FML

Manaswini

2024-02-05

R Markdown

#I took this dataset from github, I preferred this dataset because it has both quantitative and qualitative variables.

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.3.2
```

```
my_data <- read.csv("C:/Users/Winnie/Downloads/HousePrices.csv")
```

summary() function gives brief about the data

```
summary(my_data)
```

```
##     rownames         price           lotsize         bedrooms
##  Min.   :  1.0   Min.   : 25000   Min.   : 1650   Min.   :1.000
##  1st Qu.:137.2   1st Qu.: 49125   1st Qu.: 3600   1st Qu.:2.000
##  Median :273.5   Median : 62000   Median : 4600   Median :3.000
##  Mean   :273.5   Mean   : 68122   Mean   : 5150   Mean   :2.965
##  3rd Qu.:409.8   3rd Qu.: 82000   3rd Qu.: 6360   3rd Qu.:3.000
##  Max.   :546.0   Max.   :190000   Max.   :16200   Max.   :6.000
##    bathrooms        stories        driveway          recreation
##  Min.   :1.000   Min.   :1.000   Length:546         Length:546
##  1st Qu.:1.000   1st Qu.:1.000   Class :character   Class :character
##  Median :1.000   Median :2.000   Mode  :character   Mode  :character
##  Mean   :1.286   Mean   :1.808
##  3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :4.000   Max.   :4.000
##    fullbase           gasheat             aircon              garage
##  Length:546         Length:546         Length:546         Min.   :0.0000
##  Class :character   Class :character   Class :character   1st Qu.:0.0000
##  Mode  :character   Mode  :character   Mode  :character   Median :0.0000
##                                                           Mean   :0.6923
##                                                           3rd Qu.:1.0000
##                                                           Max.   :3.0000
##    prefer
##  Length:546
##  Class :character
##  Mode  :character
##
##
##
```

#QUANTITATIVE VARIABLE here I used summary() function and I have choosen “price” as my quantitative variable.

```
summary(my_data$price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25000   49125   62000   68122   82000  190000
```

#QUALITATIVE VARIABLE I have choosen lotsize as Qualitative variable.

```
table(my_data$lotsize)
```

```
##
##  1650  1700  1836  1905  1950  2000  2015  2135  2145  2160  2175  2176  2275
##     1     1     1     1     2     1     1     1     6     1     1     1     1
##  2325  2398  2400  2430  2475  2500  2520  2550  2610  2640  2650  2684  2700
##     1     1     2     1     1     1     1     1     2     1     1     1     2
##  2747  2787  2800  2817  2835  2850  2856  2870  2880  2910  2953  2970  2990
##     1     2     2     1     1     1     1     1     2     2     1     1     1
##  3000  3036  3040  3060  3069  3090  3100  3120  3150  3162  3180  3185  3210
##    14     1     1     1     1     2     1     3     4     1     7     3     1
##  3240  3264  3290  3300  3350  3360  3400  3410  3420  3450  3460  3480  3500
##     3     1     1     2     1     1     1     1     4     5     2     5     6
##  3510  3512  3520  3540  3570  3584  3600  3620  3630  3635  3640  3649  3650
##     2     1     5     1     1     1     8     1     7     1     7     1     2
##  3660  3680  3700  3720  3745  3750  3760  3780  3792  3800  3816  3840  3850
##     1     1     1     1     1     3     2     1     1     2     1     1     5
##  3880  3900  3930  3934  3960  3968  3970  3986  3990  4000  4032  4040  4046
##     1     1     1     1     3     1     2     1     1    11     1     7     1
##  4050  4075  4079  4080  4095  4100  4120  4130  4160  4200  4240  4260  4280
##     2     1     1     2     2     3     2     1     2     1     1     2     1
##  4300  4320  4340  4350  4352  4360  4370  4400  4410  4500  4510  4520  4560
##     2     4     1     2     1     1     1     4     2    13     2     1     1
##  4600  4632  4640  4646  4700  4750  4770  4775  4785  4800  4815  4820  4840
##     5     1     1     1     1     1     1     1     1     5     1     1     2
##  4880  4900  4920  4950  4960  4990  4992  4995  5000  5010  5020  5040  5076
##     2     2     1     1     3     1     1     1     4     1     1     1     1
##  5136  5150  5170  5200  5300  5320  5360  5400  5450  5495  5500  5600  5640
##     1     1     1     2     3     2     1     6     2     1     9     1     1
##  5680  5700  5720  5750  5800  5828  5830  5850  5880  5885  5900  5948  5960
##     1     1     1     1     3     1     1     2     2     1     2     1     2
##  5985  6000  6020  6040  6050  6060  6100  6210  6240  6254  6300  6321  6325
##     1    24     1     1     1     2     3     1     1     1     1     1     1
##  6350  6360  6400  6420  6440  6450  6480  6500  6525  6540  6550  6600  6615
##     1     7     1     4     1     2     2     2     1     3     2     9     2
##  6650  6660  6670  6710  6720  6750  6800  6825  6840  6862  6900  6930  7000
##     1     1     1     1     1     2     2     1     1     1     2     1     5
##  7020  7085  7152  7155  7160  7200  7231  7260  7320  7350  7410  7420  7424
##     1     1     1     1     1     2     1     1     1     1     1     2     1
##  7440  7475  7482  7500  7600  7680  7686  7700  7770  7800  7950  7980  8000
##     2     1     1     2     1     1     1     2     1     2     1     2     1
##  8050  8080  8100  8150  8250  8372  8400  8500  8520  8580  8800  8875  8880
##     2     1     3     1     4     1     3     1     1     2     1     1     2
##  8960  9000  9166  9500  9620  9667  9800  9860  9960 10240 10269 10360 10500
##     1     4     1     1     1     1     1     1     1     1     1     1     3
## 10700 11175 11410 11440 11460 12090 12900 12944 13200 15600 16200
##     1     1     1     1     1     1     1     1     2     1     1
```

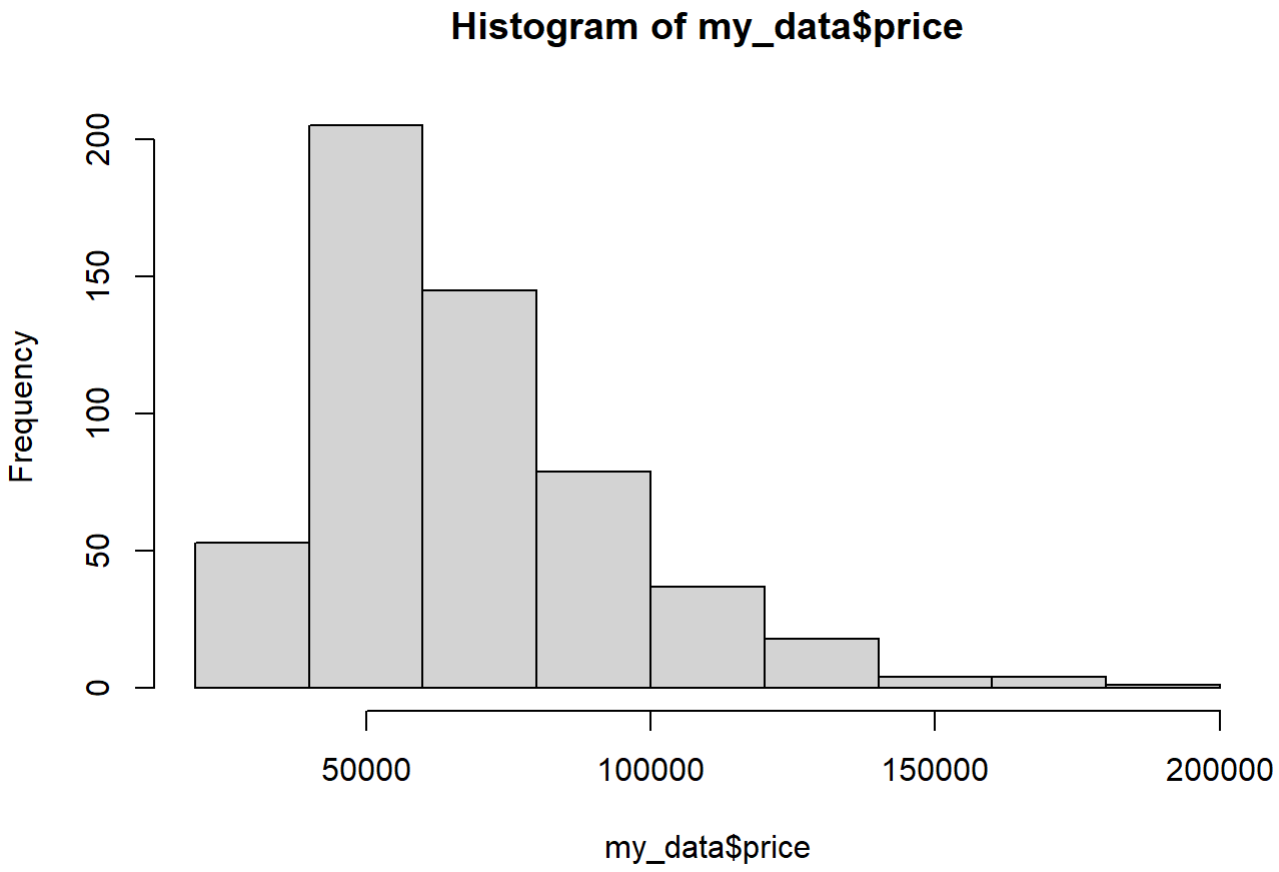
#TRANSFORMATION OF QUANTITATIVE VARIABLE I have transformed my quantitative variable which is Price.

```
my_data$log_transformed_price <- log(my_data$price)
head(my_data)
```

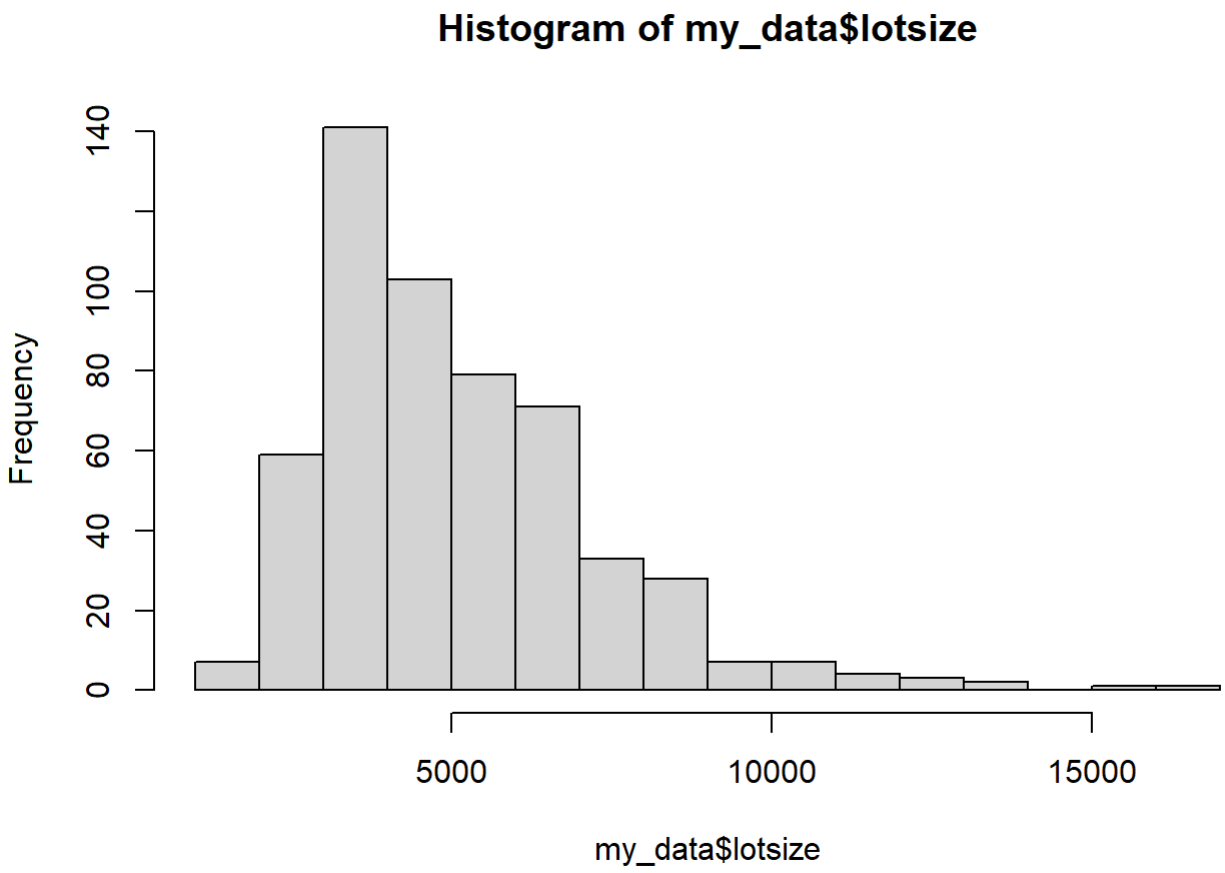
```
##   rownames price lotsize bedrooms bathrooms stories driveway recreation
## 1        1 42000    5850        3         1       2      yes         no
## 2        2 38500    4000        2         1       1      yes         no
## 3        3 49500    3060        3         1       1      yes         no
## 4        4 60500    6650        3         1       2      yes        yes
## 5        5 61000    6360        2         1       1      yes         no
## 6        6 66000    4160        3         1       1      yes        yes
##   fullbase gasheat aircon garage prefer log_transformed_price
## 1      yes      no     no      1     no              10.64542
## 2       no      no     no      0     no              10.55841
## 3       no      no     no      0     no              10.80973
## 4       no      no     no      0     no              11.01040
## 5       no      no     no      0     no              11.01863
## 6      yes      no    yes      0     no              11.09741
```

#HISTOGRAM

```
hist(my_data$price)
```

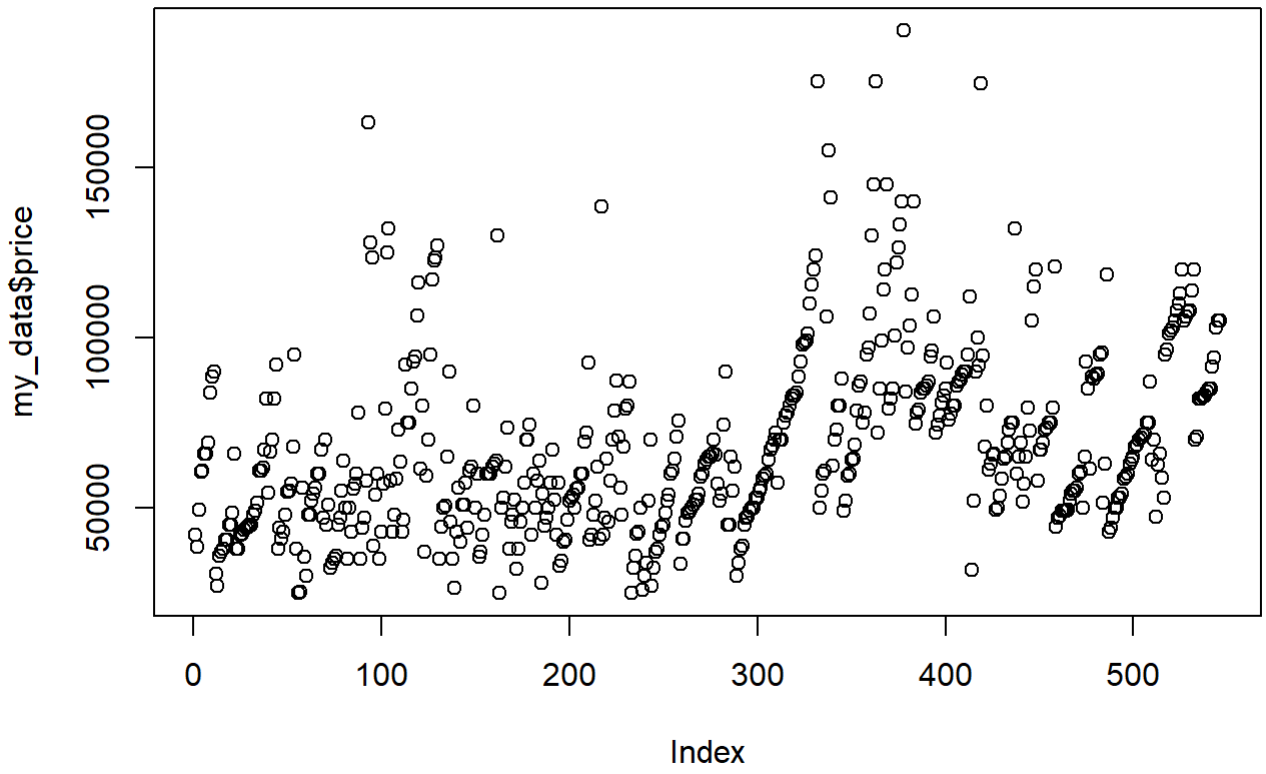



```
hist(my_data$lotsize)
```

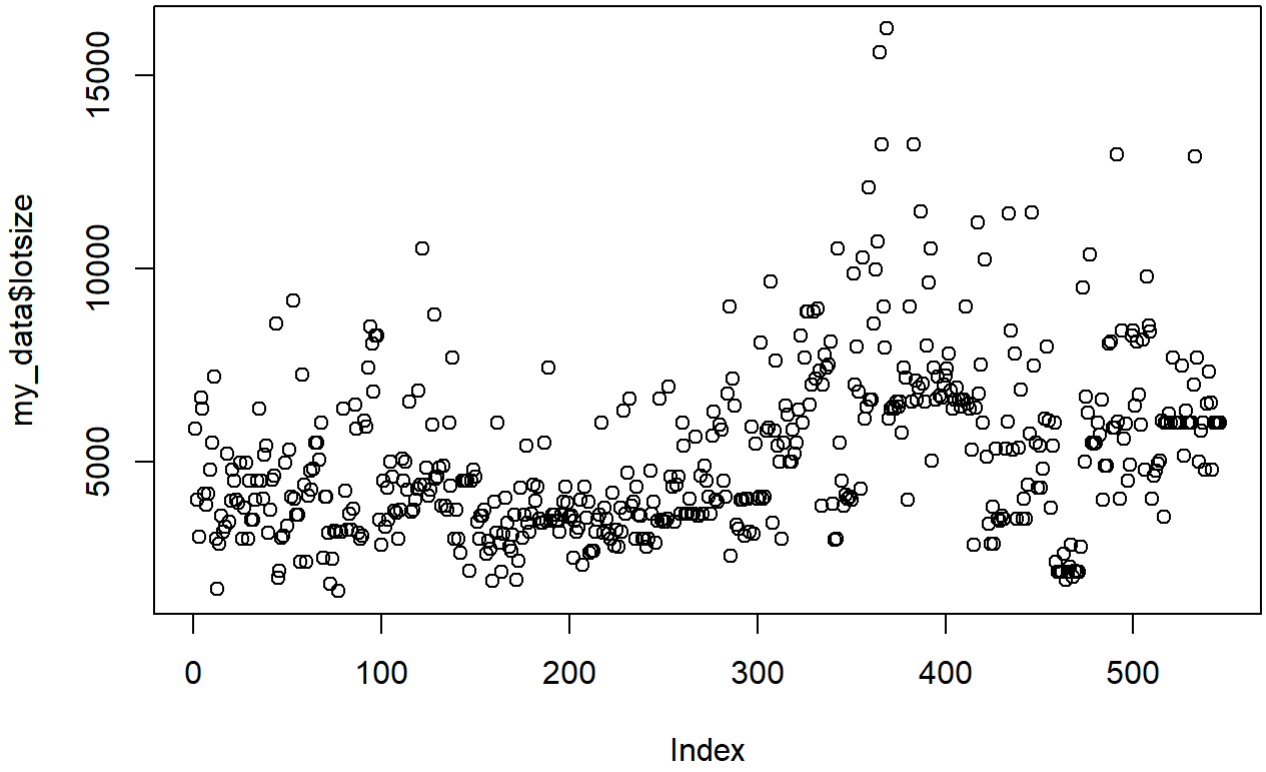


#SCATTERPLOT

```
plot(my_data$price)
```

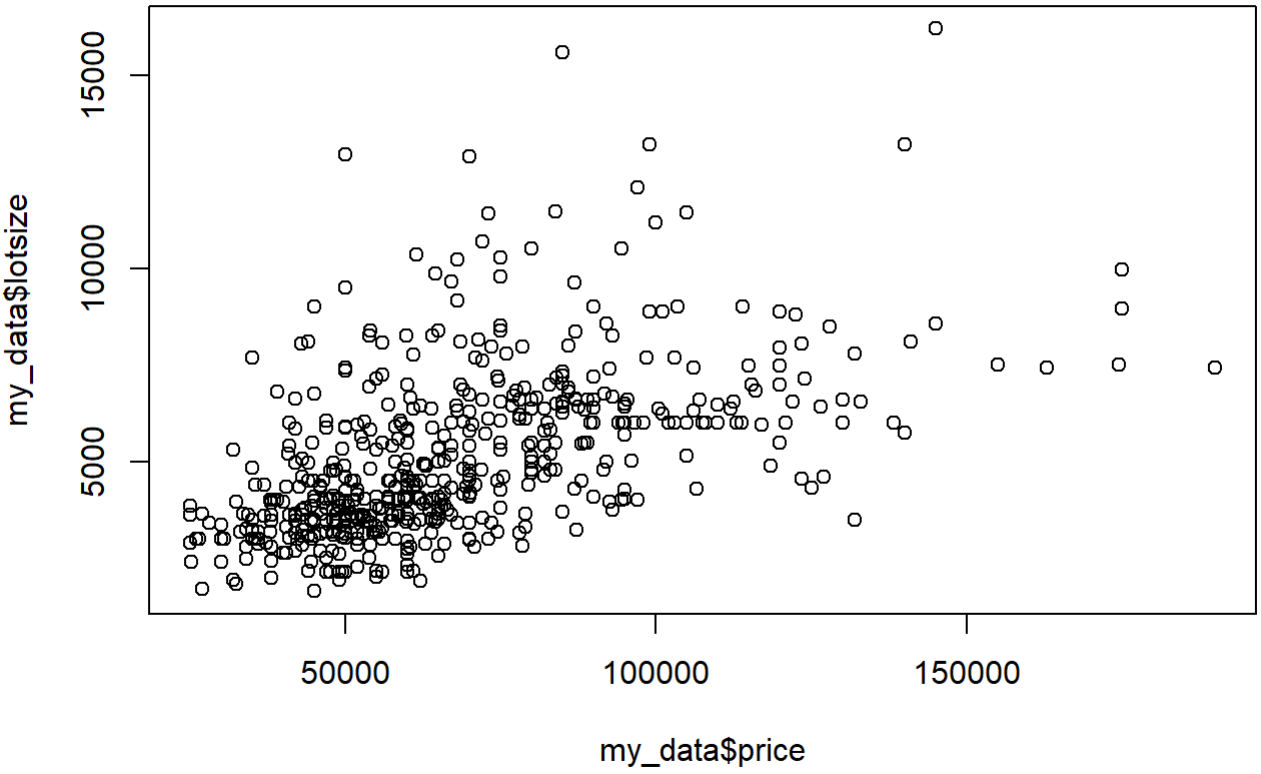



```
plot(my_data$lotsize)
```



#SCATTERPLOT OF BOTH QUANTITATIVE AND QUALITATIVE VARIABLES

```
plot(my_data$price, my_data$lotsize)
```


my_data$lotsize
5000
10000
15000
50000
100000
150000
my_data$price