

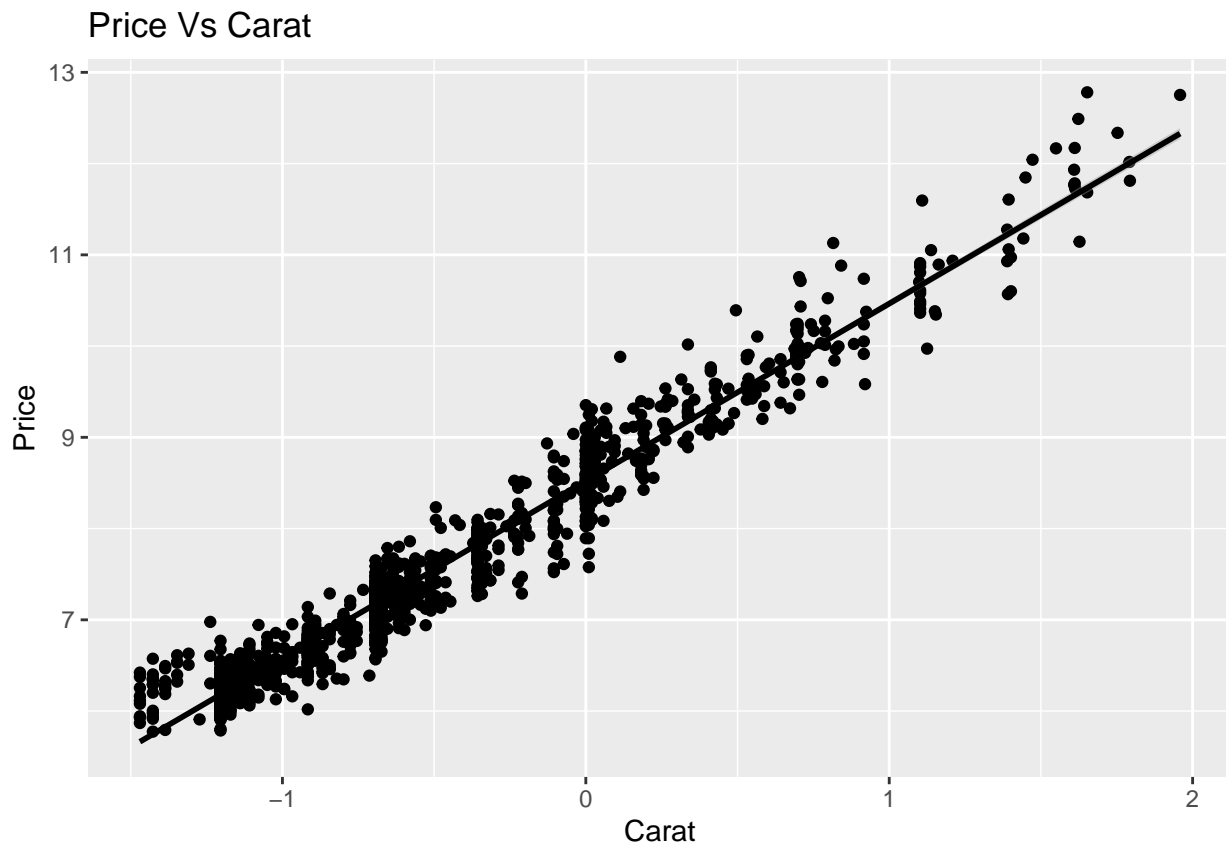
Project 1

Sirish

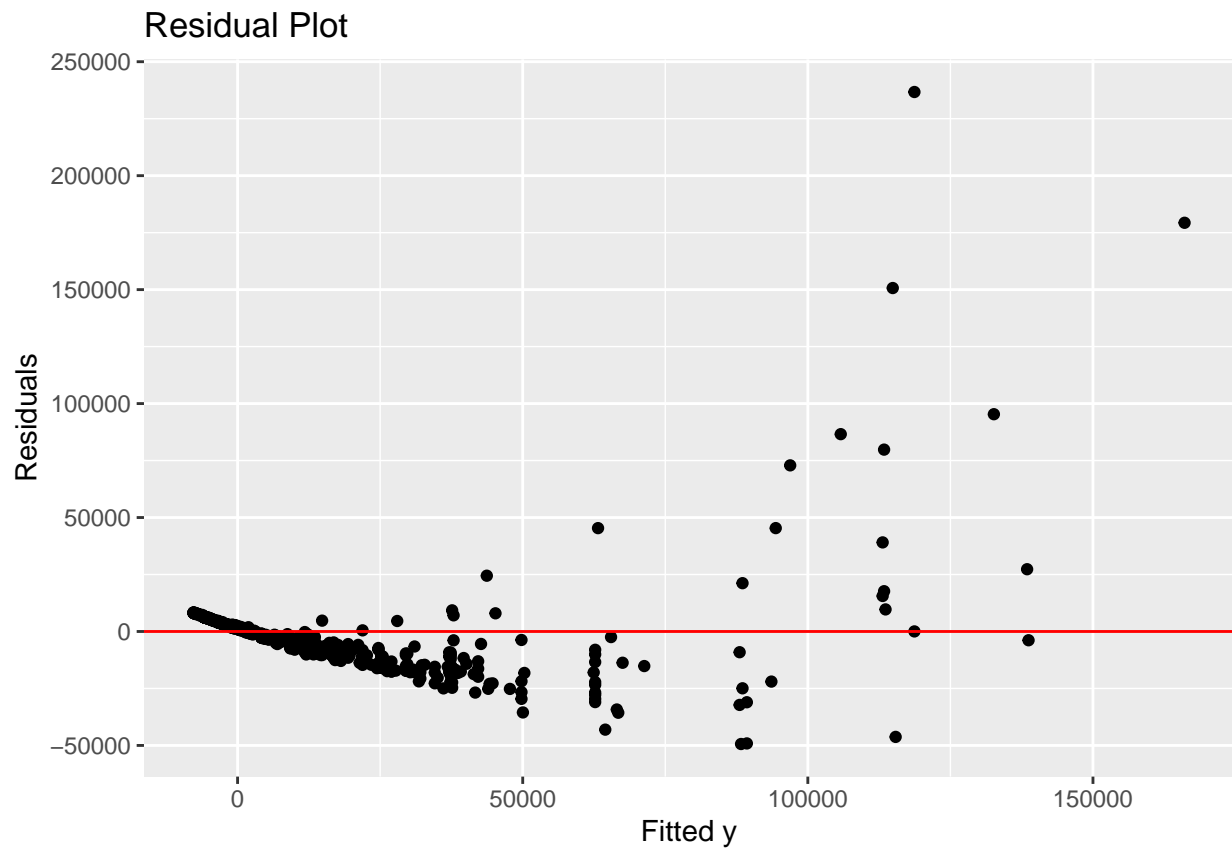
2022-10-03

```
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")
```

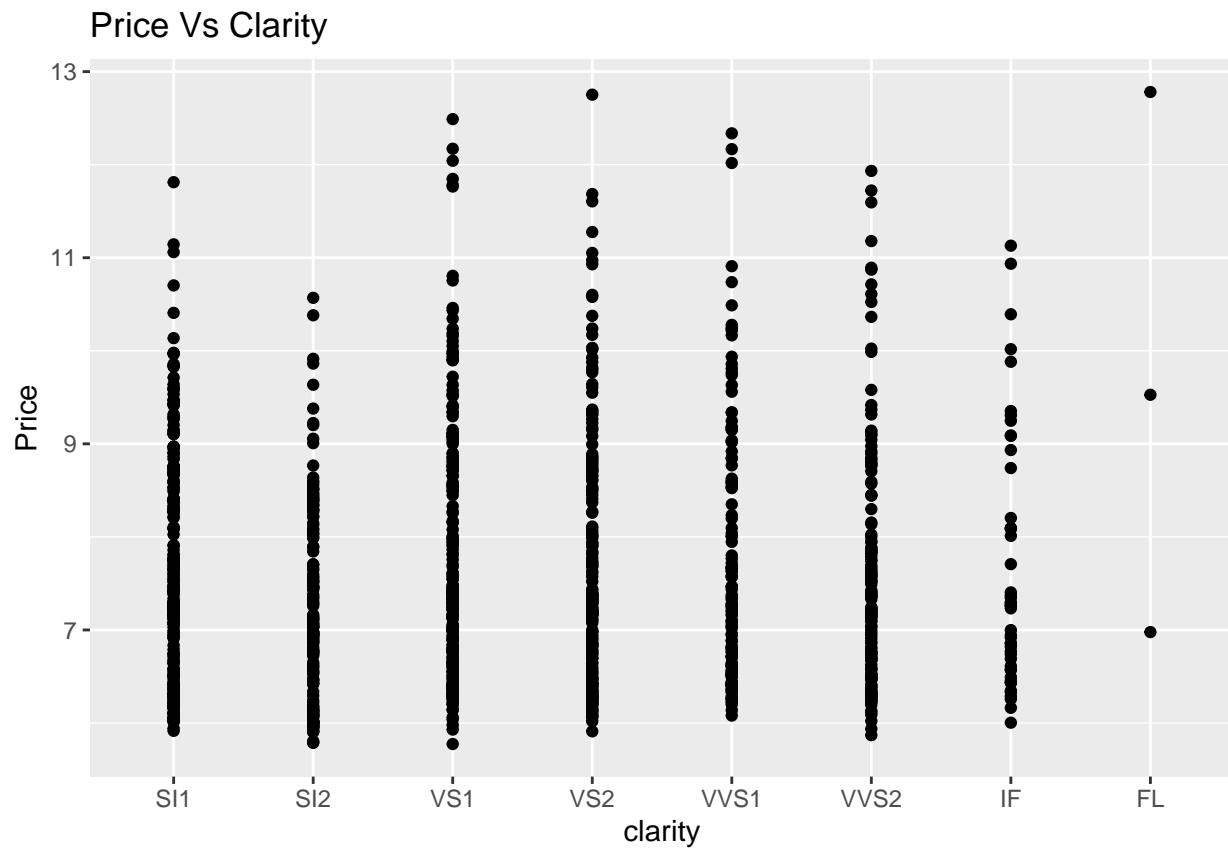
```
## `geom_smooth()` using formula 'y ~ x'
```



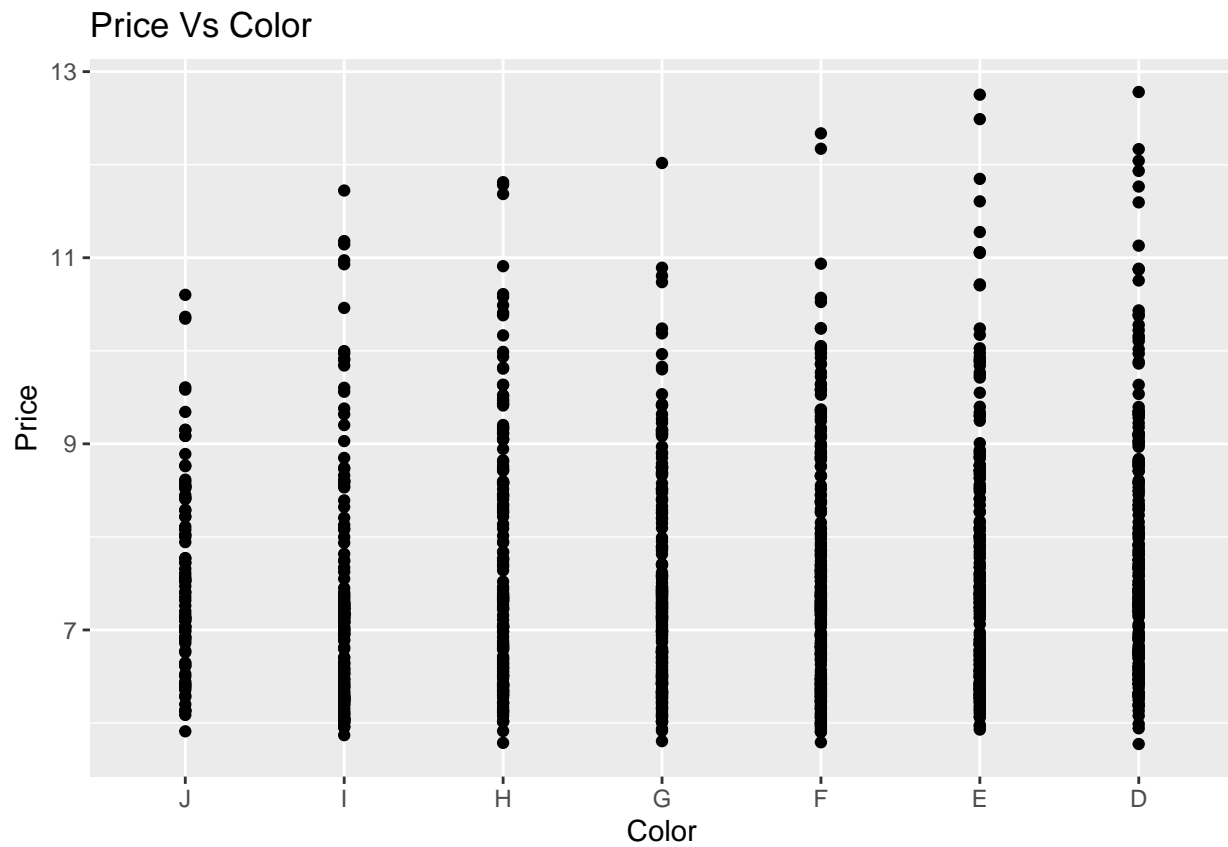
```
result <- lm(price~carat, Data)  
Data$yhat<-result$fitted.values  
Data$res<-result$residuals  
ggplot(Data, aes(x=yhat,y=res))+  
  geom_point()+  
  geom_hline(yintercept=0, color="red")+  
  labs(x="Fitted y", y="Residuals", title="Residual Plot")
```



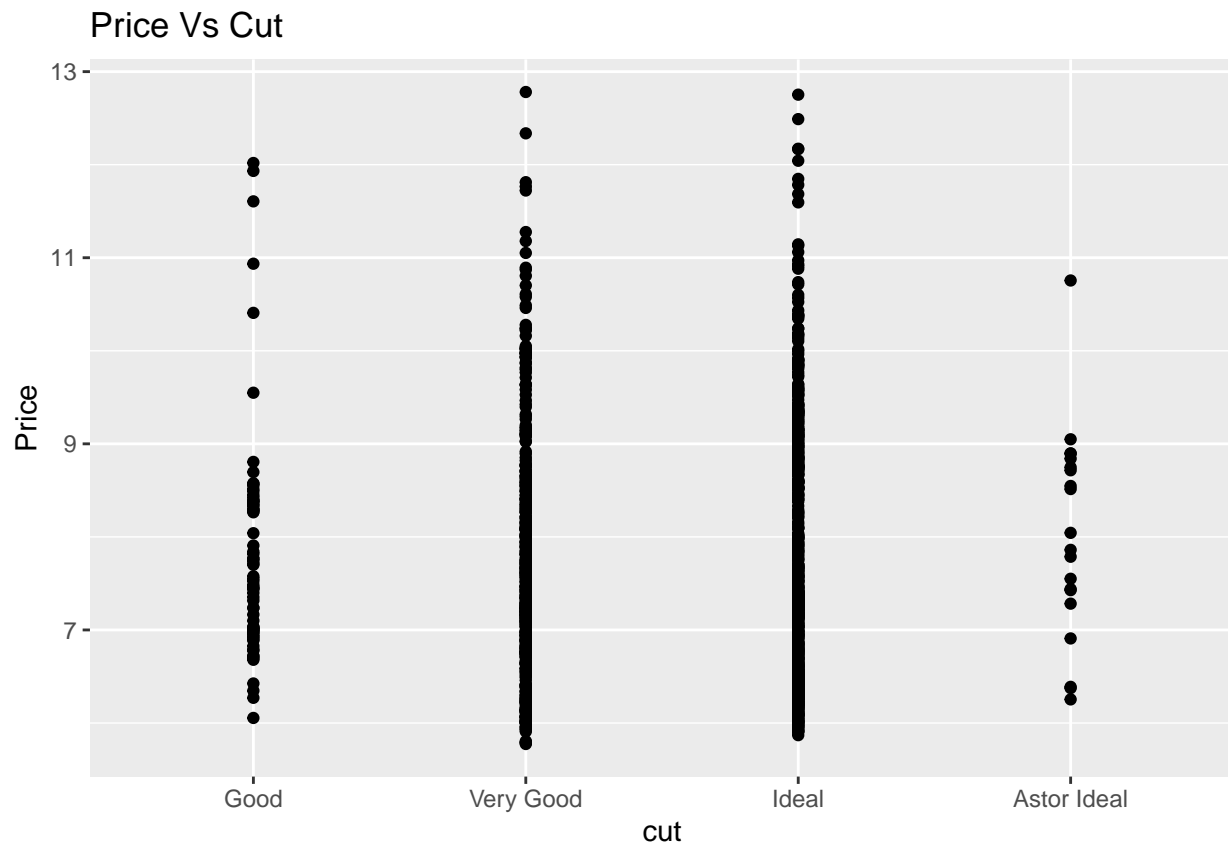
```
ggplot(Data, aes(x= clarity, y=log(price)))+  
  geom_point()+  
  labs(x= "clarity",  
        y= "Price",  
        title = "Price Vs Clarity")
```



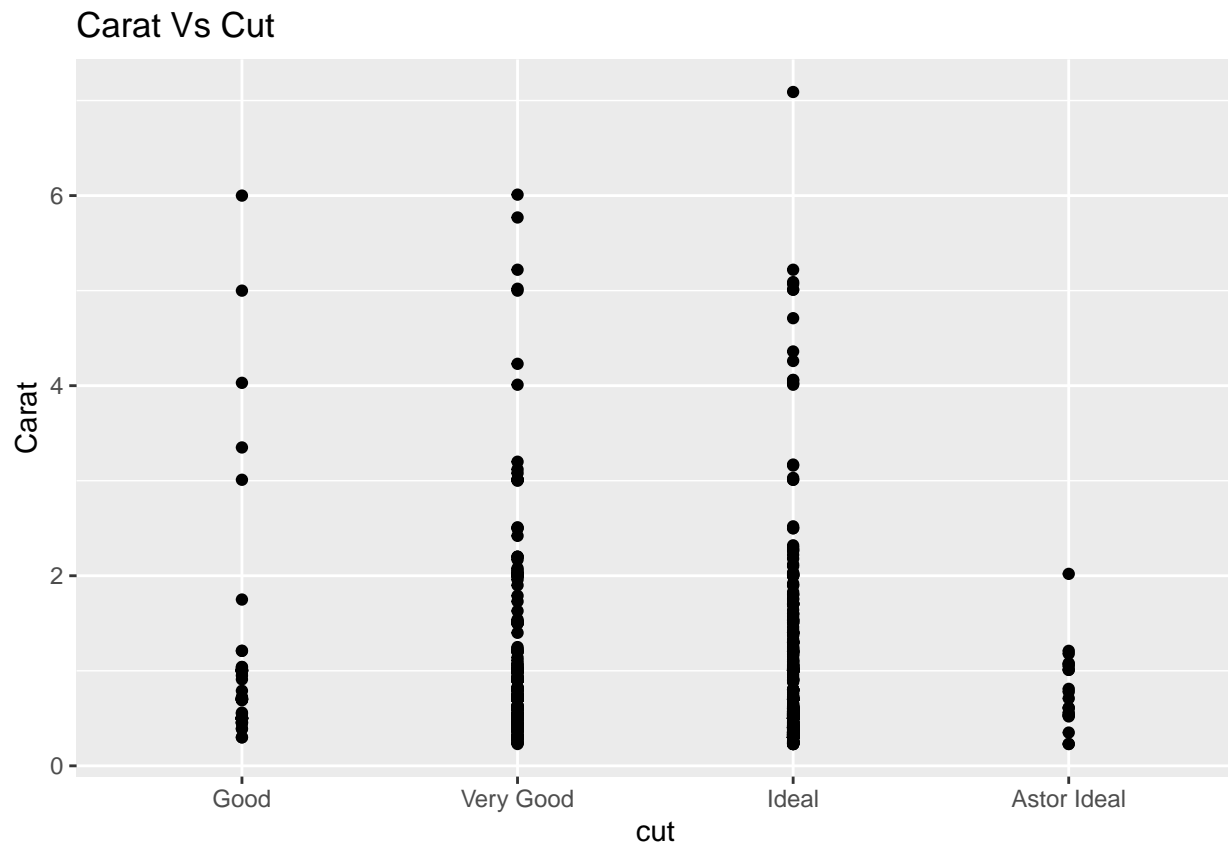
```
ggplot(Data, aes(x= color, y=log(price)))+
  geom_point()+
  labs(x= "Color",
       y= "Price",
       title = "Price Vs Color")
```



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

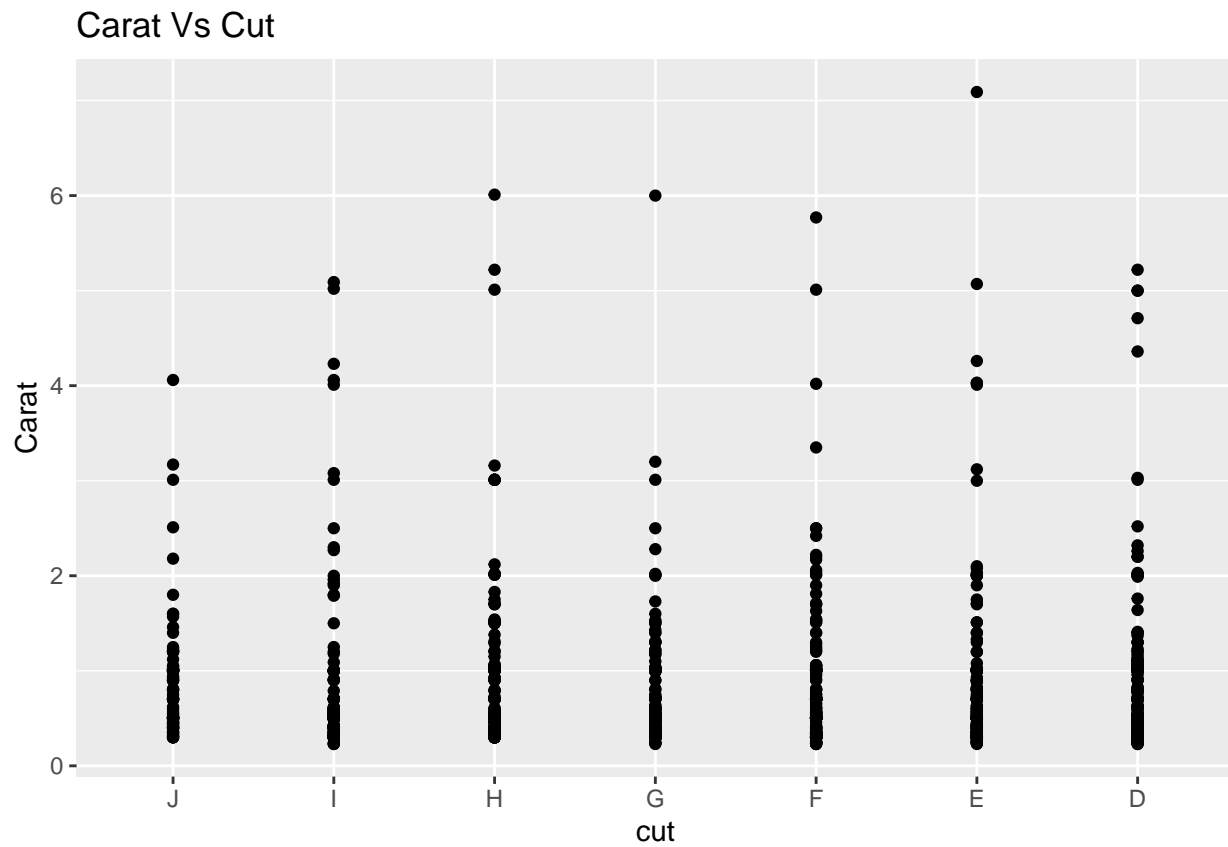


```
ggplot(Data, aes(x= cut, y=carat))+  
  geom_point()+  
  labs(x= "cut",  
        y= "Carat",  
        title = "Carat Vs Cut")
```

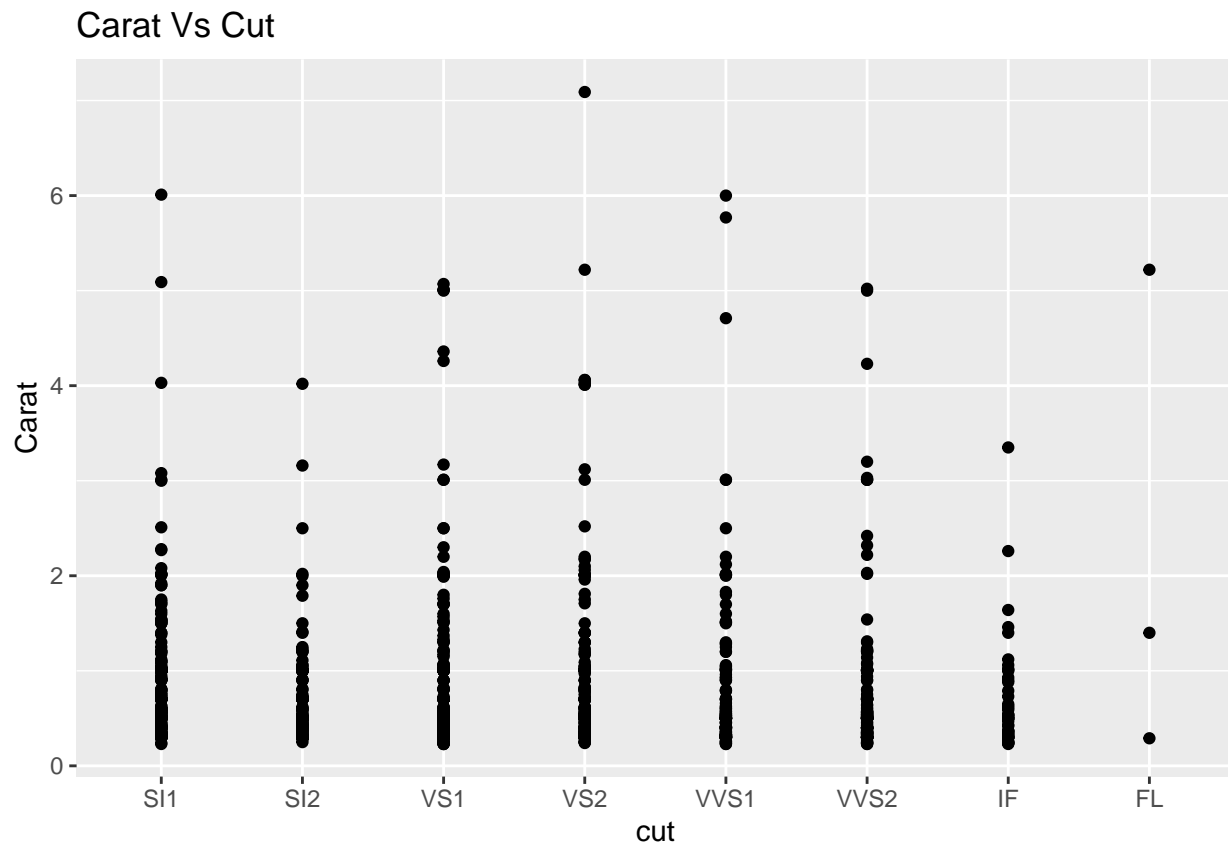


#Carat vs Cut is really low when you get to Astor Ideal

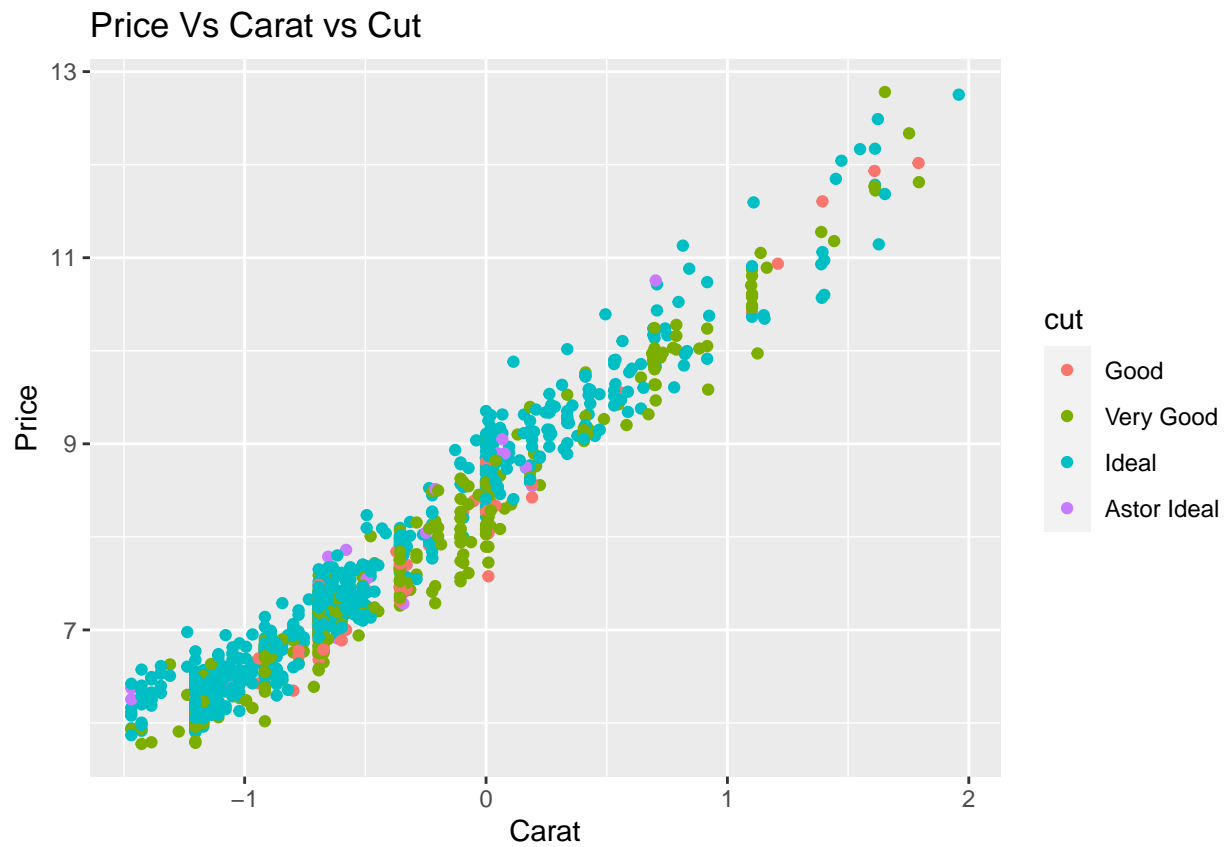
```
ggplot(Data, aes(x= color, y=carat))+  
  geom_point()+  
  labs(x= "cut",  
        y= "Carat",  
        title = "Carat Vs Cut")
```



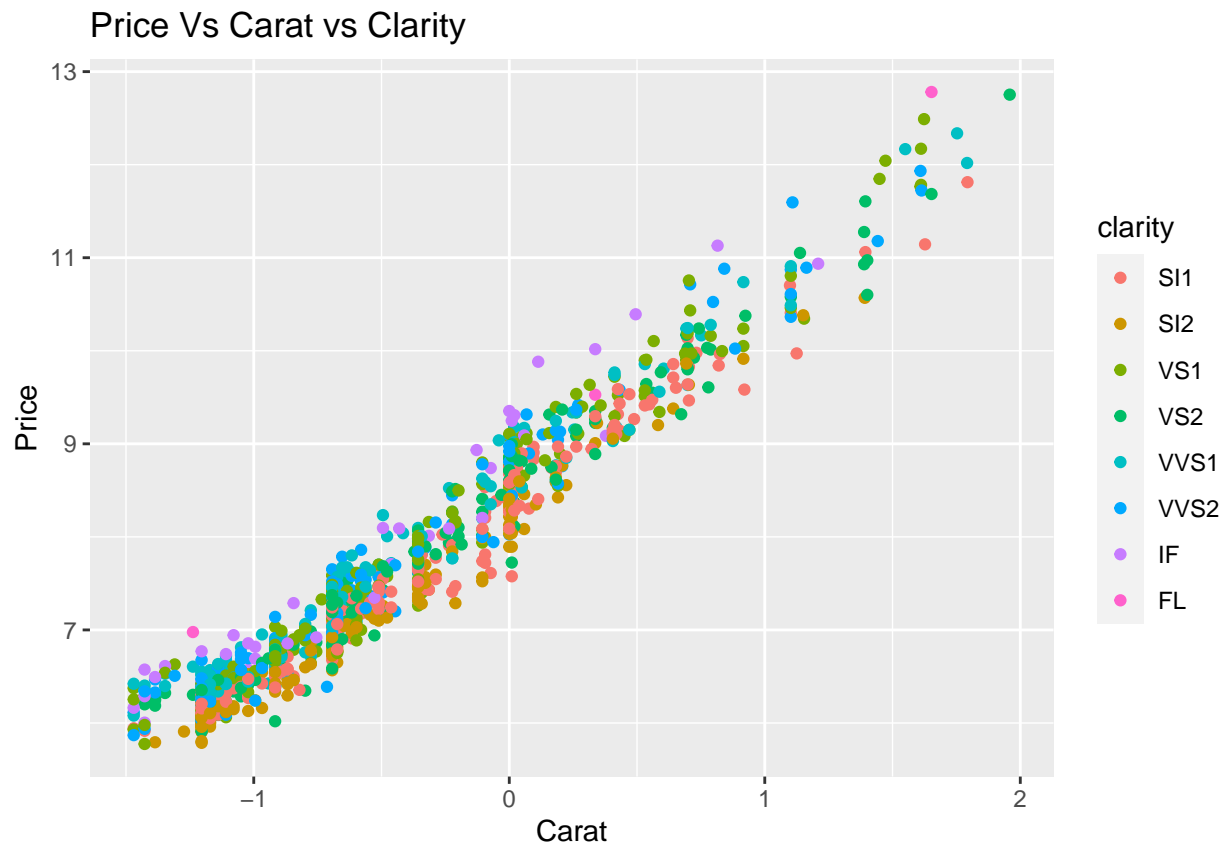
```
ggplot(Data, aes(x= clarity, y=carat))+  
  geom_point()+  
  labs(x= "cut",  
        y= "Carat",  
        title = "Carat Vs Cut")
```



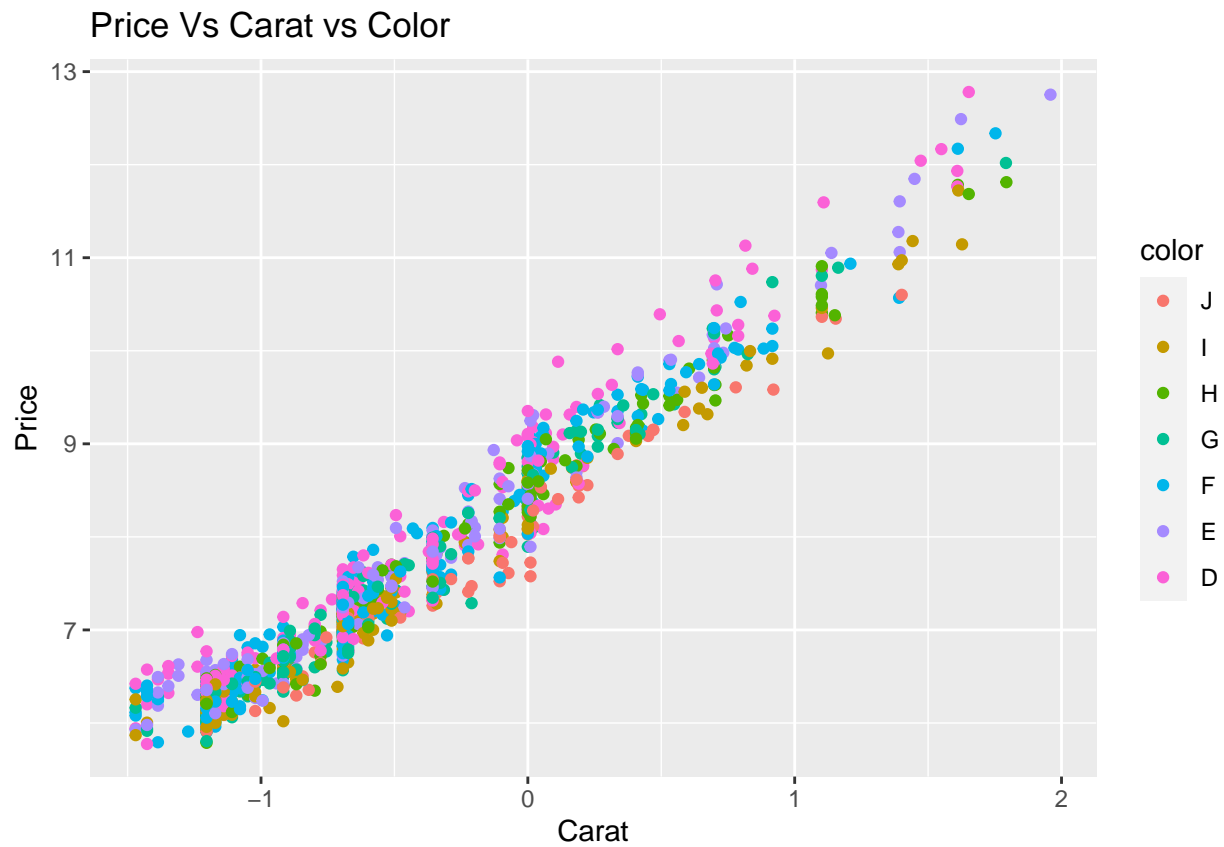
```
Data$cut <- factor(Data$cut, levels = c("Good", "Very Good", "Ideal", "Astor Ideal"))
ggplot(Data, aes(x=log(carat), y=log(price), color = cut))+
  geom_point()+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Cut")
```

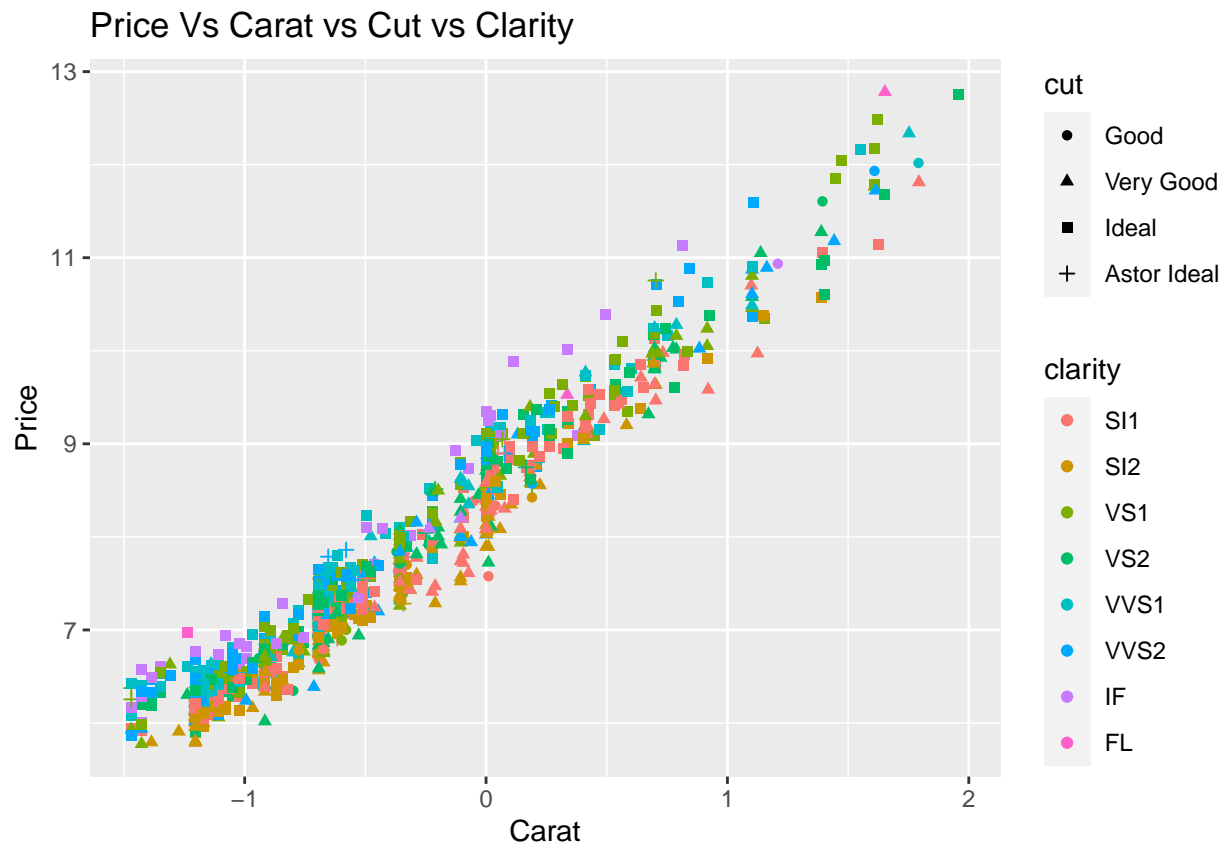
```
ggplot(Data, aes(x=log(carat), y=log(price), color = clarity))+  
  geom_point()+  
  labs(x= "Carat",  
        y= "Price",  
        title = "Price Vs Carat vs Clarity")
```



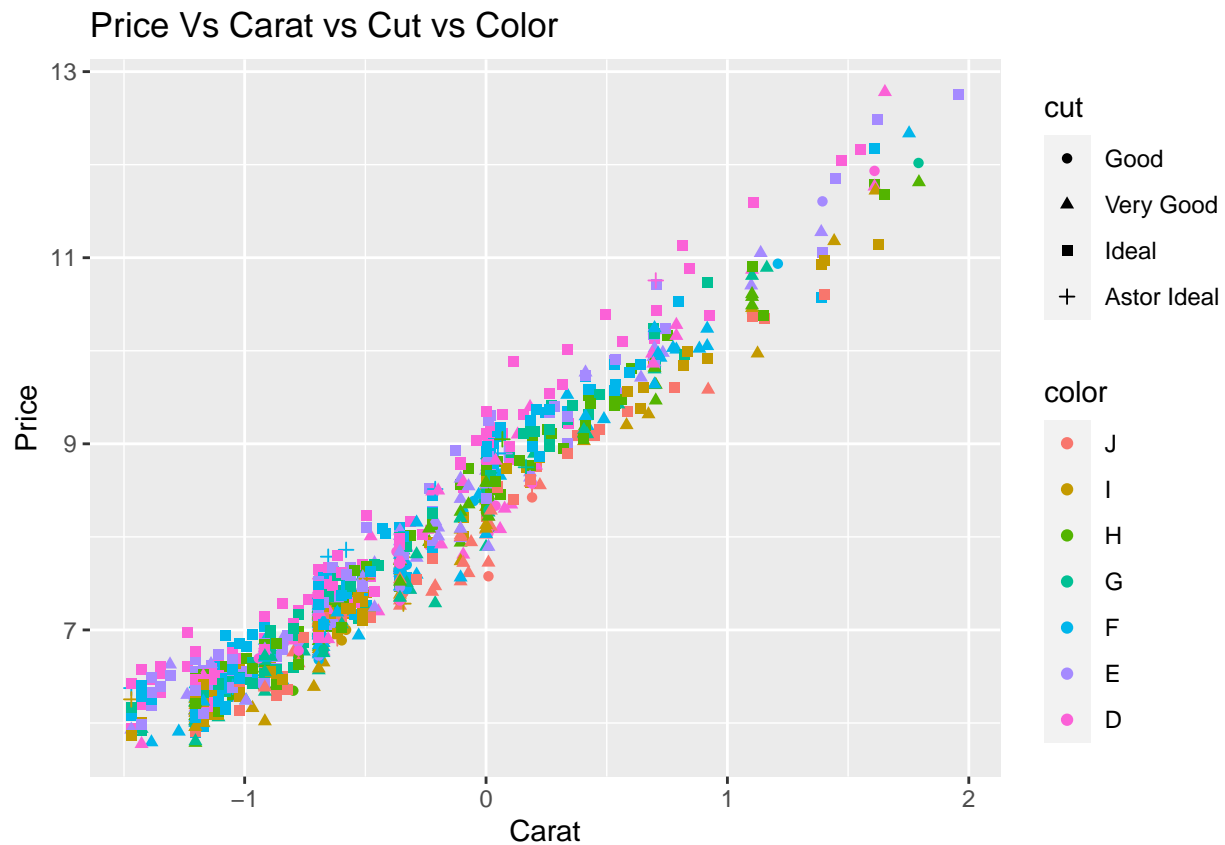
```
ggplot(Data, aes(x=log(carat), y=log(price), color = color))+  
  geom_point()+  
  labs(x= "Carat",  
        y= "Price",  
        title = "Price Vs Carat vs Color")
```



```
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")
```



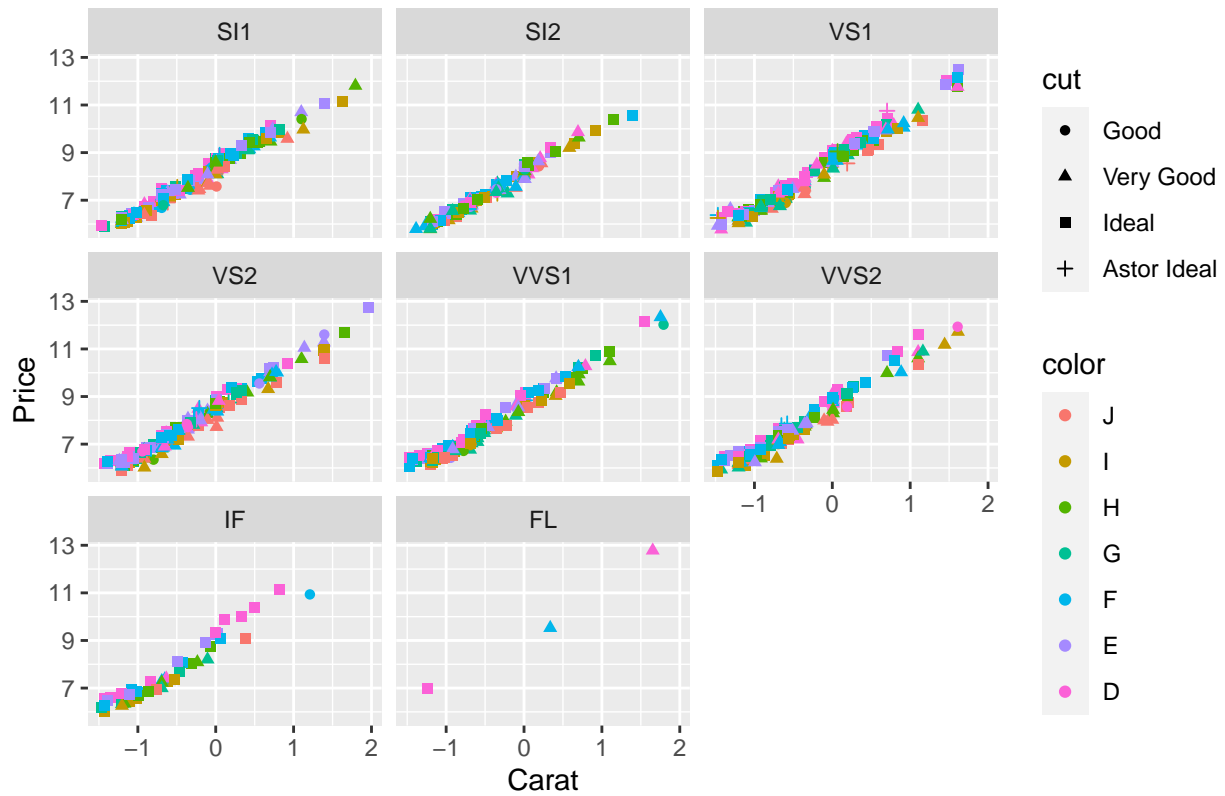
```
ggplot(Data, aes(x=log(carat), y=log(price), color = color, shape = cut))+
  geom_point()+
  labs(x= "Carat",
       y= "Price",
       title = "Price Vs Carat vs Cut vs Color")
```



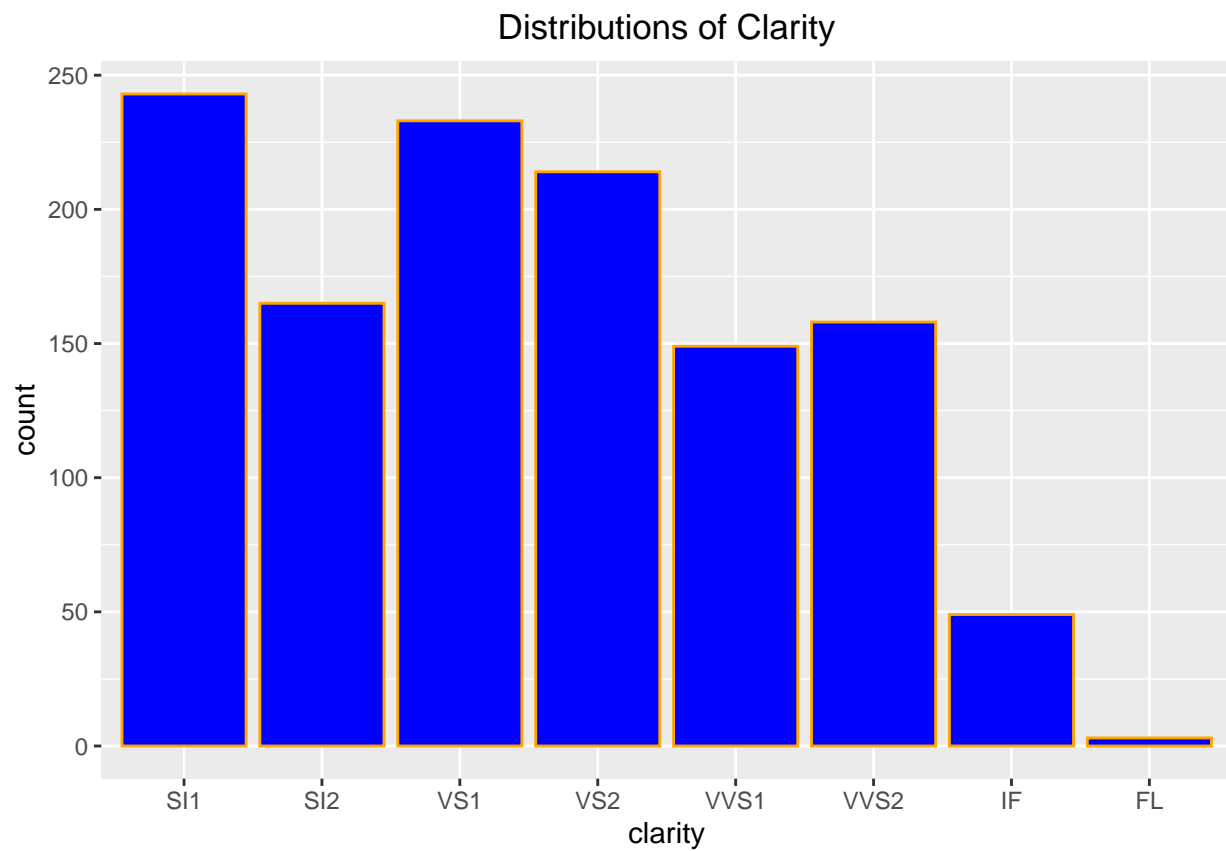
```
#ggplot(Data, aes(x=log(carat), y=log(price), shape = color, color = clarity))+
#geom_point()+
#labs(x= "Carat",
#      y= "Price",
#      title = "Price Vs Carat vs Cut vs Clarity")
# bad because I can't show all the data points

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")
```

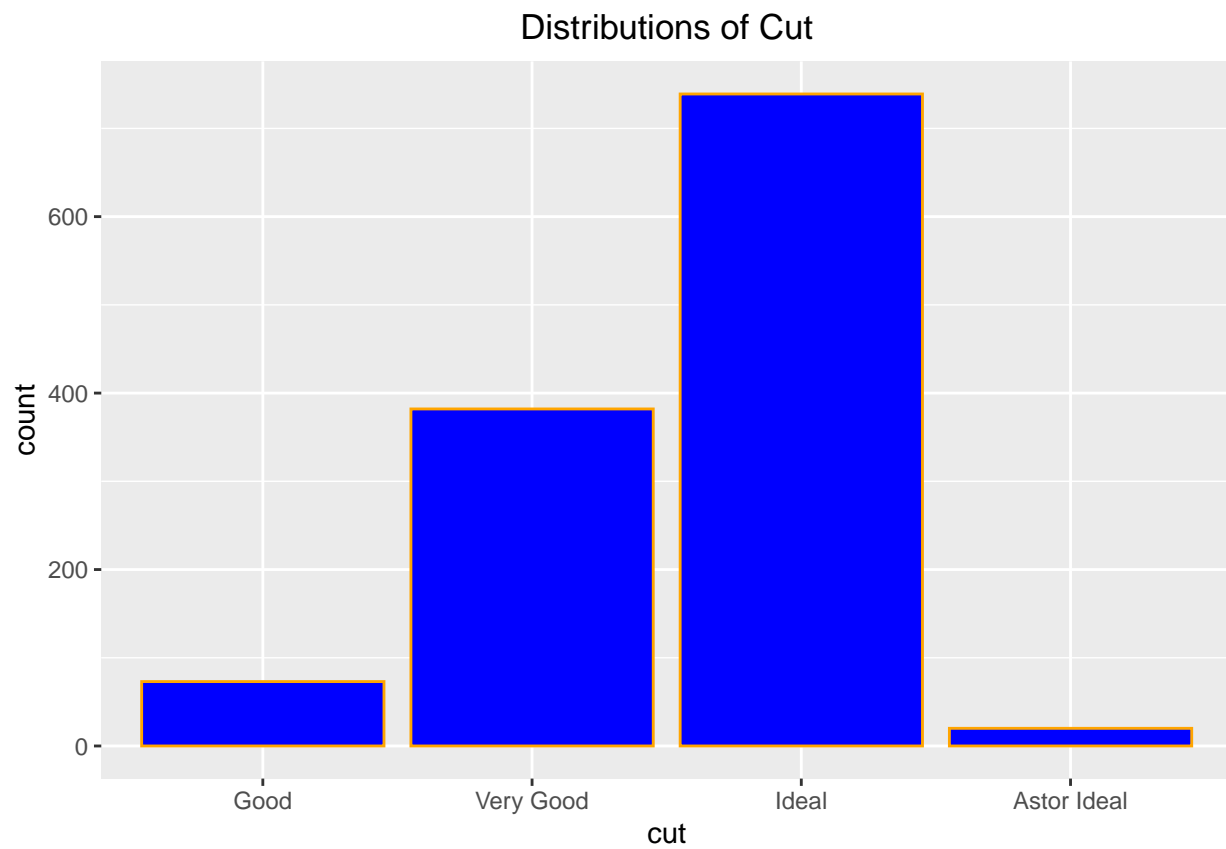
Price Vs Carat vs Cut vs Color vs Clarity



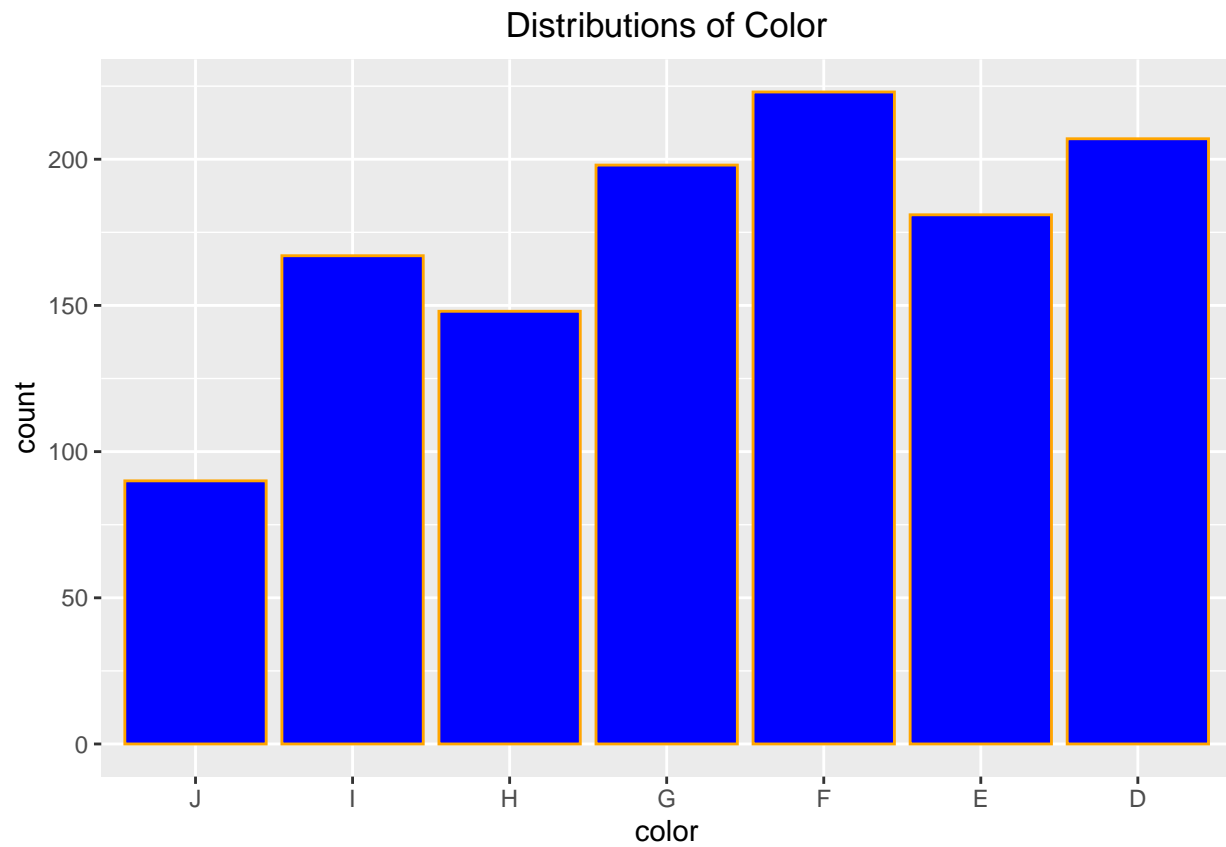
```
ggplot(Data)+
  aes(x=clarity)+
  geom_bar(fill="blue",color="orange")+
  labs(title = "Distributions of Clarity") +
  theme(
    plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 0)
  )
```



```
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=cut)+  
  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

