Part A

(List of R programs)

1. Create vector for the following
   a. (4, 6, 3, 4, 6, 3, . . . , 4, 6, 3) where there are 10 occurrences of 4.
   b. Use the function paste to create the following character vectors of length 30 ("fn1", "fn2", ..., "fn30").In this case, there is no space between fn and the number

2. **a.** Turn the vector of character items "Control", "Control", "Control", "Ear Removal", "Ear Removal", "Ear Removal", "Ear Removal", "Fake Ear Removal", "Fake Ear Removal", "Fake Ear Removal", "Fake Ear Removal" into a Factor variable and create a table from it to show the number of entries in each treatment.
   b. Create a vector of character variables that contains 25 ”a”, 15 ”b”, and 58 ”c” instances. What is the length of this vector? Create a table from the entries.
3. **a.** Create three different variables, one that is numeric type and other two are vector of characters. Use these to create data frame of student.(USN,Name,Marks)
   b. Add a new numeric data column to the existing data frame (Age). Provide summary of the data
   c. Display the list of student whose Age is less than 20 and Marks greater than 25

4. Write a program to create the csv file for storing Employee data. Containing the data (EmpID, EmpName , DOJ, EmpCode, Dept, Desig.)
   a. Read the suitable number of employee details from the user.
   b. Create a dataframe of Employee
   c. Store the dataframe in the csv file
   d. Check the difference between csv and csv2 file
   e. Read the data from csv and Display the contents
   f. Append a new row into the csv file

5. Dataset example
   a. List the data set available in your system using suitable command
   b. Select “mtcars” data set, find and display the number of rows and columns in that data set
   c. Find are there more `automatic (0)` *or* `manual (1)` transmission-type cars in the dataset? *Hint: 9th column indicate the transmission type*
   d. Get a scatter plot of ‘hp’ vs ‘weight’.
   e. Change ‘am’, ‘cyl’ and ‘vs’ to *integer* and store the new dataset as ‘newmtc’.
   f. Extract the cases where cylinder is less than 5

6. Consider “Airquality” dataset
   a. Display the dimension of the dataset
   b. Display the class of each fields in the data set
   c. Test the missing values
   d. Recode the missing values, as mean of the column values
   e. Exclude the missing values

Solutions

1. Create vector for the following

a. (4, 6, 3, 4, 6, 3, . . . , 4, 6, 3) where there are 10 occurrences of 4.

b. Use the function paste to create the following character vectors of length 30 ("fn1", "fn2", ..., "fn30").In this case, there is no space between fn and the number

a. tmp <- c(4,6,3) # Create the vector
rep(tmp,10) #Repeat the vector 10 times

b. paste("fn",1:30,sep="") # paste 1st and 2nd argument

2. **a.** Turn the vector of character items "Control", "Control", "Control", "Ear Removal", "Ear Removal", "Ear Removal", "Ear Removal", "Fake Ear Removal", "Fake Ear Removal", "Fake Ear Removal", "Fake Ear Removal" into a Factor variable and create a table from it to show the number of entries in each treatment.

b. Create a vector of character variables that contains 25 ”a”, 15 ”b”, and 58 ”c” instances. What is the length of this vector? Create a table from the entries.

```
a. # Create the vector of strings
 x<-c("Control", "Control", "Control", "Ear Removal", "Ear Removal", "Ear Rem
oval", "Ear Removal", "Fake Ear Removal", "Fake Ear Removal", "Fake Ear Remov
al", "Fake Ear Removal")

# display the vector
> x
 [1] "Control"          "Control"          "Control"          "Ear Removal"
 [5] "Ear Removal"      "Ear Removal"      "Ear Removal"      "Fake Ear Remov
al"
 [9] "Fake Ear Removal" "Fake Ear Removal" "Fake Ear Removal"

#construct factor from the vector
> xfact<- factor(x)

#Display the vector
> xfact
 [1] Control          Control          Control          Ear Removal
 [5] Ear Removal      Ear Removal      Ear Removal      Fake Ear Removal
 [9] Fake Ear Removal Fake Ear Removal Fake Ear Removal
Levels: Control Ear Removal Fake Ear Removal
> nlevels(xfact)
[1] 3
```

2b.

#Create the vector

```
> x<-c(rep("a",25),rep("b",15),rep("c",58))
> x
```

```
 [1] "a" "a" "a" "a" "a" "a" "a" "a" "a" "a" "a" "a" "a" "a" "a" "a" "a" "a"
"a" "a"
[21] "a" "a" "a" "a" "a" "b" "b" "b" "b" "b" "b" "b" "b" "b" "b" "b" "b" "b"
"b" "b"
[41] "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c"
"c" "c"
[61] "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c"
"c" "c"
[81] "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c" "c"
> length(x)  # Find the length of the vector
[1] 98
> table1<- data.frame(x) # Construct table from the vector
> table1
   x
1  a
2  a
3  a
4  a
5  a
6  a
|
|
|
|
|
93 c
94 c
95 c
96 c
97 c
98 c
```

3. **a.** Create three different variables, one that is numeric type and other two are vector of characters. Use these to create data frame of student.(USN,Name,Marks)

b. Add a new numeric data column to the existing data frame (Age). Provide summary of the data

c. Display the list of student whose Age is less than 20 and Marks greater than 25

```
n <- as.integer(readline(prompt = "Enter no of students")) # Read No of stude
nts
# Declare the vector of character of length n
USN <- vector(mode="character", length= n)
Name <- vector(mode="character", length= n)
Marks <- vector(mode="numeric", length= n)

#Read the elements of the vector
print("Enter USN")
for (i in 1:n)
  USN[i] <- as.character(readline())
print("Enter Name")
for (i in 1:n)
  Name [i] <- readline()
print("Enter Marks" )
for (i in 1:n)
  Marks[i] <- as.integer(readline())
```

```
#Construct the data frame from the vectors
student <- data.frame(USN,Name,Marks)
print("The Student detials are as follows")
print(student) # Display data frame

print("Enter Age")  # Read the vector of Age
Age <- vector(mode="integer", length=n)
for (i in 1:n)
  Age [i] <- readline()
student <- cbind(student,Age) # Append the vector to the data frame

print(student)


for(i in 1:n) # Print student age > 20 , marks > 25
  if ( student[i,][3]  > 25 )
    if (student[i,][4] > 20)
              print(student[i,])
```

4. Write a program to create the csv file for storing Employee data. Containing the data (EmpID, EmpName , DOJ, EmpCode, Dept, Desig.)

- **a.** Read the suitable number of employee details from the user.
- **b.** Create a dataframe of Employee
- **c.** Store the dataframe in the csv file
- **d.** Read the data from csv and Display the contents
- **e.** Append a new row into the csv file

a.
```
n <- as.integer(readline(prompt = "Enter no of Employee"))
EmpId <- vector(mode="character", length= n)
EmpName <- vector(mode="character", length= n)
DOJ <- vector(mode="character", length= n)
EmpCode <- vector(mode="numeric",length = n)
Desig <- vector(mode="character",length = n)
Dept <- vector(mode="character",length = n)

print("Enter EmpId")
for (i in 1:n)
          EmpId[i] <- as.character(readline())
print("Enter EmployeeName")
         for (i in 1:n)
                  EmpName [i] <- readline()
print("Enter DOJ" )
for (i in 1:n)
          DOJ[i] <- (readline())
print("Enter EmployeeCode" )
for (i in 1:n)
          EmpCode[i] <- as.integer(readline())
print("Enter Designation" )
```

```
        for (i in 1:n)
                Desig[i] <- (readline())
        print("Enter Dept" )
        for (i in 1:n)
                Dept[i] <- (readline())
```

b.

```
Emp <- data.frame(EmpId,EmpName,EmpCode,Desig,Dept,DOJ)

print("The Employee detials are as follows")
print(Emp)
```

c.

```
write.csv(Emp,"C:/Users/ARCHANA/Documents/Empfile.csv")
```

d.

```
readStudent=read.csv("C:/Users/ARCHANA/Documents/file.csv")
```

e.

```
 print("Enter a new row")
u<- readline(prompt = "EmpId")
n<- readline(prompt = "EmpName")
m<- readline(prompt = "EmpCode")
A<- readline(prompt = "Desig")
s<- readline(prompt = "Dept")
t<- readline(prompt = "DOJ")

x<- data.frame(u,n,m,A,s,t)

write.table(x,"C:/Users/ARCHANA/Documents/Empfile.csv",col.names = FALSE,  append = T,row.names
= T, quote= FALSE, sep = ",")
```

5. Dataset example

a. List the data set available in your system using suitable command

b. Select "mtcars" data set, find and display the number of rows and columns in that data set

c. Find are there more `automatic (0)` *or* `manual (1)` transmission-type cars in the dataset? *Hint: 9th column indicate the transmission type*

d. Get a scatter plot of 'hp' vs 'weight'.

e. Change 'am', 'cyl' and 'vs' to *integer* and store the new dataset as 'newmtc'.

f. Extract the cases where cylinder is less than 5

a.

```
data()
head(mtcars)
```

b.

```
# Number of rows (observations)
rownum <- nrow(mtcars)
# Number of columns (variables)
colnum <- ncol(mtcars)
```

```
c. x<- data.frame(mtcars)
   automatic <-0
   manual <-0
for (i in 1:rownum)
  ifelse( x[i,9] == 1, automatic <- automatic + 1, manual <- manual +1)
ifelse (automatic > manual,
  print("There are more automatic transmission type"),
  print("There are more manual transmission type") )

d. //The scatter plot
HorsePower <- x[,4]
Weight <- x[,6]
scatter.smooth(HorsePower,Weight, span=2/3, degree = 1, family =c("symmetric","gaussian"))

// Plot histogram of Miles/gallon
Mpg <- x[,1]
 hist(Mpg, breaks = 12, col ="lightblue", border = "pink")


e.   // Solution for e
  x[,2]<- as.integer(x[,2])
 x[,8]<- as.integer(x[,8])
 x[,9]<- as.integer(x[,9])
 x[,2] <= 5

    f.  mtcars[mtcars$cyl <=5 ]
```

6. Consider "Airquality" dataset

- **a.** Display the dimension of the dataset
- **b.** Display the class of each fields in the data set
- **c.** Test the missing values
- **d.** Recode the missing values, as mean of the column values
- **e.** Exclude the missing values

```
a. df <- airquality
   dim(df)
b. sapply(df,class)
c. #Printing the missing values
   print("The Missing values are as follows")
   xcolNames <- colnames(df)
   x<- colSums(is.na(df))
   print(x)
d. which(is.na(df))
   sum(is.na(df))
   df1<- as.data.frame(df)
e. #Recoding the missing values
```

```
for(i in 1:4)
        df1[,i]<- ifelse ( is.na(df[,i]), mean(df[,i], na.rm = TRUE), df[,i])
        # Excluding the missing values

df2<-na.omit(df)
```