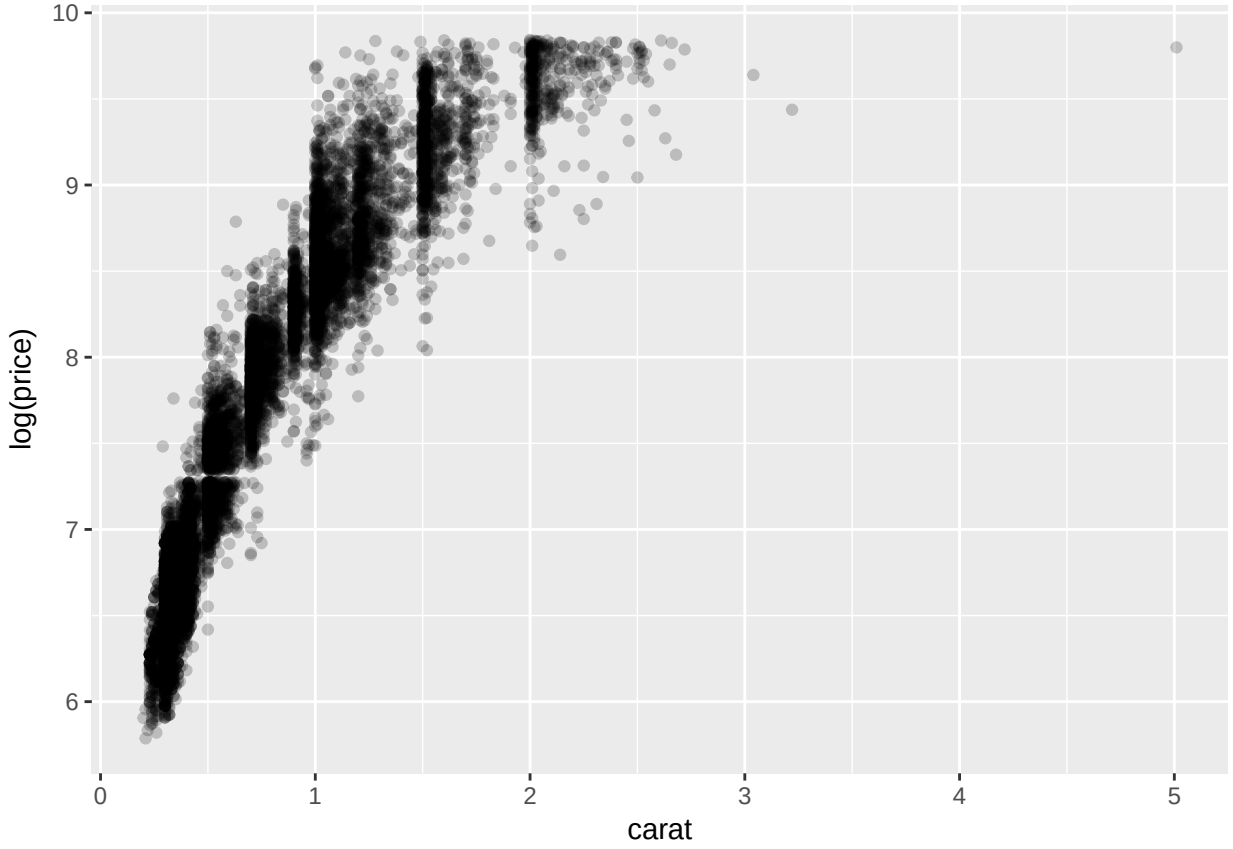
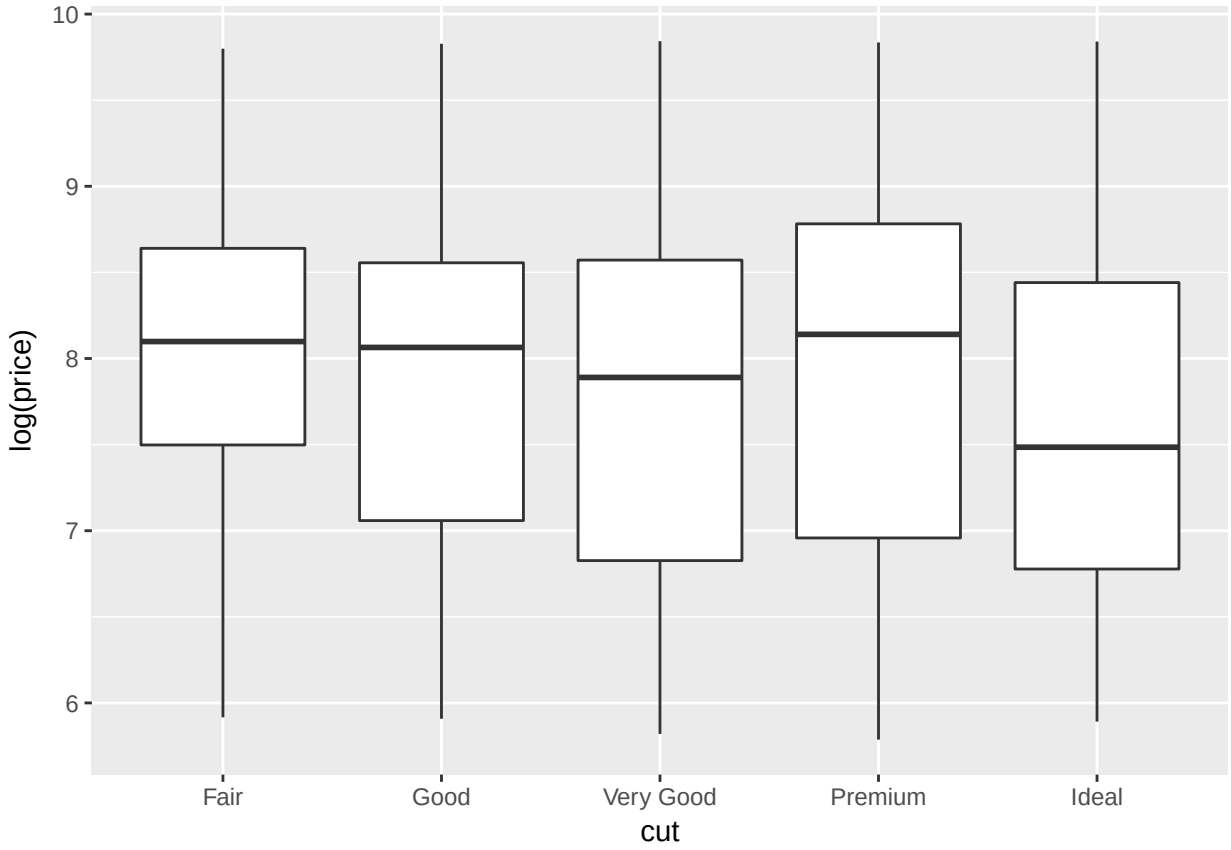
From Visualization to Modeling

The `diamonds` data set contains observations of over 50,000 diamonds. In the following, we focus on a sample of 10,000 diamonds and the relationship between their `log(price)`, their `carat` (size), and their `cut` (quality of the diamond). Look carefully at the following graphics and summarize the relationship that you see in a **single sentence**. Note that the last two plots are two ways to view the same variables, form a single sentence for both plots.

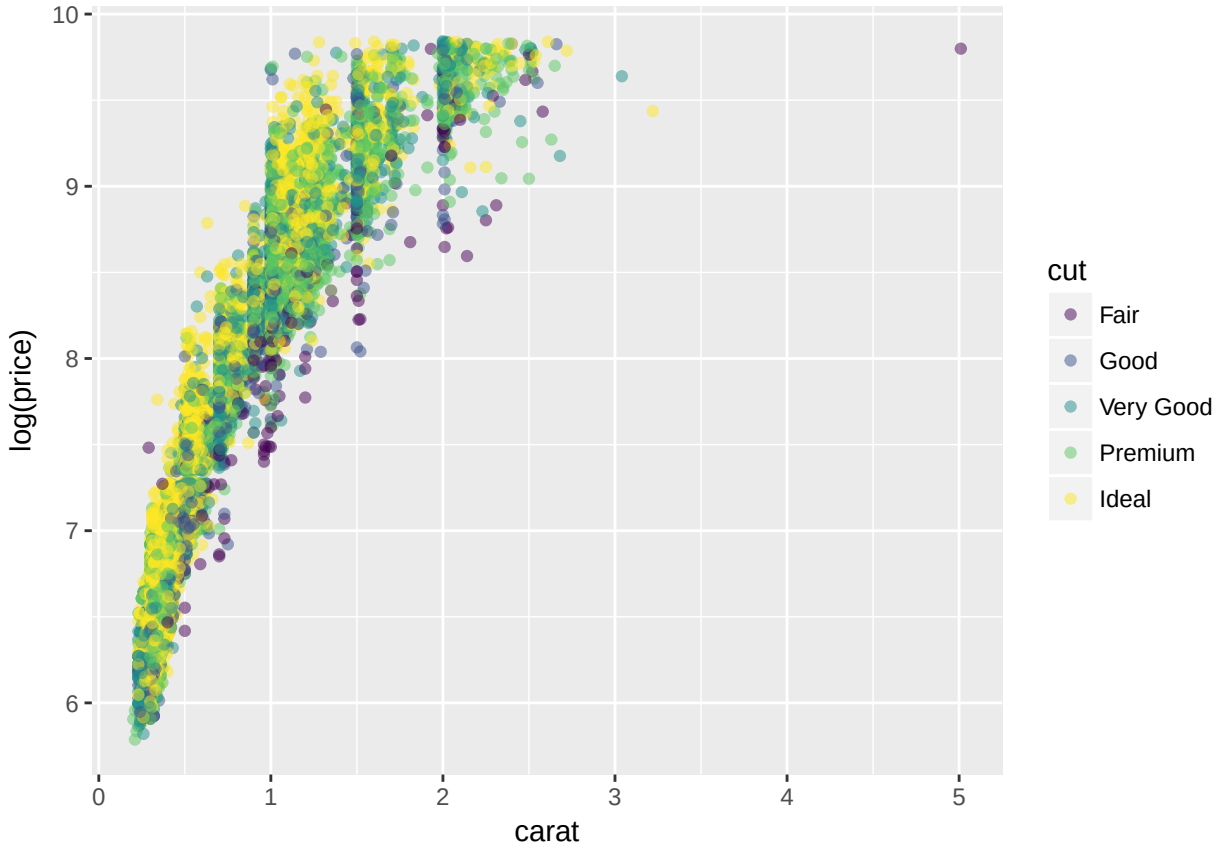
```
ggplot(diamonds, aes(x = carat, y = log(price))) +
  geom_point(alpha = .2)
```




```
ggplot(diamonds, aes(x = cut, y = log(price))) +
  geom_boxplot()
```




```
ggplot(diamonds, aes(x = carat, y = log(price))) +
  geom_point(aes(color = cut), alpha = .5)
```


10
9
8
7
6
log(price)
0
1
2
3
4
5
carat
cut
Fair
Good
Very Good
Premium
Ideal

```
ggplot(diamonds, aes(x = carat, y = log(price))) +
  geom_point(alpha = .2) +
  facet_wrap(~cut, nrow = 1)
```