Diamonds in R package

Phiranat Nawachindaphan

2023-01-13

Load library

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(patchwork)
```

Load Data

```
data(diamonds)
```

Overview Data

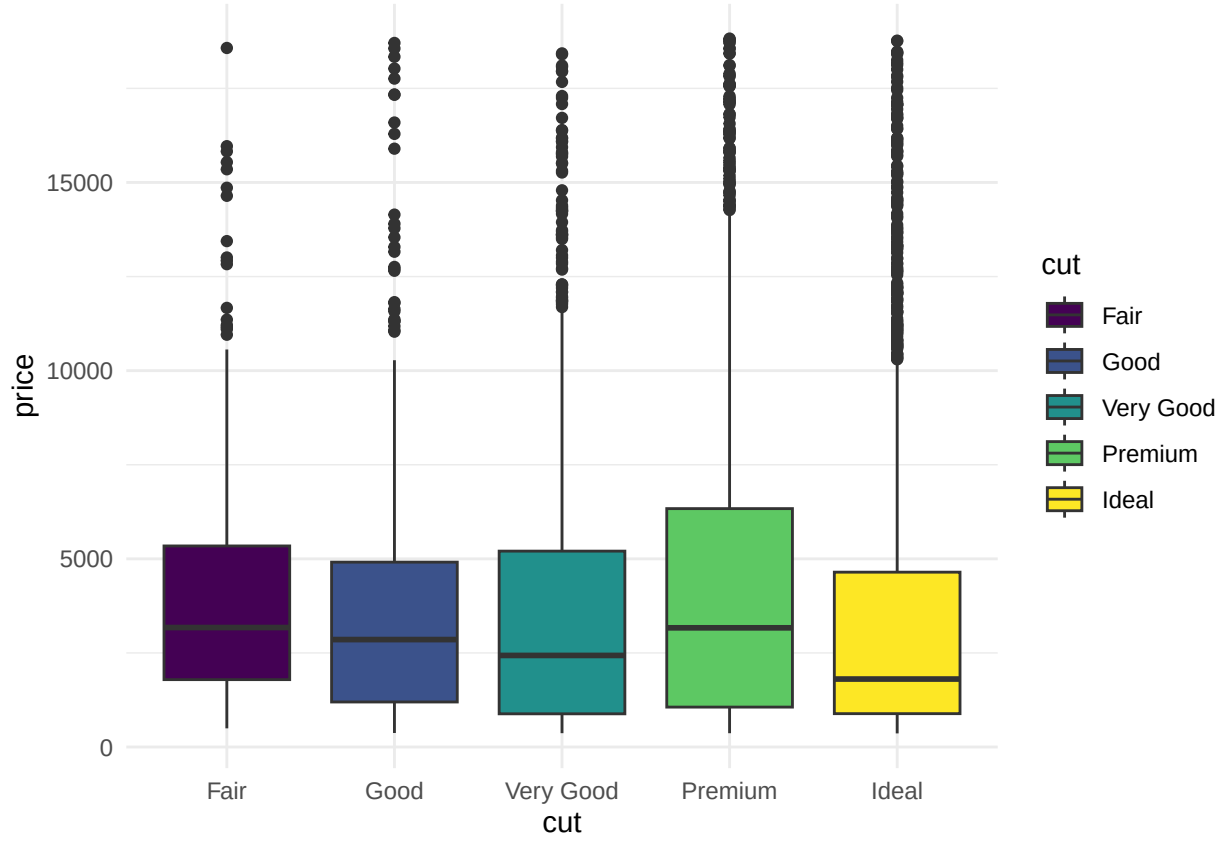
```
head(diamonds)
```

```
## # A tibble: 6 x 10
##   carat cut       color clarity depth table price     x     y     z
##   <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1  0.23 Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
## 2  0.21 Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
## 3  0.23 Good      E     VS1      56.9    65   327  4.05  4.07  2.31
## 4  0.29 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63
## 5  0.31 Good      J     SI2      63.3    58   335  4.34  4.35  2.75
## 6  0.24 Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48
```

A Dataset has 53,940 records sample 5000 records

```
diamonds_s <- sample_n(diamonds, 5000)
```

```
ggplot(diamonds_s, aes(cut, price, fill=cut)) +
  geom_boxplot() +
  theme_minimal()
```



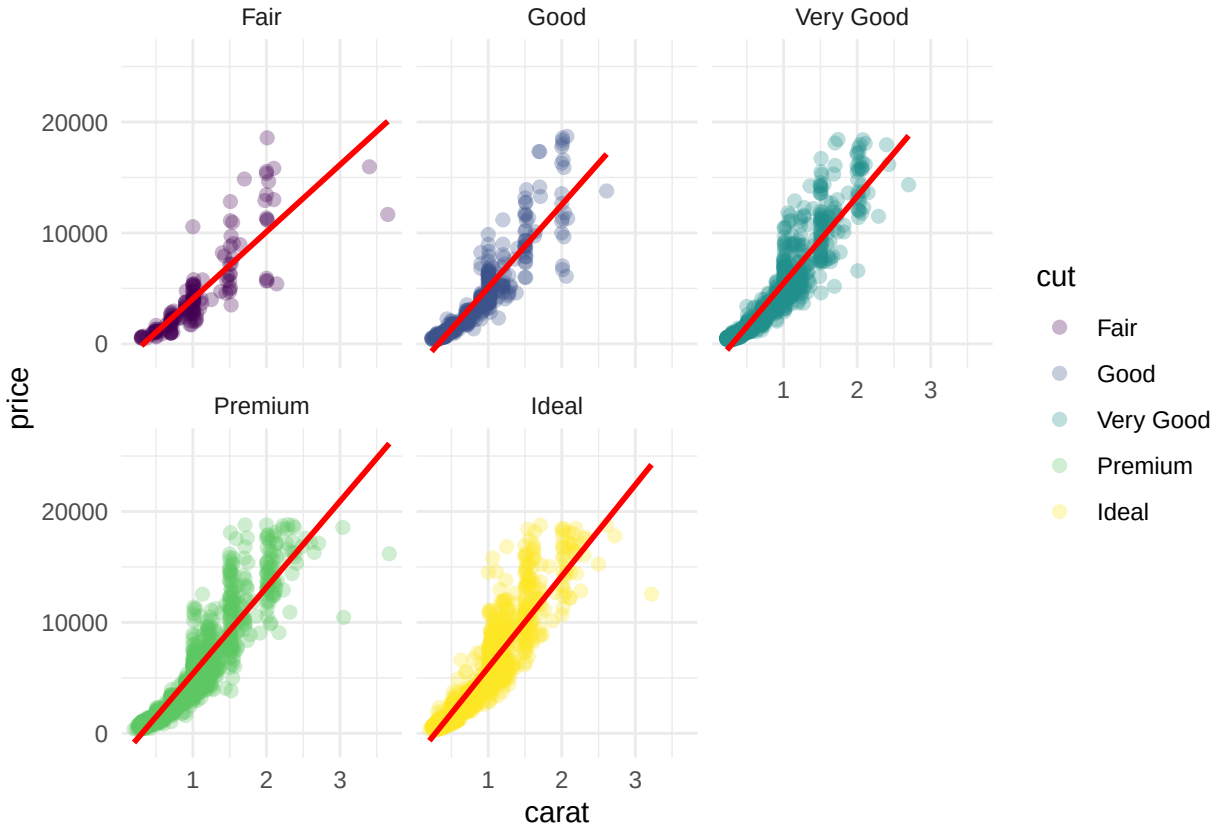

```
diamonds_s %>%
  group_by(cut) %>%
  summarise(Average_price = mean(price)) %>%
  arrange(desc(Average_price))
```

```
## # A tibble: 5 x 2
##   cut       Average_price
##   <ord>             <dbl>
## 1 Premium           4595.
## 2 Fair              4366.
## 3 Good              3871.
## 4 Very Good         3829.
## 5 Ideal             3427.
```

Summary cut type Fair has the most expensive in this group ,cut type Ideal has cheapest in this group

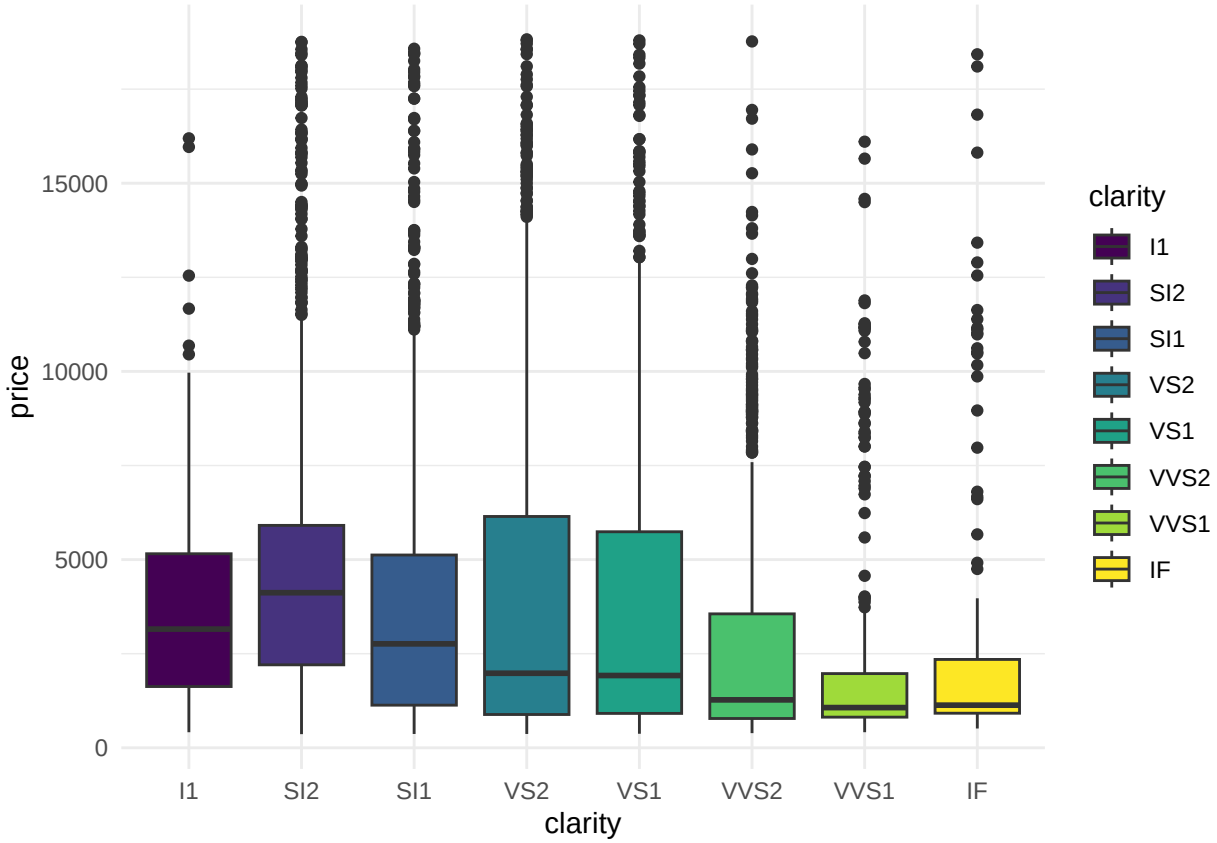
```
ggplot(diamonds_s, aes(carat, price, color=cut)) +
  geom_point(size=2, alpha=0.3) +
  theme_minimal() +
  geom_smooth(method="lm", color="red", se=F) +
  facet_wrap(~ cut, ncol = 3)
```

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



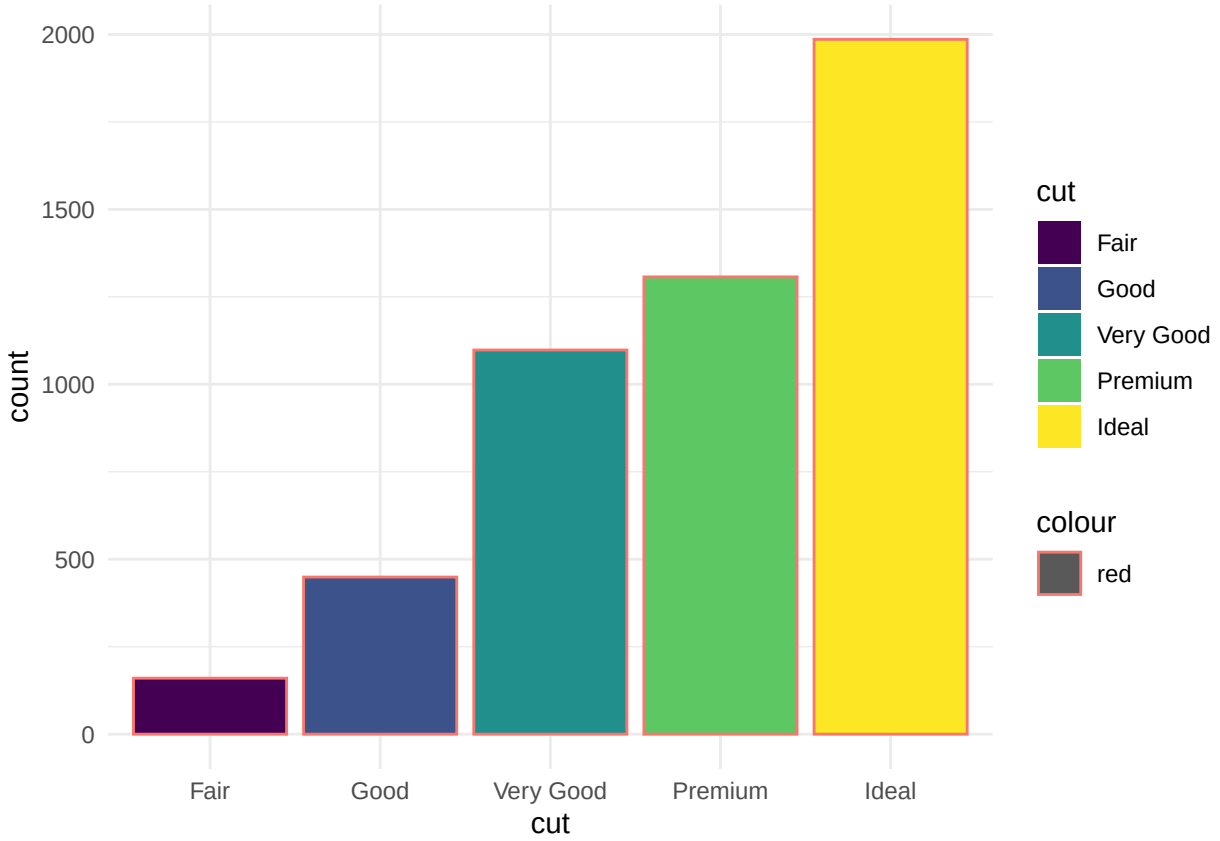
Summary carat increase, resulting in price increase (linear)

```
ggplot(diamonds_s, aes(clarity, price, fill=clarity)) +
  geom_boxplot() +
  theme_minimal()
```



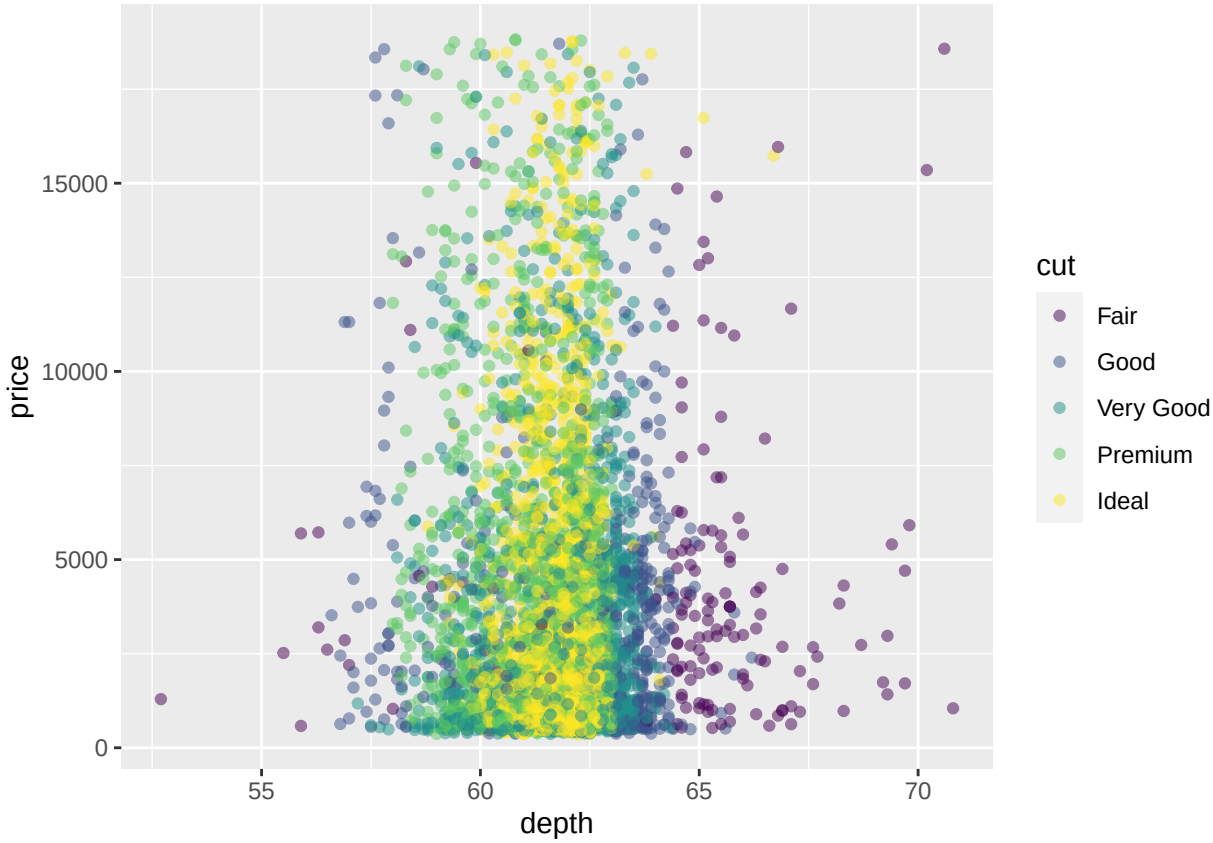
Summary clarity VS2 and VS1 have the most of range in this group

```
ggplot(diamonds_s, aes(cut, fill=cut, color="red")) +
  geom_bar() +
  theme_minimal()
```



Summary cut type ideal has the most in this group

```
ggplot(diamonds_s, aes(depth, price, color = cut)) +
  geom_point(alpha = 0.5)
```



Summary depth of diamonds **don't effect** to price