

Project 2 Visuals

Alanna Hazlett

2024-04-26

```
Data<-read.csv("kc_house_data.csv", sep=",", header=TRUE)
Data[15871,]

##           id      date  price bedrooms bathrooms sqft_living sqft_lot
## 15871 2402100895 20140625T000000 640000       33     1.75      1620      6000
##      floors waterfront view condition grade sqft_above sqft_basement yr_built
## 15871      1          0    0       5     7      1040            580      1947
##      yr_renovated zipcode      lat      long sqft_living15 sqft_lot15
## 15871          0   98103 47.6878 -122.331        1330        4700

Data[15871,4] <- 3
Data<-Data[-c(876,1150,3120,5833,6995,9774,9855,10482,14424,19453), ]

Data$house_age<-2024 - Data$yr_built
Data$yrs_reno<-ifelse(Data$yr_renovated == 0,0, 2024 - Data$yr_renovated)

set.seed(6021)
sample.data<-sample.int(nrow(Data), floor(.50*nrow(Data)), replace = F)
train<-Data[sample.data, ]
test<-Data[-sample.data, ]

train<-train[,-c(1,2,15,16,18,19,20,21)]

ystar<-log(train$price)
train<-data.frame(train,ystar)

sqft_lot_star<- log(train$sqft_lot)
train<-data.frame(train,sqft_lot_star)

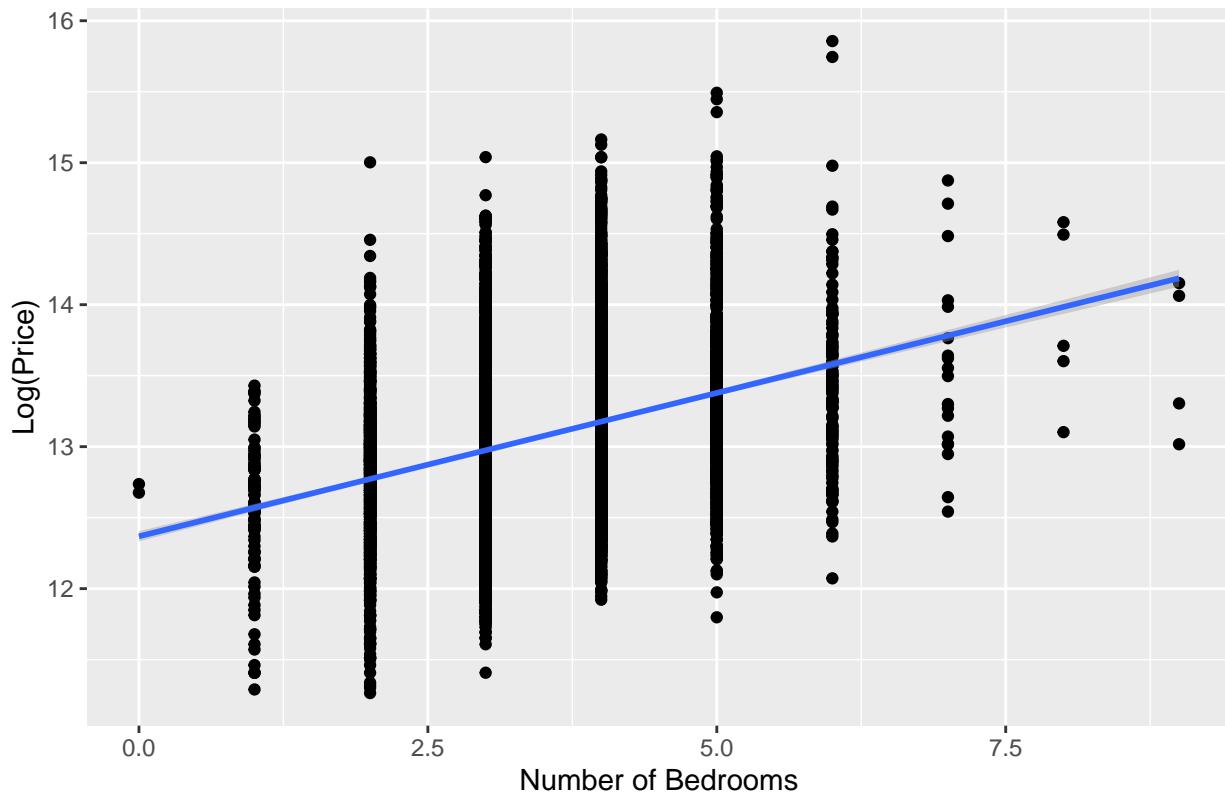
#transforming sqft_lot in test df
sqft_lot_star<- log(test$sqft_lot)
test<-data.frame(test,sqft_lot_star)
#transforming price in test df
ystar<-log(test$price)
test<-data.frame(test,ystar)
#narrowing test df down to predictors in our model only
test<-test[,c(25,6,12,22,9,10,5,4,24,11,8)]
```

Predictors VS Response for Linear Regression Model

```
ggplot2::ggplot(train, ggplot2::aes(x=bedrooms, y=ystar))+
  geom_point()+
  geom_smooth(method = lm)+
  labs(x="Number of Bedrooms", y="Log(Price)", title="Log(Price) against Number of Bedrooms")
```

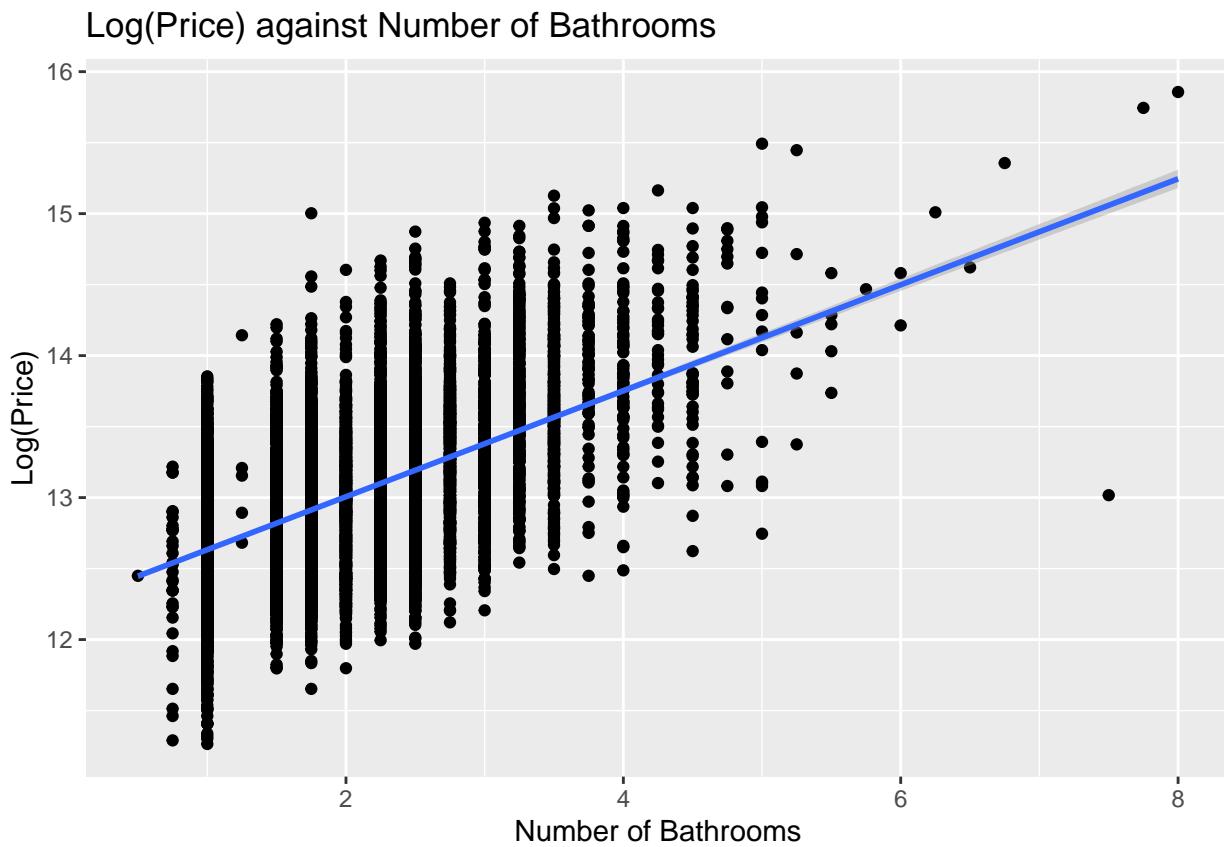
```
## `geom_smooth()` using formula = 'y ~ x'
```

Log(Price) against Number of Bedrooms



```
ggplot2::ggplot(train, ggplot2::aes(x=bathrooms, y=ystar))+  
  geom_point() +  
  geom_smooth(method = lm) +  
  labs(x="Number of Bathrooms", y="Log(Price)", title="Log(Price) against Number of Bathrooms")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



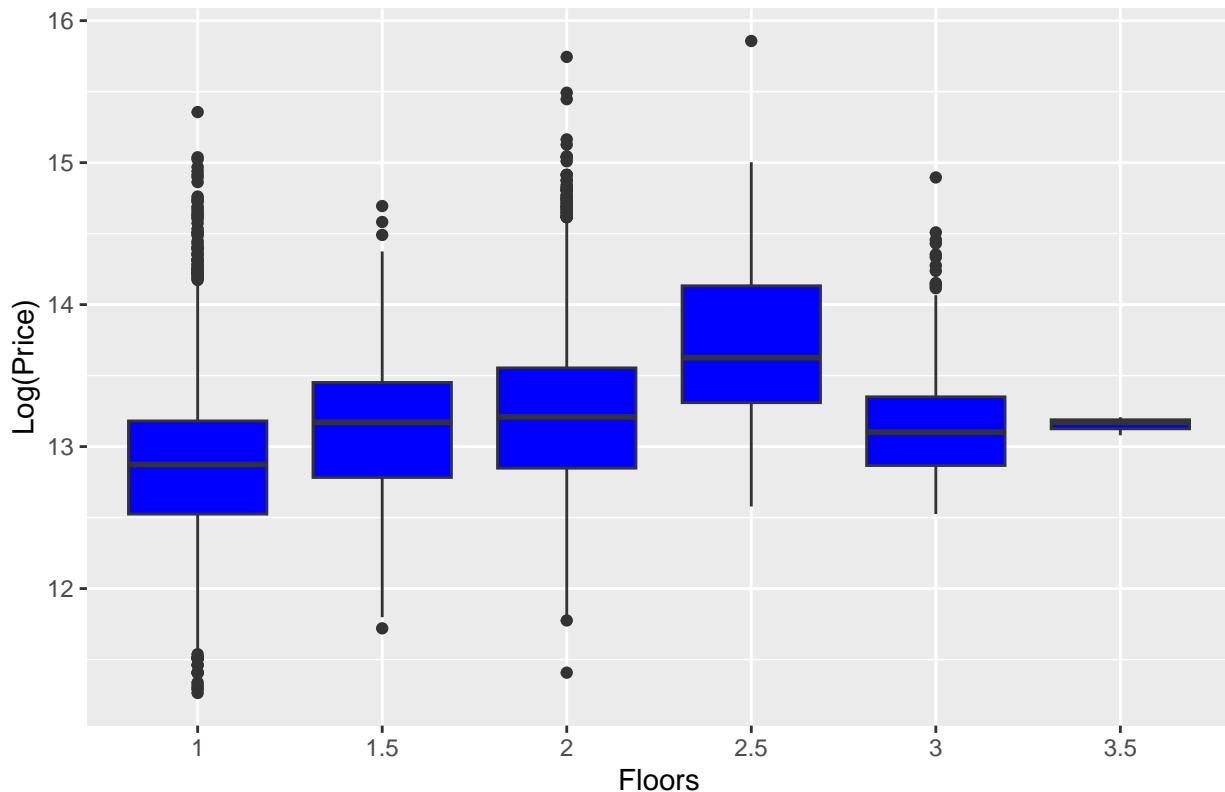
```
ggplot2::ggplot(train, ggplot2::aes(x=sqft_living, y=ystar))+
  geom_point()+
  geom_smooth(method = lm)+
  labs(x="Square Footage of Living Space", y="Log(Price)", title="Log(Price) against Square Footage of Living Space")
## `geom_smooth()` using formula = 'y ~ x'
```

Log(Price) against Square Footage of Living Space

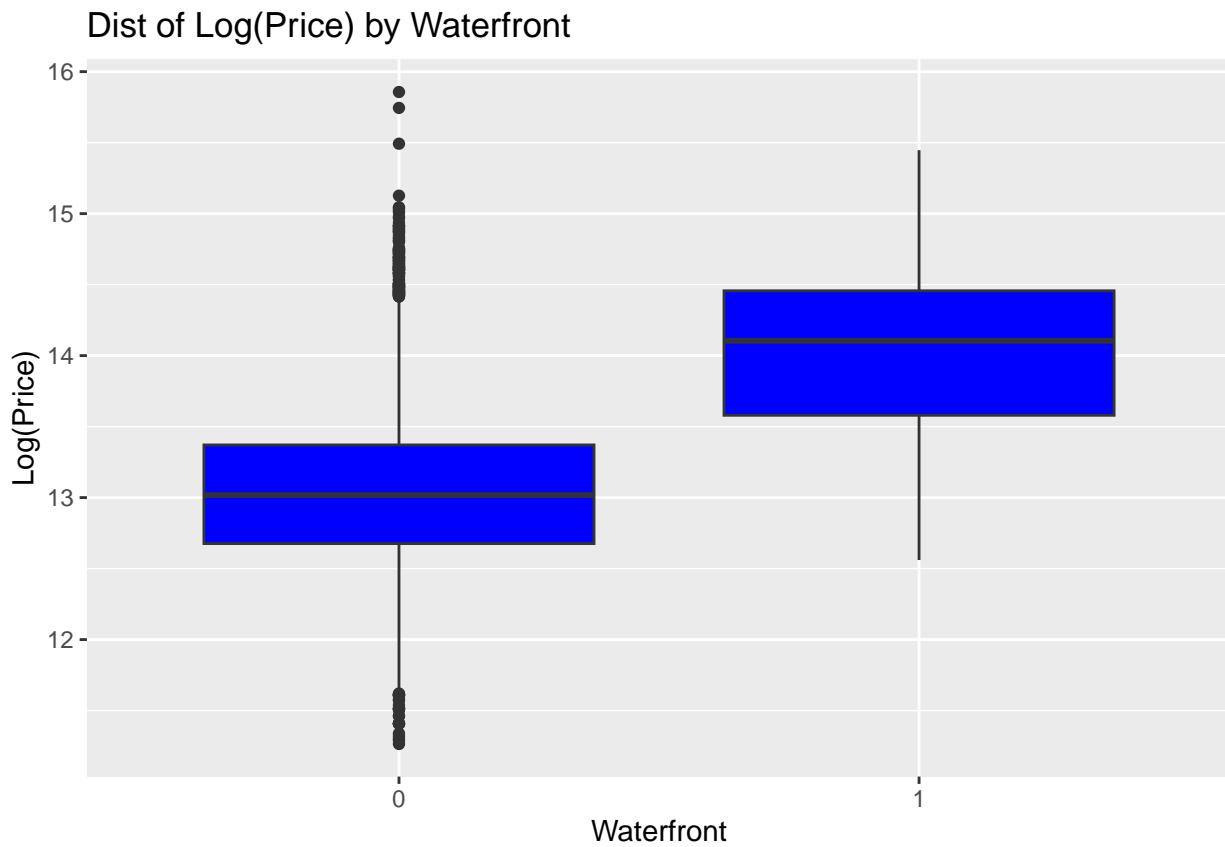


```
train$floors<-factor(train$floors)
ggplot2::ggplot(train, ggplot2::aes(x=floors, y=ystar))+
  geom_boxplot(fill="Blue")+
  labs(x="Floors", y="Log(Price)", title="Dist of Log(Price) by Floors")
```

Dist of Log(Price) by Floors

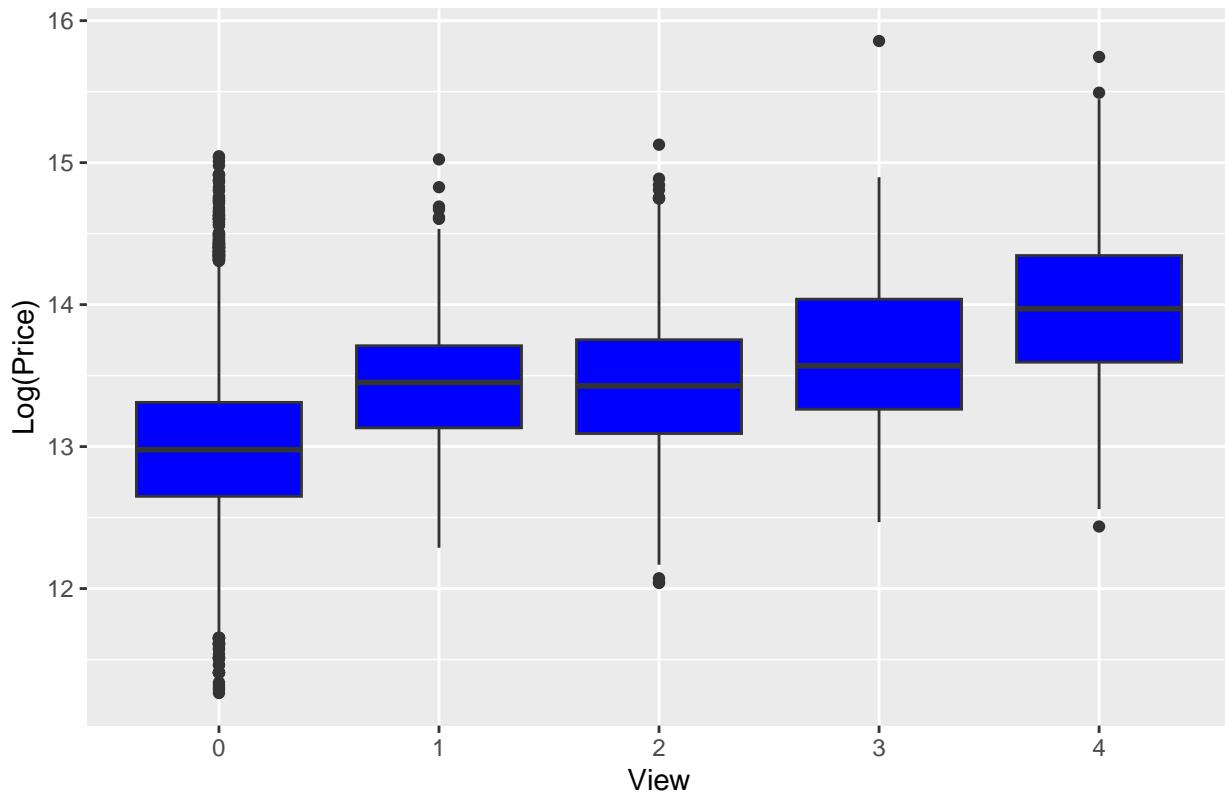


```
train$waterfront<-factor(train$waterfront)
ggplot2::ggplot(train, ggplot2::aes(x=waterfront, y=ystar))+
  geom_boxplot(fill="Blue")+
  labs(x="Waterfront", y="Log(Price)", title="Dist of Log(Price) by Waterfront")
```



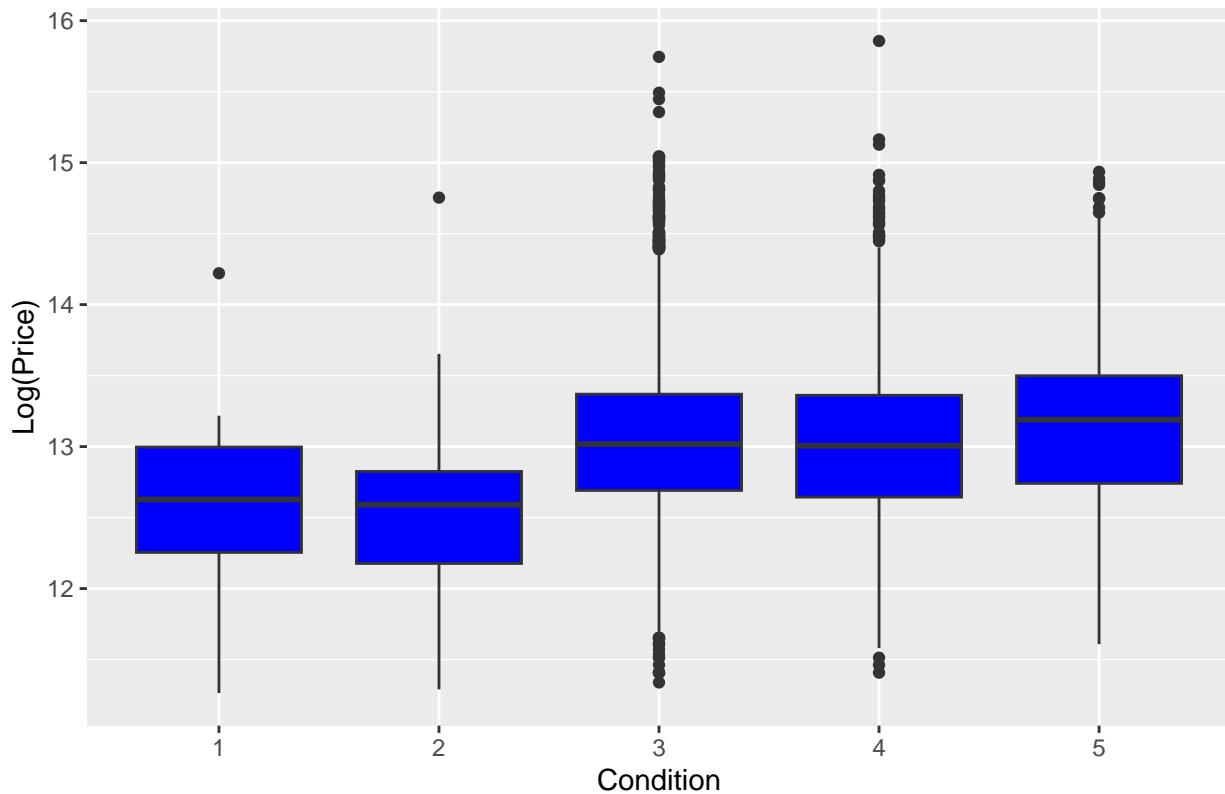
```
train$view<-factor(train$view)
ggplot2::ggplot(train, ggplot2::aes(x=view, y=ystar))+
  geom_boxplot(fill="Blue")+
  labs(x="View", y="Log(Price)", title="Dist of Log(Price) by View")
```

Dist of Log(Price) by View



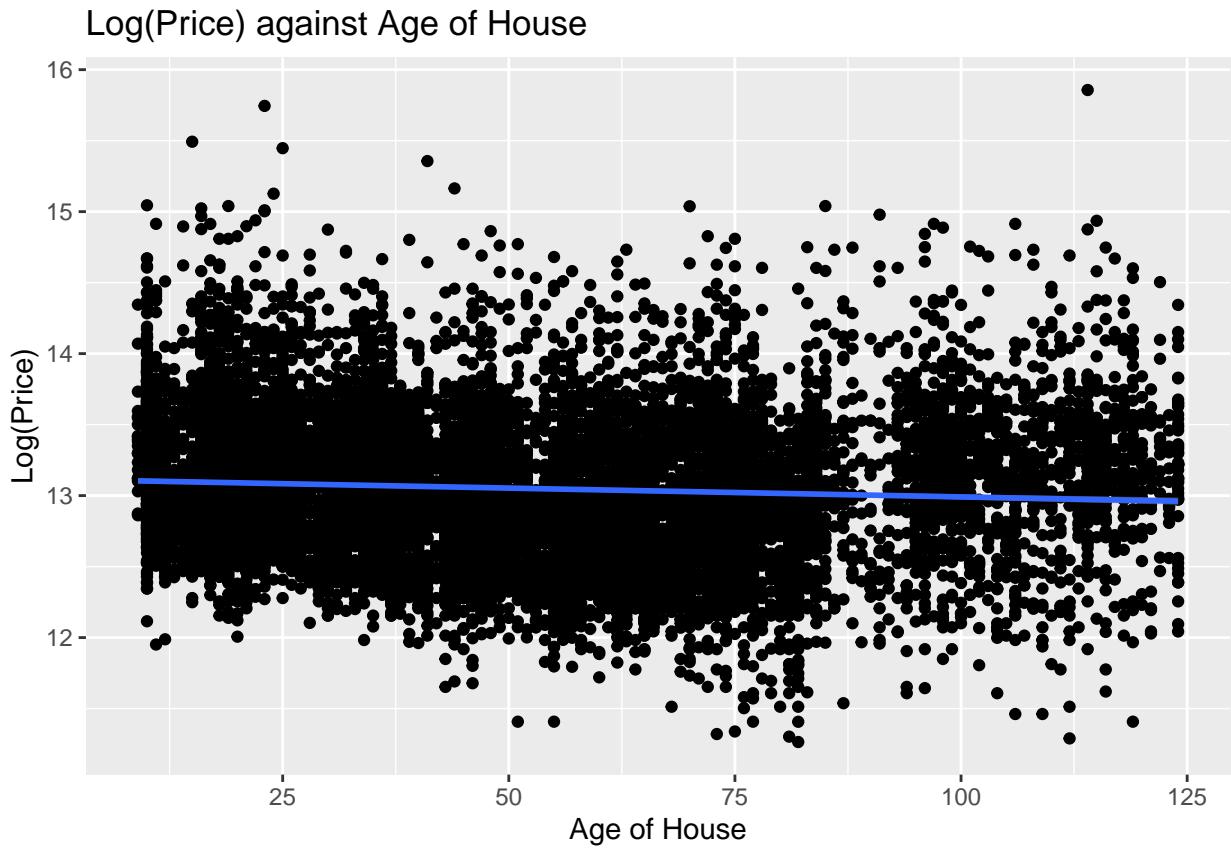
```
train$condition<-factor(train$condition)
ggplot2::ggplot(train, ggplot2::aes(x=condition, y=ystar))+
  geom_boxplot(fill="Blue")+
  labs(x="Condition", y="Log(Price)", title="Dist of Log(Price) by Condition")
```

Dist of Log(Price) by Condition



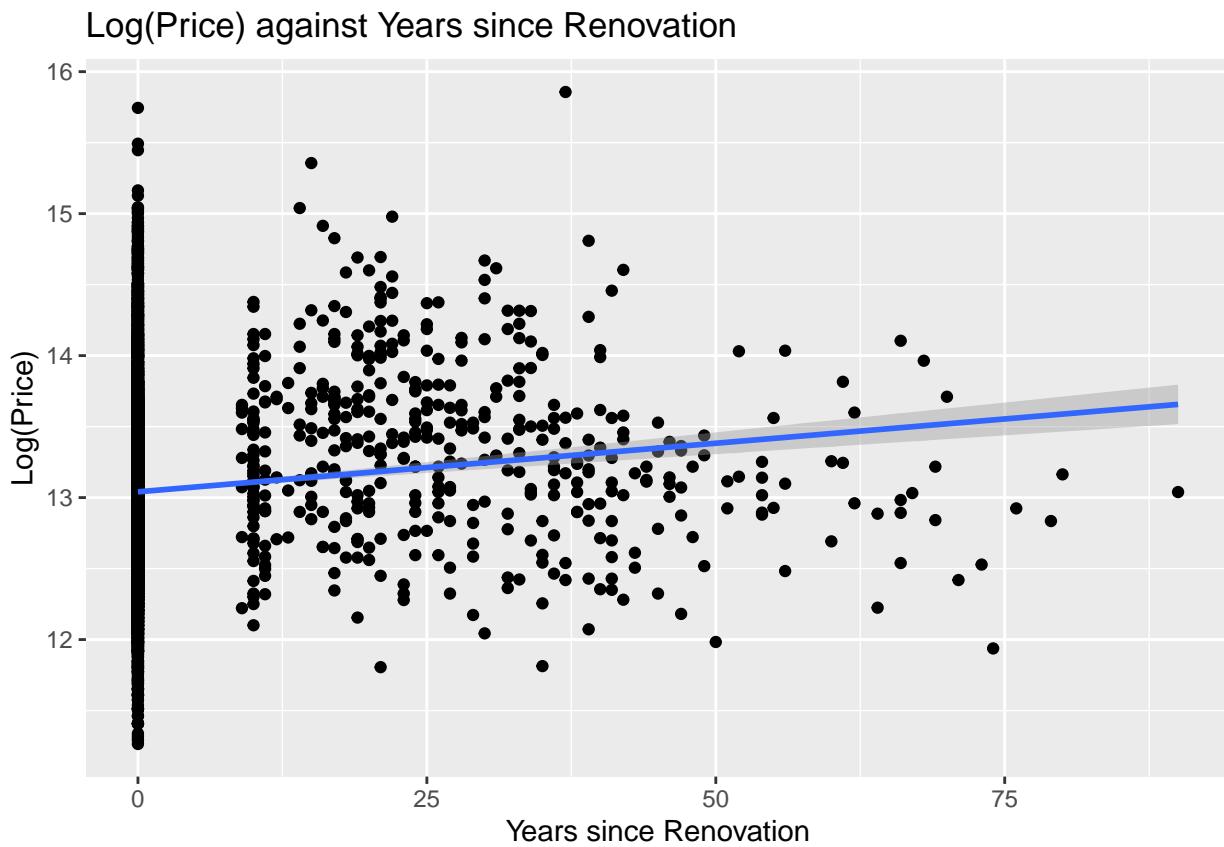
```
ggplot2::ggplot(train, ggplot2::aes(x=house_age, y=ystar))+
  geom_point()+
  geom_smooth(method = lm)+
  labs(x="Age of House", y="Log(Price)", title="Log(Price) against Age of House")

## `geom_smooth()` using formula = 'y ~ x'
```



```
ggplot2::ggplot(train, ggplot2::aes(x=yrs_reno, y=ystar))+
  geom_point()+
  geom_smooth(method = lm)+
  labs(x="Years since Renovation", y="Log(Price)", title="Log(Price) against Years since Renovation")

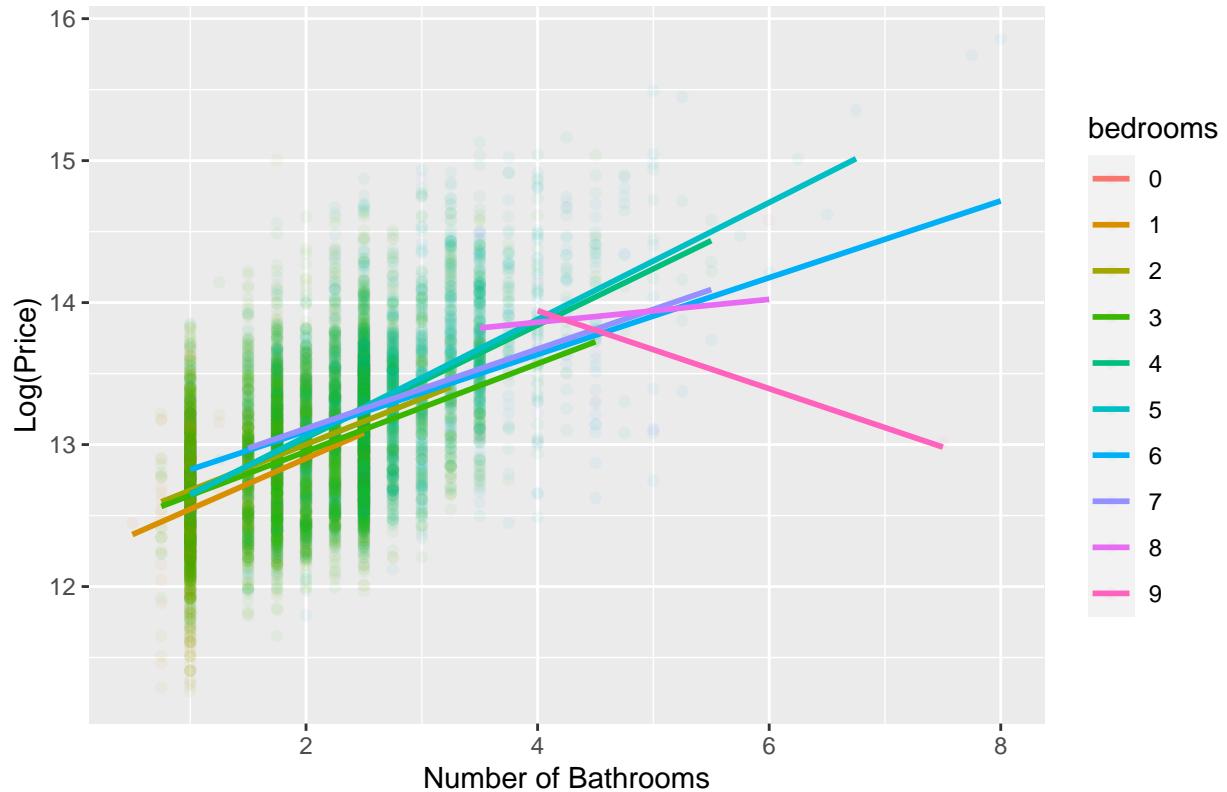
## `geom_smooth()` using formula = 'y ~ x'
```



Misc Visuals

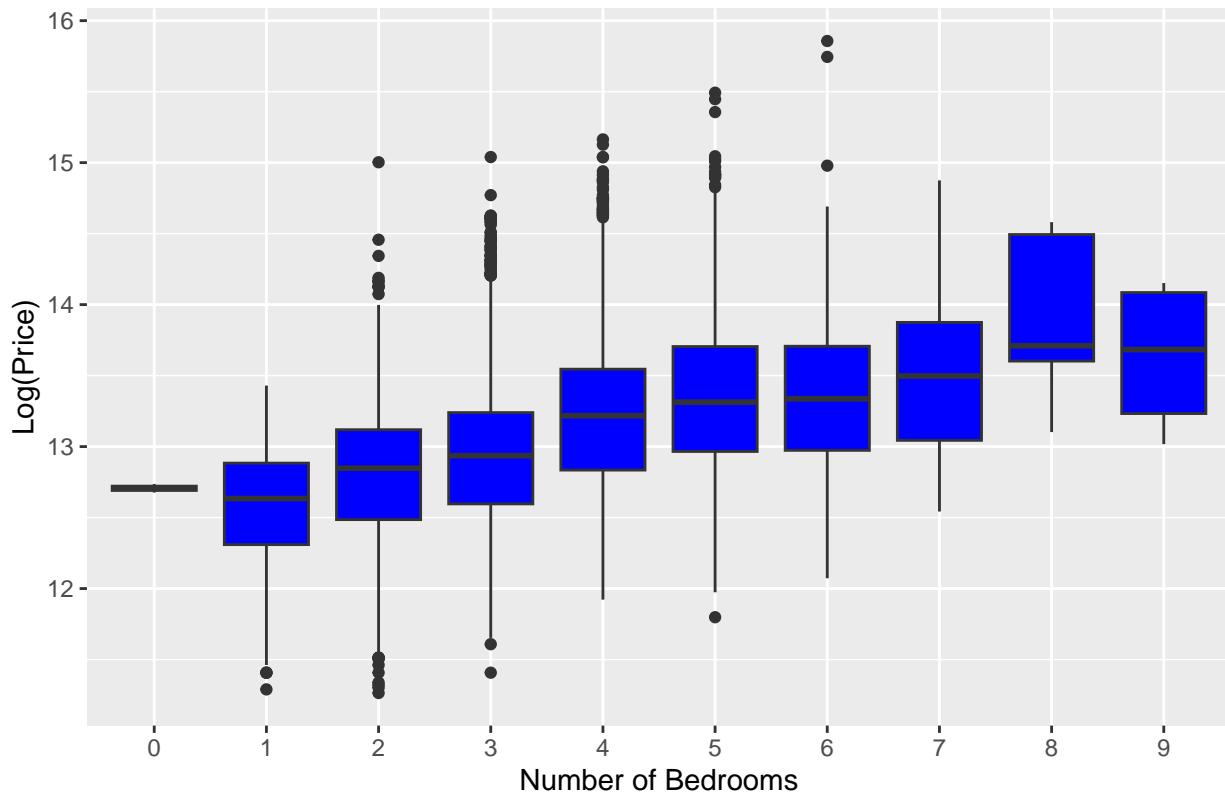
```
train$bedrooms<-factor(train$bedrooms)
ggplot2::ggplot(train, ggplot2::aes(x=bathrooms,y=ystar,color=bedrooms))+ 
  geom_point(alpha=0.05)+ 
  geom_smooth(method=lm,se=FALSE)+ 
  labs(x="Number of Bathrooms", y="Log(Price)", title="Effect of Number of Bathrooms and Bedrooms on Log Price")
## `geom_smooth()` using formula = 'y ~ x'
```

Effect of Number of Bathrooms and Bedrooms on Log(Price)



```
ggplot2::ggplot(train, ggplot2::aes(x=bedrooms, y=ystar))+  
  geom_boxplot(fill="Blue") +  
  labs(x="Number of Bedrooms", y="Log(Price)", title="Dist of Log(Price) by Number of Bedrooms")
```

Dist of Log(Price) by Number of Bedrooms



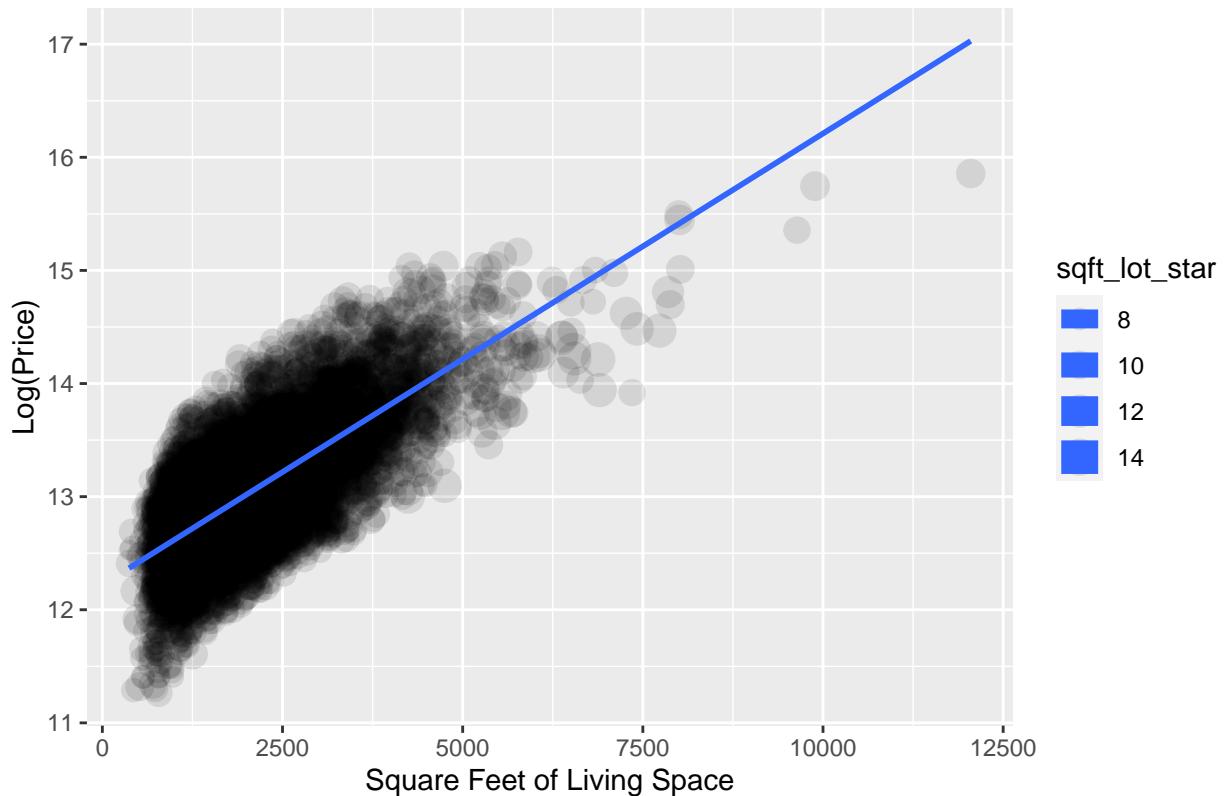
```
ggplot2::ggplot(train, ggplot2::aes(x=sqft_living,y=ystar, size=sqft_lot_star))+
  geom_point(alpha=0.1)+
  geom_smooth(method=lm,se=FALSE)+
  labs(x="Square Feet of Living Space", y= "Log(Price)", title= "Effect of Living Space and Log(Lot Size)

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## `geom_smooth()` using formula = 'y ~ x'

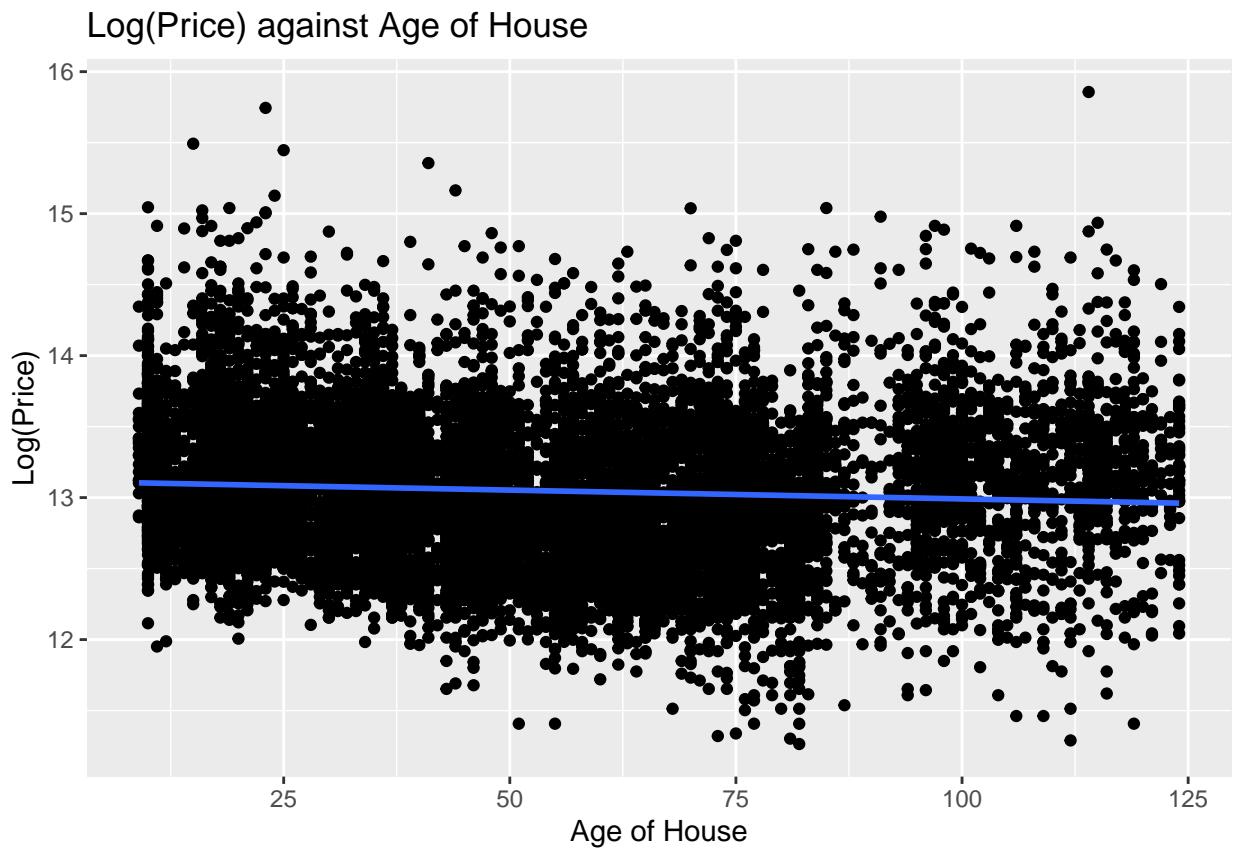
## Warning: The following aesthetics were dropped during statistical transformation: size
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```

Effect of Living Space and Log(Lot Size) on Log(Price)

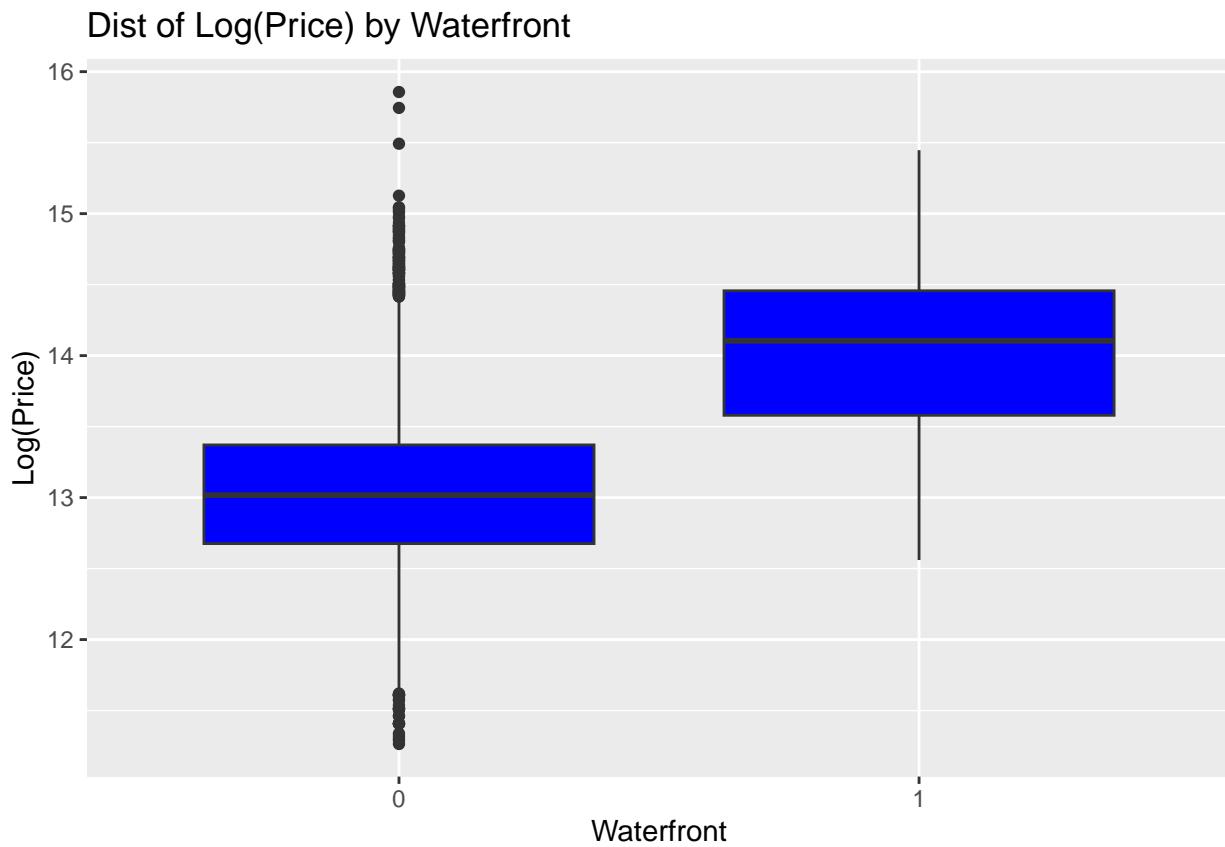


```
ggplot2::ggplot(train, ggplot2::aes(x=house_age,y=ystar))+
  geom_point()+
  geom_smooth(method=lm,se=FALSE)+
  labs(x="Age of House",y="Log(Price)", title="Log(Price) against Age of House")

## `geom_smooth()` using formula = 'y ~ x'
```



```
train$waterfront<-factor(train$waterfront)
ggplot2::ggplot(train, ggplot2::aes(x=waterfront, y=ystar))+  
  geom_boxplot(fill="Blue")+
  labs(x="Waterfront", y="Log(Price)", title="Dist of Log(Price) by Waterfront")
```



```
ggplot2::ggplot(train,ggplot2::aes(x=ystar))+  
  geom_density() +  
  labs(x="Log(Price)",y="Density",title="Density of Log(Price)")
```

Density of Log(Price)

