

Data Visualization(Sprint5-HW)

Note B

2023-02-25

Sprint5 Data Visualization Homework

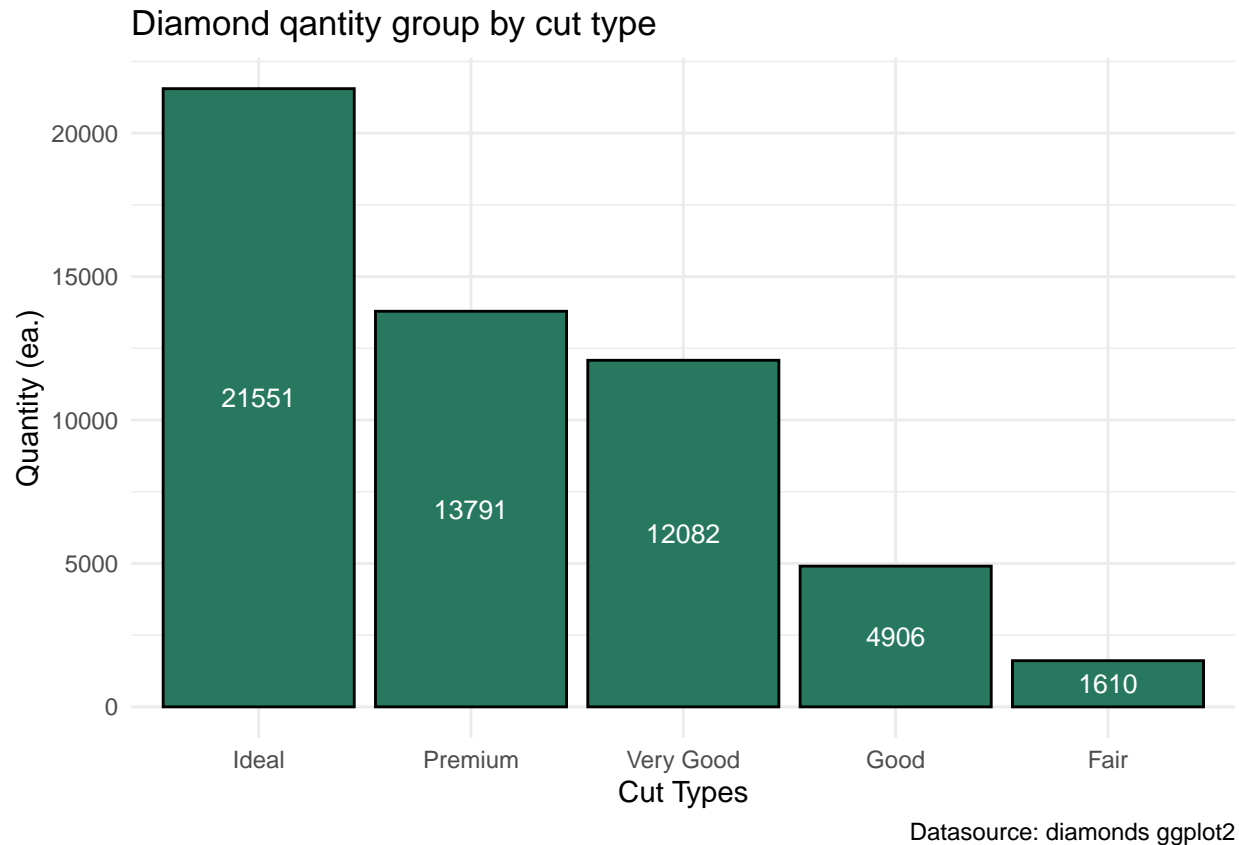
Question Explore dataframe '*diamonds*' and create 5 visualizations

R library loading

```
library (tidyverse)
library (ggplot2)
library (patchwork)
library (dplyr)
library (scales)
```

Hw1# Summary count by cut types with ascending order

```
ggplot(diamonds, aes(x = fct_infreq(cut))) +
  geom_bar(stat = "count", fill = "#287860", col = "black") +
  stat_count(geom = "text", color = "white", size=3.5,
             aes(label = after_stat(count)), position=position_stack(vjust=0.5)) +
  theme_minimal() +
  labs (
    title = "Diamond qantity group by cut type",
    x = "Cut Types",
    y = "Quantity (ea.)",
    caption = "Datasource: diamonds ggplot2"
  )
```



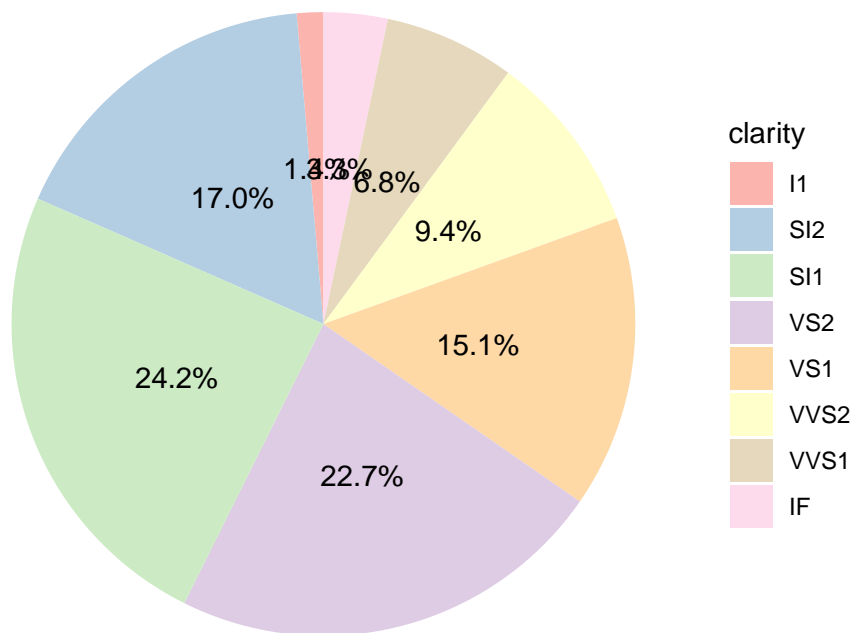
HW#1 Conclusion: The most quantity of diamond cut types is 'Ideal' with 21,551 ea and the least is 'Fair' with 1,610 ea.

Hw2# Pie chart by clarity

```
t1 <- diamonds %>%
  group_by(clarity) %>%
  count() %>%
  ungroup() %>%
  mutate(perc = n/sum(n)) %>%
  arrange(perc) %>%
  mutate(labels = scales::percent(perc))

ggplot(t1, aes(x="", y=perc, fill=clarity)) +
  geom_col()+
  geom_text (aes (label=labels),
             position = position_stack(vjust=0.5))+
  coord_polar(theta = "y")+
  theme_void()+
  scale_fill_brewer(type="qual", palette = 4) +
  labs (
    title = "% of diamonds by clarity",
    caption = "Datasource: diamonds ggplot2"
  )
```

% of diamonds by clarity

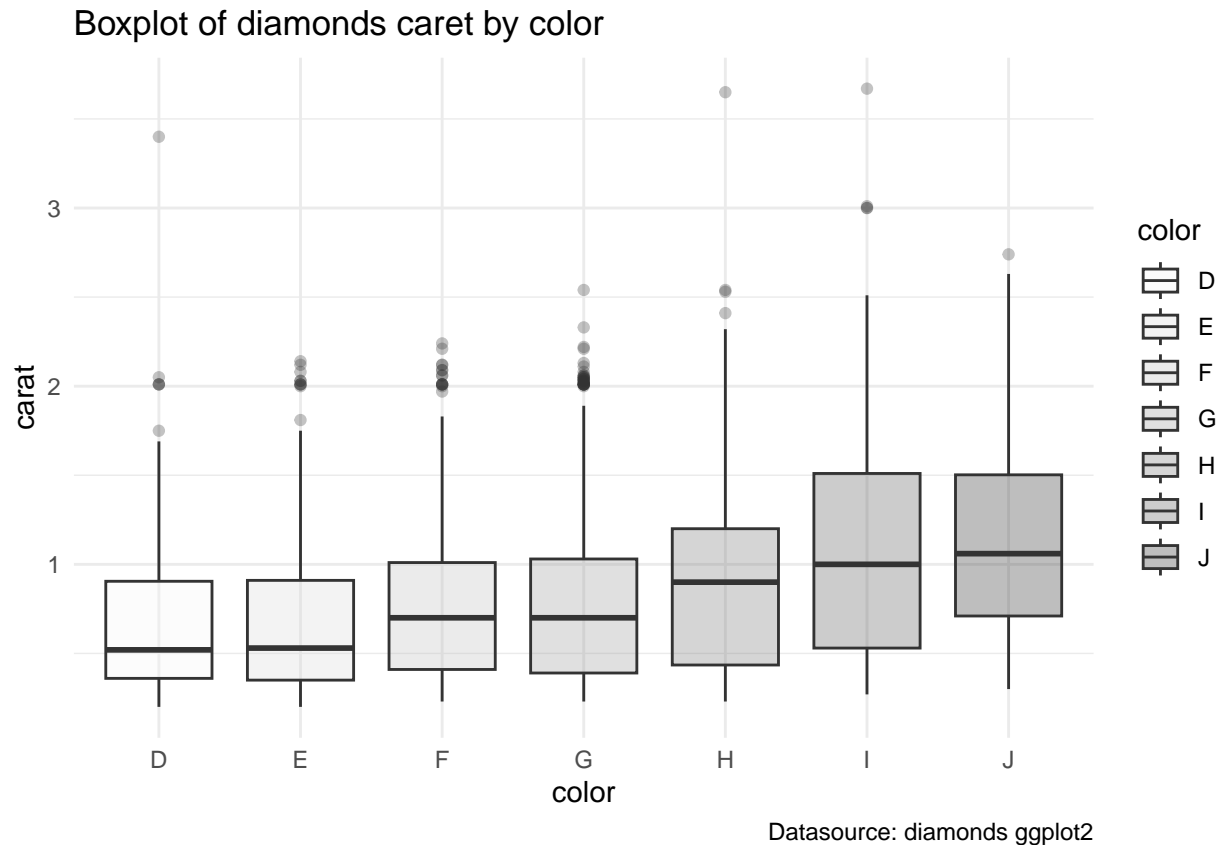


Datasource: diamonds ggplot2

HW#2 Conclusion: Total of eight(8) diamond clarity, top 3 of clarity are 1) 24.2% SI1, 22.7% VS2, and 17% SI2.

Hw3# Boxplot of caret by color

```
set.seed(07)
ggplot(diamonds %>% sample_n(5000), aes (x=color, y=carat, fill=color)) +
  geom_boxplot(alpha=0.3) +
  theme(legend.position="none") +
  scale_fill_brewer(type="seq", palette = 6) +
  theme_minimal() +
  labs (
    title = "Boxplot of diamonds caret by color",
    caption = "Datasource: diamonds ggplot2"
  )
```



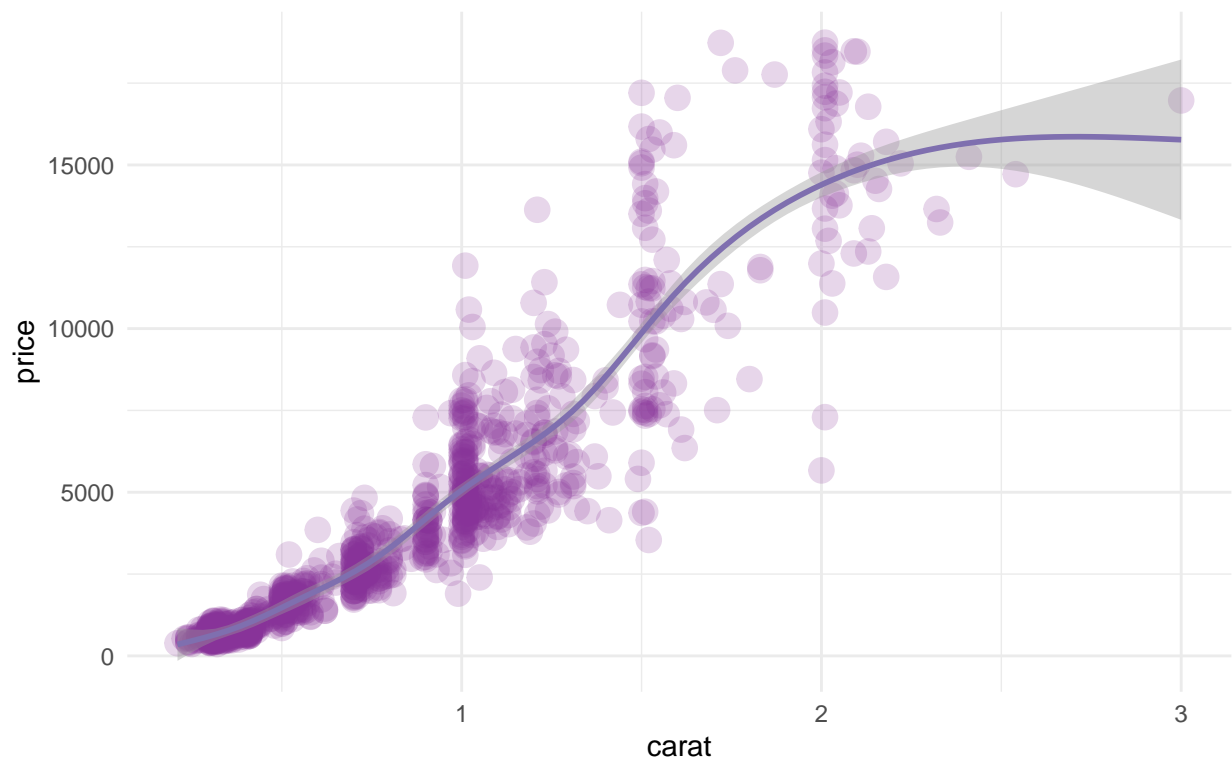
HW#3 Conclusion: Increasing trend of carat following to color ($D > E > F > G > H > I > J$)

Hw4# Scatter plot between price and caret

```
set.seed(07)
ggplot(diamonds %>% sample_n(1000), aes (x=carat, y=price)) +
  geom_point(size=4, alpha=0.2, col="#883399") +
  geom_smooth(method="auto", col="#7F6FAF") +
  theme_minimal() +
  labs (
    title = "Relationship between price and caret of diamonds",
    caption = "Datasource: diamonds ggplot2"
  )
```

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

Relationship between price and caret of diamonds



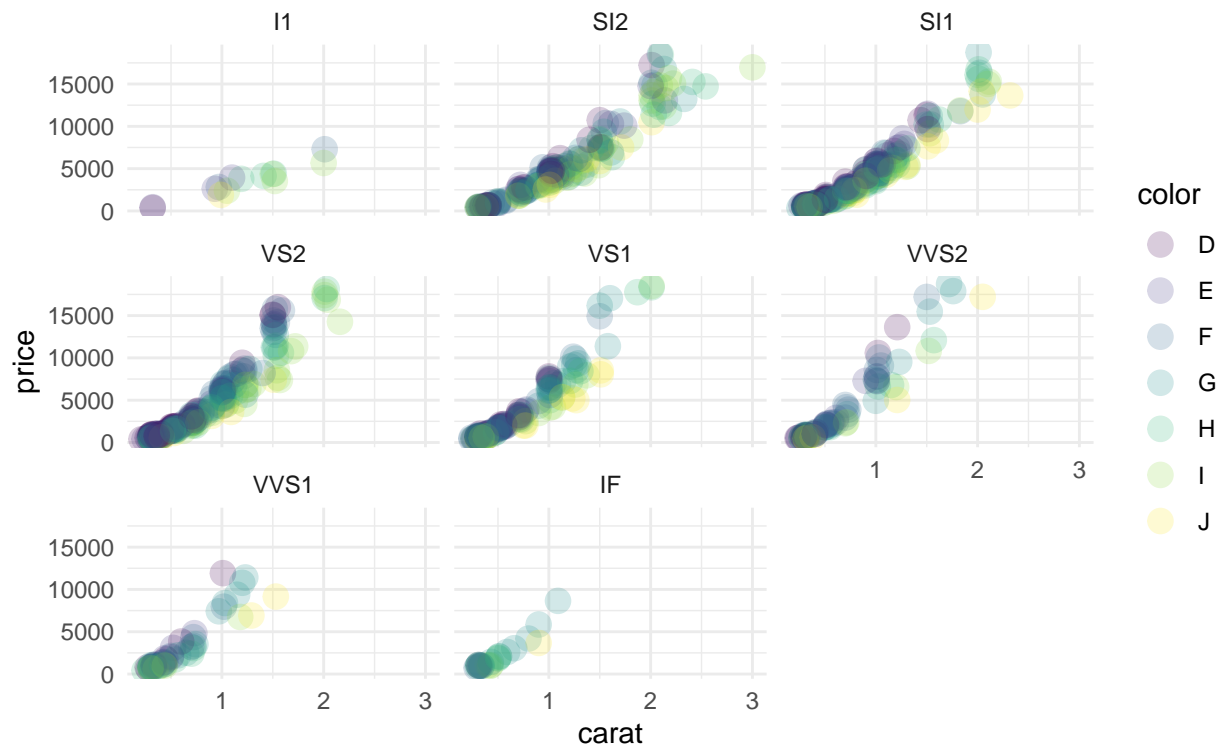
Datasource: diamonds ggplot2

HW#4 Conclusion: There is relationship between price and carat. Price is higher follow to carat value

Hw5# using facet

```
set.seed(07)
ggplot(diamonds %>% sample_n(1000), mapping =aes(x=carat, y=price, col=color)) +
  geom_point(size=4, alpha=0.2) +
  scale_fill_brewer(type="qual", palette = 6) +
  facet_wrap(~clarity) +
  theme_minimal() +
  labs (
    title = "Relationship between price and caret of diamonds, separate by clarity",
    caption = "Datasource: diamonds ggplot2"
  )
```

Relationship between price and caret of diamonds, separate by clarity



Datasource: diamonds ggplot2

HW#5 Conclusion: Price is depended on carat value for all colors
