##Exploratory/Descriptive Analsysis

###Shelton Stevens

library("tidyverse")

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.7 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library("tidymodels")

## ── Attaching packages ────────────────────────────────────── tidymodels 0.2.0 ──

## ✔ broom 0.8.0 ✔ rsample 0.1.1  
## ✔ dials 0.1.1 ✔ tune 0.2.0  
## ✔ infer 1.0.0 ✔ workflows 0.2.6  
## ✔ modeldata 0.1.1 ✔ workflowsets 0.2.1  
## ✔ parsnip 0.2.1 ✔ yardstick 0.0.9  
## ✔ recipes 0.2.0

## ── Conflicts ───────────────────────────────────────── tidymodels\_conflicts() ──  
## ✖ scales::discard() masks purrr::discard()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ recipes::fixed() masks stringr::fixed()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ yardstick::spec() masks readr::spec()  
## ✖ recipes::step() masks stats::step()  
## • Search for functions across packages at https://www.tidymodels.org/find/

library("GGally")

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library("ggcorrplot")  
library("gridExtra")

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library("dplyr")  
library("skimr")  
library("esquisse")  
library("corrr")

##   
## Attaching package: 'corrr'

## The following object is masked from 'package:skimr':  
##   
## focus

library("ranger")  
library("randomForest")

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:ranger':  
##   
## importance

## The following object is masked from 'package:gridExtra':  
##   
## combine

## The following object is masked from 'package:dplyr':  
##   
## combine

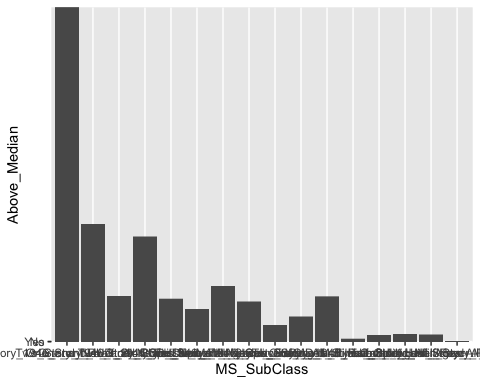
## The following object is masked from 'package:ggplot2':  
##   
## margin

library("ggplot2")

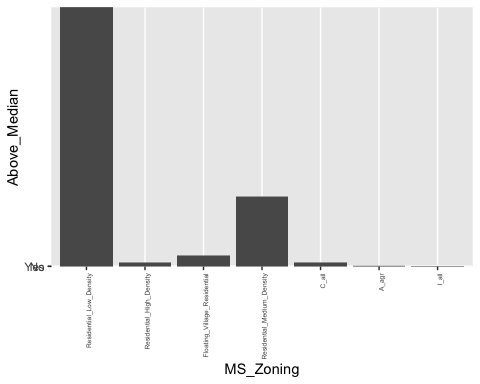
HomeSales <- read.csv("ames\_student.csv")   
HomeSales <- subset( HomeSales, select = -c(Functional, Garage\_Cars, Three\_season\_porch, Pool\_Area, Pool\_QC, Misc\_Feature, Misc\_Val))  
HomeSales = HomeSales %>% mutate\_if(is.character,as\_factor)

##Graphing Variables

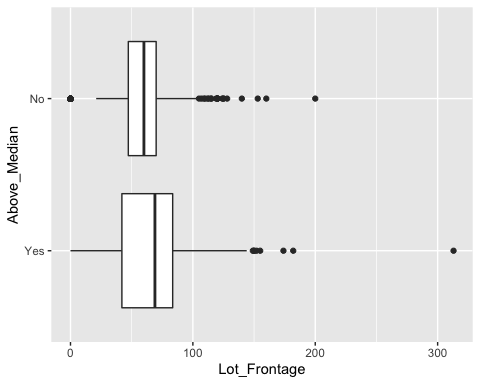
ggplot(data=HomeSales, aes(x=MS\_SubClass, y=Above\_Median)) +  
 geom\_bar(stat="identity")



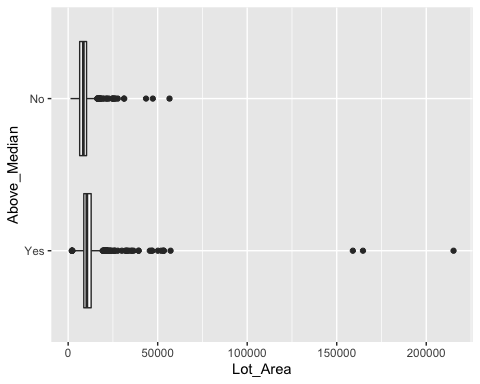
ggplot(data=HomeSales, aes(x=MS\_Zoning, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1, size=5))



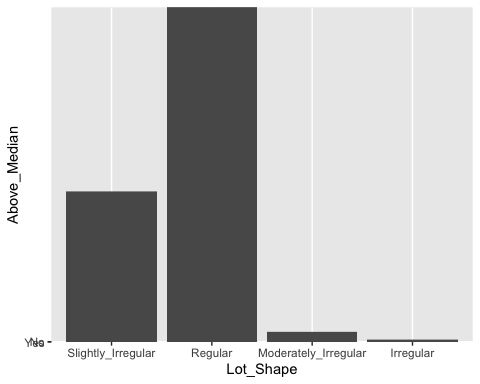
ggplot(HomeSales, aes(x=Lot\_Frontage, y=Above\_Median)) +   
 geom\_boxplot()



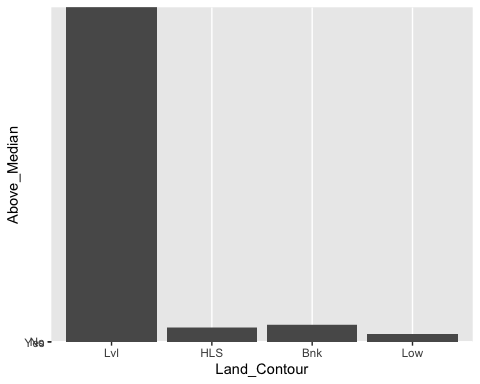
ggplot(HomeSales, aes(x=Lot\_Area, y=Above\_Median)) +   
 geom\_boxplot()



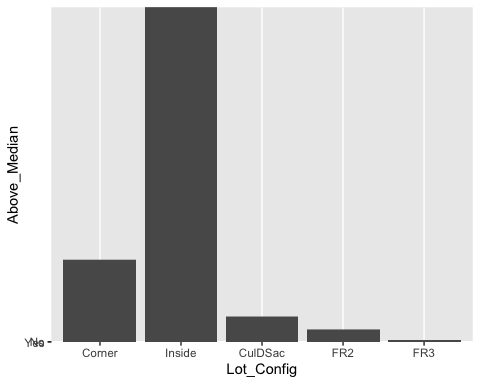
ggplot(data=HomeSales, aes(x=Lot\_Shape, y=Above\_Median)) +  
 geom\_bar(stat="identity")



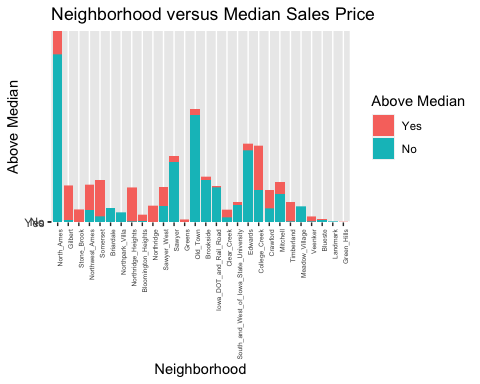
ggplot(data=HomeSales, aes(x=Land\_Contour, y=Above\_Median)) +  
 geom\_bar(stat="identity")



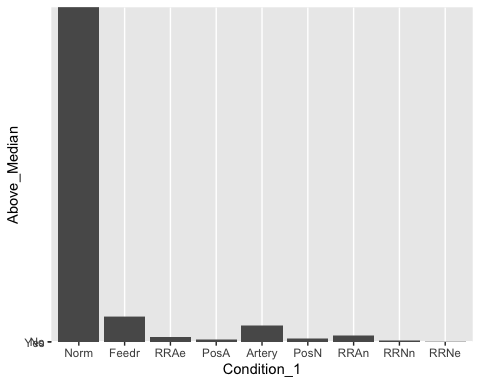
ggplot(data=HomeSales, aes(x=Lot\_Config, y=Above\_Median)) +  
 geom\_bar(stat="identity")



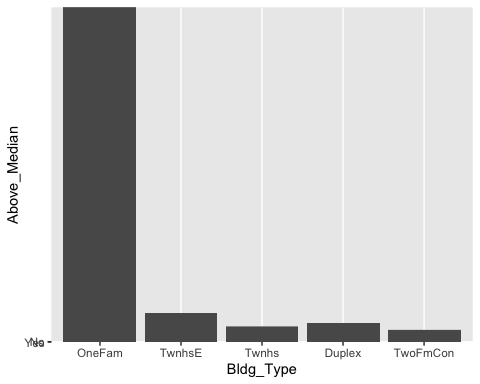
ggplot(data=HomeSales, aes(x=Neighborhood, y=Above\_Median, fill= Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1, size=5)) +  
 ggtitle("Neighborhood versus Median Sales Price") +  
 labs(y= "Above Median", x = "Neighborhood") +  
 labs(fill = "Above Median")



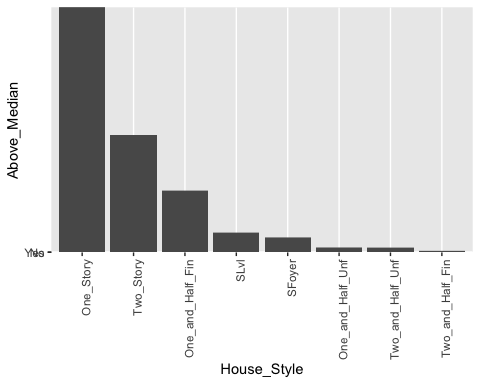
ggplot(data=HomeSales, aes(x=Condition\_1, y=Above\_Median)) +  
 geom\_bar(stat="identity")



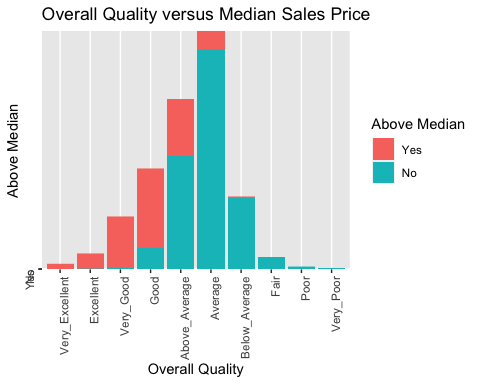
ggplot(data=HomeSales, aes(x=Bldg\_Type, y=Above\_Median)) +  
 geom\_bar(stat="identity")



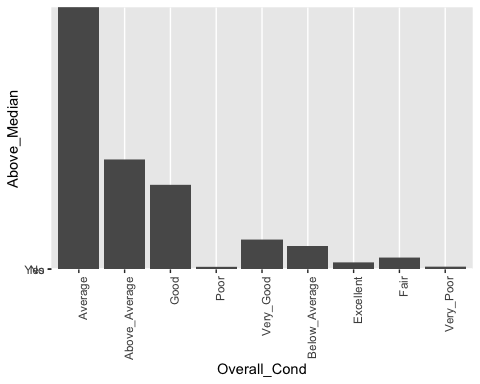
ggplot(data=HomeSales, aes(x=House\_Style, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



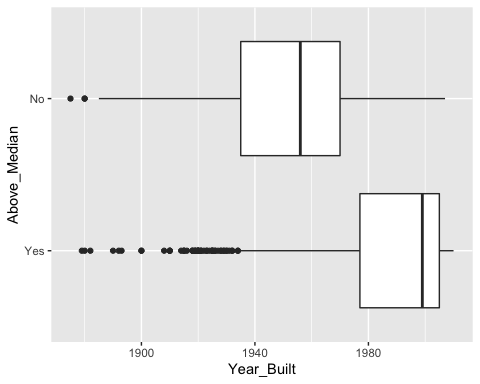
HomeSales$Overall\_Qual <-factor(HomeSales$Overall\_Qual, levels = c("Very\_Excellent", "Excellent", "Very\_Good", "Good", "Above\_Average", "Average", "Below\_Average", "Fair", "Poor", "Very\_Poor"))  
ggplot(data=HomeSales, aes(x=Overall\_Qual, y=Above\_Median, fill=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1)) +  
 theme(axis.text.y = element\_text(angle=90, vjust=.5)) +  
 ggtitle("Overall Quality versus Median Sales Price") +  
 labs(y= "Above Median", x = "Overall Quality") +  
 labs(fill = "Above Median")



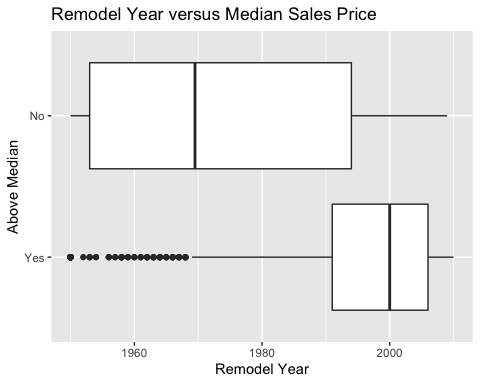
ggplot(data=HomeSales, aes(x=Overall\_Cond, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



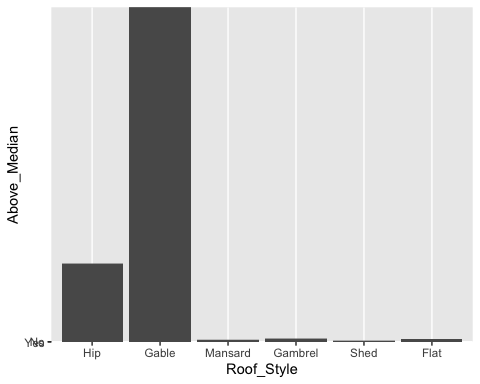
ggplot(HomeSales, aes(x=Year\_Built, y=Above\_Median)) +   
 geom\_boxplot()



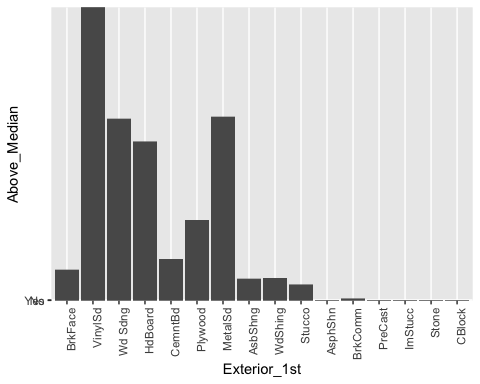
ggplot(HomeSales, aes(x=Year\_Remod\_Add, y=Above\_Median)) +   
 geom\_boxplot() +  
 ggtitle("Remodel Year versus Median Sales Price") +  
 labs(y= "Above Median", x = "Remodel Year")



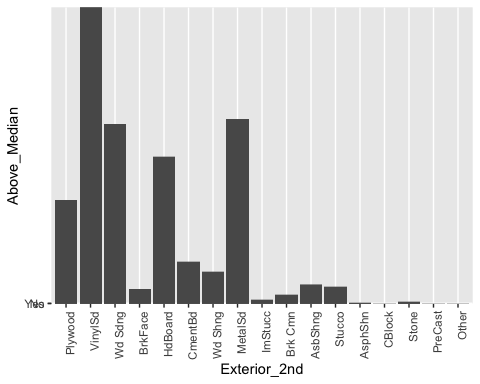
ggplot(data=HomeSales, aes(x=Roof\_Style, y=Above\_Median)) +  
 geom\_bar(stat="identity")



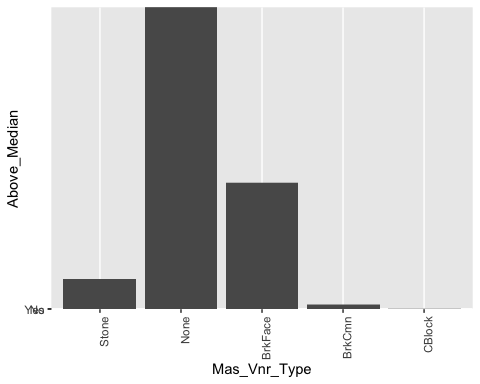
ggplot(data=HomeSales, aes(x=Exterior\_1st, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



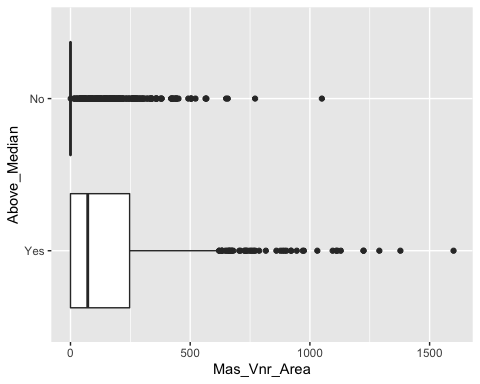
ggplot(data=HomeSales, aes(x=Exterior\_2nd, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



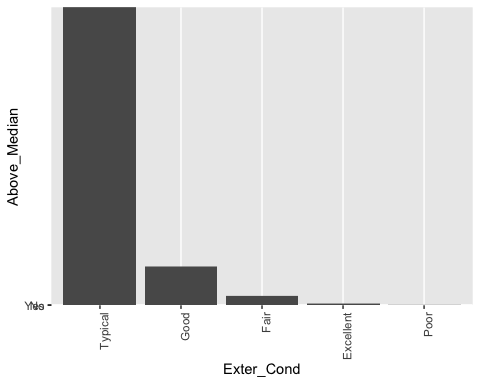
ggplot(data=HomeSales, aes(x=Mas\_Vnr\_Type, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



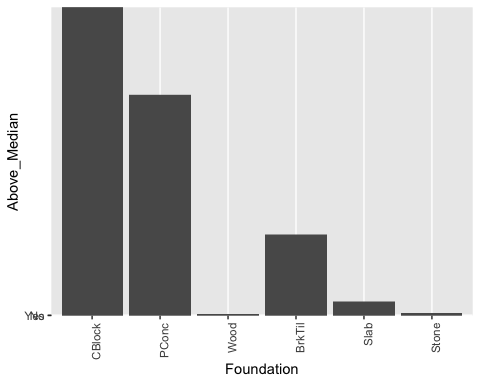
ggplot(HomeSales, aes(x=Mas\_Vnr\_Area, y=Above\_Median)) +   
 geom\_boxplot()



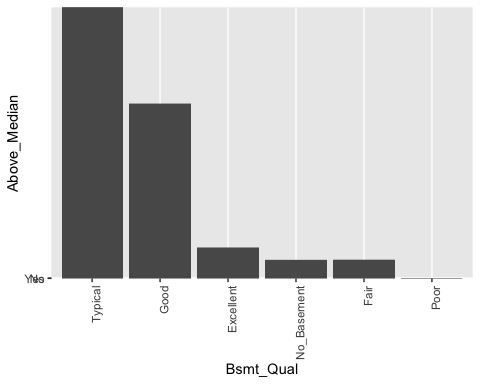
ggplot(data=HomeSales, aes(x=Exter\_Cond, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



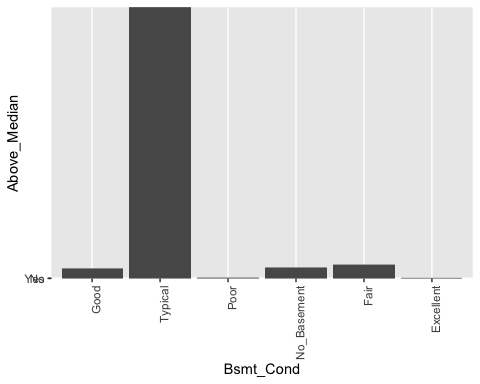
ggplot(data=HomeSales, aes(x=Foundation, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



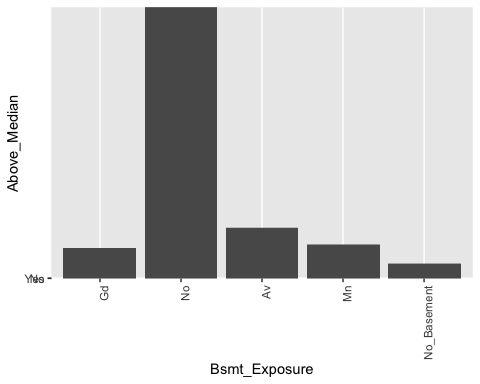
ggplot(data=HomeSales, aes(x=Bsmt\_Qual, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



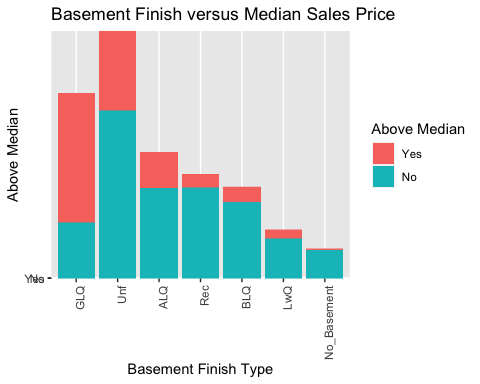
ggplot(data=HomeSales, aes(x=Bsmt\_Cond, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



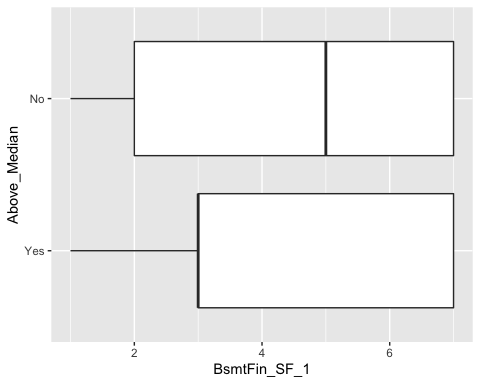
ggplot(data=HomeSales, aes(x=Bsmt\_Exposure, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



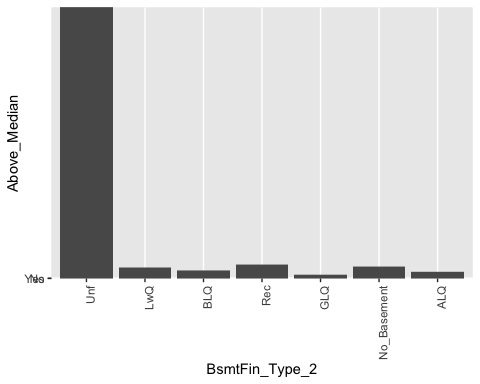
HomeSales$BsmtFin\_Type\_1 <-factor(HomeSales$BsmtFin\_Type\_1, levels = c("GLQ", "Unf", "ALQ", "Rec", "BLQ", "LwQ", "No\_Basement"))  
ggplot(data=HomeSales, aes(x=BsmtFin\_Type\_1, y=Above\_Median, fill=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1)) +  
 ggtitle("Basement Finish versus Median Sales Price") +  
 labs(y= "Above Median", x = "Basement Finish Type") +  
 labs(fill = "Above Median")



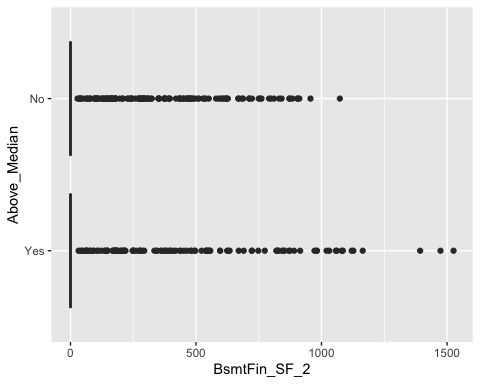
ggplot(HomeSales, aes(x=BsmtFin\_SF\_1, y=Above\_Median)) +   
 geom\_boxplot()



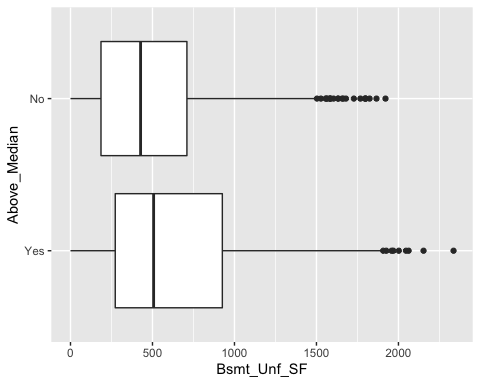
ggplot(data=HomeSales, aes(x=BsmtFin\_Type\_2, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



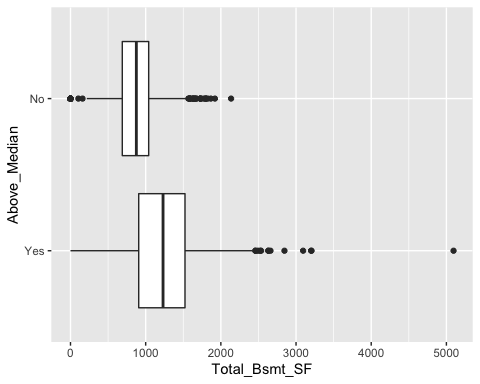
ggplot(HomeSales, aes(x=BsmtFin\_SF\_2, y=Above\_Median)) +   
 geom\_boxplot()



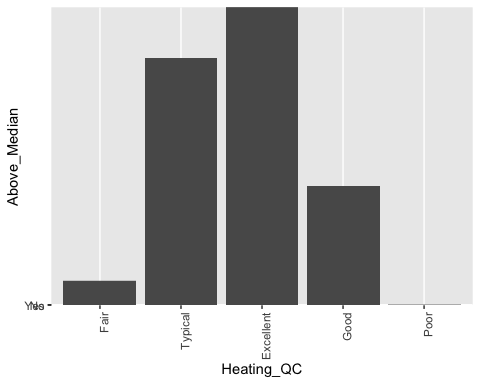
ggplot(HomeSales, aes(x=Bsmt\_Unf\_SF, y=Above\_Median)) +   
 geom\_boxplot()



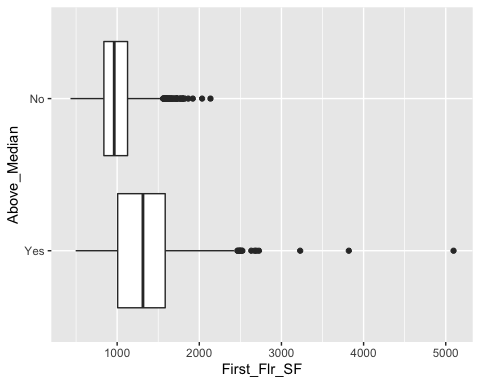
ggplot(HomeSales, aes(x=Total\_Bsmt\_SF, y=Above\_Median)) +   
 geom\_boxplot()



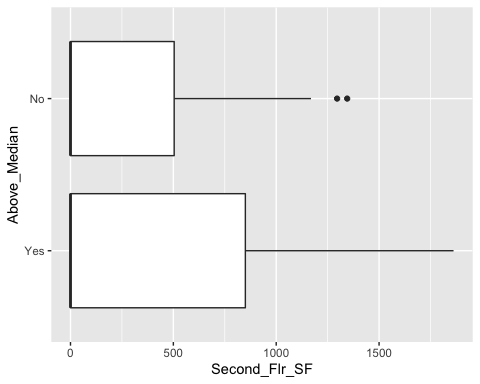
ggplot(data=HomeSales, aes(x=Heating\_QC, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



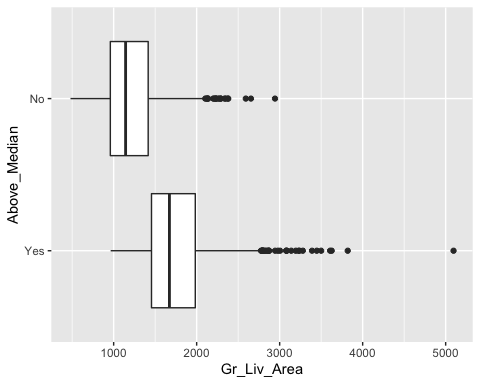
ggplot(HomeSales, aes(x=First\_Flr\_SF, y=Above\_Median)) +   
 geom\_boxplot()



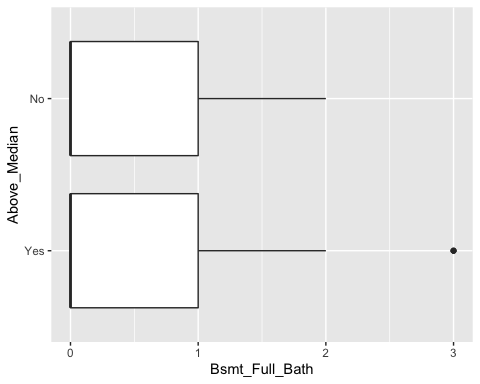
ggplot(HomeSales, aes(x=Second\_Flr\_SF, y=Above\_Median)) +   
 geom\_boxplot()



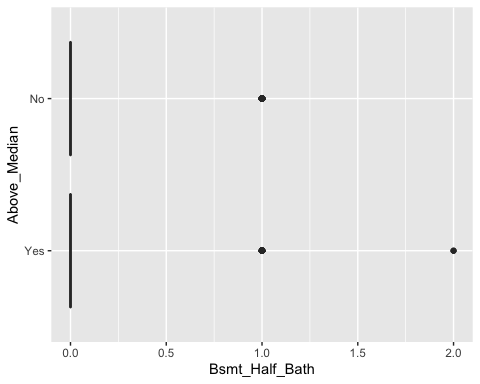
ggplot(HomeSales, aes(x=Gr\_Liv\_Area, y=Above\_Median)) +   
 geom\_boxplot()



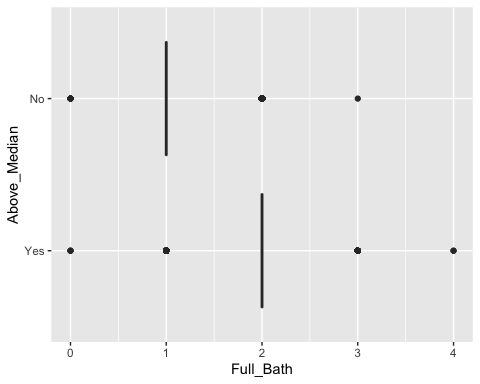
ggplot(HomeSales, aes(x=Bsmt\_Full\_Bath, y=Above\_Median)) +   
 geom\_boxplot()



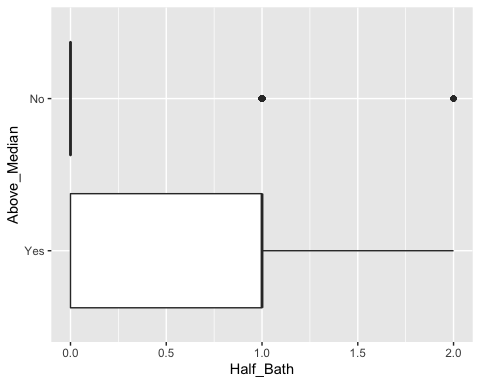
ggplot(HomeSales, aes(x=Bsmt\_Half\_Bath, y=Above\_Median)) +   
 geom\_boxplot()



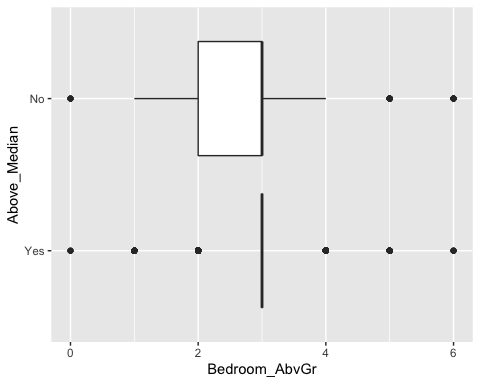
ggplot(HomeSales, aes(x=Full\_Bath, y=Above\_Median)) +   
 geom\_boxplot()



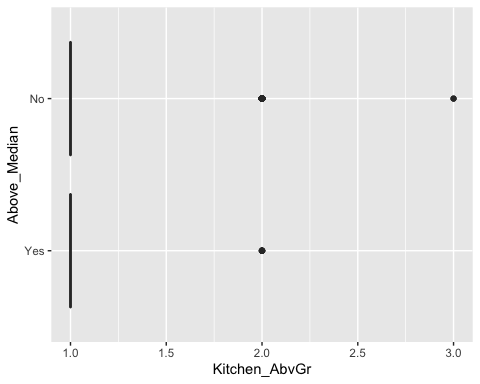
ggplot(HomeSales, aes(x=Half\_Bath, y=Above\_Median)) +   
 geom\_boxplot()



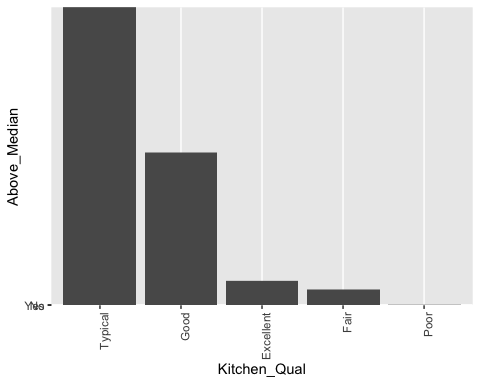
ggplot(HomeSales, aes(x=Bedroom\_AbvGr, y=Above\_Median)) +   
 geom\_boxplot()



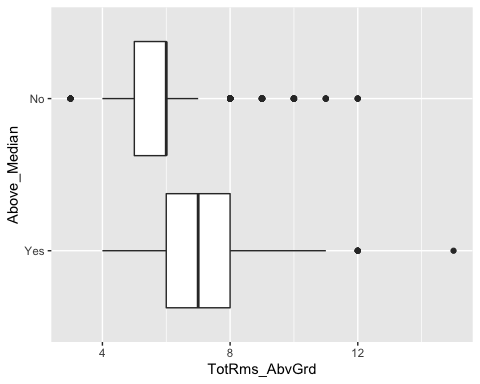
ggplot(HomeSales, aes(x=Kitchen\_AbvGr, y=Above\_Median)) +   
 geom\_boxplot()



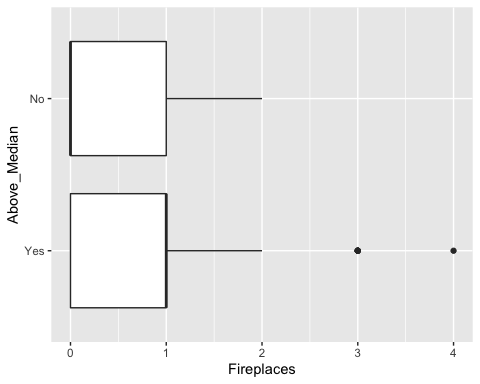
ggplot(data=HomeSales, aes(x=Kitchen\_Qual, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



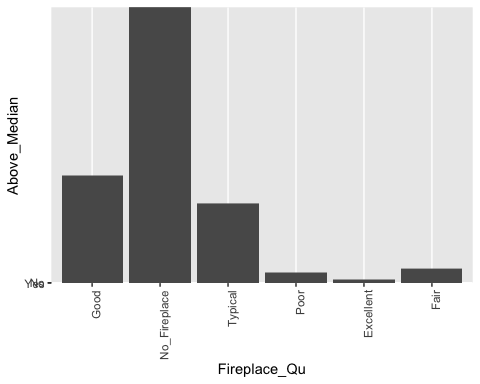
ggplot(HomeSales, aes(x=TotRms\_AbvGrd, y=Above\_Median)) +   
 geom\_boxplot()



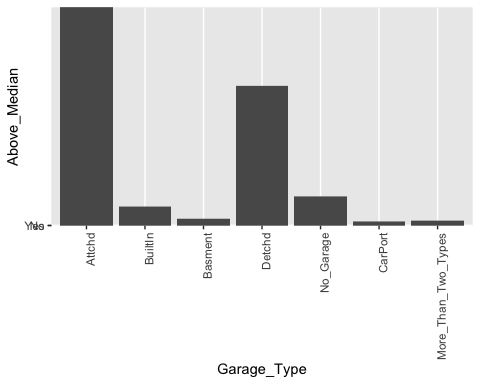
ggplot(HomeSales, aes(x=Fireplaces, y=Above\_Median)) +   
 geom\_boxplot()



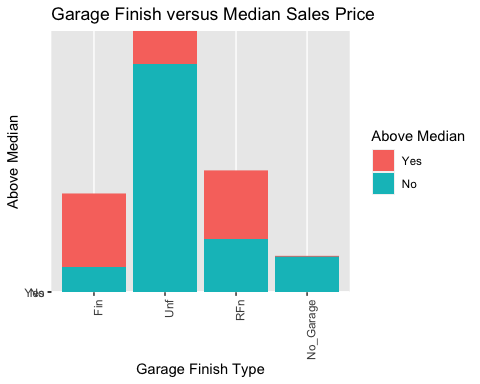
ggplot(data=HomeSales, aes(x=Fireplace\_Qu, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



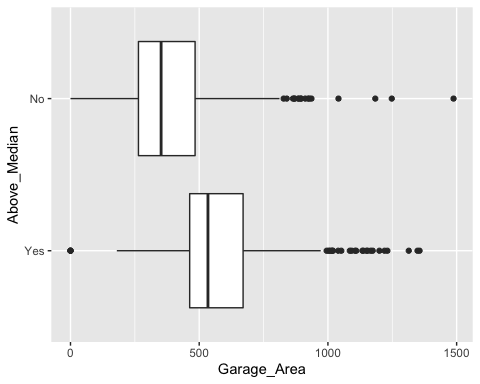
ggplot(data=HomeSales, aes(x=Garage\_Type, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



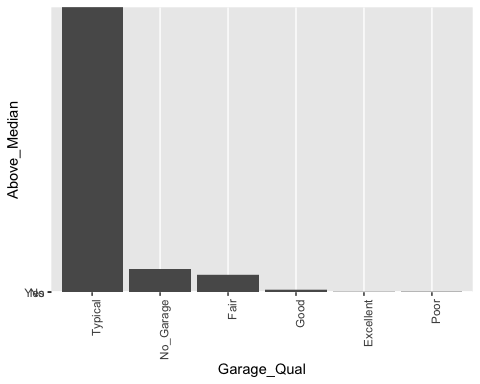
ggplot(data=HomeSales, aes(x=Garage\_Finish, y=Above\_Median, fill=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1)) +  
 ggtitle("Garage Finish versus Median Sales Price") +  
 labs(y= "Above Median", x = "Garage Finish Type") +  
 labs(fill = "Above Median")



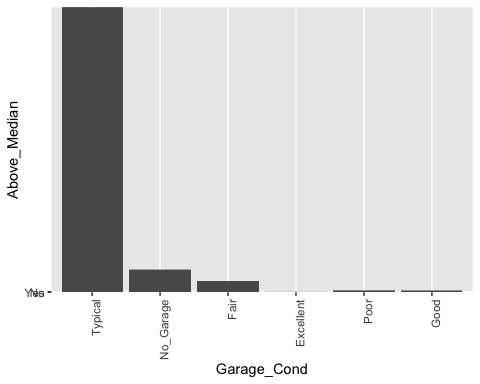
ggplot(HomeSales, aes(x=Garage\_Area, y=Above\_Median)) +   
 geom\_boxplot()



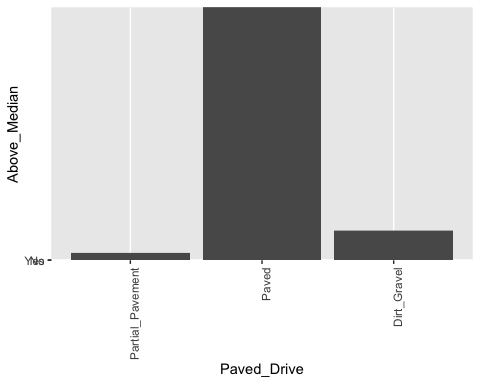
ggplot(data=HomeSales, aes(x=Garage\_Qual, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



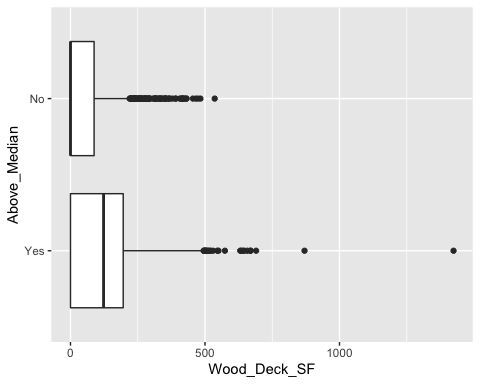
ggplot(data=HomeSales, aes(x=Garage\_Cond, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



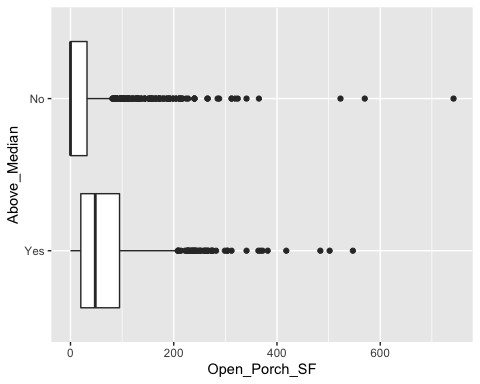
ggplot(data=HomeSales, aes(x=Paved\_Drive, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



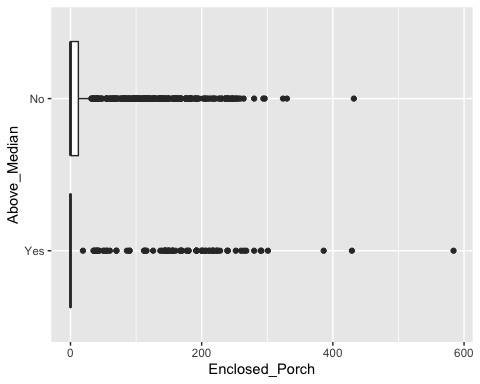
ggplot(HomeSales, aes(x=Wood\_Deck\_SF, y=Above\_Median)) +   
 geom\_boxplot()



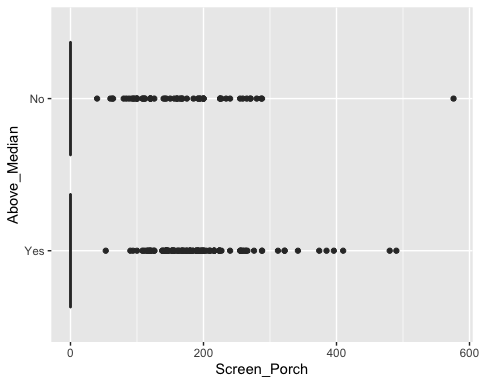
ggplot(HomeSales, aes(x=Open\_Porch\_SF, y=Above\_Median)) +   
 geom\_boxplot()



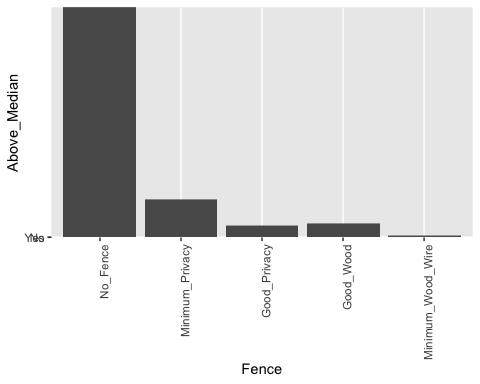
ggplot(HomeSales, aes(x=Enclosed\_Porch, y=Above\_Median)) +   
 geom\_boxplot()



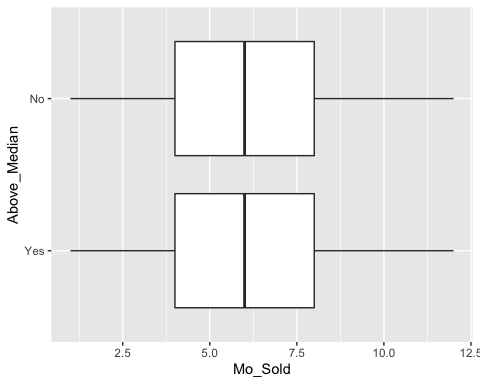
ggplot(HomeSales, aes(x=Screen\_Porch, y=Above\_Median)) +   
 geom\_boxplot()



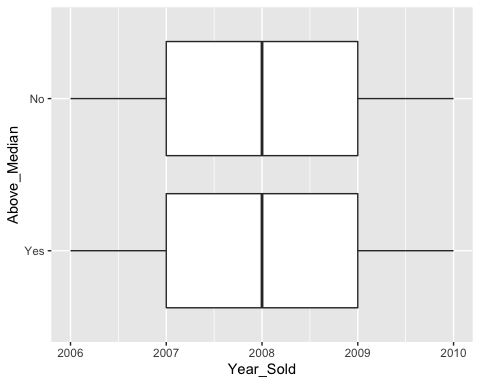
ggplot(data=HomeSales, aes(x=Fence, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



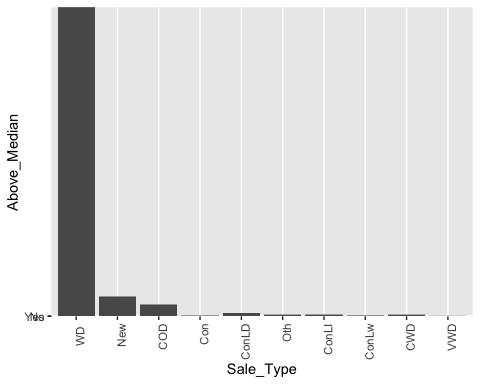
ggplot(HomeSales, aes(x=Mo\_Sold, y=Above\_Median)) +   
 geom\_boxplot()



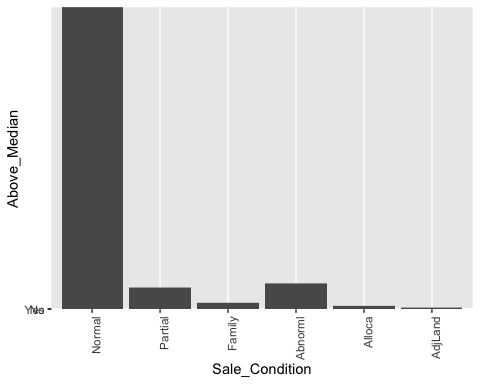
ggplot(HomeSales, aes(x=Year\_Sold, y=Above\_Median)) +   
 geom\_boxplot()



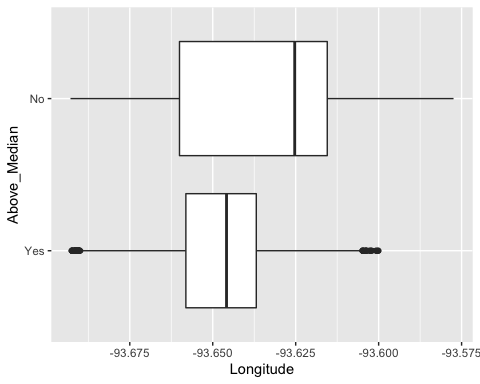
ggplot(data=HomeSales, aes(x=Sale\_Type, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



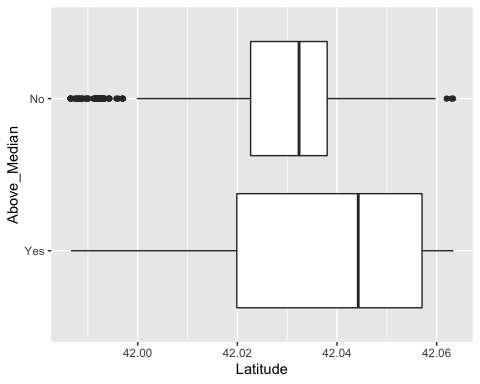
ggplot(data=HomeSales, aes(x=Sale\_Condition, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))



ggplot(HomeSales, aes(x=Longitude, y=Above\_Median)) +   
 geom\_boxplot()



ggplot(HomeSales, aes(x=Latitude, y=Above\_Median)) +   
 geom\_boxplot()



ggplot(data=HomeSales, aes(x=Above\_Median, y=Above\_Median)) +  
 geom\_bar(stat="identity") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1))

