Data Visualization Bootcamp - HW01

Arraya

2023-09-06

Instruction

Create 5 charts from diamonds dataset and knit it into PDF.

Get Start

To use the **diamonds** dataset and **ggplot2**, tidyverse package must be installed. To have more theme options, ggthemes package needed to be installed.

```
## install package
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

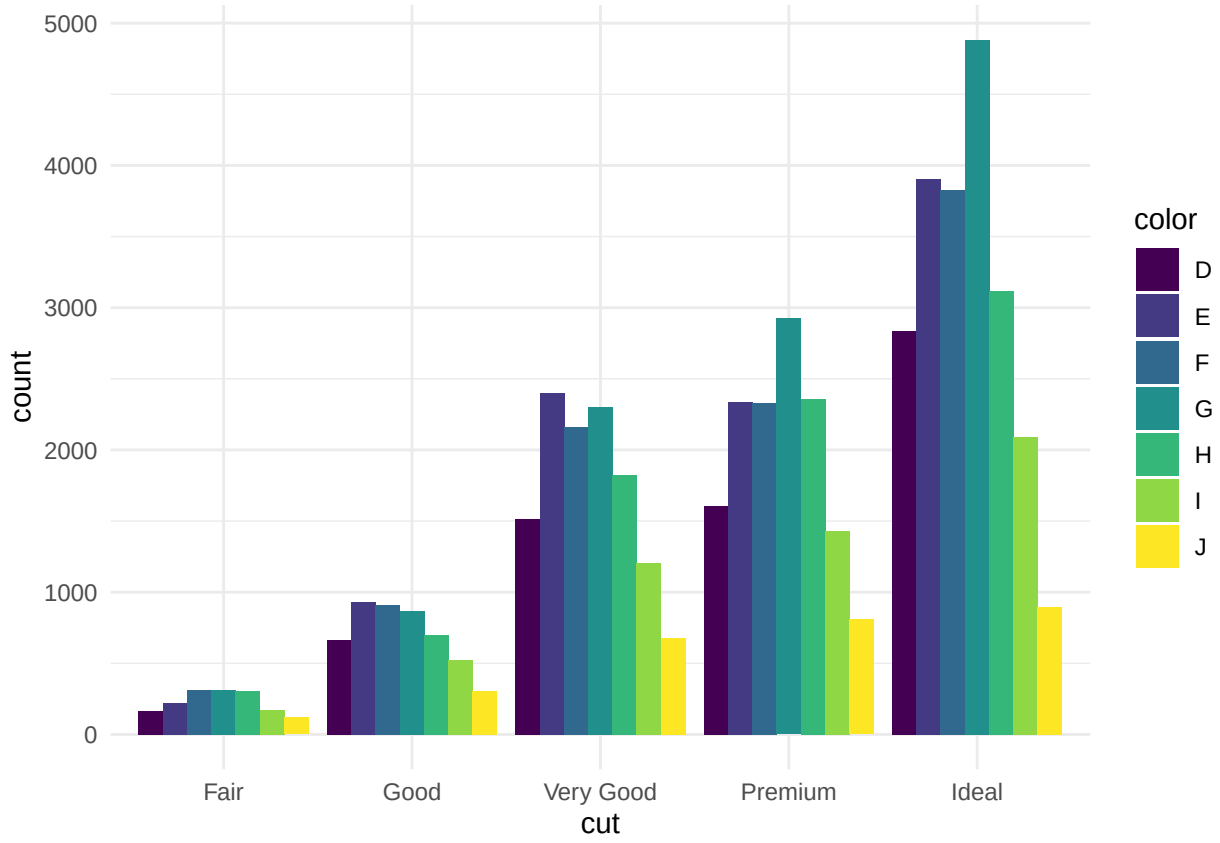
```
library(ggthemes)

## review dataset
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.~
```

1st Chart

```
ggplot(diamonds, aes(cut, fill = color)) +
  geom_bar(position = "dodge") +
  theme_minimal()
```

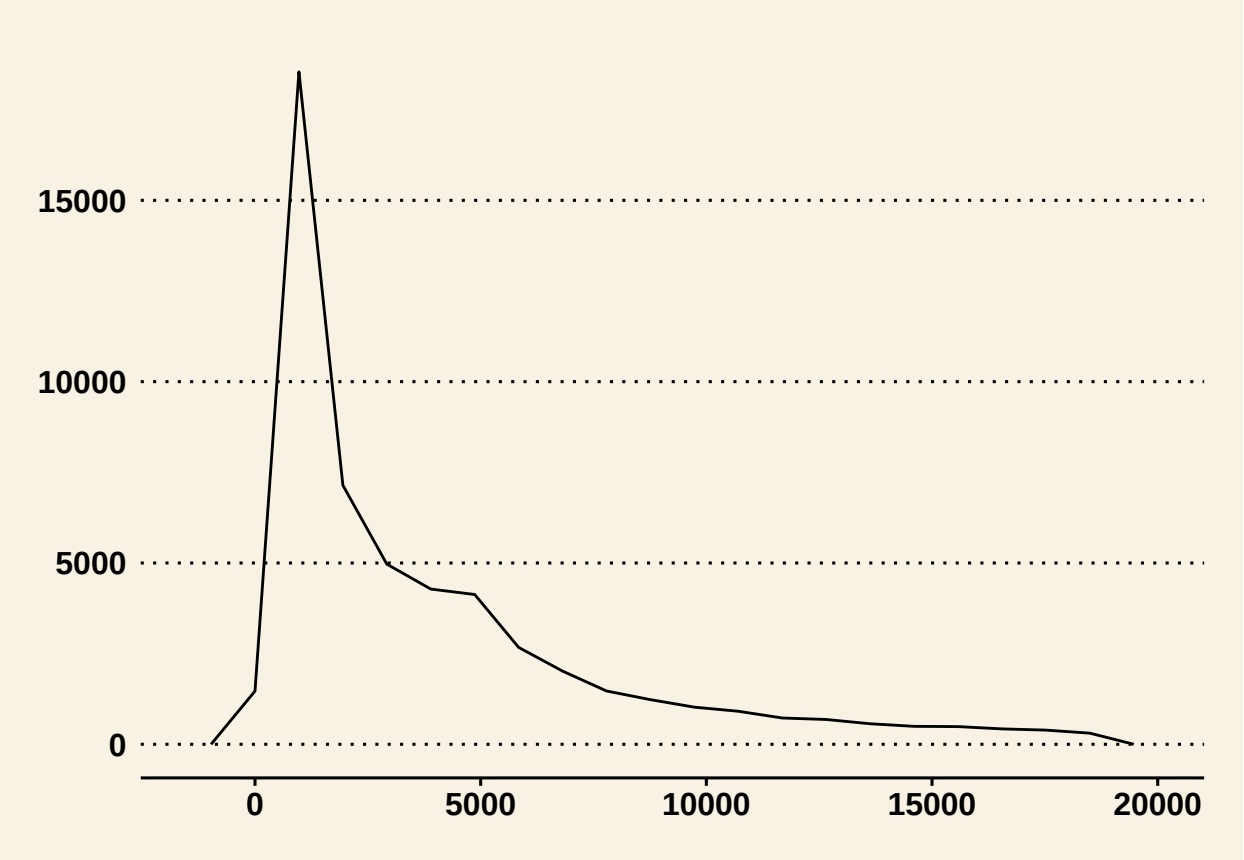


Quality of diamonds with color

- This chart shows the amount of diamonds grouped by the quality and color.

2nd Chart:

```
ggplot(diamonds, aes(price)) +
  geom_freqpoly(bins=20) +
  theme_wsj()
```



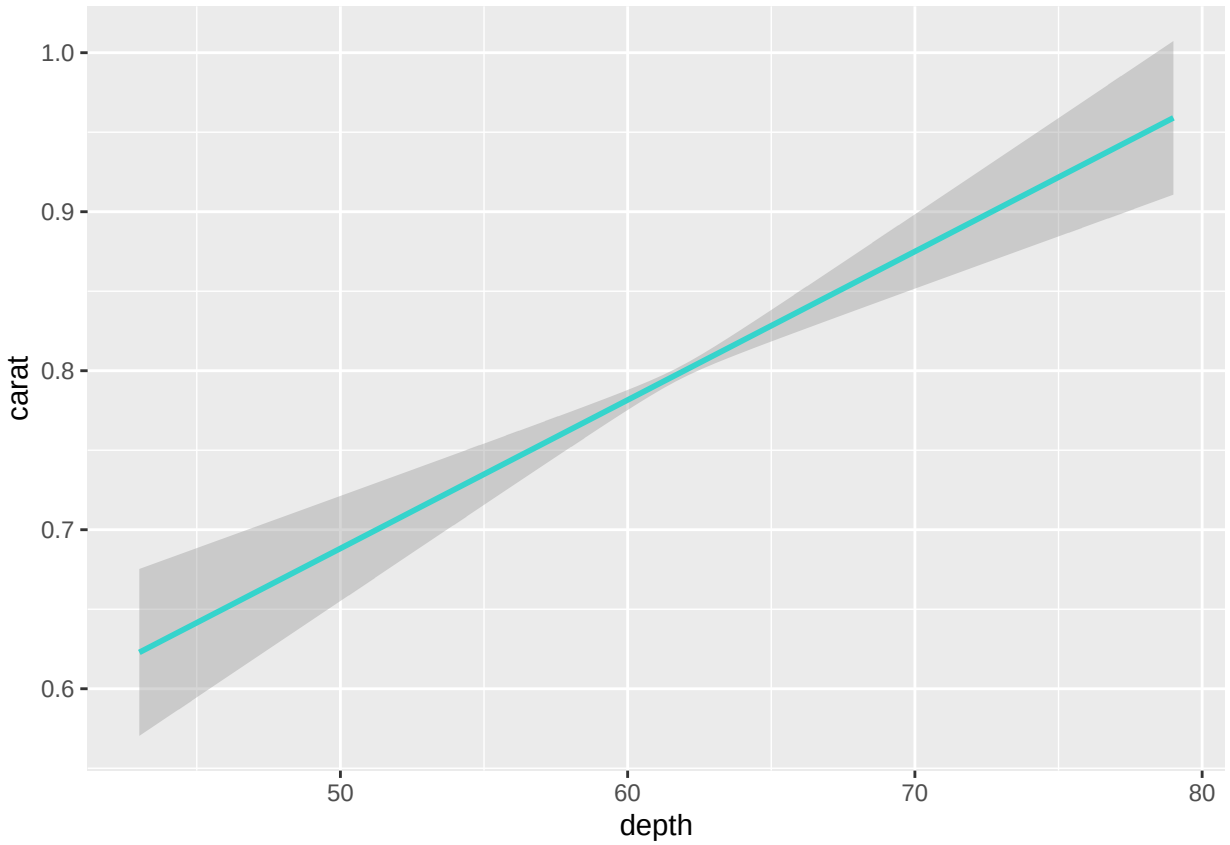
Price values

- The chart shows the price values of diamonds

3rd Chart

```
ggplot(diamonds, aes(depth, carat)) +
  geom_smooth(color = "#35d4cc", method = "lm")
```

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

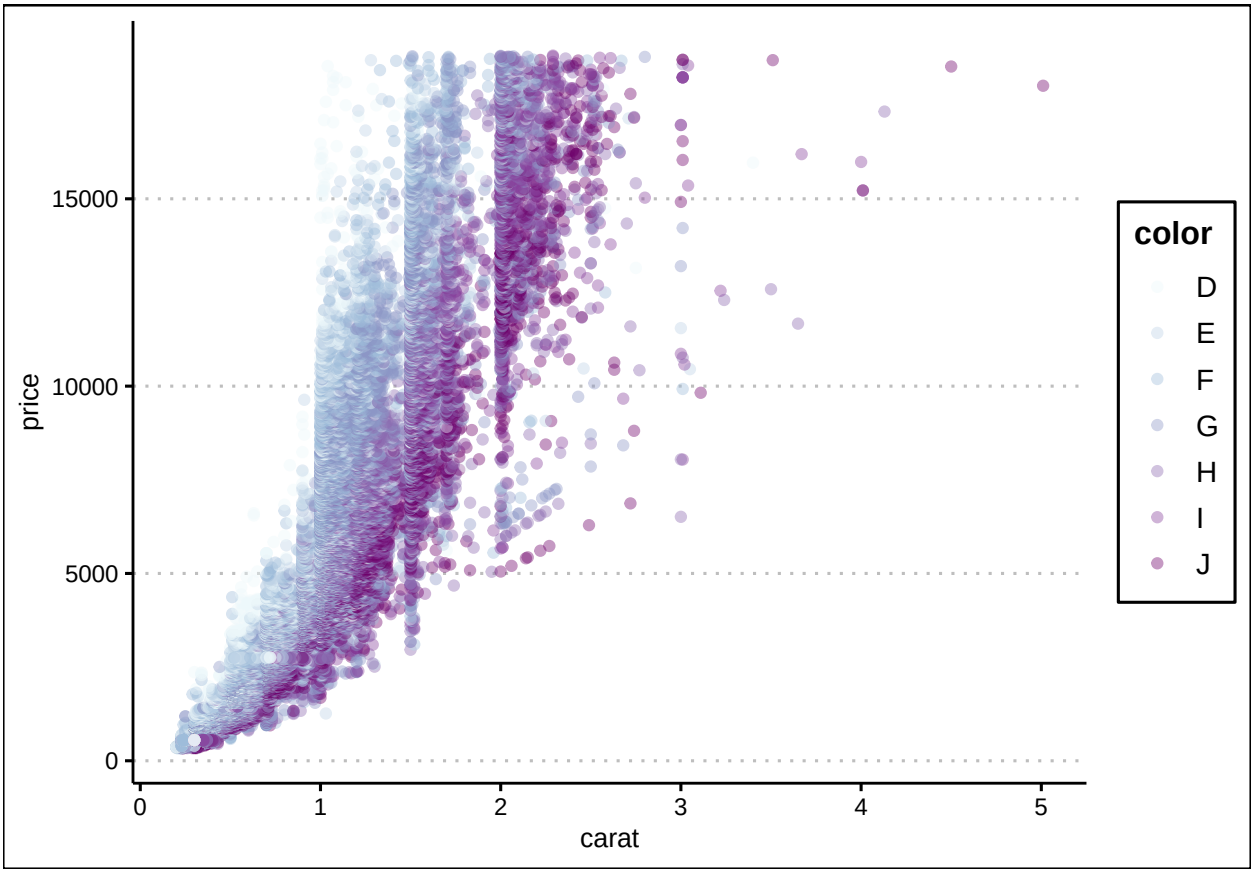


The weight and depth percentage of diamonds

- It can be implied from the chart that the depth percentage of diamonds will be higher if they are heavier.

4th Chart

```
ggplot(diamonds, aes(carat, price, col=color)) +
  geom_point(alpha = 0.4) +
  scale_color_brewer(type = "seq", palette = 3) +
  theme_clean()
```

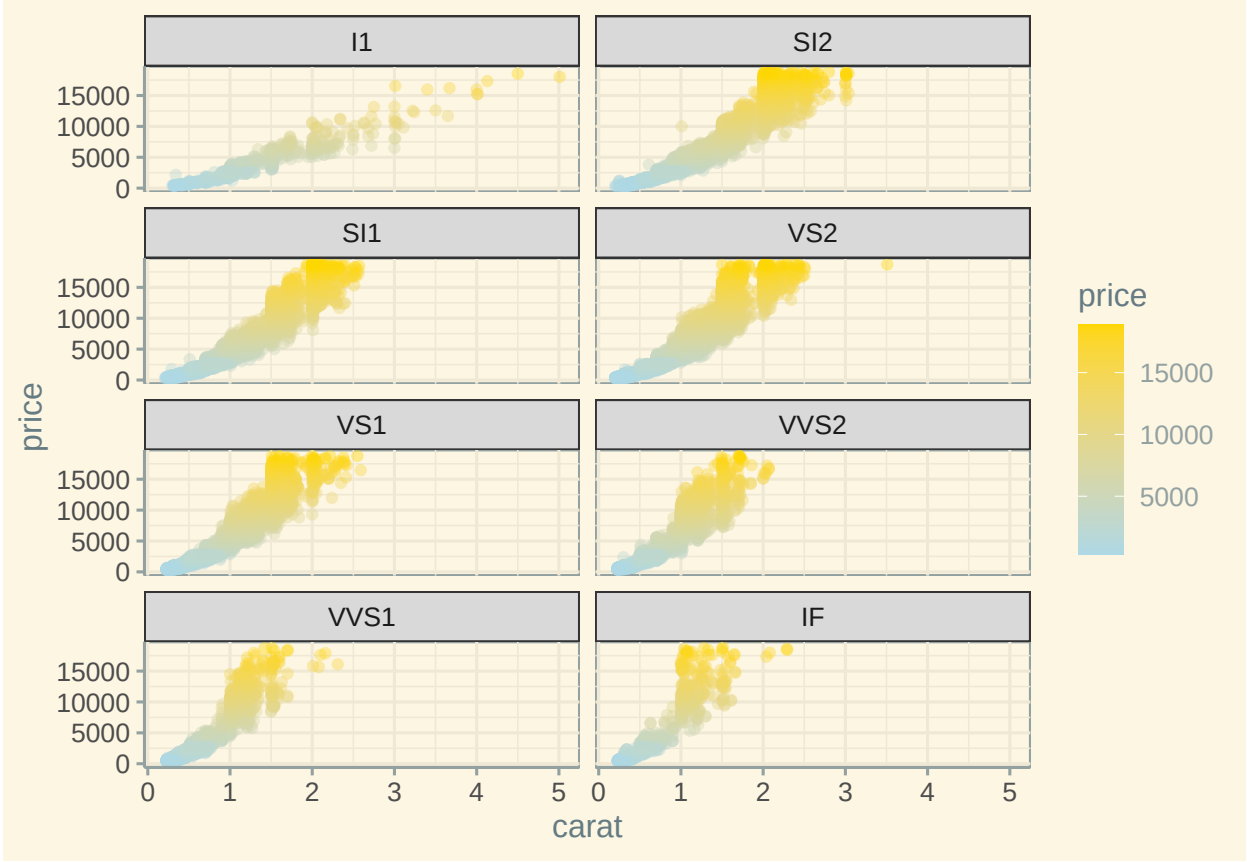


The weight and price of diamonds mapped with color

- Chart shows that carat affects the price of diamonds.

5th Chart

```
ggplot(diamonds, aes(carat, price, col=price)) +
  geom_point(alpha = 0.4) +
  scale_color_gradient(low="lightblue",
                       high="gold") +
  facet_wrap(~clarity, ncol = 2) +
  theme_solarized()
```



Weight, price, and clarity of diamonds

- This can be concluded that the heavier and clearer the diamonds, the expensive it will be.