# BAN 502

## Jacqulyn Leer and Jennafer Hinman

### Course Project Phase 1

Reading in the dataset.

ames <- 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.

Getting rid of unnecessary first column “X1” that just numbers the rows within the dataset as well as latitude and longitude.

ames = ames[-c(1, 80, 81)]

Convert the characters as well as the columns for years and months in the dataset to factors.

ames = ames %>%   
 mutate\_if(is.character, as\_factor) %>%  
 mutate(Year\_Built = as\_factor(Year\_Built)) %>%  
 mutate(Year\_Remod\_Add = as\_factor(Year\_Remod\_Add)) %>%  
 mutate(Year\_Sold = as\_factor(Year\_Sold)) %>%  
 mutate(Mo\_Sold = as\_factor(Mo\_Sold))

Look for any missing data

summary(ames)

## MS\_SubClass MS\_Zoning   
## One\_Story\_1946\_and\_Newer\_All\_Styles :772 Residential\_Low\_Density :1600   
## Two\_Story\_1946\_and\_Newer :383 Residential\_High\_Density : 20   
## One\_and\_Half\_Story\_Finished\_All\_Ages:204 Floating\_Village\_Residential: 87   
## One\_Story\_PUD\_1946\_and\_Newer :129 Residential\_Medium\_Density : 326   
## One\_Story\_1945\_and\_Older : 98 C\_all : 17   
## Two\_Story\_1945\_and\_Older : 95 A\_agr : 2   
## (Other) :372 I\_all : 1   
## Lot\_Frontage Lot\_Area Street Alley   
## Min. : 0.00 Min. : 1300 Pave:2046 No\_Alley\_Access:1914   
## 1st Qu.: 43.00 1st Qu.: 7500 Grvl: 7 Paved : 45   
## Median : 62.00 Median : 9548 Gravel : 94   
## Mean : 57.38 Mean : 10258   
## 3rd Qu.: 78.00 3rd Qu.: 11600   
## Max. :313.00 Max. :215245   
##   
## Lot\_Shape Land\_Contour Utilities Lot\_Config   
## Slightly\_Irregular : 714 Lvl:1833 AllPub:2052 Corner : 359   
## Regular :1275 HLS: 94 NoSewr: 1 Inside :1495   
## Moderately\_Irregular: 53 Bnk: 81 CulDSac: 135   
## Irregular : 11 Low: 45 FR2 : 56   
## FR3 : 8   
##   
##   
## Land\_Slope Neighborhood Condition\_1 Condition\_2 Bldg\_Type   
## Gtl:1951 North\_Ames : 327 Norm :1771 Norm :2027 OneFam :1706   
## Mod: 89 College\_Creek: 183 Feedr : 113 Feedr : 12 TwnhsE : 157   
## Sev: 13 Old\_Town : 181 Artery : 67 PosA : 4 Twnhs : 67   
## Edwards : 129 RRAn : 35 Artery : 4 Duplex : 76   
## Somerset : 119 PosN : 24 PosN : 3 TwoFmCon: 47   
## Gilbert : 109 RRAe : 19 RRNn : 1   
## (Other) :1005 (Other): 24 (Other): 2   
## House\_Style Overall\_Qual Overall\_Cond   
## One\_Story :1052 Average :587 Average :1143   
## Two\_Story : 590 Above\_Average:518 Above\_Average: 376   
## One\_and\_Half\_Fin: 225 Good :411 Good : 286   
## SLvl : 90 Very\_Good :237 Very\_Good : 98   
## SFoyer : 56 Below\_Average:169 Below\_Average: 73   
## Two\_and\_Half\_Unf: 19 Excellent : 70 Fair : 35   
## (Other) : 21 (Other) : 61 (Other) : 42   
## Year\_Built Year\_Remod\_Add Roof\_Style Roof\_Matl Exterior\_1st  
## 2005 : 104 1950 : 256 Hip : 404 CompShg:2023 VinylSd:705   
## 2006 : 93 2006 : 147 Gable :1607 WdShake: 8 MetalSd:319   
## 2007 : 76 2007 : 116 Mansard: 9 Tar&Grv: 17 Wd Sdng:313   
## 2003 : 62 2005 : 94 Gambrel: 14 WdShngl: 3 HdBoard:303   
## 2004 : 60 2004 : 80 Shed : 5 Roll : 1 Plywood:151   
## 1977 : 40 2003 : 70 Flat : 14 Metal : 1 CemntBd: 90   
## (Other):1618 (Other):1290 (Other):172   
## Exterior\_2nd Mas\_Vnr\_Type Mas\_Vnr\_Area Exter\_Qual   
## VinylSd:699 Stone : 166 Min. : 0.0 Typical :1272   
## MetalSd:317 None :1231 1st Qu.: 0.0 Good : 682   
## Wd Sdng:302 BrkFace: 638 Median : 0.0 Excellent: 78   
## HdBoard:277 BrkCmn : 17 Mean : 103.8 Fair : 21   
## Plywood:190 CBlock : 1 3rd Qu.: 164.0   
## CmentBd: 90 Max. :1600.0   
## (Other):178   
## Exter\_Cond Foundation Bsmt\_Qual Bsmt\_Cond   
## Typical :1787 CBlock:880 Typical :911 Good : 80   
## Good : 213 PConc :911 Good :849 Typical :1833   
## Fair : 43 Wood : 4 Excellent :178 Poor : 4   
## Excellent: 9 BrkTil:216 No\_Basement: 57 No\_Basement: 57   
## Poor : 1 Slab : 36 Fair : 57 Fair : 76   
## Stone : 6 Poor : 1 Excellent : 3   
##   
## Bsmt\_Exposure BsmtFin\_Type\_1 BsmtFin\_SF\_1 BsmtFin\_Type\_2  
## Gd : 199 BLQ :196 Min. :1.00 Unf :1740   
## No :1331 Rec :216 1st Qu.:3.00 LwQ : 64   
## Av : 284 ALQ :298 Median :3.00 BLQ : 47   
## Mn : 179 GLQ :578 Mean :4.21 Rec : 79   
## No\_Basement: 60 Unf :602 3rd Qu.:7.00 GLQ : 23   
## LwQ :106 Max. :7.00 No\_Basement: 58   
## No\_Basement: 57 ALQ : 42   
## BsmtFin\_SF\_2 Bsmt\_Unf\_SF Total\_Bsmt\_SF Heating   
## Min. : 0.00 Min. : 0.0 Min. : 0 GasA :2019   
## 1st Qu.: 0.00 1st Qu.: 226.0 1st Qu.: 793 GasW : 21   
## Median : 0.00 Median : 460.0 Median : 988 Grav : 6   
## Mean : 52.57 Mean : 561.2 Mean :1055 Wall : 5   
## 3rd Qu.: 0.00 3rd Qu.: 801.0 3rd Qu.:1304 Floor: 1   
## Max. :1526.00 Max. :2336.0 Max. :5095 OthW : 1   
##   
## Heating\_QC Central\_Air Electrical First\_Flr\_SF Second\_Flr\_SF   
## Fair : 61 Y:1916 SBrkr :1887 Min. : 432 Min. : 0.0   
## Typical : 618 N: 137 FuseA : 126 1st Qu.: 882 1st Qu.: 0.0   
## Excellent:1040 FuseF : 33 Median :1088 Median : 0.0   
## Good : 333 FuseP : 6 Mean :1168 Mean : 326.1   
## Poor : 1 Unknown: 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   
## Typical :1070 Min. : 3.000 Typ :1896 Min. :0.000   
## Good : 790 1st Qu.: 5.000 Min2 : 54 1st Qu.:0.000   
## Excellent: 142 Median : 6.000 Min1 : 51 Median :1.000   
## Fair : 50 Mean : 6.442 Mod : 27 Mean :0.603   
## Poor : 1 3rd Qu.: 7.000 Maj1 : 15 3rd Qu.:1.000   
## Max. :15.000 Maj2 : 6 Max. :4.000   
## (Other): 4   
## Fireplace\_Qu Garage\_Type Garage\_Finish Garage\_Cars   
## Good :538 Attchd :1204 Fin :509 Min. :0.000   
## No\_Fireplace:993 BuiltIn : 127 Unf :872 1st Qu.:1.000   
## Typical :409 Basment : 29 RFn :563 Median :2.000   
## Poor : 36 Detchd : 549 No\_Garage:109 Mean :1.774   
## Excellent : 21 No\_Garage : 108 3rd Qu.:2.000   
## Fair : 56 CarPort : 15 Max. :5.000   
## More\_Than\_Two\_Types: 21   
## Garage\_Area Garage\_Qual Garage\_Cond Paved\_Drive   
## Min. : 0 Typical :1839 Typical :1872 Partial\_Pavement: 42   
## 1st Qu.: 320 No\_Garage: 109 No\_Garage: 109 Paved :1848   
## Median : 478 Fair : 85 Fair : 53 Dirt\_Gravel : 163   
## Mean : 472 Good : 16 Excellent: 1   
## 3rd Qu.: 576 Excellent: 2 Poor : 8   
## Max. :1488 Poor : 2 Good : 10   
##   
## 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 No\_Pool :2047 No\_Fence :1661   
## 1st Qu.: 0.00 1st Qu.: 0.000 Excellent: 2 Minimum\_Privacy : 225   
## Median : 0.00 Median : 0.000 Typical : 2 Good\_Privacy : 81   
## Mean : 16.68 Mean : 1.339 Fair : 1 Good\_Wood : 77   
## 3rd Qu.: 0.00 3rd Qu.: 0.000 Good : 1 Minimum\_Wood\_Wire: 9   
## Max. :576.00 Max. :800.000   
##   
## Misc\_Feature Misc\_Val Mo\_Sold Year\_Sold Sale\_Type   
## None:1978 Min. : 0.00 6 :352 2006:442 WD :1789   
## Gar2: 5 1st Qu.: 0.00 7 :320 2007:499 New : 163   
## Shed: 66 Median : 0.00 5 :275 2008:445 COD : 54   
## Othr: 3 Mean : 60.12 4 :187 2009:456 ConLD : 16   
## Elev: 1 3rd Qu.: 0.00 8 :169 2010:211 ConLI : 8   
## Max. :17000.00 3 :164 CWD : 8   
## (Other):586 (Other): 15   
## Sale\_Condition Above\_Median  
## Normal :1712 Yes:1043   
## Partial: 169 No :1010   
## Family : 30   
## Abnorml: 121   
## Alloca : 16   
## AdjLand: 5   
##

skim(ames)

Data summary

|  |  |
| --- | --- |
| Name | ames |
| Number of rows | 2053 |
| Number of columns | 79 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 51 |
| numeric | 28 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

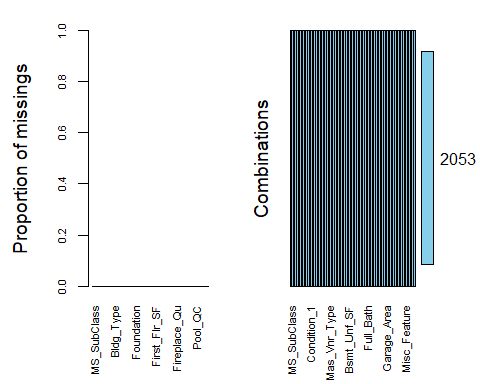
**Variable type: factor**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| MS\_SubClass | 0 | 1 | FALSE | 16 | One: 772, Two: 383, One: 204, One: 129 |
| MS\_Zoning | 0 | 1 | FALSE | 7 | Res: 1600, Res: 326, Flo: 87, Res: 20 |
| Street | 0 | 1 | FALSE | 2 | Pav: 2046, Grv: 7 |
| Alley | 0 | 1 | FALSE | 3 | No\_: 1914, Gra: 94, Pav: 45 |
| Lot\_Shape | 0 | 1 | FALSE | 4 | Reg: 1275, Sli: 714, Mod: 53, Irr: 11 |
| Land\_Contour | 0 | 1 | FALSE | 4 | Lvl: 1833, HLS: 94, Bnk: 81, Low: 45 |
| Utilities | 0 | 1 | FALSE | 2 | All: 2052, NoS: 1 |
| Lot\_Config | 0 | 1 | FALSE | 5 | Ins: 1495, Cor: 359, Cul: 135, FR2: 56 |
| Land\_Slope | 0 | 1 | FALSE | 3 | Gtl: 1951, Mod: 89, Sev: 13 |
| Neighborhood | 0 | 1 | FALSE | 28 | Nor: 327, Col: 183, Old: 181, Edw: 129 |
| Condition\_1 | 0 | 1 | FALSE | 9 | Nor: 1771, Fee: 113, Art: 67, RRA: 35 |
| Condition\_2 | 0 | 1 | FALSE | 8 | Nor: 2027, Fee: 12, Pos: 4, Art: 4 |
| Bldg\_Type | 0 | 1 | FALSE | 5 | One: 1706, Twn: 157, Dup: 76, Twn: 67 |
| House\_Style | 0 | 1 | FALSE | 8 | One: 1052, Two: 590, One: 225, SLv: 90 |
| Overall\_Qual | 0 | 1 | FALSE | 10 | Ave: 587, Abo: 518, Goo: 411, Ver: 237 |
| Overall\_Cond | 0 | 1 | FALSE | 9 | Ave: 1143, Abo: 376, Goo: 286, Ver: 98 |
| Year\_Built | 0 | 1 | FALSE | 114 | 200: 104, 200: 93, 200: 76, 200: 62 |
| Year\_Remod\_Add | 0 | 1 | FALSE | 61 | 195: 256, 200: 147, 200: 116, 200: 94 |
| Roof\_Style | 0 | 1 | FALSE | 6 | Gab: 1607, Hip: 404, Gam: 14, Fla: 14 |
| Roof\_Matl | 0 | 1 | FALSE | 6 | Com: 2023, Tar: 17, WdS: 8, WdS: 3 |
| Exterior\_1st | 0 | 1 | FALSE | 16 | Vin: 705, Met: 319, Wd : 313, HdB: 303 |
| Exterior\_2nd | 0 | 1 | FALSE | 17 | Vin: 699, Met: 317, Wd : 302, HdB: 277 |
| Mas\_Vnr\_Type | 0 | 1 | FALSE | 5 | Non: 1231, Brk: 638, Sto: 166, Brk: 17 |
| Exter\_Qual | 0 | 1 | FALSE | 4 | Typ: 1272, Goo: 682, Exc: 78, Fai: 21 |
| Exter\_Cond | 0 | 1 | FALSE | 5 | Typ: 1787, Goo: 213, Fai: 43, Exc: 9 |
| Foundation | 0 | 1 | FALSE | 6 | PCo: 911, CBl: 880, Brk: 216, Sla: 36 |
| Bsmt\_Qual | 0 | 1 | FALSE | 6 | Typ: 911, Goo: 849, Exc: 178, No\_: 57 |
| Bsmt\_Cond | 0 | 1 | FALSE | 6 | Typ: 1833, Goo: 80, Fai: 76, No\_: 57 |
| Bsmt\_Exposure | 0 | 1 | FALSE | 5 | No: 1331, Av: 284, Gd: 199, Mn: 179 |
| BsmtFin\_Type\_1 | 0 | 1 | FALSE | 7 | Unf: 602, GLQ: 578, ALQ: 298, Rec: 216 |
| BsmtFin\_Type\_2 | 0 | 1 | FALSE | 7 | Unf: 1740, Rec: 79, LwQ: 64, No\_: 58 |
| Heating | 0 | 1 | FALSE | 6 | Gas: 2019, Gas: 21, Gra: 6, Wal: 5 |
| Heating\_QC | 0 | 1 | FALSE | 5 | Exc: 1040, Typ: 618, Goo: 333, Fai: 61 |
| Central\_Air | 0 | 1 | FALSE | 2 | Y: 1916, N: 137 |
| Electrical | 0 | 1 | FALSE | 5 | SBr: 1887, Fus: 126, Fus: 33, Fus: 6 |
| Kitchen\_Qual | 0 | 1 | FALSE | 5 | Typ: 1070, Goo: 790, Exc: 142, Fai: 50 |
| Functional | 0 | 1 | FALSE | 8 | Typ: 1896, Min: 54, Min: 51, Mod: 27 |
| Fireplace\_Qu | 0 | 1 | FALSE | 6 | No\_: 993, Goo: 538, Typ: 409, Fai: 56 |
| Garage\_Type | 0 | 1 | FALSE | 7 | Att: 1204, Det: 549, Bui: 127, No\_: 108 |
| Garage\_Finish | 0 | 1 | FALSE | 4 | Unf: 872, RFn: 563, Fin: 509, No\_: 109 |
| Garage\_Qual | 0 | 1 | FALSE | 6 | Typ: 1839, No\_: 109, Fai: 85, Goo: 16 |
| Garage\_Cond | 0 | 1 | FALSE | 6 | Typ: 1872, No\_: 109, Fai: 53, Goo: 10 |
| Paved\_Drive | 0 | 1 | FALSE | 3 | Pav: 1848, Dir: 163, Par: 42 |
| Pool\_QC | 0 | 1 | FALSE | 5 | No\_: 2047, Exc: 2, Typ: 2, Fai: 1 |
| Fence | 0 | 1 | FALSE | 5 | No\_: 1661, Min: 225, Goo: 81, Goo: 77 |
| Misc\_Feature | 0 | 1 | FALSE | 5 | Non: 1978, She: 66, Gar: 5, Oth: 3 |
| Mo\_Sold | 0 | 1 | FALSE | 12 | 6: 352, 7: 320, 5: 275, 4: 187 |
| Year\_Sold | 0 | 1 | FALSE | 5 | 200: 499, 200: 456, 200: 445, 200: 442 |
| Sale\_Type | 0 | 1 | FALSE | 10 | WD: 1789, New: 163, COD: 54, Con: 16 |
| Sale\_Condition | 0 | 1 | FALSE | 6 | Nor: 1712, Par: 169, Abn: 121, Fam: 30 |
| Above\_Median | 0 | 1 | FALSE | 2 | Yes: 1043, No: 1010 |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| Lot\_Frontage | 0 | 1 | 57.38 | 33.20 | 0 | 43 | 62 | 78 | 313 | ▇▇▁▁▁ |
| Lot\_Area | 0 | 1 | 10258.40 | 8427.38 | 1300 | 7500 | 9548 | 11600 | 215245 | ▇▁▁▁▁ |
| Mas\_Vnr\_Area | 0 | 1 | 103.75 | 183.59 | 0 | 0 | 0 | 164 | 1600 | ▇▁▁▁▁ |
| BsmtFin\_SF\_1 | 0 | 1 | 4.21 | 2.24 | 1 | 3 | 3 | 7 | 7 | ▅▆▁▁▇ |
| BsmtFin\_SF\_2 | 0 | 1 | 52.57 | 175.99 | 0 | 0 | 0 | 0 | 1526 | ▇▁▁▁▁ |
| Bsmt\_Unf\_SF | 0 | 1 | 561.19 | 441.72 | 0 | 226 | 460 | 801 | 2336 | ▇▅▂▁▁ |
| Total\_Bsmt\_SF | 0 | 1 | 1054.57 | 435.33 | 0 | 793 | 988 | 1304 | 5095 | ▇▇▁▁▁ |
| First\_Flr\_SF | 0 | 1 | 1167.52 | 391.79 | 432 | 882 | 1088 | 1402 | 5095 | ▇▃▁▁▁ |
| Second\_Flr\_SF | 0 | 1 | 326.07 | 422.44 | 0 | 0 | 0 | 701 | 1862 | ▇▂▂▁▁ |
| Low\_Qual\_Fin\_SF | 0 | 1 | 4.97 | 49.09 | 0 | 0 | 0 | 0 | 1064 | ▇▁▁▁▁ |
| Gr\_Liv\_Area | 0 | 1 | 1498.56 | 487.84 | 480 | 1137 | 1447 | 1737 | 5095 | ▇▇▁▁▁ |
| Bsmt\_Full\_Bath | 0 | 1 | 0.43 | 0.53 | 0 | 0 | 0 | 1 | 3 | ▇▆▁▁▁ |
| Bsmt\_Half\_Bath | 0 | 1 | 0.06 | 0.24 | 0 | 0 | 0 | 0 | 2 | ▇▁▁▁▁ |
| Full\_Bath | 0 | 1 | 1.56 | 0.55 | 0 | 1 | 2 | 2 | 4 | ▁▇▇▁▁ |
| Half\_Bath | 0 | 1 | 0.38 | 0.50 | 0 | 0 | 0 | 1 | 2 | ▇▁▅▁▁ |
| Bedroom\_AbvGr | 0 | 1 | 2.86 | 0.82 | 0 | 2 | 3 | 3 | 6 | ▁▃▇▂▁ |
| Kitchen\_AbvGr | 0 | 1 | 1.05 | 0.22 | 1 | 1 | 1 | 1 | 3 | ▇▁▁▁▁ |
| TotRms\_AbvGrd | 0 | 1 | 6.44 | 1.54 | 3 | 5 | 6 | 7 | 15 | ▅▇▃▁▁ |
| Fireplaces | 0 | 1 | 0.60 | 0.65 | 0 | 0 | 1 | 1 | 4 | ▇▇▁▁▁ |
| Garage\_Cars | 0 | 1 | 1.77 | 0.76 | 0 | 1 | 2 | 2 | 5 | ▅▇▂▁▁ |
| Garage\_Area | 0 | 1 | 471.96 | 213.43 | 0 | 320 | 478 | 576 | 1488 | ▃▇▂▁▁ |
| Wood\_Deck\_SF | 0 | 1 | 93.52 | 127.71 | 0 | 0 | 0 | 168 | 1424 | ▇▁▁▁▁ |
| Open\_Porch\_SF | 0 | 1 | 48.17 | 69.51 | 0 | 0 | 27 | 72 | 742 | ▇▁▁▁▁ |
| Enclosed\_Porch | 0 | 1 | 23.02 | 60.59 | 0 | 0 | 0 | 0 | 584 | ▇▁▁▁▁ |
| Three\_season\_porch | 0 | 1 | 2.80 | 25.65 | 0 | 0 | 0 | 0 | 407 | ▇▁▁▁▁ |
| Screen\_Porch | 0 | 1 | 16.68 | 57.94 | 0 | 0 | 0 | 0 | 576 | ▇▁▁▁▁ |
| Pool\_Area | 0 | 1 | 1.34 | 27.74 | 0 | 0 | 0 | 0 | 800 | ▇▁▁▁▁ |
| Misc\_Val | 0 | 1 | 60.12 | 662.76 | 0 | 0 | 0 | 0 | 17000 | ▇▁▁▁▁ |

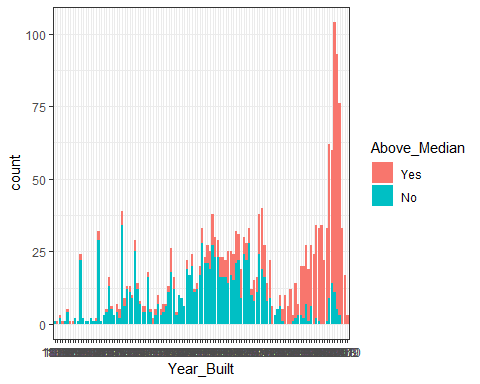
vim\_plot = aggr(ames, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)



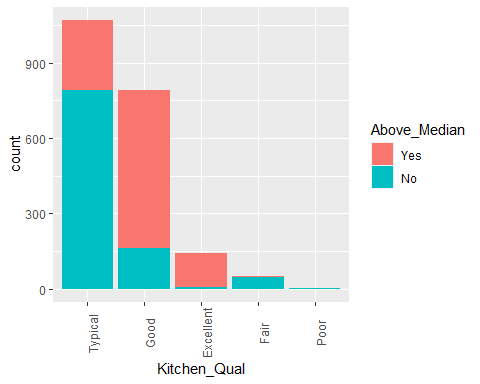
It does not appear as though we have any missing data in this dataset upon looking at the summary, skim, and VIM for this dataset.

Visualizations:

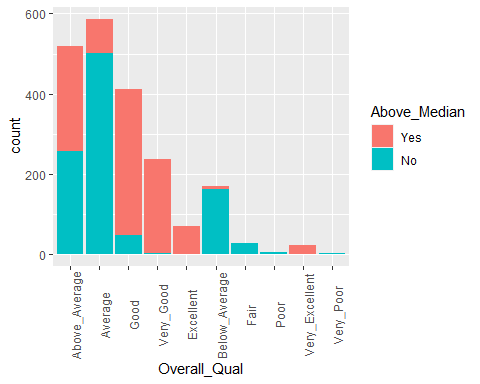
ggplot(ames, aes(x=Year\_Built, fill = Above\_Median)) + geom\_bar() + theme\_bw()



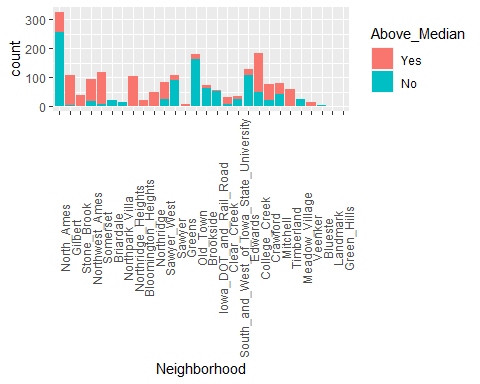
ggplot(ames, aes(x=Kitchen\_Qual, fill = Above\_Median)) + geom\_bar() + theme(axis.text.x = element\_text(angle = 90))



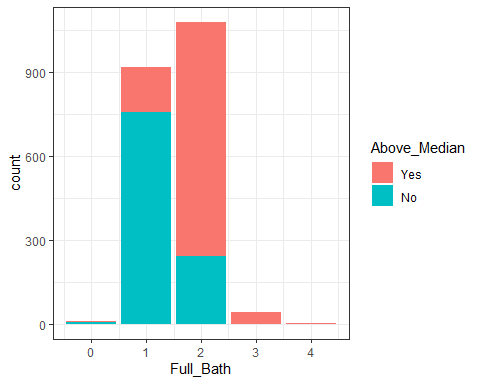
ggplot(ames, aes(x=Overall\_Qual, fill = Above\_Median)) + geom\_bar() + theme(axis.text.x = element\_text(angle = 90))



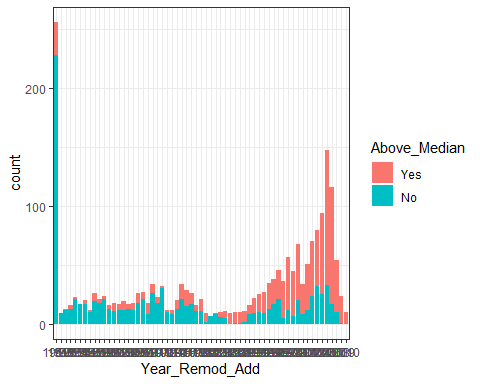
ggplot(ames, aes(x=Neighborhood, fill = Above\_Median)) + geom\_bar() + theme(axis.text.x = element\_text(angle = 90))



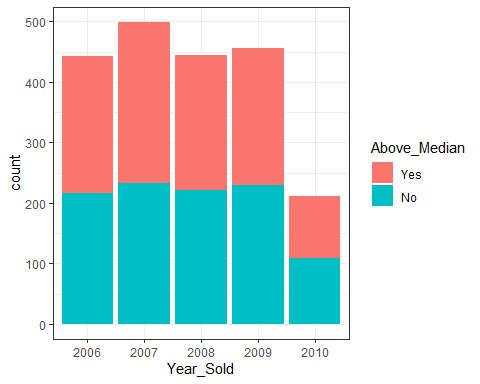
ggplot(ames, aes(x=Full\_Bath, fill = Above\_Median)) + geom\_bar() + theme\_bw()



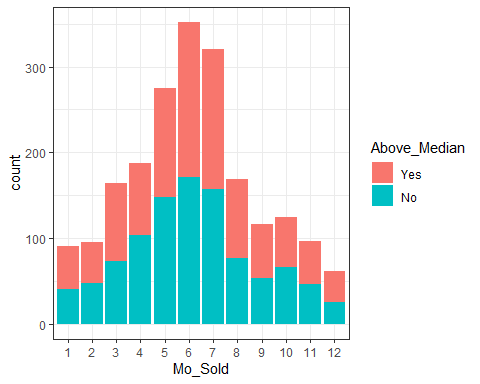
ggplot(ames, aes(x=Year\_Remod\_Add, fill = Above\_Median)) + geom\_bar() + theme\_bw()



ggplot(ames, aes(x=Year\_Sold, fill = Above\_Median)) + geom\_bar() + theme\_bw()



ggplot(ames, aes(x=Mo\_Sold, fill = Above\_Median)) + geom\_bar() + theme\_bw()



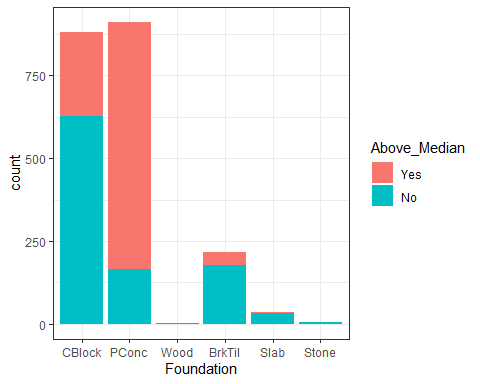
Another thing we wanted to look at was variable importance.

set.seed(1234)  
ames\_split = initial\_split(ames, prop = 0.7, strata = Above\_Median)  
train = training(ames\_split)  
test = testing(ames\_split)

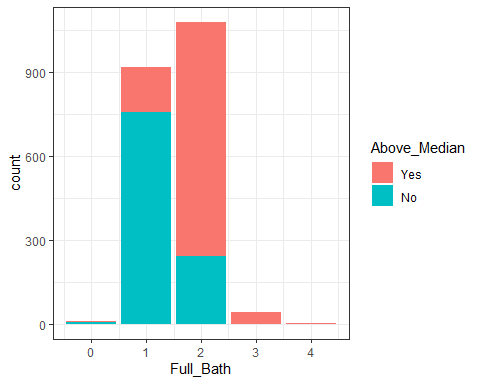
set.seed(100)  
rPartMod <- train(Above\_Median ~ ., data=train, method="rpart")  
rpartImp <- varImp(rPartMod)  
print(rpartImp)

## rpart variable importance  
##   
## only 20 most important variables shown (out of 483)  
##   
## Overall  
## FoundationPConc 100.000  
## Garage\_Cars 98.969  
## Full\_Bath 86.854  
## Gr\_Liv\_Area 82.313  
## Exter\_QualGood 75.024  
## Kitchen\_AbvGr 28.065  
## Garage\_FinishUnf 20.376  
## Garage\_TypeDetchd 18.932  
## Fireplaces 13.010  
## Fireplace\_QuNo\_Fireplace 13.010  
## Lot\_Area 8.555  
## Overall\_QualGood 6.561  
## Overall\_CondFair 0.000  
## NeighborhoodNorthwest\_Ames 0.000  
## MS\_ZoningFloating\_Village\_Residential 0.000  
## Year\_Built1991 0.000  
## Year\_Remod\_Add1974 0.000  
## Exterior\_2ndHdBoard 0.000  
## `Exterior\_2ndWd Sdng` 0.000  
## NeighborhoodBloomington\_Heights 0.000

ggplot(ames, aes(x=Foundation, fill = Above\_Median)) + geom\_bar() + theme\_bw()



ggplot(ames, aes(x=Full\_Bath, fill = Above\_Median)) + geom\_bar() + theme\_bw()



ggplot(ames, aes(x=Garage\_Cars, fill = Above\_Median)) + geom\_bar() + theme\_bw()

