Data Science II: Homework 4

Name: Jasmin Martinez (JRM2319) Date: 04/20/25

QUESTION 1: In this exercise, we will build tree-based models using the College data (see "College.csv" in Homework 2). The response variable is the out-of-state tuition (Outstate). Partition the dataset into two parts: training data (80%) and test data (20%).

```
# initial data steps--importing and partitioning
College = read.csv("College.csv")
head(College)
```

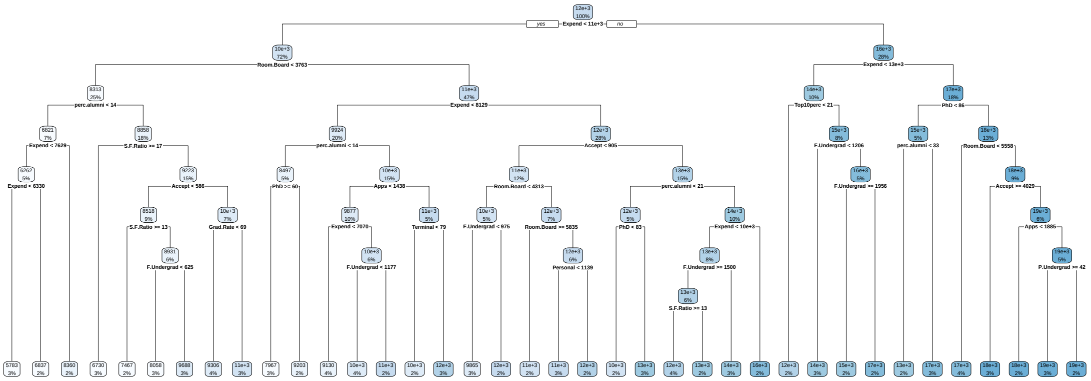
```
##                         College Apps Accept Enroll Top10perc Top25perc
## 1 Abilene Christian University 1660   1232    721        23        52
## 2           Adelphi University 2186   1924    512        16        29
## 3               Adrian College 1428   1097    336        22        50
## 4          Agnes Scott College  417    349    137        60        89
## 5    Alaska Pacific University  193    146     55        16        44
## 6            Albertson College  587    479    158        38        62
##   F.Undergrad P.Undergrad Outstate Room.Board Books Personal PhD Terminal
## 1        2885         537     7440       3300   450     2200  70       78
## 2        2683        1227    12280       6450   750     1500  29       30
## 3        1036          99    11250       3750   400     1165  53       66
## 4         510          63    12960       5450   450      875  92       97
## 5         249         869     7560       4120   800     1500  76       72
## 6         678          41    13500       3335   500      675  67       73
##   S.F.Ratio perc.alumni Expend Grad.Rate
## 1      18.1          12   7041        60
## 2      12.2          16  10527        56
## 3      12.9          30   8735        54
## 4       7.7          37  19016        59
## 5      11.9           2  10922        15
## 6       9.4          11   9727        55
```

```
datSplit = initial_split(data = College, prop = 0.8)
trainData = training(datSplit)
testData = testing(datSplit)
head(trainData)
```

```
##               College Apps Accept Enroll Top10perc Top25perc F.Undergrad
## 1  Bellarmine College  807    707    308        39        63        1198
## 2       Barat College  261    192    111        15        36         453
## 3 Columbia College MO  314    158    132        10        28         690
## 4    Augsburg College  662    513    257        12        30        2074
## 5   Morehouse College 3708   1678    722        41        66        2852
## 6   Quincy University 1025    707    297         22         66          1070
##   P.Undergrad Outstate Room.Board Books Personal PhD Terminal S.F.Ratio
## 1         605     8840       2950   750     1290  74       82      13.1
## 2         266     9690       4300   500      500  57       77       9.7
## 3        5346     8294       3700   400      900  87       87      15.3
## 4         726    11902       4372   540      950  65       65      12.8
## 5         153     7050       5490   250      600  71       74      17.8
## 6          72    10100       4140   450     1080  69       71      16.3
##   perc.alumni Expend Grad.Rate
## 1          31   6668        84
## 2          35   9337        71
## 3           2   5015        37
## 4          31   7836        58
## 5          10   8122        83
## 6          32   6880        80
```

```r
set.seed(1)
tree1 = rpart(formula = Outstate ~ . - College,
              data = trainData,
              control = rpart.control(cp=0))
rpart.plot(tree1) #this gives the full tree, but we want a more complex and smaller tree
```

1.A: Build a regression tree on the training data to predict the response (10pts). Create a plot

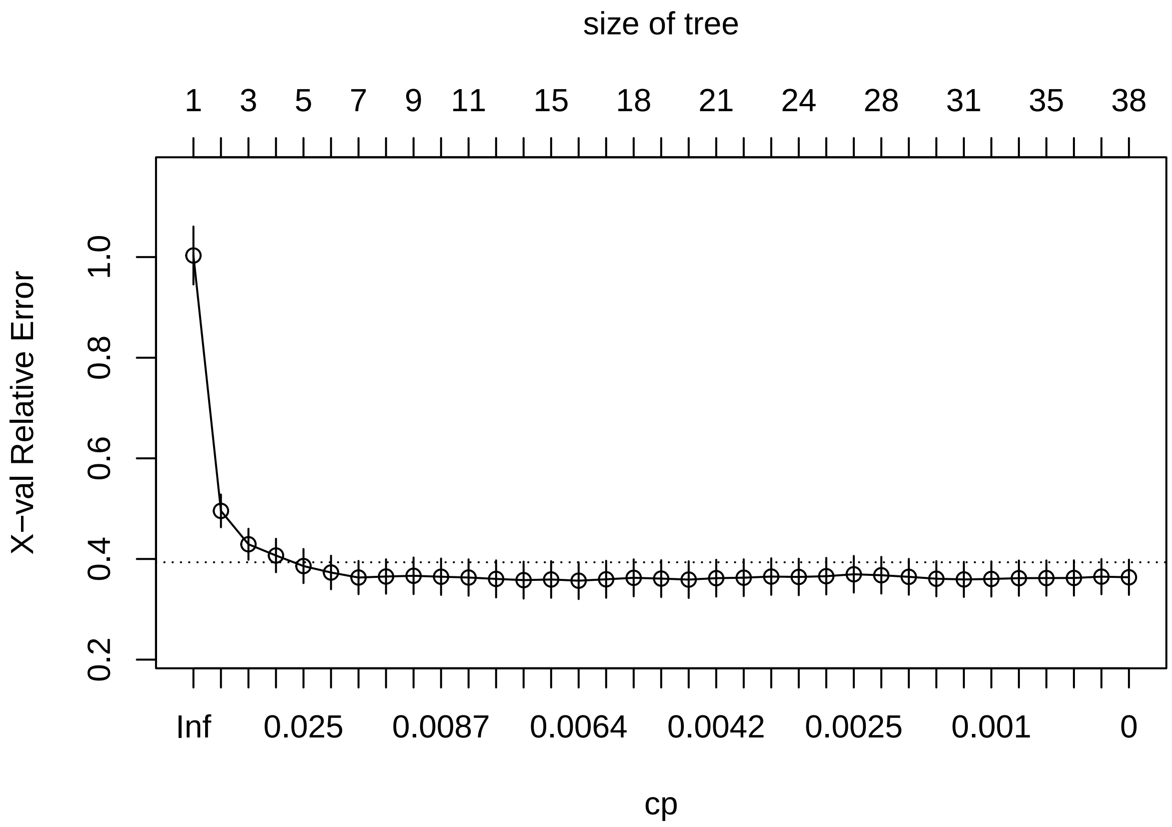


of the tree (10pts).

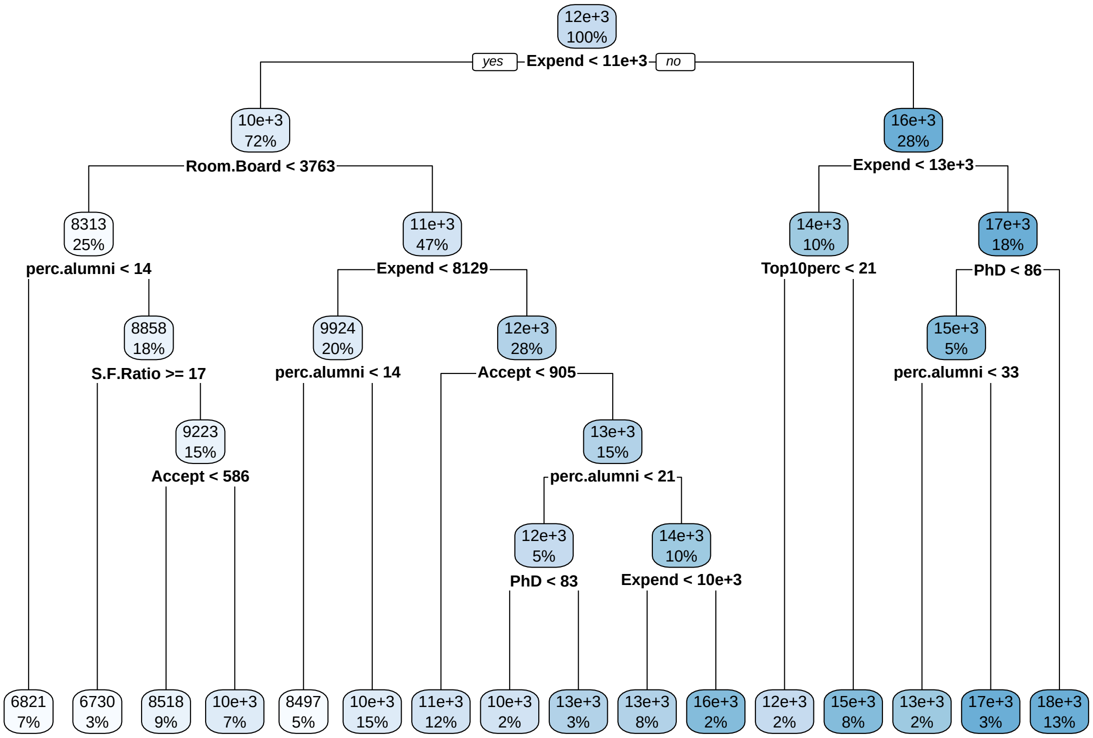
```r
printcp(tree1)
```

```
##
## Regression tree:
## rpart(formula = Outstate ~ . - College, data = trainData, control = rpart.control(cp = 0))
##
## Variables actually used in tree construction:
##  [1] Accept      Apps        Expend      F.Undergrad Grad.Rate   P.Undergrad
##  [7] perc.alumni Personal    PhD         Room.Board  S.F.Ratio   Terminal
## [13] Top10perc
##
## Root node error: 6345327363/452 = 14038335
##
## n= 452
##
##            CP nsplit rel error  xerror     xstd
## 1  0.51835135      0   1.00000 1.00308 0.057594
## 2  0.09850360      1   0.48165 0.49558 0.032540
## 3  0.04087938      2   0.38315 0.42928 0.030790
## 4  0.03428521      3   0.34227 0.40693 0.033279
## 5  0.01863879      4   0.30798 0.38599 0.033861
## 6  0.01752037      5   0.28934 0.37314 0.033396
## 7  0.01436278      6   0.27182 0.36315 0.033017
## 8  0.01003387      7   0.25746 0.36520 0.034232
## 9  0.00882200      8   0.24742 0.36663 0.036351
## 10 0.00852605      9   0.23860 0.36478 0.036147
## 11 0.00834425     10   0.23008 0.36315 0.036175
## 12 0.00723941     12   0.21339 0.36041 0.036853
## 13 0.00690162     13   0.20615 0.35804 0.036713
## 14 0.00640576     14   0.19925 0.35930 0.036468
## 15 0.00639032     15   0.19284 0.35694 0.036539
## 16 0.00534616     16   0.18645 0.35966 0.036782
## 17 0.00513089     17   0.18110 0.36259 0.036738
## 18 0.00466995     18   0.17597 0.36105 0.036650
## 19 0.00437217     19   0.17130 0.35907 0.036494
## 20 0.00407089     20   0.16693 0.36190 0.036567
## 21 0.00339007     21   0.16286 0.36284 0.036564
## 22 0.00294261     22   0.15947 0.36508 0.036358
## 23 0.00282619     23   0.15653 0.36435 0.036260
## 24 0.00279247     24   0.15370 0.36584 0.036330
## 25 0.00215660     26   0.14812 0.36962 0.036417
## 26 0.00208123     27   0.14596 0.36772 0.036467
## 27 0.00156288     28   0.14388 0.36453 0.035635
## 28 0.00123992     29   0.14232 0.36076 0.034953
## 29 0.00106361     30   0.14108 0.35942 0.034906
## 30 0.00100411     31   0.14001 0.36031 0.034953
## 31 0.00095590     33   0.13800 0.36192 0.035074
## 32 0.00091077     34   0.13705 0.36211 0.035072
## 33 0.00080953     35   0.13614 0.36228 0.035069
## 34 0.00025201     36   0.13533 0.36492 0.035007
## 35 0.00000000     37   0.13508 0.36366 0.035014
```

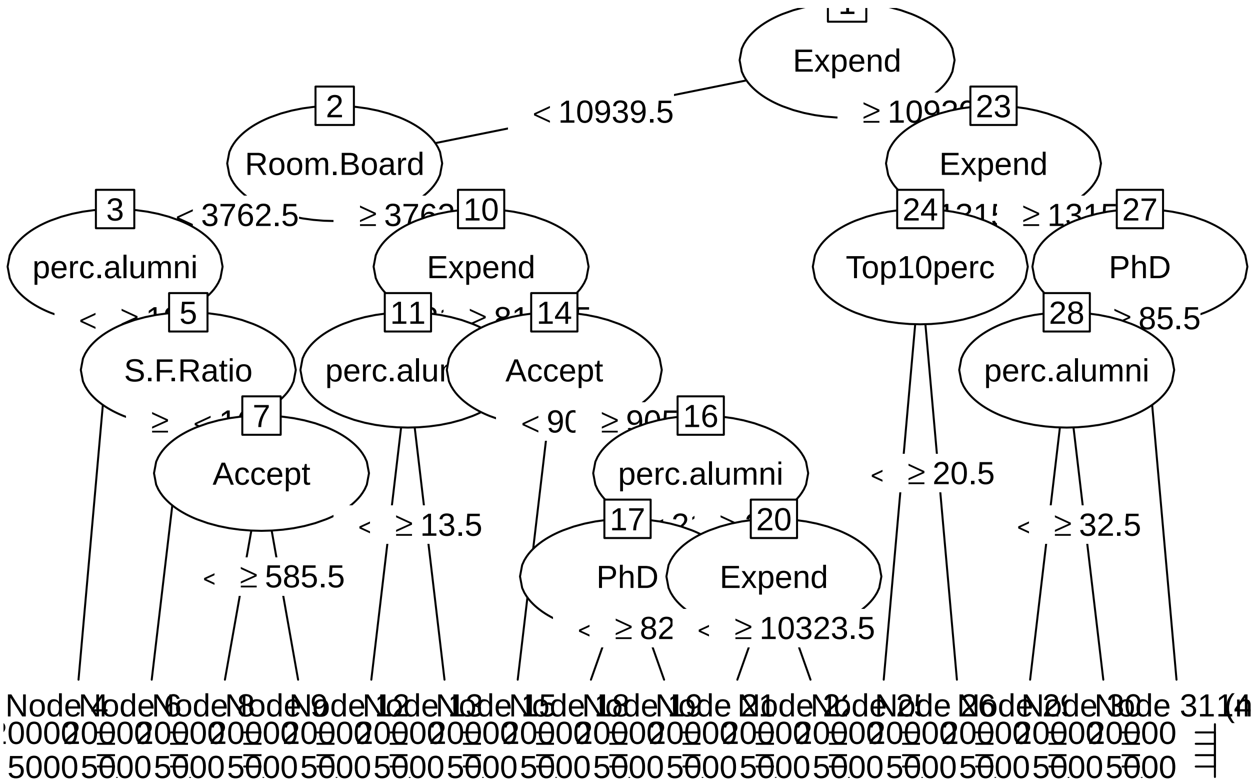
```
cpTable = tree1$cptable
plotcp(tree1)
```



```
# Picking the cp that yields the minimum cross-validation error
minErr = which.min(cpTable[,4])
tree3 = rpart::prune(tree1, cp = cpTable[minErr,1])
rpart.plot(tree3)
```



```r
plot(as.party(tree3)) #another visual
```



```
summary(tree3) # summary of Tree3 (the final condensed version of the regression tree)
```

```
## Call:
## rpart(formula = Outstate ~ . - College, data = trainData, control = rpart.control(cp = 0))
##   n= 452
##
##             CP nsplit rel error    xerror       xstd
## 1  0.518351348      0 1.0000000 1.0030811 0.05759430
## 2  0.098503601      1 0.4816487 0.4955757 0.03253971
## 3  0.040879376      2 0.3831451 0.4292833 0.03078994
## 4  0.034285206      3 0.3422657 0.4069274 0.03327941
## 5  0.018638786      4 0.3079805 0.3859914 0.03386112
## 6  0.017520372      5 0.2893417 0.3731431 0.03339643
## 7  0.014362780      6 0.2718213 0.3631521 0.03301699
## 8  0.010033868      7 0.2574585 0.3652015 0.03423193
## 9  0.008822004      8 0.2474247 0.3666336 0.03635051
## 10 0.008526053      9 0.2386027 0.3647768 0.03614748
## 11 0.008344251     10 0.2300766 0.3631521 0.03617503
## 12 0.007239407     12 0.2133881 0.3604132 0.03685270
## 13 0.006901616     13 0.2061487 0.3580436 0.03671281
## 14 0.006405762     14 0.1992471 0.3593016 0.03646769
## 15 0.006390321     15 0.1928413 0.3569353 0.03653860
##
## Variable importance
##      Expend    Terminal         PhD   Top10perc   Top25perc        Apps
##          30          14          12          11          10           8
##  Room.Board perc.alumni      Accept   S.F.Ratio      Enroll F.Undergrad
##           5           2           2           1           1           1
##   Grad.Rate P.Undergrad
##           1           1
##
## Node number 1: 452 observations,    complexity param=0.5183513
##   mean=11903.88, MSE=1.403833e+07
##   left son=2 (326 obs) right son=3 (126 obs)
##   Primary splits:
##       Expend     < 10939.5 to the left,  improve=0.5183513, (0 missing)
##       Terminal   < 84.5    to the left,  improve=0.3748716, (0 missing)
##       PhD        < 78.5    to the left,  improve=0.3667132, (0 missing)
##       Room.Board < 3961    to the left,  improve=0.2904086, (0 missing)
##       Top10perc  < 35.5    to the left,  improve=0.2793795, (0 missing)
##   Surrogate splits:
##       Terminal  < 93.5    to the left,  agree=0.850, adj=0.460, (0 split)
##       PhD       < 89.5    to the left,  agree=0.836, adj=0.413, (0 split)
##       Top10perc < 43.5    to the left,  agree=0.825, adj=0.373, (0 split)
##       Top25perc < 74.5    to the left,  agree=0.810, adj=0.317, (0 split)
##       Apps      < 2647.5  to the left,  agree=0.794, adj=0.262, (0 split)
##
## Node number 2: 326 observations,    complexity param=0.0985036
##   mean=10226.83, MSE=6554930
##   left son=4 (112 obs) right son=5 (214 obs)
##   Primary splits:
##       Room.Board < 3762.5  to the left,  improve=0.2924964, (0 missing)
##       Expend     < 8132    to the left,  improve=0.2810090, (0 missing)
##       Terminal   < 80.5    to the left,  improve=0.1938974, (0 missing)
##       Grad.Rate  < 61.5    to the left,  improve=0.1721348, (0 missing)
##       PhD        < 77.5    to the left,  improve=0.1621505, (0 missing)
##   Surrogate splits:
##       Expend      < 7115.5  to the left,  agree=0.718, adj=0.179, (0 split)
##       Terminal    < 63.5    to the left,  agree=0.709, adj=0.152, (0 split)
##       P.Undergrad < 66.5    to the left,  agree=0.702, adj=0.134, (0 split)
##       Grad.Rate   < 50.5    to the left,  agree=0.681, adj=0.071, (0 split)
##       Accept      < 184.5   to the left,  agree=0.669, adj=0.036, (0 split)
##
## Node number 3: 126 observations,    complexity param=0.03428521
##   mean=16242.92, MSE=7296120
##   left son=6 (44 obs) right son=7 (82 obs)
##   Primary splits:
##       Expend     < 13158   to the left,  improve=0.2366455, (0 missing)
##       Room.Board < 5557.5  to the left,  improve=0.2232844, (0 missing)
##       Top25perc  < 74.5    to the left,  improve=0.2225995, (0 missing)
##       PhD        < 85.5    to the left,  improve=0.2130894, (0 missing)
##       Terminal   < 91.5    to the left,  improve=0.1909604, (0 missing)
##   Surrogate splits:
##       Top25perc < 64.5    to the left,  agree=0.762, adj=0.318, (0 split)
##       Top10perc < 36.5    to the left,  agree=0.754, adj=0.295, (0 split)
##       Books     < 462.5   to the left,  agree=0.722, adj=0.205, (0 split)
##       Terminal  < 89.5    to the left,  agree=0.722, adj=0.205, (0 split)
##       S.F.Ratio < 13.25   to the right, agree=0.722, adj=0.205, (0 split)
##
```

```
## Node number 4: 112 observations,    complexity param=0.01436278
##   mean=8312.83, MSE=4930922
##   left son=8 (30 obs) right son=9 (82 obs)
##   Primary splits:
##       perc.alumni < 13.5    to the left,  improve=0.1650237, (0 missing)
##       Expend      < 6215.5  to the left,  improve=0.1359266, (0 missing)
##       S.F.Ratio   < 15.05   to the right, improve=0.1348840, (0 missing)
##       Grad.Rate   < 44.5    to the left,  improve=0.1156903, (0 missing)
##       Room.Board  < 3050    to the left,  improve=0.1096671, (0 missing)
##   Surrogate splits:
##       Top25perc  < 28.5    to the left,  agree=0.750, adj=0.067, (0 split)
##       Room.Board < 2536    to the left,  agree=0.750, adj=0.067, (0 split)
##       Apps       < 4254.5  to the right, agree=0.741, adj=0.033, (0 split)
##       Accept     < 170     to the left,  agree=0.741, adj=0.033, (0 split)
##       Enroll     < 1198.5  to the right, agree=0.741, adj=0.033, (0 split)
##
## Node number 5: 214 observations,    complexity param=0.04087938
##   mean=11228.55, MSE=4484142
##   left son=10 (89 obs) right son=11 (125 obs)
##   Primary splits:
##       Expend    < 8128.5  to the left,  improve=0.2703119, (0 missing)
##       Terminal  < 85.5    to the left,  improve=0.1910113, (0 missing)
##       Accept    < 932.5   to the left,  improve=0.1864649, (0 missing)
##       Apps      < 1181    to the left,  improve=0.1850094, (0 missing)
##       Grad.Rate < 55.5    to the left,  improve=0.1783875, (0 missing)
##   Surrogate splits:
##       PhD       < 70.5    to the left,  agree=0.715, adj=0.315, (0 split)
##       Terminal  < 73.5    to the left,  agree=0.696, adj=0.270, (0 split)
##       S.F.Ratio < 14.05   to the right, agree=0.673, adj=0.213, (0 split)
##       Top10perc < 22.5    to the left,  agree=0.654, adj=0.169, (0 split)
##       Top25perc < 43.5    to the left,  agree=0.654, adj=0.169, (0 split)
##
## Node number 6: 44 observations,    complexity param=0.007239407
##   mean=14449.11, MSE=5024178
##   left son=12 (7 obs) right son=13 (37 obs)
##   Primary splits:
##       Top10perc   < 20.5    to the left,  improve=0.2077970, (0 missing)
##       F.Undergrad < 1206    to the left,  improve=0.1726283, (0 missing)
##       Apps        < 1282    to the left,  improve=0.1672279, (0 missing)
##       Accept      < 917.5   to the left,  improve=0.1672279, (0 missing)
##       P.Undergrad < 346.5   to the right, improve=0.1283876, (0 missing)
##   Surrogate splits:
##       Top25perc   < 44      to the left,  agree=0.955, adj=0.714, (0 split)
##       Apps        < 433.5   to the left,  agree=0.886, adj=0.286, (0 split)
##       Accept      < 396     to the left,  agree=0.886, adj=0.286, (0 split)
##       Enroll      < 146     to the left,  agree=0.886, adj=0.286, (0 split)
##       F.Undergrad < 612     to the left,  agree=0.886, adj=0.286, (0 split)
##
## Node number 7: 82 observations,    complexity param=0.01863879
##   mean=17205.45, MSE=5862151
##   left son=14 (23 obs) right son=15 (59 obs)
##   Primary splits:
##       PhD        < 85.5    to the left,  improve=0.2460372, (0 missing)
##       Room.Board < 5557.5  to the left,  improve=0.1810750, (0 missing)
##       Grad.Rate  < 67.5    to the left,  improve=0.1810398, (0 missing)
##       Terminal   < 91.5    to the left,  improve=0.1701871, (0 missing)
##       Apps       < 3335.5  to the left,  improve=0.1657770, (0 missing)
##   Surrogate splits:
##       Terminal  < 91.5    to the left,  agree=0.902, adj=0.652, (0 split)
##       Top10perc < 30.5    to the left,  agree=0.805, adj=0.304, (0 split)
##       Grad.Rate < 64.5    to the left,  agree=0.805, adj=0.304, (0 split)
##       Top25perc < 67.5    to the left,  agree=0.793, adj=0.261, (0 split)
##       Apps      < 827.5   to the left,  agree=0.780, adj=0.217, (0 split)
##
## Node number 8: 30 observations
##   mean=6821.467, MSE=2964807
##
## Node number 9: 82 observations,    complexity param=0.01003387
##   mean=8858.451, MSE=4538812
##   left son=18 (12 obs) right son=19 (70 obs)
##   Primary splits:
##       S.F.Ratio < 16.8    to the right, improve=0.17106710, (0 missing)
##       Grad.Rate < 44.5    to the left,  improve=0.12521090, (0 missing)
##       Books     < 680     to the right, improve=0.10664180, (0 missing)
##       Expend    < 8128.5  to the left,  improve=0.08251206, (0 missing)
##       Terminal  < 82.5    to the left,  improve=0.08062386, (0 missing)
##   Surrogate splits:
##       Expend      < 5015    to the left,  agree=0.915, adj=0.417, (0 split)
##       P.Undergrad < 1073    to the right, agree=0.890, adj=0.250, (0 split)
##       Apps        < 2081    to the right, agree=0.878, adj=0.167, (0 split)
##       Enroll      < 645.5   to the right, agree=0.878, adj=0.167, (0 split)
##       F.Undergrad < 2744.5  to the right, agree=0.878, adj=0.167, (0 split)
##
## Node number 10: 89 observations,    complexity param=0.008822004
##   mean=9923.787, MSE=2773243
##   left son=20 (21 obs) right son=21 (68 obs)
##   Primary splits:
##       perc.alumni < 13.5    to the left,  improve=0.2268001, (0 missing)
##       Expend      < 7069.5  to the left,  improve=0.1982358, (0 missing)
##       Grad.Rate   < 57      to the left,  improve=0.1705047, (0 missing)
##       Top10perc   < 18.5    to the left,  improve=0.1342929, (0 missing)
##       Apps        < 1438    to the left,  improve=0.1302637, (0 missing)
##   Surrogate splits:
##       Grad.Rate < 50.5    to the left,  agree=0.809, adj=0.190, (0 split)
##       S.F.Ratio < 21.75   to the right, agree=0.798, adj=0.143, (0 split)
##       Apps      < 250     to the left,  agree=0.787, adj=0.095, (0 split)
##       Accept    < 226.5   to the left,  agree=0.787, adj=0.095, (0 split)
##       Personal  < 1700    to the right, agree=0.787, adj=0.095, (0 split)
##
## Node number 11: 125 observations,    complexity param=0.01752037
##   mean=12157.54, MSE=3627158
##   left son=22 (55 obs) right son=23 (70 obs)
##   Primary splits:
##       Accept      < 905     to the left,  improve=0.2452003, (0 missing)
##       Apps        < 1096    to the left,  improve=0.2382105, (0 missing)
##       Terminal    < 82.5    to the left,  improve=0.1962363, (0 missing)
##       Enroll      < 299     to the left,  improve=0.1940069, (0 missing)
##       F.Undergrad < 1064.5  to the left,  improve=0.1840590, (0 missing)
##   Surrogate splits:
##       Apps        < 1096    to the left,  agree=0.968, adj=0.927, (0 split)
##       Enroll      < 299     to the left,  agree=0.920, adj=0.818, (0 split)
##       F.Undergrad < 1064.5  to the left,  agree=0.872, adj=0.709, (0 split)
##       PhD         < 76.5    to the left,  agree=0.728, adj=0.382, (0 split)
##       Top25perc   < 53.5    to the left,  agree=0.720, adj=0.364, (0 split)
##
## Node number 12: 7 observations
##   mean=12100, MSE=3566024
##
## Node number 13: 37 observations
##   mean=14893.54, MSE=4058520
##
## Node number 14: 23 observations,    complexity param=0.008526053
##   mean=15281.96, MSE=9558413
##   left son=28 (9 obs) right son=29 (14 obs)
##   Primary splits:
##       perc.alumni < 32.5    to the left,  improve=0.24608690, (0 missing)
##       Top25perc   < 76.5    to the left,  improve=0.21677310, (0 missing)
##       P.Undergrad < 140     to the right, improve=0.21246950, (0 missing)
##       Books       < 632.5   to the right, improve=0.09221313, (0 missing)
##       Top10perc   < 44.5    to the left,  improve=0.06921854, (0 missing)
##   Surrogate splits:
##       F.Undergrad < 2774    to the right, agree=0.870, adj=0.667, (0 split)
##       P.Undergrad < 171     to the right, agree=0.870, adj=0.667, (0 split)
##       Accept      < 2801.5  to the right, agree=0.826, adj=0.556, (0 split)
##       Enroll      < 777     to the right, agree=0.826, adj=0.556, (0 split)
##       Grad.Rate   < 68      to the left,  agree=0.826, adj=0.556, (0 split)
##
## Node number 15: 59 observations
##   mean=17955.29, MSE=2416673
##
## Node number 18: 12 observations
##   mean=6730.25, MSE=6174270
##
## Node number 19: 70 observations,    complexity param=0.006901616
##   mean=9223.286, MSE=3348902
##   left son=38 (39 obs) right son=39 (31 obs)
##   Primary splits:
##       Accept    < 585.5   to the left,  improve=0.1868118, (0 missing)
##       Apps      < 924.5   to the left,  improve=0.1622362, (0 missing)
##       Enroll    < 376.5   to the left,  improve=0.1535067, (0 missing)
##       Top10perc < 28.5    to the left,  improve=0.1399150, (0 missing)
##       Top25perc < 57.5    to the left,  improve=0.1333957, (0 missing)
##   Surrogate splits:
##       Apps        < 667.5   to the left,  agree=0.943, adj=0.871, (0 split)
##       Enroll      < 234.5   to the left,  agree=0.857, adj=0.677, (0 split)
##       F.Undergrad < 1124    to the left,  agree=0.829, adj=0.613, (0 split)
##       Top25perc   < 55.5    to the left,  agree=0.757, adj=0.452, (0 split)
##       Grad.Rate   < 66.5    to the left,  agree=0.729, adj=0.387, (0 split)
##
## Node number 20: 21 observations
##   mean=8496.667, MSE=1463384
##
```

```
## Node number 21: 68 observations
##   mean=10364.51, MSE=2354545
##
## Node number 22: 55 observations
##   mean=11093.62, MSE=1595111
##
## Node number 23: 70 observations,    complexity param=0.008344251
##   mean=12993.49, MSE=3635587
##   left son=46 (23 obs) right son=47 (47 obs)
##   Primary splits:
##       perc.alumni < 21      to the left,  improve=0.1868718, (0 missing)
##       Expend      < 10712   to the left,  improve=0.1594021, (0 missing)
##       Personal    < 740     to the right, improve=0.1570990, (0 missing)
##       Terminal    < 85.5    to the left,  improve=0.1122604, (0 missing)
##       F.Undergrad < 1607.5  to the right, improve=0.0908322, (0 missing)
##   Surrogate splits:
##       Books       < 608.5   to the right, agree=0.800, adj=0.391, (0 split)
##       Personal    < 1269    to the right, agree=0.786, adj=0.348, (0 split)
##       Grad.Rate   < 66.5    to the left,  agree=0.771, adj=0.304, (0 split)
##       P.Undergrad < 1167    to the right, agree=0.757, adj=0.261, (0 split)
##       F.Undergrad < 2611    to the right, agree=0.743, adj=0.217, (0 split)
##
## Node number 28: 9 observations
##   mean=13369.11, MSE=9777604
##
## Node number 29: 14 observations
##   mean=16511.64, MSE=5553176
##
## Node number 38: 39 observations
##   mean=8518.103, MSE=2547846
##
## Node number 39: 31 observations
##   mean=10110.45, MSE=2944004
##
## Node number 46: 23 observations,    complexity param=0.006405762
##   mean=11815.22, MSE=3929265
##   left son=92 (11 obs) right son=93 (12 obs)
##   Primary splits:
##       PhD         < 82.5    to the left,  improve=0.4497651, (0 missing)
##       Terminal    < 82      to the left,  improve=0.3800468, (0 missing)
##       Room.Board  < 5924    to the left,  improve=0.2331664, (0 missing)
##       P.Undergrad < 225     to the left,  improve=0.2155932, (0 missing)
##       Grad.Rate   < 65.5    to the left,  improve=0.2081110, (0 missing)
##   Surrogate splits:
##       Terminal  < 84.5    to the left,  agree=0.870, adj=0.727, (0 split)
##       Personal  < 1125    to the left,  agree=0.739, adj=0.455, (0 split)
##       Grad.Rate < 65.5    to the left,  agree=0.739, adj=0.455, (0 split)
##       Apps      < 2106    to the left,  agree=0.696, adj=0.364, (0 split)
##       Top10perc < 27      to the left,  agree=0.696, adj=0.364, (0 split)
##
## Node number 47: 47 observations,    complexity param=0.008344251
##   mean=13570.09, MSE=2480017
##   left son=94 (38 obs) right son=95 (9 obs)
##   Primary splits:
```

```
##       Expend      < 10323.5 to the left,  improve=0.5004839, (0 missing)
##       Personal    < 860     to the right, improve=0.3234696, (0 missing)
##       perc.alumni < 36.5    to the left,  improve=0.1378371, (0 missing)
##       Top25perc   < 51.5    to the left,  improve=0.1265320, (0 missing)
##       S.F.Ratio   < 13.25   to the right, improve=0.1241689, (0 missing)
##   Surrogate splits:
##       Personal  < 350     to the right, agree=0.851, adj=0.222, (0 split)
##       Top10perc < 48.5    to the left,  agree=0.830, adj=0.111, (0 split)
##       Top25perc < 74      to the left,  agree=0.830, adj=0.111, (0 split)
##       Terminal  < 93.5    to the left,  agree=0.830, adj=0.111, (0 split)
##       S.F.Ratio < 11.35   to the right, agree=0.830, adj=0.111, (0 split)
##
## Node number 92: 11 observations
##   mean=10426.73, MSE=4073306
##
## Node number 93: 12 observations
##   mean=13088, MSE=410005.3
##
## Node number 94: 38 observations
##   mean=13027.89, MSE=1436389
##
## Node number 95: 9 observations
##   mean=15859.33, MSE=404578.9
```

1.B: Perform random forest on the training data (10pts). Report the variable importance (5pts) and the test error (5pts). Variable importance Test error

1.C: Perform boosting on the training data (10pts). Report the variable importance (5pts) and the test error (5pts).

QUESTION 2: This problem is based on the data "auto.csv" in Homework 3. Split the dataset into two parts: training data (70%) and test data (30%).

2.A: Build a classification tree using the training data, with mpg cat as the response (10pts). Which tree size corresponds to the lowest cross-validation error? Is this the same as the tree size obtained using the 1 SE rule (10pts)?

2.B: Perform boosting on the training data and report the variable importance (10pts). Report the test data performance (10pts).