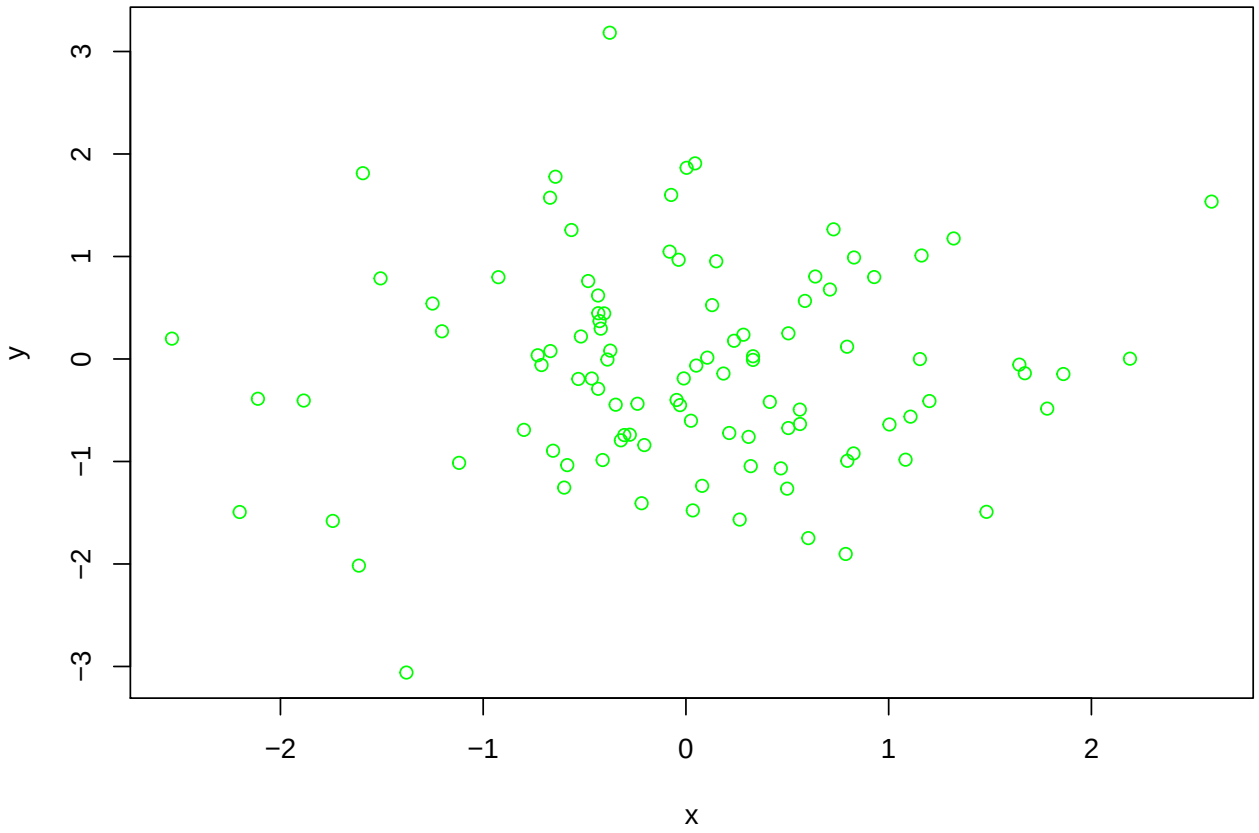
Data Analysis whith R

Henri Makika

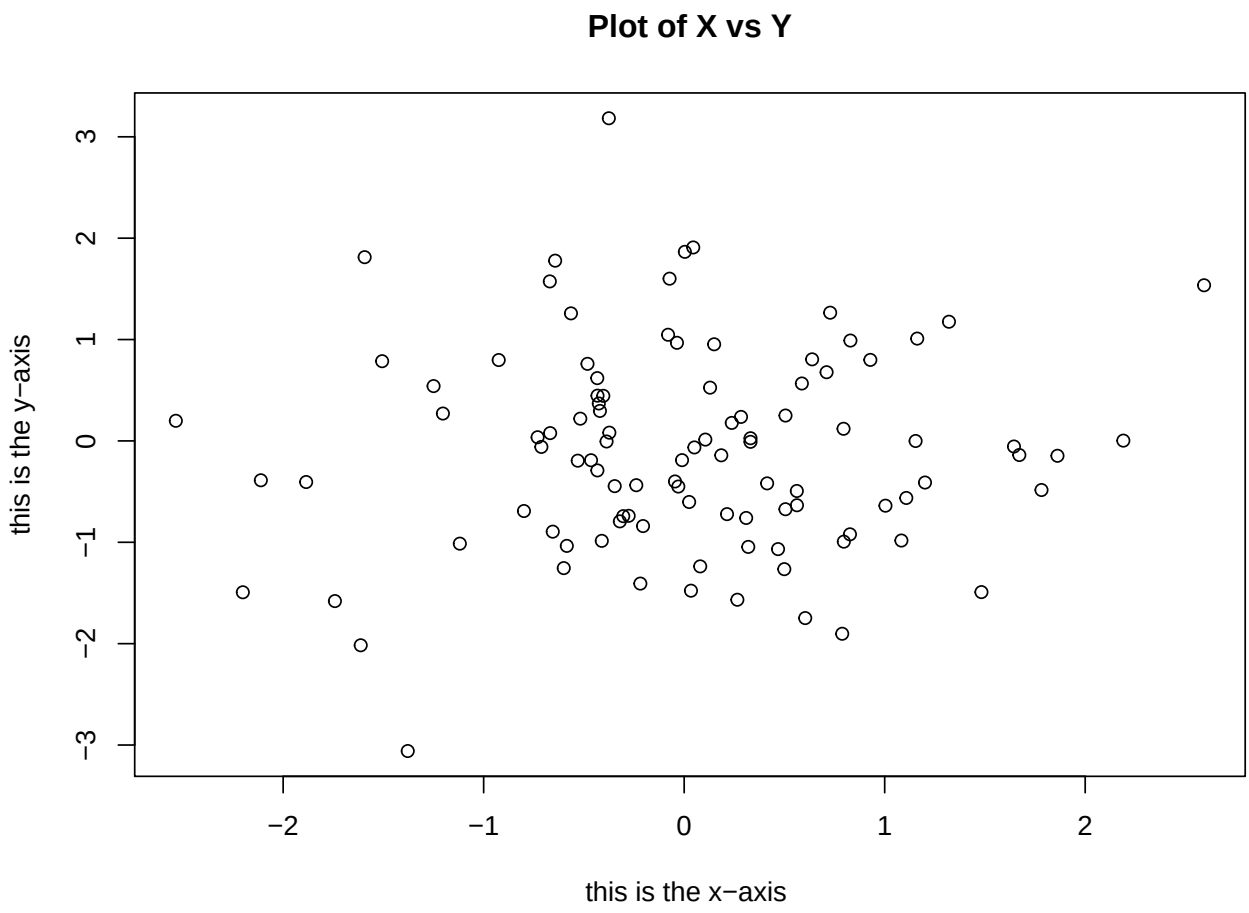
7/27/2019

Exemples plotting

```
x = rnorm (100)
y = rnorm (100)
plot(x, y, col = "green")
```




```
plot(x, y, xlab = "this is the x-axis " , ylab = "this is the y-axis" ,
      main = "Plot of X vs Y")
```



Loading Data

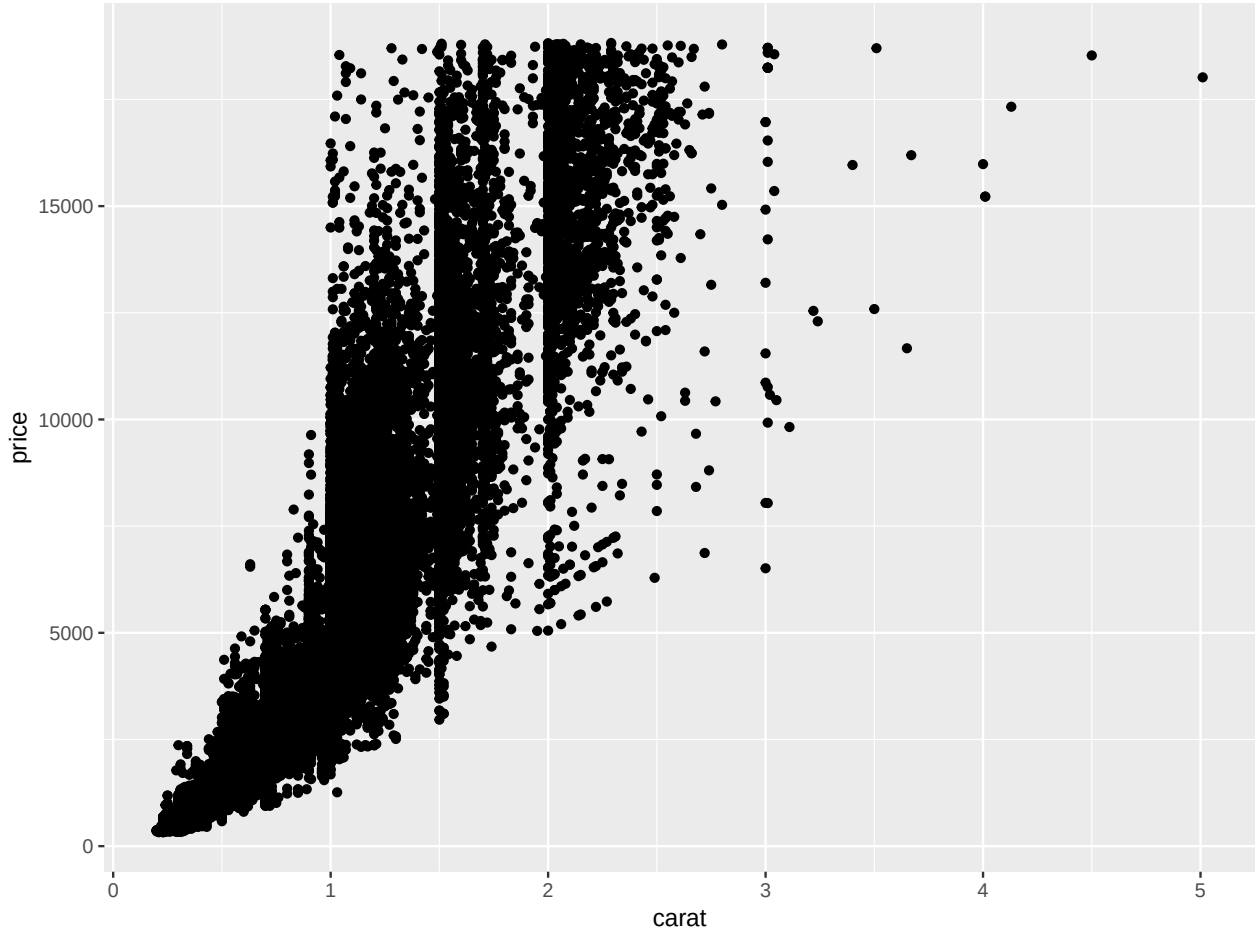
```
library(ggplot2) # Pour le graphique
library(knitr) # Para Kable
library(dplyr) # Pipe operator

data(diamonds)
summary(diamonds[,1:5]) %>% kable(digits = 0)
```

carat	cut	color	clarity	depth
Min. :0.2000	Fair : 1610	D: 6775	SI1 :13065	Min. :43.00
1st Qu.:0.4000	Good : 4906	E: 9797	VS2 :12258	1st Qu.:61.00
Median :0.7000	Very Good:12082	F: 9542	SI2 : 9194	Median :61.80
Mean :0.7979	Premium :13791	G:11292	VS1 : 8171	Mean :61.75
3rd Qu.:1.0400	Ideal :21551	H: 8304	VVS2 : 5066	3rd Qu.:62.50
Max. :5.0100	NA	I: 5422	VVS1 : 3655	Max. :79.00
NA	NA	J: 2808	(Other): 2531	NA

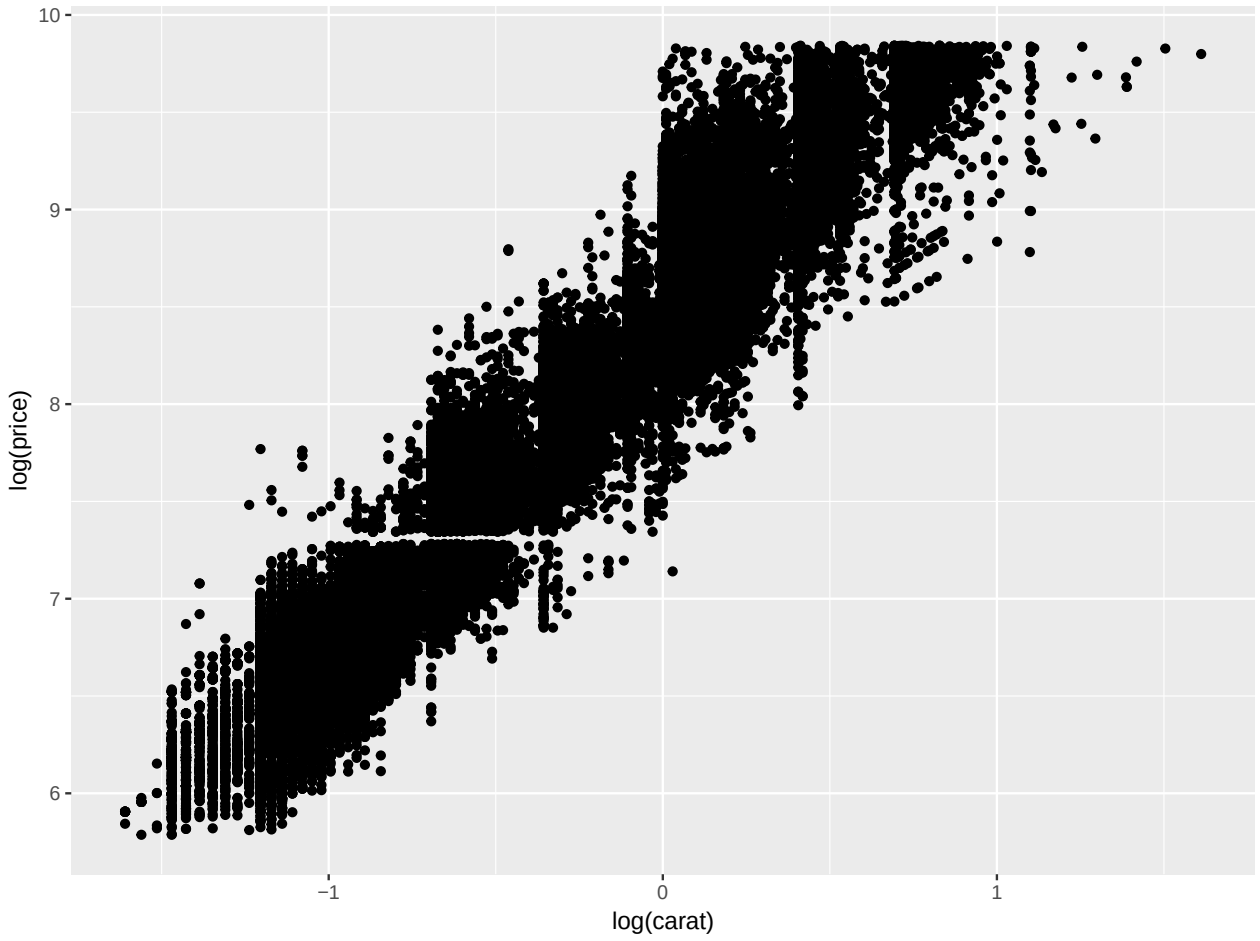
```
data(diamonds)
summary(diamonds[,6:10]) %>% kable(digits = 0)
```

table	price	x	y	z
Min. :43.00	Min. : 326	Min. : 0.000	Min. : 0.000	Min. : 0.000
1st Qu.:56.00	1st Qu.: 950	1st Qu.: 4.710	1st Qu.: 4.720	1st Qu.: 2.910
Median :57.00	Median : 2401	Median : 5.700	Median : 5.710	Median : 3.530
Mean :57.46	Mean : 3933	Mean : 5.731	Mean : 5.735	Mean : 3.539
3rd Qu.:59.00	3rd Qu.: 5324	3rd Qu.: 6.540	3rd Qu.: 6.540	3rd Qu.: 4.040
Max. :95.00	Max. :18823	Max. :10.740	Max. :58.900	Max. :31.800

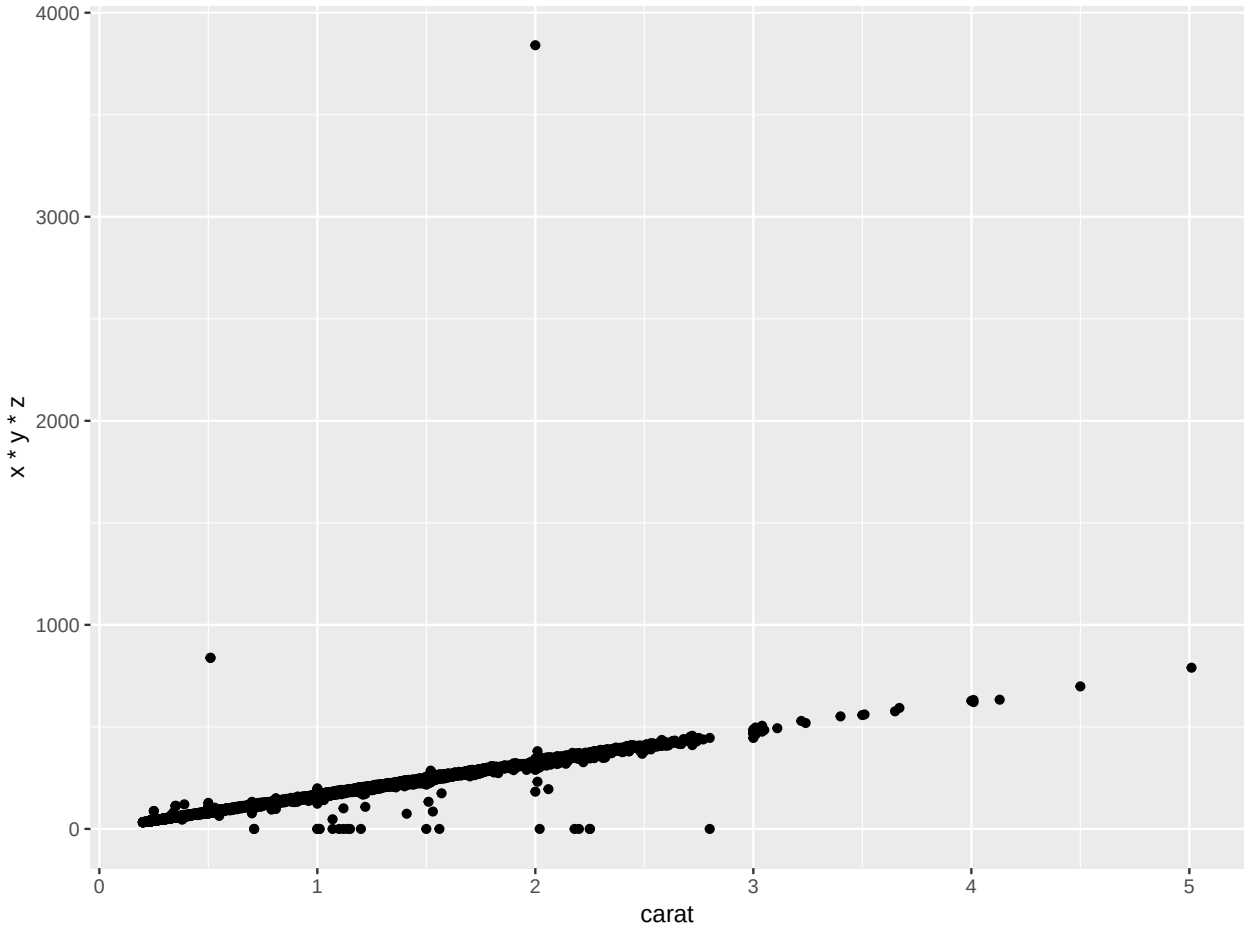
```
qplot(carat, price, data = diamonds)
```




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

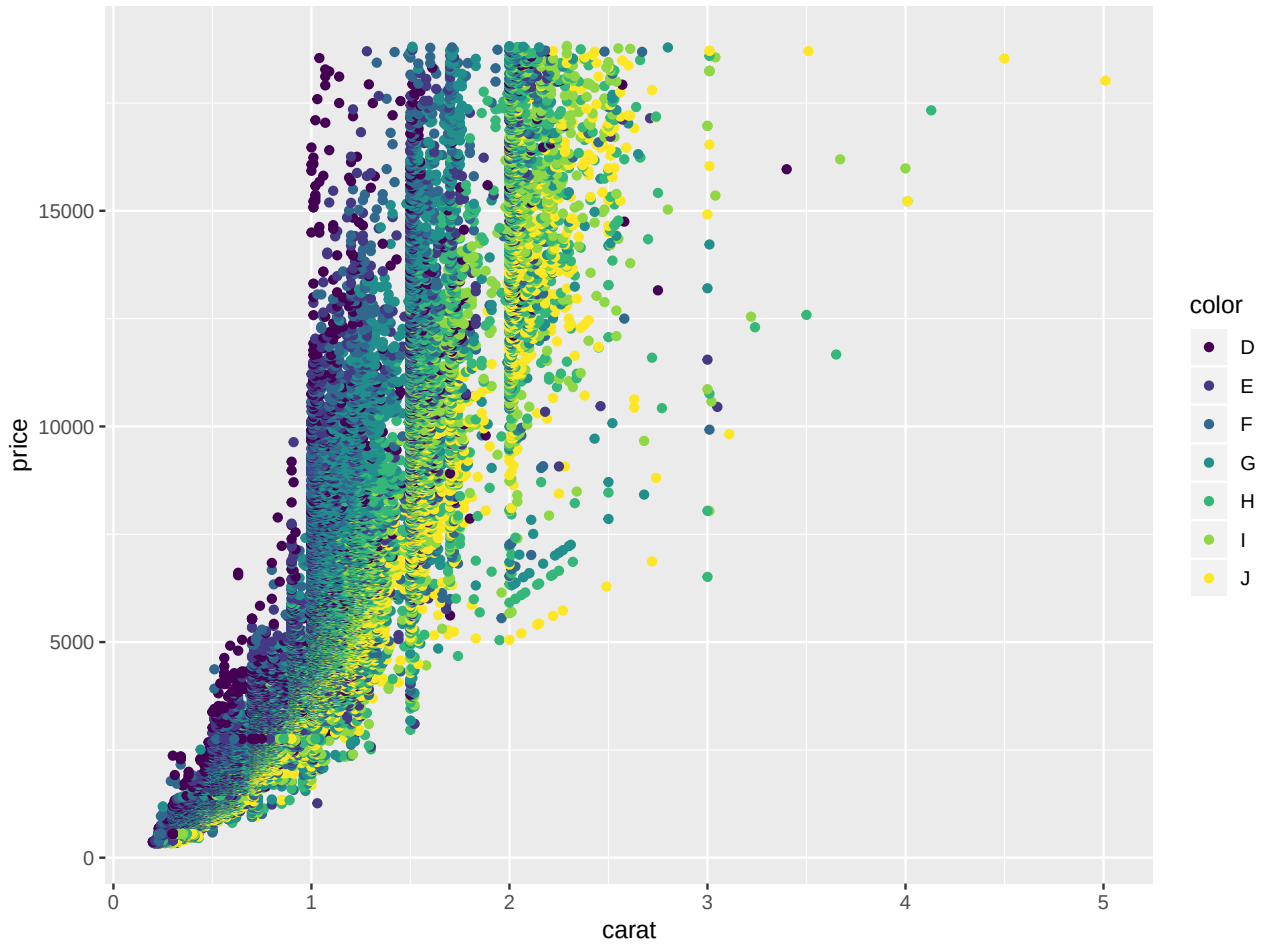



```
qplot(carat, x * y * z, data = diamonds)
```



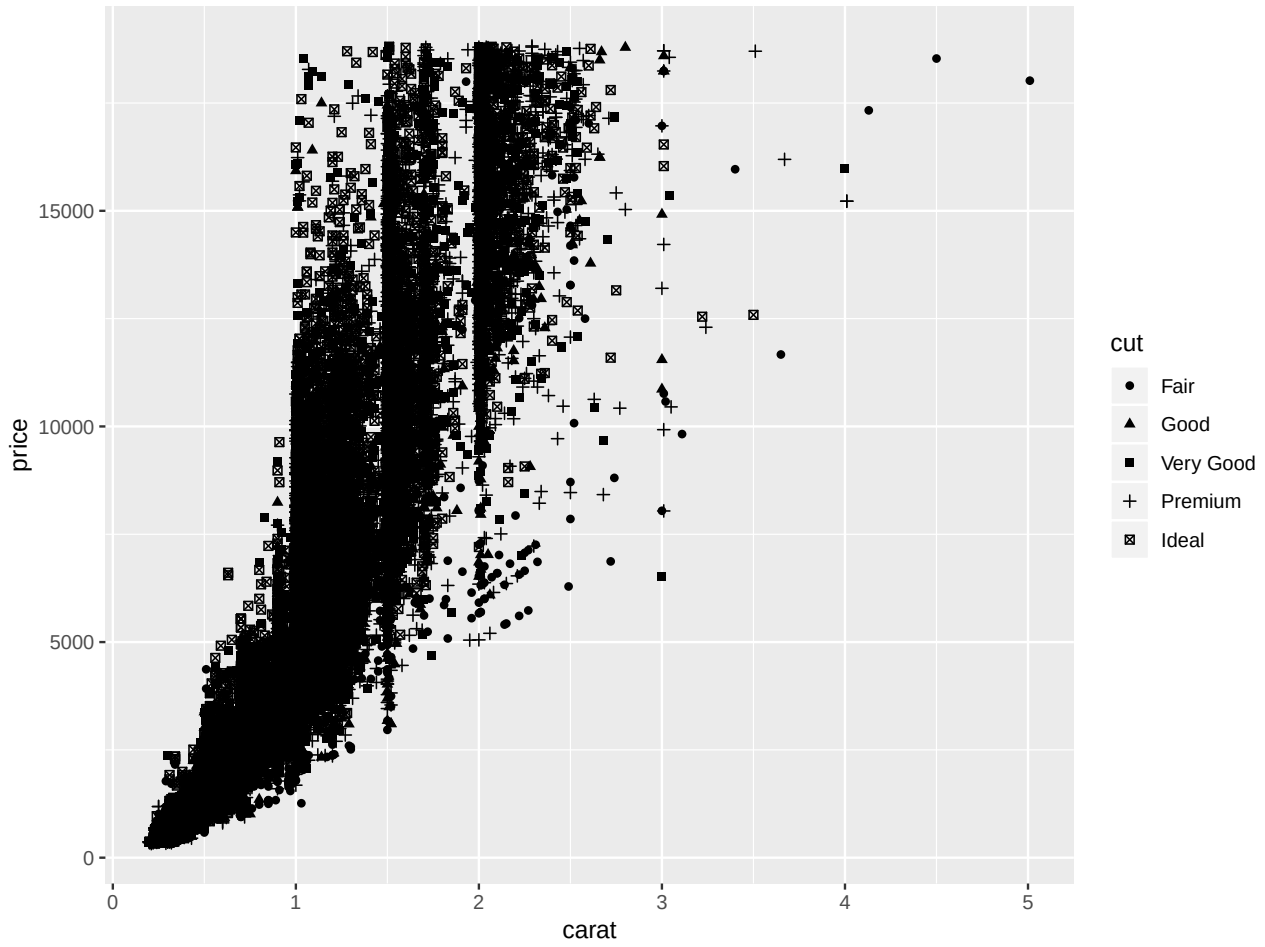
Colour, size, shape and other aesthetic attributes

```
par(mfrow = c(2,1))

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

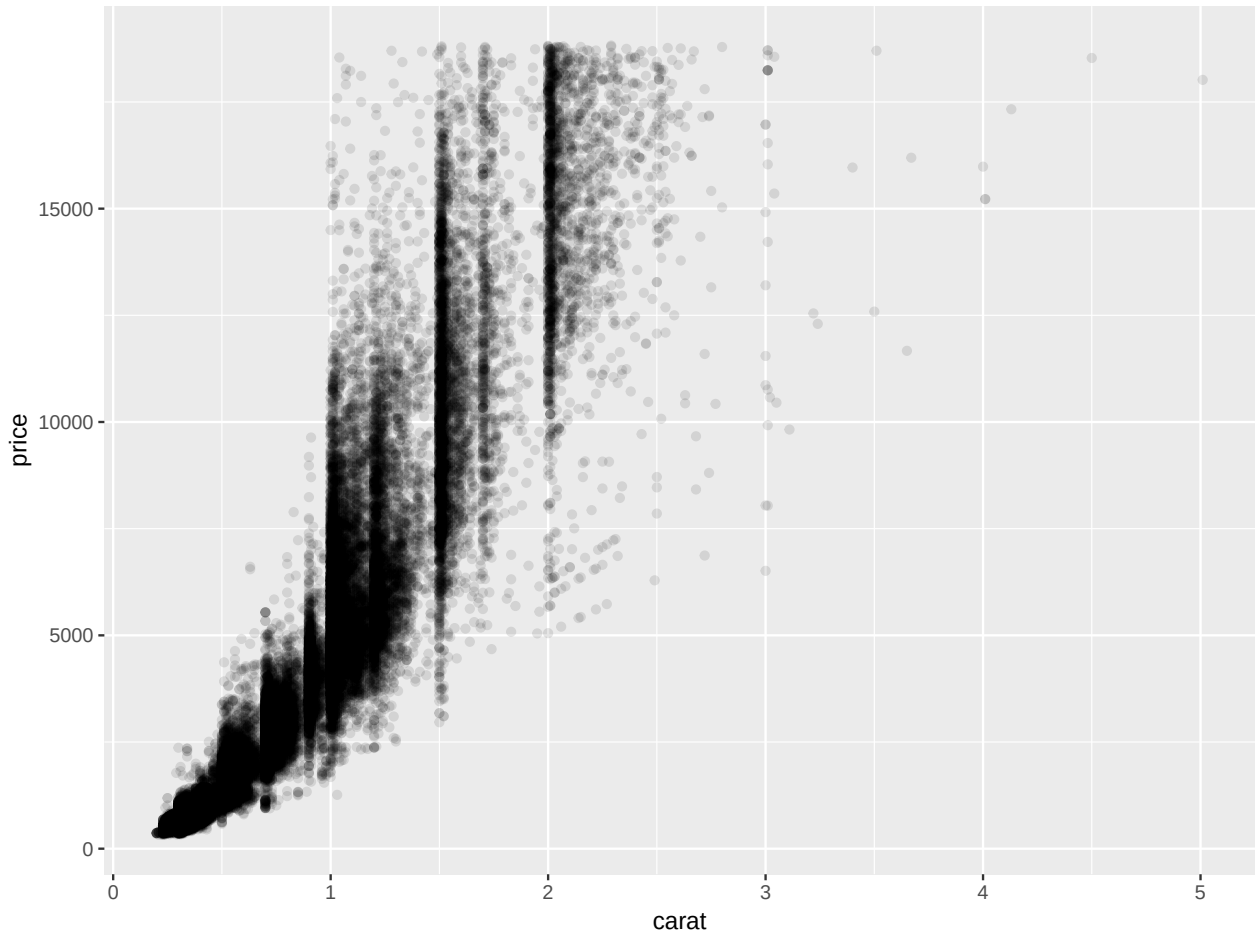



```
qplot(carat, price, data = diamonds, shape = cut)
```

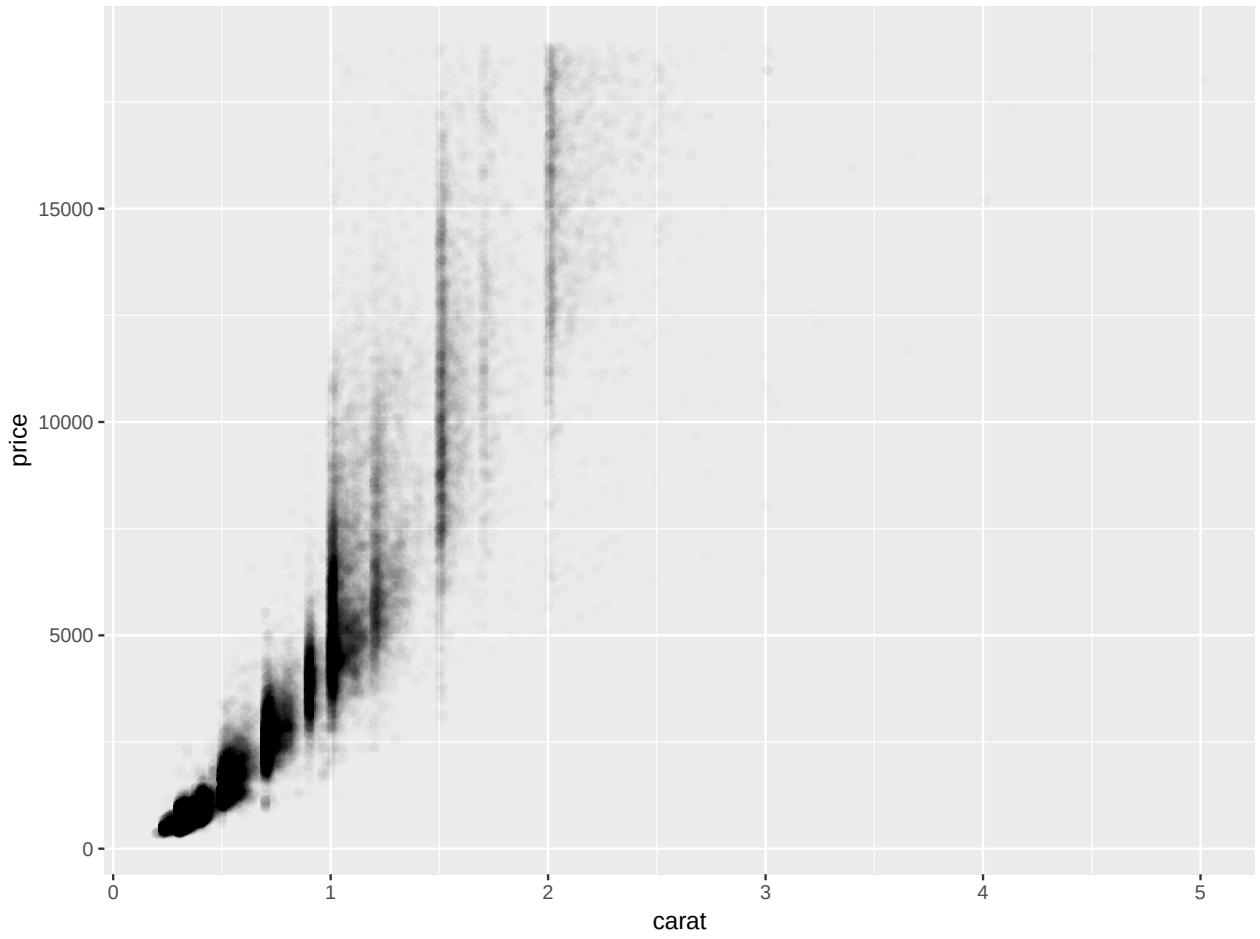
```
## Warning: Using shapes for an ordinal variable is not advised
```



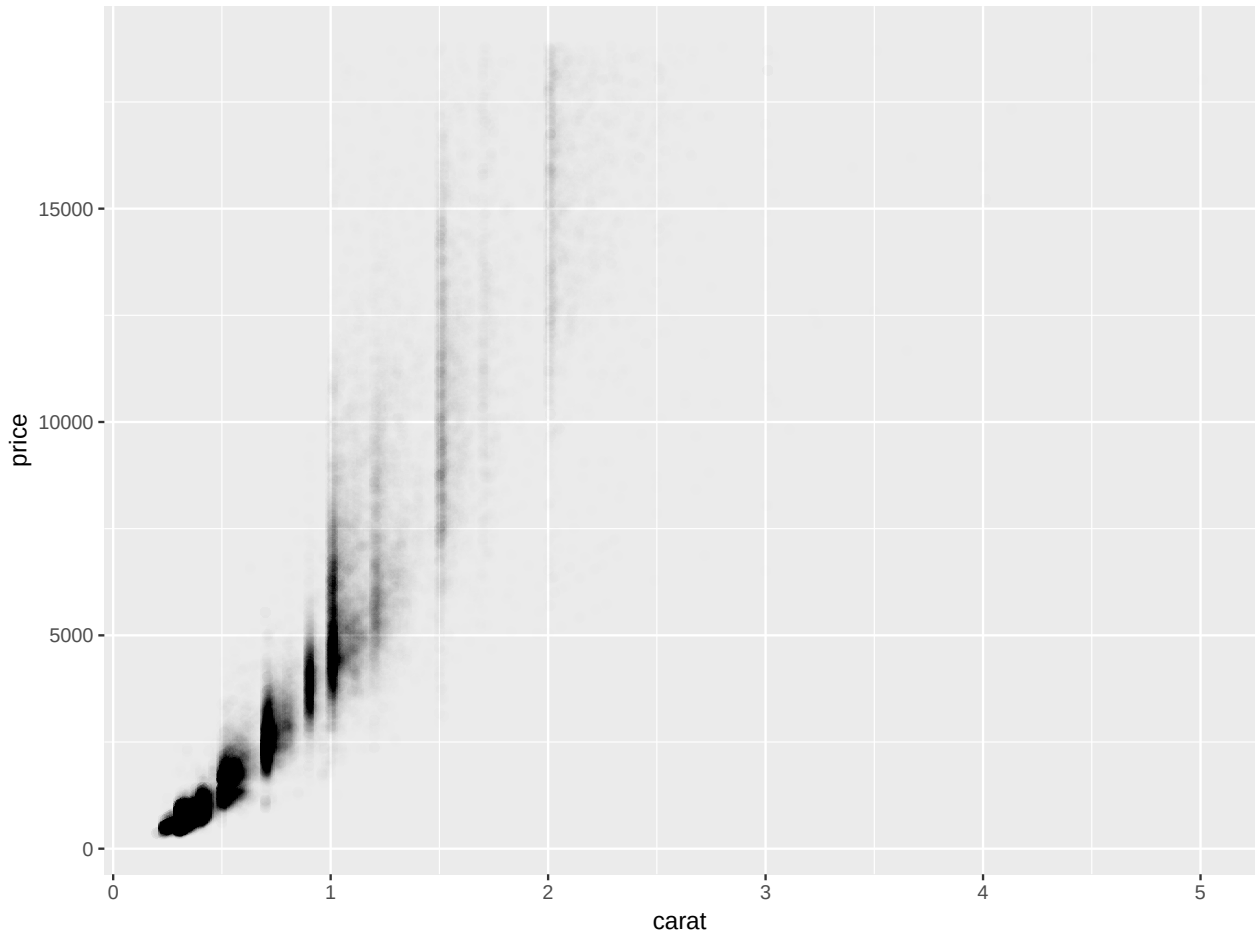

```
qplot(carat, price, data = diamonds, alpha = I(1/10))
```




```
qplot(carat, price, data = diamonds, alpha = I(1/100))
```

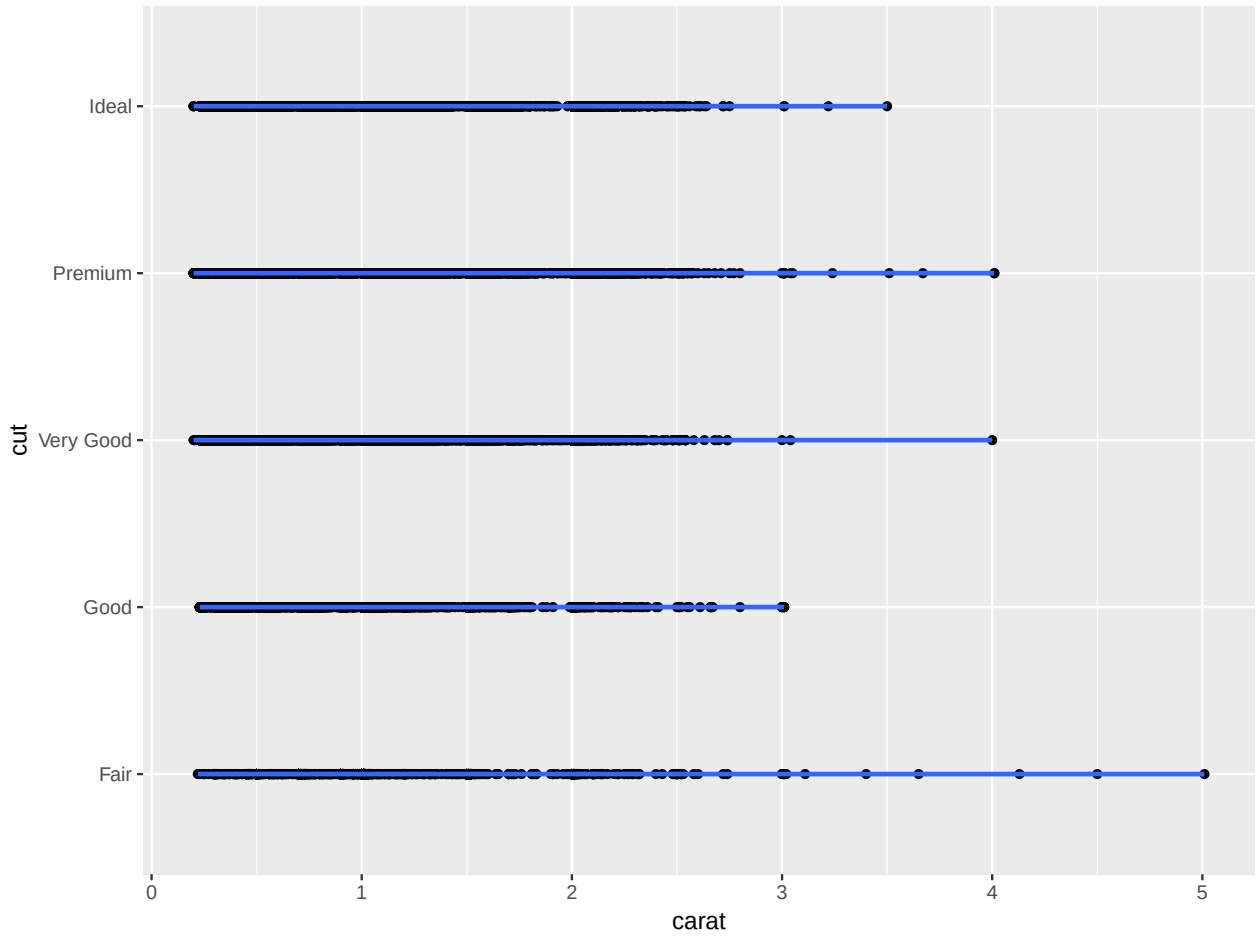



```
qplot(carat, price, data = diamonds, alpha = I(1/200))
```

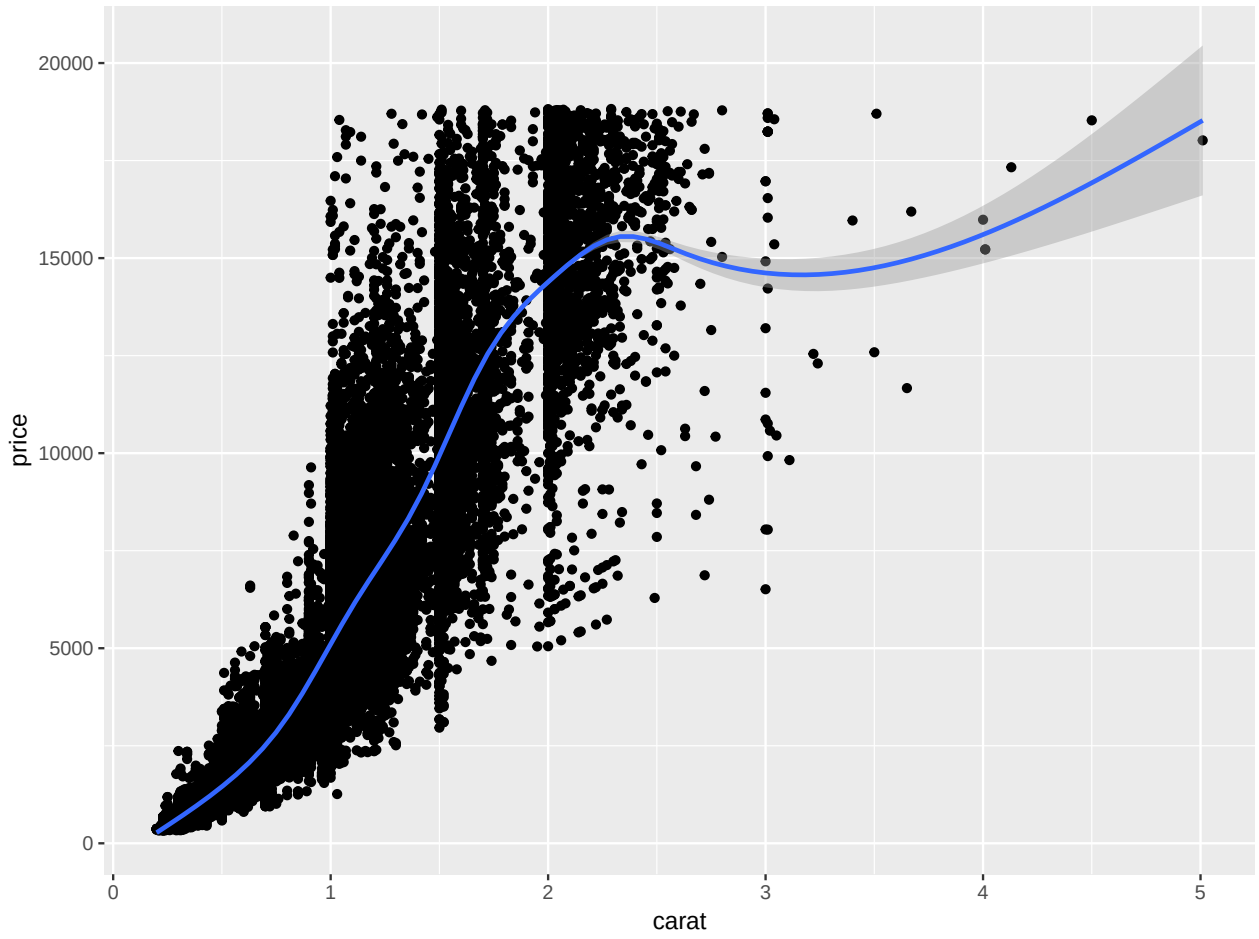


Plot geoms

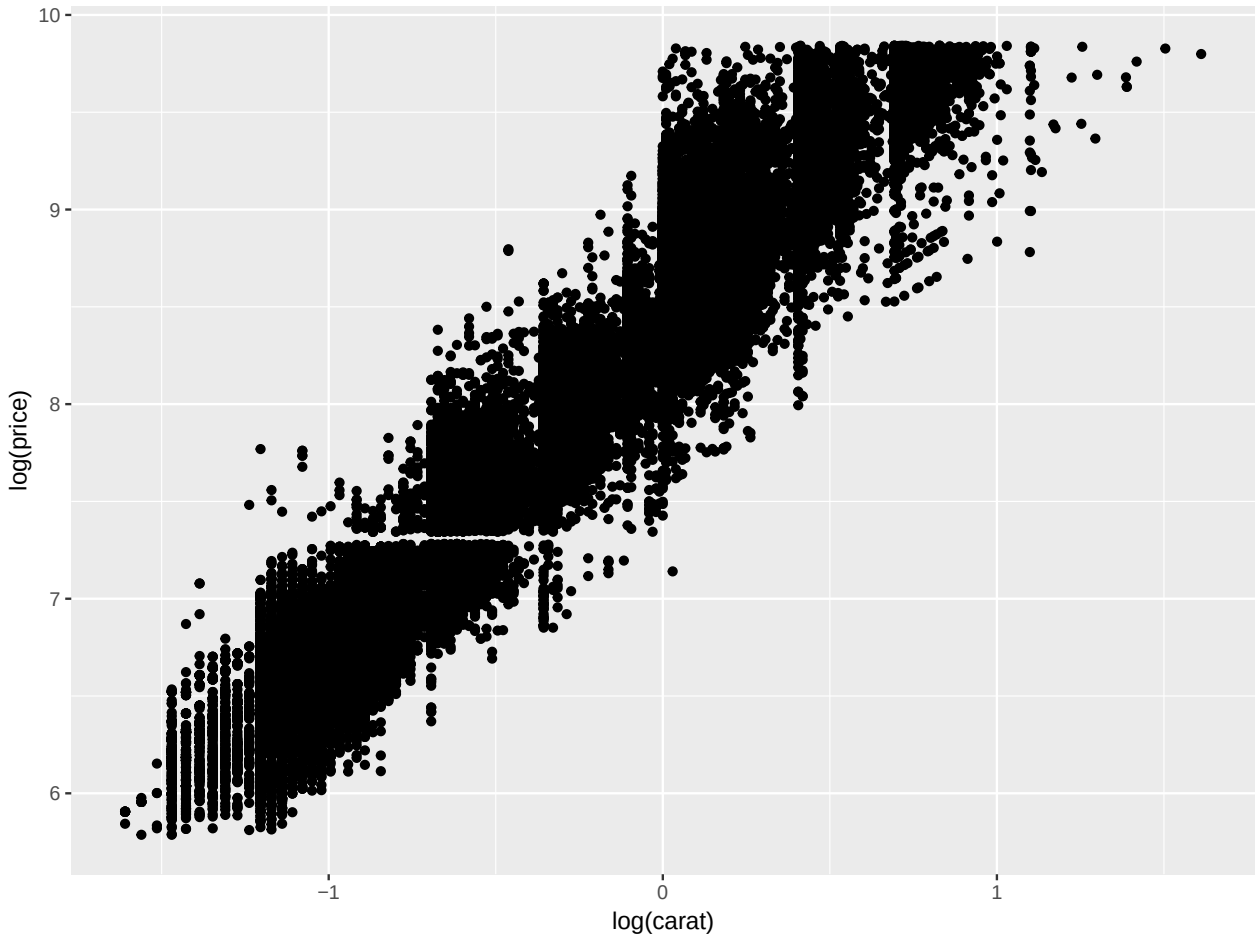
```
# Adding a smoother to a plot

qplot(carat, cut, data = diamonds, geom = c("point", "smooth"))
```



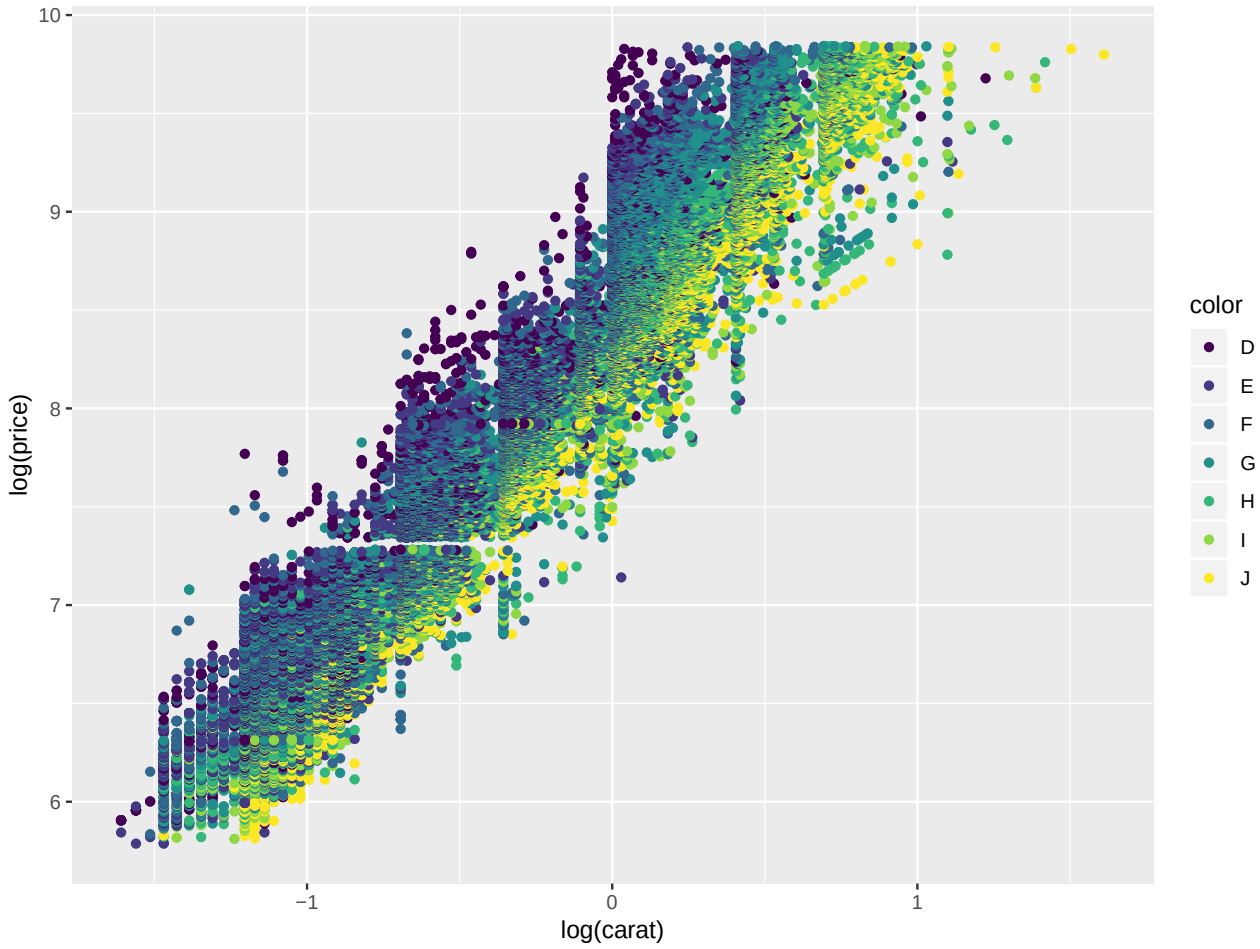

```
qplot(carat, price, data = diamonds, geom = c("point", "smooth"))
```




```
set.seed(1234)  # Fixed seed
dsmall <- diamonds[sample(nrow(diamonds), 200), ]

qplot(log(carat), log(price), data = diamonds)  # Scatter Plot
```

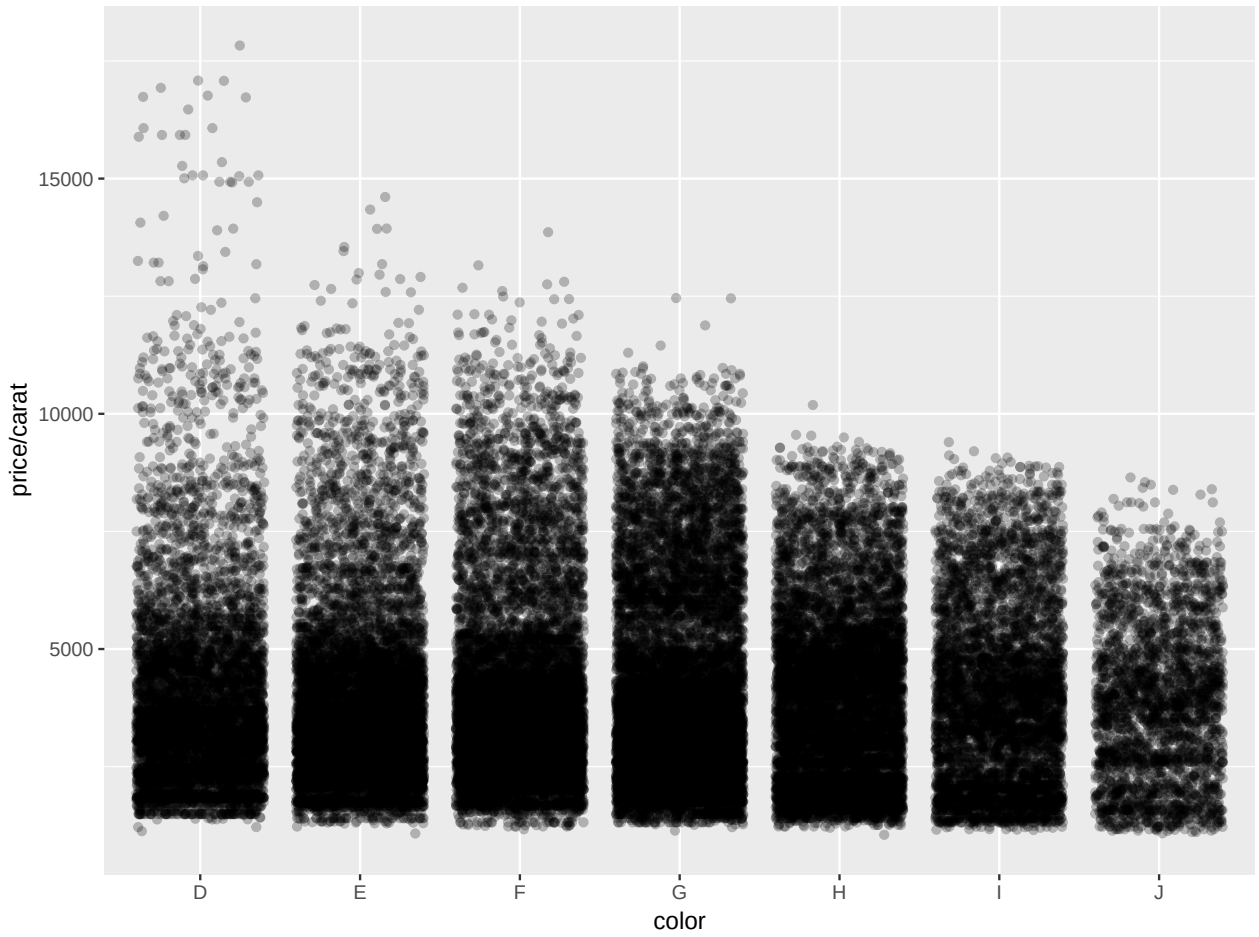



```
qplot(log(carat), log(price), data = diamonds, colour = color)
```

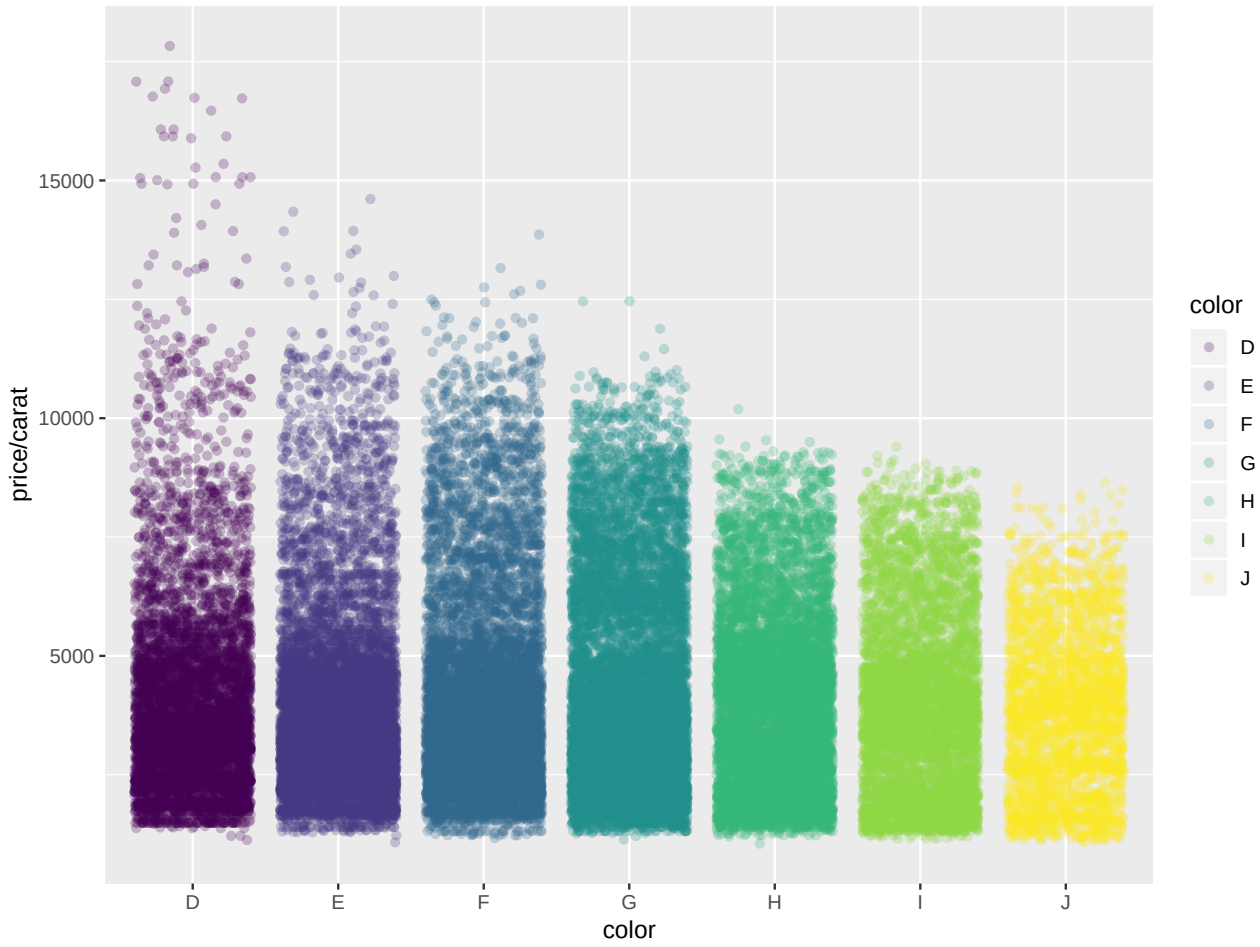


Jitter Plot

```
qplot(color, price/carat, data = diamonds, geom = "jitter", alpha = I(0.25))  # Standard Jitter plot
```

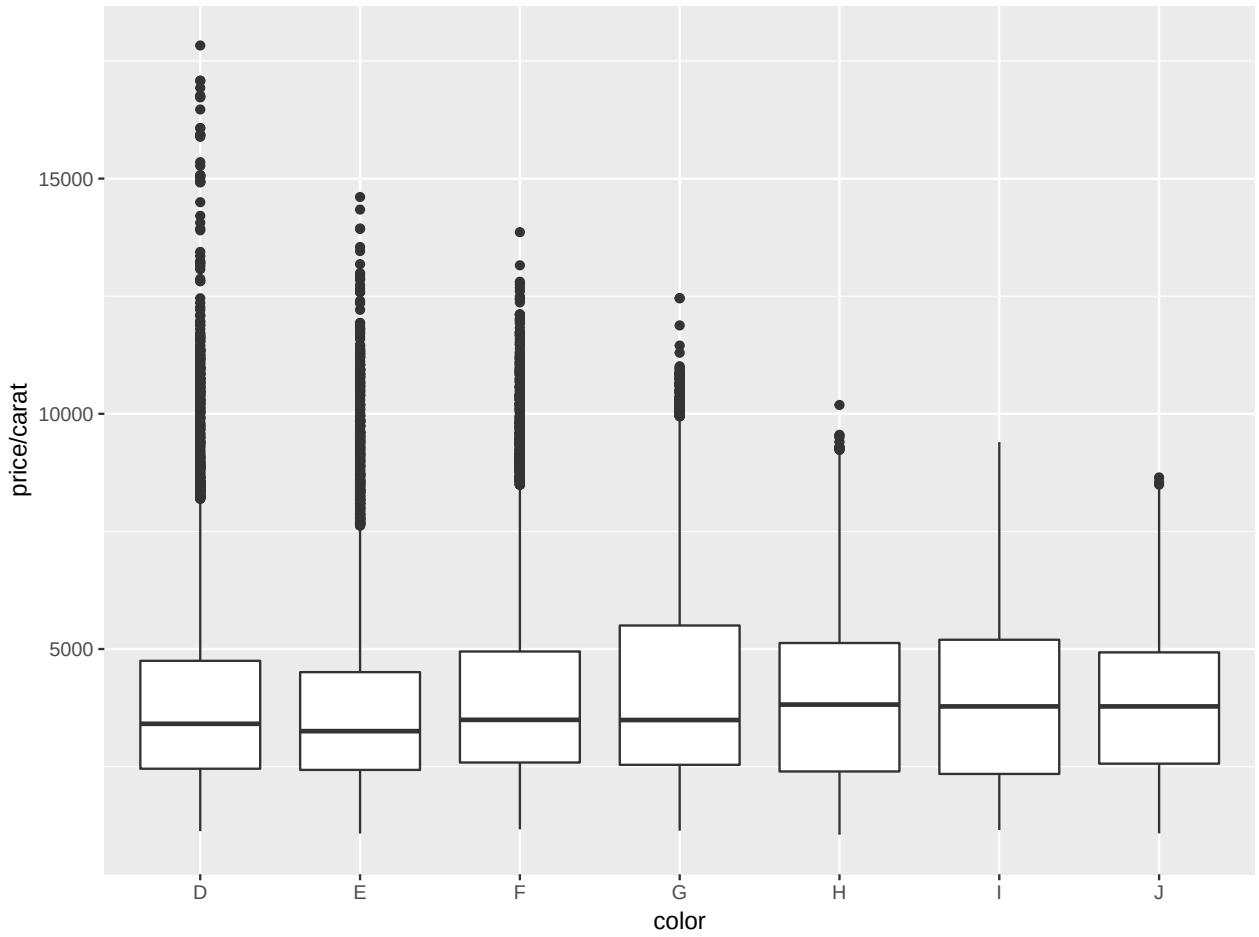



```
qplot(color, price/carat, data = diamonds, geom = "jitter", alpha = I(0.25),
      colour = color)
```

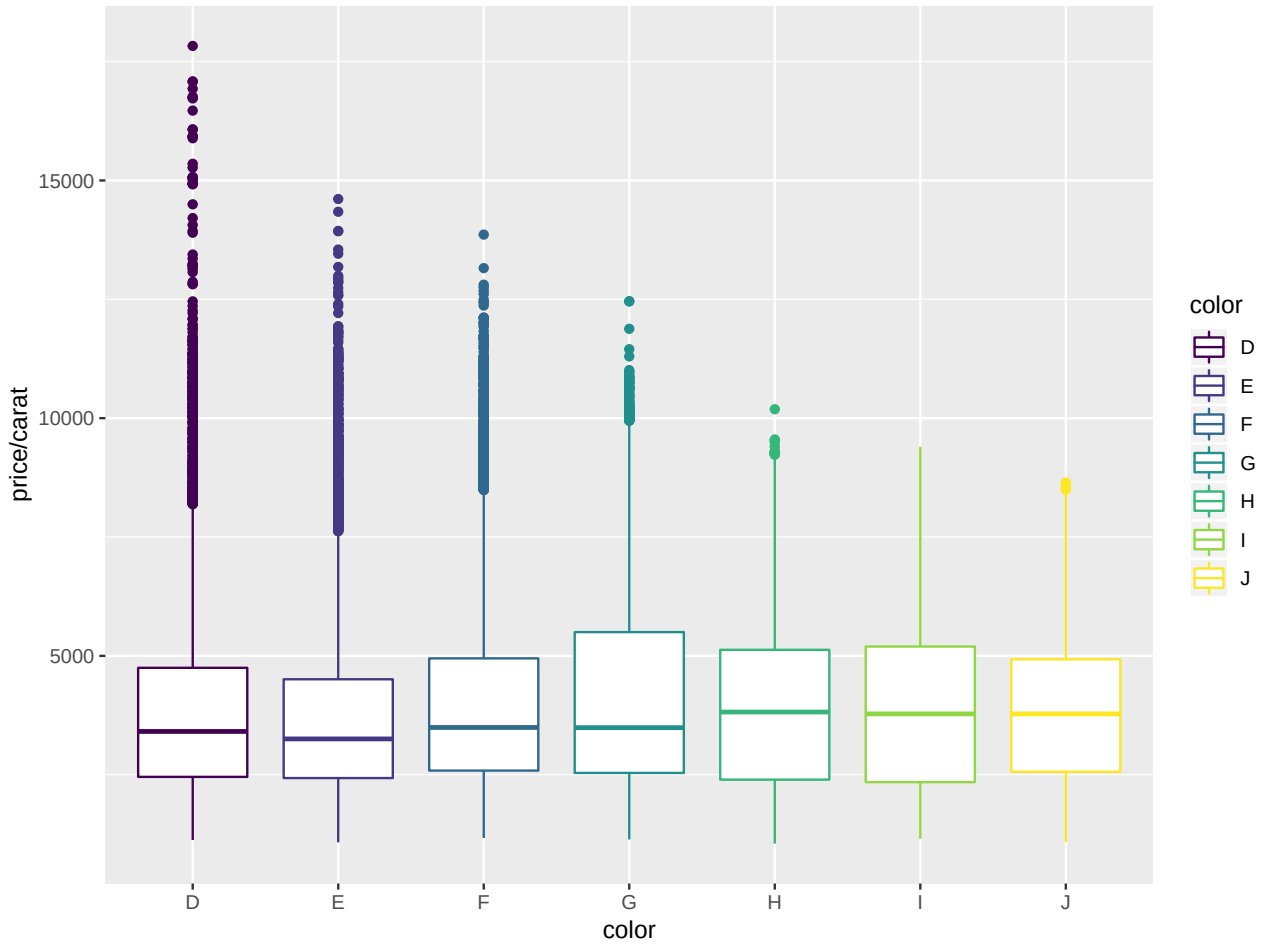


Boxplot

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

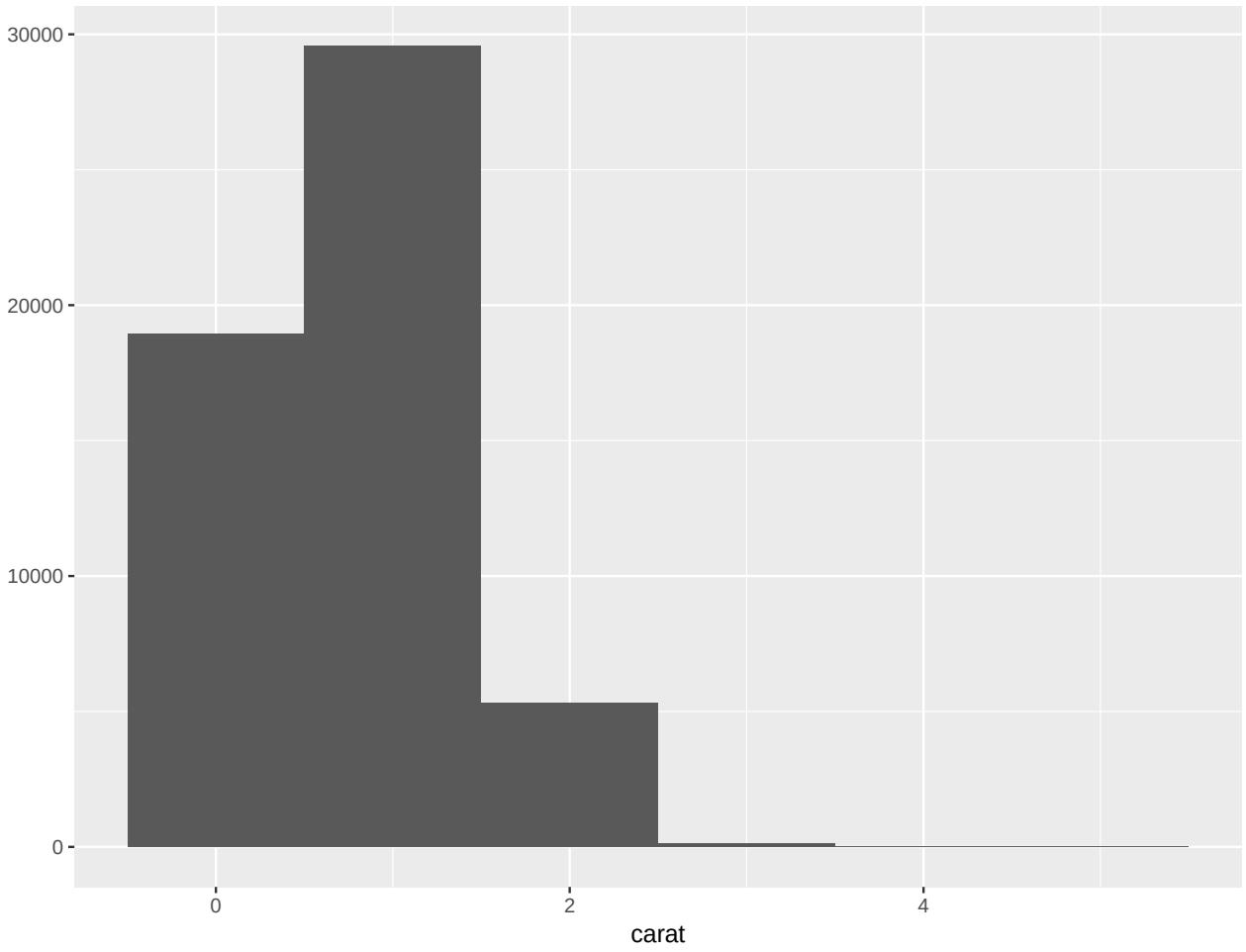



```
qplot(color, price/carat, data = diamonds, geom = "boxplot", colour = color)
```

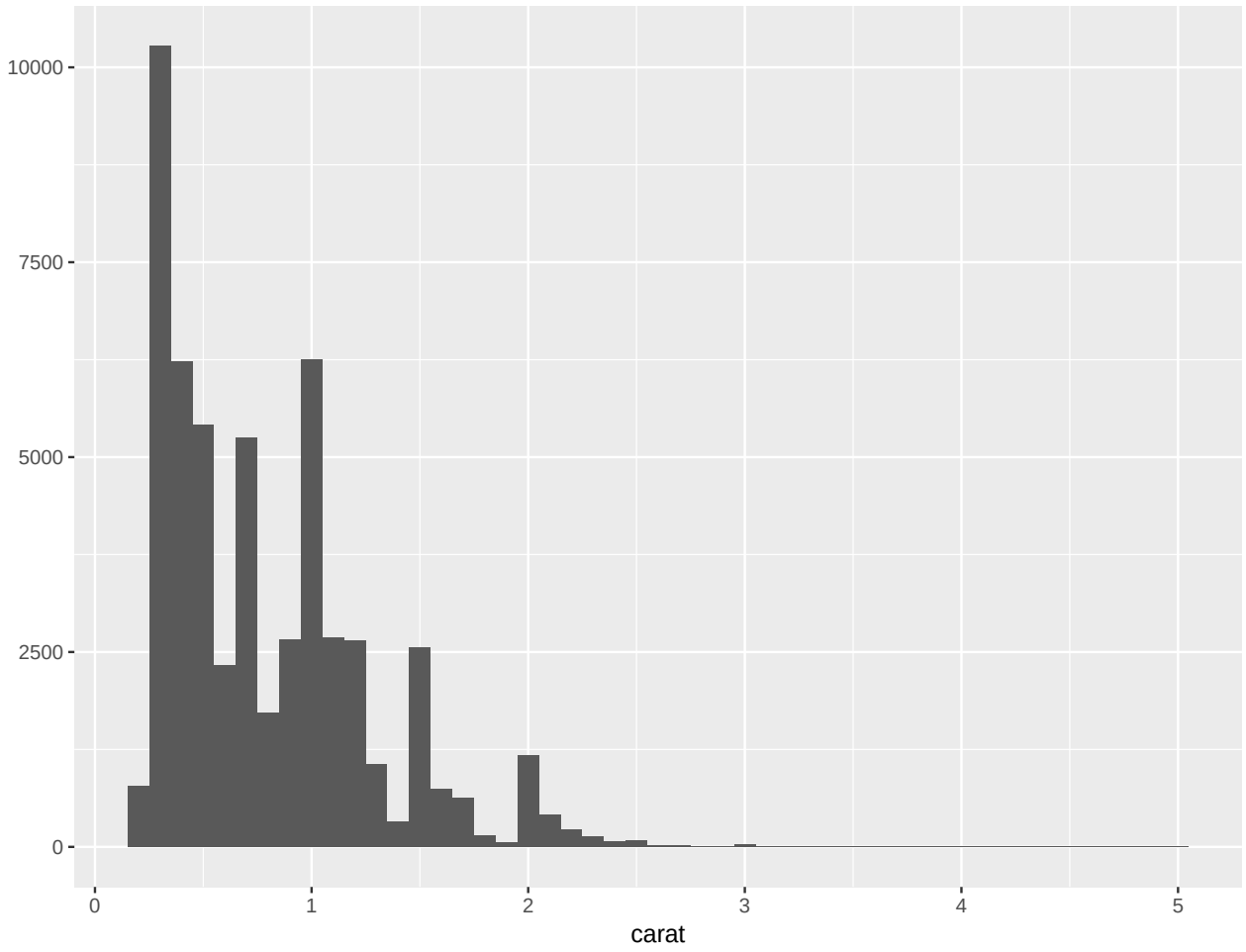


Histogram

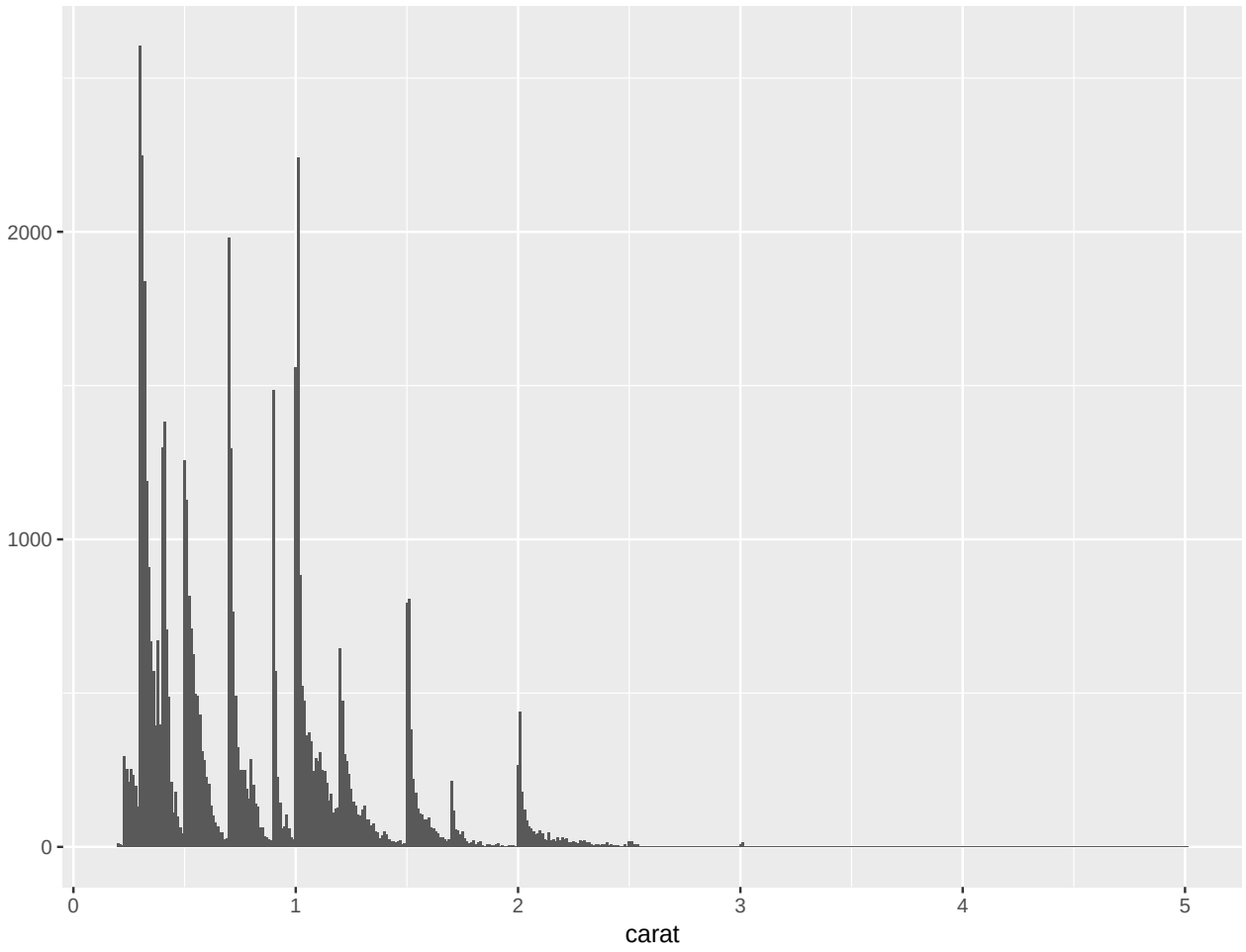
```
qplot(carat, data = diamonds, geom = "histogram", binwidth = 1)
```



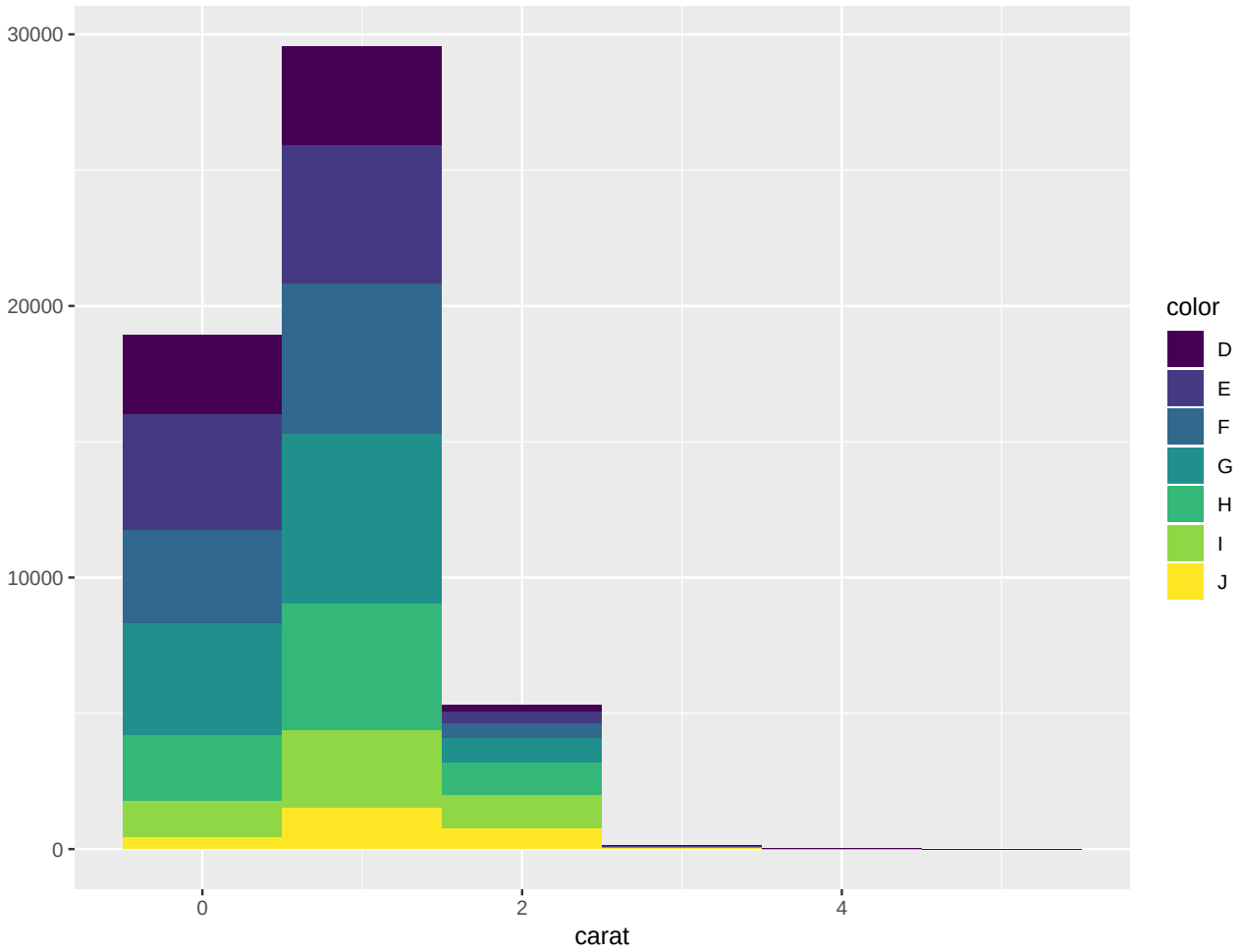

```
qplot(carat, data = diamonds, geom = "histogram", binwidth = 0.1)
```



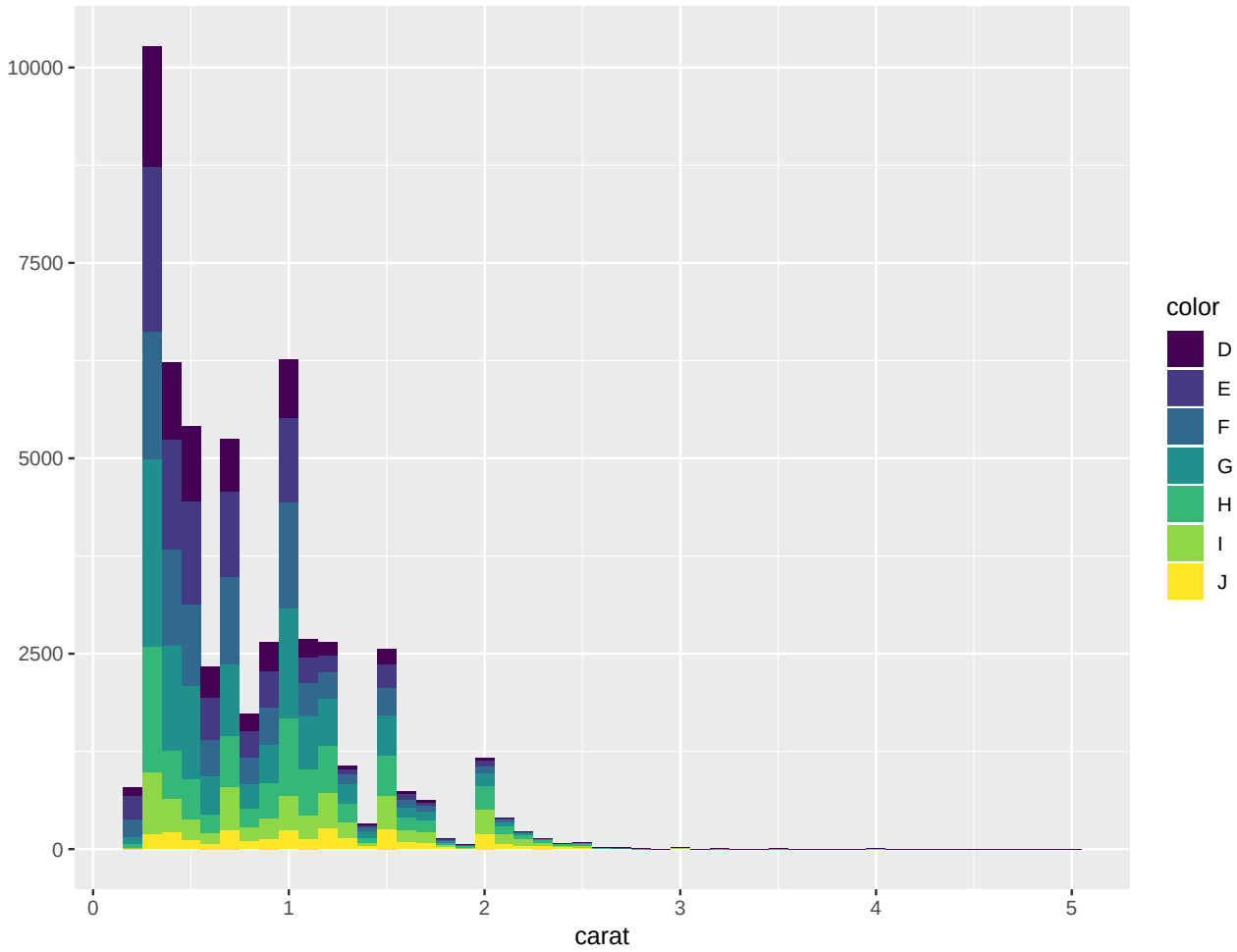

```
qplot(carat, data = diamonds, geom = "histogram", binwidth = 0.01)
```



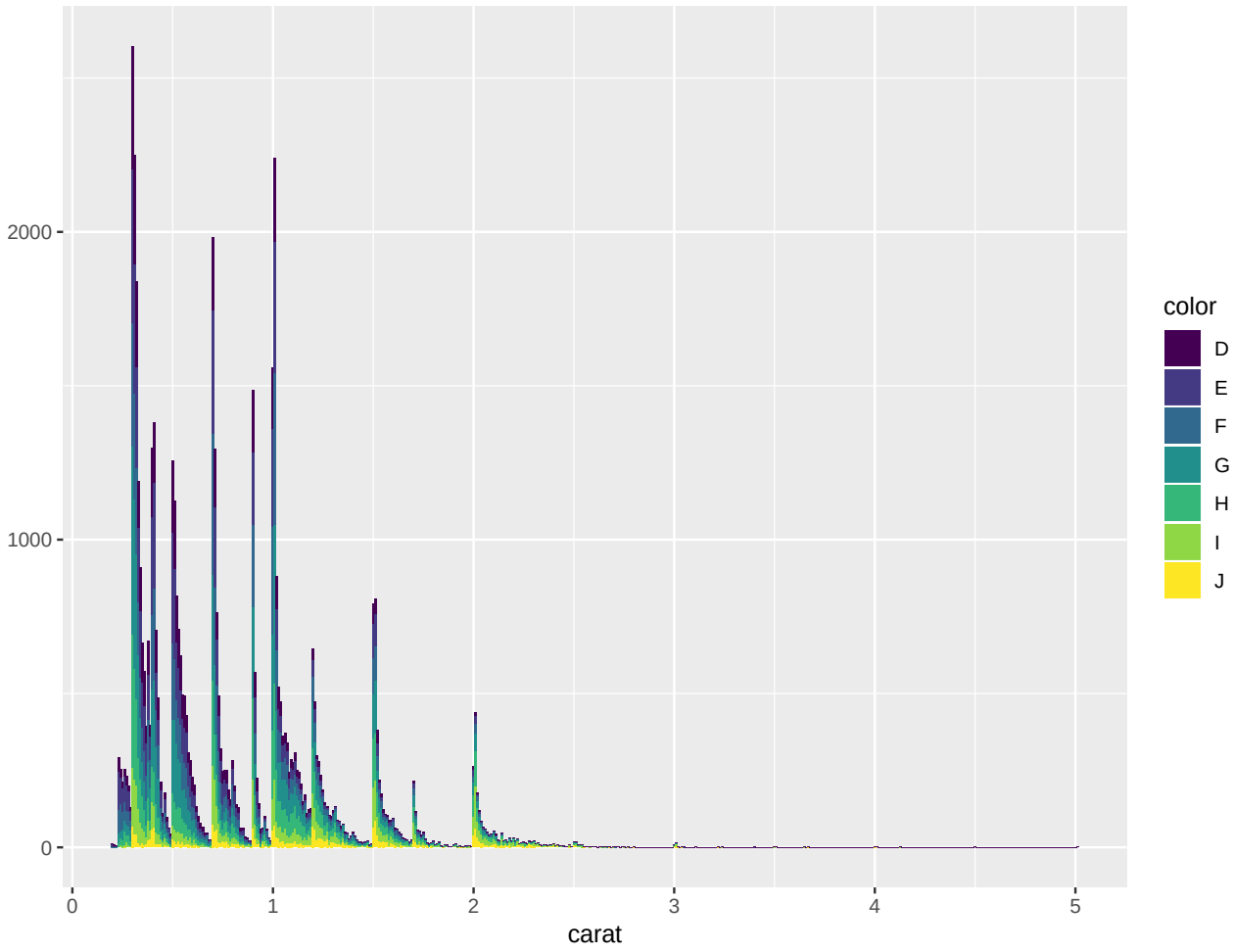

```
qplot(carat, data = diamonds, geom = "histogram", binwidth = 1, fill = color)
```




```
qplot(carat, data = diamonds, geom = "histogram", binwidth = 0.1, fill = color)
```

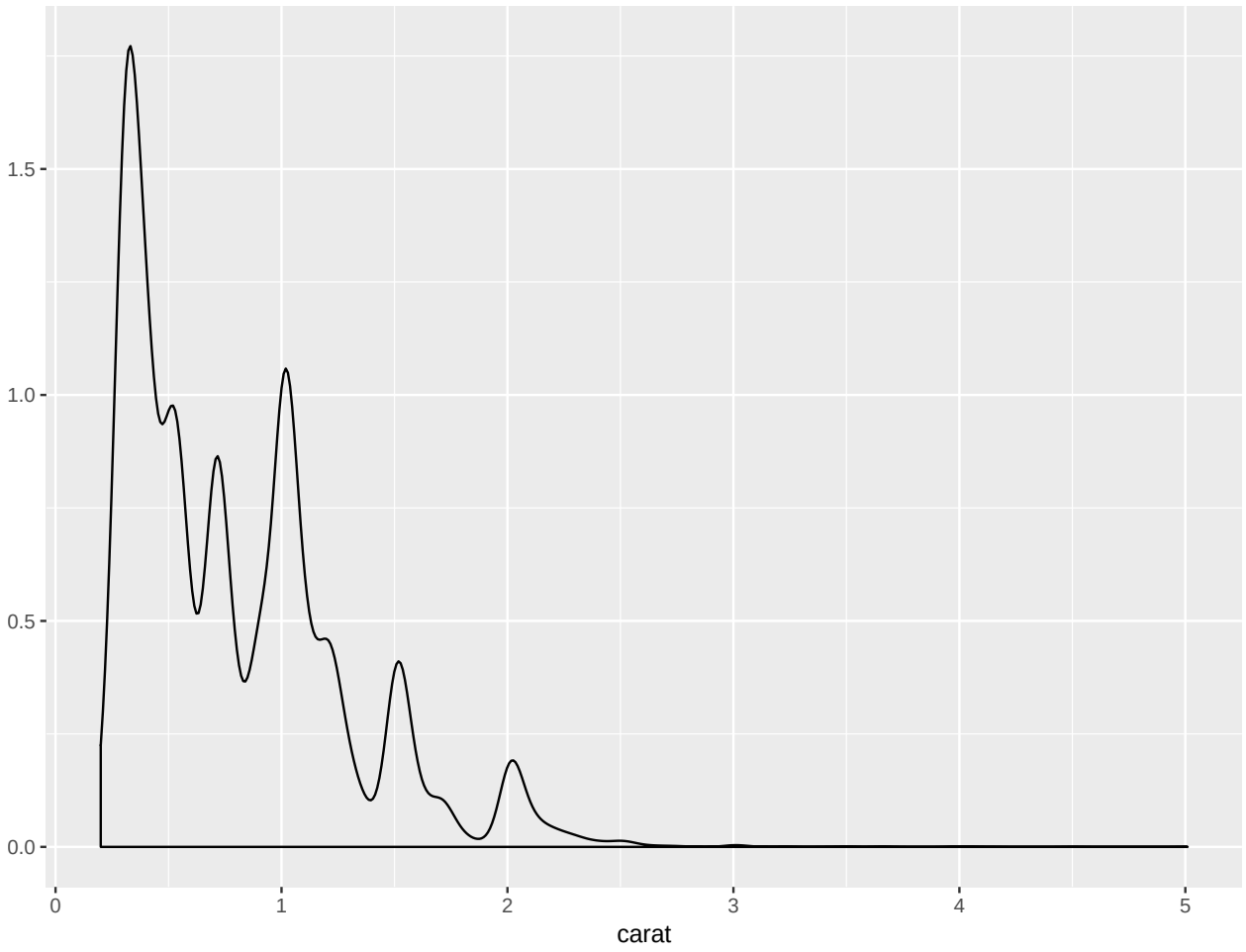



```
qplot(carat, data = diamonds, geom = "histogram", binwidth = 0.01, fill = color)
```

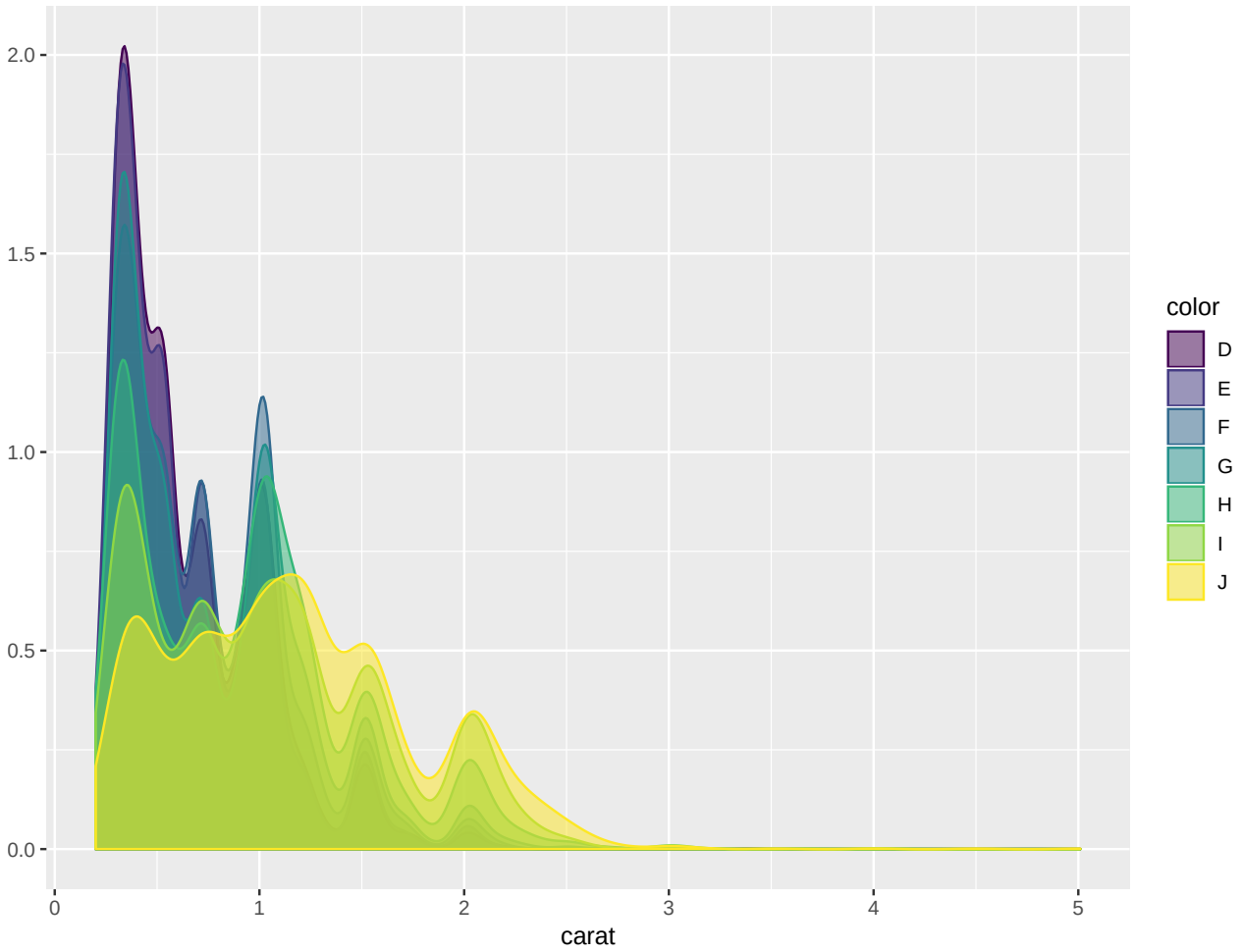


Density plot

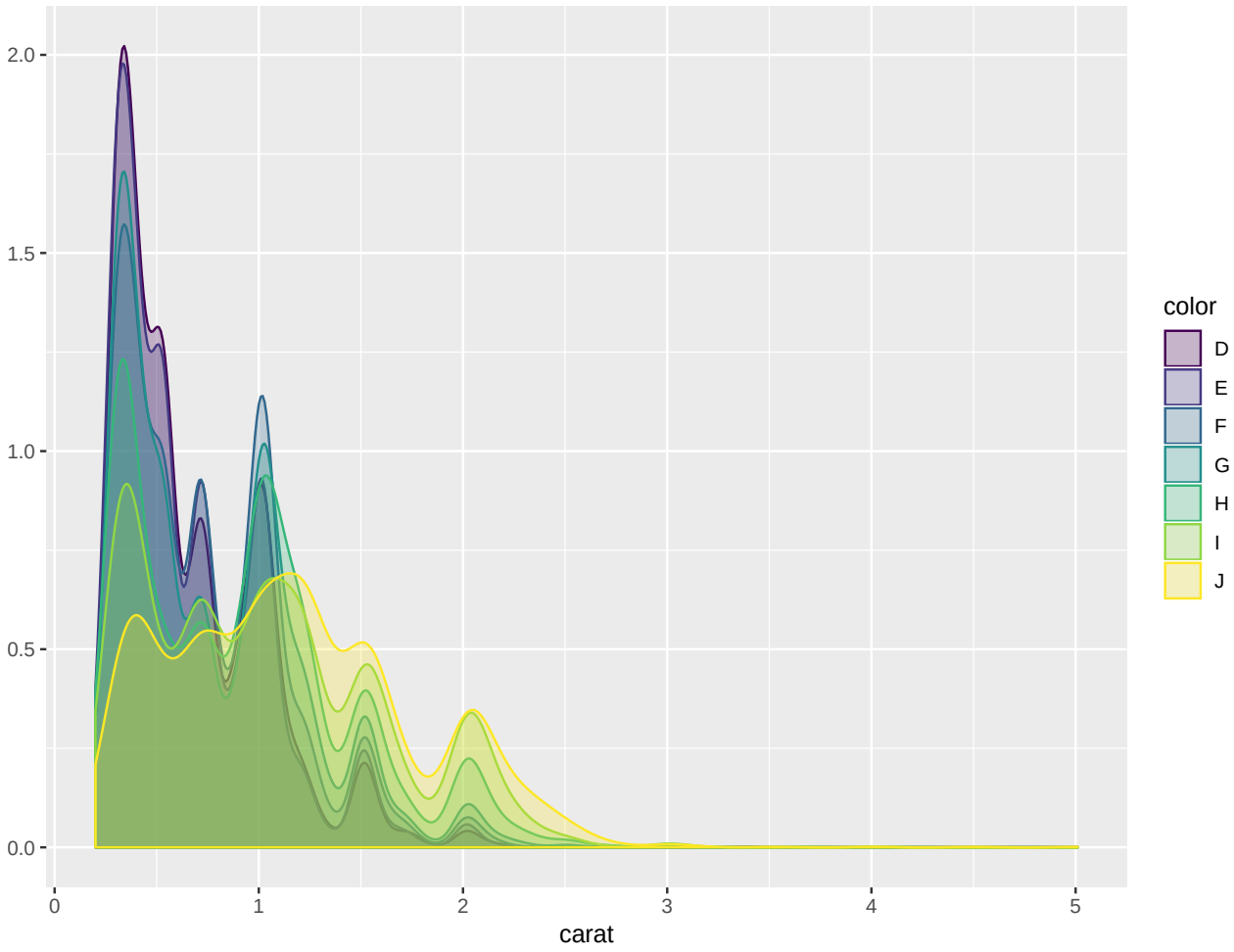
```
qplot(carat, data = diamonds, geom = "density")
```




```
qplot(carat, data = diamonds, geom = "density", colour = color,
      fill = color, alpha = I(0.5))
```

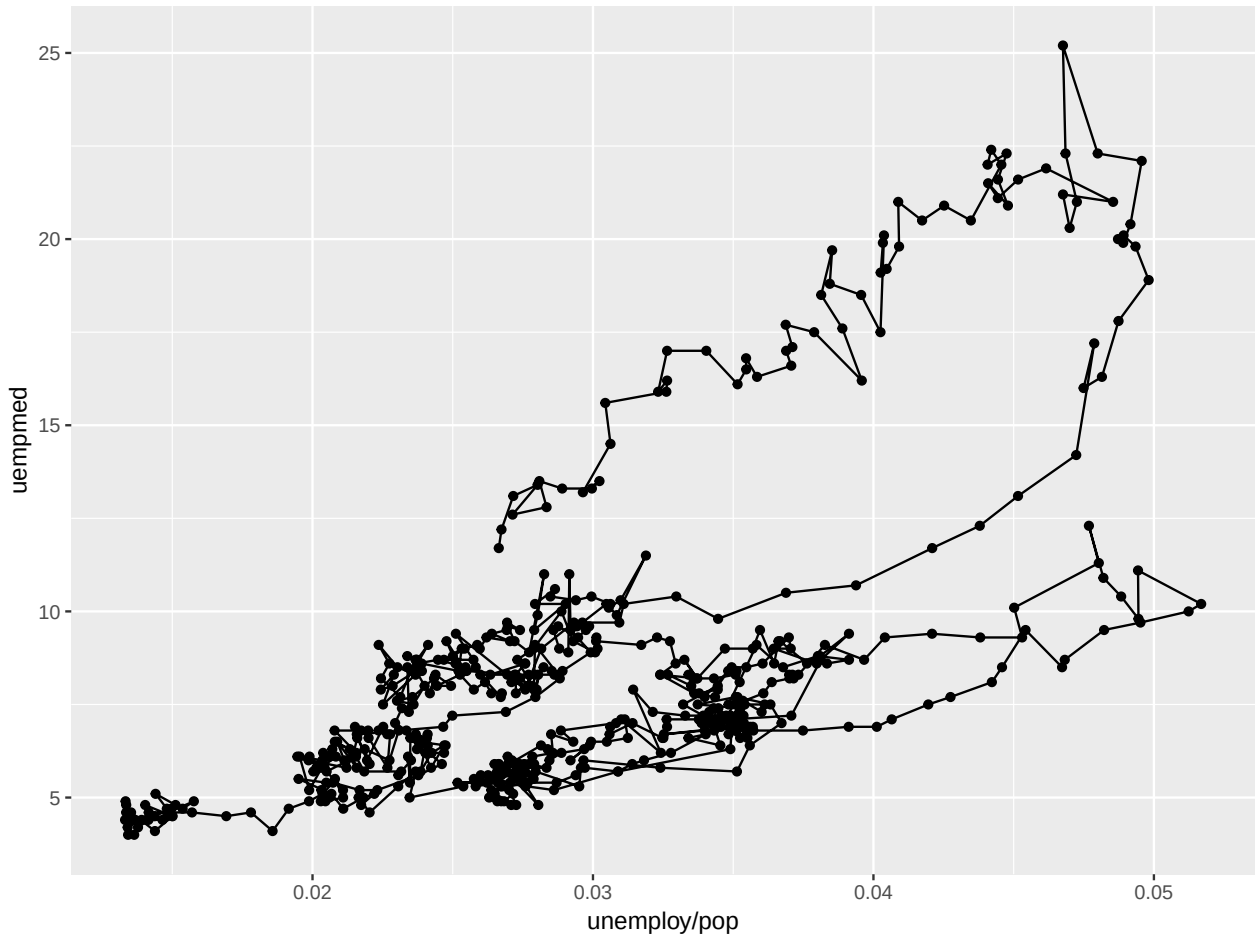



```
qplot(carat, data = diamonds, geom = "density", colour = color,
      fill = color, alpha = I(0.25))
```

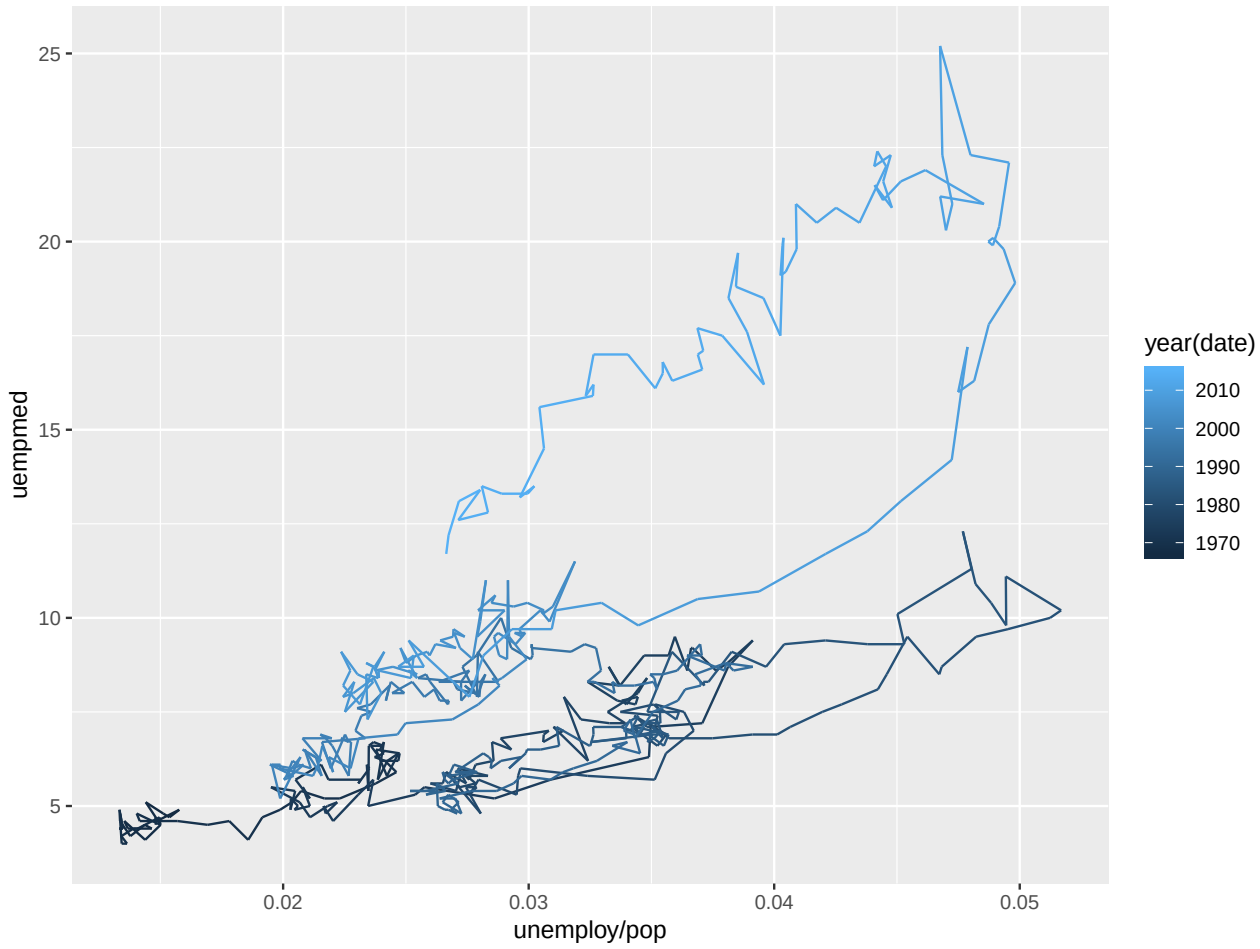


Time Series with "path"

```
year = function(x) as.POSIXlt(x)$year + 1900
qplot(unemploy/pop, uempmed, data = economics, geom = c("point", "path"))
```




```
qplot(unemploy/pop, uempmed, data = economics, geom = "path", colour = year(date)) +
  scale_size_area()
```


25
20
15
10
5
uempmed
0.02
0.03
0.04
0.05
unemploy/pop
year(date)
2010
2000
1990
1980
1970