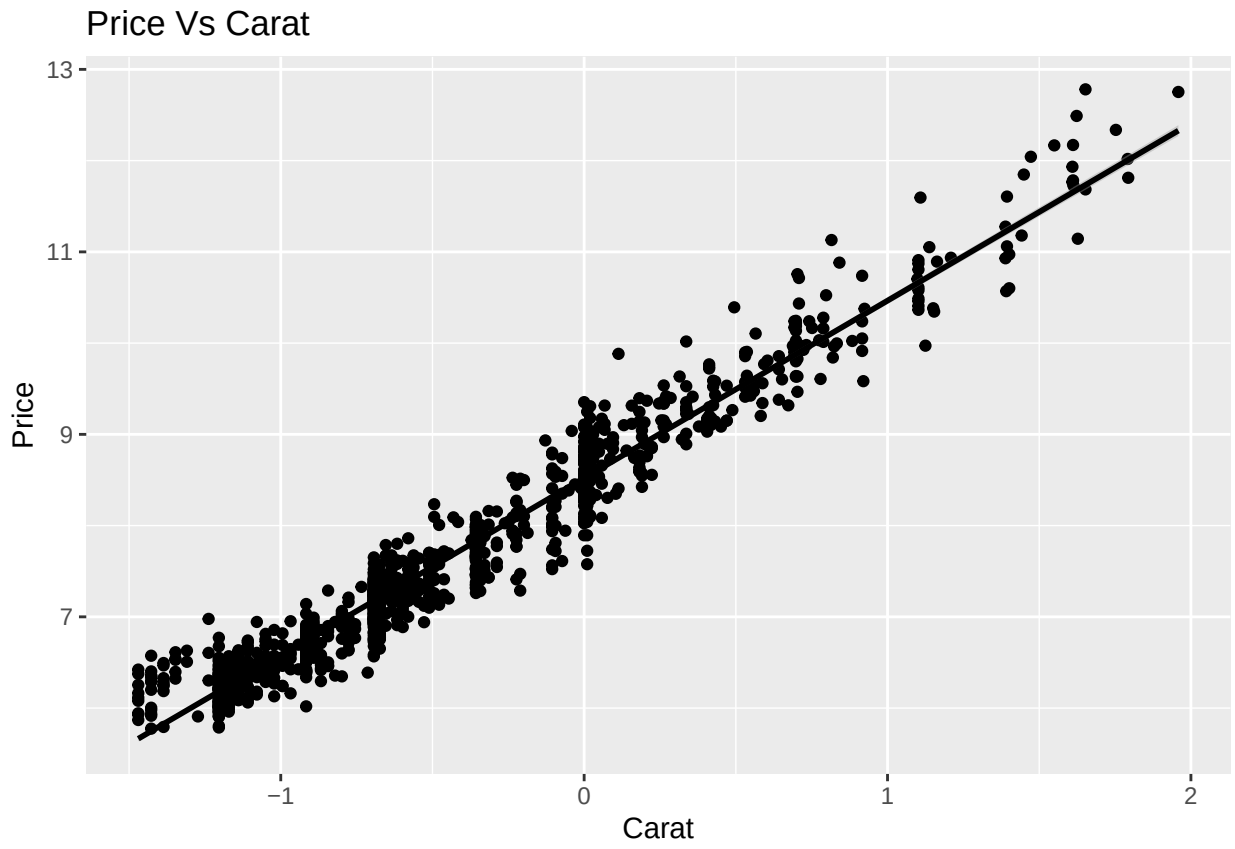
Project 1

Sirish

2022-10-03

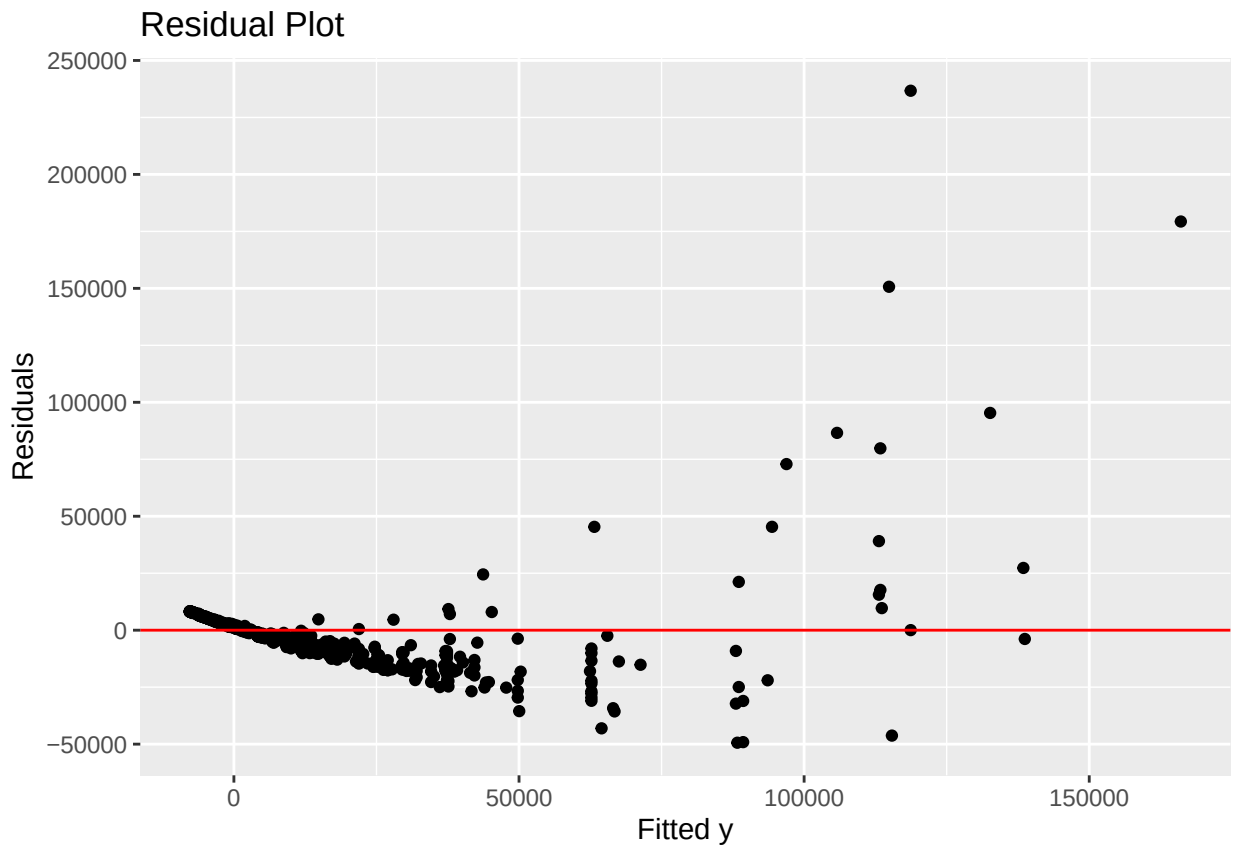
```
ggplot(Data, aes(x=log(carat), y=log(price)))+
  geom_point()+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat")+
  geom_smooth(method="lm", col="black")
```

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

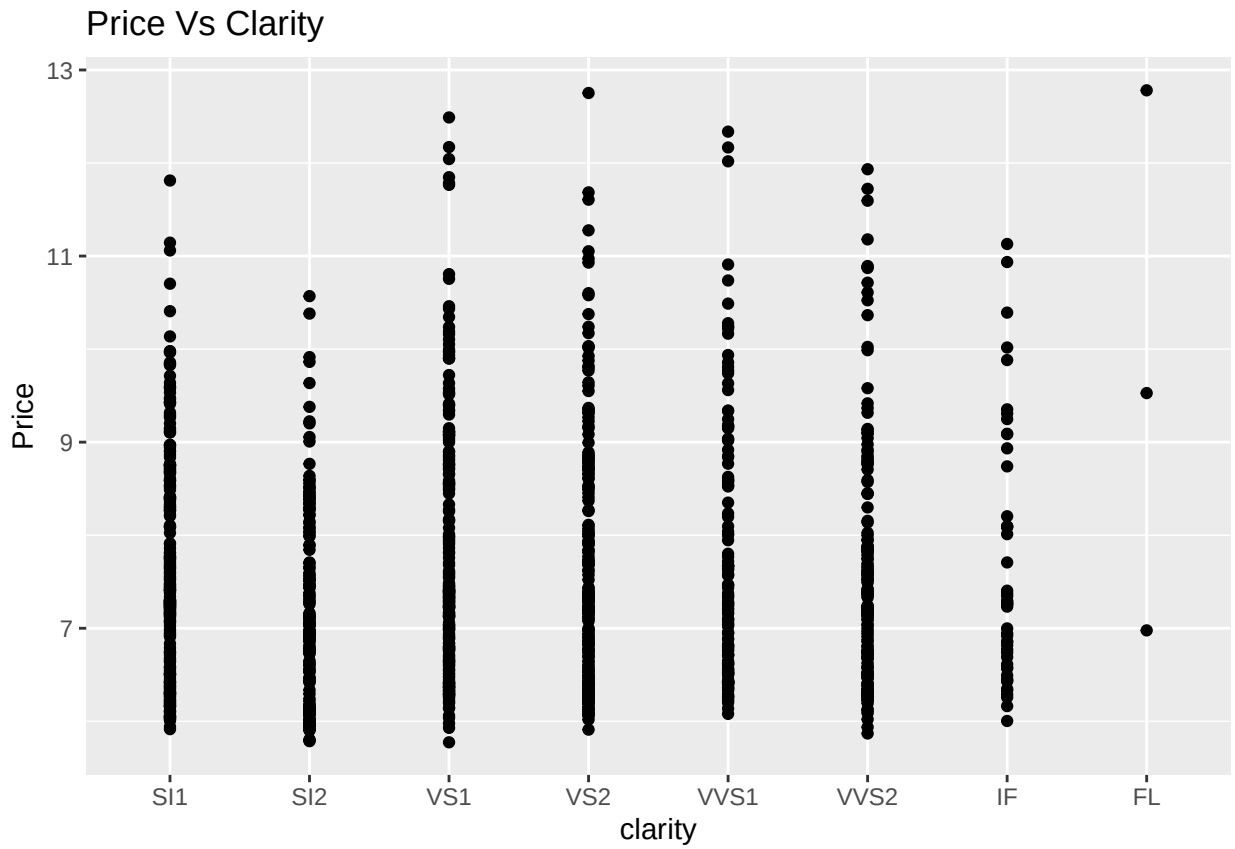



```
result <- lm(price~carat, Data)
Data$yhat<-result$fitted.values
Data$res<-result$residuals
```

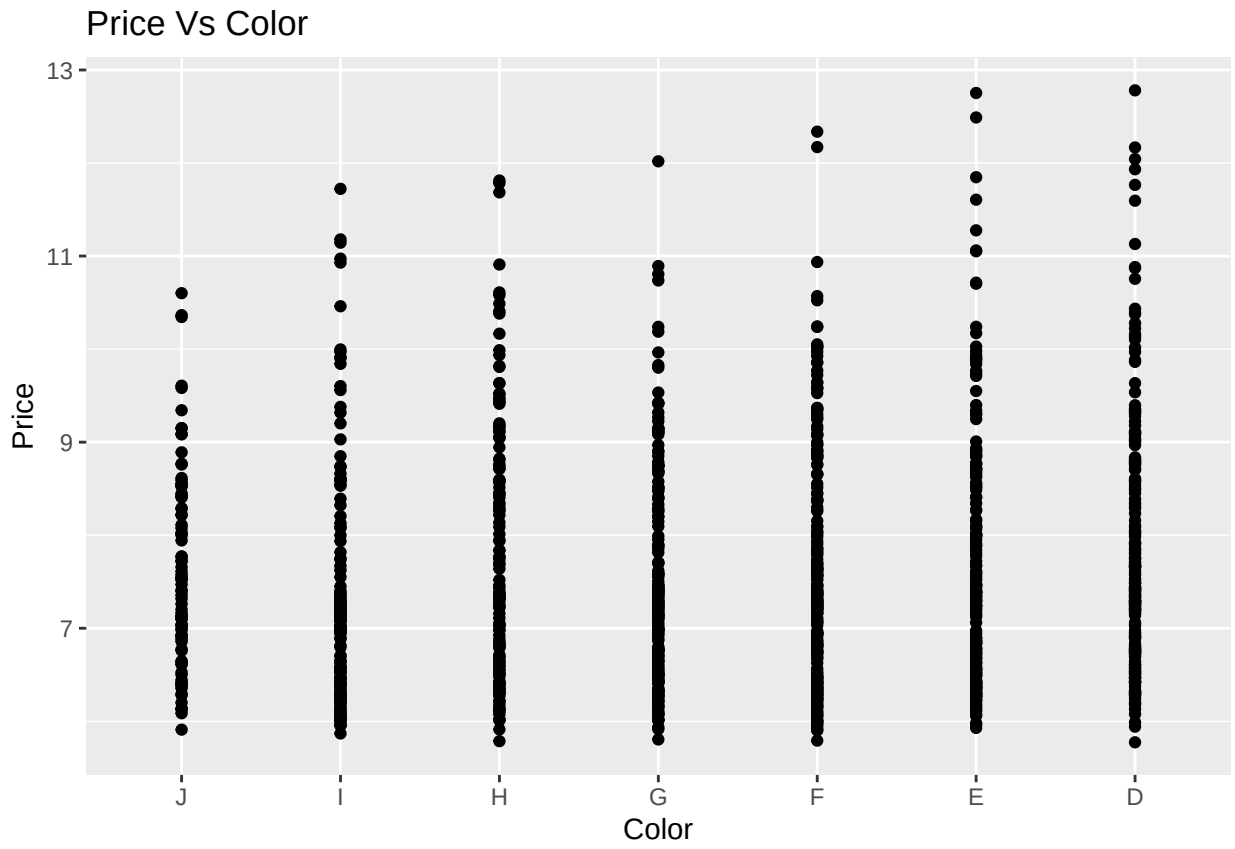
```
ggplot(Data, aes(x=yhat,y=res))+
  geom_point()+
  geom_hline(yintercept=0, color="red")+
  labs(x="Fitted y", y="Residuals", title="Residual Plot")
```



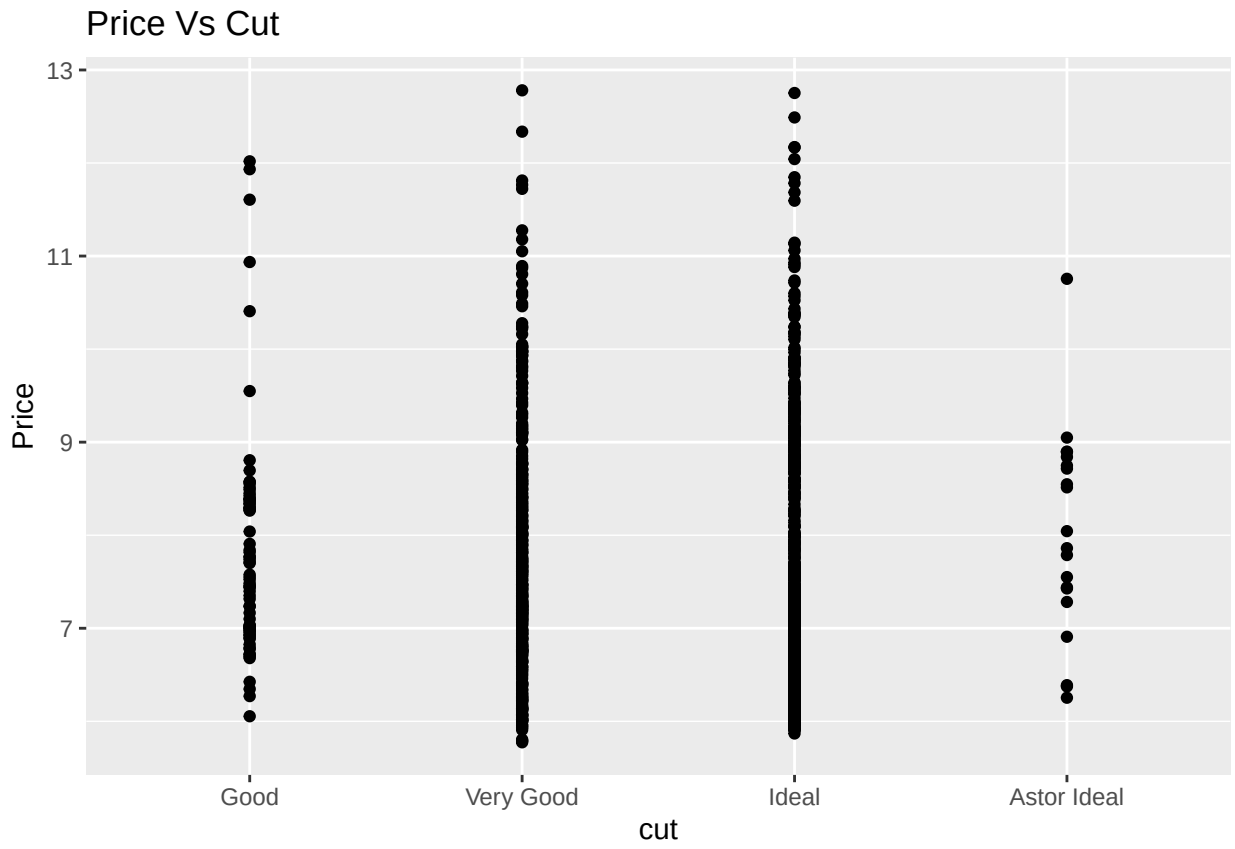

```
ggplot(Data, aes(x= clarity, y=log(price)))+
  geom_point()+
  labs(x= "clarity",
       y= "Price",
       title = "Price Vs Clarity")
```



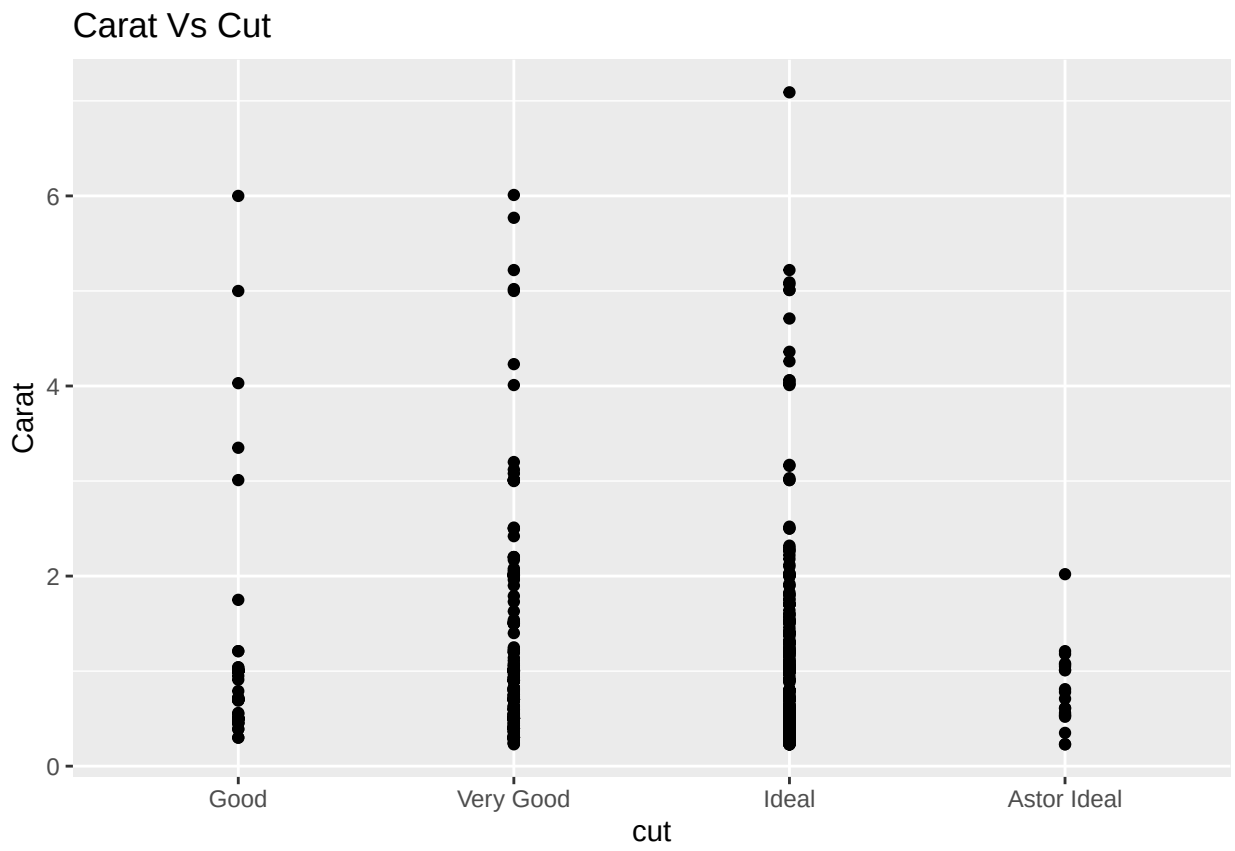

```
ggplot(Data, aes(x= color, y=log(price)))+
  geom_point()+
  labs(x= "Color",
       y= "Price",
       title = "Price Vs Color")
```




```
ggplot(Data, aes(x= cut, y=log(price)))+
  geom_point()+
  labs(x= "cut",
       y= "Price",
       title = "Price Vs Cut")
```

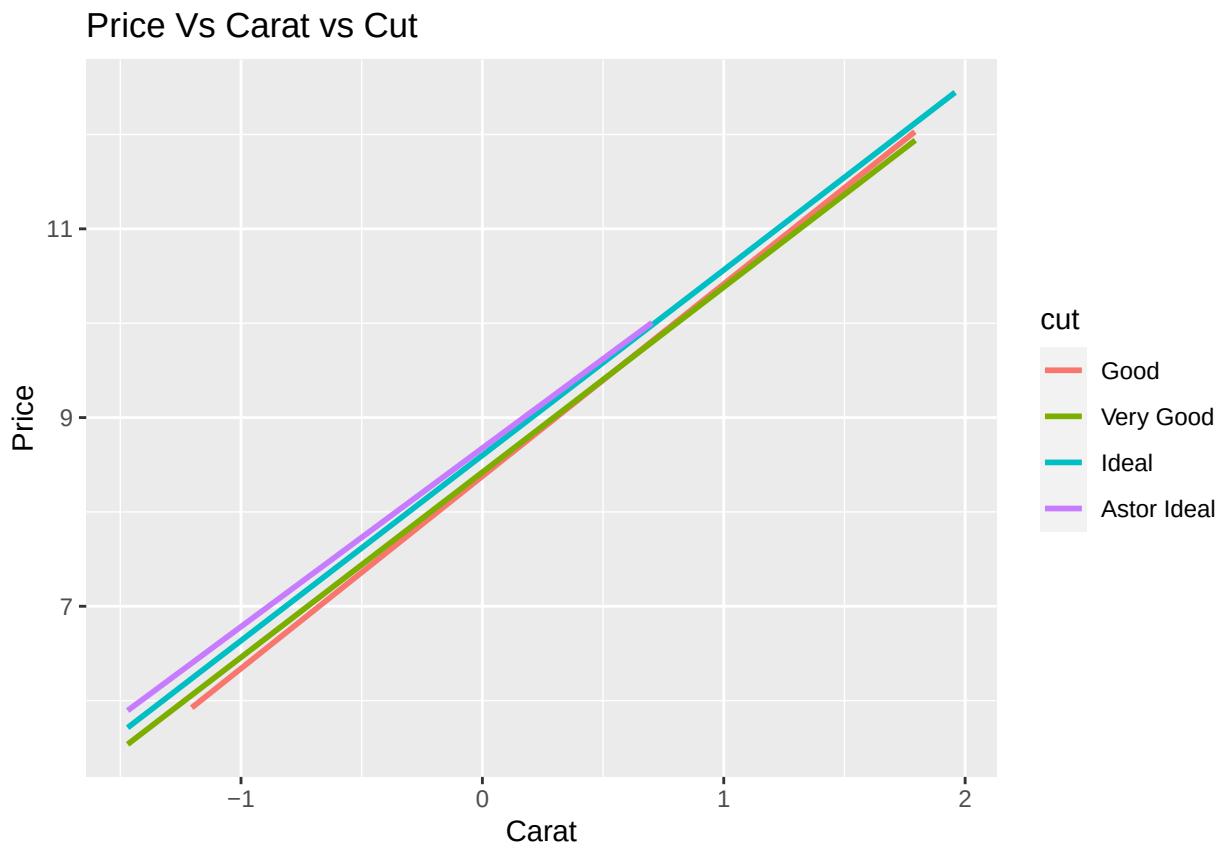



```
ggplot(Data, aes(x= cut, y=carat))+
  geom_point()+
  labs(x= "cut",
       y= "Carat",
       title = "Carat Vs Cut")
```



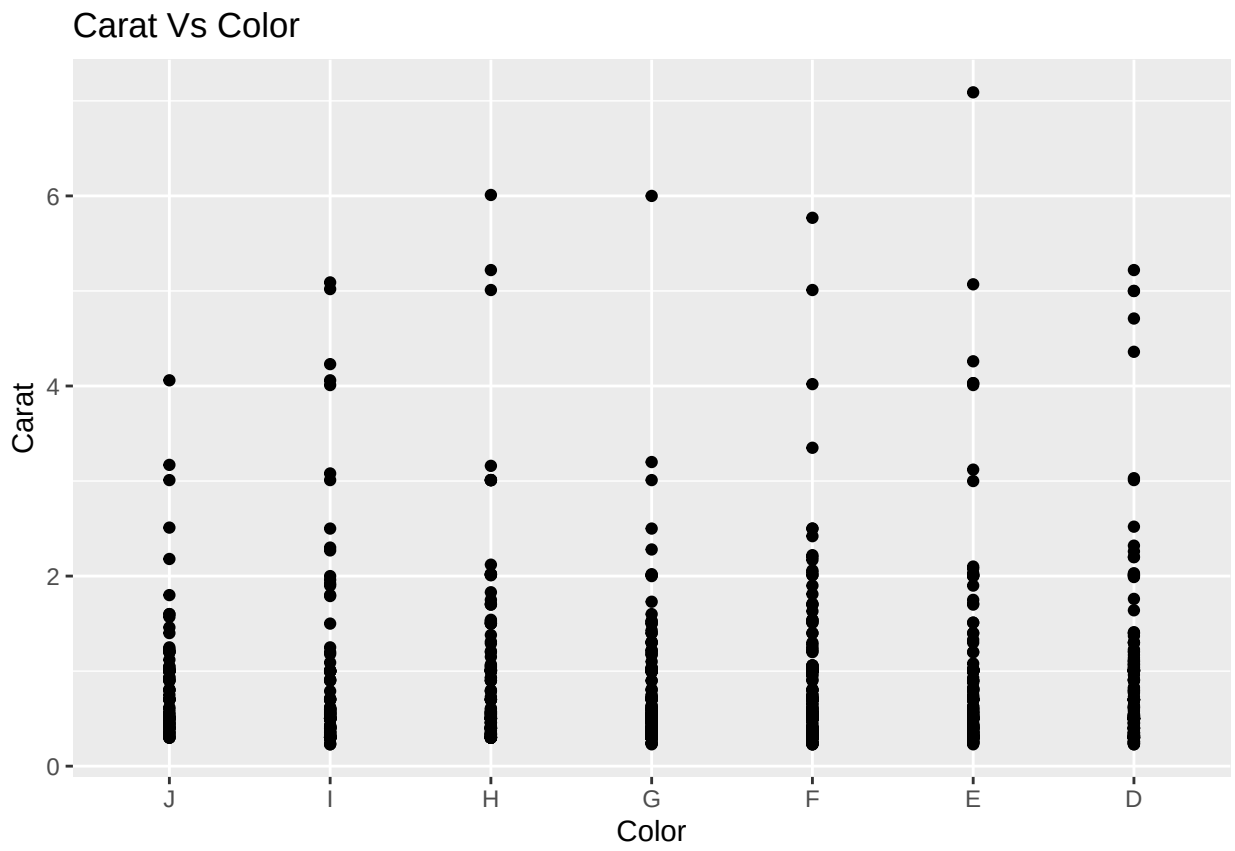

```
ggplot(Data, aes(x=log(carat), y=log(price), color = cut))+
  geom_smooth(method = lm, se=F)+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Cut")
```

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

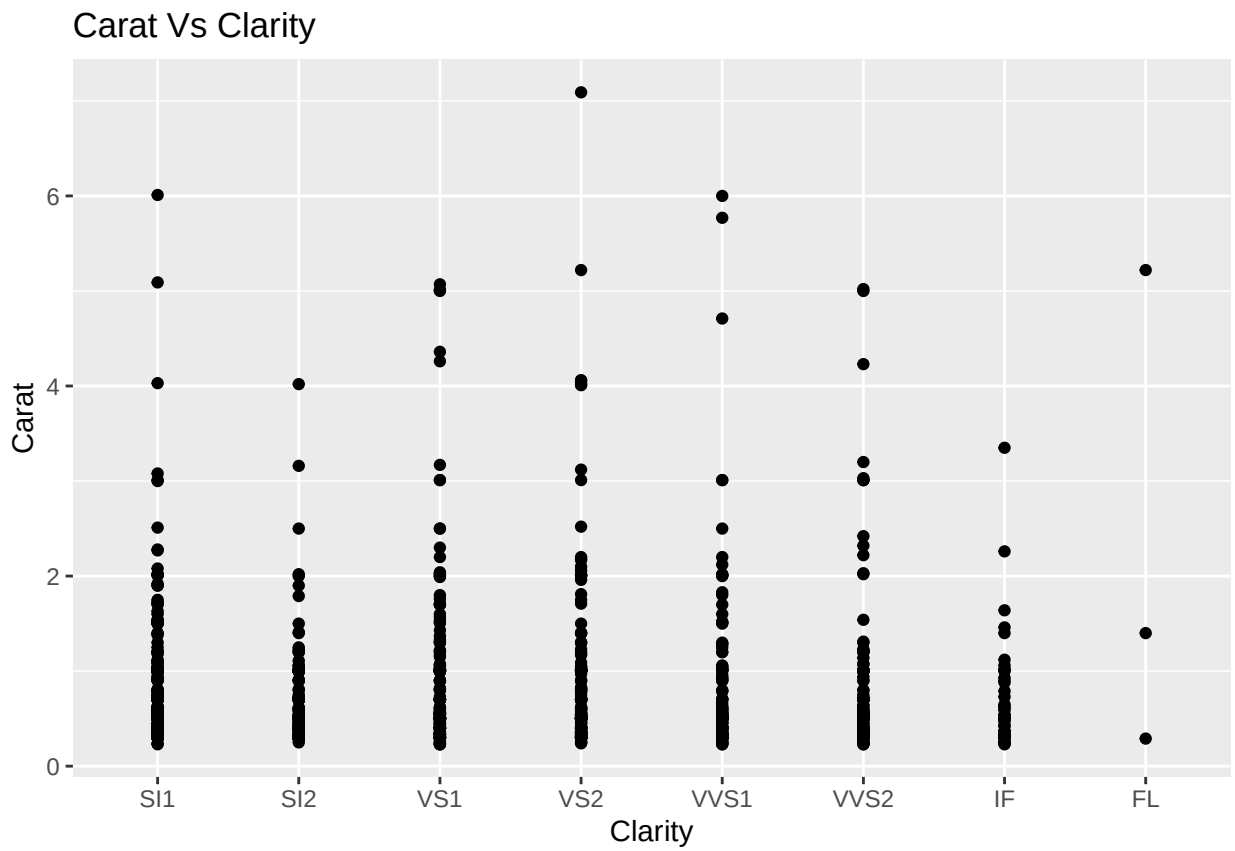



```
#Carat vs Cut is really low when you get to Astor Ideal. Even though that Astor Ideal is what Astor spe
```

```
ggplot(Data, aes(x= color, y=carat))+
  geom_point()+
  labs(x= "Color",
       y= "Carat",
       title = "Carat Vs Color")
```

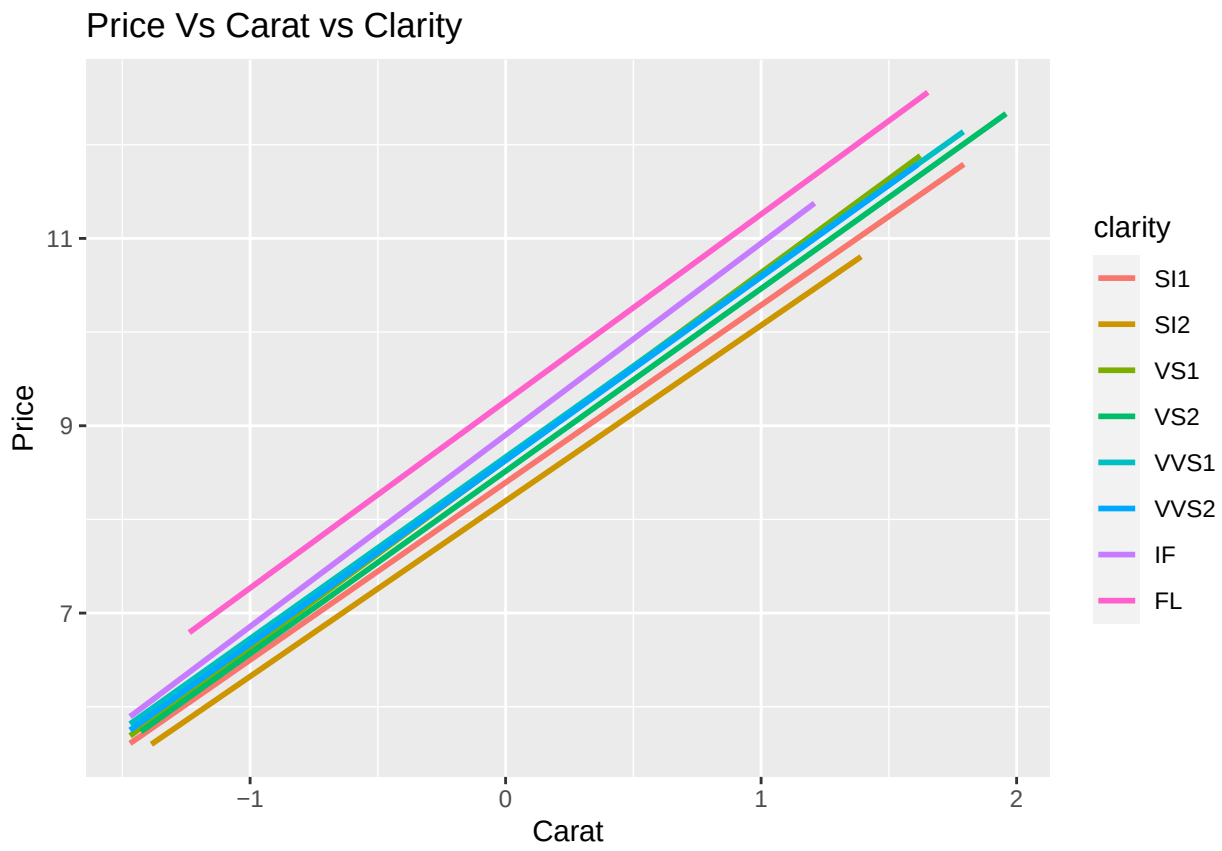



```
ggplot(Data, aes(x= clarity, y=carat))+
  geom_point()+
  labs(x= "Clarity",
       y= "Carat",
       title = "Carat Vs Clarity")
```




```
ggplot(Data, aes(x=log(carat), y=log(price), color = clarity))+
  geom_smooth(method = lm, se=F)+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Clarity")
```

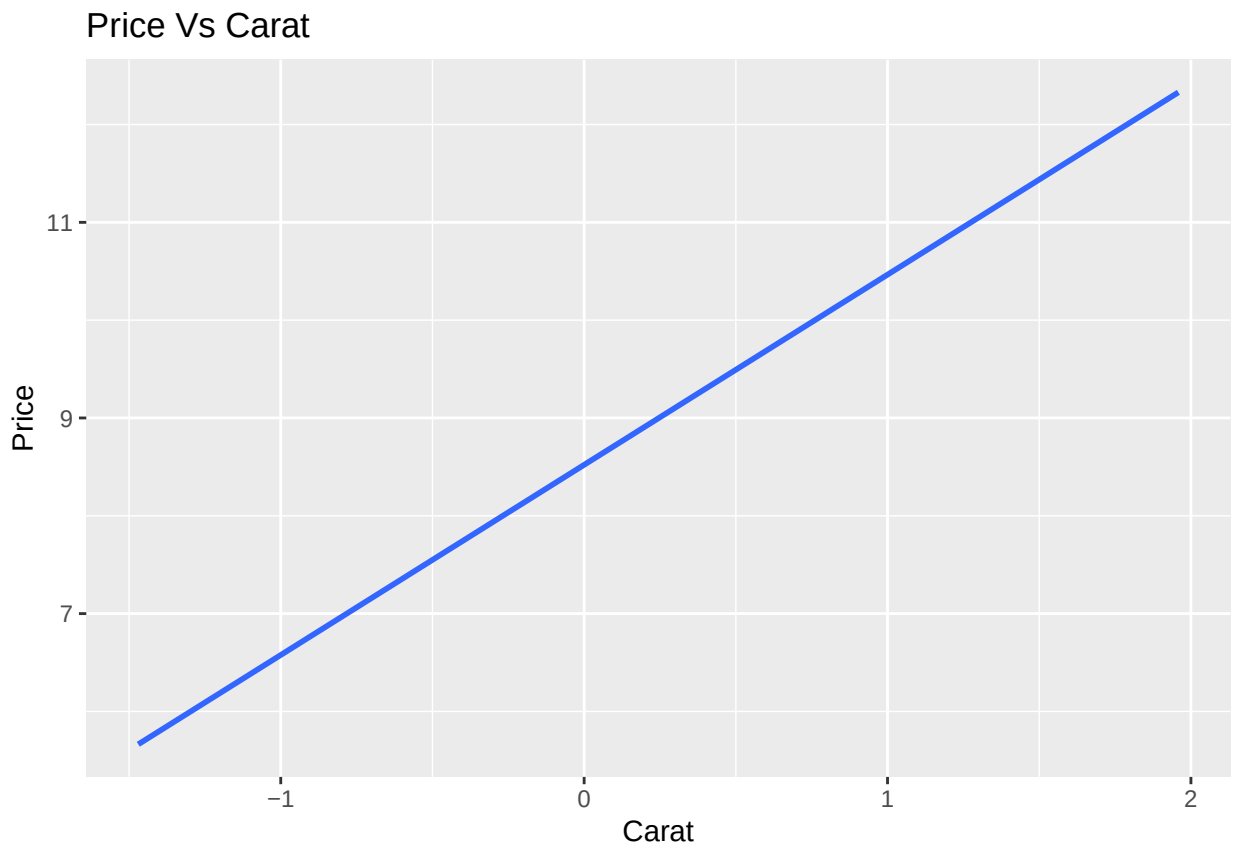
```
## `geom_smooth()` using formula 'y ~ x'
```




```
#What is surprising about this is that The general price trend tends to be higher in FL, however the Ca
```

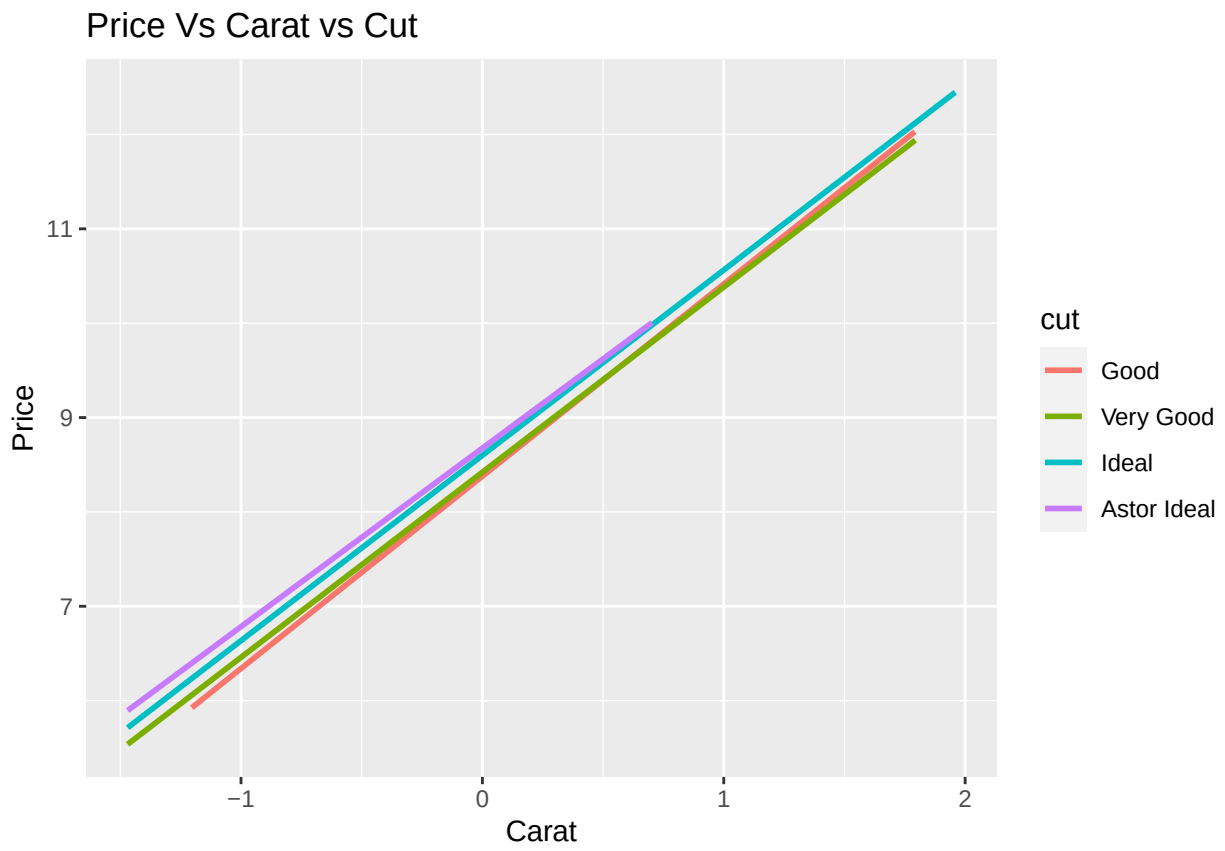
```
Data$cut <- factor(Data$cut, levels = c("Good", "Very Good", "Ideal", "Astor Ideal"))
ggplot(Data, aes(x=log(carat), y=log(price)))+
  geom_smooth(method = lm, se=F)+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat")
```

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



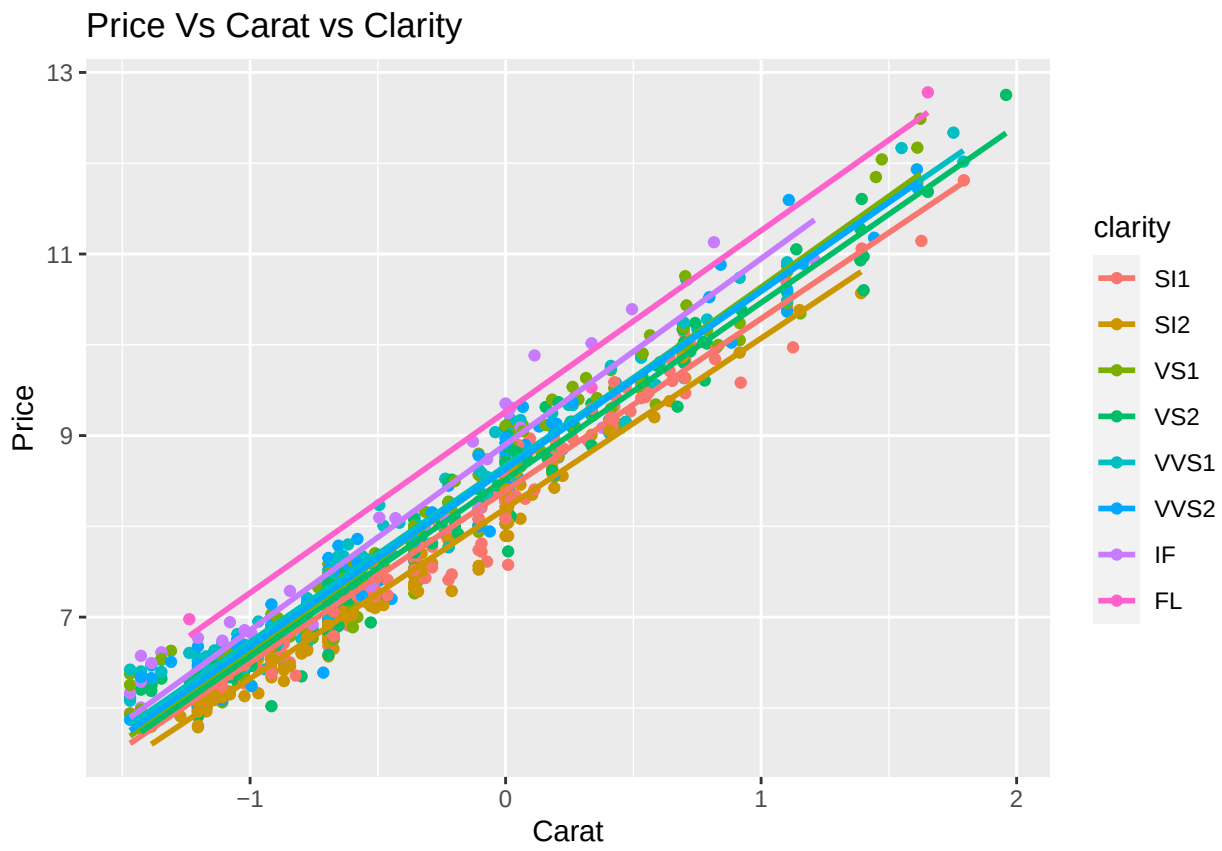

```
ggplot(Data, aes(x=log(carat), y=log(price), color = cut))+
  geom_smooth(method = lm, se=F)+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Cut")
```

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



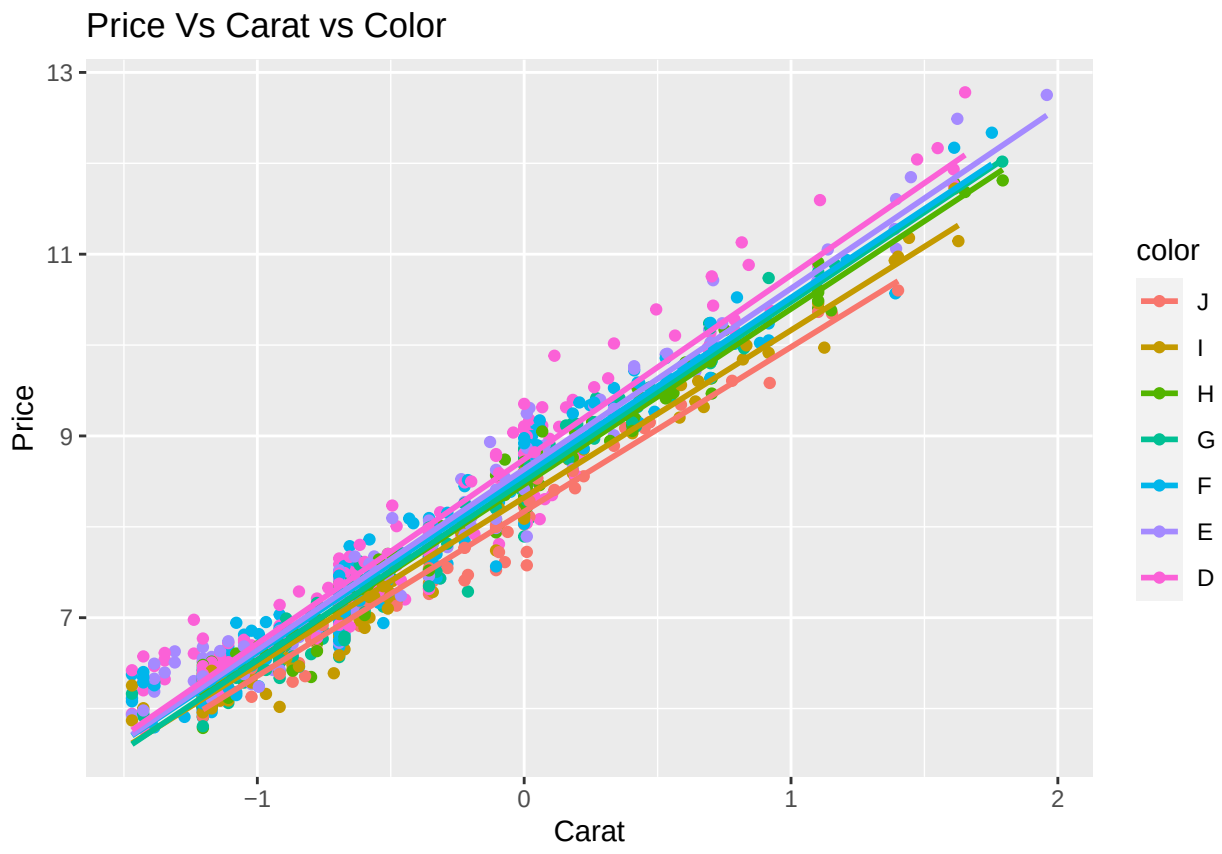

```
ggplot(Data, aes(x=log(carat), y=log(price), color = clarity))+
  geom_point()+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Clarity")+
  geom_smooth(method="lm", se=F)
```

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

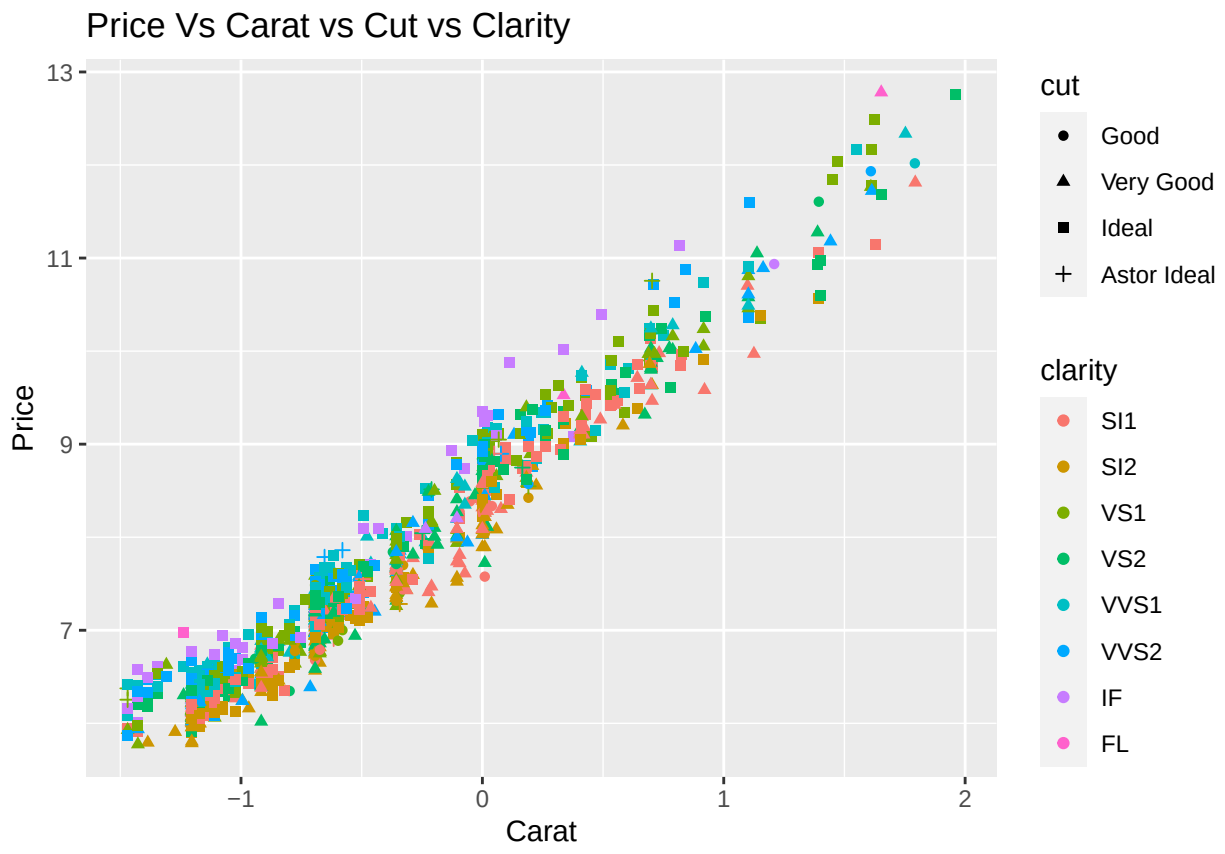



```
ggplot(Data, aes(x=log(carat), y=log(price), color = color))+
  geom_point()+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Color")+
  geom_smooth(method="lm", se=F)
```

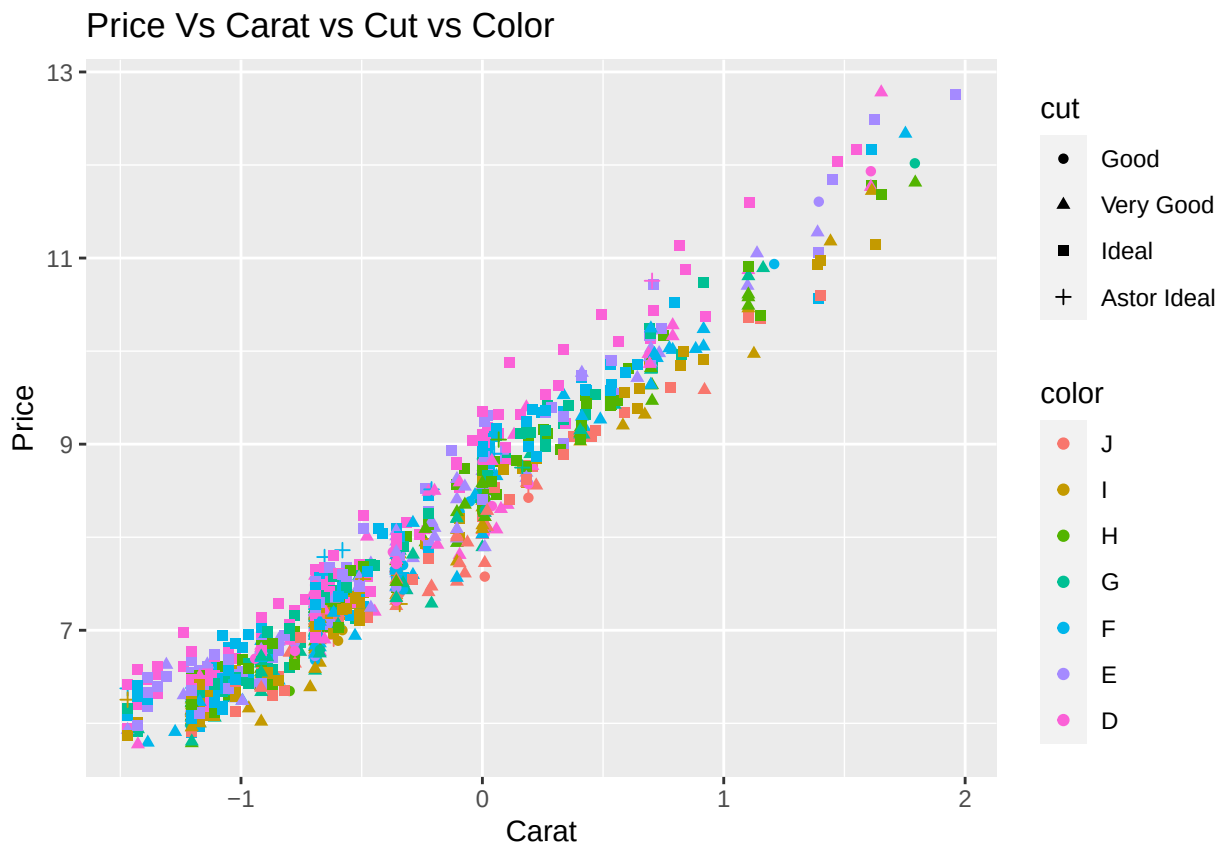
```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(Data, aes(x=log(carat), y=log(price), color = clarity, shape = cut))+
  geom_point()+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Cut vs Clarity")
```

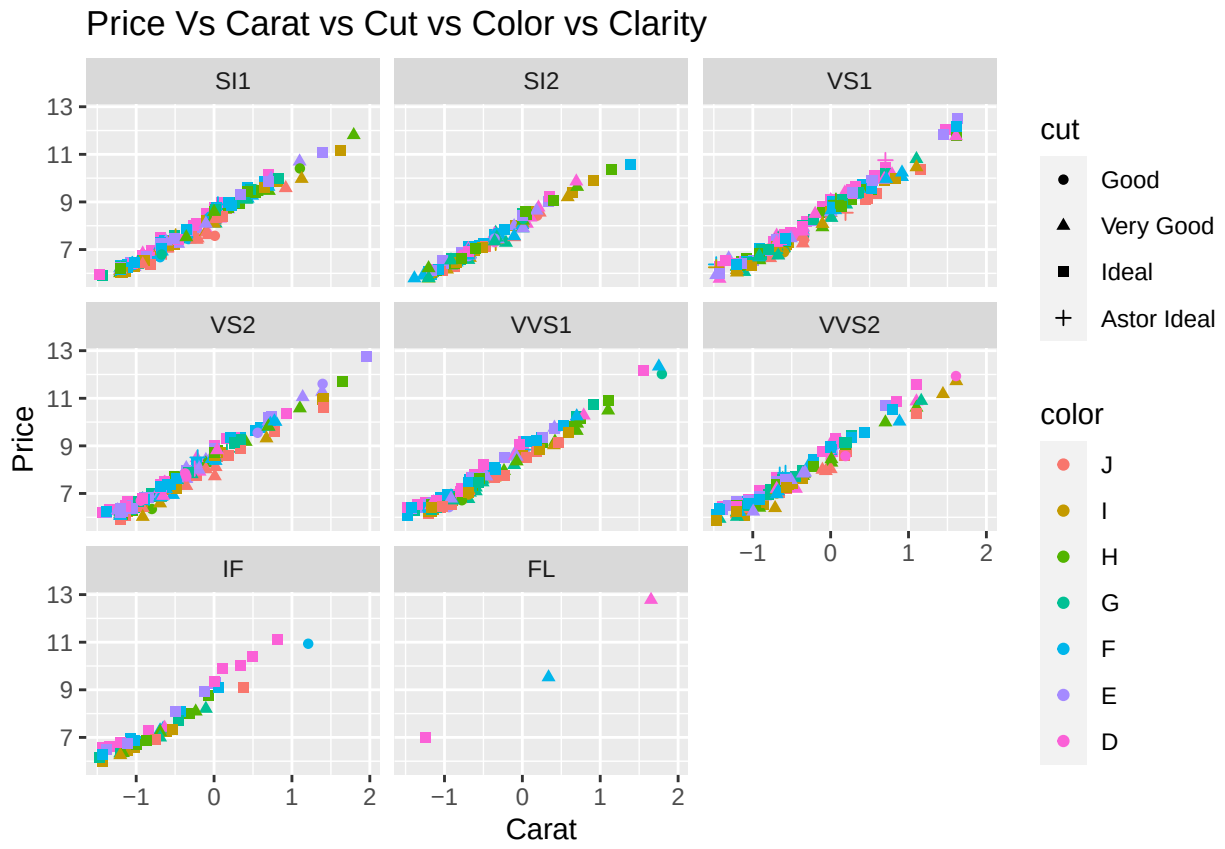



```
ggplot(Data, aes(x=log(carat), y=log(price), color = color, shape = cut))+
  geom_point()+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Cut vs Color")
```



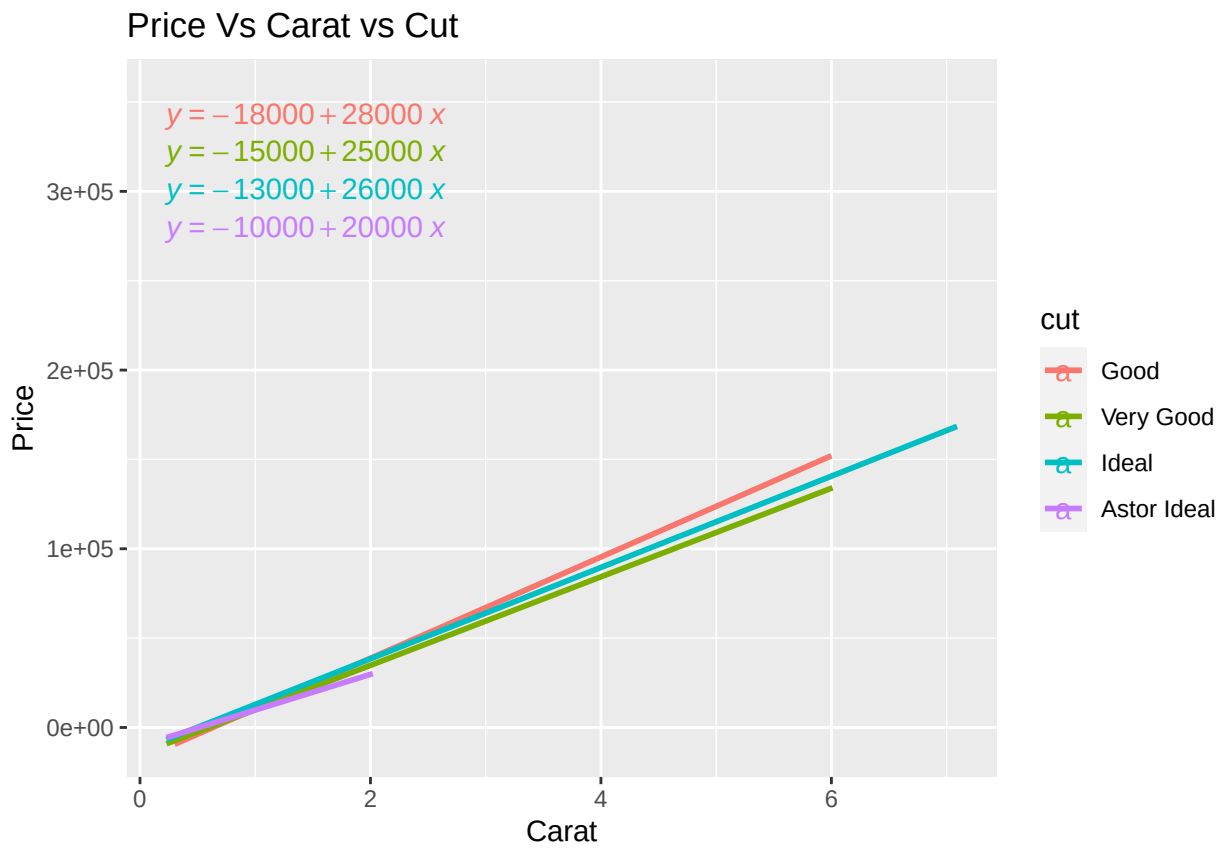

```
#ggplot(Data, aes(x=log(carat), y=log(price), shape = color, color = clarity))+
  #geom_point()+
  #labs(x= "Carat",
  #     y= "Price",
  #     title = "Price Vs Carat vs Cut vs Clarity")
# bad because I can't show all the data points

ggplot(Data, aes(x=log(carat), y=log(price), shape = cut, color = color))+
  geom_point()+
  facet_wrap(vars(clarity))+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Cut vs Color vs Clarity")
```



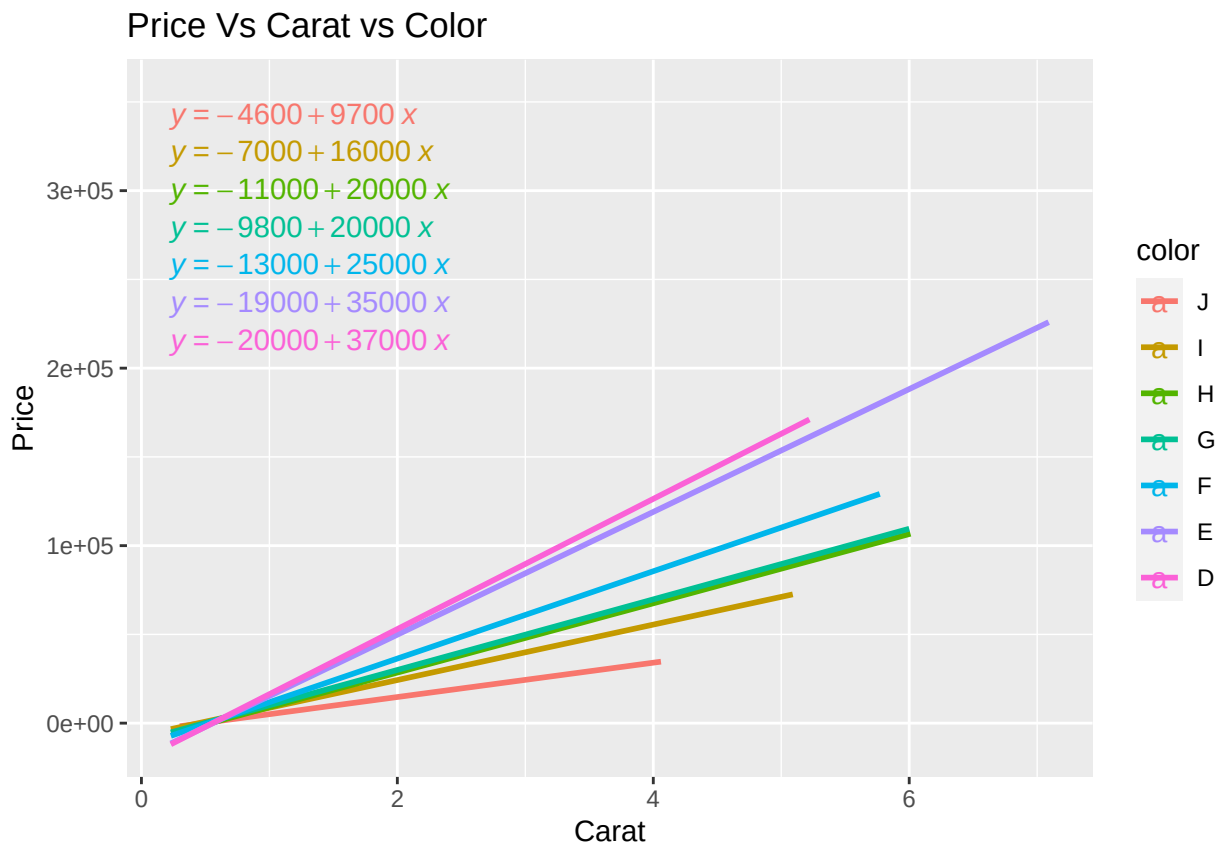

```
ggplot(Data, aes(x=(carat),  y=(price), color = cut))+
  geom_smooth(method = lm, se=F)+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Cut")+
  stat_regline_equation()
```

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



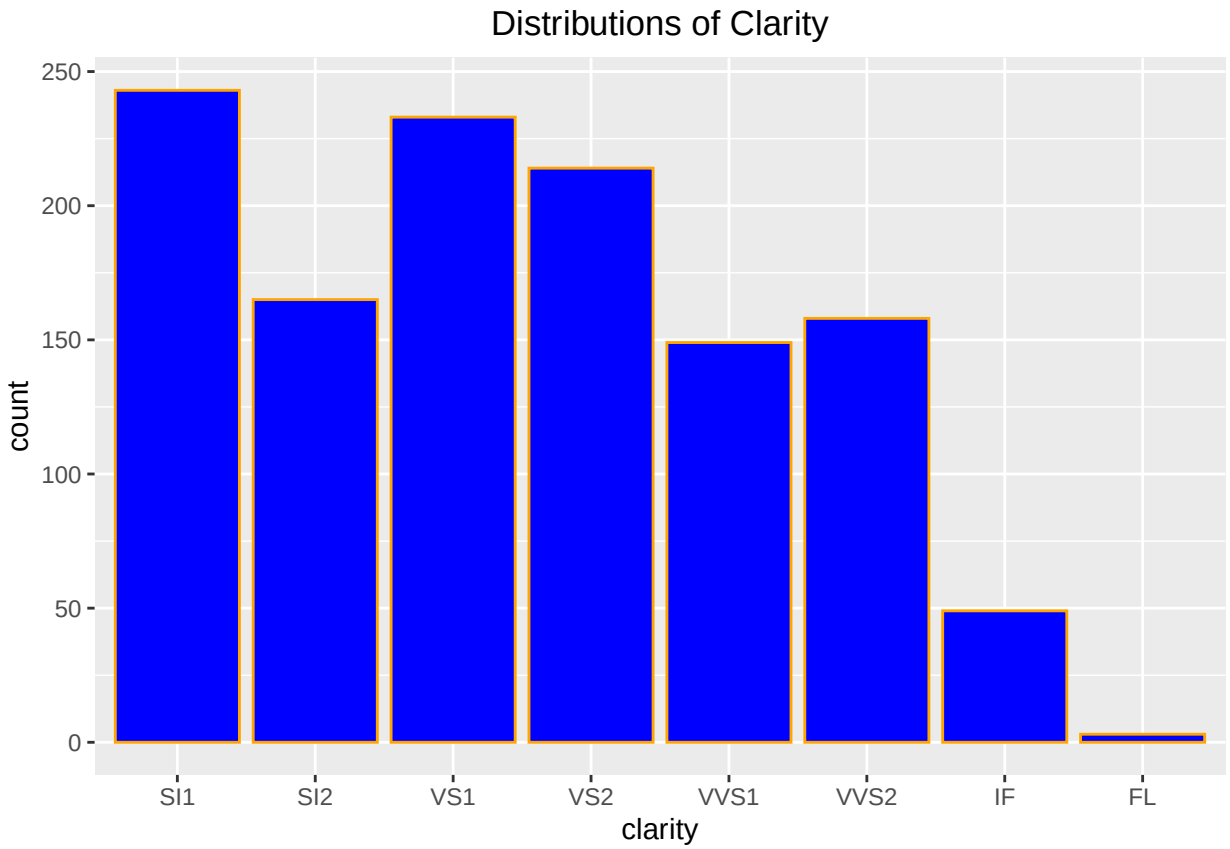

```
ggplot(Data, aes(x=(carat), y=(price), color = color))+
  geom_smooth(method = lm, se=F)+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Color")+
  stat_regline_equation()
```

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

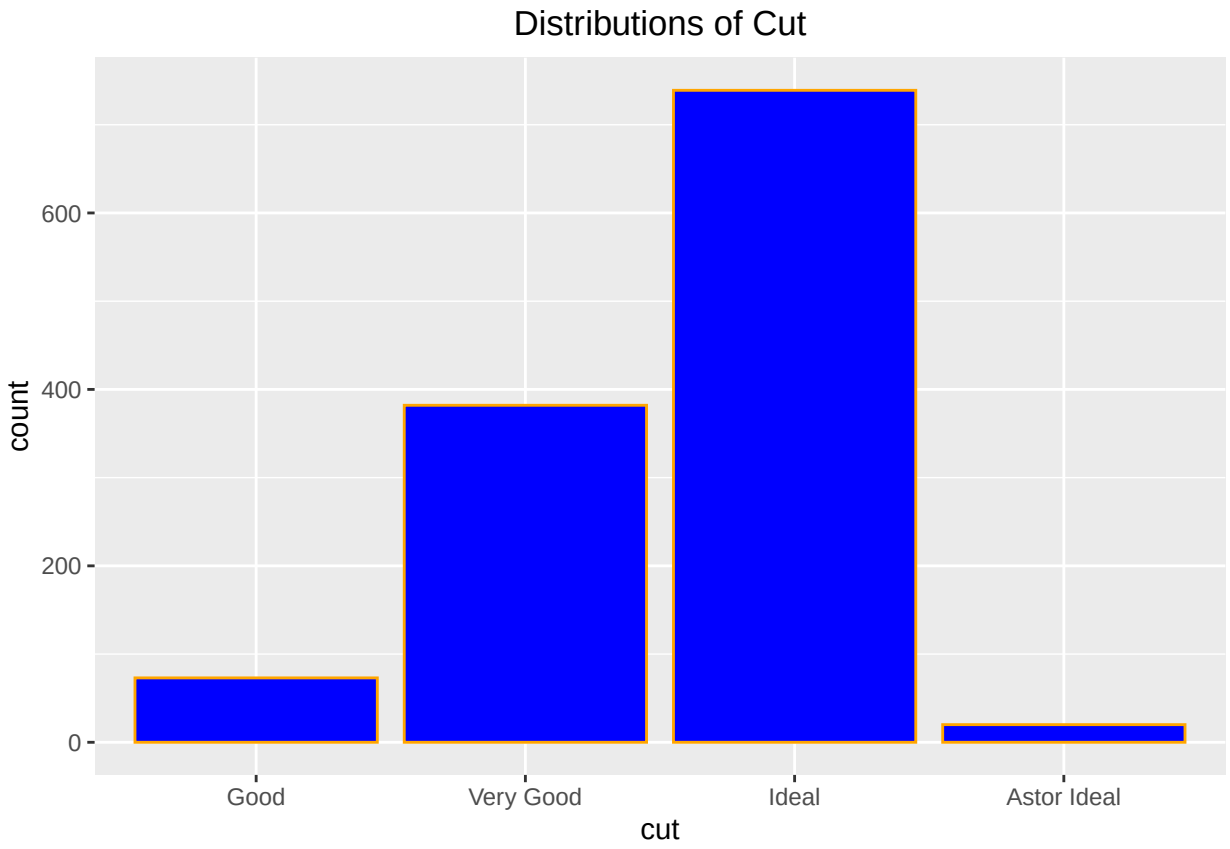



```
#From these 2 graphs we can see that for Astor Ideal and D color, that Astor Ideal starts at a higher b
```

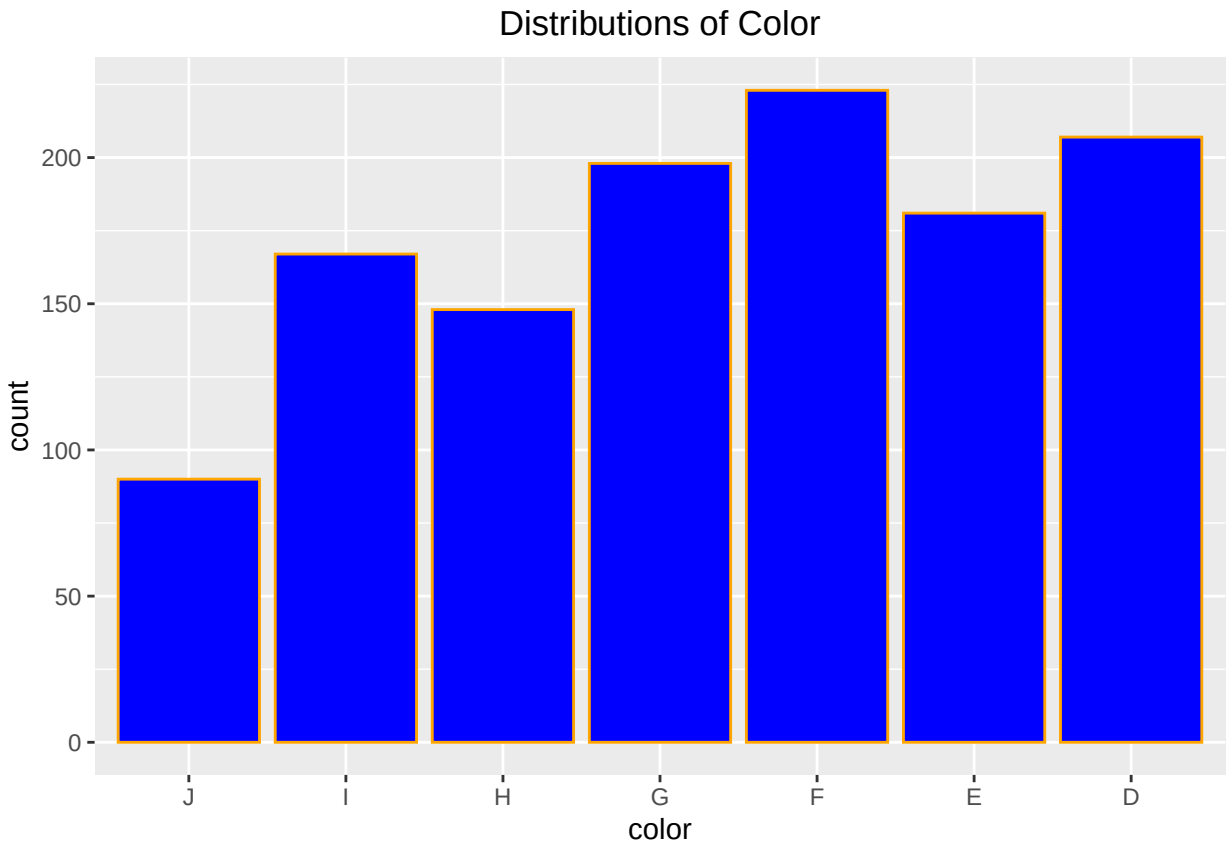
```
ggplot(Data)+
  aes(x=clarity)+
  geom_bar(fill="blue",color="orange")+
  labs(title = "Distributions of Clarity") +
    theme(
        plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 0)
    )
```



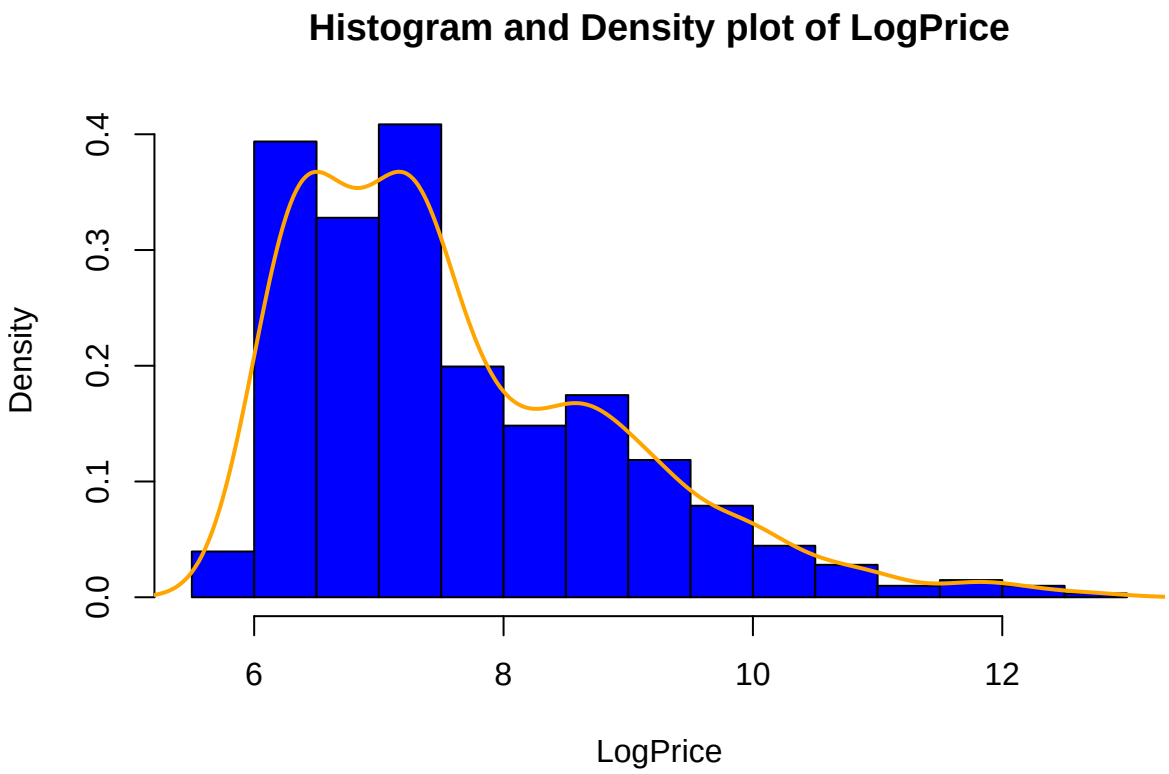

```
ggplot(Data)+
  aes(x=cut)+
  geom_bar(fill="blue",color="orange")+
  labs(title = "Distributions of Cut") +
  theme(
      plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0)
    )
```




```
ggplot(Data)+
  aes(x=color)+
  geom_bar(fill="blue",color="orange")+
  labs(title = "Distributions of Color") +
  theme(
      plot.title = element_text(hjust = 0.5),
      axis.text.x = element_text(angle = 0)
    )
```




```
Data <- Data %>%
  mutate(logprice = log(price))
hist(Data$logprice,
     col="blue",
     prob = TRUE,
     xlab = "LogPrice",
     main = "Histogram and Density plot of LogPrice")
lines(density(Data$logprice),
      lwd = 2,
      col = "orange")
```


Histogram and Density plot of LogPrice
Density
0.0
0.1
0.2
0.3
0.4
6
8
10
12
LogPrice