Batch 6 - Data Visualization_HW

Vongsakorn Ommaruek

Install package & preview Data

```
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   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(patchwork)

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.~
```

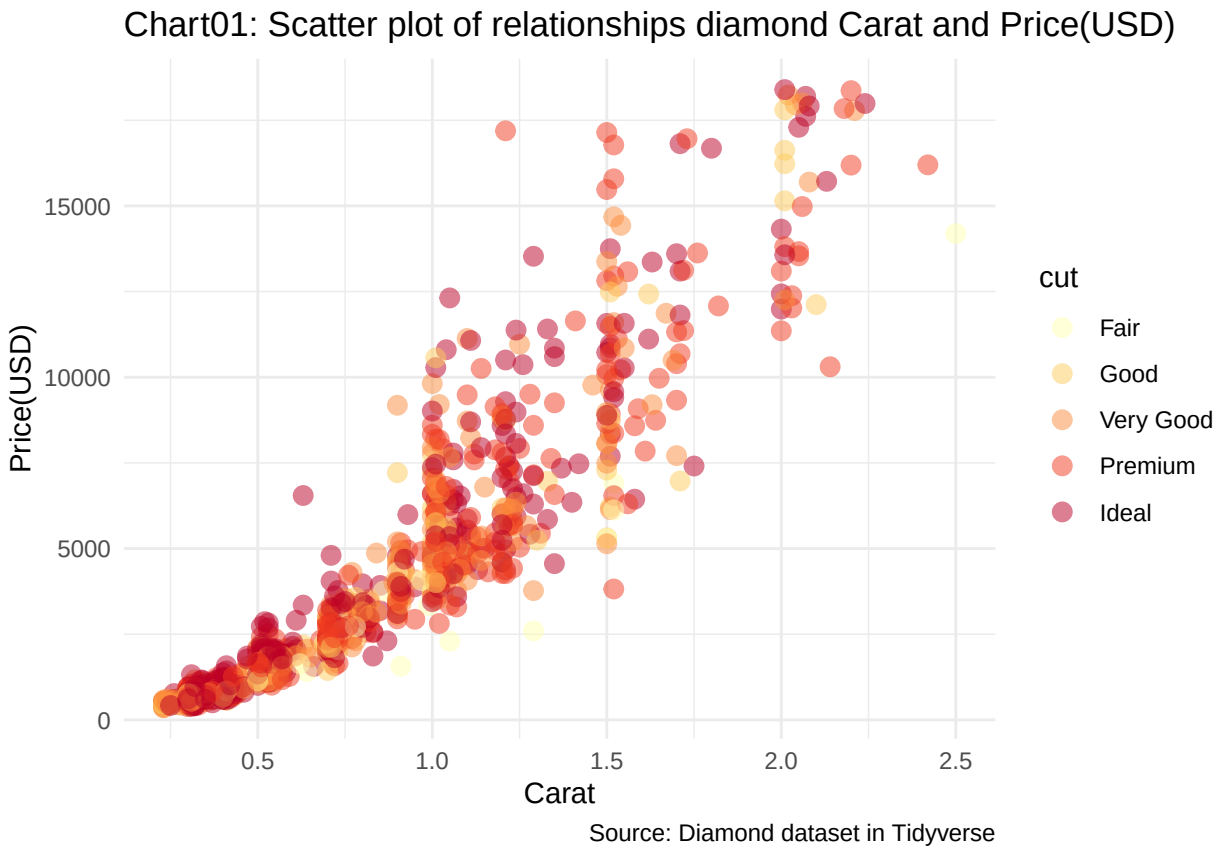
Data Description

- carat : Weight of the Diamond
- cut : Quality of the cut
- color : Diamond Color
- clarity : measurement of how clear the diamond
- depth : total depth percentage
- table : width of top of diamond relative to widest point
- price : Price in US dollars

Data Visualization

Chart01: The relationship between carat and price

```
set.seed(14)
sample_n(diamonds,1000) %>%
ggplot(aes(carat,price,color = cut))+
  geom_point(size = 3,
             alpha = 0.50)+
  theme_minimal()+
  scale_color_brewer(palette = "YlOrRd")+
  labs(
    title = "Chart01: Scatter plot of relationships diamond Carat and Price(USD)",
    x = "Carat",
    y = "Price(USD)",
    caption = "Source: Diamond dataset in Tidyverse")
```



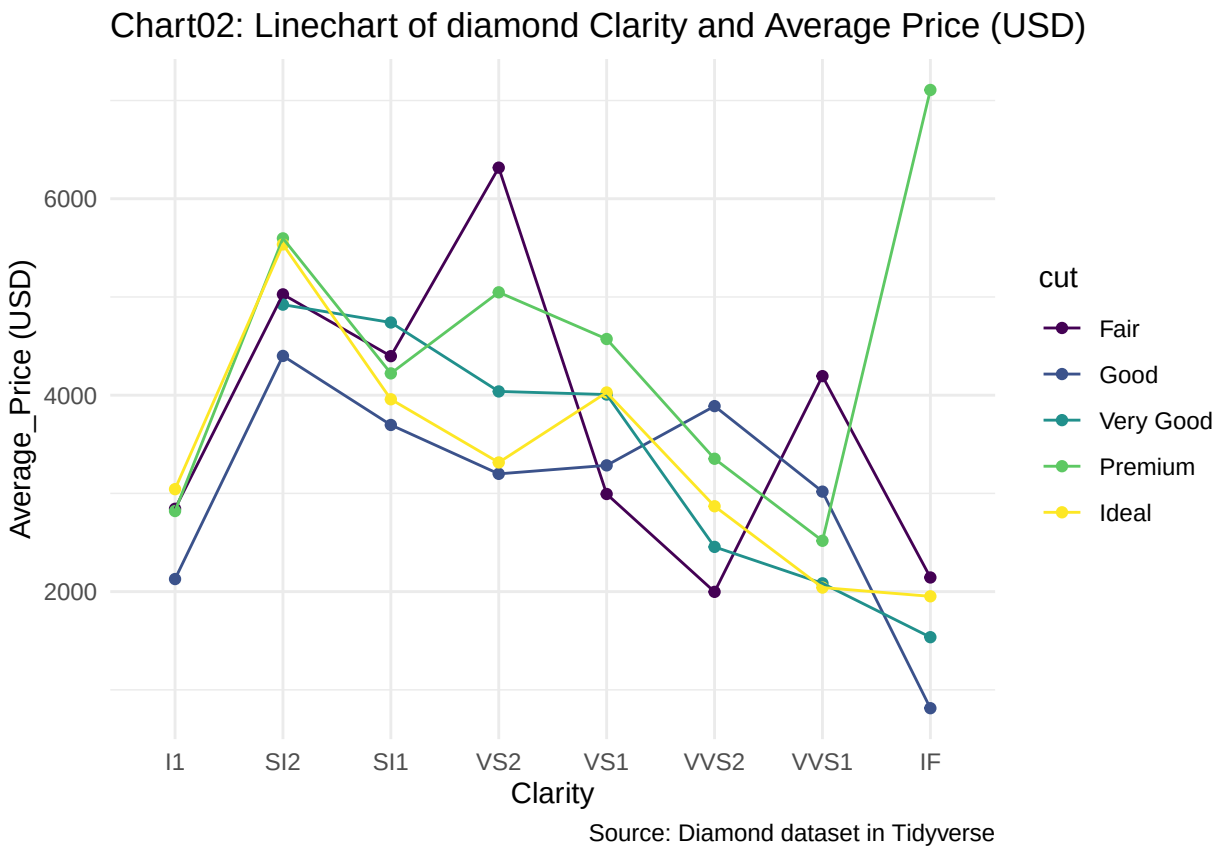
In scatter plot shows a positive correlation between carat and price.The higher the carat, the higher the price also.

Chart02: The relationship between clarity and price of diamonds sourcing by cutting

```
set.seed(03)
sample_n(diamonds,1000) %>%
group_by(clarity,cut) %>%
summarise(avg_price = mean(price)) %>%
ggplot(aes(clarity, avg_price,group = cut, col = cut))+
geom_point()+
```

```
geom_line()+
theme_minimal()+
labs(title = "Chart02: Linechart of diamond Clarity and Average Price (USD)",
x = "Clarity", y = "Average_Price (USD)",
caption = "Source: Diamond dataset in Tidyverse")
```

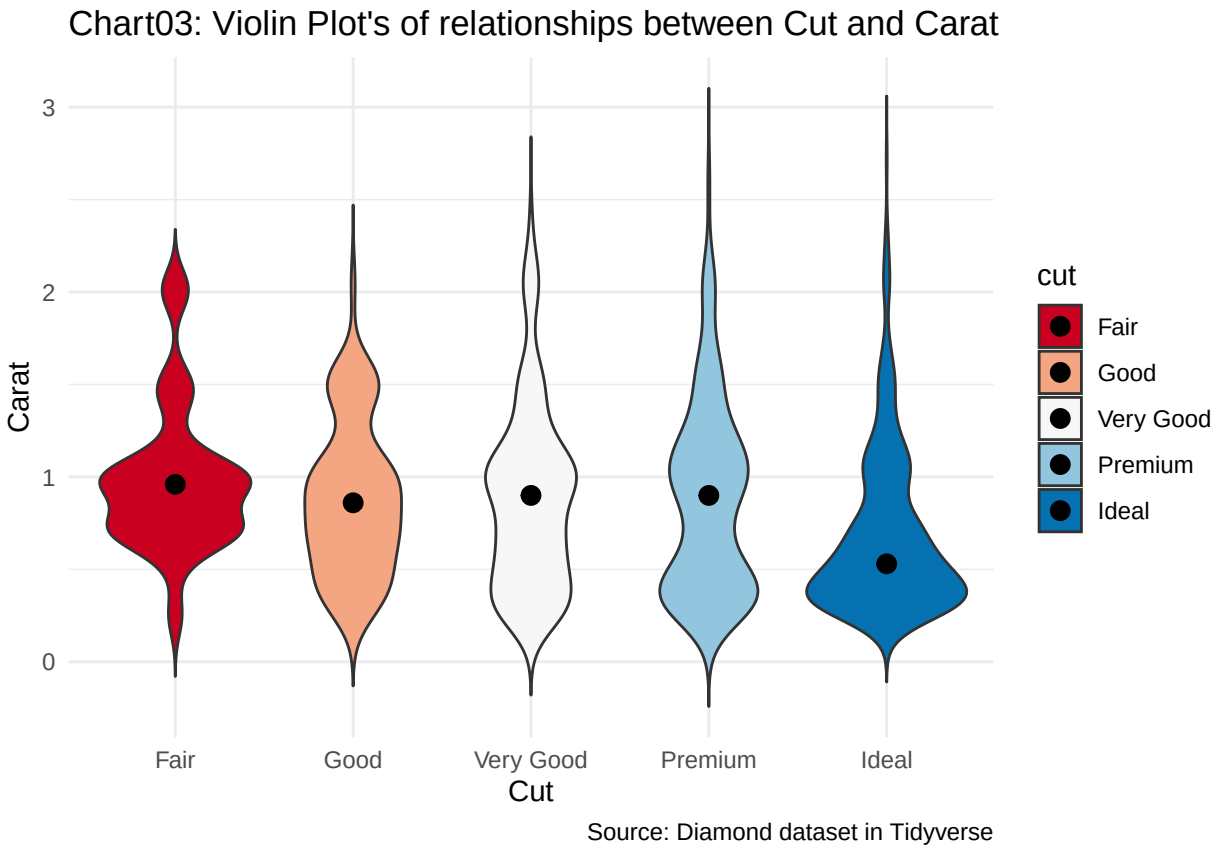
```
## `summarise()` has grouped output by 'clarity'. You can override using the
## `.groups` argument.
```



In line chart shows a negative correlation between clarity and price.but if we consider quality of cut also.we will have Fair & Premium quality cut are abnormal.

Chart03: The relationship between a cutting & carat of diamonds

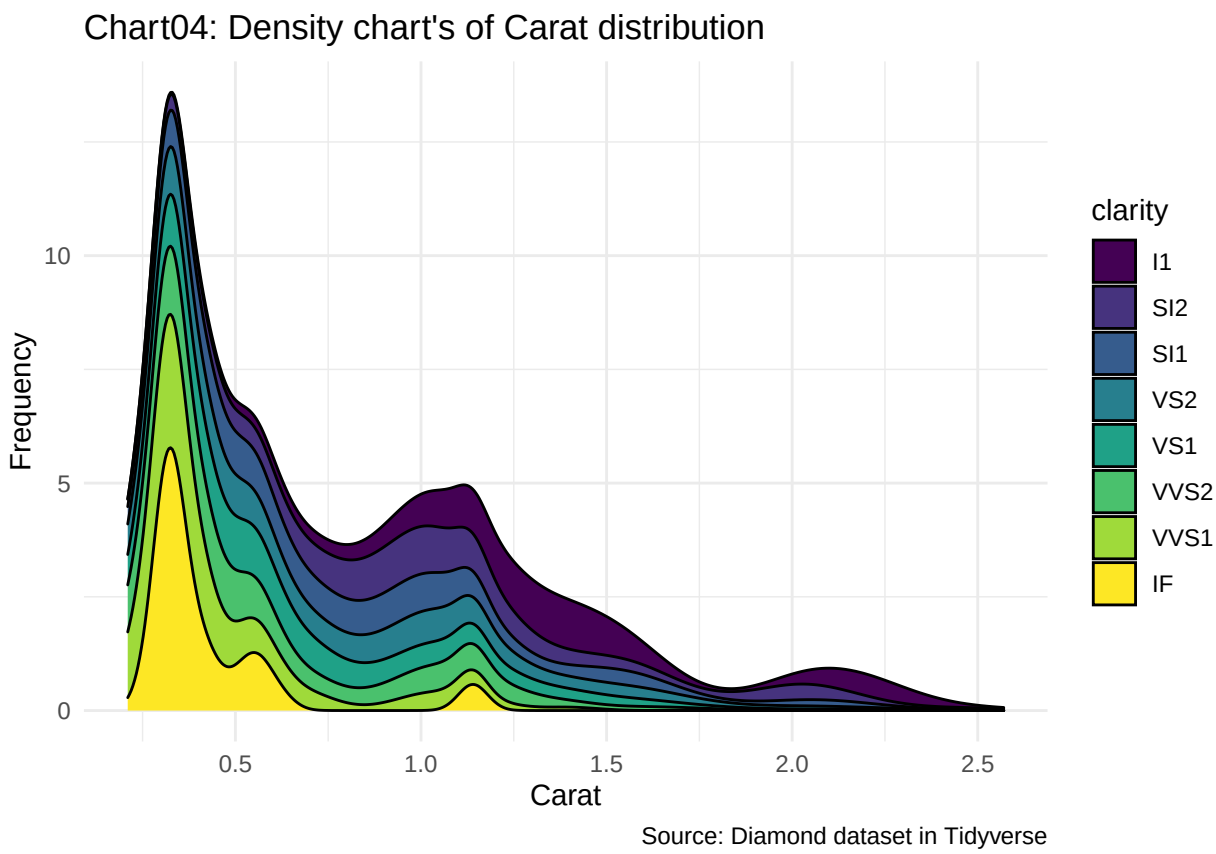
```
set.seed(25)
sample_n(diamonds, 1000) %>%
ggplot(aes(cut, carat, fill = cut)) +
geom_violin(trim = F) +
stat_summary(fun = median, geom = "point", size = 3, color = "black") +
theme_minimal() +
labs(
title = "Chart03: Violin Plot's of relationships between Cut and Carat",
x = "Cut",
y = "Carat",
caption = "Source: Diamond dataset in Tidyverse"
) +
scale_fill_brewer(palette = "RdBu")
```



In violin plot's is show destitution of Weight & quality cut of the Diamond.in chart will show in line production we have 1 carat is the most of product and has only ldeal quaility type only is less than 1 carrat.

Chart04: The relationship between a density & carat of diamonds

```
set.seed(69)
sample_n(diamonds,1000) %>%
ggplot(aes(carat, fill=clarity))+
geom_density(linetype="solid", position="stack")+
theme_minimal()+
labs(
title="Chart04: Density chart's of Carat distribution",
x ="Carat",
y = "Frequency",
caption = "Source: Diamond dataset in Tidyverse")
```

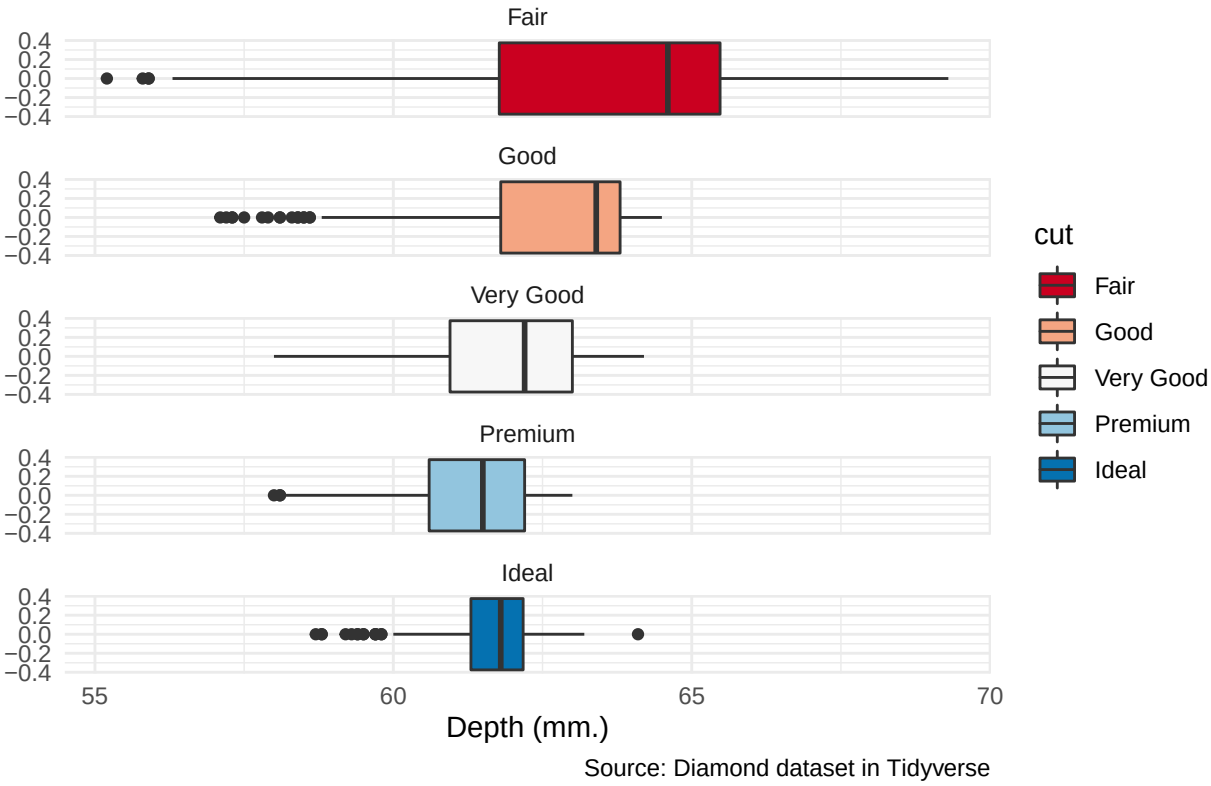


In density chart shows a negative correlation between carat and density.The higher the carat, the low the density.

Chart05: Box Plot's of relationships between Carat and Depth of cutting

```
set.seed(19)
sample_n(diamonds,2000) %>%
  ggplot(aes(y= depth,fill = cut))+
  geom_boxplot()+
  coord_flip()+
  theme_minimal()+
  facet_wrap(~cut,nrow = 5)+
  scale_fill_brewer(palette = "RdBu")+
  labs(
title="Chart05: Box Plot's of relationships between quality of cutting and Percentage of cutting",
y ="Depth (mm.)",
caption = "Source: Diamond dataset in Tidyverse")
```

Chart05: Box Plot's of relationships between quality of cutting and Percentag



Source: Diamond dataset in Tidyverse

In box plot shows a percentage of cutting as per quality cut.In Fair type we will cutting more than other quality and have a long in depth range also.