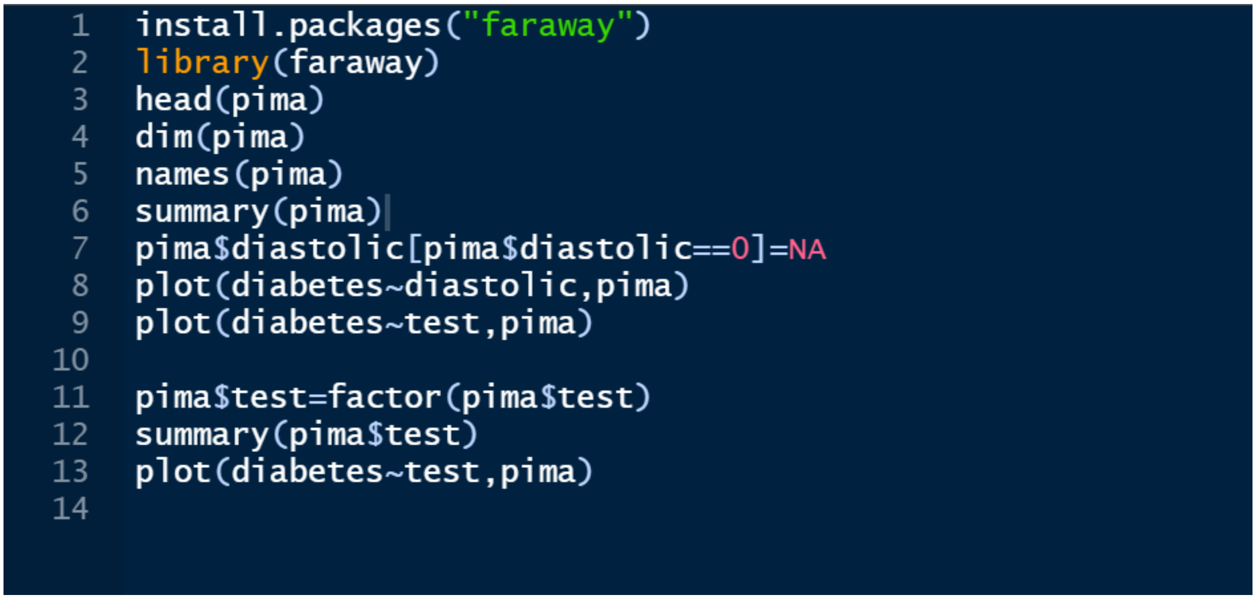
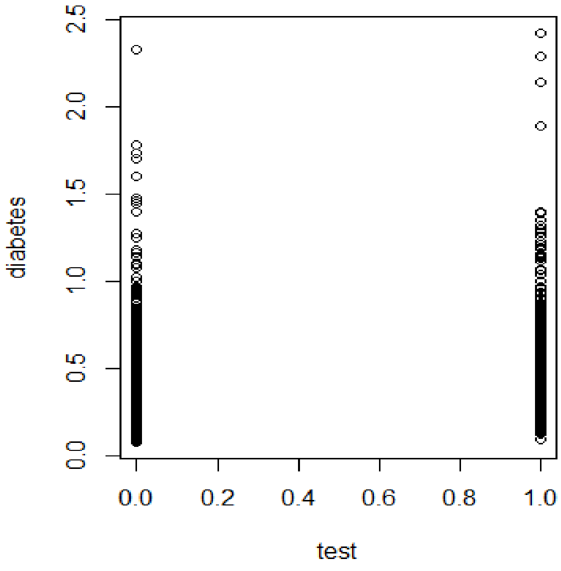
Factors in R

We continue our discussion of using the faraway package and commands in R.



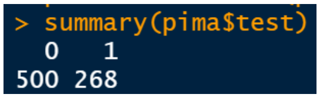
```
1   install.packages("faraway")
2   library(faraway)
3   head(pima)
4   dim(pima)
5   names(pima)
6   summary(pima)
7   pima$diastolic[pima$diastolic==0]=NA
8   plot(diabetes~diastolic,pima)
9   plot(diabetes~test,pima)
10
11  pima$test=factor(pima$test)
12  summary(pima$test)
13  plot(diabetes~test,pima)
14
```

Let us look at line 9 from the code snippet. ”plot(diabetes test,pima)” gives us the following output.



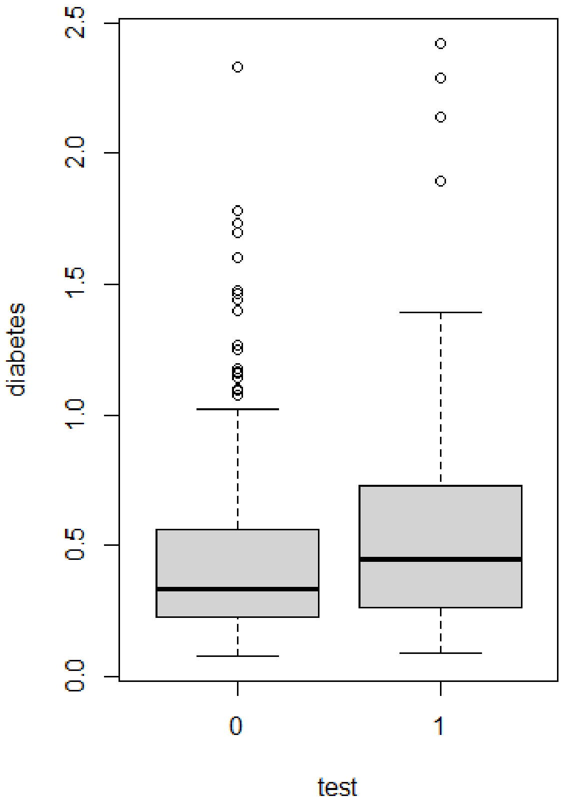
As seen from the diagram, test is a boolean variable.It returns whether or not the patient has diabetes.The numbers between 0 and 1 exist in the plot because R interprets the test variable as continuous whereas it is categorical by nature.These categories are known as factors in linear models nomenclature.

Now,we change the variable test to the factors with line 11.Line 12 generates the following output



```
> summary(pima$test)
  0   1
500 268
```

Test no more generates quartiles and means. But rather shows the number of elements of each factor; 0 has 500 elements while 1 has 268.Now,let us plot it again with line 13.



Now,we see that the plot has changed to a box and whiskers plot, which is a very useful but underrated statistical graphical tool.The bold horizontal line in the middle of each box is the median,while the horizontal boundaries of the box indicate the first(bottom one) and third(top) quartiles respectively.A popular method for calculating the whiskers is on the basis of k*IQF ,where k=1.5 and Inter-Quartile range(IQF)=Third quartile-First quartile.All those points lying beyond the whiskers are suspected outliers and are shown as dots.

In this specific scenario, this tells us that the scatter of the data is less for test=0 than it is for test=1.It also suspects 4 points in test=1 to be outliers.

Things to try: We can also find the outliers using commands in R and it should be tried out by the reader.

Note: For those new to programming, use comments whenever necessary and abundantly as it improves the readability of the code and helps in finding out bugs.