Assignment – 6

Q= 1. Write a R program using control operators to test whether following values are prime

numbers or not by providing a PRIME or NOT PRIME message as output :

A. 103

B. 82

C. 179

**Code - xy<-c(103,82,179)**

**#For prime number :**

**for(j in 1:length(xy)){**

**num<-xy[j]**

**flag=0**

**for(i in 2:(num-1)) {**

**if ((num %% i) == 0) {**

**flag = 1**

**break**

**}**

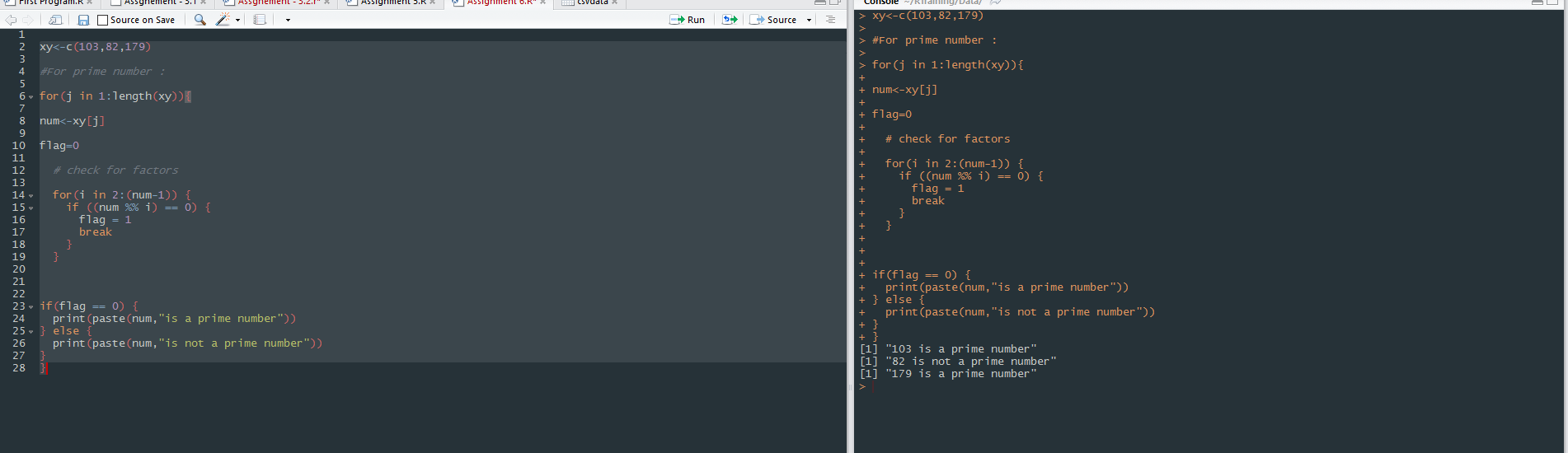
**}**

**if(flag == 0) {**

**print(paste(num,"is a prime number"))**

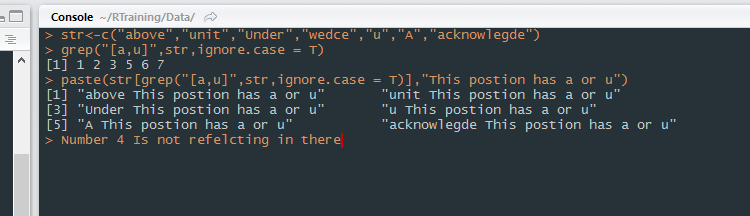
**} else {**

**print(paste(num,"is not a prime number"))}}**

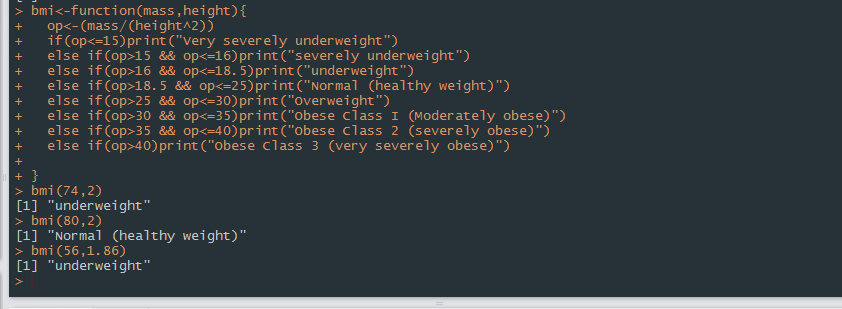


Q2 – Write a R program using control operators to identify letter u and a both occur in the

following words:



Q3 –Write a function that to calculate BMI (Body Mass Index):



Code –

**bmi<-function(mass,height){**

**op<-(mass/(height^2))**

**if(op<=15)print("Very severely underweight")**

**else if(op>15 && op<=16)print("severely underweight")**

**else if(op>16 && op<=18.5)print("underweight")**

**else if(op>18.5 && op<=25)print("Normal (healthy weight)")**

**else if(op>25 && op<=30)print("Overweight")**

**else if(op>30 && op<=35)print("Obese Class I (Moderately obese)")**

**else if(op>35 && op<=40)print("Obese Class 2 (severely obese)")**

**else if(op>40)print("Obese Class 3 (very severely obese)")**

**}**

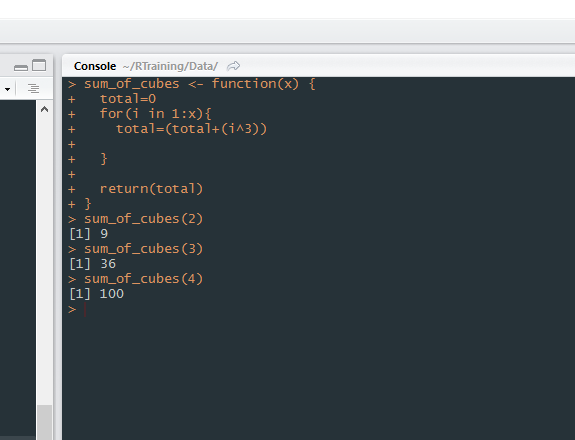
**bmi(74,2)**

**bmi(80,2)**

**bmi(56,1.86)**

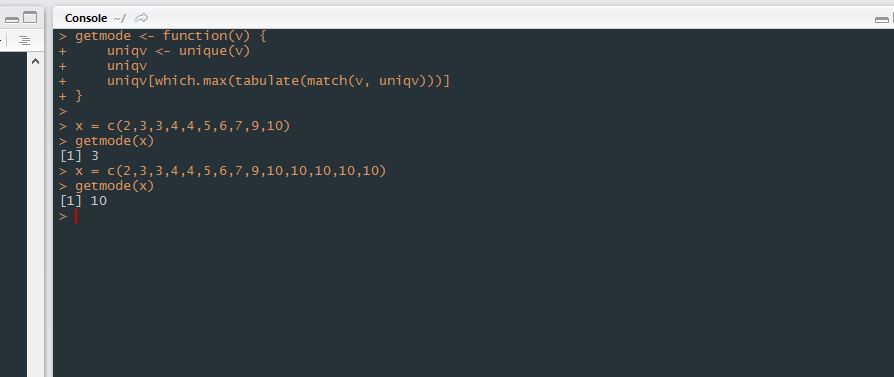
**Q 4 -Write a function called sum\_of\_cubes, that calculates the sum of cubes of the first n**

**natural numbers :**

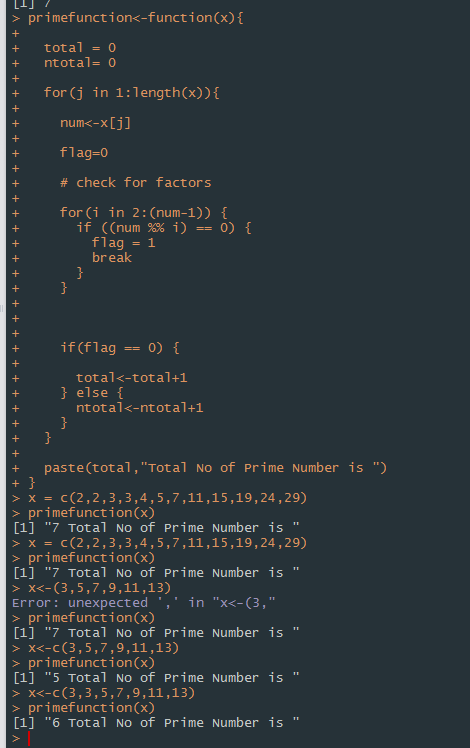


**Q 5- Write a function to calculate the mode (highest frequency) of the following vector:**

**x = c(2,3,3,4,4,5,6,7,9,10)**



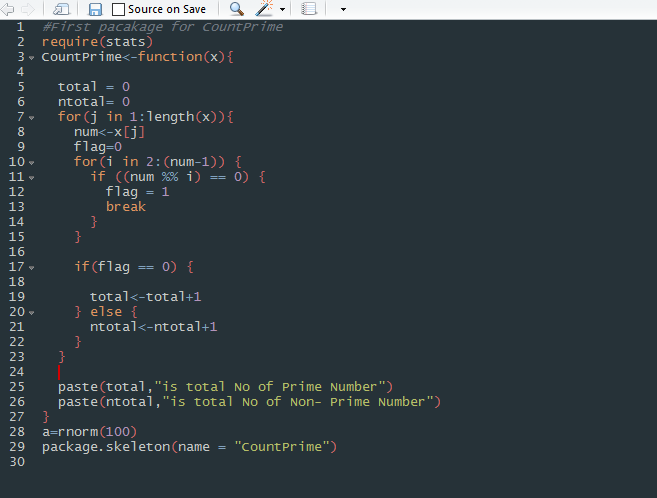
**Q 6 -Write a function to calculate the no. of prime numbers of the following vector :**

**x = c(2,2,3,3,4,5,7,11,15,19,24,29)**

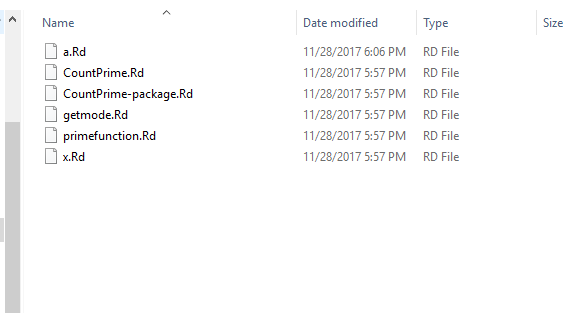
**Q7 - Create a R package for calculating the count of prime numbers , name it as**

**“CountPrime”**

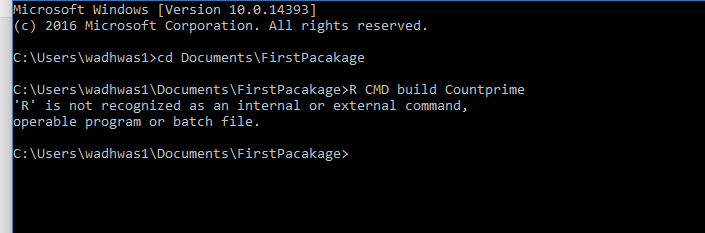
**Creating package :**



**Files for Package :**



**Error While creating package as ZIP File**

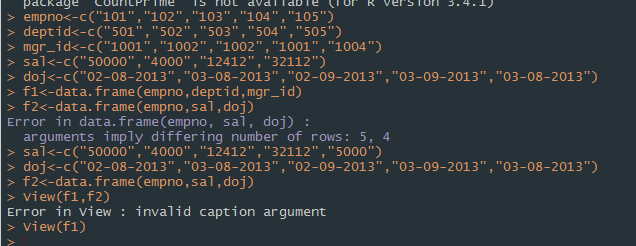


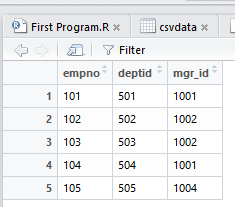
**So unable to get it install to R Studio…Please explain as well why this command not working as I have followed the all steps**

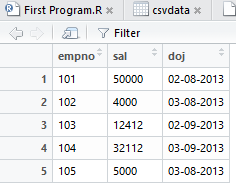
**Q8 - Perform below operations using Data.frame and Data.table**

**a. Load 2 files (.csv) and show it on screen ( F1 – empno, deptid,mgr\_id , F2 – empno,**

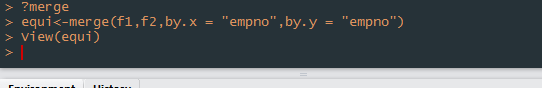
**sal, DOJ)**



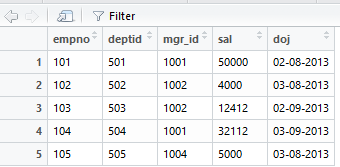
**F1 Table –**

**F2 Table : -**

**B: Perform equi join**



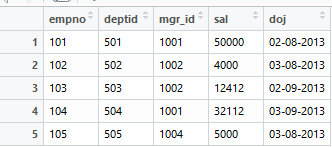
**Output:**



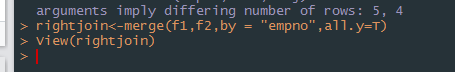
**C: Perform left outer join:**

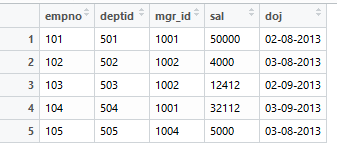


**Output :**

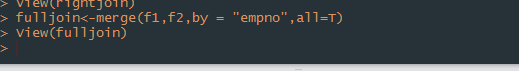


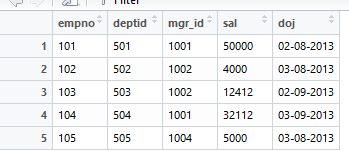
**D: Right Join**



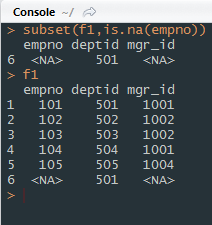


**E: Perform full outer join**

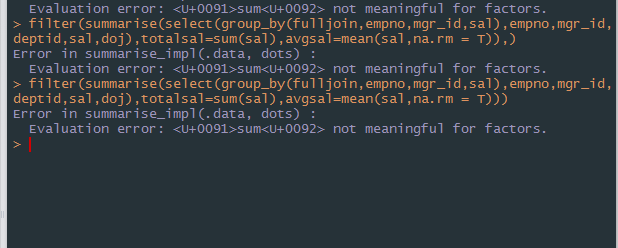




**F: Perform filter operation –Eg find all the rows for which col1 is null**

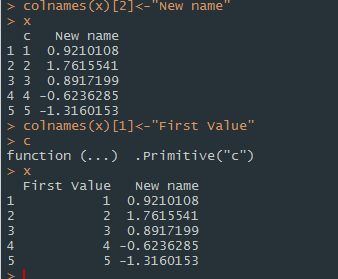


**G :Perform group by , sum, average operation**

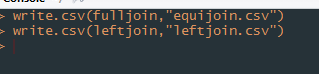


**Getting an error,tried multiple time,,please explain this one more time**

**Create a derived column (empname in F2) – do some data transformation on that**



**Write output for each in separate files (eg- equijoin.csv ,leftoutjoin.csv)**



9. **Create R functions for the following operations**

1. library(dplyr)

d1<- f1 %>% group\_by(deptid)%>%

select count(distinct mgr\_id)

1. library(dplyr)

d1<- f1 %>% group\_by(deptid,mgr\_id)%>%

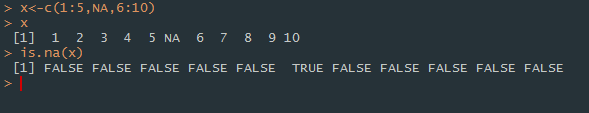
select count(distinct empno)

library(dplyr)

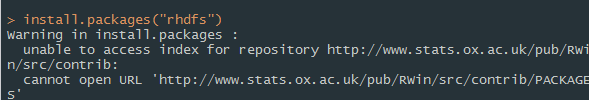
1. d2<- f1[duplicated(f1$empno)

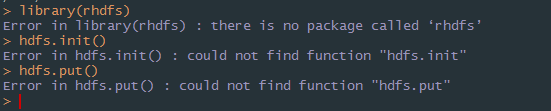
10. **Create R functions for the following operations**

1. Find out if there are any nulls in a dataset or in some specific number of columns

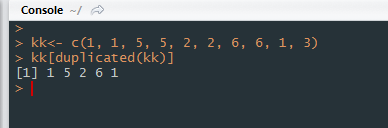


1. Write a function to read data from hdfs and dump it back to hdfs ?

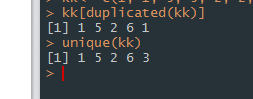




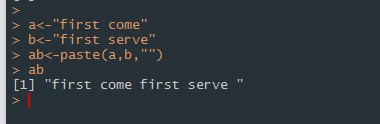
11. **.) Remove duplicates from a given vector and return it back.**



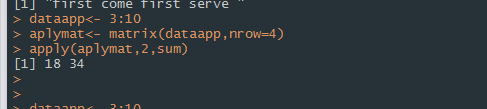
1. Compute count of distinct



1. Concatenate two strings.



1. Perform Column-wise/Row-wise sum using apply function.



1. Get list of files in an hdfs path.

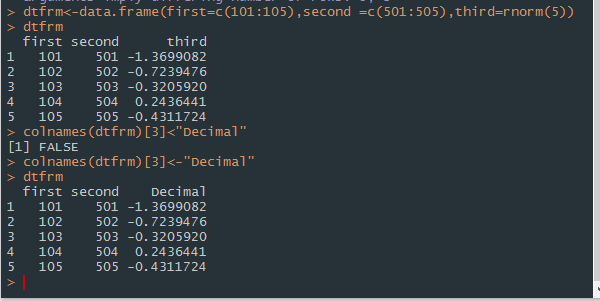
Ans= hdfs.ls(‘/’)

1. Delete a file from hdfs if it exists.

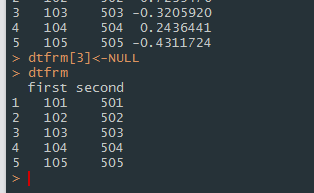
Ans= hdfs.delete(“/RHadoop”)

12

Rename column names in a dataframe.



**Drop given column from a data frame**

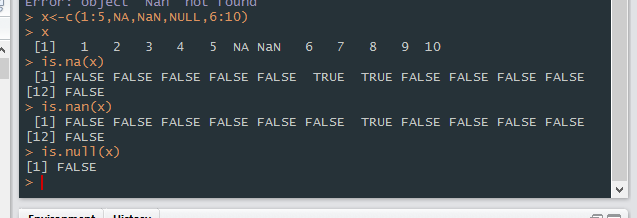


**Illustrate the difference between NA, NULL, NaN.**

**NaN : means 0/0 -- Stands for Not a Number**

**NA : is generally interpreted as a missing, does not exist**

**NULL : is for empty object.**



**Most importantly most of the question seems to be from Hadoop so it does not really make sense or either those topics has not been cover by Mento**