

# Explore data with ggplot in R programing

Jintawee.s

## Install & Load Package

```
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

## Review Data

```
head(diamonds)

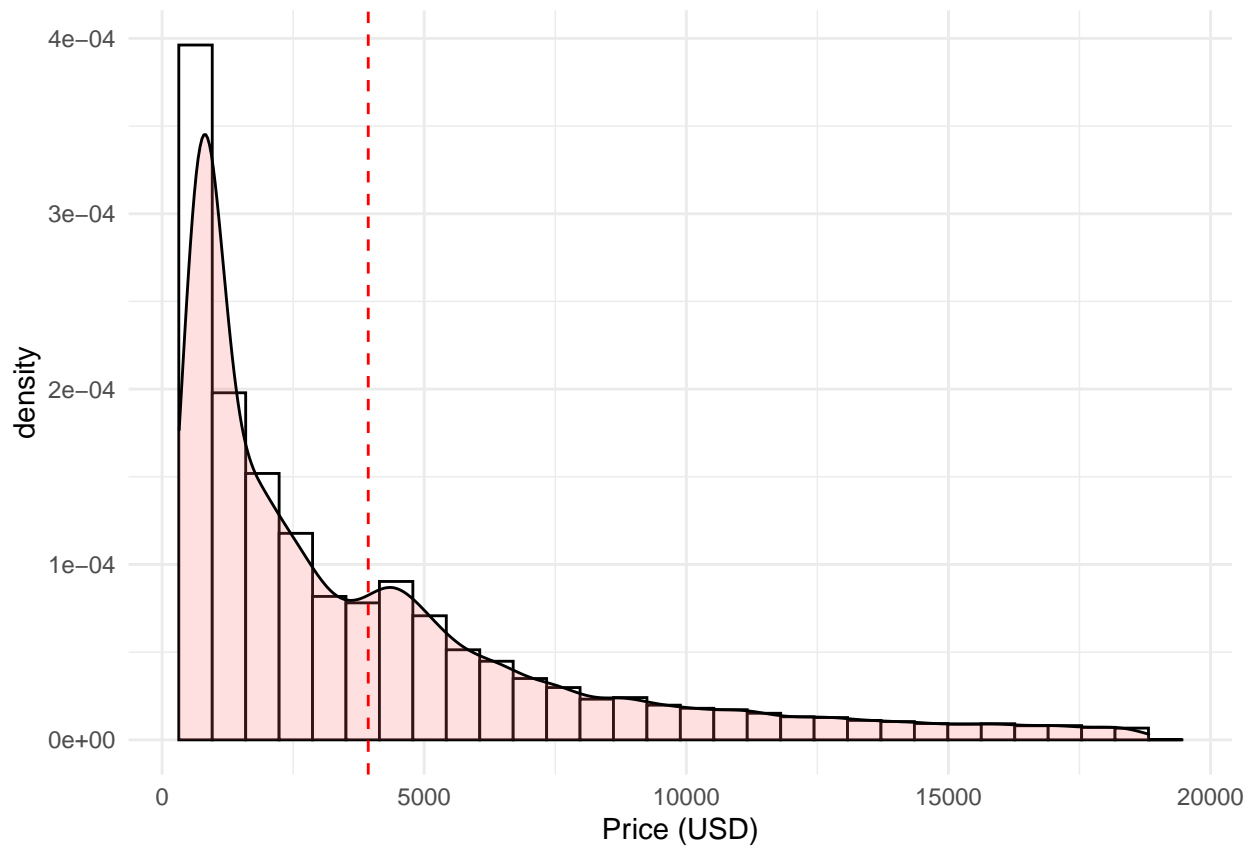
## # A tibble: 6 x 10
##   carat cut      color clarity depth table price      x      y      z
##   <dbl> <ord>    <ord> <ord>    <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1  0.23 Ideal    E     SI2     61.5   55   326  3.95  3.98  2.43
## 2  0.21 Premium E     SI1     59.8   61   326  3.89  3.84  2.31
## 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
```

## Chart 1

This is a histogram chart to show price distributin.

```
ggplot(diamonds, aes(price)) +
  geom_histogram(aes(y=..density..), color="black", fill="white")+
  geom_density(alpha=.2, fill="#FF6666") +
  geom_vline(aes(xintercept=mean(price)), color="red",
             linetype="dashed") +
  theme_minimal() +
  labs(x = "Price (USD)")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

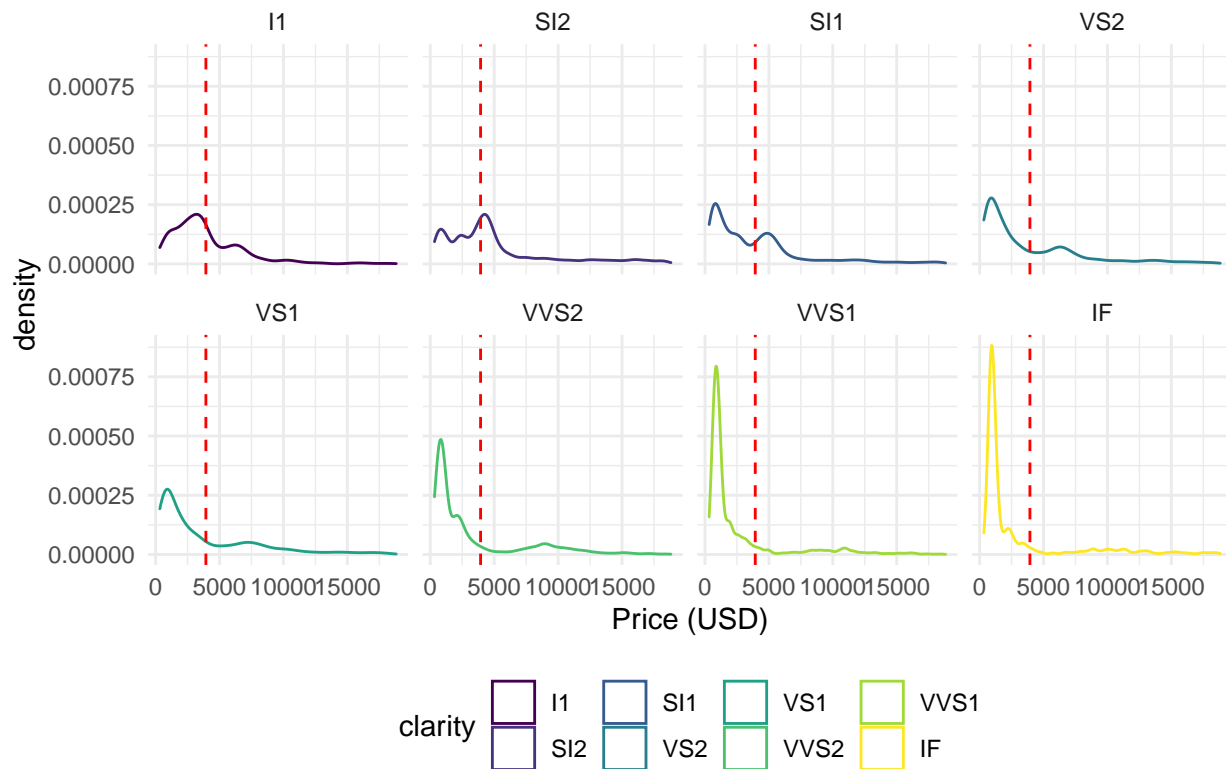


## # Chart 2

This is a line graph to show price distributin according clarity.

```
ggplot(diamonds, aes(x=price, color=clarity)) +
  geom_density() +
  geom_vline(aes(xintercept=mean(price)), color="red",
             linetype="dashed") +
  theme_minimal() +
  theme(legend.position="bottom") +
  labs(title = "Diamonds price frequency group by clarity ",
       x = "Price (USD)") +
  facet_wrap(~ clarity, ncol=4)
```

## Diamonds price frequency group by clarity



### # Chart 3

This is a scatter plot to show the relationship between carat and price.

```
set.seed(25)
ggplot(sample_n(diamonds, 1000),
  aes(carat ,price)) +
  geom_point(size= 2,
    alpha=0.6,
    mapping = aes(color = clarity)) +
  geom_smooth(method="lm") +
  scale_colour_brewer(palette = "RdYlGn") +
  theme_minimal() +
  labs(title = "Relationship between carat and price",
    x = "Carat",
    y = "Price (USD)")
```

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

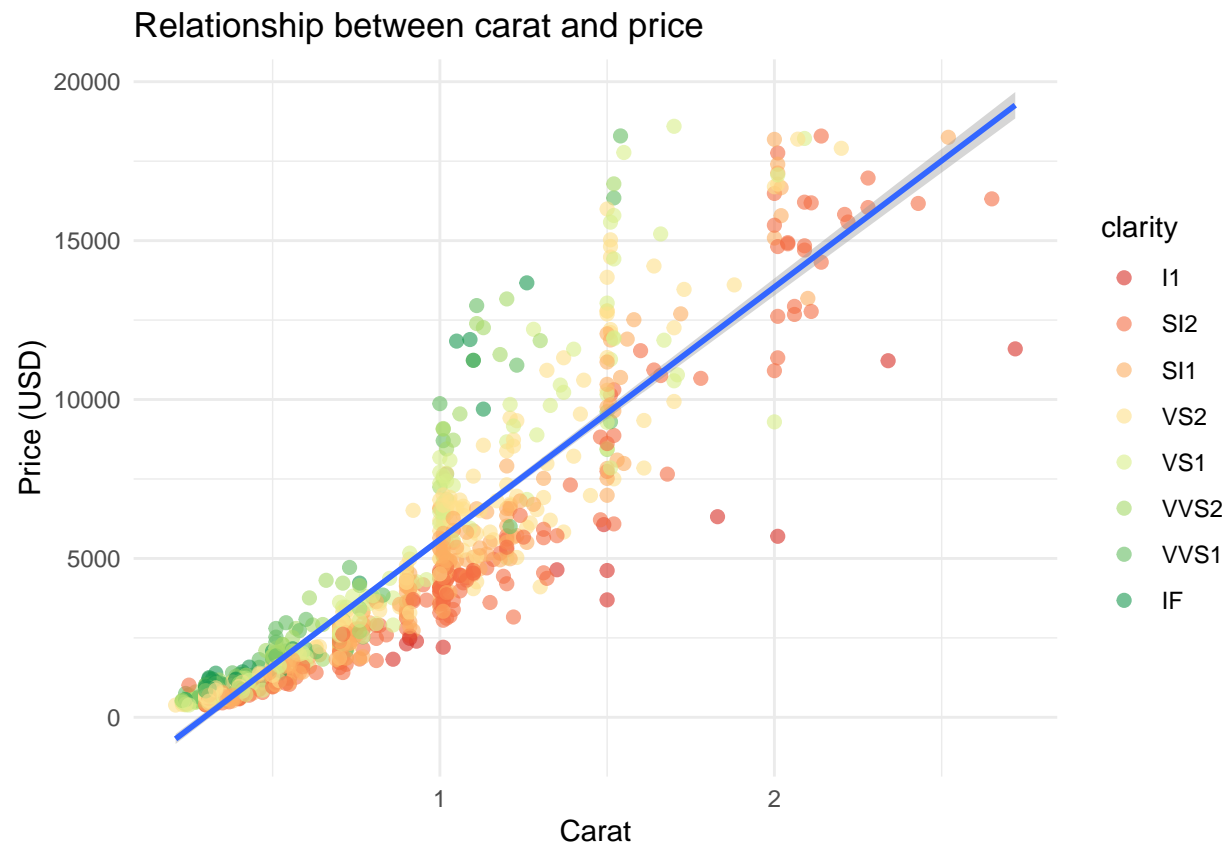


Chart 4

This is a stacked bar chart to show clarity distributin.

```
ggplot(sample_n(diamonds, 5000), aes(clarity, fill=cut)) +
  geom_bar(position = "fill") +
  scale_fill_brewer(direction = -1) +
  theme_minimal() +
  labs(title = " ",
       x = "Clarity",
       y = "Count")
```

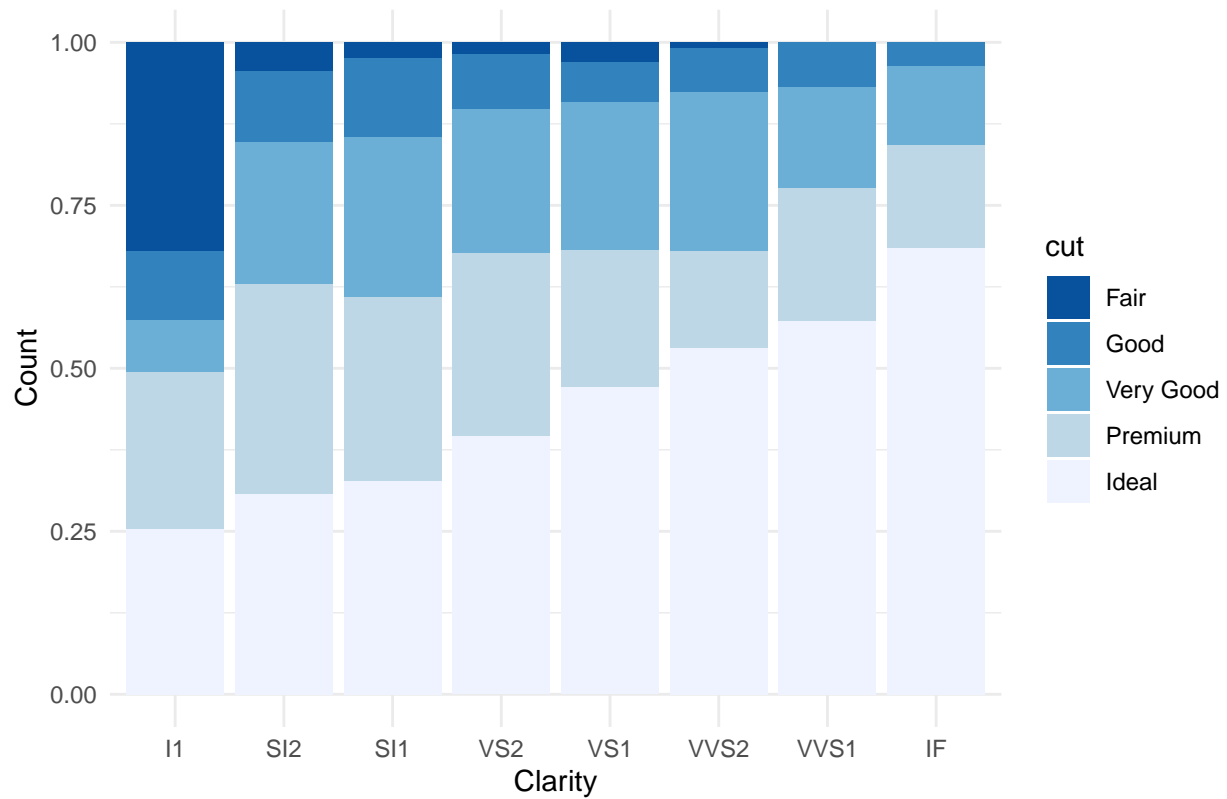


Chart 5

This is a boxplot to show price distributin according clarity.

```
set.seed(15)
ggplot(sample_n(diamonds, 1000),
  aes(clarity, price, fill=clarity)) +
  geom_boxplot(width=0.3) +
  scale_fill_brewer(palette="BuPu") +
  labs(title = "Diamonds price according Clarity",
    x = "Clarity",
    y = "Price (USD)")
```

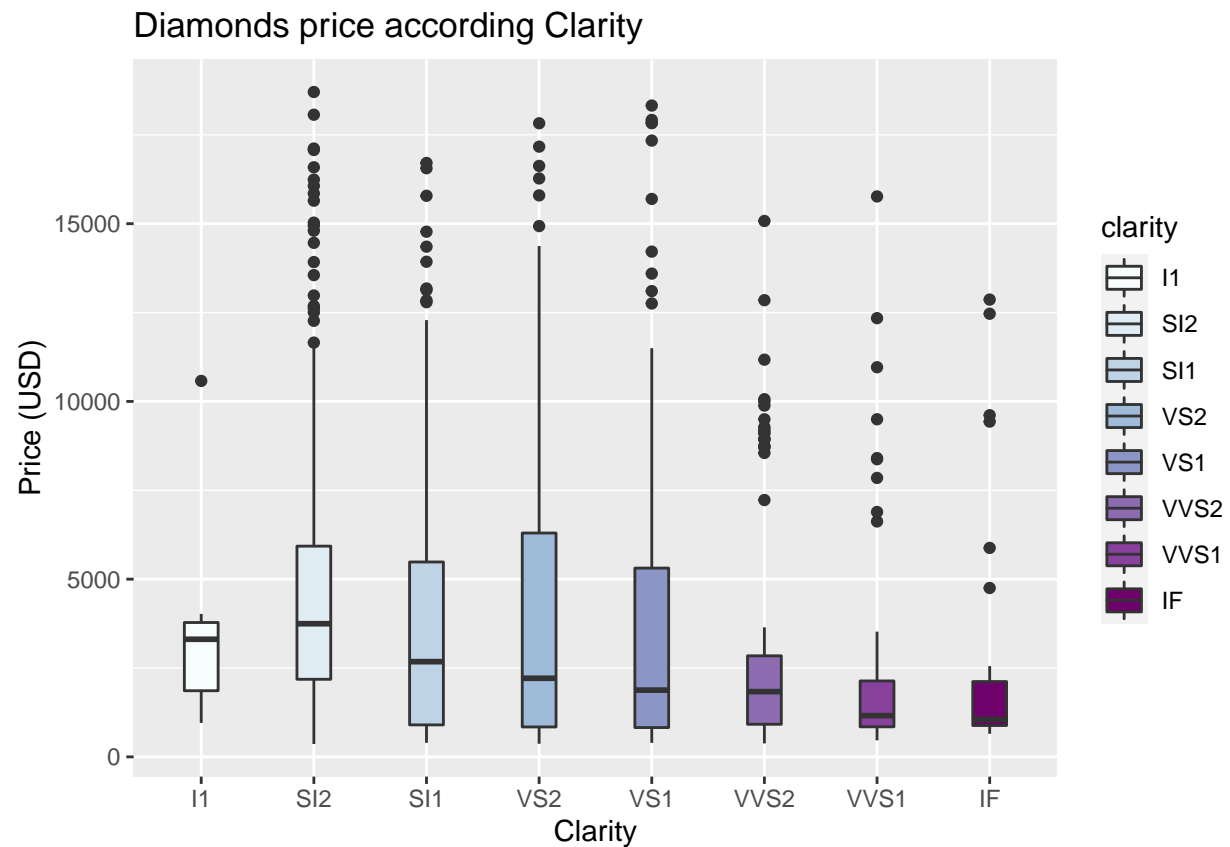
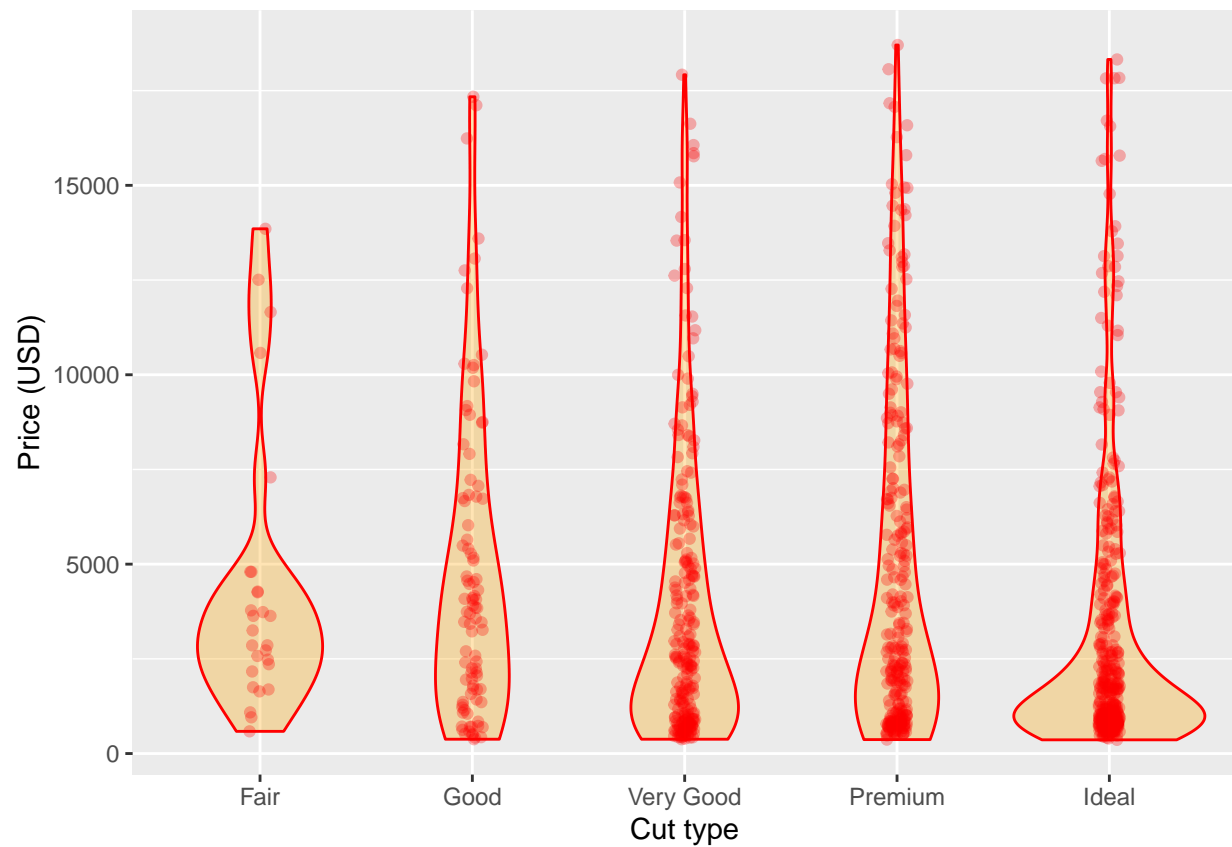


Chart 6

This is a violin plot to show the relationship between distributions of diamonds and price according cut type.

```
set.seed(15)
ggplot(sample_n(diamonds, 1000),
  aes(cut, price)) +
  geom_violin(color="red", fill="orange", alpha=0.3) +
  geom_jitter(width=0.05,color="red", alpha=0.3 ) +
  labs(x = "Cut type",
    y = "Price (USD)")
```



#### # Chart 7

This is a scatter plot to show the relationship between clarity and color according cut type.

```
ggplot(diamonds, aes(clarity, color)) +
  geom_count(aes(color = ..n.., size = ..n..)) +
  scale_colour_gradient(low = "yellow", high = "black") +
  facet_wrap(~ cut, ncol=3)
```

