Experiment1.2

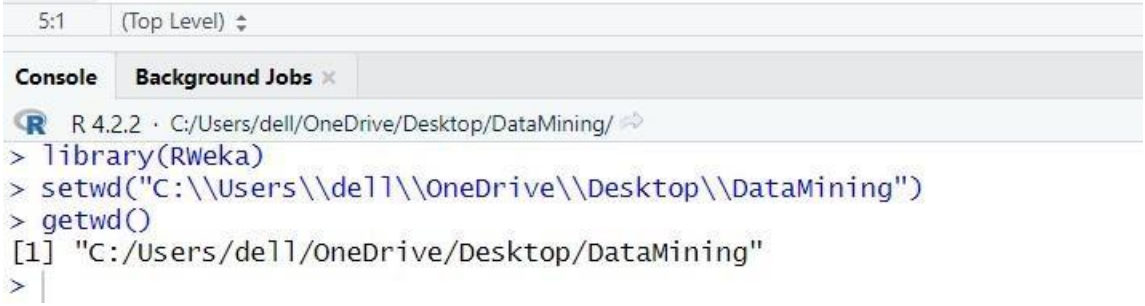
Student Name: Arif khan **UID: 20BCS4947**
Branch: BE CSE **Section/Group: 20BCS_DM_607_B**
Semester: 6TH **Subject Name: DM LAB**

1. **Aim: Statistical analysis of data.**

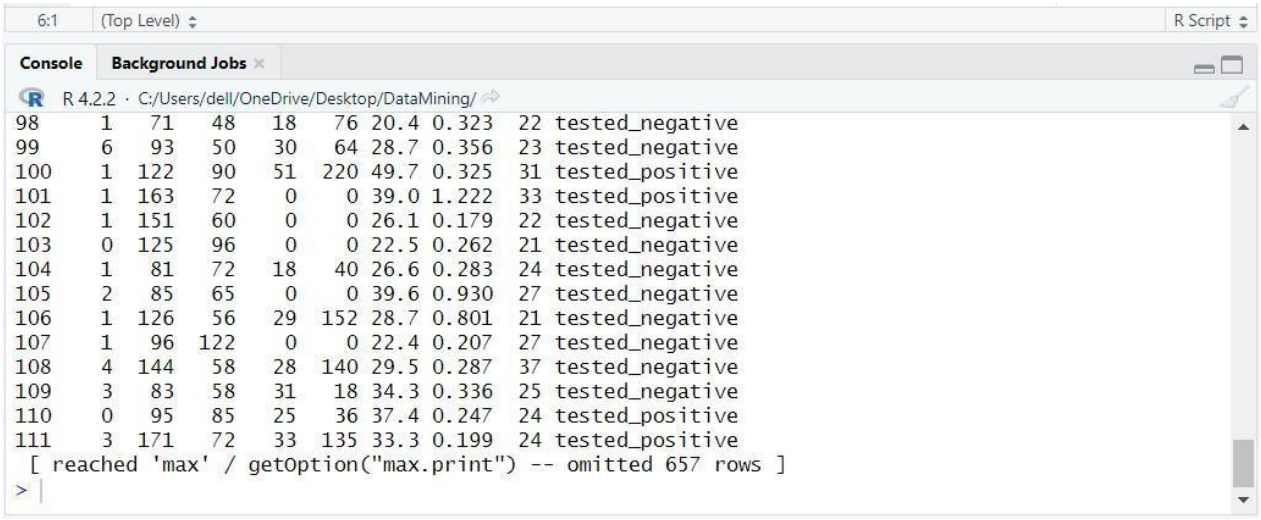
2. **Objective: To perform statistical analysis of data.**

3. **Script and Output:**

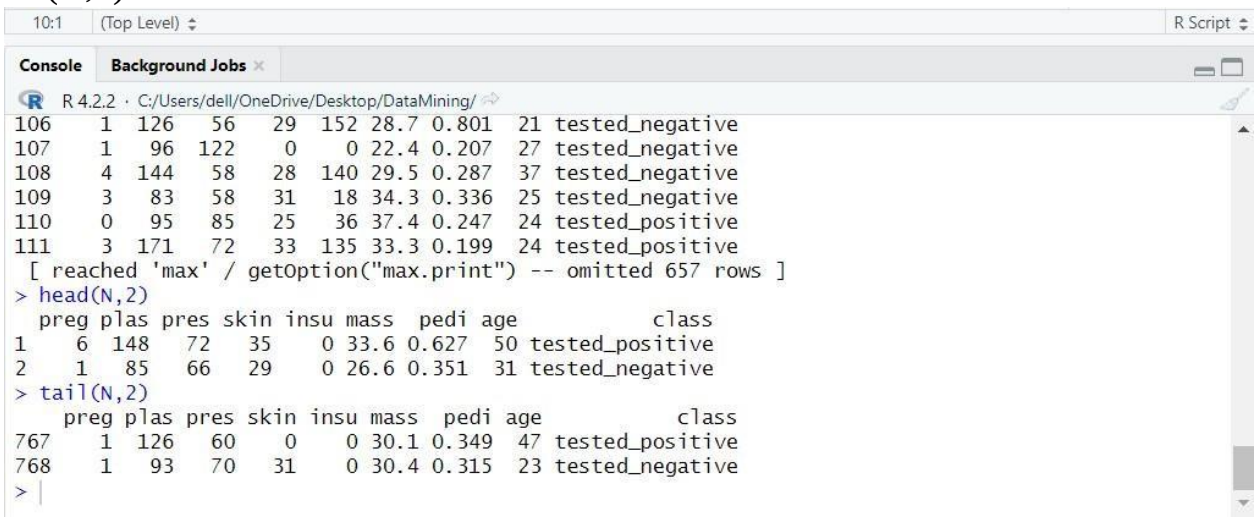
```
library(RWeka) #setting
the working directory
setwd("C:\\Users\\dell\\OneDrive\\Desktop\\DataMining")
#checking the working
directory getwd()
```




```
#reading arff file
N = read.arff("diabetes
.arff") print(N)
```



#printing first and
last 2 rows head(N,2)
tail(N,2)



#printing
dimension dim(N)

```
14:1  (Top Level)                                                    R Script
Console  Background Jobs
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
108    4  144   58   28  140 29.5 0.287  37 tested_negative
109    3   83   58   31   18 34.3 0.336  25 tested_negative
110    0   95   85   25   36 37.4 0.247  24 tested_positive
111    3  171   72   33  135 33.3 0.199  24 tested_positive
 [ reached 'max' / getOption("max.print") -- omitted 657 rows ]
> head(N,2)
  preg plas pres skin insu mass  pedi age           class
1    6  148   72   35    0 33.6 0.627  50 tested_positive
2    1   85   66   29    0 26.6 0.351  31 tested_negative
> tail(N,2)
    preg plas pres skin insu mass  pedi age           class
767    1  126   60    0    0 30.1 0.349  47 tested_positive
768    1   93   70   31    0 30.4 0.315  23 tested_negative
> dim(N)
[1] 768   9
>
```

#printing column
names names(N)

```
17:1  (Top Level)                                                    R Script
Console  Background Jobs
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
110    0   95   85   25   36 37.4 0.247  24 tested_positive
111    3  171   72   33  135 33.3 0.199  24 tested_positive
 [ reached 'max' / getOption("max.print") -- omitted 657 rows ]
> head(N,2)
  preg plas pres skin insu mass  pedi age           class
1    6  148   72   35    0 33.6 0.627  50 tested_positive
2    1   85   66   29    0 26.6 0.351  31 tested_negative
> tail(N,2)
    preg plas pres skin insu mass  pedi age           class
767    1  126   60    0    0 30.1 0.349  47 tested_positive
768    1   93   70   31    0 30.4 0.315  23 tested_negative
> dim(N)
[1] 768   9
> names(N)
[1] "preg"  "plas"  "pres"  "skin"  "insu"  "mass"  "pedi"  "age"   "class"
>
```

#printing maximum and minimum
age
max(N["age"])
min(N["age"])

```
24:10  (Top Level)                                                    R Script
Console  Background Jobs
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
  preg plas pres skin insu mass  pedi age           class
1    6  148   72   35    0 33.6 0.627  50 tested_positive
2    1   85   66   29    0 26.6 0.351  31 tested_negative
> tail(N,2)
    preg plas pres skin insu mass  pedi age           class
767    1  126   60    0    0 30.1 0.349  47 tested_positive
768    1   93   70   31    0 30.4 0.315  23 tested_negative
> dim(N)
[1] 768   9
> names(N)
[1] "preg"  "plas"  "pres"  "skin"  "insu"  "mass"  "pedi"  "age"   "class"
> max(N["age"])
[1] 81
> min(N["age"])
[1] 21
>
```

#mean age

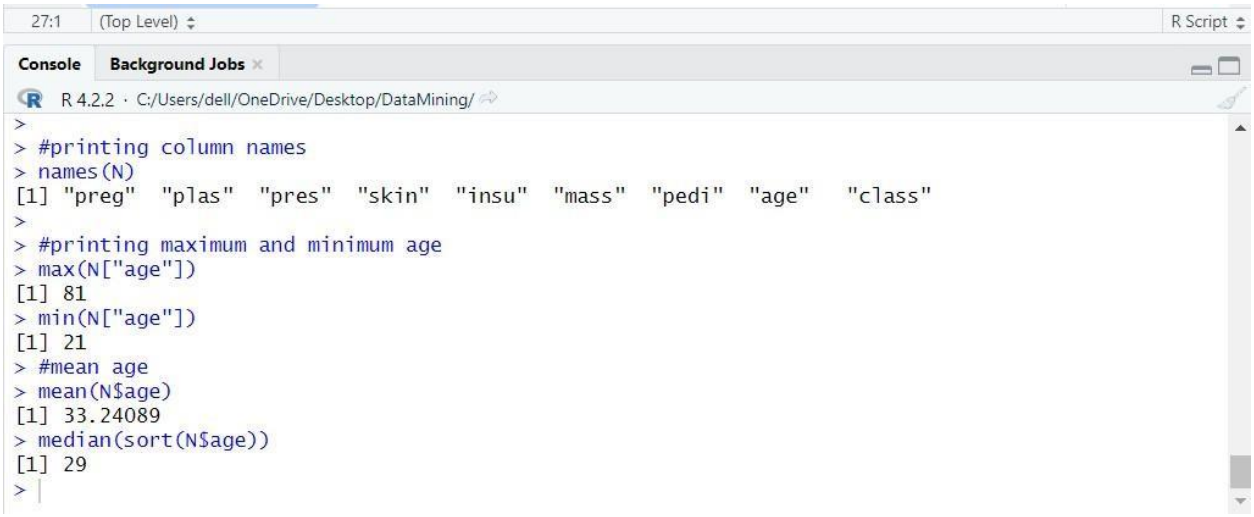
mean(N$age)

```
24:1  (Top Level)                                                     R Script
Console  Background Jobs
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
> dim(N)
[1] 768   9
>
> #printing column names
> names(N)
[1] "preg"  "plas"  "pres"  "skin"  "insu"  "mass"  "pedi"  "age"   "class"
>
> #printing maximum and minimum age
> max(N["age"])
[1] 81
> min(N["age"])
[1] 21
> #mean age
> mean(N$age)
[1] 33.24089
>
```

#median age

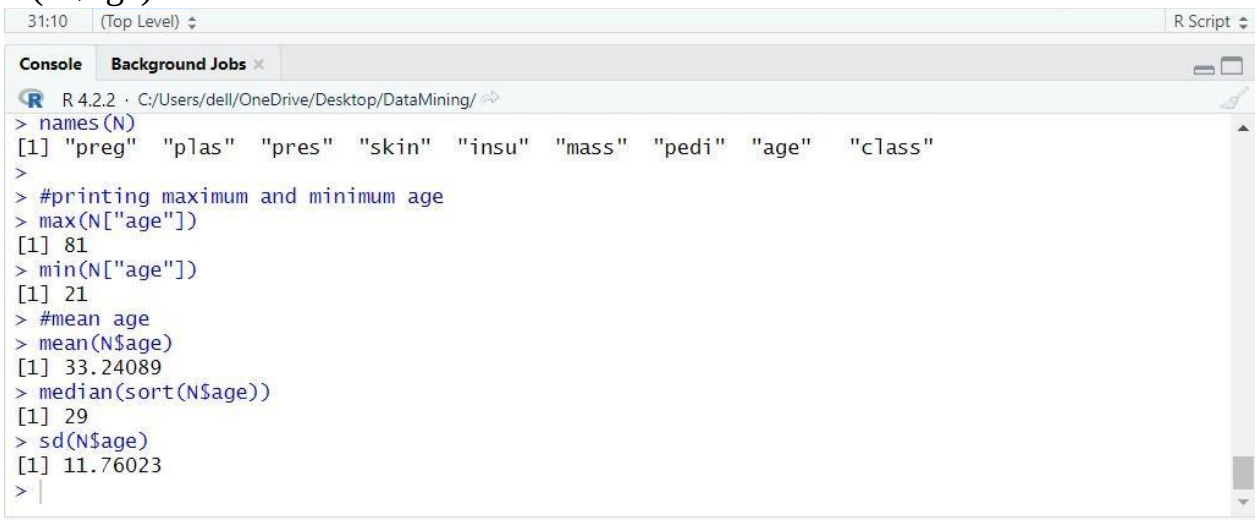
median(sort(N$ag

e))



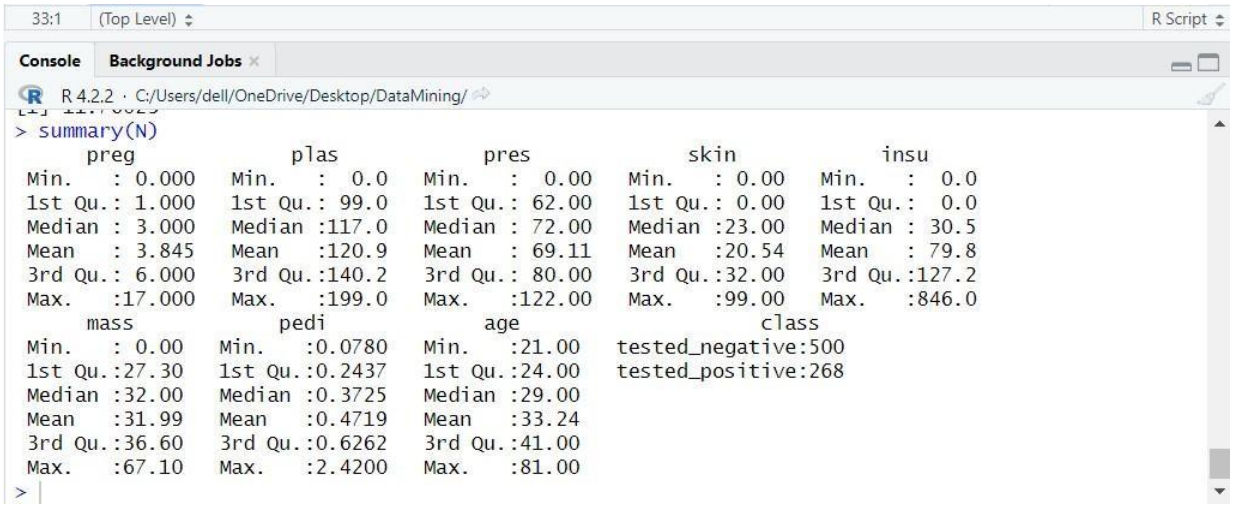
#standard deviation

sd(N$age)



#summary

summary(N)



5. Learning Outcomes:

- Creating data frames.
- Reading an .arff file.
- Worked on large dataset and analyse it.
- Printing dataset and performing statistical analysis .
- Finding mean, median, standard deviation etc of the data frame

1.