Batch_06_Homework_Data_viz

YO

Library

```
install.packages(c("tidyverse", "ggthemes"))
```

```
## Installing packages into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggthemes)
```

Review Data

```
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.~
```

Chart 1 Show relationship between price and carat

```
set.seed(10)
ggplot(sample_n(diamonds,2500), aes(price, carat)) +
  geom_point(alpha = 0.8,
             mapping = aes(color = clarity)) +
  geom_smooth(color = "red") +
```

```
  scale_colour_brewer(palette = "YlGnBu") +
  theme_minimal() +
  labs(title = "Relationship between price and carat",
       x = "Price(USD)",
       y = "Carat")
```

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

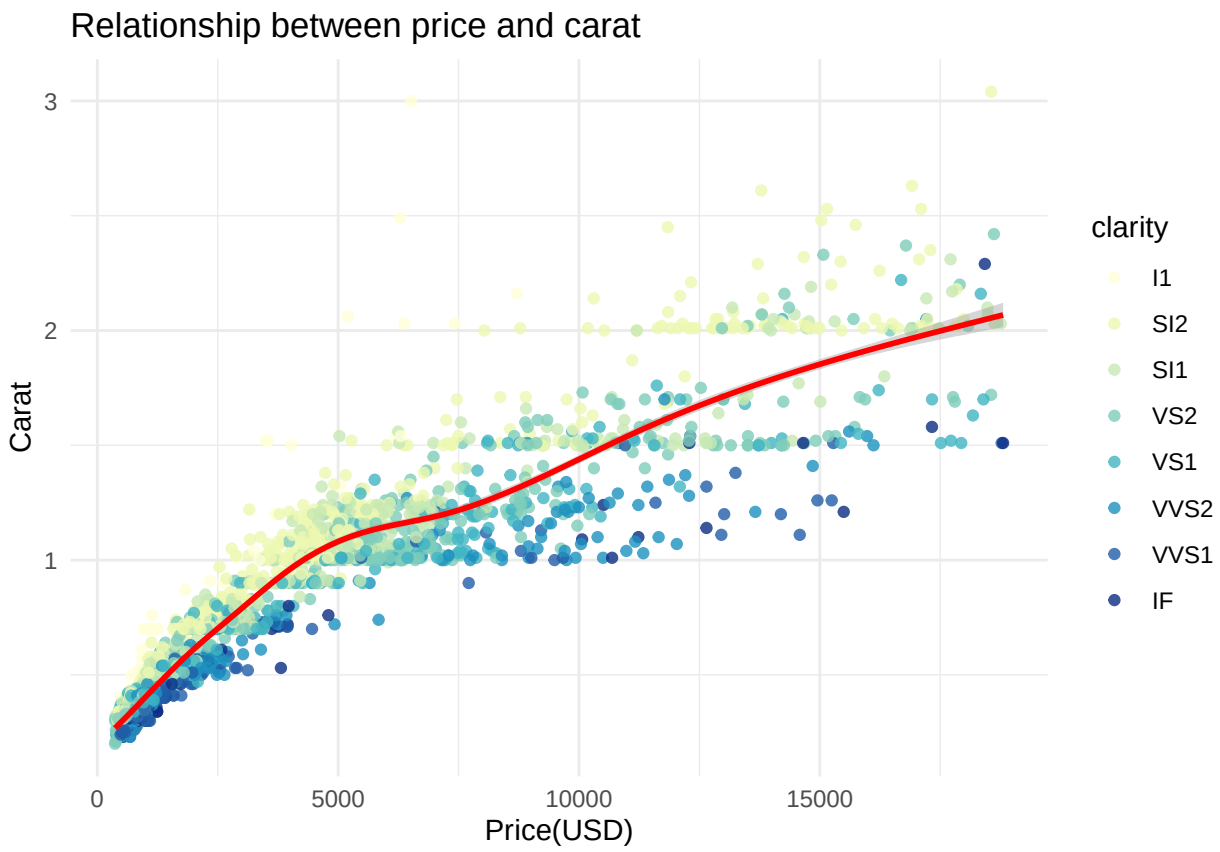


Chart 2 Relationship price and carat of each color

```
set.seed(10)
ggplot(sample_n(diamonds,2500), aes(price, carat, color = color)) +
  geom_point() +
  geom_smooth(method = "lm", col = "#05a3ff") +
  scale_color_brewer(palette = "Pastel2") +
  facet_wrap(~color) +
  labs(title = "Relationship between price and carat",
       subtitle = "Group by color",
       x = "Price(USD)",
       y = "Carat") +
  theme_minimal()
```

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

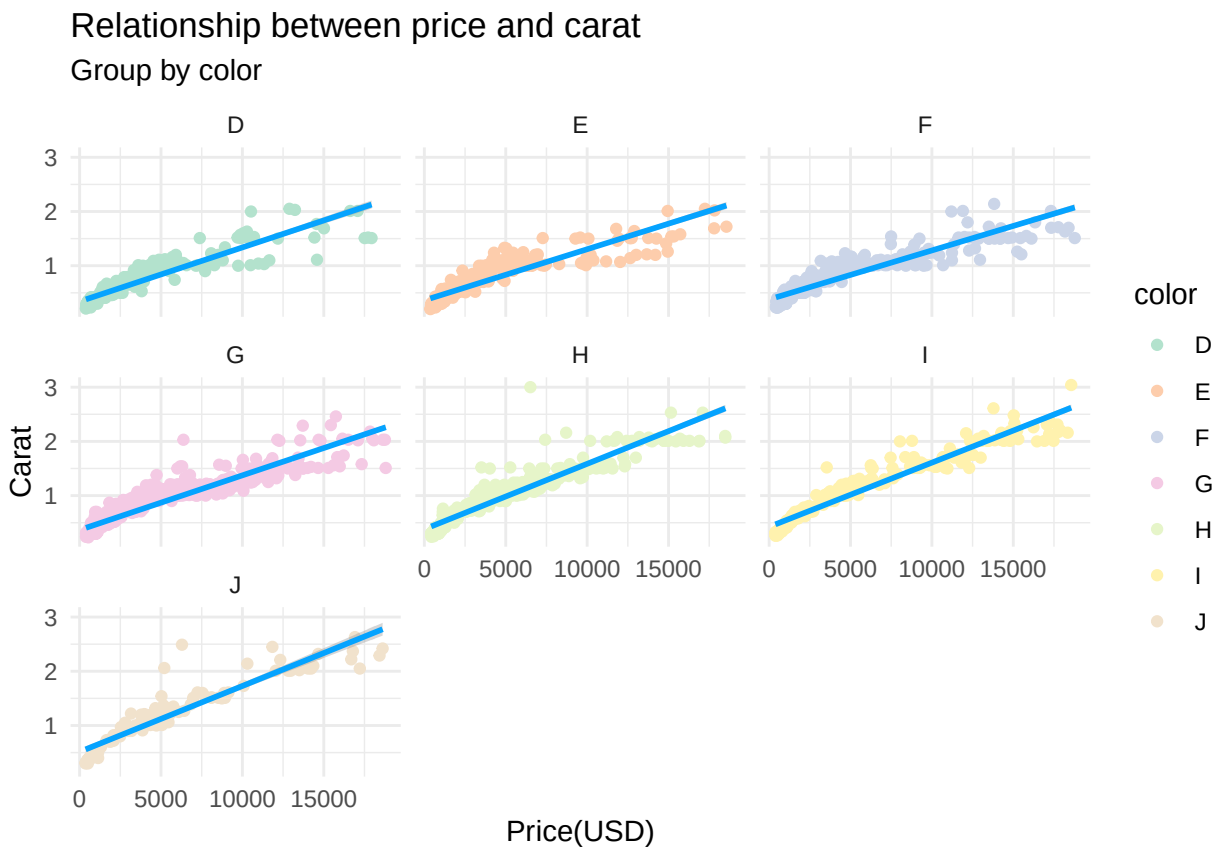


Chart 3 Counting each diamond color

```
ggplot(diamonds, aes(color, fill = color)) +
  geom_bar() +
  labs(title = "Each diamonds color",
       x = "Color",
       y = "Count")
```

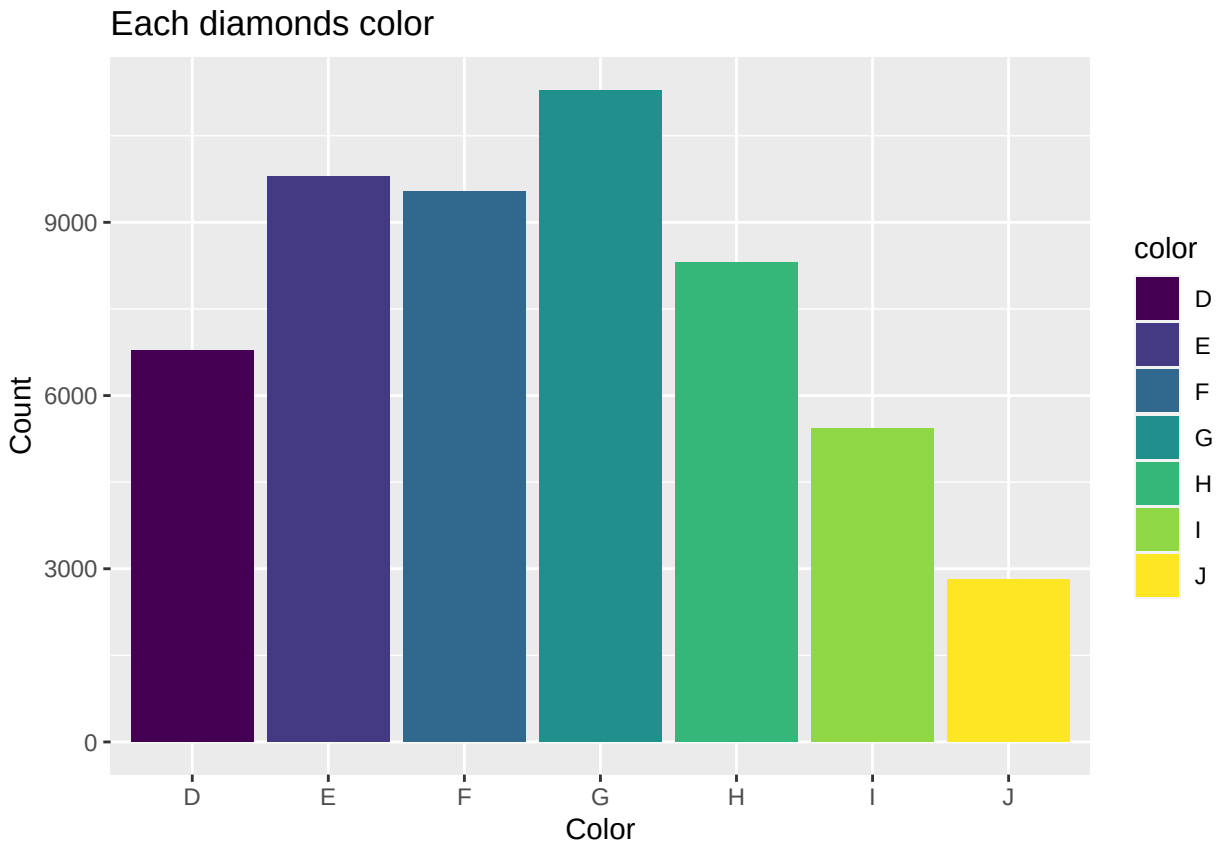


Chart 4 Relationship between cut and clarity

```
set.seed(10)
ggplot(sample_n(diamonds, 2500), aes(cut, fill = clarity)) +
  geom_bar(position = "fill") +
  labs(title = "Relationship between cut and clarity")
```

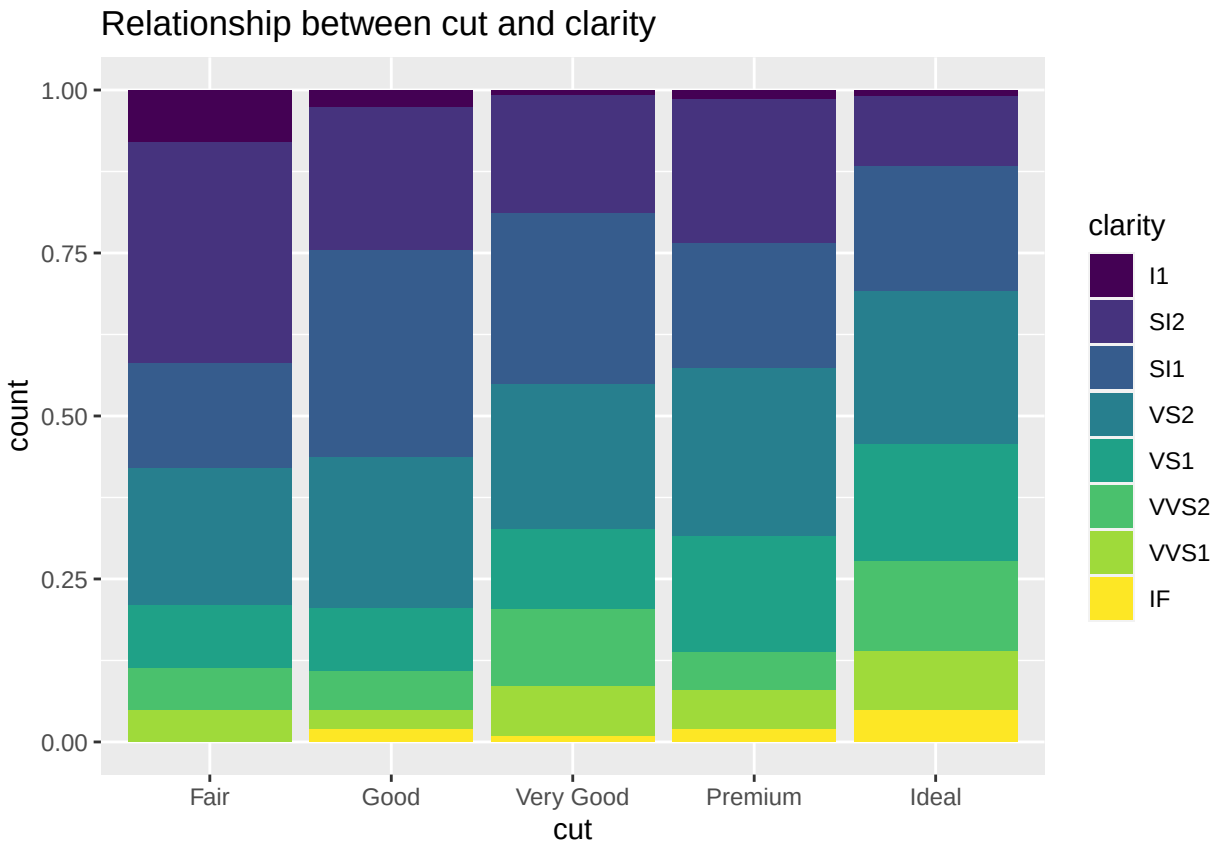
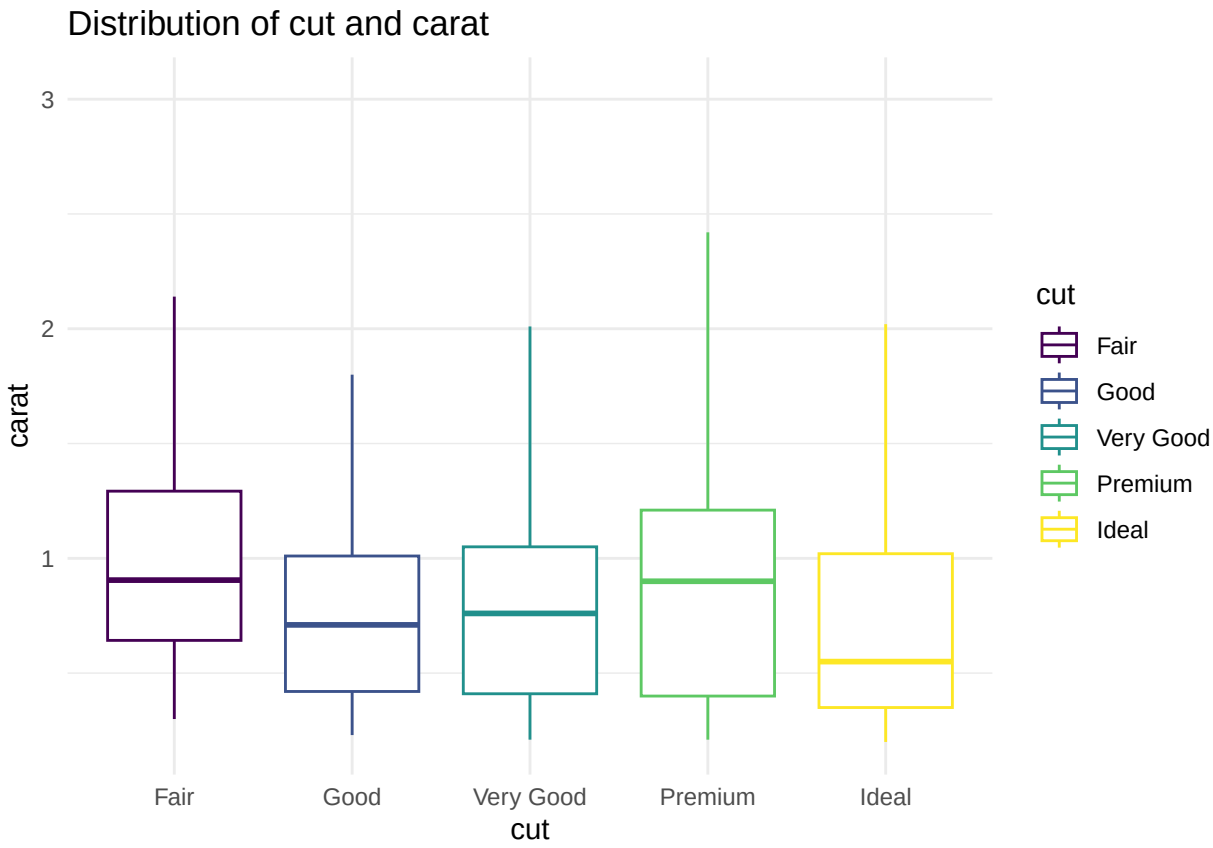


Chart 5 Distribution of cut and carat

```
set.seed(10)
ggplot(sample_n(diamonds,2500), aes(cut, carat, color = cut)) +
  geom_boxplot(outlier.shape = NA) +
  theme_minimal() +
  labs(title = "Distribution of cut and carat")
```


Distribution of cut and carat
carat
3
2
1
Fair
Good
Very Good
Premium
Ideal
cut
cut
Fair
Good
Very Good
Premium
Ideal