Homework Data Visualization

Kittipong Hor

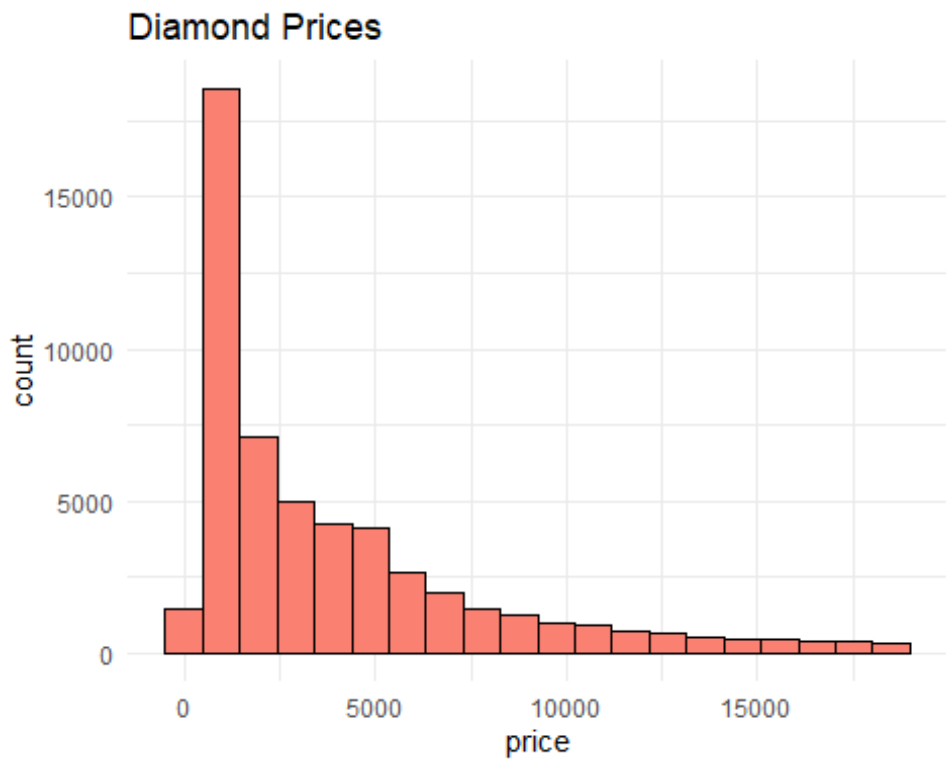
2023-03-31

Hello World

```
library(tidyverse)
library(dplyr)
library(ggplot2)
data("diamonds")
```

Chart 1 - A Histogram of the price of all the diamonds in the diamond data set

```
ggplot(diamonds, aes(price)) +
  geom_histogram(bins=20,
                 col="black",
                 fill = "salmon") +
  theme_minimal() +
  labs(title = "Diamond Prices")
```



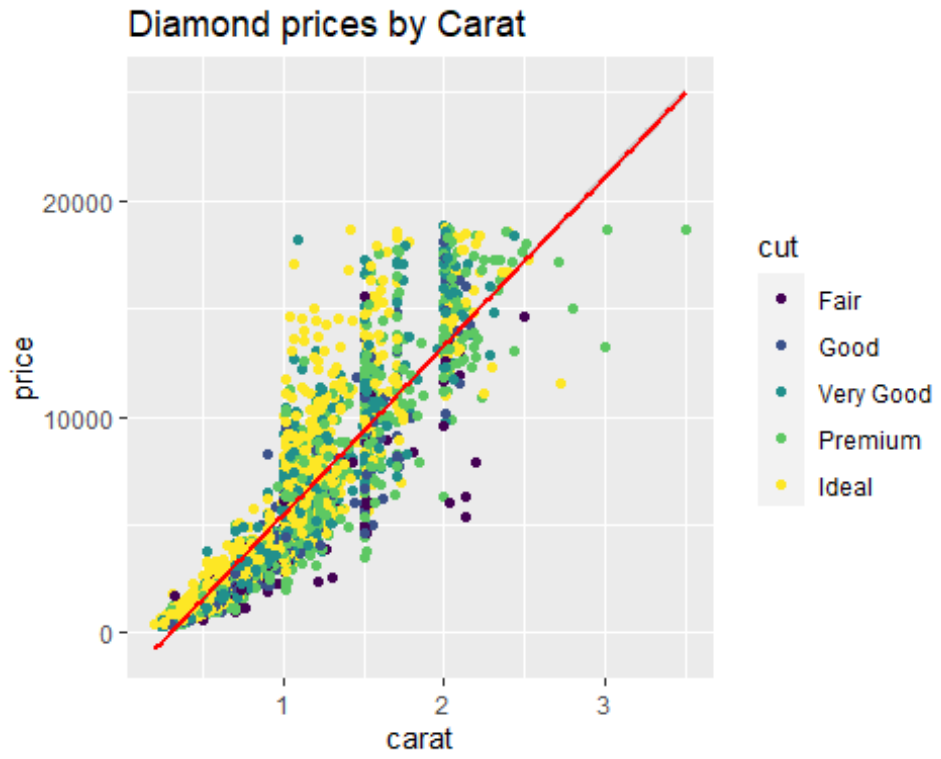

```
summary(diamonds$price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     326     950    2401    3933    5324   18823
```

We found a positive skewness in this histogram chart

Chart 2 - A Scatterplot of diamond price vs. carat

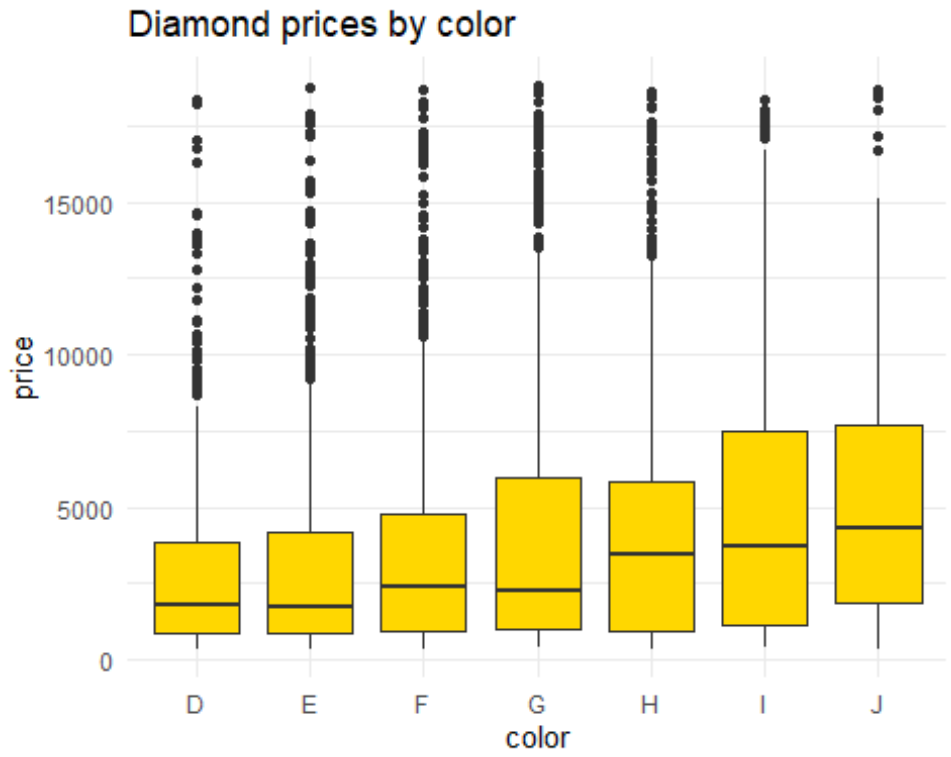
```
set.seed(18)
ggplot(diamonds %>% sample_n(5000), aes(carat,price, col=cut))+
  geom_point() +
  geom_smooth(method = "lm",
              col = "red") +
  labs(title = "Diamond prices by Carat")
```



We found a positive correlation between two variables.

Chart 3 - A Boxplot of sample diamond price vs. color

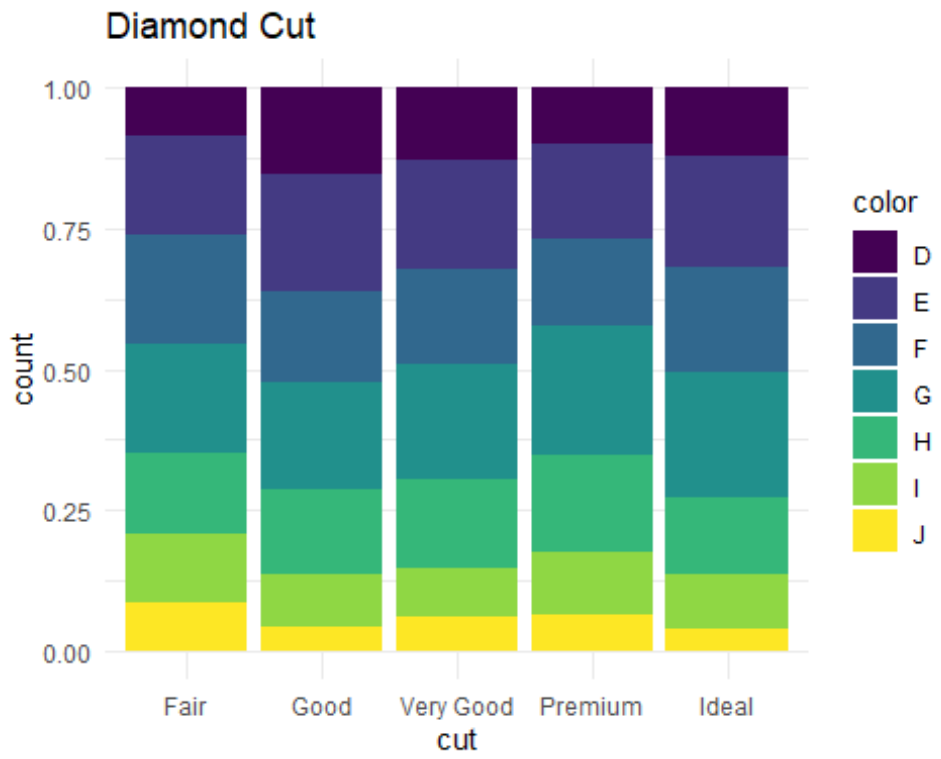
```
set.seed(18)
ggplot(diamonds %>% sample_n(5000), aes(color,price)) +
  geom_boxplot(fill= "gold") +
  theme_minimal() +
  labs(title = "Diamond prices by color")
```



We found a result summaries of diamond colors.

Chart 4 A Barplot of diamond cut by color.

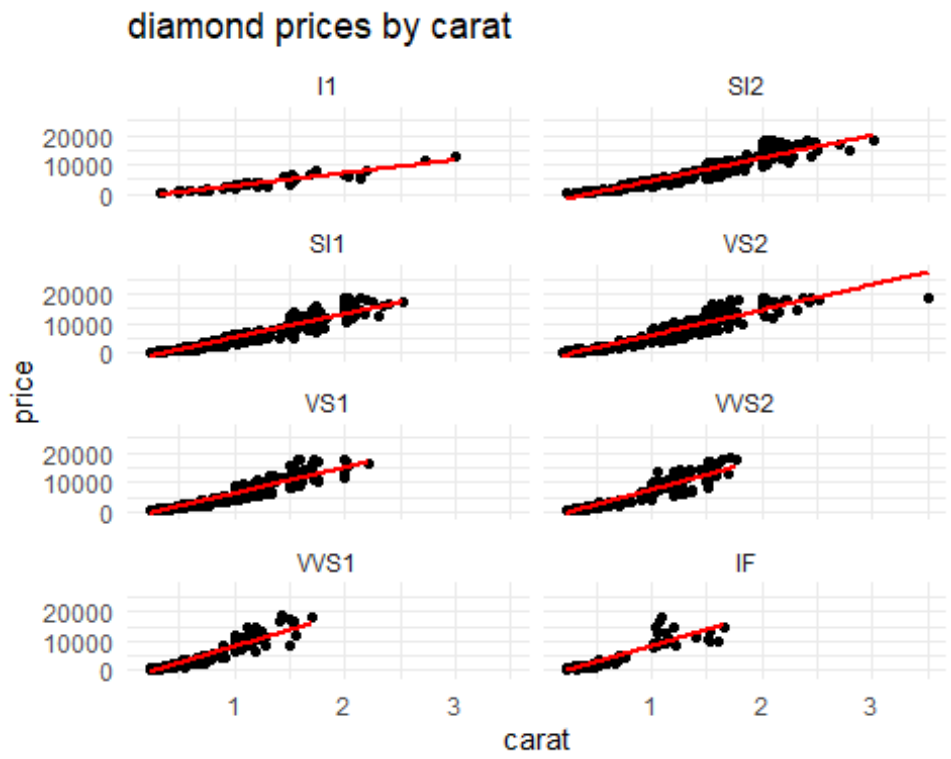
```
set.seed(18)
ggplot(diamonds %>% sample_n(5000), aes(cut, fill=color)) +
  geom_bar(position = "fill") +
  theme_minimal() +
  labs(title = "Diamond Cut")
```



We found a proportion of color diamonds for each value of cut

Chart 5 A Scatterplot of diamond price vs. carat

```
set.seed(18)
ggplot(diamonds %>% sample_n(5000), aes(carat,price)) +
         geom_point() +
         geom_smooth(method ="lm",col="red") +
         facet_wrap(~clarity, ncol = 2) +
         theme_minimal() +
         labs(title = "diamond prices by carat")
```



We found a positive correlation between two variables by clarity.