

HW Data Visualization

Patsawut Tananchai

diamonds anlytics

load lib

```
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
```

```
library(tidyverse)
```

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

```
library(ggplot2)
```

```
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 Description

?diamonds

A data frame with 53940 rows and 10 variables:

price - price in US dollars (\$326–\$18,823)

carat - weight of the diamond (0.2–5.01)

cut - quality of the cut (Fair, Good, Very Good, Premium, Ideal)

color - diamond colour, from D (best) to J (worst)

clarity - a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))

x - length in mm (0–10.74)

y - width in mm (0–58.9)

z - depth in mm (0–31.8)

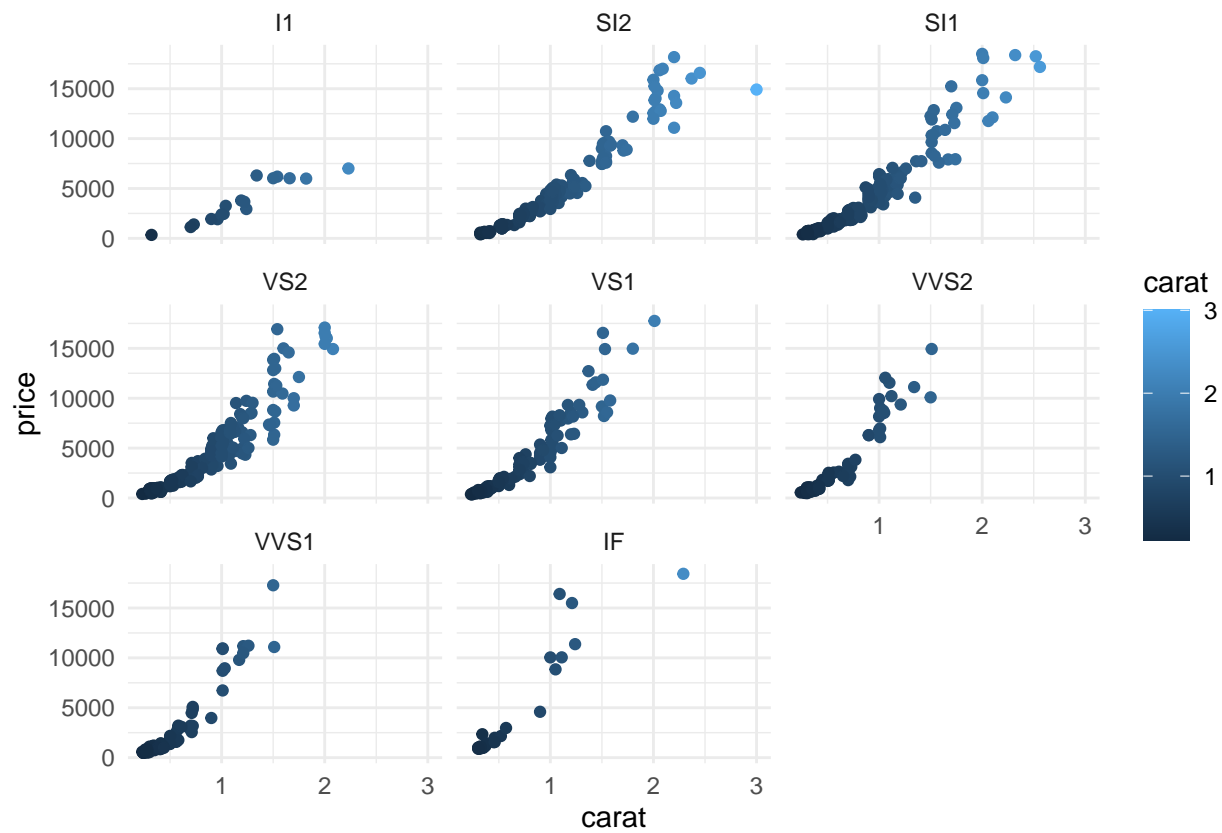
depth - total depth percentage = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43–79)

table - width of top of diamond relative to widest point (43–95)

Visualization

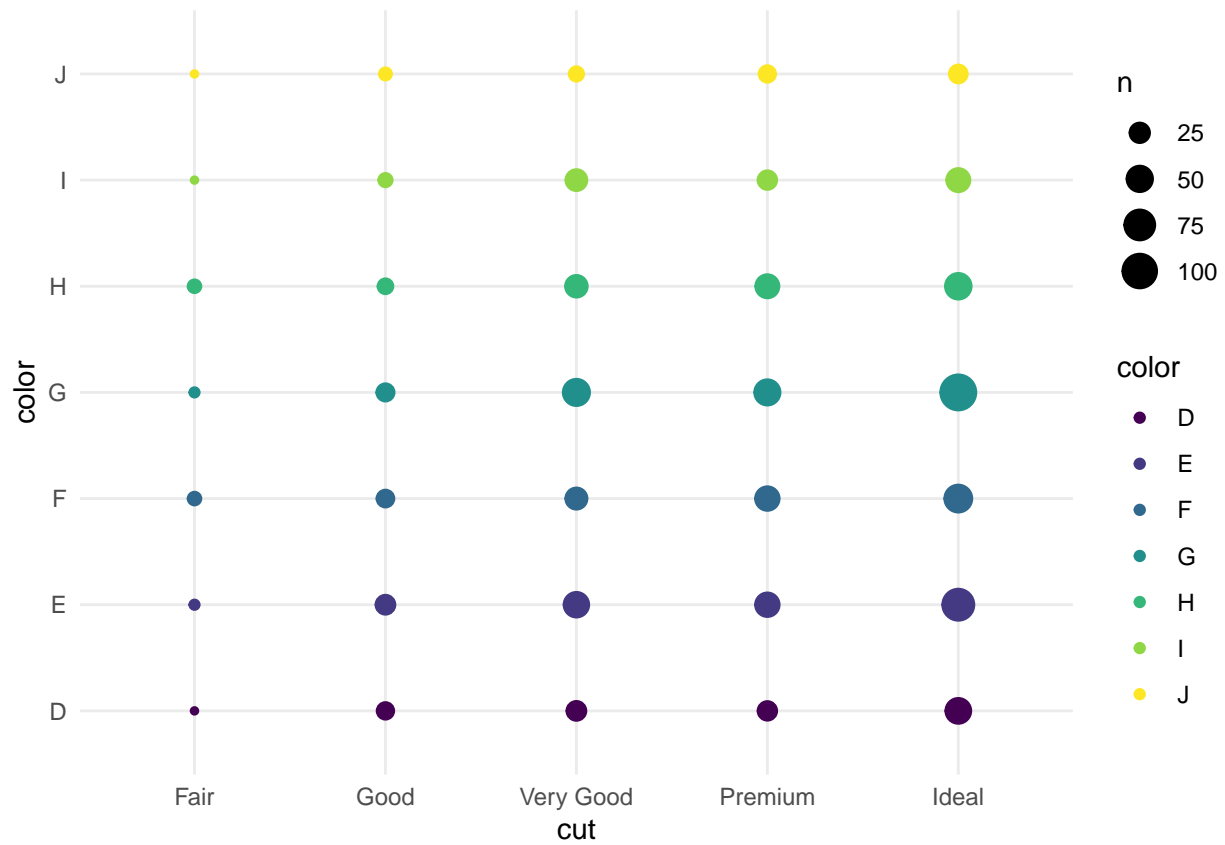
Relationships Between carat, price , color

```
set.seed(42)
sample_diamonds <- sample_n(diamonds, 1000)
ggplot(sample_diamonds, aes(carat, price, color = carat)) +
  geom_point() +
  theme_minimal() +
  facet_wrap(~ clarity)
```



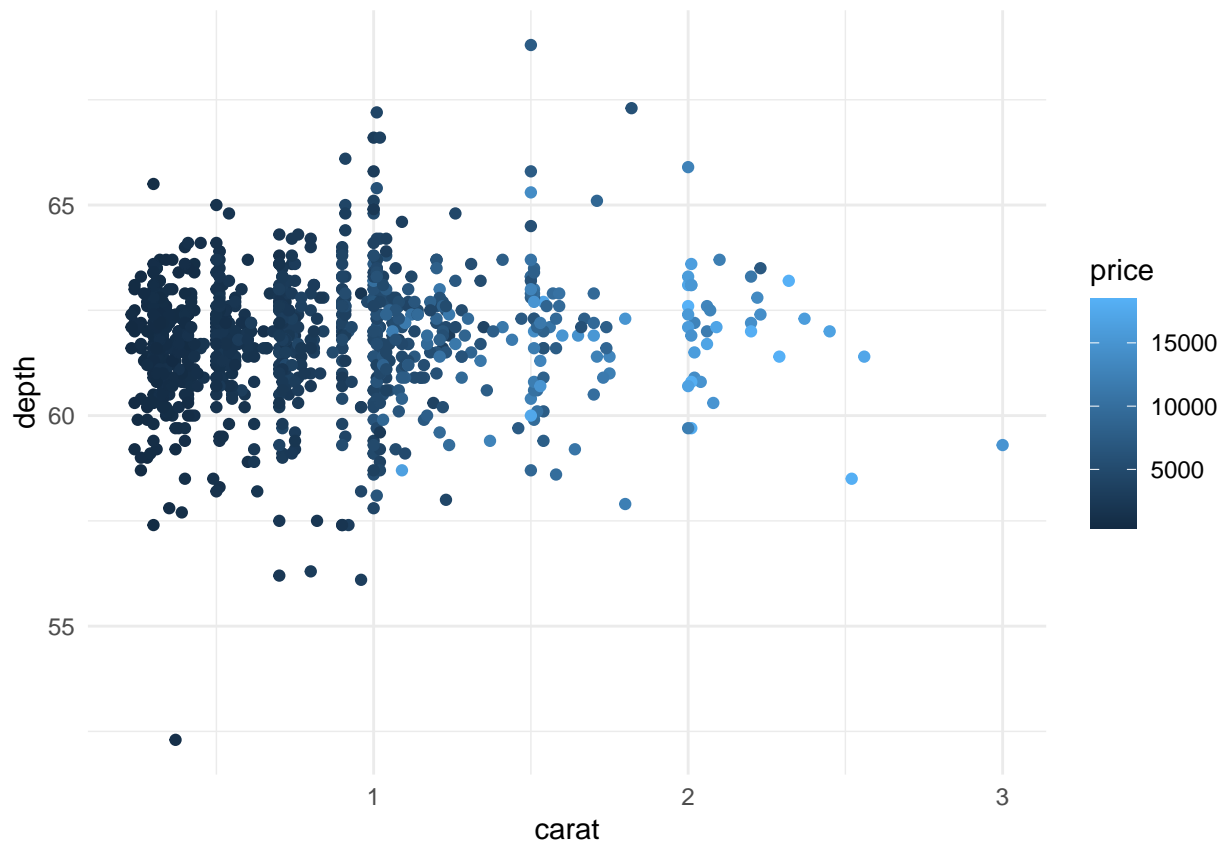
This graph shows that clarity doesn't relate with the price but carat relate with price
number of cut , color

```
ggplot(sample_daimonds, aes(cut, color , color = color )) +  
  geom_count() +  
  theme_minimal()
```



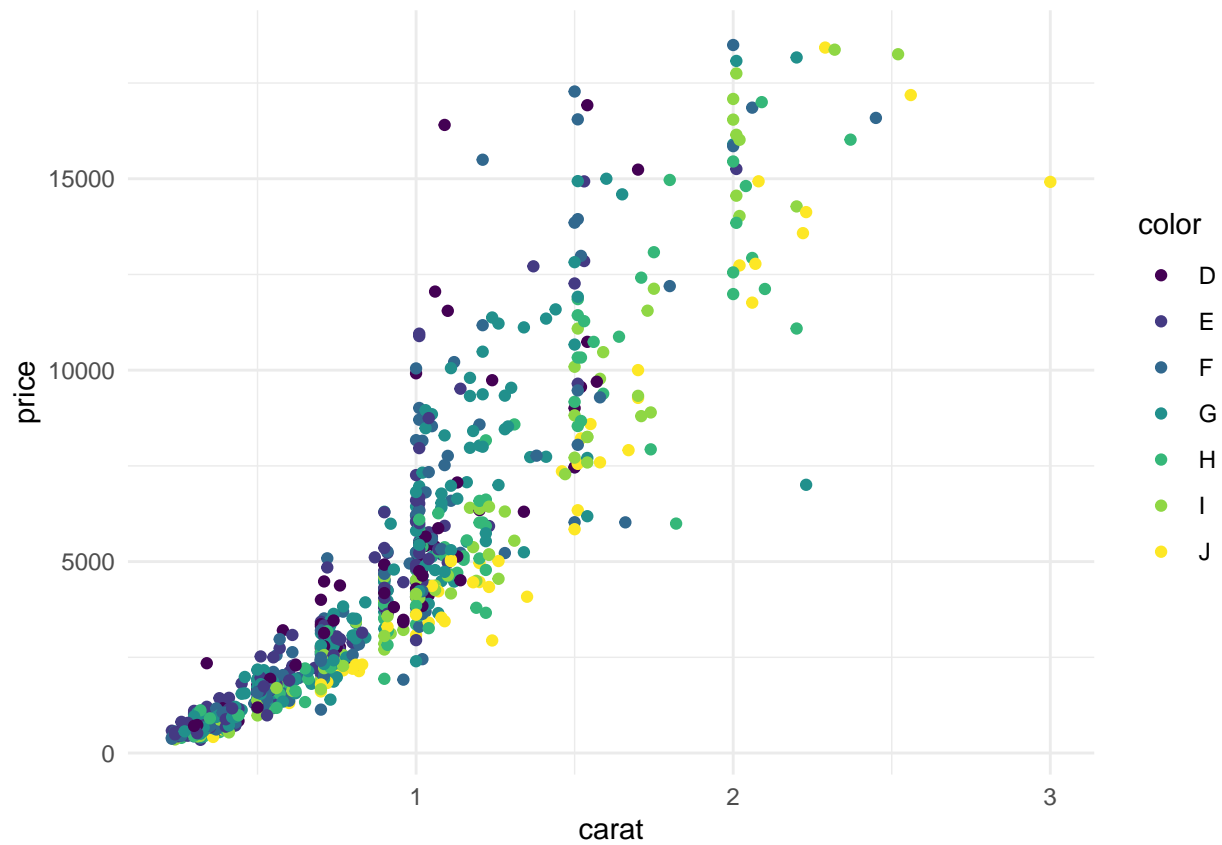
Relationships Between carat , dept

```
ggplot(sample_daimonds, aes(carat, depth , color = price )) +
  geom_point() +
  theme_minimal()
```



This graph shows that dept doesn't relate with the price but carat relate with price

```
ggplot(sample_daimonds, aes(carat, price , color = color )) +  
  geom_point() +  
  theme_minimal()
```



This graph shows that color doesn't relate with the price but carat relate with price

count number of diamonds group by color

```
sample_daimonds %>%
  ggplot(aes(color , fill = color)) +
  geom_bar()
```

