

Homework Data Viz Batch 10

Teerut Panalikul

2024-07-23

Load Library

```
library(tidyverse)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
##
## ✓ dplyr      1.1.4      ✓ readr      2.1.4
## ✓ forcats   1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2    3.4.4      ✓ tibble     3.2.1
## ✓ lubridate 1.9.3      ✓ tidyr      1.3.0
## ✓ purrr     1.0.2
## — Conflicts — tidyverse_conflicts() —
##
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggthemes)
```

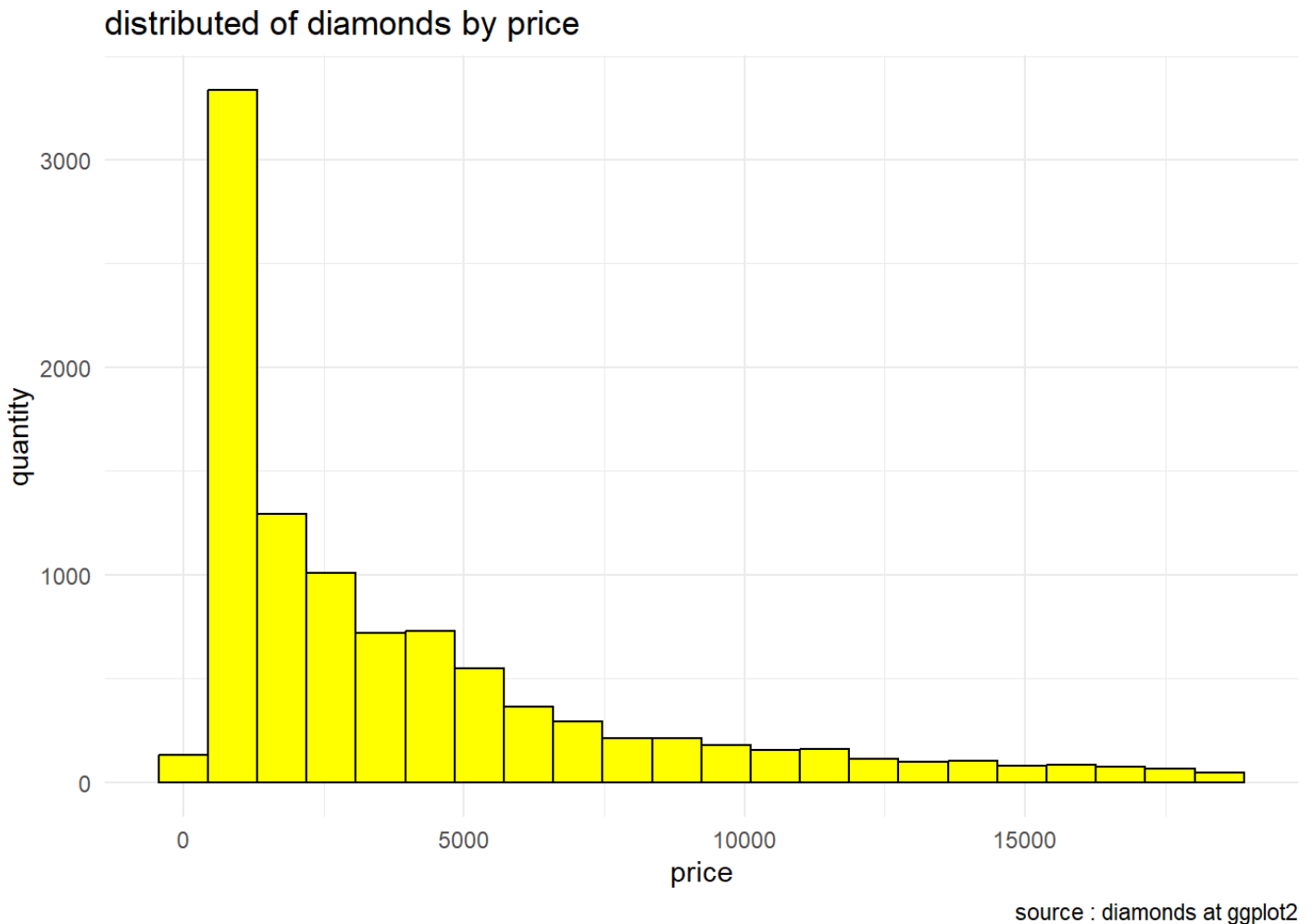
```
## Warning: package 'ggthemes' was built under R version 4.3.3
```

Prepare Data

```
set.seed(17)
small_df <- diamonds %>%
  sample_n(10000)
```

Distributed of Diamonds by Price

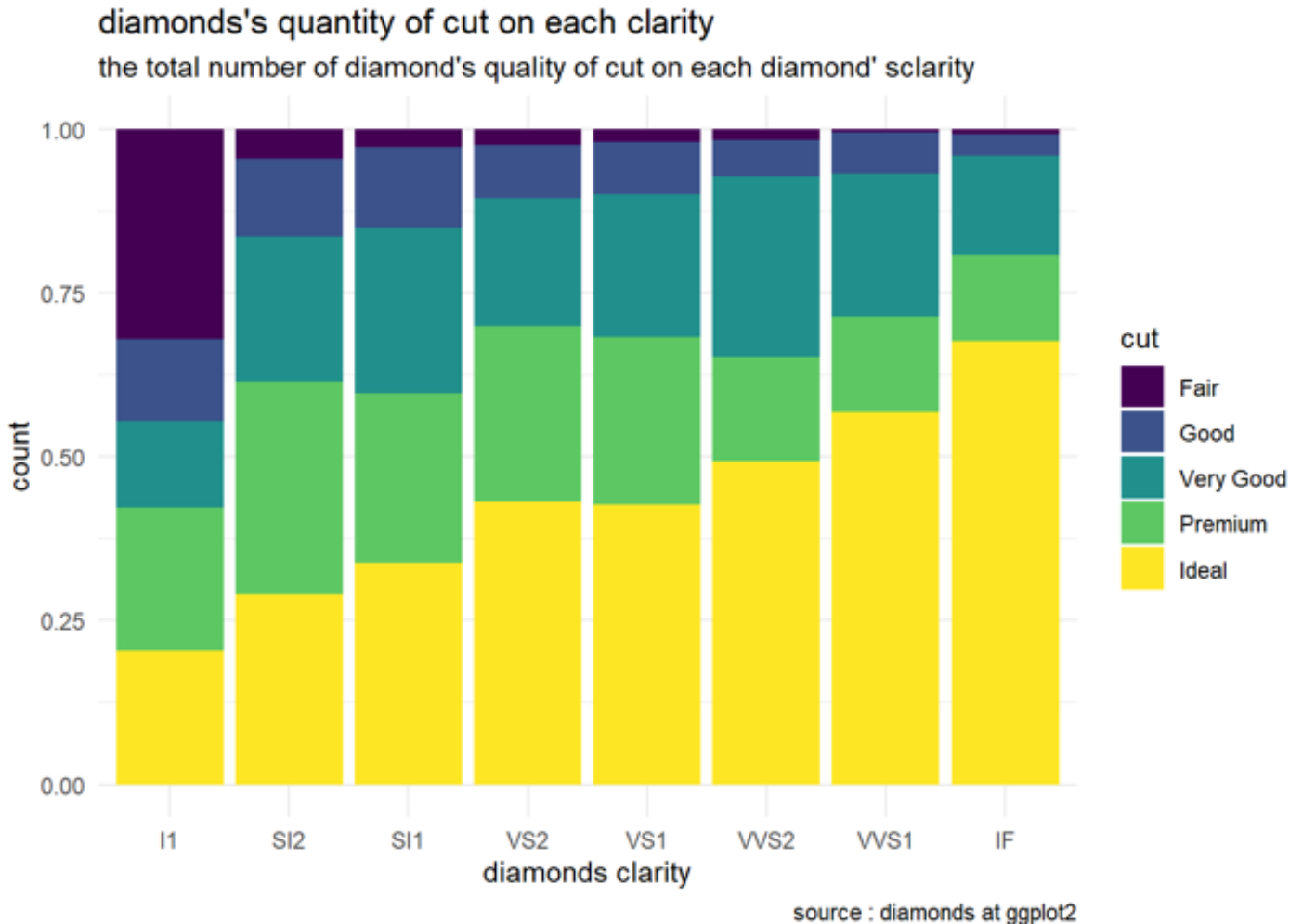
```
ggplot(small_df, aes(x=price)) +  
  geom_histogram(bins = 22, fill = "yellow", col = "black") +  
  labs(title = "distributed of diamonds by price ",  
        caption = "source : diamonds at ggplot2",  
        y = "quantity") +  
  theme_minimal()
```



from the histogram, it's found that the majority of price is around 1000 to 2000.

The total number of Diamond's Quality of Cut on each Diamond's Clarity

```
ggplot(small_df, aes(clarity, fill=cut)) +
  geom_bar(position = "fill") +
  labs(title = "diamonds's quantity of cut on each clarity",
        subtitle = "the total number of diamond's quality of cut on each diamond' s
clarity",
        caption = "source : diamonds at ggplot2",
        x = "diamonds clarity") +
  theme_minimal()
```

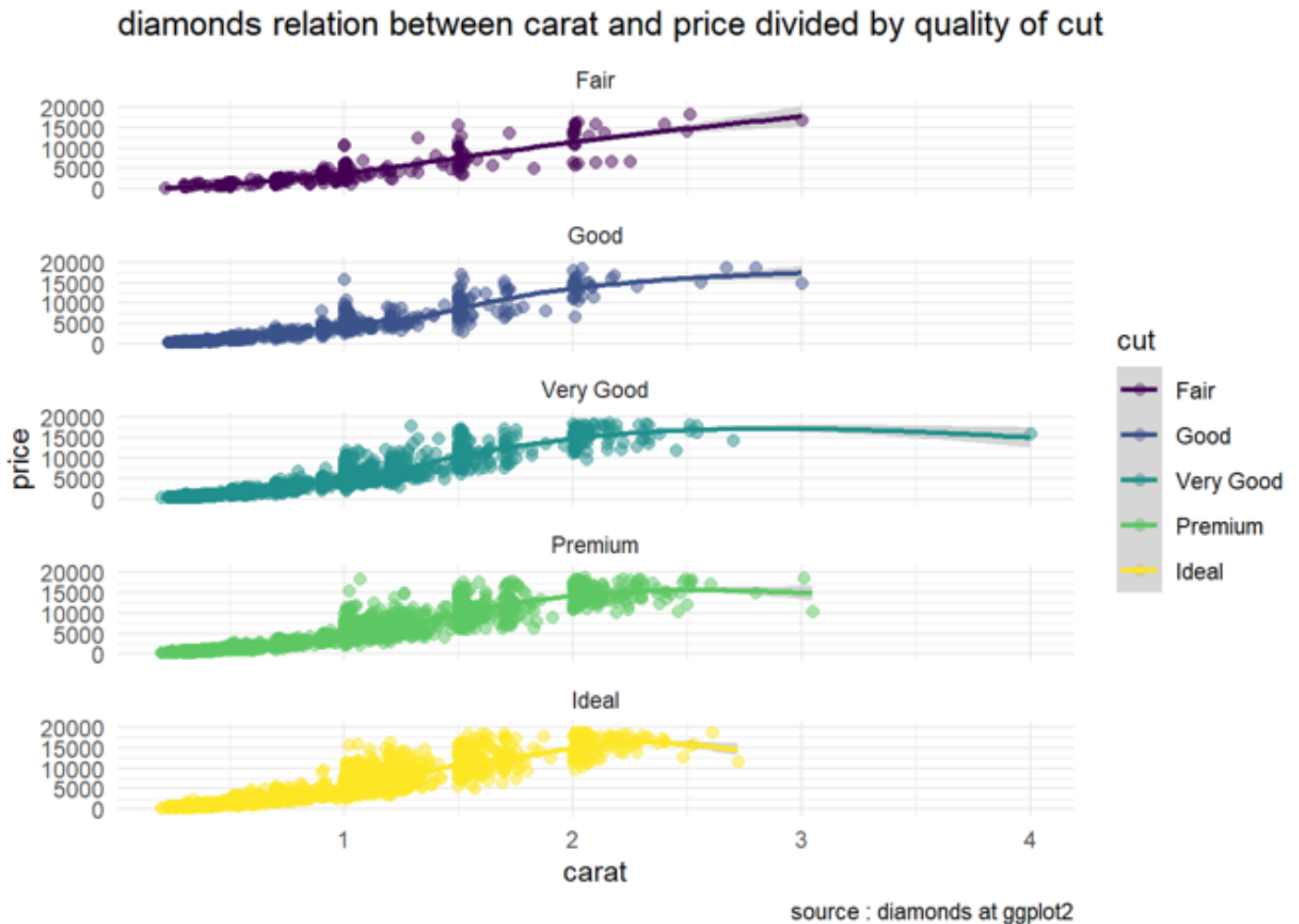


from the bar chart, it's found that the more diamonds clarity the amount of cut quality increase

Diamonds Relation between Carat and Price divided by Quality of Cut

```
ggplot(small_df, aes(carat, price, col=cut)) +
  geom_point(alpha = 0.5, size = 2) +
  geom_smooth() +
  facet_wrap(~cut, nrow = 5) +
  labs(title = "diamonds relation between carat and price divided by quality of cut",
        caption = "source : diamonds at ggplot2") +
  theme_minimal()
```

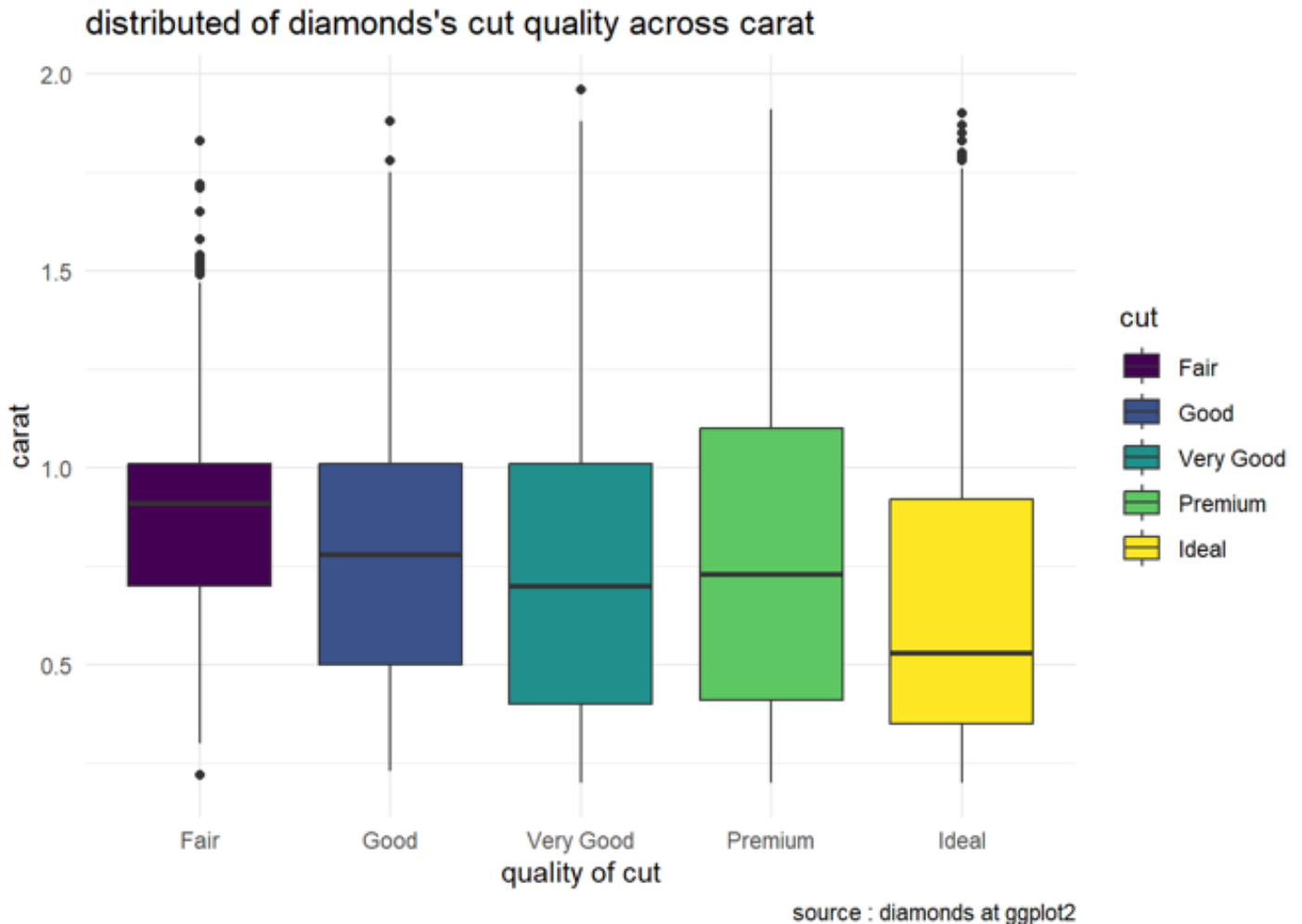
```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



from the scatter plot, it's found that price varies with carat in every cut quality

Distributed of Diamonds's Cut Quality across Carat

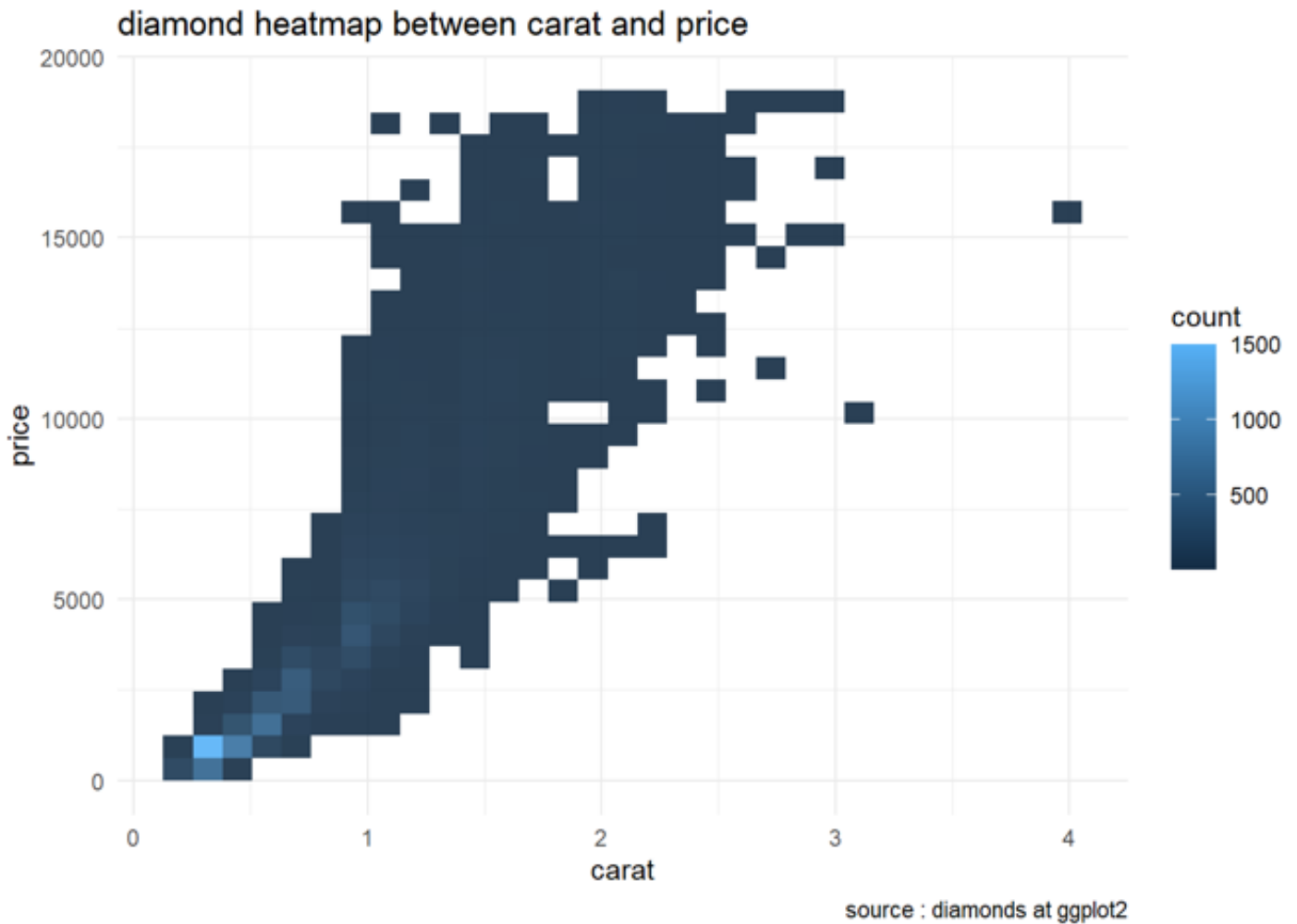
```
ggplot(small_df %>% filter(carat < 2),
       aes(cut, carat, fill = cut)) +
  geom_boxplot() +
  labs(title = "distributed of diamonds's cut quality across carat",
       caption = "source : diamonds at ggplot2",
       x = "quality of cut") +
  theme_minimal()
```



from box plot, it's found that the distribution of the lowest quality of cut is narrower than the other plots

Diamond Heatmap between Carat and Price

```
ggplot(small_df, aes(carat, price)) +
  geom_bin2d(alpha = 0.9) +
  labs(title = "diamond heatmap between carat and price",
       caption = "source : diamonds at ggplot2") +
  theme_minimal()
```



from heat map, it's found that the amount of price around 1000 and carat around 0.4 have the highest quantity