Visualization of Data

Importing the data into R console without disturbing the observations.

Input :

```
> data = read.csv(file.choose())
> data
```

Output :

```
                            Company Sales..billions.
1            Citigroup                       108.28
2     Generala Electric                      152.36
3  American Intl Group                        95.04
4      Bank of America                        65.45
5            HSBC group                       62.97
6          Exxon Mobil                       263.99
7    Royal Dutch/Shell                       265.19
8                   BP                       185.06
9            ING Group                        92.01
10        Toyota Motor                      165.68
   Profits..billions. Assets..Billions.
1               17.05            1484.10
2               16.59             750.33
3               10.91             766.42
4               14.14             110.46
5                9.52            1031.29
6               25.33             195.26
7               18.54             193.83
8               15.73             191.11
9                  NA            1175.16
10              11.13             211.15
```

```
> data = read.csv(file.choose())
> data
                Company Sales..billions.
1             Citigroup           108.28
2    Generala Electric            152.36
3  American Intl Group            95.04
4       Bank of America            65.45
5             HSBC group           62.97
6           Exxon Mobil            263.99
7    Royal Dutch/Shell             265.19
8                    BP            185.06
9             ING Group            92.01
10         Toyota Motor            165.68
   Profits..billions. Assets..Billions.
1               17.05           1484.10
2               16.59            750.33
3               10.91            766.42
4               14.14            110.46
5                9.52           1031.29
6               25.33            195.26
7               18.54            193.83
8               15.73            191.11
9                  NA           1175.16
10              11.13            211.15
```

Since, we are having one NA entry in our dataset. Therefore, we replace it by the mean value of that column.

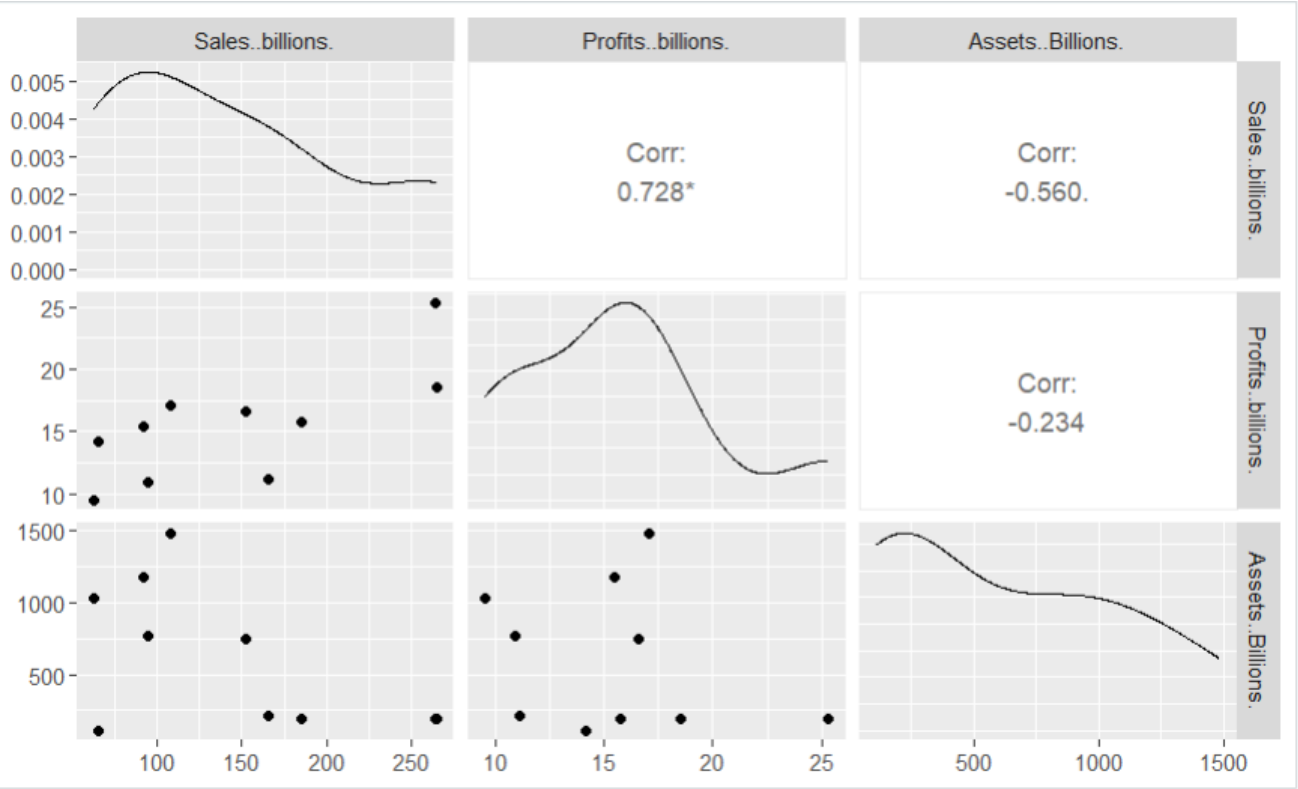
```
> data = read.csv(file.choose())
> data
              Company Sales..billions. Profits..billions. Assets..Billions.
1           Citigroup           108.28           17.05000           1484.10
2    Generala Electric          152.36           16.59000            750.33
3  American Intl Group           95.04           10.91000            766.42
4       Bank of America           65.45           14.14000            110.46
5             HSBC group          62.97            9.52000           1031.29
6           Exxon Mobil          263.99           25.33000            195.26
7    Royal Dutch/Shell           265.19           18.54000            193.83
8                    BP          185.06           15.73000            191.11
9             ING Group           92.01           15.43778           1175.16
10         Toyota Motor          165.68           11.13000            211.15
```

Visualizing the data with suitable diagrams and interpreting the findings.

Input :

```
> library(GGally)
> library(ggplot2)
> ggpairs(data[,c(2:4)])
```

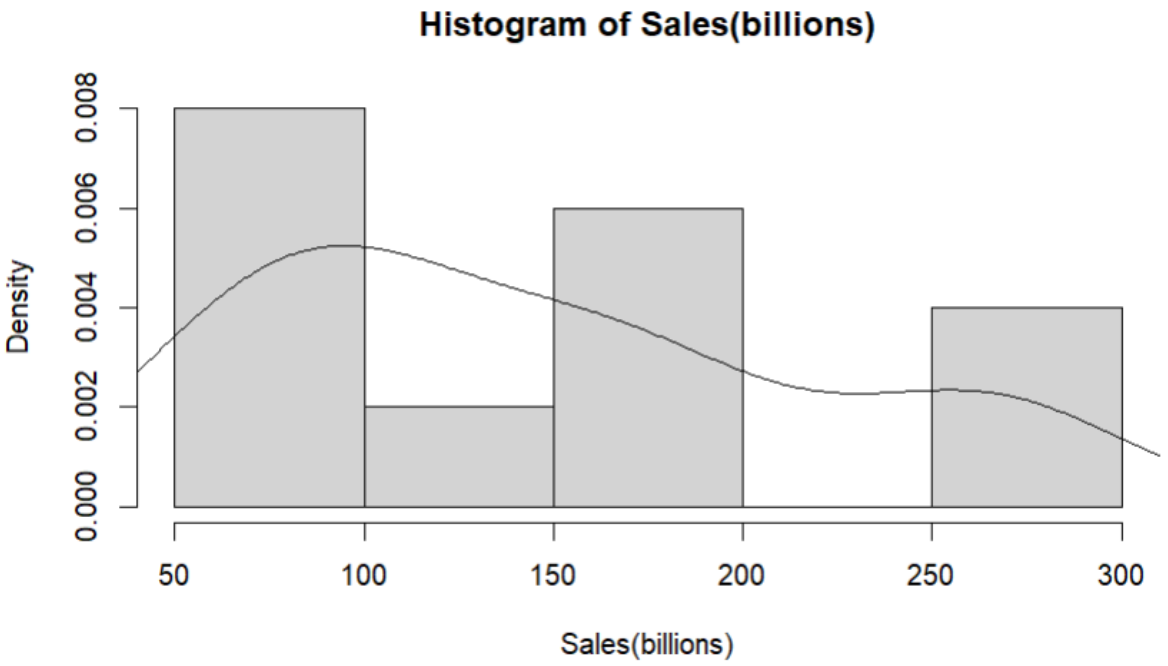
Output :



Input :

```
> hist(data[,2],probability = T,main = "Hi
stogram of Sales(billions)",xlab = "Sales(
billions)")
> lines(density(data[,2]))
```

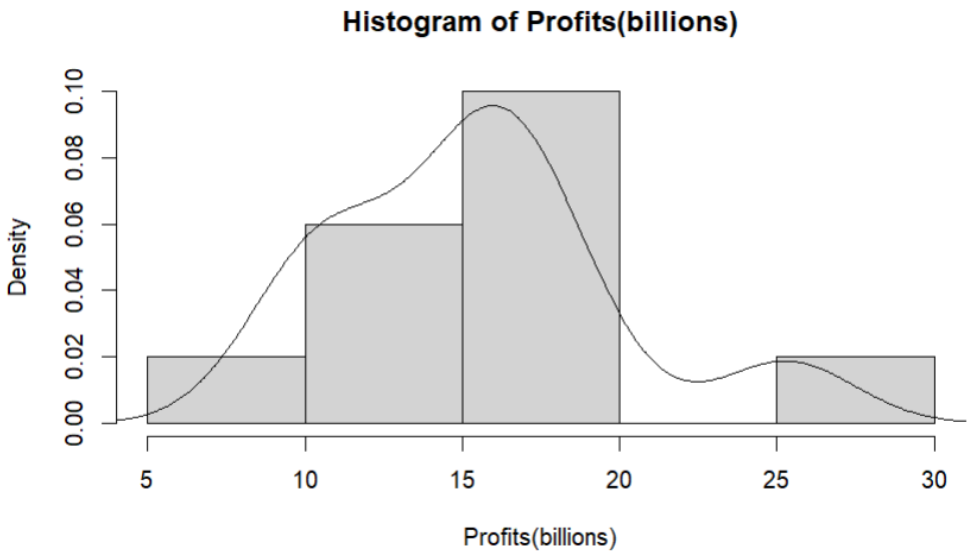
Output :



Input :

```
> hist(data[,3],probability=T,main = "Hist
ogram of Profits(billions)",xlab = "Profit
s(billions)")
> lines(density(data[,3]))
```

Output :



Interpretations from graph:

Comparing, the above histograms with the graphs on the diagonal of ggpair plot. We can say that diagonal graphs are the density curves of the respective variables.

We can also see that sales & profits are positively correlated whereas sales-assets and profit-assets are negatively correlated.

Also, we can see that profit is partially following a normal distribution.

Sales and Assets are following a right-skewed normal distribution.

suggestions to the end user of data.

Since, the profit frequency distribution is following a normal distribution pattern, we can say that companies are making a constant amount of profit (i.e. profit lies in the same range) and it's not increasing. So, the companies should work on increasing the revenue generation.