

HW-Sprint-05—Data-Visualization.R

r1617576

2023-01-02

```
#install packages
install.packages(c("tidyverse",
                  "patchwork",
                  "lubridate"))

## Installing packages into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)

#load library
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --

## v ggplot2 3.4.0      v purrr 1.0.0
## 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)
library(lubridate)

## Loading required package: timechange
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union

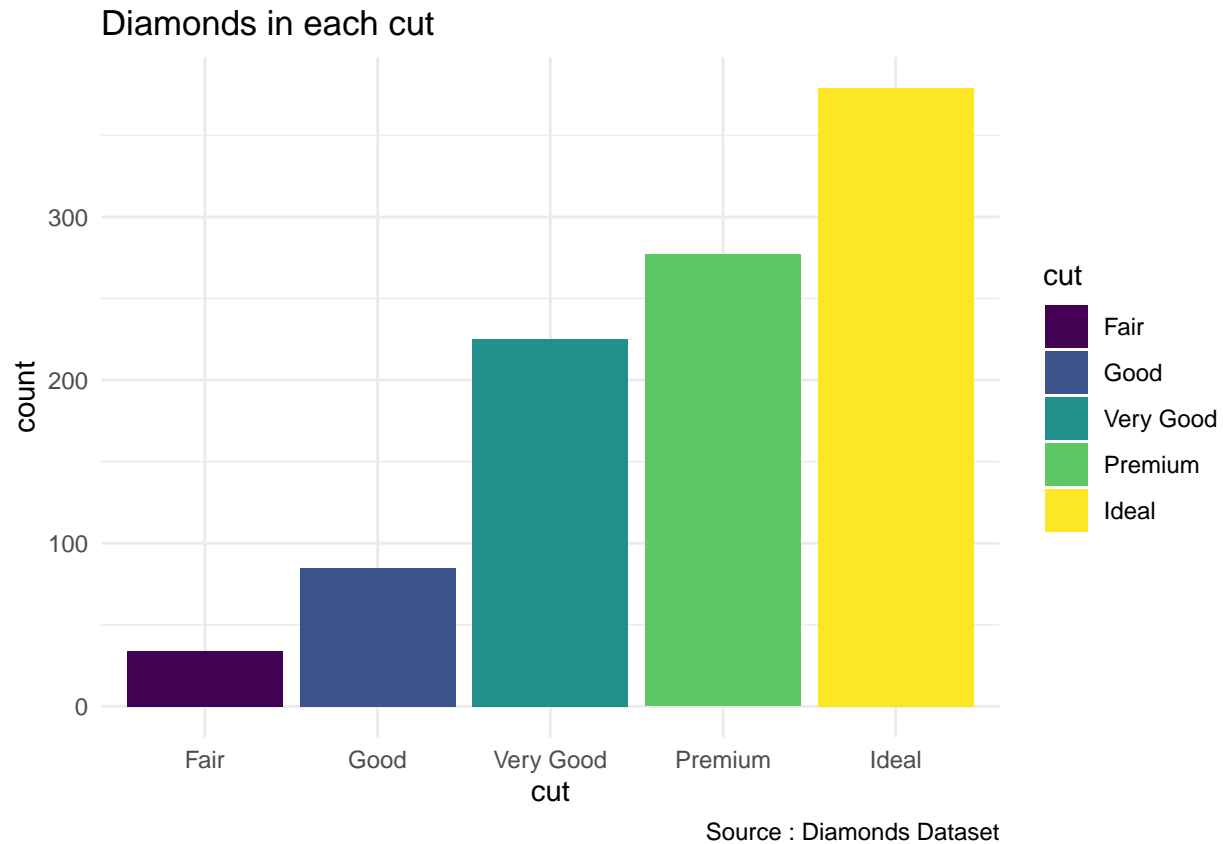
#glimpse diamonds
glimpse(diamonds)

## Rows: 53,940
## Columns: 10
## $ carat   <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, 0.23, 0.~
## $ cut     <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very Good, Ver~
## $ color   <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, J, J, I,~
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, VS1, SI1, VS1, ~
## $ depth   <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, 65.1, 59.4, 64~
## $ table   <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, 61, 54, 62, 58~
## $ price   <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338, 339, 340, 34~
## $ x       <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, 3.87, 4.00, 4.~
## $ y       <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, 3.78, 4.05, 4.~
```

```
## $ z      <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, 2.49, 2.39, 2.~
```

```
#1.Count of diamonds in each cut
```

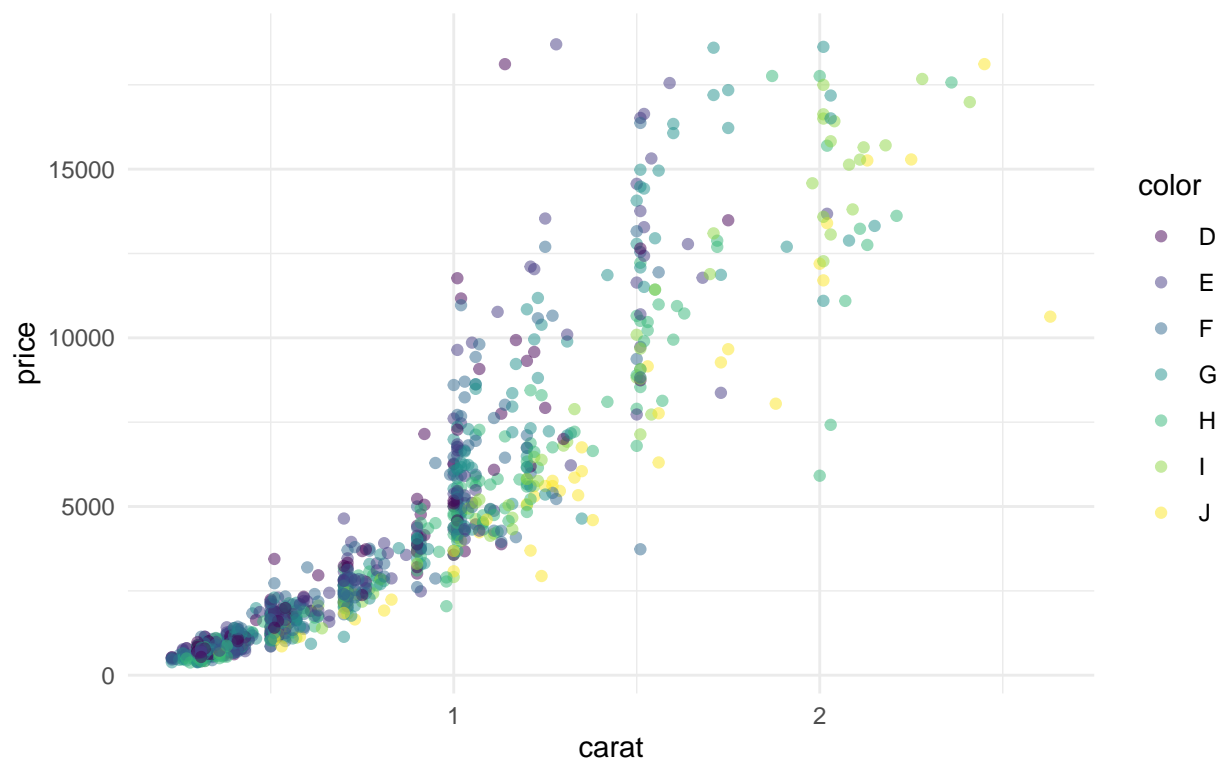
```
ggplot(sample_n(diamonds,1000),aes(cut, fill = cut))+  
  geom_bar()+  
  labs(title = "Diamonds in each cut",  
        caption = "Source : Diamonds Dataset")+  
  theme_minimal()
```



```
#2.Relationship between caret and price
```

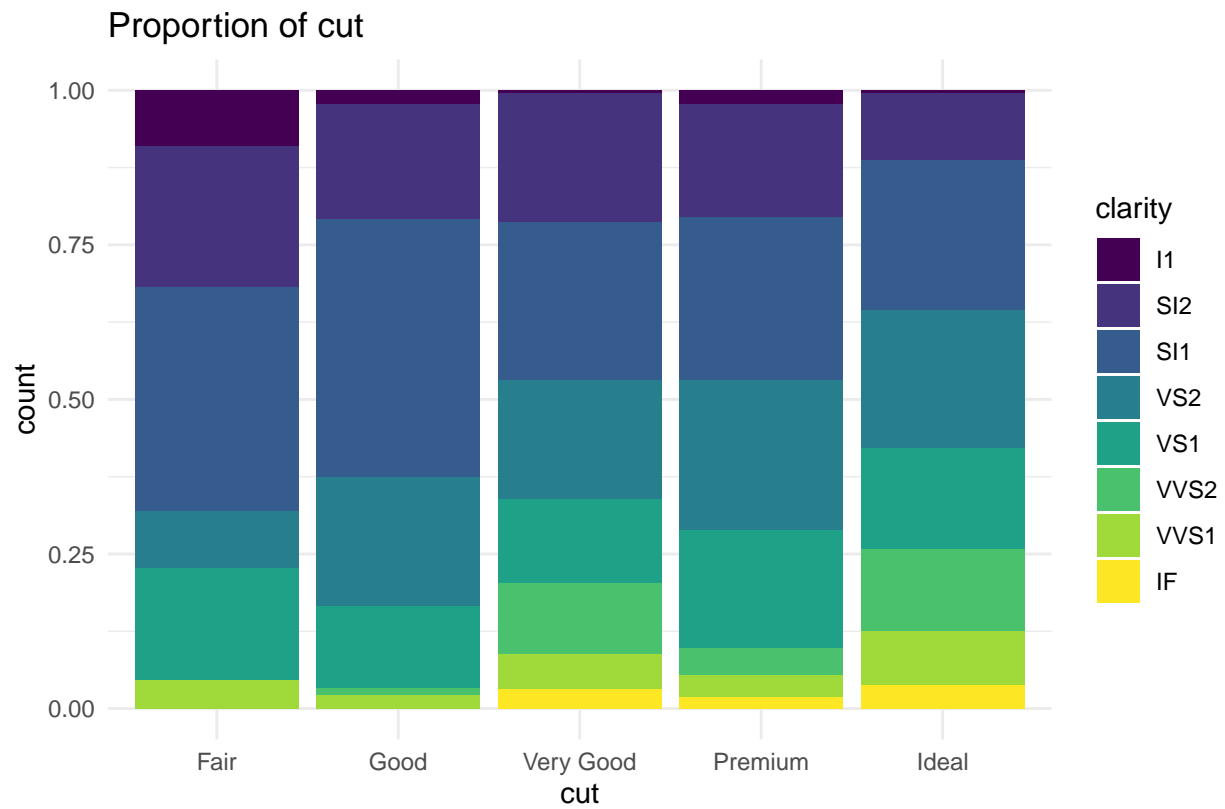
```
set.seed(24)  
ggplot(sample_n(diamonds,1000),  
  aes(carat,price,color=color))+  
  geom_point(alpha = 0.5)+  
  labs(title ="Relationship between carat and price",  
        caption ="Source : Diamonds Dataset")+  
  theme_minimal()
```

Relationship between carat and price



Source : Diamonds Dataset

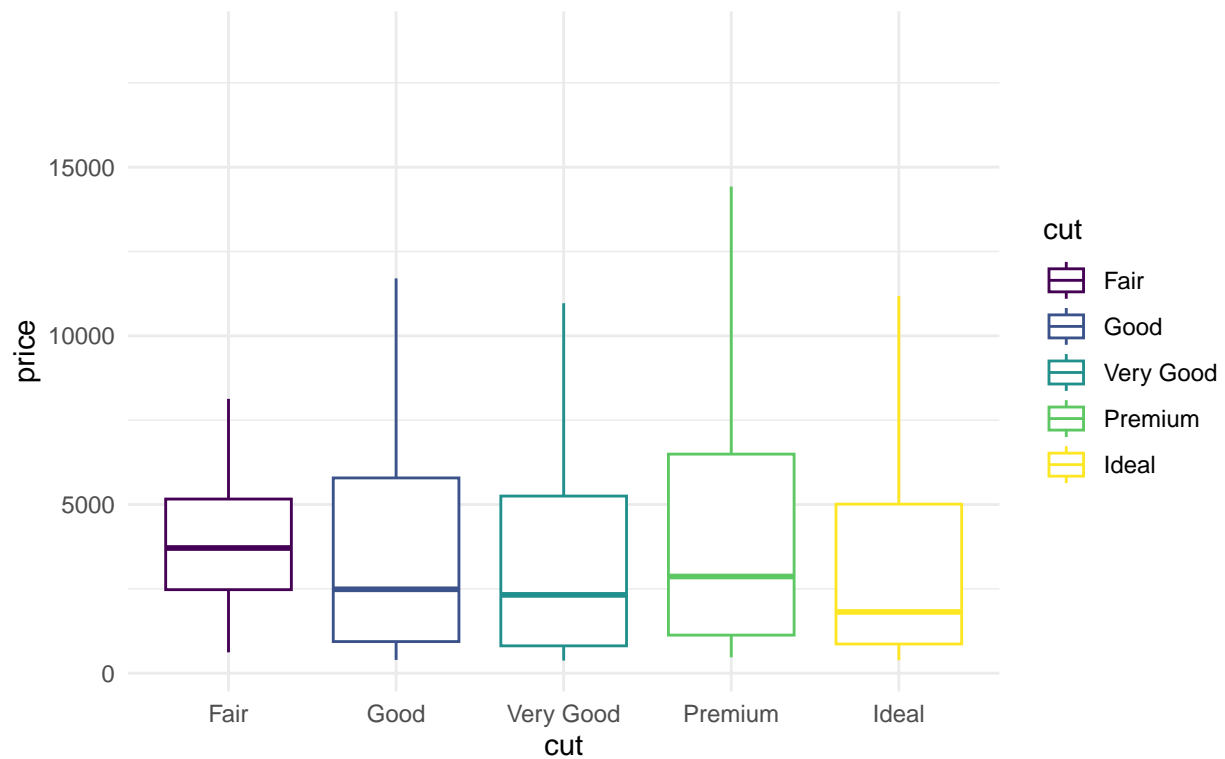
```
#3.Proportion of cut
set.seed(24)
ggplot(sample_n(diamonds,1000),
  aes(cut, fill = clarity))+
  geom_bar(position = "fill")+
  labs(title = "Proportion of cut",
    caption = "Source : Diamonds Dataset")+
  theme_minimal()
```



Source : Diamonds Dataset

```
#4.Relationship between cut and price
set.seed(24)
ggplot(sample_n(diamonds,1000),aes(cut, price, col = cut))+
  geom_boxplot(outlier.shape = NA)+
  labs(title ="Relationship between cut and price",
        caption ="Source : Diamonds Dataset")+
  theme_minimal()
```

Relationship between cut and price

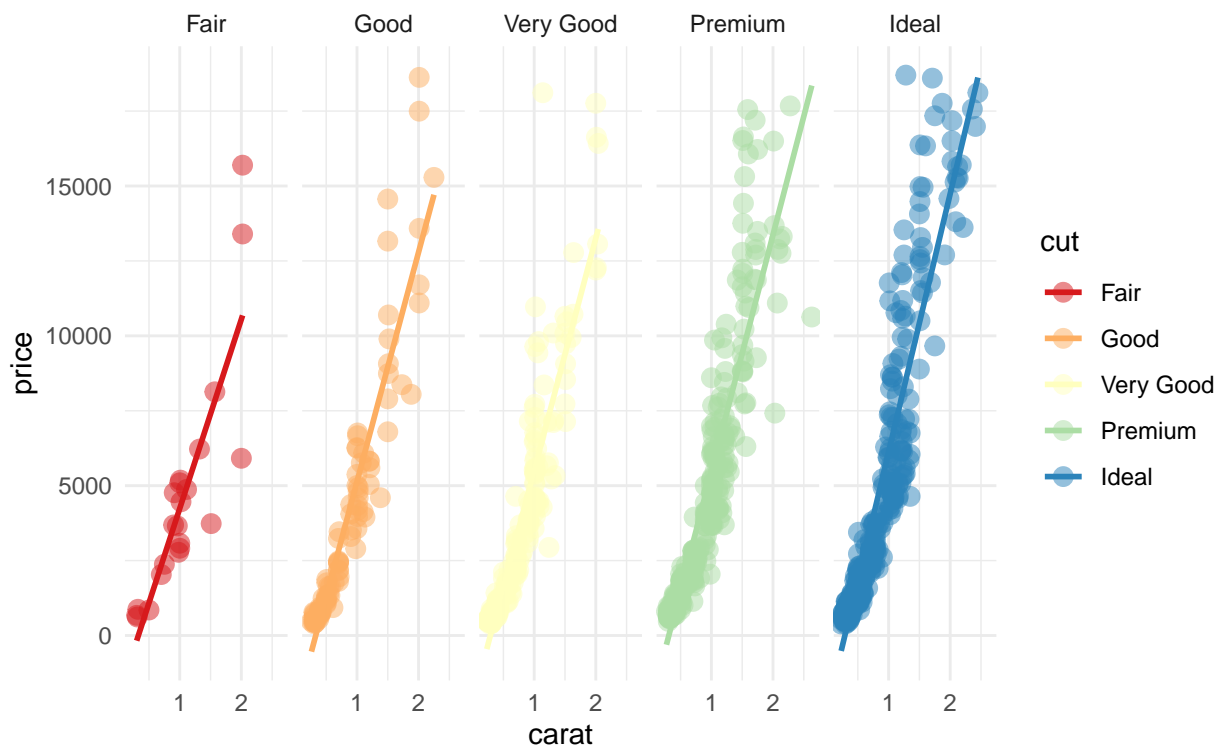


Source : Diamonds Dataset

```
#5.Relationship between carat and price by cut
set.seed(24)
ggplot(sample_n(diamonds,1000),aes(carat,price,color= cut))+
  geom_point(size = 3,alpha =0.5)+
  geom_smooth(method = "lm", se = F)+
  labs(title ="Relationship between carat and price by cut",
        caption ="Source : Diamonds Dataset")+
  theme_minimal()+
  facet_wrap(~cut, ncol = 5)+
  scale_color_brewer(palette = "Spectral")
```

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

Relationship between carat and price by cut



Source : Diamonds Dataset