Dataset Overview:

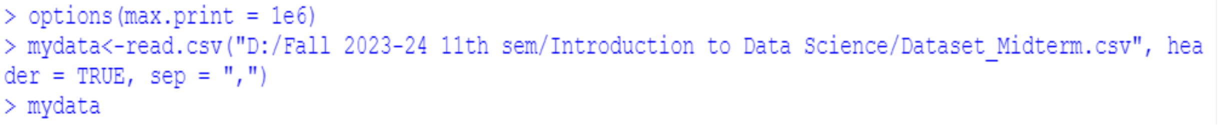
The Diabetes Prediction Dataset includes patients' medical and demographic information as well as the state of their diabetes (positive or negative). The provided dataset contains various health-related attributes. The information includes gender, age, hypertension, heart disease, smoking history, BMI, HbA1c level, blood glucose level, and diabetes status. This dataset appears to be intended for research or analysis related to health and wellness, with a focus on factors that may impact an individual's health. The project's objective seems to be to investigate the relationships between these factors and their potential impact on developing diabetes. The findings from this analysis could be valuable for understanding how these factors are associated with specific health conditions and for making informed decisions regarding diabetes identification and personalized treatment.

Reference: https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset/data

Data pre-processing:

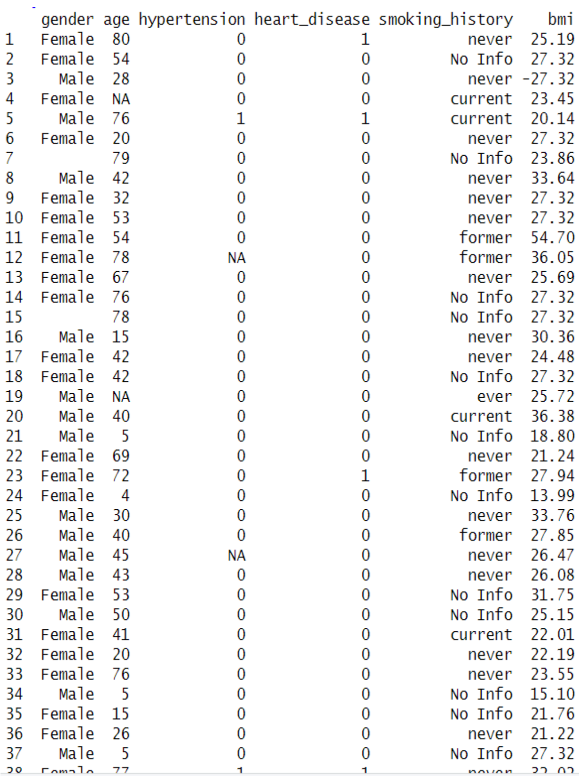
*Importing the Dataset:

At first the dataset is imported from the correct path. The data is read from a CSV file named "Dataset_Midterm.csv" using the read.csv method in R. The resulting data frame is then assigned to a variable named mydata. mydata is then used to display the contents of the dataset in the console. The output of this code displays the entire contents of the dataset, including all rows and columns.

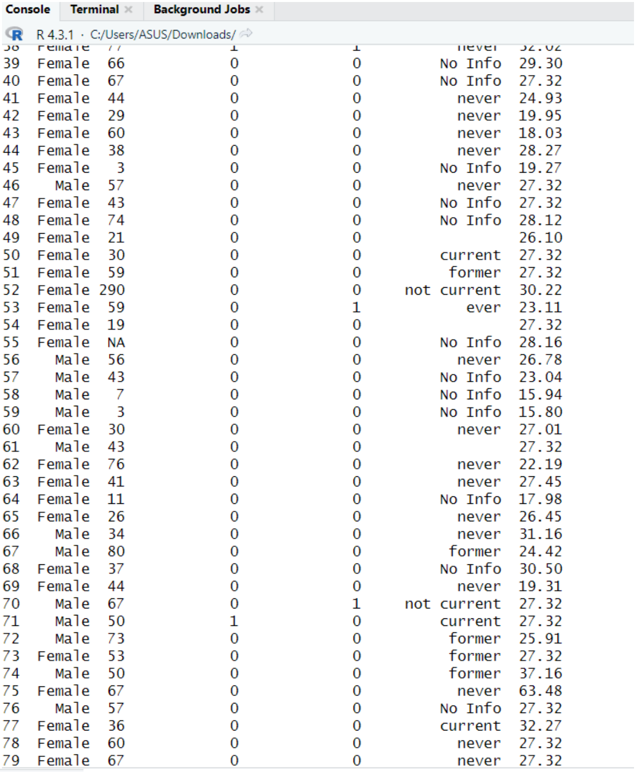


```
> options(max.print = 1e6)
> mydata<-read.csv("D:/Fall 2023-24 11th sem/Introduction to Data Science/Dataset_Midterm.csv", hea
der = TRUE, sep = ",")
> mydata
```

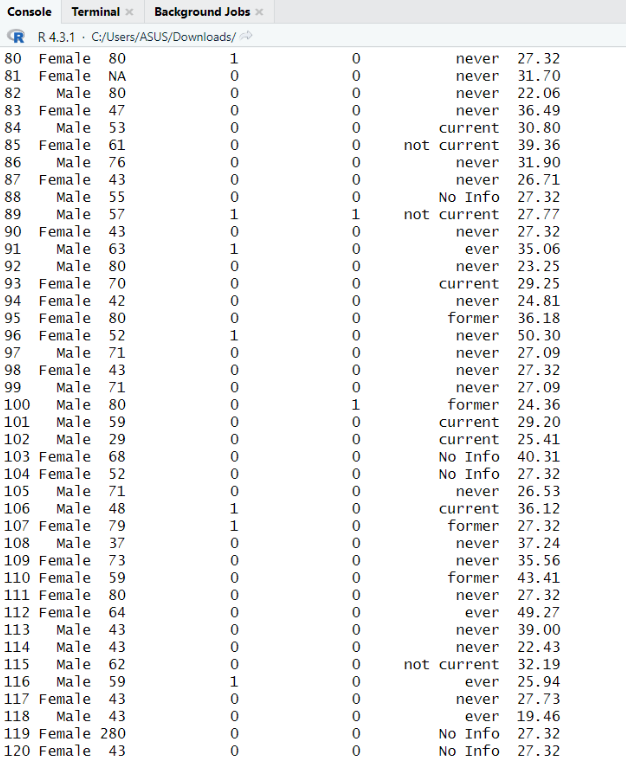
The output provides a way to verify that the data has been read correctly and allows for a quick visual inspection of the dataset. The output also provides information about the structure of the data, such as the number of columns and rows and the type of data in each column. Overall, this code and its output are a necessary first step in any data analysis project in R. The output is given below:



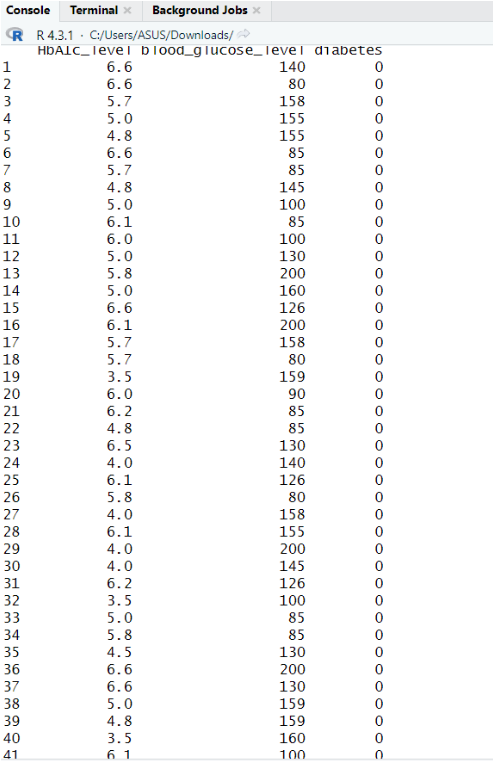
```
    gender age hypertension heart_disease smoking_history    bmi
1   Female  80            0             1           never  25.19
2   Female  54            0             0         No Info  27.32
3     Male  28            0             0           never -27.32
4   Female  NA            0             0         current  23.45
5     Male  76            1             1         current  20.14
6   Female  20            0             0           never  27.32
7           79            0             0         No Info  23.86
8     Male  42            0             0           never  33.64
9   Female  32            0             0           never  27.32
10  Female  53            0             0           never  27.32
11  Female  54            0             0          former  54.70
12  Female  78           NA             0          former  36.05
13  Female  67            0             0           never  25.69
14  Female  76            0             0         No Info  27.32
15          78            0             0         No Info  27.32
16    Male  15            0             0           never  30.36
17  Female  42            0             0           never  24.48
18  Female  42            0             0         No Info  27.32
19    Male  NA            0             0            ever  25.72
20    Male  40            0             0         current  36.38
21    Male   5            0             0         No Info  18.80
22  Female  69            0             0           never  21.24
23  Female  72            0             1          former  27.94
24  Female   4            0             0         No Info  13.99
25    Male  30            0             0           never  33.76
26    Male  40            0             0          former  27.85
27    Male  45           NA             0           never  26.47
28    Male  43            0             0           never  26.08
29  Female  53            0             0         No Info  31.75
30    Male  50            0             0         No Info  25.15
31  Female  41            0             0         current  22.01
32  Female  20            0             0           never  22.19
33  Female  76            0             0           never  23.55
34    Male   5            0             0         No Info  15.10
35  Female  15            0             0         No Info  21.76
36  Female  26            0             0           never  21.22
37    Male   5            0             0         No Info  27.32
38  Female  77            1             1           never  32.02
```



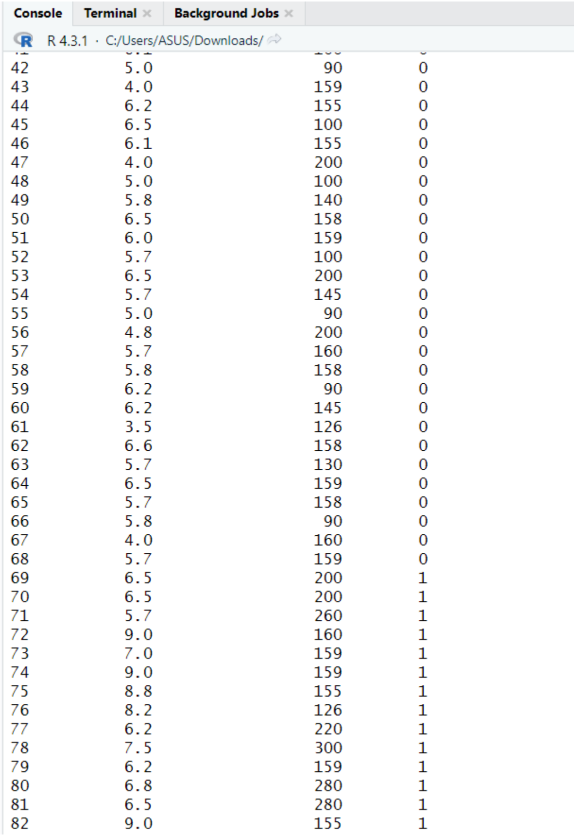
```
38  Female  77            1             1           never  32.02
39  Female  66            0             0         No Info  29.30
40  Female  67            0             0         No Info  27.32
41  Female  44            0             0           never  24.93
42  Female  29            0             0           never  19.95
43  Female  60            0             0           never  18.03
44  Female  38            0             0           never  28.27
45  Female   3            0             0         No Info  19.27
46    Male  57            0             0           never  27.32
47  Female  43            0             0         No Info  27.32
48  Female  74            0             0         No Info  28.12
49  Female  21            0             0                  26.10
50  Female  30            0             0         current  27.32
51  Female  59            0             0          former  27.32
52  Female 290            0             0     not current  30.22
53  Female  59            0             1            ever  23.11
54  Female  19            0             0                  27.32
55  Female  NA            0             0         No Info  28.16
56    Male  56            0             0           never  26.78
57    Male  43            0             0         No Info  23.04
58    Male   7            0             0         No Info  15.94
59    Male   3            0             0         No Info  15.80
60  Female  30            0             0           never  27.01
61    Male  43            0             0                  27.32
62  Female  76            0             0           never  22.19
63  Female  41            0             0           never  27.45
64  Female  11            0             0         No Info  17.98
65  Female  26            0             0           never  26.45
66    Male  34            0             0           never  31.16
67    Male  80            0             0          former  24.42
68  Female  37            0             0         No Info  30.50
69  Female  44            0             0           never  19.31
70    Male  67            0             1     not current  27.32
71    Male  50            1             0         current  27.32
72    Male  73            0             0          former  25.91
73  Female  53            0             0          former  27.32
74    Male  50            0             0          former  37.16
75  Female  67            0             0           never  63.48
76    Male  57            0             0         No Info  27.32
77  Female  36            0             0         current  32.27
78  Female  60            0             0           never  27.32
79  Female  67            0             0           never  27.32
```



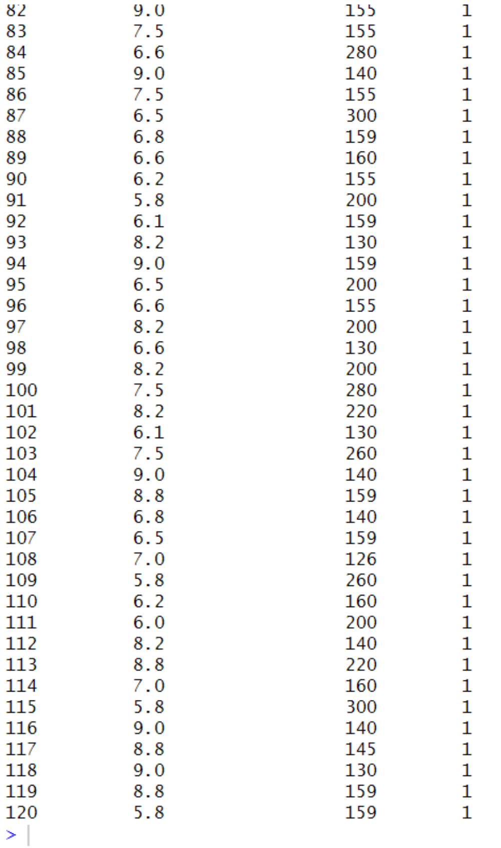
```
80  Female  80            1             0           never  27.32
81  Female  NA            0             0           never  31.70
82    Male  80            0             0           never  22.06
83  Female  47            0             0           never  36.49
84    Male  53            0             0         current  30.80
85  Female  61            0             0     not current  39.36
86    Male  76            0             0           never  31.90
87  Female  43            0             0           never  26.71
88    Male  55            0             0         No Info  27.32
89    Male  57            1             1     not current  27.77
90  Female  43            0             0           never  27.32
91    Male  63            1             0            ever  35.06
92    Male  80            0             0           never  23.25
93  Female  70            0             0         current  29.25
94  Female  42            0             0           never  24.81
95  Female  80            0             0          former  36.18
96  Female  52            1             0           never  50.30
97    Male  71            0             0           never  27.09
98  Female  43            0             0           never  27.32
99    Male  71            0             0           never  27.09
100   Male  80            0             1          former  24.36
101   Male  59            0             0         current  29.20
102   Male  29            0             0         current  25.41
103 Female  68            0             0         No Info  40.31
104 Female  52            0             0         No Info  27.32
105   Male  71            0             0           never  26.53
106   Male  48            1             0         current  36.12
107 Female  79            1             0          former  27.32
108   Male  37            0             0           never  37.24
109 Female  73            0             0           never  35.56
110 Female  59            0             0          former  43.41
111 Female  80            0             0           never  27.32
112 Female  64            0             0            ever  49.27
113   Male  43            0             0           never  39.00
114   Male  43            0             0           never  22.43
115   Male  62            0             0     not current  32.19
116   Male  59            1             0            ever  25.94
117 Female  43            0             0           never  27.73
118   Male  43            0             0            ever  19.46
119 Female 280            0             0         No Info  27.32
120 Female  43            0             0         No Info  27.32
```



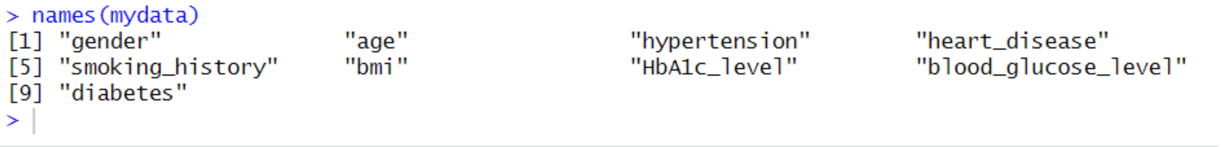
```
   HbA1c_level blood_glucose_level diabetes
1          6.6                 140        0
2          6.6                  80        0
3          5.7                 158        0
4          5.0                 155        0
5          4.8                 155        0
6          6.6                  85        0
7          5.7                  85        0
8          4.8                 145        0
9          5.0                 100        0
10         6.1                  85        0
11         6.0                 100        0
12         5.0                 130        0
13         5.8                 200        0
14         5.0                 160        0
15         6.6                 126        0
16         6.1                 200        0
17         5.7                 158        0
18         5.7                  80        0
19         3.5                 159        0
20         6.0                  90        0
21         6.2                  85        0
22         4.8                  85        0
23         6.5                 130        0
24         4.0                 140        0
25         6.1                 126        0
26         5.8                  80        0
27         4.0                 158        0
28         6.1                 155        0
29         4.0                 200        0
30         4.0                 145        0
31         6.2                 126        0
32         3.5                 100        0
33         5.0                  85        0
34         5.8                  85        0
35         4.5                 130        0
36         6.6                 200        0
37         6.6                 130        0
38         5.0                 159        0
39         4.8                 159        0
40         3.5                 160        0
41         6.1                 100        0
```



```
42          5.0                  90      0
43          4.0                 159      0
44          6.2                 155      0
45          6.5                 100      0
46          6.1                 155      0
47          4.0                 200      0
48          5.0                 100      0
49          5.8                 140      0
50          6.5                 158      0
51          6.0                 159      0
52          5.7                 100      0
53          6.5                 200      0
54          5.7                 145      0
55          5.0                  90      0
56          4.8                 200      0
57          5.7                 160      0
58          5.8                 158      0
59          6.2                  90      0
60          6.2                 145      0
61          3.5                 126      0
62          6.6                 158      0
63          5.7                 130      0
64          6.5                 159      0
65          5.7                 158      0
66          5.8                  90      0
67          4.0                 160      0
68          5.7                 159      0
69          6.5                 200      1
70          6.5                 200      1
71          5.7                 260      1
72          9.0                 160      1
73          7.0                 159      1
74          9.0                 159      1
75          8.8                 155      1
76          8.2                 126      1
77          6.2                 220      1
78          7.5                 300      1
79          6.2                 159      1
80          6.8                 280      1
81          6.5                 280      1
82          9.0                 155      1
```



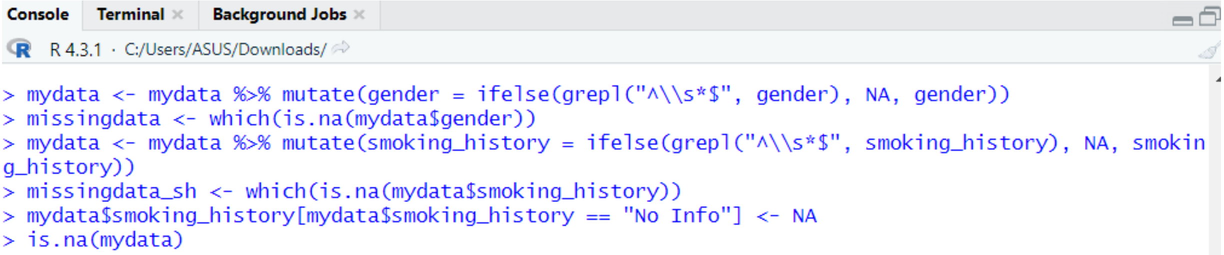
```
82          9.0                 155      1
83          7.5                 155      1
84          6.6                 280      1
85          9.0                 140      1
86          7.5                 155      1
87          6.5                 300      1
88          6.8                 159      1
89          6.6                 160      1
90          6.2                 155      1
91          5.8                 200      1
92          6.1                 159      1
93          8.2                 130      1
94          9.0                 159      1
95          6.5                 200      1
96          6.6                 155      1
97          8.2                 200      1
98          6.6                 130      1
99          8.2                 200      1
100         7.5                 280      1
101         8.2                 220      1
102         6.1                 130      1
103         7.5                 260      1
104         9.0                 140      1
105         8.8                 159      1
106         6.8                 140      1
107         6.5                 159      1
108         7.0                 126      1
109         5.8                 260      1
110         6.2                 160      1
111         6.0                 200      1
112         8.2                 140      1
113         8.8                 220      1
114         7.0                 160      1
115         5.8                 300      1
116         9.0                 140      1
117         8.8                 145      1
118         9.0                 130      1
119         8.8                 159      1
120         5.8                 159      1
> 
```



```
> names(mydata)
[1] "gender"              "age"                 "hypertension"        "heart_disease"
[5] "smoking_history"     "bmi"                 "HbA1c_level"         "blood_glucose_level"
[9] "diabetes"
> 
```

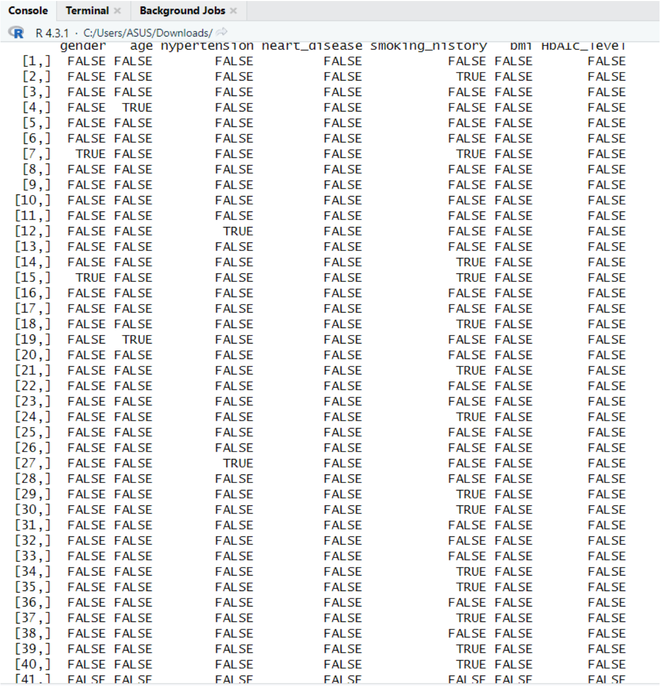
* Identifying Missing Values:

To identify missing value “is.na” function has been used to check for missing values in the variable. The is.na function returns a logical value of TRUE for each missing value in the data and FALSE for all other values.

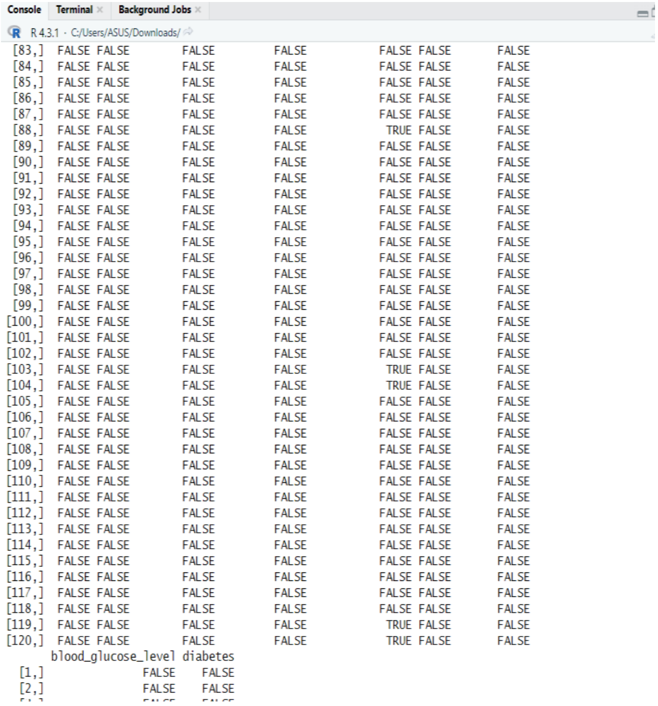


```
> mydata <- mydata %>% mutate(gender = ifelse(grepl("^\\s*$", gender), NA, gender))
> missingdata <- which(is.na(mydata$gender))
> mydata <- mydata %>% mutate(smoking_history = ifelse(grepl("^\\s*$", smoking_history), NA, smokin
g_history))
> missingdata_sh <- which(is.na(mydata$smoking_history))
> mydata$smoking_history[mydata$smoking_history == "No Info"] <- NA
> is.na(mydata)
```

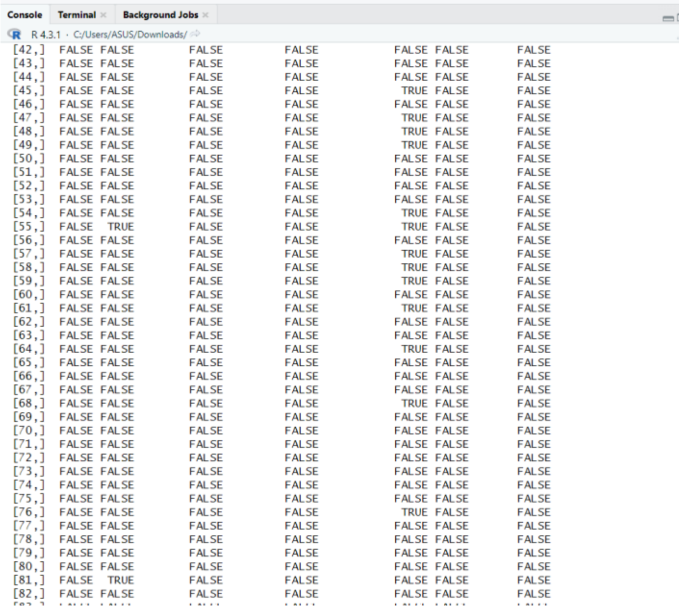
The output of this code is a matrix that has the same dimensions as the variable. The matrix contains TRUE values in the cells where the corresponding values are missing and FALSE values in all other cells. This output is useful for identifying any missing values in the dataset and determining the extent to which missing values may affect subsequent analysis:



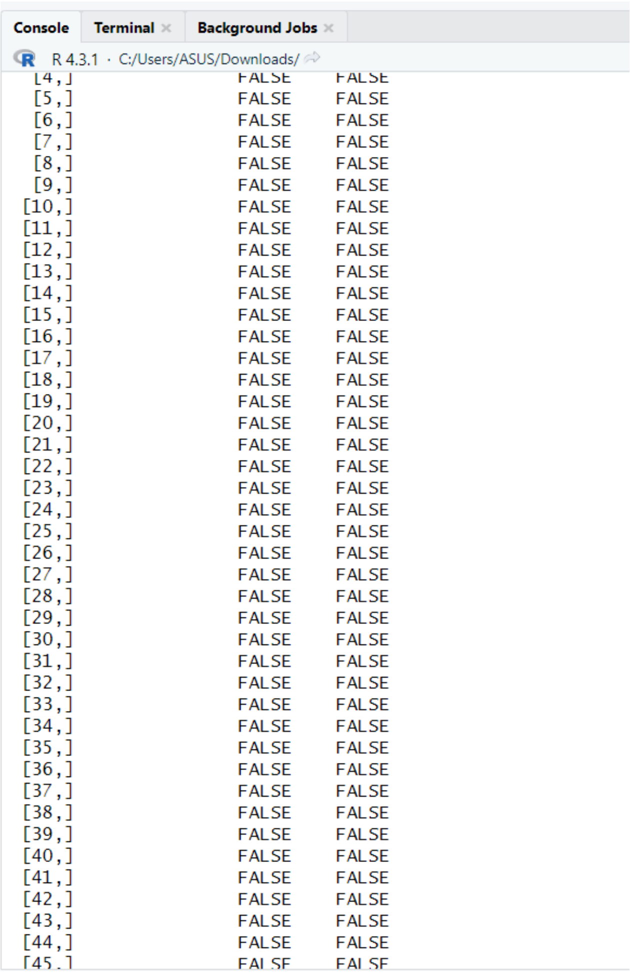
```
       gender   age hypertension heart_disease smoking_history   bmi HbA1c_level
  [1,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
  [2,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
  [3,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
  [4,]  FALSE  TRUE        FALSE         FALSE           FALSE FALSE       FALSE
  [5,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
  [6,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
  [7,]   TRUE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
  [8,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
  [9,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [10,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [11,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [12,]  FALSE FALSE         TRUE         FALSE           FALSE FALSE       FALSE
 [13,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [14,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [15,]   TRUE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [16,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [17,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [18,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [19,]  FALSE  TRUE        FALSE         FALSE           FALSE FALSE       FALSE
 [20,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [21,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [22,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [23,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [24,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [25,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [26,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [27,]  FALSE FALSE         TRUE         FALSE           FALSE FALSE       FALSE
 [28,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [29,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [30,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [31,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [32,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [33,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [34,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [35,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [36,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [37,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [38,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [39,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [40,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [41,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
```



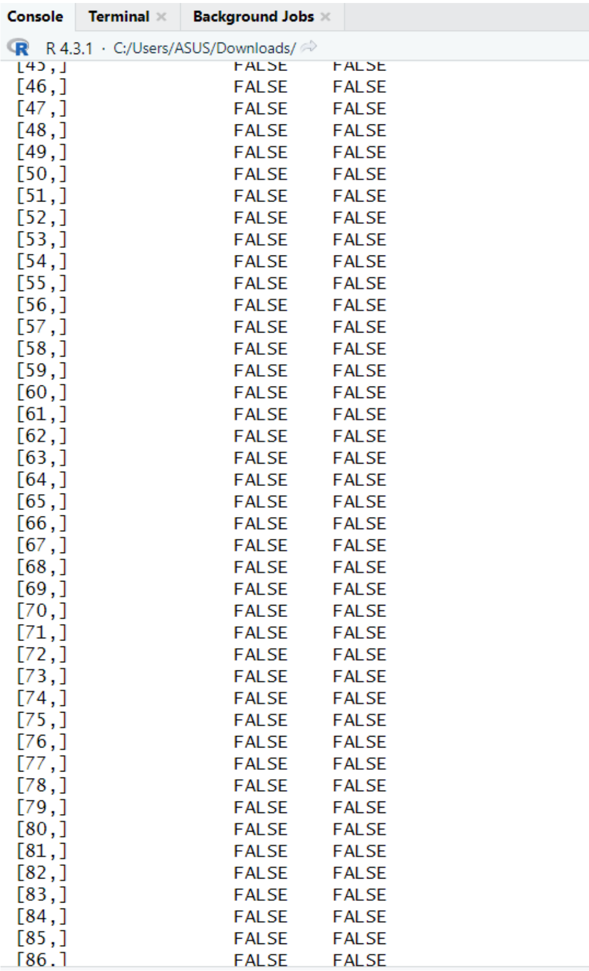
```
 [83,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [84,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [85,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [86,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [87,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [88,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [89,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [90,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [91,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [92,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [93,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [94,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [95,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [96,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [97,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [98,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [99,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[100,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[101,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[102,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[103,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
[104,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
[105,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[106,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[107,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[108,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[109,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[110,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[111,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[112,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[113,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[114,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[115,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[116,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[117,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[118,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
[119,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
[120,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
       blood_glucose_level diabetes
  [1,]               FALSE    FALSE
  [2,]               FALSE    FALSE
```



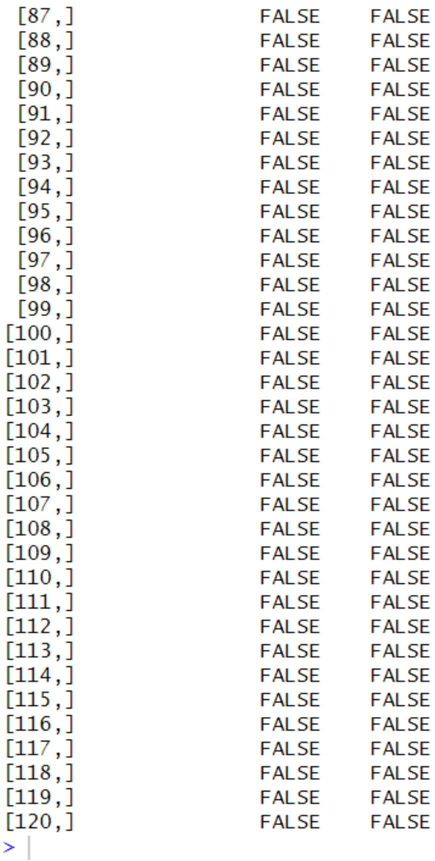
```
 [42,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [43,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [44,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [45,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [46,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [47,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [48,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [49,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [50,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [51,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [52,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [53,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [54,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [55,]  FALSE  TRUE        FALSE         FALSE            TRUE FALSE       FALSE
 [56,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [57,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [58,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [59,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [60,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [61,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [62,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [63,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [64,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [65,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [66,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [67,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [68,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [69,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [70,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [71,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [72,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [73,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [74,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [75,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [76,]  FALSE FALSE        FALSE         FALSE            TRUE FALSE       FALSE
 [77,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [78,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [79,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [80,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
 [81,]  FALSE  TRUE        FALSE         FALSE           FALSE FALSE       FALSE
 [82,]  FALSE FALSE        FALSE         FALSE           FALSE FALSE       FALSE
```



```
  [4,]                FALSE    FALSE
  [5,]                FALSE    FALSE
  [6,]                FALSE    FALSE
  [7,]                FALSE    FALSE
  [8,]                FALSE    FALSE
  [9,]                FALSE    FALSE
 [10,]                FALSE    FALSE
 [11,]                FALSE    FALSE
 [12,]                FALSE    FALSE
 [13,]                FALSE    FALSE
 [14,]                FALSE    FALSE
 [15,]                FALSE    FALSE
 [16,]                FALSE    FALSE
 [17,]                FALSE    FALSE
 [18,]                FALSE    FALSE
 [19,]                FALSE    FALSE
 [20,]                FALSE    FALSE
 [21,]                FALSE    FALSE
 [22,]                FALSE    FALSE
 [23,]                FALSE    FALSE
 [24,]                FALSE    FALSE
 [25,]                FALSE    FALSE
 [26,]                FALSE    FALSE
 [27,]                FALSE    FALSE
 [28,]                FALSE    FALSE
 [29,]                FALSE    FALSE
 [30,]                FALSE    FALSE
 [31,]                FALSE    FALSE
 [32,]                FALSE    FALSE
 [33,]                FALSE    FALSE
 [34,]                FALSE    FALSE
 [35,]                FALSE    FALSE
 [36,]                FALSE    FALSE
 [37,]                FALSE    FALSE
 [38,]                FALSE    FALSE
 [39,]                FALSE    FALSE
 [40,]                FALSE    FALSE
 [41,]                FALSE    FALSE
 [42,]                FALSE    FALSE
 [43,]                FALSE    FALSE
 [44,]                FALSE    FALSE
 [45,]                FALSE    FALSE
```

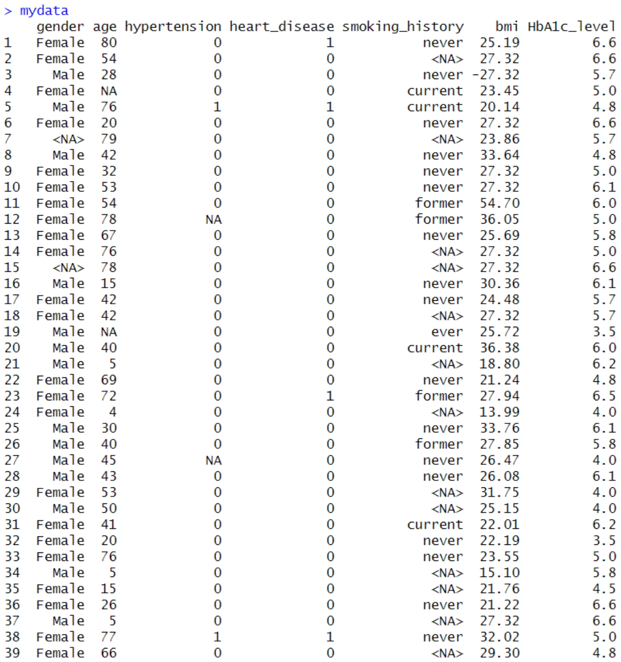


```
 [45,]                FALSE    FALSE
 [46,]                FALSE    FALSE
 [47,]                FALSE    FALSE
 [48,]                FALSE    FALSE
 [49,]                FALSE    FALSE
 [50,]                FALSE    FALSE
 [51,]                FALSE    FALSE
 [52,]                FALSE    FALSE
 [53,]                FALSE    FALSE
 [54,]                FALSE    FALSE
 [55,]                FALSE    FALSE
 [56,]                FALSE    FALSE
 [57,]                FALSE    FALSE
 [58,]                FALSE    FALSE
 [59,]                FALSE    FALSE
 [60,]                FALSE    FALSE
 [61,]                FALSE    FALSE
 [62,]                FALSE    FALSE
 [63,]                FALSE    FALSE
 [64,]                FALSE    FALSE
 [65,]                FALSE    FALSE
 [66,]                FALSE    FALSE
 [67,]                FALSE    FALSE
 [68,]                FALSE    FALSE
 [69,]                FALSE    FALSE
 [70,]                FALSE    FALSE
 [71,]                FALSE    FALSE
 [72,]                FALSE    FALSE
 [73,]                FALSE    FALSE
 [74,]                FALSE    FALSE
 [75,]                FALSE    FALSE
 [76,]                FALSE    FALSE
 [77,]                FALSE    FALSE
 [78,]                FALSE    FALSE
 [79,]                FALSE    FALSE
 [80,]                FALSE    FALSE
 [81,]                FALSE    FALSE
 [82,]                FALSE    FALSE
 [83,]                FALSE    FALSE
 [84,]                FALSE    FALSE
 [85,]                FALSE    FALSE
 [86,]                FALSE    FALSE
```

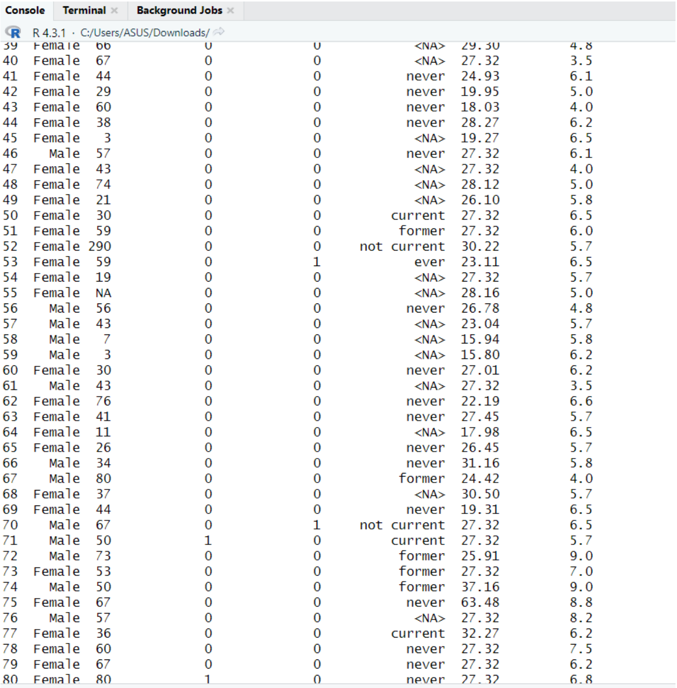


```
 [87,]                FALSE    FALSE
 [88,]                FALSE    FALSE
 [89,]                FALSE    FALSE
 [90,]                FALSE    FALSE
 [91,]                FALSE    FALSE
 [92,]                FALSE    FALSE
 [93,]                FALSE    FALSE
 [94,]                FALSE    FALSE
 [95,]                FALSE    FALSE
 [96,]                FALSE    FALSE
 [97,]                FALSE    FALSE
 [98,]                FALSE    FALSE
 [99,]                FALSE    FALSE
[100,]                FALSE    FALSE
[101,]                FALSE    FALSE
[102,]                FALSE    FALSE
[103,]                FALSE    FALSE
[104,]                FALSE    FALSE
[105,]                FALSE    FALSE
[106,]                FALSE    FALSE
[107,]                FALSE    FALSE
[108,]                FALSE    FALSE
[109,]                FALSE    FALSE
[110,]                FALSE    FALSE
[111,]                FALSE    FALSE
[112,]                FALSE    FALSE
[113,]                FALSE    FALSE
[114,]                FALSE    FALSE
[115,]                FALSE    FALSE
[116,]                FALSE    FALSE
[117,]                FALSE    FALSE
[118,]                FALSE    FALSE
[119,]                FALSE    FALSE
[120,]                FALSE    FALSE
> 
```

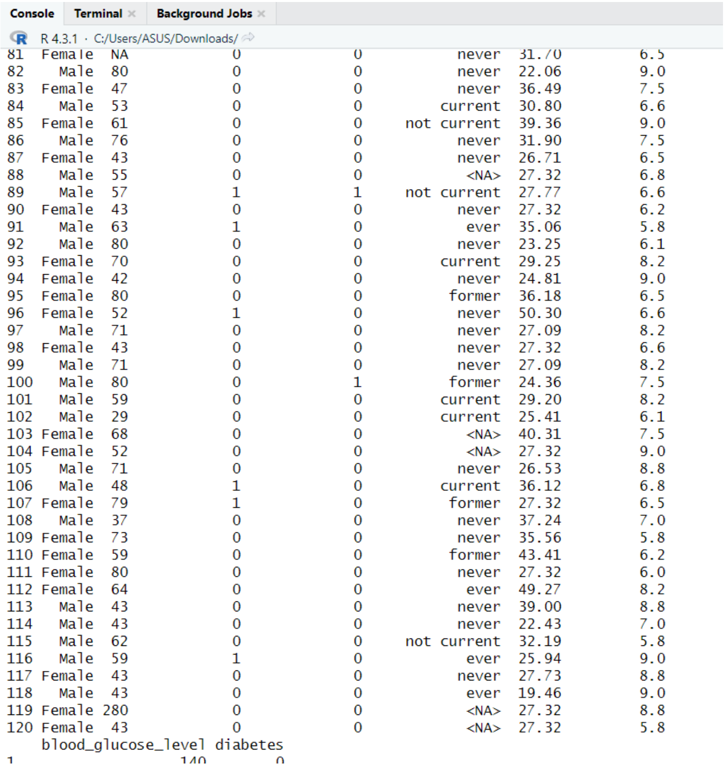
The mydata Output after replacing missing values with NA:



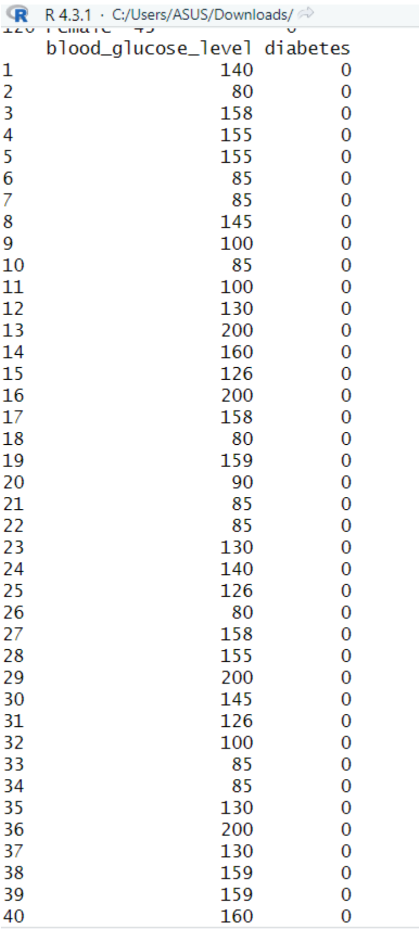
```
> mydata
    gender age hypertension heart_disease smoking_history    bmi HbA1c_level
1   Female  80            0             1           never  25.19         6.6
2   Female  54            0             0            <NA>  27.32         6.6
3     Male  28            0             0           never -27.32         5.7
4   Female  NA            0             0         current  23.45         5.0
5     Male  76            1             1         current  20.14         4.8
6   Female  20            0             0           never  27.32         6.6
7     <NA>  79            0             0            <NA>  23.86         5.7
8     Male  42            0             0           never  33.64         4.8
9   Female  32            0             0           never  27.32         5.0
10  Female  53            0             0           never  27.32         6.1
11  Female  54            0             0          former  54.70         6.0
12  Female  78           NA             0          former  36.05         5.0
13  Female  67            0             0           never  25.69         5.8
14  Female  76            0             0            <NA>  27.32         5.0
15    <NA>  78            0             0            <NA>  27.32         6.6
16    Male  15            0             0           never  30.36         6.1
17  Female  42            0             0           never  24.48         5.7
18  Female  42            0             0            <NA>  27.32         5.7
19    Male  NA            0             0            ever  25.72         3.5
20    Male  40            0             0         current  36.38         6.0
21    Male   5            0             0            <NA>  18.80         6.2
22  Female  69            0             0           never  21.24         4.8
23  Female  72            0             1          former  27.94         6.5
24  Female   4            0             0            <NA>  13.99         4.0
25    Male  30            0             0           never  33.76         6.1
26    Male  40            0             0          former  27.85         5.8
27    Male  45           NA             0           never  26.47         4.0
28    Male  43            0             0           never  26.08         6.1
29  Female  53            0             0            <NA>  31.75         4.0
30    Male  50            0             0            <NA>  25.15         4.0
31  Female  41            0             0         current  22.01         6.2
32  Female  20            0             0           never  22.19         3.5
33  Female  76            0             0           never  23.55         5.0
34    Male   5            0             0            <NA>  15.10         5.8
35  Female  15            0             0            <NA>  21.76         4.5
36  Female  26            0             0           never  21.22         6.6
37    Male   5            0             0            <NA>  27.32         6.6
38  Female  77            1             1           never  32.02         5.0
39  Female  66            0             0            <NA>  29.30         4.8
```



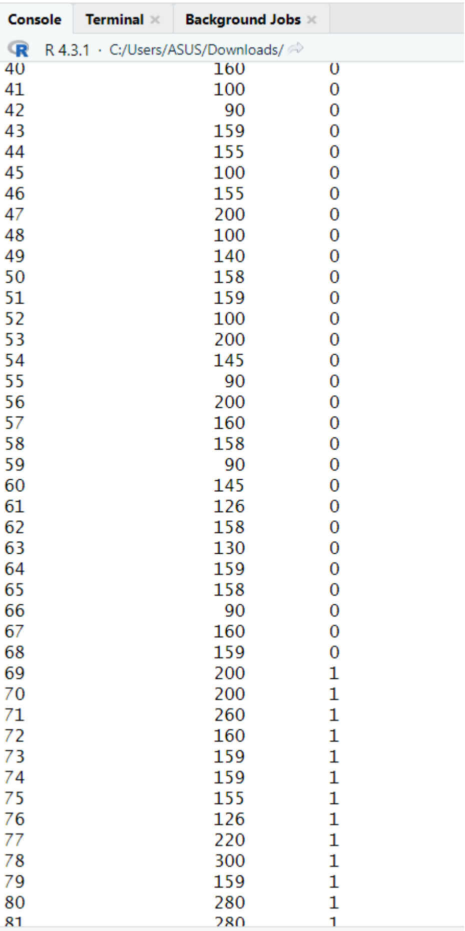
```
39  Female  66            0             0            <NA>  29.30         4.8
40  Female  67            0             0            <NA>  27.32         3.5
41  Female  44            0             0           never  24.93         6.1
42  Female  29            0             0           never  19.95         5.0
43  Female  60            0             0           never  18.03         4.0
44  Female  38            0             0           never  28.27         6.2
45  Female   3            0             0            <NA>  19.27         6.5
46    Male  57            0             0           never  27.32         6.1
47  Female  43            0             0            <NA>  27.32         4.0
48  Female  74            0             0            <NA>  28.12         5.0
49  Female  21            0             0            <NA>  26.10         5.8
50  Female  30            0             0         current  27.32         6.5
51  Female  59            0             0          former  27.32         6.0
52  Female 290            0             0     not current  30.22         5.7
53  Female  59            0             1            ever  23.11         6.5
54  Female  19            0             0            <NA>  27.32         5.7
55  Female  NA            0             0            <NA>  28.16         5.0
56    Male  56            0             0           never  26.78         4.8
57    Male  43            0             0            <NA>  23.04         5.7
58    Male   7            0             0            <NA>  15.94         5.8
59    Male   3            0             0            <NA>  15.80         6.2
60  Female  30            0             0           never  27.01         6.2
61    Male  43            0             0            <NA>  27.32         3.5
62  Female  76            0             0           never  22.19         6.6
63  Female  41            0             0           never  27.45         5.7
64  Female  11            0             0            <NA>  17.98         6.5
65  Female  26            0             0           never  26.45         5.7
66    Male  34            0             0           never  31.16         5.8
67    Male  80            0             0          former  24.42         4.0
68  Female  37            0             0            <NA>  30.50         5.7
69  Female  44            0             0           never  19.31         6.5
70    Male  67            0             1     not current  27.32         6.5
71    Male  50            1             0         current  27.32         5.7
72    Male  73            0             0          former  25.91         9.0
73  Female  53            0             0          former  27.32         7.0
74    Male  50            0             0          former  37.16         9.0
75  Female  67            0             0           never  63.48         8.8
76    Male  57            0             0            <NA>  27.32         8.2
77  Female  36            0             0         current  32.27         6.2
78  Female  60            0             0           never  27.32         7.5
79  Female  67            0             0           never  27.32         6.2
80  Female  80            1             0           never  27.32         6.8
```



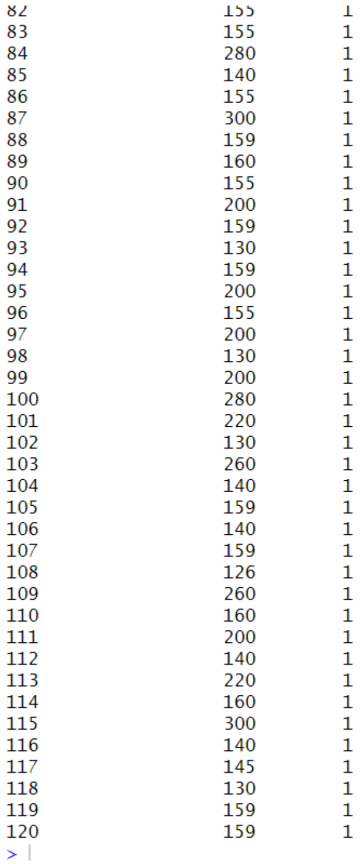
```
81  Female  NA            0             0           never  31.70         6.5
82    Male  80            0             0           never  22.06         9.0
83  Female  47            0             0           never  36.49         7.5
84    Male  53            0             0         current  30.80         6.6
85  Female  61            0             0     not current  39.36         9.0
86    Male  76            0             0           never  31.90         7.5
87  Female  43            0             0           never  26.71         6.5
88    Male  55            0             0            <NA>  27.32         6.8
89    Male  57            1             1     not current  27.77         6.6
90  Female  43            0             0           never  27.32         6.2
91    Male  63            1             0            ever  35.06         5.8
92    Male  80            0             0           never  23.25         6.1
93  Female  70            0             0         current  29.25         8.2
94  Female  42            0             0           never  24.81         9.0
95  Female  80            0             0          former  36.18         6.5
96  Female  52            1             0           never  50.30         6.6
97    Male  71            0             0           never  27.09         8.2
98  Female  43            0             0           never  27.32         6.6
99    Male  71            0             0           never  27.09         8.2
100   Male  80            0             1          former  24.36         7.5
101   Male  59            0             0         current  29.20         8.2
102   Male  29            0             0         current  25.41         6.1
103 Female  68            0             0            <NA>  40.31         7.5
104 Female  52            0             0            <NA>  27.32         9.0
105   Male  71            0             0           never  26.53         8.8
106   Male  48            1             0         current  36.12         6.8
107 Female  79            1             0          former  27.32         6.5
108   Male  37            0             0           never  37.24         7.0
109 Female  73            0             0           never  35.56         5.8
110 Female  59            0             0          former  43.41         6.2
111 Female  80            0             0           never  27.32         6.0
112 Female  64            0             0            ever  49.27         8.2
113   Male  43            0             0           never  39.00         8.8
114   Male  43            0             0           never  22.43         7.0
115   Male  62            0             0     not current  32.19         5.8
116   Male  59            1             0            ever  25.94         9.0
117 Female  43            0             0           never  27.73         8.8
118   Male  43            0             0            ever  19.46         9.0
119 Female 280            0             0            <NA>  27.32         8.8
120 Female  43            0             0            <NA>  27.32         5.8
    blood_glucose_level diabetes
1                   140        0
```



```
    blood_glucose_level diabetes
1                   140        0
2                    80        0
3                   158        0
4                   155        0
5                   155        0
6                    85        0
7                    85        0
8                   145        0
9                   100        0
10                   85        0
11                  100        0
12                  130        0
13                  200        0
14                  160        0
15                  126        0
16                  200        0
17                  158        0
18                   80        0
19                  159        0
20                   90        0
21                   85        0
22                   85        0
23                  130        0
24                  140        0
25                  126        0
26                   80        0
27                  158        0
28                  155        0
29                  200        0
30                  145        0
31                  126        0
32                  100        0
33                   85        0
34                   85        0
35                  130        0
36                  200        0
37                  130        0
38                  159        0
39                  159        0
40                  160        0
```

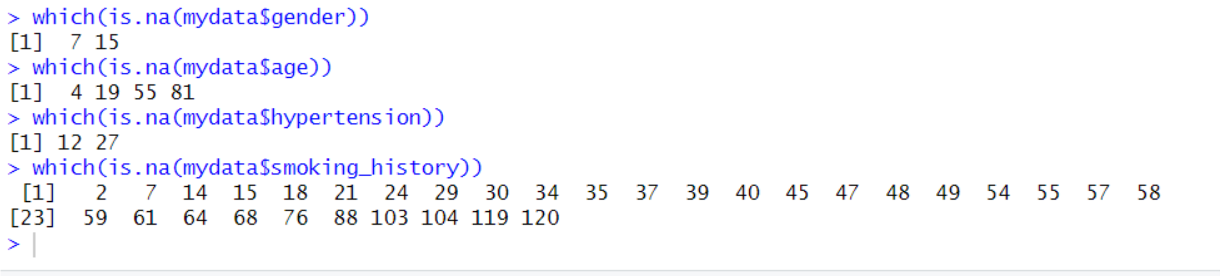


```
40                  160        0
41                  100        0
42                   90        0
43                  159        0
44                  155        0
45                  100        0
46                  155        0
47                  200        0
48                  100        0
49                  140        0
50                  158        0
51                  159        0
52                  100        0
53                  200        0
54                  145        0
55                   90        0
56                  200        0
57                  160        0
58                  158        0
59                   90        0
60                  145        0
61                  126        0
62                  158        0
63                  130        0
64                  159        0
65                  158        0
66                   90        0
67                  160        0
68                  159        0
69                  200        1
70                  200        1
71                  260        1
72                  160        1
73                  159        1
74                  159        1
75                  155        1
76                  126        1
77                  220        1
78                  300        1
79                  159        1
80                  280        1
81                  280        1
```



```
82                  155        1
83                  155        1
84                  280        1
85                  140        1
86                  155        1
87                  300        1
88                  159        1
89                  160        1
90                  155        1
91                  200        1
92                  159        1
93                  130        1
94                  159        1
95                  200        1
96                  155        1
97                  200        1
98                  130        1
99                  200        1
100                 280        1
101                 220        1
102                 130        1
103                 260        1
104                 140        1
105                 159        1
106                 140        1
107                 159        1
108                 126        1
109                 260        1
110                 160        1
111                 200        1
112                 140        1
113                 220        1
114                 160        1
115                 300        1
116                 140        1
117                 145        1
118                 130        1
119                 159        1
120                 159        1
>
```

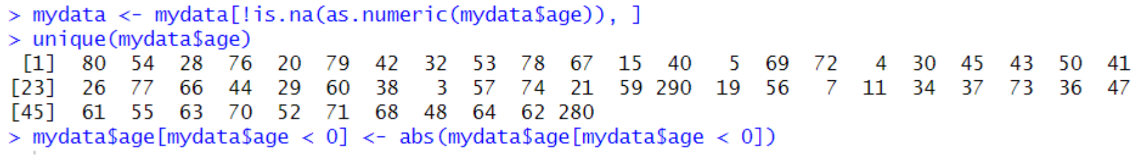
Also, there are some missing values based on specific columns and rows. The outputs are:



```
> which(is.na(mydata$gender))
[1]  7 15
> which(is.na(mydata$age))
[1]  4 19 55 81
> which(is.na(mydata$hypertension))
[1] 12 27
> which(is.na(mydata$smoking_history))
 [1]   2   7  14  15  18  21  24  29  30  34  35  37  39  40  45  47  48  49  54  55  57  58
[23]  59  61  64  68  76  88 103 104 119 120
> |
```

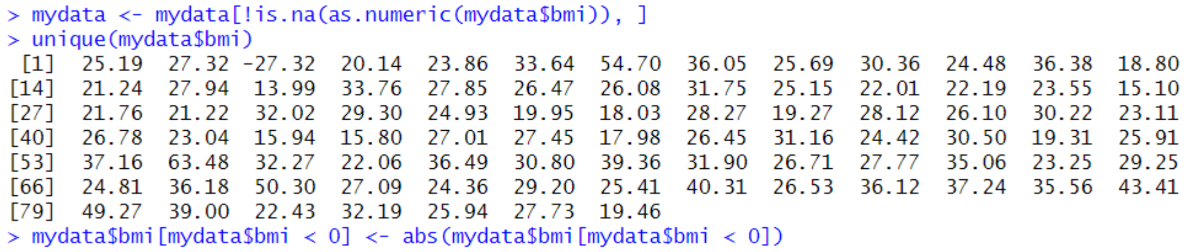
Data Exploration:

Missing age values, and any negative ages have been converted to their positive counterparts (absolute values):



```
> mydata <- mydata[!is.na(as.numeric(mydata$age)), ]
> unique(mydata$age)
 [1]  80  54  28  76  20  79  42  32  53  78  67  15  40   5  69  72   4  30  45  43  50  41
[23]  26  77  66  44  29  60  38   3  57  74  21  59 290  19  56   7  11  34  37  73  36  47
[45]  61  55  63  70  52  71  68  48  64  62 280
> mydata$age[mydata$age < 0] <- abs(mydata$age[mydata$age < 0])
```

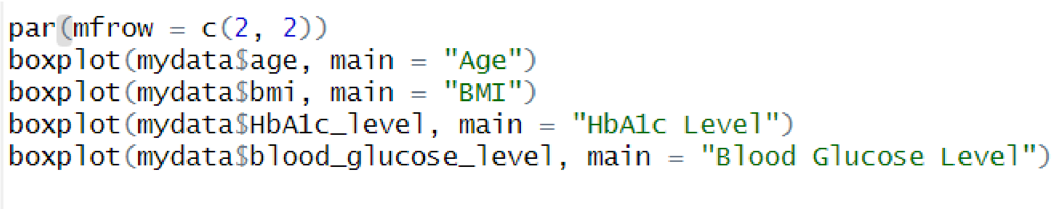
Missing bmi values, and any negative ages have been converted to their positive counterparts (absolute values):



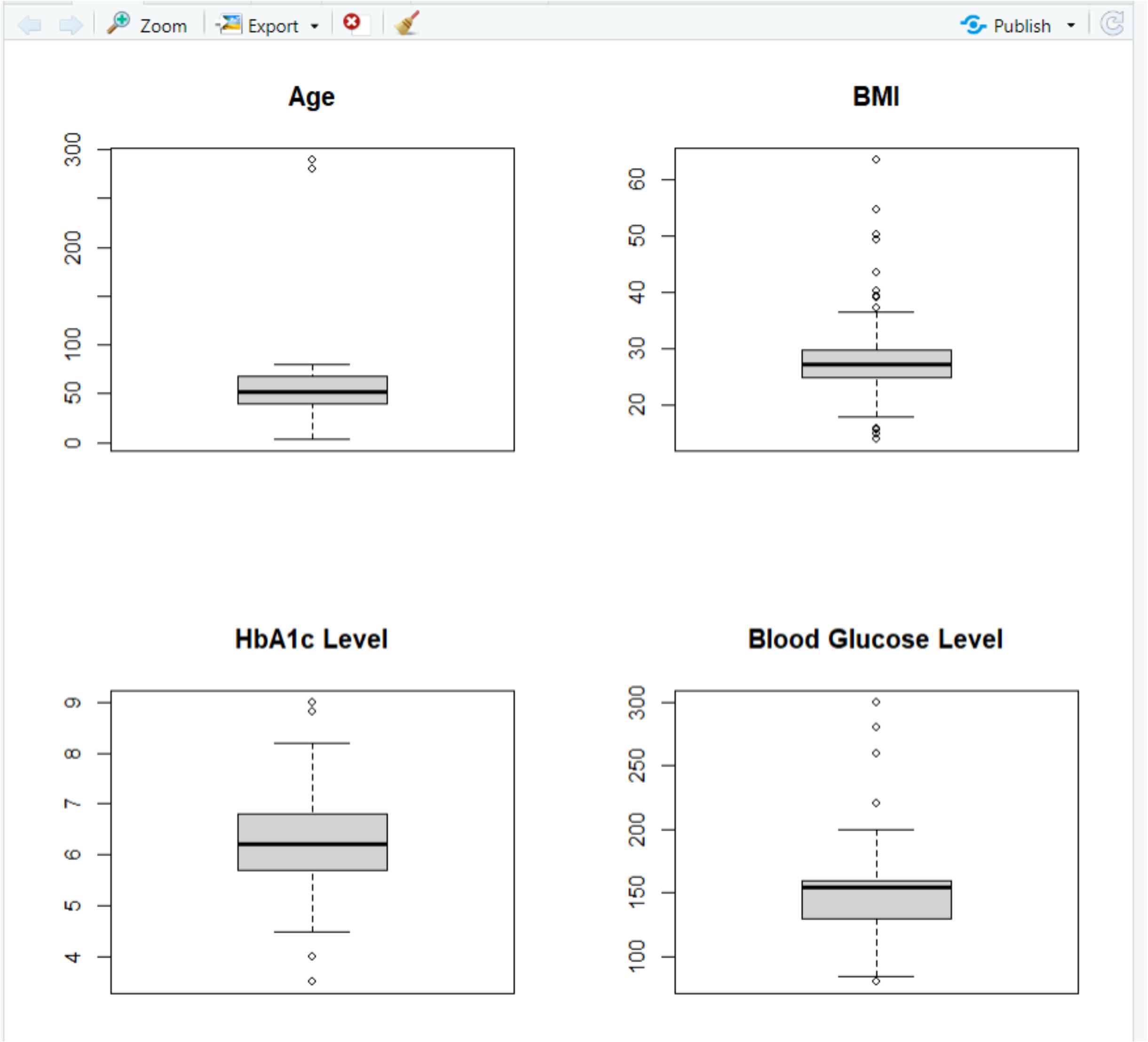
```
> mydata <- mydata[!is.na(as.numeric(mydata$bmi)), ]
> unique(mydata$bmi)
 [1]  25.19  27.32 -27.32  20.14  23.86  33.64  54.70  36.05  25.69  30.36  24.48  36.38  18.80
[14]  21.24  27.94  13.99  33.76  27.85  26.47  26.08  31.75  25.15  22.01  22.19  23.55  15.10
[27]  21.76  21.22  32.02  29.30  24.93  19.95  18.03  28.27  19.27  28.12  26.10  30.22  23.11
[40]  26.78  23.04  15.94  15.80  27.01  27.45  17.98  26.45  31.16  24.42  30.50  19.31  25.91
[53]  37.16  63.48  32.27  22.06  36.49  30.80  39.36  31.90  26.71  27.77  35.06  23.25  29.25
[66]  24.81  36.18  50.30  27.09  24.36  29.20  25.41  40.31  26.53  36.12  37.24  35.56  43.41
[79]  49.27  39.00  22.43  32.19  25.94  27.73  19.46
> mydata$bmi[mydata$bmi < 0] <- abs(mydata$bmi[mydata$bmi < 0])
```

Handling potential outliers using box plots:

This will create a 2x2 grid of boxplots to visualize the distribution and spread of data for four different variables in the mydata dataframe. The output:



```
par(mfrow = c(2, 2))
boxplot(mydata$age, main = "Age")
boxplot(mydata$bmi, main = "BMI")
boxplot(mydata$HbA1c_level, main = "HbA1c Level")
boxplot(mydata$blood_glucose_level, main = "Blood Glucose Level")
```

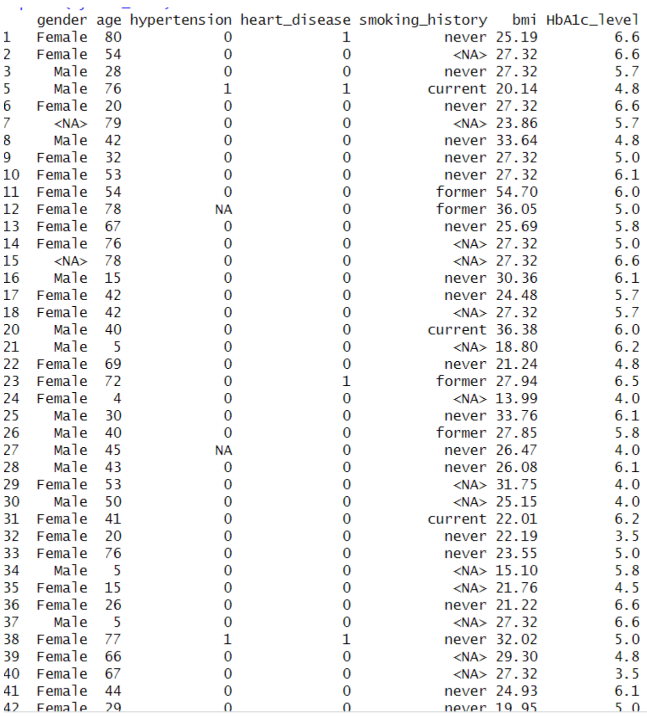


From the above boxplots it is noticible that there are few outliers present in the dataset.

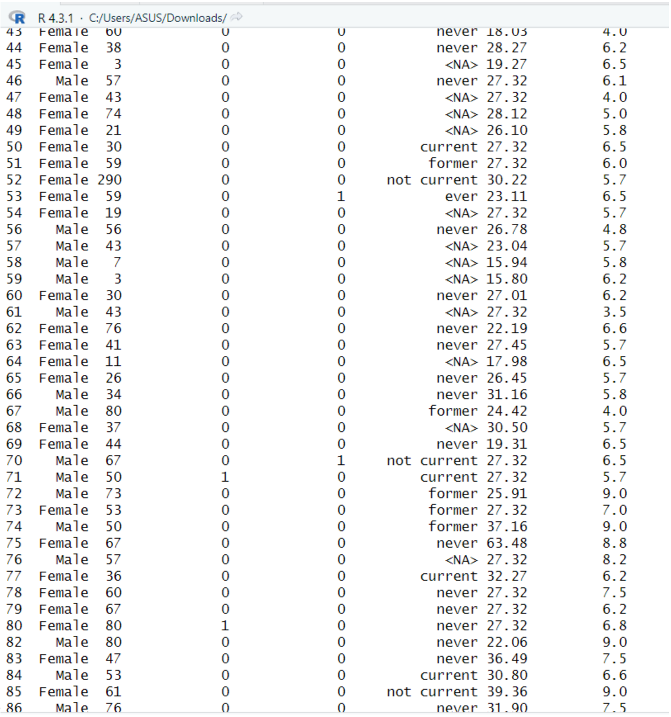
Dealing with missing values:

We replaced the main data frame “mydata” each time whenever we performed a specific task on the dataframe so that there would be backup data as mydata. Replacing missing values with the mean value.The output:

```
mydata_mean <- mydata
mean_age<- mean(mydata_mean$age, na.rm = TRUE)
mydata_mean$age[is.na(mydata_mean$age)] <- mean_age
print(mydata_mean)
```

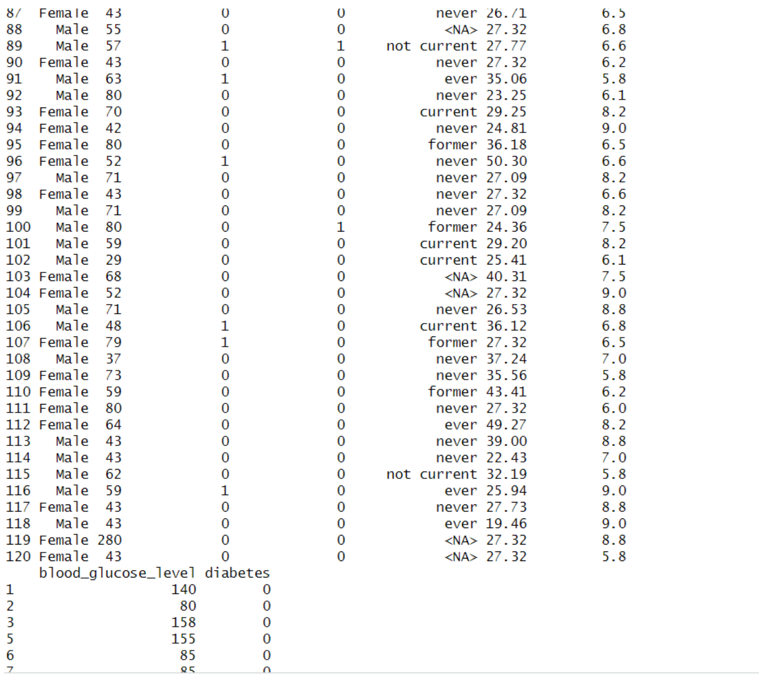


```
    gender age hypertension heart_disease smoking_history   bmi HbA1c_level
1   Female  80            0             1           never 25.19         6.6
2   Female  54            0             0            <NA> 27.32         6.6
3     Male  28            0             0           never 27.32         5.7
5     Male  76            1             1         current 20.14         4.8
6   Female  20            0             0           never 27.32         6.6
7     <NA>  79            0             0            <NA> 23.86         5.7
8     Male  42            0             0           never 33.64         4.8
9   Female  32            0             0           never 27.32         5.0
10  Female  53            0             0           never 27.32         6.1
11  Female  54            0             0          former 54.70         6.0
12  Female  78           NA             0          former 36.05         5.0
13  Female  67            0             0           never 25.69         5.8
14  Female  76            0             0            <NA> 27.32         5.0
15    <NA>  78            0             0            <NA> 27.32         6.6
16    Male  15            0             0           never 30.36         6.1
17  Female  42            0             0           never 24.48         5.7
18  Female  42            0             0            <NA> 27.32         5.7
20    Male  40            0             0         current 36.38         6.0
21    Male   5            0             0            <NA> 18.80         6.2
22  Female  69            0             0           never 21.24         4.8
23  Female  72            0             1          former 27.94         6.5
24  Female   4            0             0            <NA> 13.99         4.0
25    Male  30            0             0           never 33.76         6.1
26    Male  40            0             0          former 27.85         5.8
27    Male  45           NA             0           never 26.47         4.0
28    Male  43            0             0           never 26.08         6.1
29  Female  53            0             0            <NA> 31.75         4.0
30    Male  50            0             0            <NA> 25.15         4.0
31  Female  41            0             0         current 22.01         6.2
32  Female  20            0             0           never 22.19         3.5
33  Female  76            0             0           never 23.55         5.0
34    Male   5            0             0            <NA> 15.10         5.8
35  Female  15            0             0            <NA> 21.76         4.5
36  Female  26            0             0           never 21.22         6.6
37    Male   5            0             0            <NA> 27.32         6.6
38  Female  77            1             1           never 32.02         5.0
39  Female  66            0             0            <NA> 29.30         4.8
40  Female  67            0             0            <NA> 27.32         3.5
41  Female  44            0             0           never 24.93         6.1
42  Female  29            0             0           never 19.95         5.0
```

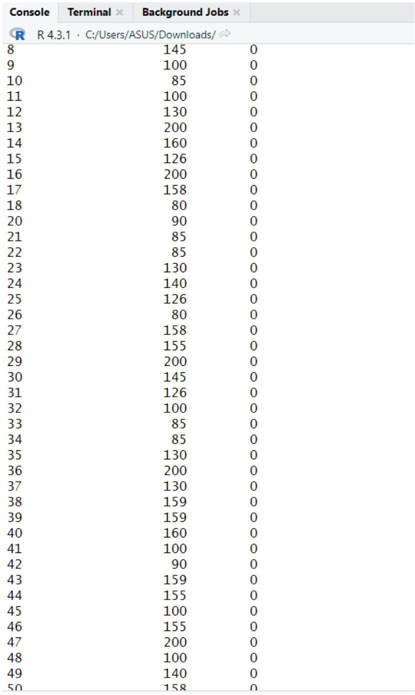


R 4.3.1 · C:/Users/ASUS/Downloads/

```
43  Female  60            0             0           never 18.03         4.0
44  Female  38            0             0           never 28.27         6.2
45  Female   3            0             0            <NA> 19.27         6.5
46    Male  57            0             0           never 27.32         6.1
47  Female  43            0             0            <NA> 27.32         4.0
48  Female  74            0             0            <NA> 28.12         5.0
49  Female  21            0             0            <NA> 26.10         5.8
50  Female  30            0             0         current 27.32         6.5
51  Female  59            0             0          former 27.32         6.0
52  Female 290            0             0     not current 30.22         5.7
53  Female  59            0             1            ever 23.11         6.5
54  Female  19            0             0            <NA> 27.32         5.7
56    Male  56            0             0           never 26.78         4.8
57    Male  43            0             0            <NA> 23.04         5.7
58    Male   7            0             0            <NA> 15.94         5.8
59    Male   3            0             0            <NA> 15.80         6.2
60  Female  30            0             0           never 27.01         6.2
61    Male  43            0             0            <NA> 27.32         3.5
62  Female  76            0             0           never 22.19         6.6
63  Female  41            0             0           never 27.45         5.7
64  Female  11            0             0            <NA> 17.98         6.5
65  Female  26            0             0           never 26.45         5.7
66    Male  34            0             0           never 31.16         5.8
67    Male  80            0             0          former 24.42         4.0
68  Female  37            0             0            <NA> 30.50         5.7
69  Female  44            0             0           never 19.31         6.5
70    Male  67            0             1     not current 27.32         6.5
71    Male  50            1             0         current 27.32         5.7
72    Male  73            0             0          former 25.91         9.0
73  Female  53            0             0          former 27.32         7.0
74    Male  50            0             0          former 37.16         9.0
75  Female  67            0             0           never 63.48         8.8
76    Male  57            0             0            <NA> 27.32         8.2
77  Female  36            0             0         current 32.27         6.2
78  Female  60            0             0           never 27.32         7.5
79  Female  67            0             0           never 27.32         6.2
80  Female  80            1             0           never 27.32         6.8
82    Male  80            0             0           never 22.06         9.0
83  Female  47            0             0           never 36.49         7.5
84    Male  53            0             0         current 30.80         6.6
85  Female  61            0             0     not current 39.36         9.0
86    Male  76            0             0           never 31.90         7.5
```

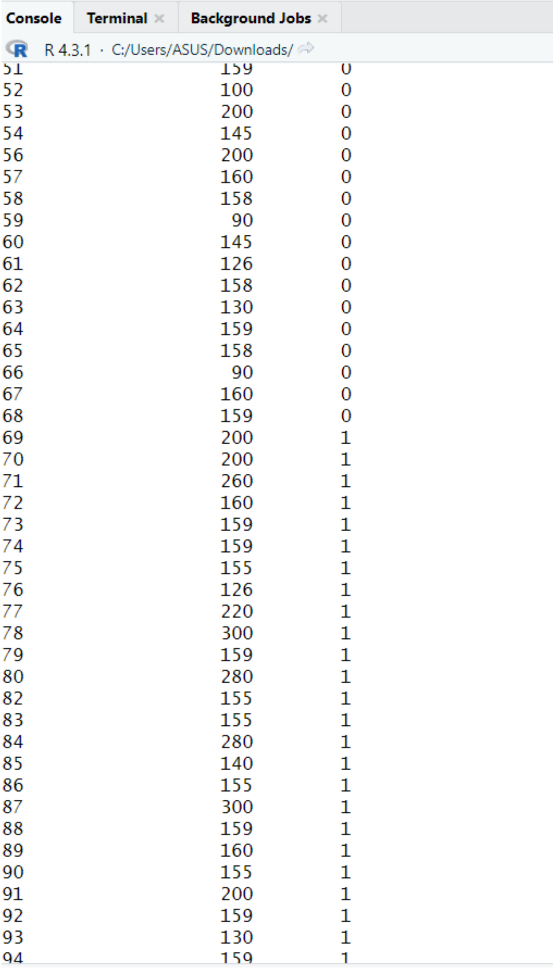


```
87  Female  43            0             0           never 26.71         6.5
88    Male  55            0             0            <NA> 27.32         6.8
89    Male  57            1             1     not current 27.77         6.6
90  Female  43            0             0           never 27.32         6.2
91    Male  63            1             0            ever 35.06         5.8
92    Male  80            0             0           never 23.25         6.1
93  Female  70            0             0         current 29.25         8.2
94  Female  42            0             0           never 24.81         9.0
95  Female  80            0             0          former 36.18         6.5
96  Female  52            1             0           never 50.30         6.6
97    Male  71            0             0           never 27.09         8.2
98  Female  43            0             0           never 27.32         6.6
99    Male  71            0             0           never 27.09         8.2
100   Male  80            0             1          former 24.36         7.5
101   Male  59            0             0         current 29.20         8.2
102   Male  29            0             0         current 25.41         6.1
103 Female  68            0             0            <NA> 40.31         7.5
104 Female  52            0             0            <NA> 27.32         9.0
105   Male  71            0             0           never 26.53         8.8
106   Male  48            1             0         current 36.12         6.8
107 Female  79            1             0          former 27.32         6.5
108   Male  37            0             0           never 37.24         7.0
109 Female  73            0             0           never 35.56         5.8
110 Female  59            0             0          former 43.41         6.2
111 Female  80            0             0           never 27.32         6.0
112 Female  64            0             0            ever 49.27         8.2
113   Male  43            0             0           never 39.00         8.8
114   Male  43            0             0           never 22.43         7.0
115   Male  62            0             0     not current 32.19         5.8
116   Male  59            1             0            ever 25.94         9.0
117 Female  43            0             0           never 27.73         8.8
118   Male  43            0             0            ever 19.46         9.0
119 Female 280            0             0            <NA> 27.32         8.8
120 Female  43            0             0            <NA> 27.32         5.8
    blood_glucose_level diabetes
1                   140        0
2                    80        0
3                   158        0
5                   155        0
6                    85        0
7                    85        0
```

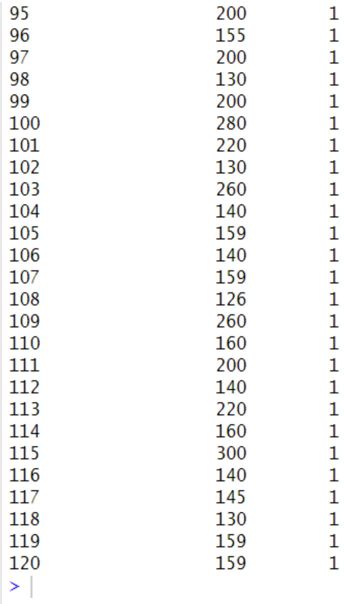


R 4.3.1 · C:/Users/ASUS/Downloads/

```
8                   145        0
9                   100        0
10                   85        0
11                  100        0
12                  130        0
13                  200        0
14                  160        0
15                  126        0
16                  200        0
17                  158        0
18                   80        0
20                   90        0
21                   85        0
22                   85        0
23                  130        0
24                  140        0
25                  126        0
26                   80        0
27                  158        0
28                  155        0
29                  200        0
30                  145        0
31                  126        0
32                  100        0
33                   85        0
34                   85        0
35                  130        0
36                  200        0
37                  130        0
38                  159        0
39                  159        0
40                  160        0
41                  100        0
42                   90        0
43                  159        0
44                  155        0
45                  100        0
46                  155        0
47                  200        0
48                  100        0
49                  140        0
50                  158        0
```

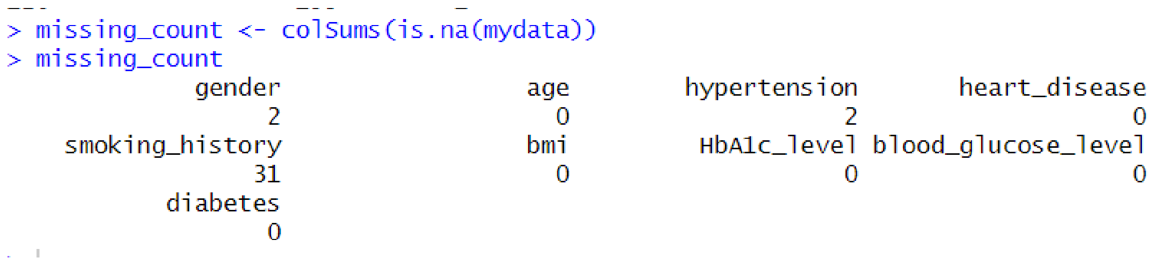


```
51                    159        0
52                    100        0
53                    200        0
54                    145        0
56                    200        0
57                    160        0
58                    158        0
59                     90        0
60                    145        0
61                    126        0
62                    158        0
63                    130        0
64                    159        0
65                    158        0
66                     90        0
67                    160        0
68                    159        0
69                    200        1
70                    200        1
71                    260        1
72                    160        1
73                    159        1
74                    159        1
75                    155        1
76                    126        1
77                    220        1
78                    300        1
79                    159        1
80                    280        1
82                    155        1
83                    155        1
84                    280        1
85                    140        1
86                    155        1
87                    300        1
88                    159        1
89                    160        1
90                    155        1
91                    200        1
92                    159        1
93                    130        1
94                    159        1
```



```
95                    200        1
96                    155        1
97                    200        1
98                    130        1
99                    200        1
100                   280        1
101                   220        1
102                   130        1
103                   260        1
104                   140        1
105                   159        1
106                   140        1
107                   159        1
108                   126        1
109                   260        1
110                   160        1
111                   200        1
112                   140        1
113                   220        1
114                   160        1
115                   300        1
116                   140        1
117                   145        1
118                   130        1
119                   159        1
120                   159        1
> 
```

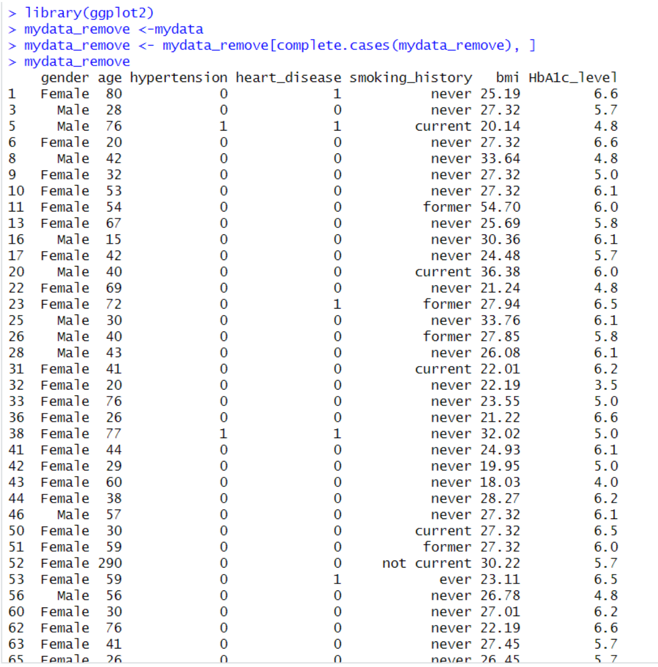
Check for missing values and count:



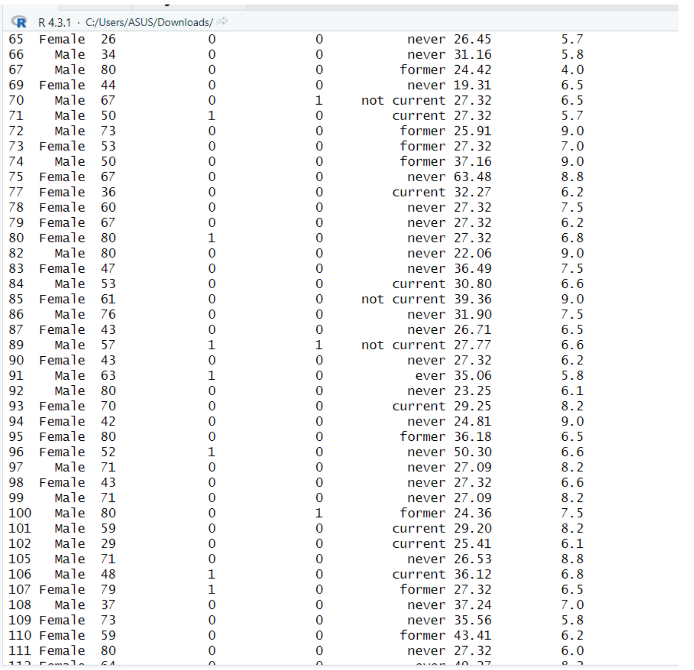
```
> missing_count <- colSums(is.na(mydata))
> missing_count
           gender               age      hypertension     heart_disease 
                2                 0                 2                 0 
  smoking_history               bmi       HbA1c_level blood_glucose_level 
               31                 0                 0                 0 
         diabetes 
                0 
```

Data Visualization:

ggplot2 library, which was used for creating data visualizations. We replaced mydata data frame with mydata_remove. Then we removed rows with missing values, ensuring that the dataset used for analysis and visualization is complete and doesn't introduce bias or inaccuracies due to missing data. The output:

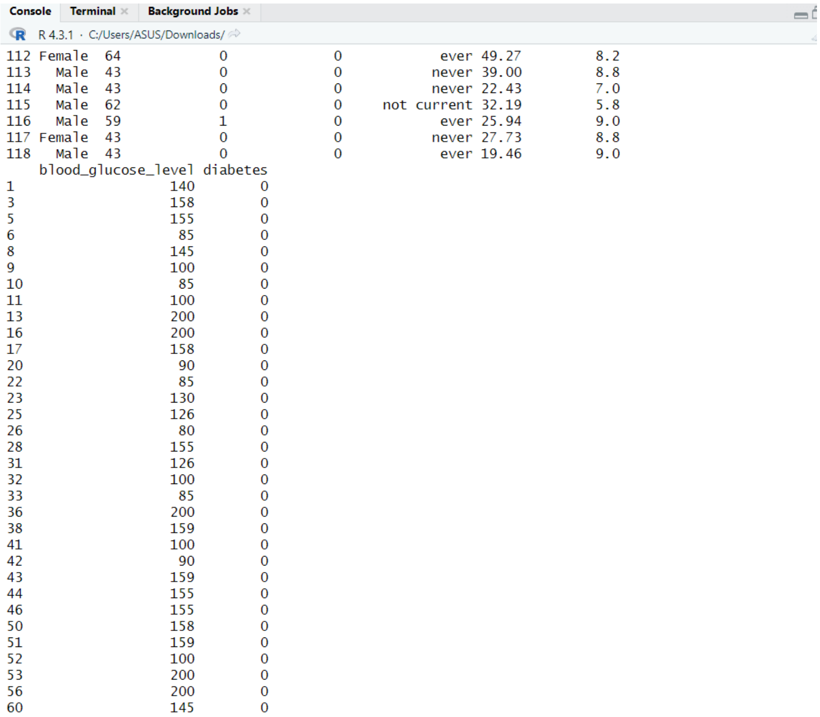


```
> library(ggplot2)
> mydata_remove <-mydata
> mydata_remove <- mydata_remove[complete.cases(mydata_remove), ]
> mydata_remove
    gender age hypertension heart_disease smoking_history   bmi HbA1c_level
1   Female  80            0             1           never 25.19         6.6
3     Male  28            0             0           never 27.32         5.7
5     Male  76            1             1         current 20.14         4.8
6   Female  20            0             0           never 27.32         6.6
8     Male  42            0             0           never 33.64         4.8
9   Female  32            0             0           never 27.32         5.0
10  Female  53            0             0           never 27.32         6.1
11  Female  54            0             0          former 54.70         6.0
13  Female  67            0             0           never 25.69         5.8
16    Male  15            0             0           never 30.36         6.1
17  Female  42            0             0           never 24.48         5.7
20    Male  40            0             0         current 36.38         6.0
22  Female  69            0             0           never 21.24         4.8
23  Female  72            0             1          former 27.94         6.5
25    Male  30            0             0           never 33.76         6.1
26    Male  40            0             0          former 27.85         5.8
28    Male  43            0             0           never 26.08         6.1
31  Female  41            0             0         current 22.01         6.2
32  Female  20            0             0           never 22.19         3.5
33  Female  76            0             0           never 23.55         5.0
36  Female  26            0             0           never 21.22         6.6
38  Female  77            1             1           never 32.02         5.0
41  Female  44            0             0           never 24.93         6.1
42  Female  29            0             0           never 19.95         5.0
43  Female  60            0             0           never 18.03         4.0
44  Female  38            0             0           never 28.27         6.2
46    Male  57            0             0           never 27.32         6.1
50  Female  30            0             0         current 27.32         6.5
51  Female  59            0             0          former 27.32         6.0
52  Female 290            0             0     not current 30.22         5.7
53  Female  59            0             1            ever 23.11         6.5
56    Male  56            0             0           never 26.78         4.8
60  Female  30            0             0           never 27.01         6.2
62  Female  76            0             0           never 22.19         6.6
63  Female  41            0             0           never 27.45         5.7
65  Female  26            0             0           never 26.45         5.7
```



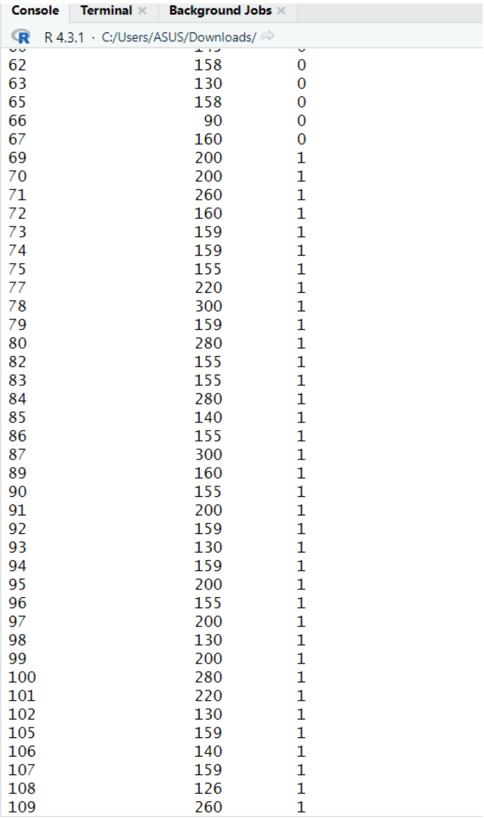
R 4.3.1 · C:/Users/ASUS/Downloads/

```
65  Female  26            0             0           never 26.45         5.7
66    Male  34            0             0           never 31.16         5.8
67    Male  80            0             0          former 24.42         4.0
69  Female  44            0             0           never 19.31         6.5
70    Male  67            0             1     not current 27.32         6.5
71    Male  50            1             0         current 27.32         5.7
72    Male  73            0             0          former 25.91         9.0
73  Female  53            0             0          former 27.32         7.0
74    Male  50            0             0          former 37.16         9.0
75  Female  67            0             0           never 63.48         8.8
77  Female  36            0             0         current 32.27         6.2
78  Female  60            0             0           never 27.32         7.5
79  Female  67            0             0           never 27.32         6.2
80  Female  80            1             0           never 27.32         6.8
82    Male  80            0             0           never 22.06         9.0
83  Female  47            0             0           never 36.49         7.5
84    Male  53            0             0         current 30.80         6.6
85  Female  61            0             0     not current 39.36         9.0
86    Male  76            0             0           never 31.90         7.5
87  Female  43            0             0           never 26.71         6.5
89    Male  57            1             1     not current 27.77         6.6
90  Female  43            0             0           never 27.32         6.2
91    Male  63            1             0            ever 35.06         5.8
92    Male  80            0             0           never 23.25         6.1
93  Female  70            0             0         current 29.25         8.2
94  Female  42            0             0           never 24.81         9.0
95  Female  80            0             0          former 36.18         6.5
96  Female  52            1             0           never 50.30         6.6
97    Male  71            0             0           never 27.09         8.2
98  Female  43            0             0           never 27.32         6.6
99    Male  71            0             0           never 27.09         8.2
100   Male  80            0             1          former 24.36         7.5
101   Male  59            0             0         current 29.20         8.2
102   Male  29            0             0         current 25.41         6.1
105   Male  71            0             0           never 26.53         8.8
106   Male  48            1             0         current 36.12         6.8
107 Female  79            1             0          former 27.32         6.5
108   Male  37            0             0           never 37.24         7.0
109 Female  73            0             0           never 35.56         5.8
110 Female  59            0             0          former 43.41         6.2
111 Female  80            0             0           never 27.32         6.0
112 Female  64            0             0            ever 49.27         8.2
```



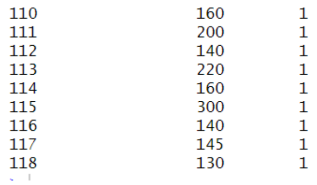
R 4.3.1 · C:/Users/ASUS/Downloads/

```
112 Female  64            0             0            ever 49.27         8.2
113   Male  43            0             0           never 39.00         8.8
114   Male  43            0             0           never 22.43         7.0
115   Male  62            0             0     not current 32.19         5.8
116   Male  59            1             0            ever 25.94         9.0
117 Female  43            0             0           never 27.73         8.8
118   Male  43            0             0            ever 19.46         9.0
    blood_glucose_level diabetes
1                   140        0
3                   158        0
5                   155        0
6                    85        0
8                   145        0
9                   100        0
10                   85        0
11                  100        0
13                  200        0
16                  200        0
17                  158        0
20                   90        0
22                   85        0
23                  130        0
25                  126        0
26                   80        0
28                  155        0
31                  126        0
32                  100        0
33                   85        0
36                  200        0
38                  159        0
41                  100        0
42                   90        0
43                  159        0
44                  155        0
46                  155        0
50                  158        0
51                  159        0
52                  100        0
53                  200        0
56                  200        0
60                  145        0
```



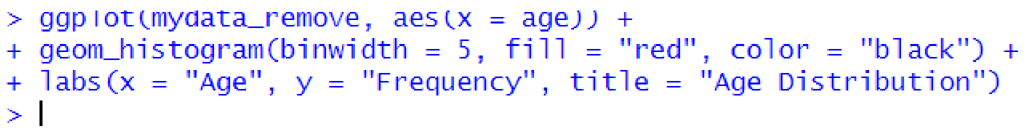
R 4.3.1 · C:/Users/ASUS/Downloads/

```
62                  158        0
63                  130        0
65                  158        0
66                   90        0
67                  160        0
69                  200        1
70                  200        1
71                  260        1
72                  160        1
73                  159        1
74                  159        1
75                  155        1
77                  220        1
78                  300        1
79                  159        1
80                  280        1
82                  155        1
83                  155        1
84                  280        1
85                  140        1
86                  155        1
87                  300        1
89                  160        1
90                  155        1
91                  200        1
92                  159        1
93                  130        1
94                  159        1
95                  200        1
96                  155        1
97                  200        1
98                  130        1
99                  200        1
100                 280        1
101                 220        1
102                 130        1
105                 159        1
106                 140        1
107                 159        1
108                 126        1
109                 260        1
```

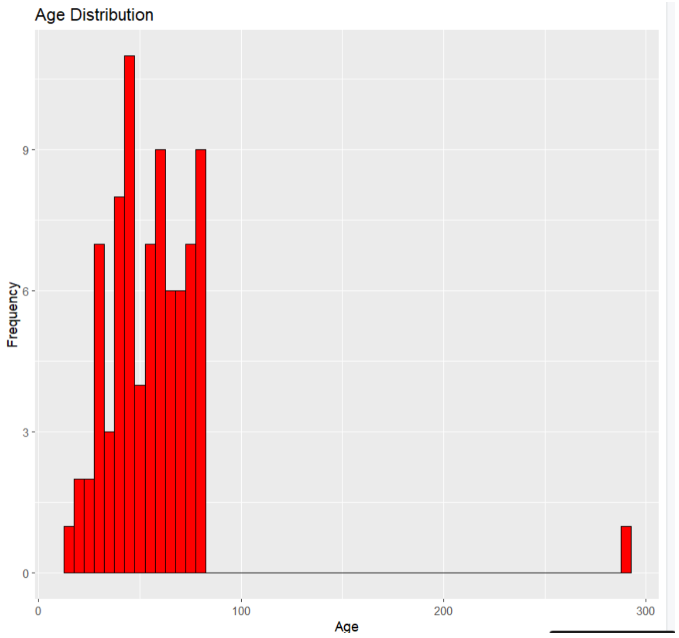


```
110                    160          1
111                    200          1
112                    140          1
113                    220          1
114                    160          1
115                    300          1
116                    140          1
117                    145          1
118                    130          1
```

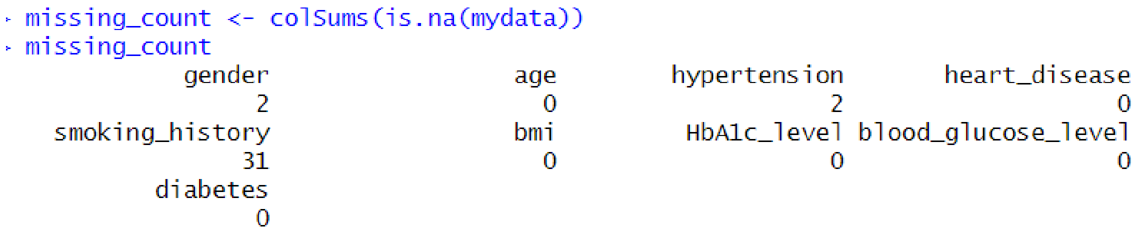
To indentify pattern trend in the dataset .The output:



```
> ggplot(mydata_remove, aes(x = age)) +
+ geom_histogram(binwidth = 5, fill = "red", color = "black") +
+ labs(x = "Age", y = "Frequency", title = "Age Distribution")
> |
```

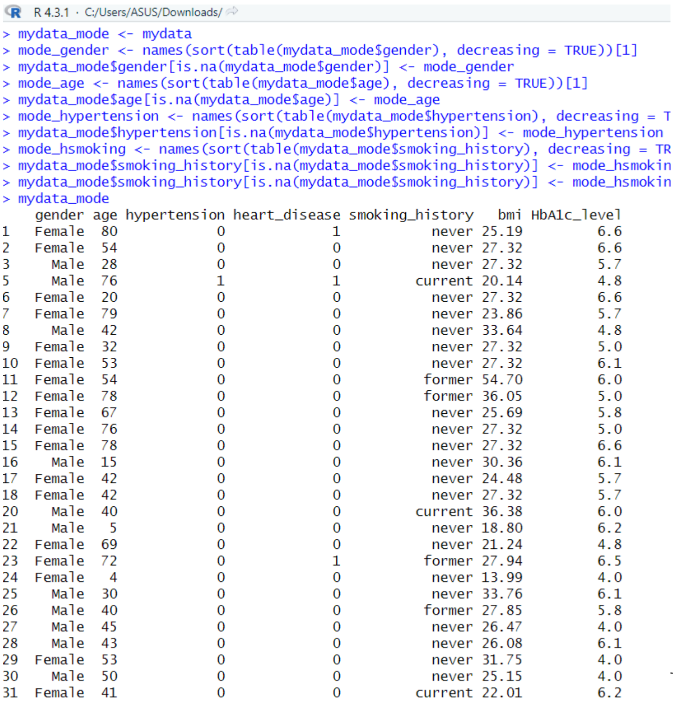


Checking for missing values and count.The output:

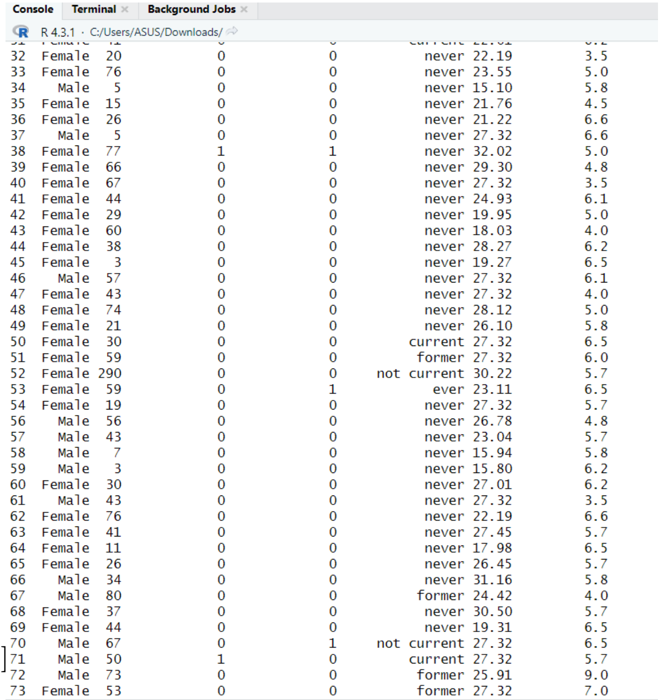


```
> missing_count <- colSums(is.na(mydata))
> missing_count
         gender                 age        hypertension       heart_disease
              2                   0                   2                   0
smoking_history                 bmi         HbA1c_level blood_glucose_level
             31                   0                   0                   0
       diabetes
              0
```

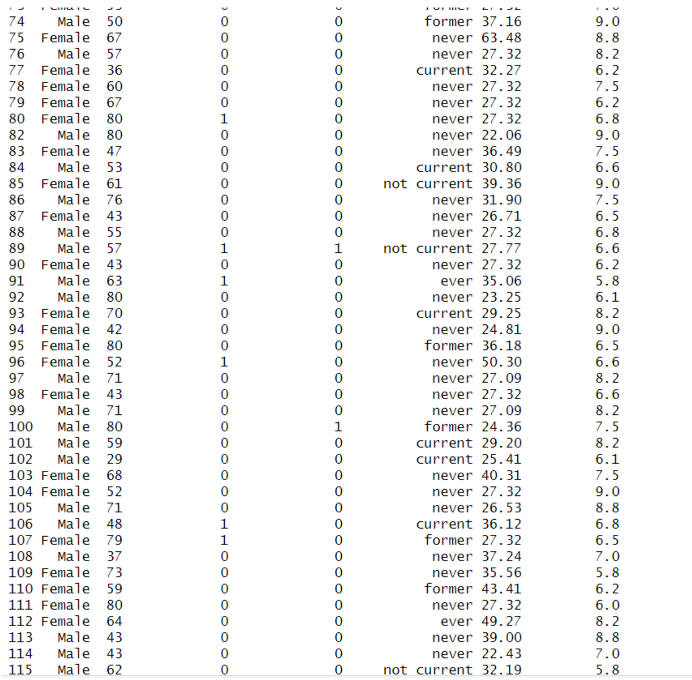
Replacing missing values in the column with the mode. The output:



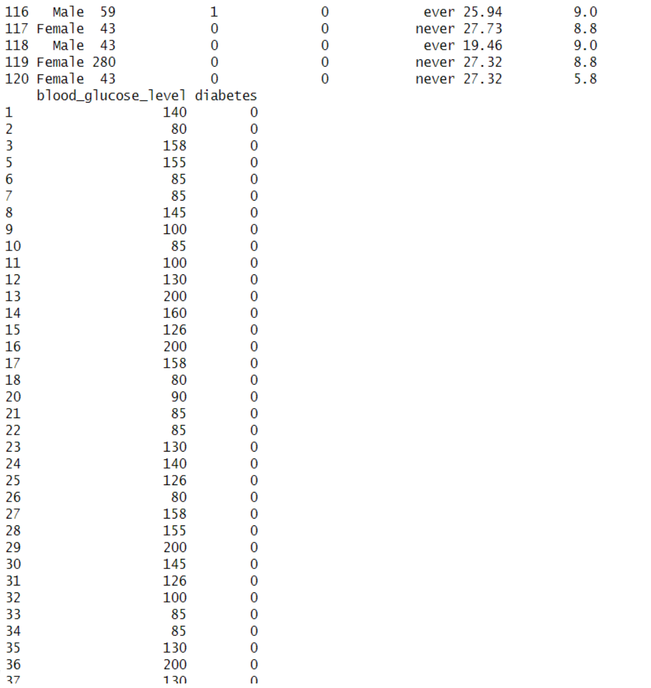
```
R 4.3.1 · C:/Users/ASUS/Downloads/
> mydata_mode <- mydata
> mode_gender <- names(sort(table(mydata_mode$gender), decreasing = TRUE))[1]
> mydata_mode$gender[is.na(mydata_mode$gender)] <- mode_gender
> mode_age <- names(sort(table(mydata_mode$age), decreasing = TRUE))[1]
> mydata_mode$age[is.na(mydata_mode$age)] <- mode_age
> mode_hypertension <- names(sort(table(mydata_mode$hypertension), decreasing = T
> mydata_mode$hypertension[is.na(mydata_mode$hypertension)] <- mode_hypertension
> mode_hsmoking <- names(sort(table(mydata_mode$smoking_history), decreasing = TR
> mydata_mode$smoking_history[is.na(mydata_mode$smoking_history)] <- mode_hsmokin
> mydata_mode$smoking_history[is.na(mydata_mode$smoking_history)] <- mode_hsmokin
> mydata_mode
    gender age hypertension heart_disease smoking_history   bmi HBA1c_level
1   Female  80            0             1           never 25.19         6.6
2   Female  54            0             0           never 27.32         6.6
3     Male  28            0             0           never 27.32         5.7
5     Male  76            1             1         current 20.14         4.8
6   Female  20            0             0           never 27.32         6.6
7   Female  79            0             0           never 23.86         5.7
8     Male  42            0             0           never 33.64         4.8
9   Female  32            0             0           never 27.32         5.0
10  Female  53            0             0           never 27.32         6.1
11  Female  54            0             0          former 54.70         6.0
12  Female  78            0             0          former 36.05         5.0
13  Female  67            0             0           never 25.69         5.8
14  Female  76            0             0           never 27.32         5.0
15  Female  78            0             0           never 27.32         6.6
16    Male  15            0             0           never 30.36         6.1
17  Female  42            0             0           never 24.48         5.7
18  Female  42            0             0           never 27.32         5.7
20    Male  40            0             0         current 36.38         6.0
21    Male   5            0             0           never 18.80         6.2
22  Female  69            0             0           never 21.24         4.8
23  Female  72            0             1          former 27.94         6.5
24  Female   4            0             0           never 13.99         4.0
25    Male  30            0             0           never 33.76         6.1
26    Male  40            0             0          former 27.85         5.8
27    Male  45            0             0           never 26.47         4.0
28    Male  43            0             0           never 26.08         6.1
29  Female  53            0             0           never 31.75         4.0
30    Male  50            0             0           never 25.15         4.0
31  Female  41            0             0         current 22.01         6.2
```



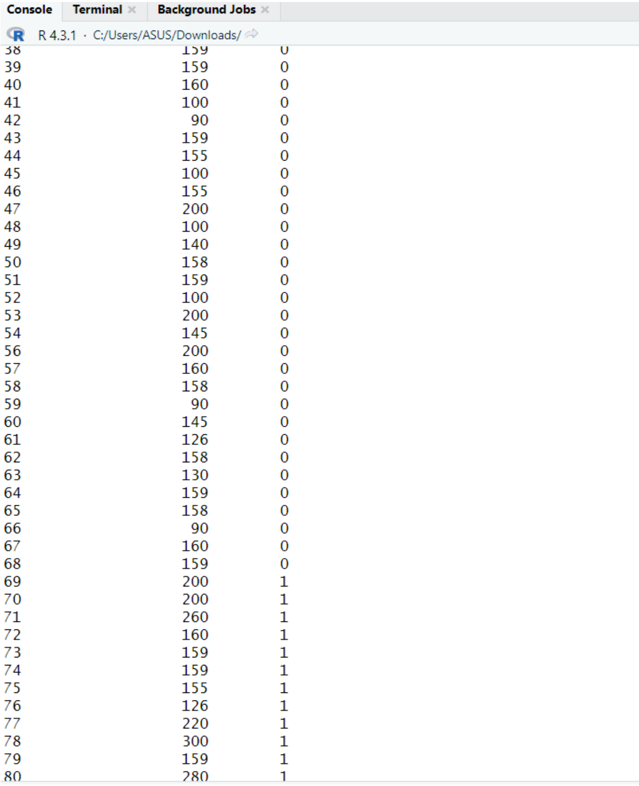
```
R 4.3.1 · C:/Users/ASUS/Downloads/
32  Female  20            0             0           never 22.19         3.5
33  Female  76            0             0           never 23.55         5.0
34    Male   5            0             0           never 15.10         5.8
35  Female  15            0             0           never 21.76         4.5
36  Female  26            0             0           never 21.22         6.6
37    Male   5            0             0           never 27.32         6.6
38  Female  77            1             1           never 32.02         5.0
39  Female  66            0             0           never 29.30         4.8
40  Female  67            0             0           never 27.32         3.5
41  Female  44            0             0           never 24.93         6.1
42  Female  29            0             0           never 19.95         5.0
43  Female  60            0             0           never 18.03         4.0
44  Female  38            0             0           never 28.27         6.2
45  Female   3            0             0           never 19.27         6.5
46    Male  57            0             0           never 27.32         6.1
47  Female  43            0             0           never 27.32         4.0
48  Female  74            0             0           never 28.12         5.0
49  Female  21            0             0           never 26.10         5.8
50  Female  30            0             0         current 27.32         6.5
51  Female  59            0             0          former 27.32         6.0
52  Female 290            0             0     not current 30.22         5.7
53  Female  59            0             1            ever 23.11         6.5
54  Female  19            0             0           never 27.32         5.7
56    Male  56            0             0           never 26.78         4.8
57    Male  43            0             0           never 23.04         5.7
58    Male   7            0             0           never 15.94         5.8
59    Male   3            0             0           never 15.80         6.2
60  Female  30            0             0           never 27.01         6.2
61    Male  43            0             0           never 27.32         3.5
62  Female  76            0             0           never 22.19         6.6
63  Female  41            0             0           never 27.45         5.7
64  Female  11            0             0           never 17.98         6.5
65  Female  26            0             0           never 26.45         5.7
66    Male  34            0             0           never 31.16         5.8
67    Male  80            0             0          former 24.42         4.0
68  Female  37            0             0           never 30.50         5.7
69  Female  44            0             0           never 19.31         6.5
70    Male  67            0             1     not current 27.32         6.5
71    Male  50            1             0         current 27.32         5.7
72    Male  73            0             0          former 25.91         9.0
73  Female  53            0             0          former 27.32         7.0
```



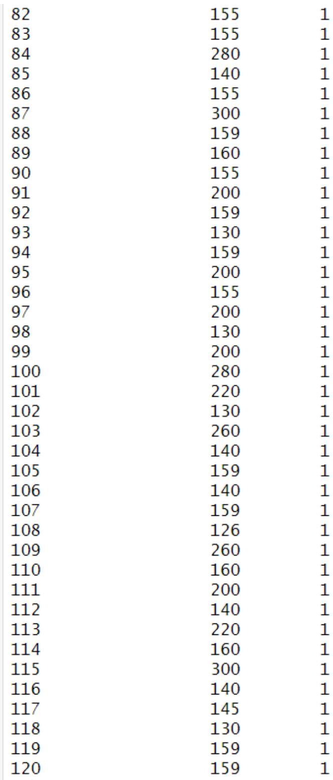
```
74    Male  50            0             0          former 37.16         9.0
75  Female  67            0             0           never 63.48         8.8
76    Male  57            0             0           never 27.32         8.2
77  Female  36            0             0         current 32.27         6.2
78  Female  60            0             0           never 27.32         7.5
79  Female  67            0             0           never 27.32         6.2
80  Female  80            1             0           never 27.32         6.8
82    Male  80            0             0           never 22.06         9.0
83  Female  47            0             0           never 36.49         7.5
84    Male  53            0             0         current 30.80         6.6
85  Female  61            0             0     not current 39.36         9.0
86    Male  76            0             0           never 31.90         7.5
87  Female  43            0             0           never 26.71         6.5
88    Male  55            0             0           never 27.32         6.8
89    Male  57            1             1     not current 27.77         6.6
90  Female  43            0             0           never 27.32         6.2
91    Male  63            1             0            ever 35.06         5.8
92    Male  80            0             0           never 23.25         6.1
93  Female  70            0             0         current 29.25         8.2
94  Female  42            0             0           never 24.81         9.0
95  Female  80            0             0          former 36.18         6.5
96  Female  52            1             0           never 50.30         6.6
97    Male  71            0             0           never 27.09         8.2
98  Female  43            0             0           never 27.32         6.6
99    Male  71            0             0           never 27.09         8.2
100   Male  80            0             1          former 24.36         7.5
101   Male  59            0             0         current 29.20         8.2
102   Male  29            0             0         current 25.41         6.1
103 Female  68            0             0           never 40.31         7.5
104 Female  52            0             0           never 27.32         9.0
105   Male  71            0             0           never 26.53         8.8
106   Male  48            1             0         current 36.12         6.8
107 Female  79            1             0          former 27.32         6.5
108   Male  37            0             0           never 37.24         7.0
109 Female  73            0             0           never 35.56         5.8
110 Female  59            0             0          former 43.41         6.2
111 Female  80            0             0           never 27.32         6.0
112 Female  64            0             0            ever 49.27         8.2
113   Male  43            0             0           never 39.00         8.8
114   Male  43            0             0           never 22.43         7.0
115   Male  62            0             0     not current 32.19         5.8
```



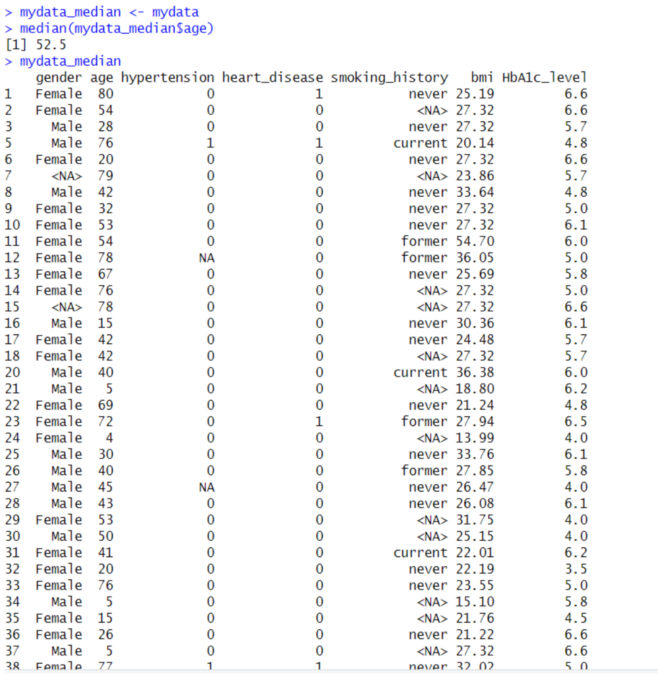
```
116   Male  59            1             0            ever 25.94         9.0
117 Female  43            0             0           never 27.73         8.8
118   Male  43            0             0            ever 19.46         9.0
119 Female 280            0             0           never 27.32         8.8
120 Female  43            0             0           never 27.32         5.8
    blood_glucose_level diabetes
1                   140        0
2                    80        0
3                   158        0
5                   155        0
6                    85        0
7                    85        0
8                   145        0
9                   100        0
10                   85        0
11                  100        0
12                  130        0
13                  200        0
14                  160        0
15                  126        0
16                  200        0
17                  158        0
18                   80        0
20                   90        0
21                   85        0
22                   85        0
23                  130        0
24                  140        0
25                  126        0
26                   80        0
27                  158        0
28                  155        0
29                  200        0
30                  145        0
31                  126        0
32                  100        0
33                   85        0
34                   85        0
35                  130        0
36                  200        0
37                  130        0
```



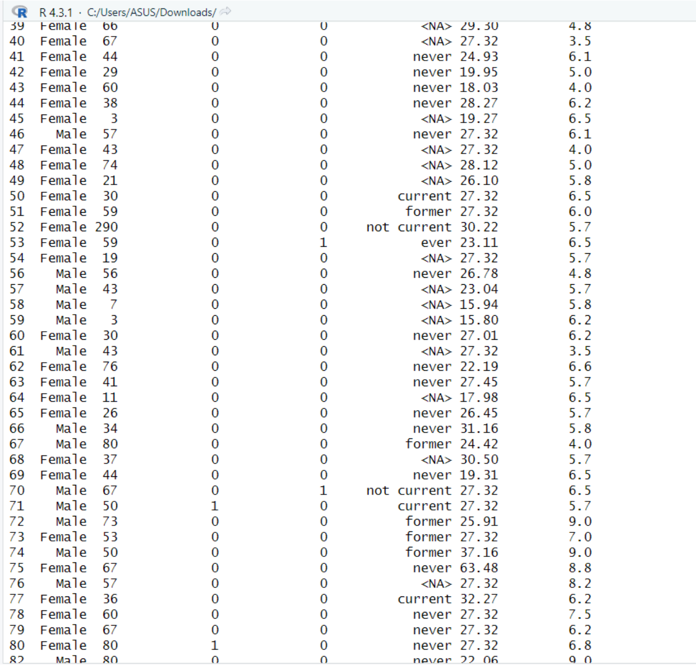
```
38                     159       0
39                     159       0
40                     160       0
41                     100       0
42                      90       0
43                     159       0
44                     155       0
45                     100       0
46                     155       0
47                     200       0
48                     100       0
49                     140       0
50                     158       0
51                     159       0
52                     100       0
53                     200       0
54                     145       0
56                     200       0
57                     160       0
58                     158       0
59                      90       0
60                     145       0
61                     126       0
62                     158       0
63                     130       0
64                     159       0
65                     158       0
66                      90       0
67                     160       0
68                     159       0
69                     200       1
70                     200       1
71                     260       1
72                     160       1
73                     159       1
74                     159       1
75                     155       1
76                     126       1
77                     220       1
78                     300       1
79                     159       1
80                     280       1
```



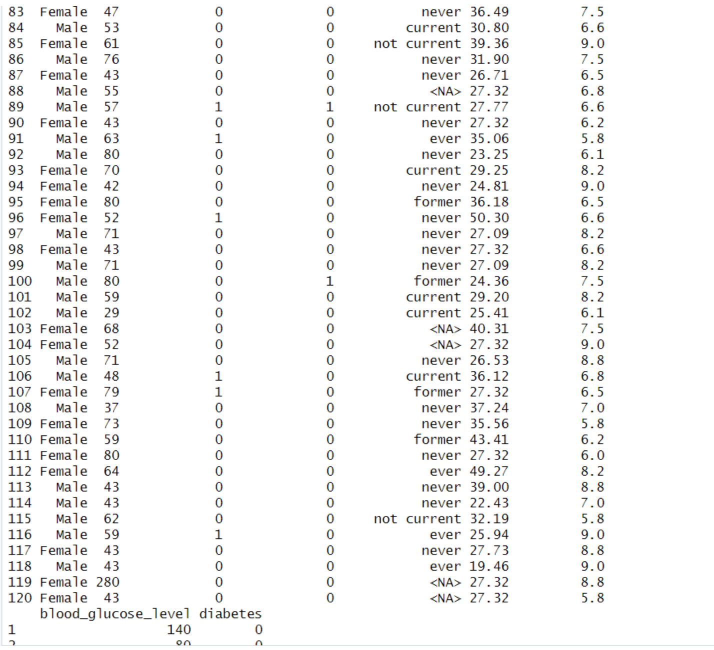
```
82                155         1
83                155         1
84                280         1
85                140         1
86                155         1
87                300         1
88                159         1
89                160         1
90                155         1
91                200         1
92                159         1
93                130         1
94                159         1
95                200         1
96                155         1
97                200         1
98                130         1
99                200         1
100               280         1
101               220         1
102               130         1
103               260         1
104               140         1
105               159         1
106               140         1
107               159         1
108               126         1
109               260         1
110               160         1
111               200         1
112               140         1
113               220         1
114               160         1
115               300         1
116               140         1
117               145         1
118               130         1
119               159         1
120               159         1
```



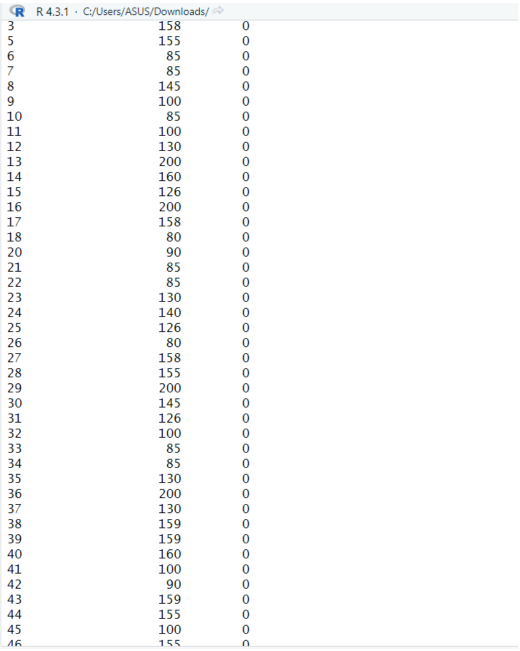
```
> mydata_median <- mydata
> median(mydata_median$age)
[1] 52.5
> mydata_median
    gender age hypertension heart_disease smoking_history   bmi HbA1c_level
1   Female  80            0             1           never 25.19         6.6
2   Female  54            0             0            <NA> 27.32         6.6
3     Male  28            0             0           never 27.32         5.7
5     Male  76            1             1         current 20.14         4.8
6   Female  20            0             0           never 27.32         6.6
7     <NA>  79            0             0            <NA> 23.86         5.7
8     Male  42            0             0           never 33.64         4.8
9   Female  32            0             0           never 27.32         5.0
10  Female  53            0             0           never 27.32         6.1
11  Female  54            0             0          former 54.70         6.0
12  Female  78           NA             0          former 36.05         5.0
13  Female  67            0             0           never 25.69         5.8
14  Female  76            0             0            <NA> 27.32         5.0
15    <NA>  78            0             0            <NA> 27.32         6.6
16    Male  15            0             0           never 30.36         6.1
17  Female  42            0             0           never 24.48         5.7
18  Female  42            0             0            <NA> 27.32         5.7
20    Male  40            0             0         current 36.38         6.0
21    Male   5            0             0            <NA> 18.80         6.2
22  Female  69            0             0           never 21.24         4.8
23  Female  72            0             1          former 27.94         6.5
24  Female   4            0             0            <NA> 13.99         4.0
25    Male  30            0             0           never 33.76         6.1
26    Male  40            0             0          former 27.85         5.8
27    Male  45           NA             0           never 26.47         4.0
28    Male  43            0             0           never 26.08         6.1
29  Female  53            0             0            <NA> 31.75         4.0
30    Male  50            0             0            <NA> 25.15         4.0
31  Female  41            0             0         current 22.01         6.2
32  Female  20            0             0           never 22.19         3.5
33  Female  76            0             0           never 23.55         5.0
34    Male   5            0             0            <NA> 15.10         5.8
35  Female  15            0             0            <NA> 21.76         4.5
36  Female  26            0             0           never 21.22         6.6
37    Male   5            0             0            <NA> 27.32         6.6
38  Female  77            1             1           never 32.02         5.0
```



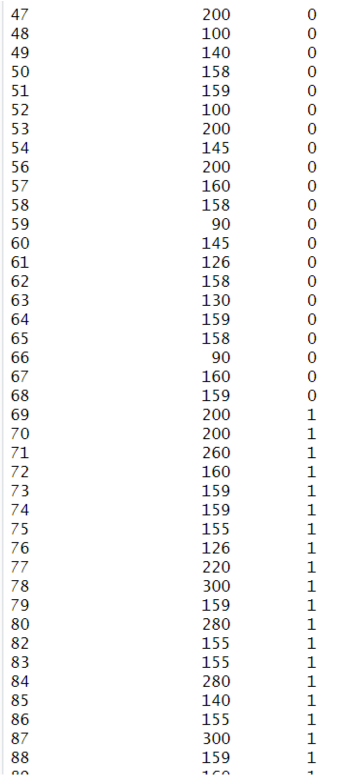
```
39  Female  66            0             0            <NA> 29.30         4.8
40  Female  67            0             0            <NA> 27.32         3.5
41  Female  44            0             0           never 24.93         6.1
42  Female  29            0             0           never 19.95         5.0
43  Female  60            0             0           never 18.03         4.0
44  Female  38            0             0           never 28.27         6.2
45  Female   3            0             0            <NA> 19.27         6.5
46    Male  57            0             0           never 27.32         6.1
47  Female  43            0             0            <NA> 27.32         4.0
48  Female  74            0             0            <NA> 28.12         5.0
49  Female  21            0             0            <NA> 26.10         5.8
50  Female  30            0             0         current 27.32         6.5
51  Female  59            0             0          former 27.32         6.0
52  Female 290            0             0     not current 30.22         5.7
53  Female  59            0             1            ever 23.11         6.5
54  Female  19            0             0            <NA> 27.32         5.7
56    Male  56            0             0           never 26.78         4.8
57    Male  43            0             0            <NA> 23.04         5.7
58    Male   7            0             0            <NA> 15.94         5.8
59    Male   3            0             0            <NA> 15.80         6.2
60  Female  30            0             0           never 27.01         6.2
61    Male  43            0             0            <NA> 27.32         3.5
62  Female  76            0             0           never 22.19         6.6
63  Female  41            0             0           never 27.45         5.7
64  Female  11            0             0            <NA> 17.98         6.5
65  Female  26            0             0           never 26.45         5.7
66    Male  34            0             0           never 31.16         5.8
67    Male  80            0             0          former 24.42         4.0
68  Female  37            0             0            <NA> 30.50         5.7
69  Female  44            0             0           never 19.31         6.5
70    Male  67            0             1     not current 27.32         6.5
71    Male  50            1             0         current 27.32         5.7
72    Male  73            0             0          former 25.91         9.0
73  Female  53            0             0          former 27.32         7.0
74    Male  50            0             0          former 37.16         9.0
75  Female  67            0             0           never 63.48         8.8
76    Male  57            0             0            <NA> 27.32         8.2
77  Female  36            0             0         current 32.27         6.2
78  Female  60            0             0           never 27.32         7.5
79  Female  67            0             0           never 27.32         6.2
80  Female  80            1             0           never 27.32         6.8
82    Male  80            0             0           never 22.06         9.0
```



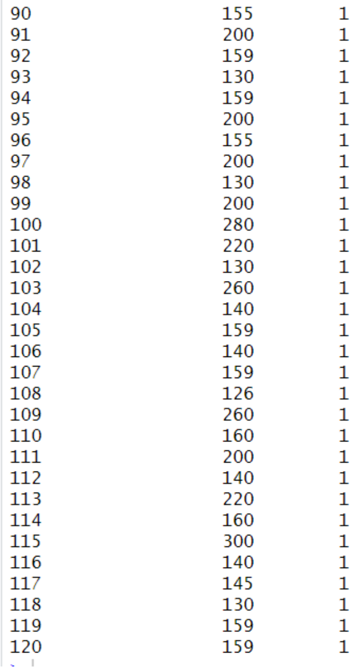
```
83  Female  47           0           0        never 36.49        7.5
84    Male  53           0           0      current 30.80        6.6
85  Female  61           0           0  not current 39.36        9.0
86    Male  76           0           0        never 31.90        7.5
87  Female  43           0           0        never 26.71        6.5
88    Male  55           0           0         <NA> 27.32        6.8
89    Male  57           1           1  not current 27.77        6.6
90  Female  43           0           0        never 27.32        6.2
91    Male  63           1           0         ever 35.06        5.8
92    Male  80           0           0        never 23.25        6.1
93  Female  70           0           0      current 29.25        8.2
94  Female  42           0           0        never 24.81        9.0
95  Female  80           0           0       former 36.18        6.5
96  Female  52           1           0        never 50.30        6.6
97    Male  71           0           0        never 27.09        8.2
98  Female  43           0           0        never 27.32        6.6
99    Male  71           0           0        never 27.09        8.2
100   Male  80           0           1       former 24.36        7.5
101   Male  59           0           0      current 29.20        8.2
102   Male  29           0           0      current 25.41        6.1
103 Female  68           0           0         <NA> 40.31        7.5
104 Female  52           0           0         <NA> 27.32        9.0
105   Male  71           0           0        never 26.53        8.8
106   Male  48           1           0      current 36.12        6.8
107 Female  79           1           0       former 27.32        6.5
108   Male  37           0           0        never 37.24        7.0
109 Female  73           0           0        never 35.56        5.8
110 Female  59           0           0       former 43.41        6.2
111 Female  80           0           0        never 27.32        6.0
112 Female  64           0           0         ever 49.27        8.2
113   Male  43           0           0        never 39.00        8.8
114   Male  43           0           0        never 22.43        7.0
115   Male  62           0           0  not current 32.19        5.8
116   Male  59           1           0         ever 25.94        9.0
117 Female  43           0           0        never 27.73        8.8
118   Male  43           0           0         ever 19.46        9.0
119 Female 280           0           0         <NA> 27.32        8.8
120 Female  43           0           0         <NA> 27.32        5.8
    blood_glucose_level diabetes
1                   140        0
2                    80        0
```



```
3                   158        0
5                   155        0
6                    85        0
7                    85        0
8                   145        0
9                   100        0
10                   85        0
11                  100        0
12                  130        0
13                  200        0
14                  160        0
15                  126        0
16                  200        0
17                  158        0
18                   80        0
20                   90        0
21                   85        0
22                   85        0
23                  130        0
24                  140        0
25                  126        0
26                   80        0
27                  158        0
28                  155        0
29                  200        0
30                  145        0
31                  126        0
32                  100        0
33                   85        0
34                   85        0
35                  130        0
36                  200        0
37                  130        0
38                  159        0
39                  159        0
40                  160        0
41                  100        0
42                   90        0
43                  159        0
44                  155        0
45                  100        0
46                  155        0
```

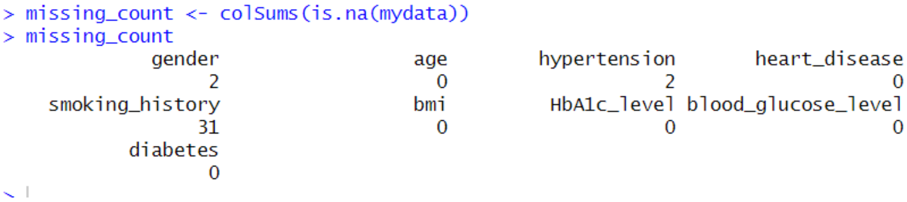


```
47                  200        0
48                  100        0
49                  140        0
50                  158        0
51                  159        0
52                  100        0
53                  200        0
54                  145        0
56                  200        0
57                  160        0
58                  158        0
59                   90        0
60                  145        0
61                  126        0
62                  158        0
63                  130        0
64                  159        0
65                  158        0
66                   90        0
67                  160        0
68                  159        0
69                  200        1
70                  200        1
71                  260        1
72                  160        1
73                  159        1
74                  159        1
75                  155        1
76                  126        1
77                  220        1
78                  300        1
79                  159        1
80                  280        1
82                  155        1
83                  155        1
84                  280        1
85                  140        1
86                  155        1
87                  300        1
88                  159        1
```



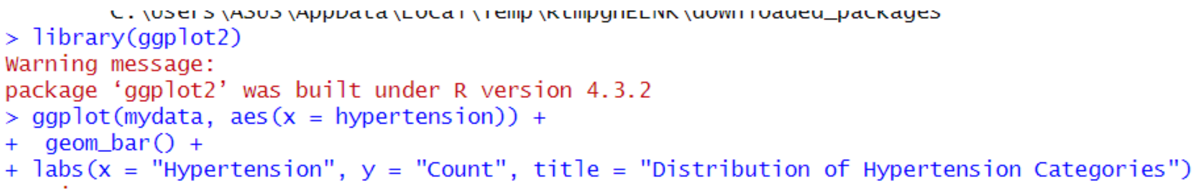
```
90                  155        1
91                  200        1
92                  159        1
93                  130        1
94                  159        1
95                  200        1
96                  155        1
97                  200        1
98                  130        1
99                  200        1
100                 280        1
101                 220        1
102                 130        1
103                 260        1
104                 140        1
105                 159        1
106                 140        1
107                 159        1
108                 126        1
109                 260        1
110                 160        1
111                 200        1
112                 140        1
113                 220        1
114                 160        1
115                 300        1
116                 140        1
117                 145        1
118                 130        1
119                 159        1
120                 159        1
```

Data cleaning by counting the total missing column. The output:

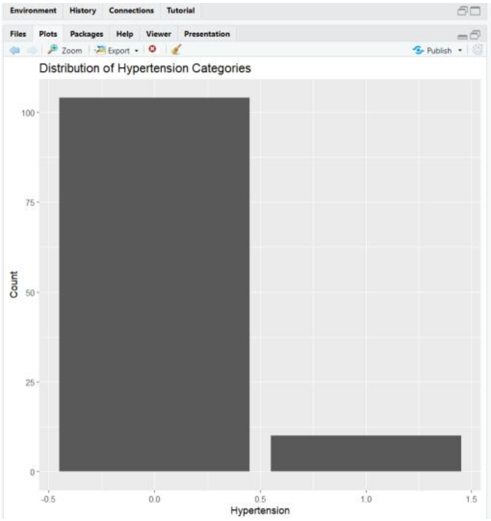


```
> missing_count <- colSums(is.na(mydata))
> missing_count
          gender              age     hypertension    heart_disease 
               2                0                2                0 
 smoking_history              bmi      HbA1c_level blood_glucose_level 
              31                0                0                0 
        diabetes 
               0 
> |
```

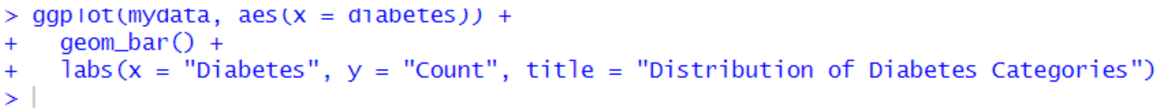
A visual representation of the counts of different hypertension categories, that will help to understand the distribution of this variable in mydata. The output:



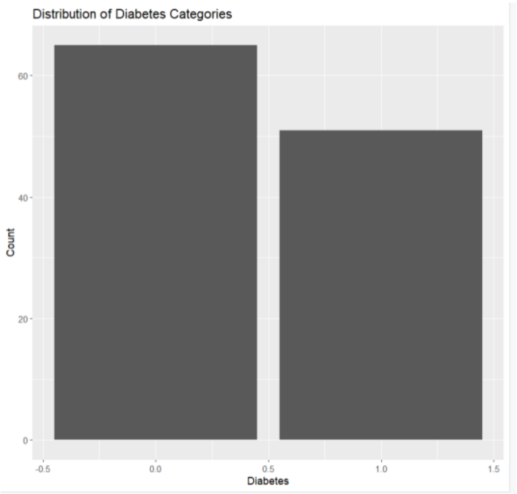
```
        C:\Users\ASUS\AppData\Local\Temp\RtmpgnELNK\downloaded_packages
> library(ggplot2)
Warning message:
package ‘ggplot2’ was built under R version 4.3.2 
> ggplot(mydata, aes(x = hypertension)) +
+   geom_bar() +
+ labs(x = "Hypertension", y = "Count", title = "Distribution of Hypertension Categories")
```



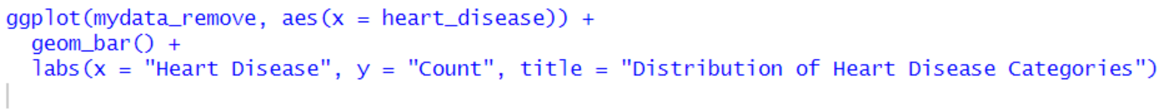
A visual representation of the counts of different diabetes categories, that will help to understand the distribution of this variable in my data. The output:



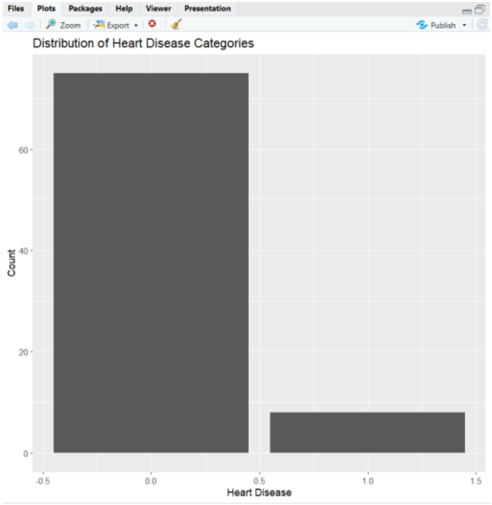
```
> ggplot(mydata, aes(x = diabetes)) +
+   geom_bar() +
+   labs(x = "Diabetes", y = "Count", title = "Distribution of Diabetes Categories")
> |
```



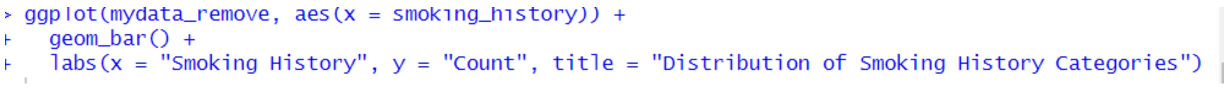
A visual representation of the counts of different heart disease categories. The output:



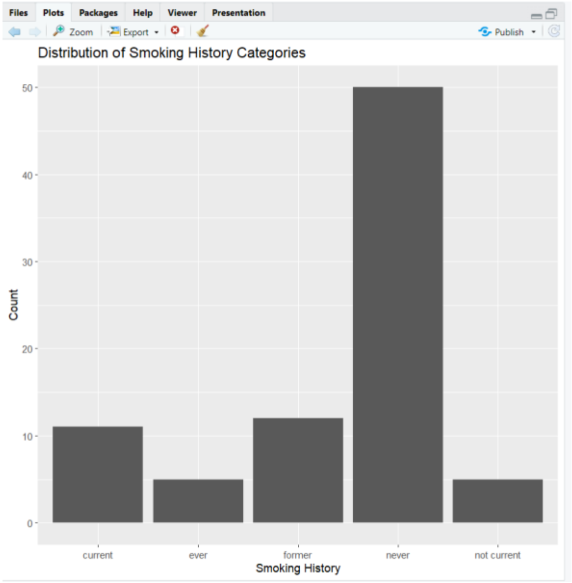
```
ggplot(mydata_remove, aes(x = heart_disease)) +
  geom_bar() +
  labs(x = "Heart Disease", y = "Count", title = "Distribution of Heart Disease Categories")
```



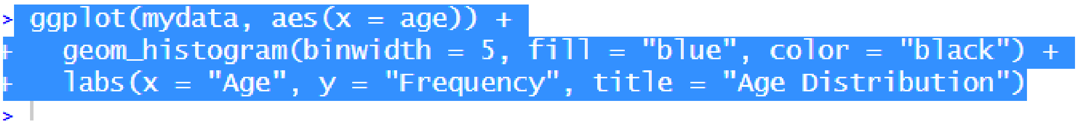
A visual representation of the counts of different smoking histories categories. The output:



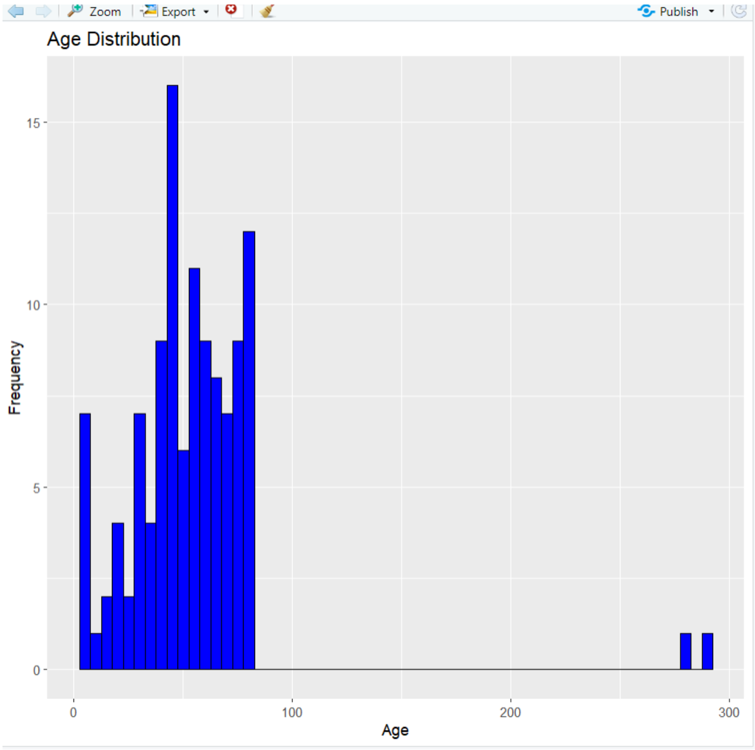
```
> ggplot(mydata_remove, aes(x = smoking_history)) +
+   geom_bar() +
+   labs(x = "Smoking History", y = "Count", title = "Distribution of Smoking History Categories")
```



Ages are distributed across different ranges. The output:

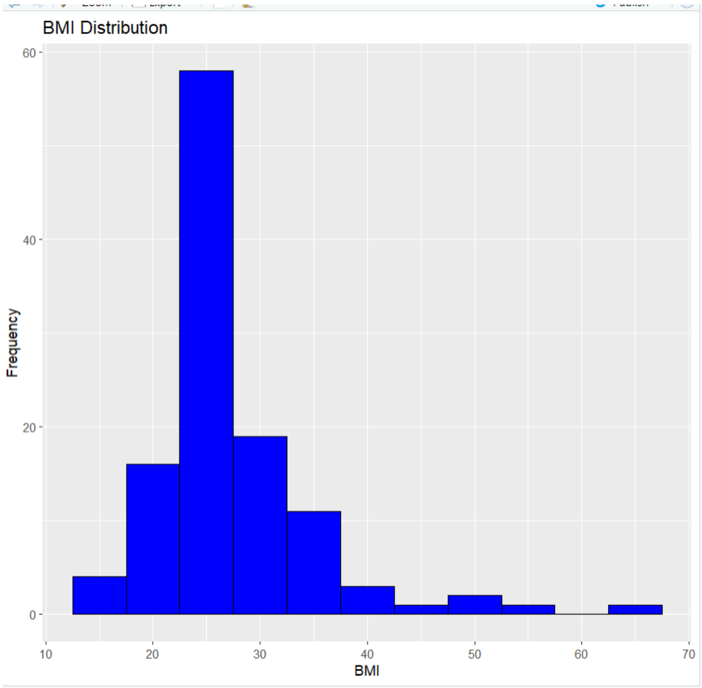


```
> ggplot(mydata, aes(x = age)) +
+   geom_histogram(binwidth = 5, fill = "blue", color = "black") +
+   labs(x = "Age", y = "Frequency", title = "Age Distribution")
>
```



The BMIs are distributed across different ranges. The output:

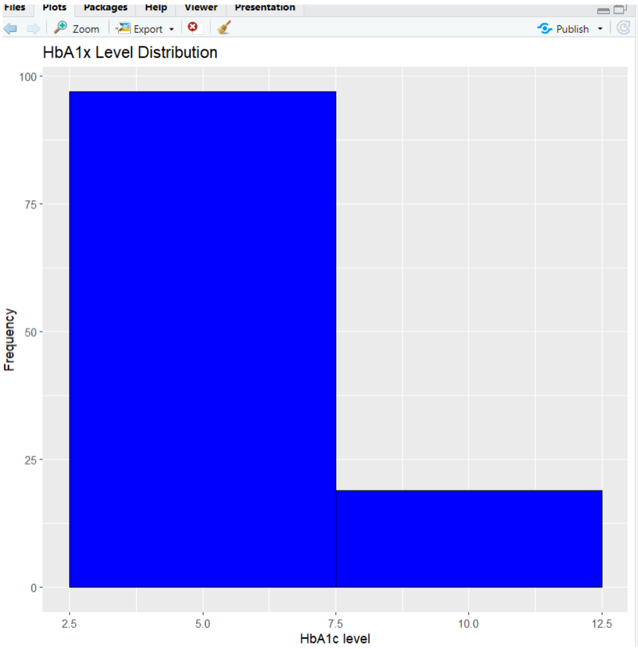
```
ggplot(mydata, aes(x = bmi)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  labs(x = "BMI", y = "Frequency", title = "BMI Distribution")
```



nges. The output:

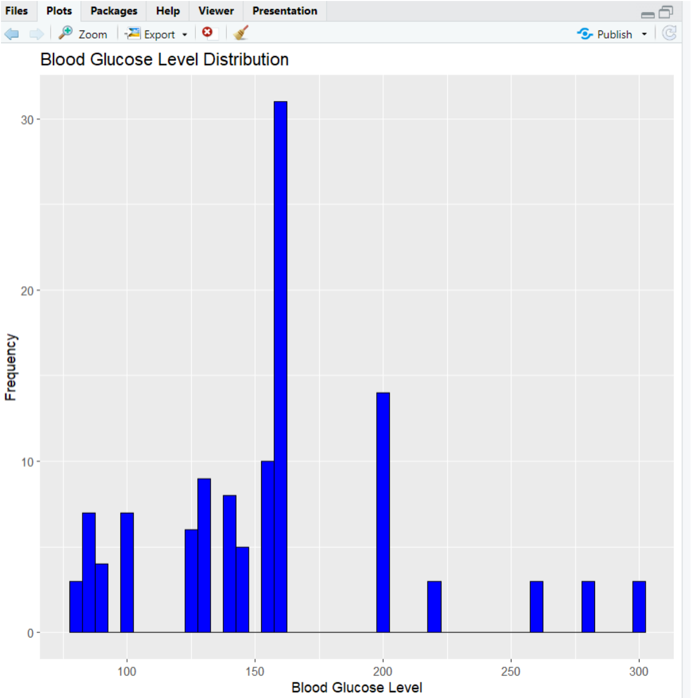
```
lor = "black") +
 = "HbA1x Level Distribution")
```

```
> |
```



The Blood glucose levels are distributed across different ranges. The output:

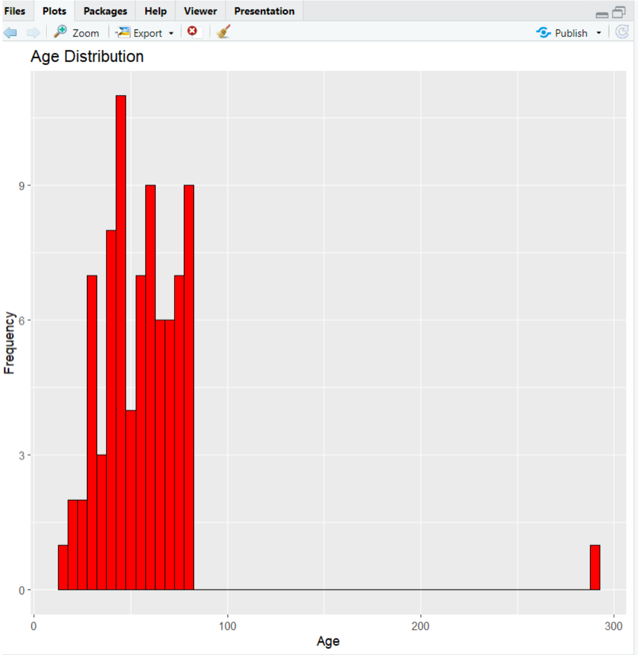
```
ggplot(mydata, aes(x = blood_glucose_level)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  labs(x = "Blood Glucose Level", y = "Frequency", title = "Blood Glucose Level Distribution")
```



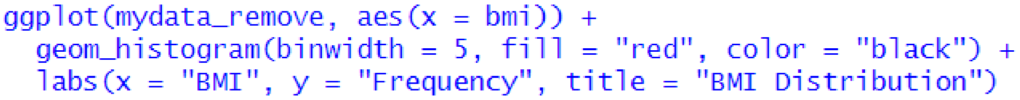
After missing values handling:

Numerical values of ages are distributed across different age ranges in 'mydata_remove' dataset. The output:

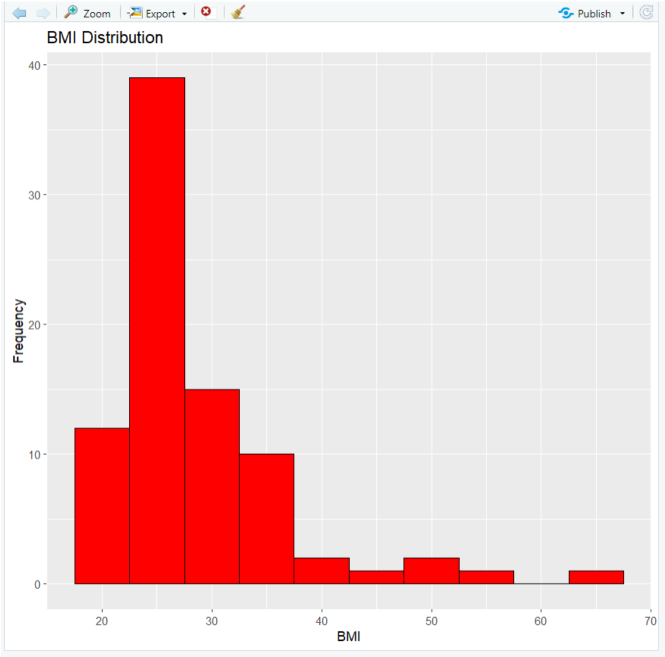
```
ggplot(mydata_remove, aes(x = age)) +
  geom_histogram(binwidth = 5, fill = "red", color = "black") +
  labs(x = "Age", y = "Frequency", title = "Age Distribution")
```



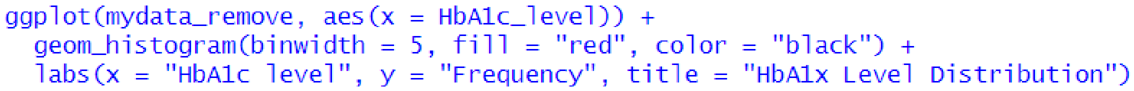
Numerical values of BMIs are distributed across different bmi ranges in my 'mydata_remove' dataset. The output:



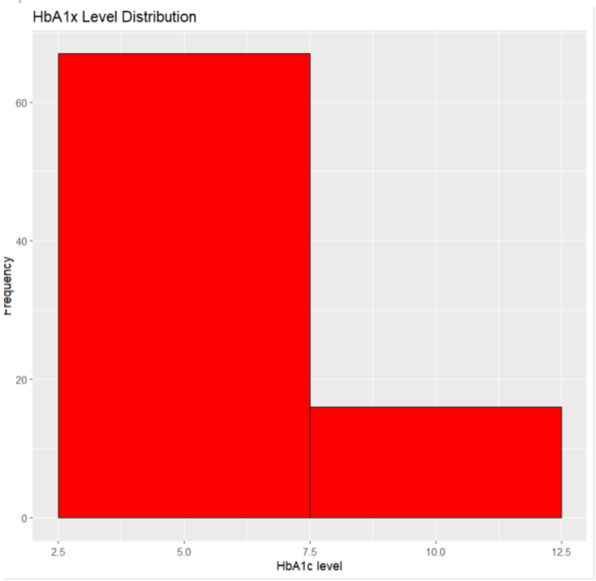
```
ggplot(mydata_remove, aes(x = bmi)) +
  geom_histogram(binwidth = 5, fill = "red", color = "black") +
  labs(x = "BMI", y = "Frequency", title = "BMI Distribution")
```



How HBA1cs are distributed between two different ranges in my 'mydata_remove' dataset. The output:



```
ggplot(mydata_remove, aes(x = HbA1c_level)) +
  geom_histogram(binwidth = 5, fill = "red", color = "black") +
  labs(x = "HbA1c level", y = "Frequency", title = "HbA1x Level Distribution")
```



This is how an effective comparison can be identified by performing data exploration, outliers' detection, missing values detection and handling the missing values by implementing mean, mode, median and row wise removing techniques.