Explore data with ggplot in R programing

using data from diamonds dataset in R package

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)

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.~
```

Chart 1 Show relationship between price and carat

```
set.seed(99)
ggplot(sample_n(diamonds, 5000), aes(price, carat))+
  geom_point(size = 1,col = "#02006c")+
  geom_smooth(col = "#c30101")+
  theme_minimal()+
  labs(title = "Relationship between price and carat",
       x = "price (USD)",
       y = "carat")
```

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

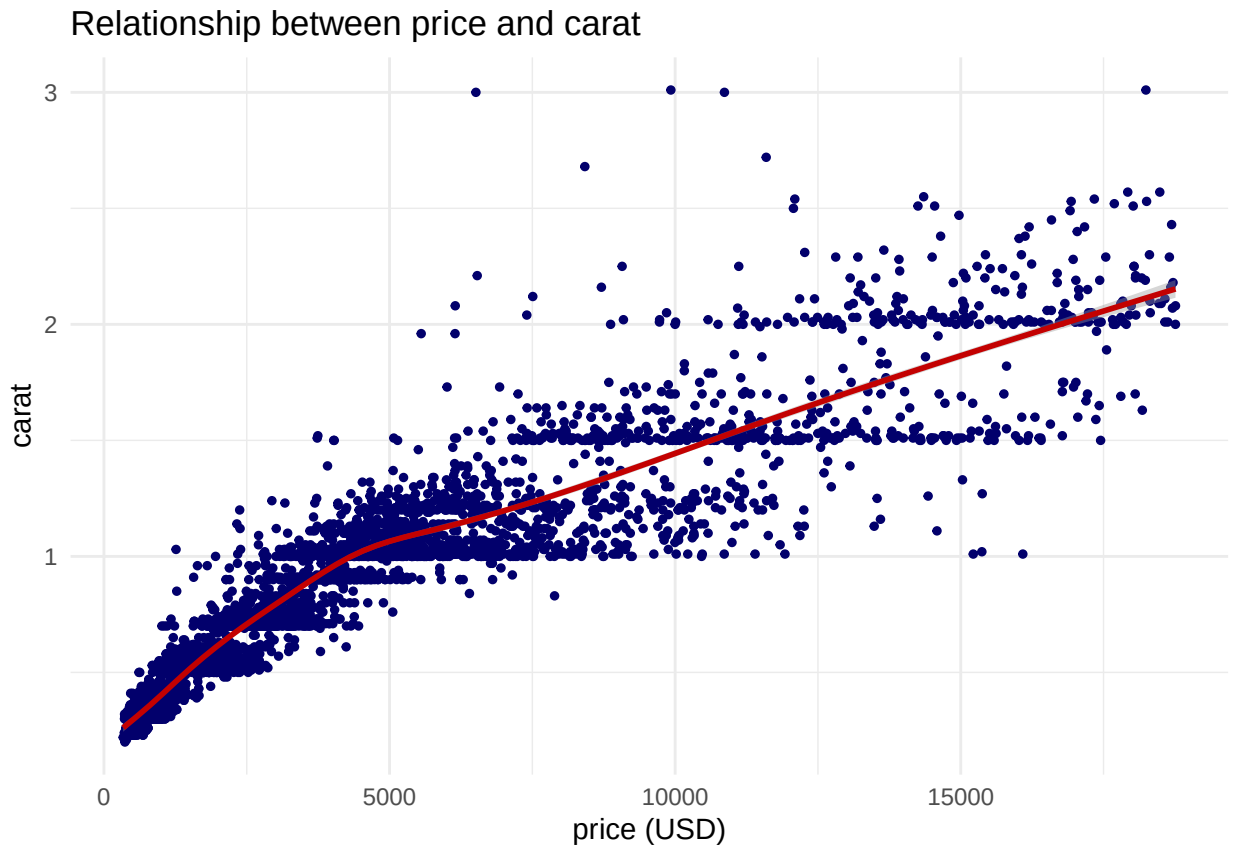


Chart 2 Proportion of cut in each clarity

```
ggplot(diamonds , aes(cut ,fill = clarity))+
  geom_bar(position = "fill")+
  theme_minimal()+
  labs(title = "Proportion of cut in each clarity")
```

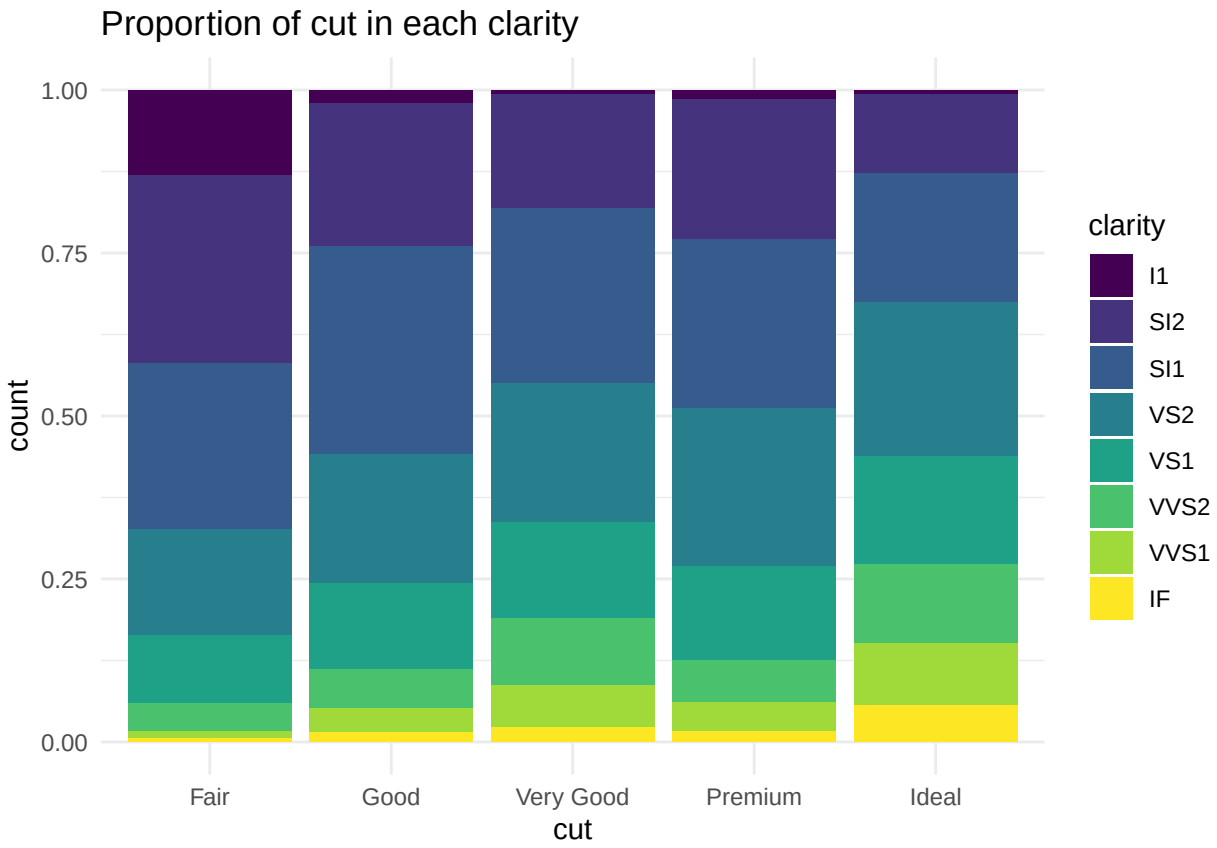


Chart 3 Distribution of price and color

```
ggplot(diamonds , aes(color ,price ,fill =color))+
  geom_boxplot()+
  theme_minimal()+
  labs(title = "Distribution of price and color",
       x= "color",
       y="price (USD)")
```

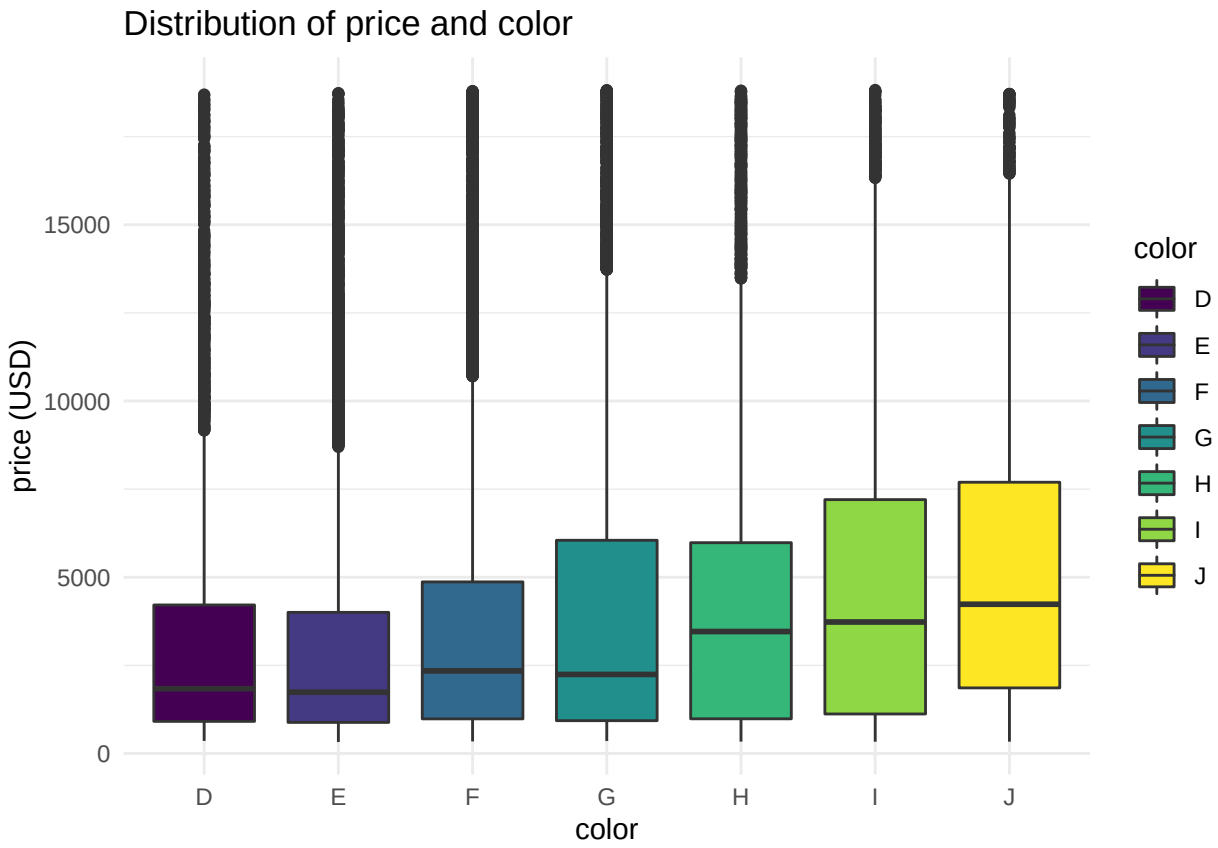


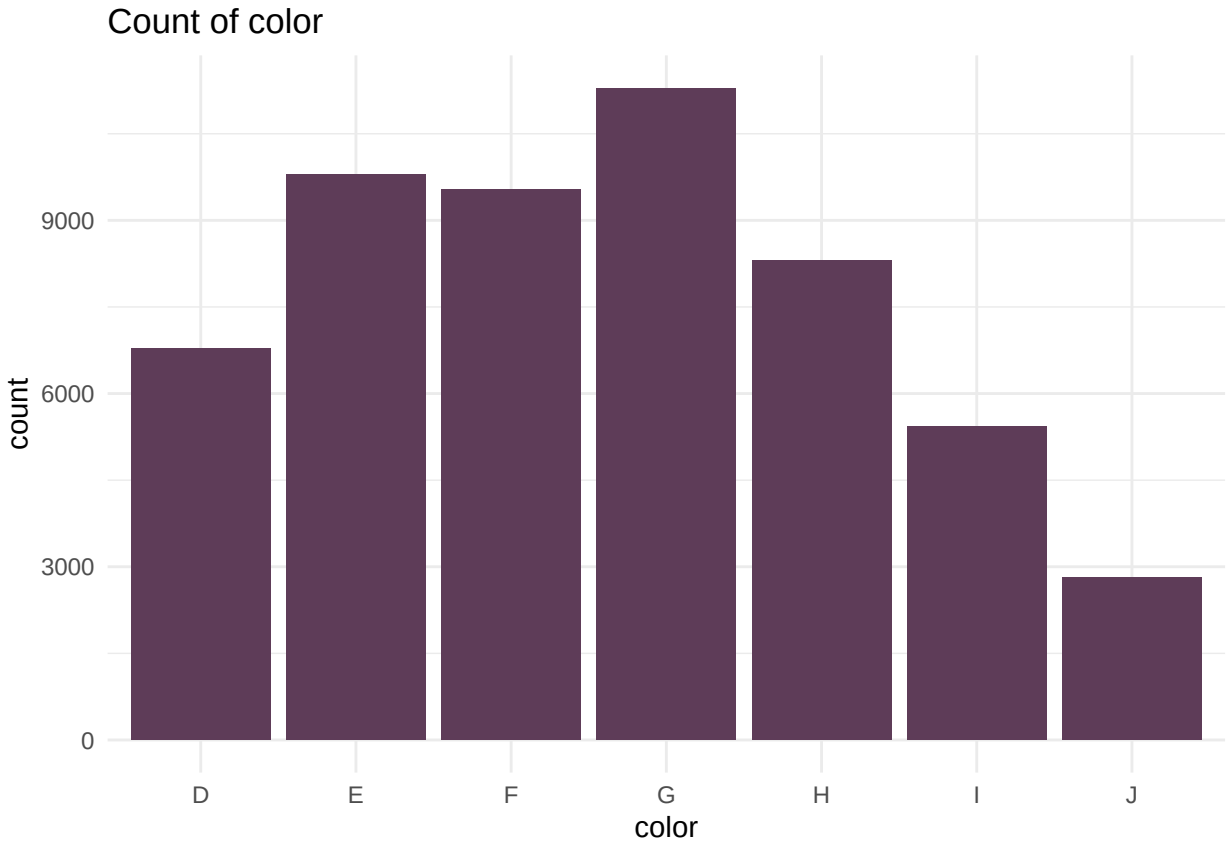
Chart 4 Relationship between price and carat group by clarity

```
ggplot(diamonds , aes(price, carat ))+
  geom_point(size = 0.5, alpha = 0.5 , col = "#65C3BA")+
  geom_smooth(size = 0.5, col = "#940000")+
  theme_minimal()+
  facet_wrap(~clarity)+
  labs(title = "Relationship between price and carat group by clarity",
       x = "price (USD)",
       y = "carat")
```

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

Chart 5 Count of color

```
ggplot(diamonds, aes(color))+
  geom_bar(fill = "#5e3c58")+
  theme_minimal()+
  labs(title = "Count of color")
```


Count of color
count
9000
6000
3000
0
D
E
F
G
H
I
J
color