

# R Notebook

[Code ▾](#)[Hide](#)

```
# To import the data
my_data <- read.table("C:/Users/Sushmita Singh/Desktop/Assignments/MachineLearning/Dataset.csv", header = TRUE ,sep=",")
my_data
```

<b>srno</b>	<b>state</b>	<b>zipcode</b>	<b>familysize</b>	<b>annualincome</b>
<int>	<fctr>	<int>	<int>	<int>
1	florida	32716	5	7000
2	alabama	35236	7	5000
3	florida	32116	4	4000
4	florida	34567	2	3000
5	california	32182	3	2000
6	florida	32415	4	6000
7	alabama	32987	6	8000
8	california	32110	8	10000
9	california	32998	1	2000

9 rows

## Compute the descriptive statistics of variables.

[Hide](#)

```
# Compute the descriptive statistics of variables.
#compute the mean value
mean(my_data$annualincome)
```

```
[1] 5222.222
```

Hide

```
mean(my_data$familysize)
```

```
[1] 4.444444
```

Hide

```
#compute the median value  
median(my_data$familysize)
```

```
[1] 4
```

Hide

```
median(my_data$annualincome)
```

```
[1] 5000
```

Hide

```
#compute the minimum value  
min(my_data$annualincome)
```

```
[1] 2000
```

Hide

```
#compute the maximum value  
max(my_data$annualincome)
```

```
[1] 10000
```

Hide

```
#compute the range of data  
range(my_data$familysize)
```

```
[1] 1 8
```

Hide

```
range(my_data$annualincome)
```

```
[1] 2000 10000
```

Hide

```
#to compute qaurtile (by default it returns minimum,maximum and
#three quartile(0.25,0.50 and ,0.75))
quantile(my_data$familysize)
```

```
0% 25% 50% 75% 100%
1 3 4 6 8
```

Hide

```
#to compute deciles
quantile(my_data$familysize, seq(0,1,0.1))
```

```
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
1.0 1.8 2.6 3.4 4.0 4.0 4.8 5.6 6.4 7.2 8.0
```

Hide

```
quantile(my_data$annualincome, seq(0,1,0.1))
```

```
0% 10% 20% 30% 40% 50% 60% 70% 80%
2000 2000 2600 3400 4200 5000 5800 6600 7400
90% 100%
8400 10000
```

Hide

```
#to compute the interquartile range
IQR(my_data$annualincome)
```

```
[1] 4000
```

[Hide](#)

```
#compute the variance  
var(my_data$familysize)
```

```
[1] 5.277778
```

[Hide](#)

```
#compute the standard deviation=square root of variance  
sd(my_data$familysize)
```

```
[1] 2.297341
```

[Hide](#)

```
#compute the overall summary  
summary(my_data$familysize)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	3.000	4.000	4.444	6.000	8.000

[Hide](#)

```
#compute the summary of dataframe:the function summary() is automatically applied to each column.  
#The format of the result depends on the type of the data contained in the column. For example:  
#If the column is a numeric variable, mean, median, min, max and quartiles are returned.  
#If the column is a factor variable, the number of observations in each group is returned.  
summary(my_data , digits = 1)
```

srno	state	zipcode	familysize
Min. :1	alabama :2	Min. :32110	Min. :1
1st Qu.:3	california:3	1st Qu.:32182	1st Qu.:3
Median :5	florida :4	Median :32716	Median :4
Mean :5		Mean :33036	Mean :4
3rd Qu.:7		3rd Qu.:32998	3rd Qu.:6
Max. :9		Max. :35236	Max. :8

annualincome

Min. : 2000

1st Qu.: 3000

Median : 5000

Mean : 5222

3rd Qu.: 7000

Max. :10000

Hide

```
#TO check the variable of data
str(my_data)
```

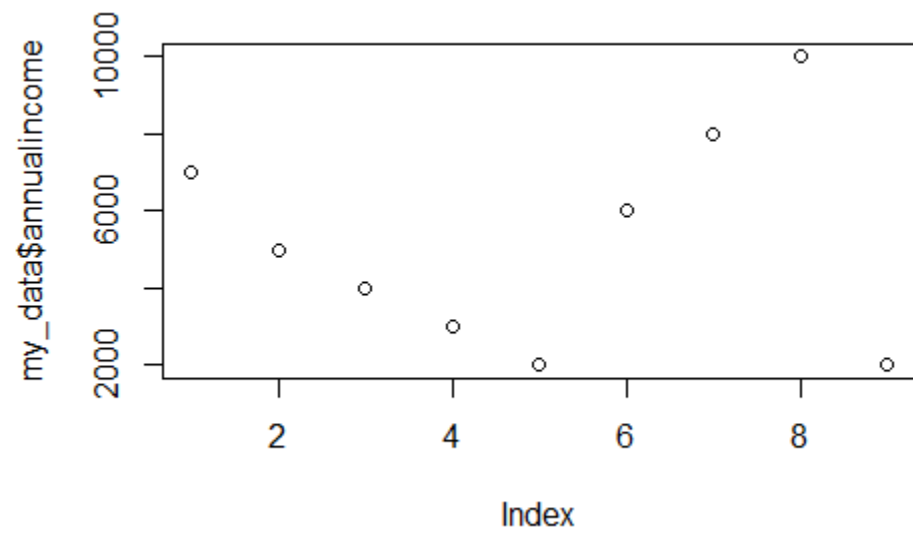
```
'data.frame': 9 obs. of 5 variables:
 $ srno      : int  1 2 3 4 5 6 7 8 9
 $ state     : Factor w/ 3 levels "alabama","california",...: 3 1 3 3 2 3 1 2 2
 $ zipcode   : int  32716 35236 32116 34567 32182 32415 32987 32110 32998
 $ familysize : int  5 7 4 2 3 4 6 8 1
 $ annualincome: int  7000 5000 4000 3000 2000 6000 8000 10000 2000
```

Hide

```
#transform familysize(numeric variable) to factor variable
my_data$familysize <- as.factor(my_data$familysize)
my_data$familysize <- (my_data$familysize)
```

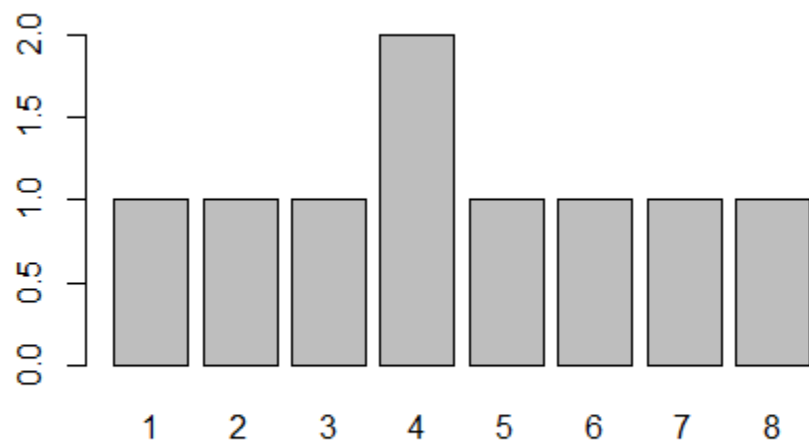
Hide

```
#plot annualsize
plot(my_data$annualincome)
```



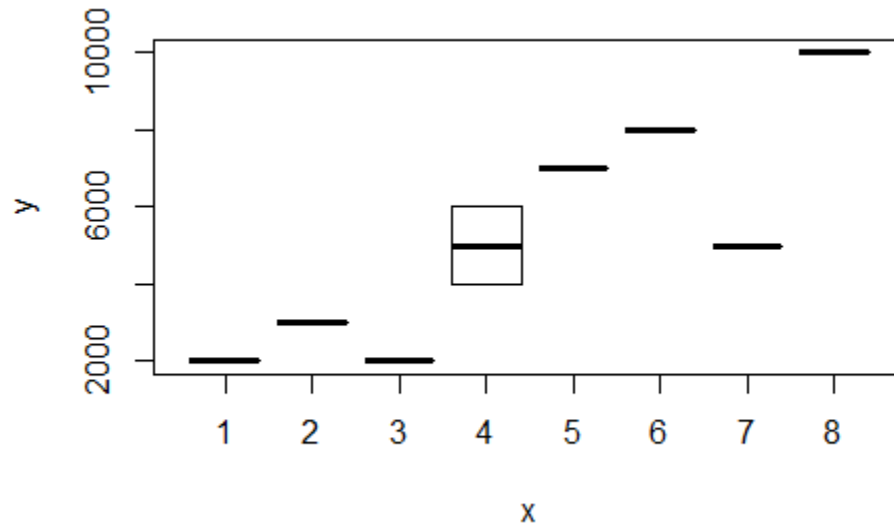
Hide

```
plot(my_data$familysize)
```



Hide

```
#scatterplot annualsize  
plot(my_data$familysize , my_data$annualincome)
```



Hide

```
#converting numeric to character  
class(my_data$familysize)
```

```
[1] "factor"
```

Hide

```
str<-as.character(my_data$familysize)  
class(str)
```

```
[1] "character"
```

Hide

```
class(my_data$annualincome)
```

```
[1] "integer"
```

[Hide](#)

```
mn<-as.matrix(my_data$annualincome)
class(mn)
```

```
[1] "matrix"
```

[Hide](#)

```
#to fetch info about data rowwise and column wise
my_data[2,]
```

	<b>srno</b> <int>	<b>state</b> <fctr>	<b>zipcode</b> <int>	<b>familysize</b> <fctr>	<b>annualincome</b> <int>
	2	2 alabama	35236	7	5000

1 row

[Hide](#)

```
my_data[,2]
```

```
[1] florida    alabama    florida    florida    california
[6] florida    alabama    california california
Levels: alabama california florida
```