> rm(list=ls())

> set.seed(123)

>

> #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/kwantlin/My Documents/Harvard 2018-2019/Econ 1152/Project 4")

>

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

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

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

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

> install.packages("haven")

Error in install.packages : Updating loaded packages

> library(haven)

>

> #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 predkfr\_pooled\_… mtrain foreignborn hscompletion veteran borninst

<dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl+l> <dbl> <dbl> <dbl> <dbl>

1 1001 Autau… 54907 22135 NA 1 0 0.370 3 [mat… 0.0160 0.179 0.108 0.678

2 1003 Baldw… 187114 104061 0.389 0 1 0.376 3 [mat… 0.0361 0.168 0.103 0.542

3 1005 Barbo… 27321 11829 0.349 0 1 0.328 3 [mat… 0.0294 0.183 0.0776 0.691

4 1007 Bibb … 22754 8981 0.363 0 1 0.389 3 [mat… 0.0118 0.238 0.0583 0.860

5 1009 Bloun… 57623 23887 NA 1 0 0.403 3 [mat… 0.0424 0.185 0.0788 0.798

6 1011 Bullo… 10746 4493 NA 1 0 0.296 3 [mat… 0.0536 0.207 0.0592 0.832

# ... with 131 more variables: female <dbl>, twoormorerac <dbl>, inc\_75plus <dbl>, owneroccupied <dbl>, belowpoverty <dbl>,

# married <dbl>, person <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>, P\_87 <dbl>, P\_88 <dbl>,

# P\_89 <dbl>, P\_90 <dbl>, P\_91 <dbl>, P\_92 <dbl>, P\_93 <dbl>, …

>

> #Storing predictor variables

> #Order data in stata so all predictors appear in right-most columns

> vars <- colnames(proj4[10:139])

> #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

-0.06676 -0.01101 -0.00043 0.01086 0.06080

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -4.331e+00 9.459e+00 -0.458 0.647147

foreignborn -8.099e-02 8.439e-02 -0.960 0.337461

hscompletion 1.316e-01 2.697e-02 4.881 1.21e-06 \*\*\*

veteran -1.639e-01 5.043e-02 -3.249 0.001191 \*\*

borninst 4.741e-03 6.491e-03 0.730 0.465306

female 2.309e-02 4.311e-02 0.536 0.592370

twoormorerac 7.595e-02 4.301e-02 1.766 0.077661 .

inc\_75plus 2.781e-01 7.638e-02 3.641 0.000284 \*\*\*

owneroccupied 3.224e-02 1.179e-02 2.735 0.006342 \*\*

belowpoverty 6.526e-02 3.563e-02 1.832 0.067282 .

married 5.286e-02 3.198e-02 1.653 0.098633 .

person 9.984e-09 4.335e-09 2.303 0.021466 \*

P\_1 -1.507e-04 7.582e-05 -1.988 0.047056 \*

P\_2 -1.079e-05 8.381e-05 -0.129 0.897575

P\_3 1.063e-04 8.578e-05 1.239 0.215521

P\_4 1.741e-05 9.619e-05 0.181 0.856431

P\_5 -1.096e-04 1.037e-04 -1.057 0.290837

P\_6 1.971e-05 1.029e-04 0.192 0.848094

P\_7 -1.325e-05 6.682e-05 -0.198 0.842896

P\_8 -5.816e-05 1.159e-04 -0.502 0.616045

P\_9 1.192e-04 1.020e-04 1.169 0.242761

P\_10 -6.492e-04 2.843e-03 -0.228 0.819436

P\_11 5.795e-03 3.800e-03 1.525 0.127481

P\_12 -5.106e-04 5.006e-04 -1.020 0.307976

P\_13 6.574e-04 5.192e-04 1.266 0.205720

P\_14 5.090e-04 5.015e-04 1.015 0.310350

P\_15 4.931e-04 5.019e-04 0.982 0.326104

P\_16 5.135e-04 5.006e-04 1.026 0.305222

P\_17 4.795e-04 8.142e-04 0.589 0.556004

P\_18 2.636e-05 5.597e-04 0.047 0.962446

P\_19 2.690e-05 5.597e-04 0.048 0.961676

P\_20 7.639e-05 5.595e-04 0.137 0.891429

P\_21 1.022e-04 4.335e-04 0.236 0.813718

P\_22 -8.693e-05 4.334e-04 -0.201 0.841066

P\_23 -1.060e-04 4.334e-04 -0.245 0.806842

P\_24 -1.121e-06 2.168e-07 -5.172 2.75e-07 \*\*\*

P\_25 -1.542e-05 2.403e-04 -0.064 0.948865

P\_26 1.548e-01 3.731e-02 4.149 3.60e-05 \*\*\*

P\_27 -1.161e-01 3.601e-02 -3.222 0.001309 \*\*

P\_28 9.675e-02 8.428e-02 1.148 0.251235

P\_29 1.053e-06 2.151e-07 4.893 1.14e-06 \*\*\*

P\_30 -2.115e-06 4.185e-07 -5.055 5.05e-07 \*\*\*

P\_31 -5.253e-02 3.289e-02 -1.597 0.110494

P\_32 1.067e-01 3.977e-02 2.683 0.007411 \*\*

P\_33 -5.431e-02 3.288e-02 -1.652 0.098882 .

P\_34 -5.834e-02 1.499e-01 -0.389 0.697110

P\_35 -2.315e-01 1.491e-01 -1.552 0.120832

P\_36 -1.569e-01 2.236e-01 -0.702 0.482911

P\_37 1.332e-01 1.582e-01 0.842 0.399880

P\_38 2.830e-01 1.556e-01 1.818 0.069280 .

P\_39 2.505e-01 1.594e-01 1.571 0.116471

P\_40 3.836e-01 2.513e-01 1.527 0.127060

P\_41 5.531e-04 1.098e-03 0.504 0.614564

P\_42 1.189e-06 8.061e-06 0.147 0.882813

P\_43 -2.823e-02 1.504e-02 -1.877 0.060851 .

P\_44 -2.676e-02 2.761e-02 -0.969 0.332685

P\_45 -1.820e-01 3.442e-02 -5.288 1.50e-07 \*\*\*

P\_46 6.935e-02 1.069e-02 6.487 1.32e-10 \*\*\*

P\_47 -9.129e-03 1.965e-02 -0.465 0.642259

P\_48 -6.365e-05 1.405e-04 -0.453 0.650623

P\_49 -1.074e-02 5.674e-03 -1.894 0.058519 .

P\_50 2.071e-01 1.481e-01 1.398 0.162323

P\_51 1.331e-05 1.304e-05 1.021 0.307602

P\_52 -9.607e-06 1.284e-05 -0.748 0.454540

P\_53 -1.005e-01 5.876e-02 -1.711 0.087422 .

P\_54 -7.633e-06 3.987e-06 -1.914 0.055822 .

P\_55 -6.326e-03 4.290e-03 -1.475 0.140573

P\_56 -1.853e-02 3.198e-03 -5.796 8.90e-09 \*\*\*

P\_57 2.746e-03 7.405e-04 3.709 0.000219 \*\*\*

P\_58 -7.789e-04 6.181e-04 -1.260 0.207879

P\_59 5.885e-06 2.888e-05 0.204 0.838603

P\_60 2.192e-06 5.614e-06 0.390 0.696301

P\_61 3.696e-05 3.334e-05 1.109 0.267856

P\_62 -2.536e-07 6.342e-07 -0.400 0.689340

P\_63 3.059e-03 8.725e-04 3.506 0.000473 \*\*\*

P\_64 4.296e-03 1.925e-03 2.231 0.025854 \*

P\_65 -9.334e-03 3.763e-03 -2.480 0.013270 \*

P\_66 -1.437e-05 1.571e-05 -0.914 0.360725

P\_67 -5.379e-05 1.007e-04 -0.534 0.593350

P\_68 1.549e-03 9.540e-04 1.624 0.104754

P\_69 -2.562e-04 1.396e-04 -1.835 0.066765 .

P\_70 -9.584e-07 5.805e-07 -1.651 0.099033 .

P\_71 -4.056e-04 8.037e-05 -5.047 5.24e-07 \*\*\*

P\_72 1.201e-05 5.390e-06 2.228 0.026105 \*

P\_73 5.133e-06 3.758e-06 1.366 0.172322

P\_74 -8.126e-04 4.616e-04 -1.760 0.078636 .

P\_75 8.480e-04 3.576e-04 2.372 0.017888 \*

P\_76 8.732e-04 2.439e-04 3.581 0.000358 \*\*\*

P\_77 -6.215e-05 4.195e-05 -1.481 0.138793

P\_78 -9.787e-05 2.394e-04 -0.409 0.682788

P\_79 -8.577e-05 2.291e-04 -0.374 0.708192

P\_80 -7.993e-05 9.071e-05 -0.881 0.378444

P\_81 4.915e-03 1.884e-03 2.609 0.009192 \*\*

P\_82 1.016e-01 6.177e-02 1.644 0.100387

P\_83 1.017e-01 6.176e-02 1.647 0.099922 .

P\_84 1.002e-01 6.179e-02 1.622 0.105049

P\_85 1.029e-01 6.177e-02 1.666 0.096041 .

P\_86 1.055e-01 6.180e-02 1.708 0.088000 .

P\_87 1.018e-01 6.177e-02 1.648 0.099553 .

P\_88 9.971e-02 6.189e-02 1.611 0.107461

P\_89 9.619e-02 6.197e-02 1.552 0.120942

P\_90 1.060e-01 6.191e-02 1.712 0.087244 .

P\_91 1.024e-01 6.188e-02 1.655 0.098212 .

P\_92 1.018e-01 6.174e-02 1.648 0.099572 .

P\_93 1.015e-01 6.177e-02 1.644 0.100565

P\_94 -5.612e-02 7.347e-02 -0.764 0.445117

P\_95 -5.606e-02 7.347e-02 -0.763 0.445659

P\_96 -5.544e-02 7.347e-02 -0.755 0.450659

P\_97 -5.613e-02 7.348e-02 -0.764 0.445059

P\_98 -5.943e-02 7.351e-02 -0.808 0.419032

P\_99 -5.587e-02 7.347e-02 -0.760 0.447155

P\_100 -5.523e-02 7.362e-02 -0.750 0.453326

P\_101 -5.321e-02 7.345e-02 -0.724 0.468948

P\_102 -6.468e-02 7.352e-02 -0.880 0.379169

P\_103 -5.661e-02 7.347e-02 -0.771 0.441168

P\_104 -5.622e-02 7.347e-02 -0.765 0.444346

P\_105 -5.596e-02 7.348e-02 -0.762 0.446519

P\_106 -7.117e-04 7.445e-04 -0.956 0.339302

P\_107 -9.550e-04 9.042e-04 -1.056 0.291093

P\_108 4.096e-04 7.895e-04 0.519 0.603980

P\_109 2.837e-05 2.113e-04 0.134 0.893234

P\_110 2.727e-03 2.104e-03 1.296 0.195242

P\_111 1.002e-05 2.980e-04 0.034 0.973176

P\_112 4.824e-04 2.916e-04 1.654 0.098392 .

P\_113 -2.601e-04 7.026e-05 -3.702 0.000225 \*\*\*

P\_114 -1.991e-05 3.732e-04 -0.053 0.957469

P\_115 -2.456e-04 2.128e-04 -1.154 0.248765

P\_116 -2.110e-03 8.004e-04 -2.636 0.008515 \*\*

P\_117 -3.419e-04 1.074e-04 -3.184 0.001491 \*\*

P\_118 1.018e-03 3.138e-04 3.244 0.001215 \*\*

P\_119 -2.481e-04 1.922e-04 -1.291 0.197042

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01814 on 1095 degrees of freedom

(33 observations deleted due to missingness)

Multiple R-squared: 0.8758, Adjusted R-squared: 0.861

F-statistic: 59.38 on 130 and 1095 DF, p-value: < 2.2e-16

> rank\_hat\_ols = predict(to\_hat, newdata=proj4)

> summary(rank\_hat\_ols); hist(rank\_hat\_ols, xlab="Predicted Rates - OLS")

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

0.2427 0.3856 0.4153 0.4166 0.4444 0.5974 71

>

> #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 validation.

>

> one\_tree #Text Representation of Tree

n= 1259

node), split, n, deviance, yval

\* denotes terminal node

1) root 1259 3.363432000 0.4132139

2) P\_57>=14.75 805 1.359447000 0.3909387

4) P\_37>=0.08508629 376 0.559827700 0.3676036

8) P\_31>=0.260393 80 0.155839400 0.3312077

16) veteran< 0.01978733 7 0.008164361 0.2516216 \*

17) veteran>=0.01978733 73 0.099085690 0.3388393 \*

9) P\_31< 0.260393 296 0.269374500 0.3774403

18) P\_34>=0.1918283 183 0.101960100 0.3653738 \*

19) P\_34< 0.1918283 113 0.097618850 0.3969817 \*

5) P\_37< 0.08508629 429 0.415427800 0.4113910

10) P\_7>=38.91737 145 0.069538510 0.3934341 \*

11) P\_7< 38.91737 284 0.275262200 0.4205592

22) P\_45>=0.2089347 261 0.206366300 0.4169893 \*

23) P\_45< 0.2089347 23 0.027825490 0.4610691 \*

3) P\_57< 14.75 454 0.896322500 0.4527106

6) P\_56>=3.05 323 0.329775900 0.4369945

12) P\_34>=0.005043071 289 0.225975900 0.4317835

24) P\_37>=0.03899463 92 0.080225390 0.4145569 \*

25) P\_37< 0.03899463 197 0.105698900 0.4398285 \*

13) P\_34< 0.005043071 34 0.029249670 0.4812873 \*

7) P\_56< 3.05 131 0.290056900 0.4914611

14) P\_37>=0.006757407 70 0.092858620 0.4613694

28) P\_105>=4.94 42 0.036846330 0.4431274 \*

29) P\_105< 4.94 28 0.021071630 0.4887323 \*

15) P\_37< 0.006757407 61 0.061074740 0.5259926 \*

> rank\_hat\_tree <- predict(one\_tree, newdata=proj4)

> table(rank\_hat\_tree)

rank\_hat\_tree

0.251621555004801 0.338839298037634 0.365373798406841 0.393434075445964 0.396981655496412 0.414556937049264 0.416989306152095

17 137 371 291 217 167 533

0.439828460591699 0.443127433458964 0.461069139449493 0.481287251500522 0.488732271960803 0.525992561070645

386 83 44 65 60 147

> hist(rank\_hat\_tree, xlab="Predicted Rates - Single Tree")

> summary(rank\_hat\_tree)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.2516 0.3934 0.4170 0.4143 0.4398 0.5260

>

> plot(one\_tree) # plot tree

> text(one\_tree) # add labels to tree

> # 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\_56 P\_57 P\_7 veteran

Root node error: 3.3634/1259 = 0.0026715

n= 1259

CP nsplit rel error xerror xstd

1 0.329325 0 1.00000 1.00090 0.049265

2 0.114226 1 0.67068 0.67874 0.033880

3 0.082205 2 0.55645 0.57966 0.029192

4 0.040472 3 0.47424 0.52295 0.027100

5 0.040023 4 0.43377 0.50836 0.026535

6 0.022165 5 0.39375 0.46584 0.023318

7 0.020999 6 0.37158 0.43492 0.021352

8 0.020751 7 0.35059 0.43180 0.021205

9 0.014446 8 0.32984 0.41498 0.020698

10 0.012211 9 0.31539 0.40548 0.020057

11 0.011908 10 0.30318 0.40429 0.019624

12 0.010388 11 0.29127 0.40177 0.019001

13 0.010000 12 0.28088 0.39797 0.018883

>

> #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,], na.action=na.exclude)

| Out-of-bag |

Tree | MSE %Var(y) |

25 | 0.0005544 23.44 |

50 | 0.0005043 21.32 |

75 | 0.0004998 21.13 |

100 | 0.0004844 20.48 |

125 | 0.0004801 20.30 |

150 | 0.0004745 20.06 |

175 | 0.0004725 19.98 |

200 | 0.0004725 19.98 |

225 | 0.0004693 19.84 |

250 | 0.0004693 19.84 |

275 | 0.0004672 19.75 |

300 | 0.0004677 19.78 |

325 | 0.0004674 19.76 |

350 | 0.0004674 19.76 |

375 | 0.0004676 19.77 |

400 | 0.0004672 19.75 |

425 | 0.0004674 19.76 |

450 | 0.0004667 19.73 |

475 | 0.0004665 19.72 |

500 | 0.0004661 19.71 |

525 | 0.0004667 19.73 |

550 | 0.0004671 19.75 |

575 | 0.0004667 19.73 |

600 | 0.0004661 19.71 |

625 | 0.000466 19.70 |

650 | 0.0004658 19.69 |

675 | 0.0004652 19.67 |

700 | 0.0004651 19.67 |

725 | 0.0004654 19.68 |

750 | 0.0004659 19.70 |

775 | 0.0004648 19.65 |

800 | 0.0004645 19.64 |

825 | 0.0004645 19.64 |

850 | 0.0004641 19.62 |

875 | 0.0004635 19.60 |

900 | 0.0004636 19.60 |

925 | 0.0004634 19.59 |

950 | 0.0004631 19.58 |

975 | 0.0004629 19.57 |

1000 | 0.000463 19.57 |

> getTree(forest\_hat, 250, labelVar = TRUE) #Text Representation of Tree

left daughter right daughter split var split point status prediction

1 2 3 P\_66 3.615500e+02 -3 0.4160989

2 4 5 P\_43 2.746568e-01 -3 0.4456030

3 6 7 P\_45 3.220091e-01 -3 0.3937807

4 8 9 P\_84 1.550000e-01 -3 0.4667058

5 10 11 P\_73 5.815548e+01 -3 0.4226651

6 12 13 P\_3 6.906212e+01 -3 0.4058821

7 14 15 P\_34 6.131369e-02 -3 0.3573372

8 16 17 P\_7 1.667493e+01 -3 0.4816520

9 18 19 P\_82 4.271500e+01 -3 0.4383868

10 20 21 P\_70 3.825000e+02 -3 0.4501386

11 22 23 P\_37 5.401682e-02 -3 0.4160661

12 24 25 P\_52 2.272588e+01 -3 0.4149077

13 26 27 P\_48 3.317486e+01 -3 0.3878826

14 28 29 P\_41 2.167863e+00 -3 0.3963770

15 30 31 P\_38 4.722328e-01 -3 0.3536542

16 32 33 P\_97 6.353500e+01 -3 0.5245813

17 34 35 P\_64 1.220375e+01 -3 0.4701639

18 36 37 P\_18 1.475355e+02 -3 0.4601592

19 38 39 P\_23 5.533300e+02 -3 0.4196028

20 40 41 P\_105 1.770000e+00 -3 0.4727381

21 42 43 P\_60 1.431951e+02 -3 0.4345527

22 44 45 P\_17 1.024768e+03 -3 0.4308996

23 46 47 P\_84 7.460000e+00 -3 0.3972771

24 48 49 P\_75 1.815000e+01 -3 0.4449781

25 50 51 P\_34 5.528096e-02 -3 0.4065364

26 0 0 <NA> 0.000000e+00 -1 0.5221364

27 52 53 P\_34 2.326384e-02 -3 0.3863306

28 0 0 <NA> 0.000000e+00 -1 0.3698320

29 54 55 foreignborn 4.653428e-02 -3 0.4030133

30 56 57 P\_17 2.375047e+03 -3 0.3389165

31 58 59 P\_60 1.642565e+01 -3 0.3606137

32 60 61 P\_38 9.320530e-01 -3 0.5221637

33 0 0 <NA> 0.000000e+00 -1 0.6140298

34 62 63 P\_55 2.950000e+00 -3 0.4949485

35 64 65 P\_43 1.866150e-01 -3 0.4607795

36 0 0 <NA> 0.000000e+00 -1 0.5051705

37 66 67 P\_43 2.430612e-01 -3 0.4568657

38 68 69 P\_40 4.642555e-02 -3 0.4182576

39 0 0 <NA> 0.000000e+00 -1 0.4868614

40 70 71 P\_51 5.472520e+01 -3 0.5002179

41 72 73 person 1.575900e+04 -3 0.4609611

42 74 75 P\_44 1.816605e-01 -3 0.4250692

43 76 77 foreignborn 1.713362e-02 -3 0.4594469

44 78 79 P\_27 1.189477e-01 -3 0.4515587

45 80 81 married 4.364245e-01 -3 0.4262346

46 82 83 P\_66 2.916500e+02 -3 0.4026312

47 84 85 P\_3 7.108452e+01 -3 0.3423977

48 86 87 P\_57 1.715000e+01 -3 0.4362150

49 88 89 P\_8 1.460931e+01 -3 0.4806148

50 90 91 P\_50 1.530116e-01 -3 0.4165913

51 92 93 P\_51 1.173201e+02 -3 0.3901972

52 94 95 P\_6 5.225006e+01 -3 0.4005160

53 96 97 P\_38 7.345310e-01 -3 0.3789890

54 98 99 P\_46 6.058751e-01 -3 0.4087116

55 0 0 <NA> 0.000000e+00 -1 0.3916166

56 100 101 P\_46 3.478728e-01 -3 0.3483270

57 102 103 P\_37 6.572331e-01 -3 0.3291297

58 104 105 P\_8 4.988511e+01 -3 0.3788615

59 106 107 P\_62 1.106192e+04 -3 0.3585443

60 108 109 P\_77 6.665687e+01 -3 0.4863904

61 110 111 P\_43 2.002380e-01 -3 0.5290876

62 112 113 P\_27 2.414894e-01 -3 0.5093572

63 114 115 P\_26 2.018224e-01 -3 0.4742359

64 116 117 P\_82 5.515500e+01 -3 0.4942396

65 118 119 P\_54 1.511688e+01 -3 0.4516915

66 120 121 P\_49 4.744670e-02 -3 0.4650419

67 122 123 P\_33 6.521234e-02 -3 0.4464182

68 124 125 P\_58 8.320290e+00 -3 0.4154947

69 0 0 <NA> 0.000000e+00 -1 0.4500315

70 0 0 <NA> 0.000000e+00 -1 0.5050718

71 0 0 <NA> 0.000000e+00 -1 0.4905102

72 0 0 <NA> 0.000000e+00 -1 0.4336847

73 126 127 P\_60 1.332146e+02 -3 0.4718716

74 128 129 P\_53 -4.253343e-03 -3 0.4397901

75 130 131 P\_28 3.212478e-02 -3 0.4088762

76 0 0 <NA> 0.000000e+00 -1 0.4333524

77 132 133 hscompletion 1.918651e-01 -3 0.4631746

78 0 0 <NA> 0.000000e+00 -1 0.3984722

79 134 135 P\_3 5.412072e+01 -3 0.4571468

80 136 137 P\_27 1.158674e-01 -3 0.4316274

81 138 139 P\_72 2.086329e+02 -3 0.4143332

82 140 141 female 5.165862e-01 -3 0.4291726

83 142 143 P\_65 1.281239e+01 -3 0.3966891

84 0 0 <NA> 0.000000e+00 -1 0.3550986

85 0 0 <NA> 0.000000e+00 -1 0.3296968

86 144 145 P\_58 7.832679e+00 -3 0.4561223

87 146 147 P\_113 7.875000e+00 -3 0.4265041

88 148 149 married 4.256757e-01 -3 0.4981020

89 0 0 <NA> 0.000000e+00 -1 0.4325250

90 150 151 hscompletion 2.763536e-01 -3 0.4136879

91 152 153 P\_55 5.050000e+00 -3 0.4306078

92 154 155 P\_43 3.767984e-01 -3 0.4007245

93 156 157 P\_93 8.600000e+00 -3 0.3719130

94 158 159 P\_36 1.745728e-03 -3 0.4138478

95 160 161 P\_75 1.615000e+01 -3 0.3913742

96 162 163 P\_95 7.695000e+00 -3 0.3593258

97 164 165 P\_52 9.088548e+01 -3 0.3824351

98 166 167 P\_62 8.393590e+03 -3 0.4082107

99 0 0 <NA> 0.000000e+00 -1 0.4102142

100 168 169 P\_47 4.110621e-01 -3 0.3546897

101 170 171 P\_33 4.112037e-01 -3 0.3381467

102 172 173 P\_35 3.270620e-02 -3 0.3339872

103 174 175 P\_72 1.310077e+03 -3 0.3137476

104 176 177 P\_34 3.411272e-01 -3 0.3916412

105 0 0 <NA> 0.000000e+00 -1 0.3635258

106 178 179 P\_66 5.771500e+02 -3 0.3557261

107 180 181 P\_48 7.437304e+01 -3 0.3739508

108 0 0 <NA> 0.000000e+00 -1 0.4648158

109 0 0 <NA> 0.000000e+00 -1 0.4971776

110 182 183 P\_50 1.879433e-02 -3 0.5469553

111 184 185 hscompletion 2.048426e-01 -3 0.5217781

112 186 187 P\_104 5.220000e+00 -3 0.5299490

113 188 189 P\_73 1.333341e+02 -3 0.4961197

114 190 191 P\_3 5.969691e+01 -3 0.4806683

115 0 0 <NA> 0.000000e+00 -1 0.4600846

116 192 193 P\_23 3.375149e+02 -3 0.5078109

117 0 0 <NA> 0.000000e+00 -1 0.4331691

118 194 195 borninst 5.732581e-01 -3 0.4757359

119 196 197 P\_9 3.958494e+01 -3 0.4409574

120 198 199 P\_53 9.469752e-04 -3 0.4702745

121 0 0 <NA> 0.000000e+00 -1 0.4502163

122 0 0 <NA> 0.000000e+00 -1 0.4277673

123 0 0 <NA> 0.000000e+00 -1 0.4517470

124 0 0 <NA> 0.000000e+00 -1 0.4175898

125 0 0 <NA> 0.000000e+00 -1 0.3854641

126 0 0 <NA> 0.000000e+00 -1 0.4677440

127 0 0 <NA> 0.000000e+00 -1 0.4815028

128 0 0 <NA> 0.000000e+00 -1 0.4304538

129 0 0 <NA> 0.000000e+00 -1 0.4509937

130 0 0 <NA> 0.000000e+00 -1 0.3960560

131 0 0 <NA> 0.000000e+00 -1 0.4174229

132 0 0 <NA> 0.000000e+00 -1 0.4581109

133 0 0 <NA> 0.000000e+00 -1 0.4758341

134 0 0 <NA> 0.000000e+00 -1 0.4718069

135 0 0 <NA> 0.000000e+00 -1 0.4408578

136 0 0 <NA> 0.000000e+00 -1 0.4774395

137 0 0 <NA> 0.000000e+00 -1 0.4301496

138 0 0 <NA> 0.000000e+00 -1 0.4286847

139 0 0 <NA> 0.000000e+00 -1 0.4055629

140 0 0 <NA> 0.000000e+00 -1 0.4415359

141 0 0 <NA> 0.000000e+00 -1 0.4044460

142 0 0 <NA> 0.000000e+00 -1 0.4148236

143 0 0 <NA> 0.000000e+00 -1 0.3900275

144 0 0 <NA> 0.000000e+00 -1 0.4452980

145 0 0 <NA> 0.000000e+00 -1 0.4885951

146 0 0 <NA> 0.000000e+00 -1 0.4429872

147 0 0 <NA> 0.000000e+00 -1 0.4148286

148 0 0 <NA> 0.000000e+00 -1 0.4567142

149 0 0 <NA> 0.000000e+00 -1 0.5217522

150 0 0 <NA> 0.000000e+00 -1 0.4075683

151 0 0 <NA> 0.000000e+00 -1 0.4405203

152 0 0 <NA> 0.000000e+00 -1 0.4339915

153 0 0 <NA> 0.000000e+00 -1 0.3849285

154 0 0 <NA> 0.000000e+00 -1 0.4059307

155 0 0 <NA> 0.000000e+00 -1 0.3813870

156 0 0 <NA> 0.000000e+00 -1 0.3488226

157 0 0 <NA> 0.000000e+00 -1 0.3780704

158 0 0 <NA> 0.000000e+00 -1 0.3791941

159 0 0 <NA> 0.000000e+00 -1 0.4169982

160 0 0 <NA> 0.000000e+00 -1 0.3860781

161 0 0 <NA> 0.000000e+00 -1 0.4125589

162 0 0 <NA> 0.000000e+00 -1 0.3959347

163 0 0 <NA> 0.000000e+00 -1 0.3514811

164 0 0 <NA> 0.000000e+00 -1 0.3922872

165 0 0 <NA> 0.000000e+00 -1 0.3749117

166 0 0 <NA> 0.000000e+00 -1 0.4088186

[ reached getOption("max.print") -- omitted 33 rows ]

> rank\_hat\_forest <- predict(forest\_hat, newdata=proj4,type="response")

> summary(rank\_hat\_forest); hist(rank\_hat\_forest, xlab="Predicted Rates - Random Forest")

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

0.3232 0.3902 0.4141 0.4166 0.4376 0.5680 71

>

> #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

> install.packages("haven")

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/haven\_2.1.0.tgz'

Content type 'application/x-gzip' length 911655 bytes (890 KB)

==================================================

downloaded 890 KB

The downloaded binary packages are in

/var/folders/n4/fp2df3b53qndktss1k37y0rh0000gn/T//Rtmpv6XNa9/downloaded\_packages