Project 4

Carra Wu

```r
#Change this to location of your data
#Can use drop down menu in R studio: file->import data set-> from stata and find stata data set
setwd(dir = "/Users/carrawu/Documents/harvard/ec1152")


if (!require(foreign)) install.packages("foreign"); library(foreign)
```

```
## Loading required package: foreign
```

```r
if (!require(haven)) install.packages("haven"); library(haven)
```

```
## Loading required package: haven
```

```r
if (!require(randomForest)) install.packages("randomForest"); library(randomForest)
```

```
## Loading required package: randomForest
## randomForest 4.6-14
## Type rfNews() to see new features/changes/bug fixes.
```

```r
if (!require(rpart)) install.packages("rpart"); library(rpart)
```

```
## Loading required package: rpart
```

```r
#Open stata data set
proj4 <- read_dta("project4.dta")
head(proj4)
```

```
## # A tibble: 6 x 144
##   geoid place    pop housing kfr_pooled_p25  test training rank_hat_ols
##   <dbl> <chr>  <dbl>   <dbl>          <dbl> <dbl>    <dbl>        <dbl>
## 1  1001 Auta~  54907   22135         NA         1        0        0.399
## 2  1003 Bald~ 187114  104061          0.389     0        1        0.395
## 3  1005 Barb~  27321   11829          0.349     0        1        0.383
## 4  1007 Bibb~  22754    8981          0.363     0        1        0.416
## 5  1009 Blou~  57623   23887         NA         1        0        0.392
## 6  1011 Bull~  10746    4493         NA         1        0        0.427
```

```
## # ... with 136 more variables: personageyears5onwardslang <dbl>, v4 <dbl>,
## #   personageyears25onwardsedu <dbl>, v6 <dbl>,
## #   personageyears18onwardsarm <dbl>, v8 <dbl>,
## #   housingunitoccupancystatus <dbl>, housingunitcashrentstatusw <dbl>,
## #   personageyears5onwardsnati <dbl>, persongenderfemale <dbl>, v13 <dbl>,
## #   v14 <dbl>, v15 <dbl>, v16 <dbl>, P_1 <dbl>, P_2 <dbl>, P_3 <dbl>,
## #   P_4 <dbl>, P_5 <dbl>, P_6 <dbl>, P_7 <dbl>, P_8 <dbl>, P_9 <dbl>,
## #   P_10 <dbl>, P_11 <dbl>, P_12 <dbl>, P_13 <dbl>, P_14 <dbl>,
## #   P_15 <dbl>, P_16 <dbl>, P_17 <dbl>, P_18 <dbl>, P_19 <dbl>,
## #   P_20 <dbl>, P_21 <dbl>, P_22 <dbl>, P_23 <dbl>, P_24 <dbl>,
## #   P_25 <dbl>, P_26 <dbl>, P_27 <dbl>, P_28 <dbl>, P_29 <dbl>,
## #   P_30 <dbl>, P_31 <dbl>, P_32 <dbl>, P_33 <dbl>, P_34 <dbl>,
## #   P_35 <dbl>, P_36 <dbl>, P_37 <dbl>, P_38 <dbl>, P_39 <dbl>,
## #   P_40 <dbl>, P_41 <dbl>, P_42 <dbl>, P_43 <dbl>, P_44 <dbl>,
## #   P_45 <dbl>, P_46 <dbl>, P_47 <dbl>, P_48 <dbl>, P_49 <dbl>,
## #   P_50 <dbl>, P_51 <dbl>, P_52 <dbl>, P_53 <dbl>, P_54 <dbl>,
## #   P_55 <dbl>, P_56 <dbl>, P_57 <dbl>, P_58 <dbl>, P_59 <dbl>,
## #   P_60 <dbl>, P_61 <dbl>, P_62 <dbl>, P_63 <dbl>, P_64 <dbl>,
## #   P_65 <dbl>, P_66 <dbl>, P_67 <dbl>, P_68 <dbl>, P_69 <dbl>,
## #   P_70 <dbl>, P_71 <dbl>, P_72 <dbl>, P_73 <dbl>, P_74 <dbl>,
## #   P_75 <dbl>, P_76 <dbl>, P_77 <dbl>, P_78 <dbl>, P_79 <dbl>,
## #   P_80 <dbl>, P_81 <dbl>, P_82 <dbl>, P_83 <dbl>, P_84 <dbl>,
## #   P_85 <dbl>, P_86 <dbl>, ...
```

```r
#Storing predictor variables
#Order data in stata so all predictors appear in right-most columns
vars <- colnames(proj4[10:ncol(proj4)])


#OLS Regression
to_hat <- with(proj4[proj4$training==1,], lm(reformulate(vars, "kfr_pooled_p25")))
summary(to_hat)
```

```
##
## Call:
```

```
## lm(formula = reformulate(vars, "kfr_pooled_p25"))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.101989 -0.010533 -0.000746  0.010740  0.116464
##
## Coefficients: (1 not defined because of singularities)
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -4.032e+00  1.025e+01  -0.393 0.694214
## v4                        -1.119e-01  5.880e-02  -1.902 0.057383 .
## personageyears25onwardsedu -5.380e-01  2.294e-01  -2.345 0.019181 *
## v6                         1.249e-01  2.981e-02   4.190 3.01e-05 ***
## personageyears18onwardsarm -1.850e-01  4.692e-02  -3.943 8.53e-05 ***
## v8                        -6.327e-02  3.999e-02  -1.582 0.113929
## housingunitoccupancystatus  2.254e-02  1.200e-02   1.878 0.060674 .
## housingunitcashrentstatusw  1.718e-04  1.681e-02   0.010 0.991848
## personageyears5onwardsnati  5.225e-02  6.501e-02   0.804 0.421719
## persongenderfemale          2.068e-02  3.772e-02   0.548 0.583615
## v13                       -2.875e-02  1.419e-01  -0.203 0.839472
## v14                       -1.239e+02  3.884e+01  -3.189 0.001465 **
## v15                       -4.914e-01  9.405e-02  -5.225 2.07e-07 ***
## v16                        1.283e-01  5.231e-02   2.452 0.014365 *
## P_1                       -1.707e-04  8.234e-05  -2.073 0.038366 *
## P_2                        1.248e-05  9.140e-05   0.137 0.891423
## P_3                        1.151e-04  9.379e-05   1.227 0.220178
## P_4                        5.625e-05  1.043e-04   0.539 0.589846
## P_5                       -1.686e-04  1.124e-04  -1.499 0.134067
## P_6                       -3.209e-05  1.126e-04  -0.285 0.775689
## P_7                        1.288e-05  7.286e-05   0.177 0.859712
## P_8                       -7.070e-05  1.262e-04  -0.560 0.575318
## P_9                        1.520e-04  1.111e-04   1.368 0.171433
## P_10                      -5.101e-03  2.851e-03  -1.789 0.073824 .
## P_11                      -1.673e-03  3.739e-03  -0.448 0.654563
```

```
## P_12                          -9.278e-04  5.465e-04  -1.698 0.089851 .
## P_13                           1.068e-03  5.671e-04   1.883 0.059945 .
## P_14                           9.469e-04  5.475e-04   1.729 0.084009 .
## P_15                           9.178e-04  5.481e-04   1.675 0.094289 .
## P_16                           9.290e-04  5.464e-04   1.700 0.089391 .
## P_17                           8.405e-04  8.880e-04   0.947 0.344079
## P_18                           8.212e-05  6.079e-04   0.135 0.892565
## P_19                           8.358e-05  6.078e-04   0.138 0.890642
## P_20                           1.260e-04  6.076e-04   0.207 0.835722
## P_21                           2.088e-04  4.709e-04   0.443 0.657584
## P_22                          -1.998e-04  4.709e-04  -0.424 0.671429
## P_23                          -2.090e-04  4.708e-04  -0.444 0.657094
## P_24                          -1.272e-06  2.376e-07  -5.354 1.04e-07 ***
## P_25                           3.737e-05  2.603e-04   0.144 0.885894
## P_26                           2.021e-01  4.154e-02   4.866 1.30e-06 ***
## P_27                          -4.584e-02  4.134e-02  -1.109 0.267728
## P_28                          -3.699e-02  6.693e-02  -0.553 0.580628
## P_29                           1.349e-06  2.134e-07   6.322 3.72e-10 ***
## P_30                          -1.604e-06  4.644e-07  -3.454 0.000572 ***
## P_31                          -5.875e-02  2.797e-02  -2.101 0.035904 *
## P_32                           1.488e-01  4.221e-02   3.526 0.000439 ***
## P_33                          -4.395e-02  3.586e-02  -1.226 0.220636
## P_34                          -8.358e-02  1.583e-01  -0.528 0.597528
## P_35                          -1.819e-01  1.541e-01  -1.180 0.238203
## P_36                          -1.245e-01  2.354e-01  -0.529 0.596908
## P_37                           1.436e-01  1.694e-01   0.848 0.396673
## P_38                           2.611e-01  1.665e-01   1.568 0.117152
## P_39                           2.321e-01  1.653e-01   1.404 0.160728
## P_40                           3.883e-01  2.642e-01   1.470 0.141955
## P_41                           9.222e-04  1.187e-03   0.777 0.437333
## P_42                           6.068e-06  8.782e-06   0.691 0.489712
## P_43                          -3.600e-02  1.537e-02  -2.342 0.019340 *
## P_44                          -3.674e-02  3.033e-02  -1.211 0.226121
```

```
## P_45                        -1.623e-01  3.655e-02  -4.441 9.83e-06 ***
## P_46                         4.888e-02  1.171e-02   4.175 3.20e-05 ***
## P_47                        -2.778e-02  2.170e-02  -1.280 0.200810
## P_48                         1.820e-04  1.514e-04   1.202 0.229671
## P_49                        -1.287e-02  6.108e-03  -2.108 0.035273 *
## P_50                         1.850e-01  1.557e-01   1.188 0.234992
## P_51                         7.369e-06  2.566e-06   2.872 0.004160 **
## P_52                        -5.980e-06  2.423e-06  -2.468 0.013737 *
## P_53                        -1.481e-01  6.425e-02  -2.305 0.021369 *
## P_54                        -2.309e-06  4.279e-06  -0.540 0.589609
## P_55                        -7.360e-03  4.733e-03  -1.555 0.120223
## P_56                        -1.274e-02  3.599e-03  -3.539 0.000418 ***
## P_57                         2.162e-03  8.067e-04   2.680 0.007470 **
## P_58                        -9.590e-04  6.794e-04  -1.411 0.158383
## P_59                         2.406e-05  3.156e-05   0.762 0.445985
## P_60                         6.726e-06  6.072e-06   1.108 0.268202
## P_61                         1.821e-05  3.700e-05   0.492 0.622737
## P_62                        -2.743e-07  6.868e-07  -0.399 0.689721
## P_63                         3.684e-03  9.602e-04   3.837 0.000131 ***
## P_64                         5.507e-03  2.113e-03   2.607 0.009264 **
## P_65                        -5.894e-03  3.897e-03  -1.513 0.130683
## P_66                        -1.741e-05  1.704e-05  -1.022 0.307101
## P_67                        -8.380e-05  1.096e-04  -0.764 0.444831
## P_68                         2.764e-03  1.016e-03   2.719 0.006644 **
## P_69                        -2.900e-04  1.513e-04  -1.916 0.055566 .
## P_70                         1.716e-07  2.394e-07   0.717 0.473810
## P_71                        -3.238e-04  8.392e-05  -3.859 0.000121 ***
## P_72                         1.468e-05  5.851e-06   2.510 0.012224 *
## P_73                         1.423e-06  4.099e-06   0.347 0.728513
## P_74                        -8.738e-04  4.846e-04  -1.803 0.071657 .
## P_75                         3.237e-04  3.918e-04   0.826 0.408933
## P_76                         7.576e-04  2.644e-04   2.865 0.004245 **
## P_77                        -6.525e-05  4.582e-05  -1.424 0.154772
```

```
## P_78                          -2.641e-04  2.598e-04  -1.017 0.309529
## P_79                          -1.684e-04  2.477e-04  -0.680 0.496866
## P_80                          -1.192e-04  9.779e-05  -1.219 0.223024
## P_81                           5.329e-03  2.053e-03   2.596 0.009559 **
## P_82                           7.086e-02  6.705e-02   1.057 0.290858
## P_83                           7.102e-02  6.705e-02   1.059 0.289716
## P_84                           6.942e-02  6.707e-02   1.035 0.300938
## P_85                           7.228e-02  6.705e-02   1.078 0.281274
## P_86                           7.434e-02  6.708e-02   1.108 0.268003
## P_87                           7.095e-02  6.706e-02   1.058 0.290263
## P_88                           6.526e-02  6.717e-02   0.972 0.331463
## P_89                           6.443e-02  6.727e-02   0.958 0.338371
## P_90                           7.434e-02  6.722e-02   1.106 0.268957
## P_91                           7.183e-02  6.716e-02   1.070 0.285066
## P_92                           7.091e-02  6.703e-02   1.058 0.290296
## P_93                           7.078e-02  6.705e-02   1.056 0.291423
## P_94                          -2.800e-02  8.006e-02  -0.350 0.726616
## P_95                          -2.803e-02  8.007e-02  -0.350 0.726357
## P_96                          -2.727e-02  8.006e-02  -0.341 0.733508
## P_97                          -2.804e-02  8.007e-02  -0.350 0.726268
## P_98                          -3.043e-02  8.012e-02  -0.380 0.704199
## P_99                          -2.772e-02  8.006e-02  -0.346 0.729279
## P_100                         -2.278e-02  8.026e-02  -0.284 0.776635
## P_101                         -2.447e-02  8.005e-02  -0.306 0.759958
## P_102                         -3.694e-02  8.010e-02  -0.461 0.644744
## P_103                         -2.923e-02  8.006e-02  -0.365 0.715071
## P_104                         -2.812e-02  8.006e-02  -0.351 0.725461
## P_105                         -2.783e-02  8.008e-02  -0.348 0.728233
## P_106                         -9.595e-04  8.144e-04  -1.178 0.238976
## P_107                         -4.662e-04  9.753e-04  -0.478 0.632724
## P_108                          6.670e-04  8.583e-04   0.777 0.437262
## P_109                          6.431e-05  2.303e-04   0.279 0.780146
## P_110                          3.128e-03  2.280e-03   1.372 0.170389
```

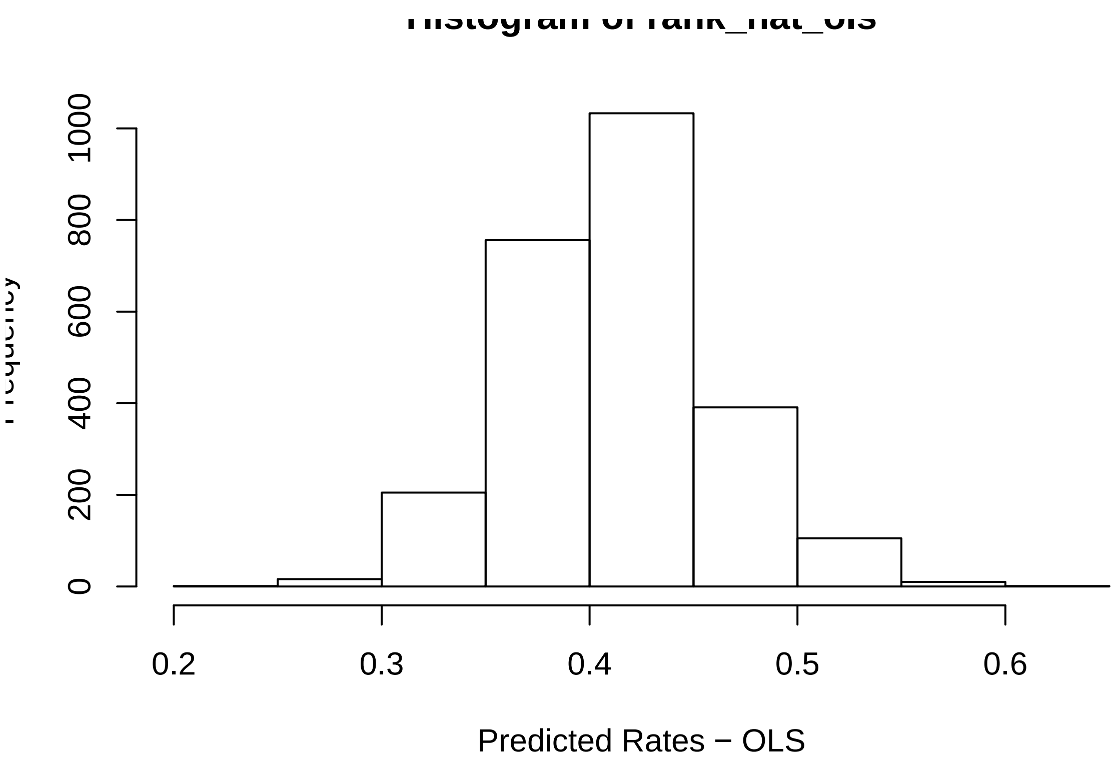
```
## P_111                          1.351e-04  3.240e-04   0.417 0.676742
## P_112                          5.077e-04  3.210e-04   1.582 0.113956
## P_113                         -2.719e-04  7.704e-05  -3.529 0.000434 ***
## P_114                          3.860e-05  4.113e-04   0.094 0.925239
## P_115                         -1.279e-04  2.321e-04  -0.551 0.581705
## P_116                         -1.893e-03  8.726e-04  -2.170 0.030248 *
## P_117                         -3.348e-04  1.170e-04  -2.862 0.004283 **
## P_118                          1.024e-03  3.403e-04   3.007 0.002694 **
## P_119                         -2.152e-04  2.094e-04  -1.028 0.304405
## P_120                          2.477e-04  2.495e-04   0.993 0.320850
## P_121                         -5.776e-04  8.697e-04  -0.664 0.506730
## mtrain                                NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01974 on 1124 degrees of freedom
## Multiple R-squared:  0.8698, Adjusted R-squared:  0.8543
## F-statistic: 56.06 on 134 and 1124 DF,  p-value: < 2.2e-16
```

```
rank_hat_ols = predict(to_hat, newdata=proj4)
```

```
## Warning in predict.lm(to_hat, newdata = proj4): prediction from a rank-
## deficient fit may be misleading
```

```
summary(rank_hat_ols); hist(rank_hat_ols, xlab="Predicted Rates - OLS")
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2339  0.3826  0.4130  0.4134  0.4422  0.6050
```




```r
#Decision Tree or Regression Tree
one_tree <- rpart(reformulate(vars, "kfr_pooled_p25")
                  , data=proj4
                  , subset = training==1
                  , control = rpart.control(xval = 10)) ## this sets the number of folds for cross vali

one_tree #Text Representation of Tree
```

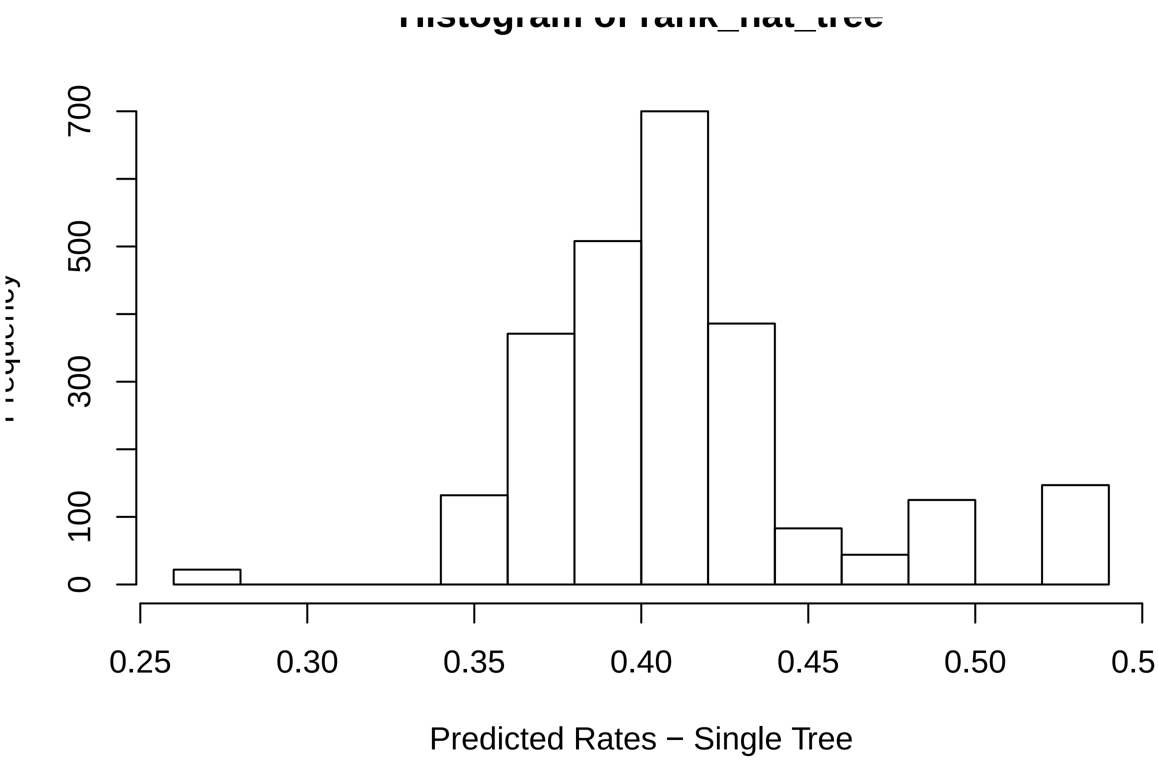
```
## n= 1259
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 1259 3.36343200 0.4132139
##    2) P_57>=14.75 805 1.35944700 0.3909387
##      4) P_37>=0.08508629 376 0.55982770 0.3676036
##        8) P_31>=0.260393 80 0.15583940 0.3312077
##          16) P_50>=0.9955153 12 0.02225725 0.2731211 *
```

```
##         17) P_50< 0.9955153 68 0.08594827 0.3414583 *
##        9) P_31< 0.260393 296 0.26937450 0.3774403
##         18) P_34>=0.1918283 183 0.10196010 0.3653738 *
##         19) P_34< 0.1918283 113 0.09761885 0.3969817 *
##      5) P_37< 0.08508629 429 0.41542780 0.4113910
##       10) P_7>=38.91737 145 0.06953851 0.3934341 *
##       11) P_7< 38.91737 284 0.27526220 0.4205592
##         22) P_45>=0.2089347 261 0.20636630 0.4169893 *
##         23) P_45< 0.2089347 23 0.02782549 0.4610691 *
##    3) P_57< 14.75 454 0.89632250 0.4527106
##      6) P_56>=3.05 323 0.32977590 0.4369945
##       12) P_34>=0.005043071 289 0.22597590 0.4317835
##         24) P_37>=0.03899463 92 0.08022539 0.4145569 *
##         25) P_37< 0.03899463 197 0.10569890 0.4398285 *
##       13) P_34< 0.005043071 34 0.02924967 0.4812873 *
##      7) P_56< 3.05 131 0.29005690 0.4914611
##       14) P_37>=0.006757407 70 0.09285862 0.4613694
##         28) P_105>=4.94 42 0.03684633 0.4431274 *
##         29) P_105< 4.94 28 0.02107163 0.4887323 *
##       15) P_37< 0.006757407 61 0.06107474 0.5259926 *
```

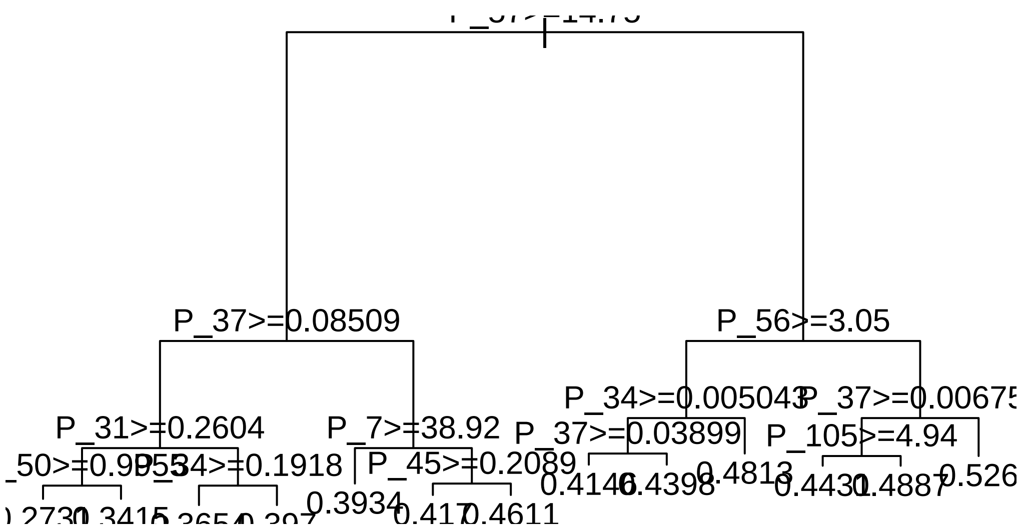
```r
rank_hat_tree <- predict(one_tree, newdata=proj4)
table(rank_hat_tree)
```

```
## rank_hat_tree
## 0.273121061424414 0.341458336833645 0.365373798406841 0.393434075445964
##                22               132               371               291
## 0.396981655496412 0.414556937049264 0.416989306152095 0.439828460591699
##               217               167               533               386
## 0.443127433458964 0.461069139449493 0.481287251500522 0.488732271960803
##                83                44                65                60
## 0.525992561070645
##               147
```

```r
hist(rank_hat_tree, xlab="Predicted Rates - Single Tree")
```




```r
plot(one_tree) # plot tree
text(one_tree) # add labels to tree
```




```r
# print complexity parameter table using cross validation
printcp(one_tree)
```

```
##
```

```
## Regression tree:
## rpart(formula = reformulate(vars, "kfr_pooled_p25"), data = proj4,
##     subset = training == 1, control = rpart.control(xval = 10))
##
## Variables actually used in tree construction:
## [1] P_105 P_31  P_34  P_37  P_45  P_50  P_56  P_57  P_7
##
## Root node error: 3.3634/1259 = 0.0026715
##
## n= 1259
##
##          CP nsplit rel error  xerror     xstd
## 1  0.329325      0   1.00000 1.00182 0.049309
## 2  0.114226      1   0.67068 0.67729 0.033871
## 3  0.082205      2   0.55645 0.57380 0.029116
## 4  0.040472      3   0.47424 0.51889 0.026479
## 5  0.040023      4   0.43377 0.50260 0.024781
## 6  0.022165      5   0.39375 0.46136 0.022345
## 7  0.020999      6   0.37158 0.43949 0.021248
## 8  0.020751      7   0.35059 0.43623 0.021192
## 9  0.014162      8   0.32984 0.40788 0.020192
## 10 0.012211      9   0.31567 0.40360 0.019854
## 11 0.011908     10   0.30346 0.39973 0.019636
## 12 0.010388     11   0.29155 0.39204 0.019580
## 13 0.010000     12   0.28117 0.38971 0.019380
```

```
#Random Forest from 1000 Bootstrapped Samples
forest_hat <- randomForest(reformulate(vars, "kfr_pooled_p25"), ntree=1000, mtry=11, maxnodes=100
                           ,importance=TRUE, do.trace=25, data=proj4[proj4$training==1,], no.action = na
```

```
##      |      Out-of-bag   |
## Tree |      MSE  %Var(y) |
##   25 | 0.0006048    22.64 |
##   50 | 0.0005575    20.87 |
```

```
##    75 | 0.0005544    20.75 |
##   100 | 0.0005579    20.88 |
##   125 | 0.0005529    20.70 |
##   150 | 0.0005468    20.47 |
##   175 | 0.0005466    20.46 |
##   200 | 0.0005493    20.56 |
##   225 | 0.0005503    20.60 |
##   250 | 0.0005469    20.47 |
##   275 | 0.0005455    20.42 |
##   300 | 0.0005459    20.43 |
##   325 | 0.0005456    20.42 |
##   350 | 0.0005441    20.37 |
##   375 | 0.0005443    20.37 |
##   400 | 0.0005443    20.38 |
##   425 | 0.0005451    20.40 |
##   450 | 0.0005433    20.34 |
##   475 | 0.0005435    20.35 |
##   500 | 0.000542    20.29 |
##   525 | 0.0005412    20.26 |
##   550 | 0.0005413    20.26 |
##   575 | 0.0005409    20.25 |
##   600 | 0.0005416    20.27 |
##   625 | 0.0005421    20.29 |
##   650 | 0.0005424    20.30 |
##   675 | 0.0005422    20.29 |
##   700 | 0.0005426    20.31 |
##   725 | 0.0005419    20.28 |
##   750 | 0.000542    20.29 |
##   775 | 0.0005416    20.27 |
##   800 | 0.0005416    20.27 |
##   825 | 0.0005419    20.29 |
##   850 | 0.0005423    20.30 |
##   875 | 0.0005425    20.31 |
```

```
## 900 | 0.0005424 20.30 |
## 925 | 0.000543 20.33 |
## 950 | 0.0005425 20.31 |
## 975 | 0.0005425 20.31 |
## 1000 | 0.0005424 20.30 |
```

```r
getTree(forest_hat, 250, labelVar = TRUE) #Text Representation of Tree
```

```
##    left daughter right daughter                          split var   split point
## 1              2              3                              P_105  3.145000e+00
## 2              4              5                               P_58  7.519521e+00
## 3              6              7                               P_71  3.287916e+01
## 4              8              9                                P_4  2.802782e+01
## 5             10             11                               P_96  3.120000e+00
## 6             12             13                               P_71  2.159720e+01
## 7             14             15                               P_43  3.847482e-01
## 8             16             17                               P_11  1.150000e+01
## 9             18             19                               P_65  8.301119e+00
## 10            20             21                               P_75  1.585000e+01
## 11            22             23                               P_51  2.712052e+01
## 12            24             25                               P_45  2.325134e-01
## 13            26             27                               P_51  3.545973e+01
## 14            28             29                               P_46  3.433064e-01
## 15            30             31                               P_84  1.755000e+00
## 16            32             33                               P_58  5.871207e+00
## 17            34             35                               P_78  3.085000e+01
## 18            36             37                               P_97  3.262000e+01
## 19            38             39                                v14 -7.014916e-05
## 20            40             41                               P_72  6.628010e+01
## 21            42             43                               P_78  2.980000e+01
## 22             0              0                               <NA>  0.000000e+00
## 23            44             45                               P_31  1.523578e-01
## 24            46             47                               P_38  9.530012e-01
## 25            48             49                               P_85  6.000000e+00
```

```
## 26          50          51                        P_55  3.050000e+00
## 27          52          53                        P_34  1.526280e-01
## 28          54          55                        P_51  5.259372e+01
## 29          56          57                        P_78  3.085000e+01
## 30          58          59                        P_10  6.100000e+00
## 31          60          61                        P_29  2.005410e+04
## 32          62          63                         P_2  2.842626e+01
## 33          64          65                        P_95  5.646000e+01
## 34           0           0                        <NA>  0.000000e+00
## 35          66          67                        P_45  2.365560e-01
## 36          68          69                        P_27  1.980439e-01
## 37           0           0                        <NA>  0.000000e+00
## 38          70          71                        P_31  2.287695e-01
## 39          72          73                        P_47  6.857073e-01
## 40           0           0                        <NA>  0.000000e+00
## 41          74          75                        P_58  8.797444e+00
## 42          76          77                        P_17  1.273779e+03
## 43          78          79                        P_34  5.122802e-02
## 44           0           0                        <NA>  0.000000e+00
## 45          80          81                         v14 -3.296223e-05
## 46          82          83                        P_62  7.213700e+03
## 47          84          85                        P_68  1.020000e+01
## 48          86          87                       P_113  6.495000e+00
## 49          88          89 personageyears5onwardsnati  3.309562e-01
## 50          90          91                        P_32  9.212385e-02
## 51          92          93                        P_20  4.858952e+01
## 52          94          95                        P_74  2.080000e+01
## 53          96          97                        P_18  6.659283e+02
## 54          98          99                       P_113  4.750000e+01
## 55         100         101 housingunitcashrentstatusw  7.398326e-02
## 56         102         103                       P_117  1.068500e+01
## 57         104         105                        P_57  1.365000e+01
## 58           0           0                        <NA>  0.000000e+00
```

```
## 59   106   107                      P_56  4.550000e+00
## 60   108   109                      P_54  2.040112e+02
## 61   110   111                      P_45  3.659864e-01
## 62   112   113                      P_55  2.400000e+00
## 63   114   115                      P_81  8.400000e+00
## 64   116   117                     P_104  3.200000e+00
## 65   118   119                      P_48  8.637618e+01
## 66     0     0                      <NA>  0.000000e+00
## 67     0     0                      <NA>  0.000000e+00
## 68     0     0                      <NA>  0.000000e+00
## 69   120   121 personageyears18onwardsarm  7.284281e-02
## 70   122   123                        v6  2.271584e-01
## 71     0     0                      <NA>  0.000000e+00
## 72   124   125                      P_61  2.650089e+01
## 73   126   127                      P_13  1.462570e-01
## 74   128   129                      P_11  9.200000e+00
## 75   130   131                      P_46  5.399068e-01
## 76     0     0                      <NA>  0.000000e+00
## 77   132   133                       v14 -3.314143e-05
## 78   134   135                       v13  4.494168e-03
## 79   136   137                      P_55  4.450000e+00
## 80   138   139 personageyears25onwardsedu  2.815344e-03
## 81     0     0                      <NA>  0.000000e+00
## 82     0     0                      <NA>  0.000000e+00
## 83   140   141                     P_103  1.220000e+00
## 84   142   143                       v14 -2.965951e-05
## 85     0     0                      <NA>  0.000000e+00
## 86     0     0                      <NA>  0.000000e+00
## 87   144   145                      P_78  2.810000e+01
## 88   146   147                      P_18  1.878525e+02
## 89     0     0                      <NA>  0.000000e+00
## 90   148   149                      P_43  3.201817e-01
## 91     0     0                      <NA>  0.000000e+00
```

```
## 92          0          0                         <NA>  0.000000e+00
## 93        150        151 personageyears25onwardsedu  2.539678e-03
## 94        152        153                         P_91  6.100000e-01
## 95        154        155                         P_56  4.200000e+00
## 96        156        157                         P_84  8.345000e+00
## 97        158        159                         P_17  2.957276e+03
## 98        160        161                         P_94  4.150500e+01
## 99        162        163                         P_41  2.238579e+00
## 100         0          0                         <NA>  0.000000e+00
## 101       164        165                         P_96  4.215000e+00
## 102       166        167                          v14 -6.321479e-05
## 103       168        169                           v4  7.705185e-04
## 104       170        171 housingunitcashrentstatusw  2.412293e-01
## 105       172        173                         P_15  6.314066e+01
## 106       174        175                          P_4  1.093886e+01
## 107       176        177                         P_40  2.086687e-03
## 108         0          0                         <NA>  0.000000e+00
## 109       178        179                         P_50  9.809754e-01
## 110       180        181                          P_4  1.862243e+01
## 111       182        183                         P_29  2.980691e+04
## 112         0          0                         <NA>  0.000000e+00
## 113         0          0                         <NA>  0.000000e+00
## 114       184        185                         P_60  3.725388e+01
## 115         0          0                         <NA>  0.000000e+00
## 116       186        187                         P_41  4.137999e+00
## 117       188        189                         P_97  9.905000e+00
## 118         0          0                         <NA>  0.000000e+00
## 119         0          0                         <NA>  0.000000e+00
## 120         0          0                         <NA>  0.000000e+00
## 121         0          0                         <NA>  0.000000e+00
## 122         0          0                         <NA>  0.000000e+00
## 123       190        191                          P_6  2.632545e+01
## 124         0          0                         <NA>  0.000000e+00
```

```
## 125           192          193                        P_41  2.897324e+00
## 126             0            0                        <NA>  0.000000e+00
## 127             0            0                        <NA>  0.000000e+00
## 128           194          195                        P_39  2.404123e-02
## 129           196          197                        P_33  1.801649e-01
## 130           198          199                        P_79  3.110000e+01
## 131             0            0                        <NA>  0.000000e+00
## 132             0            0                        <NA>  0.000000e+00
## 133             0            0                        <NA>  0.000000e+00
## 134             0            0                        <NA>  0.000000e+00
## 135             0            0                        <NA>  0.000000e+00
## 136             0            0                        <NA>  0.000000e+00
## 137             0            0                        <NA>  0.000000e+00
## 138             0            0                        <NA>  0.000000e+00
## 139             0            0                        <NA>  0.000000e+00
## 140             0            0                        <NA>  0.000000e+00
## 141             0            0                        <NA>  0.000000e+00
## 142             0            0                        <NA>  0.000000e+00
## 143             0            0                        <NA>  0.000000e+00
## 144             0            0                        <NA>  0.000000e+00
## 145             0            0                        <NA>  0.000000e+00
## 146             0            0                        <NA>  0.000000e+00
## 147             0            0                        <NA>  0.000000e+00
## 148             0            0                        <NA>  0.000000e+00
## 149             0            0                        <NA>  0.000000e+00
## 150             0            0                        <NA>  0.000000e+00
## 151             0            0                        <NA>  0.000000e+00
## 152             0            0                        <NA>  0.000000e+00
## 153             0            0                        <NA>  0.000000e+00
## 154             0            0                        <NA>  0.000000e+00
## 155             0            0                        <NA>  0.000000e+00
## 156             0            0                        <NA>  0.000000e+00
## 157             0            0                        <NA>  0.000000e+00
```

```
## 158          0          0                <NA>  0.000000e+00
## 159          0          0                <NA>  0.000000e+00
## 160          0          0                <NA>  0.000000e+00
## 161          0          0                <NA>  0.000000e+00
## 162          0          0                <NA>  0.000000e+00
## 163          0          0                <NA>  0.000000e+00
## 164          0          0                <NA>  0.000000e+00
## 165          0          0                <NA>  0.000000e+00
## 166          0          0                <NA>  0.000000e+00
## 167          0          0                <NA>  0.000000e+00
## 168          0          0                <NA>  0.000000e+00
## 169          0          0                <NA>  0.000000e+00
## 170          0          0                <NA>  0.000000e+00
## 171          0          0                <NA>  0.000000e+00
## 172          0          0                <NA>  0.000000e+00
## 173          0          0                <NA>  0.000000e+00
## 174          0          0                <NA>  0.000000e+00
## 175          0          0                <NA>  0.000000e+00
## 176          0          0                <NA>  0.000000e+00
## 177          0          0                <NA>  0.000000e+00
## 178          0          0                <NA>  0.000000e+00
## 179          0          0                <NA>  0.000000e+00
## 180          0          0                <NA>  0.000000e+00
## 181          0          0                <NA>  0.000000e+00
## 182          0          0                <NA>  0.000000e+00
## 183          0          0                <NA>  0.000000e+00
## 184          0          0                <NA>  0.000000e+00
## 185          0          0                <NA>  0.000000e+00
## 186          0          0                <NA>  0.000000e+00
## 187          0          0                <NA>  0.000000e+00
## 188          0          0                <NA>  0.000000e+00
## 189          0          0                <NA>  0.000000e+00
## 190          0          0                <NA>  0.000000e+00
```

```
## 191              0              0                       <NA>  0.000000e+00
## 192              0              0                       <NA>  0.000000e+00
## 193              0              0                       <NA>  0.000000e+00
## 194              0              0                       <NA>  0.000000e+00
## 195              0              0                       <NA>  0.000000e+00
## 196              0              0                       <NA>  0.000000e+00
## 197              0              0                       <NA>  0.000000e+00
## 198              0              0                       <NA>  0.000000e+00
## 199              0              0                       <NA>  0.000000e+00
##     status prediction
## 1       -3  0.4150675
## 2       -3  0.4596172
## 3       -3  0.4050568
## 4       -3  0.4888528
## 5       -3  0.4162352
## 6       -3  0.4328117
## 7       -3  0.3917584
## 8       -3  0.5223360
## 9       -3  0.4686072
## 10      -3  0.4289750
## 11      -3  0.3592809
## 12      -3  0.4500452
## 13      -3  0.4196814
## 14      -3  0.4035246
## 15      -3  0.3597945
## 16      -3  0.5382537
## 17      -3  0.4462848
## 18      -3  0.5060343
## 19      -3  0.4631179
## 20      -3  0.4034092
## 21      -3  0.4545408
## 22      -1  0.3897759
## 23      -3  0.3527463
```

```
## 24       -3  0.4691886
## 25       -3  0.4333881
## 26       -3  0.4615728
## 27       -3  0.4114622
## 28       -3  0.3872964
## 29       -3  0.4172743
## 30       -3  0.3981197
## 31       -3  0.3509697
## 32       -3  0.5704271
## 33       -3  0.5150889
## 34       -1  0.4009506
## 35       -3  0.4519516
## 36       -3  0.5242328
## 37       -1  0.4741869
## 38       -3  0.4366765
## 39       -3  0.4686620
## 40       -1  0.4907961
## 41       -3  0.3985544
## 42       -3  0.5012060
## 43       -3  0.4378747
## 44       -1  0.3845832
## 45       -3  0.3474402
## 46       -3  0.4570083
## 47       -3  0.5050129
## 48       -3  0.4087551
## 49       -3  0.4432413
## 50       -3  0.4999307
## 51       -3  0.4373468
## 52       -3  0.4188648
## 53       -3  0.3755459
## 54       -3  0.4087126
## 55       -3  0.3797011
## 56       -3  0.4425626
```

```
## 57       -3  0.4061528
## 58       -1  0.4605493
## 59       -3  0.3943361
## 60       -3  0.2840200
## 61       -3  0.3572312
## 62       -3  0.5981257
## 63       -3  0.5427286
## 64       -3  0.5090275
## 65       -3  0.5279693
## 66       -1  0.4579172
## 67       -1  0.4420089
## 68       -1  0.5102413
## 69       -3  0.5265647
## 70       -3  0.4436183
## 71       -1  0.3984962
## 72       -3  0.4643022
## 73       -3  0.4943365
## 74       -3  0.4110343
## 75       -3  0.3810824
## 76       -1  0.5575326
## 77       -3  0.4871243
## 78       -3  0.4655834
## 79       -3  0.4247495
## 80       -3  0.3428436
## 81       -1  0.3704231
## 82       -1  0.5068583
## 83       -3  0.4526735
## 84       -3  0.5010481
## 85       -1  0.5684507
## 86       -1  0.4419180
## 87       -3  0.3990013
## 88       -3  0.4402975
## 89       -1  0.4942672
```

```
## 90       -3  0.4770957
## 91       -1  0.5318998
## 92       -1  0.4718888
## 93       -3  0.4250104
## 94       -3  0.4213671
## 95       -3  0.3803915
## 96       -3  0.3883214
## 97       -3  0.3499950
## 98       -3  0.4160090
## 99       -3  0.3603739
## 100      -1  0.2536505
## 101      -3  0.3811841
## 102      -3  0.4331327
## 103      -3  0.5051426
## 104      -3  0.4607956
## 105      -3  0.4031339
## 106      -3  0.4027250
## 107      -3  0.3750415
## 108      -1  0.3540276
## 109      -3  0.2402652
## 110      -3  0.3698732
## 111      -3  0.3405858
## 112      -1  0.6140298
## 113      -1  0.5854024
## 114      -3  0.5293418
## 115      -1  0.5895824
## 116      -3  0.5250566
## 117      -3  0.4947794
## 118      -1  0.5209529
## 119      -1  0.5321792
## 120      -1  0.5214394
## 121      -1  0.5316899
## 122      -1  0.4610475
```

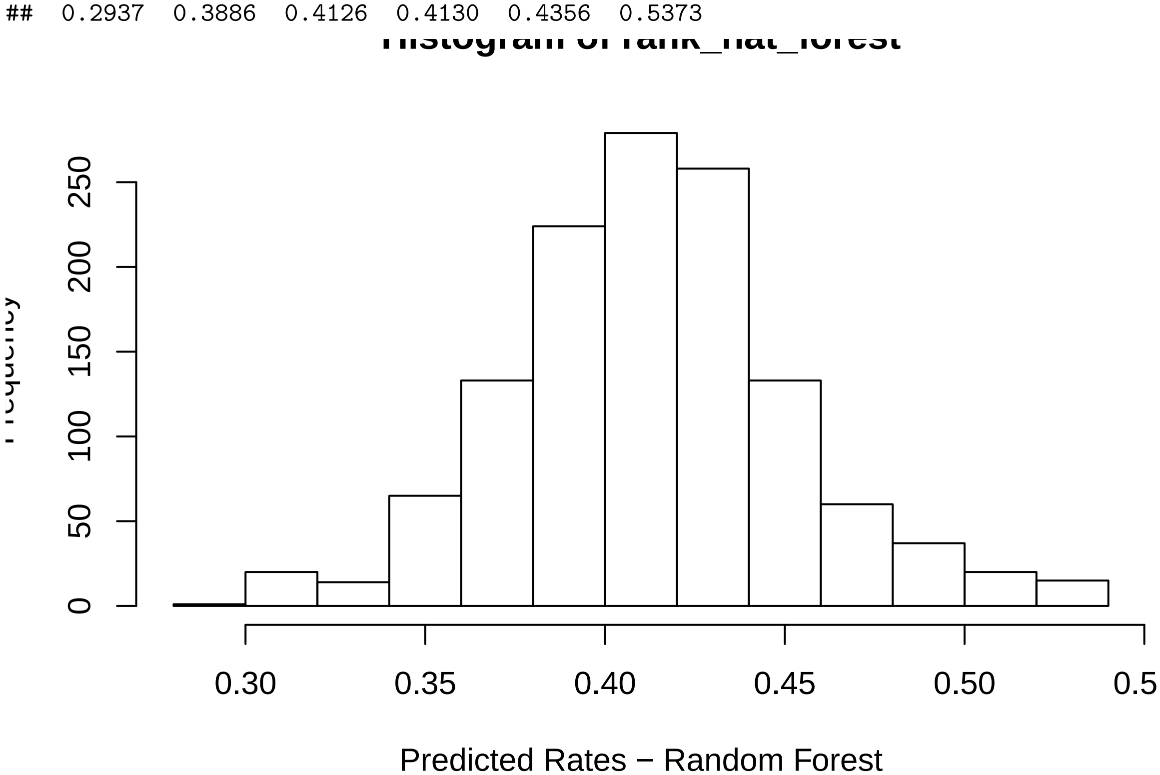
```
## 123     -3  0.4290940
## 124     -1  0.5043333
## 125     -3  0.4619004
## 126     -1  0.5156768
## 127     -1  0.4676611
## 128     -3  0.3984269
## 129     -3  0.4425528
## 130     -3  0.3717482
## 131     -1  0.4184196
## 132     -1  0.5005980
## 133     -1  0.4467032
## 134     -1  0.4783471
## 135     -1  0.4400559
## 136     -1  0.4185162
## 137     -1  0.4481243
## 138     -1  0.3466686
## 139     -1  0.3412043
## 140     -1  0.4488444
## 141     -1  0.4782005
## 142     -1  0.5125461
## 143     -1  0.4895501
## 144     -1  0.4067453
## 145     -1  0.3804155
## 146     -1  0.4753168
## 147     -1  0.4381535
## 148     -1  0.4849020
## 149     -1  0.4302574
## 150     -1  0.4606226
## 151     -1  0.4190750
## 152     -1  0.4148664
## 153     -1  0.4334615
## 154     -1  0.3883879
## 155     -1  0.3564025
```

```
## 156      -1  0.3933553
## 157      -1  0.3707028
## 158      -1  0.3779216
## 159      -1  0.3465041
## 160      -1  0.4358945
## 161      -1  0.4095462
## 162      -1  0.3801639
## 163      -1  0.3484998
## 164      -1  0.3863915
## 165      -1  0.3595654
## 166      -1  0.4891859
## 167      -1  0.4242354
## 168      -1  0.4640645
## 169      -1  0.5205469
## 170      -1  0.4768224
## 171      -1  0.4233997
## 172      -1  0.4078270
## 173      -1  0.3774895
## 174      -1  0.4624898
## 175      -1  0.4000084
## 176      -1  0.4121799
## 177      -1  0.3709150
## 178      -1  0.2279315
## 179      -1  0.2420271
## 180      -1  0.4322645
## 181      -1  0.3656576
## 182      -1  0.3645755
## 183      -1  0.3363523
## 184      -1  0.4859396
## 185      -1  0.5365755
## 186      -1  0.5306628
## 187      -1  0.5157129
## 188      -1  0.4872797
```

```
## 189     -1  0.4969222
## 190     -1  0.4359158
## 191     -1  0.4256831
## 192     -1  0.4403594
## 193     -1  0.4660034
## 194     -1  0.4033011
## 195     -1  0.3667450
## 196     -1  0.4650491
## 197     -1  0.4313046
## 198     -1  0.3798913
## 199     -1  0.3659316
```

```r
rank_hat_forest <- predict(forest_hat, data=proj4,type="response")
summary(rank_hat_forest); hist(rank_hat_forest, xlab="Predicted Rates - Random Forest")
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2937  0.3886  0.4126  0.4130  0.4356  0.5373
```




```r
proj4$rank_hat_forest = rank_hat_forest
proj4$rank_hat_ols = rank_hat_ols
```

```
proj4$rank_hat_tree = rank_hat_tree

forest_pred_error <- proj4$kfr_pooled_p25[proj4$training == 1] - proj4$rank_hat_forest[proj4$training ==
ols_pred_error <- proj4$kfr_pooled_p25[proj4$training == 1] - proj4$rank_hat_ols[proj4$training == 1]
tree_pred_error <- proj4$kfr_pooled_p25[proj4$training == 1] - proj4$rank_hat_tree[proj4$training == 1]

proj4$mse_forest = forest_pred_error^2
proj4$mse_ols = ols_pred_error^2
proj4$mse_tree = tree_pred_error^2

#mse for forest in sample
mean(proj4$mse_forest)
```

```
## [1] 0.004271803
```

```
mean(proj4$mse_ols)
```

```
## [1] 0.0003477129
```

```
mean(proj4$mse_tree)
```

```
## [1] 0.0007511369
```

```
#Export to stata
proj4$predictions_ols <- rank_hat_ols #Add OLS predictions to data set
proj4$predictions_tree <- rank_hat_tree #Add regression tree predictions to data set
proj4$predictions_forest <- rank_hat_forest #Add random forest predictions to data set
write.dta(proj4, "proj4_results.dta") #Save data as a stata .dta file

#project4$pred_error = project4$kfr_actual -project4$predictions
#project4$mse_forest = project4$pred_error^2
#mse_test <- subset(proj4, test==0, select = c(mse_forest,mse_trees,mse_ols))
#summary(mse_test)

library("haven")
```

```r
gdc<-read_dta("atlas_test.dta")
gdc<-merge(gdc,proj4,by = "geoid")

gdc$rank_hat_forest = rank_hat_forest
gdc$rank_hat_ols = rank_hat_ols
gdc$rank_hat_tree = rank_hat_tree

forest_pred_error <- gdc$kfr_actual[gdc$training == 0] - gdc$rank_hat_forest[gdc$training == 0]
ols_pred_error <- gdc$kfr_actual[gdc$training == 0] - gdc$rank_hat_ols[gdc$training == 0]
tree_pred_error <- gdc$kfr_actual[gdc$training == 0] - gdc$rank_hat_tree[gdc$training == 0]

gdc$mse_forest = forest_pred_error^2
gdc$mse_ols = ols_pred_error^2
gdc$mse_tree = tree_pred_error^2

#mse for forest in sample
mean(gdc$mse_forest)
```

```
## [1] 0.003974762
```

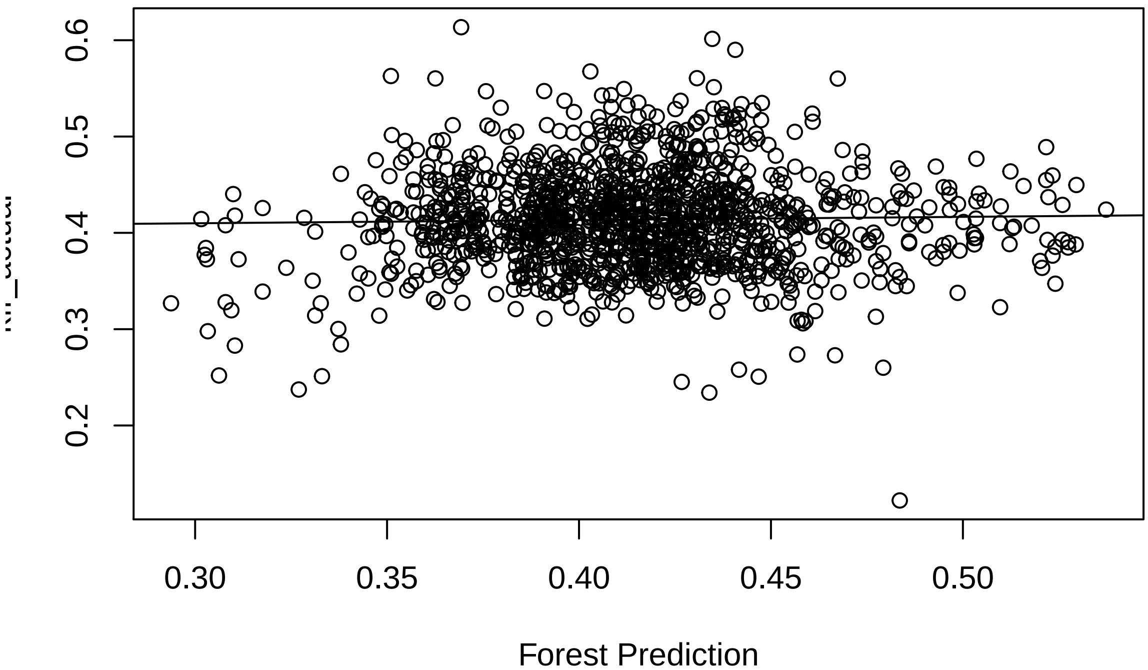
```r
mean(gdc$mse_ols)
```

```
## [1] 0.0005073468
```

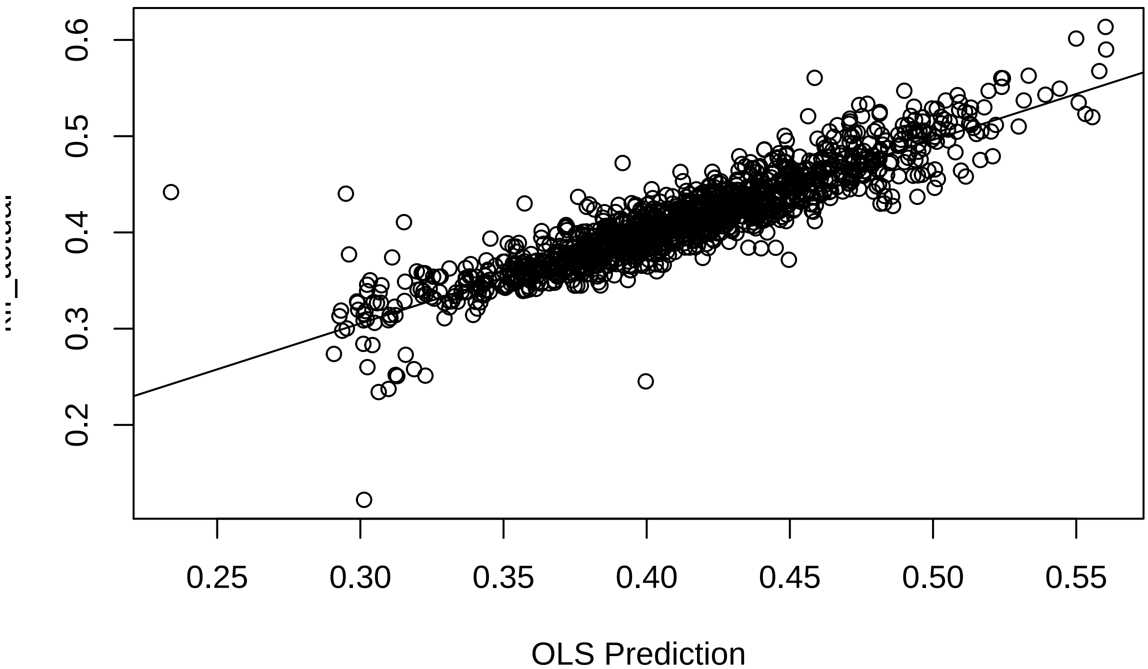
```r
mean(gdc$mse_tree)
```

```
## [1] 0.001005525
```

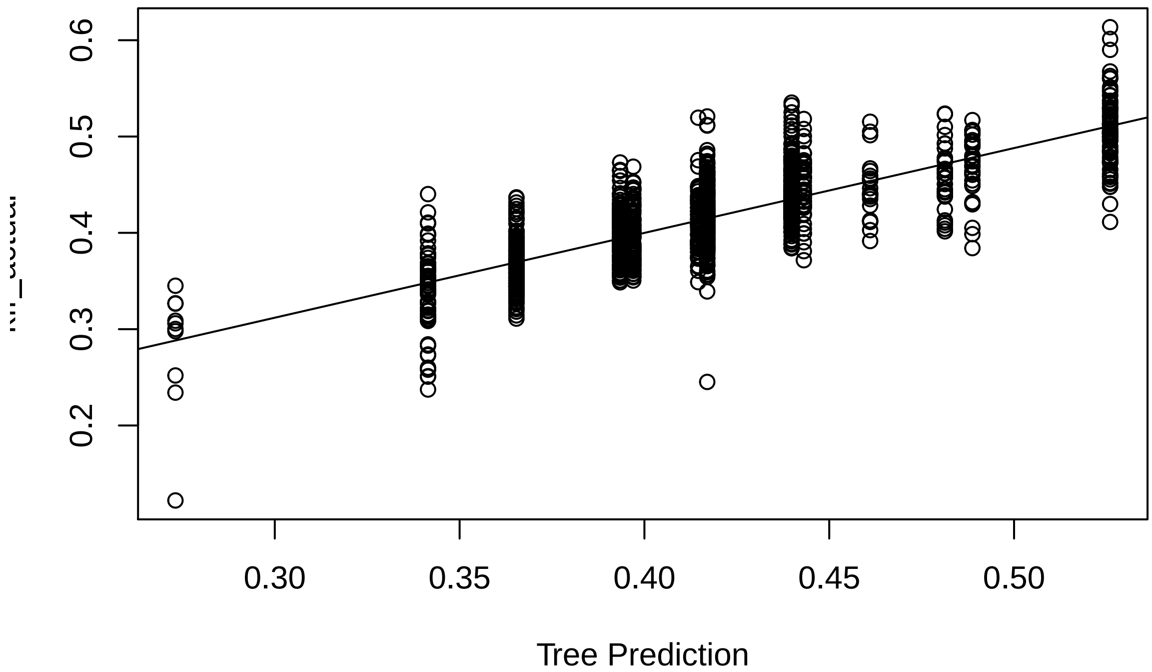
```r
gdc$predictions_ols <- rank_hat_ols #Add OLS predictions to data set
gdc$predictions_tree <- rank_hat_tree #Add regression tree predictions to data set
gdc$predictions_forest <- rank_hat_forest #Add random forest predictions to data set
write.dta(gdc, "gdc_results.dta") #Save data as a stata .dta file

plot(gdc$rank_hat_forest[gdc$test==1],gdc$kfr_actual[gdc$test==1], xlab = "Forest Prediction", ylab = "I
abline(lm(gdc$kfr_actual[gdc$test==1]~gdc$rank_hat_forest[gdc$test==1]))
```




```
plot(gdc$rank_hat_ols[gdc$test==1],gdc$kfr_actual[gdc$test==1], xlab = "OLS Prediction", ylab = "kfr_ac
abline(lm(gdc$kfr_actual[gdc$test==1]~gdc$rank_hat_ols[gdc$test==1]))
```




```
plot(gdc$rank_hat_tree[gdc$test==1],gdc$kfr_actual[gdc$test==1], xlab = "Tree Prediction", ylab = "kfr_
abline(lm(gdc$kfr_actual[gdc$test==1]~gdc$rank_hat_tree[gdc$test==1]))
```


0.2
0.3
0.4
0.5
0.6
0.30
0.35
0.40
0.45
0.50
Tree Prediction