# Homework

Nattanit im

2023-07-05

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

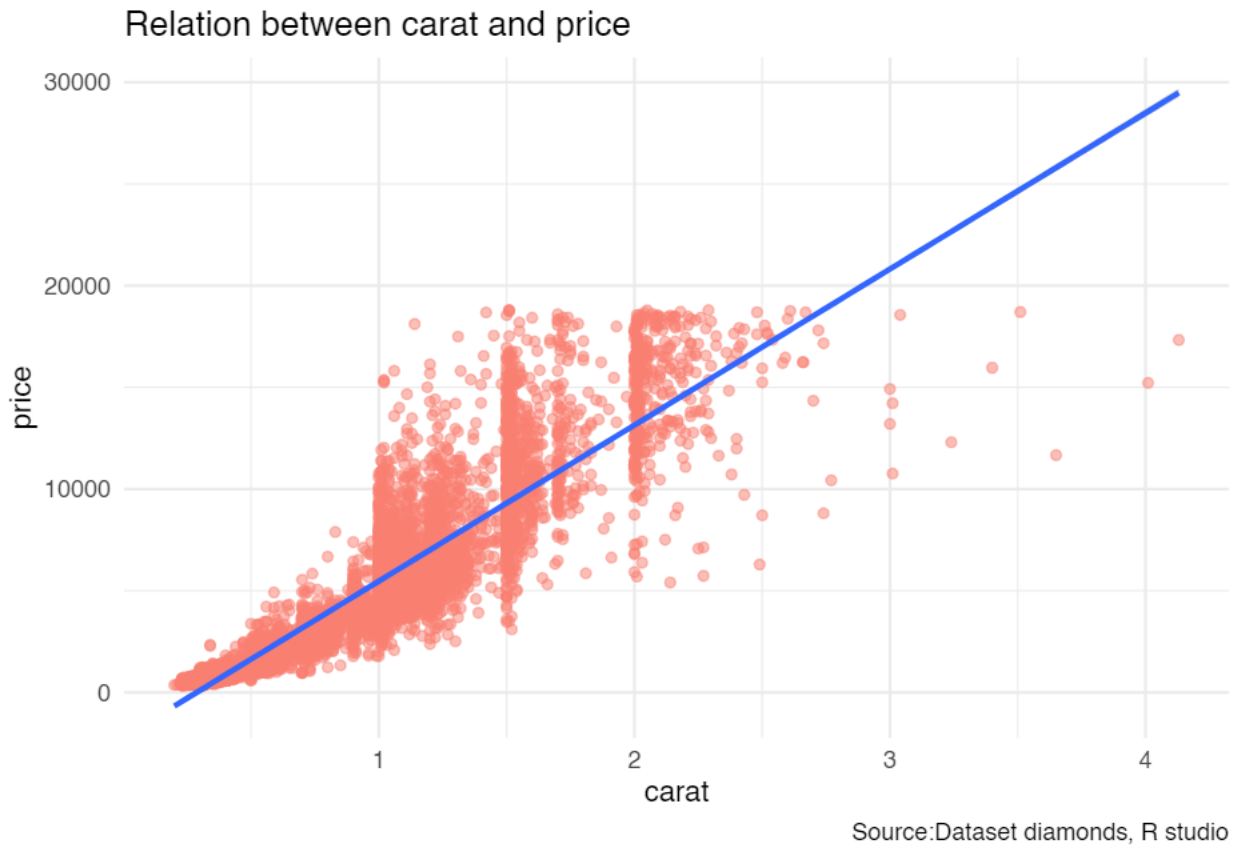
```
set.seed(12)
sample_diamonds <- sample_frac(diamonds,0.2)
```

## Question1

### What is the relation between carat and price?

```
library(tidyverse)
library(ggplot2)
ggplot(sample_diamonds, aes(carat,price))+
  geom_point(col="salmon",alpha=0.5)+
  geom_smooth(method="lm")+
  theme_minimal()+
  labs(
    title = "Relation between carat and price",
    caption = "Source:Dataset diamonds, R studio"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




## Find the price per carat

```
cor(sample_diamonds$carat, sample_diamonds$price)
```

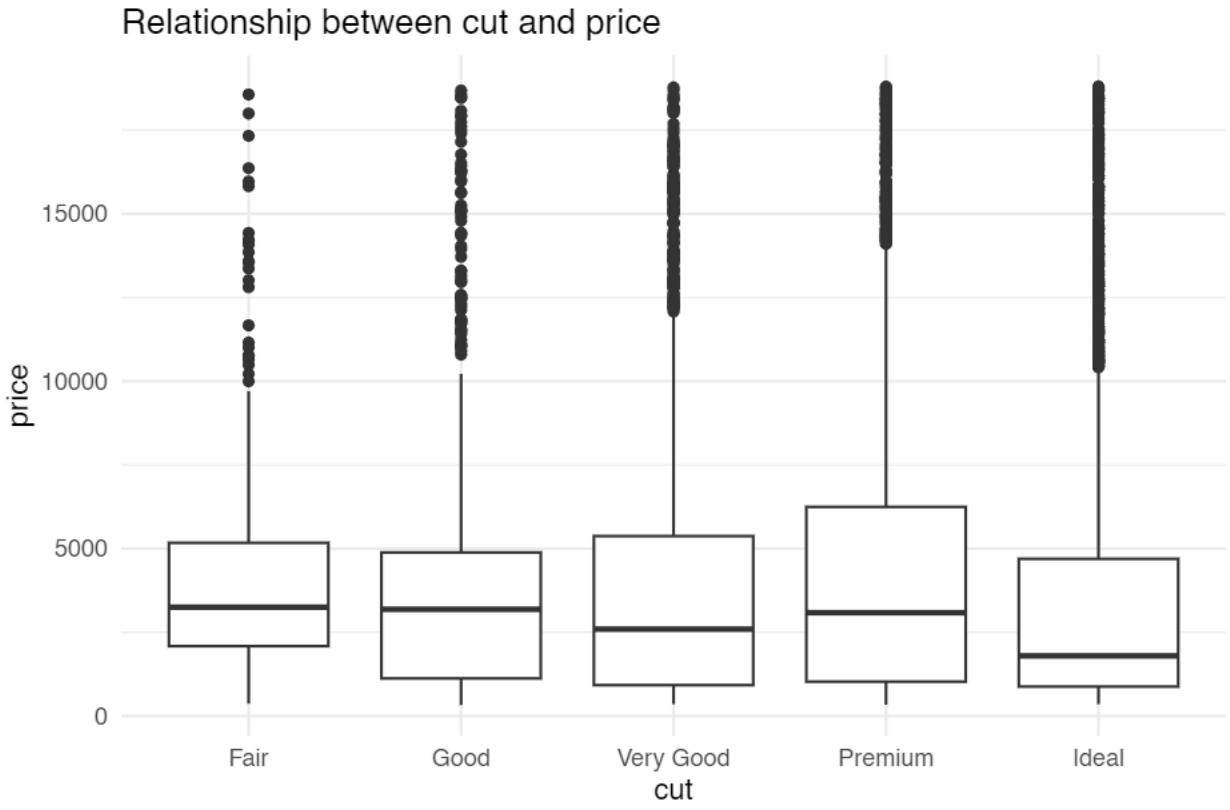
```
## [1] 0.9182729
```

From this plot, we found that it is an increasing trend with correlation = 0.9207463, that is mean the higher carat might has the higher price. Moreover, some diamonds that its carat equals 1 to 5 have the same price, means they have another options that make the price changing.

### Question 2

### What is the relationship between cutting and price of diamonds?

```
ggplot(sample_diamonds, aes(cut,price))+
  geom_boxplot()+
  theme_minimal()+
  labs(
    title = "Relationship between cut and price",
    caption = "Source: Dataset Diamonds, R studio"
  )
```




```
sample_diamonds %>%
  group_by(cut) %>%
  summarise(
    median(price),
    mean(price)
  )
```

```
## # A tibble: 5 x 3
##   cut       `median(price)` `mean(price)`
##   <ord>               <dbl>         <dbl>
## 1 Fair                3248          4273.
## 2 Good                3186.         3878.
## 3 Very Good           2593          3926.
## 4 Premium             3084          4512.
## 5 Ideal               1795          3429.
```

From chart median and average values, found that the cutting is not or less effect to the price of diamonds. Need to find other options that effect to the price of diamonds.

## Question 3

### The relationship between color and price

Use the boxplot because the color is discrete variable.

```
ggplot(sample_diamonds, aes(color,price))+
  geom_boxplot()+
  theme_minimal()+
  labs(
```

```
    title = "Relationship between color and price",
    caption = "Source: Dataset diamonds, R studio"
  )
```


```
sample_diamonds %>%
  group_by(color) %>%
  summarise(
    median(price),
    mean(price)
  )
```

```
## # A tibble: 7 x 3
##   color `median(price)` `mean(price)`
##   <ord>           <dbl>         <dbl>
## 1 D                1847          3197.
## 2 E                1746          3019.
## 3 F                2333          3756.
## 4 G                2123          3920.
## 5 H                3381          4351.
## 6 I                3780          5090.
## 7 J                4004.         5079.
```

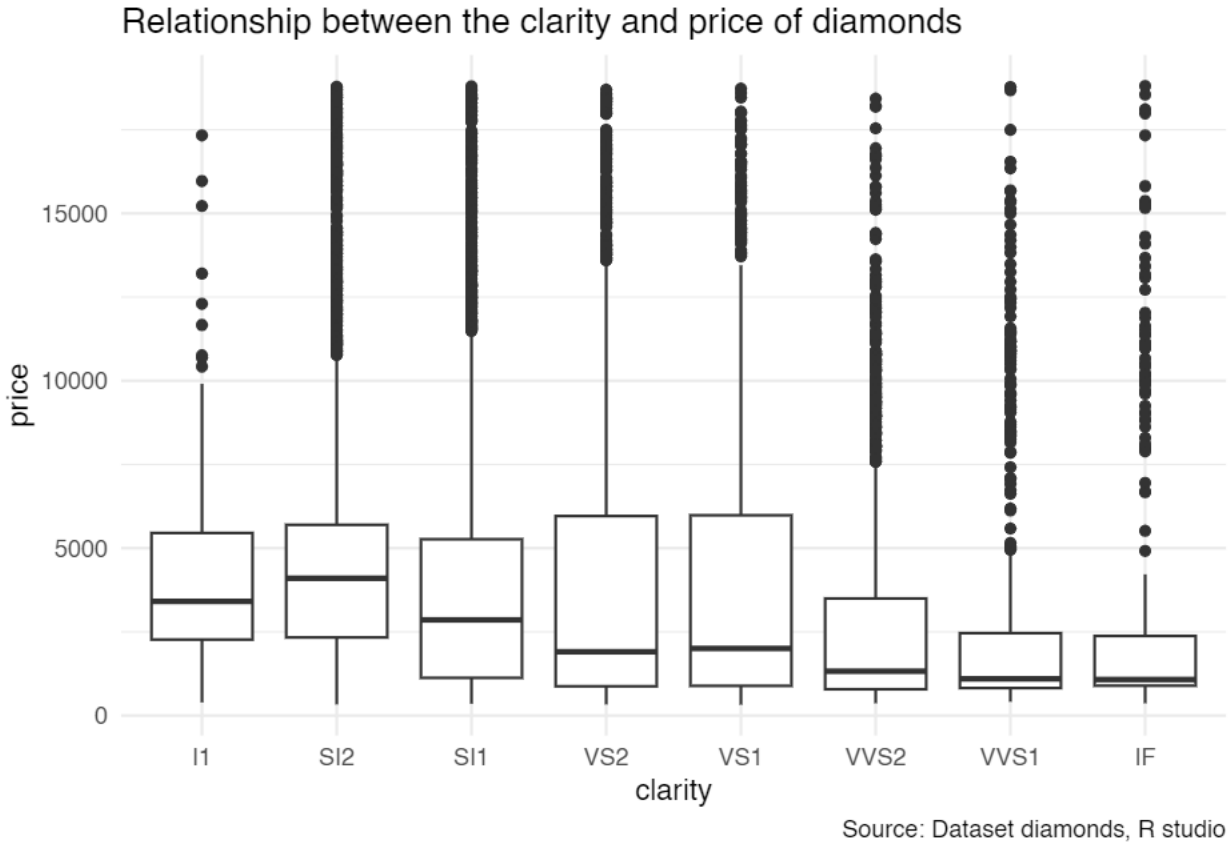
From boxplot and median and mean price above, found that the better color (Levels: D<E<F<G<H<I<J) will make the higher price too.

## Question 4

### Relationship between clarity and price of diamonds

Use boxplot because the clarity is discrete variable.

```
ggplot(sample_diamonds, aes(clarity,price))+
  geom_boxplot()+
  theme_minimal()+
  labs(
    title = "Relationship between the clarity and price of diamonds",
    caption = "Source: Dataset diamonds, R studio"
  )
```




```
sample_diamonds %>%
  group_by(clarity)%>%
  summarise(
    median(price),
    mean(price)
  )
```

```
## # A tibble: 8 x 3
##   clarity `median(price)` `mean(price)`
##   <ord>             <dbl>         <dbl>
## 1 I1                3410.         4188.
## 2 SI2               4099          4992.
## 3 SI1               2856.         3985.
## 4 VS2               1904          3831.
## 5 VS1               2003          3794.
```

```
## 6 VVS2              1323          3240.
## 7 VVS1              1096          2522.
## 8 IF                1074          2832.
```

The more clarity of diamonds, the lower price. Maybe they have negative correlation or they have another option that can make the price to be like this.

## Question 5

### What is the relation between price and carat, clarity and color

Use facet_grid to plot this case

```
ggplot(sample_diamonds,aes(carat,price))+
  geom_point()+
  geom_smooth(method="lm")+
  theme_minimal()+
  labs(
    title = "Question 5",
    caption = "Source: Dataset diamonds, R studio"
  )+
  facet_grid(clarity~color)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```