PROGRAM -1

1The intervals and corresponding frequencies are as follows. age frequency

1-5. 200

5-15 450

15-20 300

20-50 1500

50-80 700

80-110 44

Compute an approximate median value for the data

Code:

```
x <- c(1:5,5:15,15:20,20:50,50:80,80:110)
y <- c(200,450,300,1500,700,44)


median(x)
median(y)
```

output

```
> median(x)
[1] 54
> median(y)
[1] 375
>
```

Program 2 :

2.Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

(a) What is the mean of the data? What is the median?

(b) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.).

(c) What is the midrange of the data?

(d) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?

Code:

```
a <-c(13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,40,45,46,52,70)
```

mean(a)

median(a)

mode(a)

range= IQR(a)

range

Q1=quantile (a,.25)

print(Q1)

Q3=quantile (a,.75)

print(Q3)

```
> source("C:/Users/SANJIVA/OneDrive/Documents/tre.R")
> x <- c(1:5,5:15,15:20,20:50,50:80,80:110)
> y <- c(200,450,300,1500,700,44)
>
> median(x)
[1] 54
> median(y)
[1] 375
> a <-c(13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,40,45,46,52,70)
> mean(a)
[1] 29.73077
> median(a)
[1] 25
> mode=function()
+   return(names(sort(-table(a)))[1])
> mode()
[1] "25"
> range= IQR(a)
> range
[1] 14.75
> Q1=quantile (a,.25)
> print(Q1)
  25%
20.25
> Q3=quantile (a,.75)
> print(Q3)
'5%
 35
```

Program 3;
3.Data Preprocessing :Reduction and Transformation

Use the two methods below to normalize the following group of data: 200, 300, 400, 600, 1000 (a) min-max normalization by setting min = 0 and max = 1 (b) z-score normalization

Code:

```
q<-c(200, 300, 400, 600, 1000)
m=mean(q)
s=sd(q)
p=((a-m)*2/s)*0.5
p
summary(p)
print("z_score normal form")
```

h=(q-min(q)/max(q)-min(q))
h

Output:

```
> summary(p)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -1.540  -1.517  -1.502  -1.487  -1.470  -1.360
> h=(q-min(q)/max(q)-min(q))
> h
[1]  -0.2  99.8 199.8 399.8 799.8
```

Program 4

Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:

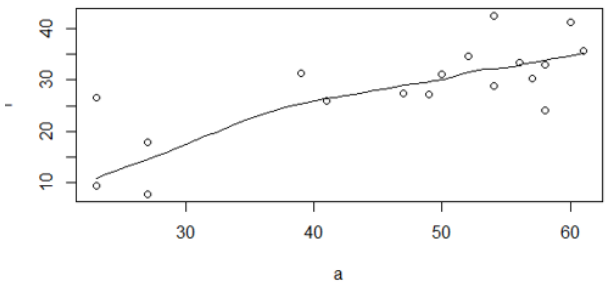
age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

Code:

```
1  a<-c(23,23,27,27,39,41,47,49,50,52,54,54,56,57,58,58,60,61)
2  f<-c(9.5,26.5,7.8,17.8,31.4,25.9,27.4,27.2,31.2,34.6,42.5,28.8,33.4,30.2,24.1,32.9,4
3  m1=mean(a)
4  m1
5  m2=mean(f)
6  m2
7  s=sd(a)
8  s
9  t=sd(f)
10 t
11 md1=median(a)
12 md1
13 md2=median(f)
14 md2
15 boxplot(a,f)
16 scatter.smooth(a,f)
17
```

Output:

```
R 4.2.0 · ~/
[1] 46.44444
> m2=mean(f)
> m2
[1] 28.22778
> s=sd(a)
> s
[1] 13.21862
> t=sd(f)
> t
[1] 9.216533
> md1=median(a)
> md1
[1] 51
> md2=median(f)
> md2
[1] 29.5
```



Program 5

The following values are the number of pencils available in the different boxes. Create a vector and find out the mean, median and mode values of set of pencils in the given data.

Box1	Box2	Box3	Box4	Box5	Box6	Box7	Box8	Box9	Box 10
9	25	23	12	11	6	7	8	9	10

Code:

```
y<-c(1:10)
y
s <-c(9,25,23,12,11,6,7,8,9,10)
s
mean(s)
median(s)
mode=function()
  return(names(sort(-table(s)))[1])
mode()
```

output

```
> s <-c(9,25,23,12,11,6,7,8,9,10)
> s
 [1]  9 25 23 12 11  6  7  8  9 10
> mean(s)
[1] 12
> median(s)
[1] 9.5
> mode=function()
+   return(names(sort(-table(s)))[1])
> mode()
[1] "9"
>
```

Program 6

the following table would be plotted as (x,y) points, with the first column being the x values as number of mobile phones sold and the second column being the y values as money. To use the scatter plot for how many mobile phones sold.

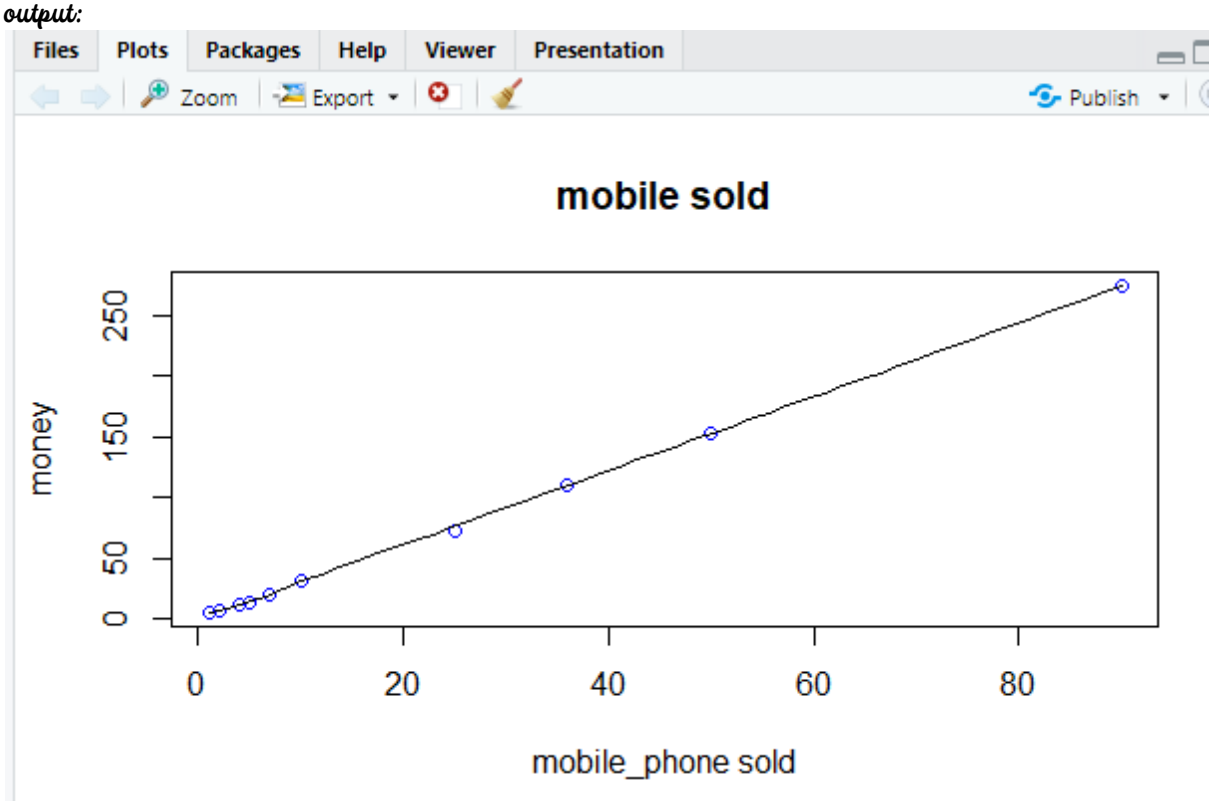
x :4 1 5 7 10 2 50 25 90 36

y :12 5 13 19 31 7 153 72 275 110

code:

```
x<-c(4,1,5,7,10,2,50,25,90,36)
y<-c(12,5,13,19,31,7,153,72,275,110)
scatter.smooth(x,y,main="mobile sold",xlab="mobile_phone sold",ylab="money",col="blue")
```

output:



Program 7

.Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?

Code;

```
k<-c(13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70)
q1=quantile(k,.25)
q1
q3=quantile(k,.75)
q3
```

output:

```
> k<-c(13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35,
35, 36, 40, 45, 46, 52, 70)
> q1=quantile(k,.25)
> q1
 25%
20.5
> q3=quantile(k,.75)
> q3
75%
 35
```

Program 8

Suppose that the speed car is mentioned in different driving style.

Regular 78.3 81.8 82 74.2 83.4 84.5 82.9 77.5 80.9 70.6 Speed

Calculate the Inter quantile and standard deviation of the given data

Code;

```
j<-c(78.3,81.8,82,74.2,83.4,84.5,82.9,77.5,80.9,70.6)
p=IQR(j)
p
s=sd(j)
s
```

output:

```
> source("C:/Users/SANJIVA/OneDrive/Documents/q1,q3.R")
> j<-c(78.3,81.8,82,74.2,83.4,84.5,82.9,77.5,80.9,70.6)
> p=IQR(j)
> p
[1] 4.975
> s=sd(j)
> s
[1] 4.445835
```