Project Phase 1

Engel, Alec

# summary(student)  
# glimpse(student)  
# skim(student)

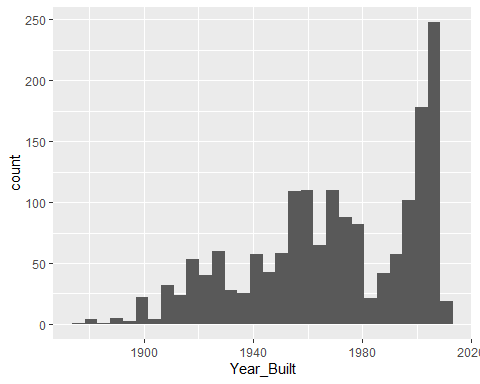
student = student %>% dplyr::select(-X1) %>%   
 mutate\_if(is.character,as\_factor) %>%   
 mutate(Mo\_Sold = as\_factor(Mo\_Sold)) %>%  
 mutate(Mo\_Sold = fct\_recode(Mo\_Sold, "Jan" = "1", "Feb" = "2", "Mar" = "3", "Apr" = "4", "May" = "5", "Jun" = "6",   
 "Jul" = "7", "Aug" = "8", "Sep" = "9", "Oct" = "10", "Nov" = "11", "Dec" = "12")) %>%  
 mutate(BsmtFin\_SF\_1 = Total\_Bsmt\_SF - BsmtFin\_SF\_2 - Bsmt\_Unf\_SF)

# ggplot(student, aes(Lot\_Frontage)) + geom\_histogram()  
# ggplot(student, aes(Lot\_Area)) + geom\_histogram()  
# ggplot(student, aes(Year\_Built)) + geom\_histogram()  
# ggplot(student, aes(Year\_Remod\_Add)) + geom\_histogram()  
# ggplot(student, aes(Mas\_Vnr\_Area)) + geom\_histogram()  
# ggplot(student, aes(BsmtFin\_SF\_1)) + geom\_histogram()  
# ggplot(student, aes(BsmtFin\_SF\_2)) + geom\_histogram()  
# ggplot(student, aes(Bsmt\_Unf\_SF)) + geom\_histogram()  
# ggplot(student, aes(Total\_Bsmt\_SF)) + geom\_histogram()  
# ggplot(student, aes(First\_Flr\_SF)) + geom\_histogram()  
# ggplot(student, aes(Second\_Flr\_SF)) + geom\_histogram()  
# ggplot(student, aes(Low\_Qual\_Fin\_SF)) + geom\_histogram()  
# ggplot(student, aes(Gr\_Liv\_Area)) + geom\_histogram()  
# ggplot(student, aes(Bsmt\_Full\_Bath)) + geom\_histogram()  
# ggplot(student, aes(Bsmt\_Half\_Bath)) + geom\_histogram()  
# ggplot(student, aes(Full\_Bath)) + geom\_histogram()  
# ggplot(student, aes(Half\_Bath)) + geom\_histogram()  
# ggplot(student, aes(Bedroom\_AbvGr)) + geom\_histogram()  
# ggplot(student, aes(Kitchen\_AbvGr)) + geom\_histogram()  
# ggplot(student, aes(TotRms\_AbvGrd)) + geom\_histogram()  
# ggplot(student, aes(Fireplaces)) + geom\_histogram()  
# ggplot(student, aes(Kitchen\_AbvGr)) + geom\_histogram()  
# ggplot(student, aes(TotRms\_AbvGrd)) + geom\_histogram()  
# ggplot(student, aes(Fireplaces)) + geom\_histogram()  
# ggplot(student, aes(Garage\_Cars)) + geom\_histogram()  
# ggplot(student, aes(Garage\_Area)) + geom\_histogram()  
# ggplot(student, aes(Wood\_Deck\_SF)) + geom\_histogram()  
# ggplot(student, aes(Open\_Porch\_SF)) + geom\_histogram()  
# ggplot(student, aes(Enclosed\_Porch)) + geom\_histogram()  
# ggplot(student, aes(Three\_season\_porch)) + geom\_histogram()  
# ggplot(student, aes(Screen\_Porch)) + geom\_histogram()  
# ggplot(student, aes(Pool\_Area)) + geom\_histogram()  
# ggplot(student, aes(Misc\_Val)) + geom\_histogram()  
# ggplot(student, aes(Year\_Sold)) + geom\_histogram()  
# ggplot(student, aes(Longitude)) + geom\_histogram()  
# ggplot(student, aes(Latitude)) + geom\_histogram()

student = student %>% filter(Lot\_Frontage < 175) %>%  
 filter(Lot\_Area < 40000) %>%  
 filter(Mas\_Vnr\_Area < 400) %>%  
 filter(BsmtFin\_SF\_1 < 1600) %>%  
 filter(BsmtFin\_SF\_2 < 400) %>%  
 filter(Bsmt\_Unf\_SF < 2250) %>%  
 filter(Total\_Bsmt\_SF < 2750) %>%  
 filter(Bsmt\_Half\_Bath < 1.1) %>%  
 filter(Full\_Bath > 0) %>%  
 filter(Half\_Bath < 1.1) %>%  
 filter(Pool\_Area < 1) %>%  
 filter(First\_Flr\_SF < 2750) %>%  
 filter(Second\_Flr\_SF < 1400) %>%  
 filter(Low\_Qual\_Fin\_SF < 600) %>%  
 filter(Gr\_Liv\_Area < 3750) %>%  
 filter(Kitchen\_AbvGr < 3) %>%  
 filter(Fireplaces < 4) %>%  
 filter(Garage\_Cars < 4) %>%  
 filter(Garage\_Area < 1250) %>%  
 filter(Wood\_Deck\_SF < 550) %>%  
 filter(Open\_Porch\_SF < 350) %>%  
 filter(Enclosed\_Porch < 300) %>%  
 filter(Three\_season\_porch < 240) %>%  
 filter(Screen\_Porch < 400) %>%  
 filter(Misc\_Val < 1000)

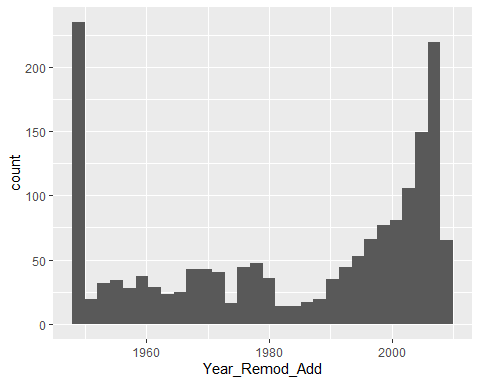
ggplot(student, aes(Year\_Built)) + geom\_histogram()

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



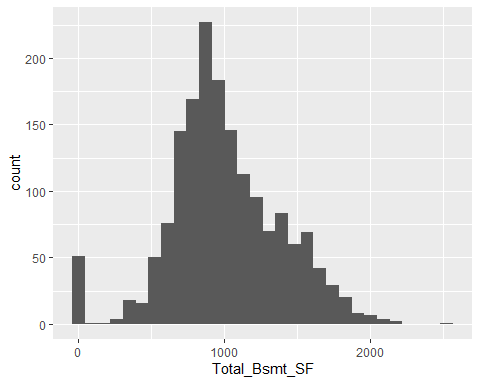
ggplot(student, aes(Year\_Remod\_Add)) + geom\_histogram()

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



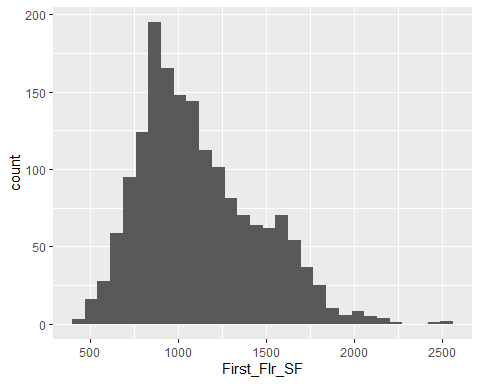
ggplot(student, aes(Total\_Bsmt\_SF)) + geom\_histogram()

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



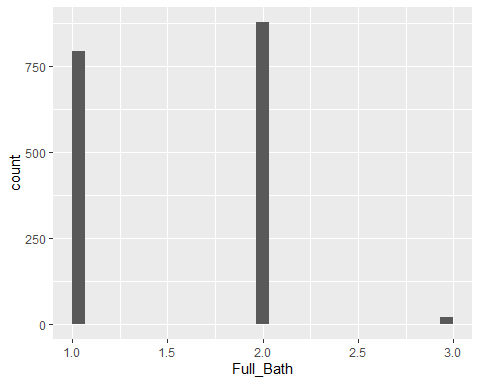
ggplot(student, aes(First\_Flr\_SF)) + geom\_histogram()

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



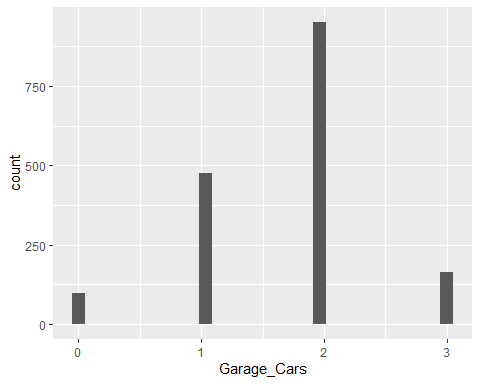
ggplot(student, aes(Full\_Bath)) + geom\_histogram()

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



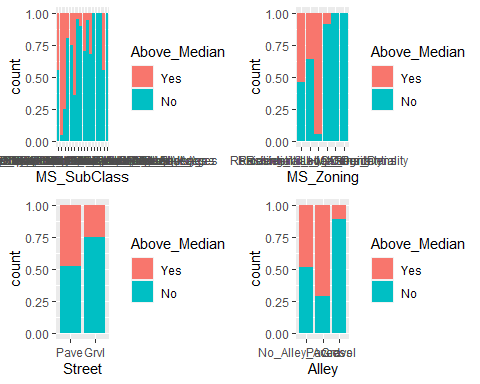
ggplot(student, aes(Garage\_Cars)) + geom\_histogram()

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

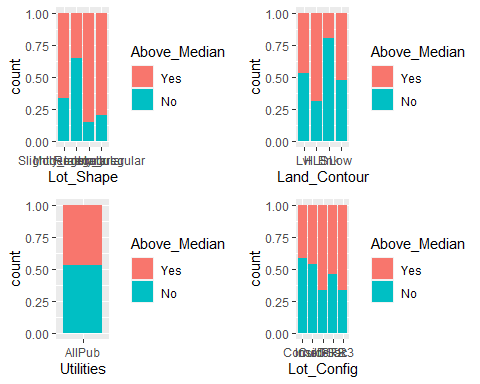


set.seed(123)  
student\_split = initial\_split(student, prob = 0.80, strata = Above\_Median)  
train = training(student\_split)  
test = testing(student\_split)

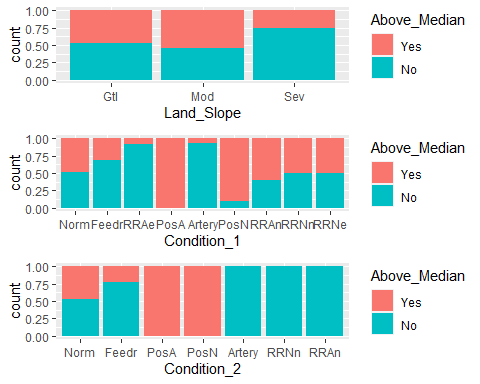
p1 = ggplot(train, aes(x = MS\_SubClass, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = MS\_Zoning, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Street, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = Alley, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



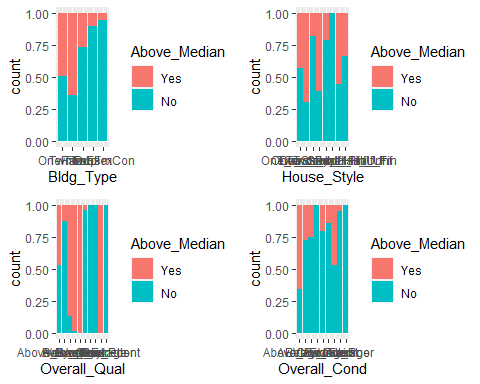
p1 = ggplot(train, aes(x = Lot\_Shape, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Land\_Contour, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Utilities, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = Lot\_Config, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



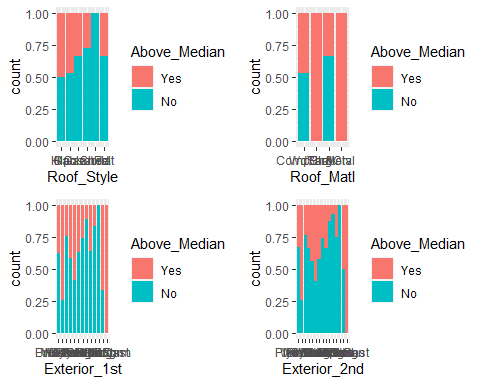
p1 = ggplot(train, aes(x = Land\_Slope, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Condition\_1, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Condition\_2, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3)



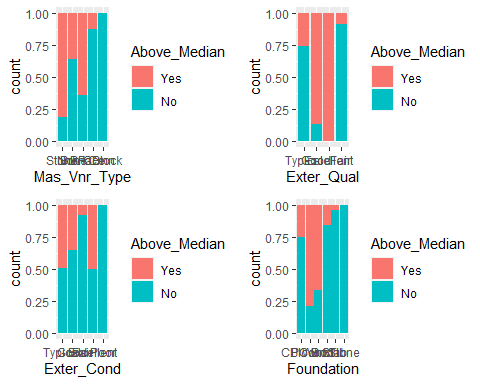
p1 = ggplot(train, aes(x = Bldg\_Type, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = House\_Style, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Overall\_Qual, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = Overall\_Cond, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



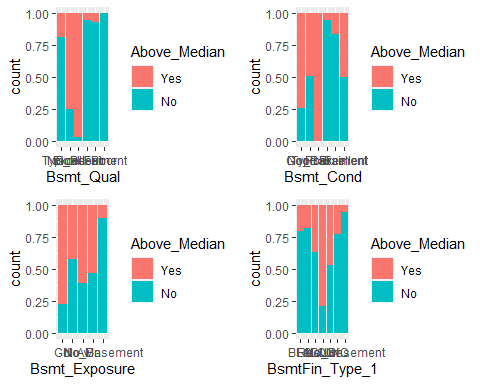
p1 = ggplot(train, aes(x = Roof\_Style, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Roof\_Matl, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Exterior\_1st, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = Exterior\_2nd, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



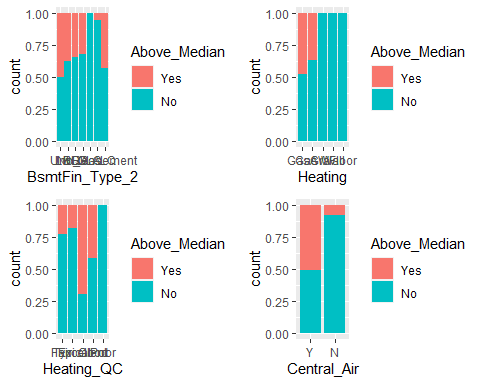
p1 = ggplot(train, aes(x = Mas\_Vnr\_Type, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Exter\_Qual, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Exter\_Cond, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = Foundation, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



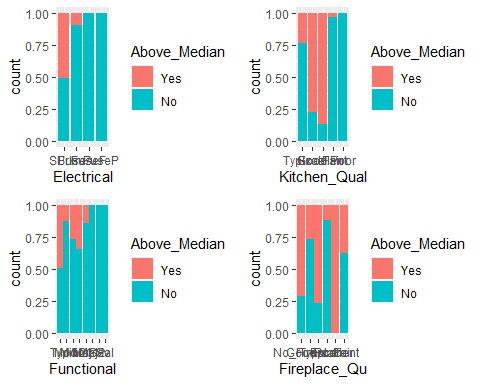
p1 = ggplot(train, aes(x = Bsmt\_Qual, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Bsmt\_Cond, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Bsmt\_Exposure, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = BsmtFin\_Type\_1, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



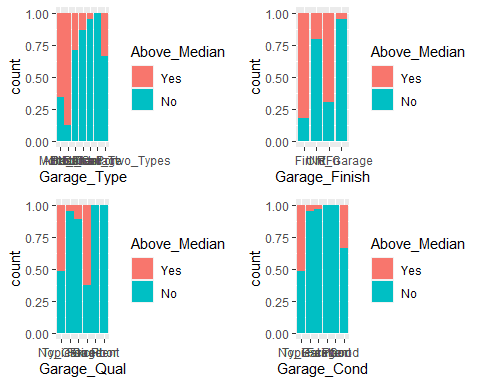
p1 = ggplot(train, aes(x = BsmtFin\_Type\_2, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Heating, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Heating\_QC, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = Central\_Air, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



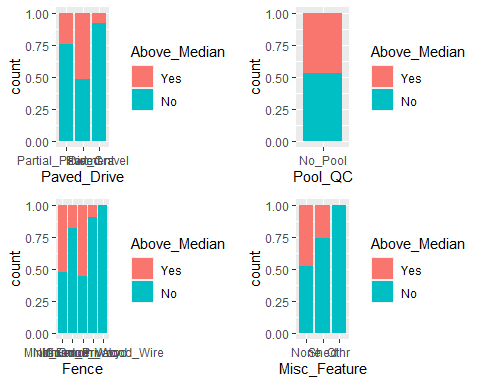
p1 = ggplot(train, aes(x = Electrical, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Kitchen\_Qual, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Functional, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = Fireplace\_Qu, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



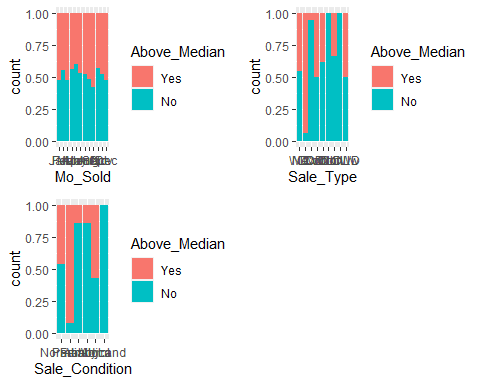
p1 = ggplot(train, aes(x = Garage\_Type, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Garage\_Finish, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Garage\_Qual, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = Garage\_Cond, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



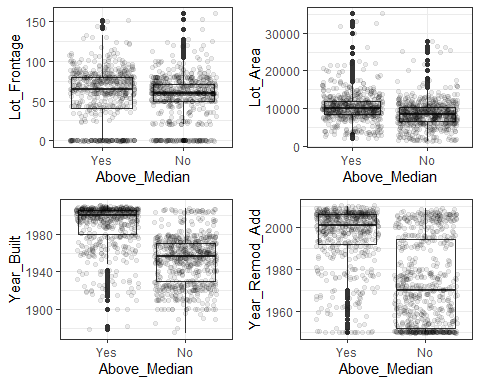
p1 = ggplot(train, aes(x = Paved\_Drive, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Pool\_QC, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Fence, fill = Above\_Median)) + geom\_bar(position = "fill")  
p4 = ggplot(train, aes(x = Misc\_Feature, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3,p4)



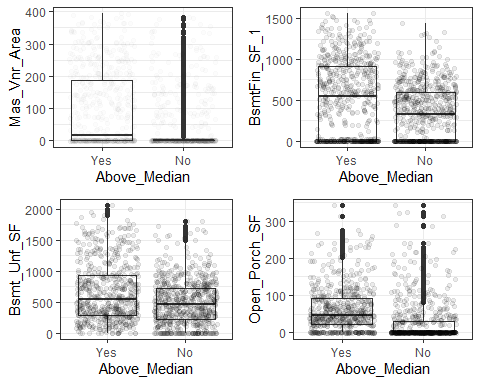
p1 = ggplot(train, aes(x = Mo\_Sold, fill = Above\_Median)) + geom\_bar(position = "fill")  
p2 = ggplot(train, aes(x = Sale\_Type, fill = Above\_Median)) + geom\_bar(position = "fill")  
p3 = ggplot(train, aes(x = Sale\_Condition, fill = Above\_Median)) + geom\_bar(position = "fill")  
grid.arrange(p1,p2,p3, ncol = 2)



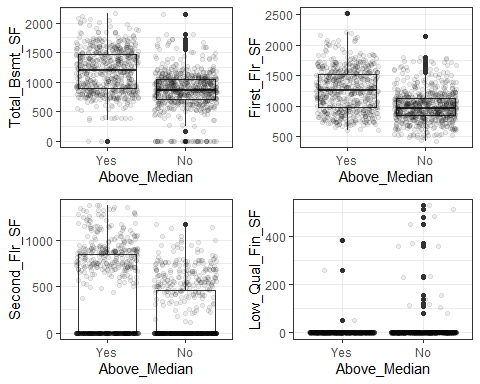
p1 = ggplot(train, aes(x = Above\_Median, y = Lot\_Frontage)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p2 = ggplot(train, aes(x = Above\_Median, y = Lot\_Area)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p3 = ggplot(train, aes(x = Above\_Median, y = Year\_Built)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p4 = ggplot(train, aes(x = Above\_Median, y = Year\_Remod\_Add)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
grid.arrange(p1,p2,p3,p4)



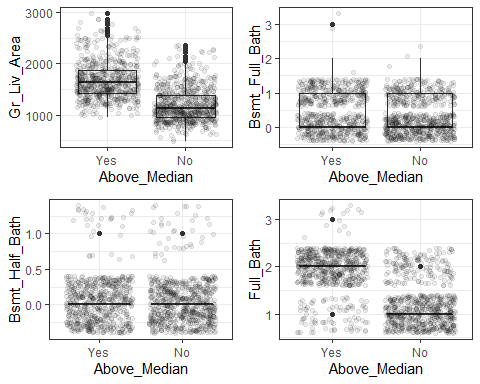
p1 = ggplot(train, aes(x = Above\_Median, y = Mas\_Vnr\_Area)) + geom\_boxplot() + geom\_jitter(alpha = 0.01) + theme\_bw()  
p2 = ggplot(train, aes(x = Above\_Median, y = BsmtFin\_SF\_1)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p3 = ggplot(train, aes(x = Above\_Median, y = Bsmt\_Unf\_SF)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p4 = ggplot(train, aes(x = Above\_Median, y = Open\_Porch\_SF)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
grid.arrange(p1,p2,p3,p4)



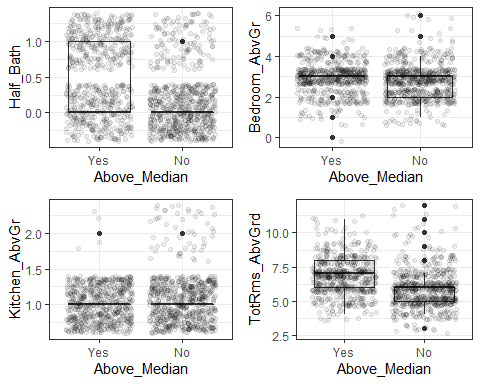
p1 = ggplot(train, aes(x = Above\_Median, y = Total\_Bsmt\_SF)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p2 = ggplot(train, aes(x = Above\_Median, y = First\_Flr\_SF)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p3 = ggplot(train, aes(x = Above\_Median, y = Second\_Flr\_SF)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p4 = ggplot(train, aes(x = Above\_Median, y = Low\_Qual\_Fin\_SF)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
grid.arrange(p1,p2,p3,p4)



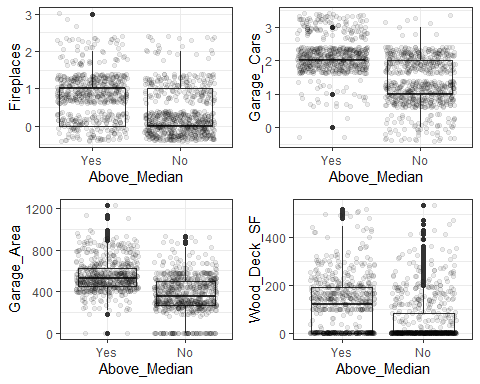
p1 = ggplot(train, aes(x = Above\_Median, y = Gr\_Liv\_Area)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p2 = ggplot(train, aes(x = Above\_Median, y = Bsmt\_Full\_Bath)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p3 = ggplot(train, aes(x = Above\_Median, y = Bsmt\_Half\_Bath)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p4 = ggplot(train, aes(x = Above\_Median, y = Full\_Bath)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
grid.arrange(p1,p2,p3,p4)



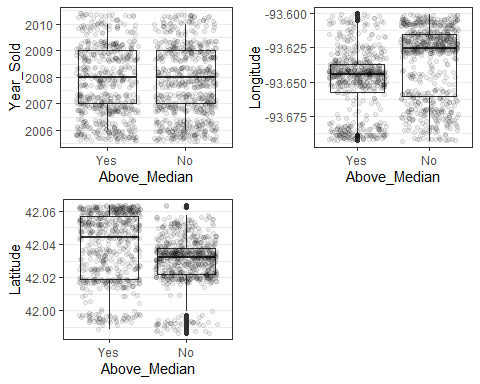
p1 = ggplot(train, aes(x = Above\_Median, y = Half\_Bath)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p2 = ggplot(train, aes(x = Above\_Median, y = Bedroom\_AbvGr)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p3 = ggplot(train, aes(x = Above\_Median, y = Kitchen\_AbvGr)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p4 = ggplot(train, aes(x = Above\_Median, y = TotRms\_AbvGrd)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
grid.arrange(p1,p2,p3,p4)



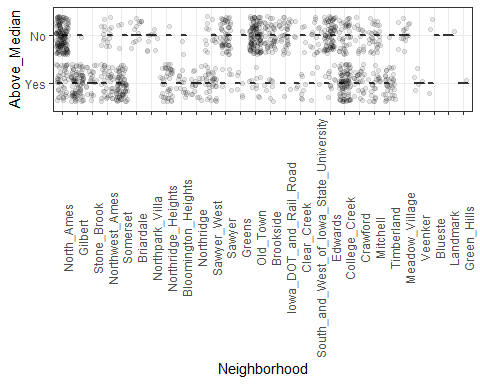
p1 = ggplot(train, aes(x = Above\_Median, y = Fireplaces)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p2 = ggplot(train, aes(x = Above\_Median, y = Garage\_Cars)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p3 = ggplot(train, aes(x = Above\_Median, y = Garage\_Area)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p4 = ggplot(train, aes(x = Above\_Median, y = Wood\_Deck\_SF)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
grid.arrange(p1,p2,p3,p4)



p1 = ggplot(train, aes(x = Above\_Median, y = Year\_Sold)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p2 = ggplot(train, aes(x = Above\_Median, y = Longitude)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p3 = ggplot(train, aes(x = Above\_Median, y = Latitude)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
grid.arrange(p1,p2,p3,ncol = 2)



ggplot(train, aes(x = Neighborhood, y = Above\_Median)) + geom\_boxplot() + geom\_jitter(alpha = 0.1) + theme\_bw() +  
 theme(axis.text.x = element\_text(angle = 90))



student\_recipe = recipe(Above\_Median ~., train) %>%  
 step\_other(Neighborhood,threshold = .02) %>%  
 step\_other(MS\_SubClass,threshold = .02) %>%  
 step\_other(Overall\_Qual,threshold = .02) %>%  
 step\_other(Overall\_Cond,threshold = .02) %>%  
 step\_other(Exterior\_1st,threshold = .02) %>%  
 step\_other(Exterior\_2nd,threshold = .02) %>%  
 step\_other(Condition\_1,threshold = .02) %>%  
 step\_other(Condition\_2,threshold = .02) %>%  
 step\_other(Functional,threshold = .02) %>%  
 step\_other(Sale\_Type,threshold = .02) %>%  
 step\_dummy(all\_nominal(), -all\_outcomes())  
  
rf\_model = rand\_forest() %>%  
 set\_engine("ranger", importance = "permutation") %>%  
 set\_mode("classification")  
  
student\_wflow =  
 workflow() %>%  
 add\_model(rf\_model) %>%  
 add\_recipe(student\_recipe)  
  
set.seed(123)  
student\_fit = fit(student\_wflow, train)

trainpredrf = predict(student\_fit, train)  
head(trainpredrf)

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

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

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Yes No  
## Yes 593 7  
## No 7 662  
##   
## Accuracy : 0.989   
## 95% CI : (0.9816, 0.994)  
## No Information Rate : 0.5272   
## P-Value [Acc > NIR] : <2e-16   
##   
## Kappa : 0.9779   
##   
## Mcnemar's Test P-Value : 1   
##   
## Sensitivity : 0.9883   
## Specificity : 0.9895   
## Pos Pred Value : 0.9883   
## Neg Pred Value : 0.9895   
## Prevalence : 0.4728   
## Detection Rate : 0.4673   
## Detection Prevalence : 0.4728   
## Balanced Accuracy : 0.9889   
##   
## 'Positive' Class : Yes   
##

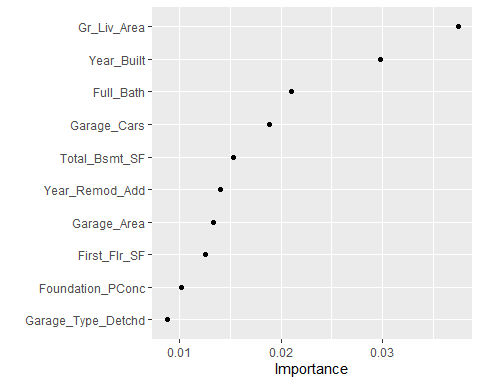
testpredrf = predict(student\_fit, test)  
head(testpredrf)

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

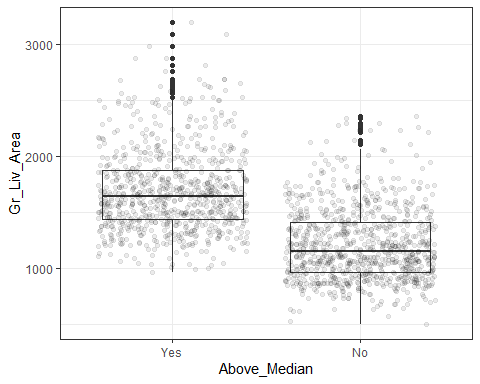
confusionMatrix(testpredrf$.pred\_class, test$Above\_Median,   
 positive = "Yes")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction Yes No  
## Yes 175 6  
## No 24 216  
##   
## Accuracy : 0.9287   
## 95% CI : (0.8998, 0.9514)  
## No Information Rate : 0.5273   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.8564   
##   
## Mcnemar's Test P-Value : 0.001911   
##   
## Sensitivity : 0.8794   
## Specificity : 0.9730   
## Pos Pred Value : 0.9669   
## Neg Pred Value : 0.9000   
## Prevalence : 0.4727   
## Detection Rate : 0.4157   
## Detection Prevalence : 0.4299   
## Balanced Accuracy : 0.9262   
##   
## 'Positive' Class : Yes   
##

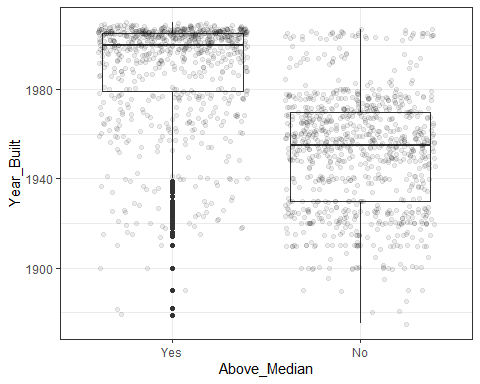
student\_fit %>% pull\_workflow\_fit() %>% vip(geom = "point")



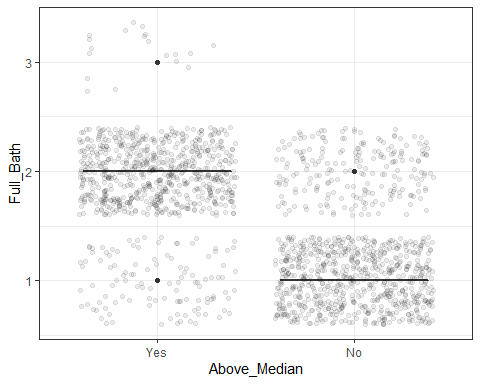
ggplot(student, aes(x = Above\_Median, y = Gr\_Liv\_Area)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()



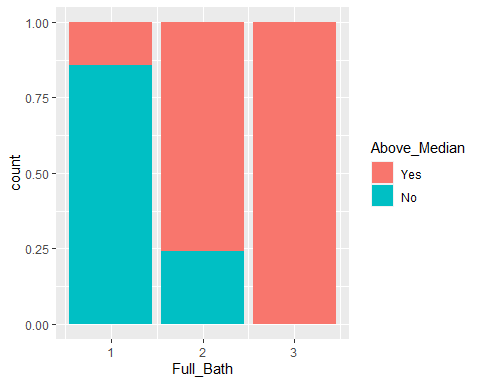
ggplot(student, aes(x = Above\_Median, y = Year\_Built)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()



ggplot(student, aes(x = Above\_Median, y = Full\_Bath)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()



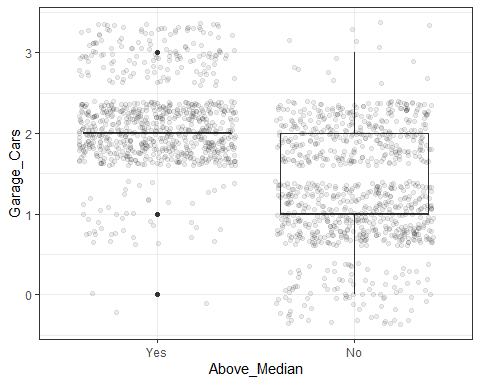
ggplot(student, aes(x = Full\_Bath, fill = Above\_Median)) + geom\_bar(position = "fill")



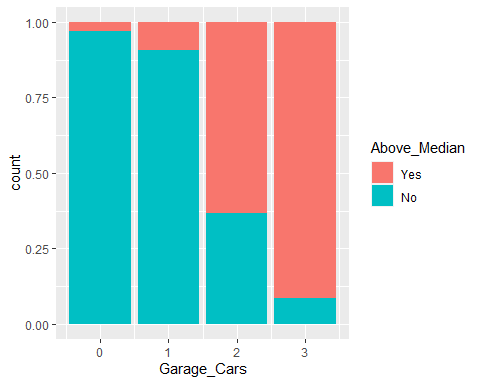
t3 = table(student$Above\_Median, student$Full\_Bath)  
prop.table(t3, margin = 2)

##   
## 1 2 3  
## Yes 0.1437579 0.7585421 1.0000000  
## No 0.8562421 0.2414579 0.0000000

ggplot(student, aes(x = Above\_Median, y = Garage\_Cars)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()



ggplot(student, aes(x = Garage\_Cars, fill = Above\_Median)) + geom\_bar(position = "fill")



t4 = table(student$Above\_Median, student$Garage\_Cars)  
prop.table(t4, margin = 2)

##   
## 0 1 2 3  
## Yes 0.03061224 0.09224319 0.63340336 0.91411043  
## No 0.96938776 0.90775681 0.36659664 0.08588957

p1 = ggplot(student, aes(x = Above\_Median, y = Gr\_Liv\_Area)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p2 = ggplot(student, aes(x = Above\_Median, y = Year\_Built)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p3 = ggplot(student, aes(x = Above\_Median, y = Garage\_Cars)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
p4 = ggplot(student, aes(x = Above\_Median, y = Full\_Bath)) + geom\_boxplot() + geom\_jitter(alpha = 0.08) + theme\_bw()  
grid.arrange(p1,p2,p3,p4)

