iris.class$Sepal.Length[i]<5)
47      Calyx.width[i]<-"short"
48    else
49      Calyx.width[i]<-"long"
50  }
51  iris.class$Calyx.width<-Calyx.width
52  head(iris.class)
```

OUTPUT:

```
> s1<-subset(iris,Species=="setosa")
> mean(s1$Sepal.Length)
[1] 5.006
> sd(s1$Sepal.Length)
[1] 0.3524897
> s2<-subset(iris,Species=="virginica")
> mean(s2$Sepal.Length)
[1] 6.588
> sd(s2$Sepal.Length)
[1] 0.6358796
> s3<-subset(iris,Species=="versicolor")
> mean(s3$Sepal.Length)
[1] 5.936
> sd(s3$Sepal.Length)
[1] 0.5161711
> mean(s1$Sepal.Width)
[1] 3.428
> sd(s1$Sepal.Width)
[1] 0.3790644
> mean(s2$Sepal.Width)
[1] 2.974
> sd(s2$Sepal.Width)
[1] 0.3224966
> mean(s3$Sepal.Width)
[1] 2.77
> sd(s3$Sepal.Width)
[1] 0.3137983
```

```
> head(iris.class)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species Calyx.width
1          5.1         3.5          1.4         0.2  setosa        long
2          4.9         3.0          1.4         0.2  setosa       short
3          4.7         3.2          1.3         0.2  setosa       short
4          4.6         3.1          1.5         0.2  setosa       short
5          5.0         3.6          1.4         0.2  setosa        long
6          5.4         3.9          1.7         0.4  setosa        long
```

Q4. Explore dataset- mtcars in R. You can get the structure and column names of data by typing the command str(mtcars) and names(mtcars) respectively. Write your code to subset the dataset- mtcars according to the following requirements (NOTE: each requirement is independent.)

A. Select cars whose cyl (a column in the dataset) value is no smaller than 5.

B. Show all the fields (columns) of the first 10 cars.

C. Find all cars matching “Honda”.

CODE:

```
54  # Q4
55  data(mtcars)
56  str(mtcars)
57  names(mtcars)
58
59  # A
60  subset(mtcars, cyl>5)
61
62  # B
63  head(mtcars,10)
64
65  # C
66  mtcars[substr(row.names(mtcars),1,5)=="Honda",]
```

OUTPUT:

```
> # Q4
> data(mtcars)
> str(mtcars)
'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 ...
> names(mtcars)
 [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear" "carb"
> # A
> subset(mtcars, cyl>5)
                   mpg cyl  disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
Merc 280C         17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
Merc 450SE        16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
Merc 450SL        17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
Merc 450SLC       15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
> # B
> head(mtcars,10)
                   mpg cyl  disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
> # C
> mtcars[substr(row.names(mtcars),1,5)=="Honda",]
            mpg cyl disp hp drat    wt  qsec vs am gear carb
Honda Civic 30.4   4 75.7 52 4.93 1.615 18.52  1  1    4    2
>
```