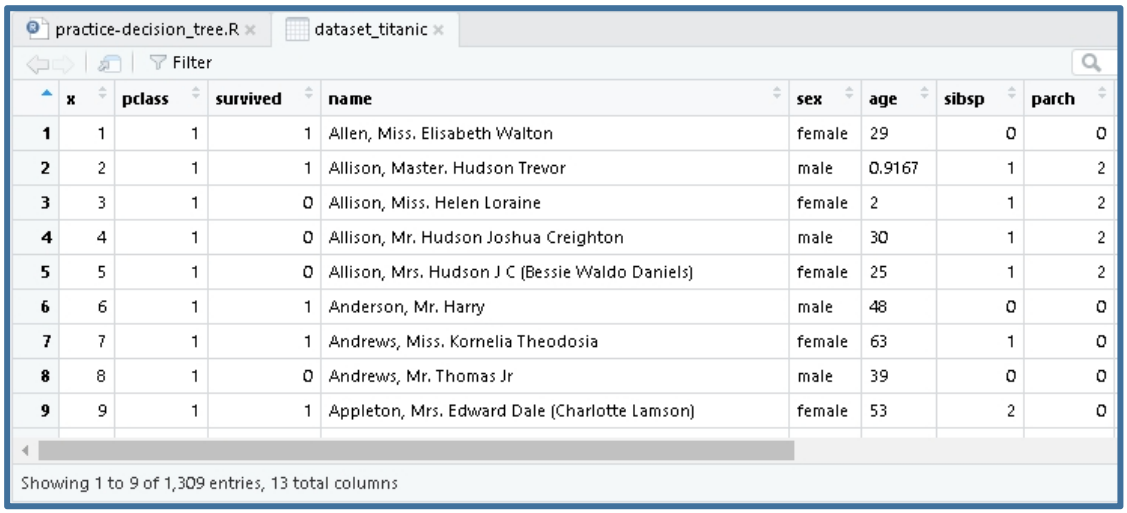
TITANIC ANALYSIS - PREDICTIVE MODEL

OPERATIONS PERFORMED :

1) set.seed(678)

library(tree)
dataset_titanic<- read.csv(file.choose())
View(dataset_titanic)



practice-decision_tree.R | dataset_titanic

Filter

	x	pclass	survived	name	sex	age	sibsp	parch
1	1	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0
2	2	1	1	Allison, Master. Hudson Trevor	male	0.9167	1	2
3	3	1	0	Allison, Miss. Helen Loraine	female	2	1	2
4	4	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2
5	5	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2
6	6	1	1	Anderson, Mr. Harry	male	48	0	0
7	7	1	1	Andrews, Miss. Kornelia Theodosia	female	63	1	0
8	8	1	0	Andrews, Mr. Thomas Jr	male	39	0	0
9	9	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53	2	0

Showing 1 to 9 of 1,309 entries, 13 total columns

2) head(dataset_titanic)

```
x pclass survived                                            name    sex    age sibsp
parch ticket
1 1      1        1                  Allen, Miss. Elisabeth Walton female     29
0     0  24160
2 2      1        1                 Allison, Master. Hudson Trevor   male 0.9167
1     2 113781
3 3      1        0                   Allison, Miss. Helen Loraine female      2     1
2 113781
4 4      1        0            Allison, Mr. Hudson Joshua Creighton   male     30
1     2 113781
5 5      1        0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female
25     1     2 113781
6 6      1        1                          Anderson, Mr. Harry   male     48     0
0  19952
      fare   cabin embarked                    home.dest
1 211.3375      B5        S                  St Louis, MO
2   151.55 C22 C26        S Montreal, PQ / Chesterville, ON
3   151.55 C22 C26        S Montreal, PQ / Chesterville, ON
4   151.55 C22 C26        S Montreal, PQ / Chesterville, ON
5   151.55 C22 C26        S Montreal, PQ / Chesterville, ON
```

6 26.55 E12 S New York, NY

3) tail(dataset_titanic)

```
x pclass survived                  name    sex  age sibsp parch ticket    fare
cabin
1304 1304      3       0     Yousseff, Mr. Gerious    male    ?     0     0
2627 14.4583     ?
1305 1305      3       0      Zabour, Miss. Hileni female 14.5     1     0
2665 14.4542     ?
1306 1306      3       0     Zabour, Miss. Thamine female    ?     1     0
2665 14.4542     ?
1307 1307      3       0 Zakarian, Mr. Mapriededer    male 26.5     0     0
2656   7.225     ?
1308 1308      3       0       Zakarian, Mr. Ortin    male   27     0     0
2670   7.225     ?
1309 1309      3       0       Zimmerman, Mr. Leo    male   29     0     0
315082   7.875     ?
    embarked home.dest
1304        C        ?
1305        C        ?
1306        C        ?
1307        C        ?
1308        C        ?
1309        S        ?
```

4) shuffle_index <- sample(1:nrow(dataset_titanic))
head(shuffle_index)

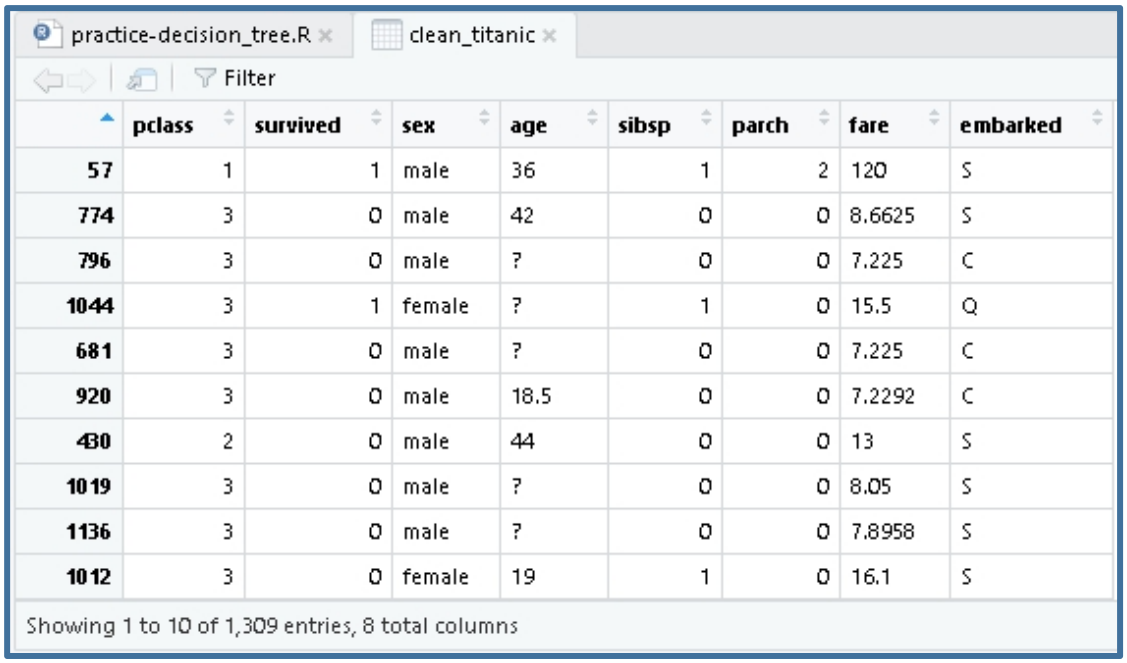
```
[1]  215   13 1036  355  146 1234
```

5) dataset_titanic <- dataset_titanic[shuffle_index,]
head(dataset_titanic)

```
x pclass survived                                        name    sex age sibsp
parch
215   215      1       1                          Newell, Miss. Marjorie female  23
1     0
13     13      1       1                   Aubart, Mme. Leontine Pauline female
24     0     0
1036 1036      3       1 Moubarek, Master. Halim Gonios ('William George')
male   ?     1     1
355   355      2       0                   Bryhl, Mr. Kurt Arnold Gottfrid    male
25     1     0
146   146      1       1                        Harper, Mr. Henry Sleeper    male
48     1     0
```

1234 1234 3 1 Sundman, Mr. Johan Julian male
44 0 0
ticket fare cabin embarked home.dest
215 35273 113.275 D36 C Lexington, MA
13 PC 17477 69.3 B35 C Paris, France
1036 2661 15.2458 ? C ?
355 236853 26 ? S Skara, Sweden / Rockford, IL
146 PC 17572 76.7292 D33 C New York, NY
1234 STON/O 2. 3101269 7.925 ? S ?

6) library(dplyr)
clean_titanic <- select(.data = dataset_titanic , -c("home.dest", "cabin", "name", "x", "ticket"))
View(clean_titanic)


practice-decision_tree.R
clean_titanic
Filter

	pclass	survived	sex	age	sibsp	parch	fare	embarked
57	1	1	male	36	1	2	120	S
774	3	0	male	42	0	0	8.6625	S
796	3	0	male	?	0	0	7.225	C
1044	3	1	female	?	1	0	15.5	Q
681	3	0	male	?	0	0	7.225	C
920	3	0	male	18.5	0	0	7.2292	C
430	2	0	male	44	0	0	13	S
1019	3	0	male	?	0	0	8.05	S
1136	3	0	male	?	0	0	7.8958	S
1012	3	0	female	19	1	0	16.1	S

Showing 1 to 10 of 1,309 entries, 8 total columns

7) clean_titanic<- mutate(.data = clean_titanic, pclass = factor(pclass, levels = c(1, 2, 3), labels = c('Upper', 'Middle', 'Lower')), survived = factor(survived, levels = c(0, 1), labels = c('No', 'Yes')))
na.omit(clean_titanic)

pclass survived sex age sibsp parch fare embarked
215 Upper Yes female 23 1 0 113.275 C
13 Upper Yes female 24 0 0 69.3 C
1036 Lower Yes male ? 1 1 15.2458 C
355 Middle No male 25 1 0 26 S
146 Upper Yes male 48 1 0 76.7292 C
1234 Lower Yes male 44 0 0 7.925 S

490	Middle	Yes	female	42	1	0	26	S
871	Lower	Yes	female	27	0	0	7.925	S
217	Upper	Yes	female	19	0	2	26.2833	S
629	Lower	No	female	11	4	2	31.275	S
117	Upper	Yes	female	60	1	4	263	S
226	Upper	No	male	23	0	0	93.5	S
505	Middle	No	male	39	0	0	13	S
357	Middle	No	male	25	0	0	13	S
84	Upper	Yes	female	64	1	1	26.55	S
328	Middle	No	male	25	0	0	10.5	S
744	Lower	No	female	22	0	0	10.5167	S
349	Middle	No	male	27	0	0	13	S
597	Middle	Yes	male	31	0	0	13	S
94	Upper	Yes	male	53	1	1	81.8583	S
874	Lower	No	male	42	0	0	7.65	S
177	Upper	Yes	female	?	1	0	51.8625	S
135	Upper	Yes	female	?	1	0	89.1042	C
926	Lower	No	male	44	0	0	8.05	S
1048	Lower	Yes	female	15	0	0	7.225	C
943	Lower	No	male	?	0	0	7.225	C
1008	Lower	Yes	female	15	0	0	8.0292	Q
747	Lower	Yes	male	29	0	0	7.75	Q
1258	Lower	Yes	female	9	1	1	15.2458	C
857	Lower	Yes	female	?	0	0	7.75	Q
78	Upper	No	male	37	1	1	83.1583	C
755	Lower	No	male	21	2	0	24.15	S
354	Middle	Yes	female	20	1	0	26	S
468	Middle	Yes	female	24	1	0	26	S
1082	Lower	Yes	female	?	0	0	7.8292	Q
1268	Lower	No	female	30	1	1	24.15	S
324	Middle	No	male	30	1	0	24	C
1150	Lower	Yes	female	?	0	0	7.7208	Q
389	Middle	Yes	female	28	0	0	13	S
1067	Lower	No	male	28.5	0	0	7.2292	C
313	Upper	No	male	50	1	1	211.5	C
145	Upper	Yes	female	25	1	0	55.4417	C
1024	Lower	Yes	female	?	0	0	7.8792	Q
1179	Lower	No	male	?	8	2	69.55	S
900	Lower	Yes	female	27	0	2	11.1333	S
1113	Lower	No	female	3	1	1	13.775	S
1088	Lower	No	male	28	0	0	7.8542	S
45	Upper	Yes	female	41	0	0	134.5	C
253	Upper	No	male	61	1	3	262.375	C
686	Lower	No	male	21	0	0	16.1	S
1228	Lower	No	female	22	0	0	9.8375	S
999	Lower	No	male	?	0	0	7.8958	C
708	Lower	No	male	28	0	0	7.7958	S
266	Upper	Yes	male	28	0	0	35.5	S
1148	Lower	No	female	22	0	0	39.6875	S
107	Upper	No	male	?	0	0	221.7792	S
483	Middle	Yes	female	17	0	0	12	C
486	Middle	No	male	36	0	0	12.875	C

571	Middle	Yes	female	50	0	0	10.5	S
1018	Lower	Yes	male	21	0	0	7.775	S
613	Lower	Yes	female	18	0	1	9.35	S
435	Middle	Yes	female	7	0	2	26.25	S
89	Upper	Yes	female	33	0	0	151.55	S
673	Lower	No	male	?	0	0	7.2292	C
546	Middle	Yes	female	30	3	0	21	S
671	Lower	No	male	26	0	0	7.775	S
707	Lower	No	female	?	1	0	14.4583	C
1068	Lower	Yes	female	22	0	0	7.75	S
507	Middle	No	male	70	0	0	10.5	S
694	Lower	Yes	male	21	0	0	7.8208	Q
1136	Lower	No	male	?	0	0	7.8958	S
1266	Lower	No	female	10	0	2	24.15	S
790	Lower	No	male	?	0	0	7.225	C
741	Lower	No	male	17	0	0	8.6625	S
887	Lower	Yes	female	?	0	0	7.75	Q
1237	Lower	Yes	male	14	0	0	9.225	S
136	Upper	No	male	71	0	0	34.6542	C
138	Upper	Yes	female	19	0	0	30	S
1011	Lower	No	male	24	1	0	16.1	S
941	Lower	No	male	?	0	0	7.8958	C
76	Upper	No	male	47	0	0	25.5875	S
1117	Lower	No	male	?	0	0	8.05	S
401	Middle	Yes	female	34	1	1	32.5	S
1245	Lower	Yes	female	16	1	1	8.5167	C
499	Middle	No	male	32	0	0	13.5	S
403	Middle	Yes	female	30	1	0	13.8583	C
155	Upper	No	male	55	1	1	93.5	S
565	Middle	Yes	female	40	0	0	13	S
1109	Lower	No	male	19	0	0	14.5	S
27	Upper	Yes	male	25	1	0	91.0792	C
99	Upper	Yes	female	48	1	0	106.425	C
602	Lower	No	male	13	0	2	20.25	S
1064	Lower	No	male	41	0	0	7.125	S
7	Upper	Yes	female	63	1	0	77.9583	S
668	Lower	No	female	27	0	0	7.8792	Q
333	Middle	No	male	23	0	0	10.5	S
488	Middle	No	male	61	0	0	12.35	Q
214	Upper	Yes	female	31	1	0	113.275	C
812	Lower	No	female	48	1	3	34.375	S
736	Lower	Yes	female	36	0	2	15.9	S
742	Lower	No	male	22.5	0	0	7.225	C
475	Middle	No	male	31	0	0	10.5	S
1051	Lower	Yes	female	19	1	1	15.7417	C
430	Middle	No	male	44	0	0	13	S
491	Middle	No	female	57	0	0	10.5	S
994	Lower	Yes	female	?	0	0	7.7375	Q
633	Lower	No	female	39	1	5	31.275	S
1079	Lower	Yes	female	?	0	0	7.8792	Q
722	Lower	No	male	36	0	0	7.4958	S
683	Lower	No	female	?	0	2	7.75	Q

```
296   Upper      Yes    male   17     0     2 110.8833       C
43    Upper      Yes female   59     2     0  51.4792       S
1152  Lower       No female    ?     0     0     14.5       S
970   Lower       No female   30     1     0    15.55       S
835   Lower       No    male  32     0     0   8.3625       S
175   Upper       No    male  58     0     0     29.7       C
77    Upper      Yes female   39     1     1  83.1583       C
347   Middle      No    male  26     0     0       13       S
1084  Lower       No    male  28     0     0   22.525       S
717   Lower       No    male  26     1     0  14.4542       C
819   Lower      Yes female   16     0     0   7.7333       Q
279   Upper      Yes    male  32     0     0     30.5       C
1247  Lower       No    male   ?     1     0     16.1       S
1028  Lower       No    male   ?     0     0     8.05       S
1244  Lower       No    male   ?     0     0    7.225       C
 [ reached 'max' / getOption("max.print") -- omitted 1184 rows ]
```

8) View(clean_titanic)
glimpse(clean_titanic)

Rows: 1,309
Columns: 8

```
> glimpse(clean_titanic)
Rows: 1,309
Columns: 8
$ pclass   <fct> Upper, Upper, Lower, Middle, Upper, Lower, Middle, Lower, Upper, Lower, Upper, ...
$ survived <fct> Yes, Yes, Yes, No, Yes, Yes, Yes, Yes, Yes, No, Yes, No, No, No, Yes, No, No, N...
$ sex      <chr> "female", "female", "male", "male", "male", "male", "female", "female", "female...
$ age      <chr> "23", "24", "?", "25", "48", "44", "42", "27", "19", "11", "60", "23", "39", "2...
$ sibsp    <int> 1, 0, 1, 1, 1, 0, 1, 0, 0, 4, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0...
$ parch    <int> 0, 0, 1, 0, 0, 0, 0, 0, 2, 2, 4, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0...
$ fare     <chr> "113.275", "69.3", "15.2458", "26", "76.7292", "7.925", "26", "7.925", "26.2833...
$ embarked <chr> "C", "C", "C", "S", "C", "S", "S", "S", "S", "S", "S", "S", "S", "S", "S", "S",...
```

9) clean_titanic<- select(.data = clean_titanic, -c("age"))
n<- nrow(clean_titanic)
n1<- floor(0.8 * n)
n2<- n - n1
n2

[1] 262

10) train<- sample(1:n , n1)
data_train<- clean_titanic[train,]
data_test<- clean_titanic[-train,]

dim(data_train)

[1] 1047   7

11) dim(data_test)

[1] 262 7

12) prop.table(table(data_train$survived))

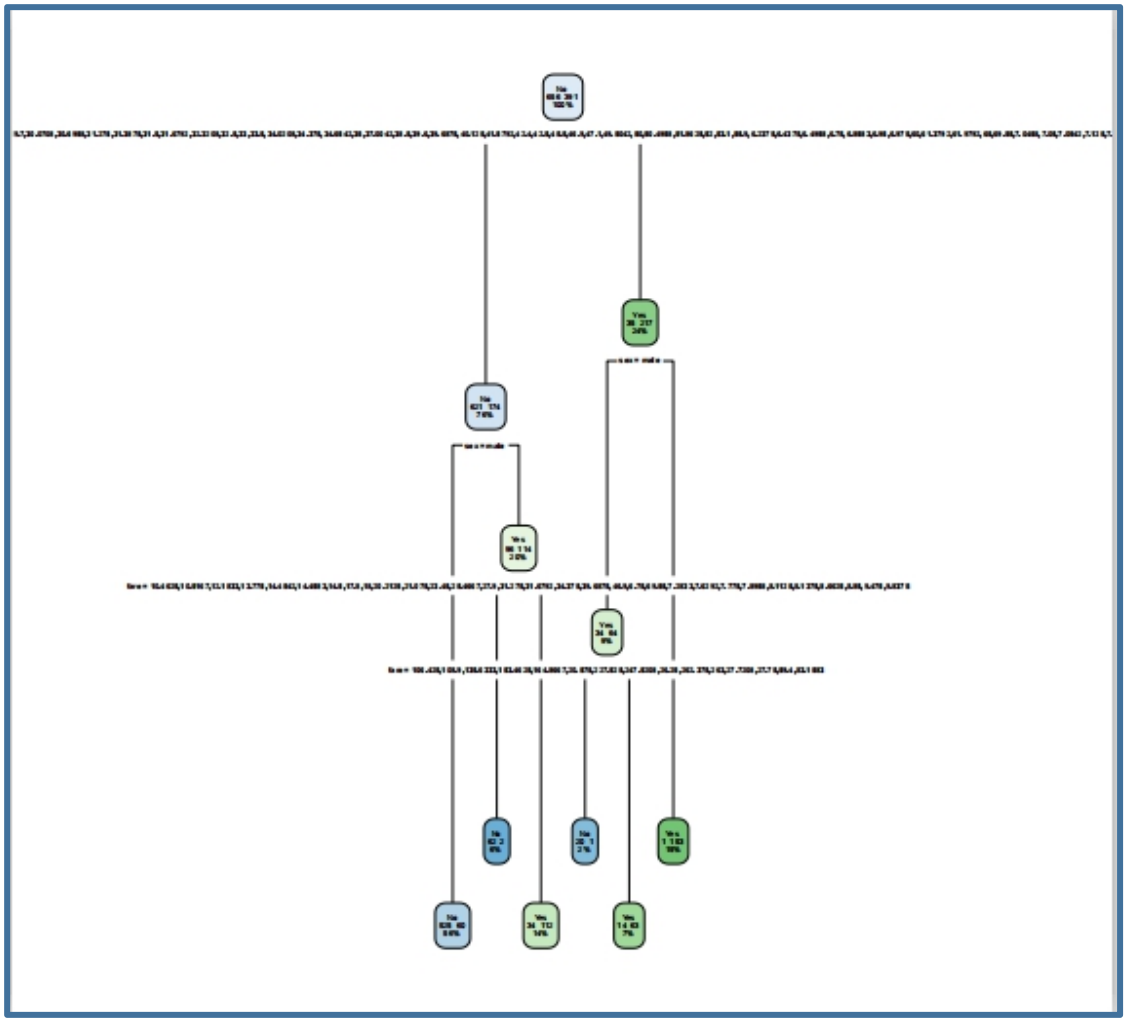
No Yes
0.6265521 0.3734479

13) prop.table(table(data_test$survived))

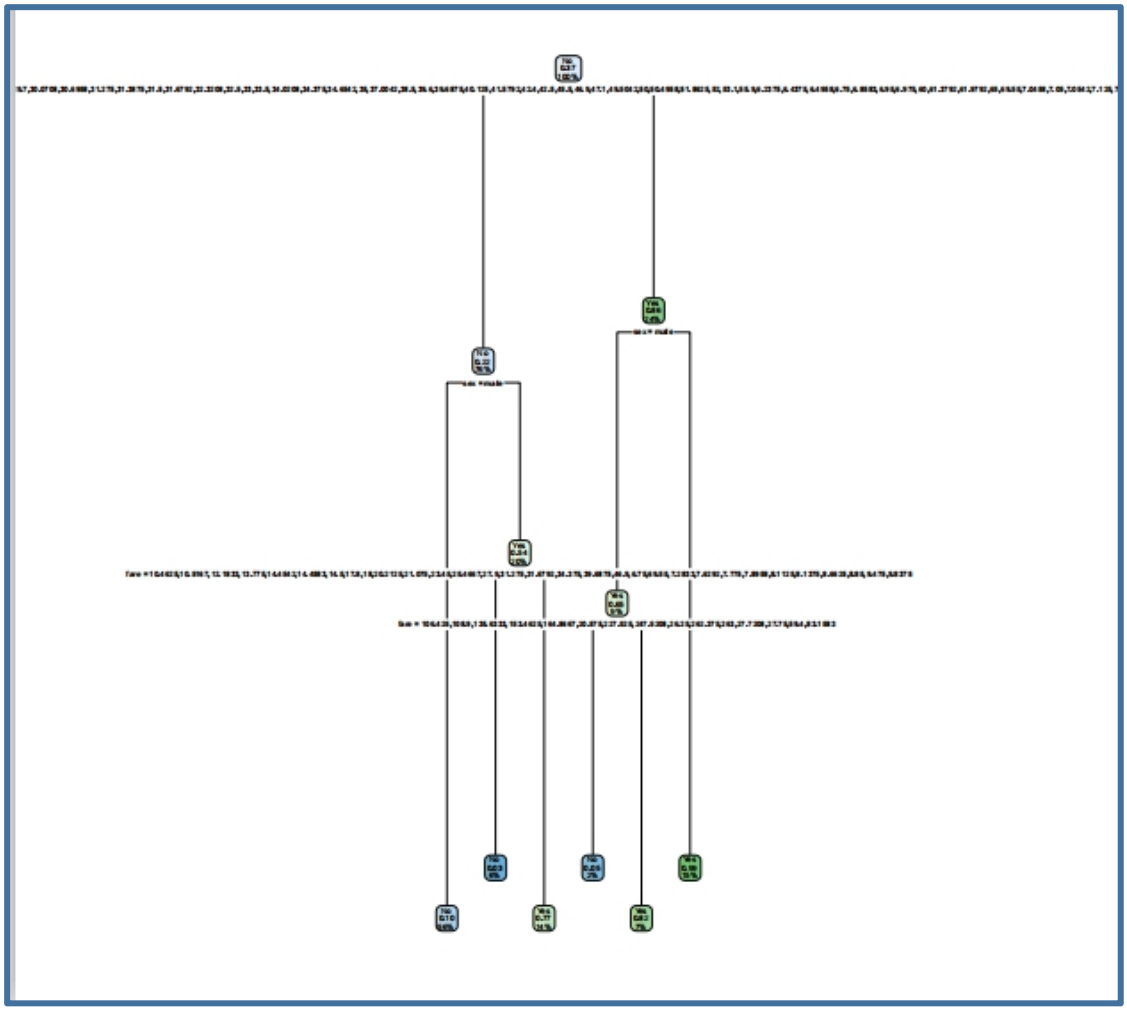
No Yes
0.5839695 0.4160305

14) library(rpart)
install.packages("rpart.plot")
library(rpart.plot)
fit<- rpart(survived~., data = data_train, method = "class")

15) rpart.plot(fit, extra = 101)



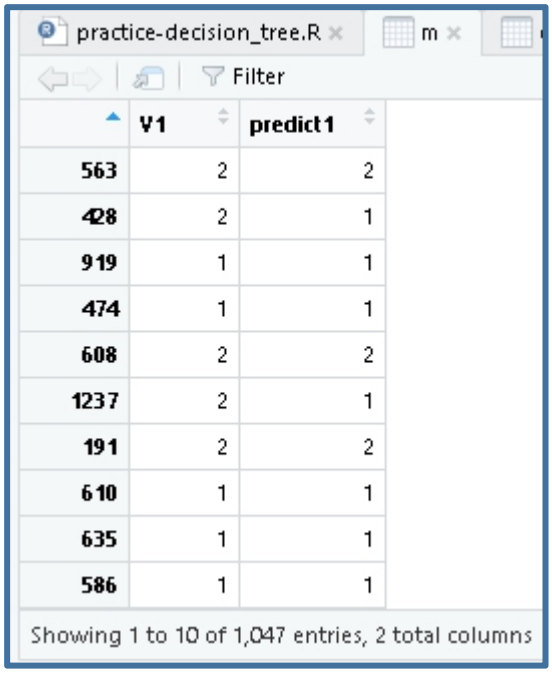
16) rpart.plot(fit, extra = 106)



17) predict1<- predict(fit, data_train, type = "class")
m<- cbind(data_train$survived, predict1)
table_combine<- table(data_train$survived, predict1)
table_combine

```
predict1
      No Yes
  No  607  49
  Yes  63 328
```

18) View(m)


practice-decision_tree.R
m
Filter
V1
predict1
563 2 2
428 2 1
919 1 1
474 1 1
608 2 2
1237 2 1
191 2 2
610 1 1
635 1 1
586 1 1
Showing 1 to 10 of 1,047 entries, 2 total columns