DataViz_Homework

PakkapornCh

2023-07-10

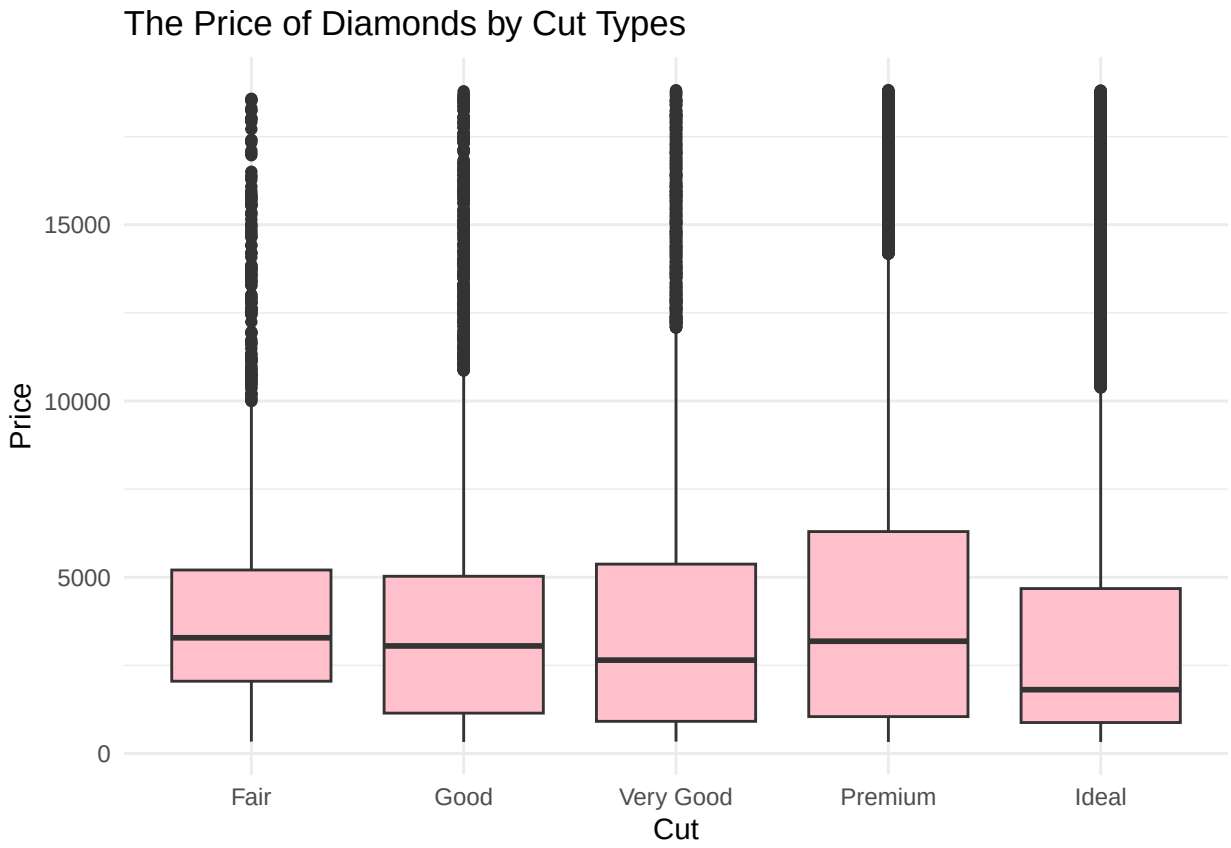
```
library(tidyverse)
```

```
## -- 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 errors
```

```
library(ggplot2)
```

1 The Price of Diamonds by Cut Types

```
ggplot(diamonds, aes(x = cut, y = price)) +
  geom_boxplot(fill="pink") +
  labs(title = "The Price of Diamonds by Cut Types",
        x = "Cut",
        y = "Price") +
        theme_minimal()
```

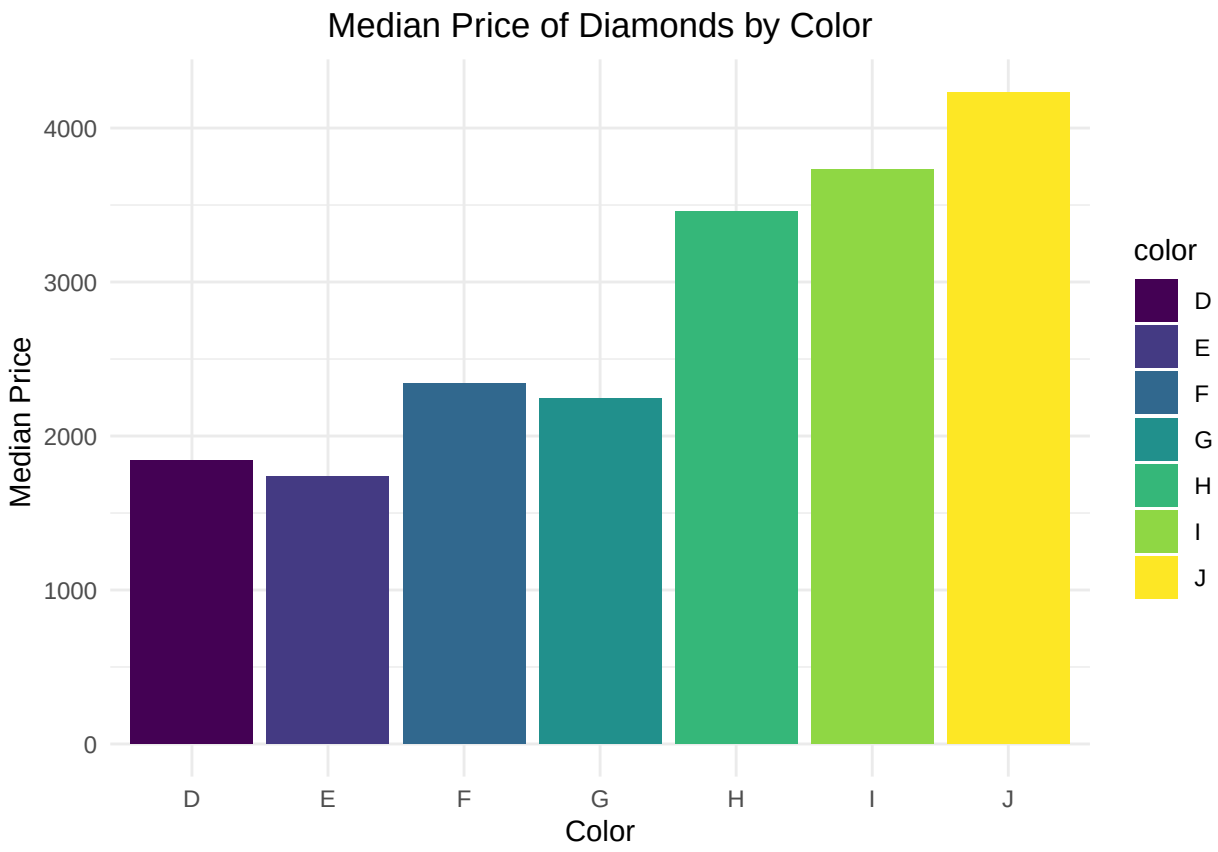


Meaning:

There are many outliers in diamond prices for each cut type, which means factors other than cut type influence the prices of diamonds.

2 Median Price of Diamonds by Color

```
diamonds %>%
   group_by(color) %>%
   summarise (
     med_price = median(price)) %>%
   ggplot(aes(color, med_price, fill = color)) +
   geom_col() +
   labs(title = "Median Price of Diamonds by Color",
        x = "Color",
        y = "Median Price ") +
   theme_minimal() +
   theme(plot.title = element_text(hjust = 0.5))
```



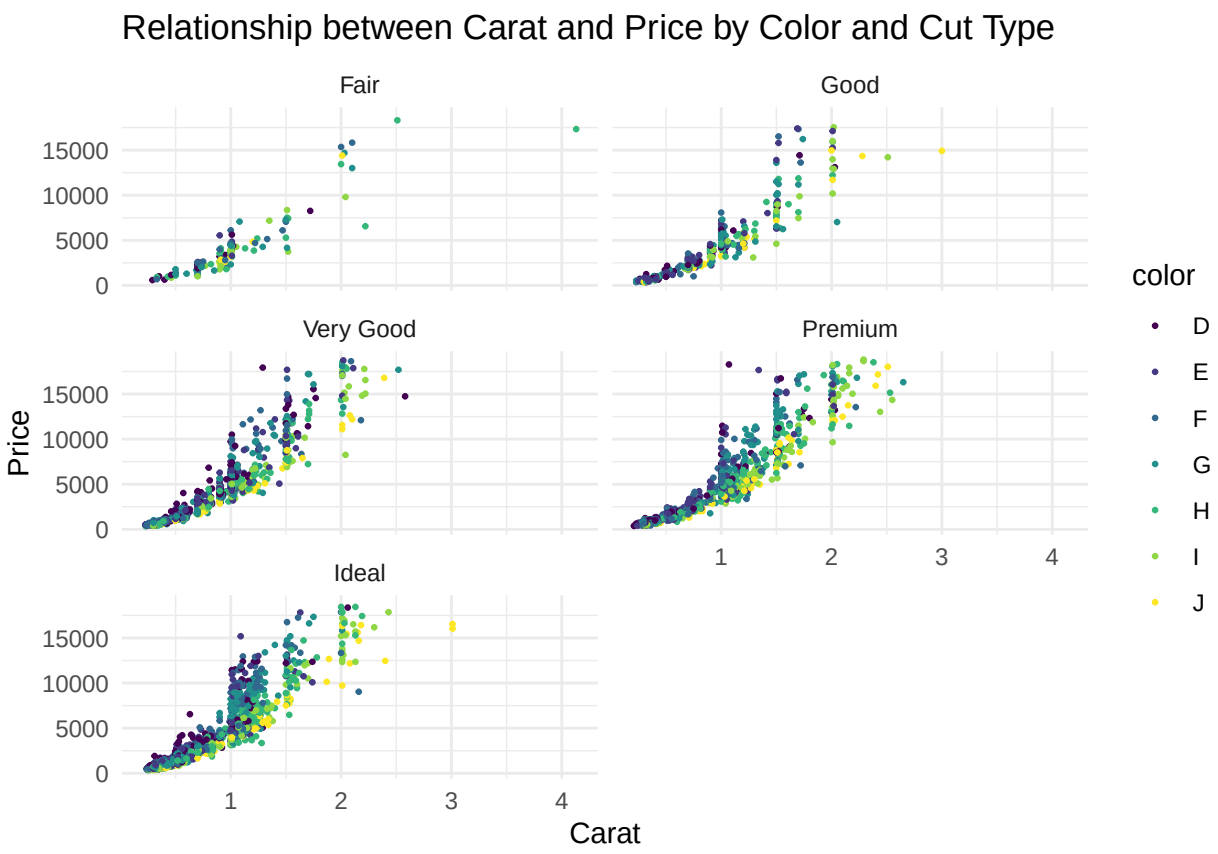
Meaning:

Color J has the highest median price, and color E has the lowest median price.

3 Relationship between Carat and Price by Color and Cut Type

```
small_diamonds <- sample_frac(diamonds, 0.08)

ggplot(small_diamonds, aes(carat, price, col=color)) +
  geom_point(size = 0.5, alpha = 1) +
  facet_wrap(~ cut, ncol = 2) +
  labs(title = "Relationship between Carat and Price by Color and Cut Type",
       x = "Carat",
       y = "Price") +
  theme_minimal()
```

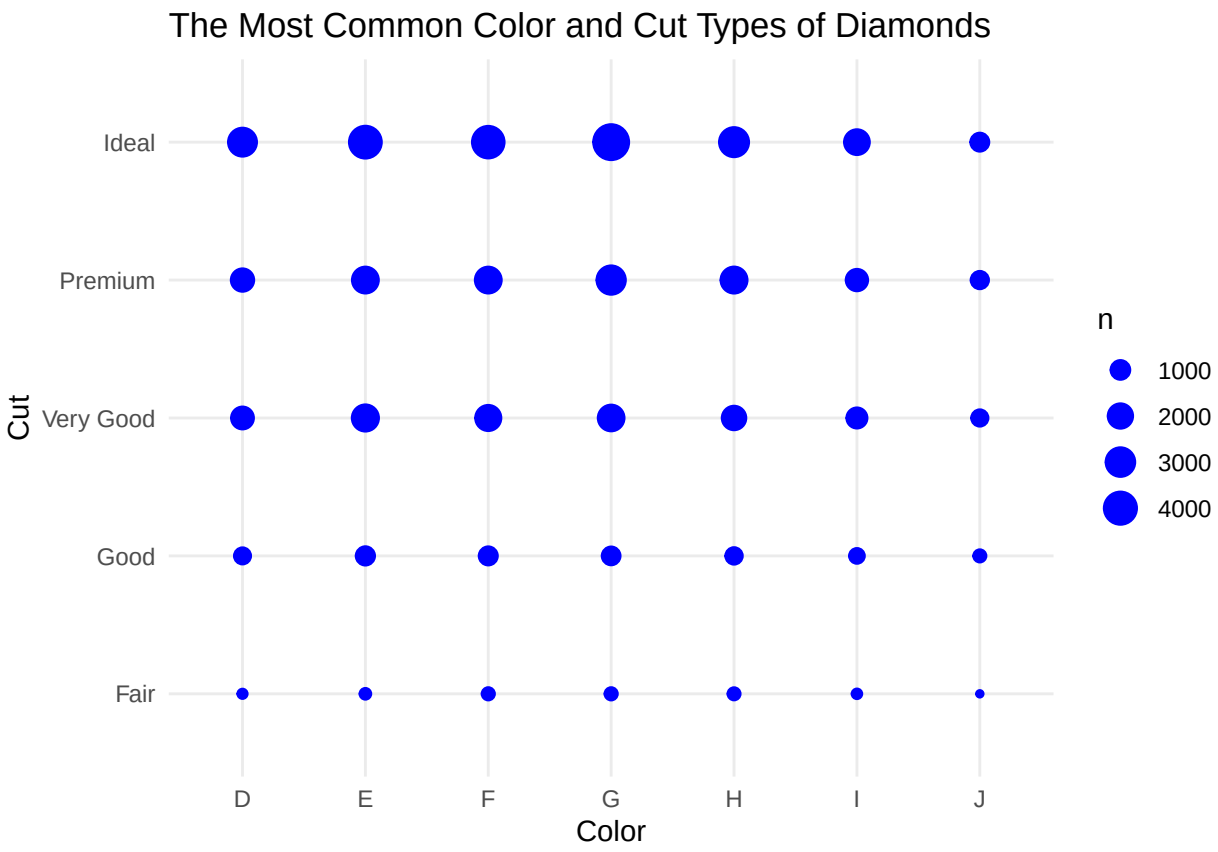


Meaning:

Prices vary according to carat. Color J tends to have a lower price than other colors at the same carat. Color D tends to have a higher price than other colors at the same carat.

4 The Most Common Color and Cut Types of Diamonds

```
ggplot(diamonds, aes(color, cut))+
   geom_count(color = "blue")+
   labs(title = "The Most Common Color and Cut Types of Diamonds",
        x = "Color",
        y = "Cut") +
   theme_minimal()
```

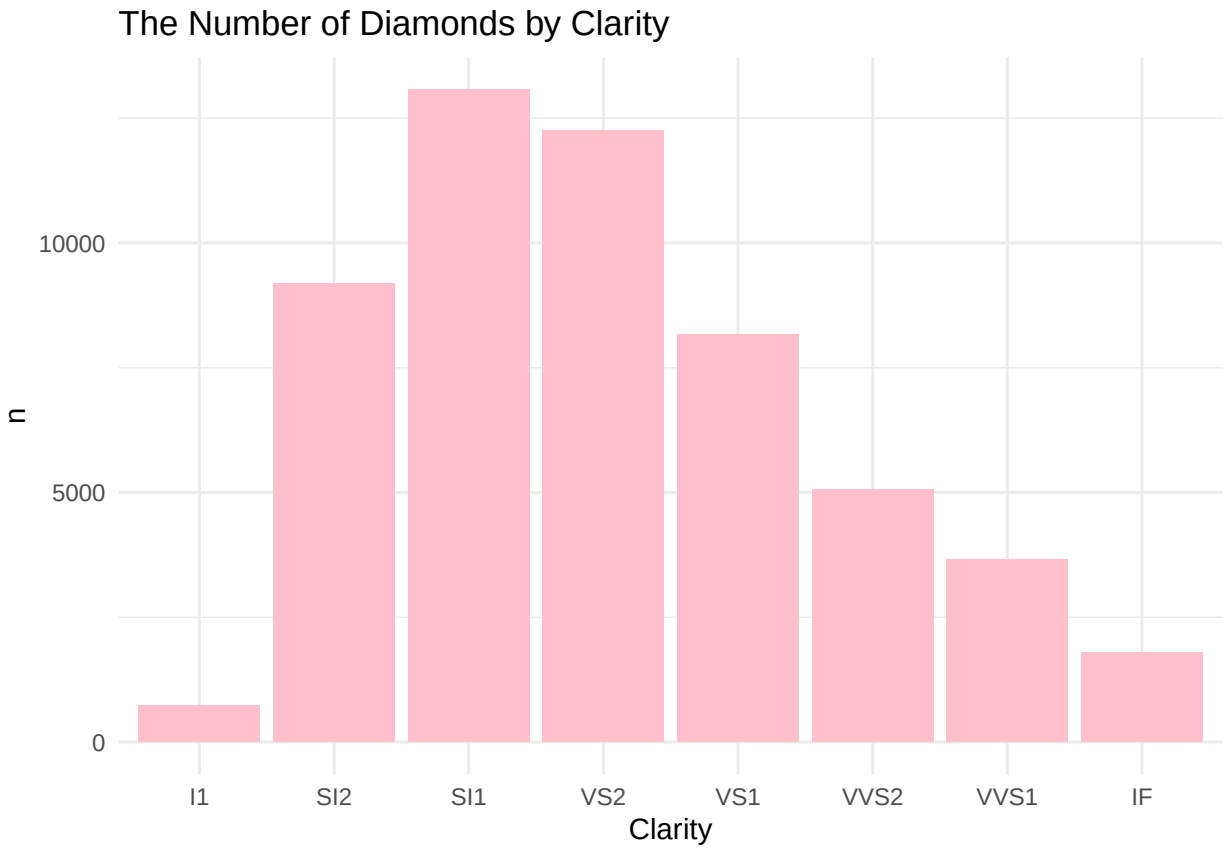


Meaning:

The group of ideal-cut-type and G-color diamonds has the largest number.

5 The Number of Diamonds by Clarity

```
ggplot(diamonds, aes(clarity))+
  geom_bar(fill = "pink")+
  labs(title = "The Number of Diamonds by Clarity",
       x = "Clarity",
       y = "n") +
  theme_minimal()
```



Meaning:

Clarity SI1 has the largest number, followed by VS2 and SI1. Clarity I1 has the smallest number.