Problem statement1:

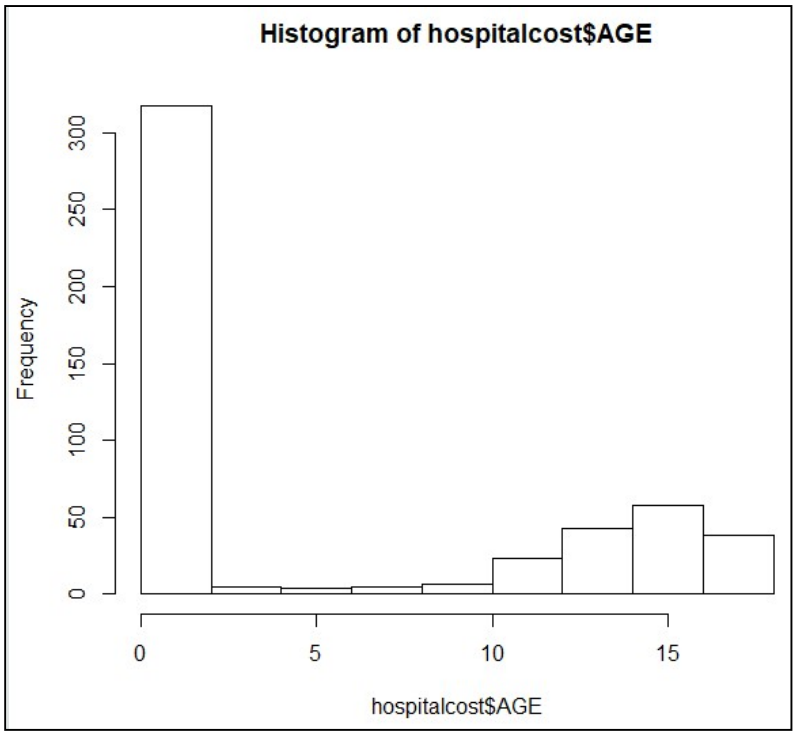
To record the patient statistics, the agency wants to find the age category of people who frequent the hospital and has the maximum expenditure.

R code

```
********** PROJECT HEALTH CARE **********

QUESTION 1

Healthcare <- read.csv("HospitalCosts.csv")
View(Healthcare)

COUNT_AGE = table(Healthcare$AGE)
COUNT_AGE

Avg_of_TOTCHG = aggregate(TOTCHG~AGE,Healthcare,mean)
Avg_of_TOTCHG

QUEST1 = cbind(COUNT_AGE,Avg_of_TOTCHG)
QUEST1

hist(hospitalcost$AGE)

max(COUNT_AGE)

max(Avg_of_TOTCHG)
```

```
> Healthcare <- read.csv("HospitalCosts.csv")
> View(Healthcare)
> COUNT_AGE = table(Healthcare$AGE)
> COUNT_AGE

  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
307  10   1   3   2   2   2   3   2   2   4   8  15  18  25  29  29  38
> Avg_of_TOTCHG = aggregate(TOTCHG~AGE,Healthcare,mean)
> Avg_of_TOTCHG
   AGE    TOTCHG
1    0  2208.853
2    1  3774.400
3    2  7298.000
4    3 10183.333
5    4  7996.000
6    5  9253.500
7    6  8964.000
8    7  3362.333
9    8  2370.500
10   9 10573.500
11  10  6117.250
12  11  1781.250
13  12  3660.800
14  13  1729.722
15  14  2585.720
16  15  3853.345
17  16  2384.448
18  17  4599.395
>
```

```
> QUEST1 = cbind(COUNT_AGE,Avg_of_TOTCHG)
> QUEST1
   Var1 Freq AGE    TOTCHG
1     0  307   0  2208.853
2     1   10   1  3774.400
3     2    1   2  7298.000
4     3    3   3 10183.333
5     4    2   4  7996.000
6     5    2   5  9253.500
7     6    2   6  8964.000
8     7    3   7  3362.333
9     8    2   8  2370.500
10    9    2   9 10573.500
11   10    4  10  6117.250
12   11    8  11  1781.250
13   12   15  12  3660.800
14   13   18  13  1729.722
15   14   25  14  2585.720
16   15   29  15  3853.345
17   16   29  16  2384.448
18   17   38  17  4599.395
> hist(hospitalcost$AGE)
> hist(hospitalcost$AGE)
> max(COUNT_AGE)
[1] 307
> max(Avg_of_TOTCHG)
[1] 10573.5
```



Based on the output we can see that Age wise Hospital Visit and Expenses

Maximum Hospital Visit – 0-1 yrs age group: 307

Maximum Expenditure – 0-1yrs age group: 10573.5

Problem statement2:

In order of severity of the diagnosis and treatments and to find out the expensive treatments, the agency wants to find the diagnosis related group that has maximum hospitalization and expenditure.

```
27  QUESTION 2
28
29  APRDRG_LOS = aggregate(LOS~APRDRG,Healthcare,mean)
30  APRDRG_LOS
31
32  APRDRG_TOTCHG = aggregate(TOTCHG~APRDRG,Healthcare,mean)
33  APRDRG_TOTCHG
34
35  QUEST2 = cbind(APRDRG_LOS,APRDRG_TOTCHG)
36  QUEST2
37
38  max(APRDRG_LOS)
39
40  max(APRDRG_TOTCHG)
```

```
> APRDRG_LOS = aggregate(LOS~APRDRG,Healthcare,mean)
> APRDRG_LOS
   APRDRG        LOS
1      21  2.0000000
2      23  2.0000000
3      49  6.0000000
4      50  2.0000000
5      51  3.0000000
6      53  2.9000000
7      54  1.0000000
8      57  1.0000000
9      58  1.0000000
10     92  1.0000000
11     97  3.0000000
12    114  3.0000000
13    115  4.0000000
14    137 12.0000000
15    138  1.5000000
16    139  1.4000000
17    141  2.0000000
18    143  2.0000000
19    204  3.0000000
20    206  3.0000000
21    225  5.5000000
22    249  1.3333333
23    254  0.0000000
24    308  1.0000000
25    313  1.0000000
26    317  7.0000000
27    344  2.5000000
28    347  0.3333333
29    420  1.5000000
30    421 39.0000000
31    422  1.6666667
```

```
31    422  1.6666667
32    560  2.0000000
33    561  4.0000000
34    566  2.0000000
35    580  1.0000000
36    581  0.3333333
37    602 41.0000000
38    614 15.0000000
39    626  6.0000000
40    633  6.2500000
41    634  4.5000000
42    636  6.6666667
43    639  4.5000000
44    640  2.4419476
45    710  4.0000000
46    720  5.0000000
47    723  1.0000000
48    740  5.0000000
49    750  3.0000000
50    751  2.2857143
51    753  3.0277778
52    754  2.2972973
53    755  1.0000000
54    756  1.0000000
55    758  2.6500000
56    760  7.0000000
57    776  2.0000000
58    811  0.5000000
59    812  1.6666667
60    863 24.0000000
61    911  7.0000000
62    930  5.0000000
63    952  1.0000000
```

```
> APRDRG_TOTCHG = aggregate(TOTCHG~APRDRG,Healthcare,mean)
> APRDRG_TOTCHG
   APRDRG     TOTCHG
1      21 10002.0000
2      23 14174.0000
3      49 20195.0000
4      50  3908.0000
5      51  3023.0000
6      53  8227.1000
7      54   851.0000
8      57  7254.5000
9      58  2117.0000
10     92 12024.0000
11     97  9530.0000
12    114 10562.0000
13    115 12916.0000
14    137 15129.0000
15    138  3405.5000
16    139  3553.2000
17    141  2860.0000
18    143  1393.0000
19    204  8439.0000
20    206  9230.0000
21    225 12824.5000
22    249  2773.6667
23    254   615.0000
24    308 10585.0000
25    313  8159.0000
26    317 17524.0000
27    344  7401.0000
28    347  4199.0000
29    420  3178.5000
30    421 26356.0000
31    422  1725.6667
32    560  2438.5000
```

```
28    347  4199.0000
29    420  3178.5000
30    421 26356.0000
31    422  1725.6667
32    560  2438.5000
33    561  2296.0000
34    566  2129.0000
35    580  2825.0000
36    581  2484.3333
37    602 29188.0000
38    614  9177.0000
39    626  3881.5000
40    633  4397.7500
41    634  4976.0000
42    636  7741.3333
43    639  3153.0000
44    640  1640.3670
45    710  8223.0000
46    720 14243.0000
47    723  2644.5000
48    740 11125.0000
49    750  1753.0000
50    751  1547.5714
51    753  2209.5000
52    754  1598.6486
53    755   859.0769
54    756   747.0000
55    758  1747.6500
56    760  4136.5000
57    776  1193.0000
58    811  1919.0000
59    812  3174.6667
60    863 13040.0000
61    911 48388.0000
62    930 13327.0000
63    952  4833.0000
```

```
> QUEST2 = cbind(APRDRG_LOS,APRDRG_TOTCHG)
> QUEST2
   APRDRG        LOS APRDRG     TOTCHG
1      21  2.0000000     21 10002.0000
2      23  2.0000000     23 14174.0000
3      49  6.0000000     49 20195.0000
4      50  2.0000000     50  3908.0000
5      51  3.0000000     51  3023.0000
6      53  2.9000000     53  8227.1000
7      54  1.0000000     54   851.0000
8      57  1.0000000     57  7254.5000
9      58  1.0000000     58  2117.0000
10     92  1.0000000     92 12024.0000
11     97  3.0000000     97  9530.0000
12    114  3.0000000    114 10562.0000
13    115  4.0000000    115 12916.0000
14    137 12.0000000    137 15129.0000
15    138  1.5000000    138  3405.5000
16    139  1.4000000    139  3553.2000
17    141  2.0000000    141  2860.0000
18    143  2.0000000    143  1393.0000
19    204  3.0000000    204  8439.0000
20    206  3.0000000    206  9230.0000
21    225  5.5000000    225 12824.5000
22    249  1.3333333    249  2773.6667
23    254  0.0000000    254   615.0000
24    308  1.0000000    308 10585.0000
25    313  1.0000000    313  8159.0000
26    317  7.0000000    317 17524.0000
27    344  2.5000000    344  7401.0000
28    347  0.3333333    347  4199.0000
29    420  1.5000000    420  3178.5000
30    421 39.0000000    421 26356.0000
```

```
30    421 39.0000000    421 26356.0000
31    422  1.6666667    422  1725.6667
32    560  2.0000000    560  2438.5000
33    561  4.0000000    561  2296.0000
34    566  2.0000000    566  2129.0000
35    580  1.0000000    580  2825.0000
36    581  0.3333333    581  2484.3333
37    602 41.0000000    602 29188.0000
38    614 15.0000000    614  9177.0000
39    626  6.0000000    626  3881.5000
40    633  6.2500000    633  4397.7500
41    634  4.5000000    634  4976.0000
42    636  6.6666667    636  7741.3333
43    639  4.5000000    639  3153.0000
44    640  2.4419476    640  1640.3670
45    710  4.0000000    710  8223.0000
46    720  5.0000000    720 14243.0000
47    723  1.0000000    723  2644.5000
48    740  5.0000000    740 11125.0000
49    750  3.0000000    750  1753.0000
50    751  2.2857143    751  1547.5714
51    753  3.0277778    753  2209.5000
52    754  2.2972973    754  1598.6486
53    755  1.0000000    755   859.0769
54    756  1.0000000    756   747.0000
55    758  2.6500000    758  1747.6500
56    760  7.0000000    760  4136.5000
57    776  2.0000000    776  1193.0000
58    811  0.5000000    811  1919.0000
59    812  1.6666667    812  3174.6667
60    863 24.0000000    863 13040.0000
61    911  7.0000000    911 48388.0000
62    930  5.0000000    930 13327.0000
63    952  1.0000000    952  4833.0000
```

```
> max(APRDRG_LOS)
[1] 952
> max(APRDRG_TOTCHG)
[1] 48388
>
```

Based on the output we can see the list of Expenditure based on the Diagnosis and treatment.

TOTCHG - Hospital discharge costs: 48388

APRDRG – All patient Refined Diagnosis Related Groups: 952

Problem statement 3:

To make sure that there is no malpractice, the agency needs to analyze if the race of the patient is related to the hospitalization costs.

```
43  QUESTION 3
44
45  #### ANOVA
46  Race_mean = aggregate(TOTCHG~RACE,Healthcare,mean)
47  Race_mean
48
49
50  Aov_Hospitalcost = aov(TOTCHG~RACE,Race_mean)
51  summary(Aov_Hospitalcost)
```

```
> Race_mean = aggregate(TOTCHG~RACE,Healthcare,mean)
> Race_mean
  RACE   TOTCHG
1    1 2772.669
2    2 4202.167
3    3 3041.000
4    4 2344.667
5    5 2026.667
6    6 1349.000
> Aov_Hospitalcost = aov(TOTCHG~RACE,Race_mean)
> summary(Aov_Hospitalcost)
            Df  Sum Sq Mean Sq F value Pr(>F)
RACE         1 2938135 2938135   6.497 0.0634 .
Residuals    4 1808916  452229
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

There is no extra charge from one race than another.

Problem statement 4:

To properly utilize the costs, the agency has to analyze the severity of the hospital costs by age and gender for proper allocation of resources.

```
53  QUESTION 4
54
55
56  AGE_GENDER = aggregate(TOTCHG~AGE+FEMALE,Healthcare,mean)
57  AGE_GENDER
```

```
> AGE_GENDER = aggregate(TOTCHG~AGE+FEMALE,Healthcare,mean)
> AGE_GENDER
   AGE FEMALE    TOTCHG
1    0      0  2198.435
2    1      0  4327.750
3    2      0  7298.000
4    3      0 11163.500
5    4      0  9230.000
6    5      0  7923.000
7    6      0  8964.000
8    7      0  3362.333
9    8      0  2370.500
10   9      0 10573.500
11  10      0  7769.667
12  11      0  1468.000
13  12      0  2592.167
14  13      0  1054.000
15  14      0  5741.000
16  15      0  7223.000
17  16      0  4629.833
18  17      0  3961.154
19   0      1  2221.781
20   1      1  1561.000
21   3      1  8223.000
22   4      1  6762.000
23   5      1 10584.000
24  10      1  1160.000
25  11      1  2721.000
26  12      1  4373.222
27  13      1  1922.786
28  14      1  1984.714
29  15      1  2079.842
30  16      1  1798.696
31  17      1  4931.280
```

Total Charges obtained based on AGE / GENDER (MALR or FEMALE).

Problem statement 5:

Since the length of stay is the crucial factor for inpatients, the agency wants to find if the length of stay can be predicted from age, gender, and race.

```
QUESTION 5

LOS_lm = lm(LOS~AGE+FEMALE+RACE,Healthcare)
LOS_lm
summary(LOS_lm)


## Make predictions
Predicted_Price = predict(LOS_lm,Healthcare)
View(Predicted_Price)
Healthcare_Final = cbind(Healthcare,Predicted_Price)
View(Healthcare_Final)


install.packages("dplyr")

Healthcare_Error_pct = transform(Healthcare_Final,Error_pct = (abs(LOS-Predicted_Price)/LOS))

library(dplyr)

Healthcare_Error_pct_Final =  filter(Healthcare_Error_pct, LOS != 0)

OverallErrorRate = mean(Healthcare_Error_pct_Final$Error_pct,na.rm=T)
OverallErrorRate

AccuracyRate = 1- OverallErrorRate
AccuracyRate

View(Healthcare_Error_pct_Final)
```

```
> LOS_lm = lm(LOS~AGE+FEMALE+RACE,Healthcare)
> LOS_lm

Call:
lm(formula = LOS ~ AGE + FEMALE + RACE, data = Healthcare)

Coefficients:
(Intercept)          AGE       FEMALE         RACE
    2.94377     -0.03960      0.37011     -0.09408

> summary(LOS_lm)

Call:
lm(formula = LOS ~ AGE + FEMALE + RACE, data = Healthcare)

Residuals:
   Min     1Q Median     3Q    Max
 -3.22  -1.22  -0.85   0.15  37.78

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.94377    0.39318   7.487 3.25e-13 ***
AGE         -0.03960    0.02231  -1.775   0.0766 .
FEMALE       0.37011    0.31024   1.193   0.2334
RACE        -0.09408    0.29312  -0.321   0.7484
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.363 on 495 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.007898,  Adjusted R-squared:  0.001886
F-statistic: 1.314 on 3 and 495 DF,  p-value: 0.2692
```

For inpatients, Length of stay calculated based on AGE / FEMALE / RACE.

```
> ## Make predictions
> Predicted_Price = predict(LOS_lm,Healthcare)
> View(Predicted_Price)
> Healthcare_Final = cbind(Healthcare,Predicted_Price)
> View(Healthcare_Final)
```

Filter

	AGE	FEMALE	LOS	RACE	TOTCHG	APRDRG	Predicted_Price
1	17	1	2	1	2660	560	2.546578
2	17	0	2	1	1689	753	2.176465
3	17	1	7	1	20060	930	2.546578
4	17	1	1	1	736	758	2.546578
5	17	1	1	1	1194	754	2.546578
6	17	0	0	1	3305	347	2.176465
7	17	1	4	1	2205	754	2.546578
8	16	1	2	1	1167	754	2.586180
9	16	1	1	1	532	753	2.586180
10	17	1	2	1	1363	758	2.546578
11	17	1	2	1	1245	758	2.546578
12	15	0	2	1	1656	753	2.255668
13	15	1	2	1	1379	751	2.625782
14	15	1	4	1	2346	758	2.625782
15	15	1	7	1	4006	753	2.625782
16	15	1	4	1	2181	758	2.625782
17	14	1	1	1	628	754	2.665383
18	14	1	4	1	2463	758	2.665383

19	15	1	3	1	1956	753	2.625782
20	14	1	3	1	1802	758	2.665383
21	13	1	1	1	3188	812	2.704985
22	17	1	2	1	2129	566	2.546578
23	12	0	1	1	7421	249	2.374473
24	15	1	1	1	1122	422	2.625782
25	13	1	2	4	1173	754	2.422759
26	12	0	2	1	3625	812	2.374473
27	11	1	2	1	3908	50	2.784188
28	15	0	1	1	3994	139	2.255668
29	11	0	0	1	1033	753	2.414075
30	10	0	2	1	2860	141	2.453677
31	11	0	2	1	3814	420	2.414075
32	7	0	0	1	1132	139	2.572482
33	16	1	2	6	1163	751	2.115803
34	17	1	1	1	610	751	2.546578

```
> library(dplyr)
>
> Healthcare_Error_pct_Final =  filter(Healthcare_Error_pct, LOS != 0)
>
> OverallErrorRate = mean(Healthcare_Error_pct_Final$Error_pct,na.rm=T)
> OverallErrorRate
[1] 0.5690605
>
> AccuracyRate = 1- OverallErrorRate
> AccuracyRate
[1] 0.4309395
```

HEALTHCARE PROJECT.R | Healthcare_Error_pct_Final | Predicted_Price | Healthcare_Final | 6.

Filter

	AGE	FEMALE	LOS	RACE	TOTCHG	APRDRG	Predicted_Price	Error_pct
1	17	1	2	1	2660	560	2.546578	0.27328916
2	17	0	2	1	1689	753	2.176465	0.08823234
3	17	1	7	1	20060	930	2.546578	0.63620310
4	17	1	1	1	736	758	2.546578	1.54657831
5	17	1	1	1	1194	754	2.546578	1.54657831
6	17	1	4	1	2205	754	2.546578	0.36335542
7	16	1	2	1	1167	754	2.586180	0.29309001
8	16	1	1	1	532	753	2.586180	1.58618001
9	17	1	2	1	1363	758	2.546578	0.27328916
10	17	1	2	1	1245	758	2.546578	0.27328916
11	15	0	2	1	1656	753	2.255668	0.12783404
12	15	1	2	1	1379	751	2.625782	0.31289085
13	15	1	4	1	2346	758	2.625782	0.34355457
14	15	1	7	1	4006	753	2.625782	0.62488833
15	15	1	4	1	2181	758	2.625782	0.34355457
16	14	1	1	1	628	754	2.665383	1.66538341
17	14	1	4	1	2463	758	2.665383	0.33365415

HEALTHCARE PROJECT.R | Healthcare_Error_pct_Final | Predicted_Price | Healthcare_Final | 6.

Filter

	AGE	FEMALE	LOS	RACE	TOTCHG	APRDRG	Predicted_Price	Error_pct
17	14	1	4	1	2463	758	2.665383	0.33365415
18	15	1	3	1	1956	753	2.625782	0.12473943
19	14	1	3	1	1802	758	2.665383	0.11153886
20	13	1	1	1	3188	812	2.704985	1.70498510
21	17	1	2	1	2129	566	2.546578	0.27328916
22	12	0	1	1	7421	249	2.374473	1.37447317
23	15	1	1	1	1122	422	2.625782	1.62578171
24	13	1	2	4	1173	754	2.422759	0.21137958
25	12	0	2	1	3625	812	2.374473	0.18723659
26	11	1	2	1	3908	50	2.784188	0.39209425
27	15	0	1	1	3994	139	2.255668	1.25566808
28	10	0	2	1	2860	141	2.453677	0.22683828
29	11	0	2	1	3814	420	2.414075	0.20703744
30	16	1	2	6	1163	751	2.115803	0.05790172
31	17	1	1	1	610	751	2.546578	1.54657831
32	6	0	3	1	9530	97	2.612083	0.12930555
33	15	1	1	1	1268	811	2.625782	1.62578171

Problem statement 6:

To perform a complete analysis, the agency wants to find the variable that mainly affects the hospital costs.

```
QUESTION 6

ALL_VAR_lm = lm(TOTCHG~AGE+FEMALE+LOS+RACE+APRDRG,Healthcare)
ALL_VAR_lm
summary(ALL_VAR_lm)


## Make predictions
Predicted_Price = predict(ALL_VAR_lm,Healthcare)
View(Predicted_Price)

Healthcare_Final_All = cbind(Healthcare,Predicted_Price)
View(Healthcare_Final_All)


Healthcare_Error_pct_All = transform(Healthcare_Final_All,Error_pct = (abs(TOTCHG-Predicted_Price)/TOTCHG))

library(dplyr)

Healthcare_Error_pct_All_Final =  filter(Healthcare_Error_pct_All, LOS != 0)

OverallErrorRate_All = mean(Healthcare_Error_pct_All_Final$Error_pct,na.rm=T)
OverallErrorRate_All

AccuracyRate_All = 1- OverallErrorRate_All
AccuracyRate_All

View(Healthcare_Error_pct_Final)
```

Console | Terminal | Jobs

~/Apr Batch/

```
> ALL_VAR_lm = lm(TOTCHG~AGE+FEMALE+LOS+RACE+APRDRG,Healthcare)
> ALL_VAR_lm

Call:
lm(formula = TOTCHG ~ AGE + FEMALE + LOS + RACE + APRDRG, data = Healthcare)

Coefficients:
(Intercept)          AGE       FEMALE          LOS         RACE       APRDRG
   5218.677      134.695     -390.692      743.152     -212.429       -7.791

> summary(ALL_VAR_lm)

Call:
lm(formula = TOTCHG ~ AGE + FEMALE + LOS + RACE + APRDRG, data = Healthcare)

Residuals:
   Min     1Q Median     3Q    Max
 -6377   -700   -174    122  43378

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5218.6769   507.6475  10.280  < 2e-16 ***
AGE          134.6949    17.4711   7.710 7.02e-14 ***
FEMALE      -390.6924   247.7390  -1.577    0.115
LOS          743.1521    34.9225  21.280  < 2e-16 ***
RACE        -212.4291   227.9326  -0.932    0.352
APRDRG        -7.7909     0.6816 -11.430  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2613 on 493 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5536,    Adjusted R-squared:  0.5491
F-statistic: 122.3 on 5 and 493 DF,  p-value: < 2.2e-16
```

Console | Terminal | Jobs

~/Apr Batch/

```
> ## Make predictions
> Predicted_Price = predict(ALL_VAR_lm,Healthcare)
> View(Predicted_Price)
>
> Healthcare_Final_All = cbind(Healthcare,Predicted_Price)
> View(Healthcare_Final_All)
```

HEALTHCARE PROJECT.R | Predicted_Price | Healthcare_Final_All

Filter

	AGE	FEMALE	LOS	RACE	TOTCHG	APRDRG	Predicted_Price
1	17	1	2	1	2660	560	4028.7415
2	17	0	2	1	1689	753	2915.7810
3	17	1	7	1	20060	930	4861.8511
4	17	1	1	1	736	758	1742.9817
5	17	1	1	1	1194	754	1774.1455
6	17	0	0	1	3305	347	4592.6017
7	17	1	4	1	2205	754	4003.6017
8	16	1	2	1	1167	754	2382.6027
9	16	1	1	1	532	753	1647.2416
10	17	1	2	1	1363	758	2486.1338
11	17	1	2	1	1245	758	2486.1338
12	15	0	2	1	1656	753	2646.3912
13	15	1	2	1	1379	751	2271.2807
14	15	1	4	1	2346	758	3703.0482
15	15	1	7	1	4006	753	5971.4591
16	15	1	4	1	2181	758	3703.0482
17	14	1	1	1	628	754	1370.0609
18	14	1	4	1	2463	758	3568.3533
19	15	1	3	1	1956	753	2998.8508
20	14	1	3	1	1802	758	2825.2012
21	13	1	1	1	3188	812	783.4910

HEALTHCARE PROJECT.R × | Predicted_Price × | Healthcare_Final_All ×

Filter

	AGE	FEMALE	LOS	RACE	TOTCHG	APRDRG	Predicted_Price
21	13	1	1	1	3188	812	783.4910
22	17	1	2	1	2129	566	3981.9958
23	12	0	1	1	7421	249	5425.7924
24	15	1	1	1	1122	422	4091.3505
25	13	1	2	4	1173	754	1341.2309
26	12	0	2	1	3625	812	1782.6406
27	11	1	2	1	3908	50	7193.9558
28	15	0	1	1	3994	139	6686.8813
29	11	0	0	1	1033	753	621.3076
30	10	0	2	1	2860	141	6740.9771
31	11	0	2	1	3814	420	4701.9974
32	7	0	0	1	1132	139	4866.1702
33	16	1	2	1	1163	751	1343.8303
34	17	1	1	1	610	751	1797.5183
35	6	0	3	1	9530	97	7288.1514
36	15	1	1	1	1268	811	1060.6717
37	17	1	4	1	2582	753	4011.3927
38	16	1	2	1	1287	755	2374.8117
39	17	1	3	1	6594	930	1889.2428
40	13	1	0	1	909	755	484.4230
41	7	0	0	1	2530	347	3245.6530

```
> Healthcare_Final_All = cbind(Healthcare,Predicted_Price)
> View(Healthcare_Final_All)
> Healthcare_Error_pct_All = transform(Healthcare_Final_All,Error_pct = (abs(TOTCHG-Predicted_Price)/TOTCHG))
> Healthcare_Error_pct_All_Final =  filter(Healthcare_Error_pct_All, LOS != 0)
> OverallErrorRate_All = mean(Healthcare_Error_pct_All_Final$Error_pct,na.rm=T)
> OverallErrorRate_All
[1] 0.3877528
> AccuracyRate_All = 1- OverallErrorRate_All
> AccuracyRate_All
[1] 0.6122472
> View(Healthcare_Error_pct_Final)
```

HEALTHCARE PROJECT.R | Healthcare_Error_pct_Final

Filter

	AGE	FEMALE	LOS	RACE	TOTCHG	APRDRG	Predicted_Price	Error_pct
1	17	1	2	1	2660	560	2.546578	0.27328916
2	17	0	2	1	1689	753	2.176465	0.08823234
3	17	1	7	1	20060	930	2.546578	0.63620310
4	17	1	1	1	736	758	2.546578	1.54657831
5	17	1	1	1	1194	754	2.546578	1.54657831
6	17	1	4	1	2205	754	2.546578	0.36335542
7	16	1	2	1	1167	754	2.586180	0.29309001
8	16	1	1	1	532	753	2.586180	1.58618001
9	17	1	2	1	1363	758	2.546578	0.27328916
10	17	1	2	1	1245	758	2.546578	0.27328916
11	15	0	2	1	1656	753	2.255668	0.12783404
12	15	1	2	1	1379	751	2.625782	0.31289085
13	15	1	4	1	2346	758	2.625782	0.34355457
14	15	1	7	1	4006	753	2.625782	0.62488833
15	15	1	4	1	2181	758	2.625782	0.34355457
16	14	1	1	1	628	754	2.665383	1.66538341
17	14	1	4	1	2463	758	2.665383	0.33365415
18	15	1	3	1	1956	753	2.625782	0.12473943
19	14	1	3	1	1802	758	2.665383	0.11153886

HEALTHCARE PROJECT.R × Healthcare_Error_pct_Final ×

Filter

	AGE	FEMALE	LOS	RACE	TOTCHG	APRDRG	Predicted_Price	Error_pct
19	14	1	3	1	1802	758	2.665383	0.11153886
20	13	1	1	1	3188	812	2.704985	1.70498510
21	17	1	2	1	2129	566	2.546578	0.27328916
22	12	0	1	1	7421	249	2.374473	1.37447317
23	15	1	1	1	1122	422	2.625782	1.62578171
24	13	1	2	4	1173	754	2.422759	0.21137958
25	12	0	2	1	3625	812	2.374473	0.18723659
26	11	1	2	1	3908	50	2.784188	0.39209425
27	15	0	1	1	3994	139	2.255668	1.25566808
28	10	0	2	1	2860	141	2.453677	0.22683828
29	11	0	2	1	3814	420	2.414075	0.20703744
30	16	1	2	6	1163	751	2.115803	0.05790172
31	17	1	1	1	610	751	2.546578	1.54657831
32	6	0	3	1	9530	97	2.612083	0.12930555
33	15	1	1	1	1268	811	2.625782	1.62578171
34	17	1	4	1	2582	753	2.546578	0.36335542
35	16	1	2	1	1287	755	2.586180	0.29309001
36	17	1	3	1	6594	930	2.546578	0.15114056
37	11	1	2	2	1534	753	2.690113	0.34505659

HEALTHCARE PROJECT.R | Healthcare_Error_pct_Final

Filter

	AGE	FEMALE	LOS	RACE	TOTCHG	APRDRG	Predicted_Price	Error_pct
38	3	0	5	1	14243	720	2.730888	0.45382231
39	16	1	3	1	1699	754	2.586180	0.13794000
40	2	0	2	1	7298	53	2.770490	0.38524507
41	16	1	1	1	636	754	2.586180	1.58618001
42	15	1	1	1	626	754	2.625782	1.62578171
43	1	0	2	1	3782	53	2.810092	0.40504592
44	14	1	2	1	1444	753	2.665383	0.33269170
45	14	1	2	1	1183	754	2.665383	0.33269170
46	14	1	5	1	3045	754	2.665383	0.46692332
47	14	1	5	1	3624	754	2.665383	0.46692332
48	14	1	12	1	6810	760	2.665383	0.77788472
49	1	0	1	1	1409	249	2.810092	1.81009185
50	13	0	2	1	1211	754	2.334871	0.16743574
51	1	0	4	1	9606	53	2.810092	0.29747704
52	1	1	1	1	1411	249	3.180205	2.18020548
53	1	0	1	1	2932	249	2.810092	1.81009185
54	1	0	3	1	5075	139	2.810092	0.06330272
55	14	1	1	1	762	753	2.665383	1.66538341
56	16	1	6	1	6329	753	2.586180	0.56897000