# Elizabeth Taylor BAN 502 Predictive Analytics

library(tidyverse) # read in the packages I need  
library(tidymodels)  
library(GGally) #ggcorr and ggpairs  
library(ggcorrplot) #correlation plot alternative  
library(gridExtra) #create grids of plots  
library(esquisse)

library(readr) #read in the data  
ames\_student <- read\_csv("ames\_student.csv")

str(ames\_student) #review the data

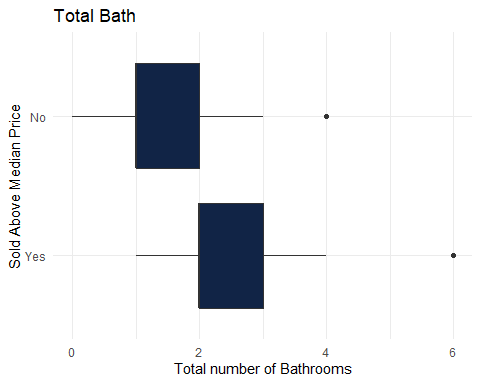
#esquisser(ames\_student) Helped me visualize which variables might be strong predictors of Above Average

ames = ames\_student%>%dplyr::select(Above\_Median,Year\_Built,Gr\_Liv\_Area, Lot\_Area, Full\_Bath,Half\_Bath, TotRms\_AbvGrd, Year\_Built,Year\_Sold, Three\_season\_porch,Screen\_Porch )  
#These were the variables that I found that seemed to be good predictions of the variable Above\_Median

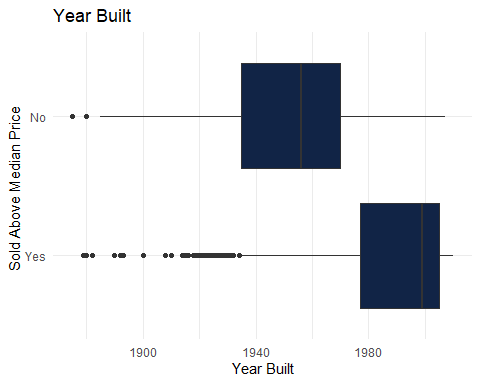
ames = ames %>% mutate(Above\_Median = as\_factor(Above\_Median))   
   
ames$TotBath <- ames$Full\_Bath+ ames$Half\_Bath  
ames$Porch <- ames$Three\_season\_porch+ames$Screen\_Porch  
ames <- ames %>% select(-Full\_Bath,-Half\_Bath,-Three\_season\_porch,-Screen\_Porch)  
#I decide to add both half bath and full bath together to create total bath as one variable.   
#I decided that a three season porch and screen porch were similar enough to add them together as one variable  
#Once I created the new variables I subtracted the old variables out of the data set.

#ggpairs(ames) #So I can visualize my variables on my secdata set

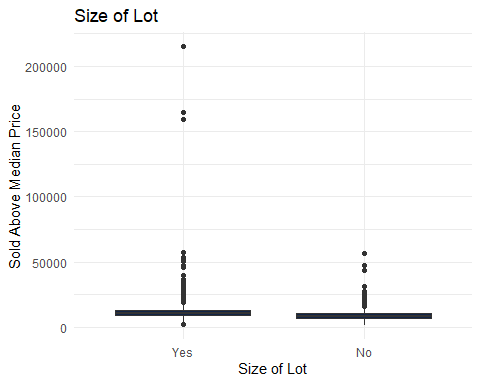
ggplot(ames) +  
 aes(x = TotBath, y = Above\_Median) +  
 geom\_boxplot(fill = "#112446") +  
 labs(  
 x = "Total number of Bathrooms",  
 y = "Sold Above Median Price",  
 title = "Total Bath"  
 ) +  
 theme\_minimal()



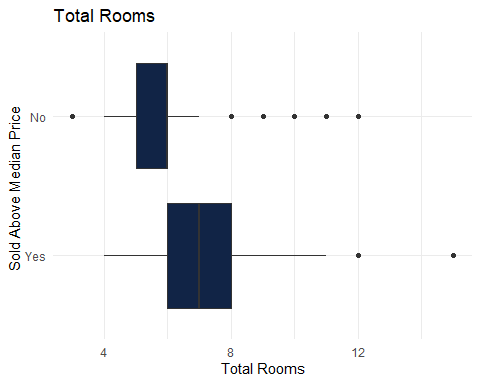
ggplot(ames) +  
 aes(x = Year\_Built, y = Above\_Median) +  
 geom\_boxplot(fill = "#112446") +  
 labs(  
 x = "Year Built",  
 y = "Sold Above Median Price ",  
 title = "Year Built"  
 ) +  
 theme\_minimal()



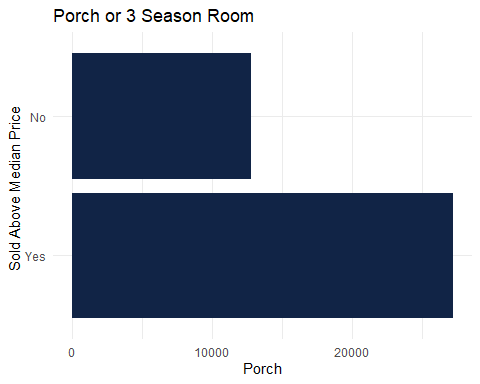
ggplot(ames) +  
 aes(x = Above\_Median, y = Lot\_Area) +  
 geom\_boxplot(fill = "#112446") +  
 labs(  
 x = "Size of Lot",  
 y = "Sold Above Median Price",  
 title = "Size of Lot"  
 ) +  
 theme\_minimal()



ggplot(ames) +  
 aes(x = TotRms\_AbvGrd, y = Above\_Median) +  
 geom\_boxplot(fill = "#112446") +  
 labs(  
 x = "Total Rooms",  
 y = "Sold Above Median Price ",  
 title = "Total Rooms"  
 ) +  
 theme\_minimal()



ggplot(ames) +aes(x = Porch, y = Above\_Median) +  
 geom\_col(fill = "#112446") +  
 labs(  
 x = "Porch",  
 y = "Sold Above Median Price ",  
 title = "Porch or 3 Season Room"  
 ) +  
 theme\_minimal()



theme\_minimal()  
ggplot(ames) +  
 aes(x = Above\_Median, y = Gr\_Liv\_Area, ) +  
 geom\_col(fill = "#112446") +  
 labs(  
 x = "Total Sq Footage",  
 y = "Sold Above Median Price ",  
 title = "Size of House"  
 ) +  
 theme\_minimal()

