## Course Project Part 2

### Kenny Buckenmeyer

Data Cleanup

ames\_student <- read\_csv("ames\_student.csv")

## Warning: Missing column names filled in: 'X1' [1]

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## .default = col\_character(),  
## X1 = col\_double(),  
## Lot\_Frontage = col\_double(),  
## Lot\_Area = col\_double(),  
## Year\_Built = col\_double(),  
## Year\_Remod\_Add = col\_double(),  
## Mas\_Vnr\_Area = col\_double(),  
## BsmtFin\_SF\_1 = col\_double(),  
## BsmtFin\_SF\_2 = col\_double(),  
## Bsmt\_Unf\_SF = col\_double(),  
## Total\_Bsmt\_SF = col\_double(),  
## First\_Flr\_SF = col\_double(),  
## Second\_Flr\_SF = col\_double(),  
## Low\_Qual\_Fin\_SF = col\_double(),  
## Gr\_Liv\_Area = col\_double(),  
## Bsmt\_Full\_Bath = col\_double(),  
## Bsmt\_Half\_Bath = col\_double(),  
## Full\_Bath = col\_double(),  
## Half\_Bath = col\_double(),  
## Bedroom\_AbvGr = col\_double(),  
## Kitchen\_AbvGr = col\_double()  
## # ... with 15 more columns  
## )  
## i Use `spec()` for the full column specifications.

ames\_student <- ames\_student %>% mutate\_if(is.character, as\_factor)  
  
summary(ames\_student)

## X1 MS\_SubClass   
## Min. : 1 One\_Story\_1946\_and\_Newer\_All\_Styles :772   
## 1st Qu.: 514 Two\_Story\_1946\_and\_Newer :383   
## Median :1027 One\_and\_Half\_Story\_Finished\_All\_Ages:204   
## Mean :1027 One\_Story\_PUD\_1946\_and\_Newer :129   
## 3rd Qu.:1540 One\_Story\_1945\_and\_Older : 98   
## Max. :2053 Two\_Story\_1945\_and\_Older : 95   
## (Other) :372   
## MS\_Zoning Lot\_Frontage Lot\_Area   
## Residential\_Low\_Density :1600 Min. : 0.00 Min. : 1300   
## Residential\_High\_Density : 20 1st Qu.: 43.00 1st Qu.: 7500   
## Floating\_Village\_Residential: 87 Median : 62.00 Median : 9548   
## Residential\_Medium\_Density : 326 Mean : 57.38 Mean : 10258   
## C\_all : 17 3rd Qu.: 78.00 3rd Qu.: 11600   
## A\_agr : 2 Max. :313.00 Max. :215245   
## I\_all : 1   
## Street Alley Lot\_Shape Land\_Contour  
## Pave:2046 No\_Alley\_Access:1914 Slightly\_Irregular : 714 Lvl:1833   
## Grvl: 7 Paved : 45 Regular :1275 HLS: 94   
## Gravel : 94 Moderately\_Irregular: 53 Bnk: 81   
## Irregular : 11 Low: 45   
##   
##   
##   
## Utilities Lot\_Config Land\_Slope Neighborhood Condition\_1   
## AllPub:2052 Corner : 359 Gtl:1951 North\_Ames : 327 Norm :1771   
## NoSewr: 1 Inside :1495 Mod: 89 College\_Creek: 183 Feedr : 113   
## CulDSac: 135 Sev: 13 Old\_Town : 181 Artery : 67   
## FR2 : 56 Edwards : 129 RRAn : 35   
## FR3 : 8 Somerset : 119 PosN : 24   
## Gilbert : 109 RRAe : 19   
## (Other) :1005 (Other): 24   
## Condition\_2 Bldg\_Type House\_Style Overall\_Qual  
## Norm :2027 OneFam :1706 One\_Story :1052 Average :587   
## Feedr : 12 TwnhsE : 157 Two\_Story : 590 Above\_Average:518   
## PosA : 4 Twnhs : 67 One\_and\_Half\_Fin: 225 Good :411   
## Artery : 4 Duplex : 76 SLvl : 90 Very\_Good :237   
## PosN : 3 TwoFmCon: 47 SFoyer : 56 Below\_Average:169   
## RRNn : 1 Two\_and\_Half\_Unf: 19 Excellent : 70   
## (Other): 2 (Other) : 21 (Other) : 61   
## Overall\_Cond Year\_Built Year\_Remod\_Add Roof\_Style   
## Average :1143 Min. :1875 Min. :1950 Hip : 404   
## Above\_Average: 376 1st Qu.:1953 1st Qu.:1965 Gable :1607   
## Good : 286 Median :1972 Median :1993 Mansard: 9   
## Very\_Good : 98 Mean :1971 Mean :1984 Gambrel: 14   
## Below\_Average: 73 3rd Qu.:2000 3rd Qu.:2004 Shed : 5   
## Fair : 35 Max. :2010 Max. :2010 Flat : 14   
## (Other) : 42   
## Roof\_Matl Exterior\_1st Exterior\_2nd Mas\_Vnr\_Type Mas\_Vnr\_Area   
## CompShg:2023 VinylSd:705 VinylSd:699 Stone : 166 Min. : 0.0   
## WdShake: 8 MetalSd:319 MetalSd:317 None :1231 1st Qu.: 0.0   
## Tar&Grv: 17 Wd Sdng:313 Wd Sdng:302 BrkFace: 638 Median : 0.0   
## WdShngl: 3 HdBoard:303 HdBoard:277 BrkCmn : 17 Mean : 103.8   
## Roll : 1 Plywood:151 Plywood:190 CBlock : 1 3rd Qu.: 164.0   
## Metal : 1 CemntBd: 90 CmentBd: 90 Max. :1600.0   
## (Other):172 (Other):178   
## Exter\_Qual Exter\_Cond Foundation Bsmt\_Qual   
## Typical :1272 Typical :1787 CBlock:880 Typical :911   
## Good : 682 Good : 213 PConc :911 Good :849   
## Excellent: 78 Fair : 43 Wood : 4 Excellent :178   
## Fair : 21 Excellent: 9 BrkTil:216 No\_Basement: 57   
## Poor : 1 Slab : 36 Fair : 57   
## Stone : 6 Poor : 1   
##   
## Bsmt\_Cond Bsmt\_Exposure BsmtFin\_Type\_1 BsmtFin\_SF\_1   
## Good : 80 Gd : 199 BLQ :196 Min. :1.00   
## Typical :1833 No :1331 Rec :216 1st Qu.:3.00   
## Poor : 4 Av : 284 ALQ :298 Median :3.00   
## No\_Basement: 57 Mn : 179 GLQ :578 Mean :4.21   
## Fair : 76 No\_Basement: 60 Unf :602 3rd Qu.:7.00   
## Excellent : 3 LwQ :106 Max. :7.00   
## No\_Basement: 57   
## BsmtFin\_Type\_2 BsmtFin\_SF\_2 Bsmt\_Unf\_SF Total\_Bsmt\_SF   
## Unf :1740 Min. : 0.00 Min. : 0.0 Min. : 0   
## LwQ : 64 1st Qu.: 0.00 1st Qu.: 226.0 1st Qu.: 793   
## BLQ : 47 Median : 0.00 Median : 460.0 Median : 988   
## Rec : 79 Mean : 52.57 Mean : 561.2 Mean :1055   
## GLQ : 23 3rd Qu.: 0.00 3rd Qu.: 801.0 3rd Qu.:1304   
## No\_Basement: 58 Max. :1526.00 Max. :2336.0 Max. :5095   
## ALQ : 42   
## Heating Heating\_QC Central\_Air Electrical First\_Flr\_SF   
## GasA :2019 Fair : 61 Y:1916 SBrkr :1887 Min. : 432   
## GasW : 21 Typical : 618 N: 137 FuseA : 126 1st Qu.: 882   
## Grav : 6 Excellent:1040 FuseF : 33 Median :1088   
## Wall : 5 Good : 333 FuseP : 6 Mean :1168   
## Floor: 1 Poor : 1 Unknown: 1 3rd Qu.:1402   
## OthW : 1 Max. :5095   
##   
## Second\_Flr\_SF Low\_Qual\_Fin\_SF Gr\_Liv\_Area Bsmt\_Full\_Bath   
## Min. : 0.0 Min. : 0.000 Min. : 480 Min. :0.0000   
## 1st Qu.: 0.0 1st Qu.: 0.000 1st Qu.:1137 1st Qu.:0.0000   
## Median : 0.0 Median : 0.000 Median :1447 Median :0.0000   
## Mean : 326.1 Mean : 4.973 Mean :1499 Mean :0.4301   
## 3rd Qu.: 701.0 3rd Qu.: 0.000 3rd Qu.:1737 3rd Qu.:1.0000   
## Max. :1862.0 Max. :1064.000 Max. :5095 Max. :3.0000   
##   
## Bsmt\_Half\_Bath Full\_Bath Half\_Bath Bedroom\_AbvGr   
## Min. :0.00000 Min. :0.000 Min. :0.0000 Min. :0.000   
## 1st Qu.:0.00000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000   
## Median :0.00000 Median :2.000 Median :0.0000 Median :3.000   
## Mean :0.05796 Mean :1.564 Mean :0.3751 Mean :2.855   
## 3rd Qu.:0.00000 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000   
## Max. :2.00000 Max. :4.000 Max. :2.0000 Max. :6.000   
##   
## Kitchen\_AbvGr Kitchen\_Qual TotRms\_AbvGrd Functional   
## Min. :1.000 Typical :1070 Min. : 3.000 Typ :1896   
## 1st Qu.:1.000 Good : 790 1st Qu.: 5.000 Min2 : 54   
## Median :1.000 Excellent: 142 Median : 6.000 Min1 : 51   
## Mean :1.047 Fair : 50 Mean : 6.442 Mod : 27   
## 3rd Qu.:1.000 Poor : 1 3rd Qu.: 7.000 Maj1 : 15   
## Max. :3.000 Max. :15.000 Maj2 : 6   
## (Other): 4   
## Fireplaces Fireplace\_Qu Garage\_Type Garage\_Finish  
## Min. :0.000 Good :538 Attchd :1204 Fin :509   
## 1st Qu.:0.000 No\_Fireplace:993 BuiltIn : 127 Unf :872   
## Median :1.000 Typical :409 Basment : 29 RFn :563   
## Mean :0.603 Poor : 36 Detchd : 549 No\_Garage:109   
## 3rd Qu.:1.000 Excellent : 21 No\_Garage : 108   
## Max. :4.000 Fair : 56 CarPort : 15   
## More\_Than\_Two\_Types: 21   
## Garage\_Cars Garage\_Area Garage\_Qual Garage\_Cond   
## Min. :0.000 Min. : 0 Typical :1839 Typical :1872   
## 1st Qu.:1.000 1st Qu.: 320 No\_Garage: 109 No\_Garage: 109   
## Median :2.000 Median : 478 Fair : 85 Fair : 53   
## Mean :1.774 Mean : 472 Good : 16 Excellent: 1   
## 3rd Qu.:2.000 3rd Qu.: 576 Excellent: 2 Poor : 8   
## Max. :5.000 Max. :1488 Poor : 2 Good : 10   
##   
## Paved\_Drive Wood\_Deck\_SF Open\_Porch\_SF Enclosed\_Porch   
## Partial\_Pavement: 42 Min. : 0.00 Min. : 0.00 Min. : 0.00   
## Paved :1848 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00   
## Dirt\_Gravel : 163 Median : 0.00 Median : 27.00 Median : 0.00   
## Mean : 93.52 Mean : 48.17 Mean : 23.02   
## 3rd Qu.: 168.00 3rd Qu.: 72.00 3rd Qu.: 0.00   
## Max. :1424.00 Max. :742.00 Max. :584.00   
##   
## Three\_season\_porch Screen\_Porch Pool\_Area Pool\_QC   
## Min. : 0.000 Min. : 0.00 Min. : 0.000 No\_Pool :2047   
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.000 Excellent: 2   
## Median : 0.000 Median : 0.00 Median : 0.000 Typical : 2   
## Mean : 2.799 Mean : 16.68 Mean : 1.339 Fair : 1   
## 3rd Qu.: 0.000 3rd Qu.: 0.00 3rd Qu.: 0.000 Good : 1   
## Max. :407.000 Max. :576.00 Max. :800.000   
##   
## Fence Misc\_Feature Misc\_Val Mo\_Sold   
## No\_Fence :1661 None:1978 Min. : 0.00 Min. : 1.000   
## Minimum\_Privacy : 225 Gar2: 5 1st Qu.: 0.00 1st Qu.: 4.000   
## Good\_Privacy : 81 Shed: 66 Median : 0.00 Median : 6.000   
## Good\_Wood : 77 Othr: 3 Mean : 60.12 Mean : 6.189   
## Minimum\_Wood\_Wire: 9 Elev: 1 3rd Qu.: 0.00 3rd Qu.: 8.000   
## Max. :17000.00 Max. :12.000   
##   
## Year\_Sold Sale\_Type Sale\_Condition Longitude Latitude   
## Min. :2006 WD :1789 Normal :1712 Min. :-93.69 Min. :41.99   
## 1st Qu.:2007 New : 163 Partial: 169 1st Qu.:-93.66 1st Qu.:42.02   
## Median :2008 COD : 54 Family : 30 Median :-93.64 Median :42.03   
## Mean :2008 ConLD : 16 Abnorml: 121 Mean :-93.64 Mean :42.03   
## 3rd Qu.:2009 ConLI : 8 Alloca : 16 3rd Qu.:-93.62 3rd Qu.:42.05   
## Max. :2010 CWD : 8 AdjLand: 5 Max. :-93.58 Max. :42.06   
## (Other): 15   
## Above\_Median  
## Yes:1043   
## No :1010   
##   
##   
##   
##   
##

str(ames\_student)

## tibble [2,053 x 82] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ X1 : num [1:2053] 1 2 3 4 5 6 7 8 9 10 ...  
## $ MS\_SubClass : Factor w/ 16 levels "One\_Story\_1946\_and\_Newer\_All\_Styles",..: 1 1 1 1 2 2 3 3 1 1 ...  
## $ MS\_Zoning : Factor w/ 7 levels "Residential\_Low\_Density",..: 1 2 1 1 1 1 1 1 1 1 ...  
## $ Lot\_Frontage : num [1:2053] 141 80 81 93 74 78 43 39 0 85 ...  
## $ Lot\_Area : num [1:2053] 31770 11622 14267 11160 13830 ...  
## $ Street : Factor w/ 2 levels "Pave","Grvl": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Alley : Factor w/ 3 levels "No\_Alley\_Access",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Lot\_Shape : Factor w/ 4 levels "Slightly\_Irregular",..: 1 2 1 2 1 1 1 1 1 2 ...  
## $ Land\_Contour : Factor w/ 4 levels "Lvl","HLS","Bnk",..: 1 1 1 1 1 1 2 1 1 1 ...  
## $ Utilities : Factor w/ 2 levels "AllPub","NoSewr": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Lot\_Config : Factor w/ 5 levels "Corner","Inside",..: 1 2 1 1 2 2 2 2 2 2 ...  
## $ Land\_Slope : Factor w/ 3 levels "Gtl","Mod","Sev": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Neighborhood : Factor w/ 28 levels "North\_Ames","Gilbert",..: 1 1 1 1 2 2 3 3 2 2 ...  
## $ Condition\_1 : Factor w/ 9 levels "Norm","Feedr",..: 1 2 1 1 1 1 1 1 1 1 ...  
## $ Condition\_2 : Factor w/ 8 levels "Norm","Feedr",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Bldg\_Type : Factor w/ 5 levels "OneFam","TwnhsE",..: 1 1 1 1 1 1 2 2 1 1 ...  
## $ House\_Style : Factor w/ 8 levels "One\_Story","Two\_Story",..: 1 1 1 1 2 2 1 1 1 1 ...  
## $ Overall\_Qual : Factor w/ 10 levels "Above\_Average",..: 1 2 1 3 2 1 4 4 1 3 ...  
## $ Overall\_Cond : Factor w/ 9 levels "Average","Above\_Average",..: 1 2 2 1 1 2 1 1 3 1 ...  
## $ Year\_Built : num [1:2053] 1960 1961 1958 1968 1997 ...  
## $ Year\_Remod\_Add : num [1:2053] 1960 1961 1958 1968 1998 ...  
## $ Roof\_Style : Factor w/ 6 levels "Hip","Gable",..: 1 2 1 1 2 2 2 2 2 2 ...  
## $ Roof\_Matl : Factor w/ 6 levels "CompShg","WdShake",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Exterior\_1st : Factor w/ 16 levels "BrkFace","VinylSd",..: 1 2 3 1 2 2 4 5 4 4 ...  
## $ Exterior\_2nd : Factor w/ 17 levels "Plywood","VinylSd",..: 1 2 3 4 2 2 5 6 5 5 ...  
## $ Mas\_Vnr\_Type : Factor w/ 5 levels "Stone","None",..: 1 2 3 2 2 3 2 2 2 2 ...  
## $ Mas\_Vnr\_Area : num [1:2053] 112 0 108 0 0 20 0 0 0 0 ...  
## $ Exter\_Qual : Factor w/ 4 levels "Typical","Good",..: 1 1 1 2 1 1 2 2 1 1 ...  
## $ Exter\_Cond : Factor w/ 5 levels "Typical","Good",..: 1 1 1 1 1 1 1 1 2 1 ...  
## $ Foundation : Factor w/ 6 levels "CBlock","PConc",..: 1 1 1 1 2 2 2 2 2 2 ...  
## $ Bsmt\_Qual : Factor w/ 6 levels "Typical","Good",..: 1 1 1 1 2 1 2 2 2 2 ...  
## $ Bsmt\_Cond : Factor w/ 6 levels "Good","Typical",..: 1 2 2 2 2 2 2 2 2 2 ...  
## $ Bsmt\_Exposure : Factor w/ 5 levels "Gd","No","Av",..: 1 2 2 2 2 2 2 2 2 1 ...  
## $ BsmtFin\_Type\_1 : Factor w/ 7 levels "BLQ","Rec","ALQ",..: 1 2 3 3 4 4 3 4 3 4 ...  
## $ BsmtFin\_SF\_1 : num [1:2053] 2 6 1 1 3 3 1 3 1 3 ...  
## $ BsmtFin\_Type\_2 : Factor w/ 7 levels "Unf","LwQ","BLQ",..: 1 2 1 1 1 1 1 1 1 1 ...  
## $ BsmtFin\_SF\_2 : num [1:2053] 0 144 0 0 0 0 0 0 0 0 ...  
## $ Bsmt\_Unf\_SF : num [1:2053] 441 270 406 1045 137 ...  
## $ Total\_Bsmt\_SF : num [1:2053] 1080 882 1329 2110 928 ...  
## $ Heating : Factor w/ 6 levels "GasA","GasW",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Heating\_QC : Factor w/ 5 levels "Fair","Typical",..: 1 2 2 3 4 3 3 3 3 4 ...  
## $ Central\_Air : Factor w/ 2 levels "Y","N": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Electrical : Factor w/ 5 levels "SBrkr","FuseA",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ First\_Flr\_SF : num [1:2053] 1656 896 1329 2110 928 ...  
## $ Second\_Flr\_SF : num [1:2053] 0 0 0 0 701 678 0 0 0 0 ...  
## $ Low\_Qual\_Fin\_SF : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Gr\_Liv\_Area : num [1:2053] 1656 896 1329 2110 1629 ...  
## $ Bsmt\_Full\_Bath : num [1:2053] 1 0 0 1 0 0 0 1 1 1 ...  
## $ Bsmt\_Half\_Bath : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Full\_Bath : num [1:2053] 1 1 1 2 2 2 2 2 2 1 ...  
## $ Half\_Bath : num [1:2053] 0 0 1 1 1 1 0 0 0 1 ...  
## $ Bedroom\_AbvGr : num [1:2053] 3 2 3 3 3 3 2 2 3 2 ...  
## $ Kitchen\_AbvGr : num [1:2053] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Kitchen\_Qual : Factor w/ 5 levels "Typical","Good",..: 1 1 2 3 1 2 2 2 1 2 ...  
## $ TotRms\_AbvGrd : num [1:2053] 7 5 6 8 6 7 5 5 6 5 ...  
## $ Functional : Factor w/ 8 levels "Typ","Mod","Min2",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Fireplaces : num [1:2053] 2 0 0 2 1 1 0 1 0 1 ...  
## $ Fireplace\_Qu : Factor w/ 6 levels "Good","No\_Fireplace",..: 1 2 2 3 3 1 2 3 2 4 ...  
## $ Garage\_Type : Factor w/ 7 levels "Attchd","BuiltIn",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Garage\_Finish : Factor w/ 4 levels "Fin","Unf","RFn",..: 1 2 2 1 1 1 3 3 1 2 ...  
## $ Garage\_Cars : num [1:2053] 2 1 1 2 2 2 2 2 2 2 ...  
## $ Garage\_Area : num [1:2053] 528 730 312 522 482 470 506 608 420 506 ...  
## $ Garage\_Qual : Factor w/ 6 levels "Typical","No\_Garage",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Garage\_Cond : Factor w/ 6 levels "Typical","No\_Garage",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Paved\_Drive : Factor w/ 3 levels "Partial\_Pavement",..: 1 2 2 2 2 2 2 2 2 2 ...  
## $ Wood\_Deck\_SF : num [1:2053] 210 140 393 0 212 360 0 237 483 192 ...  
## $ Open\_Porch\_SF : num [1:2053] 62 0 36 0 34 36 82 152 21 0 ...  
## $ Enclosed\_Porch : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Three\_season\_porch: num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Screen\_Porch : num [1:2053] 0 120 0 0 0 0 144 0 0 0 ...  
## $ Pool\_Area : num [1:2053] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_QC : Factor w/ 5 levels "No\_Pool","Excellent",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Fence : Factor w/ 5 levels "No\_Fence","Minimum\_Privacy",..: 1 2 1 1 2 1 1 1 3 1 ...  
## $ Misc\_Feature : Factor w/ 5 levels "None","Gar2",..: 1 1 2 1 1 1 1 1 3 1 ...  
## $ Misc\_Val : num [1:2053] 0 0 12500 0 0 0 0 0 500 0 ...  
## $ Mo\_Sold : num [1:2053] 5 6 6 4 3 6 1 3 3 2 ...  
## $ Year\_Sold : num [1:2053] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ Sale\_Type : Factor w/ 10 levels "WD","New","COD",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Sale\_Condition : Factor w/ 6 levels "Normal","Partial",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Longitude : num [1:2053] -93.6 -93.6 -93.6 -93.6 -93.6 ...  
## $ Latitude : num [1:2053] 42.1 42.1 42.1 42.1 42.1 ...  
## $ Above\_Median : Factor w/ 2 levels "Yes","No": 1 2 1 1 1 1 1 1 1 1 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. X1 = col\_double(),  
## .. MS\_SubClass = col\_character(),  
## .. MS\_Zoning = col\_character(),  
## .. Lot\_Frontage = col\_double(),  
## .. Lot\_Area = col\_double(),  
## .. Street = col\_character(),  
## .. Alley = col\_character(),  
## .. Lot\_Shape = col\_character(),  
## .. Land\_Contour = col\_character(),  
## .. Utilities = col\_character(),  
## .. Lot\_Config = col\_character(),  
## .. Land\_Slope = col\_character(),  
## .. Neighborhood = col\_character(),  
## .. Condition\_1 = col\_character(),  
## .. Condition\_2 = col\_character(),  
## .. Bldg\_Type = col\_character(),  
## .. House\_Style = col\_character(),  
## .. Overall\_Qual = col\_character(),  
## .. Overall\_Cond = col\_character(),  
## .. Year\_Built = col\_double(),  
## .. Year\_Remod\_Add = col\_double(),  
## .. Roof\_Style = col\_character(),  
## .. Roof\_Matl = col\_character(),  
## .. Exterior\_1st = col\_character(),  
## .. Exterior\_2nd = col\_character(),  
## .. Mas\_Vnr\_Type = col\_character(),  
## .. Mas\_Vnr\_Area = col\_double(),  
## .. Exter\_Qual = col\_character(),  
## .. Exter\_Cond = col\_character(),  
## .. Foundation = col\_character(),  
## .. Bsmt\_Qual = col\_character(),  
## .. Bsmt\_Cond = col\_character(),  
## .. Bsmt\_Exposure = col\_character(),  
## .. BsmtFin\_Type\_1 = col\_character(),  
## .. BsmtFin\_SF\_1 = col\_double(),  
## .. BsmtFin\_Type\_2 = col\_character(),  
## .. BsmtFin\_SF\_2 = col\_double(),  
## .. Bsmt\_Unf\_SF = col\_double(),  
## .. Total\_Bsmt\_SF = col\_double(),  
## .. Heating = col\_character(),  
## .. Heating\_QC = col\_character(),  
## .. Central\_Air = col\_character(),  
## .. Electrical = col\_character(),  
## .. First\_Flr\_SF = col\_double(),  
## .. Second\_Flr\_SF = col\_double(),  
## .. Low\_Qual\_Fin\_SF = col\_double(),  
## .. Gr\_Liv\_Area = col\_double(),  
## .. Bsmt\_Full\_Bath = col\_double(),  
## .. Bsmt\_Half\_Bath = col\_double(),  
## .. Full\_Bath = col\_double(),  
## .. Half\_Bath = col\_double(),  
## .. Bedroom\_AbvGr = col\_double(),  
## .. Kitchen\_AbvGr = col\_double(),  
## .. Kitchen\_Qual = col\_character(),  
## .. TotRms\_AbvGrd = col\_double(),  
## .. Functional = col\_character(),  
## .. Fireplaces = col\_double(),  
## .. Fireplace\_Qu = col\_character(),  
## .. Garage\_Type = col\_character(),  
## .. Garage\_Finish = col\_character(),  
## .. Garage\_Cars = col\_double(),  
## .. Garage\_Area = col\_double(),  
## .. Garage\_Qual = col\_character(),  
## .. Garage\_Cond = col\_character(),  
## .. Paved\_Drive = col\_character(),  
## .. Wood\_Deck\_SF = col\_double(),  
## .. Open\_Porch\_SF = col\_double(),  
## .. Enclosed\_Porch = col\_double(),  
## .. Three\_season\_porch = col\_double(),  
## .. Screen\_Porch = col\_double(),  
## .. Pool\_Area = col\_double(),  
## .. Pool\_QC = col\_character(),  
## .. Fence = col\_character(),  
## .. Misc\_Feature = col\_character(),  
## .. Misc\_Val = col\_double(),  
## .. Mo\_Sold = col\_double(),  
## .. Year\_Sold = col\_double(),  
## .. Sale\_Type = col\_character(),  
## .. Sale\_Condition = col\_character(),  
## .. Longitude = col\_double(),  
## .. Latitude = col\_double(),  
## .. Above\_Median = col\_character()  
## .. )

ames <- ames\_student %>% select(-X1, -Longitude, -Latitude, -Electrical, -Utilities, -Street, -Garage\_Qual, -Garage\_Cond, -Paved\_Drive, -Functional, -Heating, -Bsmt\_Cond, -Roof\_Matl, -Condition\_1, -Condition\_2, -Land\_Slope, -Land\_Contour, -Alley, -BsmtFin\_SF\_2, -Misc\_Val, -Screen\_Porch, -Low\_Qual\_Fin\_SF, -Kitchen\_AbvGr, -Three\_season\_porch, -Pool\_Area, -Pool\_QC, -Enclosed\_Porch, -Open\_Porch\_SF, -Wood\_Deck\_SF, -Mas\_Vnr\_Type, -Misc\_Feature)  
  
ames <- select(ames, MS\_Zoning, everything()) %>%  
 filter(MS\_Zoning %in% c("Residential\_Low\_Density", "Residential\_High\_Density", "Floating\_Village\_Residential", "Residential\_Medium\_Density"))  
  
summary(ames)

## MS\_Zoning MS\_SubClass   
## Residential\_Low\_Density :1600 One\_Story\_1946\_and\_Newer\_All\_Styles :769   
## Residential\_High\_Density : 20 Two\_Story\_1946\_and\_Newer :383   
## Floating\_Village\_Residential: 87 One\_and\_Half\_Story\_Finished\_All\_Ages:199   
## Residential\_Medium\_Density : 326 One\_Story\_PUD\_1946\_and\_Newer :129   
## C\_all : 0 One\_Story\_1945\_and\_Older : 92   
## A\_agr : 0 Two\_Story\_1945\_and\_Older : 92   
## I\_all : 0 (Other) :369   
## Lot\_Frontage Lot\_Area Lot\_Shape Lot\_Config   
## Min. : 0.00 Min. : 1300 Slightly\_Irregular : 711 Corner : 354   
## 1st Qu.: 43.00 1st Qu.: 7500 Regular :1258 Inside :1480   
## Median : 62.00 Median : 9555 Moderately\_Irregular: 53 CulDSac: 135   
## Mean : 57.34 Mean : 10236 Irregular : 11 FR2 : 56   
## 3rd Qu.: 79.00 3rd Qu.: 11600 FR3 : 8   
## Max. :313.00 Max. :215245   
##   
## Neighborhood Bldg\_Type House\_Style   
## North\_Ames :327 OneFam :1689 One\_Story :1043   
## College\_Creek:183 TwnhsE : 157 Two\_Story : 587   
## Old\_Town :179 Twnhs : 67 One\_and\_Half\_Fin: 220   
## Edwards :129 Duplex : 76 SLvl : 90   
## Somerset :119 TwoFmCon: 44 SFoyer : 56   
## Gilbert :109 Two\_and\_Half\_Unf: 17   
## (Other) :987 (Other) : 20   
## Overall\_Qual Overall\_Cond Year\_Built Year\_Remod\_Add  
## Average :584 Average :1140 Min. :1875 Min. :1950   
## Above\_Average:517 Above\_Average: 370 1st Qu.:1954 1st Qu.:1966   
## Good :411 Good : 284 Median :1972 Median :1993   
## Very\_Good :237 Very\_Good : 98 Mean :1971 Mean :1984   
## Below\_Average:161 Below\_Average: 72 3rd Qu.:2000 3rd Qu.:2004   
## Excellent : 70 Fair : 30 Max. :2010 Max. :2010   
## (Other) : 53 (Other) : 39   
## Roof\_Style Exterior\_1st Exterior\_2nd Mas\_Vnr\_Area Exter\_Qual   
## Hip : 401 VinylSd:704 VinylSd:696 Min. : 0.0 Typical :1257   
## Gable :1592 MetalSd:313 MetalSd:311 1st Qu.: 0.0 Good : 682   
## Mansard: 8 Wd Sdng:308 Wd Sdng:298 Median : 0.0 Excellent: 78   
## Gambrel: 13 HdBoard:303 HdBoard:277 Mean : 104.8 Fair : 16   
## Shed : 5 Plywood:149 Plywood:188 3rd Qu.: 165.0   
## Flat : 14 CemntBd: 90 CmentBd: 90 Max. :1600.0   
## (Other):166 (Other):173   
## Exter\_Cond Foundation Bsmt\_Qual Bsmt\_Exposure   
## Typical :1774 CBlock:870 Typical :897 Gd : 199   
## Good : 212 PConc :910 Good :849 No :1319   
## Fair : 38 Wood : 4 Excellent :178 Av : 283   
## Excellent: 9 BrkTil:211 No\_Basement: 54 Mn : 175   
## Poor : 0 Slab : 34 Fair : 55 No\_Basement: 57   
## Stone : 4 Poor : 0   
##   
## BsmtFin\_Type\_1 BsmtFin\_SF\_1 BsmtFin\_Type\_2 Bsmt\_Unf\_SF   
## BLQ :195 Min. :1.00 Unf :1723 Min. : 0.0   
## Rec :212 1st Qu.:3.00 LwQ : 64 1st Qu.: 226.0   
## ALQ :298 Median :3.00 BLQ : 47 Median : 460.0   
## GLQ :578 Mean :4.19 Rec : 79 Mean : 561.8   
## Unf :590 3rd Qu.:7.00 GLQ : 23 3rd Qu.: 801.0   
## LwQ :106 Max. :7.00 No\_Basement: 55 Max. :2336.0   
## No\_Basement: 54 ALQ : 42   
## Total\_Bsmt\_SF Heating\_QC Central\_Air First\_Flr\_SF Second\_Flr\_SF   
## Min. : 0 Fair : 58 Y:1909 Min. : 442 Min. : 0.0   
## 1st Qu.: 796 Typical : 608 N: 124 1st Qu.: 884 1st Qu.: 0.0   
## Median : 989 Excellent:1036 Median :1091 Median : 0.0   
## Mean :1059 Good : 331 Mean :1170 Mean : 326.6   
## 3rd Qu.:1309 Poor : 0 3rd Qu.:1405 3rd Qu.: 702.0   
## Max. :5095 Max. :5095 Max. :1862.0   
##   
## Gr\_Liv\_Area Bsmt\_Full\_Bath Bsmt\_Half\_Bath Full\_Bath   
## Min. : 498 Min. :0.0000 Min. :0.00000 Min. :0.000   
## 1st Qu.:1141 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:1.000   
## Median :1452 Median :0.0000 Median :0.00000 Median :2.000   
## Mean :1502 Mean :0.4338 Mean :0.05804 Mean :1.568   
## 3rd Qu.:1738 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:2.000   
## Max. :5095 Max. :3.0000 Max. :2.00000 Max. :4.000   
##   
## Half\_Bath Bedroom\_AbvGr Kitchen\_Qual TotRms\_AbvGrd   
## Min. :0.0000 Min. :0.000 Typical :1054 Min. : 3.000   
## 1st Qu.:0.0000 1st Qu.:2.000 Good : 790 1st Qu.: 5.000   
## Median :0.0000 Median :3.000 Excellent: 142 Median : 6.000   
## Mean :0.3768 Mean :2.858 Fair : 46 Mean : 6.447   
## 3rd Qu.:1.0000 3rd Qu.:3.000 Poor : 1 3rd Qu.: 7.000   
## Max. :2.0000 Max. :6.000 Max. :15.000   
##   
## Fireplaces Fireplace\_Qu Garage\_Type Garage\_Finish  
## Min. :0.000 Good :536 Attchd :1202 Fin :509   
## 1st Qu.:0.000 No\_Fireplace:975 BuiltIn : 127 Unf :863   
## Median :1.000 Typical :409 Basment : 29 RFn :561   
## Mean :0.608 Poor : 36 Detchd : 541 No\_Garage:100   
## 3rd Qu.:1.000 Excellent : 21 No\_Garage : 99   
## Max. :4.000 Fair : 56 CarPort : 14   
## More\_Than\_Two\_Types: 21   
## Garage\_Cars Garage\_Area Fence Mo\_Sold   
## Min. :0.000 Min. : 0.0 No\_Fence :1643 Min. : 1.000   
## 1st Qu.:1.000 1st Qu.: 330.0 Minimum\_Privacy : 224 1st Qu.: 4.000   
## Median :2.000 Median : 480.0 Good\_Privacy : 81 Median : 6.000   
## Mean :1.784 Mean : 474.3 Good\_Wood : 76 Mean : 6.182   
## 3rd Qu.:2.000 3rd Qu.: 576.0 Minimum\_Wood\_Wire: 9 3rd Qu.: 8.000   
## Max. :5.000 Max. :1488.0 Max. :12.000   
##   
## Year\_Sold Sale\_Type Sale\_Condition Above\_Median  
## Min. :2006 WD :1773 Normal :1701 Yes:1043   
## 1st Qu.:2007 New : 163 Partial: 169 No : 990   
## Median :2008 COD : 53 Family : 30   
## Mean :2008 ConLD : 13 Abnorml: 112   
## 3rd Qu.:2009 ConLI : 8 Alloca : 16   
## Max. :2010 CWD : 8 AdjLand: 5   
## (Other): 15

str(ames)

## tibble [2,033 x 51] (S3: tbl\_df/tbl/data.frame)  
## $ MS\_Zoning : Factor w/ 7 levels "Residential\_Low\_Density",..: 1 2 1 1 1 1 1 1 1 1 ...  
## $ MS\_SubClass : Factor w/ 16 levels "One\_Story\_1946\_and\_Newer\_All\_Styles",..: 1 1 1 1 2 2 3 3 1 1 ...  
## $ Lot\_Frontage : num [1:2033] 141 80 81 93 74 78 43 39 0 85 ...  
## $ Lot\_Area : num [1:2033] 31770 11622 14267 11160 13830 ...  
## $ Lot\_Shape : Factor w/ 4 levels "Slightly\_Irregular",..: 1 2 1 2 1 1 1 1 1 2 ...  
## $ Lot\_Config : Factor w/ 5 levels "Corner","Inside",..: 1 2 1 1 2 2 2 2 2 2 ...  
## $ Neighborhood : Factor w/ 28 levels "North\_Ames","Gilbert",..: 1 1 1 1 2 2 3 3 2 2 ...  
## $ Bldg\_Type : Factor w/ 5 levels "OneFam","TwnhsE",..: 1 1 1 1 1 1 2 2 1 1 ...  
## $ House\_Style : Factor w/ 8 levels "One\_Story","Two\_Story",..: 1 1 1 1 2 2 1 1 1 1 ...  
## $ Overall\_Qual : Factor w/ 10 levels "Above\_Average",..: 1 2 1 3 2 1 4 4 1 3 ...  
## $ Overall\_Cond : Factor w/ 9 levels "Average","Above\_Average",..: 1 2 2 1 1 2 1 1 3 1 ...  
## $ Year\_Built : num [1:2033] 1960 1961 1958 1968 1997 ...  
## $ Year\_Remod\_Add: num [1:2033] 1960 1961 1958 1968 1998 ...  
## $ Roof\_Style : Factor w/ 6 levels "Hip","Gable",..: 1 2 1 1 2 2 2 2 2 2 ...  
## $ Exterior\_1st : Factor w/ 16 levels "BrkFace","VinylSd",..: 1 2 3 1 2 2 4 5 4 4 ...  
## $ Exterior\_2nd : Factor w/ 17 levels "Plywood","VinylSd",..: 1 2 3 4 2 2 5 6 5 5 ...  
## $ Mas\_Vnr\_Area : num [1:2033] 112 0 108 0 0 20 0 0 0 0 ...  
## $ Exter\_Qual : Factor w/ 4 levels "Typical","Good",..: 1 1 1 2 1 1 2 2 1 1 ...  
## $ Exter\_Cond : Factor w/ 5 levels "Typical","Good",..: 1 1 1 1 1 1 1 1 2 1 ...  
## $ Foundation : Factor w/ 6 levels "CBlock","PConc",..: 1 1 1 1 2 2 2 2 2 2 ...  
## $ Bsmt\_Qual : Factor w/ 6 levels "Typical","Good",..: 1 1 1 1 2 1 2 2 2 2 ...  
## $ Bsmt\_Exposure : Factor w/ 5 levels "Gd","No","Av",..: 1 2 2 2 2 2 2 2 2 1 ...  
## $ BsmtFin\_Type\_1: Factor w/ 7 levels "BLQ","Rec","ALQ",..: 1 2 3 3 4 4 3 4 3 4 ...  
## $ BsmtFin\_SF\_1 : num [1:2033] 2 6 1 1 3 3 1 3 1 3 ...  
## $ BsmtFin\_Type\_2: Factor w/ 7 levels "Unf","LwQ","BLQ",..: 1 2 1 1 1 1 1 1 1 1 ...  
## $ Bsmt\_Unf\_SF : num [1:2033] 441 270 406 1045 137 ...  
## $ Total\_Bsmt\_SF : num [1:2033] 1080 882 1329 2110 928 ...  
## $ Heating\_QC : Factor w/ 5 levels "Fair","Typical",..: 1 2 2 3 4 3 3 3 3 4 ...  
## $ Central\_Air : Factor w/ 2 levels "Y","N": 1 1 1 1 1 1 1 1 1 1 ...  
## $ First\_Flr\_SF : num [1:2033] 1656 896 1329 2110 928 ...  
## $ Second\_Flr\_SF : num [1:2033] 0 0 0 0 701 678 0 0 0 0 ...  
## $ Gr\_Liv\_Area : num [1:2033] 1656 896 1329 2110 1629 ...  
## $ Bsmt\_Full\_Bath: num [1:2033] 1 0 0 1 0 0 0 1 1 1 ...  
## $ Bsmt\_Half\_Bath: num [1:2033] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Full\_Bath : num [1:2033] 1 1 1 2 2 2 2 2 2 1 ...  
## $ Half\_Bath : num [1:2033] 0 0 1 1 1 1 0 0 0 1 ...  
## $ Bedroom\_AbvGr : num [1:2033] 3 2 3 3 3 3 2 2 3 2 ...  
## $ Kitchen\_Qual : Factor w/ 5 levels "Typical","Good",..: 1 1 2 3 1 2 2 2 1 2 ...  
## $ TotRms\_AbvGrd : num [1:2033] 7 5 6 8 6 7 5 5 6 5 ...  
## $ Fireplaces : num [1:2033] 2 0 0 2 1 1 0 1 0 1 ...  
## $ Fireplace\_Qu : Factor w/ 6 levels "Good","No\_Fireplace",..: 1 2 2 3 3 1 2 3 2 4 ...  
## $ Garage\_Type : Factor w/ 7 levels "Attchd","BuiltIn",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Garage\_Finish : Factor w/ 4 levels "Fin","Unf","RFn",..: 1 2 2 1 1 1 3 3 1 2 ...  
## $ Garage\_Cars : num [1:2033] 2 1 1 2 2 2 2 2 2 2 ...  
## $ Garage\_Area : num [1:2033] 528 730 312 522 482 470 506 608 420 506 ...  
## $ Fence : Factor w/ 5 levels "No\_Fence","Minimum\_Privacy",..: 1 2 1 1 2 1 1 1 3 1 ...  
## $ Mo\_Sold : num [1:2033] 5 6 6 4 3 6 1 3 3 2 ...  
## $ Year\_Sold : num [1:2033] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ Sale\_Type : Factor w/ 10 levels "WD","New","COD",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Sale\_Condition: Factor w/ 6 levels "Normal","Partial",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Above\_Median : Factor w/ 2 levels "Yes","No": 1 2 1 1 1 1 1 1 1 1 ...

Decision Trees

set.seed(1234)  
ames\_split <- initial\_split(ames, prob=0.70, strata=Above\_Median)  
ames\_train <- training(ames\_split)  
ames\_test <- testing(ames\_split)

ames\_recipe <- recipe(Above\_Median ~., ames\_train) %>%  
 step\_other(all\_nominal(), threshold=0.1) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes()) %>%  
 step\_corr(all\_predictors()) %>%  
 step\_center(all\_predictors()) %>%  
 step\_scale(all\_predictors())  
  
ames\_model <-  
 decision\_tree() %>%  
 set\_engine("rpart", model=TRUE) %>%  
 set\_mode("classification")  
  
ames\_wf <-  
 workflow() %>%  
 add\_recipe(ames\_recipe) %>%  
 add\_model(ames\_model)

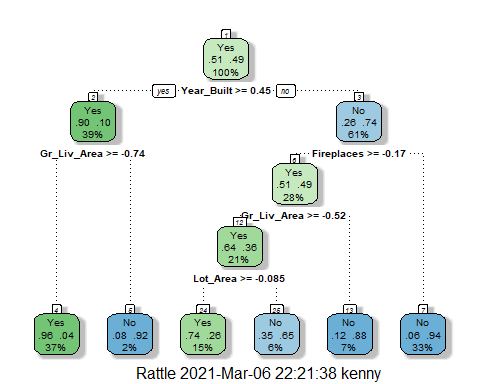
ames\_fit <-  
 fit(ames\_wf, ames\_train)

ames\_fit %>%  
 pull\_workflow\_fit() %>%  
 pluck("fit")

## n= 1526   
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 1526 743 Yes (0.51310616 0.48689384)   
## 2) Year\_Built>=0.445364 595 58 Yes (0.90252101 0.09747899)   
## 4) Gr\_Liv\_Area>=-0.7412937 558 24 Yes (0.95698925 0.04301075) \*  
## 5) Gr\_Liv\_Area< -0.7412937 37 3 No (0.08108108 0.91891892) \*  
## 3) Year\_Built< 0.445364 931 246 No (0.26423201 0.73576799)   
## 6) Fireplaces>=-0.1731311 424 209 Yes (0.50707547 0.49292453)   
## 12) Gr\_Liv\_Area>=-0.5236626 318 116 Yes (0.63522013 0.36477987)   
## 24) Lot\_Area>=-0.08464355 234 61 Yes (0.73931624 0.26068376) \*  
## 25) Lot\_Area< -0.08464355 84 29 No (0.34523810 0.65476190) \*  
## 13) Gr\_Liv\_Area< -0.5236626 106 13 No (0.12264151 0.87735849) \*  
## 7) Fireplaces< -0.1731311 507 31 No (0.06114398 0.93885602) \*

ames\_tree <-  
 ames\_fit %>%  
 pull\_workflow\_fit() %>%  
 pluck("fit")

fancyRpartPlot(ames\_tree)



ames\_fit$fit$fit$fit$cptable

## CP nsplit rel error xerror xstd  
## 1 0.59084791 0 1.0000000 1.0000000 0.02627902  
## 2 0.05787349 1 0.4091521 0.4091521 0.02099937  
## 3 0.04172275 3 0.2934051 0.3203230 0.01907569  
## 4 0.03499327 4 0.2516824 0.2893674 0.01829172  
## 5 0.01000000 5 0.2166891 0.2543742 0.01731931

amespred <- predict(ames\_fit, ames\_train, type="class")  
head(amespred)

## # A tibble: 6 x 1  
## .pred\_class  
## <fct>   
## 1 Yes   
## 2 No   
## 3 No   
## 4 Yes   
## 5 Yes   
## 6 Yes

confusionMatrix(amespred$.pred\_class, ames\_train$Above\_Median, positive="Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Yes No  
## Yes 707 85  
## No 76 658  
##   
## Accuracy : 0.8945   
## 95% CI : (0.878, 0.9095)  
## No Information Rate : 0.5131   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.7888   
##   
## Mcnemar's Test P-Value : 0.5284   
##   
## Sensitivity : 0.9029   
## Specificity : 0.8856   
## Pos Pred Value : 0.8927   
## Neg Pred Value : 0.8965   
## Prevalence : 0.5131   
## Detection Rate : 0.4633   
## Detection Prevalence : 0.5190   
## Balanced Accuracy : 0.8943   
##   
## 'Positive' Class : Yes   
##

set.seed(234)  
folds <- vfold\_cv(ames\_train, v=5)

ames2\_recipe <- recipe(Above\_Median ~., ames\_train) %>%  
 step\_other(all\_nominal(), threshold=0.1) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes()) %>%  
 step\_center(all\_predictors()) %>%  
 step\_scale(all\_predictors())  
  
ames2\_model <-  
 decision\_tree(cost\_complexity=tune()) %>%  
 set\_engine("rpart", model=TRUE) %>%  
 set\_mode("classification")  
  
ames2\_grid <- grid\_regular(cost\_complexity(),  
 levels=20)  
  
ames2\_wf <-  
 workflow() %>%  
 add\_recipe(ames2\_recipe) %>%  
 add\_model(ames2\_model)  
  
ames2\_res <-  
 ames2\_wf %>%  
 tune\_grid(  
 resamples=folds,  
 grid=ames2\_grid  
 )

##   
## Attaching package: 'rlang'

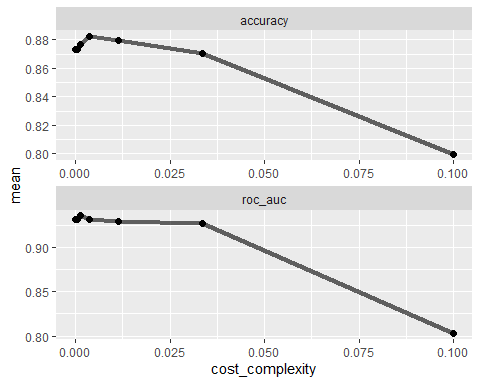
## The following objects are masked from 'package:purrr':  
##   
## %@%, as\_function, flatten, flatten\_chr, flatten\_dbl, flatten\_int,  
## flatten\_lgl, flatten\_raw, invoke, list\_along, modify, prepend,  
## splice

##   
## Attaching package: 'vctrs'

## The following object is masked from 'package:dplyr':  
##   
## data\_frame

## The following object is masked from 'package:tibble':  
##   
## data\_frame

ames2\_res %>%  
 collect\_metrics() %>%  
 ggplot(aes(cost\_complexity, mean)) +  
 geom\_line(size=1.5, alpha=0.6) +  
 geom\_point(size=2) +  
 facet\_wrap(~ .metric, scales="free", nrow=2)

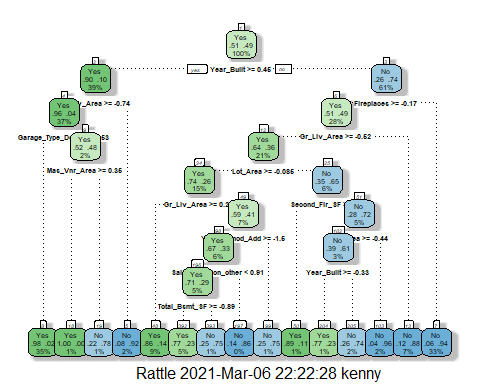


best\_ames\_tree <- ames2\_res %>%  
 select\_best("accuracy")

final\_ames2\_wf <-   
 ames2\_wf %>%  
 finalize\_workflow(best\_ames\_tree)

final\_ames2\_fit <-fit(final\_ames2\_wf, ames\_train)  
  
ames2\_tree <- final\_ames2\_fit %>%  
 pull\_workflow\_fit() %>%  
 pluck("fit")

fancyRpartPlot(ames2\_tree, tweak=1.5)



final\_ames2\_fit$fit$fit$fit$cptable

## CP nsplit rel error xerror xstd  
## 1 0.590847914 0 1.0000000 1.0000000 0.02627902  
## 2 0.057873486 1 0.4091521 0.4091521 0.02099937  
## 3 0.041722746 3 0.2934051 0.3216689 0.01910830  
## 4 0.034993271 4 0.2516824 0.2987887 0.01853742  
## 5 0.009421265 5 0.2166891 0.2557201 0.01735857  
## 6 0.006729475 6 0.2072678 0.2341857 0.01671079  
## 7 0.005383580 11 0.1736205 0.2409152 0.01691781  
## 8 0.004710633 12 0.1682369 0.2328398 0.01666886  
## 9 0.003792690 14 0.1588156 0.2355316 0.01675254

ames2pred <- predict(final\_ames2\_fit, ames\_train, type="class")  
head(amespred)

## # A tibble: 6 x 1  
## .pred\_class  
## <fct>   
## 1 Yes   
## 2 No   
## 3 No   
## 4 Yes   
## 5 Yes   
## 6 Yes

confusionMatrix(ames2pred$.pred\_class, ames\_train$Above\_Median, positive="Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Yes No  
## Yes 713 48  
## No 70 695  
##   
## Accuracy : 0.9227   
## 95% CI : (0.9081, 0.9356)  
## No Information Rate : 0.5131   
## P-Value [Acc > NIR] : < 2e-16   
##   
## Kappa : 0.8454   
##   
## Mcnemar's Test P-Value : 0.05321   
##   
## Sensitivity : 0.9106   
## Specificity : 0.9354   
## Pos Pred Value : 0.9369   
## Neg Pred Value : 0.9085   
## Prevalence : 0.5131   
## Detection Rate : 0.4672   
## Detection Prevalence : 0.4987   
## Balanced Accuracy : 0.9230   
##   
## 'Positive' Class : Yes   
##

Random Forest

ames3\_recipe <- recipe(Above\_Median ~., ames\_train) %>%  
 step\_other(all\_nominal(), threshold=0.1) %>%  
 step\_nzv(all\_predictors()) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes()) %>%  
 step\_center(all\_predictors()) %>%  
 step\_scale(all\_predictors())  
  
ames3\_model <-  
 rand\_forest() %>%  
 set\_engine("ranger") %>%  
 set\_mode("classification")  
  
ames3\_wf <-  
 workflow() %>%  
 add\_recipe(ames3\_recipe) %>%  
 add\_model(ames3\_model)  
  
set.seed(123)  
ames3\_fit <- fit(ames3\_wf, ames\_train)

ames3pred <- predict(ames3\_fit, ames\_train)  
head(ames3pred)

## # A tibble: 6 x 1  
## .pred\_class  
## <fct>   
## 1 Yes   
## 2 No   
## 3 Yes   
## 4 Yes   
## 5 Yes   
## 6 Yes

confusionMatrix(ames3pred$.pred\_class, ames\_train$Above\_Median, positive="Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Yes No  
## Yes 776 10  
## No 7 733  
##   
## Accuracy : 0.9889   
## 95% CI : (0.9822, 0.9935)  
## No Information Rate : 0.5131   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.9777   
##   
## Mcnemar's Test P-Value : 0.6276   
##   
## Sensitivity : 0.9911   
## Specificity : 0.9865   
## Pos Pred Value : 0.9873   
## Neg Pred Value : 0.9905   
## Prevalence : 0.5131   
## Detection Rate : 0.5085   
## Detection Prevalence : 0.5151   
## Balanced Accuracy : 0.9888   
##   
## 'Positive' Class : Yes   
##

XGBoost

set.seed(123)  
folds2 <- vfold\_cv(ames\_train, v=5)

xgboost\_recipe <-   
 recipe(formula = Above\_Median ~ ., data=ames\_train) %>%   
 step\_dummy(all\_nominal(), -all\_outcomes(), one\_hot = TRUE) %>%   
 step\_zv(all\_predictors())   
  
xgboost\_spec <-   
 boost\_tree(trees=tune(), min\_n=tune(), tree\_depth=tune(), learn\_rate=tune(),   
 loss\_reduction=tune(), sample\_size=tune()) %>%   
 set\_mode("classification") %>%   
 set\_engine("xgboost")   
  
xgboost\_workflow <-   
 workflow() %>%   
 add\_recipe(xgboost\_recipe) %>%   
 add\_model(xgboost\_spec)   
  
set.seed(2986)  
xgboost\_tune <-  
 tune\_grid(xgboost\_workflow, resamples=folds2, grid=25)

best\_xgb <- select\_best(xgboost\_tune, "accuracy")

ames\_xgb <-   
 finalize\_workflow(  
 xgboost\_workflow,  
 best\_xgb  
 )

xgb\_fit <- fit(ames\_xgb, ames\_train)

## [22:36:54] WARNING: amalgamation/../src/learner.cc:1061: Starting in XGBoost 1.3.0, the default evaluation metric used with the objective 'binary:logistic' was changed from 'error' to 'logloss'. Explicitly set eval\_metric if you'd like to restore the old behavior.

xgbpred <- predict(xgb\_fit, ames\_train)  
head(xgbpred)

## # A tibble: 6 x 1  
## .pred\_class  
## <fct>   
## 1 Yes   
## 2 No   
## 3 No   
## 4 Yes   
## 5 Yes   
## 6 Yes

confusionMatrix(xgbpred$.pred\_class, ames\_train$Above\_Median, positive="Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Yes No  
## Yes 744 38  
## No 39 705  
##   
## Accuracy : 0.9495   
## 95% CI : (0.9373, 0.96)  
## No Information Rate : 0.5131   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.899   
##   
## Mcnemar's Test P-Value : 1   
##   
## Sensitivity : 0.9502   
## Specificity : 0.9489   
## Pos Pred Value : 0.9514   
## Neg Pred Value : 0.9476   
## Prevalence : 0.5131   
## Detection Rate : 0.4875   
## Detection Prevalence : 0.5125   
## Balanced Accuracy : 0.9495   
##   
## 'Positive' Class : Yes   
##

Neural Network

set.seed(123)  
folds3 <- vfold\_cv(ames\_train, v=5)

ames4\_recipe <- recipe(Above\_Median ~., ames\_train) %>%  
 step\_normalize(all\_predictors(), -all\_nominal()) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes())  
  
ames4\_model <-  
 mlp(hidden\_units=tune(), penalty=tune(),  
 epochs=tune()) %>%  
 set\_mode("classification") %>%  
 set\_engine("nnet", verbose=0)  
  
ames4\_wf <-   
 workflow() %>%  
 add\_recipe(ames4\_recipe) %>%  
 add\_model(ames4\_model)  
  
set.seed(1234)  
ames\_neural\_tune <-   
 tune\_grid(ames4\_wf, resamples=folds3, grid=25)

## x Fold1: preprocessor 1/1, model 1/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold1: preprocessor 1/1, model 2/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold1: preprocessor 1/1, model 5/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold1: preprocessor 1/1, model 8/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold1: preprocessor 1/1, model 10/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 11/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 12/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 13/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 18/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 19/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 20/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 21/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 22/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 23/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold1: preprocessor 1/1, model 24/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 1/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold2: preprocessor 1/1, model 2/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold2: preprocessor 1/1, model 5/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold2: preprocessor 1/1, model 8/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold2: preprocessor 1/1, model 10/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 11/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 12/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 13/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 18/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 19/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 20/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 21/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 22/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 23/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold2: preprocessor 1/1, model 24/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 1/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold3: preprocessor 1/1, model 2/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold3: preprocessor 1/1, model 5/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold3: preprocessor 1/1, model 8/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold3: preprocessor 1/1, model 10/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 11/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 12/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 13/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 18/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 19/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 20/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 21/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 22/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 23/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold3: preprocessor 1/1, model 24/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 1/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold4: preprocessor 1/1, model 2/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold4: preprocessor 1/1, model 5/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold4: preprocessor 1/1, model 8/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold4: preprocessor 1/1, model 10/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 11/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 12/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 13/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 18/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 19/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 20/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 21/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 22/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 23/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold4: preprocessor 1/1, model 24/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 1/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold5: preprocessor 1/1, model 2/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold5: preprocessor 1/1, model 5/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold5: preprocessor 1/1, model 8/25: Error in nnet.default(x, y, w, entropy = T...

## x Fold5: preprocessor 1/1, model 10/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 11/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 12/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 13/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 18/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 19/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 20/25: Error in nnet.default(x, y, w, entropy = ...

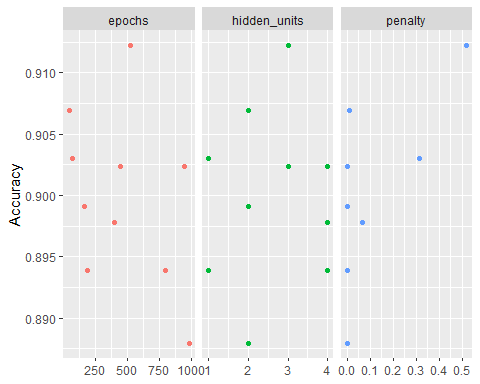
## x Fold5: preprocessor 1/1, model 21/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 22/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 23/25: Error in nnet.default(x, y, w, entropy = ...

## x Fold5: preprocessor 1/1, model 24/25: Error in nnet.default(x, y, w, entropy = ...

ames\_neural\_tune %>%  
 collect\_metrics() %>%  
 filter(.metric=="accuracy") %>%  
 select(mean, hidden\_units, penalty, epochs) %>%  
 pivot\_longer(hidden\_units:epochs,  
 values\_to="value",  
 names\_to="parameter") %>%  
 ggplot(aes(value, mean, color=parameter)) +  
 geom\_point(show.legend=FALSE) +  
 facet\_wrap(~parameter, scales="free\_x") +  
 labs(x=NULL, y="Accuracy")



best\_ames\_nn <- select\_best(ames\_neural\_tune, "accuracy")

final\_ames\_nn <- finalize\_workflow(  
 ames4\_wf,  
 best\_ames\_nn)

ames4\_fit <- fit(final\_ames\_nn, ames\_train)

ames4pred <- predict(ames4\_fit, ames\_train)  
head(ames4pred)

## # A tibble: 6 x 1  
## .pred\_class  
## <fct>   
## 1 Yes   
## 2 No   
## 3 Yes   
## 4 Yes   
## 5 Yes   
## 6 Yes

confusionMatrix(ames4pred$.pred\_class, ames\_train$Above\_Median, positive="Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Yes No  
## Yes 771 11  
## No 12 732  
##   
## Accuracy : 0.9849   
## 95% CI : (0.9775, 0.9904)  
## No Information Rate : 0.5131   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.9698   
##   
## Mcnemar's Test P-Value : 1   
##   
## Sensitivity : 0.9847   
## Specificity : 0.9852   
## Pos Pred Value : 0.9859   
## Neg Pred Value : 0.9839   
## Prevalence : 0.5131   
## Detection Rate : 0.5052   
## Detection Prevalence : 0.5125   
## Balanced Accuracy : 0.9849   
##   
## 'Positive' Class : Yes   
##