R_visual_HW

Kittitat

Basic information of diamonds

```
library(tidyverse)
library(patchwork)
glimpse(diamonds)
```

```
## Rows: 53,940
## Columns: 10
## $ carat   <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, 0.23, 0.~
## $ cut     <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very Good, Ver~
## $ color   <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, J, J, I,~
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, VS1, SI1, VS1, ~
## $ depth   <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, 65.1, 59.4, 64~
## $ table   <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, 61, 54, 62, 58~
## $ price   <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338, 339, 340, 34~
## $ x       <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, 3.87, 4.00, 4.~
## $ y       <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, 3.78, 4.05, 4.~
## $ z       <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, 2.49, 2.39, 2.~
```
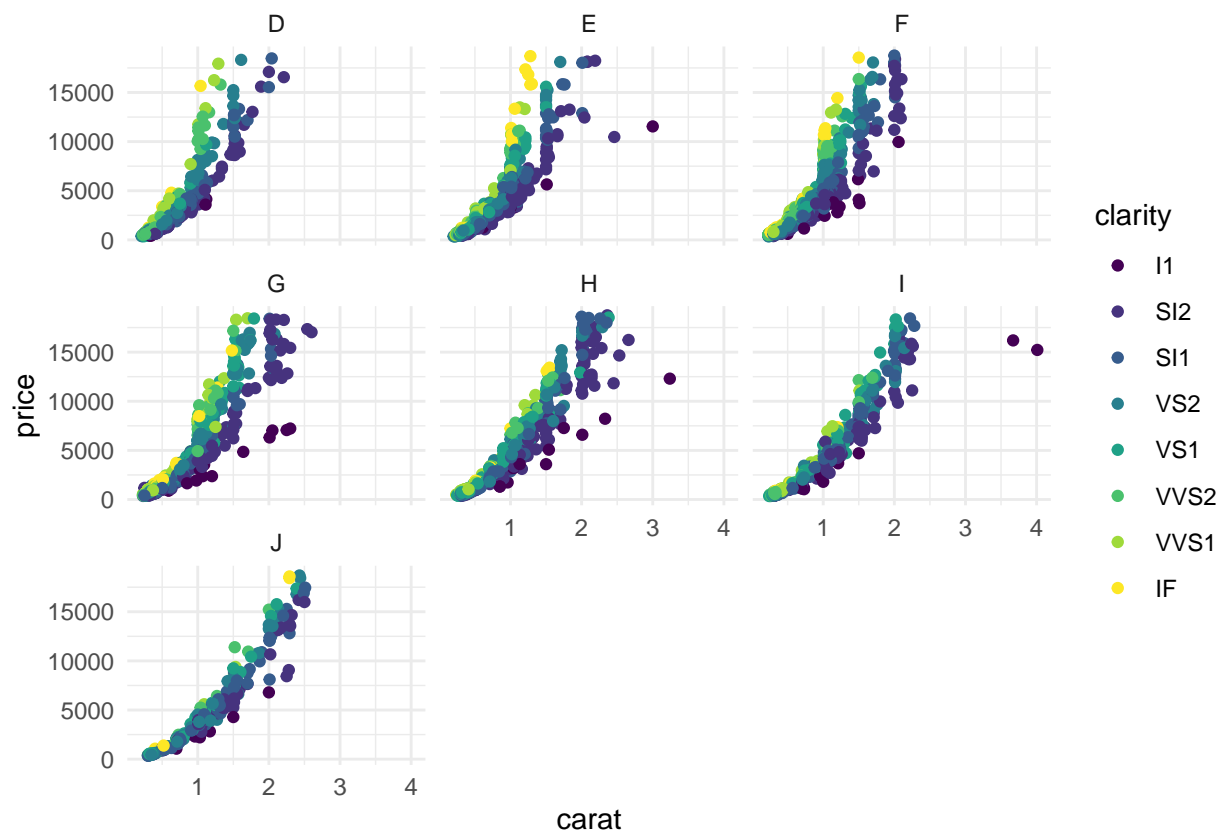
Data sampling

Get 10% the sample of data to analyse.

```
n_diamonds <- diamonds %>%
    sample_n(5000)
```

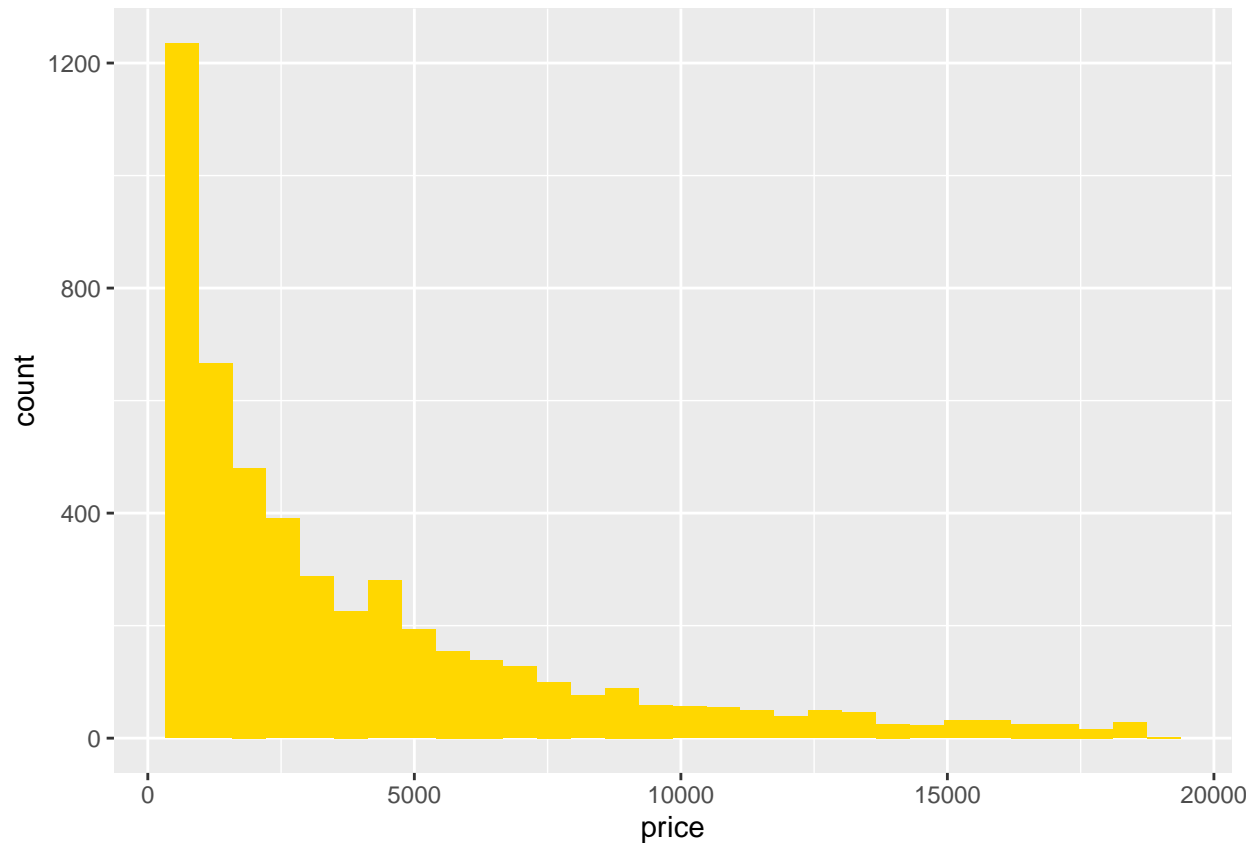## Diamon realtionship

The increase price of diamonds

Price of diamonds depend on carat and clarity which happen the same to all color of diamonds as the price increases when the size is bigger, the color and the clarity is more clear.

Diamond selection

People seem to pay more attention to carat, clarity, color and cut respectively. However, we could see that all cut have almost the same distribution of clarity as Ideal pose the greatest amount while Premium is the biggest figure for SI2 with Ideal VS2 clarity and E color has the largest proportion

```
ggplot(n_diamonds , aes(clarity, fill=color)) +
  geom_bar(position= "stack") +
  facet_wrap(~cut)
```

Diamon price distribution

The price distribution displays large amount of diamonds cost less than than 5000 and that of sample is 3671 which is more than half.

```
ggplot(n_diamonds , aes(price)) +
  geom_histogram(fill= "gold")
```
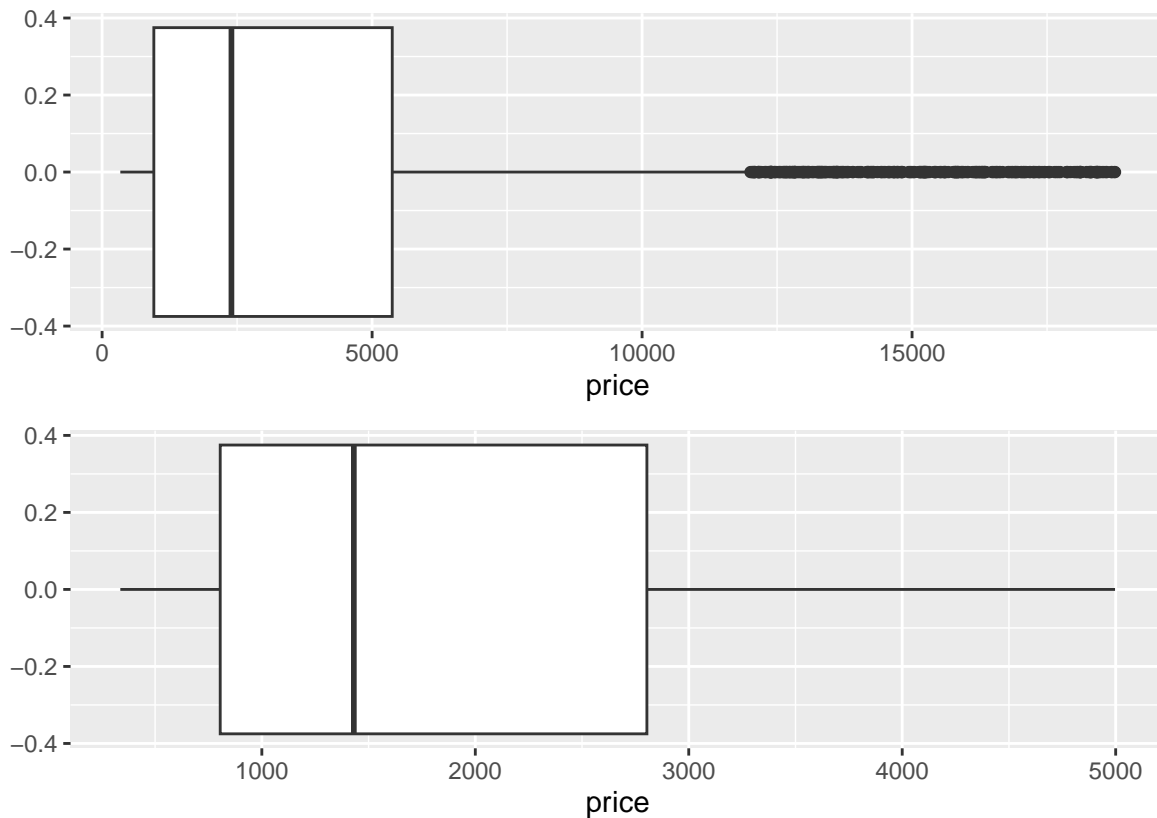
```
n_diamonds %>%
  filter(price<5000) %>%
  count()
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1  3641
```

So I took the opportunity to look into diamonds that has the price less than 5000

```
lp_n_diamonds <- n_diamonds %>%
  filter(price<5000)
p1 <- qplot(price,data=n_diamonds,geom="boxplot")
p2 <- qplot(price,data=lp_n_diamonds,geom="boxplot")
p1/p2
```

Acceptable grade of diamond in good value

GIA said that there is a range of diamonds clarity which human eyes could not detect and the yellowish color as the diamonds will be seemed to be clear or acceptable or indistinguishable but not flawless. What is more, the perfect size of the diamonds is more than 1 carat, but in this sample the are a certain diamonds is in between that size, so I would show that in the chart. Additionally, the clarity of VS2, SI1 and SI2 illustrate themselves as clear as others with the same reason of people eyes color detection.

```
ggplot(lp_n_diamonds %>%
        filter(clarity %in% c("VS2","SI1","SI2"),
               color != "J",
               carat > 0.8), aes(carat, fill=color)) +
  geom_histogram( stat = "bin")
```

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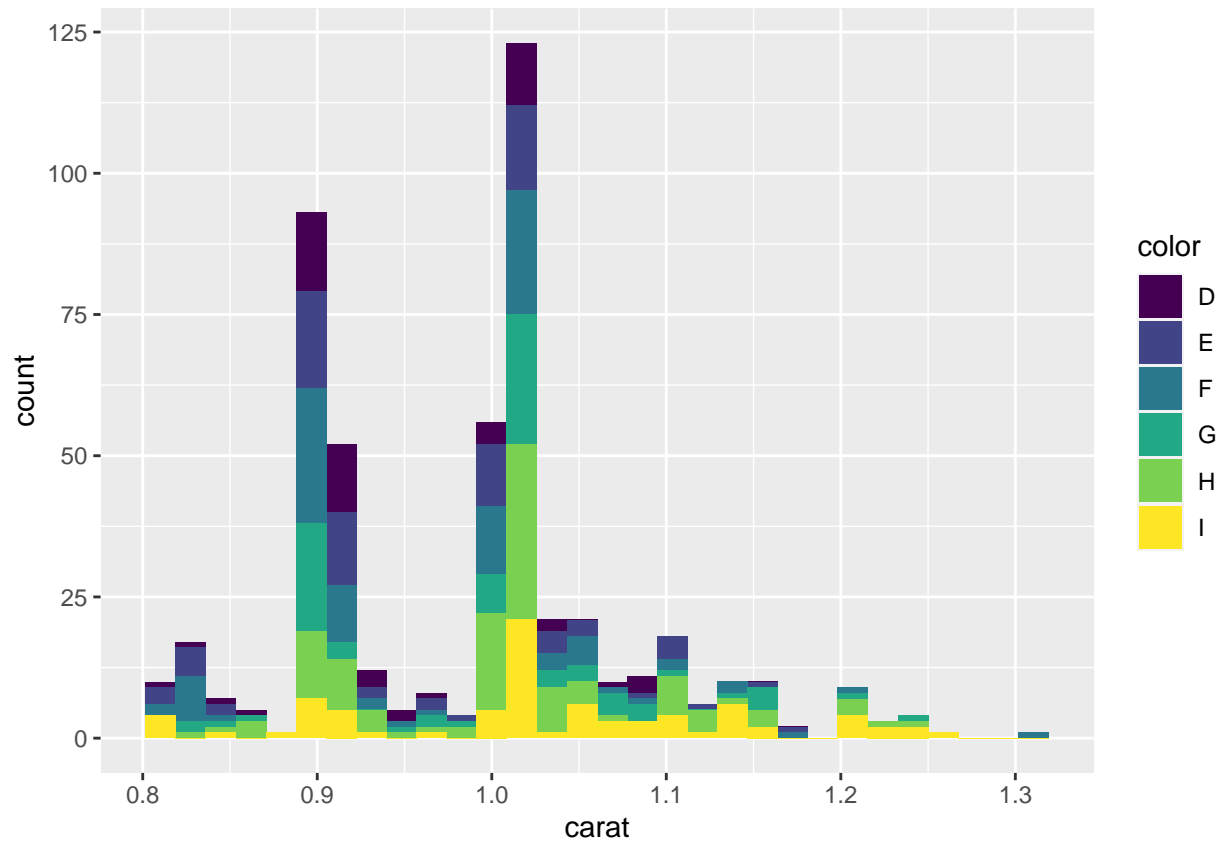
Summary

There are still various diamonds that has the cost of less than 5000 with great conditions and could not be separated to the one that has the cost of double or triple.