

Data Visualization Homework

Anonymcat

2023-08-06

Data Visualization Homework

Instruction

Use Diamonds dataset to create 5 charts.

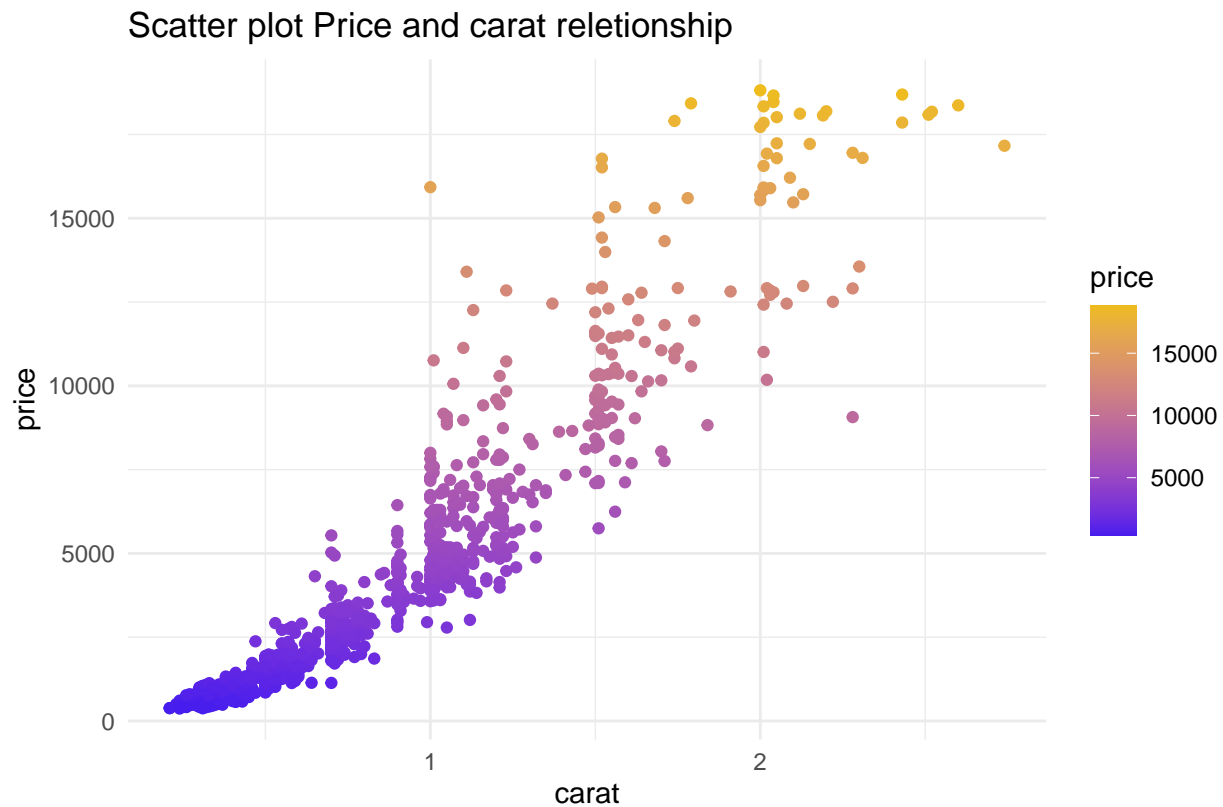
first, load library.

```
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
```

(No.1) Is the price depend on carat?

```
ggplot(sample_n(diamonds, 1000), aes(carat, price, col=price)) +
  geom_point() +
  theme_minimal() +
  scale_color_gradient(low = "#461DEE",
                      high = "#EEBC1D") +
  labs(title = "Scatter plot Price and carat relationship",
       caption = "Data Viz HW 01, - Anonymcat")
```



Data Viz HW 01, – Anonymcat

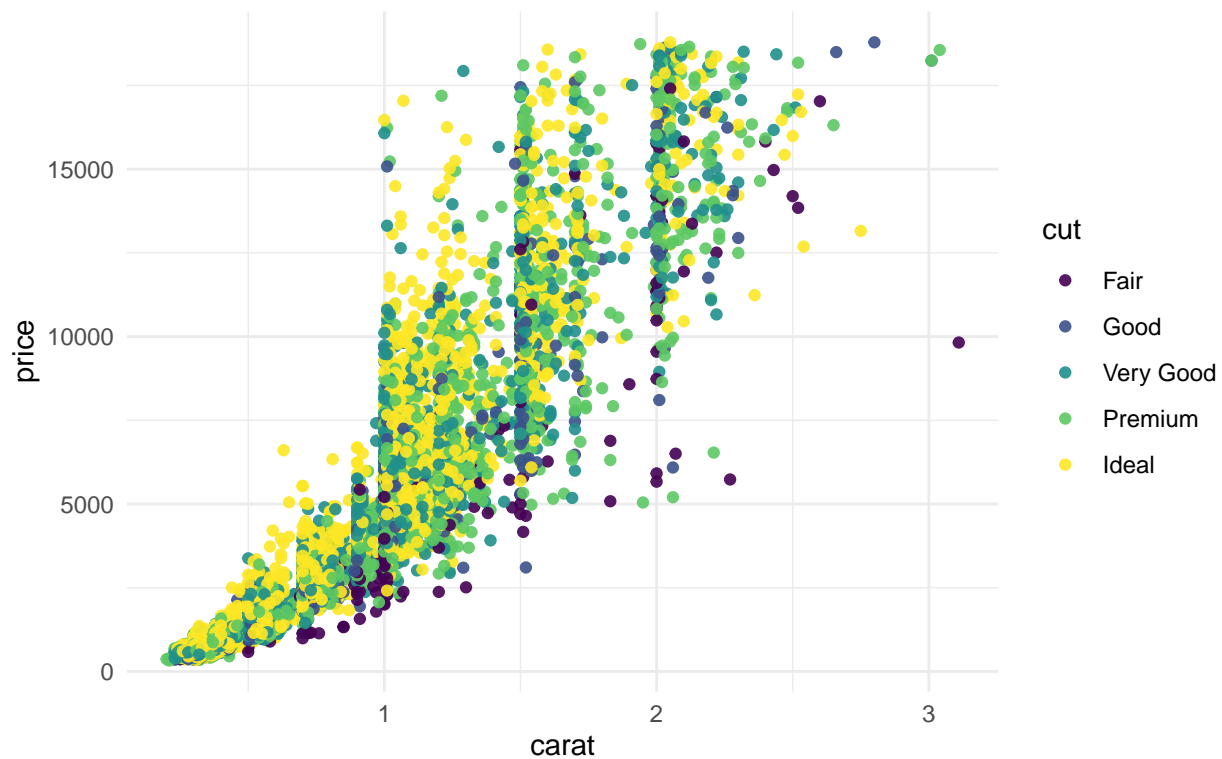
- the chart show that the price is depend on carat, high carat brings high price.
- but we can not assume that all high carat always brings high price because

(No.2)

```
set.seed(55)
mini_diamonds <- sample_n(diamonds,10000)

ggplot(mini_diamonds, aes(carat ,price, col=cut)) +
  geom_point(alpha=0.9) +
  theme_minimal() +
  labs( title = "Scatter plot relationship between prices and carat by Cut",
        caption = "Data Viz HW 01, – Anonymcat")
```

Scatter plot relationship between prices and carat by Cut



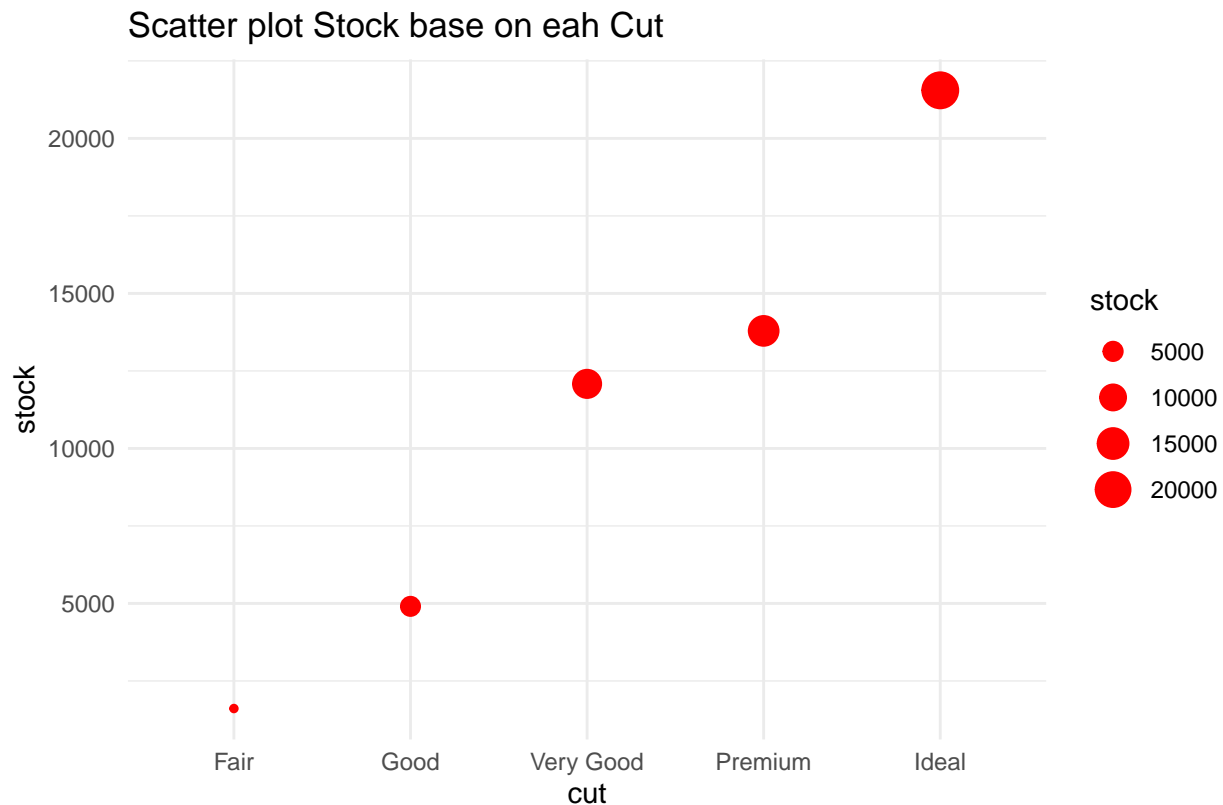
Data Viz HW 01, – Anonymcat

- As you see in Chart no.2, there are 3 diamonds that have 3 carat, but the price is different. why?
- because the price not only depend on carat but also depen on cut

(No.3) How many stock for each cut do we have? which cut we have the most?

```
cut_stock <- diamonds %>%
  group_by(cut) %>%
  summarise(
    stock = n())

ggplot(cut_stock, aes(cut, stock, size=stock))+
  geom_point( col="red" ) +
  theme_minimal() +
  labs( title = "Scatter plot Stock base on eah Cut",
        caption = "Data Viz HW 01, – Anonymcat")
```



Data Viz HW 01, – Anonymcat

- we have stock in total [53940].

```
sum(cut_stock$stock)
```

```
## [1] 53940
```

- by each cut we have...

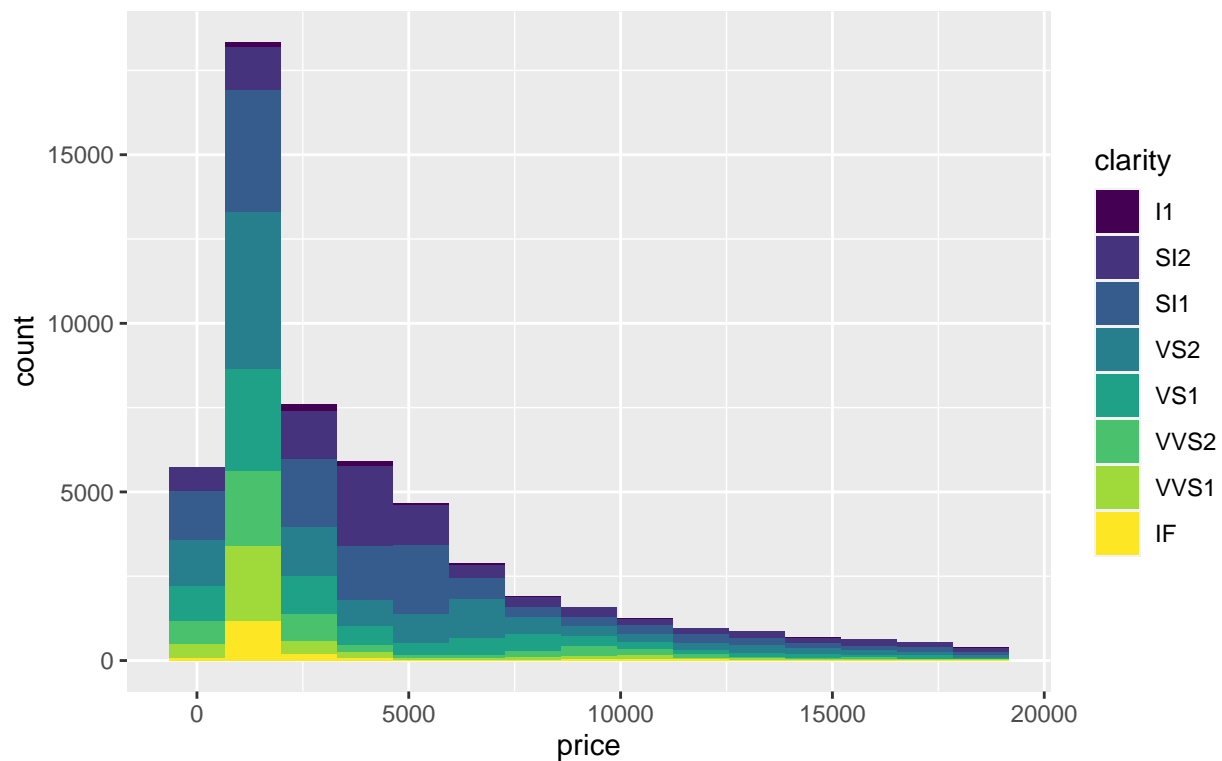
```
cut_stock
```

```
## # A tibble: 5 x 2
##   cut      stock
##   <ord>    <int>
## 1 Fair      1610
## 2 Good      4906
## 3 Very Good 12082
## 4 Premium   13791
## 5 Ideal     21551
```

(N0.4) Stock base on clarity and average price

```
ggplot(diamonds, aes(price, fill=clarity)) +
  geom_histogram(bins =15) +
  labs( title = "Histogram price and clarity",
        caption = "Data Viz HW 01, – Anonymcat")
```

Histogram price and clarity



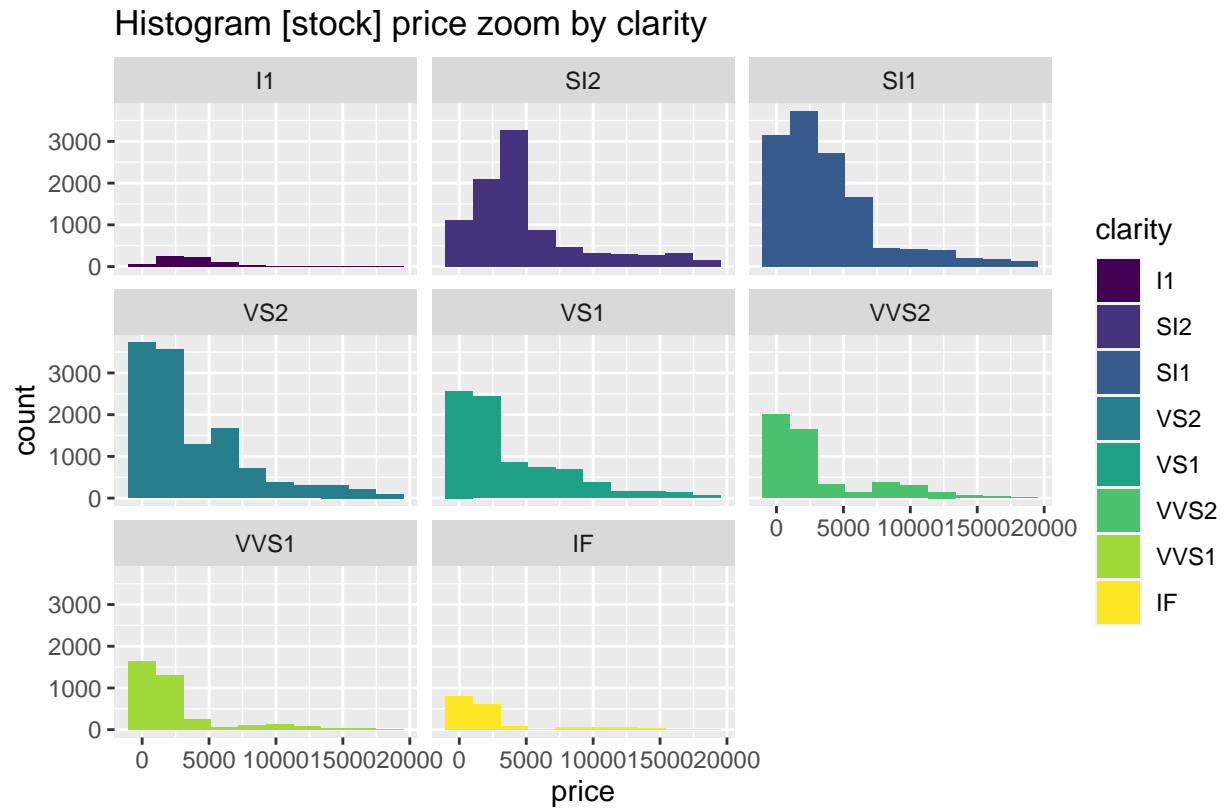
Data Viz HW 01, – Anonymcat

```
diamonds %>%
  summarise(
    mean(price)
  )
```

```
## # A tibble: 1 x 1
##   `mean(price)`
##   <dbl>
## 1       3933.
```

- the average price of diamonds in our stock is 3933, and the IF clarity is the cheap one

```
ggplot(diamonds, aes(price, fill=clarity)) +
  geom_histogram(bins = 10) +
  facet_wrap(~clarity) +
  labs( title = "Histogram [stock] price zoom by clarity",
        caption = "Data Viz HW 01, – Anonymcat")
```



Data Viz HW 01, – Anonymcat

(no.5) as the histogram above show that

I1 and IF. clarity is have less than 1000 in stock

Data Viz HW 01

Thanks