Data Visualization Homework

Miss Natthamon Ratthanasurakarn

Diamonds Data Visualization

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggthemes)
```

Preview 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.~
```

```
mean(complete.cases(diamonds))
```

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

Data Description

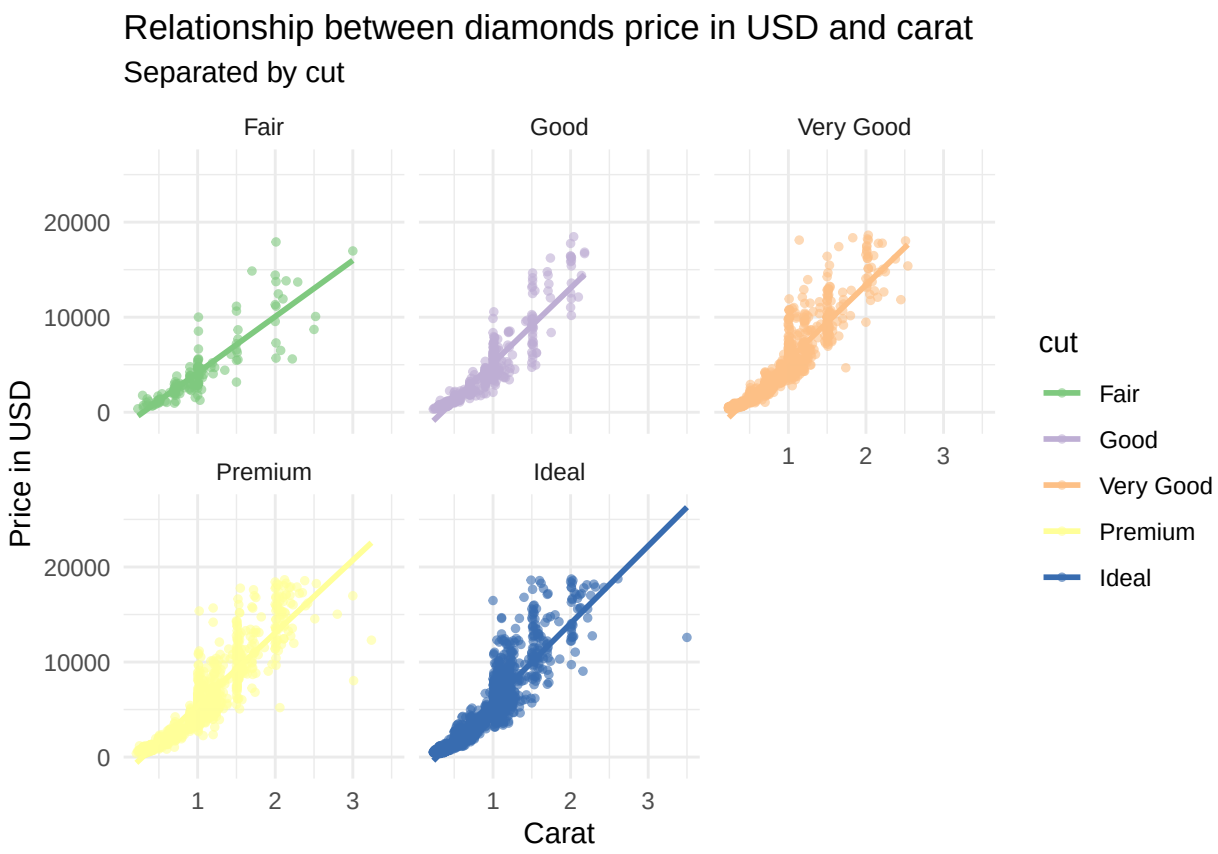
Variable	Description	Values
price	price in US dollars	$326-$18,823
carat	weight of the diamond	0.2-5.01
cut	quality of the cut	Fair, Good, Very Good, Premium, Ideal

Variable	Description	Values
color	diamond color	J (worst) to D (best)
clarity	measurement of how clear the diamond is	I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best)
x	length in mm	0-10.74
y	width in mm	0-58.9
z	depth in mm	0-31.8
depth	total depth percentage	43-79
table	width of top of diamond relative to widest point	43-95

Plot 1 : Scatter plot of Relationship between price(USD) and carat

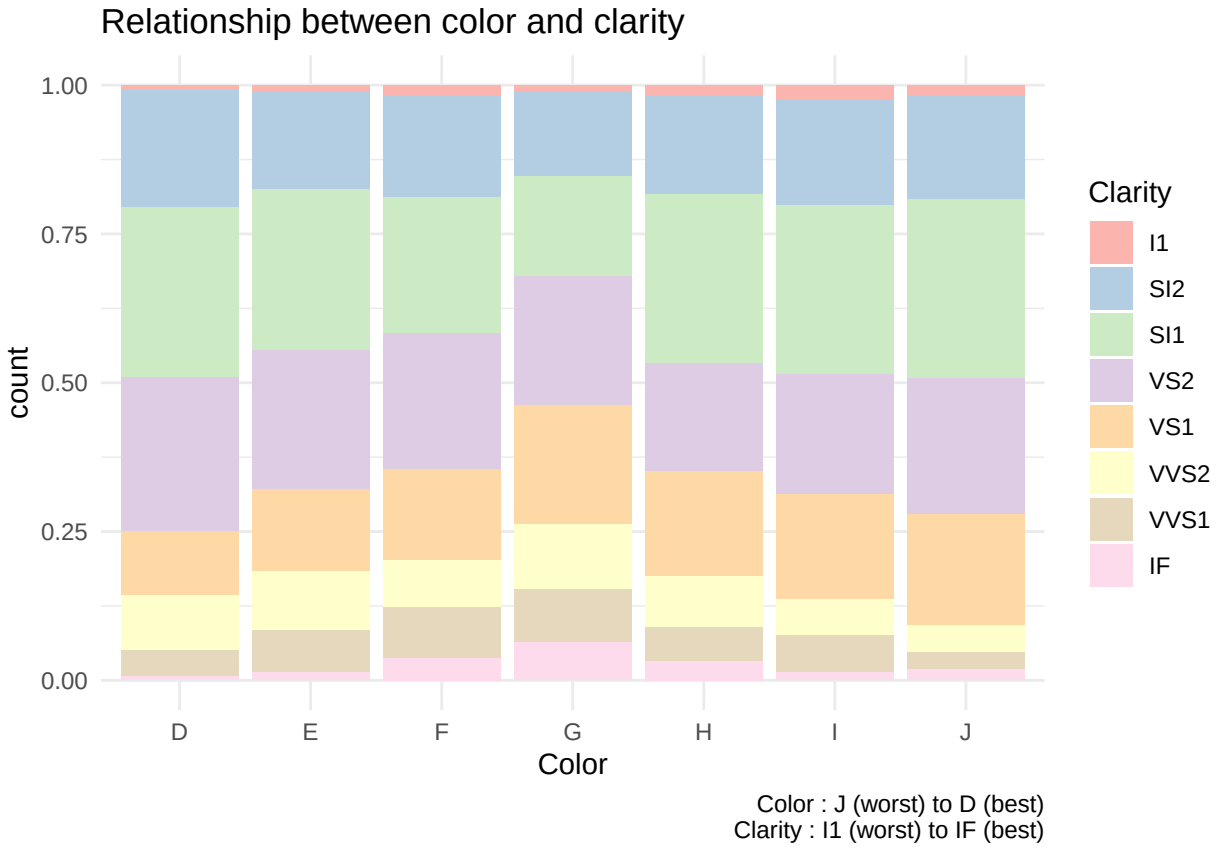
```
set.seed(33)
diamonds %>%
  sample_n(5000) %>%
  ggplot(aes(carat,price,color = cut)) +
    geom_point(alpha = 0.6, size = 1) +
    geom_smooth(method = "lm",se = F) +
    facet_wrap(~cut, ncol = 3) +
    theme_minimal() +
    scale_color_brewer(palette = "Accent") +
    labs(title = "Relationship between diamonds price in USD and carat",
         x = "Carat",
         y = "Price in USD",
         subtitle = "Separated by cut",
         fill = "Cut")
```

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



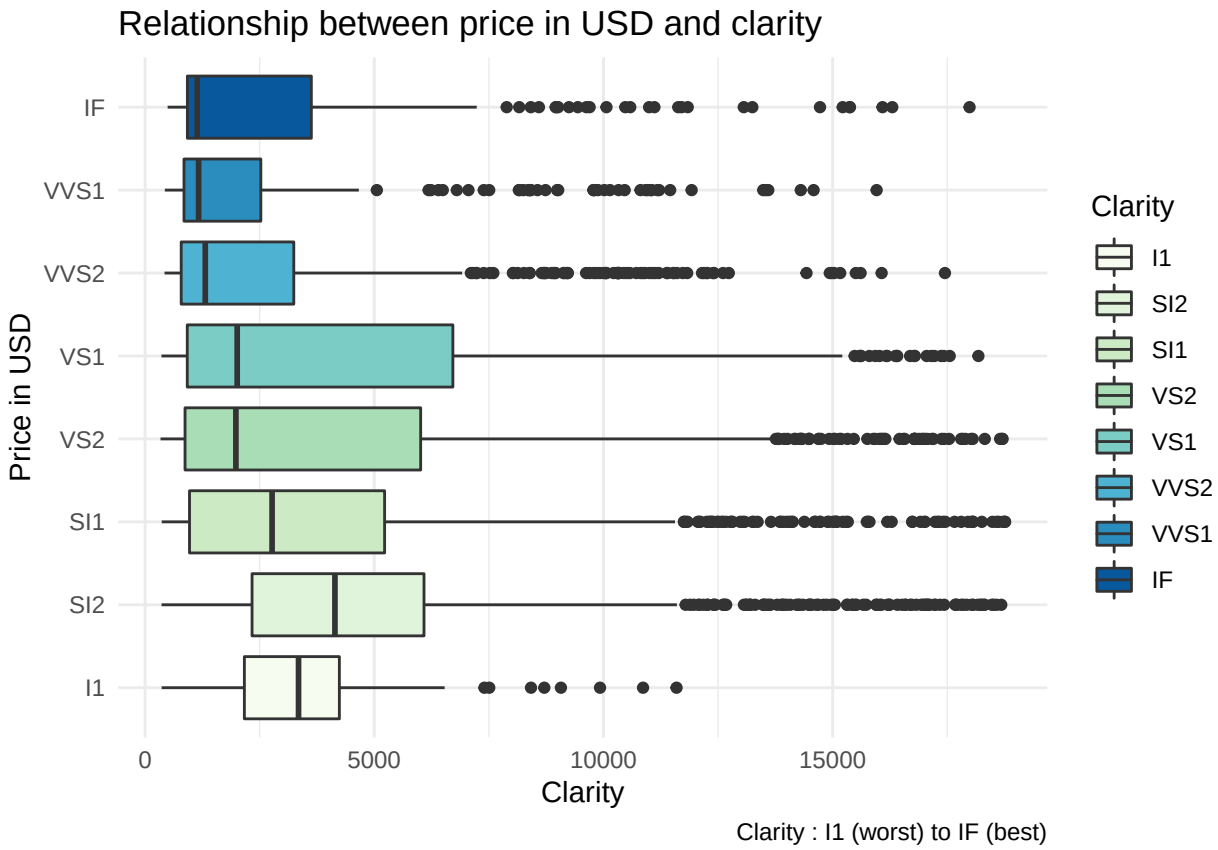
Plot 2 : Bar plot of Relationship between diamonds color and clarity

```
set.seed(99)
diamonds %>%
  sample_n(5000) %>%
  ggplot(aes(color,fill = clarity)) +
    geom_bar(position = "fill") +
    scale_fill_brewer(palette = "Pastel1") +
    theme_minimal() +
    labs(title = "Relationship between color and clarity",
         x = "Color",
         fill = "Clarity",
         caption = "Color : J (worst) to D (best)\n Clarity : I1 (worst) to IF (best)")
```



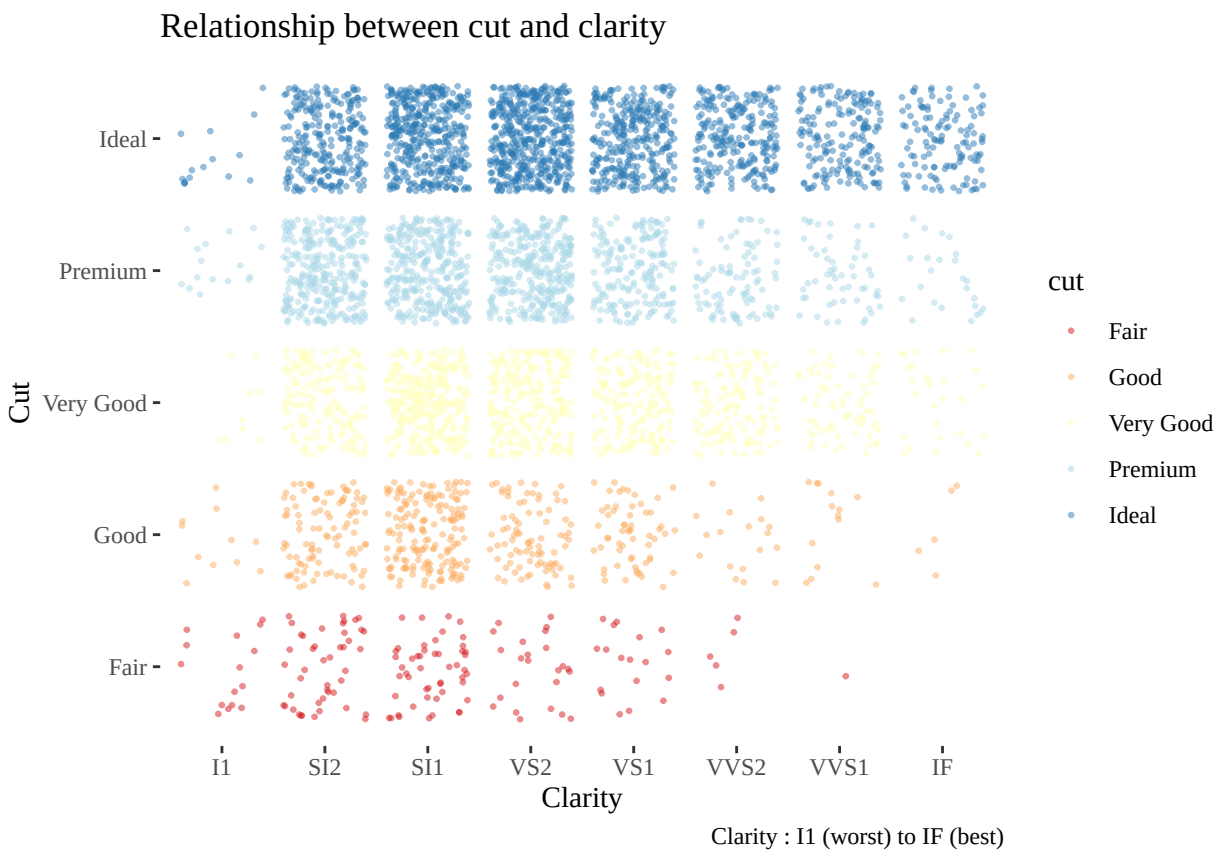
Plot 3 : Box plot of Relationship between price(USD) and clarity

```
set.seed(99)
diamonds %>%
  sample_n(5000) %>%
  ggplot(aes(clarity,price,fill = clarity)) +
    geom_boxplot() +
    coord_flip() +
    scale_fill_brewer(palette = "GnBu") +
    theme_minimal() +
    labs(title = "Relationship between price in USD and clarity",
         y = "Clarity",
         x = "Price in USD",
         caption = "Clarity : I1 (worst) to IF (best)",
         fill = "Clarity")
```



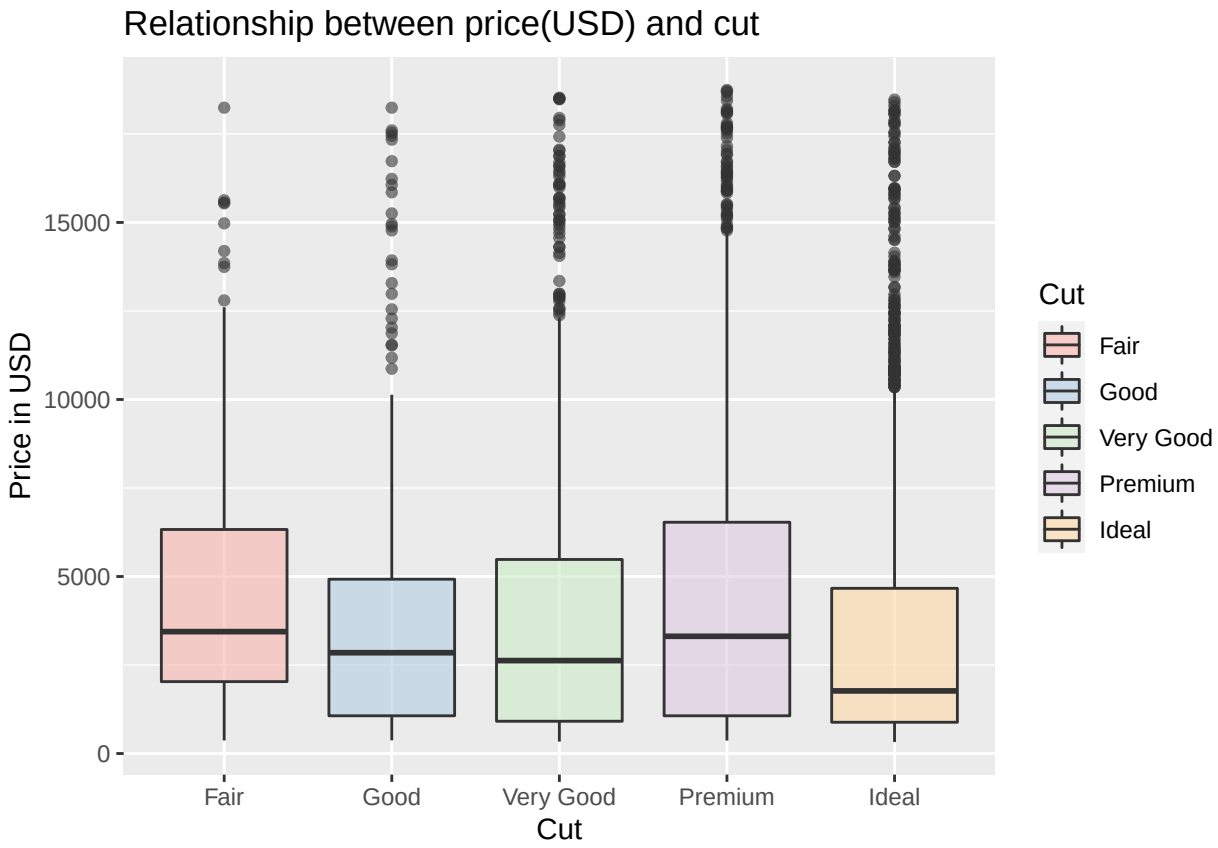
Plot 4 : Jitter plot of Relationship between clarity and cut

```
set.seed(88)
diamonds %>%
  sample_n(5000) %>%
  ggplot(aes(clarity,cut,color = cut)) +
    geom_jitter(alpha = 0.5,size = 0.5) +
    scale_color_brewer(palette = "RdYlBu") +
    theme_tufte() +
    labs(x = "Clarity",
         y = "Cut",
         fill = "Cut",
         title = "Relationship between cut and clarity",
         caption = "Clarity : I1 (worst) to IF (best)")
```



Plot 5 : Box plot of Relationship between price(USD) and cut

```
set.seed(66)
diamonds %>%
  sample_n(5000) %>%
  ggplot(aes(cut,price,fill = cut)) +
    geom_boxplot(alpha = 0.6) +
    scale_fill_brewer(palette = "Pastel1") +
    labs(x = "Cut",
         y = "Price in USD",
         title = "Relationship between price(USD) and cut",
         fill = "Cut")
```


Relationship between price(USD) and cut
Price in USD
15000
10000
5000
0
Fair
Good
Very Good
Premium
Ideal
Cut
Cut
Fair
Good
Very Good
Premium
Ideal