VIT-AP UNIVERSITY, ANDHRA PRADESH

CSE4027– Data Analytics - Lab Sheet :6

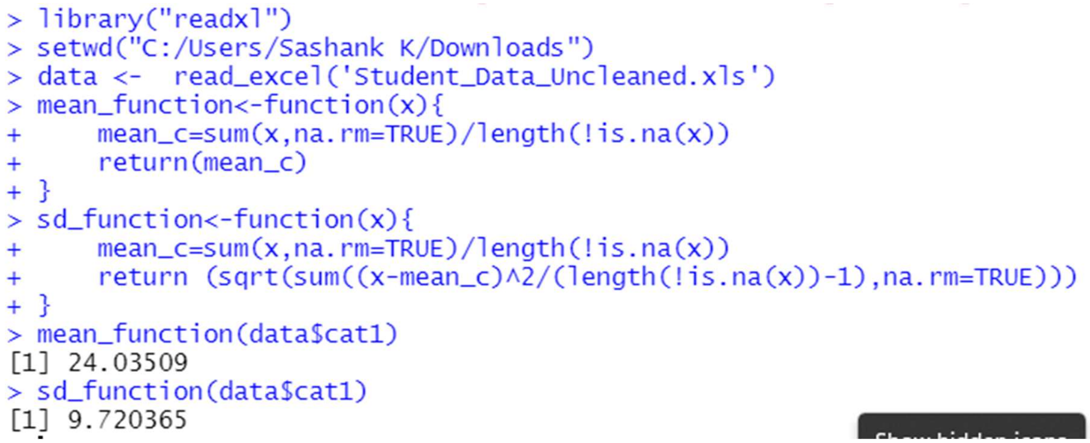
Academic year: 2022-2023	**Branch/ Class:** B.Tech
Semester: Fall	**Date:** 14-10-2022
Faculty Name: Dr Syed Khasim	**School:** SCOPE
Student name: MAJJIGA JASWANTH	**Reg. no.:20BCD7171**

---

1. Create a student result dataset with numeric values.
   a. write a function for calculating the mean.
   b. Write a function to compute std.deviation.

Code:

```
> library("readxl")
> setwd("C:/Users/Sashank K/Downloads")
> data <-  read_excel('Student_Data_Uncleaned.xls')
> mean_function<-function(x){
+     mean_c=sum(x,na.rm=TRUE)/length(!is.na(x))
+     return(mean_c)
+ }
> sd_function<-function(x){
+     mean_c=sum(x,na.rm=TRUE)/length(!is.na(x))
+     return (sqrt(sum((x-mean_c)^2/(length(!is.na(x))-1),na.rm=TRUE)))
+ }
```



```
> library("readxl")
> setwd("C:/Users/Sashank K/Downloads")
> data <-  read_excel('Student_Data_Uncleaned.xls')
> mean_function<-function(x){
+     mean_c=sum(x,na.rm=TRUE)/length(!is.na(x))
+     return(mean_c)
+ }
> sd_function<-function(x){
+     mean_c=sum(x,na.rm=TRUE)/length(!is.na(x))
+     return (sqrt(sum((x-mean_c)^2/(length(!is.na(x))-1),na.rm=TRUE)))
+ }
> mean_function(data$cat1)
[1] 24.03509
> sd_function(data$cat1)
[1] 9.720365
```

2. UseCovid.csv and weather.csv. Do all observations (min, max, mean, variance, SD, range) in both dataframe.

   Code:

```
library("readxl")

data_1 <-  read_excel('COVID_country_wise_latest.xls')

data_2 <-  read_excel('weatherHistory.xls')

cat("Covid\n\n")

cat("Mean Values\n")

colMeans(data_1[sapply(data_1, is.numeric)])

cat("\nMinimum Values\n")

apply(data_1[sapply(data_1, is.numeric)],2,min)

cat("\nMaximum Values\n")

apply(data_1[sapply(data_1, is.numeric)],2,max)

cat("\nVariance\n")

sapply(data_1[sapply(data_1, is.numeric)],var)

cat("\nStandard Deviation\n")

sapply(data_1[sapply(data_1, is.numeric)],sd)

cat("\nRange\n")

sapply(data_1[sapply(data_1, is.numeric)],range)

cat("\nWeather\n\n")

cat("Mean Values\n")

colMeans(data_2[sapply(data_2, is.numeric)])

cat("\nMinimum Values\n")

apply(data_2[sapply(data_2, is.numeric)],2,min)

cat("\nMaximum Values\n")

apply(data_2[sapply(data_2, is.numeric)],2,max)
```

cat("\nVariance\n")

sapply(data_2[sapply(data_2, is.numeric)],var)

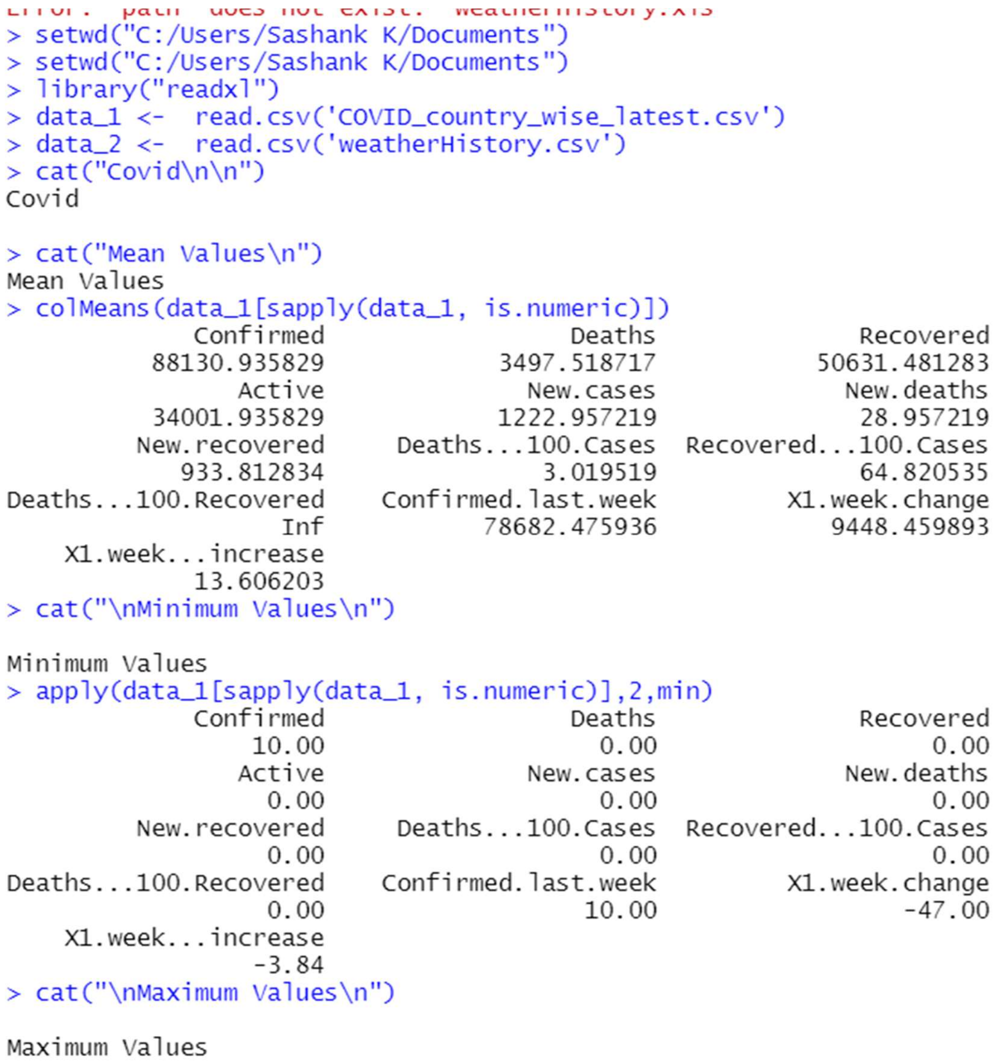
cat("\nStandard Deviation\n")

sapply(data_2[sapply(data_2, is.numeric)],sd)

cat("\nRange\n")

sapply(data_2[sapply(data_2, is.numeric)],range)

Output:

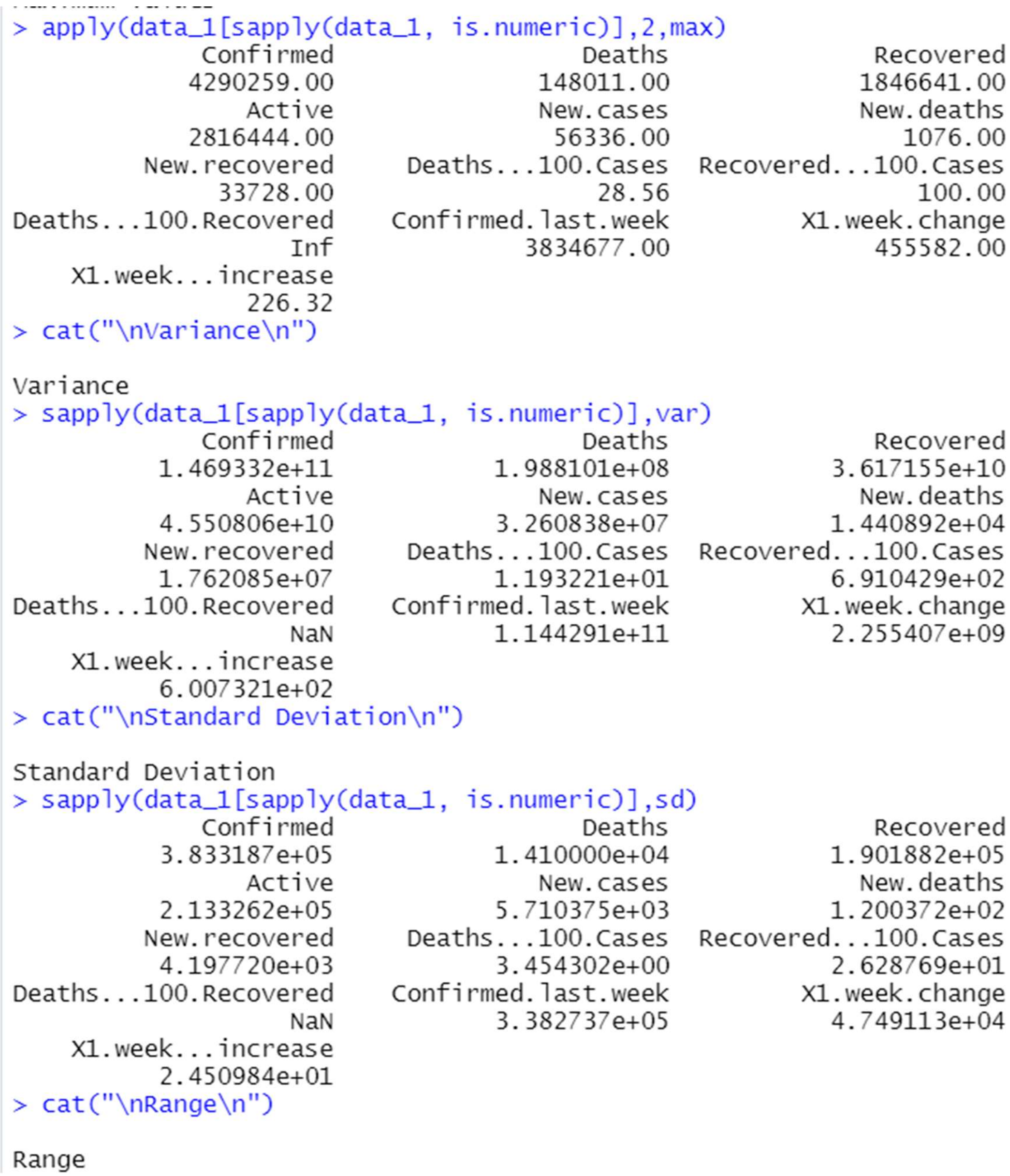


```
Error: path does not exist: 'weatherHistory.xls'
> setwd("C:/Users/Sashank K/Documents")
> setwd("C:/Users/Sashank K/Documents")
> library("readxl")
> data_1 <-  read.csv('COVID_country_wise_latest.csv')
> data_2 <-  read.csv('weatherHistory.csv')
> cat("Covid\n\n")
Covid

> cat("Mean Values\n")
Mean Values
> colMeans(data_1[sapply(data_1, is.numeric)])
             Confirmed                 Deaths              Recovered
          88130.935829            3497.518717           50631.481283
                Active              New.cases             New.deaths
          34001.935829            1222.957219              28.957219
         New.recovered     Deaths...100.Cases  Recovered...100.Cases
            933.812834               3.019519              64.820535
Deaths...100.Recovered    Confirmed.last.week         X1.week.change
                   Inf           78682.475936            9448.459893
    X1.week...increase
             13.606203
> cat("\nMinimum Values\n")

Minimum Values
> apply(data_1[sapply(data_1, is.numeric)],2,min)
             Confirmed                 Deaths              Recovered
                 10.00                   0.00                   0.00
                Active              New.cases             New.deaths
                  0.00                   0.00                   0.00
         New.recovered     Deaths...100.Cases  Recovered...100.Cases
                  0.00                   0.00                   0.00
Deaths...100.Recovered    Confirmed.last.week         X1.week.change
                  0.00                  10.00                 -47.00
    X1.week...increase
                 -3.84
> cat("\nMaximum Values\n")

Maximum Values
```

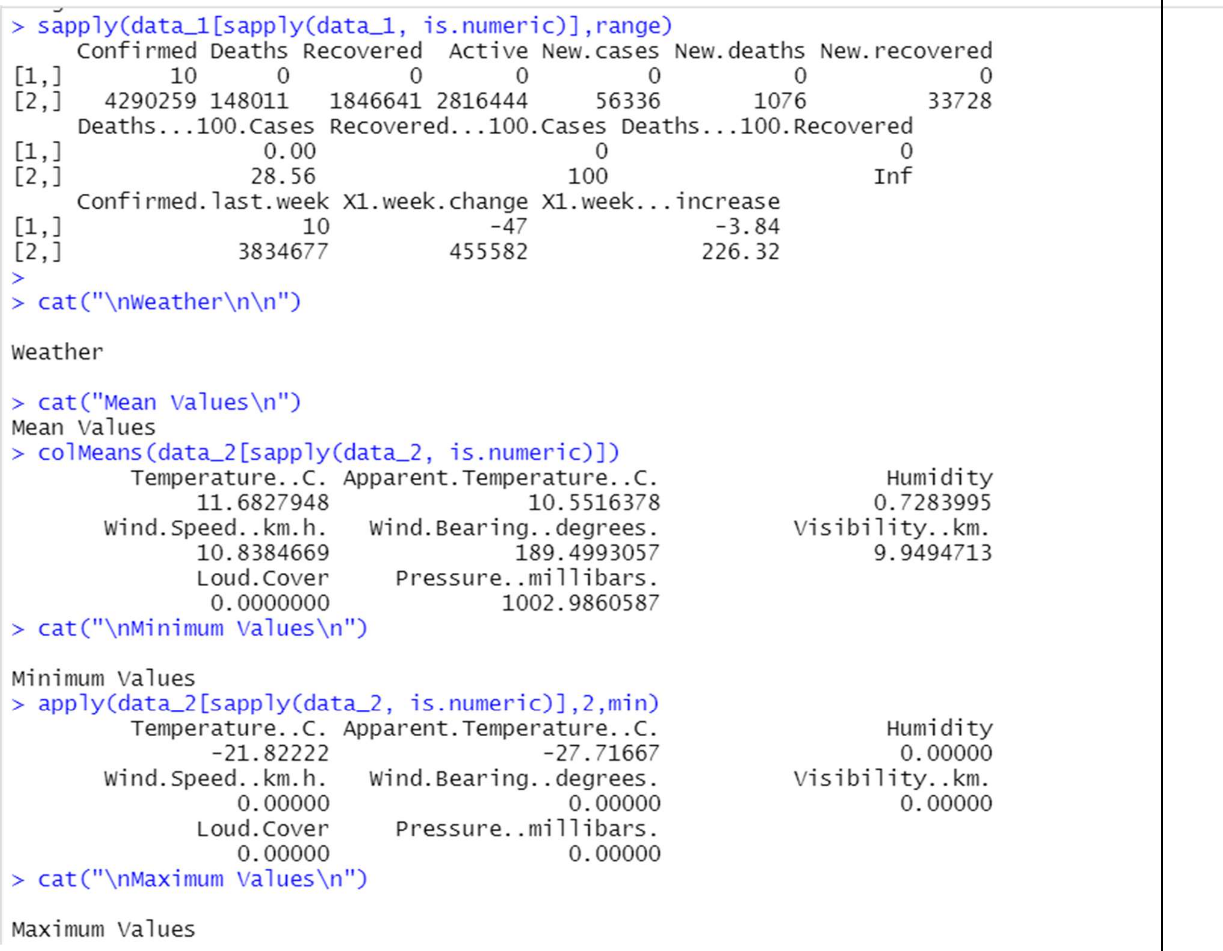


```
> apply(data_1[sapply(data_1, is.numeric)],2,max)
             Confirmed                 Deaths              Recovered
            4290259.00              148011.00             1846641.00
                Active              New.cases             New.deaths
            2816444.00               56336.00                1076.00
         New.recovered     Deaths...100.Cases  Recovered...100.Cases
              33728.00                  28.56                 100.00
Deaths...100.Recovered    Confirmed.last.week         X1.week.change
                   Inf             3834677.00              455582.00
    X1.week...increase
                226.32
> cat("\nVariance\n")

Variance
> sapply(data_1[sapply(data_1, is.numeric)],var)
             Confirmed                 Deaths              Recovered
          1.469332e+11           1.988101e+08           3.617155e+10
                Active              New.cases             New.deaths
          4.550806e+10           3.260838e+07           1.440892e+04
         New.recovered     Deaths...100.Cases  Recovered...100.Cases
          1.762085e+07           1.193221e+01           6.910429e+02
Deaths...100.Recovered    Confirmed.last.week         X1.week.change
                   NaN           1.144291e+11           2.255407e+09
    X1.week...increase
          6.007321e+02
> cat("\nStandard Deviation\n")

Standard Deviation
> sapply(data_1[sapply(data_1, is.numeric)],sd)
             Confirmed                 Deaths              Recovered
          3.833187e+05           1.410000e+04           1.901882e+05
                Active              New.cases             New.deaths
          2.133262e+05           5.710375e+03           1.200372e+02
         New.recovered     Deaths...100.Cases  Recovered...100.Cases
          4.197720e+03           3.454302e+00           2.628769e+01
Deaths...100.Recovered    Confirmed.last.week         X1.week.change
                   NaN           3.382737e+05           4.749113e+04
    X1.week...increase
          2.450984e+01
> cat("\nRange\n")

Range
```



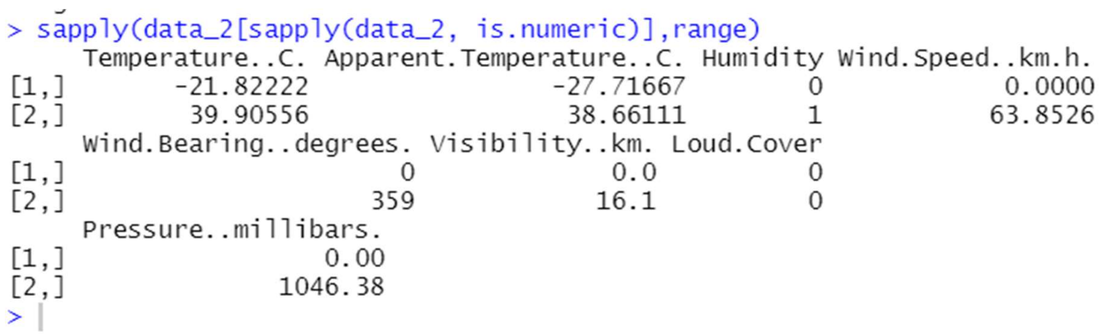
```
> sapply(data_1[sapply(data_1, is.numeric)],range)
     Confirmed Deaths Recovered  Active New.cases New.deaths New.recovered
[1,]        10      0         0       0         0          0             0
[2,]   4290259 148011   1846641 2816444     56336       1076         33728
     Deaths...100.Cases Recovered...100.Cases Deaths...100.Recovered
[1,]               0.00                     0                      0
[2,]              28.56                   100                    Inf
     Confirmed.last.week X1.week.change X1.week...increase
[1,]                  10            -47              -3.84
[2,]             3834677         455582             226.32
> 
> cat("\nWeather\n\n")

Weather

> cat("Mean Values\n")
Mean Values
> colMeans(data_2[sapply(data_2, is.numeric)])
         Temperature..C. Apparent.Temperature..C.                 Humidity 
              11.6827948               10.5516378                0.7283995 
      Wind.Speed..km.h.   Wind.Bearing..degrees.          Visibility..km. 
              10.8384669              189.4993057                9.9494713 
              Loud.Cover     Pressure..millibars. 
               0.0000000             1002.9860587 
> cat("\nMinimum Values\n")

Minimum Values
> apply(data_2[sapply(data_2, is.numeric)],2,min)
         Temperature..C. Apparent.Temperature..C.                 Humidity 
               -21.82222                -27.71667                  0.00000 
      Wind.Speed..km.h.   Wind.Bearing..degrees.          Visibility..km. 
                 0.00000                  0.00000                  0.00000 
              Loud.Cover     Pressure..millibars. 
                 0.00000                  0.00000 
> cat("\nMaximum Values\n")

Maximum Values
```



```
> sapply(data_2[sapply(data_2, is.numeric)],range)
     Temperature..C. Apparent.Temperature..C. Humidity Wind.Speed..km.h.
[1,]       -21.82222                -27.71667        0            0.0000
[2,]        39.90556                 38.66111        1           63.8526
     Wind.Bearing..degrees. Visibility..km. Loud.Cover
[1,]                      0             0.0          0
[2,]                    359            16.1          0
     Pressure..millibars.
[1,]                 0.00
[2,]              1046.38
> 
```

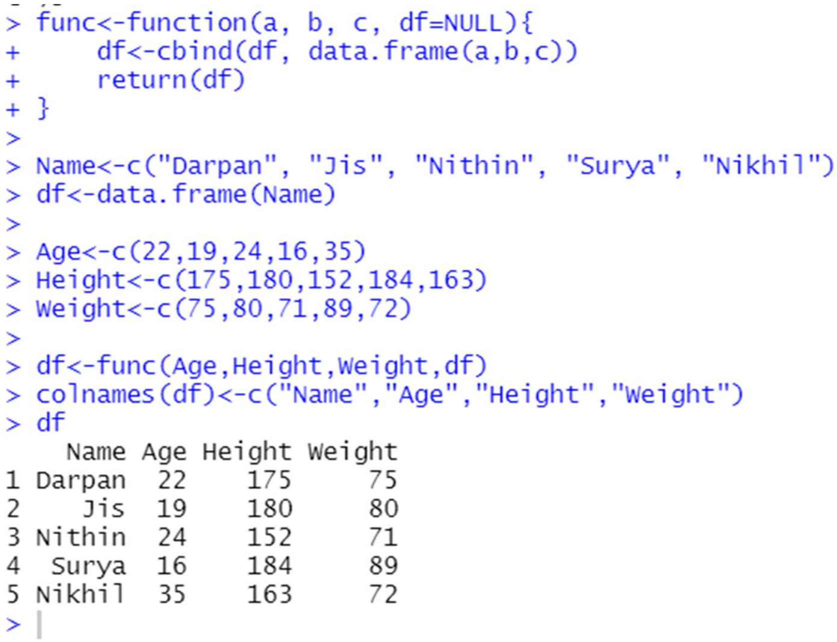
3. Write a function that has three vector arguments for merging the into an existing dataframe.

   Code:

```
 > func<-function(a, b, c, df=NULL){
+     df<-cbind(df, data.frame(a,b,c))
+     return(df)
+ }
>
```

```
> Name<-c("Darpan", "Jis", "Nithin", "Surya", "Nikhil")
> df<-data.frame(Name)
>
> Age<-c(22,19,24,16,35)
> Height<-c(175,180,152,184,163)
> Weight<-c(75,80,71,89,72)
>
> df<-func(Age,Height,Weight,df)
> colnames(df)<-c("Name","Age","Height","Weight")
> df
```

Output:



```
> func<-function(a, b, c, df=NULL){
+     df<-cbind(df, data.frame(a,b,c))
+     return(df)
+ }
>
> Name<-c("Darpan", "Jis", "Nithin", "Surya", "Nikhil")
> df<-data.frame(Name)
>
> Age<-c(22,19,24,16,35)
> Height<-c(175,180,152,184,163)
> Weight<-c(75,80,71,89,72)
>
> df<-func(Age,Height,Weight,df)
> colnames(df)<-c("Name","Age","Height","Weight")
> df
    Name Age Height Weight
1 Darpan  22    175     75
2    Jis  19    180     80
3 Nithin  24    152     71
4  Surya  16    184     89
5 Nikhil  35    163     72
>
```

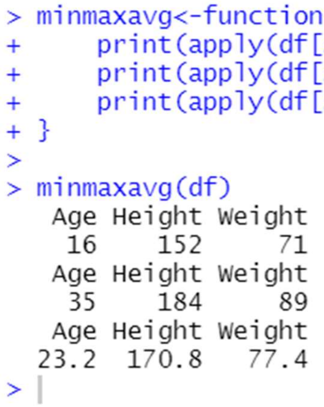
4. After merging create a function compute to find out min,max and avg of all numeric columns.

Code:

```
minmaxavg<-function(df){
  print(apply(df[sapply(df,is.numeric)],2,min))
  print(apply(df[sapply(df,is.numeric)],2,max))
  print(apply(df[sapply(df,is.numeric)],2,mean))
}

minmaxavg(df)
```

Output:



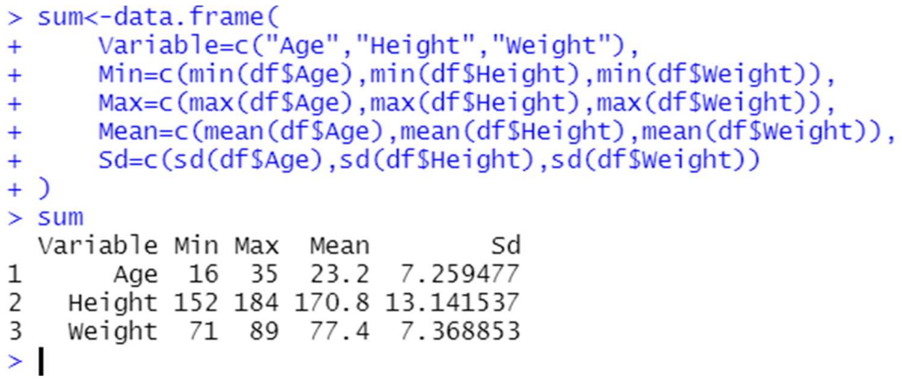
```
> minmaxavg<-function(df){
+     print(apply(df[sapply(df,is.numeric)],2,min))
+     print(apply(df[sapply(df,is.numeric)],2,max))
+     print(apply(df[sapply(df,is.numeric)],2,mean))
+ }
>
> minmaxavg(df)
   Age Height Weight
    16    152     71
   Age Height Weight
    35    184     89
   Age Height Weight
  23.2  170.8   77.4
> |
```

5. The summary values should be in a single data frame with the following columns: variable name, mean, sd, minimum, and maximum.

   Code:

   ```
   > sum<-data.frame(
   +     Variable=c("Age","Height","Weight"),
   +     Min=c(min(df$Age),min(df$Height),min(df$Weight)),
   +     Max=c(max(df$Age),max(df$Height),max(df$Weight)),
   +     Mean=c(mean(df$Age),mean(df$Height),mean(df$Weight)),
   +     Sd=c(sd(df$Age),sd(df$Height),sd(df$Weight))
   + )
   > sum
   ```

   output:



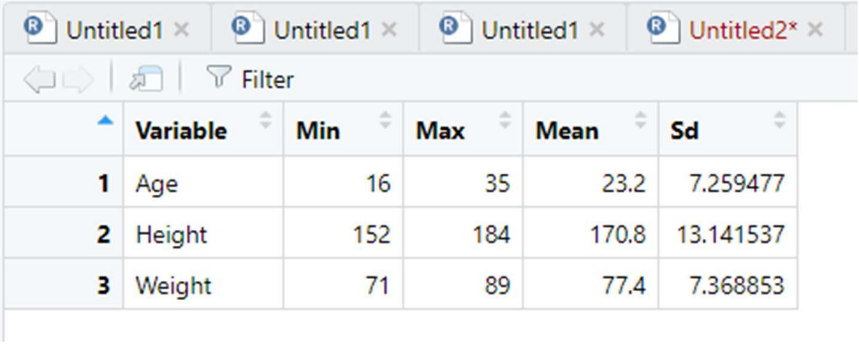
```
> sum<-data.frame(
+     Variable=c("Age","Height","Weight"),
+     Min=c(min(df$Age),min(df$Height),min(df$Weight)),
+     Max=c(max(df$Age),max(df$Height),max(df$Weight)),
+     Mean=c(mean(df$Age),mean(df$Height),mean(df$Weight)),
+     Sd=c(sd(df$Age),sd(df$Height),sd(df$Weight))
+ )
> sum
  Variable Min Max  Mean        Sd
1      Age  16  35  23.2  7.259477
2   Height 152 184 170.8 13.141537
3   Weight  71  89  77.4  7.368853
> |
```

6. Write a function so that the summary of the dataframe should be written to a csv file and to R.

   Code:

```
> write.csv(summ,"Summary.csv",row.names=FALSE)
> data_3<-read.csv("Summary.csv")
> View(data_3)
```

   Output:



	Variable	Min	Max	Mean	Sd
1	Age	16	35	23.2	7.259477
2	Height	152	184	170.8	13.141537
3	Weight	71	89	77.4	7.368853