##Vicki Hartley Course Project Phase 2 Delivery 2##

ames = read\_csv("ames\_student-1.csv")

## Rows: 2053 Columns: 81  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (47): MS\_SubClass, MS\_Zoning, Street, Alley, Lot\_Shape, Land\_Contour, Ut...  
## dbl (34): Lot\_Frontage, Lot\_Area, Year\_Built, Year\_Remod\_Add, Mas\_Vnr\_Area, ...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

glimpse(ames)

## Rows: 2,053  
## Columns: 81  
## $ MS\_SubClass <chr> "One\_Story\_1946\_and\_Newer\_All\_Styles", "One\_Story\_1…  
## $ MS\_Zoning <chr> "Residential\_Low\_Density", "Residential\_High\_Densit…  
## $ Lot\_Frontage <dbl> 141, 80, 81, 93, 74, 78, 43, 39, 0, 85, 0, 47, 152,…  
## $ Lot\_Area <dbl> 31770, 11622, 14267, 11160, 13830, 9978, 5005, 5389…  
## $ Street <chr> "Pave", "Pave", "Pave", "Pave", "Pave", "Pave", "Pa…  
## $ Alley <chr> "No\_Alley\_Access", "No\_Alley\_Access", "No\_Alley\_Acc…  
## $ Lot\_Shape <chr> "Slightly\_Irregular", "Regular", "Slightly\_Irregula…  
## $ Land\_Contour <chr> "Lvl", "Lvl", "Lvl", "Lvl", "Lvl", "Lvl", "HLS", "L…  
## $ Utilities <chr> "AllPub", "AllPub", "AllPub", "AllPub", "AllPub", "…  
## $ Lot\_Config <chr> "Corner", "Inside", "Corner", "Corner", "Inside", "…  
## $ Land\_Slope <chr> "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", "G…  
## $ Neighborhood <chr> "North\_Ames", "North\_Ames", "North\_Ames", "North\_Am…  
## $ Condition\_1 <chr> "Norm", "Feedr", "Norm", "Norm", "Norm", "Norm", "N…  
## $ Condition\_2 <chr> "Norm", "Norm", "Norm", "Norm", "Norm", "Norm", "No…  
## $ Bldg\_Type <chr> "OneFam", "OneFam", "OneFam", "OneFam", "OneFam", "…  
## $ House\_Style <chr> "One\_Story", "One\_Story", "One\_Story", "One\_Story",…  
## $ Overall\_Qual <chr> "Above\_Average", "Average", "Above\_Average", "Good"…  
## $ Overall\_Cond <chr> "Average", "Above\_Average", "Above\_Average", "Avera…  
## $ Year\_Built <dbl> 1960, 1961, 1958, 1968, 1997, 1998, 1992, 1995, 199…  
## $ Year\_Remod\_Add <dbl> 1960, 1961, 1958, 1968, 1998, 1998, 1992, 1996, 200…  
## $ Roof\_Style <chr> "Hip", "Gable", "Hip", "Hip", "Gable", "Gable", "Ga…  
## $ Roof\_Matl <chr> "CompShg", "CompShg", "CompShg", "CompShg", "CompSh…  
## $ Exterior\_1st <chr> "BrkFace", "VinylSd", "Wd Sdng", "BrkFace", "VinylS…  
## $ Exterior\_2nd <chr> "Plywood", "VinylSd", "Wd Sdng", "BrkFace", "VinylS…  
## $ Mas\_Vnr\_Type <chr> "Stone", "None", "BrkFace", "None", "None", "BrkFac…  
## $ Mas\_Vnr\_Area <dbl> 112, 0, 108, 0, 0, 20, 0, 0, 0, 0, 0, 603, 0, 350, …  
## $ Exter\_Qual <chr> "Typical", "Typical", "Typical", "Good", "Typical",…  
## $ Exter\_Cond <chr> "Typical", "Typical", "Typical", "Typical", "Typica…  
## $ Foundation <chr> "CBlock", "CBlock", "CBlock", "CBlock", "PConc", "P…  
## $ Bsmt\_Qual <chr> "Typical", "Typical", "Typical", "Typical", "Good",…  
## $ Bsmt\_Cond <chr> "Good", "Typical", "Typical", "Typical", "Typical",…  
## $ Bsmt\_Exposure <chr> "Gd", "No", "No", "No", "No", "No", "No", "No", "No…  
## $ BsmtFin\_Type\_1 <chr> "BLQ", "Rec", "ALQ", "ALQ", "GLQ", "GLQ", "ALQ", "G…  
## $ BsmtFin\_SF\_1 <dbl> 2, 6, 1, 1, 3, 3, 1, 3, 1, 3, 3, 1, 3, 3, 2, 3, 1, …  
## $ BsmtFin\_Type\_2 <chr> "Unf", "LwQ", "Unf", "Unf", "Unf", "Unf", "Unf", "U…  
## $ BsmtFin\_SF\_2 <dbl> 0, 144, 0, 0, 0, 0, 0, 0, 0, 0, 1120, 0, 0, 0, 0, 0…  
## $ Bsmt\_Unf\_SF <dbl> 441, 270, 406, 1045, 137, 324, 1017, 415, 233, 663,…  
## $ Total\_Bsmt\_SF <dbl> 1080, 882, 1329, 2110, 928, 926, 1280, 1595, 1168, …  
## $ Heating <chr> "GasA", "GasA", "GasA", "GasA", "GasA", "GasA", "Ga…  
## $ Heating\_QC <chr> "Fair", "Typical", "Typical", "Excellent", "Good", …  
## $ Central\_Air <chr> "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "…  
## $ Electrical <chr> "SBrkr", "SBrkr", "SBrkr", "SBrkr", "SBrkr", "SBrkr…  
## $ First\_Flr\_SF <dbl> 1656, 896, 1329, 2110, 928, 926, 1280, 1616, 1187, …  
## $ Second\_Flr\_SF <dbl> 0, 0, 0, 0, 701, 678, 0, 0, 0, 0, 0, 1589, 672, 0, …  
## $ Low\_Qual\_Fin\_SF <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Gr\_Liv\_Area <dbl> 1656, 896, 1329, 2110, 1629, 1604, 1280, 1616, 1187…  
## $ Bsmt\_Full\_Bath <dbl> 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, …  
## $ Bsmt\_Half\_Bath <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Full\_Bath <dbl> 1, 1, 1, 2, 2, 2, 2, 2, 2, 1, 1, 3, 2, 1, 2, 2, 1, …  
## $ Half\_Bath <dbl> 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, …  
## $ Bedroom\_AbvGr <dbl> 3, 2, 3, 3, 3, 3, 2, 2, 3, 2, 1, 4, 4, 1, 3, 3, 2, …  
## $ Kitchen\_AbvGr <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …  
## $ Kitchen\_Qual <chr> "Typical", "Typical", "Good", "Excellent", "Typical…  
## $ TotRms\_AbvGrd <dbl> 7, 5, 6, 8, 6, 7, 5, 5, 6, 5, 4, 12, 8, 8, 7, 7, 5,…  
## $ Functional <chr> "Typ", "Typ", "Typ", "Typ", "Typ", "Typ", "Typ", "T…  
## $ Fireplaces <dbl> 2, 0, 0, 2, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, …  
## $ Fireplace\_Qu <chr> "Good", "No\_Fireplace", "No\_Fireplace", "Typical", …  
## $ Garage\_Type <chr> "Attchd", "Attchd", "Attchd", "Attchd", "Attchd", "…  
## $ Garage\_Finish <chr> "Fin", "Unf", "Unf", "Fin", "Fin", "Fin", "RFn", "R…  
## $ Garage\_Cars <dbl> 2, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 3, 2, 2, 2, …  
## $ Garage\_Area <dbl> 528, 730, 312, 522, 482, 470, 506, 608, 420, 506, 5…  
## $ Garage\_Qual <chr> "Typical", "Typical", "Typical", "Typical", "Typica…  
## $ Garage\_Cond <chr> "Typical", "Typical", "Typical", "Typical", "Typica…  
## $ Paved\_Drive <chr> "Partial\_Pavement", "Paved", "Paved", "Paved", "Pav…  
## $ Wood\_Deck\_SF <dbl> 210, 140, 393, 0, 212, 360, 0, 237, 483, 192, 0, 50…  
## $ Open\_Porch\_SF <dbl> 62, 0, 36, 0, 34, 36, 82, 152, 21, 0, 54, 36, 12, 0…  
## $ Enclosed\_Porch <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Three\_season\_porch <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Screen\_Porch <dbl> 0, 120, 0, 0, 0, 0, 144, 0, 0, 0, 140, 210, 0, 0, 0…  
## $ Pool\_Area <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Pool\_QC <chr> "No\_Pool", "No\_Pool", "No\_Pool", "No\_Pool", "No\_Poo…  
## $ Fence <chr> "No\_Fence", "Minimum\_Privacy", "No\_Fence", "No\_Fenc…  
## $ Misc\_Feature <chr> "None", "None", "Gar2", "None", "None", "None", "No…  
## $ Misc\_Val <dbl> 0, 0, 12500, 0, 0, 0, 0, 0, 500, 0, 0, 0, 0, 0, 0, …  
## $ Mo\_Sold <dbl> 5, 6, 6, 4, 3, 6, 1, 3, 3, 2, 6, 6, 6, 6, 1, 1, 3, …  
## $ Year\_Sold <dbl> 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 201…  
## $ Sale\_Type <chr> "WD", "WD", "WD", "WD", "WD", "WD", "WD", "WD", "WD…  
## $ Sale\_Condition <chr> "Normal", "Normal", "Normal", "Normal", "Normal", "…  
## $ Longitude <dbl> -93.61975, -93.61976, -93.61939, -93.61732, -93.638…  
## $ Latitude <dbl> 42.05403, 42.05301, 42.05266, 42.05125, 42.06090, 4…  
## $ Above\_Median <chr> "Yes", "No", "Yes", "Yes", "Yes", "Yes", "Yes", "Ye…

summary(ames)

## MS\_SubClass MS\_Zoning Lot\_Frontage Lot\_Area   
## Length:2053 Length:2053 Min. : 0.00 Min. : 1300   
## Class :character Class :character 1st Qu.: 43.00 1st Qu.: 7500   
## Mode :character Mode :character Median : 62.00 Median : 9548   
## Mean : 57.38 Mean : 10258   
## 3rd Qu.: 78.00 3rd Qu.: 11600   
## Max. :313.00 Max. :215245   
## Street Alley Lot\_Shape Land\_Contour   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Utilities Lot\_Config Land\_Slope Neighborhood   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Condition\_1 Condition\_2 Bldg\_Type House\_Style   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Overall\_Qual Overall\_Cond Year\_Built Year\_Remod\_Add  
## Length:2053 Length:2053 Min. :1875 Min. :1950   
## Class :character Class :character 1st Qu.:1953 1st Qu.:1965   
## Mode :character Mode :character Median :1972 Median :1993   
## Mean :1971 Mean :1984   
## 3rd Qu.:2000 3rd Qu.:2004   
## Max. :2010 Max. :2010   
## Roof\_Style Roof\_Matl Exterior\_1st Exterior\_2nd   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Mas\_Vnr\_Type Mas\_Vnr\_Area Exter\_Qual Exter\_Cond   
## Length:2053 Min. : 0.0 Length:2053 Length:2053   
## Class :character 1st Qu.: 0.0 Class :character Class :character   
## Mode :character Median : 0.0 Mode :character Mode :character   
## Mean : 103.8   
## 3rd Qu.: 164.0   
## Max. :1600.0   
## Foundation Bsmt\_Qual Bsmt\_Cond Bsmt\_Exposure   
## Length:2053 Length:2053 Length:2053 Length:2053   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## BsmtFin\_Type\_1 BsmtFin\_SF\_1 BsmtFin\_Type\_2 BsmtFin\_SF\_2   
## Length:2053 Min. :1.00 Length:2053 Min. : 0.00   
## Class :character 1st Qu.:3.00 Class :character 1st Qu.: 0.00   
## Mode :character Median :3.00 Mode :character Median : 0.00   
## Mean :4.21 Mean : 52.57   
## 3rd Qu.:7.00 3rd Qu.: 0.00   
## Max. :7.00 Max. :1526.00   
## Bsmt\_Unf\_SF Total\_Bsmt\_SF Heating Heating\_QC   
## Min. : 0.0 Min. : 0 Length:2053 Length:2053   
## 1st Qu.: 226.0 1st Qu.: 793 Class :character Class :character   
## Median : 460.0 Median : 988 Mode :character Mode :character   
## Mean : 561.2 Mean :1055   
## 3rd Qu.: 801.0 3rd Qu.:1304   
## Max. :2336.0 Max. :5095   
## Central\_Air Electrical First\_Flr\_SF Second\_Flr\_SF   
## Length:2053 Length:2053 Min. : 432 Min. : 0.0   
## Class :character Class :character 1st Qu.: 882 1st Qu.: 0.0   
## Mode :character Mode :character Median :1088 Median : 0.0   
## Mean :1168 Mean : 326.1   
## 3rd Qu.:1402 3rd Qu.: 701.0   
## Max. :5095 Max. :1862.0   
## Low\_Qual\_Fin\_SF Gr\_Liv\_Area Bsmt\_Full\_Bath Bsmt\_Half\_Bath   
## Min. : 0.000 Min. : 480 Min. :0.0000 Min. :0.00000   
## 1st Qu.: 0.000 1st Qu.:1137 1st Qu.:0.0000 1st Qu.:0.00000   
## Median : 0.000 Median :1447 Median :0.0000 Median :0.00000   
## Mean : 4.973 Mean :1499 Mean :0.4301 Mean :0.05796   
## 3rd Qu.: 0.000 3rd Qu.:1737 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1064.000 Max. :5095 Max. :3.0000 Max. :2.00000   
## Full\_Bath Half\_Bath Bedroom\_AbvGr Kitchen\_AbvGr   
## Min. :0.000 Min. :0.0000 Min. :0.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000 1st Qu.:1.000   
## Median :2.000 Median :0.0000 Median :3.000 Median :1.000   
## Mean :1.564 Mean :0.3751 Mean :2.855 Mean :1.047   
## 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000 3rd Qu.:1.000   
## Max. :4.000 Max. :2.0000 Max. :6.000 Max. :3.000   
## Kitchen\_Qual TotRms\_AbvGrd Functional Fireplaces   
## Length:2053 Min. : 3.000 Length:2053 Min. :0.000   
## Class :character 1st Qu.: 5.000 Class :character 1st Qu.:0.000   
## Mode :character Median : 6.000 Mode :character Median :1.000   
## Mean : 6.442 Mean :0.603   
## 3rd Qu.: 7.000 3rd Qu.:1.000   
## Max. :15.000 Max. :4.000   
## Fireplace\_Qu Garage\_Type Garage\_Finish Garage\_Cars   
## Length:2053 Length:2053 Length:2053 Min. :0.000   
## Class :character Class :character Class :character 1st Qu.:1.000   
## Mode :character Mode :character Mode :character Median :2.000   
## Mean :1.774   
## 3rd Qu.:2.000   
## Max. :5.000   
## Garage\_Area Garage\_Qual Garage\_Cond Paved\_Drive   
## Min. : 0 Length:2053 Length:2053 Length:2053   
## 1st Qu.: 320 Class :character Class :character Class :character   
## Median : 478 Mode :character Mode :character Mode :character   
## Mean : 472   
## 3rd Qu.: 576   
## Max. :1488   
## Wood\_Deck\_SF Open\_Porch\_SF Enclosed\_Porch Three\_season\_porch  
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 27.00 Median : 0.00 Median : 0.000   
## Mean : 93.52 Mean : 48.17 Mean : 23.02 Mean : 2.799   
## 3rd Qu.: 168.00 3rd Qu.: 72.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :1424.00 Max. :742.00 Max. :584.00 Max. :407.000   
## Screen\_Porch Pool\_Area Pool\_QC Fence   
## Min. : 0.00 Min. : 0.000 Length:2053 Length:2053   
## 1st Qu.: 0.00 1st Qu.: 0.000 Class :character Class :character   
## Median : 0.00 Median : 0.000 Mode :character Mode :character   
## Mean : 16.68 Mean : 1.339   
## 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :576.00 Max. :800.000   
## Misc\_Feature Misc\_Val Mo\_Sold Year\_Sold   
## Length:2053 Min. : 0.00 Min. : 1.000 Min. :2006   
## Class :character 1st Qu.: 0.00 1st Qu.: 4.000 1st Qu.:2007   
## Mode :character Median : 0.00 Median : 6.000 Median :2008   
## Mean : 60.12 Mean : 6.189 Mean :2008   
## 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009   
## Max. :17000.00 Max. :12.000 Max. :2010   
## Sale\_Type Sale\_Condition Longitude Latitude   
## Length:2053 Length:2053 Min. :-93.69 Min. :41.99   
## Class :character Class :character 1st Qu.:-93.66 1st Qu.:42.02   
## Mode :character Mode :character Median :-93.64 Median :42.03   
## Mean :-93.64 Mean :42.03   
## 3rd Qu.:-93.62 3rd Qu.:42.05   
## Max. :-93.58 Max. :42.06   
## Above\_Median   
## Length:2053   
## Class :character   
## Mode :character   
##   
##   
##

str(ames)

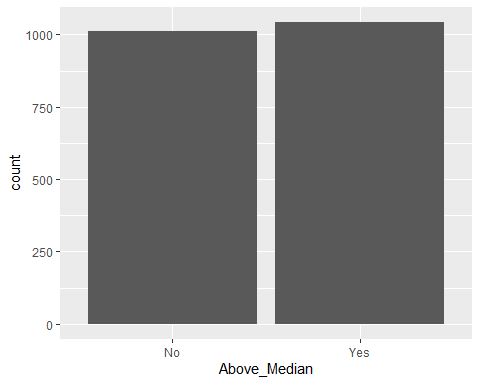
## spc\_tbl\_ [2,053 × 81] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ MS\_SubClass : chr [1:2053] "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" ...  
## $ MS\_Zoning : chr [1:2053] "Residential\_Low\_Density" "Residential\_High\_Density" "Residential\_Low\_Density" "Residential\_Low\_Density" ...  
## $ 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 : chr [1:2053] "Pave" "Pave" "Pave" "Pave" ...  
## $ Alley : chr [1:2053] "No\_Alley\_Access" "No\_Alley\_Access" "No\_Alley\_Access" "No\_Alley\_Access" ...  
## $ Lot\_Shape : chr [1:2053] "Slightly\_Irregular" "Regular" "Slightly\_Irregular" "Regular" ...  
## $ Land\_Contour : chr [1:2053] "Lvl" "Lvl" "Lvl" "Lvl" ...  
## $ Utilities : chr [1:2053] "AllPub" "AllPub" "AllPub" "AllPub" ...  
## $ Lot\_Config : chr [1:2053] "Corner" "Inside" "Corner" "Corner" ...  
## $ Land\_Slope : chr [1:2053] "Gtl" "Gtl" "Gtl" "Gtl" ...  
## $ Neighborhood : chr [1:2053] "North\_Ames" "North\_Ames" "North\_Ames" "North\_Ames" ...  
## $ Condition\_1 : chr [1:2053] "Norm" "Feedr" "Norm" "Norm" ...  
## $ Condition\_2 : chr [1:2053] "Norm" "Norm" "Norm" "Norm" ...  
## $ Bldg\_Type : chr [1:2053] "OneFam" "OneFam" "OneFam" "OneFam" ...  
## $ House\_Style : chr [1:2053] "One\_Story" "One\_Story" "One\_Story" "One\_Story" ...  
## $ Overall\_Qual : chr [1:2053] "Above\_Average" "Average" "Above\_Average" "Good" ...  
## $ Overall\_Cond : chr [1:2053] "Average" "Above\_Average" "Above\_Average" "Average" ...  
## $ Year\_Built : num [1:2053] 1960 1961 1958 1968 1997 ...  
## $ Year\_Remod\_Add : num [1:2053] 1960 1961 1958 1968 1998 ...  
## $ Roof\_Style : chr [1:2053] "Hip" "Gable" "Hip" "Hip" ...  
## $ Roof\_Matl : chr [1:2053] "CompShg" "CompShg" "CompShg" "CompShg" ...  
## $ Exterior\_1st : chr [1:2053] "BrkFace" "VinylSd" "Wd Sdng" "BrkFace" ...  
## $ Exterior\_2nd : chr [1:2053] "Plywood" "VinylSd" "Wd Sdng" "BrkFace" ...  
## $ Mas\_Vnr\_Type : chr [1:2053] "Stone" "None" "BrkFace" "None" ...  
## $ Mas\_Vnr\_Area : num [1:2053] 112 0 108 0 0 20 0 0 0 0 ...  
## $ Exter\_Qual : chr [1:2053] "Typical" "Typical" "Typical" "Good" ...  
## $ Exter\_Cond : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Foundation : chr [1:2053] "CBlock" "CBlock" "CBlock" "CBlock" ...  
## $ Bsmt\_Qual : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Bsmt\_Cond : chr [1:2053] "Good" "Typical" "Typical" "Typical" ...  
## $ Bsmt\_Exposure : chr [1:2053] "Gd" "No" "No" "No" ...  
## $ BsmtFin\_Type\_1 : chr [1:2053] "BLQ" "Rec" "ALQ" "ALQ" ...  
## $ BsmtFin\_SF\_1 : num [1:2053] 2 6 1 1 3 3 1 3 1 3 ...  
## $ BsmtFin\_Type\_2 : chr [1:2053] "Unf" "LwQ" "Unf" "Unf" ...  
## $ 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 : chr [1:2053] "GasA" "GasA" "GasA" "GasA" ...  
## $ Heating\_QC : chr [1:2053] "Fair" "Typical" "Typical" "Excellent" ...  
## $ Central\_Air : chr [1:2053] "Y" "Y" "Y" "Y" ...  
## $ Electrical : chr [1:2053] "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...  
## $ 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 : chr [1:2053] "Typical" "Typical" "Good" "Excellent" ...  
## $ TotRms\_AbvGrd : num [1:2053] 7 5 6 8 6 7 5 5 6 5 ...  
## $ Functional : chr [1:2053] "Typ" "Typ" "Typ" "Typ" ...  
## $ Fireplaces : num [1:2053] 2 0 0 2 1 1 0 1 0 1 ...  
## $ Fireplace\_Qu : chr [1:2053] "Good" "No\_Fireplace" "No\_Fireplace" "Typical" ...  
## $ Garage\_Type : chr [1:2053] "Attchd" "Attchd" "Attchd" "Attchd" ...  
## $ Garage\_Finish : chr [1:2053] "Fin" "Unf" "Unf" "Fin" ...  
## $ 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 : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Garage\_Cond : chr [1:2053] "Typical" "Typical" "Typical" "Typical" ...  
## $ Paved\_Drive : chr [1:2053] "Partial\_Pavement" "Paved" "Paved" "Paved" ...  
## $ 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 : chr [1:2053] "No\_Pool" "No\_Pool" "No\_Pool" "No\_Pool" ...  
## $ Fence : chr [1:2053] "No\_Fence" "Minimum\_Privacy" "No\_Fence" "No\_Fence" ...  
## $ Misc\_Feature : chr [1:2053] "None" "None" "Gar2" "None" ...  
## $ 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 : chr [1:2053] "WD" "WD" "WD" "WD" ...  
## $ Sale\_Condition : chr [1:2053] "Normal" "Normal" "Normal" "Normal" ...  
## $ 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 : chr [1:2053] "Yes" "No" "Yes" "Yes" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. 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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

There does not appear to be missing data

Using Geom Box Plots and Bars to better visualize the data

*It looks like the data is just about split in half with a little over 1,000 of homes being above the median and ~1,000 falling below median*

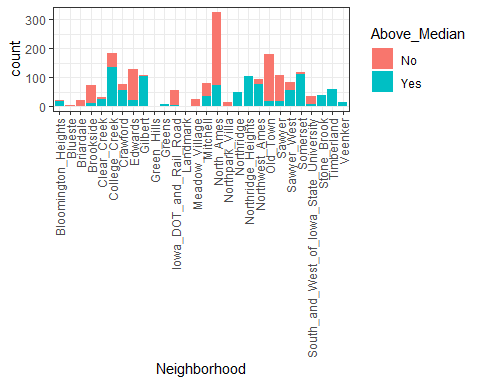
ggplot(ames,aes(Above\_Median)) + geom\_bar() + theme\_get()



*The below shows what neighborhoods have the most houses above median* *North Ames seems to have the most houses below median whereas College Creek looks to have the most above but this is only a visual inspection prediction*

ggplot(ames,aes(x = Neighborhood, fill = Above\_Median)) + geom\_bar(postion="fill") + theme\_bw() + scale\_x\_discrete(guide = guide\_axis(angle = 90))

## Warning in geom\_bar(postion = "fill"): Ignoring unknown parameters: `postion`



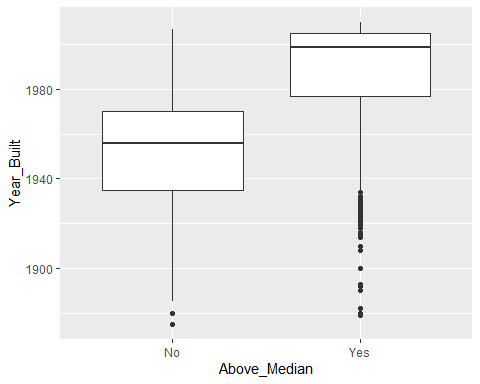
*The below chart shows what neighborhoods have the most and least houses above median* 99% of homes withing the Northridge Heights neighborhood are above median  
92% of homes within the Iowa\_DOT and Rail Road neighborhood are below median

t1 = table(ames$Above\_Median, ames$Neighborhood)   
prop.table(t1, margin = 2 )

##   
## Bloomington\_Heights Blueste Briardale Brookside Clear\_Creek  
## No 0.09523810 0.80000000 1.00000000 0.86486486 0.22580645  
## Yes 0.90476190 0.20000000 0.00000000 0.13513514 0.77419355  
##   
## College\_Creek Crawford Edwards Gilbert Green\_Hills Greens  
## No 0.26775956 0.27272727 0.84496124 0.02752294 0.00000000 0.14285714  
## Yes 0.73224044 0.72727273 0.15503876 0.97247706 1.00000000 0.85714286  
##   
## Iowa\_DOT\_and\_Rail\_Road Landmark Meadow\_Village Mitchell North\_Ames  
## No 0.92982456 1.00000000 1.00000000 0.54430380 0.77981651  
## Yes 0.07017544 0.00000000 0.00000000 0.45569620 0.22018349  
##   
## Northpark\_Villa Northridge Northridge\_Heights Northwest\_Ames Old\_Town  
## No 1.00000000 0.00000000 0.00952381 0.20000000 0.90055249  
## Yes 0.00000000 1.00000000 0.99047619 0.80000000 0.09944751  
##   
## Sawyer Sawyer\_West Somerset South\_and\_West\_of\_Iowa\_State\_University  
## No 0.84403670 0.30487805 0.07563025 0.74285714  
## Yes 0.15596330 0.69512195 0.92436975 0.25714286  
##   
## Stone\_Brook Timberland Veenker  
## No 0.00000000 0.01666667 0.06250000  
## Yes 1.00000000 0.98333333 0.93750000

*It looks like the average age of homes below median is ~1960 and the average age of homes selling above median is ~2000s*

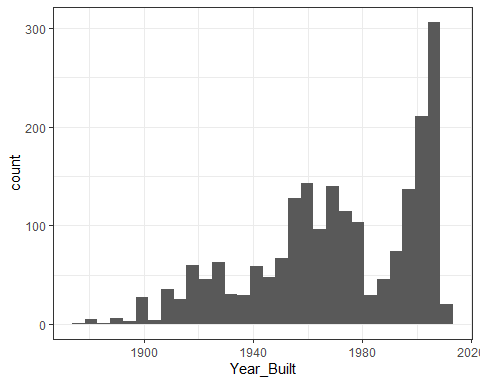
ggplot(ames, aes(x = Above\_Median, y = Year\_Built)) + geom\_boxplot() + theme\_get()



*It looks like most homes were built after the 2000s*

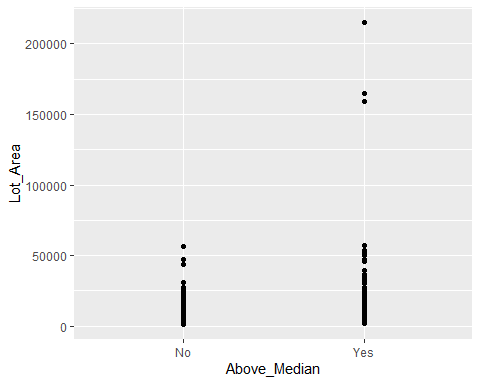
ggplot(ames, aes(x = Year\_Built)) + geom\_histogram() + theme\_bw()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



*There is some spread for homes above median by lot area*

ggplot(ames, aes(x = Above\_Median, y = Lot\_Area)) + geom\_point() + theme\_update()



#Predictions#

Split the data into testing (20%) and training (80%) stratified by the “Above\_Median” variable

set.seed(123)  
ames\_split = initial\_split (ames, prop = 0.80, strata = Above\_Median)  
train\_train = training(ames\_split)  
test\_train = testing(ames\_split)  
str(train\_train)

## spc\_tbl\_ [1,642 × 81] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ MS\_SubClass : chr [1:1642] "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" "One\_Story\_1946\_and\_Newer\_All\_Styles" "Two\_Story\_PUD\_1946\_and\_Newer" ...  
## $ MS\_Zoning : chr [1:1642] "Residential\_Low\_Density" "Residential\_Low\_Density" "Residential\_Low\_Density" "Residential\_Medium\_Density" ...  
## $ Lot\_Frontage : num [1:1642] 0 65 70 21 21 55 81 65 30 30 ...  
## $ Lot\_Area : num [1:1642] 11241 8450 8400 1680 1680 ...  
## $ Street : chr [1:1642] "Pave" "Pave" "Pave" "Pave" ...  
## $ Alley : chr [1:1642] "No\_Alley\_Access" "No\_Alley\_Access" "No\_Alley\_Access" "No\_Alley\_Access" ...  
## $ Lot\_Shape : chr [1:1642] "Slightly\_Irregular" "Regular" "Regular" "Regular" ...  
## $ Land\_Contour : chr [1:1642] "Lvl" "Lvl" "Lvl" "Lvl" ...  
## $ Utilities : chr [1:1642] "AllPub" "AllPub" "AllPub" "AllPub" ...  
## $ Lot\_Config : chr [1:1642] "CulDSac" "Inside" "Corner" "Inside" ...  
## $ Land\_Slope : chr [1:1642] "Gtl" "Gtl" "Gtl" "Gtl" ...  
## $ Neighborhood : chr [1:1642] "North\_Ames" "North\_Ames" "North\_Ames" "Briardale" ...  
## $ Condition\_1 : chr [1:1642] "Norm" "Norm" "Norm" "Norm" ...  
## $ Condition\_2 : chr [1:1642] "Norm" "Norm" "Norm" "Norm" ...  
## $ Bldg\_Type : chr [1:1642] "OneFam" "OneFam" "OneFam" "Twnhs" ...  
## $ House\_Style : chr [1:1642] "One\_Story" "One\_Story" "One\_Story" "Two\_Story" ...  
## $ Overall\_Qual : chr [1:1642] "Above\_Average" "Average" "Below\_Average" "Above\_Average" ...  
## $ Overall\_Cond : chr [1:1642] "Good" "Above\_Average" "Average" "Average" ...  
## $ Year\_Built : num [1:1642] 1970 1968 1970 1971 1971 ...  
## $ Year\_Remod\_Add : num [1:1642] 1970 1968 1970 1971 1971 ...  
## $ Roof\_Style : chr [1:1642] "Gable" "Gable" "Gable" "Gable" ...  
## $ Roof\_Matl : chr [1:1642] "CompShg" "CompShg" "CompShg" "CompShg" ...  
## $ Exterior\_1st : chr [1:1642] "Wd Sdng" "VinylSd" "Plywood" "HdBoard" ...  
## $ Exterior\_2nd : chr [1:1642] "Wd Sdng" "VinylSd" "Plywood" "HdBoard" ...  
## $ Mas\_Vnr\_Type : chr [1:1642] "BrkFace" "None" "None" "BrkFace" ...  
## $ Mas\_Vnr\_Area : num [1:1642] 180 0 0 504 381 0 0 0 120 0 ...  
## $ Exter\_Qual : chr [1:1642] "Typical" "Typical" "Typical" "Typical" ...  
## $ Exter\_Cond : chr [1:1642] "Typical" "Typical" "Typical" "Typical" ...  
## $ Foundation : chr [1:1642] "CBlock" "CBlock" "CBlock" "CBlock" ...  
## $ Bsmt\_Qual : chr [1:1642] "Typical" "Typical" "Typical" "Typical" ...  
## $ Bsmt\_Cond : chr [1:1642] "Typical" "Typical" "Typical" "Typical" ...  
## $ Bsmt\_Exposure : chr [1:1642] "No" "No" "No" "No" ...  
## $ BsmtFin\_Type\_1 : chr [1:1642] "ALQ" "BLQ" "ALQ" "Rec" ...  
## $ BsmtFin\_SF\_1 : num [1:1642] 1 2 1 6 7 7 3 3 3 7 ...  
## $ BsmtFin\_Type\_2 : chr [1:1642] "Unf" "Unf" "Rec" "Unf" ...  
## $ BsmtFin\_SF\_2 : num [1:1642] 0 0 78 0 0 0 0 117 0 0 ...  
## $ Bsmt\_Unf\_SF : num [1:1642] 426 281 0 327 525 918 702 224 320 600 ...  
## $ Total\_Bsmt\_SF : num [1:1642] 1004 1056 882 483 525 ...  
## $ Heating : chr [1:1642] "GasA" "GasA" "GasA" "GasA" ...  
## $ Heating\_QC : chr [1:1642] "Excellent" "Excellent" "Typical" "Typical" ...  
## $ Central\_Air : chr [1:1642] "Y" "Y" "Y" "Y" ...  
## $ Electrical : chr [1:1642] "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...  
## $ First\_Flr\_SF : num [1:1642] 1004 1056 882 483 525 ...  
## $ Second\_Flr\_SF : num [1:1642] 0 0 0 504 567 0 0 0 600 600 ...  
## $ Low\_Qual\_Fin\_SF : num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Gr\_Liv\_Area : num [1:1642] 1004 1056 882 987 1092 ...  
## $ Bsmt\_Full\_Bath : num [1:1642] 1 1 1 0 0 0 0 1 0 0 ...  
## $ Bsmt\_Half\_Bath : num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Full\_Bath : num [1:1642] 1 1 1 1 1 2 2 1 2 2 ...  
## $ Half\_Bath : num [1:1642] 0 0 0 1 1 0 0 0 1 1 ...  
## $ Bedroom\_AbvGr : num [1:1642] 2 3 2 2 3 2 3 3 2 2 ...  
## $ Kitchen\_AbvGr : num [1:1642] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Kitchen\_Qual : chr [1:1642] "Typical" "Typical" "Typical" "Typical" ...  
## $ TotRms\_AbvGrd : num [1:1642] 5 6 4 5 6 5 6 5 4 4 ...  
## $ Functional : chr [1:1642] "Typ" "Typ" "Typ" "Typ" ...  
## $ Fireplaces : num [1:1642] 1 1 0 0 0 1 0 1 0 0 ...  
## $ Fireplace\_Qu : chr [1:1642] "Typical" "Fair" "No\_Fireplace" "No\_Fireplace" ...  
## $ Garage\_Type : chr [1:1642] "Attchd" "Attchd" "Attchd" "Detchd" ...  
## $ Garage\_Finish : chr [1:1642] "Fin" "Unf" "Fin" "Unf" ...  
## $ Garage\_Cars : num [1:1642] 2 1 2 1 1 1 2 1 2 2 ...  
## $ Garage\_Area : num [1:1642] 480 304 525 264 264 264 480 336 480 480 ...  
## $ Garage\_Qual : chr [1:1642] "Typical" "Typical" "Typical" "Typical" ...  
## $ Garage\_Cond : chr [1:1642] "Typical" "Typical" "Typical" "Typical" ...  
## $ Paved\_Drive : chr [1:1642] "Paved" "Paved" "Paved" "Paved" ...  
## $ Wood\_Deck\_SF : num [1:1642] 0 0 240 275 0 28 0 416 0 0 ...  
## $ Open\_Porch\_SF : num [1:1642] 0 85 0 0 0 0 0 144 172 172 ...  
## $ Enclosed\_Porch : num [1:1642] 0 184 0 0 0 0 0 0 0 0 ...  
## $ Three\_season\_porch: num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Screen\_Porch : num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_Area : num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_QC : chr [1:1642] "No\_Pool" "No\_Pool" "No\_Pool" "No\_Pool" ...  
## $ Fence : chr [1:1642] "No\_Fence" "No\_Fence" "Minimum\_Privacy" "No\_Fence" ...  
## $ Misc\_Feature : chr [1:1642] "Shed" "None" "None" "None" ...  
## $ Misc\_Val : num [1:1642] 700 0 0 0 0 0 0 0 0 0 ...  
## $ Mo\_Sold : num [1:1642] 3 7 4 2 3 4 5 4 4 6 ...  
## $ Year\_Sold : num [1:1642] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ Sale\_Type : chr [1:1642] "WD" "WD" "WD" "COD" ...  
## $ Sale\_Condition : chr [1:1642] "Normal" "Normal" "Normal" "Normal" ...  
## $ Longitude : num [1:1642] -93.6 -93.6 -93.6 -93.6 -93.6 ...  
## $ Latitude : num [1:1642] 42.1 42.1 42.1 42.1 42.1 ...  
## $ Above\_Median : chr [1:1642] "No" "No" "No" "No" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. 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()  
## .. )  
## - attr(\*, "problems")=<externalptr>

Preparing data

train\_train <- train\_train %>% mutate\_if(is.character,as.factor)%>%   
 mutate(Above\_Median = as\_factor(Above\_Median)) %>%   
 mutate(Above\_Median = fct\_recode(Above\_Median, "No" = "0", "Yes" = "1" ))

## Warning: There was 1 warning in `mutate()`.  
## ℹ In argument: `Above\_Median = fct\_recode(Above\_Median, No = "0", Yes = "1")`.  
## Caused by warning:  
## ! Unknown levels in `f`: 0, 1

str(train\_train)

## tibble [1,642 × 81] (S3: tbl\_df/tbl/data.frame)  
## $ MS\_SubClass : Factor w/ 16 levels "Duplex\_All\_Styles\_and\_Ages",..: 6 6 6 16 16 7 6 6 16 16 ...  
## $ MS\_Zoning : Factor w/ 7 levels "A\_agr","C\_all",..: 6 6 6 7 7 6 6 6 3 3 ...  
## $ Lot\_Frontage : num [1:1642] 0 65 70 21 21 55 81 65 30 30 ...  
## $ Lot\_Area : num [1:1642] 11241 8450 8400 1680 1680 ...  
## $ Street : Factor w/ 2 levels "Grvl","Pave": 2 2 2 2 2 2 2 2 2 2 ...  
## $ Alley : Factor w/ 3 levels "Gravel","No\_Alley\_Access",..: 2 2 2 2 2 2 2 2 3 3 ...  
## $ Lot\_Shape : Factor w/ 4 levels "Irregular","Moderately\_Irregular",..: 4 3 3 3 3 3 3 3 3 3 ...  
## $ Land\_Contour : Factor w/ 4 levels "Bnk","HLS","Low",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ Utilities : Factor w/ 2 levels "AllPub","NoSewr": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Lot\_Config : Factor w/ 5 levels "Corner","CulDSac",..: 2 5 1 5 5 5 1 5 5 5 ...  
## $ Land\_Slope : Factor w/ 3 levels "Gtl","Mod","Sev": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Neighborhood : Factor w/ 28 levels "Bloomington\_Heights",..: 16 16 16 3 3 23 23 22 24 24 ...  
## $ Condition\_1 : Factor w/ 9 levels "Artery","Feedr",..: 3 3 3 3 3 3 3 6 3 3 ...  
## $ Condition\_2 : Factor w/ 8 levels "Artery","Feedr",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ Bldg\_Type : Factor w/ 5 levels "Duplex","OneFam",..: 2 2 2 3 3 4 2 2 4 4 ...  
## $ House\_Style : Factor w/ 8 levels "One\_and\_Half\_Fin",..: 3 3 3 8 8 3 3 3 8 8 ...  
## $ Overall\_Qual : Factor w/ 10 levels "Above\_Average",..: 1 2 3 1 1 1 1 2 6 6 ...  
## $ Overall\_Cond : Factor w/ 9 levels "Above\_Average",..: 6 1 2 2 2 2 2 8 2 2 ...  
## $ Year\_Built : num [1:1642] 1970 1968 1970 1971 1971 ...  
## $ Year\_Remod\_Add : num [1:1642] 1970 1968 1970 1971 1971 ...  
## $ Roof\_Style : Factor w/ 6 levels "Flat","Gable",..: 2 2 2 2 2 2 4 2 2 2 ...  
## $ Roof\_Matl : Factor w/ 6 levels "CompShg","Metal",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Exterior\_1st : Factor w/ 16 levels "AsbShng","AsphShn",..: 15 14 10 7 7 10 7 9 9 9 ...  
## $ Exterior\_2nd : Factor w/ 17 levels "AsbShng","AsphShn",..: 16 15 11 7 8 11 11 9 9 9 ...  
## $ Mas\_Vnr\_Type : Factor w/ 5 levels "BrkCmn","BrkFace",..: 2 4 4 2 2 4 4 4 2 4 ...  
## $ Mas\_Vnr\_Area : num [1:1642] 180 0 0 504 381 0 0 0 120 0 ...  
## $ Exter\_Qual : Factor w/ 4 levels "Excellent","Fair",..: 4 4 4 4 4 4 4 4 3 3 ...  
## $ Exter\_Cond : Factor w/ 4 levels "Excellent","Fair",..: 4 4 4 4 4 4 4 3 4 4 ...  
## $ Foundation : Factor w/ 6 levels "BrkTil","CBlock",..: 2 2 2 2 2 2 3 2 3 3 ...  
## $ Bsmt\_Qual : Factor w/ 6 levels "Excellent","Fair",..: 6 6 6 6 6 3 3 6 3 3 ...  
## $ Bsmt\_Cond : Factor w/ 6 levels "Excellent","Fair",..: 6 6 6 6 6 6 6 6 6 6 ...  
## $ Bsmt\_Exposure : Factor w/ 5 levels "Av","Gd","Mn",..: 4 4 4 4 4 4 4 4 1 4 ...  
## $ BsmtFin\_Type\_1 : Factor w/ 7 levels "ALQ","BLQ","GLQ",..: 1 2 1 6 7 7 3 3 3 7 ...  
## $ BsmtFin\_SF\_1 : num [1:1642] 1 2 1 6 7 7 3 3 3 7 ...  
## $ BsmtFin\_Type\_2 : Factor w/ 7 levels "ALQ","BLQ","GLQ",..: 7 7 6 7 7 7 7 2 7 7 ...  
## $ BsmtFin\_SF\_2 : num [1:1642] 0 0 78 0 0 0 0 117 0 0 ...  
## $ Bsmt\_Unf\_SF : num [1:1642] 426 281 0 327 525 918 702 224 320 600 ...  
## $ Total\_Bsmt\_SF : num [1:1642] 1004 1056 882 483 525 ...  
## $ Heating : Factor w/ 6 levels "Floor","GasA",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ Heating\_QC : Factor w/ 4 levels "Excellent","Fair",..: 1 1 4 4 4 4 4 1 1 1 ...  
## $ Central\_Air : Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 2 ...  
## $ Electrical : Factor w/ 5 levels "FuseA","FuseF",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ First\_Flr\_SF : num [1:1642] 1004 1056 882 483 525 ...  
## $ Second\_Flr\_SF : num [1:1642] 0 0 0 504 567 0 0 0 600 600 ...  
## $ Low\_Qual\_Fin\_SF : num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Gr\_Liv\_Area : num [1:1642] 1004 1056 882 987 1092 ...  
## $ Bsmt\_Full\_Bath : num [1:1642] 1 1 1 0 0 0 0 1 0 0 ...  
## $ Bsmt\_Half\_Bath : num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Full\_Bath : num [1:1642] 1 1 1 1 1 2 2 1 2 2 ...  
## $ Half\_Bath : num [1:1642] 0 0 0 1 1 0 0 0 1 1 ...  
## $ Bedroom\_AbvGr : num [1:1642] 2 3 2 2 3 2 3 3 2 2 ...  
## $ Kitchen\_AbvGr : num [1:1642] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Kitchen\_Qual : Factor w/ 5 levels "Excellent","Fair",..: 5 5 5 5 5 5 5 5 3 3 ...  
## $ TotRms\_AbvGrd : num [1:1642] 5 6 4 5 6 5 6 5 4 4 ...  
## $ Functional : Factor w/ 8 levels "Maj1","Maj2",..: 8 8 8 8 8 8 8 8 8 8 ...  
## $ Fireplaces : num [1:1642] 1 1 0 0 0 1 0 1 0 0 ...  
## $ Fireplace\_Qu : Factor w/ 6 levels "Excellent","Fair",..: 6 2 4 4 4 6 4 3 4 4 ...  
## $ Garage\_Type : Factor w/ 7 levels "Attchd","Basment",..: 1 1 1 5 5 1 1 5 5 5 ...  
## $ Garage\_Finish : Factor w/ 4 levels "Fin","No\_Garage",..: 1 4 1 4 4 4 4 4 3 3 ...  
## $ Garage\_Cars : num [1:1642] 2 1 2 1 1 1 2 1 2 2 ...  
## $ Garage\_Area : num [1:1642] 480 304 525 264 264 264 480 336 480 480 ...  
## $ Garage\_Qual : Factor w/ 6 levels "Excellent","Fair",..: 6 6 6 6 6 6 6 6 6 6 ...  
## $ Garage\_Cond : Factor w/ 5 levels "Fair","Good",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ Paved\_Drive : Factor w/ 3 levels "Dirt\_Gravel",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ Wood\_Deck\_SF : num [1:1642] 0 0 240 275 0 28 0 416 0 0 ...  
## $ Open\_Porch\_SF : num [1:1642] 0 85 0 0 0 0 0 144 172 172 ...  
## $ Enclosed\_Porch : num [1:1642] 0 184 0 0 0 0 0 0 0 0 ...  
## $ Three\_season\_porch: num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Screen\_Porch : num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_Area : num [1:1642] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_QC : Factor w/ 5 levels "Excellent","Fair",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ Fence : Factor w/ 5 levels "Good\_Privacy",..: 5 5 3 5 5 5 1 3 5 5 ...  
## $ Misc\_Feature : Factor w/ 5 levels "Elev","Gar2",..: 5 3 3 3 3 3 3 3 3 3 ...  
## $ Misc\_Val : num [1:1642] 700 0 0 0 0 0 0 0 0 0 ...  
## $ Mo\_Sold : num [1:1642] 3 7 4 2 3 4 5 4 4 6 ...  
## $ Year\_Sold : num [1:1642] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ Sale\_Type : Factor w/ 9 levels "COD","Con","ConLD",..: 9 9 9 1 9 9 9 9 3 9 ...  
## $ Sale\_Condition : Factor w/ 6 levels "Abnorml","AdjLand",..: 5 5 5 5 4 5 5 5 5 5 ...  
## $ Longitude : num [1:1642] -93.6 -93.6 -93.6 -93.6 -93.6 ...  
## $ Latitude : num [1:1642] 42.1 42.1 42.1 42.1 42.1 ...  
## $ Above\_Median : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...

Predicting log odds of Neighborhood corresponding to Above\_Median Some neighborhoods are very statistically significant

ames\_n\_model =   
 logistic\_reg() %>%   
 set\_engine("glm")   
  
ames\_n\_recipe = recipe(Above\_Median ~ Neighborhood, train\_train) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes())   
  
logreg\_wf = workflow() %>%  
 add\_recipe(ames\_n\_recipe) %>%   
 add\_model(ames\_n\_model)  
  
ames\_n\_fit = fit(logreg\_wf, train\_train)  
summary(ames\_n\_fit$fit$fit$fit)

##   
## Call:  
## stats::glm(formula = ..y ~ ., family = stats::binomial, data = data)  
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 2.7726 1.0308  
## Neighborhood\_Blueste -20.3387 1978.0904  
## Neighborhood\_Briardale -20.3387 1021.4819  
## Neighborhood\_Brookside -4.6007 1.0915  
## Neighborhood\_Clear\_Creek -1.3863 1.1273  
## Neighborhood\_College\_Creek -1.6841 1.0502  
## Neighborhood\_Crawford -1.6305 1.0726  
## Neighborhood\_Edwards -4.3931 1.0626  
## Neighborhood\_Gilbert 0.5715 1.1864  
## Neighborhood\_Green\_Hills 14.7935 3956.1805  
## Neighborhood\_Greens -0.9808 1.4930  
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road -5.9081 1.2587  
## Neighborhood\_Landmark -20.3387 3956.1805  
## Neighborhood\_Meadow\_Village -20.3387 989.0456  
## Neighborhood\_Mitchell -2.8978 1.0608  
## Neighborhood\_North\_Ames -4.1688 1.0428  
## Neighborhood\_Northpark\_Villa -20.3387 1142.0514  
## Neighborhood\_Northridge 14.7935 603.3127  
## Neighborhood\_Northridge\_Heights 1.6582 1.4403  
## Neighborhood\_Northwest\_Ames -1.5358 1.0650  
## Neighborhood\_Old\_Town -4.9698 1.0661  
## Neighborhood\_Sawyer -4.2341 1.0675  
## Neighborhood\_Sawyer\_West -2.0035 1.0675  
## Neighborhood\_Somerset -0.2412 1.1031  
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -3.8712 1.1194  
## Neighborhood\_Stone\_Brook 14.7935 710.5517  
## Neighborhood\_Timberland 1.0986 1.4434  
## Neighborhood\_Veenker -0.2076 1.4627  
## z value Pr(>|z|)   
## (Intercept) 2.690 0.007149 \*\*   
## Neighborhood\_Blueste -0.010 0.991796   
## Neighborhood\_Briardale -0.020 0.984114   
## Neighborhood\_Brookside -4.215 2.50e-05 \*\*\*  
## Neighborhood\_Clear\_Creek -1.230 0.218797   
## Neighborhood\_College\_Creek -1.604 0.108810   
## Neighborhood\_Crawford -1.520 0.128474   
## Neighborhood\_Edwards -4.134 3.56e-05 \*\*\*  
## Neighborhood\_Gilbert 0.482 0.630049   
## Neighborhood\_Green\_Hills 0.004 0.997016   
## Neighborhood\_Greens -0.657 0.511223   
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road -4.694 2.68e-06 \*\*\*  
## Neighborhood\_Landmark -0.005 0.995898   
## Neighborhood\_Meadow\_Village -0.021 0.983594   
## Neighborhood\_Mitchell -2.732 0.006300 \*\*   
## Neighborhood\_North\_Ames -3.998 6.40e-05 \*\*\*  
## Neighborhood\_Northpark\_Villa -0.018 0.985791   
## Neighborhood\_Northridge 0.025 0.980437   
## Neighborhood\_Northridge\_Heights 1.151 0.249599   
## Neighborhood\_Northwest\_Ames -1.442 0.149270   
## Neighborhood\_Old\_Town -4.662 3.14e-06 \*\*\*  
## Neighborhood\_Sawyer -3.966 7.29e-05 \*\*\*  
## Neighborhood\_Sawyer\_West -1.877 0.060545 .   
## Neighborhood\_Somerset -0.219 0.826937   
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -3.458 0.000543 \*\*\*  
## Neighborhood\_Stone\_Brook 0.021 0.983389   
## Neighborhood\_Timberland 0.761 0.446573   
## Neighborhood\_Veenker -0.142 0.887113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2275.9 on 1641 degrees of freedom  
## Residual deviance: 1241.5 on 1614 degrees of freedom  
## AIC: 1297.5  
##   
## Number of Fisher Scoring iterations: 16

Adding Year\_Built lowered my AIC from 1297.5 to 1254.4 Year built is statistically significant

ames\_ny\_model =   
 logistic\_reg() %>%   
 set\_engine("glm")   
  
ames\_ny\_recipe = recipe(Above\_Median ~ Neighborhood + Year\_Built, train\_train) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes())  
  
logreg\_wf = workflow() %>%  
 add\_recipe(ames\_ny\_recipe) %>%   
 add\_model(ames\_ny\_model)  
  
ames\_ny\_fit = fit(logreg\_wf, train\_train)  
summary(ames\_ny\_fit$fit$fit$fit)

##   
## Call:  
## stats::glm(formula = ..y ~ ., family = stats::binomial, data = data)  
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) -6.125e+01 1.012e+01  
## Year\_Built 3.193e-02 5.022e-03  
## Neighborhood\_Blueste -1.954e+01 1.978e+03  
## Neighborhood\_Briardale -1.927e+01 1.021e+03  
## Neighborhood\_Brookside -2.256e+00 1.151e+00  
## Neighborhood\_Clear\_Creek -3.785e-02 1.156e+00  
## Neighborhood\_College\_Creek -1.425e+00 1.052e+00  
## Neighborhood\_Crawford 3.942e-01 1.128e+00  
## Neighborhood\_Edwards -3.153e+00 1.078e+00  
## Neighborhood\_Gilbert 8.334e-01 1.190e+00  
## Neighborhood\_Green\_Hills 1.540e+01 3.956e+03  
## Neighborhood\_Greens -1.654e-01 1.499e+00  
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road -3.528e+00 1.310e+00  
## Neighborhood\_Landmark -1.995e+01 3.956e+03  
## Neighborhood\_Meadow\_Village -1.927e+01 9.887e+02  
## Neighborhood\_Mitchell -2.098e+00 1.070e+00  
## Neighborhood\_North\_Ames -2.716e+00 1.067e+00  
## Neighborhood\_Northpark\_Villa -1.941e+01 1.142e+03  
## Neighborhood\_Northridge 1.509e+01 6.029e+02  
## Neighborhood\_Northridge\_Heights 1.647e+00 1.440e+00  
## Neighborhood\_Northwest\_Ames -5.778e-01 1.076e+00  
## Neighborhood\_Old\_Town -2.534e+00 1.126e+00  
## Neighborhood\_Sawyer -2.945e+00 1.086e+00  
## Neighborhood\_Sawyer\_West -1.503e+00 1.073e+00  
## Neighborhood\_Somerset -2.186e-01 1.103e+00  
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -1.607e+00 1.181e+00  
## Neighborhood\_Stone\_Brook 1.495e+01 7.066e+02  
## Neighborhood\_Timberland 1.526e+00 1.450e+00  
## Neighborhood\_Veenker 6.133e-01 1.469e+00  
## z value Pr(>|z|)   
## (Intercept) -6.050 1.45e-09 \*\*\*  
## Year\_Built 6.357 2.06e-10 \*\*\*  
## Neighborhood\_Blueste -0.010 0.99212   
## Neighborhood\_Briardale -0.019 0.98495   
## Neighborhood\_Brookside -1.961 0.04994 \*   
## Neighborhood\_Clear\_Creek -0.033 0.97389   
## Neighborhood\_College\_Creek -1.355 0.17553   
## Neighborhood\_Crawford 0.349 0.72677   
## Neighborhood\_Edwards -2.924 0.00345 \*\*   
## Neighborhood\_Gilbert 0.701 0.48358   
## Neighborhood\_Green\_Hills 0.004 0.99689   
## Neighborhood\_Greens -0.110 0.91214   
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road -2.693 0.00709 \*\*   
## Neighborhood\_Landmark -0.005 0.99598   
## Neighborhood\_Meadow\_Village -0.019 0.98445   
## Neighborhood\_Mitchell -1.961 0.04983 \*   
## Neighborhood\_North\_Ames -2.546 0.01090 \*   
## Neighborhood\_Northpark\_Villa -0.017 0.98644   
## Neighborhood\_Northridge 0.025 0.98003   
## Neighborhood\_Northridge\_Heights 1.143 0.25295   
## Neighborhood\_Northwest\_Ames -0.537 0.59137   
## Neighborhood\_Old\_Town -2.251 0.02440 \*   
## Neighborhood\_Sawyer -2.712 0.00669 \*\*   
## Neighborhood\_Sawyer\_West -1.401 0.16125   
## Neighborhood\_Somerset -0.198 0.84289   
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -1.361 0.17359   
## Neighborhood\_Stone\_Brook 0.021 0.98312   
## Neighborhood\_Timberland 1.053 0.29256   
## Neighborhood\_Veenker 0.417 0.67642   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2275.9 on 1641 degrees of freedom  
## Residual deviance: 1196.4 on 1613 degrees of freedom  
## AIC: 1254.4  
##   
## Number of Fisher Scoring iterations: 16

Adding Lot information lowered AIC amount to 1180.3 Lot is also statistically significant

ames\_nyl\_model =   
 logistic\_reg() %>%   
 set\_engine("glm")   
  
ames\_nyl\_recipe = recipe(Above\_Median ~ Neighborhood + Year\_Built + Lot\_Area, train\_train) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes())  
  
logreg\_wf = workflow() %>%  
 add\_recipe(ames\_nyl\_recipe) %>%   
 add\_model(ames\_nyl\_model)  
  
ames\_nyl\_fit = fit(logreg\_wf, train\_train)  
summary(ames\_nyl\_fit$fit$fit$fit)

##   
## Call:  
## stats::glm(formula = ..y ~ ., family = stats::binomial, data = data)  
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) -8.459e+01 1.121e+01  
## Year\_Built 4.331e-02 5.559e-03  
## Lot\_Area 1.573e-04 2.146e-05  
## Neighborhood\_Blueste -1.898e+01 1.978e+03  
## Neighborhood\_Briardale -1.866e+01 1.021e+03  
## Neighborhood\_Brookside -2.062e+00 1.161e+00  
## Neighborhood\_Clear\_Creek -1.388e+00 1.175e+00  
## Neighborhood\_College\_Creek -2.363e+00 1.060e+00  
## Neighborhood\_Crawford -7.464e-02 1.144e+00  
## Neighborhood\_Edwards -3.848e+00 1.089e+00  
## Neighborhood\_Gilbert -2.473e-01 1.195e+00  
## Neighborhood\_Green\_Hills 1.484e+01 3.956e+03  
## Neighborhood\_Greens 3.208e-02 1.500e+00  
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road -4.260e+00 1.450e+00  
## Neighborhood\_Landmark -1.987e+01 3.956e+03  
## Neighborhood\_Meadow\_Village -1.869e+01 9.867e+02  
## Neighborhood\_Mitchell -3.309e+00 1.084e+00  
## Neighborhood\_North\_Ames -3.329e+00 1.074e+00  
## Neighborhood\_Northpark\_Villa -1.905e+01 1.139e+03  
## Neighborhood\_Northridge 1.378e+01 5.844e+02  
## Neighborhood\_Northridge\_Heights 5.991e-01 1.447e+00  
## Neighborhood\_Northwest\_Ames -1.503e+00 1.085e+00  
## Neighborhood\_Old\_Town -2.584e+00 1.133e+00  
## Neighborhood\_Sawyer -3.708e+00 1.096e+00  
## Neighborhood\_Sawyer\_West -2.370e+00 1.082e+00  
## Neighborhood\_Somerset -7.042e-01 1.106e+00  
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -1.492e+00 1.192e+00  
## Neighborhood\_Stone\_Brook 1.405e+01 6.393e+02  
## Neighborhood\_Timberland 1.088e-01 1.455e+00  
## Neighborhood\_Veenker -7.023e-01 1.485e+00  
## z value Pr(>|z|)   
## (Intercept) -7.543 4.58e-14 \*\*\*  
## Year\_Built 7.792 6.61e-15 \*\*\*  
## Lot\_Area 7.332 2.27e-13 \*\*\*  
## Neighborhood\_Blueste -0.010 0.992343   
## Neighborhood\_Briardale -0.018 0.985418   
## Neighborhood\_Brookside -1.777 0.075557 .   
## Neighborhood\_Clear\_Creek -1.182 0.237235   
## Neighborhood\_College\_Creek -2.229 0.025788 \*   
## Neighborhood\_Crawford -0.065 0.947992   
## Neighborhood\_Edwards -3.534 0.000409 \*\*\*  
## Neighborhood\_Gilbert -0.207 0.836124   
## Neighborhood\_Green\_Hills 0.004 0.997007   
## Neighborhood\_Greens 0.021 0.982936   
## Neighborhood\_Iowa\_DOT\_and\_Rail\_Road -2.937 0.003313 \*\*   
## Neighborhood\_Landmark -0.005 0.995994   
## Neighborhood\_Meadow\_Village -0.019 0.984889   
## Neighborhood\_Mitchell -3.053 0.002269 \*\*   
## Neighborhood\_North\_Ames -3.100 0.001933 \*\*   
## Neighborhood\_Northpark\_Villa -0.017 0.986659   
## Neighborhood\_Northridge 0.024 0.981180   
## Neighborhood\_Northridge\_Heights 0.414 0.678890   
## Neighborhood\_Northwest\_Ames -1.385 0.166127   
## Neighborhood\_Old\_Town -2.282 0.022512 \*   
## Neighborhood\_Sawyer -3.385 0.000712 \*\*\*  
## Neighborhood\_Sawyer\_West -2.190 0.028526 \*   
## Neighborhood\_Somerset -0.637 0.524419   
## Neighborhood\_South\_and\_West\_of\_Iowa\_State\_University -1.252 0.210628   
## Neighborhood\_Stone\_Brook 0.022 0.982469   
## Neighborhood\_Timberland 0.075 0.940420   
## Neighborhood\_Veenker -0.473 0.636315   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2275.9 on 1641 degrees of freedom  
## Residual deviance: 1120.3 on 1612 degrees of freedom  
## AIC: 1180.3  
##   
## Number of Fisher Scoring iterations: 16

#Predictions#

*There is a 79% chance a home in the North Ames neighborhood built in 1961 sold below median*

newdata = data.frame(Neighborhood = "North\_Ames", Year\_Built = 1961)  
predict(ames\_ny\_fit, newdata, type="prob")

## # A tibble: 1 × 2  
## .pred\_No .pred\_Yes  
## <dbl> <dbl>  
## 1 0.794 0.206

*There is a 50/50 chance on whether a home built in the year 2000 in the Edwards neighborhood with a large lot area will sell above median*

newdata = data.frame(Neighborhood = "Edwards", Year\_Built = 2000, Lot\_Area = 11679)  
predict(ames\_nyl\_fit, newdata, type="prob")

## # A tibble: 1 × 2  
## .pred\_No .pred\_Yes  
## <dbl> <dbl>  
## 1 0.494 0.506

*There is a 95% chance that a home in the Bloomington Heights built in 2007 with a lot area between 3000- 3500 will sell above median*

newdata = data.frame(Neighborhood = "Bloomington\_Heights", Year\_Built = 2007, Lot\_Area = 3000-3500)  
predict(ames\_nyl\_fit, newdata, type="prob")

## # A tibble: 1 × 2  
## .pred\_No .pred\_Yes  
## <dbl> <dbl>  
## 1 0.0946 0.905

Classification Tree by Lot

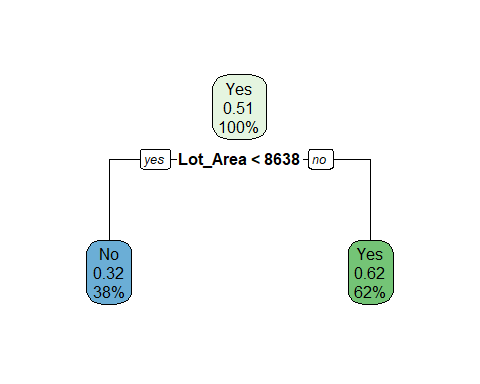
ames\_lrecipe = recipe(Above\_Median ~ Lot\_Area, train\_train)  
  
tree\_model = decision\_tree() %>%   
 set\_engine("rpart", model = TRUE) %>%   
 set\_mode("classification")  
  
ames\_wflow =   
 workflow() %>%   
 add\_model(tree\_model) %>%   
 add\_recipe(ames\_lrecipe)  
  
ames\_lfit = fit(ames\_wflow, train\_train)

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

## Warning: `pull\_workflow\_fit()` was deprecated in workflows 0.2.3.  
## ℹ Please use `extract\_fit\_parsnip()` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

## n= 1642   
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 1642 808 Yes (0.4920828 0.5079172)   
## 2) Lot\_Area< 8637.5 621 199 No (0.6795491 0.3204509) \*  
## 3) Lot\_Area>=8637.5 1021 386 Yes (0.3780607 0.6219393) \*

tree = ames\_lfit %>%   
 pull\_workflow\_fit() %>%   
 pluck("fit")  
   
rpart.plot(tree)



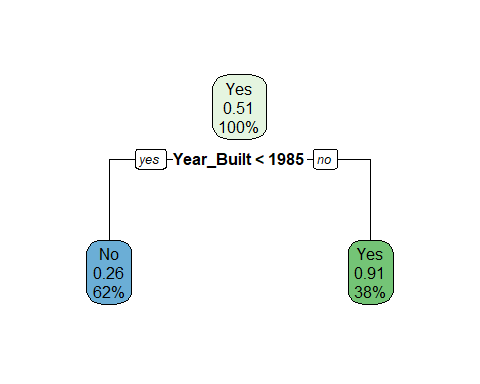
Classification Tree by Year Built

ames\_yrecipe = recipe(Above\_Median ~ Year\_Built, train\_train)  
  
tree\_model = decision\_tree() %>%   
 set\_engine("rpart", model = TRUE) %>%   
 set\_mode("classification")  
  
ames\_wflow =   
 workflow() %>%   
 add\_model(tree\_model) %>%   
 add\_recipe(ames\_yrecipe)  
  
ames\_yfit = fit(ames\_wflow, train\_train)

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

## n= 1642   
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 1642 808 Yes (0.4920828 0.5079172)   
## 2) Year\_Built< 1984.5 1015 261 No (0.7428571 0.2571429) \*  
## 3) Year\_Built>=1984.5 627 54 Yes (0.0861244 0.9138756) \*

tree = ames\_yfit %>%   
 pull\_workflow\_fit() %>%   
 pluck("fit")  
   
rpart.plot(tree)



Random Forests and Performance on Testing and Training

train\_recipe = recipe(Above\_Median ~., train\_train) %>%   
 step\_dummy(all\_nominal(), -all\_outcomes())  
  
rf\_model = rand\_forest() %>%   
 set\_engine("ranger") %>%   
 set\_mode("classification")   
  
ames\_wflow =   
 workflow() %>%   
 add\_model(rf\_model) %>%   
 add\_recipe(train\_recipe)  
  
set.seed(123)  
train\_fit = fit(ames\_wflow, train\_train)  
train\_fit

## ══ Workflow [trained] ══════════════════════════════════════════════════════════  
## Preprocessor: Recipe  
## Model: rand\_forest()  
##   
## ── Preprocessor ────────────────────────────────────────────────────────────────  
## 1 Recipe Step  
##   
## • step\_dummy()  
##   
## ── Model ───────────────────────────────────────────────────────────────────────  
## Ranger result  
##   
## Call:  
## ranger::ranger(x = maybe\_data\_frame(x), y = y, num.threads = 1, verbose = FALSE, seed = sample.int(10^5, 1), probability = TRUE)   
##   
## Type: Probability estimation   
## Number of trees: 500   
## Sample size: 1642   
## Number of independent variables: 297   
## Mtry: 17   
## Target node size: 10   
## Variable importance mode: none   
## Splitrule: gini   
## OOB prediction error (Brier s.): 0.06288818

predRF = predict(train\_fit, train\_train)  
head(predRF)

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

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

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 800 7  
## Yes 8 827  
##   
## Accuracy : 0.9909   
## 95% CI : (0.985, 0.9949)  
## No Information Rate : 0.5079   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.9817   
##   
## Mcnemar's Test P-Value : 1   
##   
## Sensitivity : 0.9916   
## Specificity : 0.9901   
## Pos Pred Value : 0.9904   
## Neg Pred Value : 0.9913   
## Prevalence : 0.5079   
## Detection Rate : 0.5037   
## Detection Prevalence : 0.5085   
## Balanced Accuracy : 0.9909   
##   
## 'Positive' Class : Yes   
##

test\_train <- test\_train %>% mutate\_if(is.character,as.factor)%>%   
 mutate(Above\_Median = as\_factor(Above\_Median)) %>%   
 mutate(Above\_Median = fct\_recode(Above\_Median, "No" = "0", "Yes" = "1" ))

## Warning: There was 1 warning in `mutate()`.  
## ℹ In argument: `Above\_Median = fct\_recode(Above\_Median, No = "0", Yes = "1")`.  
## Caused by warning:  
## ! Unknown levels in `f`: 0, 1

str(test\_train)

## tibble [411 × 81] (S3: tbl\_df/tbl/data.frame)  
## $ MS\_SubClass : Factor w/ 15 levels "Duplex\_All\_Styles\_and\_Ages",..: 5 5 6 5 14 5 6 15 6 5 ...  
## $ MS\_Zoning : Factor w/ 6 levels "A\_agr","C\_all",..: 4 5 5 5 3 5 4 6 5 5 ...  
## $ Lot\_Frontage : num [1:411] 80 93 39 88 0 0 26 21 24 100 ...  
## $ Lot\_Area : num [1:411] 11622 11160 5389 11394 7500 ...  
## $ Street : Factor w/ 2 levels "Grvl","Pave": 2 2 2 2 2 2 2 2 2 2 ...  
## $ Alley : Factor w/ 3 levels "Gravel","No\_Alley\_Access",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ Lot\_Shape : Factor w/ 4 levels "Irregular","Moderately\_Irregular",..: 3 3 4 3 3 4 4 3 3 4 ...  
## $ Land\_Contour : Factor w/ 4 levels "Bnk","HLS","Low",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ Utilities : Factor w/ 1 level "AllPub": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Lot\_Config : Factor w/ 5 levels "Corner","CulDSac",..: 5 1 5 1 5 2 3 5 3 5 ...  
## $ Land\_Slope : Factor w/ 3 levels "Gtl","Mod","Sev": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Neighborhood : Factor w/ 25 levels "Bloomington\_Heights",..: 13 13 23 23 21 13 13 3 14 16 ...  
## $ Condition\_1 : Factor w/ 9 levels "Artery","Feedr",..: 2 3 3 3 3 3 3 3 3 3 ...  
## $ Condition\_2 : Factor w/ 5 levels "Artery","Feedr",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ Bldg\_Type : Factor w/ 5 levels "Duplex","OneFam",..: 2 2 4 2 2 2 4 3 3 2 ...  
## $ House\_Style : Factor w/ 7 levels "One\_and\_Half\_Fin",..: 3 3 3 3 7 3 3 7 3 3 ...  
## $ Overall\_Qual : Factor w/ 9 levels "Above\_Average",..: 2 6 8 4 6 2 6 2 6 4 ...  
## $ Overall\_Cond : Factor w/ 9 levels "Above\_Average",..: 1 2 2 7 2 1 2 2 1 2 ...  
## $ Year\_Built : num [1:411] 1961 1968 1995 2010 2000 ...  
## $ Year\_Remod\_Add : num [1:411] 1961 1968 1996 2010 2000 ...  
## $ Roof\_Style : Factor w/ 5 levels "Flat","Gable",..: 2 4 2 4 2 2 2 2 2 4 ...  
## $ Roof\_Matl : Factor w/ 3 levels "CompShg","Tar&Grv",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Exterior\_1st : Factor w/ 10 levels "AsbShng","BrkFace",..: 8 2 3 8 8 8 5 4 6 8 ...  
## $ Exterior\_2nd : Factor w/ 12 levels "AsbShng","Brk Cmn",..: 10 3 4 10 10 10 7 5 2 10 ...  
## $ Mas\_Vnr\_Type : Factor w/ 4 levels "BrkCmn","BrkFace",..: 3 3 3 4 3 3 3 2 3 4 ...  
## $ Mas\_Vnr\_Area : num [1:411] 0 0 0 350 0 0 0 492 0 760 ...  
## $ Exter\_Qual : Factor w/ 4 levels "Excellent","Fair",..: 4 3 3 3 3 4 3 4 4 1 ...  
## $ Exter\_Cond : Factor w/ 5 levels "Excellent","Fair",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ Foundation : Factor w/ 6 levels "BrkTil","CBlock",..: 2 2 3 3 3 2 3 2 2 3 ...  
## $ Bsmt\_Qual : Factor w/ 5 levels "Excellent","Fair",..: 5 5 3 1 3 5 3 5 3 1 ...  
## $ Bsmt\_Cond : Factor w/ 4 levels "Fair","Good",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ Bsmt\_Exposure : Factor w/ 5 levels "Av","Gd","Mn",..: 4 4 4 1 4 4 4 4 4 2 ...  
## $ BsmtFin\_Type\_1 : Factor w/ 7 levels "ALQ","BLQ","GLQ",..: 6 1 3 3 3 3 3 6 7 3 ...  
## $ BsmtFin\_SF\_1 : num [1:411] 6 1 3 3 3 3 3 6 7 3 ...  
## $ BsmtFin\_Type\_2 : Factor w/ 7 levels "ALQ","BLQ","GLQ",..: 4 7 7 7 7 7 2 7 7 7 ...  
## $ BsmtFin\_SF\_2 : num [1:411] 144 0 0 0 0 0 0 0 0 0 ...  
## $ Bsmt\_Unf\_SF : num [1:411] 270 1045 415 411 281 ...  
## $ Total\_Bsmt\_SF : num [1:411] 882 2110 1595 1856 814 ...  
## $ Heating : Factor w/ 4 levels "GasA","GasW",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Heating\_QC : Factor w/ 5 levels "Excellent","Fair",..: 5 1 1 1 1 1 1 5 1 1 ...  
## $ Central\_Air : Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 2 ...  
## $ Electrical : Factor w/ 4 levels "FuseA","FuseF",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ First\_Flr\_SF : num [1:411] 896 2110 1616 1856 814 ...  
## $ Second\_Flr\_SF : num [1:411] 0 0 0 0 860 0 0 567 0 0 ...  
## $ Low\_Qual\_Fin\_SF : num [1:411] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Gr\_Liv\_Area : num [1:411] 896 2110 1616 1856 1674 ...  
## $ Bsmt\_Full\_Bath : num [1:411] 0 1 1 1 1 1 1 0 0 1 ...  
## $ Bsmt\_Half\_Bath : num [1:411] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Full\_Bath : num [1:411] 1 2 2 1 2 1 2 1 1 2 ...  
## $ Half\_Bath : num [1:411] 0 1 0 1 1 1 0 1 0 1 ...  
## $ Bedroom\_AbvGr : num [1:411] 2 3 2 1 3 3 2 3 2 2 ...  
## $ Kitchen\_AbvGr : num [1:411] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Kitchen\_Qual : Factor w/ 4 levels "Excellent","Fair",..: 4 1 3 1 3 4 3 4 4 1 ...  
## $ TotRms\_AbvGrd : num [1:411] 5 8 5 8 7 6 5 6 4 11 ...  
## $ Functional : Factor w/ 8 levels "Maj1","Maj2",..: 8 8 8 8 8 8 8 8 8 8 ...  
## $ Fireplaces : num [1:411] 0 2 1 1 0 1 1 0 0 2 ...  
## $ Fireplace\_Qu : Factor w/ 6 levels "Excellent","Fair",..: 4 6 6 1 4 2 2 4 4 3 ...  
## $ Garage\_Type : Factor w/ 7 levels "Attchd","Basment",..: 1 1 1 1 1 1 1 5 1 1 ...  
## $ Garage\_Finish : Factor w/ 4 levels "Fin","No\_Garage",..: 4 1 3 1 3 1 1 4 4 1 ...  
## $ Garage\_Cars : num [1:411] 1 2 2 3 2 2 2 1 1 3 ...  
## $ Garage\_Area : num [1:411] 730 522 608 834 663 500 511 320 308 820 ...  
## $ Garage\_Qual : Factor w/ 5 levels "Excellent","Fair",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ Garage\_Cond : Factor w/ 5 levels "Excellent","Fair",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ Paved\_Drive : Factor w/ 3 levels "Dirt\_Gravel",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ Wood\_Deck\_SF : num [1:411] 140 0 237 113 0 0 203 0 0 0 ...  
## $ Open\_Porch\_SF : num [1:411] 0 0 152 0 96 0 68 0 30 67 ...  
## $ Enclosed\_Porch : num [1:411] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Three\_season\_porch: num [1:411] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Screen\_Porch : num [1:411] 120 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_Area : num [1:411] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Pool\_QC : Factor w/ 2 levels "No\_Pool","Typical": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Fence : Factor w/ 5 levels "Good\_Privacy",..: 3 5 5 5 5 5 5 5 5 5 ...  
## $ Misc\_Feature : Factor w/ 3 levels "None","Othr",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Misc\_Val : num [1:411] 0 0 0 0 0 0 0 0 0 0 ...  
## $ Mo\_Sold : num [1:411] 6 4 3 6 1 4 6 3 6 3 ...  
## $ Year\_Sold : num [1:411] 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...  
## $ Sale\_Type : Factor w/ 9 levels "COD","Con","ConLD",..: 9 9 9 6 9 9 9 9 9 6 ...  
## $ Sale\_Condition : Factor w/ 6 levels "Abnorml","AdjLand",..: 5 5 5 6 5 5 5 5 5 6 ...  
## $ Longitude : num [1:411] -93.6 -93.6 -93.6 -93.6 -93.6 ...  
## $ Latitude : num [1:411] 42.1 42.1 42.1 42.1 42 ...  
## $ Above\_Median : Factor w/ 2 levels "No","Yes": 1 2 2 2 2 1 2 1 1 2 ...

test\_recipe = recipe(Above\_Median ~., test\_train) %>%   
 step\_dummy()  
  
rf\_model = rand\_forest() %>%   
 set\_engine("ranger") %>%   
 set\_mode("classification")   
  
ames\_wflow =   
 workflow() %>%   
 add\_model(rf\_model) %>%   
 add\_recipe(test\_recipe)  
  
set.seed(123)  
test\_fit = fit(ames\_wflow, test\_train)  
test\_fit

## ══ Workflow [trained] ══════════════════════════════════════════════════════════  
## Preprocessor: Recipe  
## Model: rand\_forest()  
##   
## ── Preprocessor ────────────────────────────────────────────────────────────────  
## 1 Recipe Step  
##   
## • step\_dummy()  
##   
## ── Model ───────────────────────────────────────────────────────────────────────  
## Ranger result  
##   
## Call:  
## ranger::ranger(x = maybe\_data\_frame(x), y = y, num.threads = 1, verbose = FALSE, seed = sample.int(10^5, 1), probability = TRUE)   
##   
## Type: Probability estimation   
## Number of trees: 500   
## Sample size: 411   
## Number of independent variables: 80   
## Mtry: 8   
## Target node size: 10   
## Variable importance mode: none   
## Splitrule: gini   
## OOB prediction error (Brier s.): 0.06677222

predRF = predict(test\_fit, test\_train)  
head(predRF)

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

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

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction No Yes  
## No 200 0  
## Yes 2 209  
##   
## Accuracy : 0.9951   
## 95% CI : (0.9825, 0.9994)  
## No Information Rate : 0.5085   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.9903   
##   
## Mcnemar's Test P-Value : 0.4795   
##   
## Sensitivity : 1.0000   
## Specificity : 0.9901   
## Pos Pred Value : 0.9905   
## Neg Pred Value : 1.0000   
## Prevalence : 0.5085   
## Detection Rate : 0.5085   
## Detection Prevalence : 0.5134   
## Balanced Accuracy : 0.9950   
##   
## 'Positive' Class : Yes   
##