# Diamonds Data Visualization : ggplot

Mr.Tanakorn

**Library**

```
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()
```

**Database**

```
glimpse(diamonds)
```

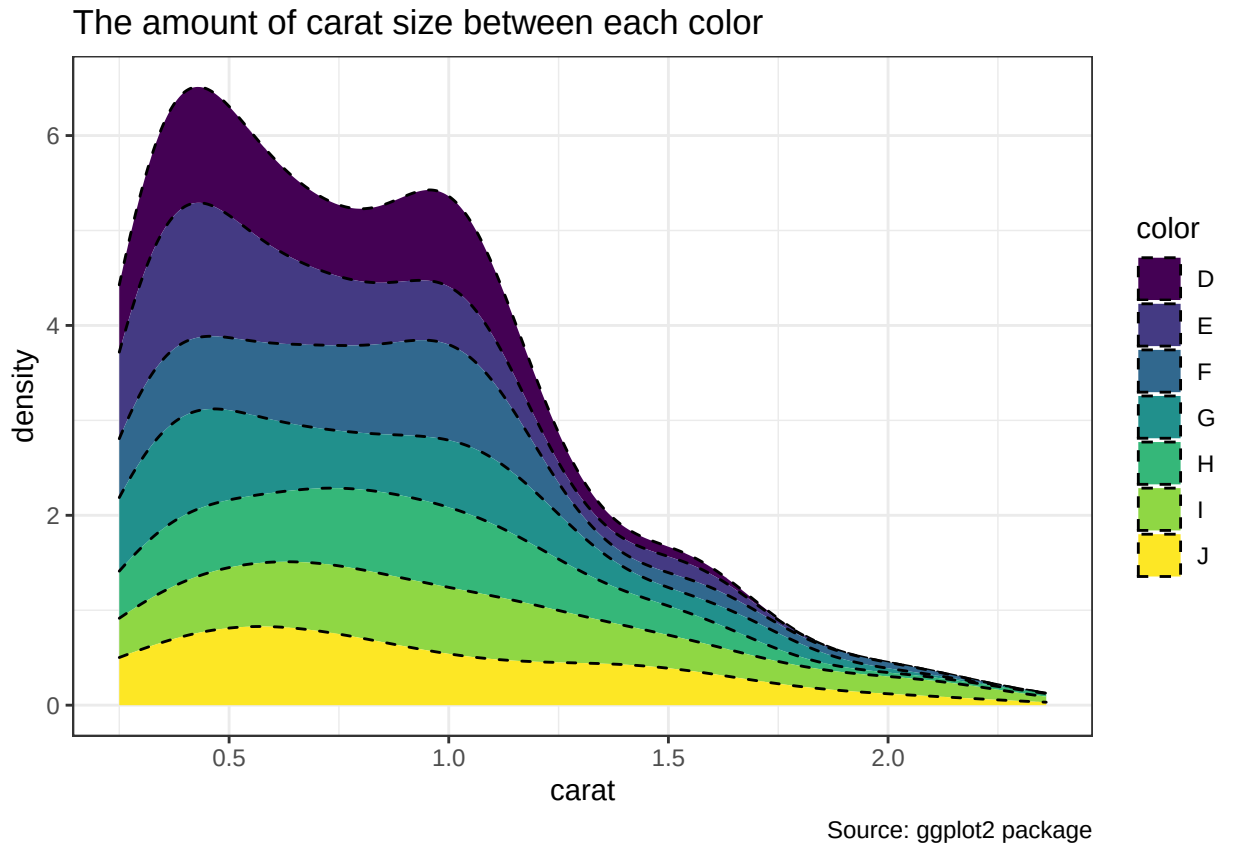
```
## Rows: 53,940
## Columns: 10
## $ carat   <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, 0.23, 0.~
## $ cut     <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very Good, Ver~
## $ color   <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, J, J, I,~
## $ clarity <ord> 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.~
```

## Data Visualization

### The amount of carat size between each color

```
sample_df <- diamonds |>
  sample_n(size = 500)
sample_df  |>
```

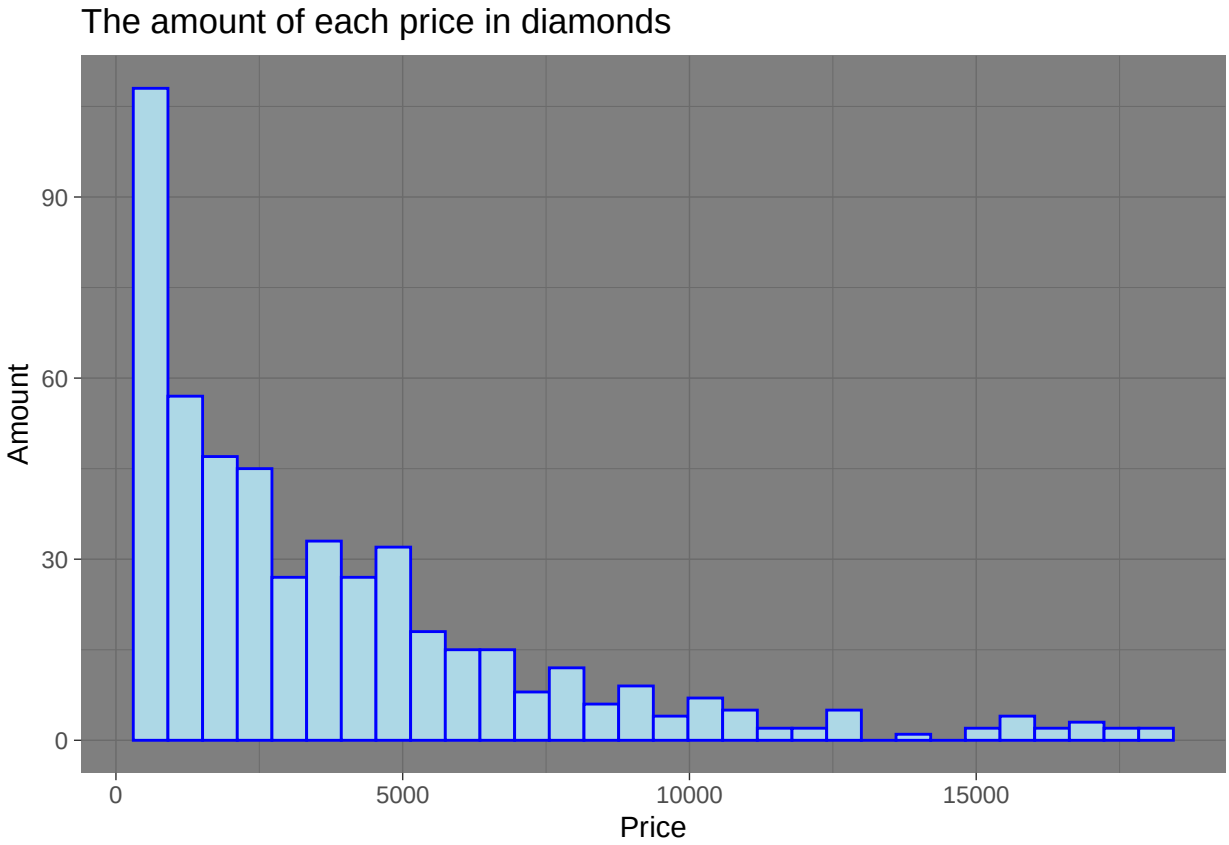
```
  ggplot(aes(carat, after_stat(density), fill = color)) +
  geom_density(linetype="dashed", position = "stack") +
  theme_bw() +
  labs(title = "The amount of carat size between each color",
       caption = "Source: ggplot2 package")
```




### The amount of each price in diamonds

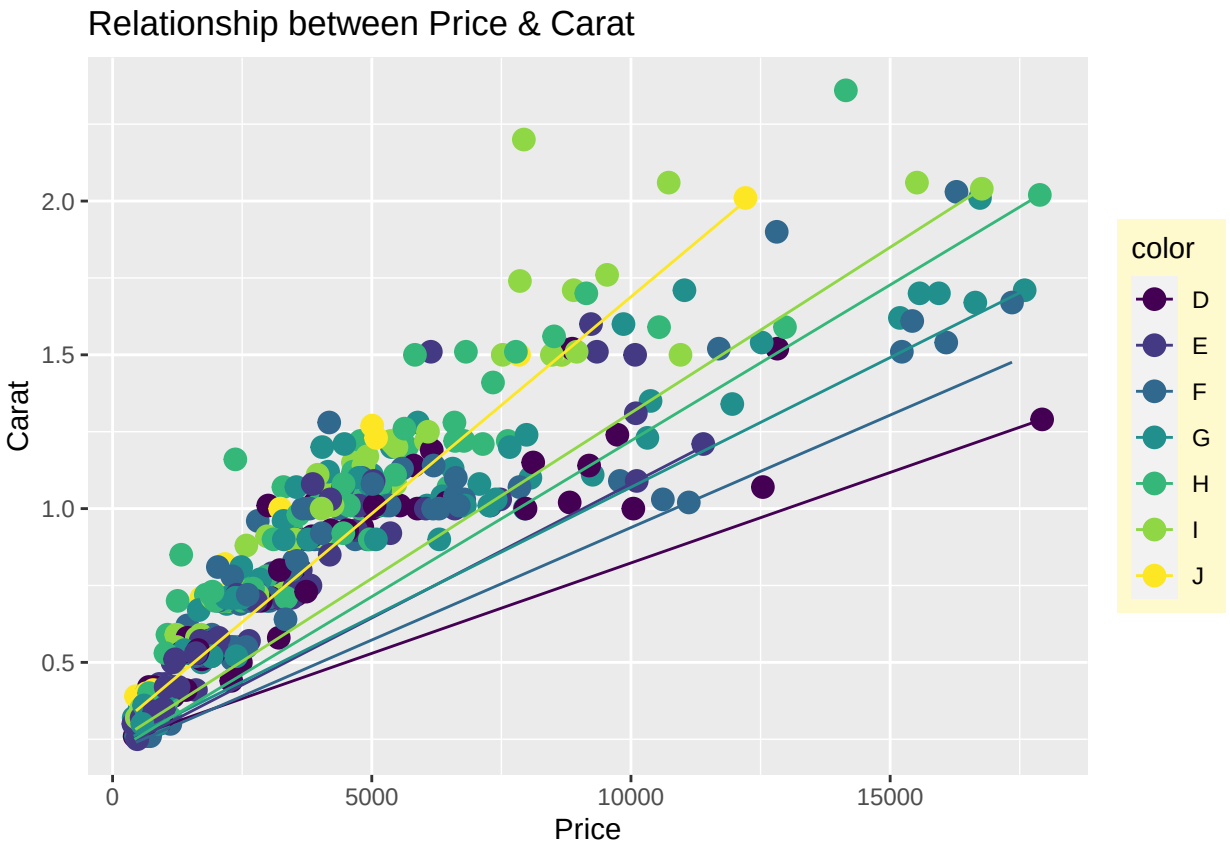
```
sample_df  |>
  ggplot(aes(x= price)) +
  geom_histogram(color="blue", fill="lightblue") +
  theme_dark() +
  labs(title = "The amount of each price in diamonds",
       x = "Price",
       y = "Amount")
```

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



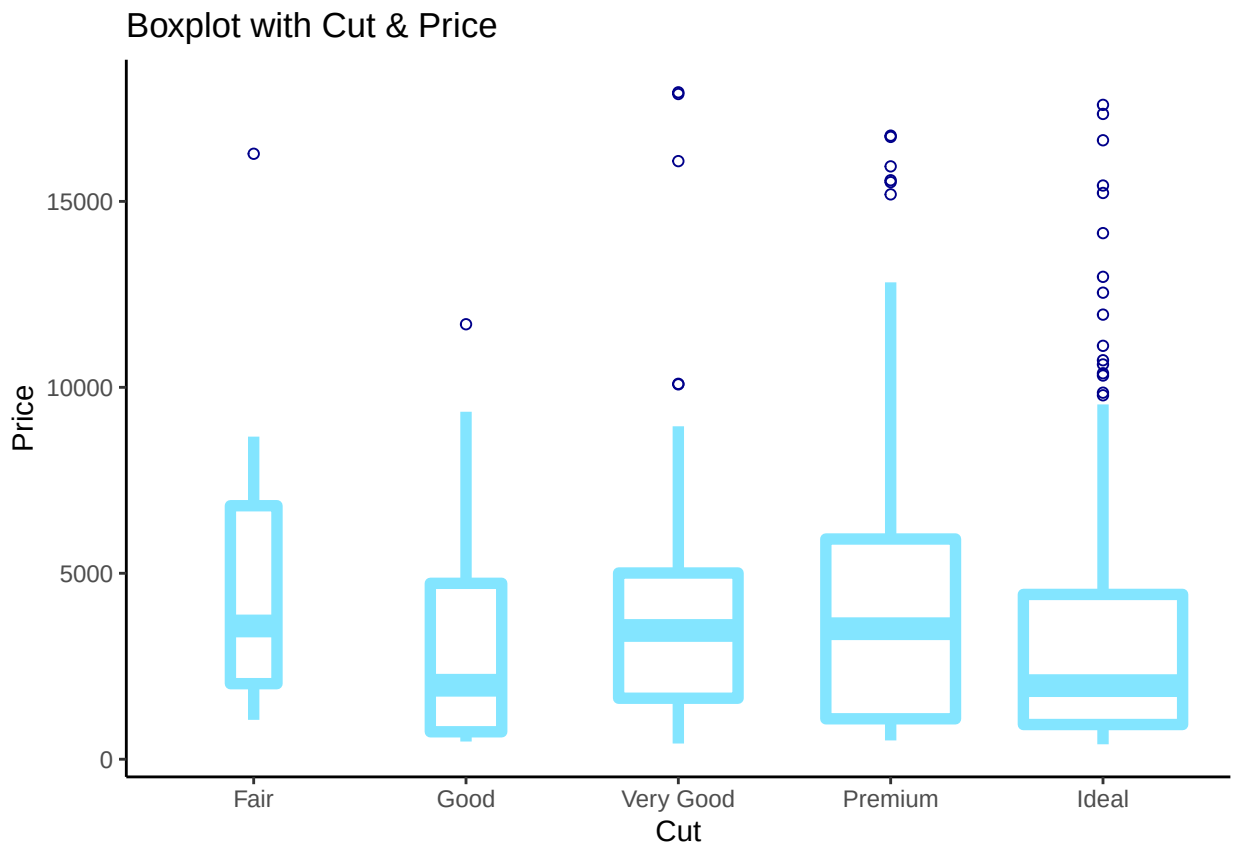

### Relationship between Price & Carat

```
sample_df |>
  ggplot(aes(price, carat, colour=color)) +
  geom_point(size=3.5, alpha = 1.5) +
  geom_quantile(quantiles = 1.5, aes(group = color)) +
  theme(legend.background = element_rect(fill = "lemonchiffon")) +
  labs(title = "Relationship between Price & Carat",
       x = "Price",
       y = "Carat")
```



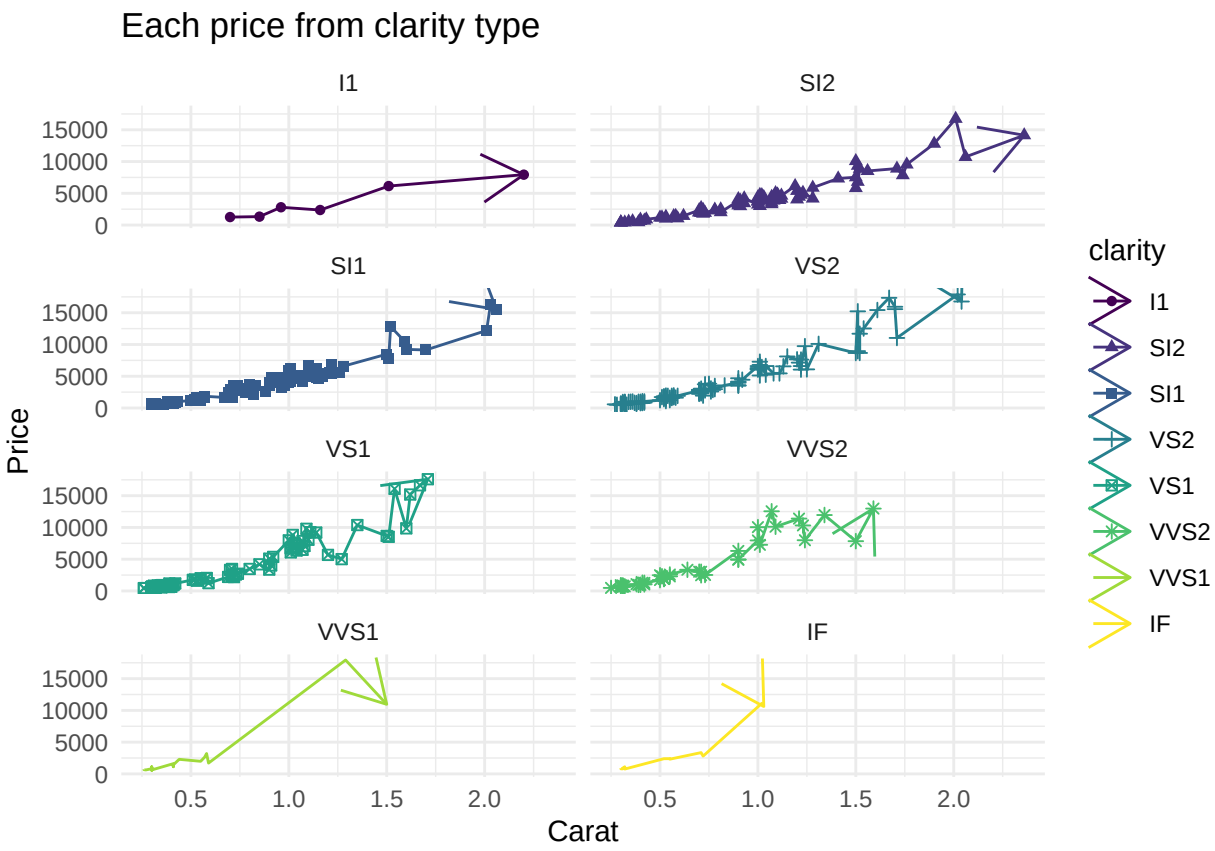

### Boxplot with Cut & Price

```
sample_df |>
  ggplot(aes(cut,price)) +
  geom_boxplot(varwidth = TRUE,
               fill = "white", colour = "#83e5ff",size=2,
               outlier.colour = "Darkblue", outlier.shape = 1)+
  theme_classic()+
  labs(title = "Boxplot with Cut & Price",
       x = "Cut",
       y = "Price")
```




### Each price from clarity type

```
sample_df |>
  ggplot(aes(carat,price, color=clarity)) +
  geom_point(aes(shape=clarity))+
  geom_line(arrow=arrow())+
  facet_wrap(~ clarity , ncol = 2)+
  theme_minimal()+
  labs(title = "Each price from clarity type",
       x = "Carat",
       y = "Price")
```


Each price from clarity type
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF
15000
10000
5000
0
Price
0.5
1.0
1.5
2.0
Carat
clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF