

Data visualization

Bibek Sapkota

Data Visualization Lab ggplot2(qplot)

Task 1: Importing ggplot2 library

```
library(ggplot2)
```

Diamonds Dataset with qplot function

Task: loading the diamonds dataset

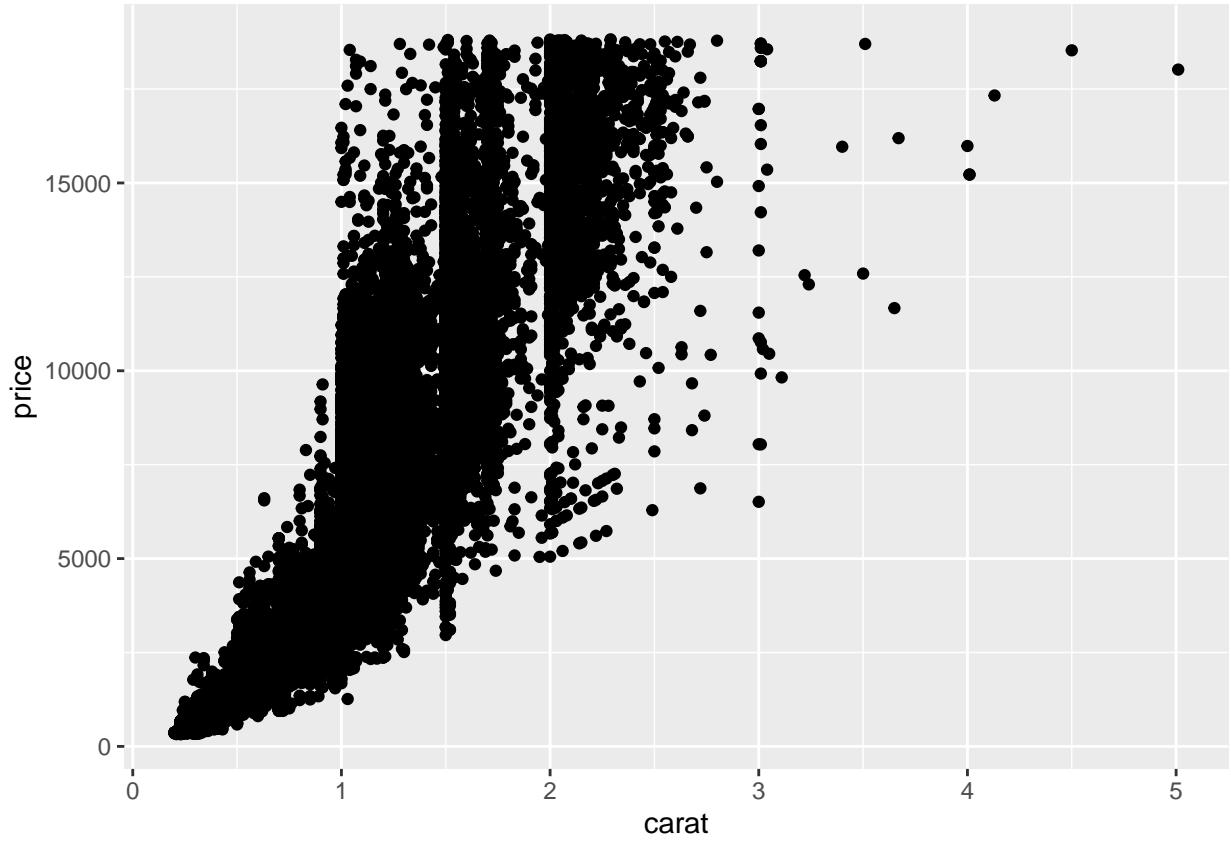
```
data("diamonds")
head(diamonds)
```

```
## # 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: Using qplot() to produce a scatterplot showing the relationship between the price and carats of a diamond

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

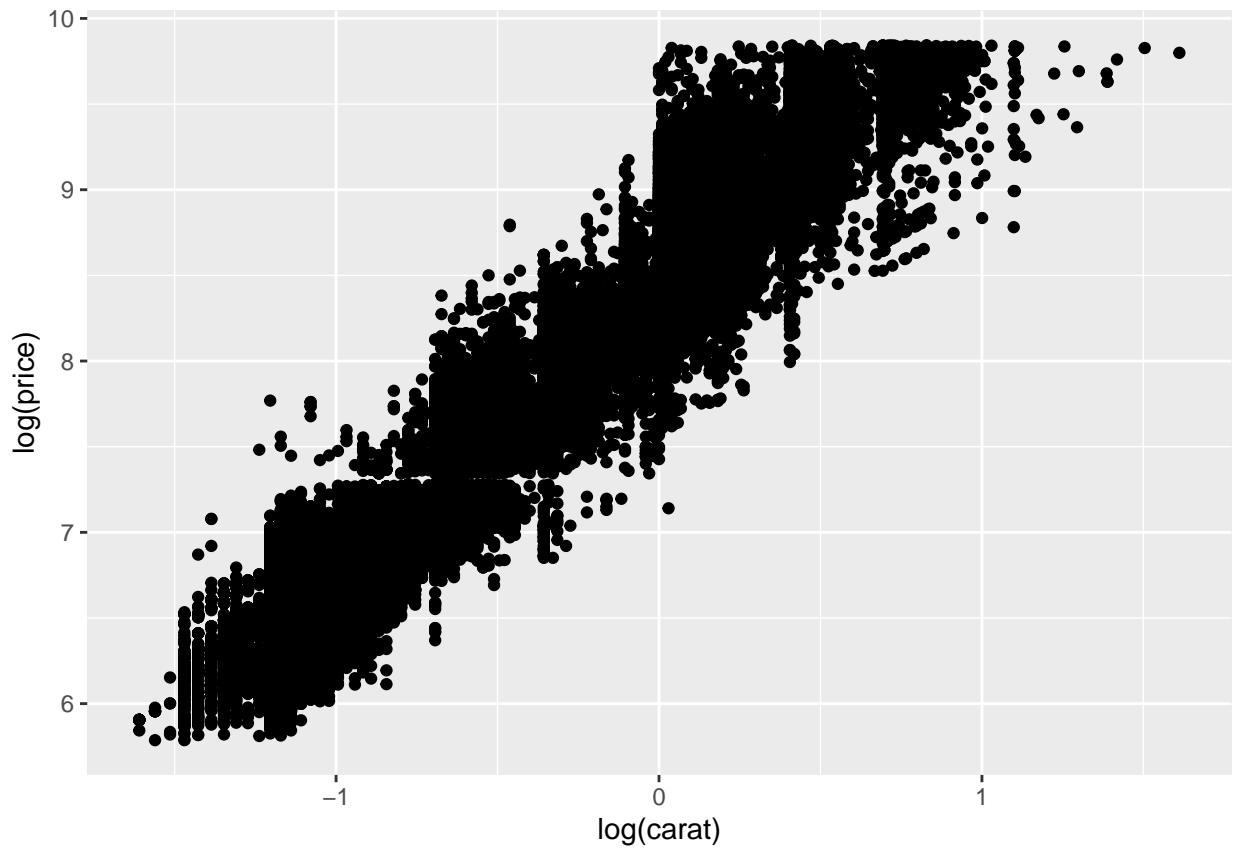
```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



```
## Variables Transformation
```

Task: using log function to transform the variables in a way to make the relationship between variables less sensitive to outliers.

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



Aesthetic attributes (color,shape,size)

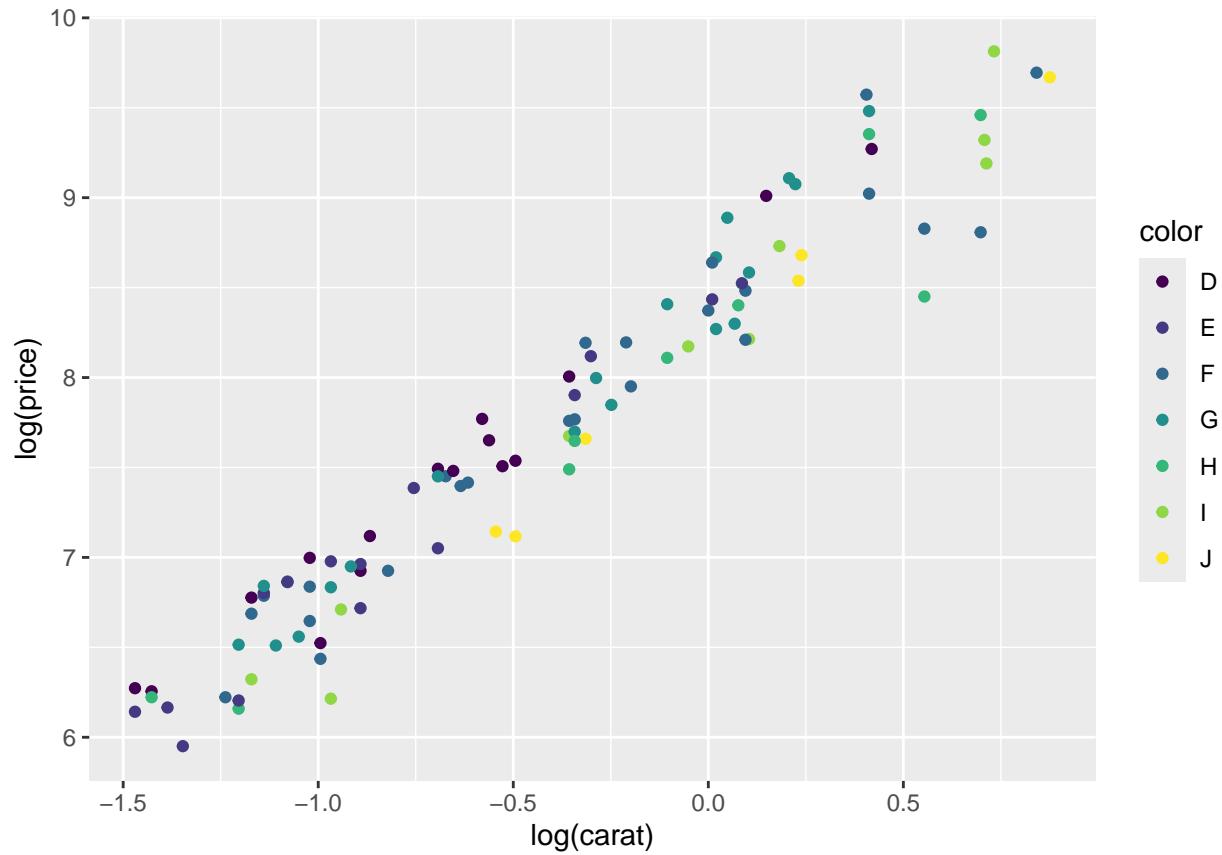
Task: creating a smaller dataset from diaminds by randomly sample.

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

Color aesthetic

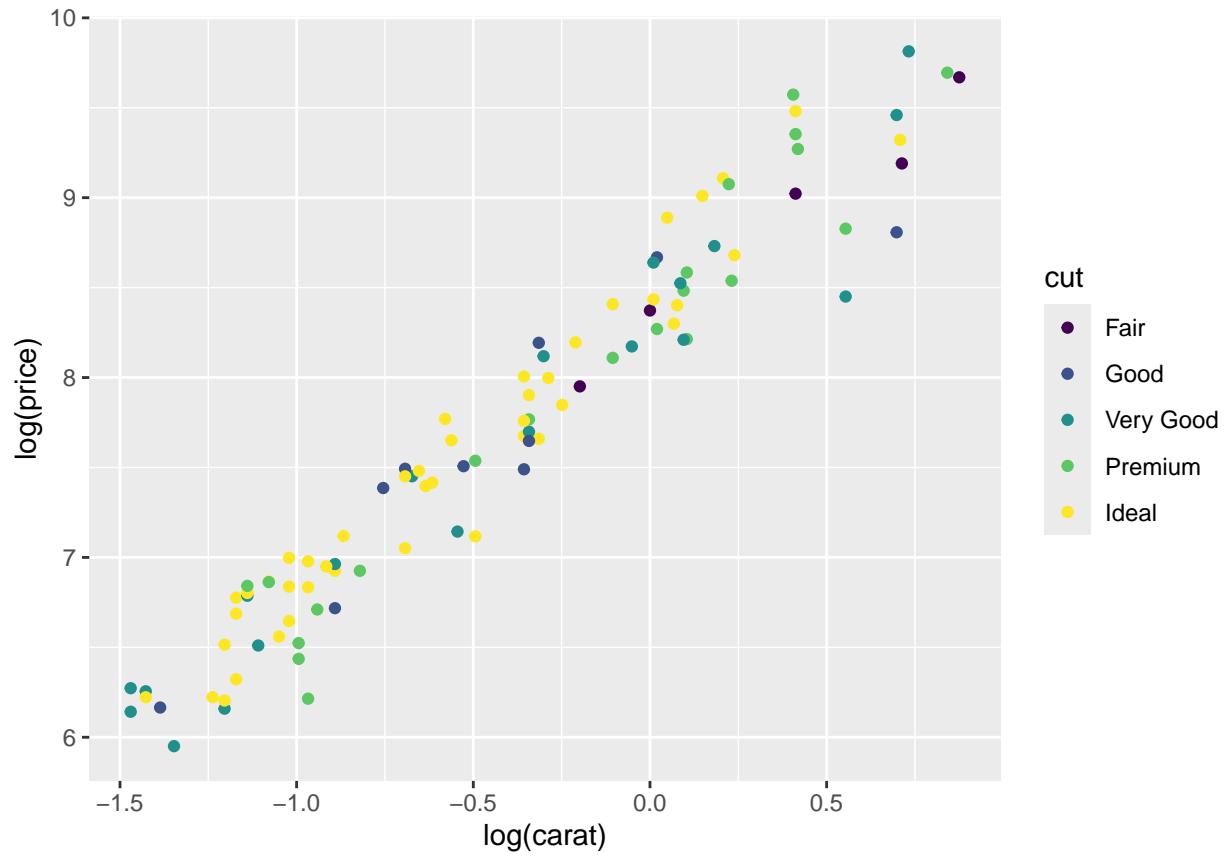
Task: Creating a scatterplot using qplot function to show the relationship between the 'log(carat)' and 'log(price)' of different colours(i.e. 'color' variable).

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



Task: create a scatter plot to visualize the relationship between ‘ $\log(\text{carat})$ ’ and ‘ $\log(\text{price})$ ’ of different cuts (i.e. ‘cut’ variable). This is by mapping ‘cut’ variable to colour aesthetic attribute.

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

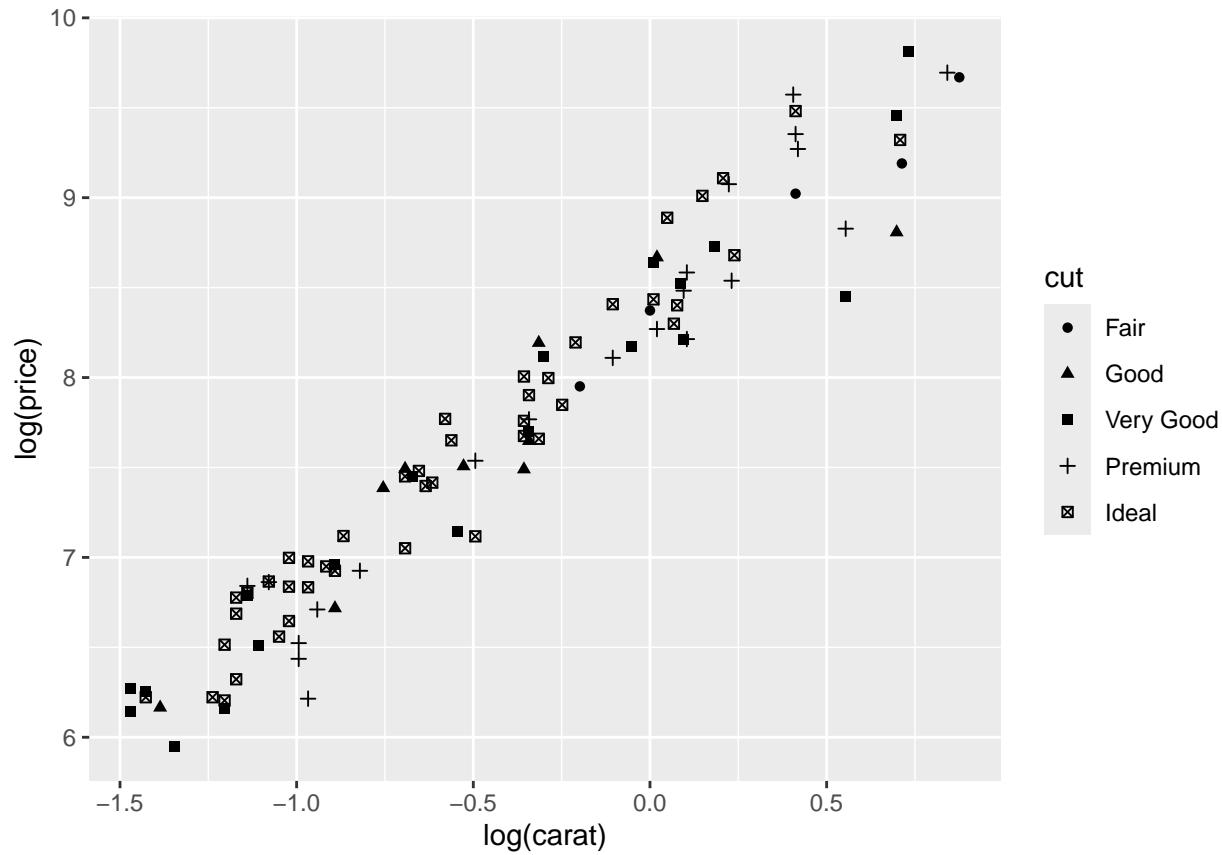


Shape aesthetic

Task:create a scatter plot to visualize the relationship between 'log(carat)' and 'log(price)' of different cuts(i.e.'cut' variable)

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

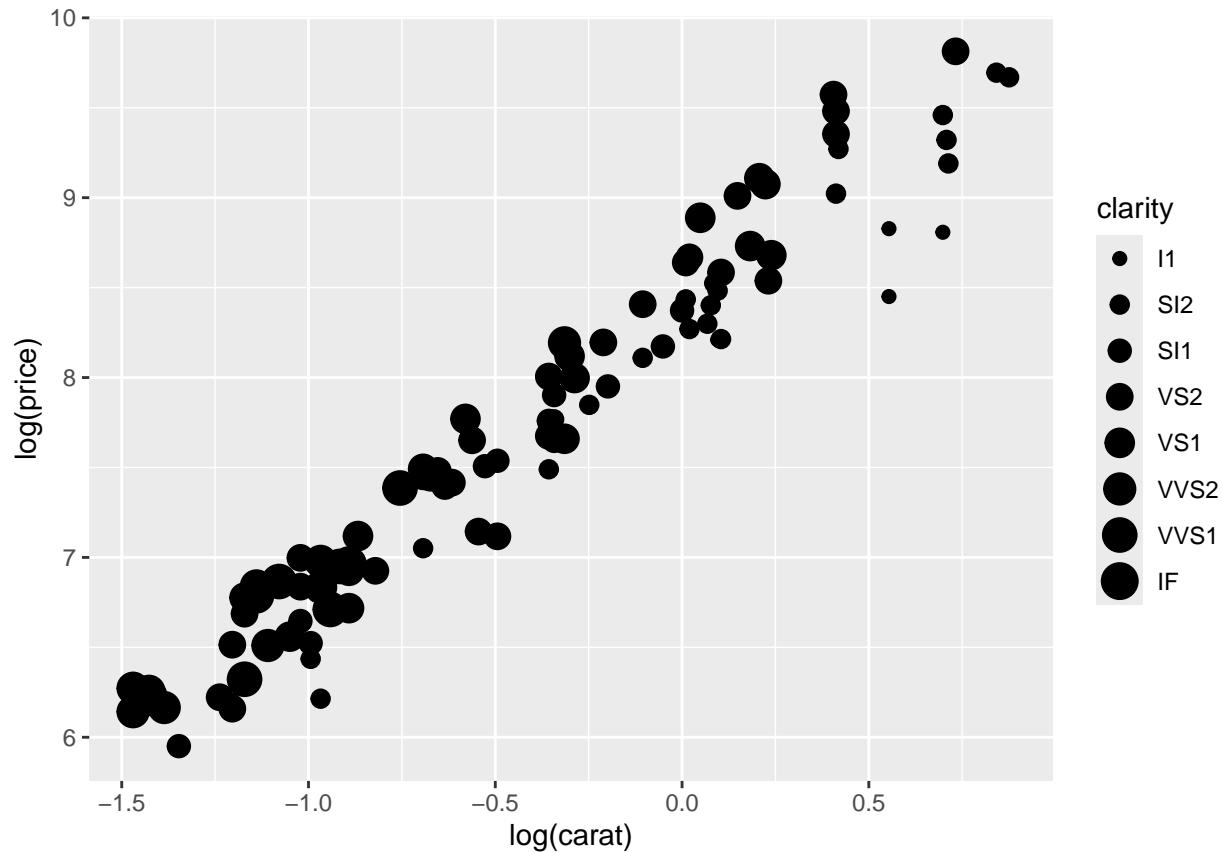
```
## Warning: Using shapes for an ordinal variable is not advised
```



Exercise 1

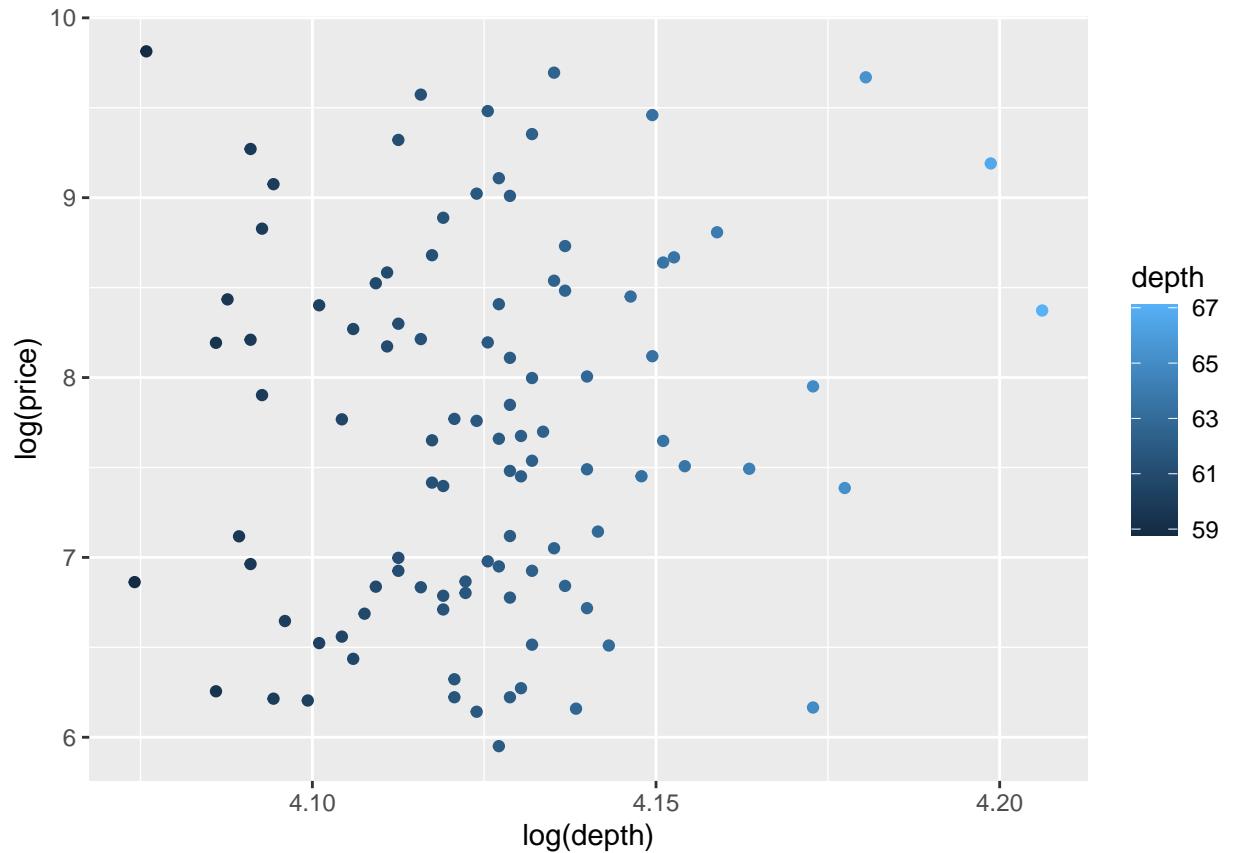
- a. Apply what you have done previously to dsmall dataset with other aesthetic attribute such as 'size'.

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

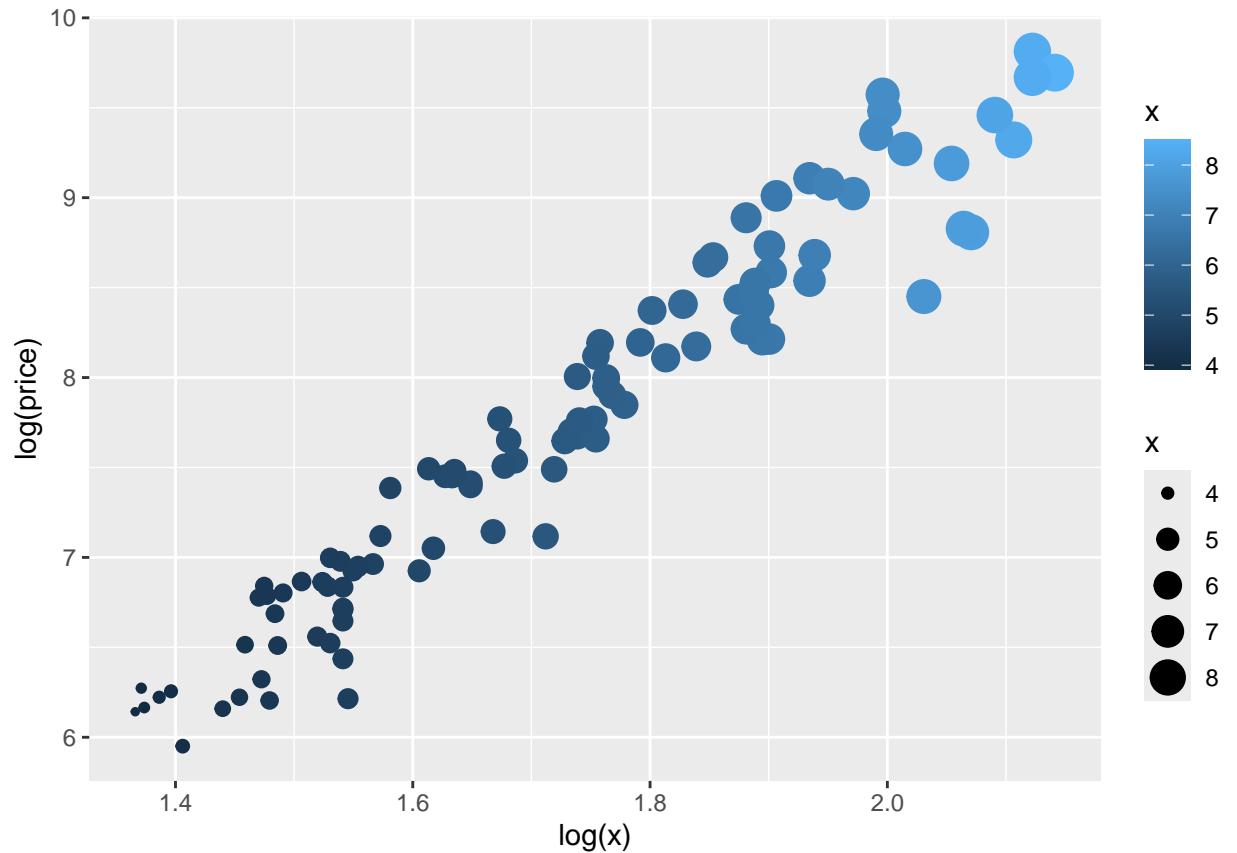


b. Investigate the relationship between other variables (using colour, shape and size aesthetics) to see if you can find interesting information.

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



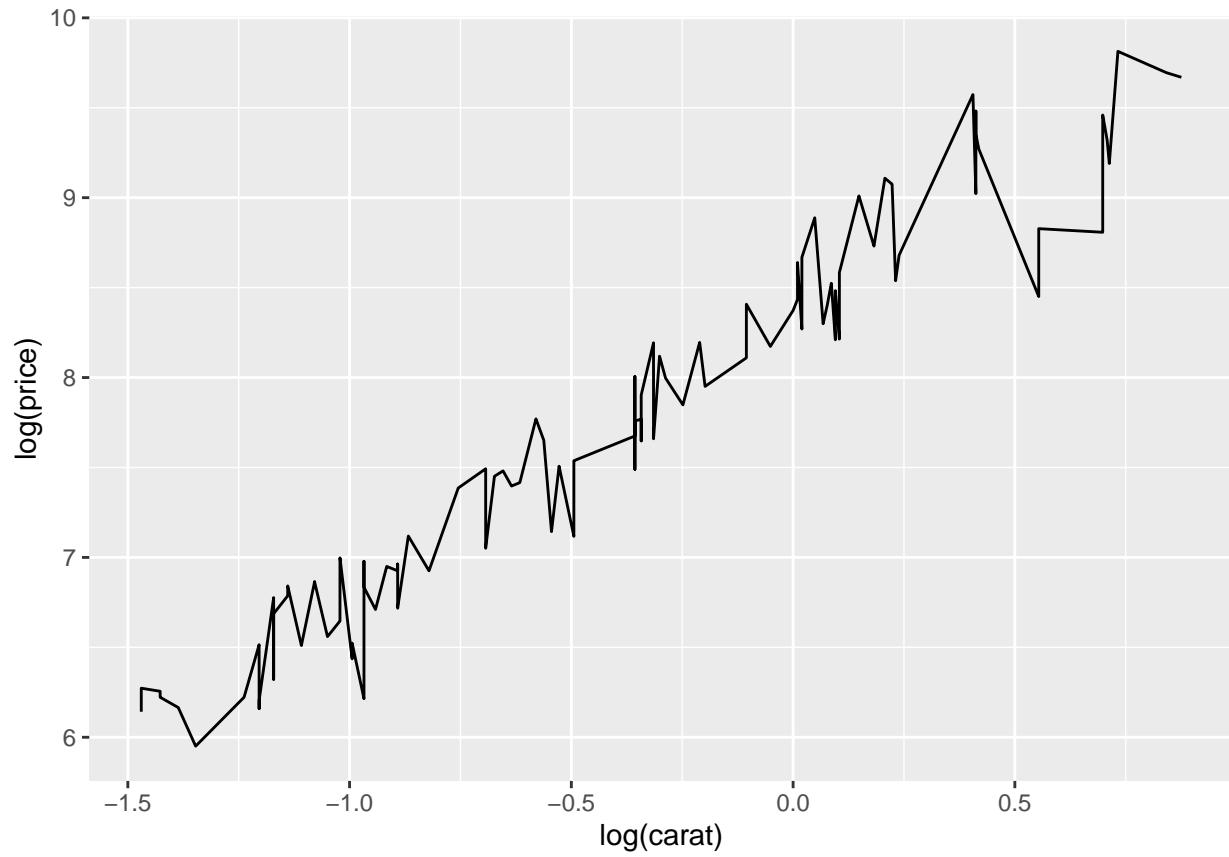
```
qplot(log(x),log(price),data = dsmall,size=x, colour=x)
```



Geom 2D

Task: Visualizing the relationship between the natural log of carat and price using a line plot.

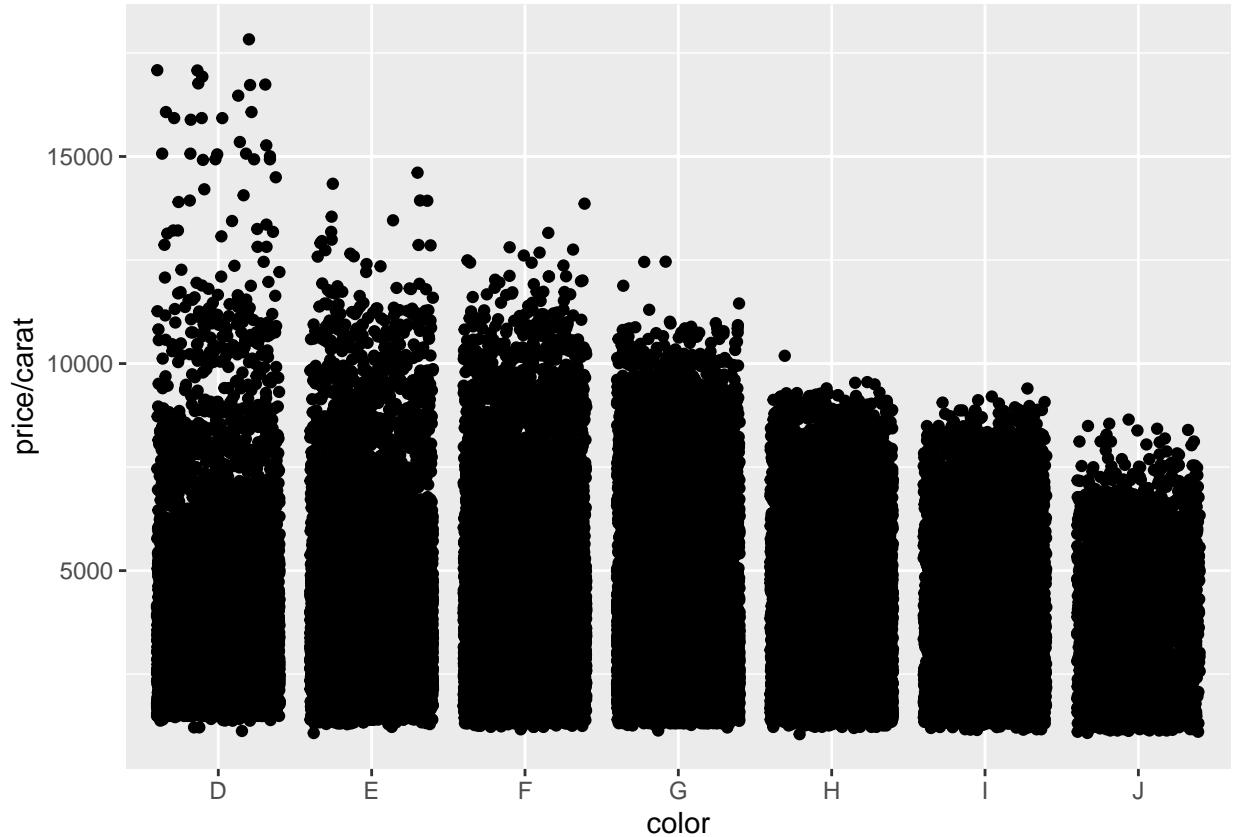
```
qplot(log(carat),log(price), data = dsmall,geom="line")
```



Boxplots ans Jitttered

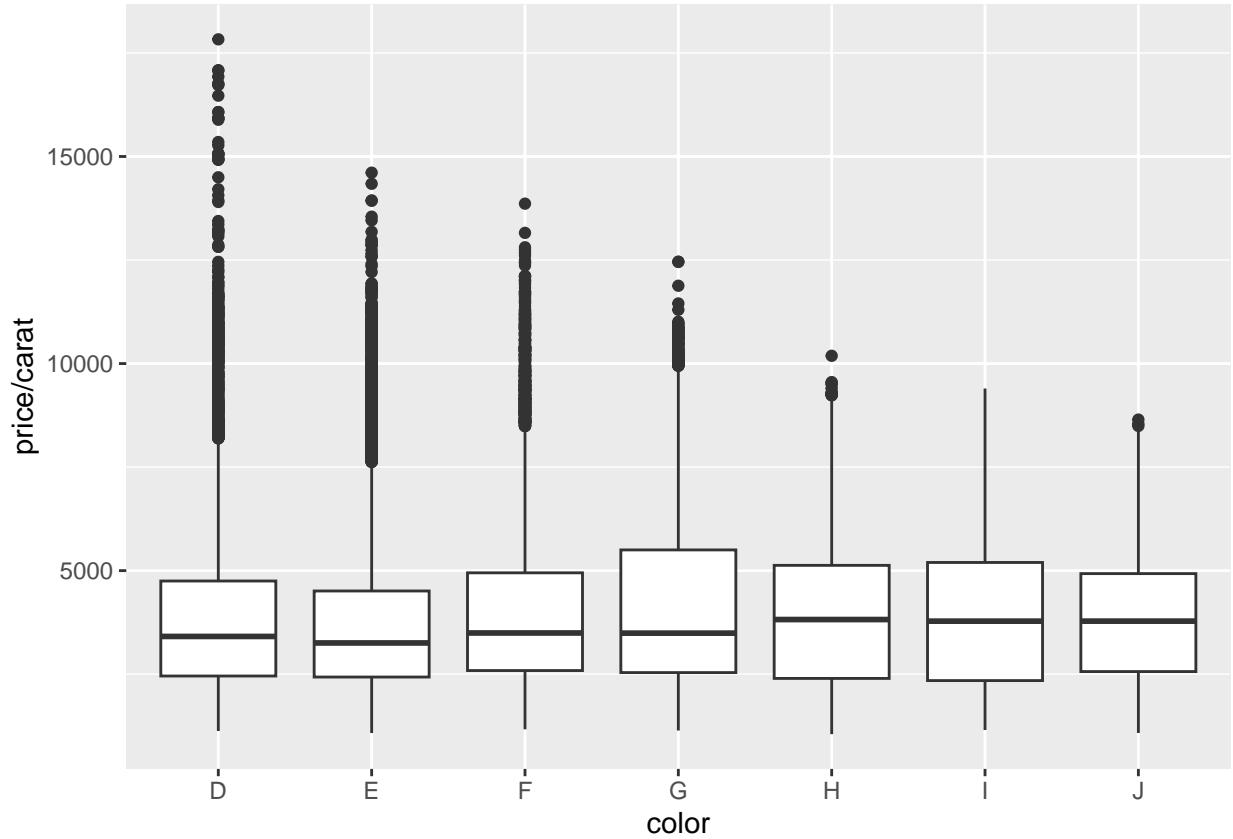
Task: Visualizing how the price per carat (price/carat) varies across different color grades (color) of diamonds using a jitter plot.

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



Task: Visualizing how the price per carat (price/carat) varies across different color grades (color) of diamonds using a boxplot

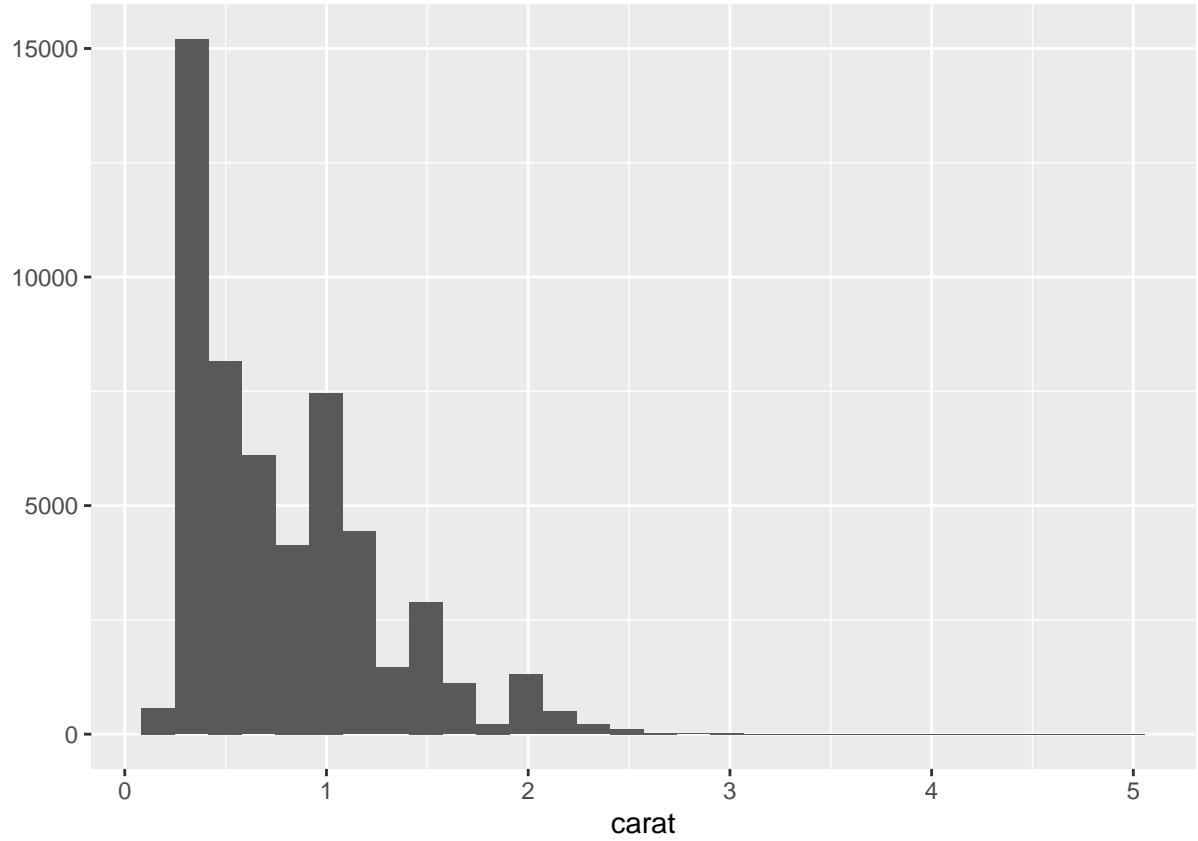
```
qplot(color,price/carat,data = diamonds,geom = "boxplot")
```



Geom 1D

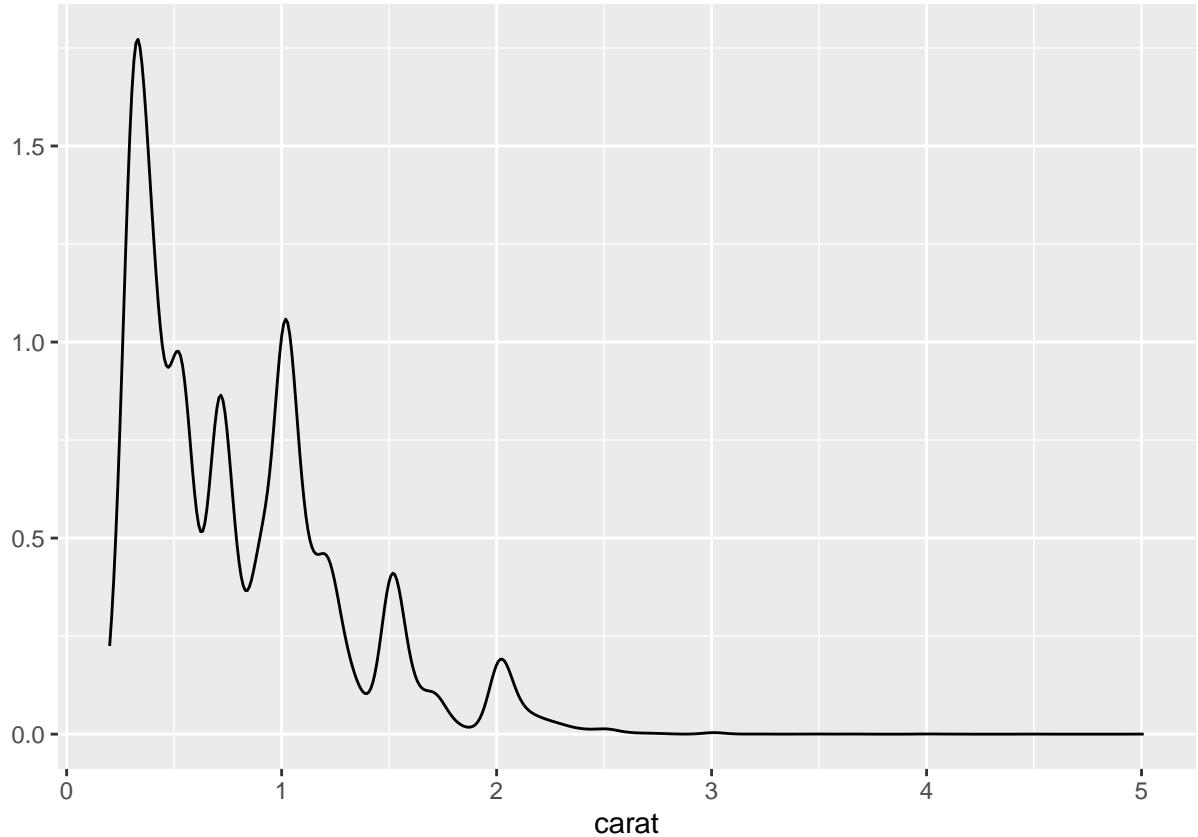
Task: Visualizing the distribution of diamond carat using a histogram

```
qplot(carat,data = diamonds,geom = "histogram")  
  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



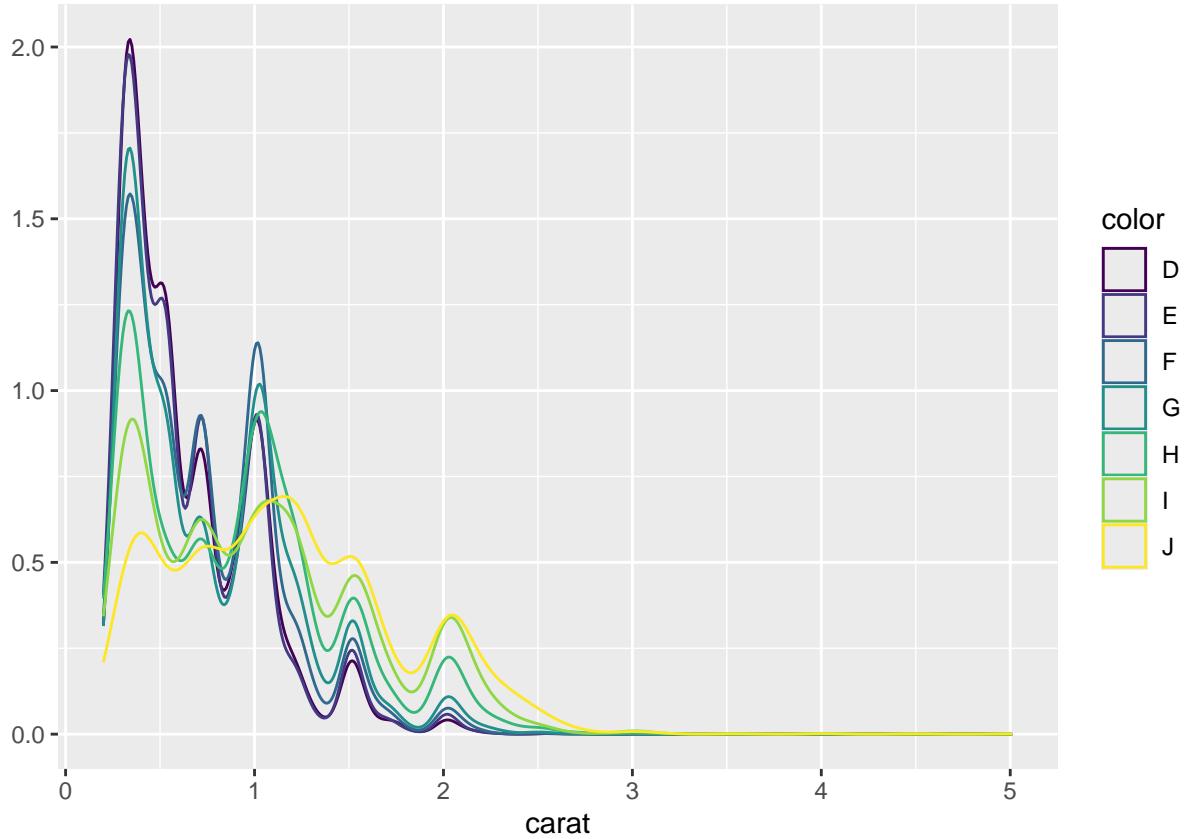
Task: Visualizing the distribution of diamond carat using a density plot

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



Task: Visualizing how the distribution of diamond carat varies across different color using a density plot

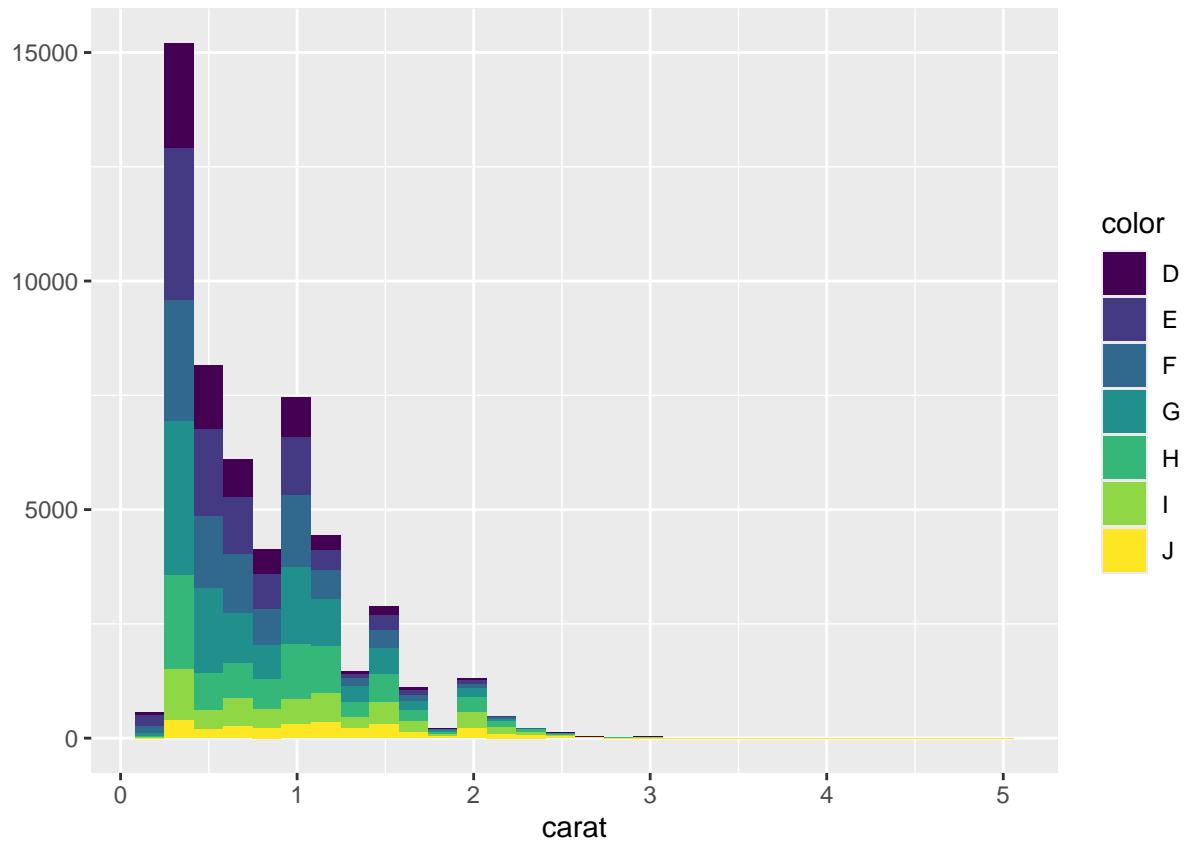
```
qplot(carat,data = diamonds,geom = "density",colour=color)
```



Task: Visualizing the distribution of diamond carat varies across different color using a filled histogram

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

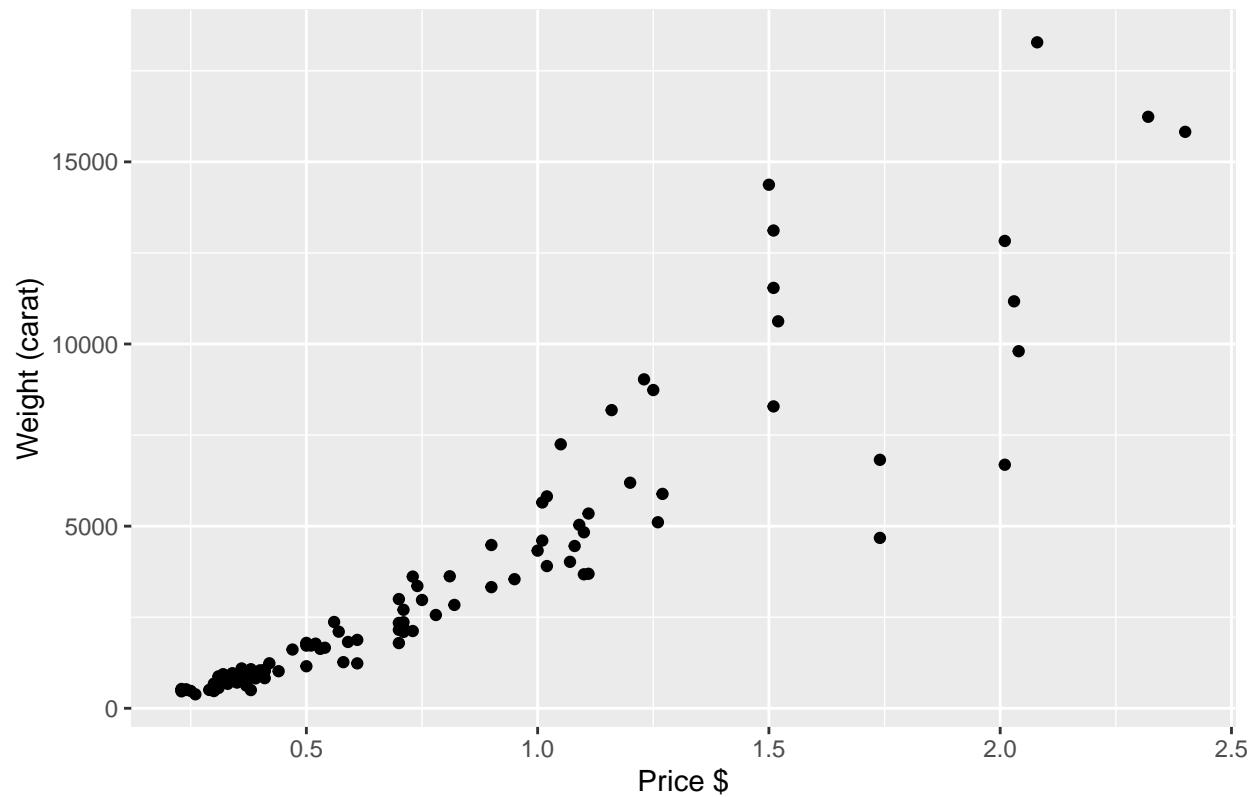
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Task: Visualizing the relationship between diamond carat and price using a scatter plot,

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

Price–Weight Relationship



Exercise 2

a. 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

