

Data Visualization Bootcamp Homework

Naphon Seeluang

2023-07-01

Instruction

Use diamonds dataset to create 5 charts. knit pdf and submit in discord.

```
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()
```

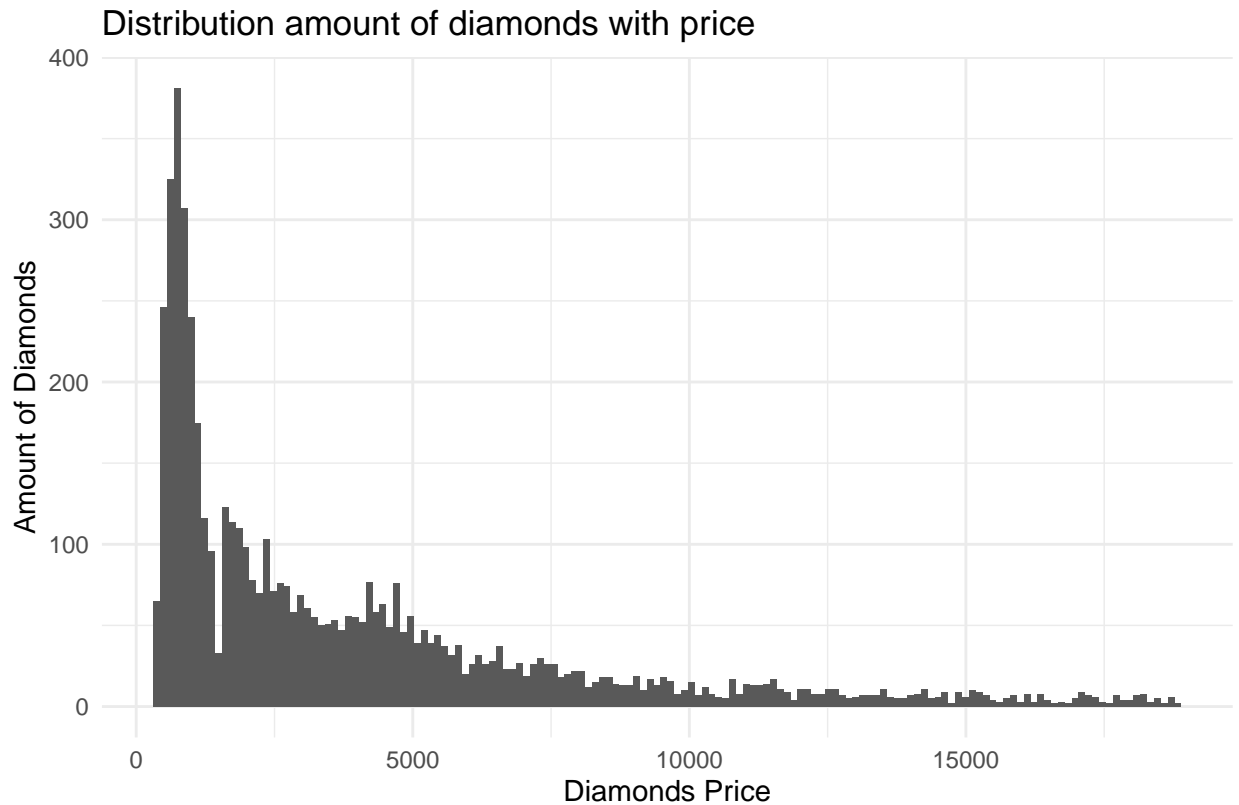
```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
set.seed(44)
```

```
mini_diamonds <- sample_frac(diamonds, 0.1)
```

1. Distribution amount of diamonds with price

```
ggplot(mini_diamonds,
  aes(price)) +
  geom_histogram(bins = 150) +
  theme_minimal() +
  labs(title = "Distribution amount of diamonds with price",
    x = "Diamonds Price",
    y = "Amount of Diamonds",
    caption = "Source: R Studio")
```



Source: R Studio

```
mean(diamonds$price)
```

```
## [1] 3932.8
```

```
median(diamonds$price)
```

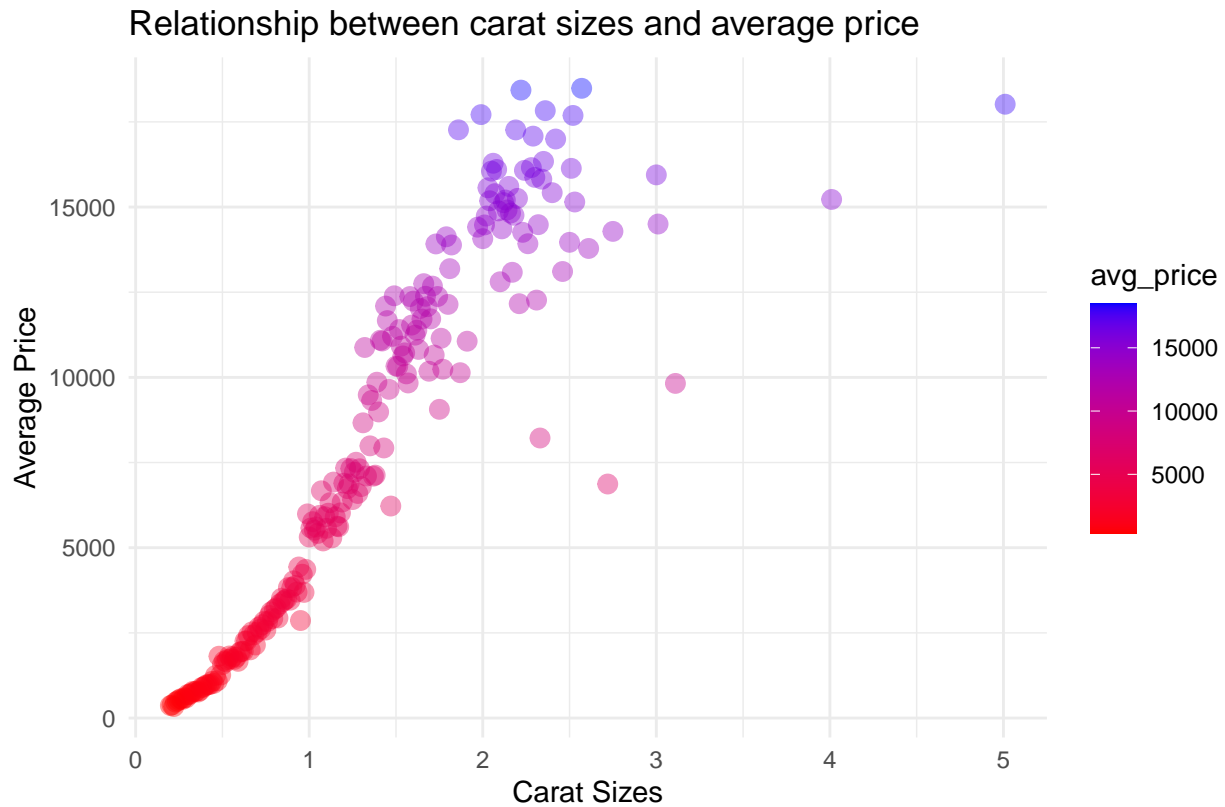
```
## [1] 2401
```

Amount of diamonds is right-skewed which mean that the expensive diamond have less than the cheap diamond.

Average price is 3943.819 and Median price is 2401.

2. Relationship between carat sizes and average price

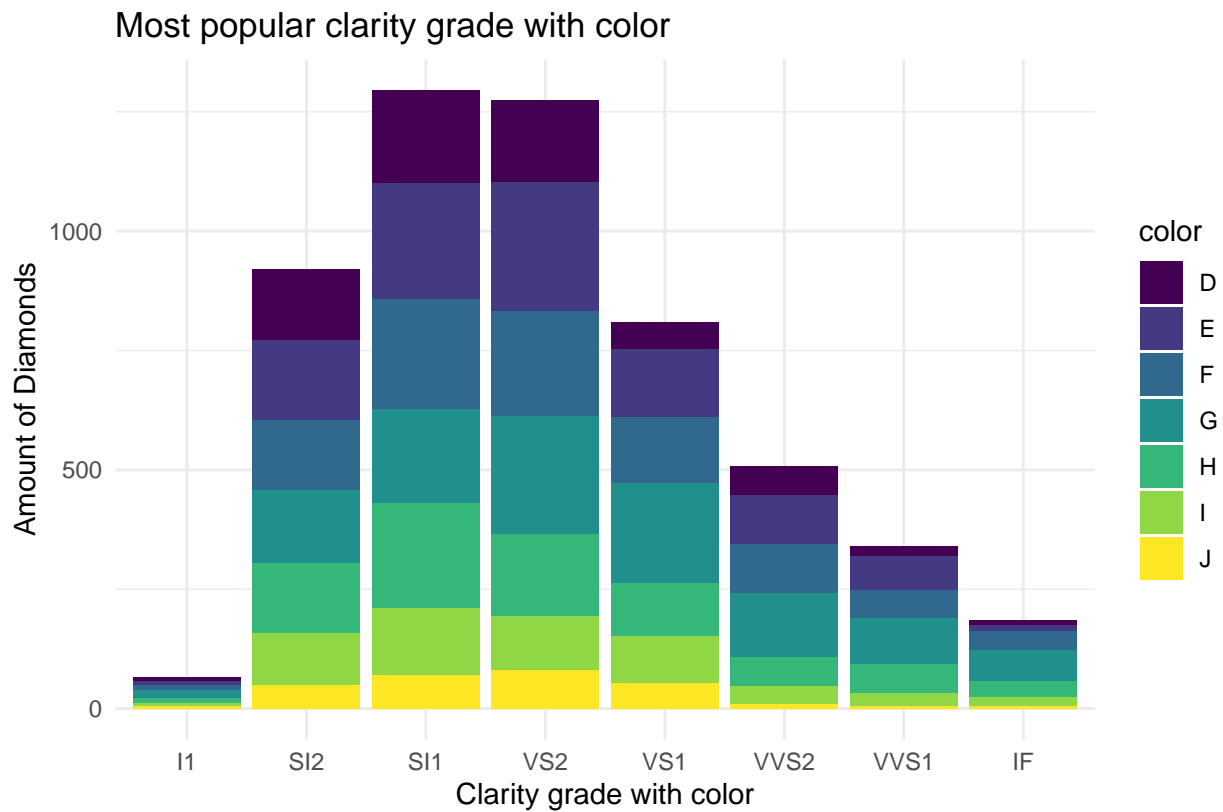
```
mini_diamonds %>%  
  group_by(carat) %>%  
  summarise(avg_price = mean(price)) %>%  
  ggplot(aes(carat, avg_price, col=avg_price)) +  
  geom_point(size=3, alpha=0.4) +  
  theme_minimal() +  
  scale_color_gradient(low="red", high="blue") +  
  labs(title = "Relationship between carat sizes and average price",  
       x = "Carat Sizes",  
       y = "Average Price",  
       caption = "Source: R Studio")
```



Size of carat affect to price. You can observe from this chart the average price will start to increase from 1 carat on wards.

3. Most popular clarity grade with color

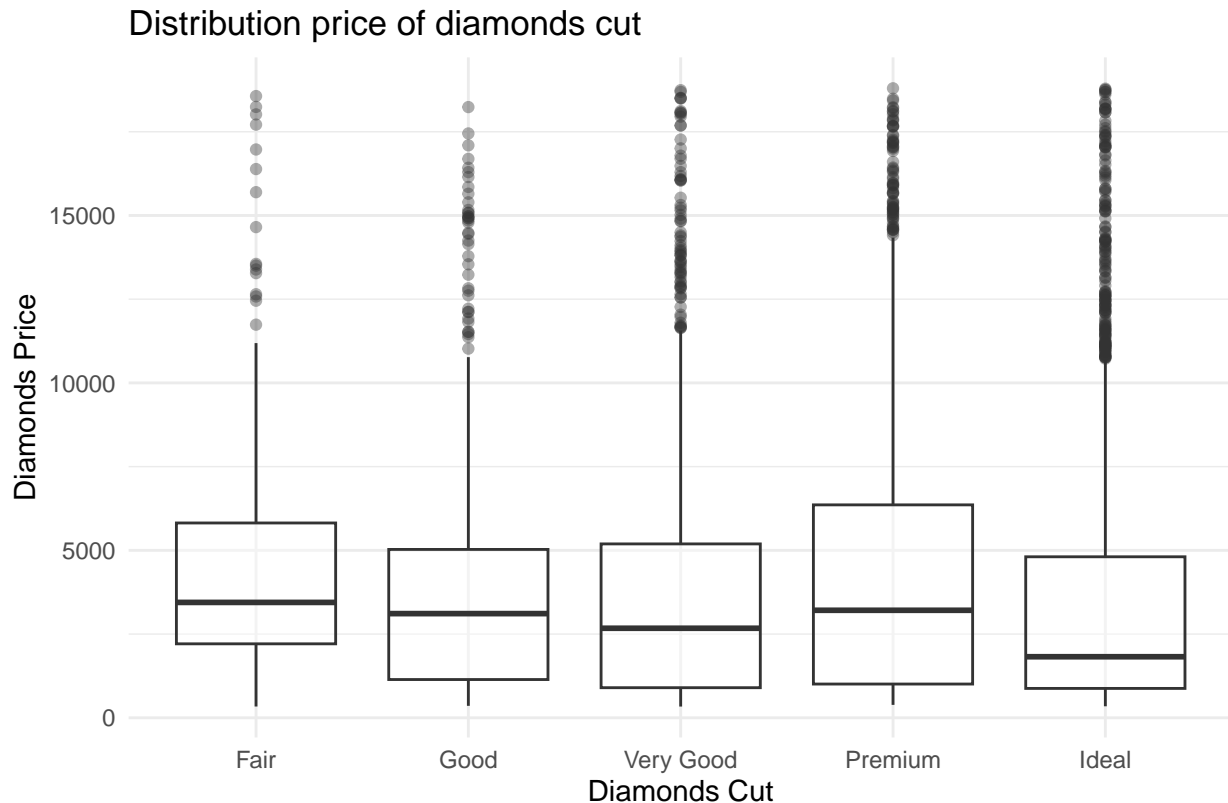
```
ggplot(mini_diamonds, aes(clarity, fill=color)) +
  geom_bar(position = "stack") +
  theme_minimal() +
  labs(title = "Most popular clarity grade with color",
       x = "Clarity grade with color",
       y = "Amount of Diamonds",
       caption = "Source: R Studio")
```



This chart show popular clarity that people bought which SI1 and VS2 are a famous grade and the most people like color E with these clarity.

4. Distribution price of diamonds cut

```
ggplot(mini_diamonds, aes(cut, price)) +
  geom_boxplot(alpha = 0.4) +
  theme_minimal() +
  labs(title = "Distribution price of diamonds cut",
       x = "Diamonds Cut",
       y = "Diamonds Price",
       caption = "Source: R Studio")
```



Source: R Studio

This chart show price distribution of diamonds cut and outliers of each cut which Ideal cut has a lot of outliers appear on the chart.

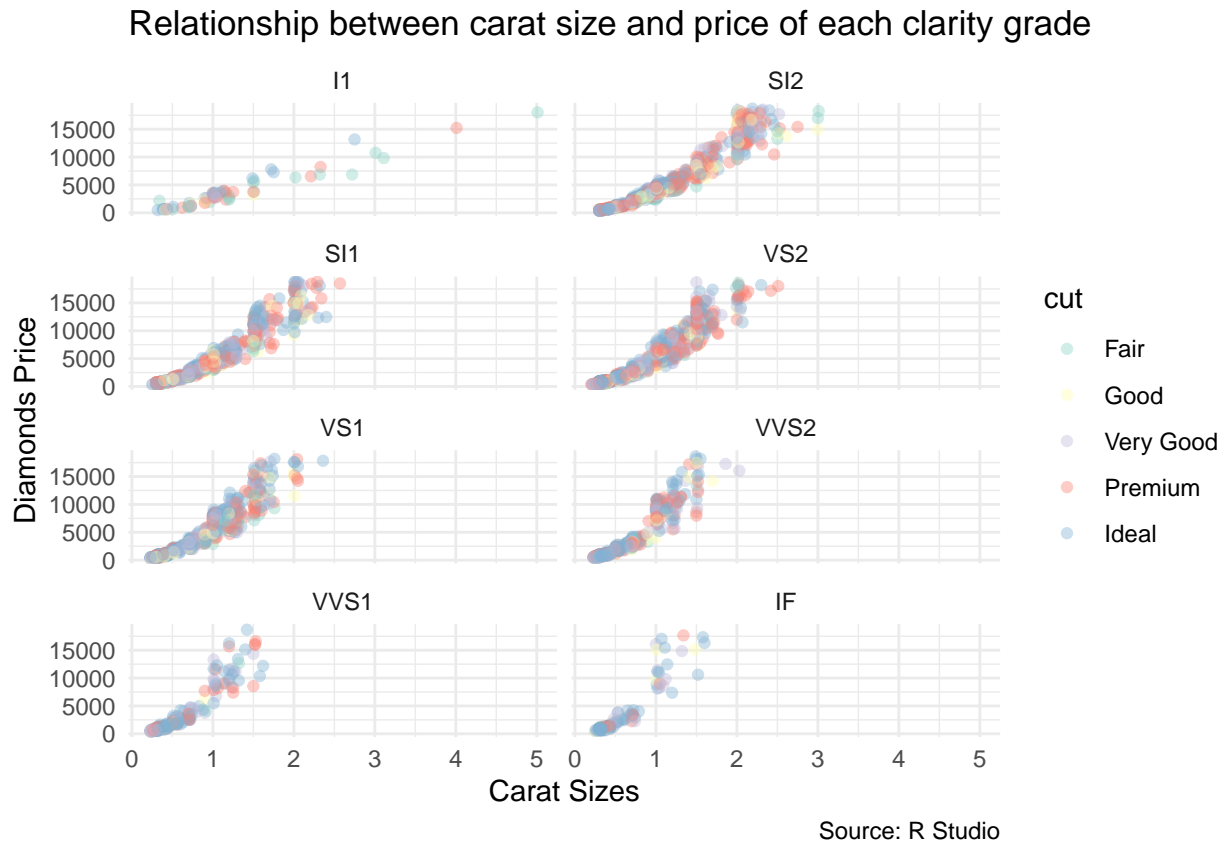
The price position table of each cut.

```
mini_diamonds %>%
  group_by(cut) %>%
  summarise(avg_price = mean(price),
            med_price = median(price),
            min_price = min(price),
            max_price = max(price),
            first_qr = quantile(price, 0.25),
            second_qr = quantile(price, 0.5),
            tird_qr = quantile(price, 0.75),
            n = n()) %>%
  arrange(-n)
```

```
## # A tibble: 5 x 9
##   cut    avg_price med_price min_price max_price first_qr second_qr tird_qr    n
##   <ord>    <dbl>    <dbl>    <int>    <int>    <dbl>    <dbl>    <dbl> <int>
## 1 Ideal     3494.     1824.     344    18787     878.     1824.     4808  2116
## 2 Prem~     4516.     3210      386    18797    1008     3210     6358.  1419
## 3 Very~     3908.     2674      336    18741     899     2674     5192.  1163
## 4 Good      3998.     3111      357    18236    1143.     3111     5027.   534
## 5 Fair      4887.     3446.      337    18565    2209.     3446.     5818.   162
```

5. Relationship between carat size and price of each clarity grade

```
ggplot(mini_diamonds, aes(carat, price, col=cut)) +  
  geom_point(size=1.5, alpha=0.4) +  
  theme_minimal() +  
  scale_color_brewer(type="qual", palette=8) +  
  facet_wrap(~clarity, ncol = 2) +  
  labs(title = "Relationship between carat size and price of each clarity grade",  
       x = "Carat Sizes",  
       y = "Diamonds Price",  
       caption = "Source: R Studio")
```



This chart shows carat sizes compared with price of each clarity grade by using cut color to separate for identifying price trend of each clarity grade with cut.

Done! Thank you.