20BCE1214\_Aryaan\_Chauhan\_L39-40

FDA

Lab1  
Code:

#data type vector

die<-c(1,2,3,4,5,6)

die

is.vector(die)

length(die)

#data type int

int<-c(-1L,2L,3L,4L) # by default numbers are double L signifies intger

int

is.integer(int)

typeof(int)

#data type string

text<-c("Hello","world")

text

typeof(text)

typeof("Hello")

#data type logical

logic<-c(TRUE,FALSE<TRUE)

logic

typeof(logic)

typeof(F)

#logical comparisons

a=4

b=2

c=3

d=1

a>b && a>c #comparing 3 variables

a>b & a>c

a>b && a>d

#data types and addition complex numbers

comp<-c(1+1i,2+4i,3+10i) #necessary to write number before i even if 1 or 0

comp

typeof(comp)

c1<-c(4+1i)

c2<-c(5+3i)

c3<-c1+c2 #addition

c3

c4=c2\*c1 #multiplication

c4

#create an atomic vector that stores just the face names of cards in royal flush,for example. the ace of spades,king of spades, queen of spades and ten of spades

hand<-c("ace","king","queen","ten")

hand

typeof(hand)

#IDENTIFY the types of R objects in the given code

rm(list=ls()) #rm is used to remove previous objects and perform the actions in () #list stores variable values even if it is of different data types

#assign a1 with a value

a1<-10L #int

a1

s1<-"Hello" #String

s1

cmp1<-21+10i #complex

cmp1

lg1<-TRUE #logical

class(a1)

typeof(a1)

class(s1)

typeof(s1)

class(cmp1)

typeof(cmp1)

class(lg1)

typeof(lg1)

#Identify the difference between list and vector and give example of each

#list is also an array like vector, but unlike vector it can store different data types in one list and can also be used as variable

comp2<-c(1+1i,2+4i,3+10i) #vector

listdemo<-list(1L,2+3i)#list

rm(list=ls(1L,2+3i))#also creating list just another method

#create 4 lists of reg\_no, name, dept and university\_affiliated

list1<-list("U001","A007","B003","I075")

list2<-list("Tom","Dick","Harry","Sita")

list3<-list("CSE","ECE","BRS","ME")

list4<-list("MIT","AIT","Oxford","IIT")

list1[4]

list2[4]

list3[4]

list4[4]

#Ram Rs.A bought B chocolates, find cost of each chocolate

firstname="Ram"

lastname="Naresh"

A=100.5

B=20L

A

D=(100.5)\*100

G=D%%100

H=D-G

C=H/100 #C storing the rupee part

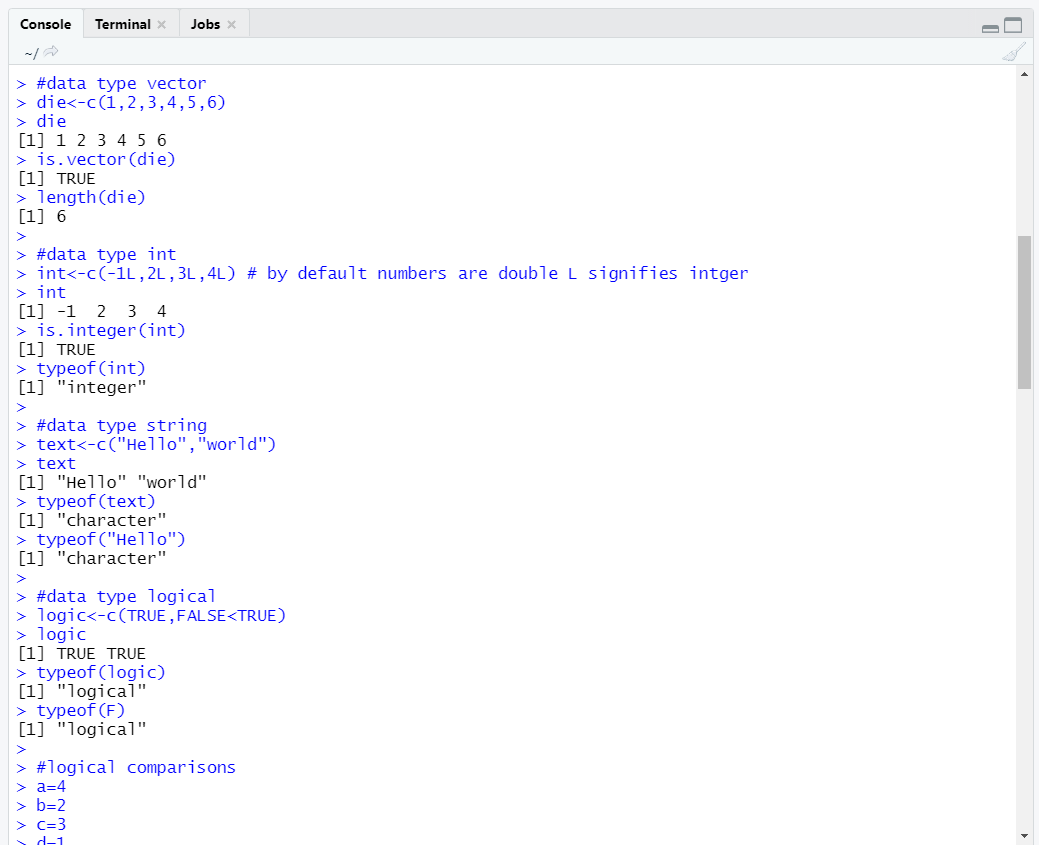
C

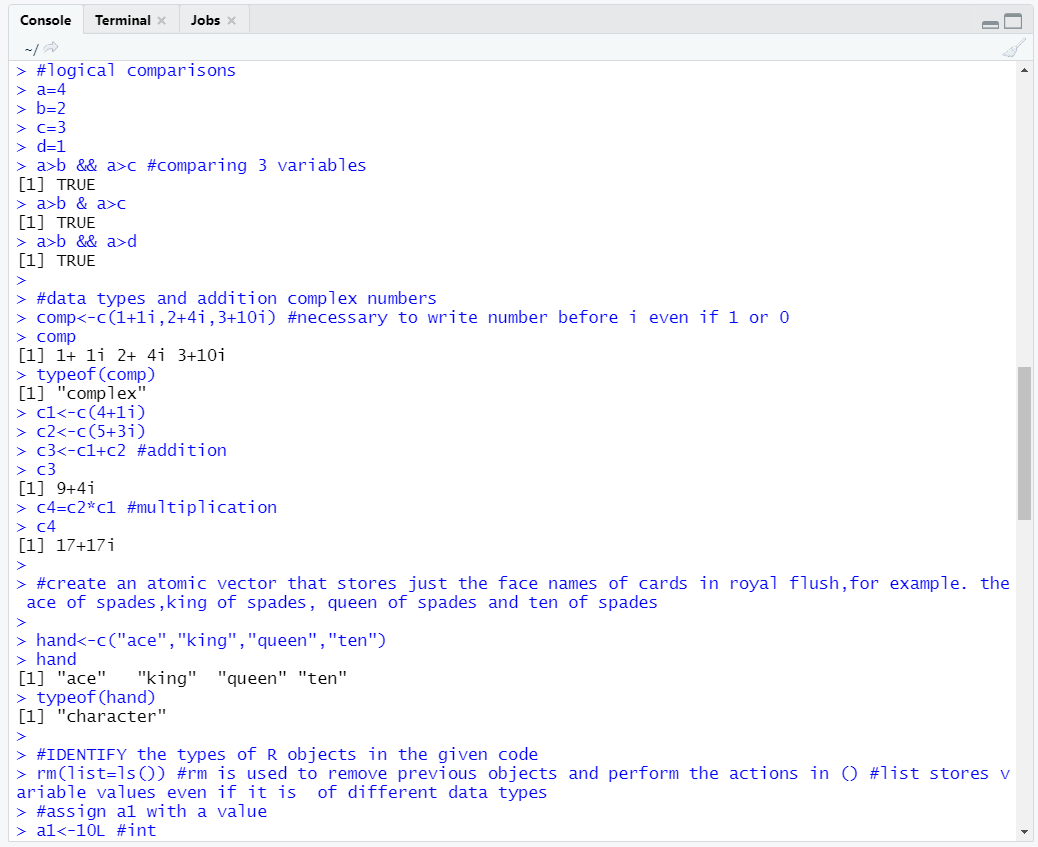
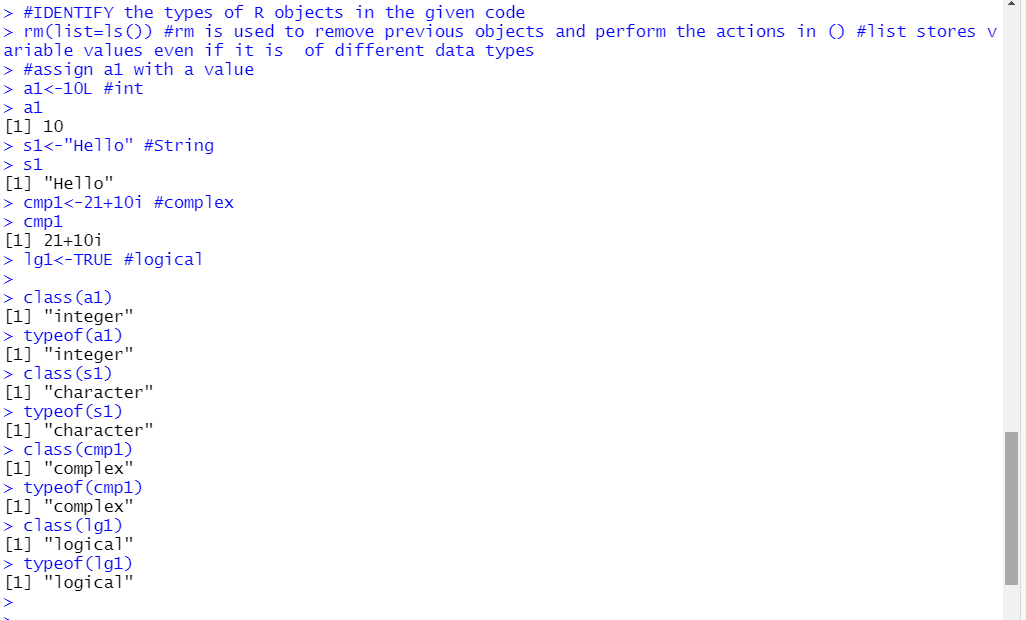
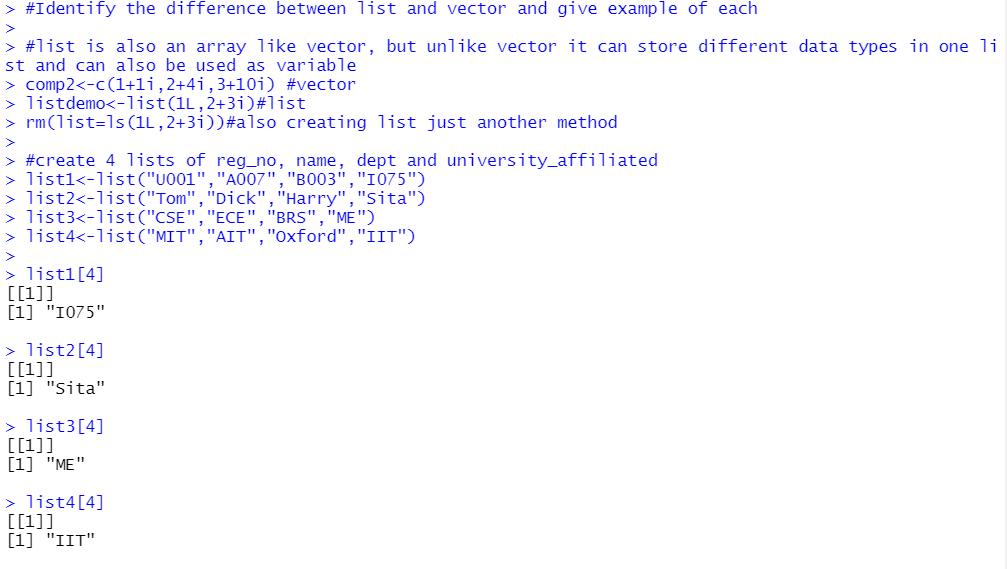
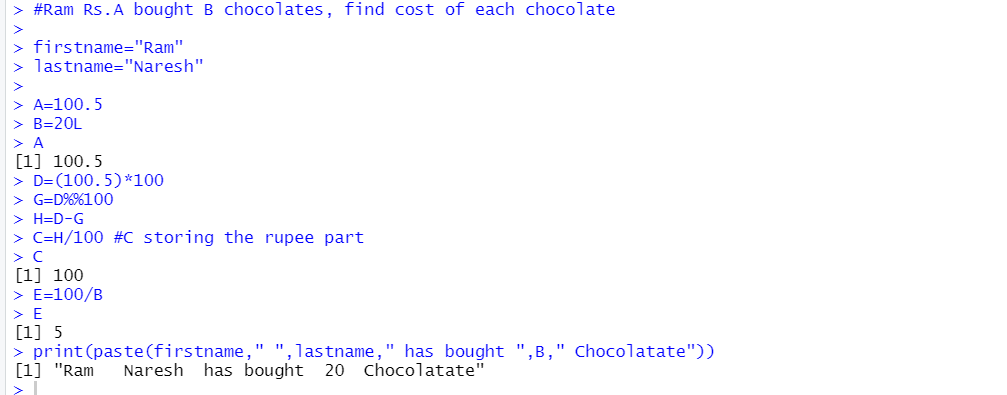
E=100/B

E

print(paste(firstname," ",lastname," has bought ",B," Chocolatate"))

Output:



lab-2

Code:

#array manipulation

die<-c(1,2,3,4,5,6)#creating object of die

dim(die)<-c(2,3)

die

dim(die)<-c(3,2)

die

#matrix

die<-c(0,2,3,4,5,6,7,8)

m<-matrix(die,nrow=4)

m

die<-c(0,2,3,4,5,6,7,8,9,10)

m<-matrix(die,nrow=5)

m

die<-c(0,2,3,4,5,6,7,8)

m<-matrix(die,nrow=3,byrow=TRUE)

m

die<-c(0,2,3,4,5,6,7,8)

m<-matrix(die,nrow=2)

m

die<-c(0,2,3,4,5,6,7,8)

m<-matrix(die,nrow=2,bycol=TRUE)#this will print undesired (3\*3)

m

#array

ar<-array(c(11:14,21:24,31:34),dim=c(2,2,3))

ar

ar<-array(c(11:19,21:29,31:39),dim=c(3,3,3))

ar

#display only the diagonal elements of a matrix

die<-c(0,1,1,1,0,1,1,1,0)

m<-matrix(die,nrow=3)

m

class(m)

#class unclass

gender<-c("female","male")

class(gender)

unclass(gender)

#Question 1

v1<-c(1,2,3)

v2<-c(11,12,13,14,15,16)

arr<-array(c(v1,v2),dim=c(3,3,2))

arr

#q2

rname<-c("r1","r2","r3")

cname<-c("c1","c2","c3")

mname<-c("mat1","mat2")

v1<-c(1,2,3,4,5,6,7,8,9)

v2<-c(1,2,3,4,5,6,7,8,9)

ar1<-array(c(v1,v2),dim=c(3,3,2),dimnames = list(rname,cname,mname))

ar1

#q3

print(ar1[2,,2])#printing 2nd row of 2nd matrix of previous case

print(ar1[,2,1])#printing 2nd column of 1st matrix of previous case

print(ar1[2,3,2])#printing element at 2nd row,3rd column and 2nd matrix

print(ar1[,,2])#printing the 2nd matrix

#q4

M1<-ar1[,,1]

M2<-ar1[,,2]

M3<-M1+M2

M3

M1

apply(M1,1,sum)#1-along row

apply(M2,2,sum)#2-along column

ar1

apply(ar1,1,sum)#1-along row

apply(ar1,2,mean)#mean along column

#q5

h1<-c("ace","king","queen","jack","ten","spades","spades","spades","spades","spades")

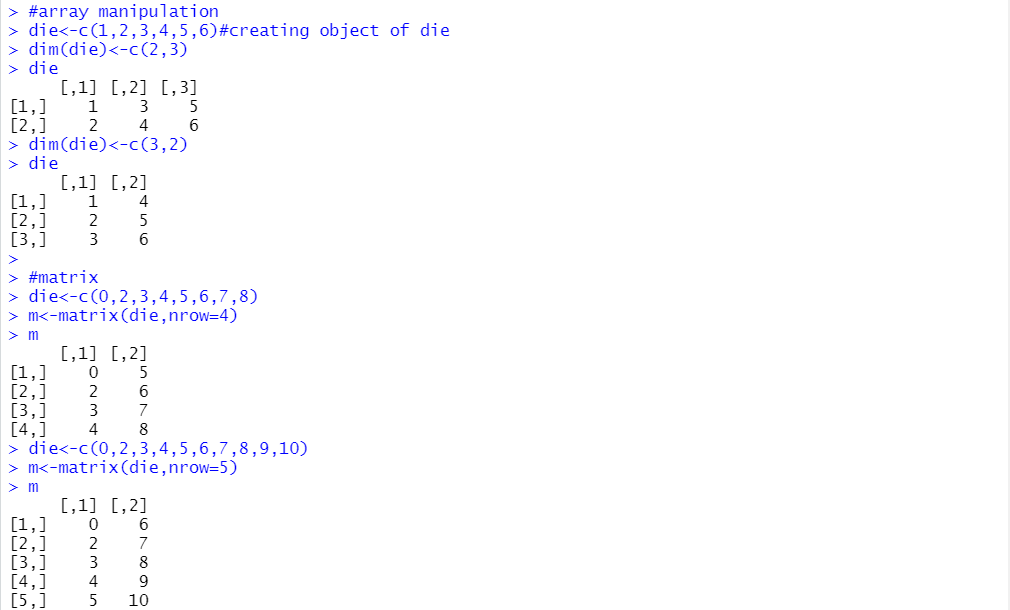
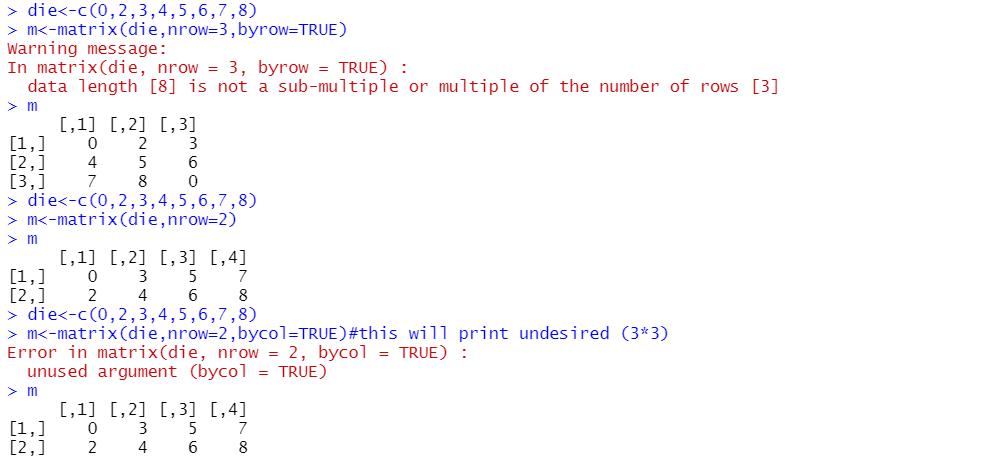
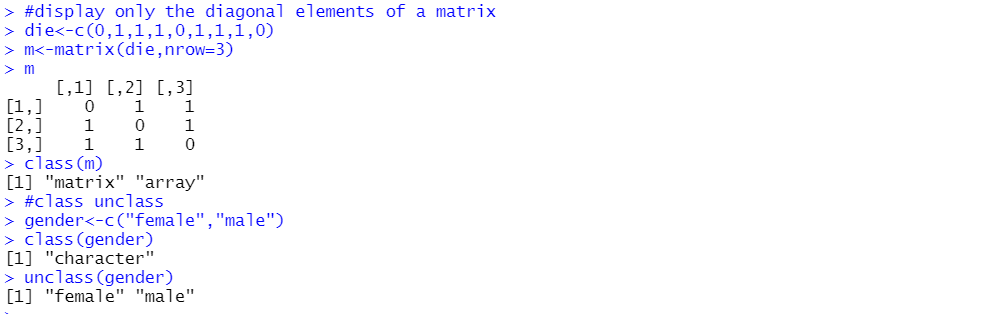
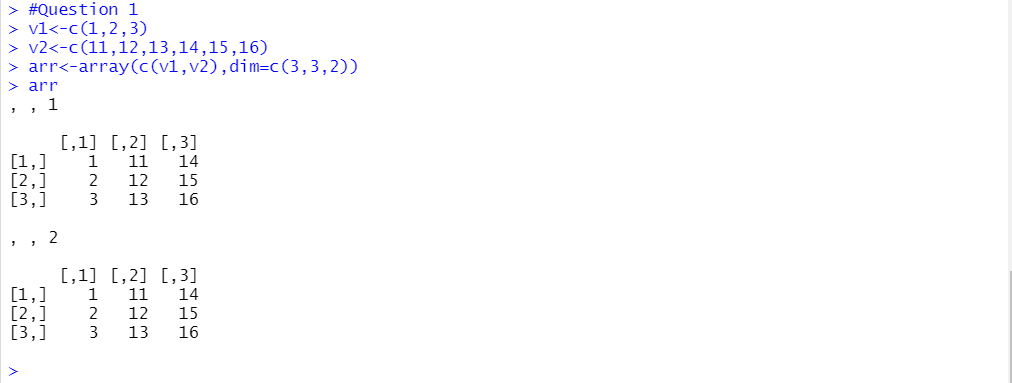
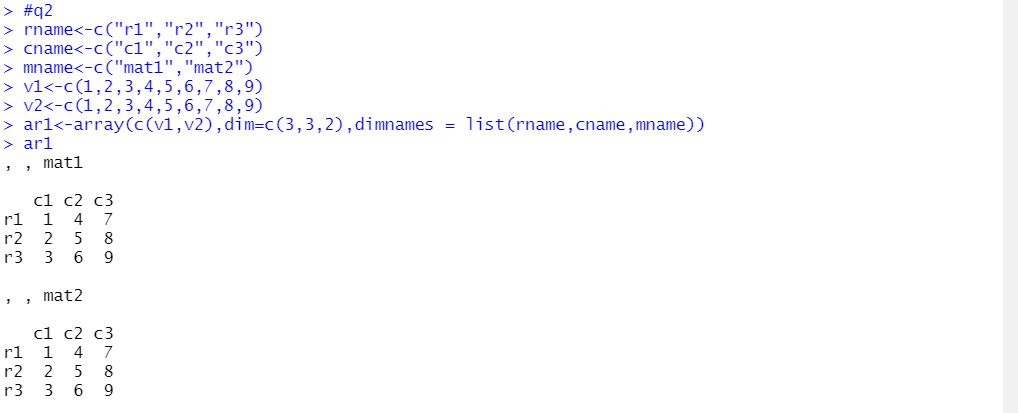
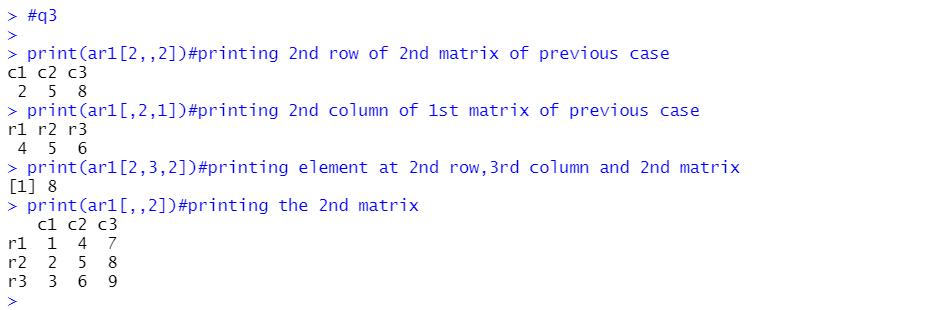
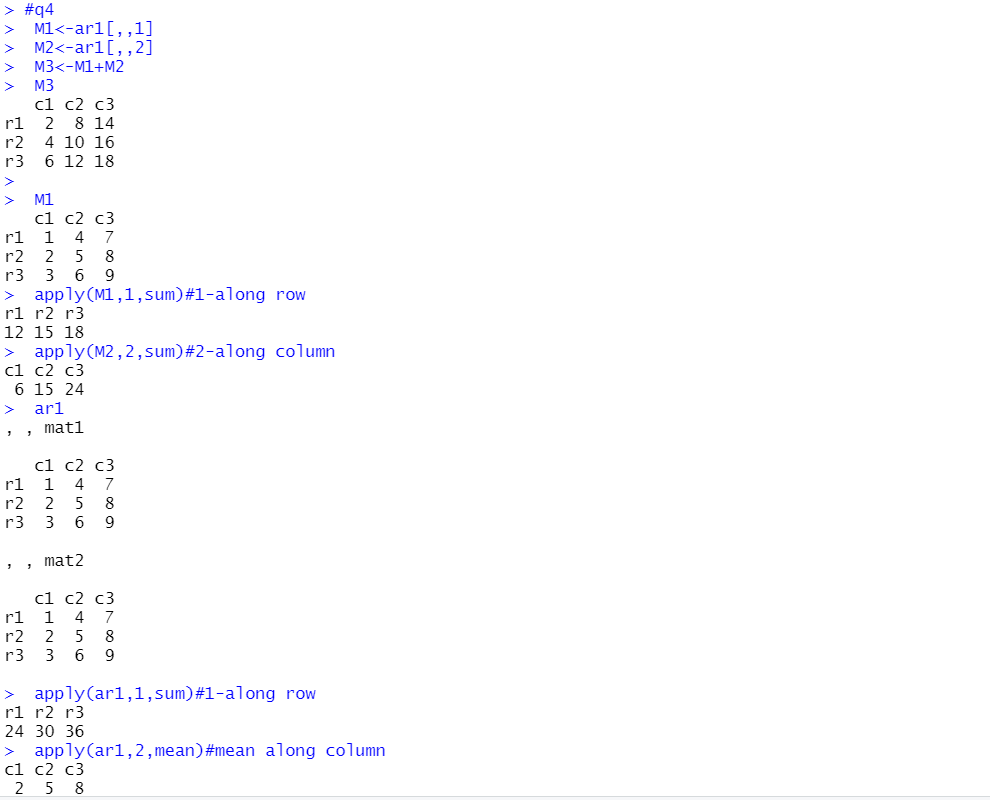
M2<-matrix(h1,ncol=2)

M2

hand2<-c("ace","spades","king","spades","queen","spades","jack","spades","ten","spades")

M1<-matrix(hand2,ncol=2)

M1  
  
  
Outputs:

  
  
  
  
  
  
Below is output for the questions  
  
  
  
  


Code:

#dataframe

df<-data.frame(ace=c("ace","two","six"),suit=c("clubs","clubs","clubs"),value=c(1,2,3))

df

typeof(df)

class(df)

str(df)#used to assign objects of size

summary(df)

head(df)

tail(df)#top6 rows

tail(df)#bottom 5 rows

data1<-data.frame(id=1:3,weight=c(20,27,24))

names(df)

#ways to add columns and initialize with 0

df$rank<-c(3,2,1)

df

#create a data frame which has 3 columns mark1,mark2&mark3 named by 5 students,print name &highest mark in each category(mark)

mdf<-data.frame(mark1=c(99,98,97,95,94),mark2=c(97,98,99,100,95),mark3=c(99,98,100,94,95),name=c("Tom","Dick","harry","Sita","Ram"))

i=which.max(mdf$mark1)

mdf$mark1[i]

mdf$name[i]

j=which.max(mdf$mark2)

mdf$mark2[j]

mdf$name[j]

k=which.max(mdf$mark3)

mdf$mark3[k]

mdf$name[k]

mdf$total<-rowSums(mdf[1:3])

mdf

mdf$average<-rowMeans(mdf[1:3])

mdf

l=which.max(mdf$average)

mdf[l]

mdf

mdf[which.max(mdf$average),1:6]

#write.csv(mdf,"marks20bce1214lab3.csv")

#q1 define a table within a column and diferent types colums using data.frame() using

df1=data.frame(Flavours=c("Chocalate","Vanilla","Butter-Scotch","Fruit-Mix"),Unit\_Cost\_Rs=c(25,25,30,30),Units\_sold=c(94,97,70,75))

df1

#q2

data(mtcars)

car<-mtcars

mtcars #viewing from mtcars

v1<-car$mpg

v2<-car$cyl

v3<-car$disp

v4<-car$hp

nvar<-mtcars$disp+mtcars$hp

mtcars1<-rbind(v1,v2,v3,v4)#horizontal join

mtcars1

mtcars2<-cbind(v1,v2,v3,v4)#vertical join

mtcars2

ob\_subset<-mtcars[4:10,]

ob\_subset

#q3

subset3<-subset(mtcars,mpg>18)[,c(1,2)]

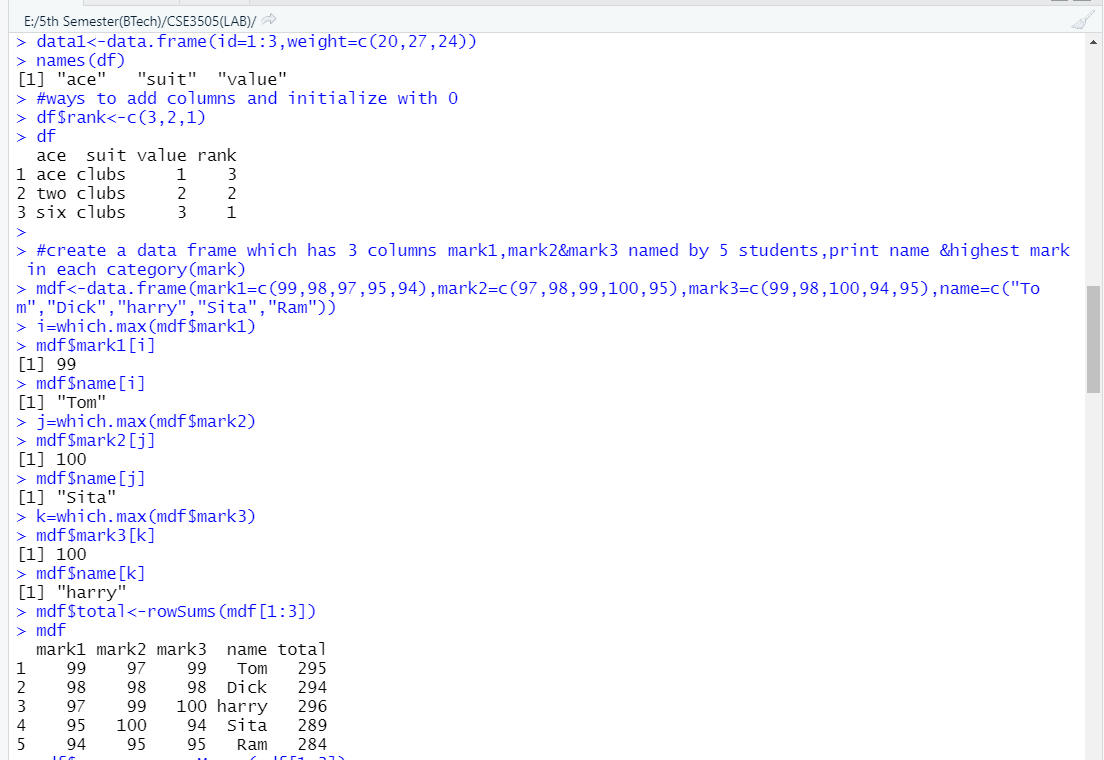
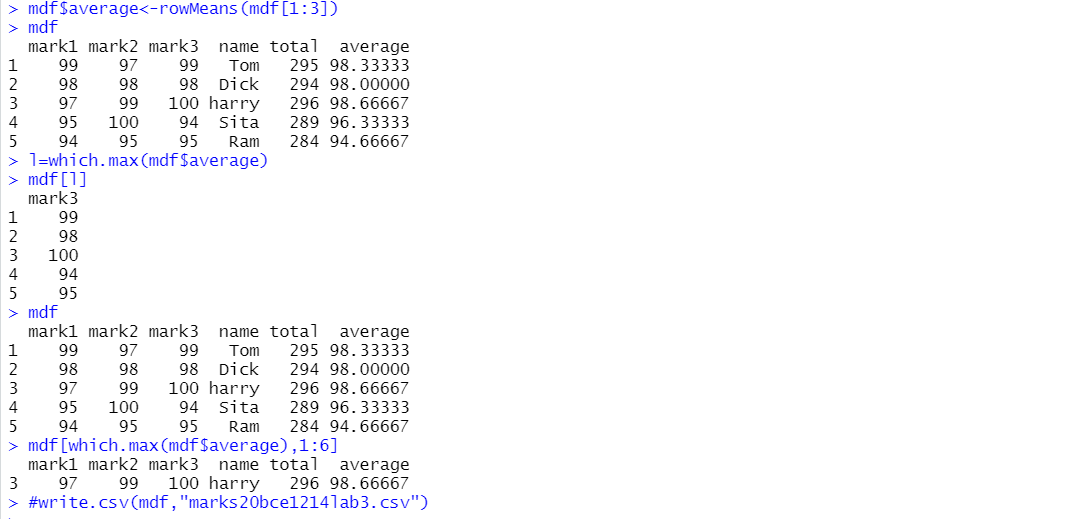
subset3

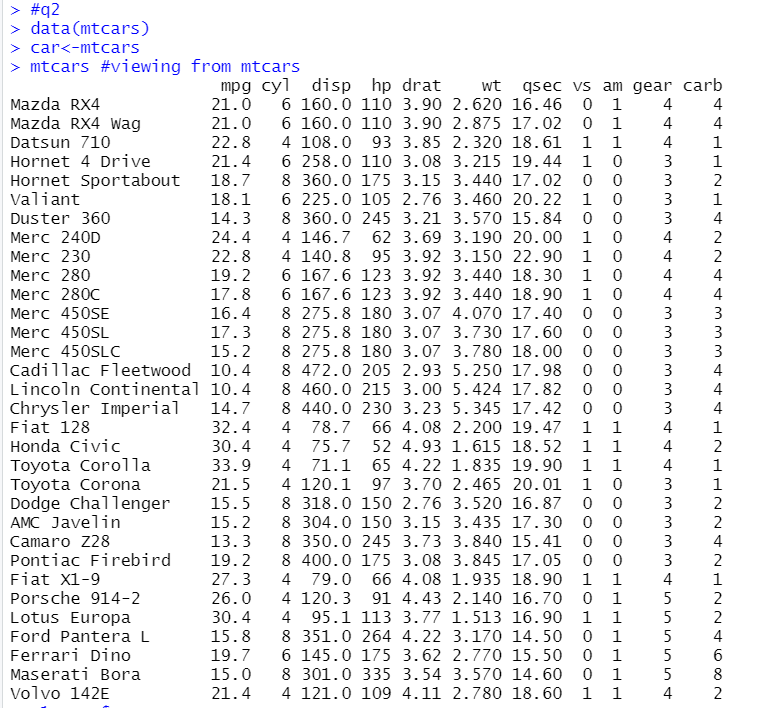
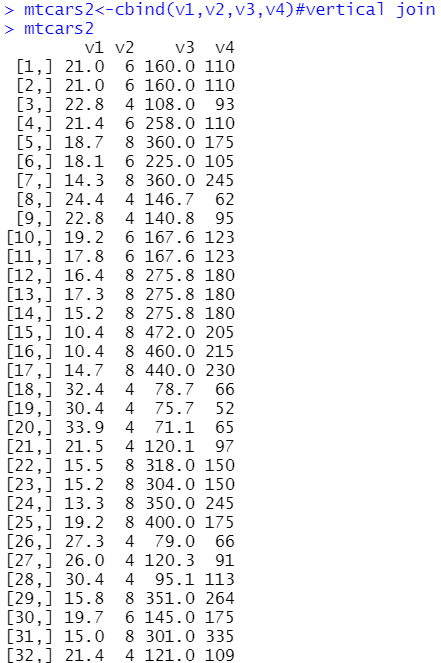
subset4<-subset(mtcars, mpg>18 & cyl>5)[,c(1,2)]

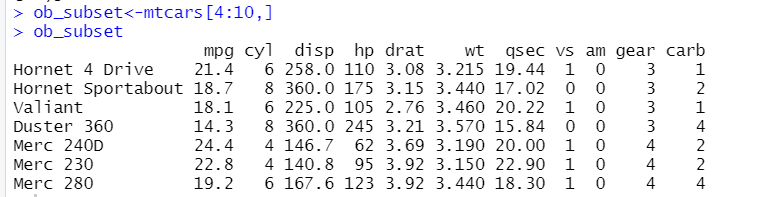
head(subset4)

subset5<-subset(mtcars, mpg>18 & cyl>5, select=c(-mpg, -cyl))

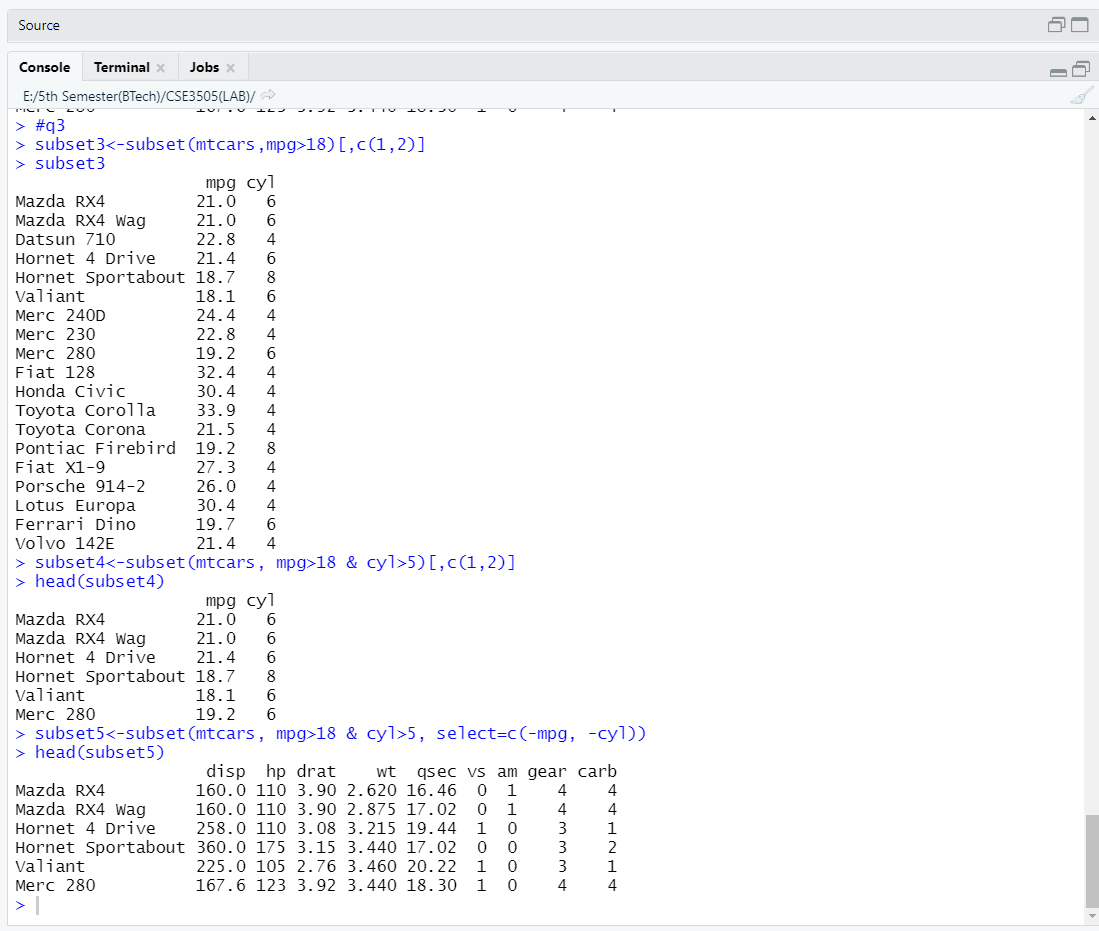
head(subset5)  
  
output:

  
  
  
  
q1  


Q2  
  
  
  




Q3



Lab-4  
Code:

install.packages("nycflights13")

install.packages("data.table")

library(nycflights13)

library(data.table)

flights<-data.table(flights)

ans<-flights[origin=="JFK" & month==6L]

ans

ans1<-flights[origin=="EWR" & dest=="IAH"]

ans1

#data.table package

install.packages("data.table")

library(data.table)

A=rnorm(6)

df<-data.frame(A=rnorm(6),B=rep(c("x","y"),each=3),C=rnorm(6))

df

dt<-data.table(A=rnorm(6),B=rep(c("x","y"),each=3),C=norm(6))

dt

#2

x<-c(10,4,5,6,8,1,7,3,2)

order(x)

list(x)

mean(x)

length(x)

rep(x)

rnorm(x)

#3

df1<-data.frame(A=c(1,2,3,4,5,6),B=c(10,9,8,10,5,6),C=c(45,6,7,97,6,7))

names(df1)[1]<-"column1"

names(df1)[2]<-"column2"

names(df1)[3]<-"column3"

df1

df1$column4<-c(60,61,62,63,64,65)

df1

df1[["column3"]]<-NULL

df1

df1[["column3"]]<-df1[["column1"]]+df1[["column4"]]

df1<-df1[-c(1,4)]

df1

#4

library(data.table)

A=rnorm(6)

dt<-data.table(A=rnorm(6),B=rep(c("x","y"),each=3),C=rnorm(6))

dt

w1<-dt[with(dt,C>0)]

w1

w2<-dt[which(B=="x")]

w2

(dt1 <- data.table(A = letters[rep(1:3, 2)], X = 1:6, key = "A"))

(dt2 <- data.table(A = letters[rep(2:4, 2)], Y = 6:1, key = "A"))

merge(dt1, dt2, allow.cartesian=TRUE)

dt3<-data.table(A=c("A","B","C"),B=c(1,3,6))

dt4<-data.table(C=c("p","q","r"),D=c(6,7,9))

setkey(dt3,B)

setkey(dt4,D)

dt3[dt4,roll=TRUE]

dt3[dt4,roll=TRUE,rollends=c(TRUE,FALSE)]

dt[,.SD]

dt[,.SD,.SDcols=c("A","C")]

dt3["A",on="A"]

DT <- data.table(x=rep(c("b","a","c"),each=3), y=c(1,3,6), v=1:9)

X = data.table(x=c("c","b"), v=8:7, foo=c(4,2))

DT[X, on="x", mult="first"]

DT[X, on="x", mult="last"]

DT[.("b", 1:2), on=.(x, y), nomatch=NULL]

#q5

library(nycflights13)

ans1 <- nycflights13::flights[origin = "JFK" & month==6L]

ans1

ans2<- flights[origin=="JFK" & month==6L]

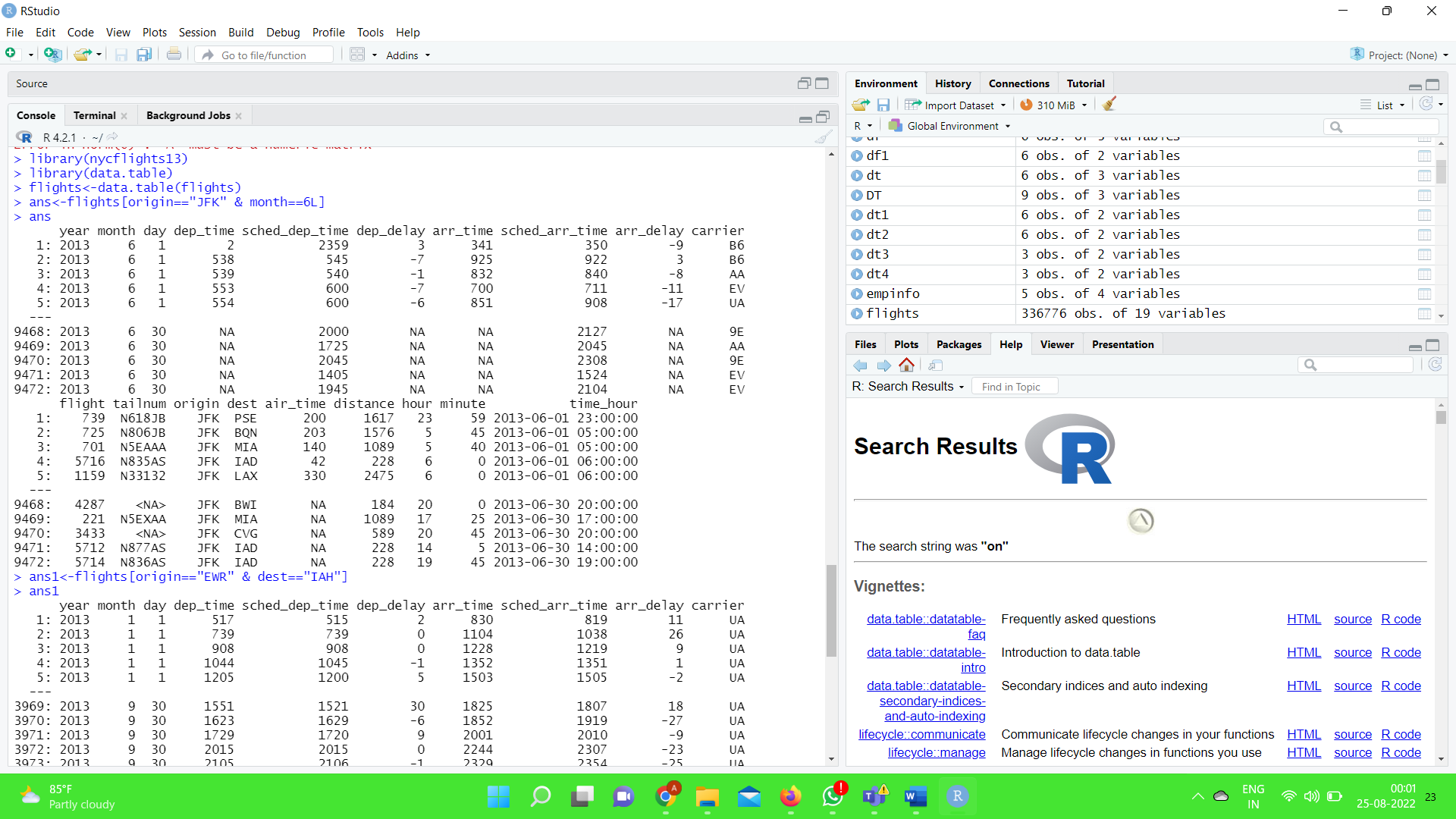
ans2

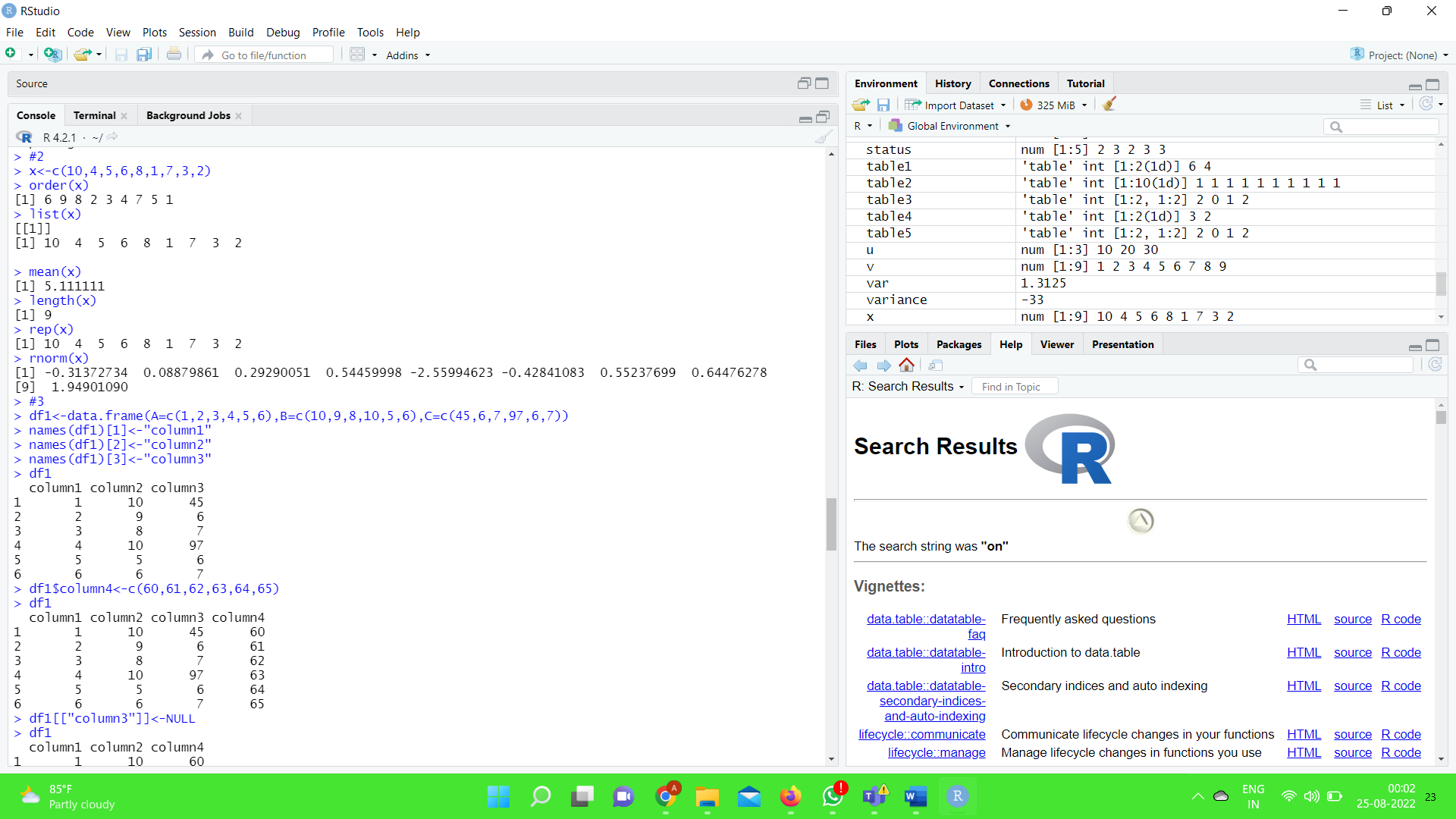
ans3 = flights[origin = "EWR", dest = "IAH"]

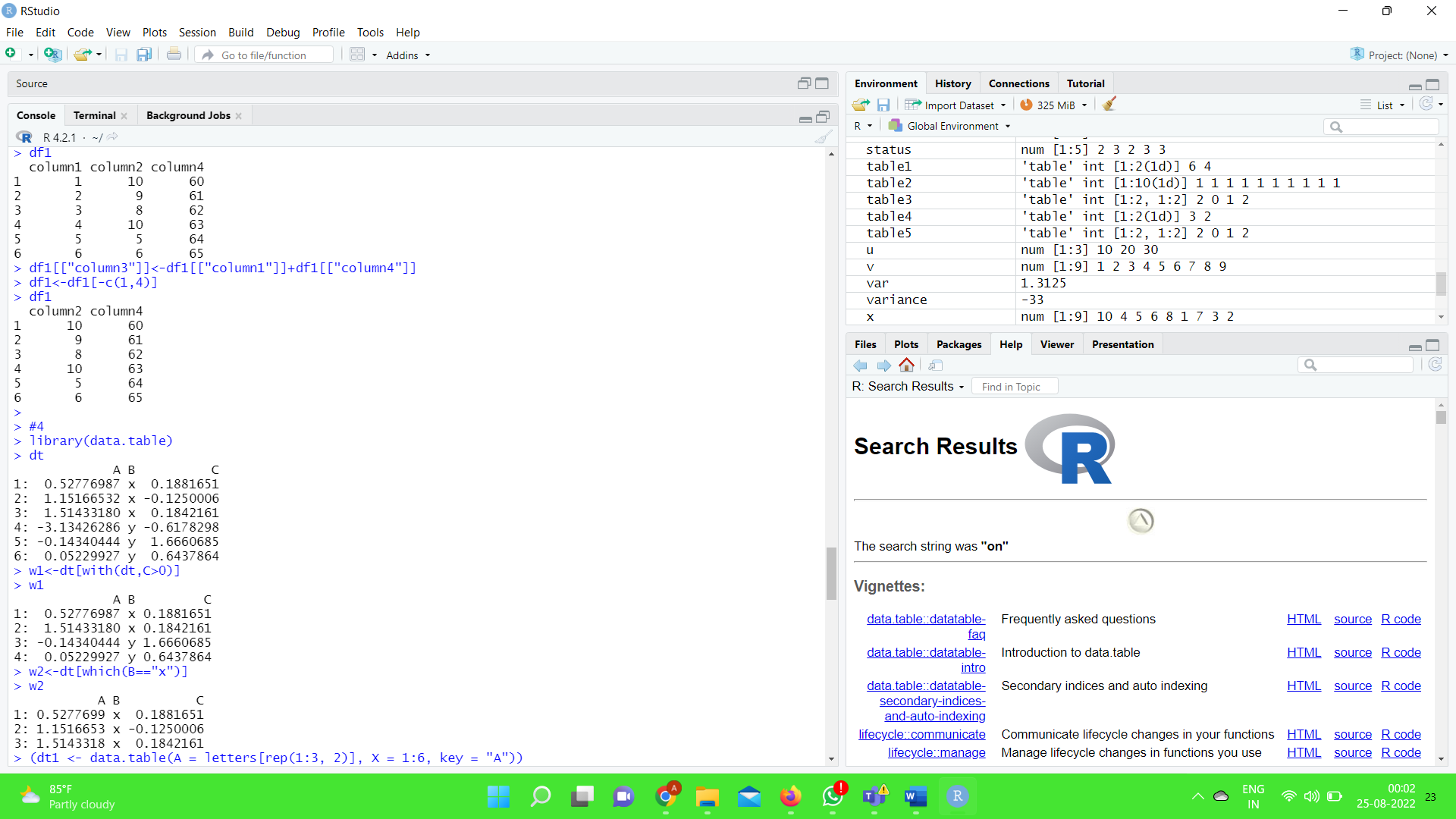
head(ans3)

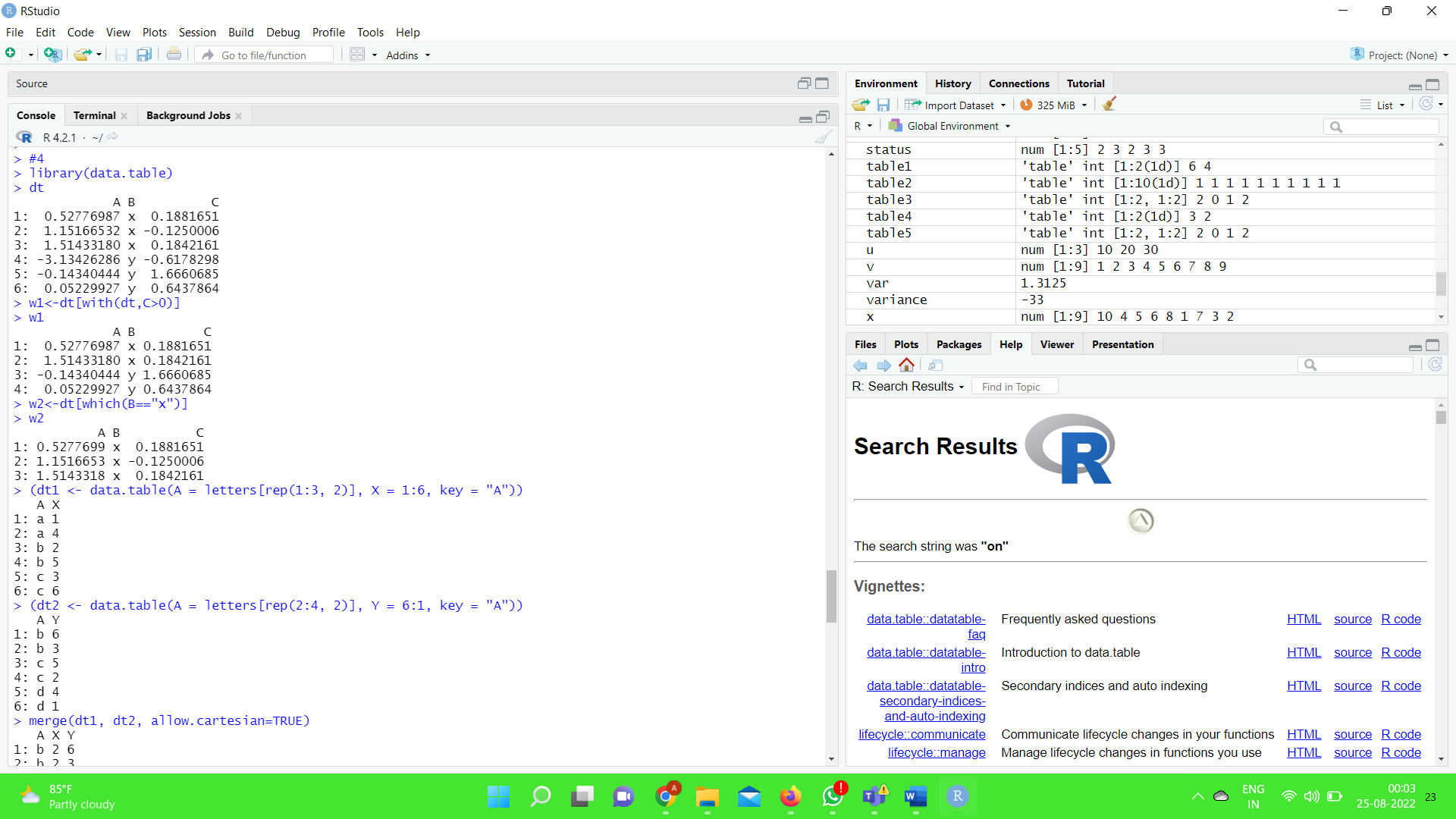
ans4 <- flights[order(origin,-dest)]

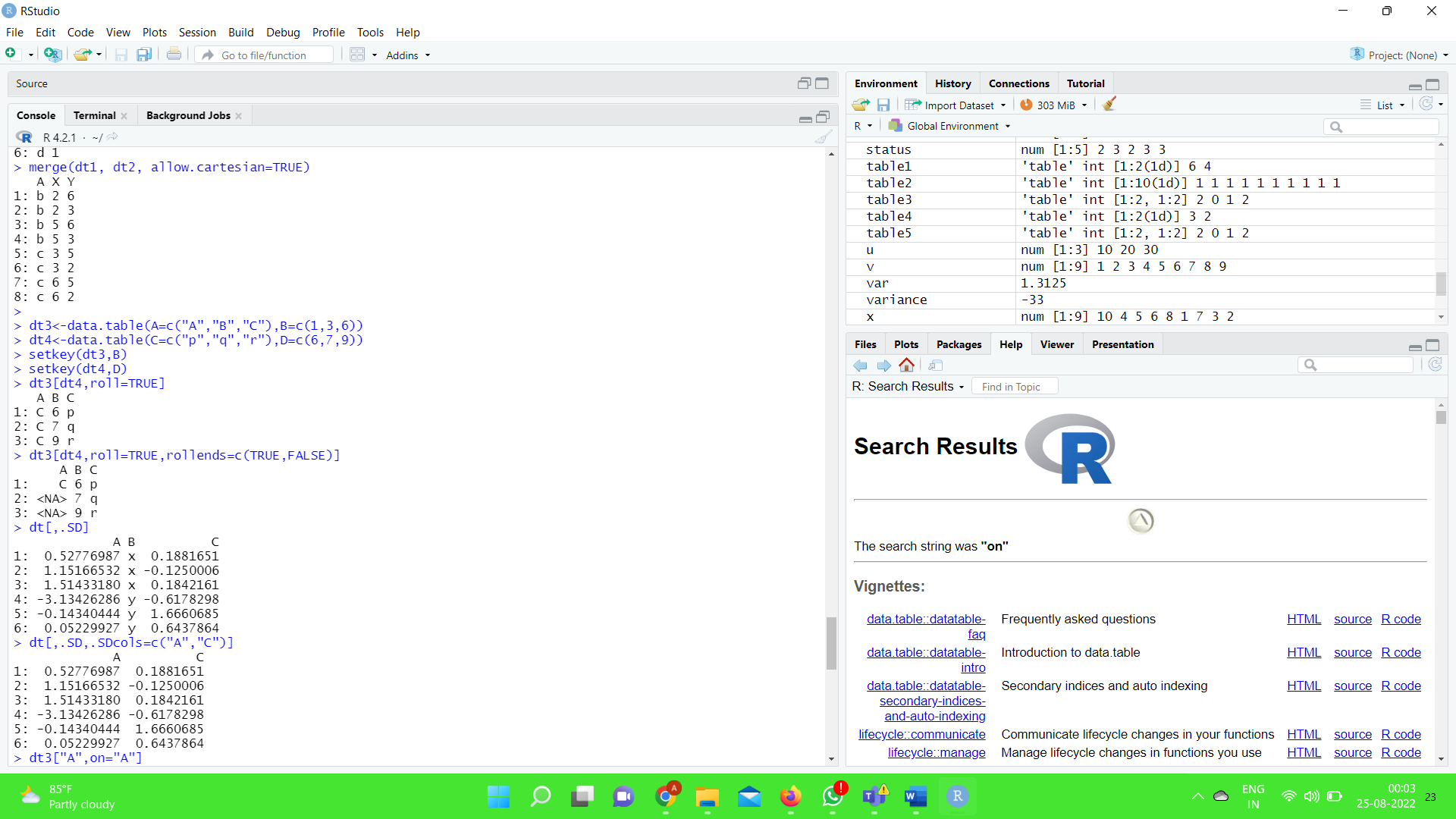
head(ans4)  
  
Outputs:

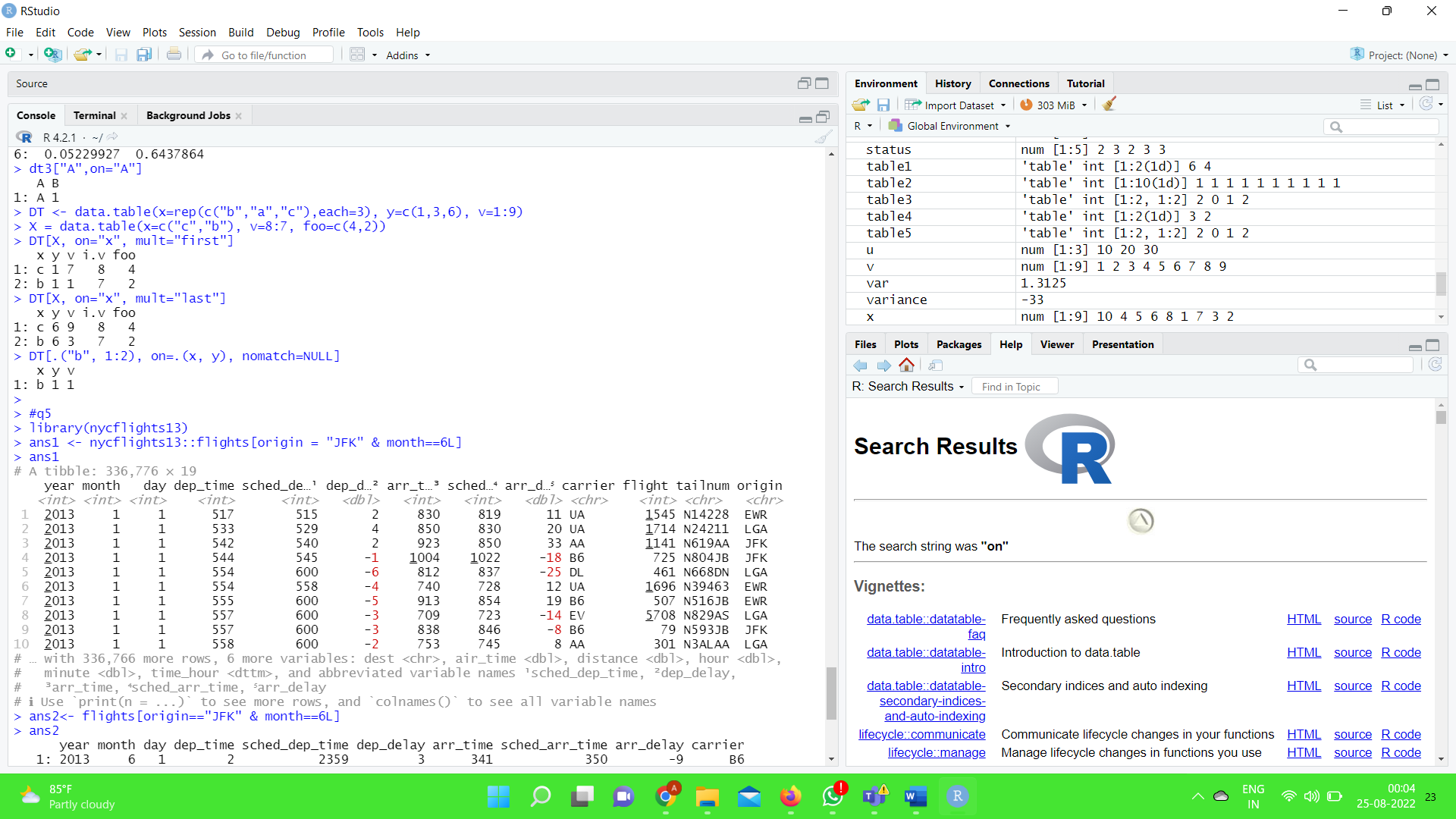


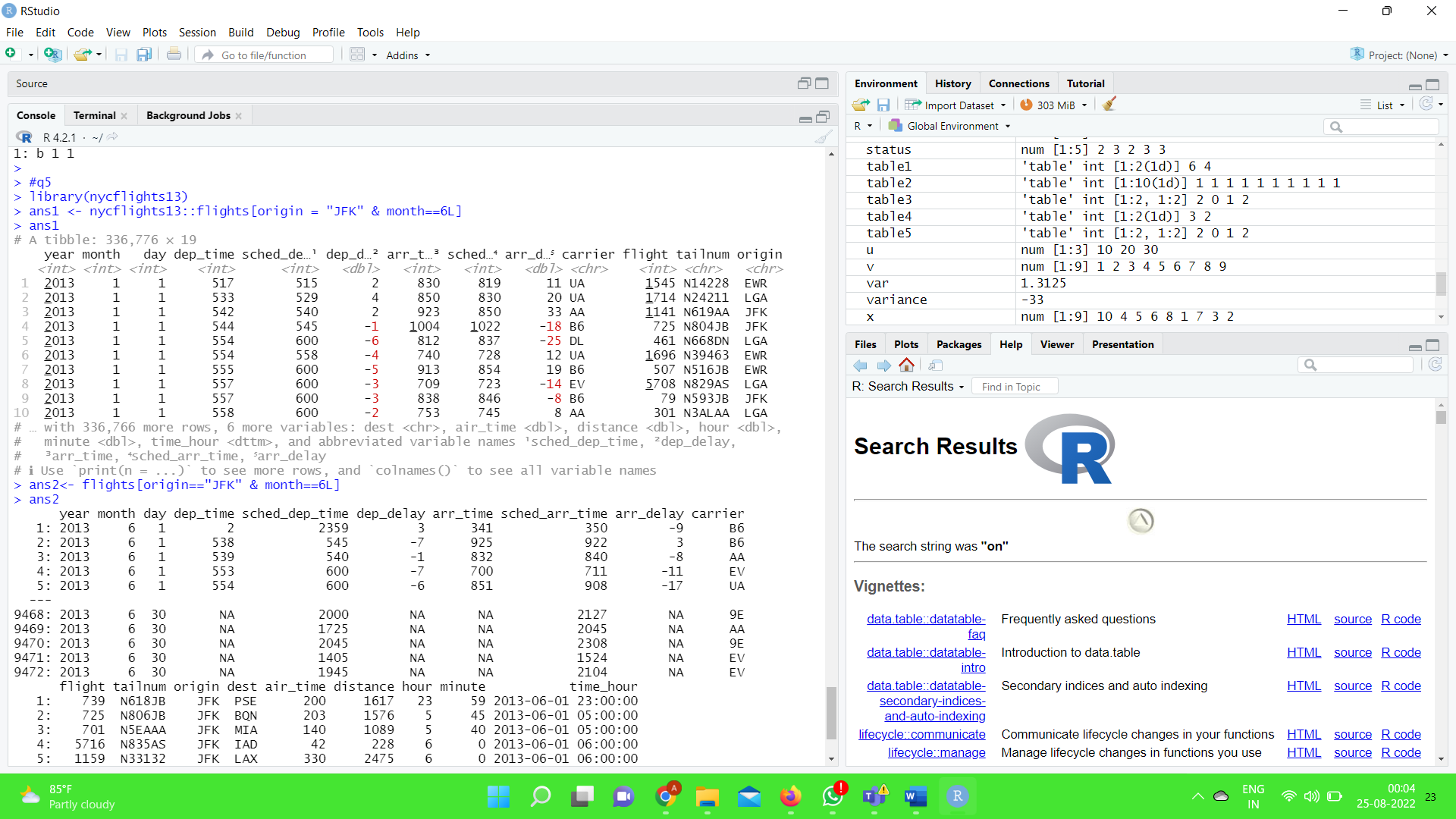


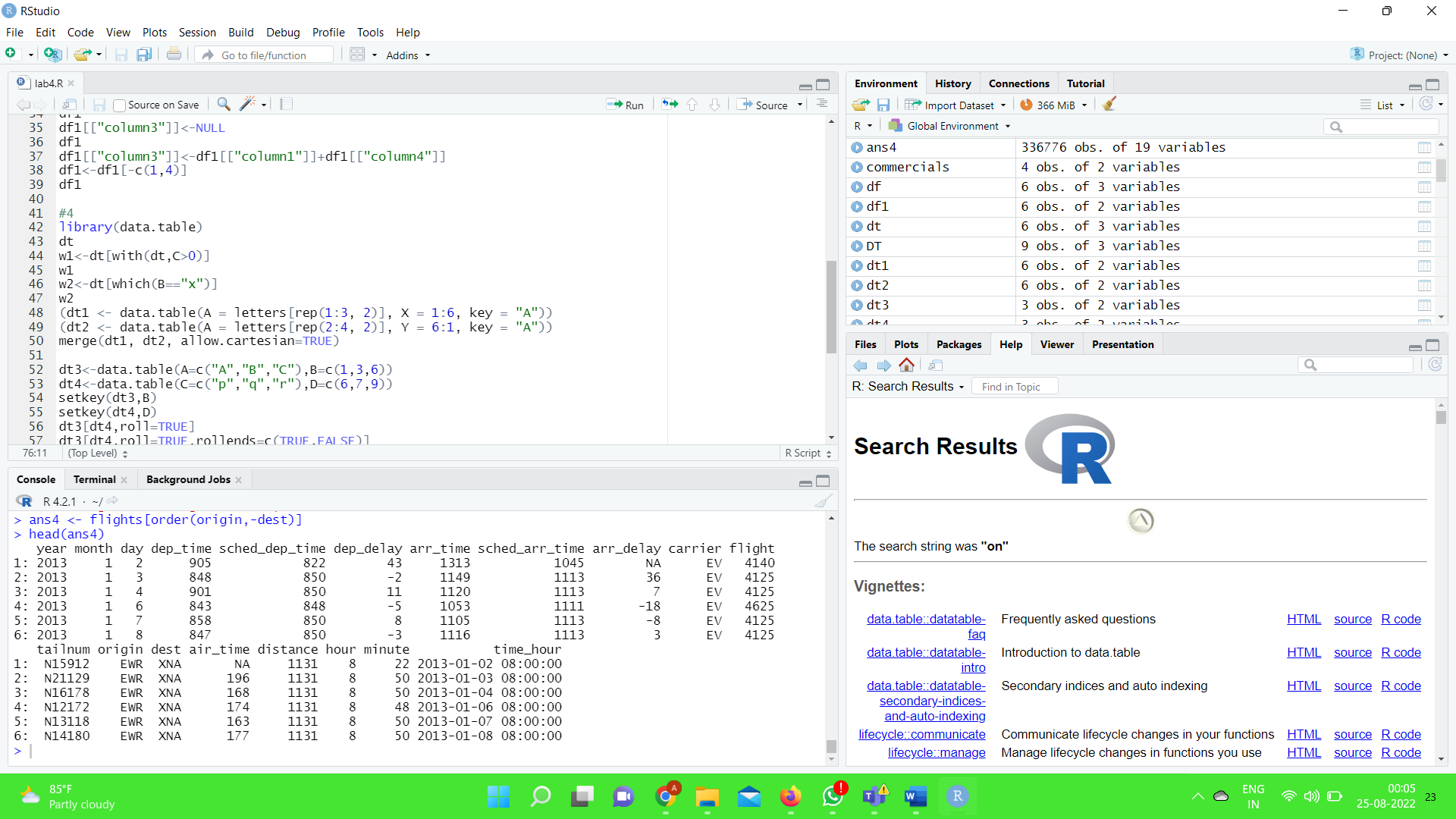












Lab-5

Code

#reading a table like file

x1<-read.delim('data1.txt')

x1

#reading .xls file

install.packages("xlsx")

library(xlsx)

x2<-read.xlsx('data2l5.xls',sheetIndex = 1)

x2

#reading a .txt file line by line

x3<-readLines("dataline.txt")

x3

#reading .xml file

install.packages("XML")

library(XML)

x4<-xmlParse("nysk.xml")

x4

#data manipulation on .xls file which above has been stored in 2

install.packages("dplyr")

library(dplyr)

filter(x2,COST>50000)

select(x2,starts\_with("COST"))

select(x2,-starts\_with("ID"))

select(x2,contains("A"))

arrange(x2,COST)

arrange(x2,desc(ID))

mutate(x2,INTEREST=COST\*0.01)

summarize(x2,med=mean(COST))

#date

d<-as.Date("2021-6-15")

d

weekdays(d)

d+5

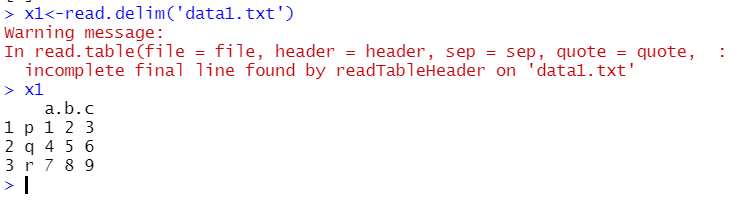
d+1:5

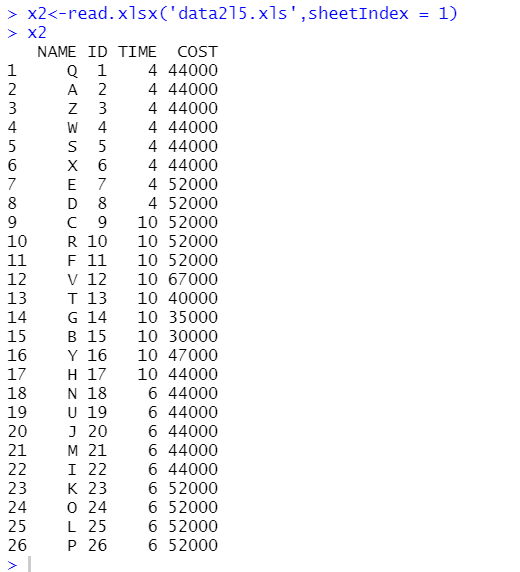
weekdays(d+1:5)

dt<-seq(d,by="2 months",length.out=6)

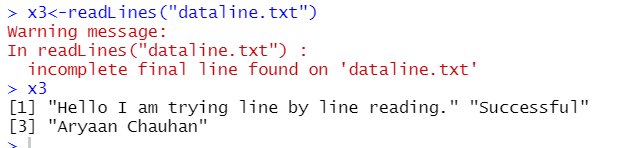
dt

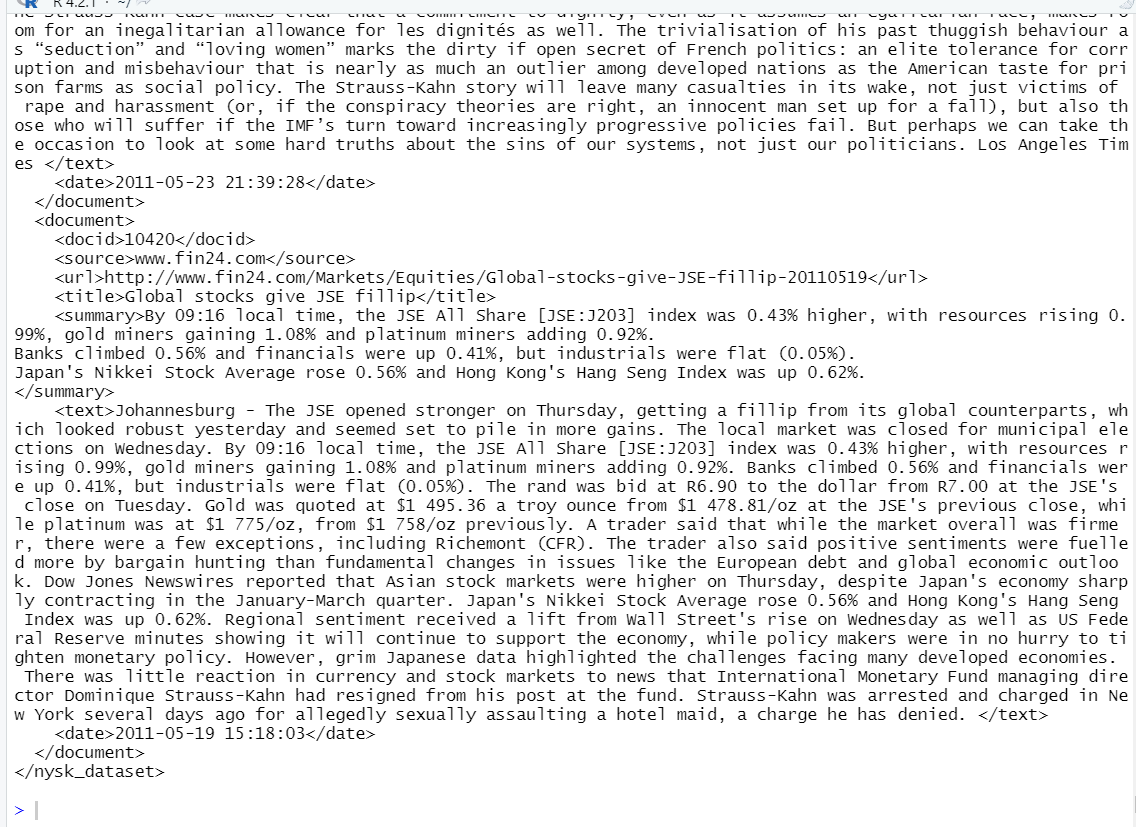
outputs:

1)read a table like file  


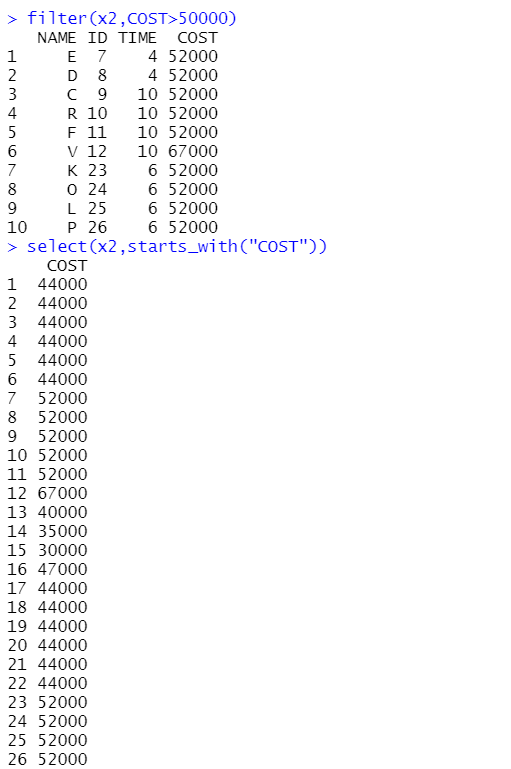
2)reading .xls file  


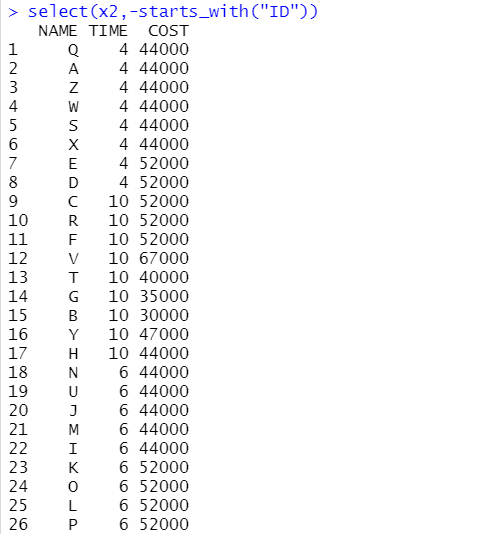
3)reading contents of .txt line by line

  
4)Reading an .xml

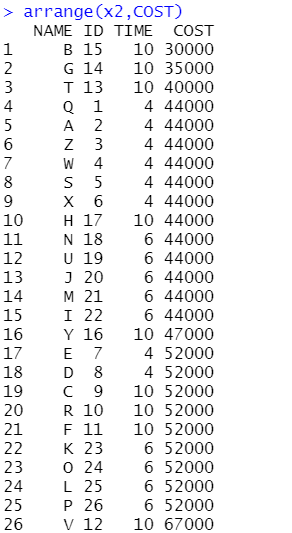


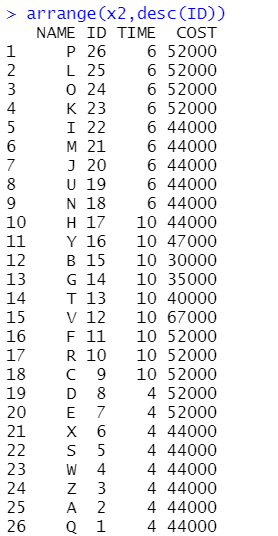
5)operating on .xls file

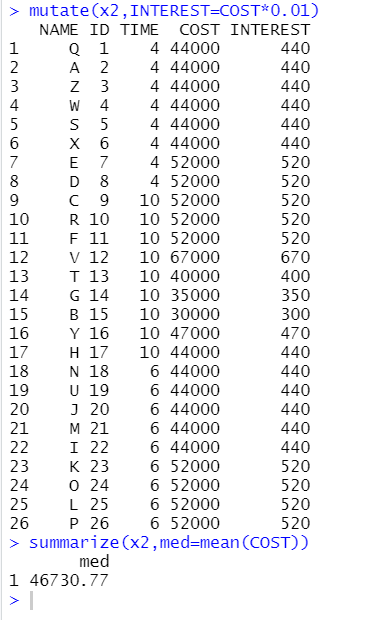


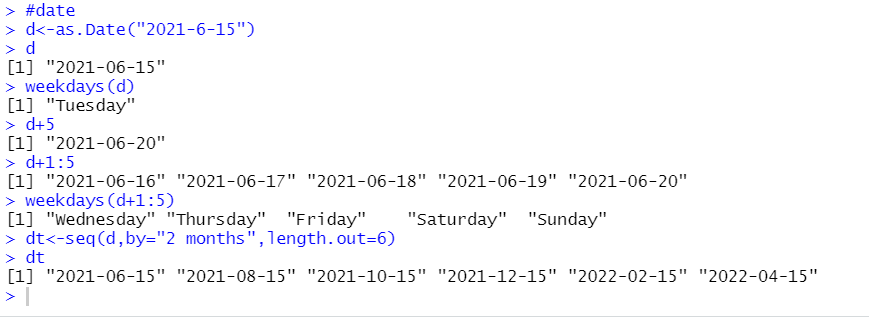








  
5)Date operation



Lab-6

Code:

#import package MASS

install.packages("MASS")

library(MASS)

survey

str(survey)

print(colnames(survey))

#list of rows with missing values

demo<-survey

missingrow<-subset(demo,is.na(demo$Sex)|is.na(demo$Wr.Hnd)|is.na(demo$NW.Hnd)|is.na(demo$W.Hnd)|is.na(demo$Fold)|is.na(demo$Pulse)|is.na(demo$Clap)|is.na(demo$Exer)|is.na(demo$Smoke)|is.na(demo$Height)|is.na(demo$M.I)|is.na(demo$Age))

missingrow

#data frame containing the survey data after removing NA values

d=data.frame(survey)

d1=na.omit(d)

d1

#getting the count of different categories of sex in data

install.packages("plyr")

library("plyr")

count(d,'Sex')#1 method

s<-d$Sex

table(s)#another method

#getting the count of left and right handed people

table(d$W.Hnd)#1 method

count(d,'W.Hnd')#2nd method

#getting relative frequency distribution of left & right handedness  
rf<-table(d$W.Hnd)/length(d$W.Hnd)

print(format(round(rf,2),nsmall = 2))

#display male left handed and female left handed

sub1<-subset(d,d$W.Hnd=="Left")

sub1

#percentage of male left handed who never smoked

sub2<-subset(d,d$Sex=="Male" & d$W.Hnd=="Left")

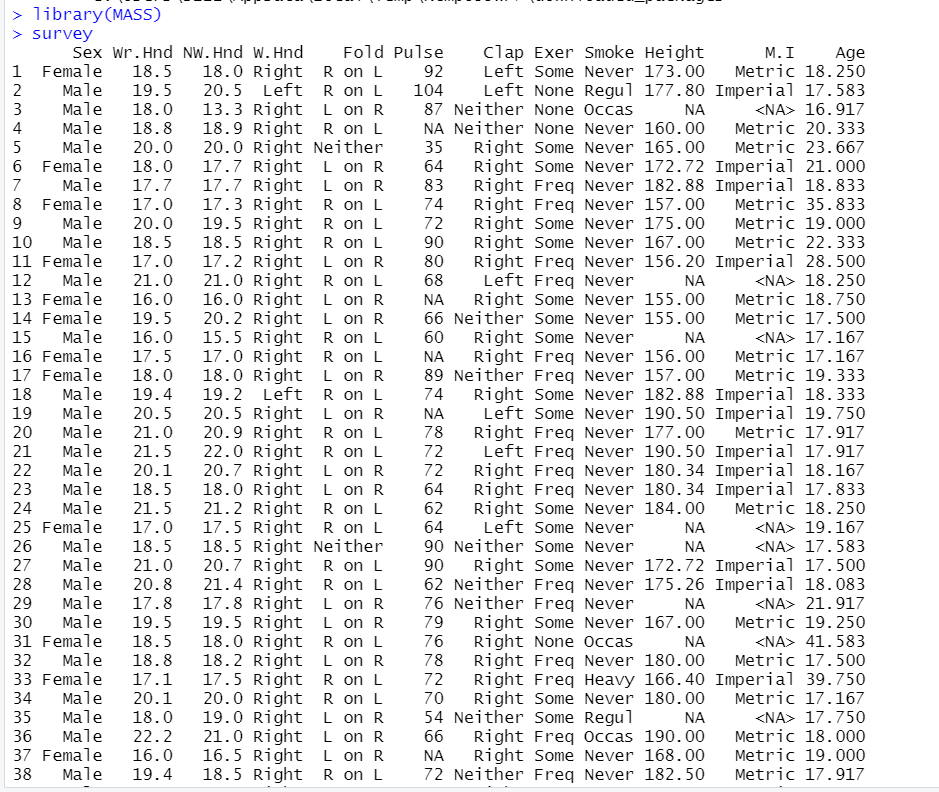
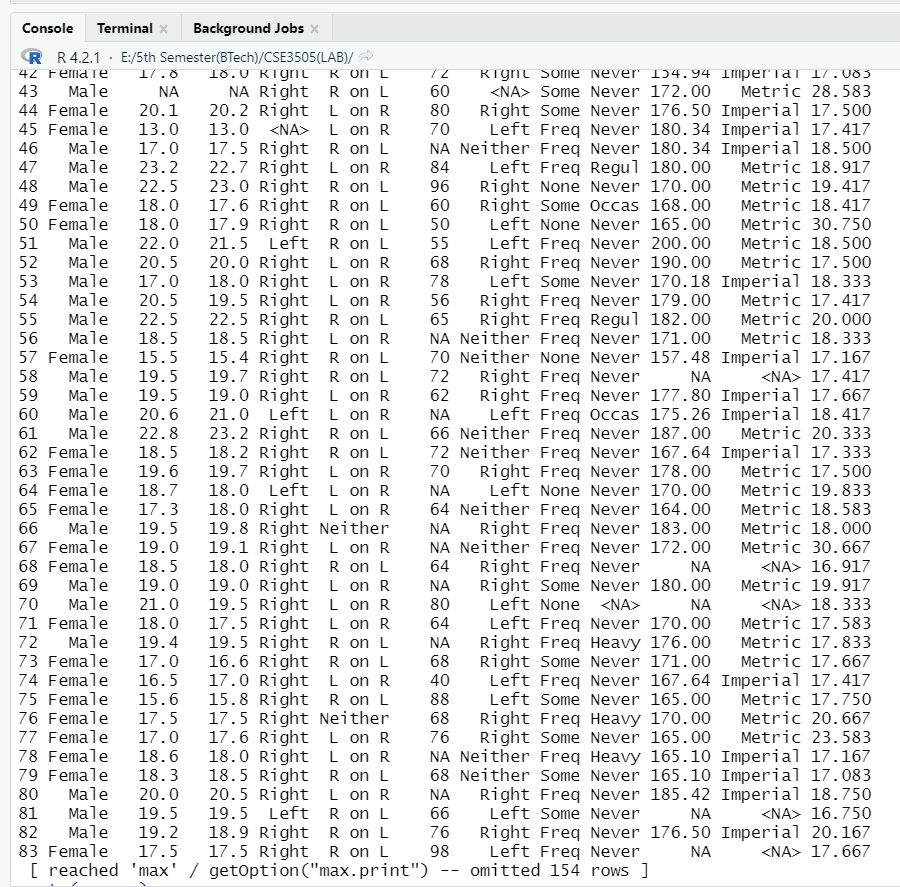
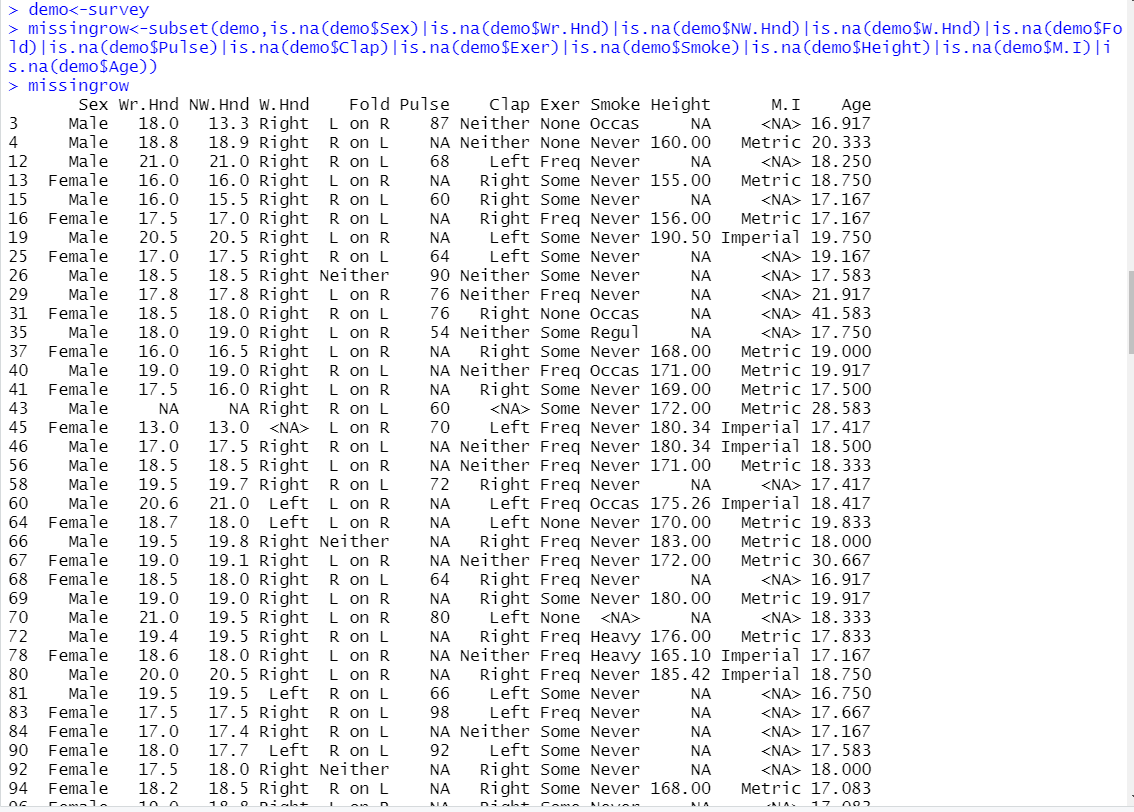
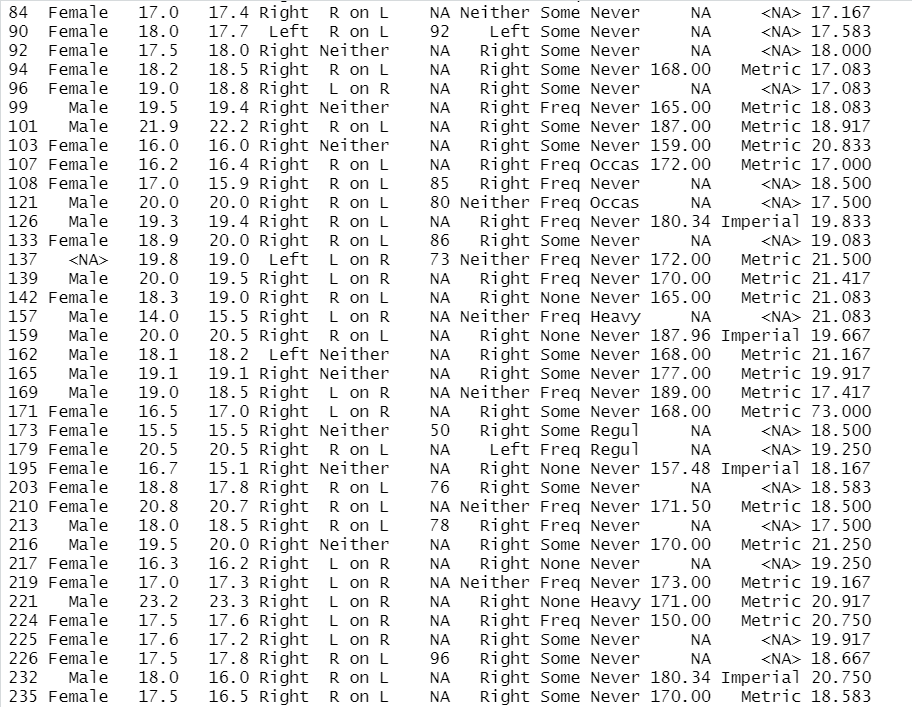
sub3<-subset(sub2,sub2$Smoke=="Never")

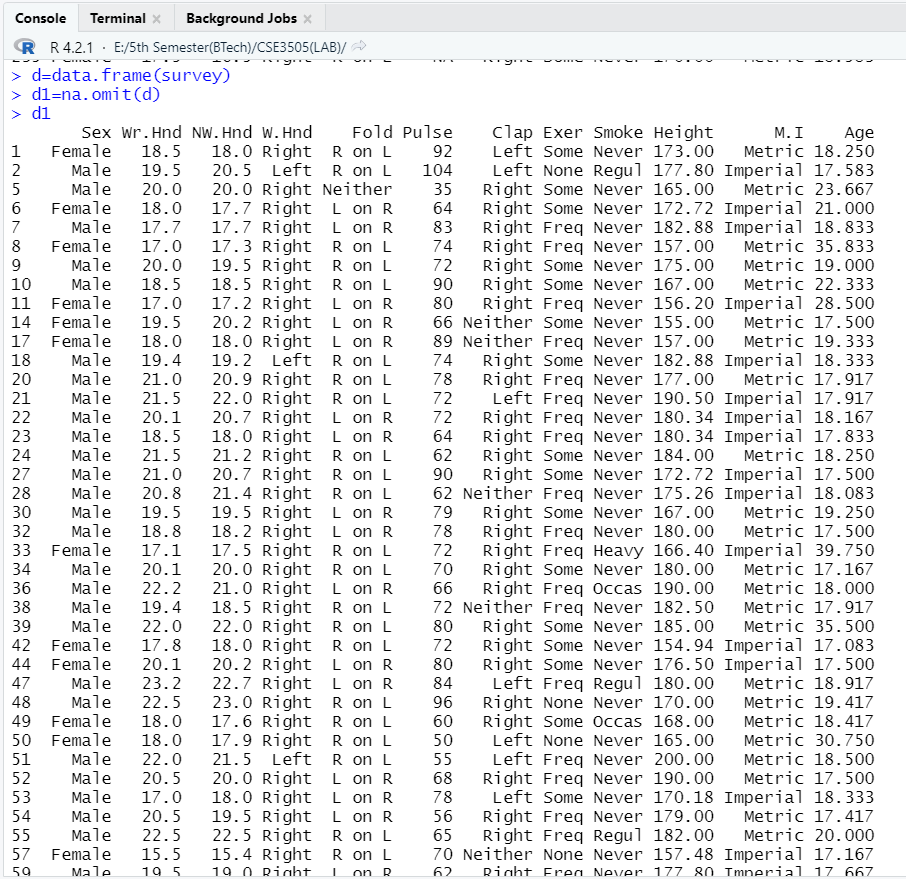
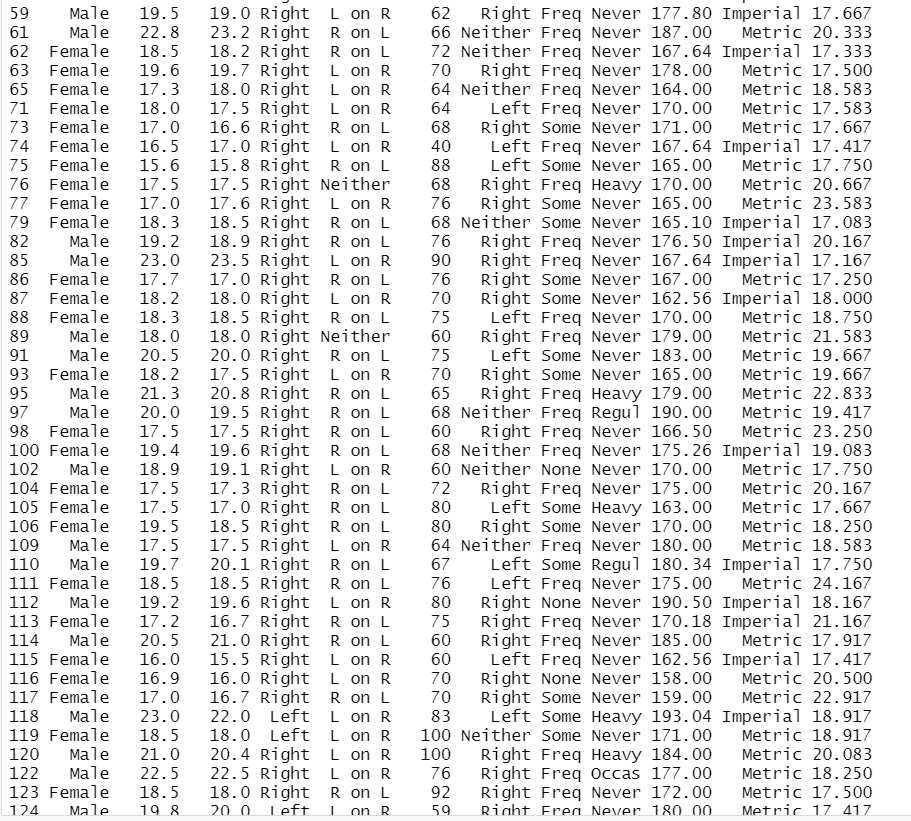
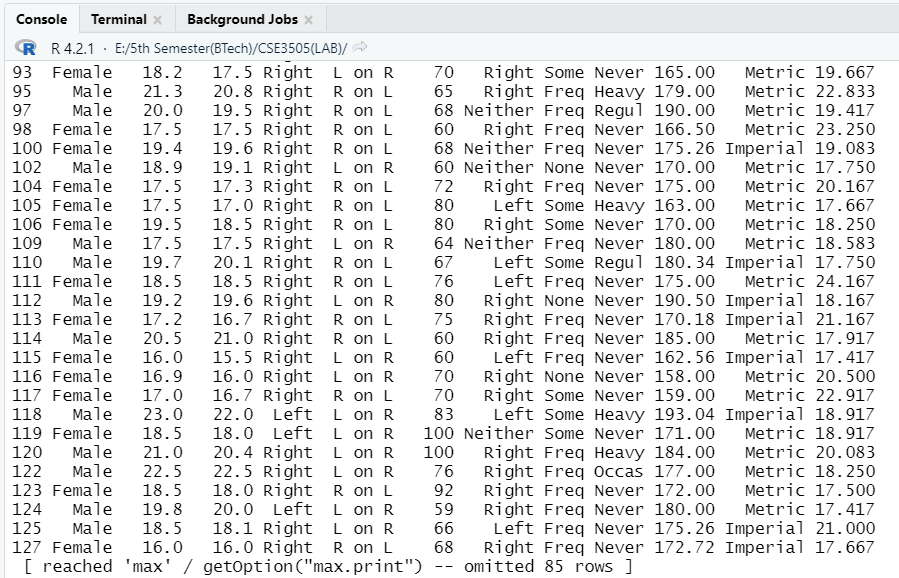
c1<-nrow(sub2)

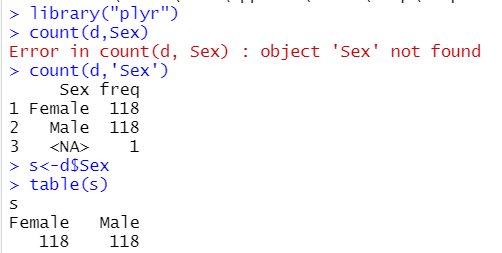
c2<-nrow(sub3)

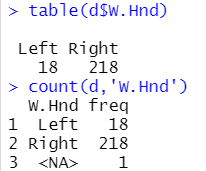
r<-(c2/c1)\*100

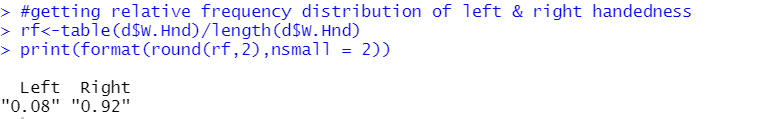
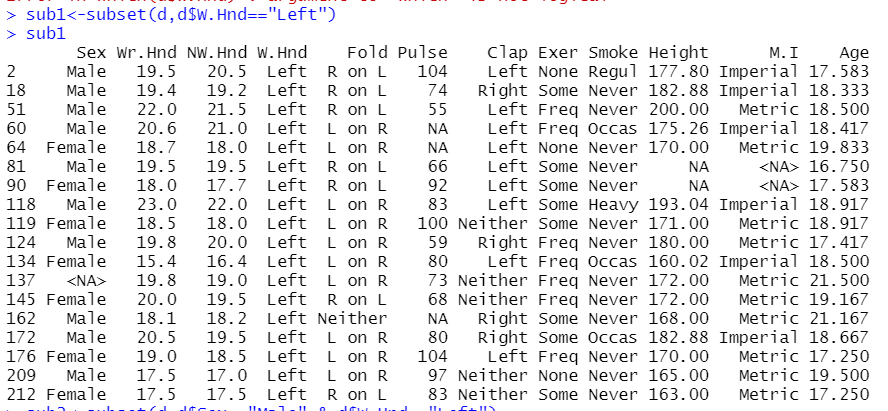
print(format(round(r,2),nsmall=2))  
  
Output:

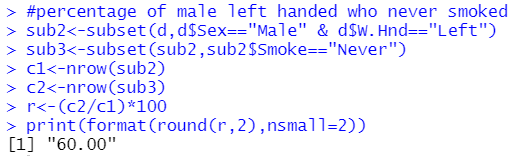
  
  
2)  
  
  
3)

  
  
  
4)2 methods to get count of sex in survey



5)Number of left and right handedness  


6)Relative frequency distribution of left & right handedness, with two decimal precisions  
  
7)Male and female left handed  
  
8)Percentage of male left handed who never smoked(upto 2 decimal places precision)



Lab-7

Code :

#function to calculate the area of a circle

carea<-function(r)

{

area=pi\*r^2

return(area)

}

carea(4)

#function to calculate area of rectangle

rarea<-function(l,b)

{

area=l\*b

return(area)

}

rarea(5,9)

#function reverse orders of numbers from 1 to 20

reverse<-function()

{

for(i in 20:1)

{

print(i)

}

}

reverse()

#function to swap two numbers

swap<-function(x,y)

{

t<-x

x<-y

y<-t

sprintf("After swap x=%d and y=%d",x,y)

}

x<-6

y<-7

swap(6,7)

#function to print squares of the numbers from 1 to a number

sqr<-function(a)

{

i<-1

while(i<=a){

b<-i^2

print(b)

i=i+1

}

}

sqr(100)

#functions with default arguments and perform as calculaator

calculator<-function(a,b)

{

a=readline("number a :")

b=readline("number b :")

a=as.integer(a)

b=as.integer(b)

sum<-a+b

d<-a-b

mul<-a\*b

div<-a/b

mod<-a%%b

sprintf("sum=%f subtraction=%f mul=%f div=%f remainder=%f",sum,d,mul,div,mod)

}

calculator(5,4)

#matrix multiplication of 3x3 matrix

matmul<-function(a,b)

{

r<-a%\*%b

return(r)

}

v2<-c(1,2,3,4,5,6,7,8,9)

v3<-c(11,12,13,14,15,16,17,18,19)

x<-matrix(v2,nrow=3,ncol=3)

y<-matrix(v3,nrow=3,ncol=3)

x

y

matmul(x,y)

#a function to print elements that are in vector a but not in vector b

f1<-function(a,b){

d1<-setdiff(a,b)

return(d1)

}

x<-c(6,20,35,40,12,17)

y<-c(20,97,16,12,35,96)

x

y

r1<-f1(x,y)

r1#present in x but not in y

# function to get three vectors A,B&C and perform AUB, A(intersection)B,A-B,A(intersection)(B U C),AU(B(intersection)C)

f2<-function(a,b,c)

{

u1<-union(a,b)#AUB

print(u1)

i2<-intersect(a,b)#A(intersection)B

print(i2)

d2<-setdiff(a,b)#A-B

print(d2)

c1<-union(b,c)

cc1<-intersect(a,c1)#A(intersection)(B U C)

print(cc1)

c2<-intersect(b,c)

cc2<-union(a,c2)#AU(B(intersection)C)

print(cc2)

}

x<-c(1,2,3,4,5)

y<-c(3,4,5,6,7,9)

z<-c(1,6,2,8,9)

r2<-f2(x,y,z)

#Read the content of ‘Events.csv’ in a data frame and view it.

mdata<-read.csv("event.csv")

mdata

#Access the scores of participants in event2(which in this case is basketball) using the column name.

mdata$basketball

#Use index number to retrieve the same data. here basket ball is saved in column number 3

mdata[,3]

#Extract the score of third participant in event3.i.e. row 1 and column 4

mdata[3,4]

#Extract the scores of the first and second participant in all the events. <=25 score is not counted else it is not counted

sc<-function(a,b)

{

print(a[b,1])

for(i in 2:8){

if(a[b,i]<=25){

print(0);

}else{

print(a[b,i])

}

}

}

p1<-sc(mdata,1)

p2<-sc(mdata,2)

#Display the names and total scores of all participants. where score is considered only for >25

ts<-function(a,b)

{

s<-0

for(i in 2:8){

if(a[b,i]<=25){

s=s+0

}else{

s=s+a[b,i]

}

}

return(s)

}

y<-c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)#an empty vector to store sum for each 20 participant

for(i in 1:20)

{

y[i]<-ts(mdata,i)

}

mdata$total<-y

mdata[,c(1,9)]

#Make the column “name” as the row index of the data frame.

row.names(mdata)<-mdata$Name

mdata

# Display the names of the students participated in event3.

par3<-function(a){

for(i in 1:20){

if(a[i,4]>25){

print(a[i,1])

}

}

}

par3(mdata)

#Obtain the names whose total score is above its average.

r4<-function(a,m)

{

for(i in 1:20){

if(a[i,9]>m){

print(a[i,1])

}

}

}

m1<-mean(mdata$total)

r4(mdata,m1)

#Mass survey exercise

install.packages("MASS")#installing package MASS

library(MASS)#loading pakcage MASS

str(survey)#structur of survey

print(class(survey))#class of survey

print(typeof(survey))#data type in survey

print(nrow(survey))#number of rows in survey

print(ncol(survey))#number of columns in survey

dim(survey)#dimension of survey

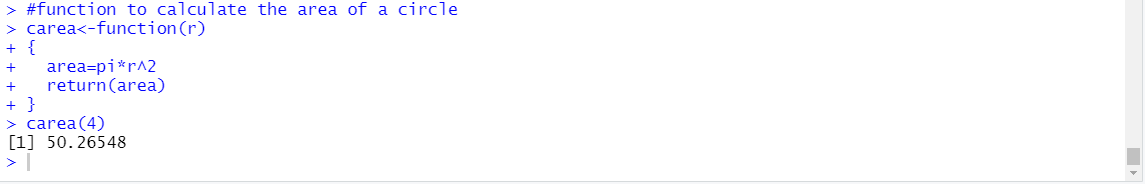
summary(survey)#summary of observation in survey

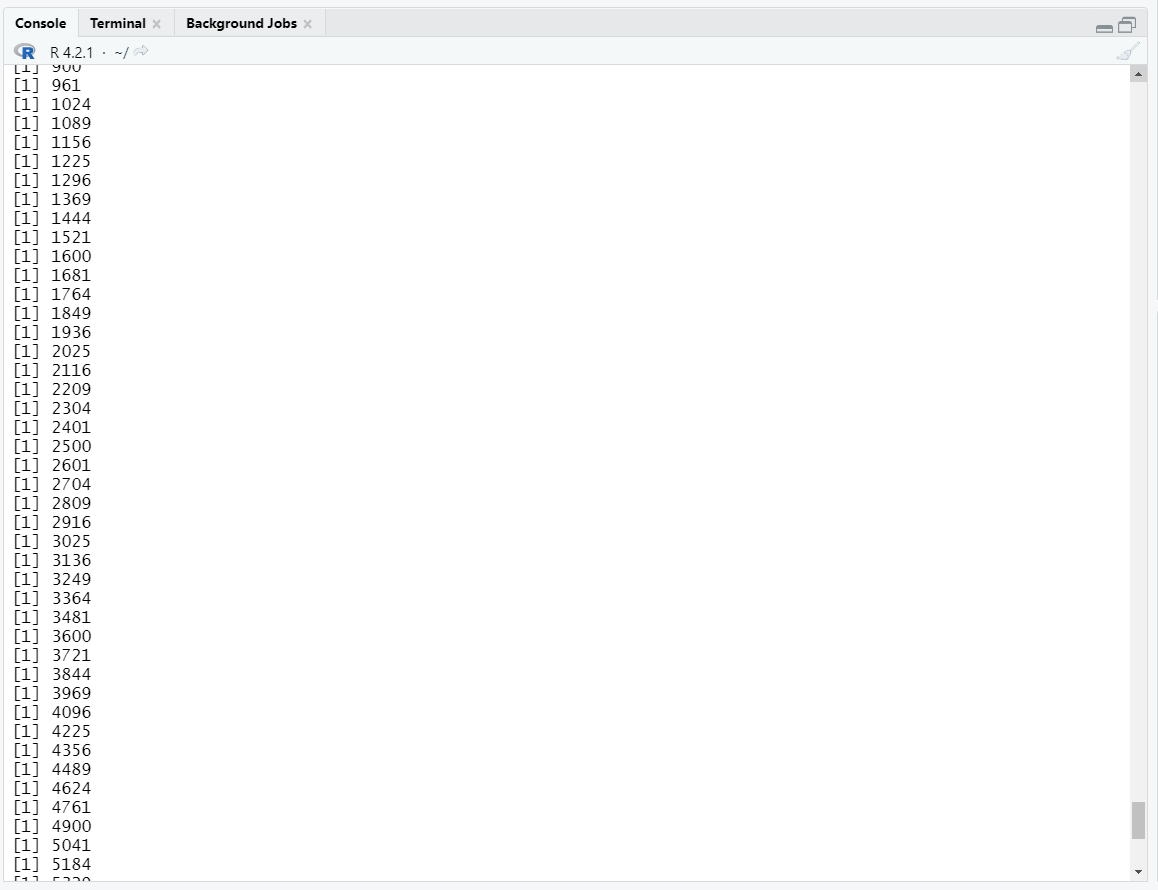
colnames(survey)#column names in survey

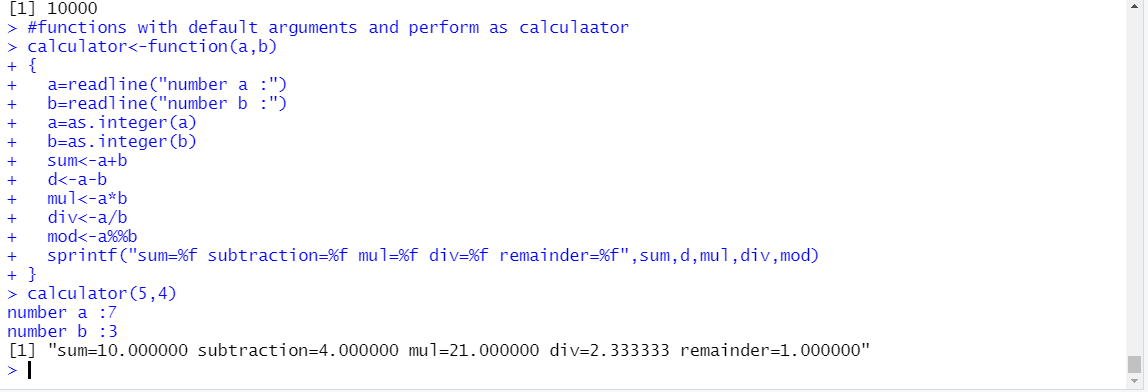
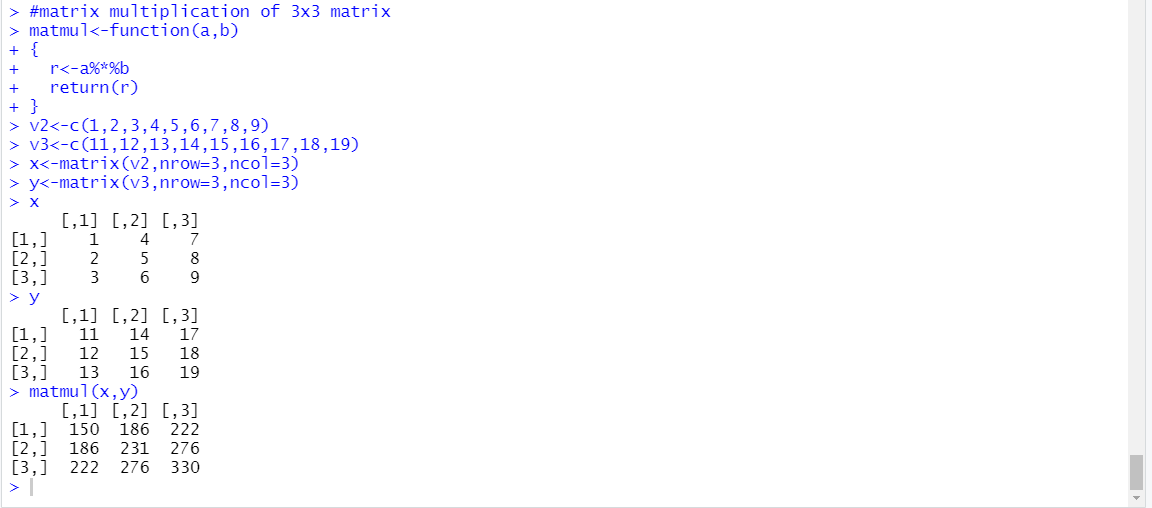
head(survey,3)#first 3 heads in survey

tail(survey,2)#last 3 elements in survey  
  
**Output:**

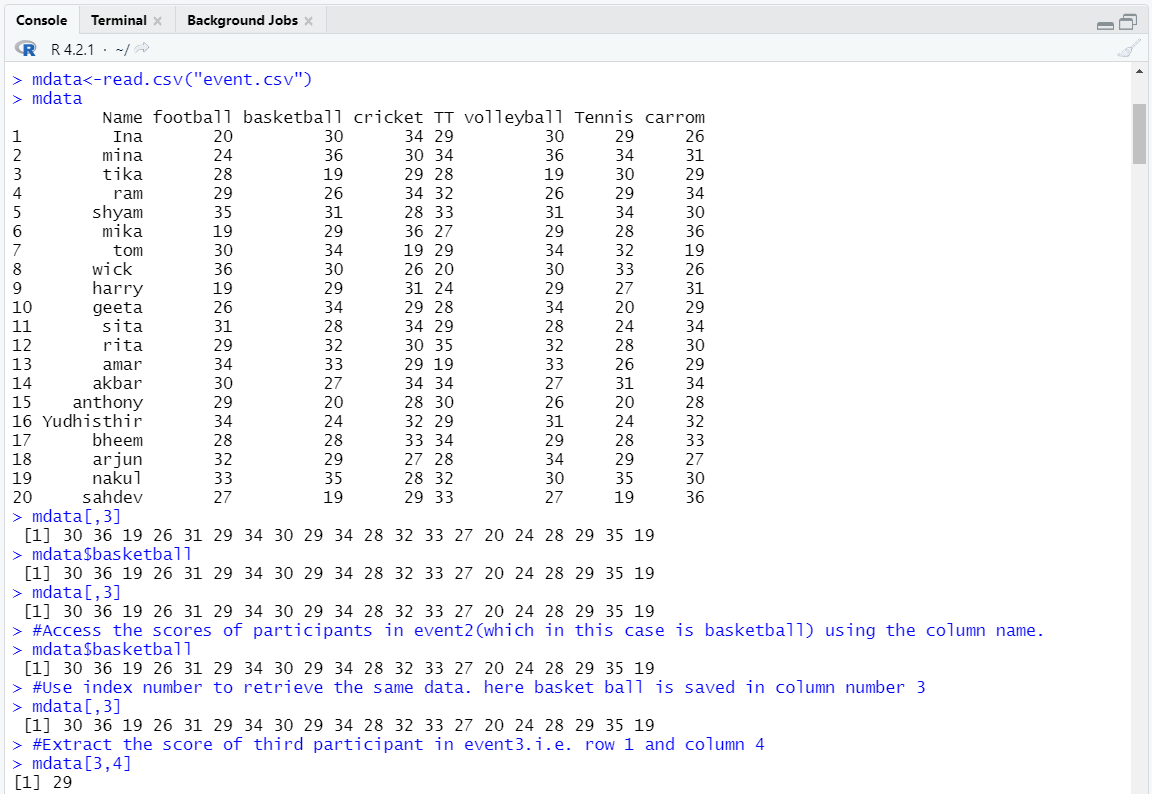
**1)**

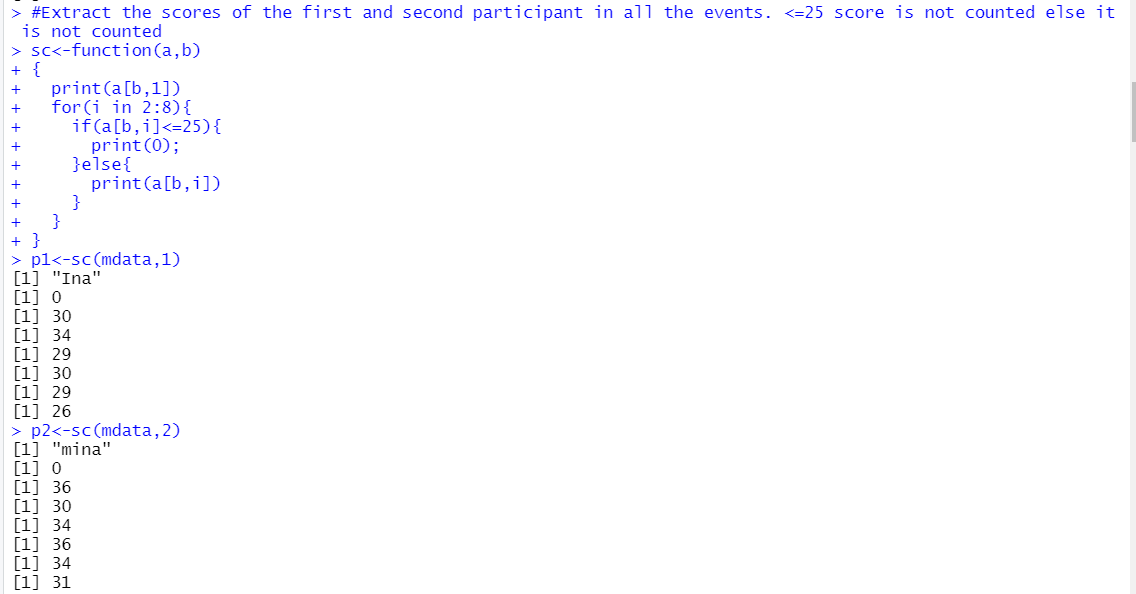
**  
  
  
  
**

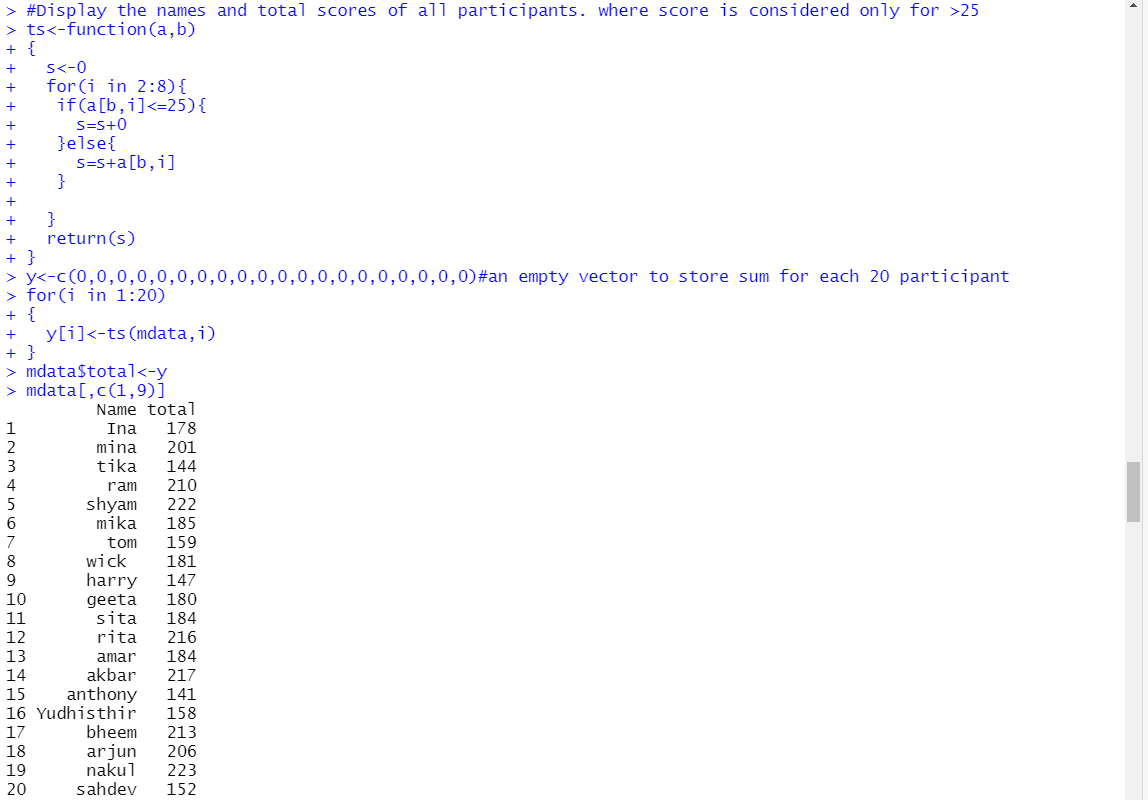
**  
**

**  
**

**2)**



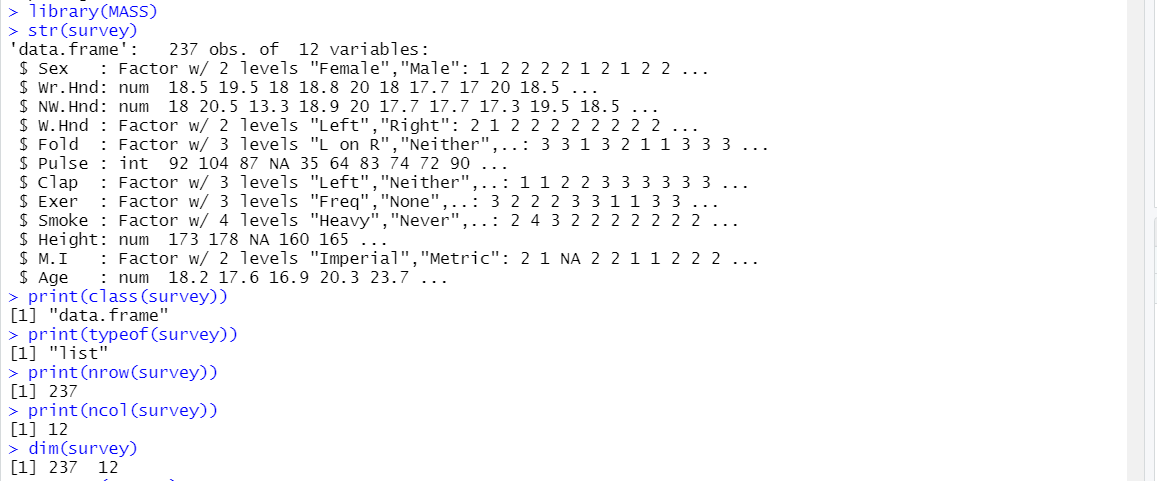


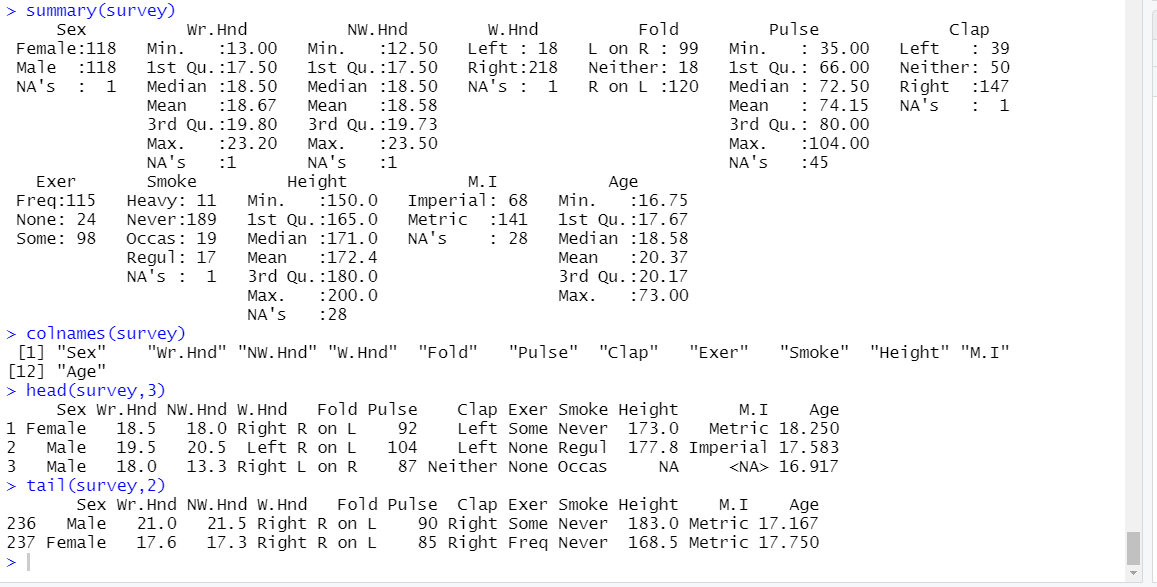


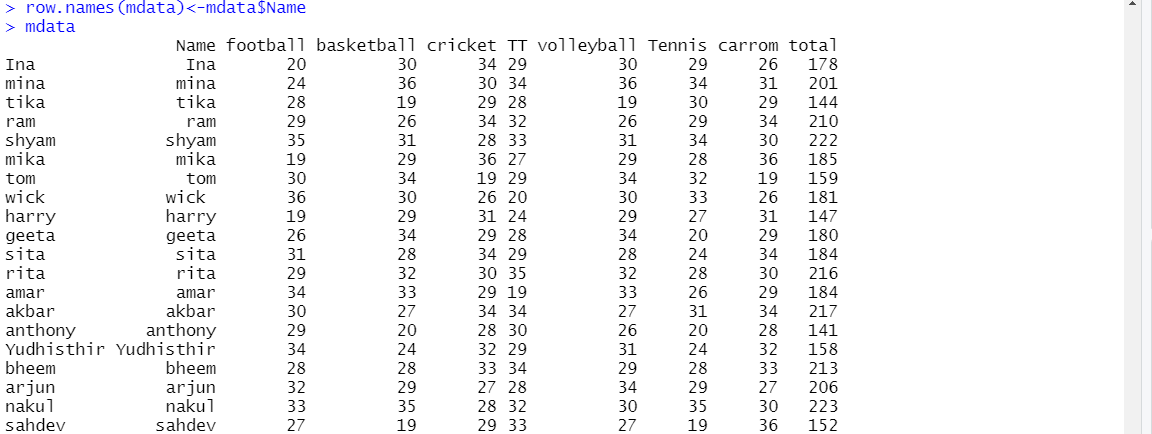
3)

Installing,loading of mass package,structure of survey,class and data types of survey,number of rows and number of columns in survey and dimension of survey

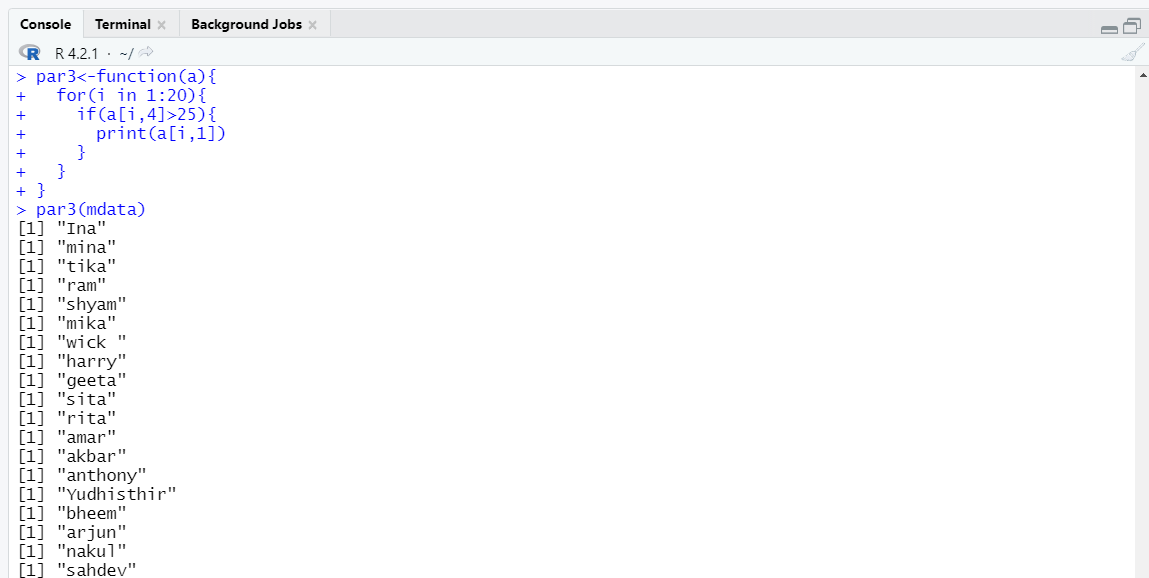


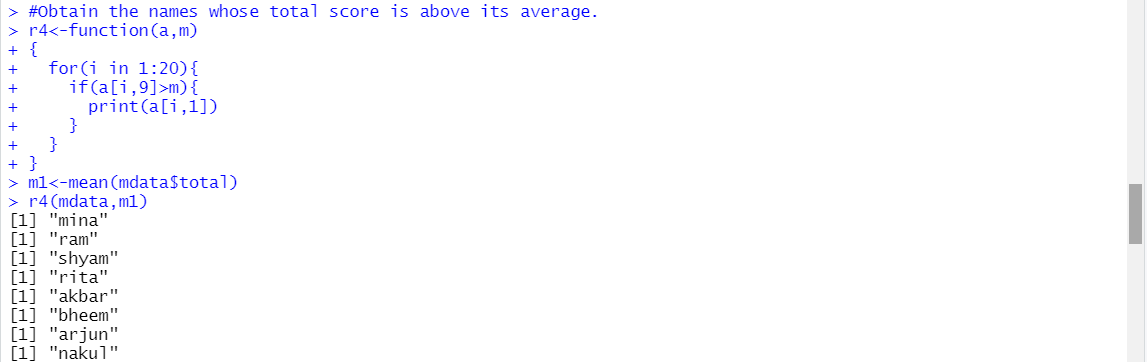


Summary of survey,name of columns in survey, 1st three elements and bottom three elements in survey  
  
  
  
  
#the name column is assigned as row name



#those who participated in event 3 ,i.e. cricket and their score is greater than 25,





Lab-8  
  
Code:

install.packages("MASS")

library(MASS)

survey

ndata<-na.omit(survey)

ndata

range(ndata$Age)#range of students in the survey participating in the survey

#method 2

m1=min(ndata$Age)

m2=max(ndata$Age)

r1=m2-m1

r1

#Break the age range into non-overlapping sub-intervals by defining a sequence of equal distance break points of 10 by rounding the range to nearest integer

ranges = as.data.frame(table(cut(ndata$Age,seq(m1,m2,10))))

ranges$Var1

#Find the distribution of the age range according to the sub-intervals with cut with its right boundary opened. Display it in column form.

as.data.frame(table(cut(ndata$Age,seq(m1,m2,10))))

#Which age range of students has mostly participated in the survey.

ranges$Var1[which.max(ranges$Freq)]

#Similarly, find the frequency distribution of Wr.Hnd span and display it in column format.

m3=min(ndata$Wr.Hnd)

m4=max(ndata$Wr.Hnd)

as.data.frame(table(cut(ndata$Wr.Hnd,seq(m3,m4,3))))

#Find the relative frequency of Wr.Hnd and display it by correcting to 3 decimal places.

round(table(ndata$Wr.Hnd)/length(ndata$Wr.Hnd), 3)

# Part B

#Find the average age of the students participated in the survey

mean(ndata$Age)

#Compute the standard deviation and variance of the height of the students participated in the survey

sd(ndata$Height)

var(ndata$Height)

#Compute the quartile of wirte hand span

quantile(ndata$Wr.Hnd)

#Find the correlation between write hand span and pulse rate of the students

cor(ndata$Wr.Hnd,ndata$Pulse)

#Find the average age of the students based on how often the student exercises.

s1 <- subset(ndata,Exer=="Freq",c(Age))

s1

mean(s1$Age)

#Find the standard deviation of height of the students under different categories of span of non-writing hand.

s2 <- subset(ndata,NW.Hnd>=17 && NW.Hnd<20,c(Height))

s3<- subset(ndata,NW.Hnd>=20,c(Height))

sd(s2$Height)

sd(s3$Height)

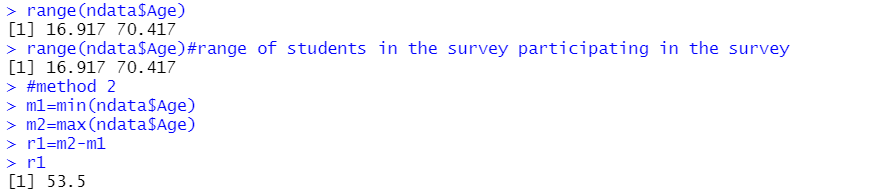
#Obtain the summary statistics of pulse rate of the students

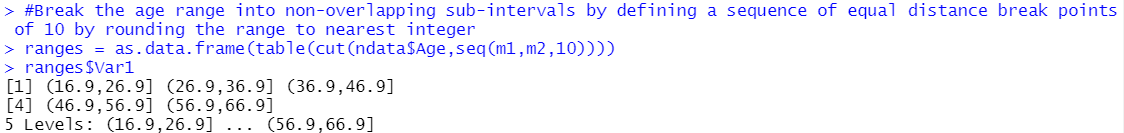
summary(ndata$Pulse)

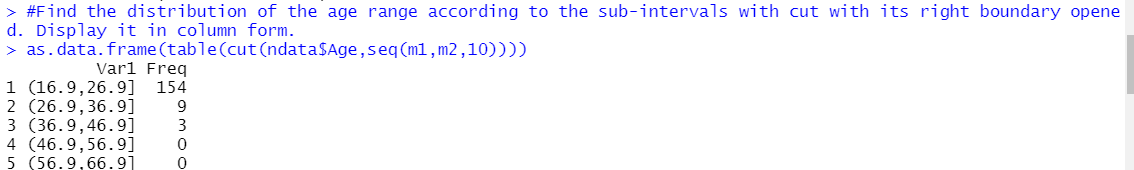
Outputs:

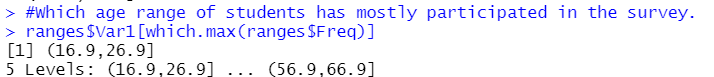
Part-A

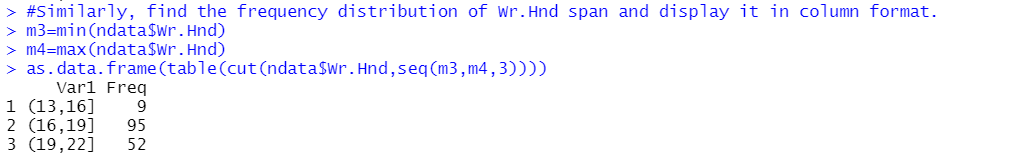
1. Find the range of students’ age participated in the survey.



2.     Break the age range into non-overlapping sub-intervals by defining a sequence of equal distance break points of 10 by rounding the range to nearest integer.  


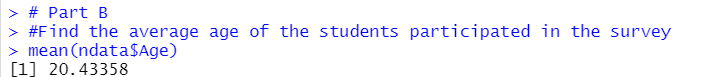
3.     Find the distribution of the age range according to the sub-intervals with cut with its right boundary opened. Display it in column form.  


4.     Which age range of students has mostly participated in the survey.  


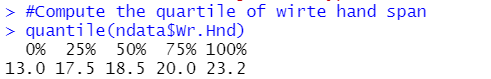
5.     Similarly, find the frequency distribution of Wr.Hnd span and display it in column format.  


6.     Find the relative frequency of Wr.Hnd and display it by correcting to 3 decimal places.  


PART -B

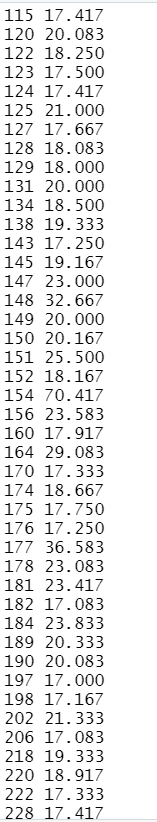
1.     Find the average age of the students participated in the survey.  


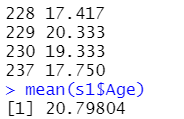
2.     Compute the standard deviation and variance of the height of the students participated in the survey.  
  


3.     Compute the quartile of wirte hand span.  


4.     Find the correlation between write hand span and pulse rate of the students.  

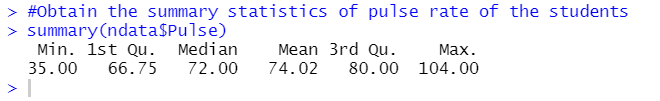

5.     Find the average age of the students based on how often the student exercises.  



6.     Find the standard deviation of height of the students under different categories of span of non-writing hand.  
  


7.     Obtain the summary statistics of pulse rate of the students.



Lab-9  
Code:

mdf<-read.csv('Inputdata.csv')

mdf

cor(mdf$T,mdf$RH,method="pearson")

cor(mdf$T,mdf$RH,method="spearman")

cor(mdf$T,mdf$RH,method="kendall")

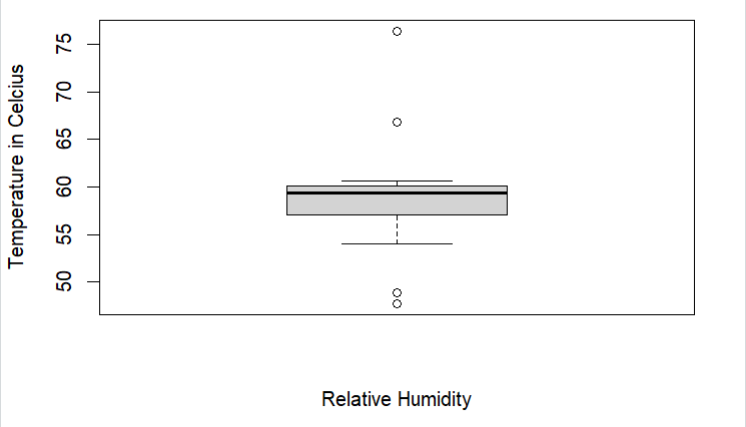
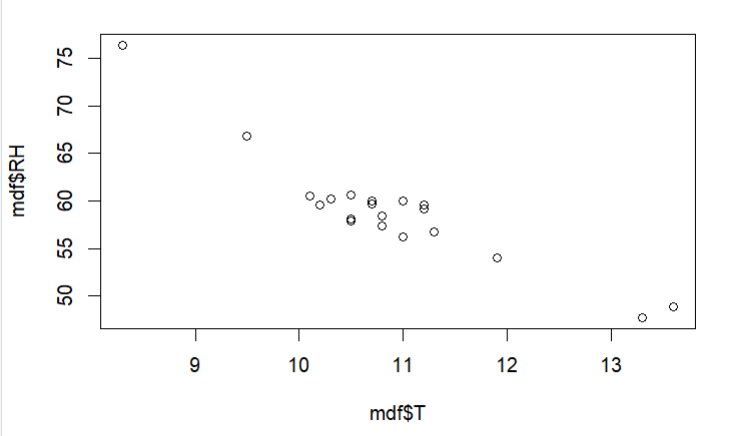
plot(mdf$T,mdf$RH)

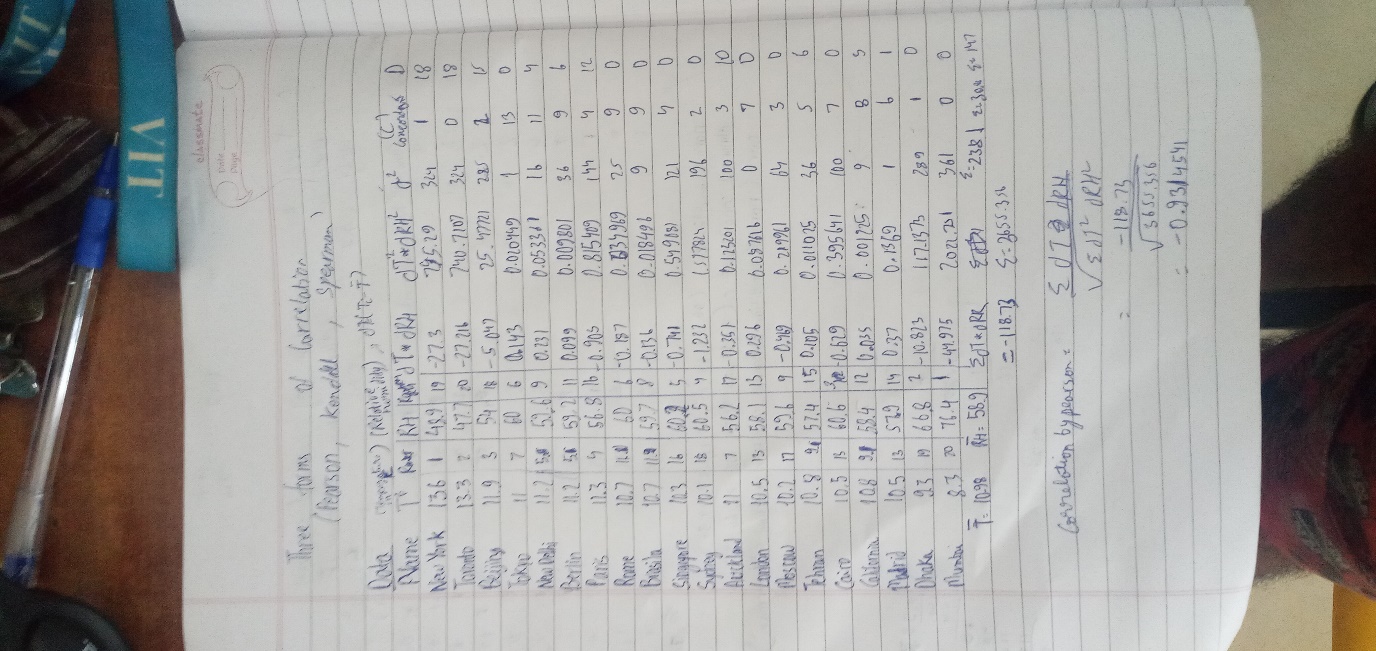
boxplot(mdf$RH,data=mdf,xlab="Relative Humidity",ylab="Temperature in Celcius")

# there fore it can be observed that Relative humidity decreases with increases in temperature

Output:



  
  
manual



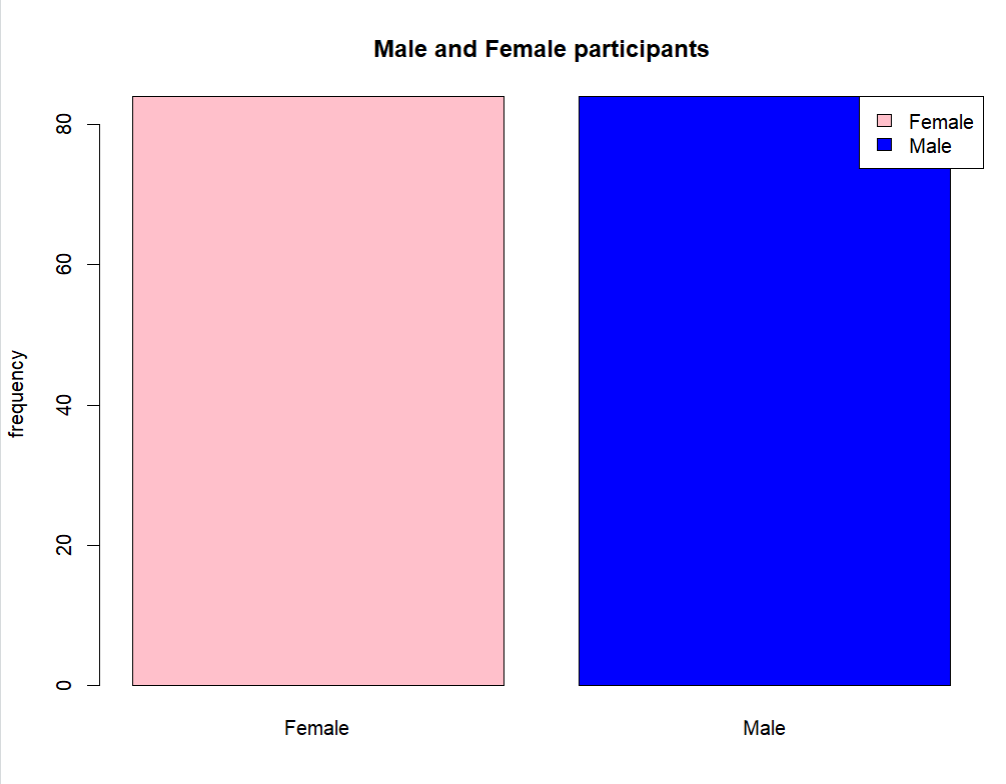
lab-10  
**PART A**

Use the newsurvey data obtained by cleaning ‘na’ values in survey data of MASS package to do the following:

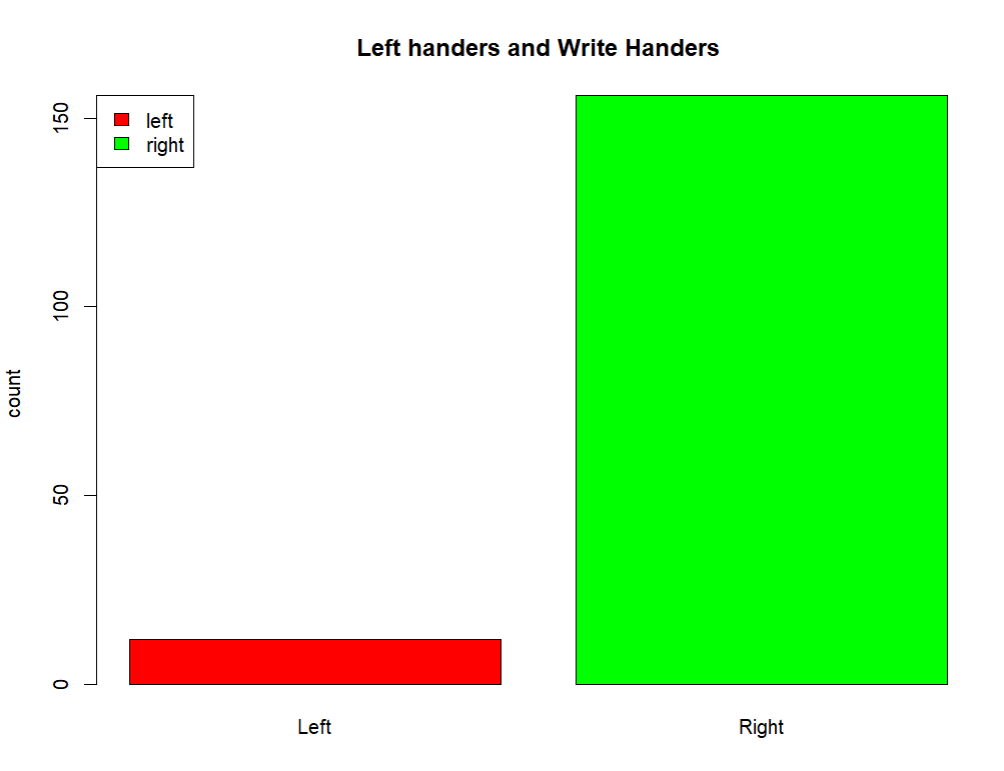
1. Plot a bar graph for the number of male and female participants in the survey. Provide the title as “Male and Female participants”, y-axis label as “frequency” and specify the colours for the bars.

Code:

barplot(table(df$Sex),main="Male and Female participants",ylab="frequency",col=c("pink","blue"))

legend("topright",c("Female","Male"),fill=c("pink","blue"))  


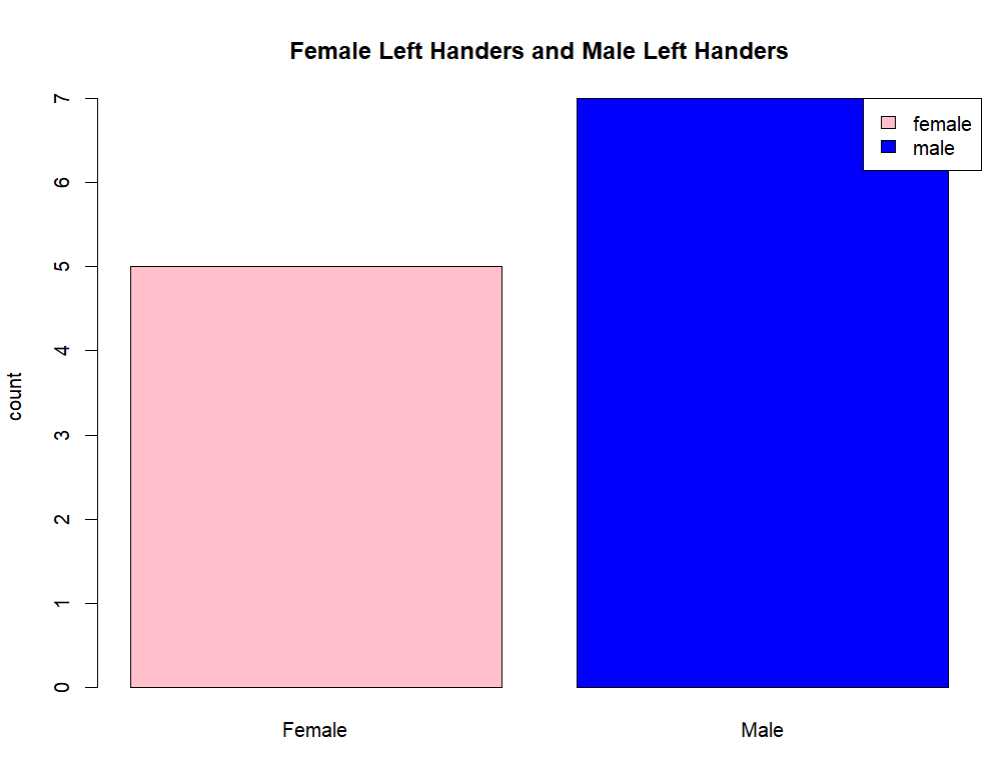
2.     Plot a bar graph for the number of left handers and right handers in the survey. Provide the title as “Left Handers and Right Handers”, y-axis label as “count” and specify the colours for the bars.  
code:  
barplot(table(df$W.Hnd),main="Left handers and Write Handers",ylab="count",col=c("red","green"))

legend("topleft",c("left","right"),fill=c("red","green"))  


3.     Plot the distribution between male left handers and female left handers using bar chart. Provide the title as “Female Left Handers and Male Left Handers , y-axis label as “count” and specify the colours for the bars.  
code:  
df3=subset(df,W.Hnd=="Left")

df3

barplot(table(df3$Sex),main="Female Left Handers and Male Left Handers",ylab="count",col=c("pink","blue"))

legend("topright",c("female","male"),fill=c("pink","blue"))  


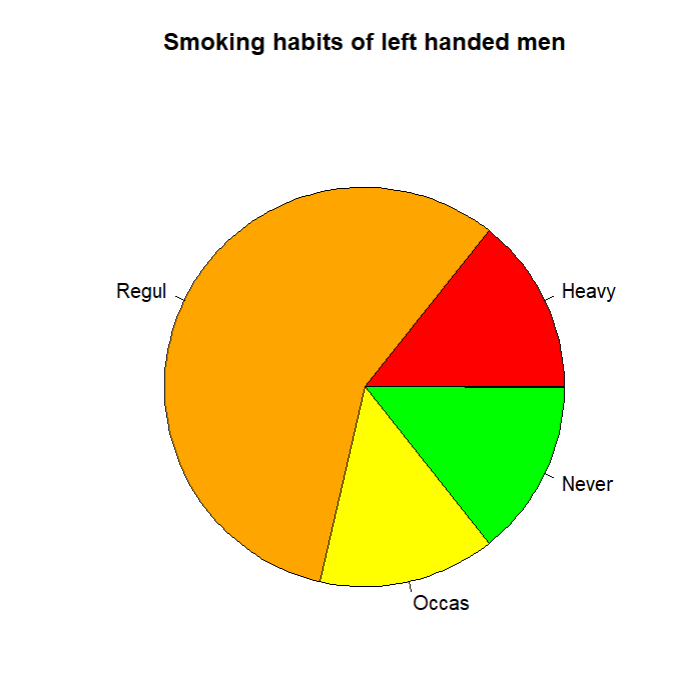
4.     Draw the distribution of smoking habits of male left handers using pie chart.

Code:

df4=subset(df,W.Hnd=="Left"& Sex=="Male")

df4

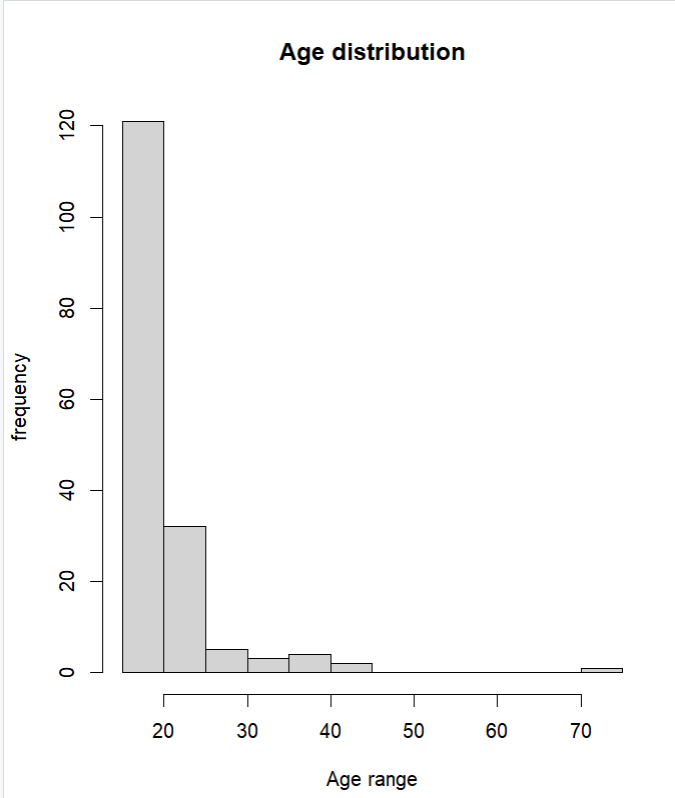
labels=c("Heavy","Regul","Occas","Never")

pie(table(df4$Smoke),labels,main="Smoking habits of left handed men",col=c("Red","Orange","Yellow","Green"))  


5.     Draw the histogram of age distribution with the title as ‘Age distribution’ and xlabel as ‘Age range’ and ylabel as ‘frequency’.

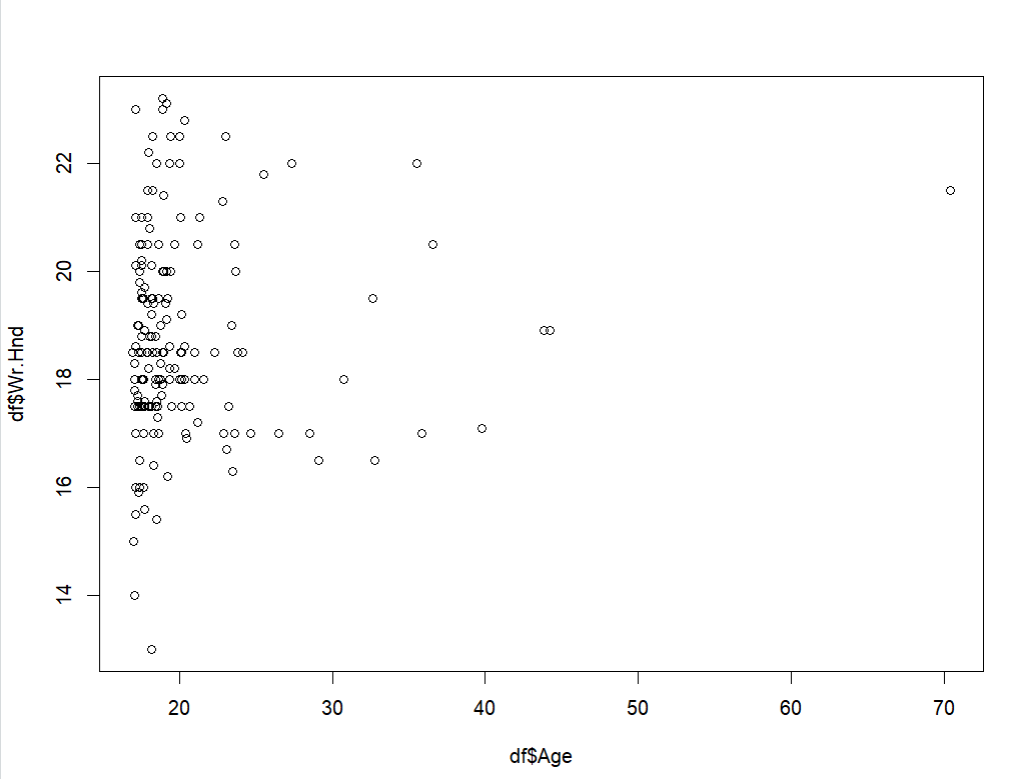
Code:

hist(df$Age,main="Age distribution",xlab="Age range",ylab="frequency")



6.     Reveal the relationship between the age and writing hand span using scatter plot.

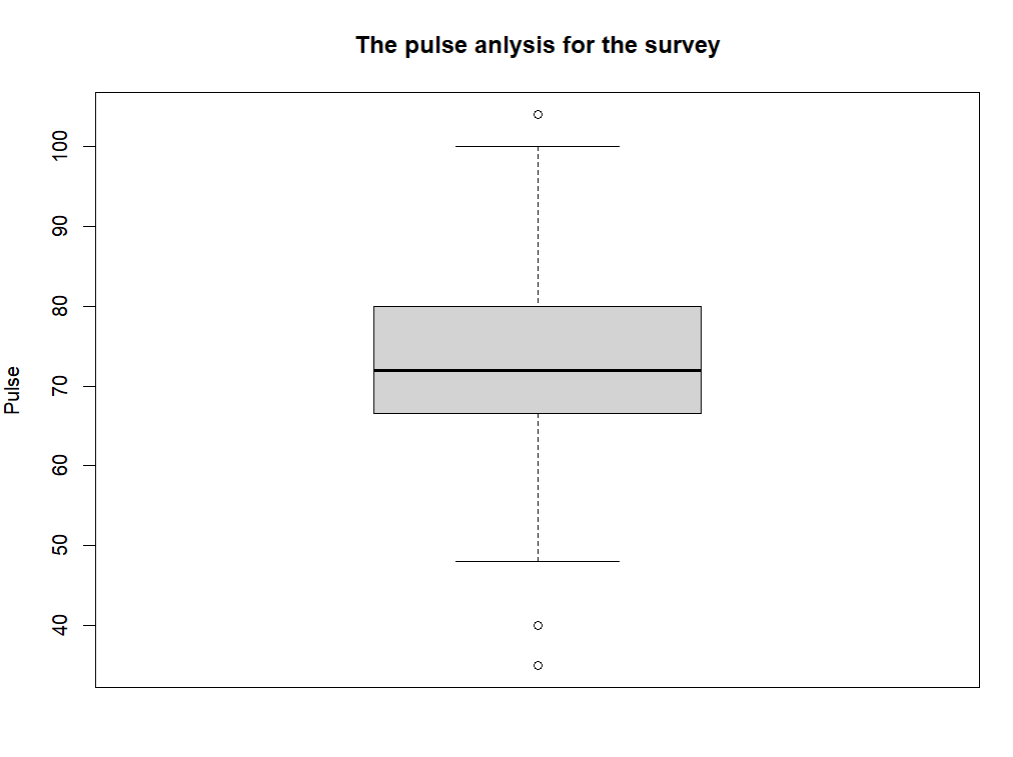
Code:

plot(df$Age,df$Wr.Hnd)  


7.     Draw the boxplot for pulse rate to analyse the five summary statistics. Provide appropriate title and label.

Code:

boxplot(df$Pulse,main="The pulse anlysis for the survey",ylab="Pulse")



PART B

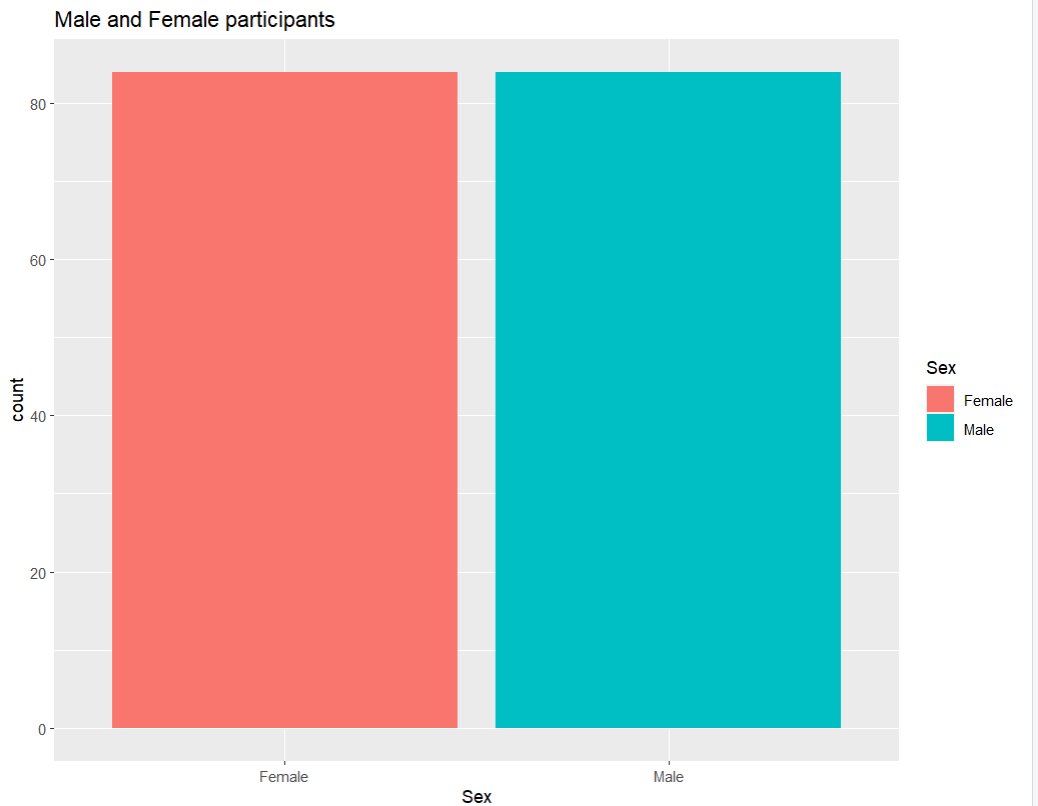
Use the newsurvey data obtained by cleaning ‘na’ values in survey data of MASS package and ggplot2 package to do the following:

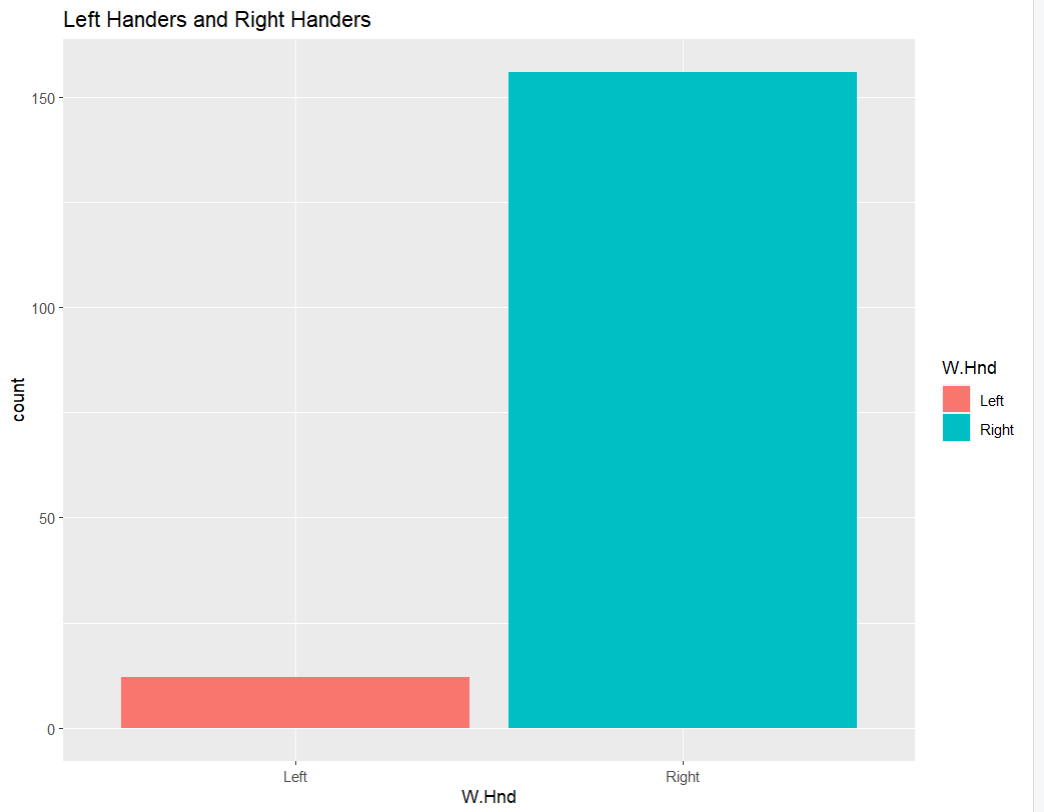
1. Install the package ggplot2 and import it.  
   code:

install.packages("ggplot2")

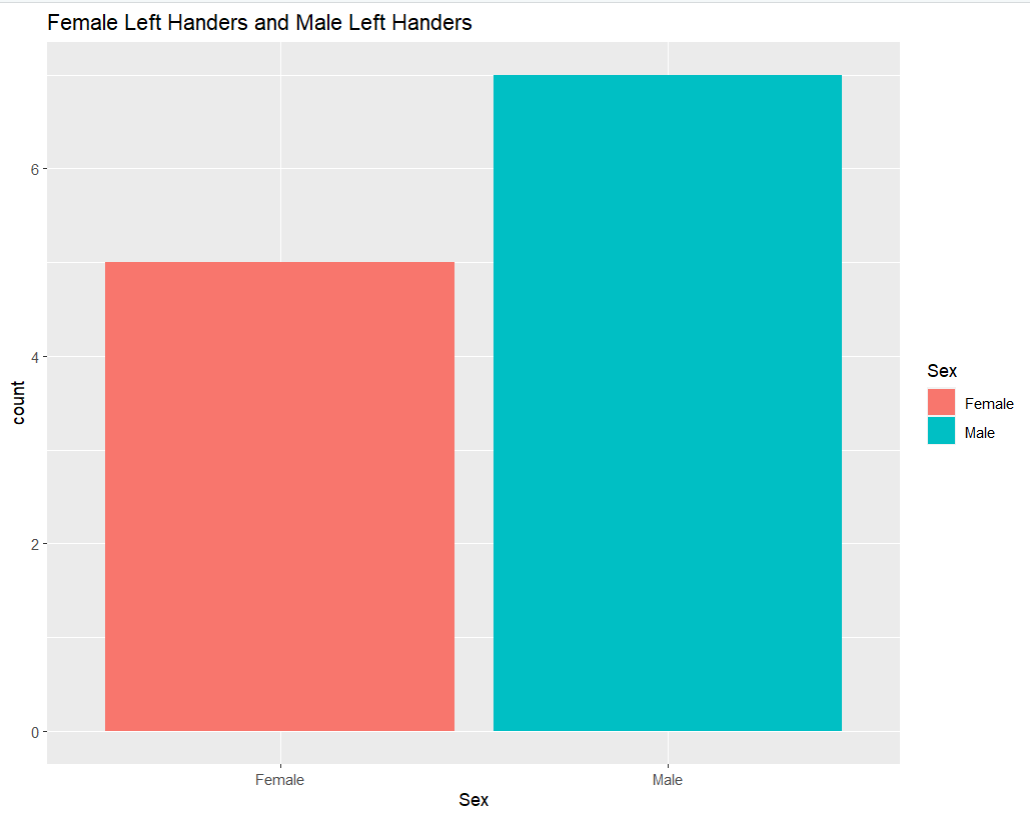
library(ggplot2)

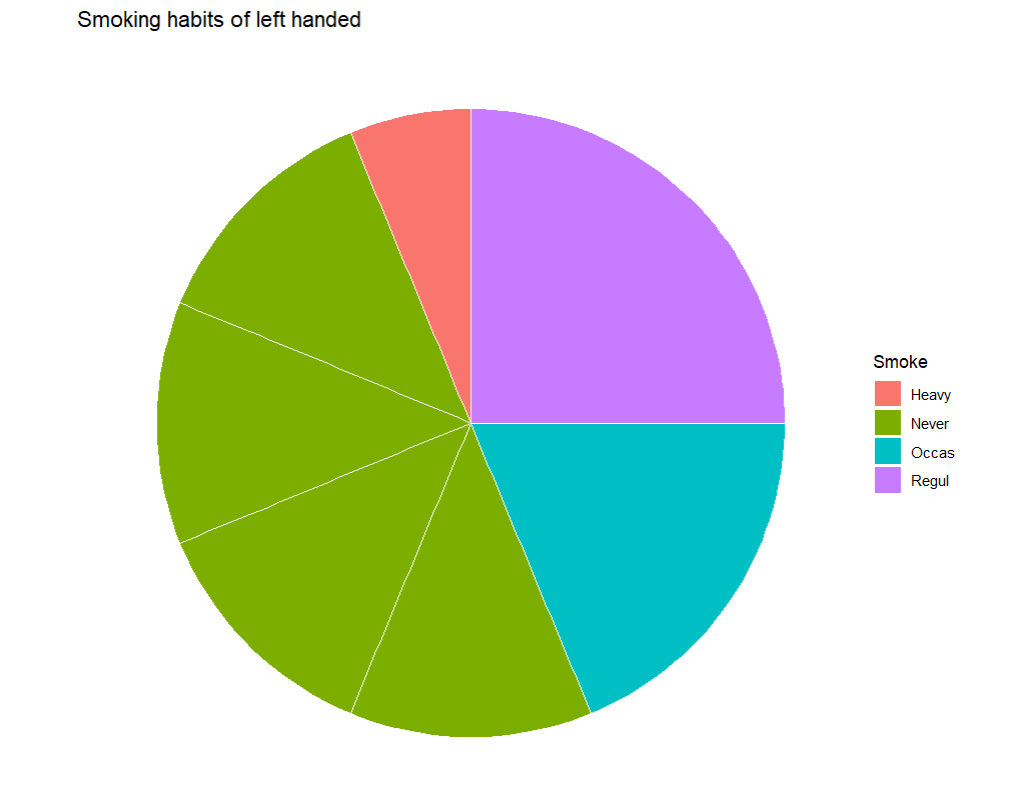
1. Plot a bar graph for the number of male and female participants in the survey. Provide the title as “Male and Female participants” and specify the colours for the bars.  
    Code  
   ggplot(df,aes(x=Sex,fill=Sex))+geom\_bar()+ggtitle("Male and Female participants")



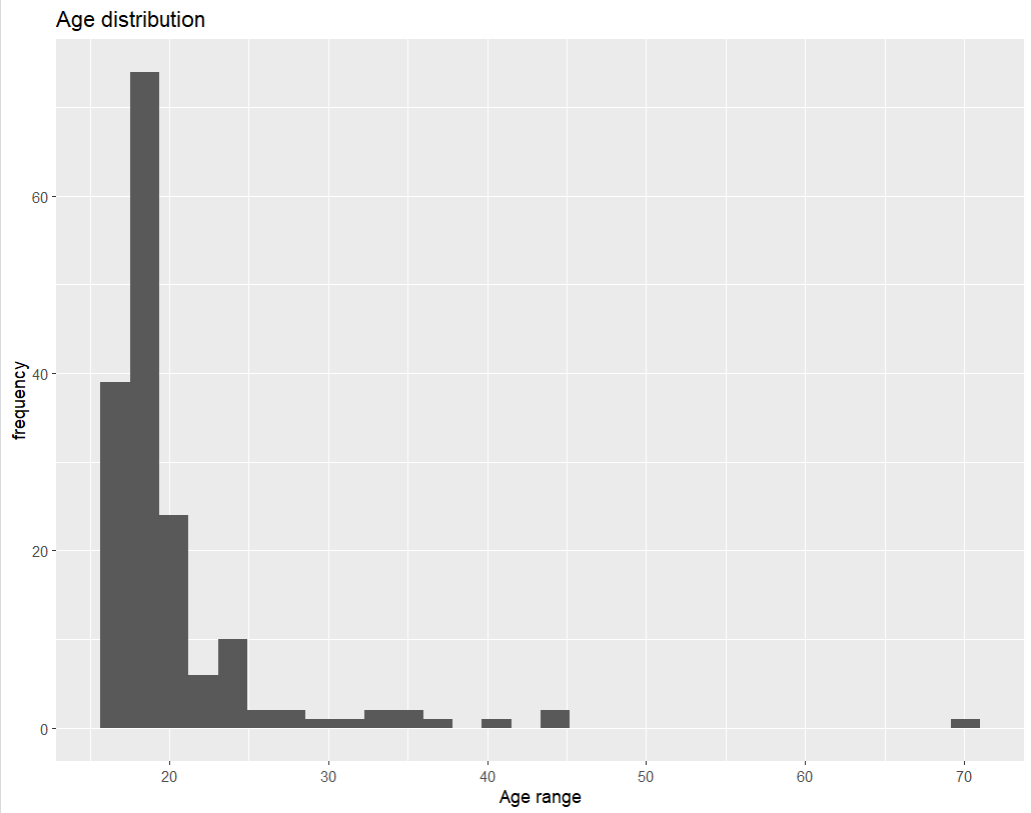
3.     Plot a bar graph for the number of left handers and right handers in the survey. Provide the title as “Left Handers and Right Handers” and specify the colours for the bars.  
Code:  
ggplot(df,aes(x=W.Hnd,fill=W.Hnd))+geom\_bar()+ggtitle("Left Handers and Right Handers")  


4.     Plot the distribution between male left handers and female left handers using bar chart. Provide the title as “Female Left Handers and Male Left Handers” and specify the colours for the bars.  
Code:

ggplot(df3,aes(x=Sex,fill=Sex))+geom\_bar()+ggtitle("Female Left Handers and Male Left Handers")  


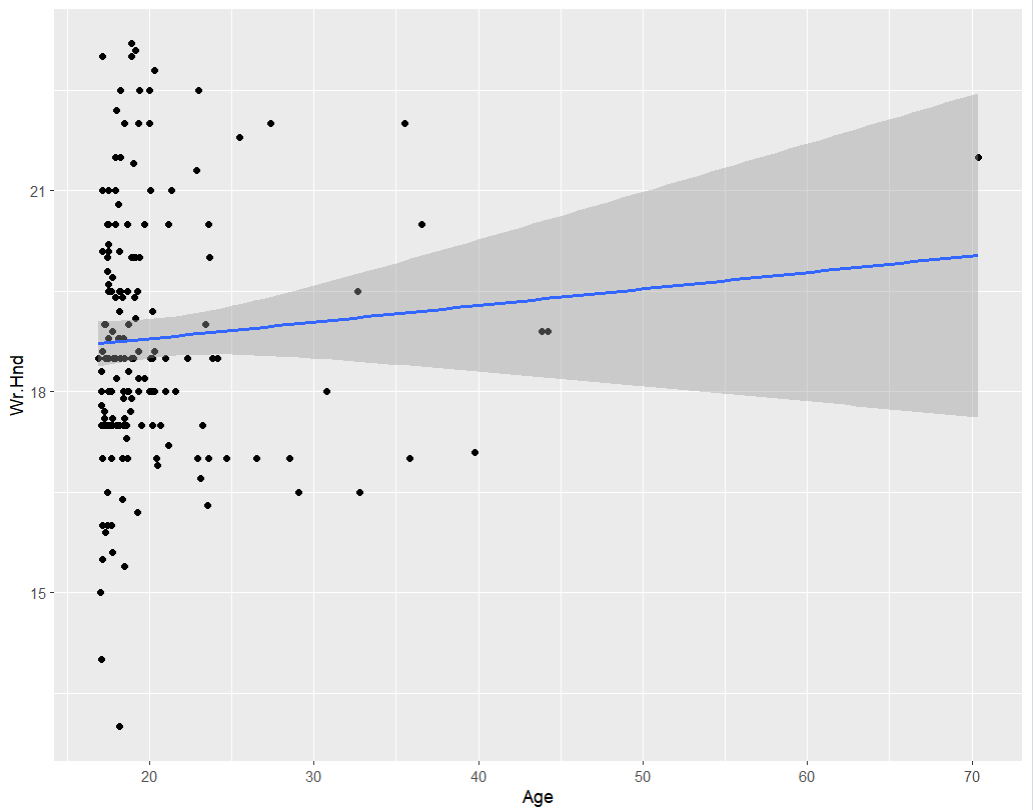
5.     Draw the distribution of smoking habits of male left handers based on age using pie chart.  
Code:  
ggplot(df4,aes(x="",y=Smoke,fill=Smoke))+ggtitle("Smoking habits of left handed")+geom\_bar(stat="Identity",width=1,color="white")+coord\_polar("y",start=0)+theme\_void()  


6.     Draw the histogram of age distribution with the title as ‘Age distribution’ and xlabel as ‘Age range’ and ylabel as ‘frequency’.  
Code:  
ggplot(df,aes(x=Age))+geom\_histogram()+labs(title="Age distribution",x="Age range",y="frequency")



7.     Reveal the relationship between the age and writing hand span using scatter plot.

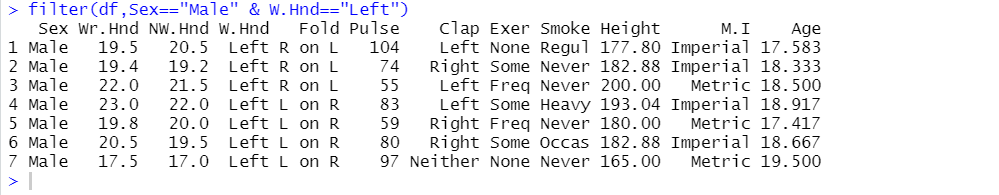
Code:  
ggplot(df,aes(x=Age,y=Wr.Hnd))+geom\_point()+geom\_smooth(method='lm')

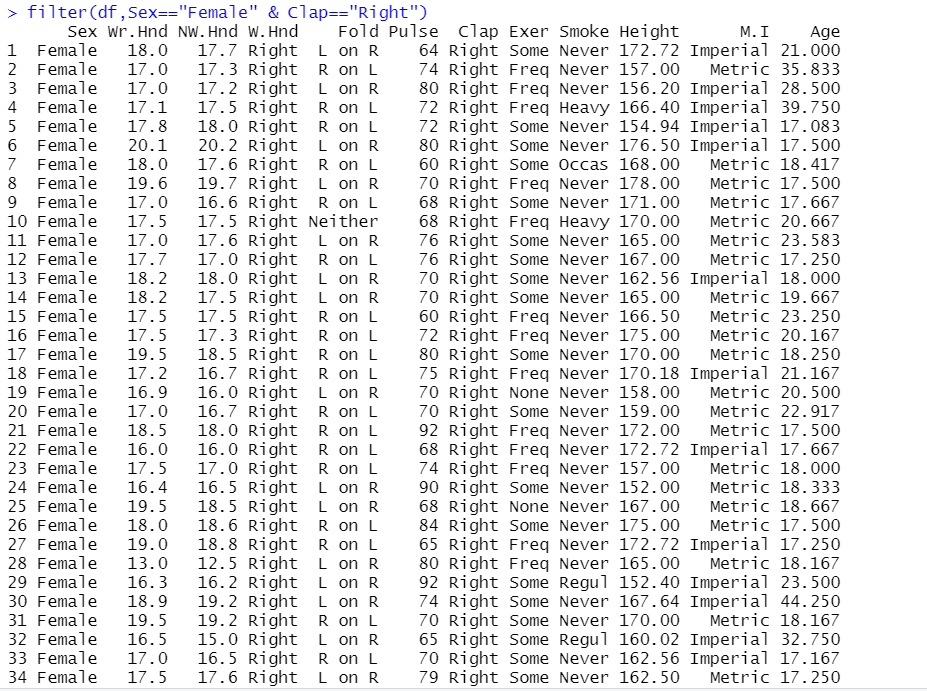
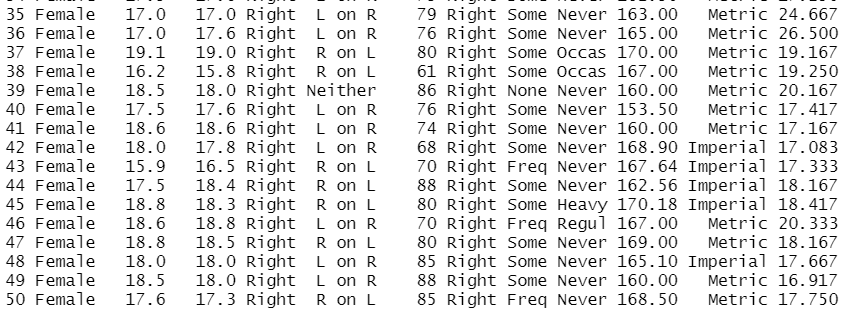


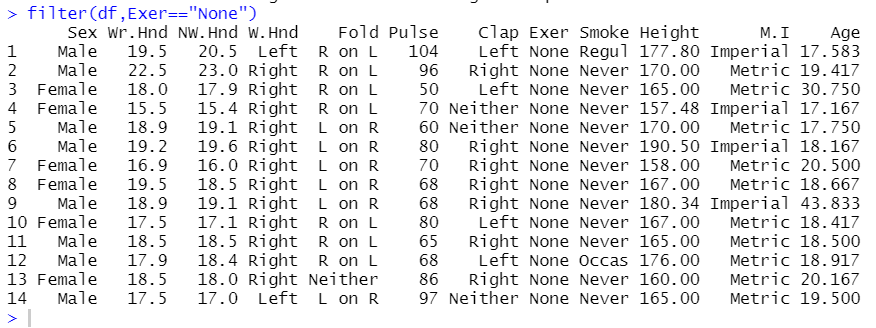
Lab-11  
  
Use the newsurvey data obtained by cleaning ‘na’ values in survey data of MASS package to do the following:

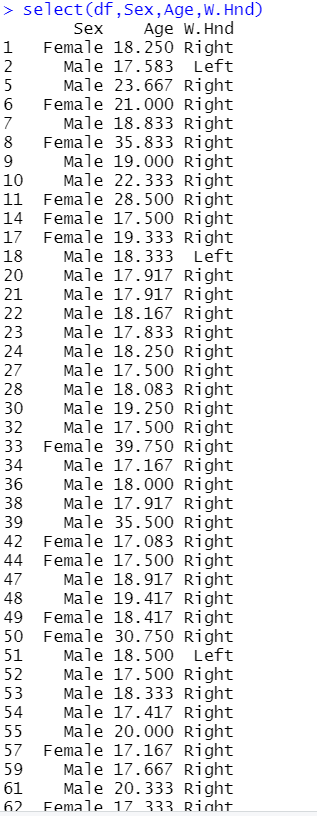
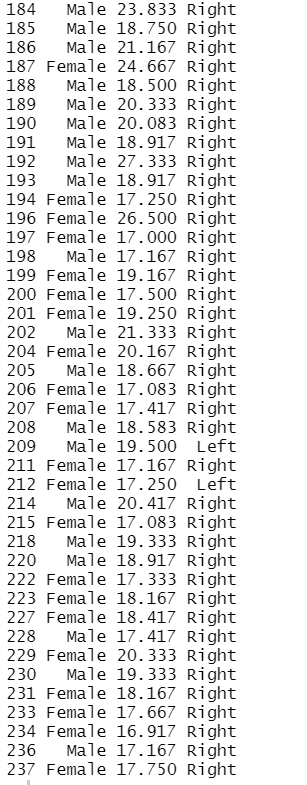
1. Install the dplyr package and import it.

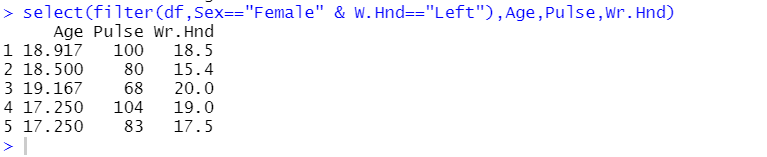
library(dplyr)  


2.     Filter all male left handers.  
  
filter(df,Sex=="Male" & W.Hnd=="Left")  


3.     Display all female right handers who keep right on top while clapping.  
  
filter(df,Sex=="Female" & Clap=="Right")  
  


4.     Display all students who never exercise.  
  
filter(df,Exer=="None")  


5.     Display only the gender, age and writing hand of the students.  
  
select(df,Sex,Age,W.Hnd)  
 

6.     Display the age, pulse rate and writing hand span of female left handers.  
  
select(filter(df,Sex=="Female" & W.Hnd=="Left"),Age,Pulse,Wr.Hnd)  
#in the code above filter is used to get only left handed females  
  


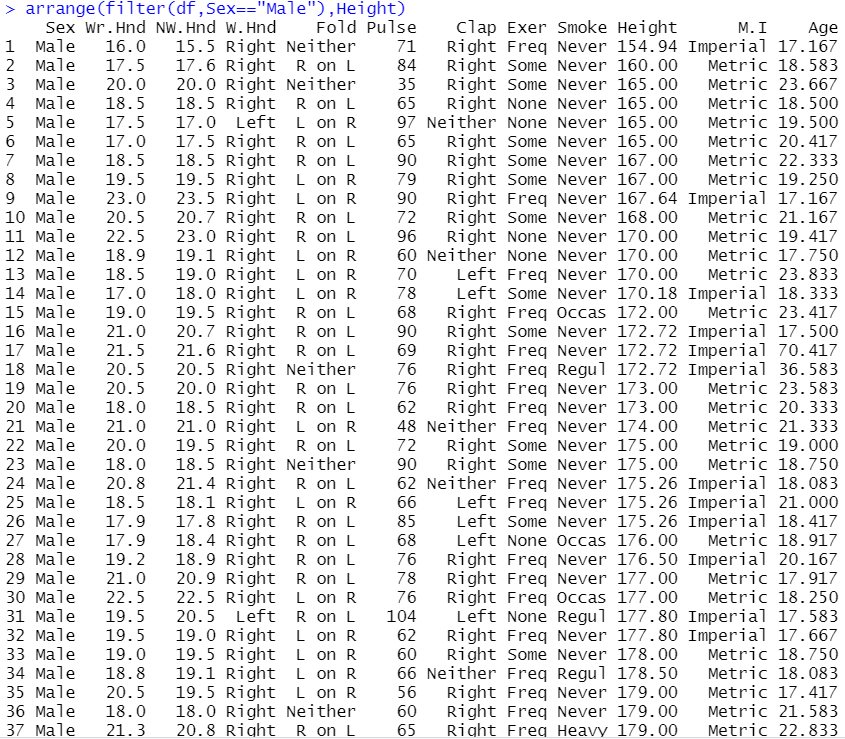
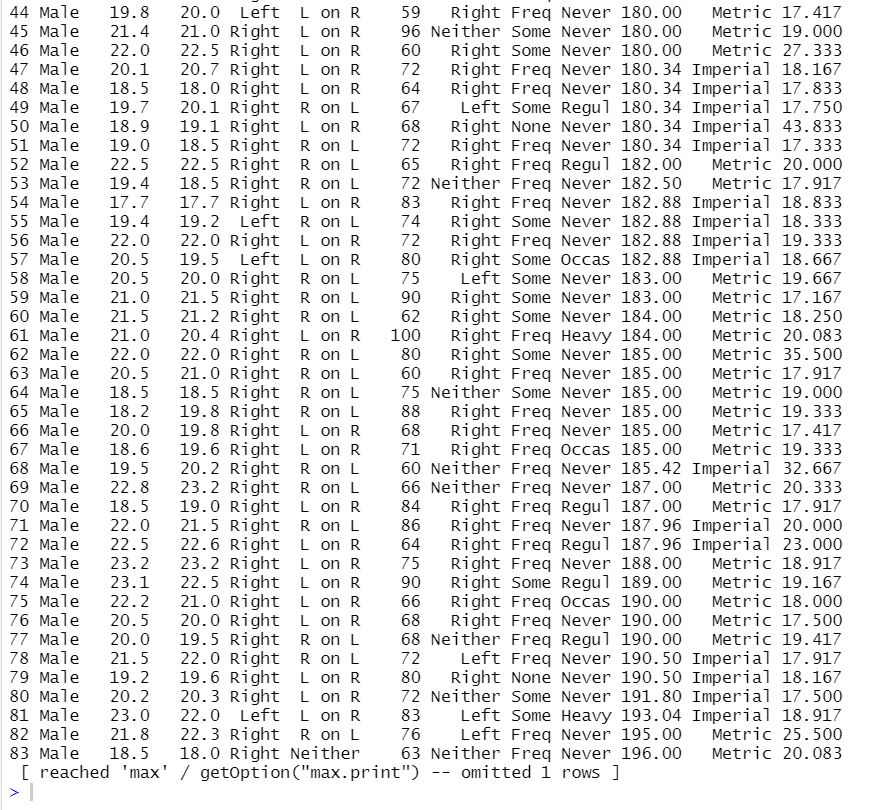
Note: use dplyr::select while using select operator as dplyr package conflicts with MASS package

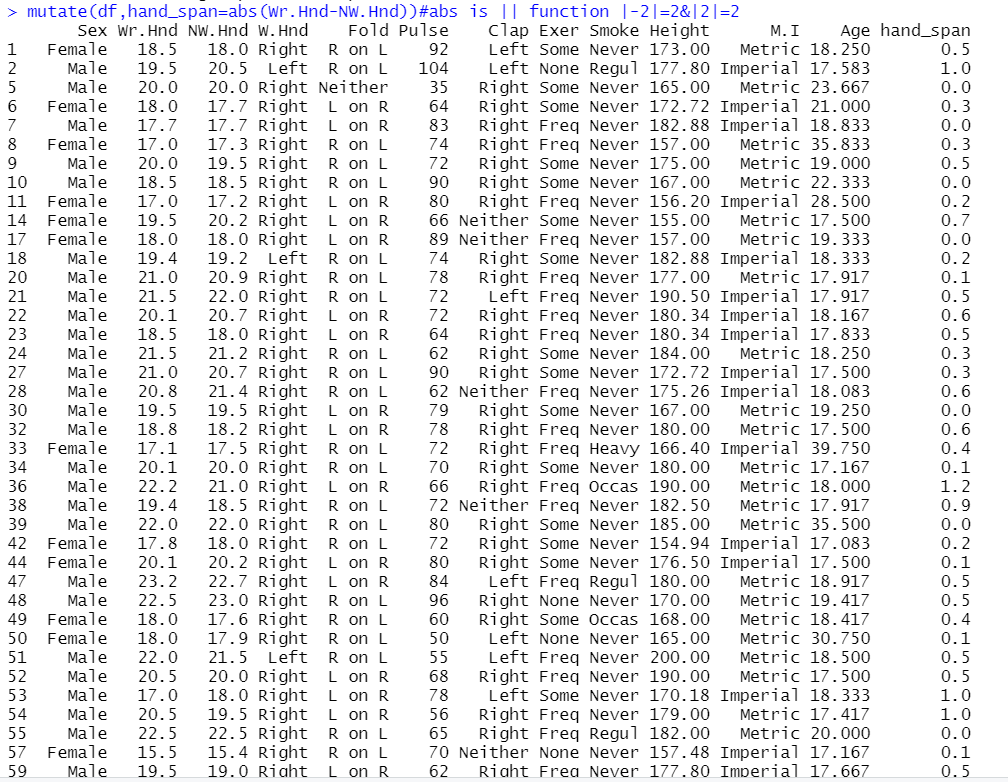
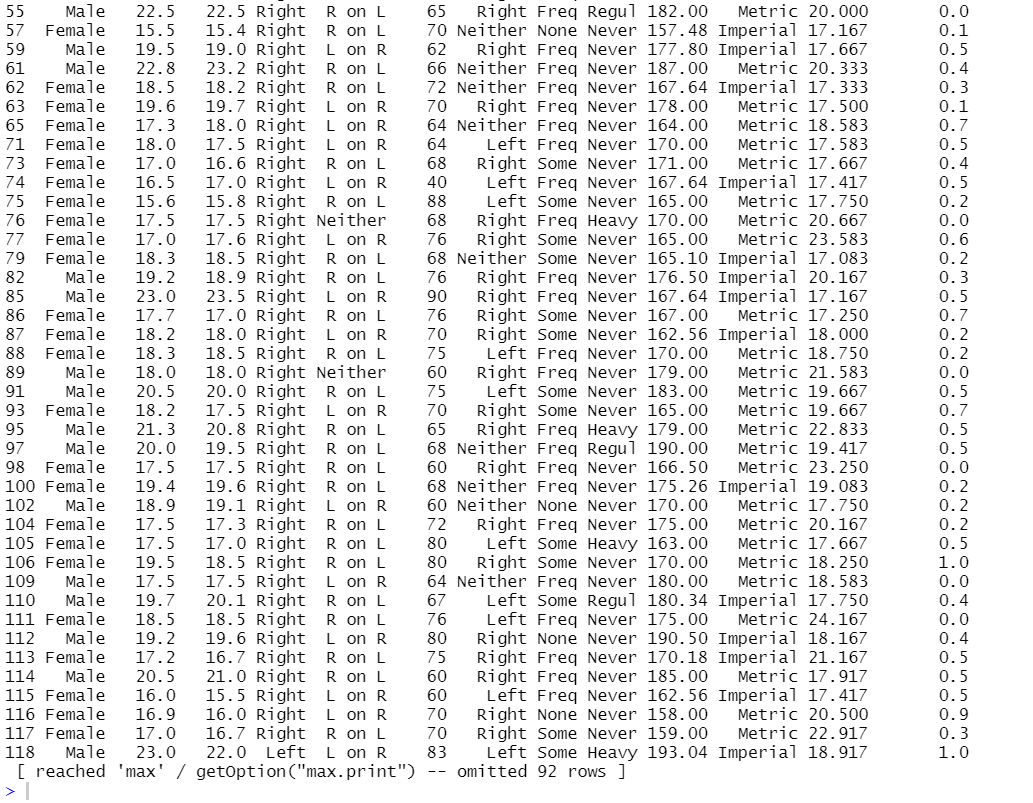
**More functions in dplyr package**

Use the newsurvey data obtained by cleaning ‘na’ values in survey data of MASS package to do the following:

1. 1.     Install the dplyr package and import it.  
     
   Install the dplyr package and import it.

library(dplyr)  


2.     Arrange all male left handers according to descending order of their heights.  
  
arrange(filter(df,Sex=="Male"),Height)  
  


3.     Introduce a new column hand\_span which contains the value as difference between the span of writing hand and non-writing hand and display it along with gender, writing hand and non-writing hand span.  
  
mutate(df,hand\_span=abs(Wr.Hnd-NW.Hnd))#abs is || function |-2|=2&|2|=2  
  


4.     Display the average writing span of male and female left handers.  
  
summarize(filter(df,W.Hnd=="Left"),mean(Wr.Hnd))  


5.     Find the maximum pulse rate of male left and right handers.  
  
summarize(filter(df,Sex=="Male"),max(Pulse))  
  
