

data visualization (qplot)

Pradip Basnet

task 1: importing the packages

```
library(ggplot2)
```

#diamonds dataset with qplot

task 2: loading the diamonds datset

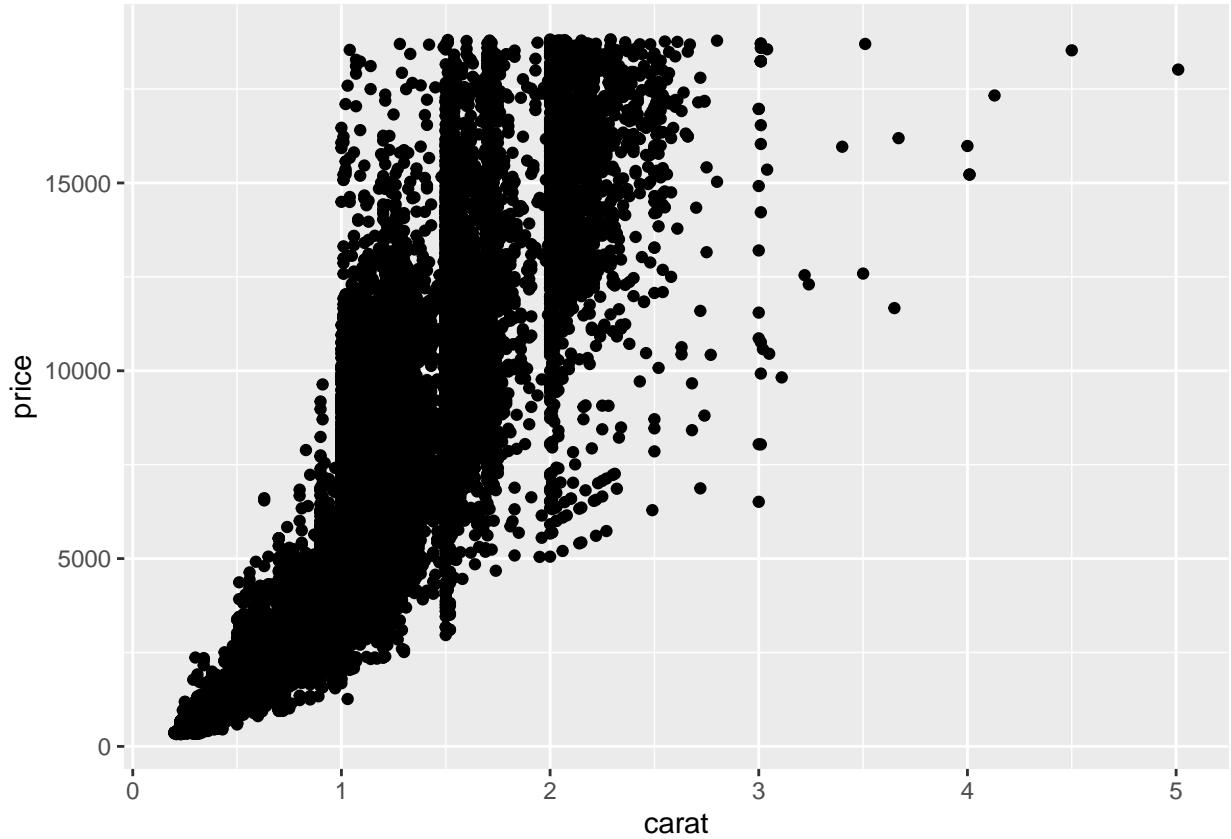
```
data("diamonds")
head(diamonds)
```

```
## # A tibble: 6 x 10
##   carat cut      color clarity depth table price     x     y     z
##   <dbl> <ord>    <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1  0.23 Ideal    E     SI2     61.5    55    326  3.95  3.98  2.43
## 2  0.21 Premium  E     SI1     59.8    61    326  3.89  3.84  2.31
## 3  0.23 Good     E     VS1     56.9    65    327  4.05  4.07  2.31
## 4  0.29 Premium  I     VS2     62.4    58    334  4.2   4.23  2.63
## 5  0.31 Good     J     SI2     63.3    58    335  4.34  4.35  2.75
## 6  0.24 Very Good J     VVS2    62.8    57    336  3.94  3.96  2.48
```

task 3: loading the data directly using the qplot function, use of qplot to produce the scatter plot showing the relaationship of the carat and price

```
qplot(carat,price,data=diamonds)
```

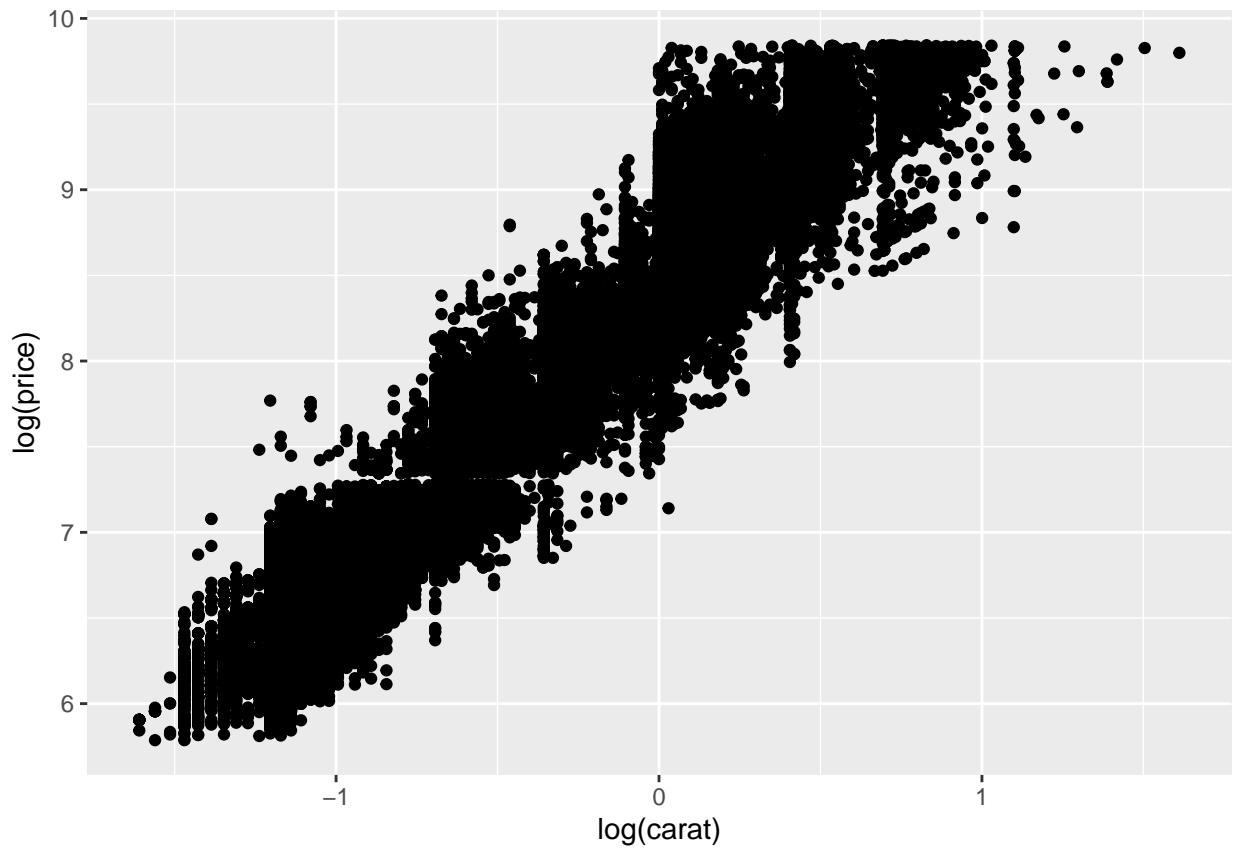
```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



#Variables Transformation

task 4: use of a log function to transform variables in a way to make the relationship between variables less sensitive. Outliers in the data become more apparent after the transformation.

```
qplot(log(carat), log(price), data=diamonds)
```



```
#aesthetic attributes (color,shape,size)
```

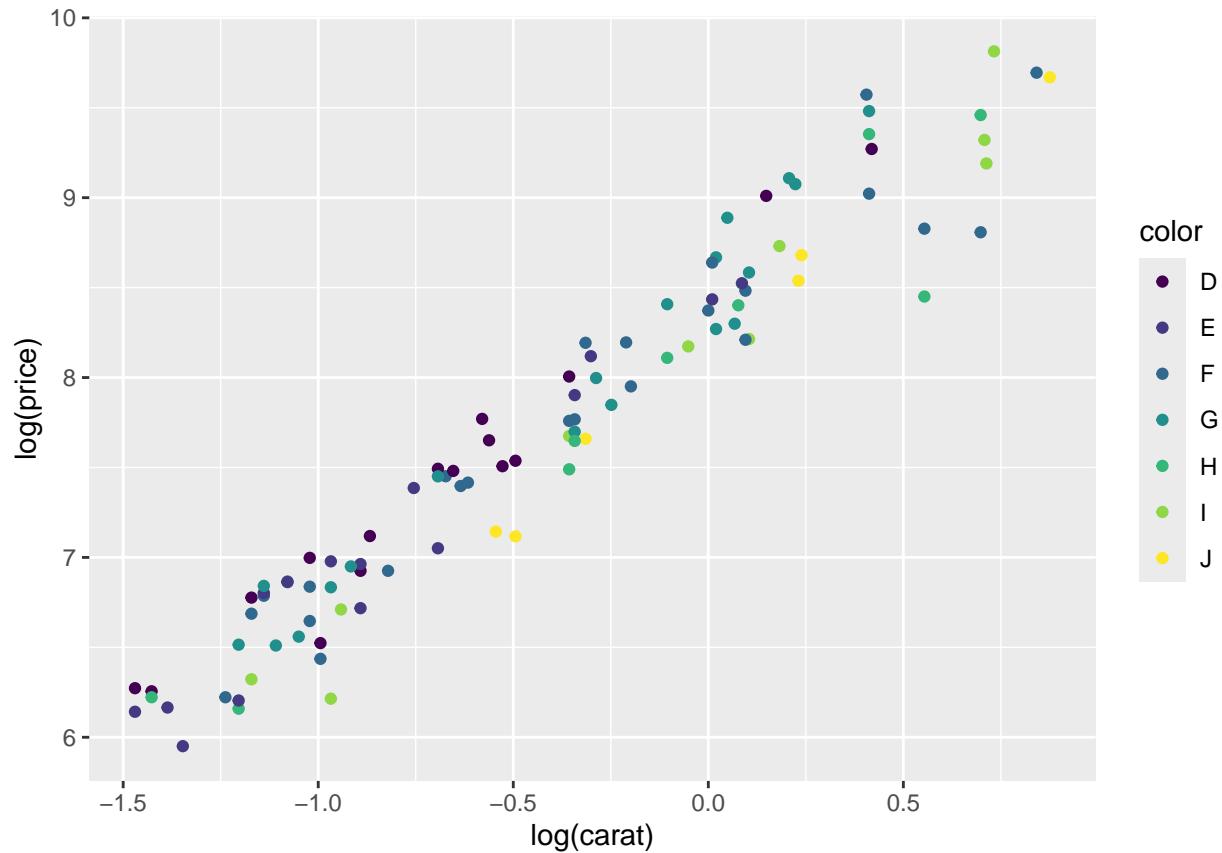
task 5: create a smaller dataset from diaminds by randomly sample. create a reproducible random sample of 100 rows from the diamonds dataset

```
set.seed(1000)
dsmall<-diamonds[sample(nrow(diamonds),100),]
```

```
#Color aesthetic
```

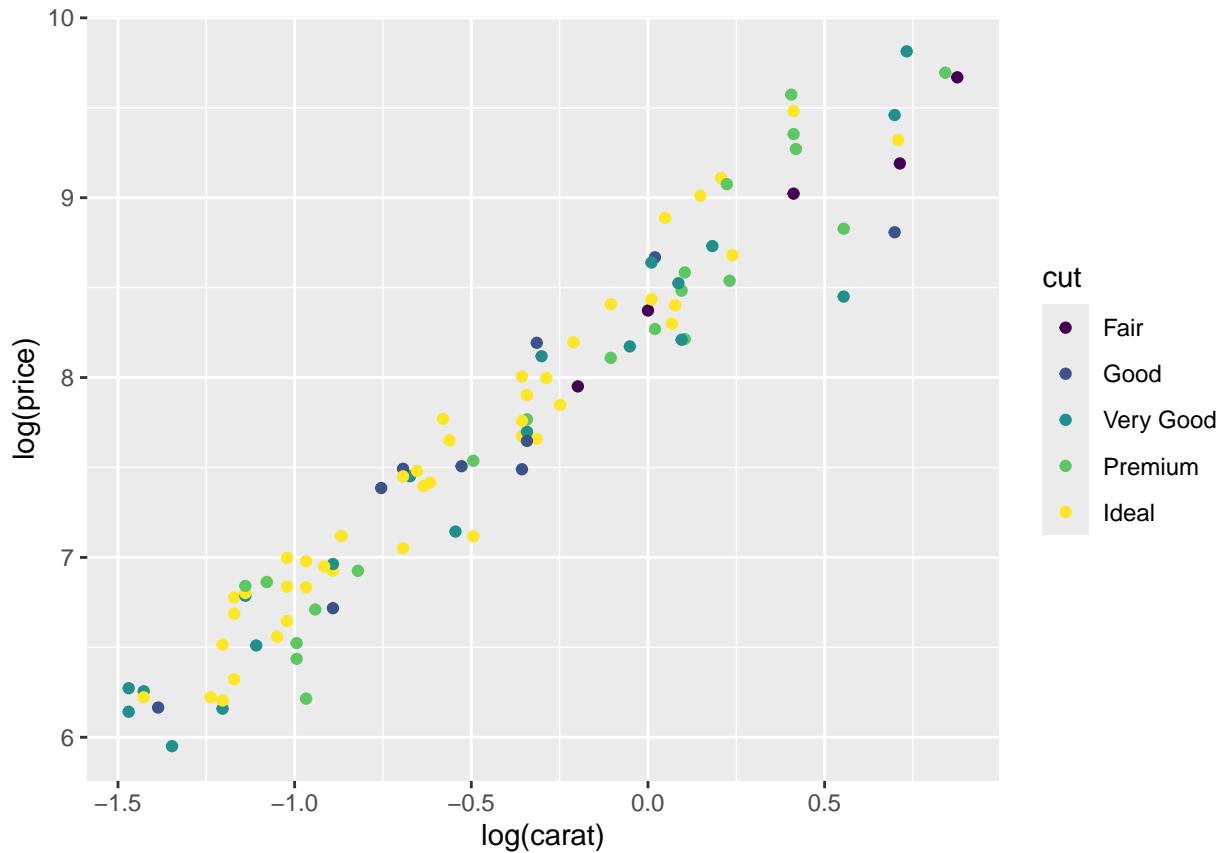
task 6: create a scatterplot usiing qplot function to show the relationship between the logarithm of carat and the logarithm of the price for a random sample of 100 diamonds

```
qplot(log(carat),log(price),data=dsmall,colour=color)
```



task 7: create a scatter plot to visualize the relationship between the logarithm of carat weight and the logarithm of price for a random sample of 100 diamonds. The points are colored based on the cut variable

```
qplot(log(carat), log(price), data = dsmall, colour=cut)
```

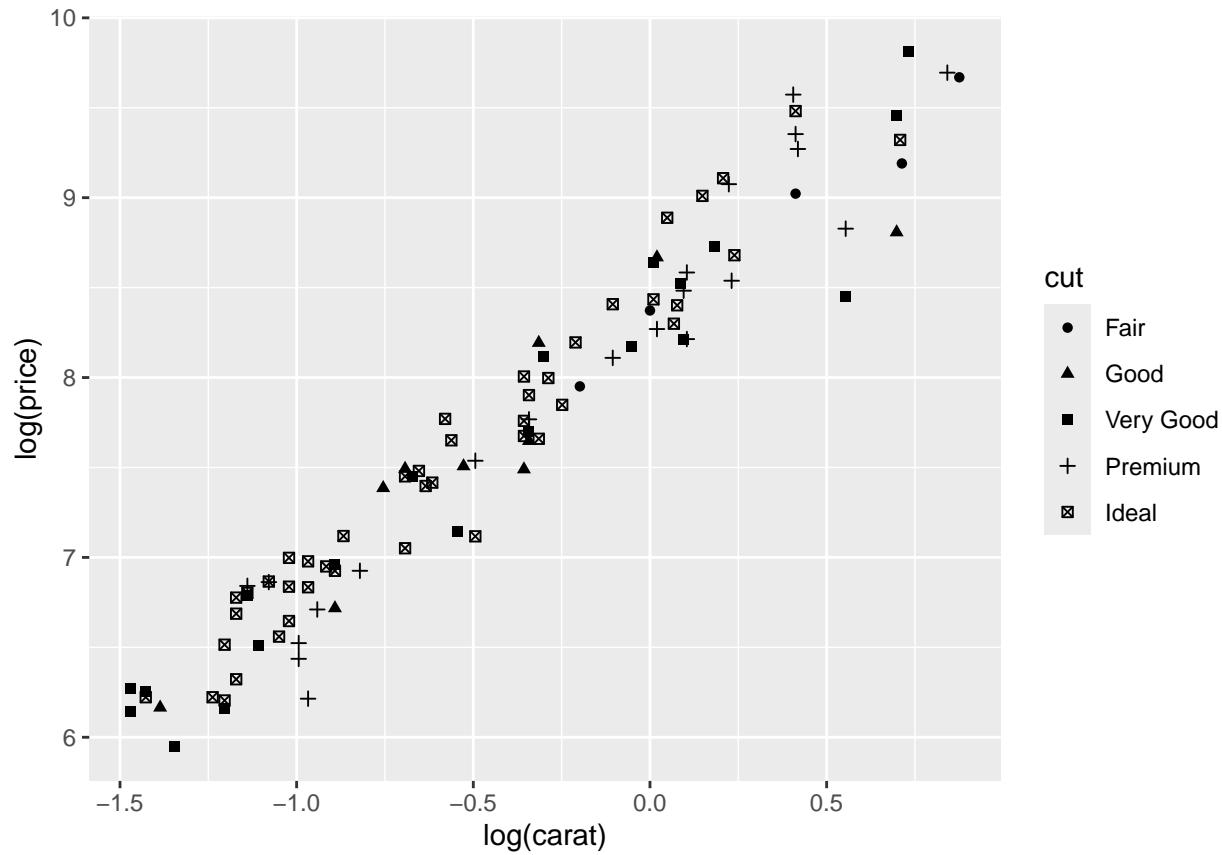


#Shape aesthetic

task 8:create a scatter plot to visualize the relationship between the logarithm of carat weight and the logarithm of price for a random sample of 100 diamonds. The points are shaped based on the cut variable

```
qplot(log(carat),log(price),data = dsmall,shape=cut)
```

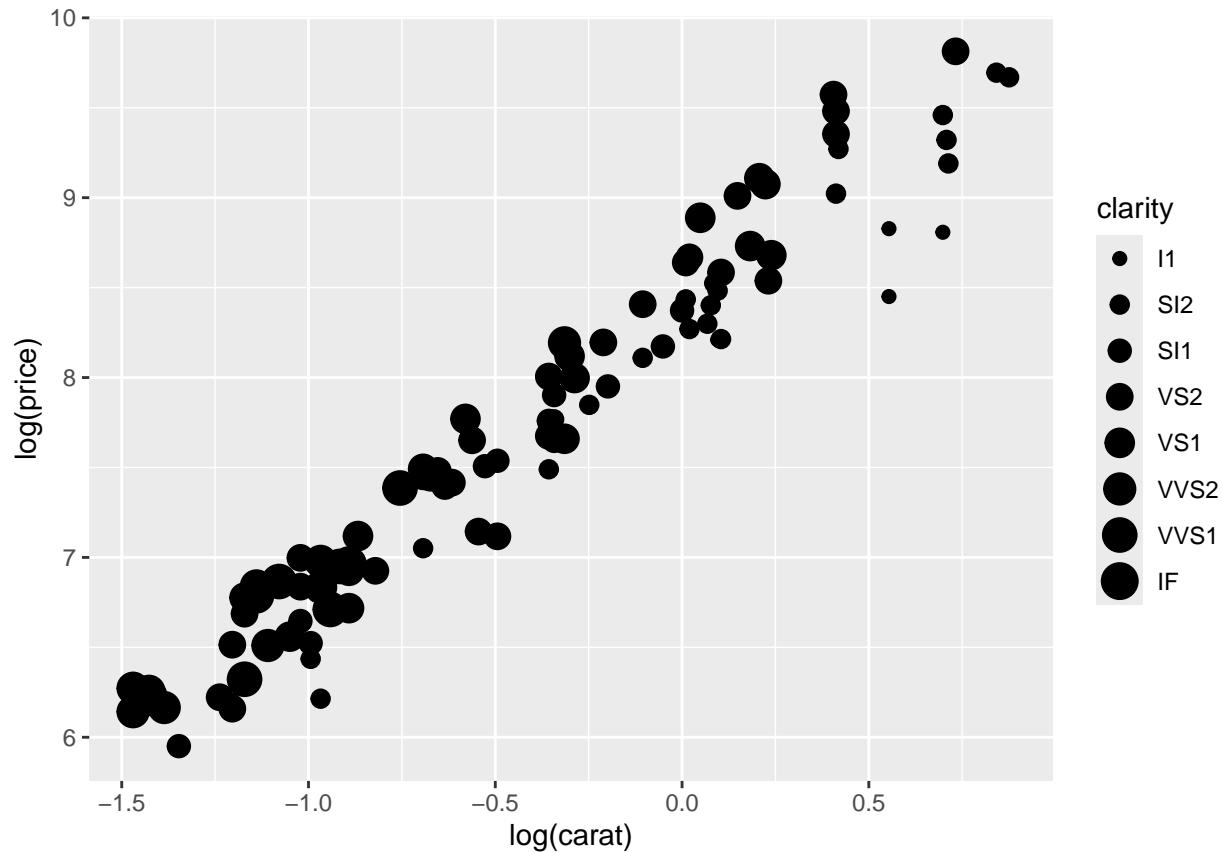
```
## Warning: Using shapes for an ordinal variable is not advised
```



```
#Exercise 1
```

```
##a. Apply what you have done previously to dsmall dataset with other aesthetic attribute such as 'size'.
```

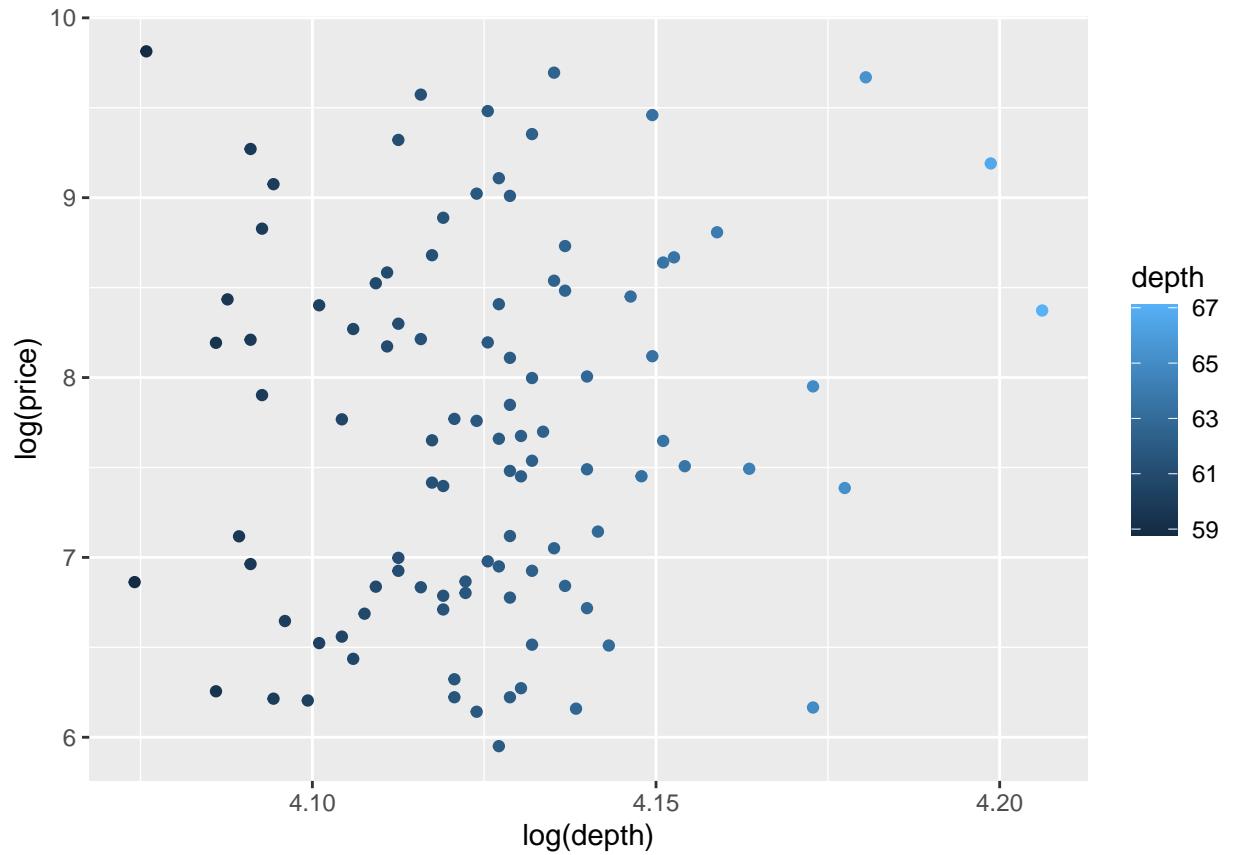
```
qplot(log(carat), log(price), data = dsmall, size=clarity)
```



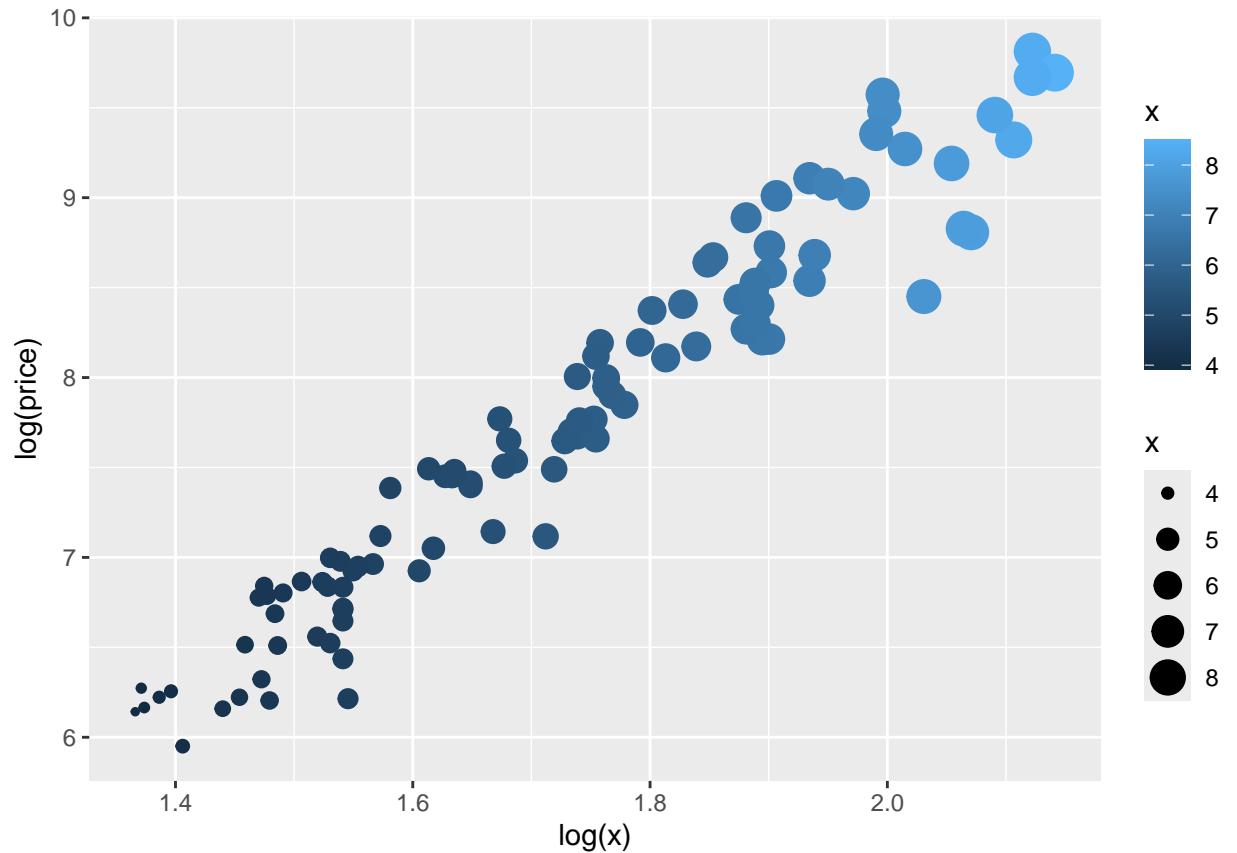
```
#View(diamonds)
```

```
##Investigate the relationship between other variables (using colour, shape and size aesthetics) to see if you can find interesting information.
```

```
qplot(log(depth),log(price),data = dsmall,colour=depth)
```



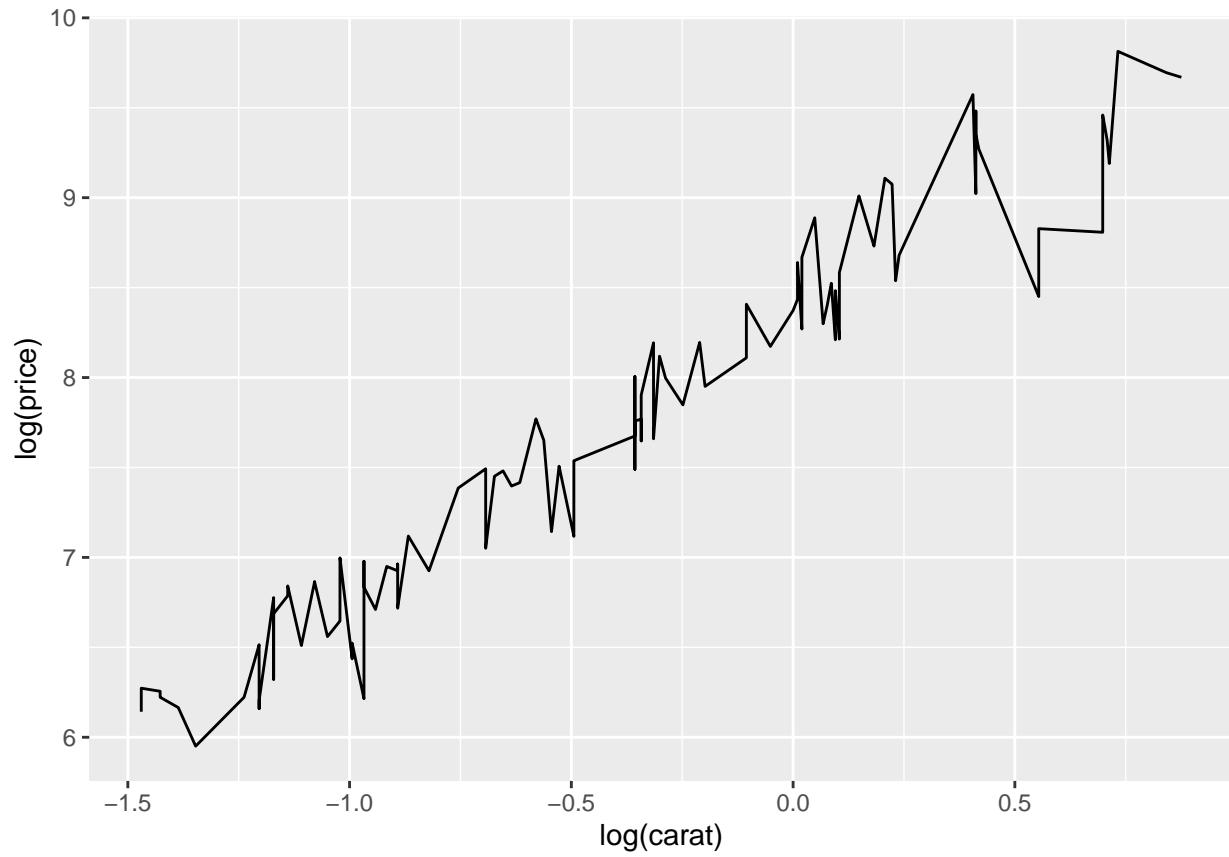
```
qplot(log(x),log(price),data = dsmall,size=x, colour=x)
```



#Geom 2D You can change the kind of plot by explicitly define geom that “describes the type of object that is used to display the data”

task 9: visualize the relationship between the natural logarithm of carat and price using a line plot.

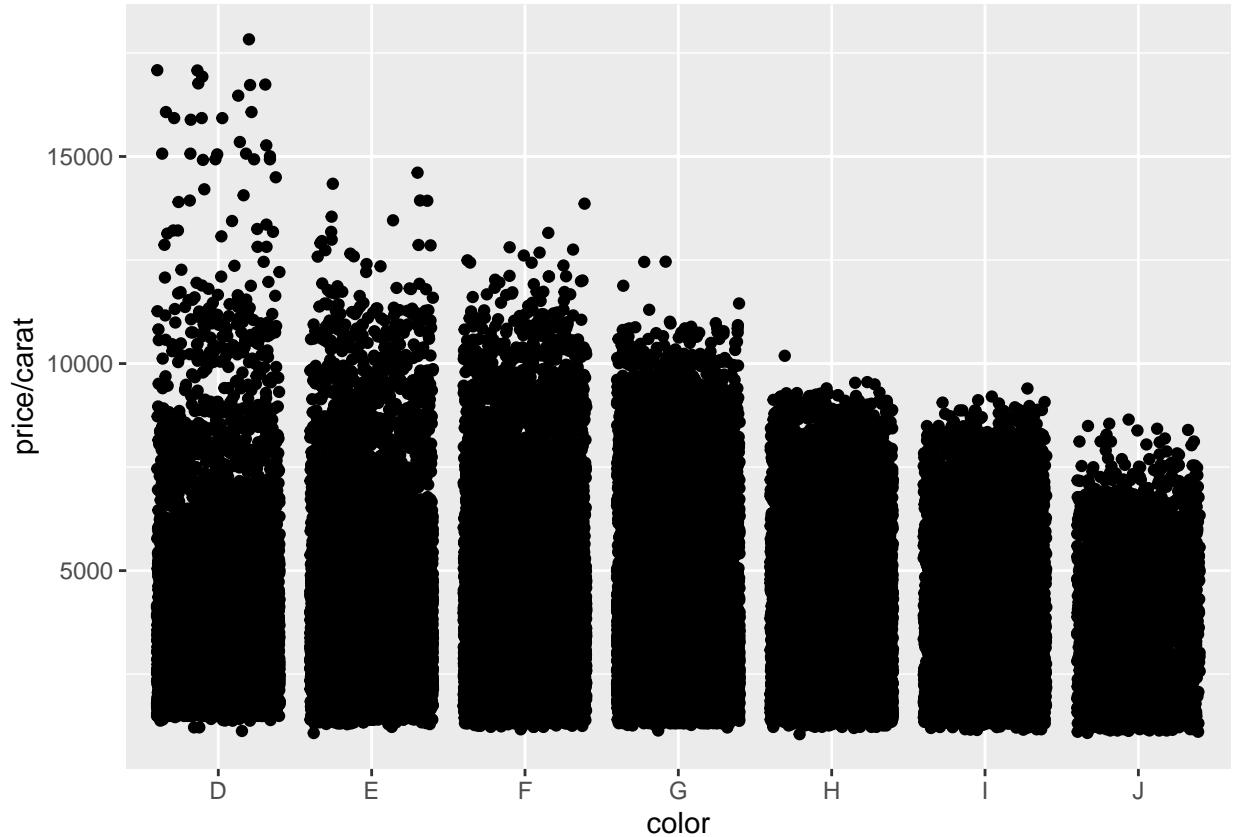
```
qplot(log(carat),log(price), data = dsmall,geom="line")
```



```
#Boxplots ans Jittered
```

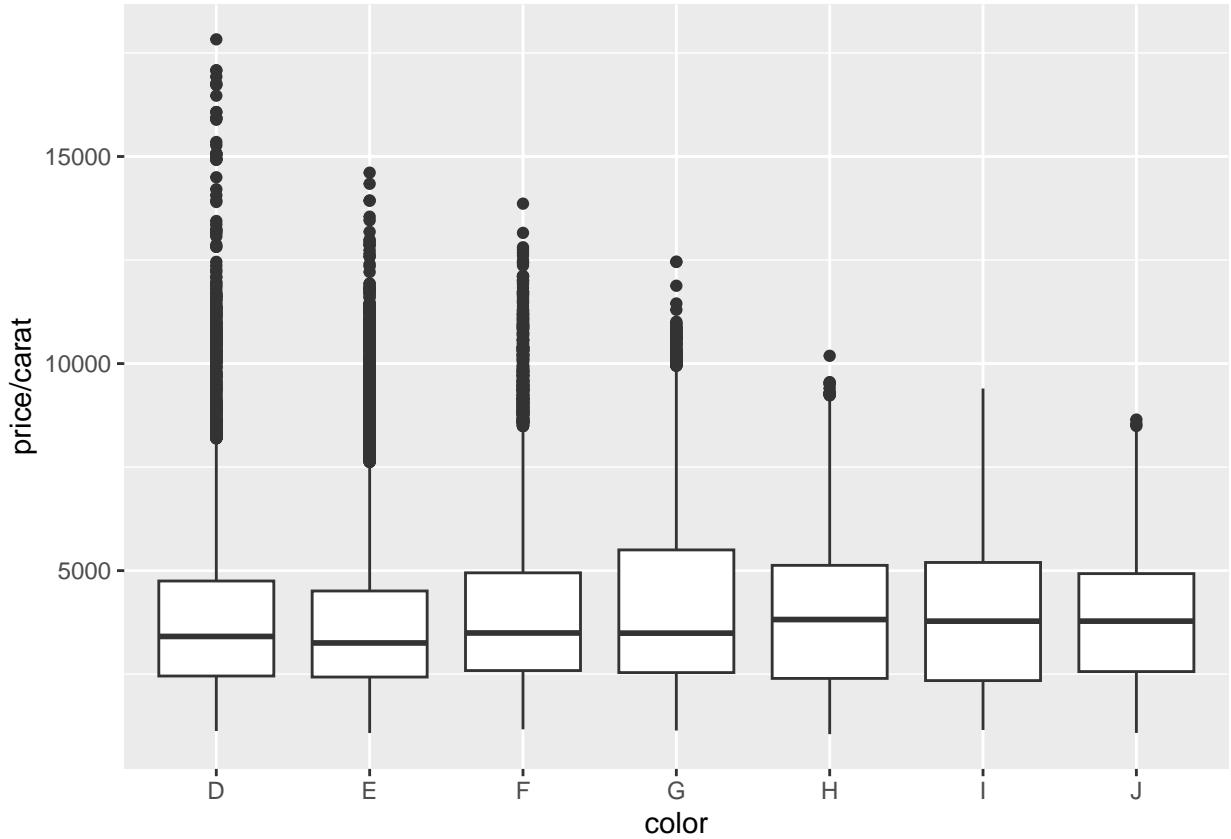
task 10: visualize how the price per carat (price/carat) varies across different color grades (color) of diamonds using a jitter plot

```
qplot(color,price/carat, data = diamonds, geom = "jitter")
```



task 11: visualize how the price per carat (price/carat) varies across different color grades (color) of diamonds using a boxplot

```
qplot(color, price/carat, data = diamonds, geom = "boxplot")
```

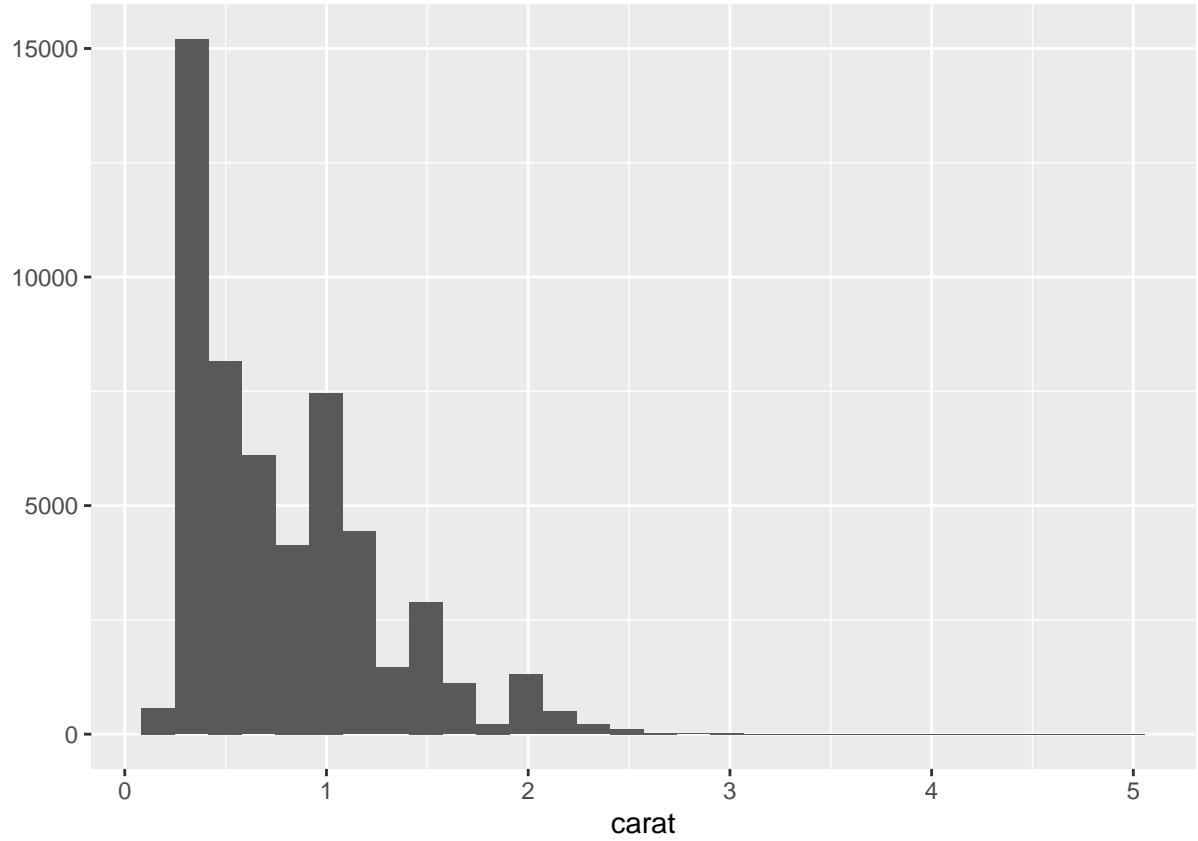


#Geom 1D

task 12: visualize the distribution of diamond carat weights (carat) using a histogram

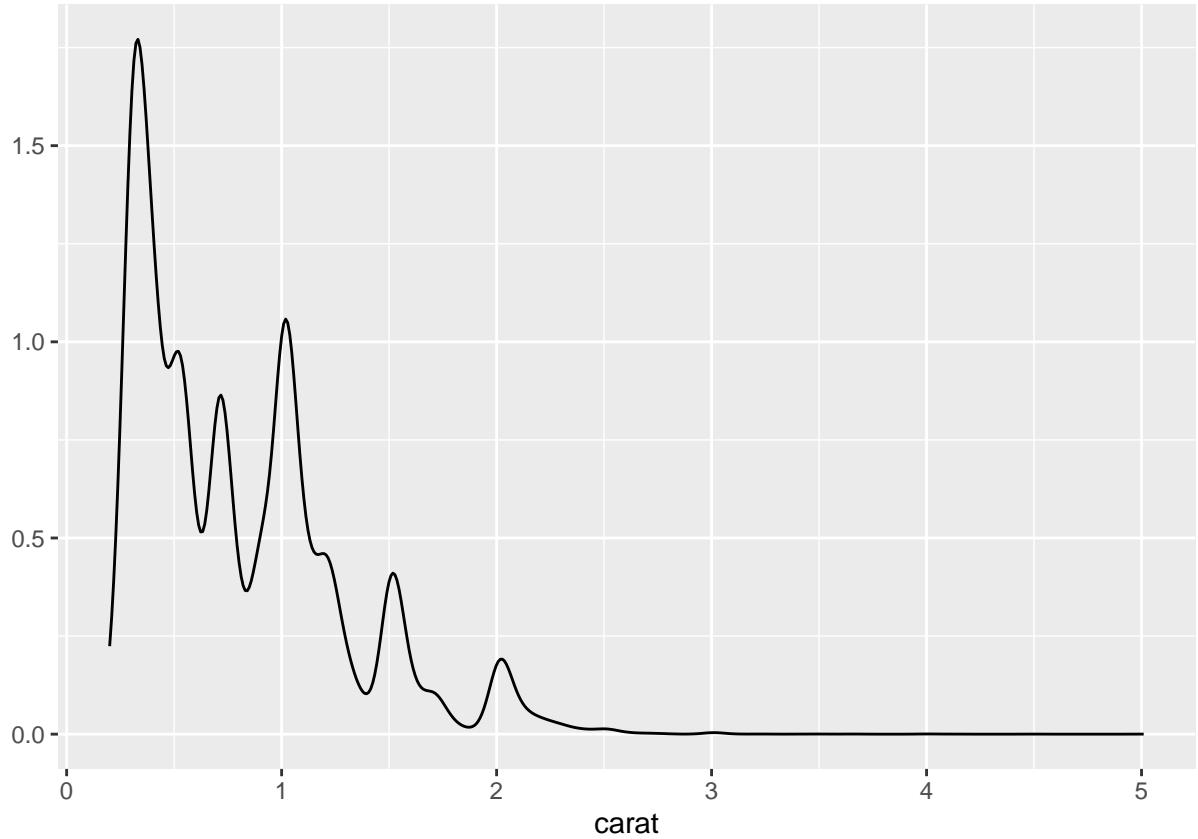
```
qplot(carat,data = diamonds,geom = "histogram")
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



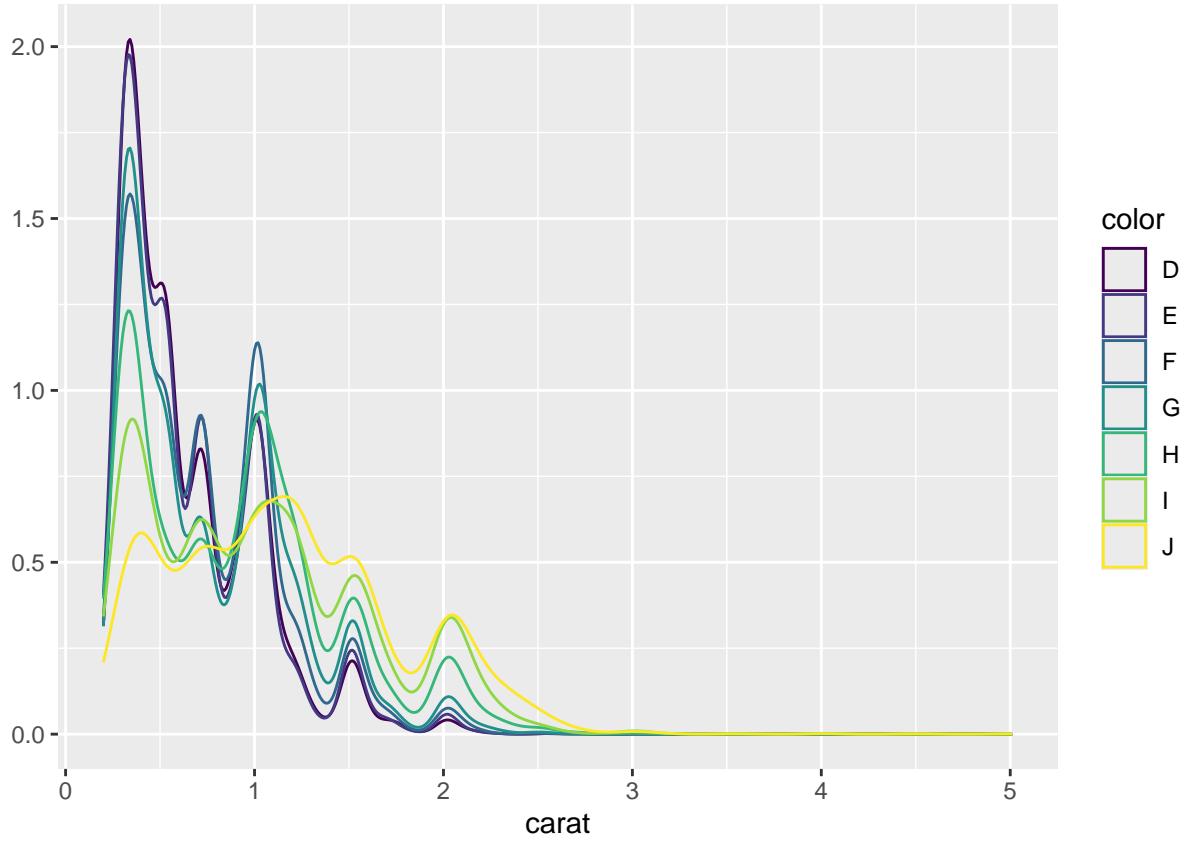
task 13: visualize the distribution of diamond carat weights (carat) using a density plot

```
qplot(carat,data = diamonds,geom = "density")
```



task 14: visualize how the distribution of diamond carat weights (carat) varies across different color grades (color) using a density plot

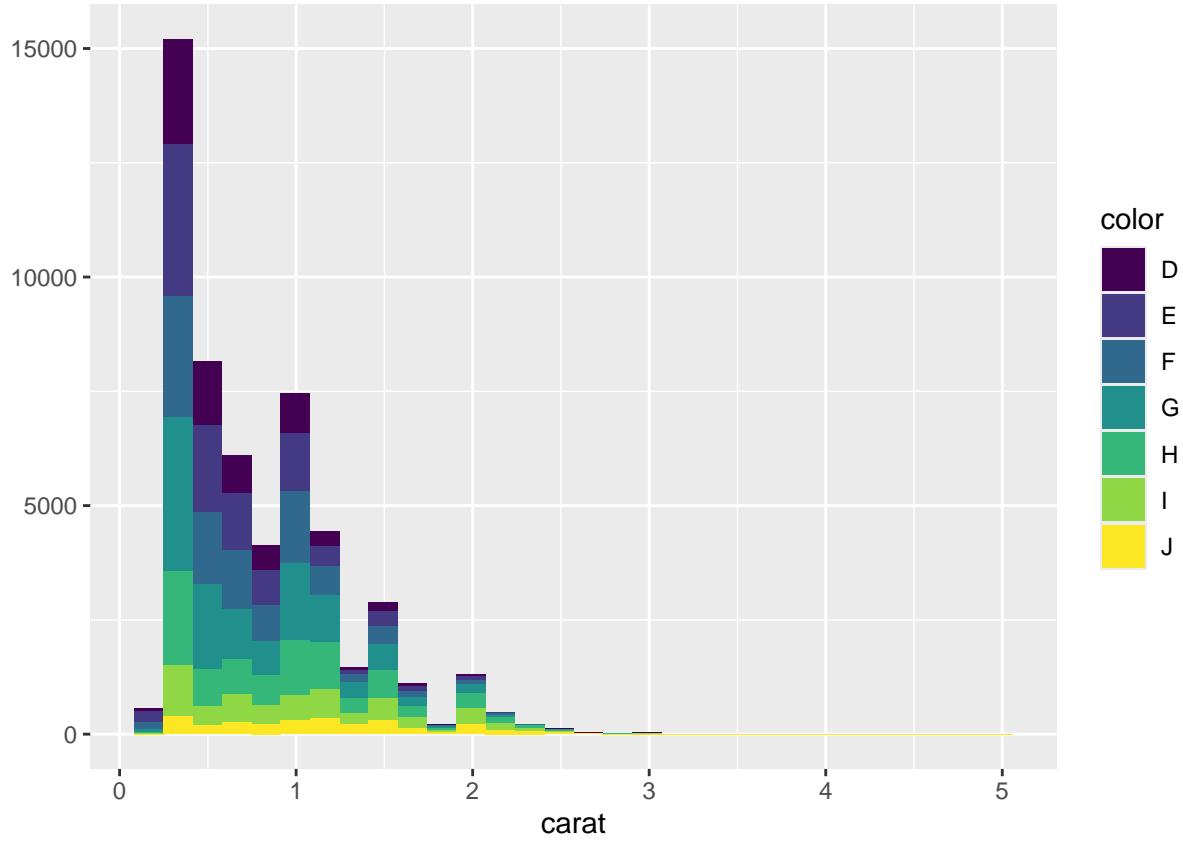
```
qplot(carat,data = diamonds,geom = "density",colour=color)
```



task 15: The purpose of this code is to visualize how the distribution of diamond carat weights (carat) varies across different color grades (color) using a filled histogram

```
qplot(carat,data = diamonds,geom = "histogram",fill=color)
```

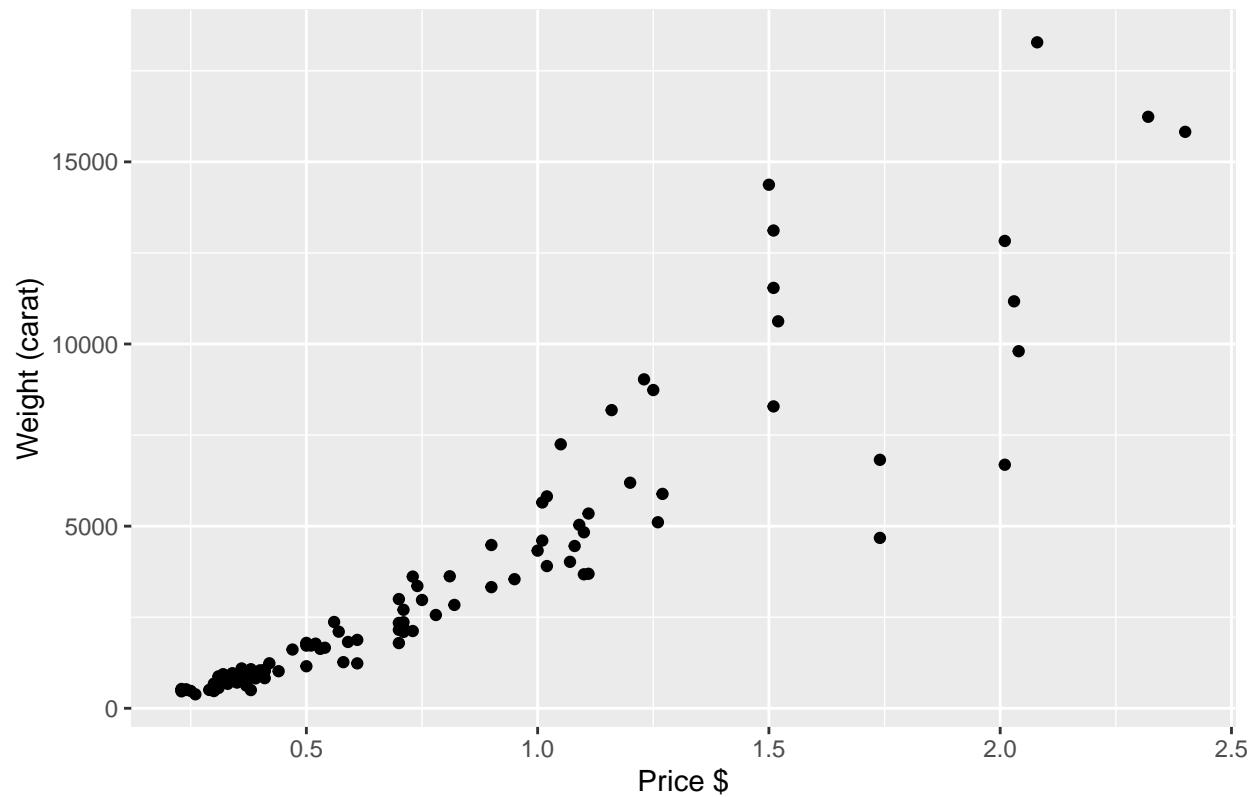
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



task 16: visualize the relationship between diamond carat weight (carat) and price (price) using a scatter plot, while also customizing the plot with meaningful axis labels and a descriptive title

```
qplot(carat,price,data = dsmall,xlab = "Price $", ylab = "Weight (carat)", main = "Price-Weight Relation")
```

Price–Weight Relationship



```
#Exercise 2
```

```
##Apply bar geoms to show the colour distribution of diamonds in dsmall dataset to see if you can find interesting relationships.
```

```
qplot(color,data = dsmall,geom = "bar",xlab="Diamond Color Grade",ylab = "Count",main = "Color Distribu
```

Color Distribution of Diamonds in dsmall

