Experiment 1.2

Student Name: Mayank Kumar
UID: 20BCS1353
Branch: BE-CSE
Section/Group: 20BCS_DM_705 A
Semester: 6th
Date of Performance:
Subject Name: Data Mining Lab
Subject Code: 20CSP-376

Aim: To perform the statistical analysis of data.

Objective: we will be learning about ARFF files and how to read ARFF file and perform statistical operation.

Script and Output:

read.arff() – this function is used to read the student.arff file from the specified location which contains the data of student.

head (name object, number) – used to print the amount of data from the top.

tail (name object, number) – used to print the amount of data from the bottom

dim () – used to find the dimension of the data frame

names () – names of column

max() – maximum value in data

min() – minimum value in data

sum() – sum of all the value

mean() – mean of the data

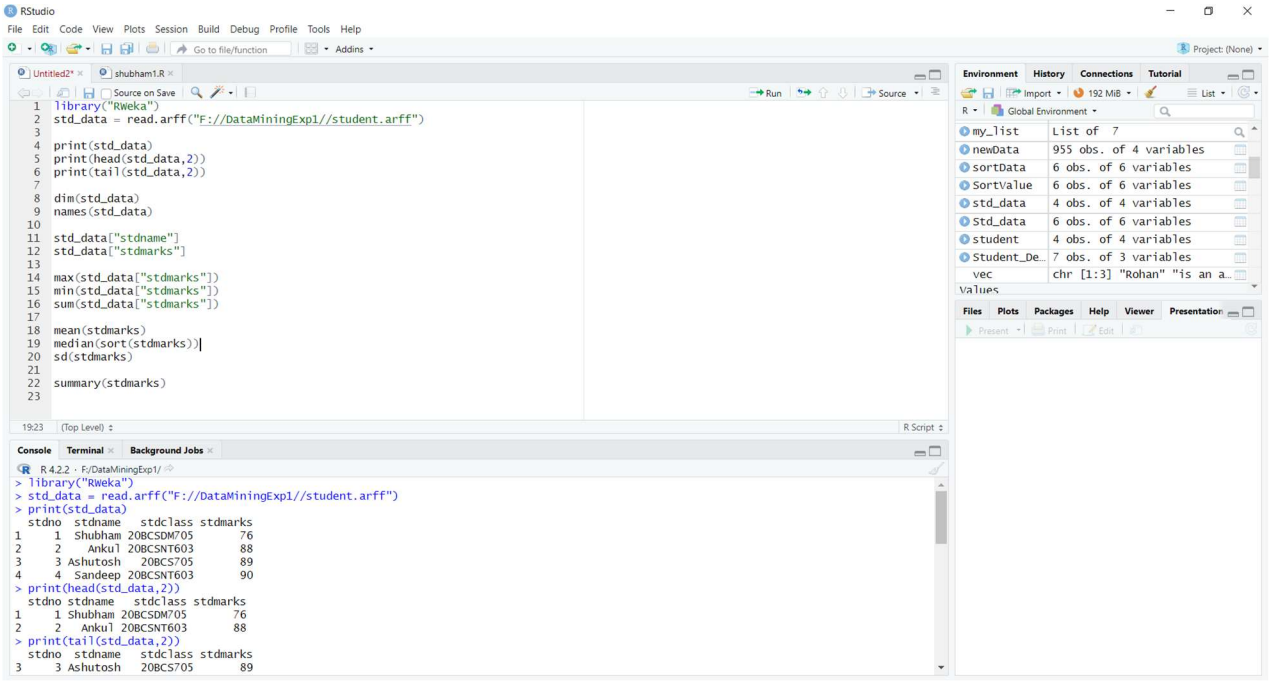
median() – median of the data

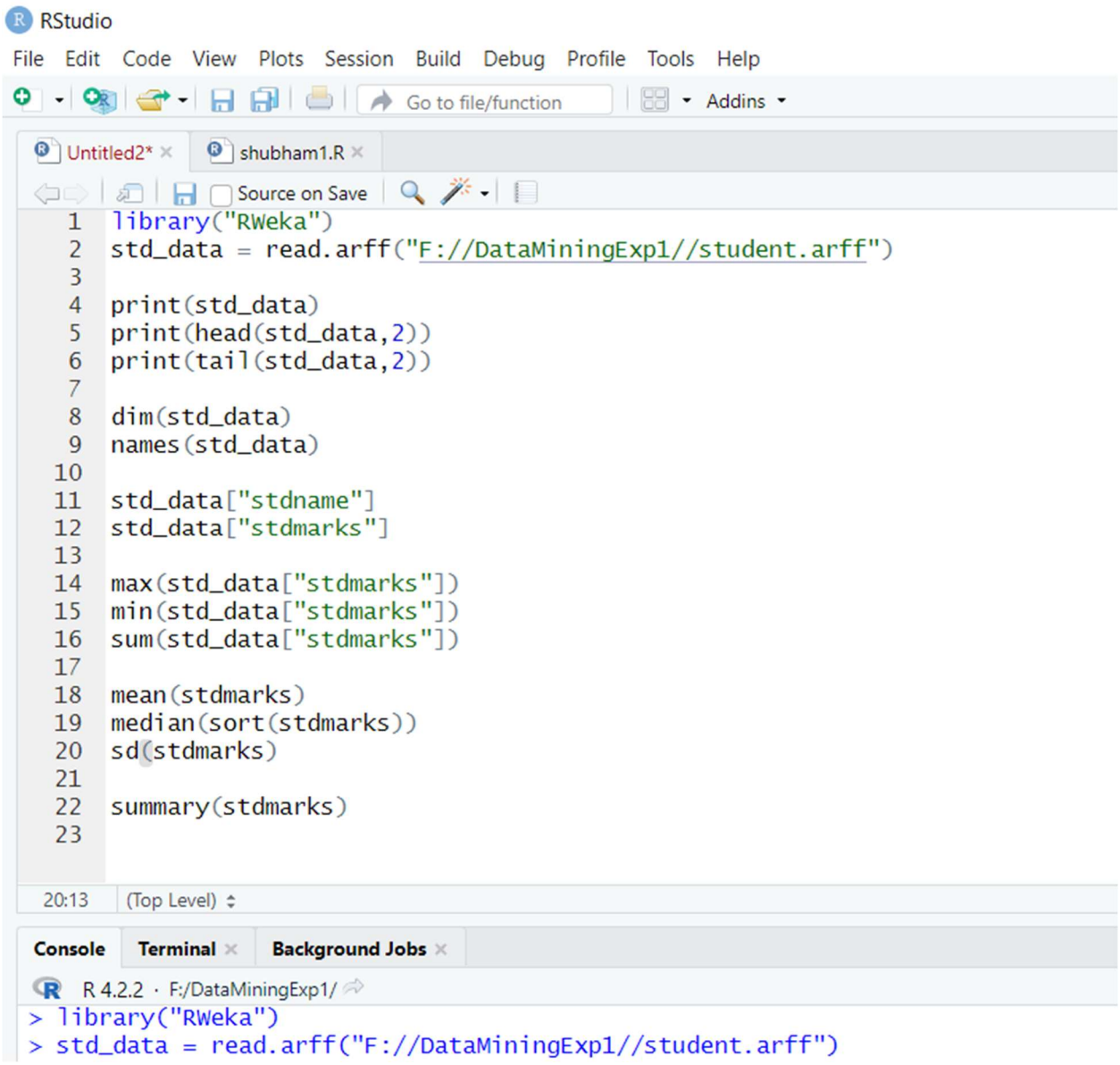
sd() - standard deviation of data

summary() – summary of the data

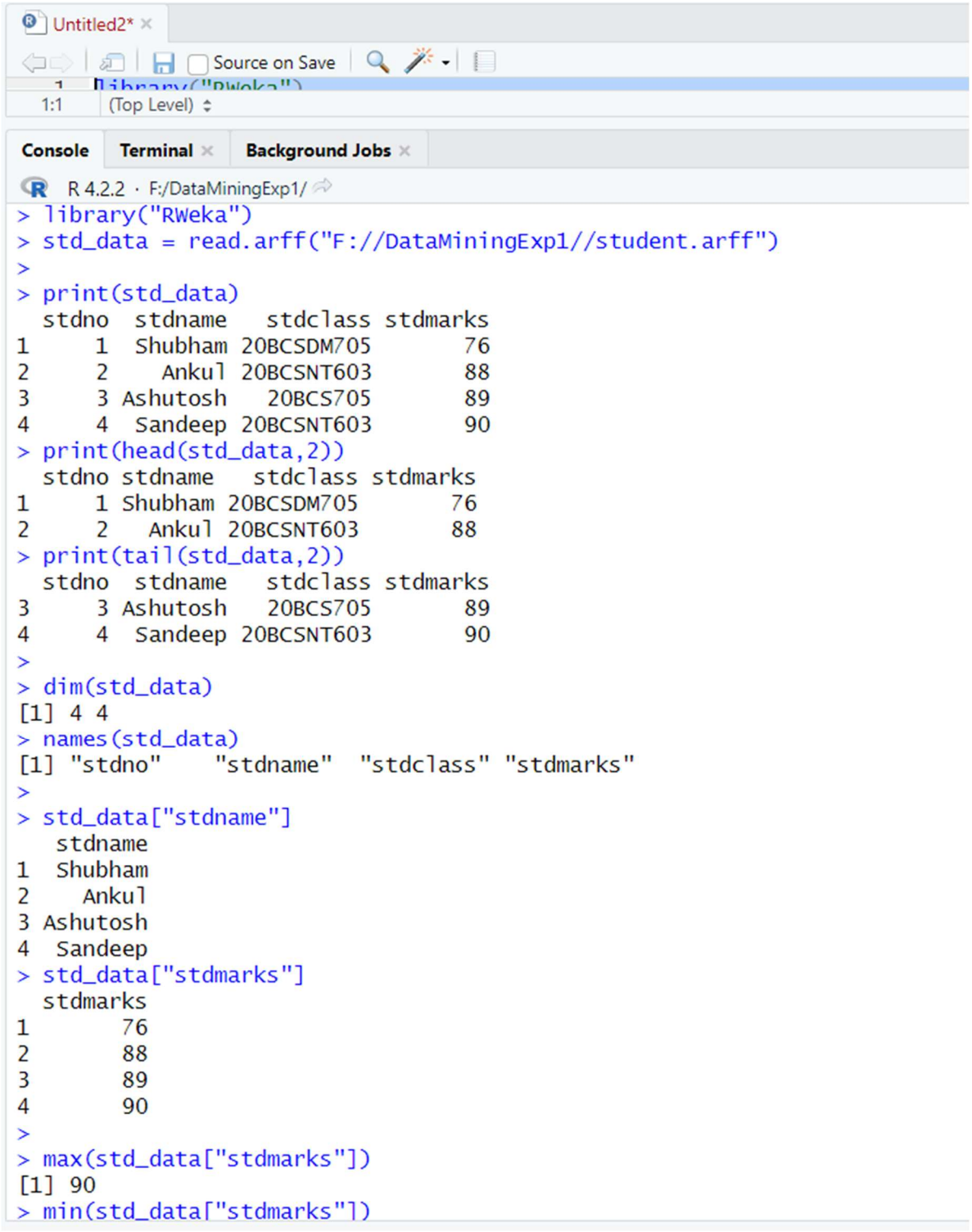
Code:

```
library("RWeka")
std_data = read.arff("F://DataMiningExp1//student.arff")

print(std_data)
print(head(std_data,2))
print(tail(std_data,2))

dim(std_data)
names(std_data)

std_data["stdname"]
std_data["stdmarks"]

max(std_data["stdmarks"])
min(std_data["stdmarks"])
sum(std_data["stdmarks"])

mean(stdmarks)
median(sort(stdmarks))
sd(stdmarks)

summary(stdmarks)
```


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Untitled2* shubham1.R
Source on Save Run Source
1 library("RWeka")
2 std_data = read.arff("F://DataMiningExp1//student.arff")
3
4 print(std_data)
5 print(head(std_data,2))
6 print(tail(std_data,2))
7
8 dim(std_data)
9 names(std_data)
10
11 std_data["stdname"]
12 std_data["stdmarks"]
13
14 max(std_data["stdmarks"])
15 min(std_data["stdmarks"])
16 sum(std_data["stdmarks"])
17
18 mean(stdmarks)
19 median(sort(stdmarks))
20 sd(stdmarks)
21
22 summary(stdmarks)
23
19:23 (Top Level) R Script
Console Terminal Background Jobs
R 4.2.2 · F:/DataMiningExp1/
> library("RWeka")
> std_data = read.arff("F://DataMiningExp1//student.arff")
> print(std_data)
stdno stdname stdclass stdmarks
1 1 Shubham 20BCSDM705 76
2 2 Ankul 20BCSNT603 88
3 3 Ashutosh 20BCS705 89
4 4 Sandeep 20BCSNT603 90
> print(head(std_data,2))
stdno stdname stdclass stdmarks
1 1 Shubham 20BCSDM705 76
2 2 Ankul 20BCSNT603 88
> print(tail(std_data,2))
stdno stdname stdclass stdmarks
3 3 Ashutosh 20BCS705 89
Environment History Connections Tutorial
Import 192 MiB List
R Global Environment
my_list List of 7
newData 955 obs. of 4 variables
sortData 6 obs. of 6 variables
SortValue 6 obs. of 6 variables
std_data 4 obs. of 4 variables
Std_data 6 obs. of 6 variables
student 4 obs. of 4 variables
Student_De... 7 obs. of 3 variables
vec chr [1:3] "Rohan" "is an a...
Values
Files Plots Packages Help Viewer Presentation
Present Print Edit


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Untitled2*
shubham1.R
Source on Save
1 library("RWeka")
2 std_data = read.arff("F://DataMiningExp1//student.arff")
3
4 print(std_data)
5 print(head(std_data,2))
6 print(tail(std_data,2))
7
8 dim(std_data)
9 names(std_data)
10
11 std_data["stdname"]
12 std_data["stdmarks"]
13
14 max(std_data["stdmarks"])
15 min(std_data["stdmarks"])
16 sum(std_data["stdmarks"])
17
18 mean(stdmarks)
19 median(sort(stdmarks))
20 sd(stdmarks)
21
22 summary(stdmarks)
23
20:13
(Top Level)
Console
Terminal
Background Jobs
R 4.2.2 · F:/DataMiningExp1/
> library("RWeka")
> std_data = read.arff("F://DataMiningExp1//student.arff")

Output:


Untitled2*
Source on Save
library("RWeka")
1:1
(Top Level)
Console
Terminal
Background Jobs
R 4.2.2 · F:/DataMiningExp1/
> library("RWeka")
> std_data = read.arff("F://DataMiningExp1//student.arff")
>
> print(std_data)
stdno stdname stdclass stdmarks
1 1 Shubham 20BCSDM705 76
2 2 Ankul 20BCSNT603 88
3 3 Ashutosh 20BCS705 89
4 4 Sandeep 20BCSNT603 90
> print(head(std_data,2))
stdno stdname stdclass stdmarks
1 1 Shubham 20BCSDM705 76
2 2 Ankul 20BCSNT603 88
> print(tail(std_data,2))
stdno stdname stdclass stdmarks
3 3 Ashutosh 20BCS705 89
4 4 Sandeep 20BCSNT603 90
>
> dim(std_data)
[1] 4 4
> names(std_data)
[1] "stdno" "stdname" "stdclass" "stdmarks"
>
> std_data["stdname"]
stdname
1 Shubham
2 Ankul
3 Ashutosh
4 Sandeep
> std_data["stdmarks"]
stdmarks
1 76
2 88
3 89
4 90
>
> max(std_data["stdmarks"])
[1] 90
> min(std_data["stdmarks"])

```
> min(std_data["stdmarks"])
[1] 76
> sum(std_data["stdmarks"])
[1] 343
>
> mean(stdmarks)
[1] 85.75
> median(sort(stdmarks))
[1] 88.5
> sd(stdmarks)
[1] 6.551081
>
> summary(stdmarks)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  76.00   85.00   88.50   85.75   89.25   90.00
>
```