Graph Output


```
                                         0
                                       0.24
                                       100%
  ┌─────┐      relationship = Not-in-family, Other-relative, Own-child, Unmarried      ┌─────┐
  │ yes │                                                                              │ no  │
  └─────┘                                                                              └─────┘

                                                                    0
                                                                  0.45
                                                                  45%

                                    education = 10th, 11th, 12th, 1st-4th, 5th-6th, 7th-8th, 9th, Assoc-acdm, Assoc-voc, HS-grad, Preschool, Some-college

            0
          0.07
          55%

     capital.gain < 7074                                          0
                                                               0.33
                                                               32%

                                                          capital.gain < 5096

                          1
                        0.97
                        1%                                                 1
                                                                         0.98
                                                                         2%

    0                                      0                                               1
  0.05                                   0.30                                            0.73
  54%                                    30%                                             13%
```

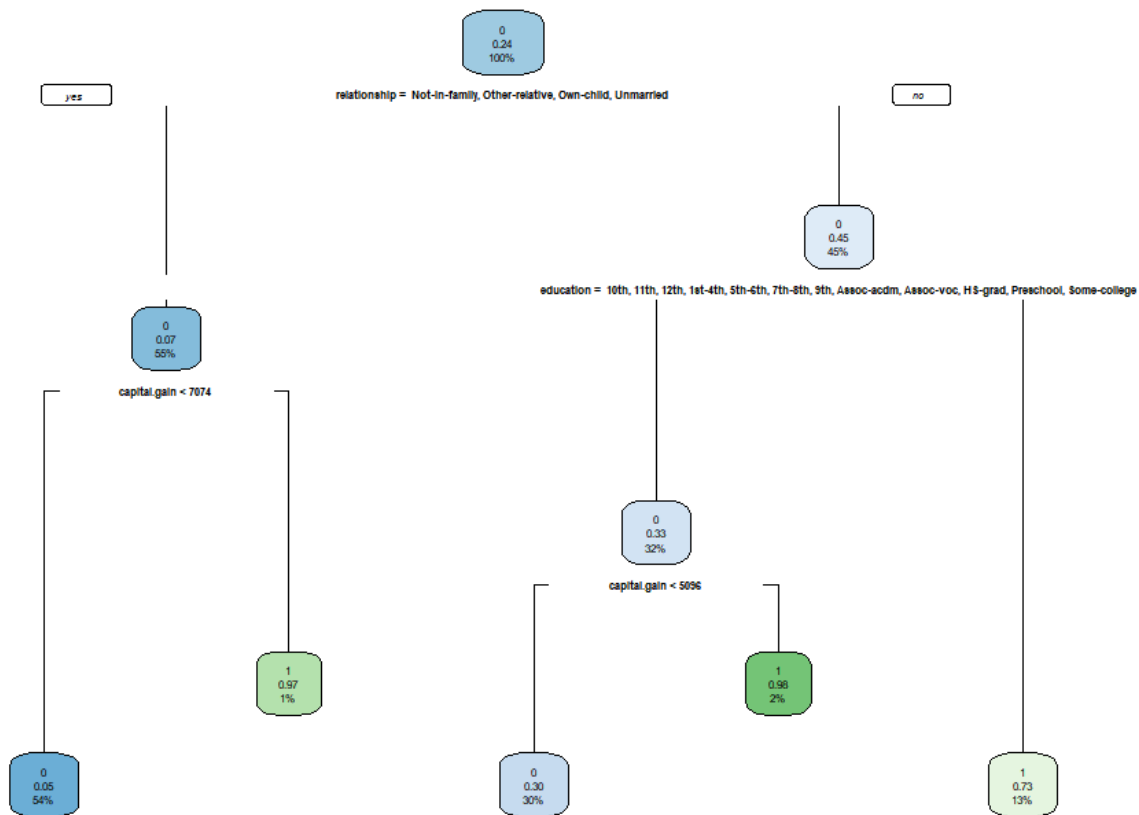## Console Output

```
> input = read.csv("D:/vaibhav/trend nxt/topgear/R Community/adult_census_inc
ome/Adult Census Income Binary Classification dataset.csv", header=TRUE)
> head(input)
  age         workclass fnlwgt education education.num      marital.status
1  39         State-gov  77516 Bachelors            13       Never-married
2  50  Self-emp-not-inc  83311 Bachelors            13  Married-civ-spouse
3  38           Private 215646   HS-grad             9            Divorced
4  53           Private 234721      11th             7  Married-civ-spouse
5  28           Private 338409 Bachelors            13  Married-civ-spouse
6  37           Private 284582   Masters            14  Married-civ-spouse
          occupation    relationship   race     sex capital.gain capital.loss
1       Adm-clerical   Not-in-family  White    Male         2174            0
2    Exec-managerial         Husband  White    Male            0            0
3  Handlers-cleaners   Not-in-family  White    Male            0            0
4  Handlers-cleaners         Husband  Black    Male            0            0
5     Prof-specialty            Wife  Black  Female            0            0
6    Exec-managerial            Wife  White  Female            0            0
  hours.per.week native.country income
1             40  United-States  <=50K
2             13  United-States  <=50K
3             40  United-States  <=50K
4             40  United-States  <=50K
```

```
5              40           Cuba   <=50K
6              40  United-States   <=50K
> summary(input)
      age                  workclass         fnlwgt                   educatio
n
 Min.   :17.00   Private       :22696   Min.   :  12285   HS-grad      :10
501
 1st Qu.:28.00   Self-emp-not-inc: 2541   1st Qu.: 117827   Some-college: 7
291
 Median :37.00   Local-gov     : 2093   Median : 178356   Bachelors    : 5
355
 Mean   :38.58   ?             : 1836   Mean   : 189778   Masters      : 1
723
 3rd Qu.:48.00   State-gov     : 1298   3rd Qu.: 237051   Assoc-voc    : 1
382
 Max.   :90.00   Self-emp-inc  : 1116   Max.   :1484705   11th         : 1
175
                 (Other)       :  981                     (Other)      : 5
134
 education.num             marital.status           occupation
 Min.   : 1.00   Divorced          : 4443   Prof-specialty :4140
 1st Qu.: 9.00   Married-AF-spouse :   23   Craft-repair   :4099
 Median :10.00   Married-civ-spouse :14976   Exec-managerial:4066
 Mean   :10.08   Married-spouse-absent:  418   Adm-clerical   :3770
 3rd Qu.:12.00   Never-married     :10683   Sales          :3650
 Max.   :16.00   Separated         : 1025   Other-service  :3295
                 Widowed           :  993   (Other)        :9541
        relationship                race              sex
 Husband       :13193   Amer-Indian-Eskimo:  311   Female:10771
 Not-in-family : 8305   Asian-Pac-Islander: 1039   Male  :21790
 Other-relative:  981   Black             : 3124
 Own-child     : 5068   Other             :  271
 Unmarried     : 3446   White             :27816
 Wife          : 1568

  capital.gain     capital.loss     hours.per.week           native.country
 Min.   :    0   Min.   :   0.0   Min.   : 1.00   United-States:29170
 1st Qu.:    0   1st Qu.:   0.0   1st Qu.:40.00   Mexico       :  643
 Median :    0   Median :   0.0   Median :40.00   ?            :  583
 Mean   : 1078   Mean   :  87.3   Mean   :40.44   Philippines  :  198
 3rd Qu.:    0   3rd Qu.:   0.0   3rd Qu.:45.00   Germany      :  137
 Max.   :99999   Max.   :4356.0   Max.   :99.00   Canada       :  121
                                                  (Other)      : 1709
    income
 <=50K:24720
 >50K : 7841




> str(input)
'data.frame':   32561 obs. of  15 variables:
 $ age          : int  39 50 38 53 28 37 49 52 31 42 ...
 $ workclass    : Factor w/ 9 levels " ?"," Federal-gov",..: 8 7 5 5 5 5 5 7
5 5 ...
 $ fnlwgt       : int  77516 83311 215646 234721 338409 284582 160187 209642
45781 159449 ...
 $ education    : Factor w/ 16 levels " 10th"," 11th",..: 10 10 12 2 10 13 7
12 13 10 ...
 $ education.num : int  13 13 9 7 13 14 5 9 14 13 ...
 $ marital.status: Factor w/ 7 levels " Divorced"," Married-AF-spouse",..: 5
3 1 3 3 3 4 3 5 3 ...
```

```
 $ occupation    : Factor w/ 15 levels " ?"," Adm-clerical",..: 2 5 7 7 11 5
9 5 11 5 ...
 $ relationship  : Factor w/ 6 levels " Husband"," Not-in-family",..: 2 1 2 1
6 6 2 1 2 1 ...
 $ race          : Factor w/ 5 levels " Amer-Indian-Eskimo",..: 5 5 5 3 3 5 3
5 5 5 ...
 $ sex           : Factor w/ 2 levels " Female"," Male": 2 2 2 2 1 1 1 2 1 2
...
 $ capital.gain  : int  2174 0 0 0 0 0 0 14084 5178 ...
 $ capital.loss  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ hours.per.week: int  40 13 40 40 40 40 16 45 50 40 ...
 $ native.country: Factor w/ 42 levels " ?"," Cambodia",..: 40 40 40 40 6 40
24 40 40 40 ...
 $ income        : Factor w/ 2 levels " <=50K"," >50K": 1 1 1 1 1 1 1 2 2 2 .
..
> input$income2 = ifelse(input$income == " <=50K",0, 1 )
> head(input)
  age          workclass fnlwgt  education education.num      marital.status
1  39          State-gov  77516  Bachelors           13       Never-married
2  50   Self-emp-not-inc  83311  Bachelors           13  Married-civ-spouse
3  38            Private 215646    HS-grad            9            Divorced
4  53            Private 234721       11th            7  Married-civ-spouse
5  28            Private 338409  Bachelors           13  Married-civ-spouse
6  37            Private 284582    Masters           14  Married-civ-spouse
          occupation    relationship   race     sex capital.gain capital.loss
1      Adm-clerical   Not-in-family  white    Male         2174            0
2   Exec-managerial         Husband  white    Male            0            0
3 Handlers-cleaners   Not-in-family  white    Male            0            0
4 Handlers-cleaners         Husband  Black    Male            0            0
5    Prof-specialty            Wife  Black  Female            0            0
6   Exec-managerial            Wife  white  Female            0            0
  hours.per.week native.country income income2
1             40  United-States  <=50K        0
2             13  United-States  <=50K        0
3             40  United-States  <=50K        0
4             40  United-States  <=50K        0
5             40           Cuba  <=50K        0
6             40  United-States  <=50K        0
> input1 = read.csv("D:/vaibhav/trend nxt/topgear/R Community/adult_census_in
come/Adult Census Income Binary Classification dataset.csv", header=TRUE, na.
strings = c(" ?", " "))
> na.omit(input1)
   age          workclass fnlwgt    education education.num      marital.s
tatus
1   39          State-gov  77516    Bachelors           13       Never-ma
rried
2   50   Self-emp-not-inc  83311    Bachelors           13      Married-civ-s
pouse
3   38            Private 215646      HS-grad            9           Div
orced
4   53            Private 234721         11th            7      Married-civ-s
pouse
5   28            Private 338409    Bachelors           13      Married-civ-s
pouse
6   37            Private 284582      Masters           14      Married-civ-s
pouse
7   49            Private 160187          9th            5  Married-spouse-a
bsent
8   52   Self-emp-not-inc 209642      HS-grad            9      Married-civ-s
pouse
9   31            Private  45781      Masters           14       Never-ma
rried
10  42            Private 159449    Bachelors           13      Married-civ-s
pouse
```

11	37	Private	280464	Some-college	10	Married-civ-spouse
12	30	State-gov	141297	Bachelors	13	Married-civ-spouse
13	23	Private	122272	Bachelors	13	Never-married
14	32	Private	205019	Assoc-acdm	12	Never-married
16	34	Private	245487	7th-8th	4	Married-civ-spouse
17	25	Self-emp-not-inc	176756	HS-grad	9	Never-married
18	32	Private	186824	HS-grad	9	Never-married
19	38	Private	28887	11th	7	Married-civ-spouse
20	43	Self-emp-not-inc	292175	Masters	14	Divorced
21	40	Private	193524	Doctorate	16	Married-civ-spouse
22	54	Private	302146	HS-grad	9	Separated
23	35	Federal-gov	76845	9th	5	Married-civ-spouse
24	43	Private	117037	11th	7	Married-civ-spouse
25	59	Private	109015	HS-grad	9	Divorced
26	56	Local-gov	216851	Bachelors	13	Married-civ-spouse
27	19	Private	168294	HS-grad	9	Never-married
29	39	Private	367260	HS-grad	9	Divorced
30	49	Private	193366	HS-grad	9	Married-civ-spouse
31	23	Local-gov	190709	Assoc-acdm	12	Never-married
32	20	Private	266015	Some-college	10	Never-married
33	45	Private	386940	Bachelors	13	Divorced
34	30	Federal-gov	59951	Some-college	10	Married-civ-spouse
35	22	State-gov	311512	Some-college	10	Married-civ-spouse
36	48	Private	242406	11th	7	Never-married
37	21	Private	197200	Some-college	10	Never-married
38	19	Private	544091	HS-grad	9	Married-AF-spouse
40	48	Self-emp-not-inc	265477	Assoc-acdm	12	Married-civ-spouse
41	31	Private	507875	9th	5	Married-civ-spouse
42	53	Self-emp-not-inc	88506	Bachelors	13	Married-civ-spouse
43	24	Private	172987	Bachelors	13	Married-civ-spouse
44	49	Private	94638	HS-grad	9	Separated
45	25	Private	289980	HS-grad	9	Never-married

46	57	Federal-gov	337895	Bachelors	13	Married-civ-spouse
47	53	Private	144361	HS-grad	9	Married-civ-spouse
48	44	Private	128354	Masters	14	Divorced
49	41	State-gov	101603	Assoc-voc	11	Married-civ-spouse
50	29	Private	271466	Assoc-voc	11	Never-married
51	25	Private	32275	Some-college	10	Married-civ-spouse
53	47	Private	51835	Prof-school	15	Married-civ-spouse
54	50	Federal-gov	251585	Bachelors	13	Divorced
55	47	Self-emp-inc	109832	HS-grad	9	Divorced
56	43	Private	237993	Some-college	10	Married-civ-spouse
57	46	Private	216666	5th-6th	3	Married-civ-spouse
58	35	Private	56352	Assoc-voc	11	Married-civ-spouse
59	41	Private	147372	HS-grad	9	Married-civ-spouse
60	30	Private	188146	HS-grad	9	Married-civ-spouse
61	30	Private	59496	Bachelors	13	Married-civ-spouse
63	48	Private	149640	HS-grad	9	Married-civ-spouse
64	42	Private	116632	Doctorate	16	Married-civ-spouse
65	29	Private	105598	Some-college	10	Divorced
66	36	Private	155537	HS-grad	9	Married-civ-spouse
67	28	Private	183175	Some-college	10	Divorced
68	53	Private	169846	HS-grad	9	Married-civ-spouse
69	49	Self-emp-inc	191681	Some-college	10	Married-civ-spouse
71	19	Private	101509	Some-college	10	Never-married
72	31	Private	309974	Bachelors	13	Separated

	occupation	relationship	race	sex	capital.gain
1	Adm-clerical	Not-in-family	White	Male	2174
2	Exec-managerial	Husband	White	Male	0
3	Handlers-cleaners	Not-in-family	White	Male	0
4	Handlers-cleaners	Husband	Black	Male	0
5	Prof-specialty	Wife	Black	Female	0
6	Exec-managerial	Wife	White	Female	0
7	Other-service	Not-in-family	Black	Female	0
8	Exec-managerial	Husband	White	Male	0
9	Prof-specialty	Not-in-family	White	Female	14084
10	Exec-managerial	Husband	White	Male	5178
11	Exec-managerial	Husband	Black	Male	0
12	Prof-specialty	Husband	Asian-Pac-Islander	Male	0
13	Adm-clerical	Own-child	White	Female	0
14	Sales	Not-in-family	Black	Male	0
16	Transport-moving	Husband	Amer-Indian-Eskimo	Male	0

17	Farming-fishing	Own-child	White	Male	0
18	Machine-op-inspct	Unmarried	White	Male	0
19	Sales	Husband	White	Male	0
20	Exec-managerial	Unmarried	White	Female	0
21	Prof-specialty	Husband	White	Male	0
22	Other-service	Unmarried	Black	Female	0
23	Farming-fishing	Husband	Black	Male	0
24	Transport-moving	Husband	White	Male	0
25	Tech-support	Unmarried	White	Female	0
26	Tech-support	Husband	White	Male	0
27	Craft-repair	Own-child	White	Male	0
29	Exec-managerial	Not-in-family	White	Male	0
30	Craft-repair	Husband	White	Male	0
31	Protective-serv	Not-in-family	White	Male	0
32	Sales	Own-child	Black	Male	0
33	Exec-managerial	Own-child	White	Male	0
34	Adm-clerical	Own-child	White	Male	0
35	Other-service	Husband	Black	Male	0
36	Machine-op-inspct	Unmarried	White	Male	0
37	Machine-op-inspct	Own-child	White	Male	0
38	Adm-clerical	Wife	White	Female	0
40	Prof-specialty	Husband	White	Male	0
41	Machine-op-inspct	Husband	White	Male	0
42	Prof-specialty	Husband	White	Male	0
43	Tech-support	Husband	White	Male	0
44	Adm-clerical	Unmarried	White	Female	0
45	Handlers-cleaners	Not-in-family	White	Male	0
46	Prof-specialty	Husband	Black	Male	0
47	Machine-op-inspct	Husband	White	Male	0
48	Exec-managerial	Unmarried	White	Female	0
49	Craft-repair	Husband	White	Male	0
50	Prof-specialty	Not-in-family	White	Male	0
51	Exec-managerial	Wife	Other	Female	0
53	Prof-specialty	Wife	White	Female	0
54	Exec-managerial	Not-in-family	White	Male	0
55	Exec-managerial	Not-in-family	White	Male	0
56	Tech-support	Husband	White	Male	0
57	Machine-op-inspct	Husband	White	Male	0
58	Other-service	Husband	White	Male	0
59	Adm-clerical	Husband	White	Male	0
60	Machine-op-inspct	Husband	White	Male	5013
61	Sales	Husband	White	Male	2407
63	Transport-moving	Husband	White	Male	0
64	Prof-specialty	Husband	White	Male	0
65	Tech-support	Not-in-family	White	Male	0
66	Craft-repair	Husband	White	Male	0
67	Adm-clerical	Not-in-family	White	Female	0
68	Adm-clerical	Wife	White	Female	0
69	Exec-managerial	Husband	White	Male	0
71	Prof-specialty	Own-child	White	Male	0
72	Sales	Own-child	Black	Female	0

	capital.loss	hours.per.week	native.country	income
1	0	40	United-States	<=50K
2	0	13	United-States	<=50K
3	0	40	United-States	<=50K
4	0	40	United-States	<=50K
5	0	40	Cuba	<=50K
6	0	40	United-States	<=50K
7	0	16	Jamaica	<=50K
8	0	45	United-States	>50K
9	0	50	United-States	>50K
10	0	40	United-States	>50K
11	0	80	United-States	>50K
12	0	40	India	>50K

```
13              0       30  United-States  <=50K
14              0       50  United-States  <=50K
16              0       45         Mexico  <=50K
17              0       35  United-States  <=50K
18              0       40  United-States  <=50K
19              0       50  United-States  <=50K
20              0       45  United-States   >50K
21              0       60  United-States   >50K
22              0       20  United-States  <=50K
23              0       40  United-States  <=50K
24           2042       40  United-States  <=50K
25              0       40  United-States  <=50K
26              0       40  United-States   >50K
27              0       40  United-States  <=50K
29              0       80  United-States  <=50K
30              0       40  United-States  <=50K
31              0       52  United-States  <=50K
32              0       44  United-States  <=50K
33           1408       40  United-States  <=50K
34              0       40  United-States  <=50K
35              0       15  United-States  <=50K
36              0       40    Puerto-Rico  <=50K
37              0       40  United-States  <=50K
38              0       25  United-States  <=50K
40              0       40  United-States  <=50K
41              0       43  United-States  <=50K
42              0       40  United-States  <=50K
43              0       50  United-States  <=50K
44              0       40  United-States  <=50K
45              0       35  United-States  <=50K
46              0       40  United-States   >50K
47              0       38  United-States  <=50K
48              0       40  United-States  <=50K
49              0       40  United-States  <=50K
50              0       43  United-States  <=50K
51              0       40  United-States  <=50K
53           1902       60        Honduras   >50K
54              0       55  United-States   >50K
55              0       60  United-States  <=50K
56              0       40  United-States   >50K
57              0       40         Mexico  <=50K
58              0       40    Puerto-Rico  <=50K
59              0       48  United-States  <=50K
60              0       40  United-States  <=50K
61              0       40  United-States  <=50K
63              0       40  United-States  <=50K
64              0       45  United-States   >50K
65              0       58  United-States  <=50K
66              0       40  United-States  <=50K
67              0       40  United-States  <=50K
68              0       40  United-States   >50K
69              0       50  United-States   >50K
71              0       32  United-States  <=50K
72              0       40  United-States  <=50K
 [ reached 'max' / getOption("max.print") -- omitted 30096 rows ]
> head(input1)
  age          workclass fnlwgt  education education.num      marital.status
1  39          State-gov  77516  Bachelors            13       Never-married
2  50   Self-emp-not-inc  83311  Bachelors            13  Married-civ-spouse
3  38            Private 215646    HS-grad             9            Divorced
4  53            Private 234721       11th             7  Married-civ-spouse
5  28            Private 338409  Bachelors            13  Married-civ-spouse
6  37            Private 284582    Masters            14  Married-civ-spouse
         occupation   relationship   race    sex capital.gain capital.loss
```

```
1       Adm-clerical   Not-in-family  White    Male          2174            0
2     Exec-managerial        Husband  White    Male             0            0
3   Handlers-cleaners  Not-in-family  White    Male             0            0
4   Handlers-cleaners        Husband  Black    Male             0            0
5       Prof-specialty           Wife  Black  Female             0            0
6     Exec-managerial           Wife  White  Female             0            0
  hours.per.week native.country income
1             40  United-States  <=50K
2             13  United-States  <=50K
3             40  United-States  <=50K
4             40  United-States  <=50K
5             40           Cuba  <=50K
6             40  United-States  <=50K
> summary(input1)
      age                  workclass          fnlwgt                    educatio
n
 Min.   :17.00    Private        :22696    Min.   :  12285    HS-grad     :10
501
 1st Qu.:28.00    Self-emp-not-inc: 2541    1st Qu.: 117827    Some-college: 7
291
 Median :37.00    Local-gov      : 2093    Median : 178356    Bachelors   : 5
355
 Mean   :38.58    State-gov      : 1298    Mean   : 189778    Masters     : 1
723
 3rd Qu.:48.00    Self-emp-inc   : 1116    3rd Qu.: 237051    Assoc-voc   : 1
382
 Max.   :90.00    (Other)        :  981    Max.   :1484705    11th        : 1
175
                  NA's           : 1836                       (Other)     : 5
134
 education.num                marital.status            occupation
 Min.   : 1.00    Divorced          : 4443    Prof-specialty : 4140
 1st Qu.: 9.00    Married-AF-spouse  :   23    Craft-repair   : 4099
 Median :10.00    Married-civ-spouse :14976    Exec-managerial: 4066
 Mean   :10.08    Married-spouse-absent:  418    Adm-clerical   : 3770
 3rd Qu.:12.00    Never-married      :10683    Sales          : 3650
 Max.   :16.00    Separated          : 1025    (Other)        :10993
                  Widowed            :  993    NA's           : 1843
         relationship              race              sex
 Husband       :13193    Amer-Indian-Eskimo:  311    Female:10771
 Not-in-family : 8305    Asian-Pac-Islander: 1039    Male  :21790
 Other-relative:  981    Black             : 3124
 Own-child     : 5068    Other             :  271
 Unmarried     : 3446    White             :27816
 Wife          : 1568

  capital.gain      capital.loss     hours.per.week        native.country
 Min.   :    0    Min.   :   0.0    Min.   : 1.00    United-States:29170
 1st Qu.:    0    1st Qu.:   0.0    1st Qu.:40.00    Mexico       :  643
 Median :    0    Median :   0.0    Median :40.00    Philippines  :  198
 Mean   : 1078    Mean   :  87.3    Mean   :40.44    Germany      :  137
 3rd Qu.:    0    3rd Qu.:   0.0    3rd Qu.:45.00    Canada       :  121
 Max.   :99999    Max.   :4356.0    Max.   :99.00    (Other)      : 1709
                                                     NA's         :  583
    income
 <=50K:24720
 >50K : 7841




> str(input1)
'data.frame':   32561 obs. of  15 variables:
```

```
 $ age          : int  39 50 38 53 28 37 49 52 31 42 ...
 $ workclass    : Factor w/ 8 levels " Federal-gov",..: 7 6 4 4 4 4 4 6 4 4
...
 $ fnlwgt       : int  77516 83311 215646 234721 338409 284582 160187 209642
45781 159449 ...
 $ education    : Factor w/ 16 levels " 10th"," 11th",..: 10 10 12 2 10 13 7
12 13 10 ...
 $ education.num : int  13 13 9 7 13 14 5 9 14 13 ...
 $ marital.status: Factor w/ 7 levels " Divorced"," Married-AF-spouse",..: 5
3 1 3 3 3 4 3 5 3 ...
 $ occupation   : Factor w/ 14 levels " Adm-clerical",..: 1 4 6 6 10 4 8 4 1
0 4 ...
 $ relationship : Factor w/ 6 levels " Husband"," Not-in-family",..: 2 1 2 1
6 6 2 1 2 1 ...
 $ race         : Factor w/ 5 levels " Amer-Indian-Eskimo",..: 5 5 5 3 3 5 3
5 5 5 ...
 $ sex          : Factor w/ 2 levels " Female"," Male": 2 2 2 2 1 1 1 2 1 2
...
 $ capital.gain : int  2174 0 0 0 0 0 0 14084 5178 ...
 $ capital.loss : int  0 0 0 0 0 0 0 0 0 0 ...
 $ hours.per.week: int  40 13 40 40 40 40 16 45 50 40 ...
 $ native.country: Factor w/ 41 levels " Cambodia"," Canada",..: 39 39 39 39
5 39 23 39 39 39 ...
 $ income       : Factor w/ 2 levels " <=50K"," >50K": 1 1 1 1 1 1 1 2 2 2 .
..
> input1$income2 = ifelse(input1$income == " <=50K",0, 1 )
> head(input1)
  age        workclass fnlwgt  education education.num      marital.status
1  39        State-gov  77516  Bachelors            13       Never-married
2  50 Self-emp-not-inc  83311  Bachelors            13  Married-civ-spouse
3  38          Private 215646    HS-grad             9            Divorced
4  53          Private 234721       11th             7  Married-civ-spouse
5  28          Private 338409  Bachelors            13  Married-civ-spouse
6  37          Private 284582    Masters            14  Married-civ-spouse
          occupation    relationship   race     sex capital.gain capital.loss
1       Adm-clerical   Not-in-family  White    Male         2174            0
2    Exec-managerial         Husband  White    Male            0            0
3  Handlers-cleaners   Not-in-family  White    Male            0            0
4  Handlers-cleaners         Husband  Black    Male            0            0
5     Prof-specialty            Wife  Black  Female            0            0
6    Exec-managerial            Wife  White  Female            0            0
  hours.per.week native.country income income2
1             40  United-States  <=50K       0
2             13  United-States  <=50K       0
3             40  United-States  <=50K       0
4             40  United-States  <=50K       0
5             40           Cuba  <=50K       0
6             40  United-States  <=50K       0
> shuffle_index = sample(1:nrow(input1))
> head(shuffle_index)
[1] 29654 10906 26268 10410 10818   355
> input1 = input1[shuffle_index,]
> head(input1)
      age workclass fnlwgt     education education.num      marital.status
29654  34   Private 185556  Some-college            10       Never-married
10906  74   Private 209454       5th-6th             3  Married-civ-spouse
26268  37   Private 217838       5th-6th             3           Separated
10410  39   Private 286789     Bachelors            13  Married-civ-spouse
10818  35   Private 158046  Some-college            10  Married-civ-spouse
355    59      <NA> 372020     Bachelors            13  Married-civ-spouse
          occupation    relationship   race     sex capital.gain capital.lo
ss
29654    Prof-specialty  Not-in-family  White  Female            0
0
```

```
10906   Transport-moving        Husband  Black    Male            0
0
26268      Other-service     Unmarried  White  Female            0
0
10410   Exec-managerial         Husband  White    Male            0
0
10818             Sales         Husband  White    Male         5013
0
355                 <NA>        Husband  White    Male            0
0
        hours.per.week native.country income income2
29654               12  United-States   >50K        1
10906               40  United-States  <=50K        0
26268               40         Mexico  <=50K        0
10410               40  United-States   >50K        1
10818               70  United-States  <=50K        0
355                 40  United-States   >50K        1
> test_data_size = 0.8
> data_size = nrow(input1)
> train_sample = 1:(test_data_size*data_size)
> train_data = input1[train_sample, ]
> test_data = input1[-train_sample, ]
> library(rpart)
> library(rpart.plot)
Error in library(rpart.plot) : there is no package called 'rpart.plot'
> fit <- rpart(income2 ~ age + workclass + fnlwgt + education + marital.statu
s + occupation + relationship + race + sex + capital.gain + capital.loss + ho
urs.per.week + native.country, data = train_data, method = 'class')
> rpart.plot(fit, extra = 106)
Error in rpart.plot(fit, extra = 106) :
  could not find function "rpart.plot"
> install.packages(rpart.plot)
Error in install.packages : object 'rpart.plot' not found
> install.packages("rpart.plot")
WARNING: Rtools is required to build R packages but is not currently installe
d. Please download and install the appropriate version of Rtools before proce
eding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/va20025797/Documents/R/win-library/3.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/rpart.plot_3.0.7
.zip'
Content type 'application/zip' length 1071354 bytes (1.0 MB)
downloaded 1.0 MB

package 'rpart.plot' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\va20025797\AppData\Local\Temp\RtmpELS4bj\downloaded_packages
> library(rpart)
> library(rpart.plot)
> fit <- rpart(income2 ~ age + workclass + fnlwgt + education + marital.statu
s + occupation + relationship + race + sex + capital.gain + capital.loss + ho
urs.per.week + native.country, data = train_data, method = 'class')
> rpart.plot(fit, extra = 106)
> predict_unseen <-predict(fit, test_data, type = 'class')
> table_mat <- table(test_data$income2, predict_unseen)
> table_mat
   predict_unseen
        0     1
  0 4613   277
  1  787   836
> accuracy_Test <- sum(diag(table_mat)) / sum(table_mat)
```

```
> accuracy_Test
[1] 0.8366344
```