Untitled

Marian Hudson

6/19/2022

Load in data

ames\_student <- read\_csv("C:/Users/m698f/Downloads/ames\_student.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.

View(ames\_student)  
ames <- ames\_student

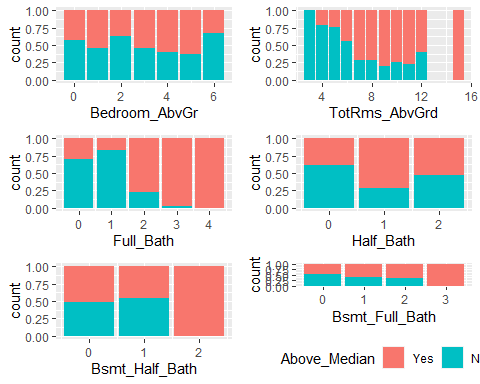
Review and clean up

ames<- ames %>% mutate\_if(is.character, as\_factor)  
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  
## Min. :1875 Min. :1950 Hip : 404 CompShg:2023 VinylSd:705   
## 1st Qu.:1953 1st Qu.:1965 Gable :1607 WdShake: 8 MetalSd:319   
## Median :1972 Median :1993 Mansard: 9 Tar&Grv: 17 Wd Sdng:313   
## Mean :1971 Mean :1984 Gambrel: 14 WdShngl: 3 HdBoard:303   
## 3rd Qu.:2000 3rd Qu.:2004 Shed : 5 Roll : 1 Plywood:151   
## Max. :2010 Max. :2010 Flat : 14 Metal : 1 CemntBd: 90   
## (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 Min. : 1.000 Min. :2006 WD :1789   
## Gar2: 5 1st Qu.: 0.00 1st Qu.: 4.000 1st Qu.:2007 New : 163   
## Shed: 66 Median : 0.00 Median : 6.000 Median :2008 COD : 54   
## Othr: 3 Mean : 60.12 Mean : 6.189 Mean :2008 ConLD : 16   
## Elev: 1 3rd Qu.: 0.00 3rd Qu.: 8.000 3rd Qu.:2009 ConLI : 8   
## Max. :17000.00 Max. :12.000 Max. :2010 CWD : 8   
## (Other): 15   
## Sale\_Condition Longitude Latitude Above\_Median  
## Normal :1712 Min. :-93.69 Min. :41.99 Yes:1043   
## Partial: 169 1st Qu.:-93.66 1st Qu.:42.02 No :1010   
## Family : 30 Median :-93.64 Median :42.03   
## Abnorml: 121 Mean :-93.64 Mean :42.03   
## Alloca : 16 3rd Qu.:-93.62 3rd Qu.:42.05   
## AdjLand: 5 Max. :-93.58 Max. :42.06   
##

Investigating how number of rooms, bedrooms and bathrooms affect sales

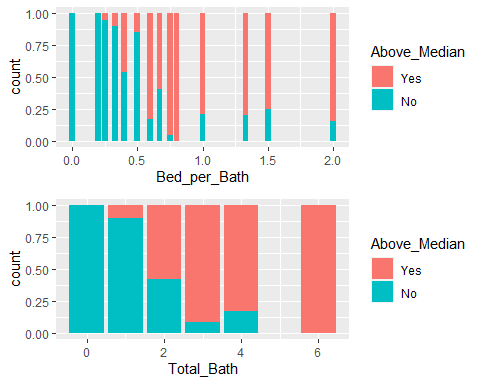
g1 = ggplot(ames, aes(x = Bedroom\_AbvGr, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(legend.position = "none")  
   
g2 = ggplot(ames, aes(x = TotRms\_AbvGrd, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(legend.position = "none") #surprisingly little correlation  
g3 = ggplot(ames, aes(x = Full\_Bath, fill = Above\_Median)) + geom\_bar(position = "fill")+ theme(legend.position = "none") #more bathrooms = more above median  
g4 = ggplot(ames, aes(x = Half\_Bath, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(legend.position = "none") #more half baths = more above median  
g5 = ggplot(ames, aes(x = Bsmt\_Half\_Bath, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(legend.position = "none") #more half baths in the basement= more above median  
g6 = ggplot(ames, aes(x = Bsmt\_Full\_Bath, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(legend.position = "bottom") #more full baths in the basement = more above median  
  
grid.arrange(g1,g2,g3,g4,g5,g6)



Number of bathrooms independently may not be that important so I am going to investigate if the number of bedrooms per bathroom shows more correlation. I am going to create two new variables. Total\_Bath willshow the total number of bathrooms, half plus full plus the basement bathrooms. Bed\_per\_Bath will show the bedroom to bathroom ratio. I believe this may be important because a house that has 5 beds but only 1 bath is likely to be valued lower than a 5 bed/5 bath home. I expect there will be more houses above average that have a lower or 1:1 Bed/Bath ratio.

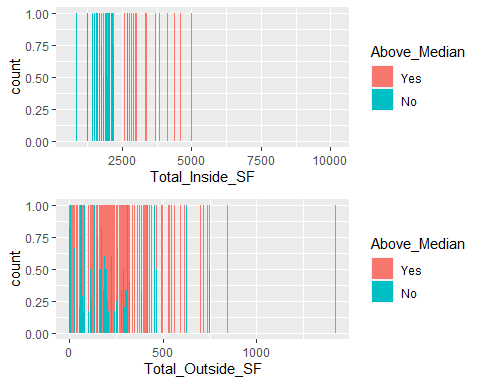
#New variables first  
ames$Total\_Bath <- with(ames, Full\_Bath + Half\_Bath)  
ames$Bed\_per\_Bath <- with(ames, Total\_Bath / Bedroom\_AbvGr)  
ames$Bed\_per\_Bath <- round(ames$Bed\_per\_Bath, 2)  
#Graphs  
g7 = ggplot(ames, aes(x = Bed\_per\_Bath, fill = Above\_Median)) + geom\_bar(position = "fill") #This isn't what I expected  
g8 = ggplot(ames, aes(x = Total\_Bath, fill = Above\_Median)) + geom\_bar(position = "fill") #This shows that more bathrooms = more houses above median  
grid.arrange(g7,g8)

## Warning: Removed 7 rows containing non-finite values (stat\_count).

 It looks like Bed\_per\_Bath may have strong correlation. Houses with more than .5 bedrooms per bathroom (2bed/3bath or “better” ratio) were found to be sold above median prices more often than not.

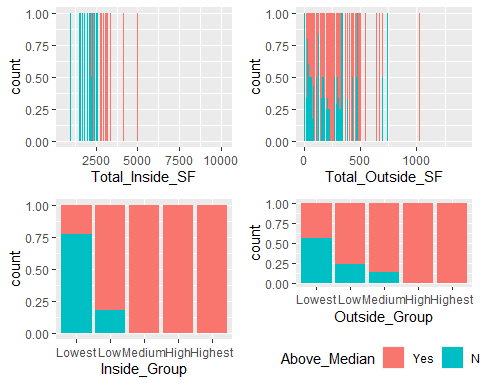
Next let’s look at house size. I want to see how above ground living space affects sales, as well as the basement sqft and the sum of above ground plus basement. I will create a new variable called Total\_Inside\_SF to represent above ground and basement square footage. I will also make a variable called Total\_Outside\_SF to represent the sum of wood deck, open porch, screened porch, and three season porch square footage.

#New variable first  
ames$Total\_Inside\_SF <- with(ames, Total\_Bsmt\_SF + First\_Flr\_SF + Second\_Flr\_SF)  
ames$Total\_Outside\_SF <- with(ames, Screen\_Porch + Open\_Porch\_SF + Wood\_Deck\_SF + Three\_season\_porch)  
  
#Graphs  
g9 = ggplot(ames, aes(x = Total\_Inside\_SF, fill = Above\_Median)) + geom\_bar(position = "fill")   
g10 = ggplot(ames, aes(x = Total\_Outside\_SF, fill = Above\_Median)) + geom\_bar(position = "fill")   
grid.arrange(g9,g10)



The previous graphs look messy and confusing because there are so many possible square footages available. I’m going to group the possible square footages into 5 categories of lowest, low, medium, high, and highest SF called Inside\_Group and Outside\_Group.

#New variables first  
ames$Inside\_Group <- cut(ames$Total\_Inside\_SF, breaks = 5, labels = c("Lowest", "Low", "Medium","High", "Highest"))  
ames$Outside\_Group <- cut(ames$Total\_Outside\_SF, breaks = 5, labels = c("Lowest", "Low", "Medium","High", "Highest"))  
#graphs  
  
g11 = ggplot(ames, aes(x = Inside\_Group, fill = Above\_Median)) + geom\_bar(position = "fill") #This is what I expected - higher SF = more homes above median  
  
g12 = ggplot(ames, aes(x = Outside\_Group, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(legend.position = "bottom") #This is what I expected - higher SF = more homes above median  
  
##Compare new groups to old graphs  
  
grid.arrange(g9 +theme(legend.position='hidden'),g10 +theme(legend.position='hidden'),g11 +theme(legend.position='hidden'),g12)

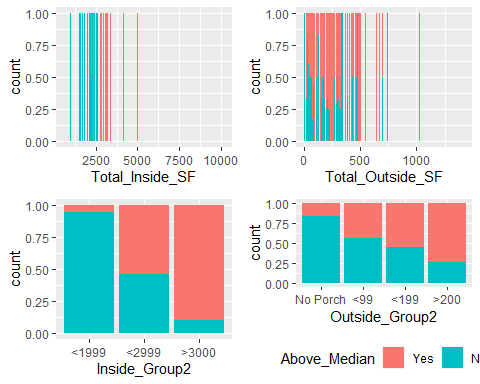


These graphs still aren’t very clear because there appear to be outliers in both data sets. Instead of cutting into 5 chunks, I am going to group the inside SF into a new variable called Inside\_Group2 cut by 0-1999, 2000-2999, and 3000+. I chose these break values because there are only 19 homes smaller than 1000 sqft, and only 93 homes larger than 4000 sqft. These groupings will help to minimize the effect of outliers. I will use a similar strategy for Outside\_Group2.

summary(ames$Total\_Outside\_SF)

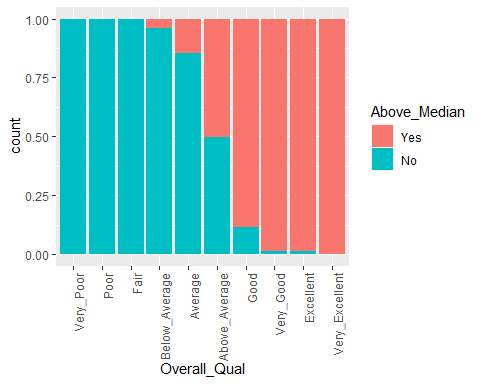
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0 20.0 138.0 161.2 250.0 1424.0

ames$Inside\_Group2 <- cut(ames$Total\_Inside\_SF, breaks = c(0,1999,2999,Inf), labels = c("<1999", "<2999", ">3000"))  
ames$Outside\_Group2 <- cut(ames$Total\_Outside\_SF, breaks = c(-Inf,1,99,199,Inf), labels = c("No Porch", "<99", "<199",">200"))  
  
g13 = ggplot(ames, aes(x = Inside\_Group2, fill = Above\_Median)) + geom\_bar(position = "fill")  
g14 = ggplot(ames, aes(x = Outside\_Group2, fill = Above\_Median)) + geom\_bar(position = "fill")  
  
grid.arrange(g9 +theme(legend.position='hidden'),g10 +theme(legend.position='hidden'),g13 +theme(legend.position='hidden'),g14 +theme(legend.position='bottom'))

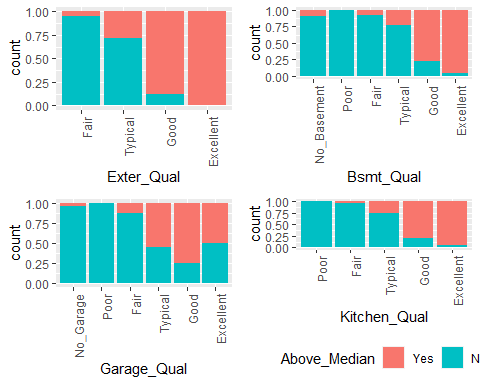
 These new groups look better, and we can see that houses greater than 2000 sqft inside are more likely to sell over median price. The outside sq foot does show a trend that homes with bigger porches do tend to sell above the median price.

Next I want to look at the quality of the homes.

g15\_order <- c("Very\_Poor", "Poor", "Fair","Below\_Average","Average", "Above\_Average", "Good", "Very\_Good", "Excellent", "Very\_Excellent")  
g15 = ggplot(ames, aes(x = Overall\_Qual, fill = Above\_Median)) + scale\_x\_discrete(limits = g15\_order) + geom\_bar(position = "fill") + theme(axis.text.x=element\_text(angle=90, hjust=1))  
g16\_order <-c("Fair", "Typical", "Good", "Excellent")   
g16 = ggplot(ames, aes(x = Exter\_Qual, fill = Above\_Median)) + scale\_x\_discrete(limits = g16\_order) + geom\_bar(position = "fill") + theme(axis.text.x=element\_text(angle=90, hjust=1))  
g17\_order <- c("No\_Basement", "Poor", "Fair", "Typical", "Good", "Excellent")  
g17 = ggplot(ames, aes(x = Bsmt\_Qual, fill = Above\_Median))+ scale\_x\_discrete(limits = g17\_order) + geom\_bar(position = "fill") + theme(axis.text.x=element\_text(angle=90, hjust=1))  
g18\_order <- c("No\_Garage", "Poor", "Fair", "Typical", "Good", "Excellent")  
g18 = ggplot(ames, aes(x = Garage\_Qual, fill = Above\_Median)) + scale\_x\_discrete(limits = g18\_order) + geom\_bar(position = "fill") + theme(axis.text.x=element\_text(angle=90, hjust=1))  
g19\_order <- c("Poor", "Fair", "Typical", "Good", "Excellent")  
g19 = ggplot(ames, aes(x = Kitchen\_Qual, fill = Above\_Median)) + scale\_x\_discrete(limits = g19\_order) + geom\_bar(position = "fill") + theme(axis.text.x=element\_text(angle=90, hjust=1))  
  
grid.arrange(g15)

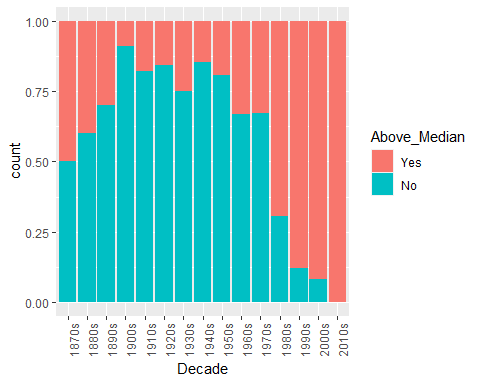


grid.arrange(g16+theme(legend.position='hidden'),g17+theme(legend.position='hidden'),g18+theme(legend.position='hidden'),g19+theme(legend.position='bottom'))

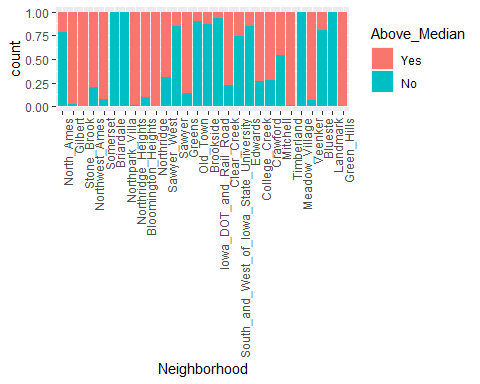


Finally I want to look into the effect of age, neighborhood, and house type. I will make a new variable named Decade to represent the decade that the home was built in.

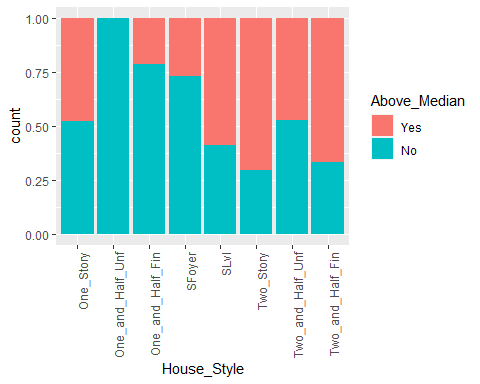
#new variables first  
ames$Decade <- cut(ames$Year\_Built, breaks = c(0,1879,1889,1899,1909,1919,1929,1939,1949,1959,1969,1979,1989,1999,2009,Inf), labels = c("1870s", "1880s", "1890s", "1900s","1910s","1920s","1930s", "1940s", "1950s", "1960s","1970s","1980s","1990s", "2000s", "2010s"))  
#graphs  
g20 = ggplot(ames, aes(x = Decade, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x=element\_text(angle=90, hjust=1))  
grid.arrange(g20)



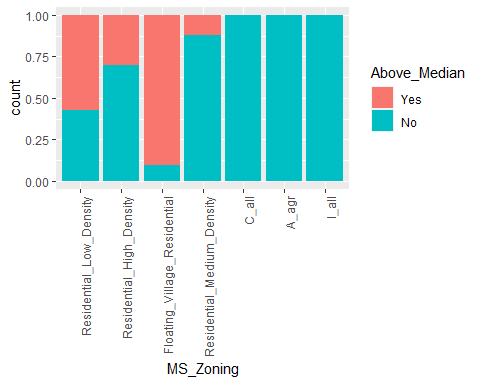
g21 = ggplot(ames, aes(x = Neighborhood, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x=element\_text(angle=90, hjust=1))  
grid.arrange(g21)



g22\_order <- c("One\_Story","One\_and\_Half\_Unf", "One\_and\_Half\_Fin","SFoyer","SLvl", "Two\_Story","Two\_and\_Half\_Unf","Two\_and\_Half\_Fin")  
g22 = ggplot(ames, aes(x = House\_Style, fill = Above\_Median)) + scale\_x\_discrete(limits = g22\_order) + geom\_bar(position = "fill") + theme(axis.text.x=element\_text(angle=90, hjust=1))  
g23 = ggplot(ames, aes(x = MS\_Zoning, fill = Above\_Median)) + geom\_bar(position = "fill") + theme(axis.text.x=element\_text(angle=90, hjust=1))  
grid.arrange(g22)



grid.arrange(g23)



Next I decided to change Decade into Century.

#new variables first  
ames$Century<- cut(ames$Year\_Built, breaks = c(0,1899,1999,Inf), labels = c("1800s","1900s", "2000s"))

I created a new data set that is only the “Yes” values for “Above\_Median” called ames\_over. When We graph century and quality together, it shows us that older homes have to be a higher quality (Very excellent, very good, or good) to be above the median sales price. Newer homes seem to still sell above the median even if they are not of the highest quality. Surprisingly, smaller homes sell above median price even if they are not the highest quality.

#New data source  
ames\_over <- ames %>% filter(Above\_Median == "Yes")  
  
g24 = ggplot(ames\_over, aes(x = Century, fill = Overall\_Qual)) + geom\_bar(position = "fill") + scale\_fill\_manual(values=c("#c25757", "#ed9898", "#56B4E9", "#5b56e9", "#7dd4c4", "#238c82", "#2f8c23"))  
g25 = ggplot(ames\_over, aes(x = Inside\_Group2, fill = Overall\_Qual)) + geom\_bar(position = "fill") + scale\_fill\_manual(values=c("#c25757", "#ed9898", "#56B4E9", "#5b56e9", "#7dd4c4", "#238c82", "#2f8c23"))  
grid.arrange(g24,g25 +theme(legend.position='hidden'))

