

Homework Data Viz

Prateep P

2023-05-16

Dataset: Diamonds

```
library(tidyverse)
library(patchwork)
```

```
diamonds %>%
  head(5)
```

```
## # A tibble: 5 × 10
##   carat cut      color clarity depth table price      x      y      z
##   <dbl> <ord>   <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1  0.23 Ideal    E     SI2     61.5    55   326   3.95   3.98   2.43
## 2  0.21 Premium E     SI1     59.8    61   326   3.89   3.84   2.31
## 3  0.23 Good     E     VS1     56.9    65   327   4.05   4.07   2.31
## 4  0.29 Premium I     VS2     62.4    58   334   4.2    4.23   2.63
## 5  0.31 Good     J     SI2     63.3    58   335   4.34   4.35   2.75
```

Chart & Table 1: Relationship Between Depth, Price, And Cut

Chart 1.1: Relationship Between Depth And Cut

```
set.seed(99)
ggplot(diamonds %>% sample_n(20000), aes(depth, cut)) +
  geom_boxplot(fill="salmon") +
  theme_minimal() +
  labs(
    title = "Relationship Between Depth And Cut",
    x = "Depth",
    y = "Cut",
    caption = "Datasource: Diamonds from Base R"
  )
```

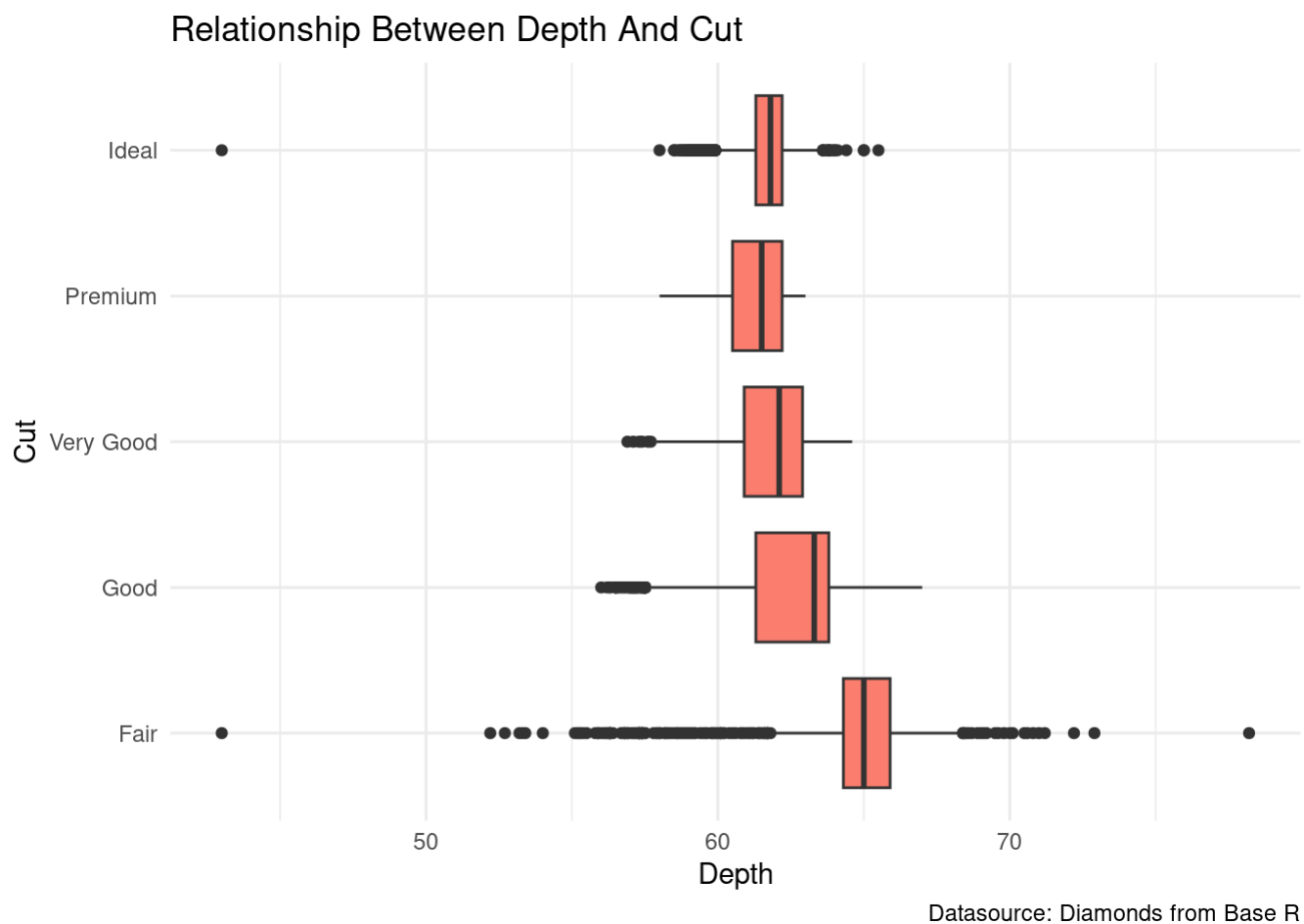
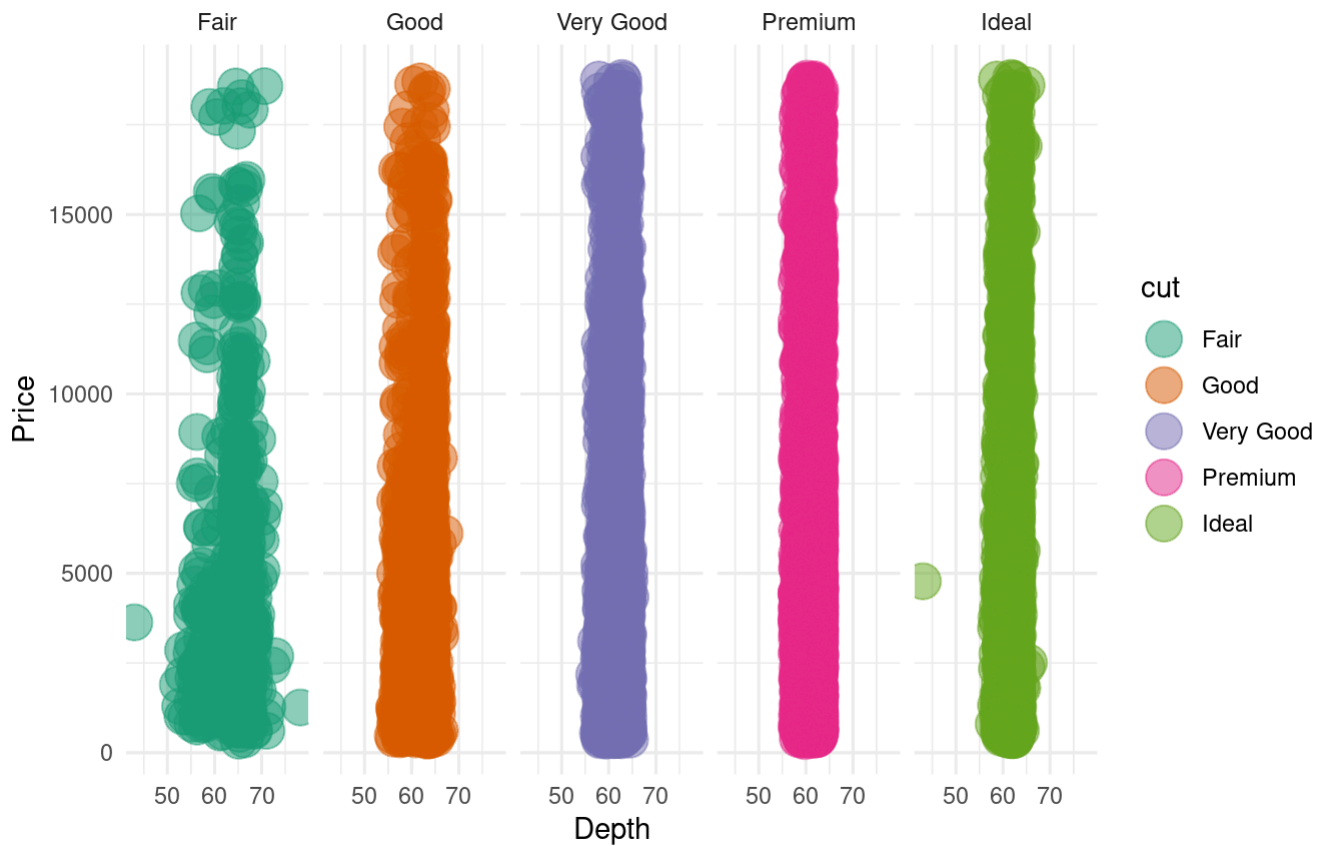


Chart 1.2: Relationship Between Depth And Price Separate By Cut

```
set.seed(99)
ggplot(diamonds %>% sample_n(20000), aes(depth,price,col=cut)) +
  geom_point(size=6,alpha=0.5) +
  theme_minimal() +
  facet_wrap(~cut,ncol=5) +
  scale_color_brewer(type="qual",palette=2) +
  labs(
    title = "Relationship Between Depth And Price Separate By Cut",
    x = "Depth",
    y = "Price",
    caption = "Datasource: Diamonds from Base R"
  )
```

Relationship Between Depth And Price Separate By Cut

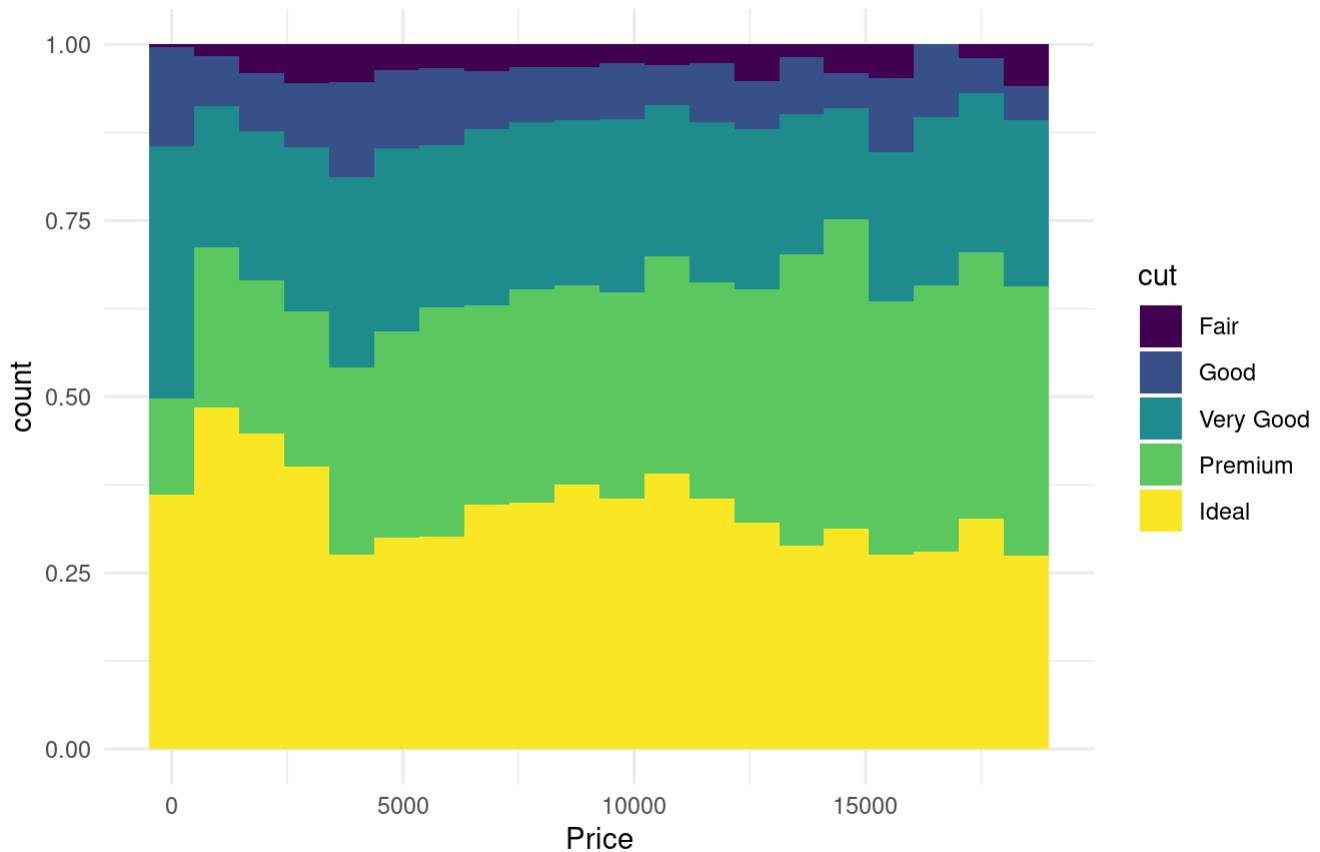


Datasource: Diamonds from Base R

Chart 1.3: Distribution Of Price Separate By Cut

```
set.seed(99)
ggplot(diamonds %>% sample_n(20000), aes(price,fill=cut)) +
  geom_histogram(bins=20,position="fill") +
  theme_minimal() +
  labs(
    title = "Distribution Of Price Separate By Cut",
    x = "Price",
    caption = "Datasource: Diamonds from Base R"
  )
```

Distribution Of Price Separate By Cut



Datasource: Diamonds from Base R

Table 1.1: Statics Information Of Price And Depth Group By Cut

```
diamonds %>%
  group_by(cut) %>%
  summarize(
    max_price = max(price),
    min_price = min(price),
    average_price = mean(price),
    quantity_price = n(),
    average_depth = mean(depth))
```

```
## # A tibble: 5 × 6
##   cut      max_price min_price average_price quantity_price average_depth
##   <ord>      <int>    <int>      <dbl>         <int>      <dbl>
## 1 Fair      18574      337      4359.          1610        64.0
## 2 Good      18788      327      3929.          4906        62.4
## 3 Very Good  18818      336      3982.         12082        61.8
## 4 Premium   18823      326      4584.         13791        61.3
## 5 Ideal     18806      326      3458.         21551        61.7
```

Result:

Regarding to charts no. 1.1 and 1.2, It shows that price will highest when depth is from 58 and near 65 and same as cut that it has the high level when it is between 61 - 63.

Moreover from chart no. 1.3 and table no. 1.1, It shows max of price that it seems to higher when its level high even if mean of Ideal level seems to lowest but I think because price of Ideal level has the high distribution since min price until max price and also has the most quantity as chart no 1.3.

In addition, mean of depth is also shows the results which support the chart no. 1.1 and 1.2 that it will have the high cut when it is between 61 - 63 as its mean shows in the table and refer to *datasource from petchcharat website about how to clarify depth of diamond in Idel cut, it indicated that depth in ideal cut will be between 58 - 62.3%.

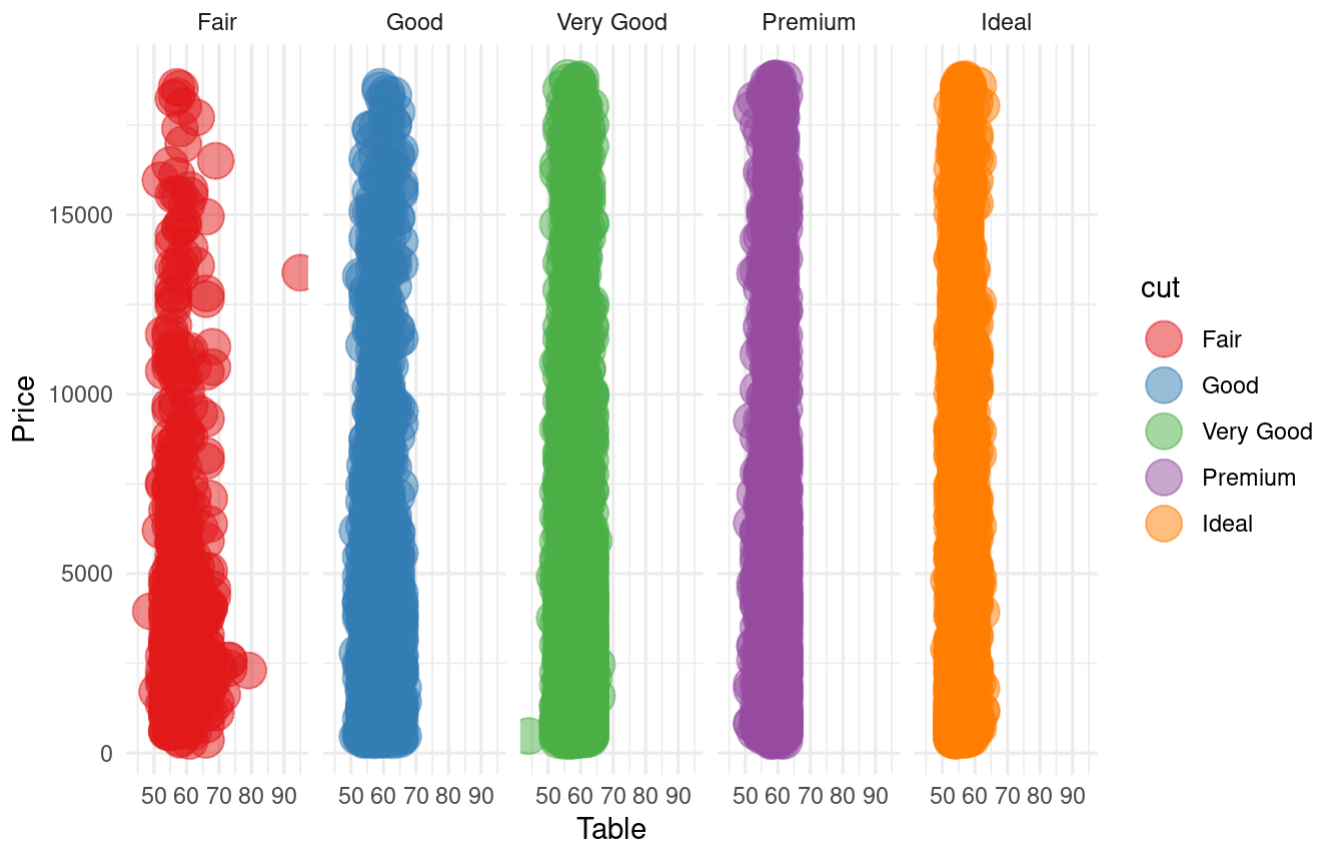
So the chart as above can explain that depth between 58 - 62.3% will have the highest price and will be indicated the high level of cut as well.

*Datasource: 1)Diamonds Base R 2)[petchcharat] (<http://www.petchcharat.com/subdiamond.php?pid=13>
(<http://www.petchcharat.com/subdiamond.php?pid=13>))

Chart 2: Relationship Between Table And Price Separate By Cut

```
set.seed(44)
ggplot(diamonds %>% sample_n(20000), aes(table,price,col=cut)) +
  geom_point(size=6,alpha=0.5) +
  theme_minimal() +
  facet_wrap(~cut,ncol=5) +
  scale_color_brewer(type="qual",palette=6) +
  labs(
    title = "Relationship Between Table And Price Separate By Cut",
    x = "Table",
    y = "Price",
    caption = "Datasource: Diamonds from Base R"
  )
```

Relationship Between Table And Price Separate By Cut



Datasource: Diamonds from Base R

Correlation Between Table And Price

```
cor(diamonds$table,diamonds$price)
```

```
## [1] 0.1271339
```

Result:

Regarding to charts no. 2, It shows that price will highest when table is around 55 - 60 in almost of cut and also have to relation in the positive trend (value=0.1271339).

Moreover refer to *datasource from petcharat website about how to clarify table of diamond in Ideal cut, it indicated that table in ideal cut will be between 52 - 60%.

So the chart as above can explain that if table have the higher price and level of cut will higher also, And table and price is also related together with positive value = 0.1271339.

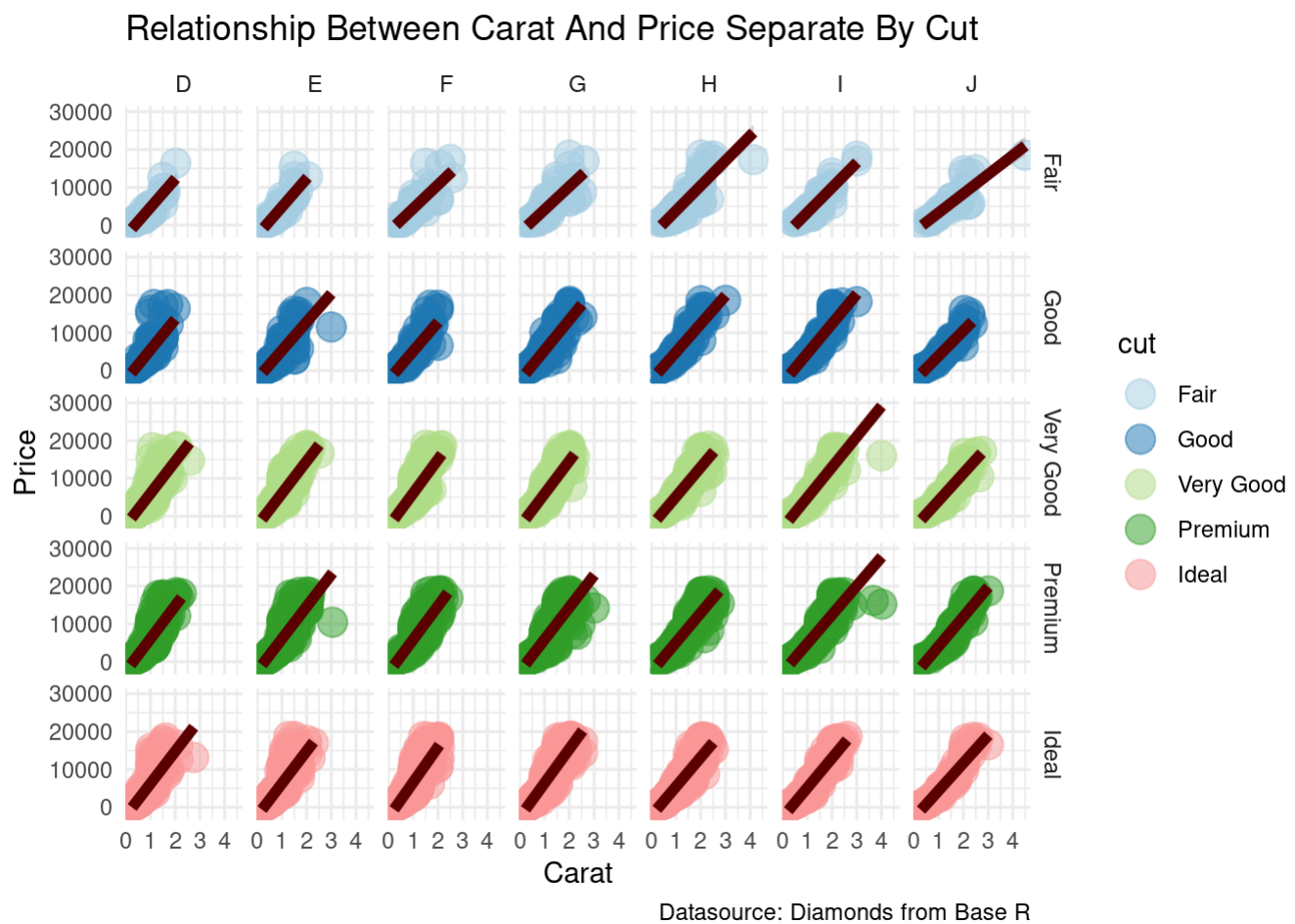
*Datasource: 1)Diamonds Base R 2)[petcharat] (<http://www.petcharat.com/subdiamond.php?pid=13>) (<http://www.petcharat.com/subdiamond.php?pid=13>)

Chart 3: Relationship Between Carat And Price Separate By Cut

```

set.seed(88)
ggplot(diamonds %>% sample_n(20000), aes(carat,price,col=cut)) +
  geom_point(size = 5 , alpha = 0.5) +
  geom_smooth(size = 2,method="lm",col="#5C0000") +
  theme_minimal() +
  facet_grid(cut~color) +
  scale_color_brewer(type="qual",palette=3) +
  labs(
    title = "Relationship Between Carat And Price Separate By Cut",
    x = "Carat",
    y = "Price",
    caption = "Datasource: Diamonds from Base R"
  )

```



Correlation Of Carat And Price

```
cor(diamonds$carat,diamonds$price)
```

```
## [1] 0.9215913
```

Result:

Regarding to charts no. 3, It shows that when quantity of carat increases, price will increase as well. Although diamond is different cut but price still increases as chart shows.

Refer to correlation of carat and price which is positive with value = 0.9215913, This value means that when carat increases 1 unit, it will make price increases 0.9215913 unit.

So the chart as above can explain that carat is related to price by positive value = 0.9215913 and cut is not effect to this relation.

*Datasource: 1)Diamonds Base R