DS311 - R Lab Assignment

Quoc Hung Nguyen

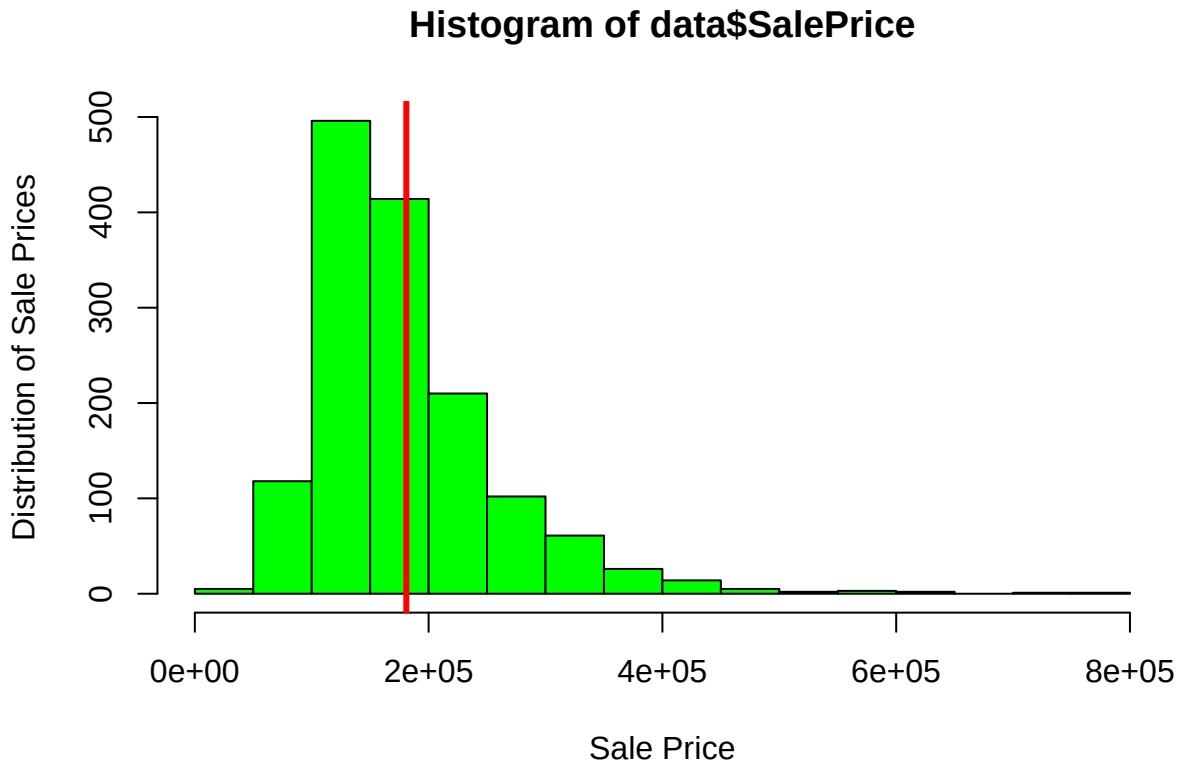
12/15/2022

Reading data

```
data = read.csv('C:/Users/nqhun/DS311-Technologies-in-Data-Analytic/Week_9_Exploratory_Data_Analysis/EDA
View(data)
```

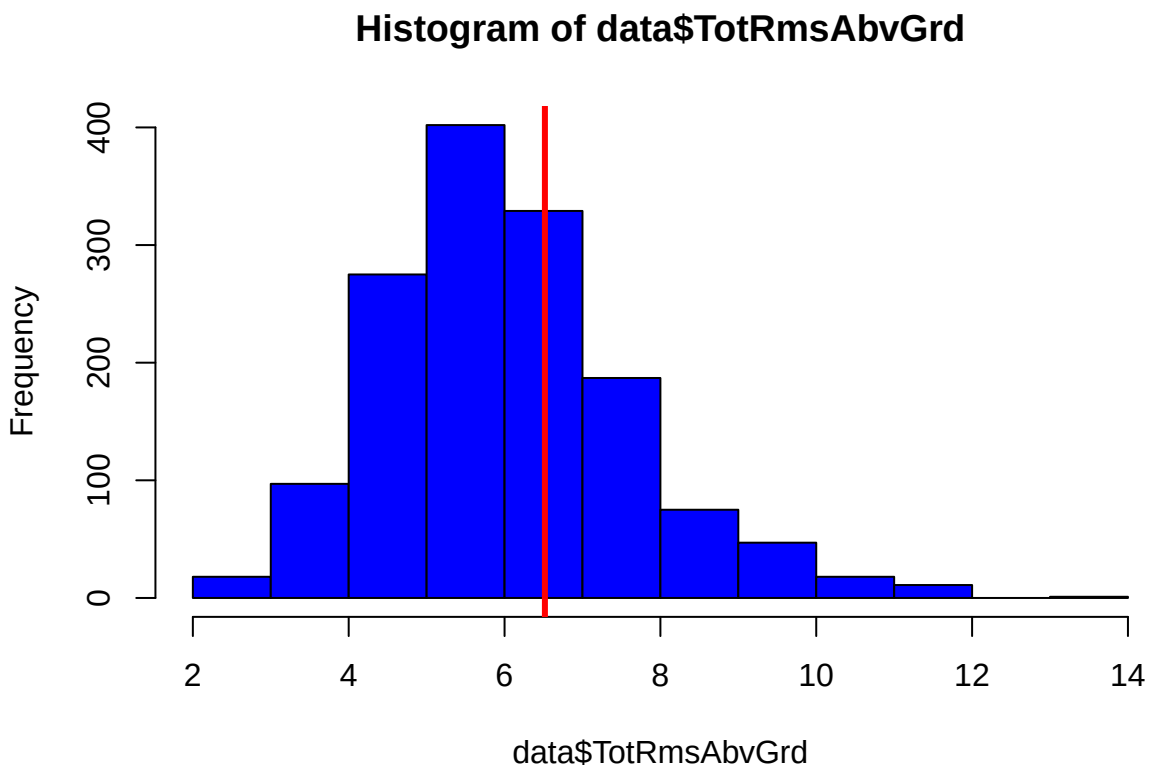
EDA plot 1

```
hist(data$SalePrice,col = "green",xlab = 'Sale Price',ylab = 'Distribution of Sale Prices')
abline(v = mean(data$SalePrice),col="red",lwd=3)
```



EDA plot 2

```
hist(data$TotRmsAbvGrd,col = 'blue')
abline(v = mean(data$TotRmsAbvGrd),col="red",lwd=3)
```



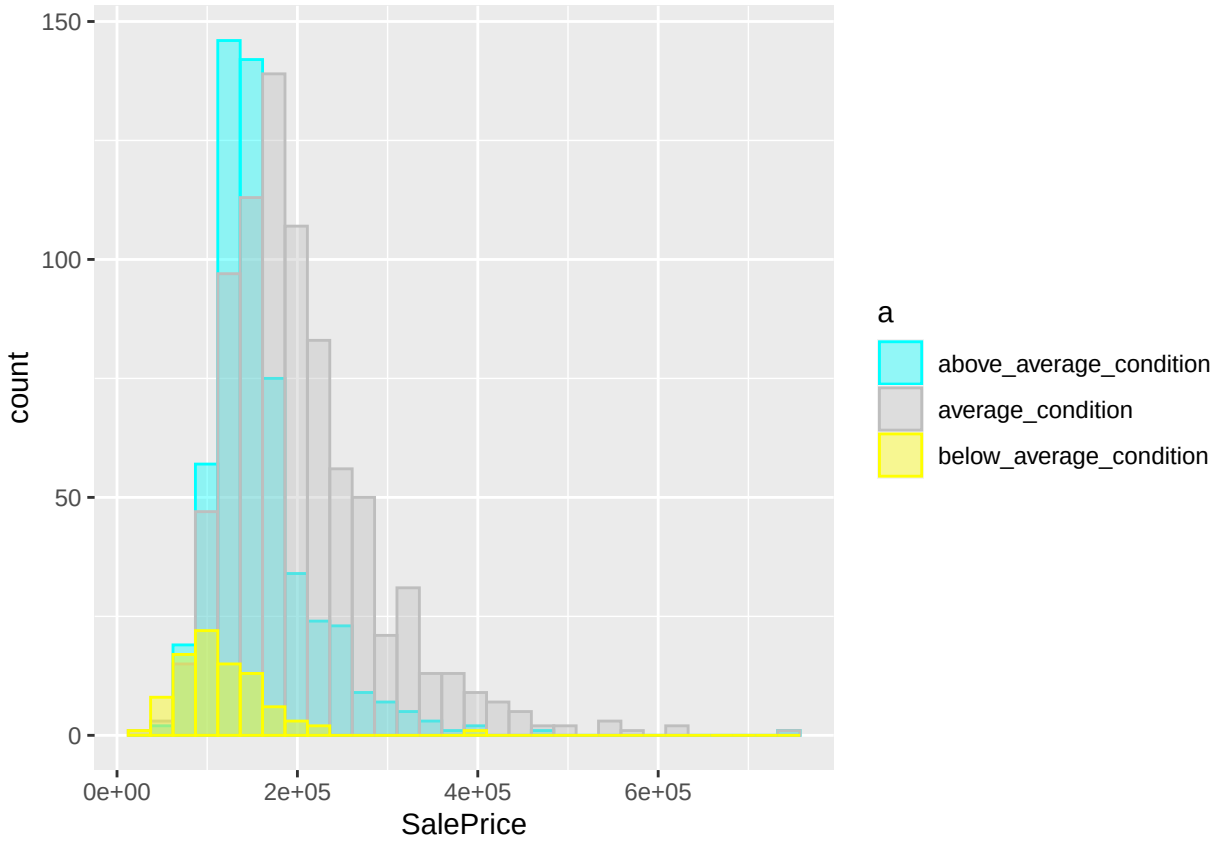
EDA plot 3

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

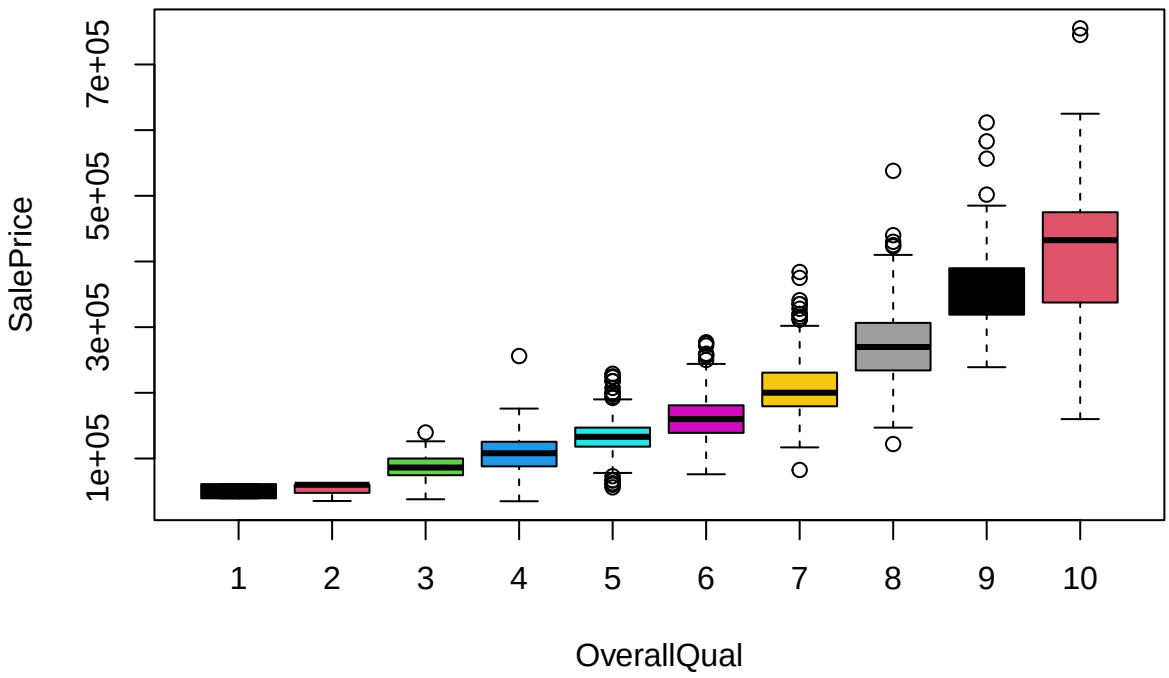
```
a = rep(NA,length(data$OverallCond))
for (i in 1:length(data$OverallCond))
{
  if (data$OverallCond[i]< 5){
    a[i] = "below_average_condition"
  }
  if (data$OverallCond[i]==5){
    a[i] = "average_condition"
  }
  if (data$OverallCond[i]> 5){
    a[i] = "above_average_condition"
```

```
  }
}
ggplot(data = data, mapping = aes(x = SalePrice))+
  geom_histogram(aes(color=a,alpha = 0.2,fill=a),position = "identity", bins = 30, alpha = 0.4)+
  scale_color_manual(values = c("cyan", "gray","yellow")) +
  scale_fill_manual(values = c("cyan", "gray","yellow"))
```



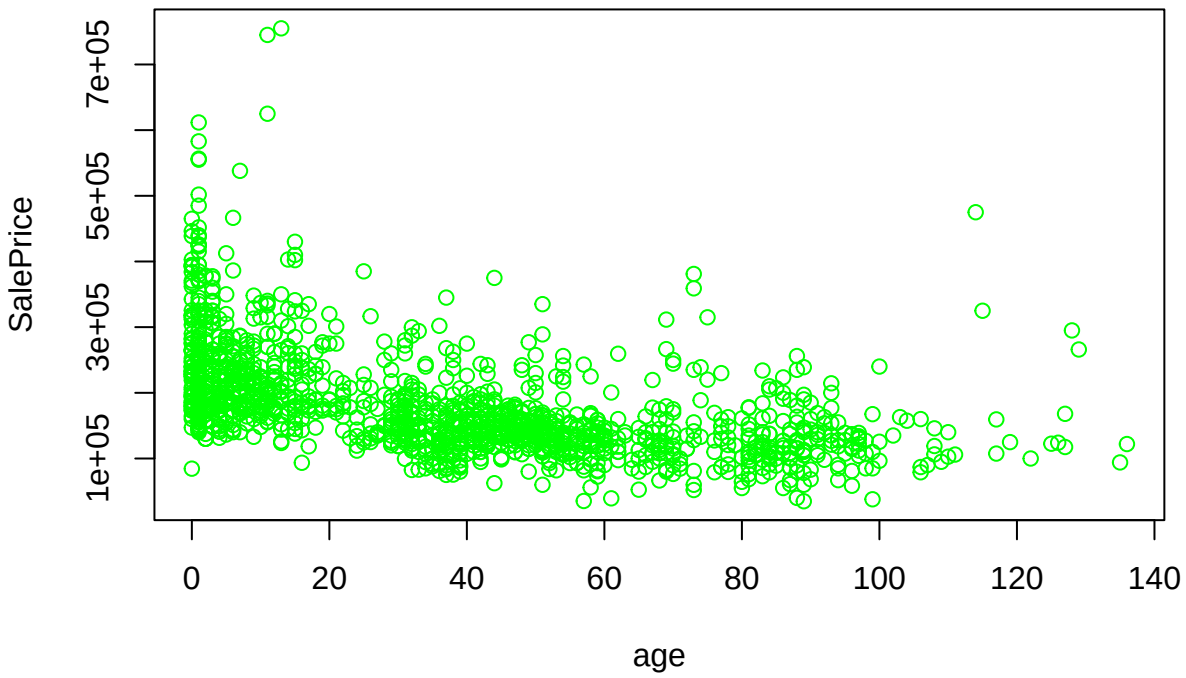
EDA plot 4

```
boxplot(SalePrice~OverallQual,data = data,col=c(1:10))
```



EDA plot 5

```
data$age = data$YrSold-data$YearBuilt
plot(data$age,data$SalePrice,col = 'green',ylab = 'SalePrice',xlab = 'age')
```


SalePrice
1e+05
3e+05
5e+05
7e+05
0
20
40
60
80
100
120
140
age