

Exploratory diamonds data and plot with ggplot

By: Morakot Teanchai

Preparation library

```
## library tidyverse
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()
```

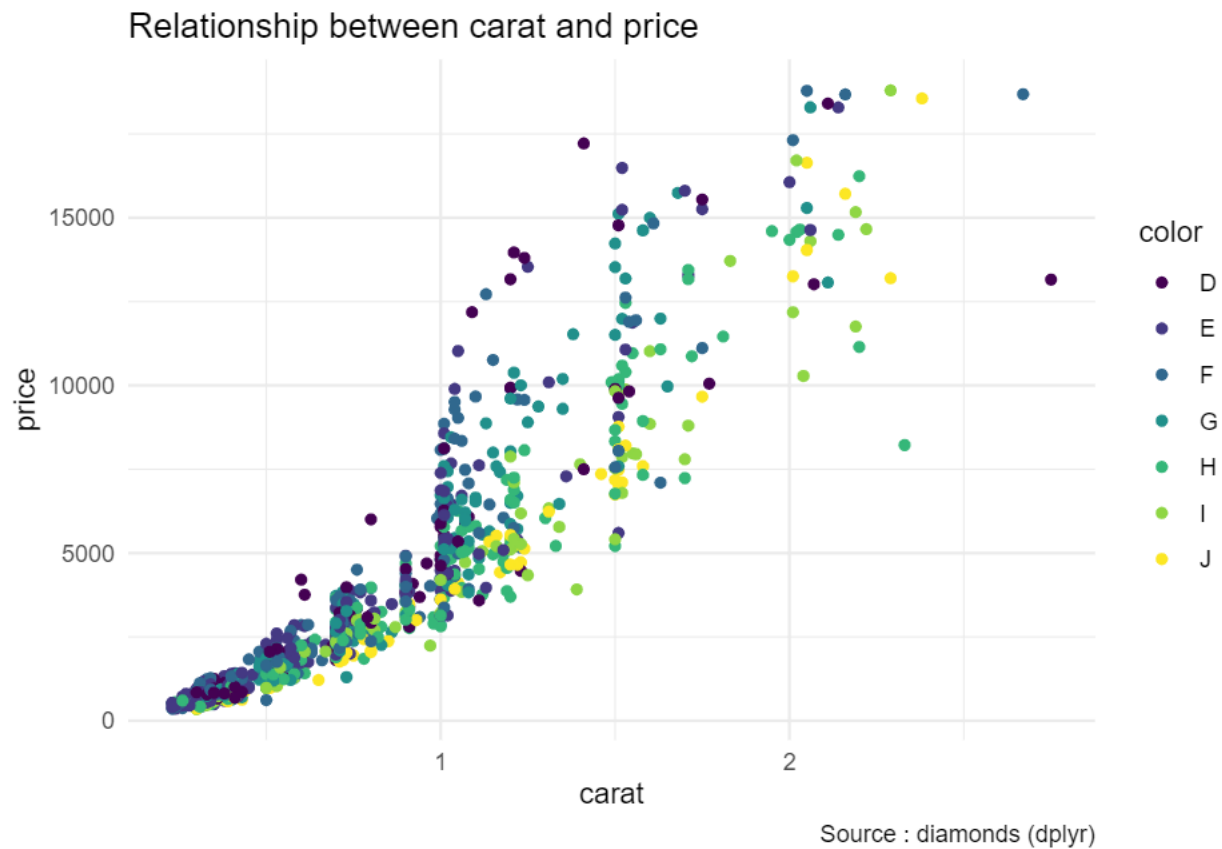
Data Overview

```
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.~
```

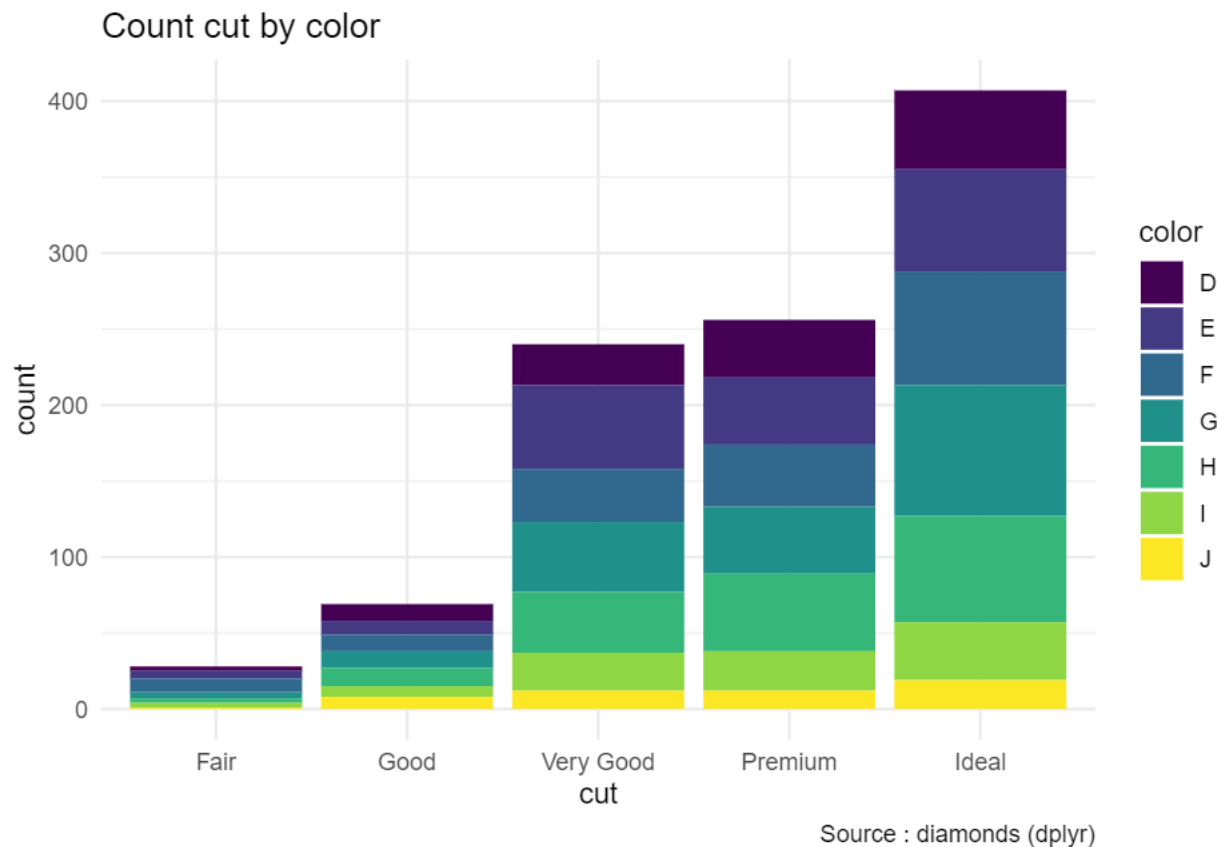
1. Relationship between weight of the diamond (carat) and table

```
set.seed(13)
ggplot(sample_n(diamonds, 1000), aes( carat,price, colour=color)) +
  geom_point() +
  labs(title = "Relationship between carat and price",
       caption = "Source : diamonds (dplyr)") +
  theme_minimal()
```



2. Count quality of the cut by color

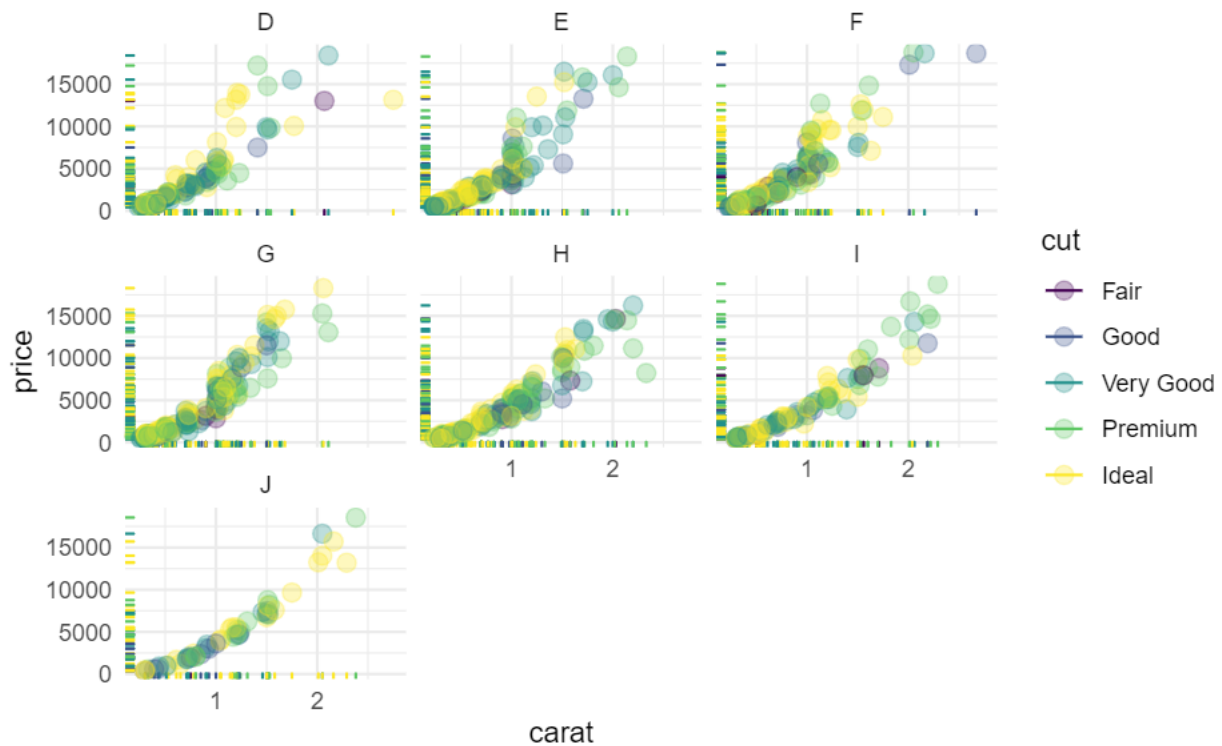
```
set.seed(13)
ggplot(sample_n(diamonds, 1000)) +
  geom_bar(mapping = aes(x=cut, fill = color)) +
  labs(title = "Count cut by color",
       caption = "Source : diamonds (dplyr)") +
  theme_minimal()
```



3. Relationship between weight of the diamond(carat) and price by color

```
set.seed(13)
ggplot(sample_n(diamonds, 1000), aes(carat, price, colour = cut)) +
  geom_rug() +
  geom_point(size = 3, alpha = 0.3) +
  facet_wrap(~color, ncol = 3) +
  labs(title = "Relationship between carat and price by color",
       caption = "Source : diamonds (dplyr)") +
  theme_minimal()
```

Relationship between carat and price by color



Source : diamonds (dplyr)

4. Relationship between quality of the cut and price

```
set.seed(13)
ggplot(sample_n(diamonds, 1000), aes(cut, price, colour=cut)) +
  geom_boxplot() +
  labs(title = "Relationship between cut and price",
       caption = "Source : diamonds (dplyr)") +
  theme_minimal()
```



5. Density price by diamond color

```
set.seed(13)
ggplot(sample_n(diamonds, 1000), aes(price, fill = color, colour = color)) +
  geom_density(alpha = 0.1) +
  labs(title = "Density price by color",
       caption = "Source : diamonds (dplyr)") +
  theme_minimal()
```

