MOHAMMED RAFIK.M
192124179

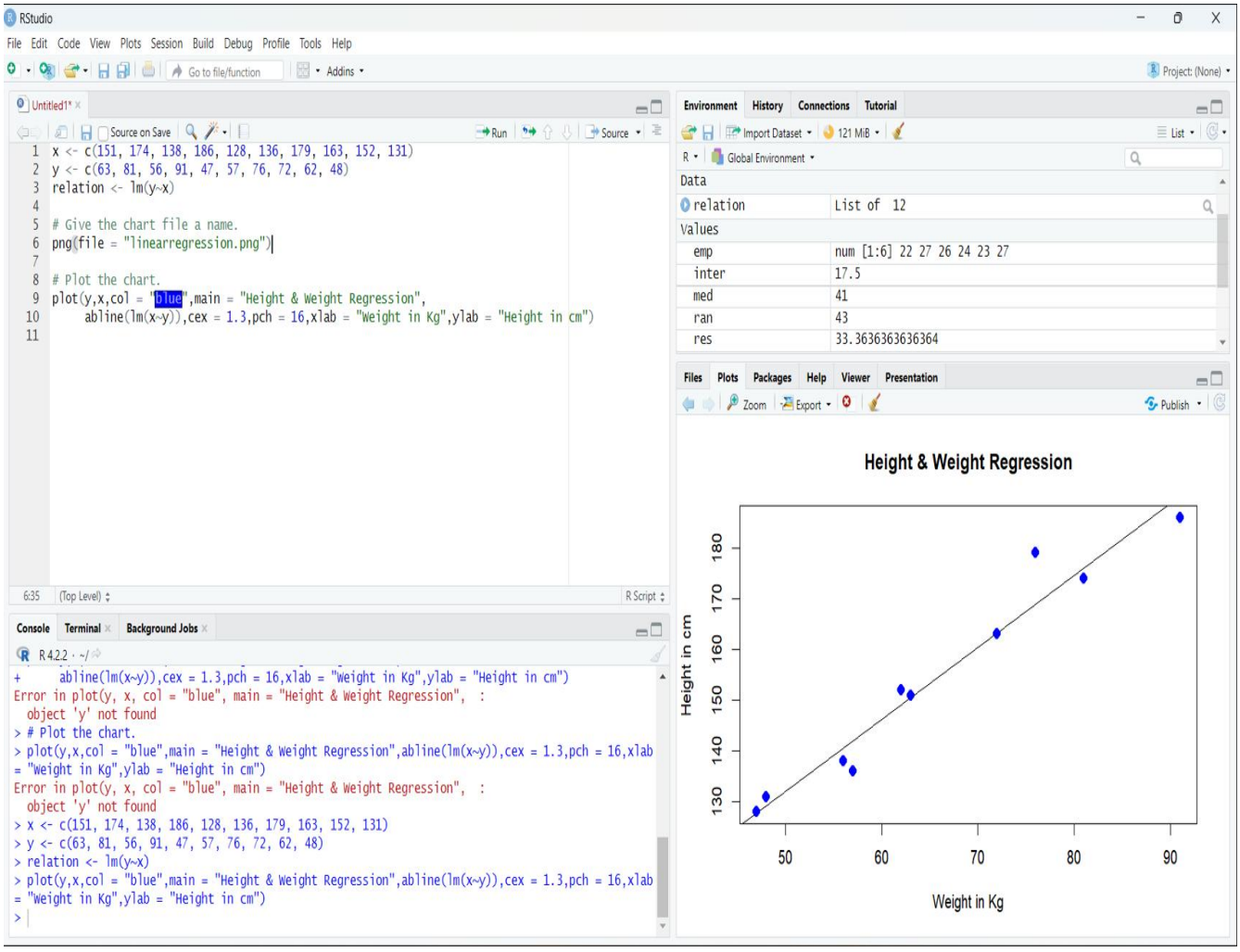
CSA1668 - DWDM FOR PATTERN ANALYSIS

1. PREDICTION ANALYSIS USING LINEAR REGRESSION THROUGH TOOL.

LINEAR REGRESSION:

Linear regression is a kind of statistical analysis that attempts to show a relationship between two variables. Linear regression looks at various data points and plots a trend line. Linear regression can create a predictive model on apparently random data, showing trends in data, such as in cancer diagnoses or in stock price.

RESULT:

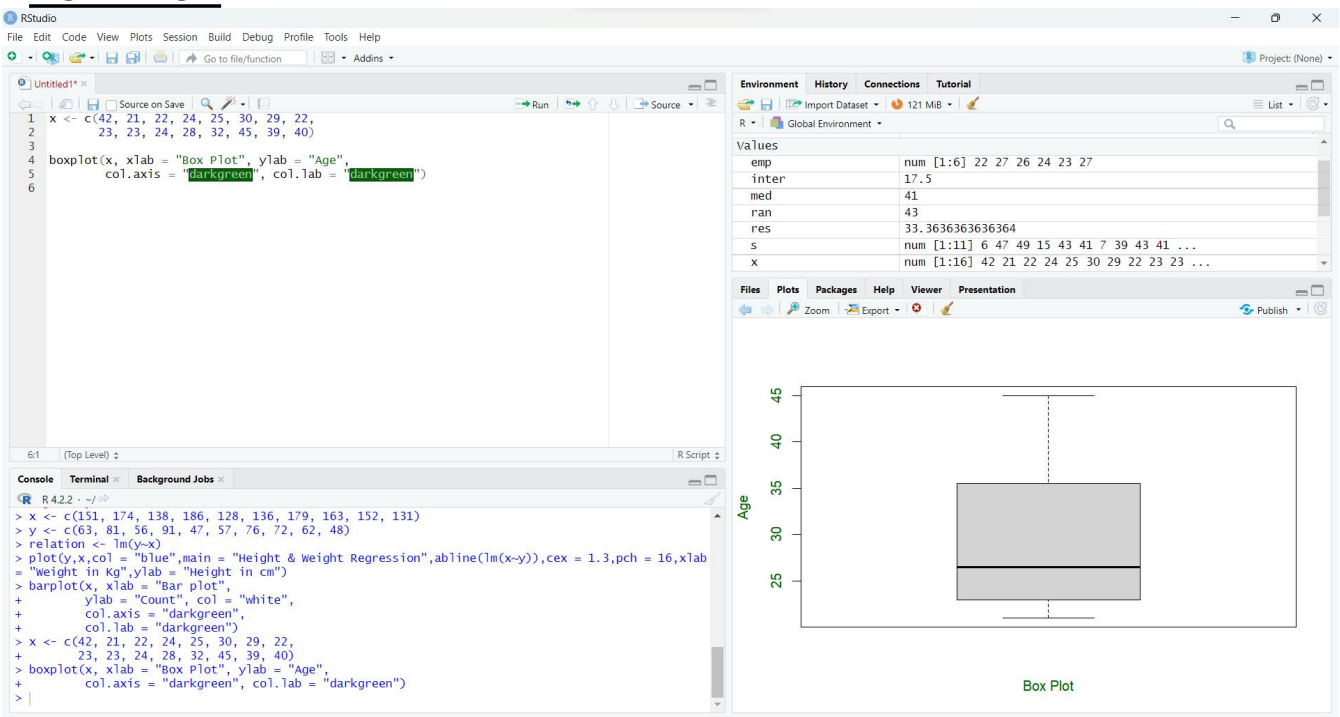


2. PLOTTING GRAPHS USING R TOOL.

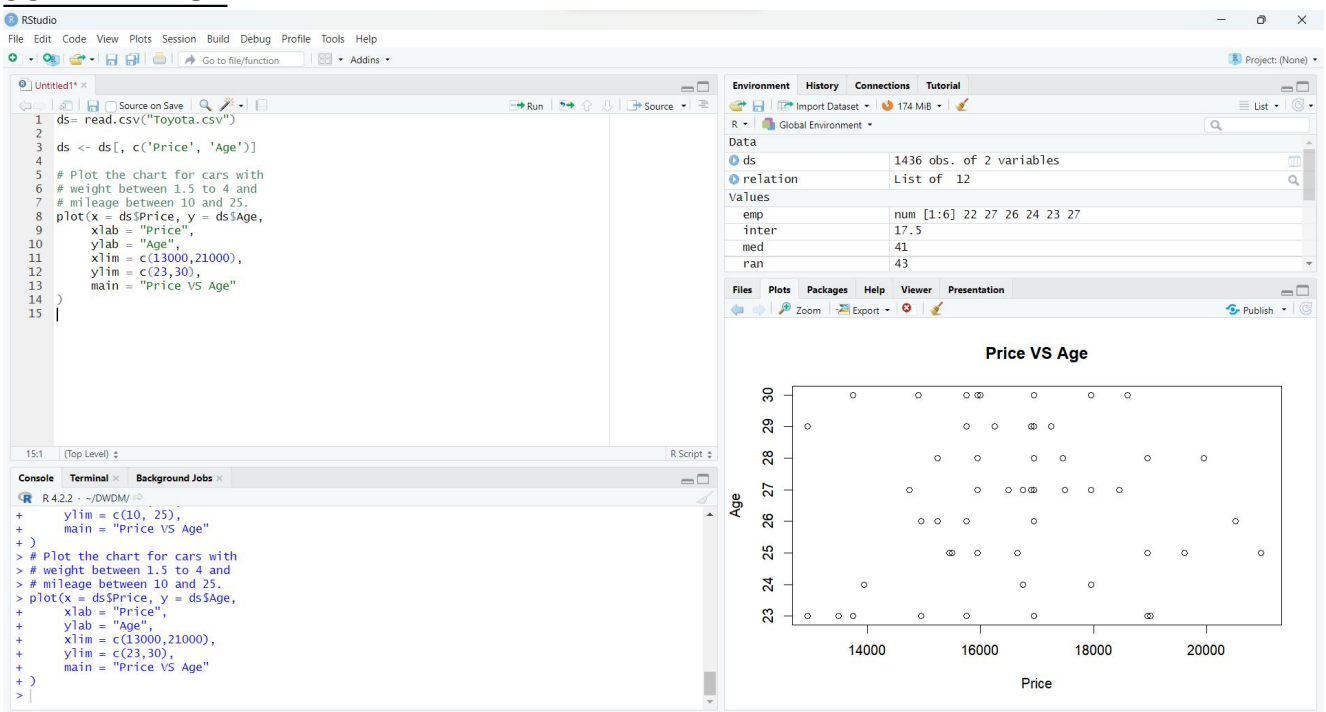
GRAPHS:

* Box plot
* scatter plot
* histogram

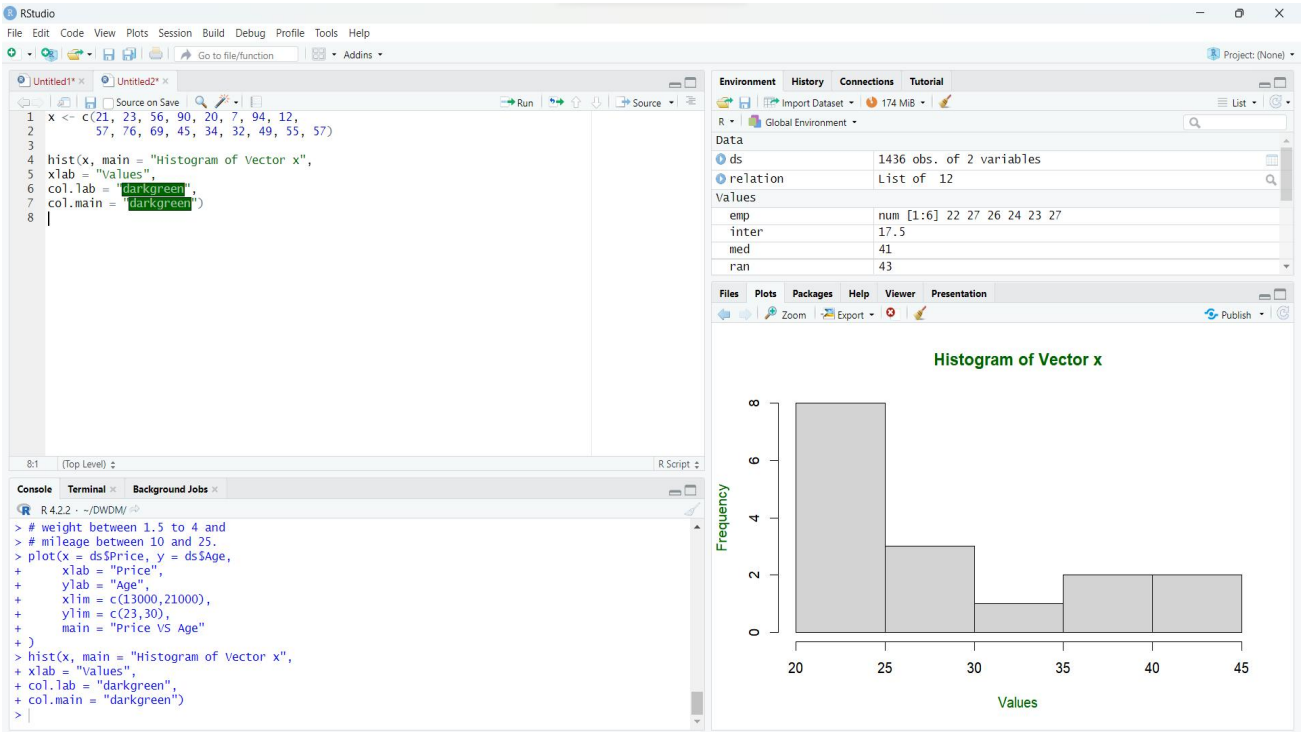
BOX PLOT



SCATTER PLOT



HISTOGRAM



3. CENTRAL TENDENCY AND DATA DISPERSION MEASURES USING R-TOOL.

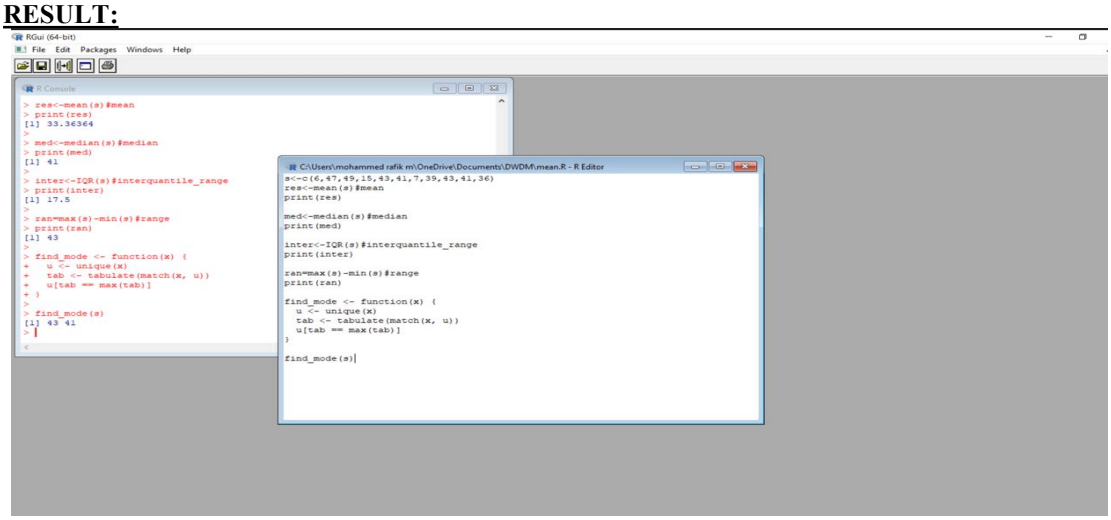
CENTRAL TENDENCY:

i. Mean :The mean is the average of the numbers: a calculated "central" value of a set of numbers.

ii. Median : The median is a statistical term that is one way of finding the 'average' of a set of data points.

iii. Mode : The mode of a set of data values is the value that appears most often.

RESULT:



4. PERFORM CORRECTION ANALYSIS AND NORMALIZATION.

CORRELATION ANALYSIS:

STEPS INVOLVED:

i. Create a new table with required dataframes.

ii. After that apply the formula or query for the chi-square test.

RESULT:

```
1  ds= read.csv("diabetes_dataset.csv")
2  ds1<-table(ds$Age,ds$Insulin)
3  ds1
4
```

4:1 (Top Level)

Console Terminal Background Jobs

R 4.2.2 · ~/DWDM/

```
    0 14 15 16 18 22 23 25 29 32 36 37 38 40 41 42 43 44 45 46 48 49 50 51 52 53 54 55 56 57
21 28  0  0  0  1  0  1  1  0  0  0  0  0  1  0  0  0  0  1  0  0  0  1  1  0  0  0  0  1  0
22 29  0  0  1  0  0  0  0  0  1  1  1  0  0  0  0  0  1  1  0  0  0  1  0  0  1  1  0  0  0
23 10  0  1  0  0  0  0  0  0  0  0  1  0  0  0  0  0  1  1  0  2  0  0  0  0  0  0  0  0  0
24 15  0  0  0  0  0  0  0  0  0  1  0  0  1  0  0  1  0  0  0  0  0  1  0  0  1  0  1  1  0
25 18  1  0  0  1  0  0  0  0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  1  0  0  0  1  1

   58 59 60 61 63 64 65 66 67 68 70 71 72 73 74 75 76 77 78 79 81 82 83 84 85 86 87 88 89 90
21  0  0  0  1  0  1  0  1  0  0  0  1  0  0  0  0  2  0  1  0  0  1  0  0  0  0  1  0  1  0
22  1  0  1  0  1  1  0  1  0  0  0  0  0  0  0  1  1  0  0  0  0  1  1  1  0  0  0  2  0  1
23  0  0  0  0  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
24  0  0  0  0  0  0  0  0  1  0  0  0  1  0  0  0  0  0  1  0  0  0  1  0  0  0  1  0  0  0  1
25  0  1  0  0  1  0  0  1  0  0  0  0  0  0  0  1  0  1  0  0  0  0  0  0  0  1  0  0  0  0  0

   91 92 94 95 96 99 100 105 106 108 110 112 114 115 116 119 120 122 125 126 127 128 129 130
21  0  0  2  0  0  0   0   1   0   0   0   0   0   1   0   0   1   0   0   0   0   0   0   0   0
22  0  1  0  0  0  0   0   2   1   0   0   0   0   0   0   1   0   1   0   0   0   0   0   0   0
23  0  1  1  2  0  0   2   0   0   0   0   0   0   0   0   1   0   1   0   1   0   0   0   0   0
24  0  0  0  0  0  0   0   2   0   0   0   0   0   0   0   0   0   2   0   0   0   0   1   0   1
25  0  0  1  0  0  0   2   0   1   0   1   0   0   1   0   0   0   0   0   0   0   0   0   1   0

   132 135 140 142 144 145 146 148 150 152 155 156 158 159 160 165 166 167 168 170 171 175
21   0   0   0   0   0   0   0   1   0   1   0   0   0   0   1   0   0   1   0   1   0   0
22   0   1   1   1   0   0   0   1   0   0   0   0   0   1   0   1   0   0   0   0   0   0
23   0   1   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0
24   1   1   1   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0   1   0   0   0
25   0   1   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0

   176 178 180 182 183 184 185 188 190 191 192 193 194 196 200 204 205 207 210 215 220 225
21   0   0   0   1   0   0   0   0   0   0   0   0   0   1   0   0   0   1   0   0   0
22   0   0   0   0   1   0   1   1   0   0   0   0   0   0   0   0   0   0   0   0   0
```

NORMALIZATION:

*MIN_MAX

*Z- score

*Decimal scaling

* MIN_MAX:

```
> minimum=min(diabetes$Age)
> maximum=max(diabetes$Age)
> minmax=(A-minimum)/(maximum-minimum)
> minmax
  [1] 0.48333333 0.16666667 0.18333333 0.00000000 0.20000000 0.15000000 0.08333333
  [8] 0.13333333 0.53333333 0.55000000 0.15000000 0.21666667 0.60000000 0.63333333
 [15] 0.50000000 0.18333333 0.16666667 0.16666667 0.20000000 0.18333333 0.10000000
 [22] 0.48333333 0.33333333 0.13333333 0.50000000 0.33333333 0.36666667 0.01666667
 [29] 0.60000000 0.28333333 0.65000000 0.11666667 0.01666667 0.11666667 0.40000000
 [36] 0.20000000 0.23333333 0.41666667 0.10000000 0.58333333 0.08333333 0.26666667
 [43] 0.45000000 0.55000000 0.31666667 0.06666667 0.13333333 0.01666667 0.16666667
 [50] 0.05000000 0.01666667 0.08333333 0.15000000 0.61666667 0.35000000 0.00000000
```

* Z-score

```
> View(diabetes)
> A<-c(diabetes$Age)
> mean=mean(A)
> std=sd(A)
> zscore=(A-mean)/std
> zscore
  [1]  1.42506672 -0.19054773 -0.10551539 -1.04087112 -0.02048305 -0.27558007
  [7] -0.61570943 -0.36061241  1.68016374  1.76519608 -0.27558007  0.06454929
 [13]  2.02029310  2.19035777  1.51009906 -0.10551539 -0.19054773 -0.19054773
 [19] -0.02048305 -0.10551539 -0.53067709  1.42506672  0.65977566 -0.36061241
 [25]  1.51009906  0.65977566  0.82984034 -0.95583878  2.02029310  0.40467865
 [31]  2.27539011 -0.44564475 -0.95583878 -0.44564475  0.99990502 -0.02048305
 [37]  0.14958163  1.08493736 -0.53067709  1.93526076 -0.61570943  0.31964631
```

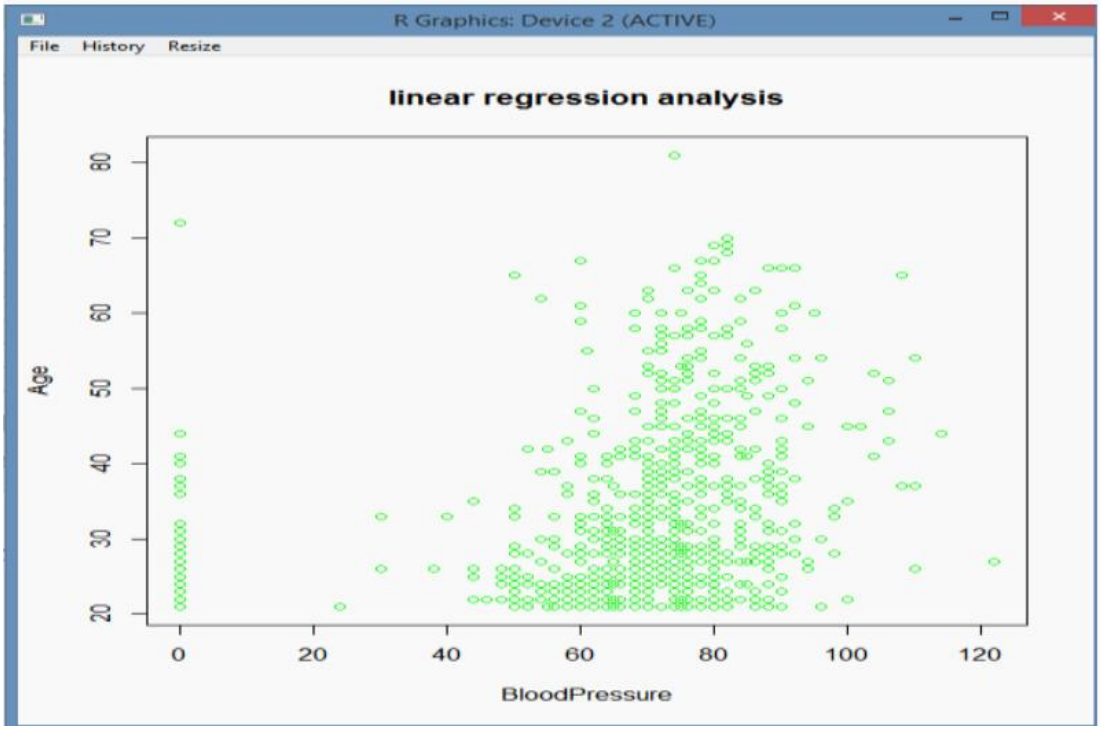
* DECIMAL SCALING

```
> decimascaling=(A/100)
> decimascaling
  [1] 0.50 0.31 0.32 0.21 0.33 0.30 0.26 0.29 0.53 0.54 0.30 0.34 0.57 0.59 0.51 0.32
 [17] 0.31 0.31 0.33 0.32 0.27 0.50 0.41 0.29 0.51 0.41 0.43 0.22 0.57 0.38 0.60 0.28
 [33] 0.22 0.28 0.45 0.33 0.35 0.46 0.27 0.56 0.26 0.37 0.48 0.54 0.40 0.25 0.29 0.22
 [49] 0.31 0.24 0.22 0.26 0.30 0.58 0.42 0.21 0.41 0.31 0.44 0.22 0.21 0.39 0.36 0.24
 [65] 0.42 0.32 0.38 0.54 0.25 0.27 0.28 0.26 0.42 0.23 0.22 0.22 0.41 0.27 0.26 0.24
 [81] 0.22 0.22 0.36 0.22 0.37 0.27 0.45 0.26 0.43 0.24 0.21 0.34 0.42 0.60 0.21 0.40
 [97] 0.24 0.22 0.23 0.31 0.33 0.22 0.21 0.24 0.27 0.21 0.27 0.37 0.25 0.24 0.24 0.46
[113] 0.23 0.25 0.39 0.61 0.38 0.25 0.22 0.21 0.25 0.24 0.23 0.69 0.23 0.26 0.30 0.23
[129] 0.40 0.62 0.33 0.33 0.30 0.39 0.26 0.31 0.21 0.22 0.29 0.28 0.55 0.38 0.22 0.42
[145] 0.23 0.21 0.41 0.34 0.65 0.22 0.24 0.37 0.42 0.23 0.43 0.36 0.21 0.23 0.22 0.47
```

5. REGRESSION ANALYSIS USING R TOOL.

REGRESSION ANALYSIS:

It is a method that is used to estimate the relationship between one or more independent variables and a dependent variable. These independent variables can be defined as an assumption or driver that is altered to evaluate its influence on a dependent variable which is the result or the outcome.


R Graphics: Device 2 (ACTIVE)
File History Resize
linear regression analysis
Age
20 30 40 50 60 70 80
0 20 40 60 80 100 120
BloodPressure