

Data Viz Batch 11

Chanchawat Pakdeesri

2025-02-22

Homework

My ggplot2 project

First, we need to call library 'ggplot2' and 'dplyr'

```
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
```

check the columns with 'glimpse'

```
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.~
```

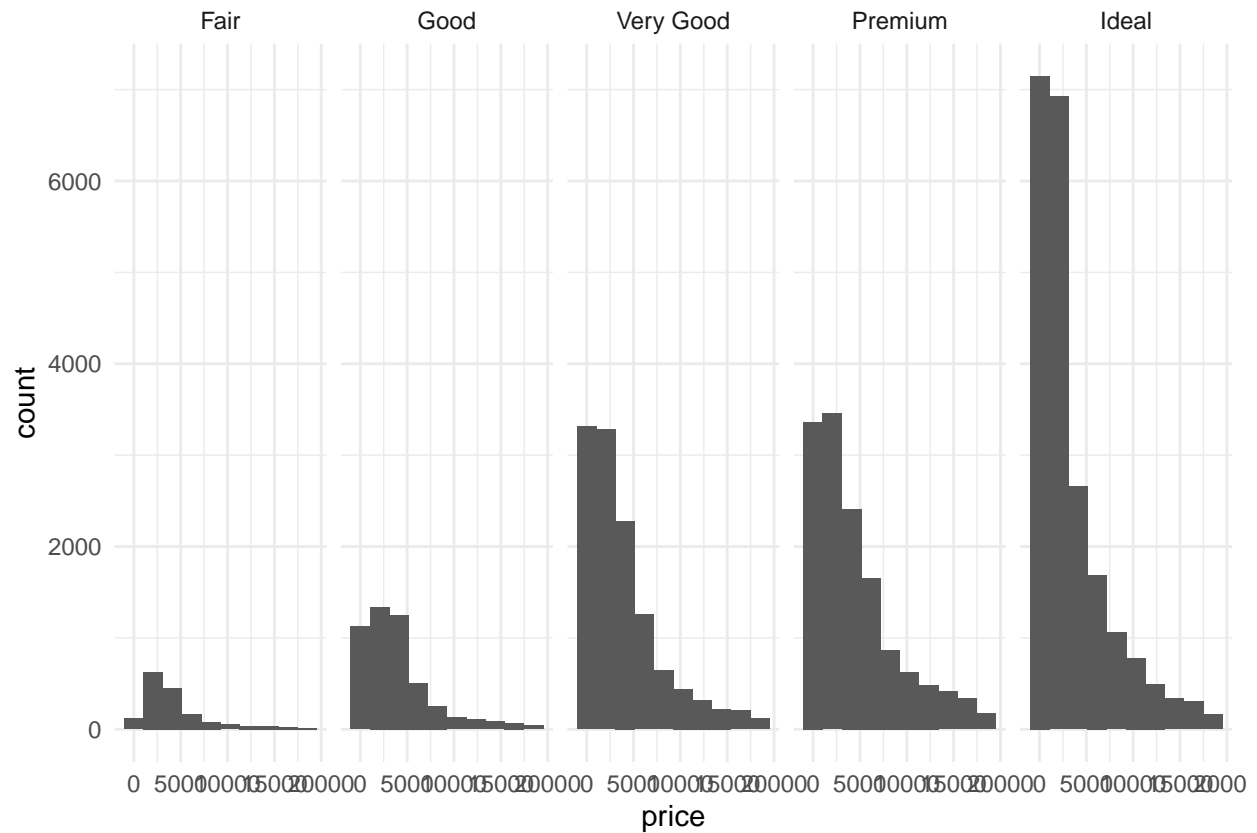
1.Create graph with ONE VARIABLE

Continuous (with faceting)

Q: How does the distribution of 'price' vary across different 'cut' categories

```
diamonds %>%
ggplot()+
  geom_histogram(aes(price),
                 bins=10) +
```

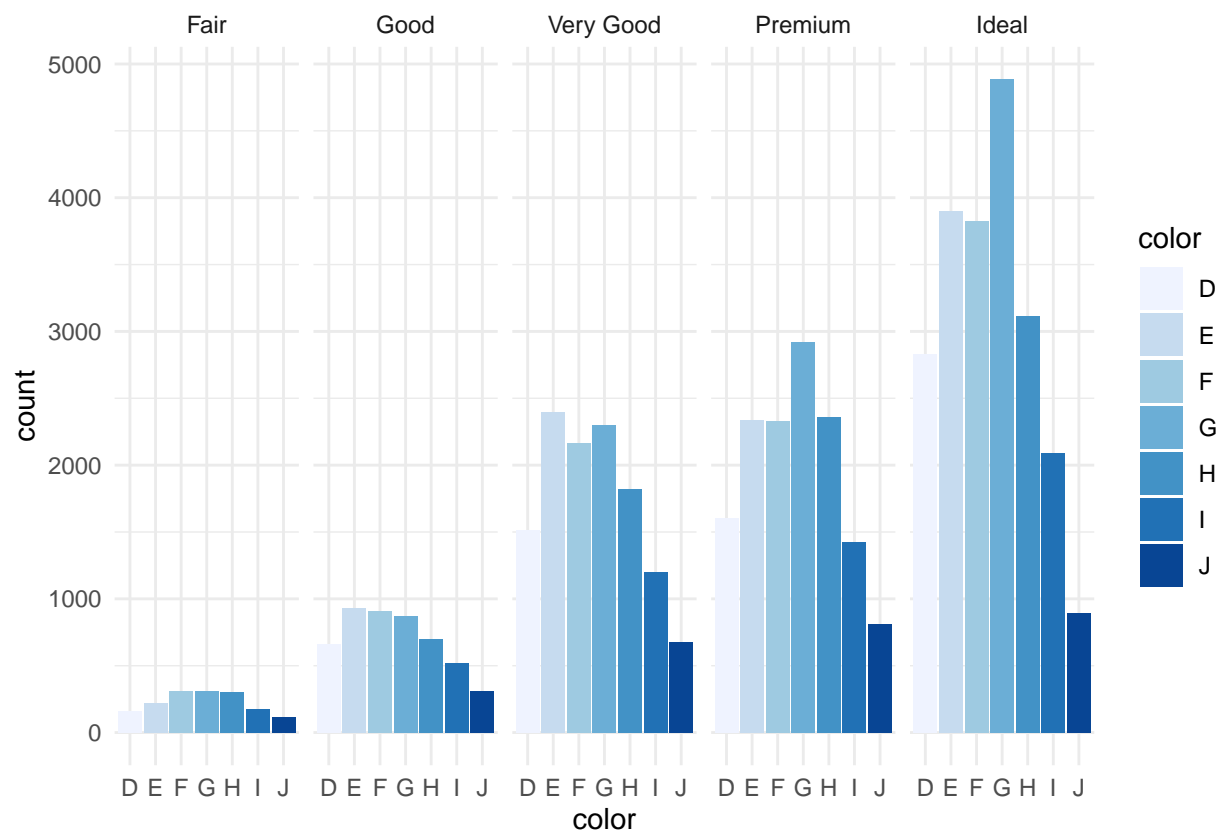
```
facet_grid(cols = vars(cut)) +  
theme_minimal()
```



Discrete (with faceting)

Q: How does the distribution of 'color' vary across different 'cut' categories?

```
diamonds %>%  
ggplot()+  
  geom_bar(aes(color,  
                fill=color)) +  
  facet_grid(cols = vars(cut)) +  
  scale_fill_brewer(palette='Blues') +  
  theme_minimal()
```



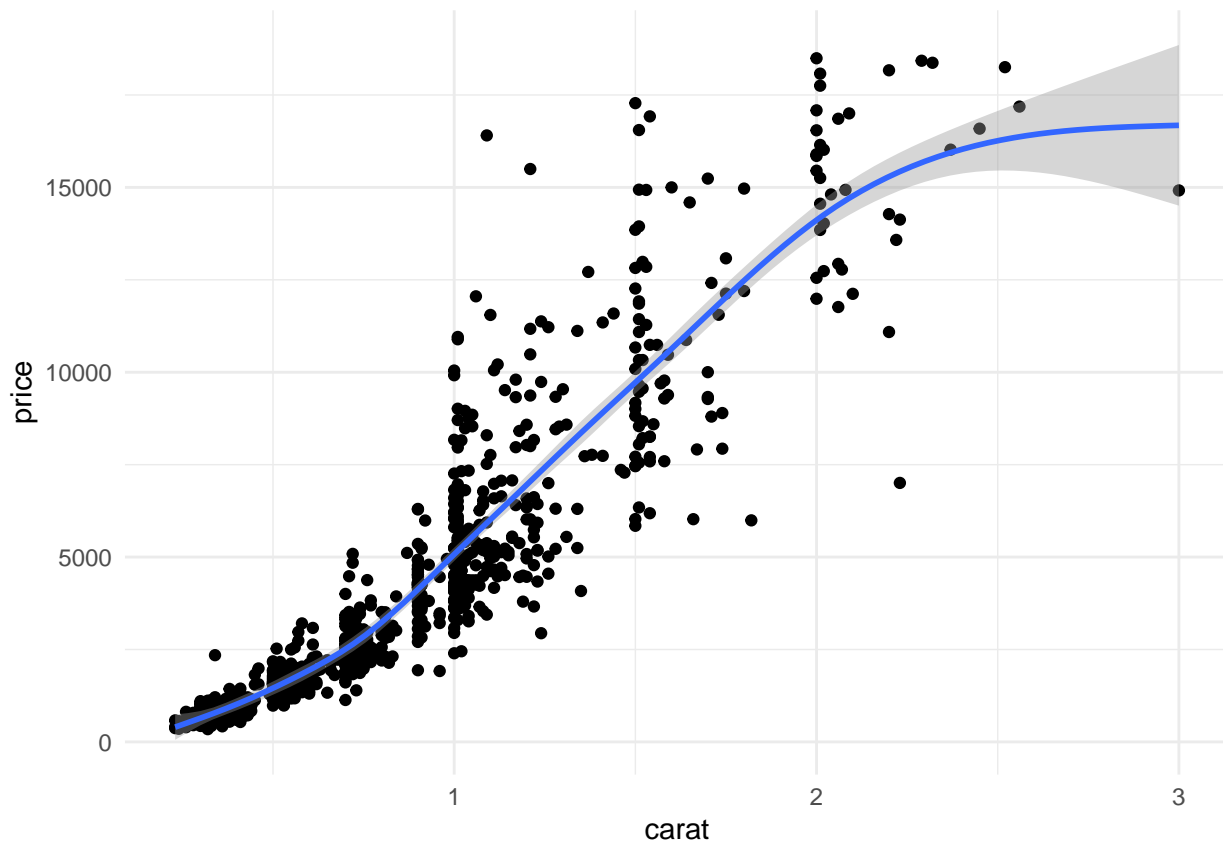
2.TWO VARIABLE

Both Continuous

Q:What is the relationship between 'carat' and 'price' as shown in this plot?

```
set.seed(42)
diamonds %>%
  sample_n(1000) %>%
  ggplot()+
    geom_point(aes(x=carat,
                   y=price)) +
    geom_smooth(aes(x=carat,
                   y=price)) +
    theme_minimal()

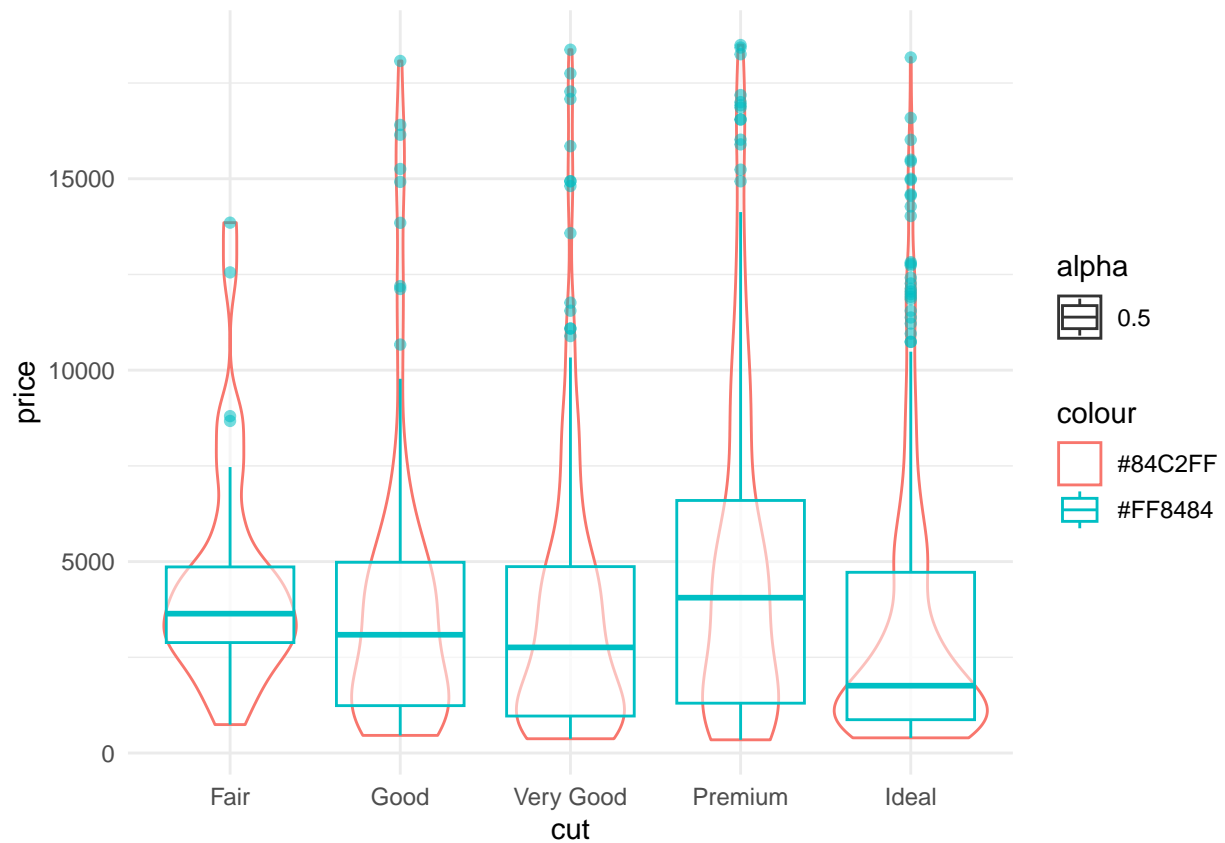
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



ONE Discrete and ONE Continuous

Q:How does the distribution of 'price' vary across different 'cut' categories?

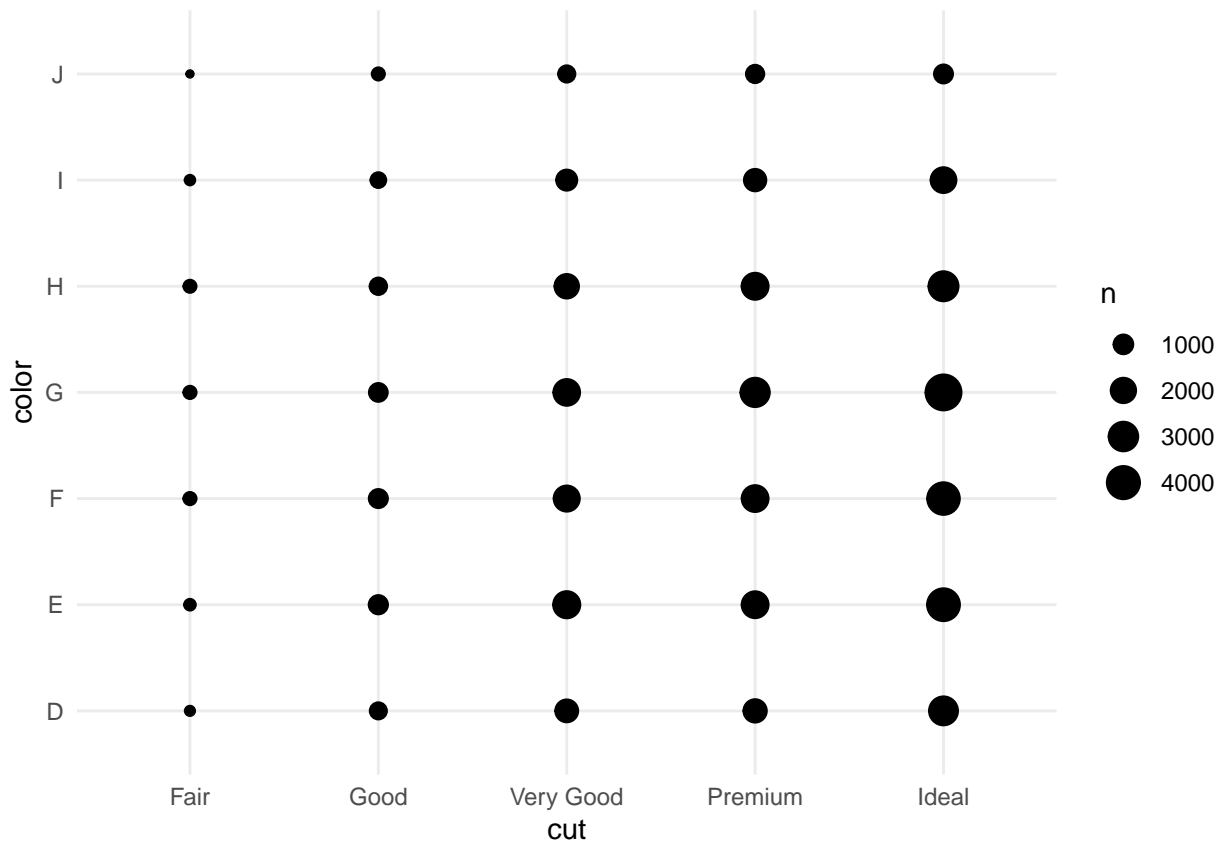
```
set.seed(42)
diamonds %>%
  sample_n(1000) %>%
  ggplot()+
    geom_violin(aes(x=cut,
                    y=price,
                    alpha=0.5,
                    color="#84C2FF")) +
    geom_boxplot(aes(x=cut,
                     y=price,
                     alpha=0.5,
                     color="#FF8484")) +
    theme_minimal()
```



Both Discrete

Q: What does this plot tell us about the relationship between 'cut' and 'color'?

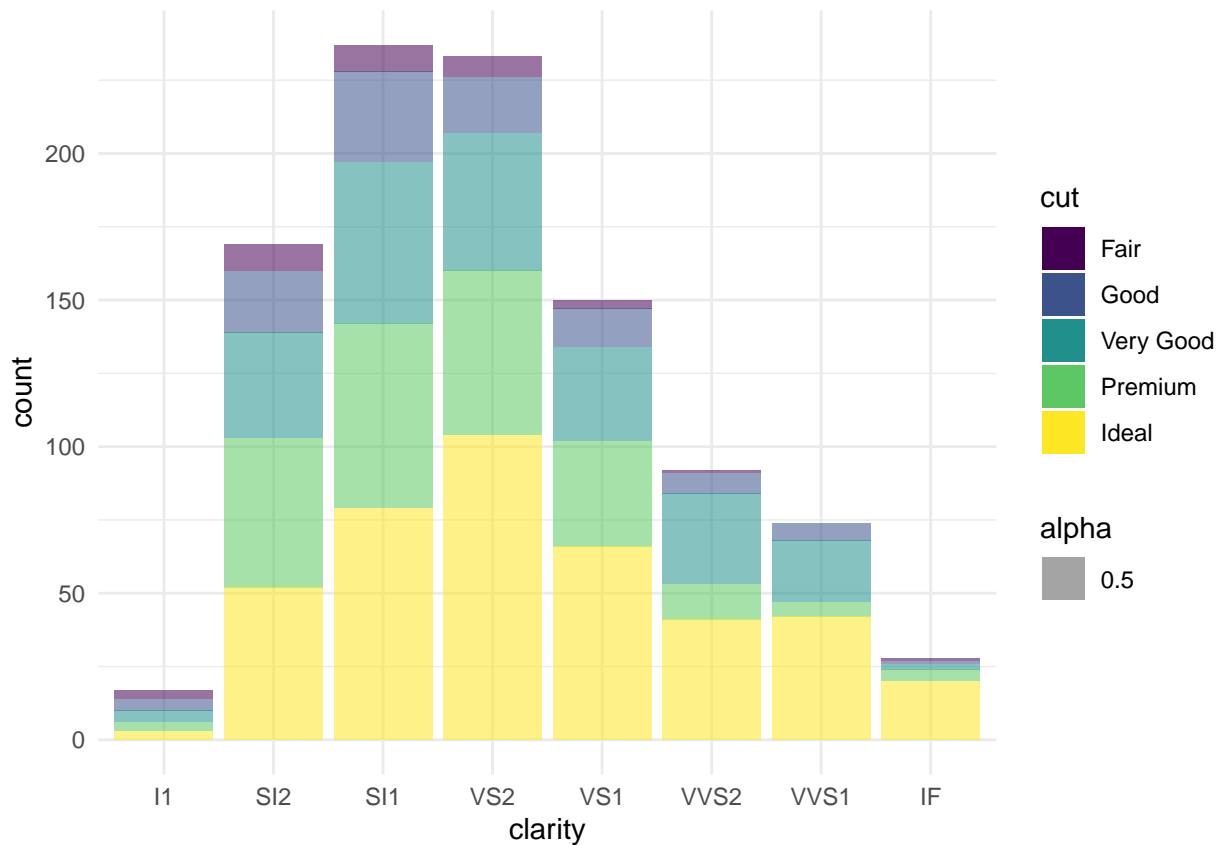
```
set.seed(42)
diamonds %>%
  #(1000) %>%
  ggplot()+
  geom_count(aes(cut,color)) +
  theme_minimal()
```



Both Discrete 2

Q: How does the distribution of 'cut' vary across different 'clarity' categories?

```
set.seed(42)
diamonds %>%
  sample_n(1000) %>%
  ggplot()+
    geom_bar(aes(x=clarity,
                 fill=cut,
                 alpha=0.5)) +
  theme_minimal()
```

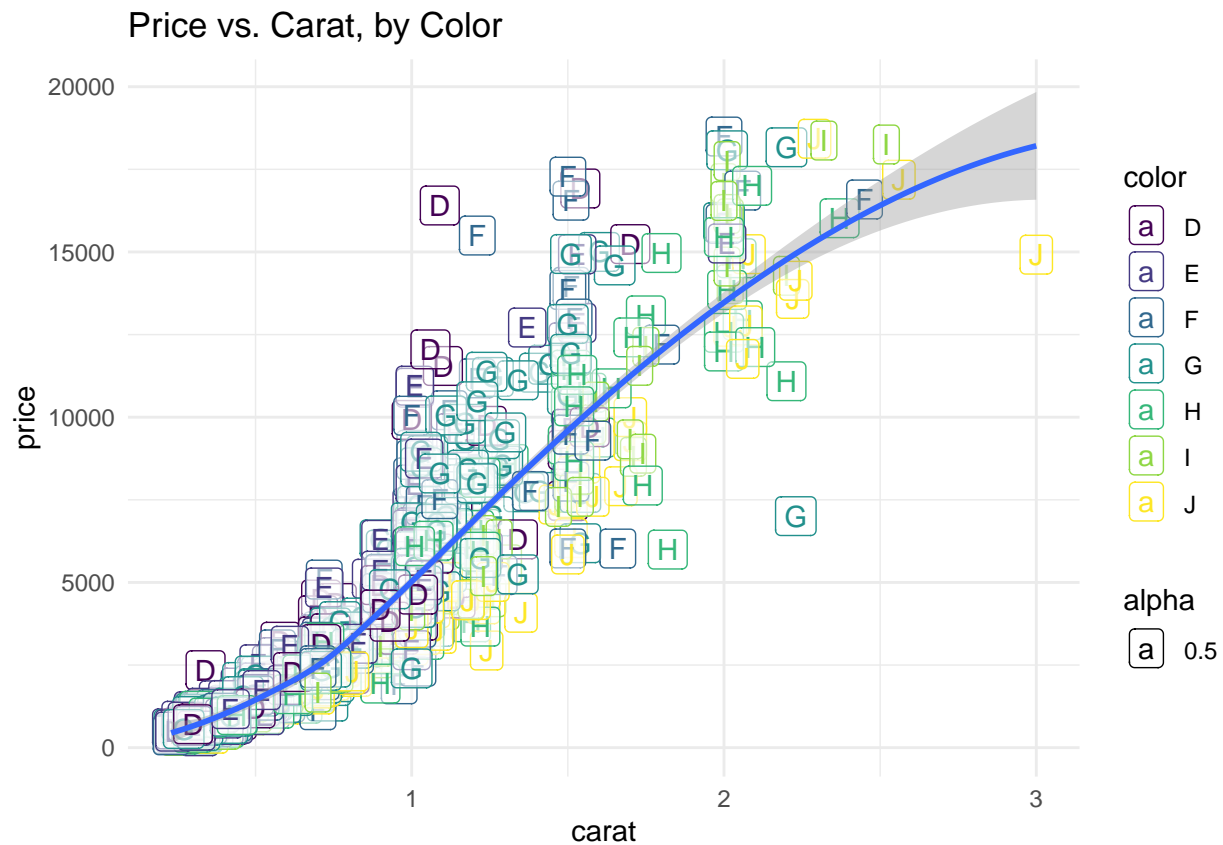


3.THREE VARIABLE

Q: How does the price of diamonds vary with carat, and how does this relationship differ across different colors?

```
set.seed(42)
diamonds %>%
  sample_n(1000) %>%
  ggplot()+
    geom_label(aes(x=carat,
                  y=price,
                  label=color,
                  color=color,
                  alpha=0.5)) +
    geom_smooth(aes(x=carat,
                  y=price),
                method = 'loess') +
    theme_minimal()+
    labs(title = "Price vs. Carat, by Color")
```

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

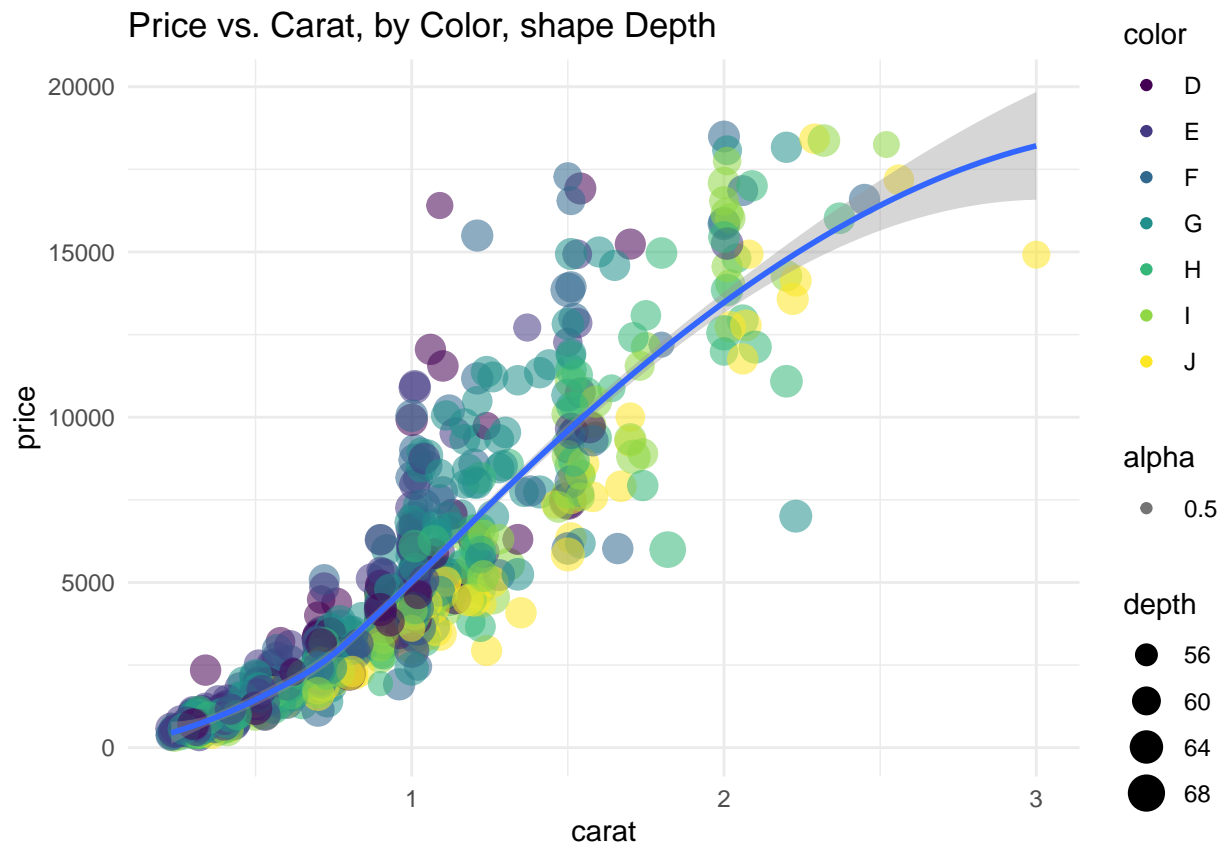


4.FOUR VARIABLE

Q:How does the price of diamonds relate to carat, considering the size and color of the points?

```
# Scatter plot: price vs. carat, colored by color
set.seed(42)
ggplot(diamonds %>% sample_n(1000), aes(x = carat,
                                         y = price)) +
  geom_point(mapping = aes(size = depth,
                           color = color,
                           alpha = 0.5)) +
  geom_smooth(method = 'loess') +
  theme_minimal() +
  labs(title = "Price vs. Carat, by Color, shape Depth")
```

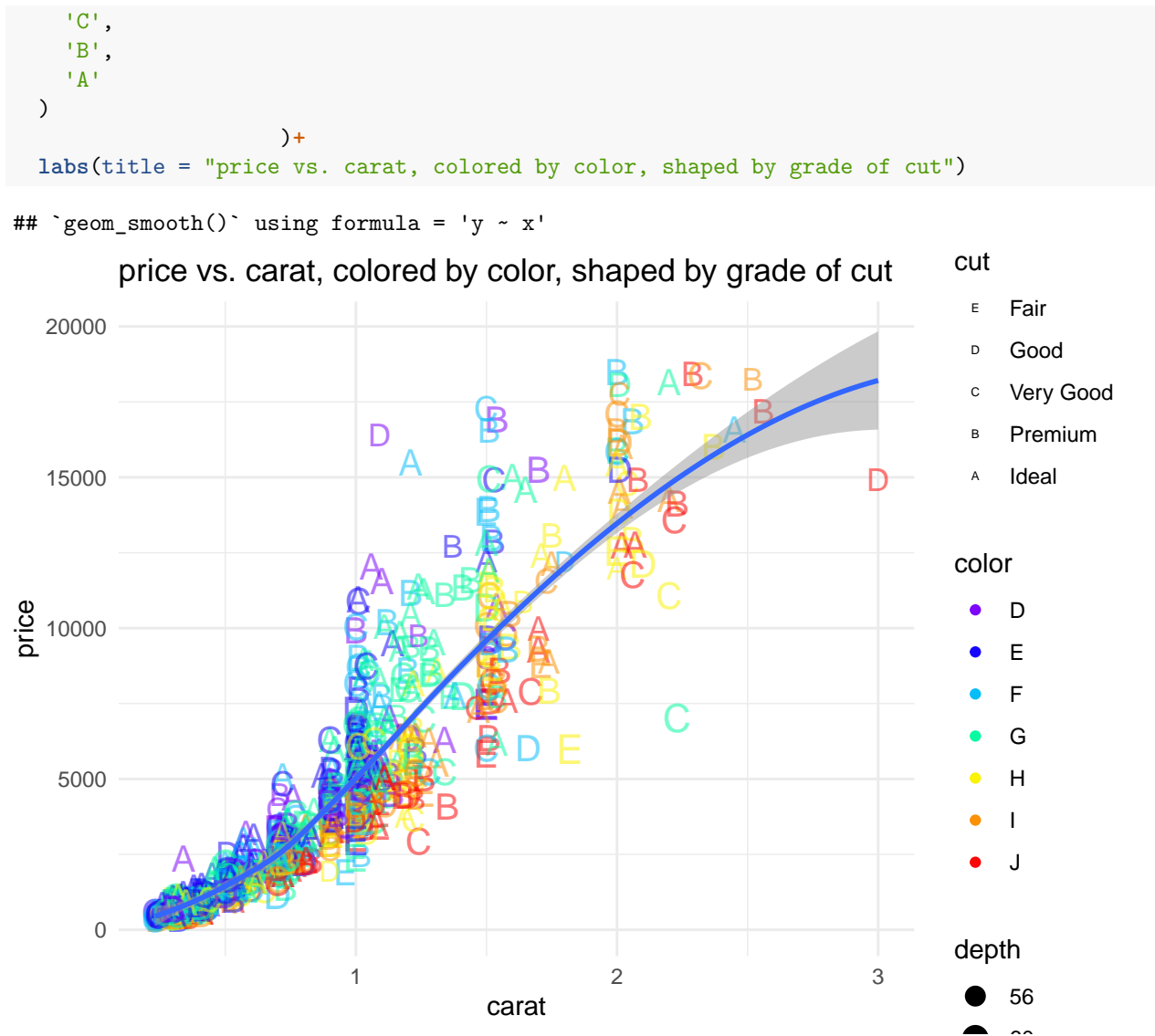
```
## `geom_smooth()` using formula = 'y ~ x'
```

5.FIVE VARIABLE

Q:How does the distribution of diamond 'color' relate to 'carat' and 'price', while also considering 'depth' and 'cut'

```
# Scatter plot: price vs. carat, colored by color, shaped by grade of cut
set.seed(42)
ggplot(diamonds %>% sample_n(1000), aes(x = carat,
                                         y = price)) +
  geom_point(mapping = aes(size = depth,
                           color = color,
                           shape = cut,
                           alpha = 0.4)) +
  geom_smooth(method = 'loess', alpha = 0.5) +
  theme_minimal() +
  scale_color_manual(values = c(
    '#7D07FA',
    '#1707FA',
    '#07BDFA',
    '#07FAA1',
    '#FAF207',
    '#FA9107',
    '#FA0707'
  )) +
  scale_shape_manual(values = c(
    'E',
    'D',
```



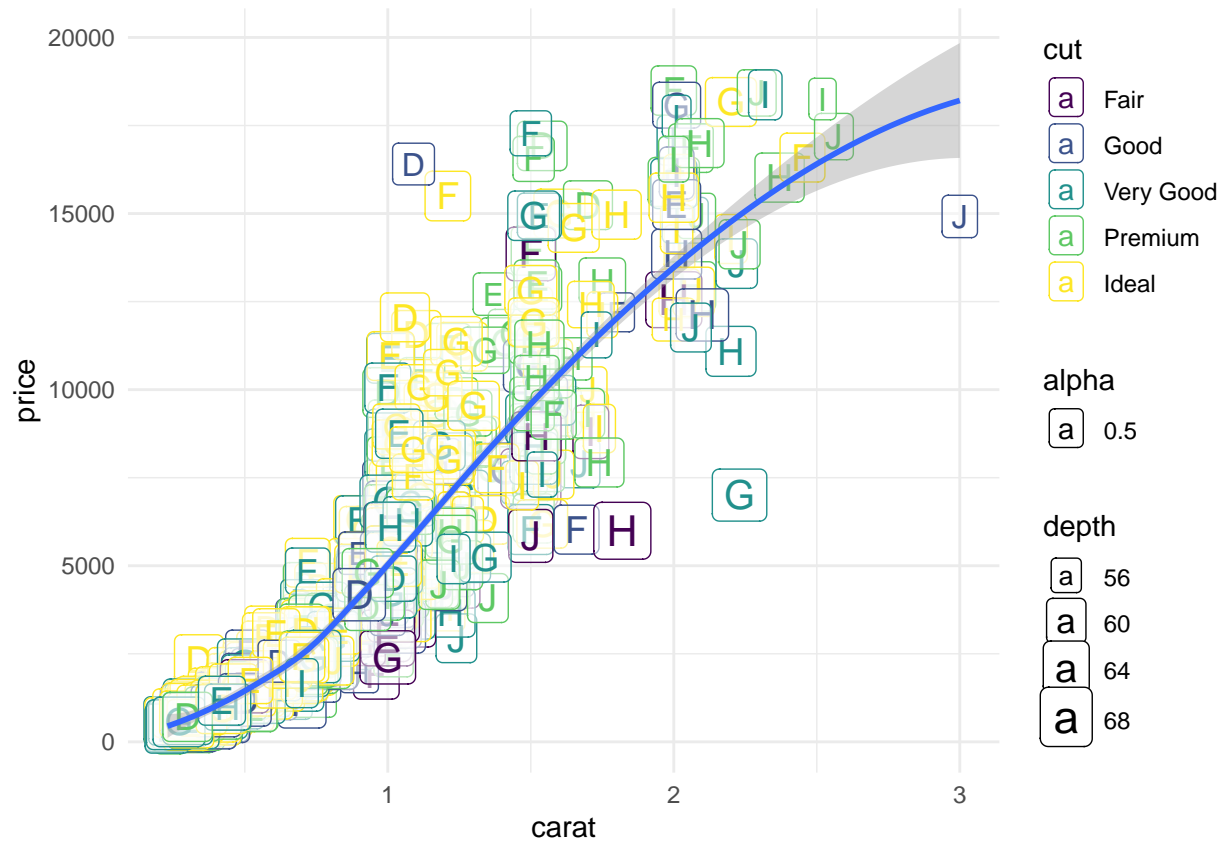
Q:How does the distribution of diamond 'color' relate to 'carat' and 'price', while also considering 'depth' and 'cut'

```

set.seed(42)
diamonds %>%
  sample_n(1000) %>%
  ggplot()+
    geom_label(aes(x=carat,
                  y=price,
                  label=color,
                  size = depth,
                  color=cut,
                  alpha=0.5)) +
    geom_smooth(aes(x=carat,
                  y=price),
                method = 'loess') +
    theme_minimal()

```

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



This is the end of the project.