Data Visualization Bootcamp Homework

Chanyaphas Nina

2023-09-15

Intruction

Use diamonds dataset to create 5 charts. knit pdf and submit in discord.

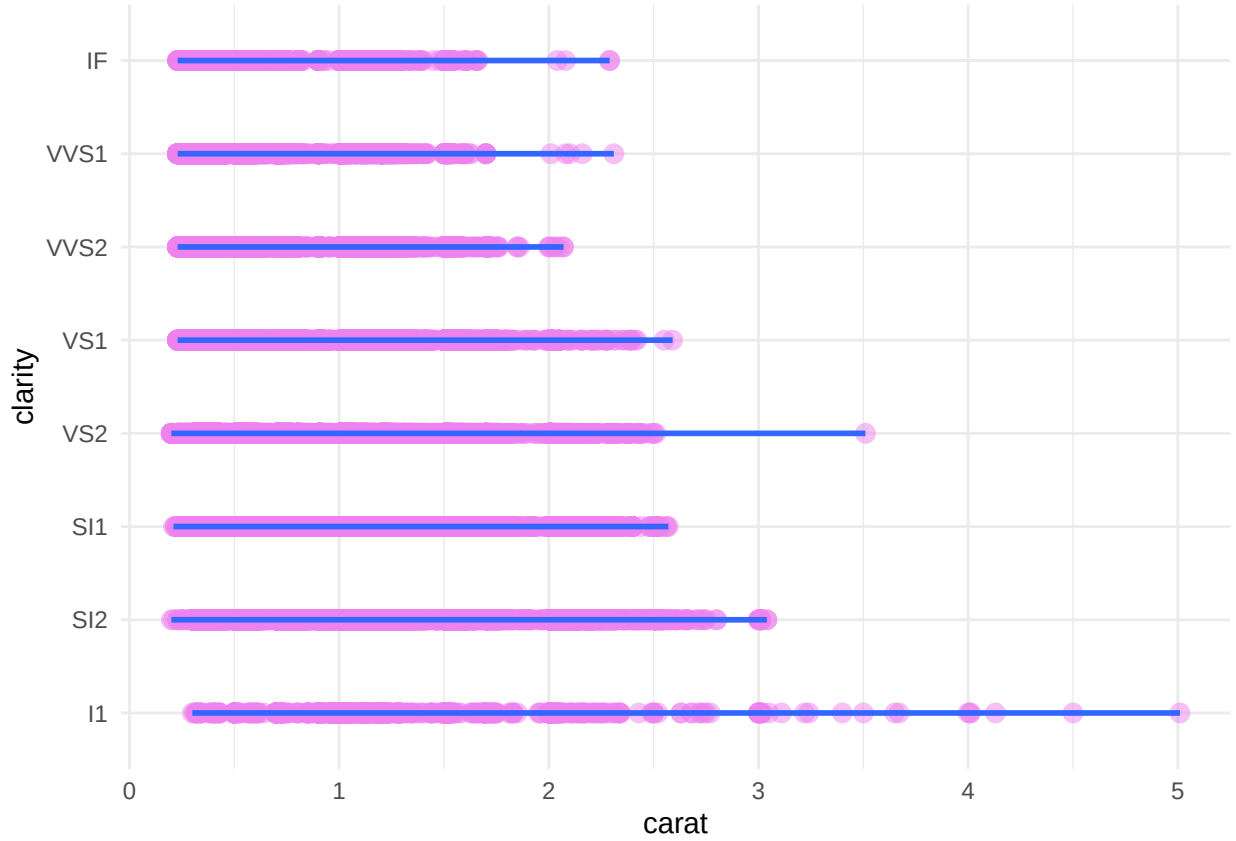
Import Library

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

1. Count the number of diamonds by clarity compared to the number of diamonds by carat

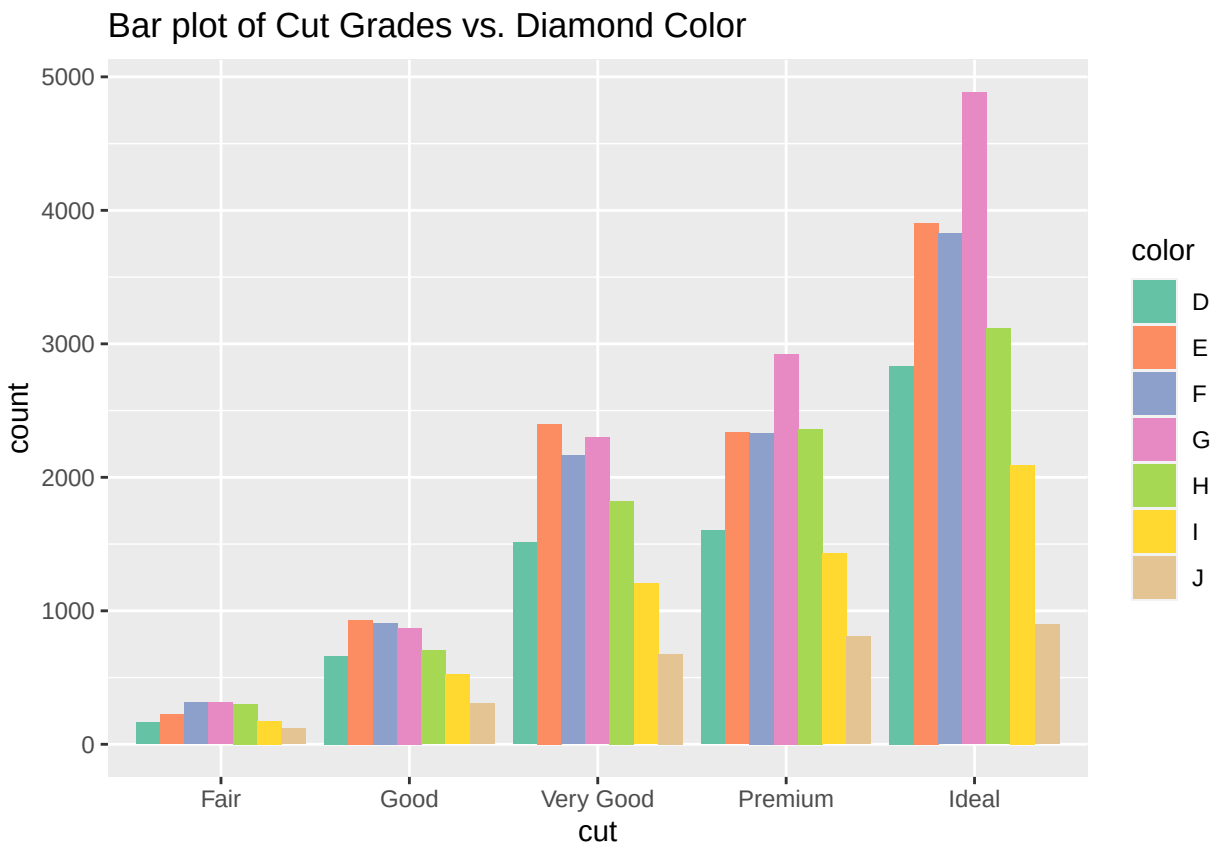
```
ggplot(diamonds, aes(carat, clarity)) +
  geom_point(size = 3, alpha = 0.5, col = "violet") +
  geom_smooth(method = 'lm', formula = 'y ~ x') +
  theme_minimal()
```



Most clarity with a carat weight ranging from 0.3 to 2.5 but Diamonds with a clarity level of 'I1' have a carat weight distribution ranging from 0.4 to 5.

2. Separate the diamond cut grade by diamond color.

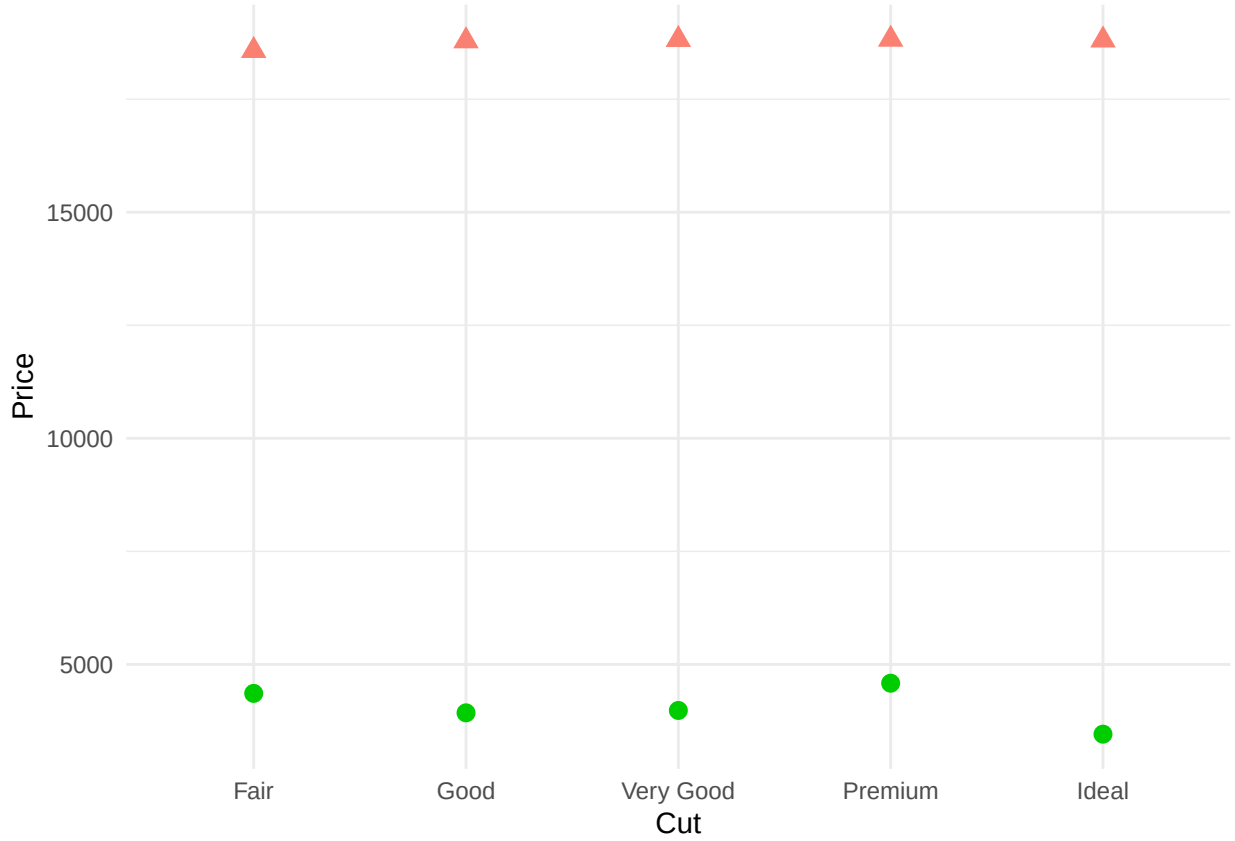
```
ggplot(diamonds, mapping = aes(cut, fill=color)) +
  geom_bar(position = "dodge") +
  labs(title = "Bar plot of Cut Grades vs. Diamond Color") +
  scale_fill_manual(values = c("#66c2a5", "#fc8d62", "#8da0cb",
                               "#e78ac3", "#a6d854", "#ffd92f", "#e5c494"))
```



In the diamond cuts, 'ideal' and 'premium' grades have the most 'G' colored diamonds, while in every diamond cut grade, 'J' and 'I' colors are the least common, respectively, according to the graph.

3. The highest and Average prices of diamonds for each cut grade.

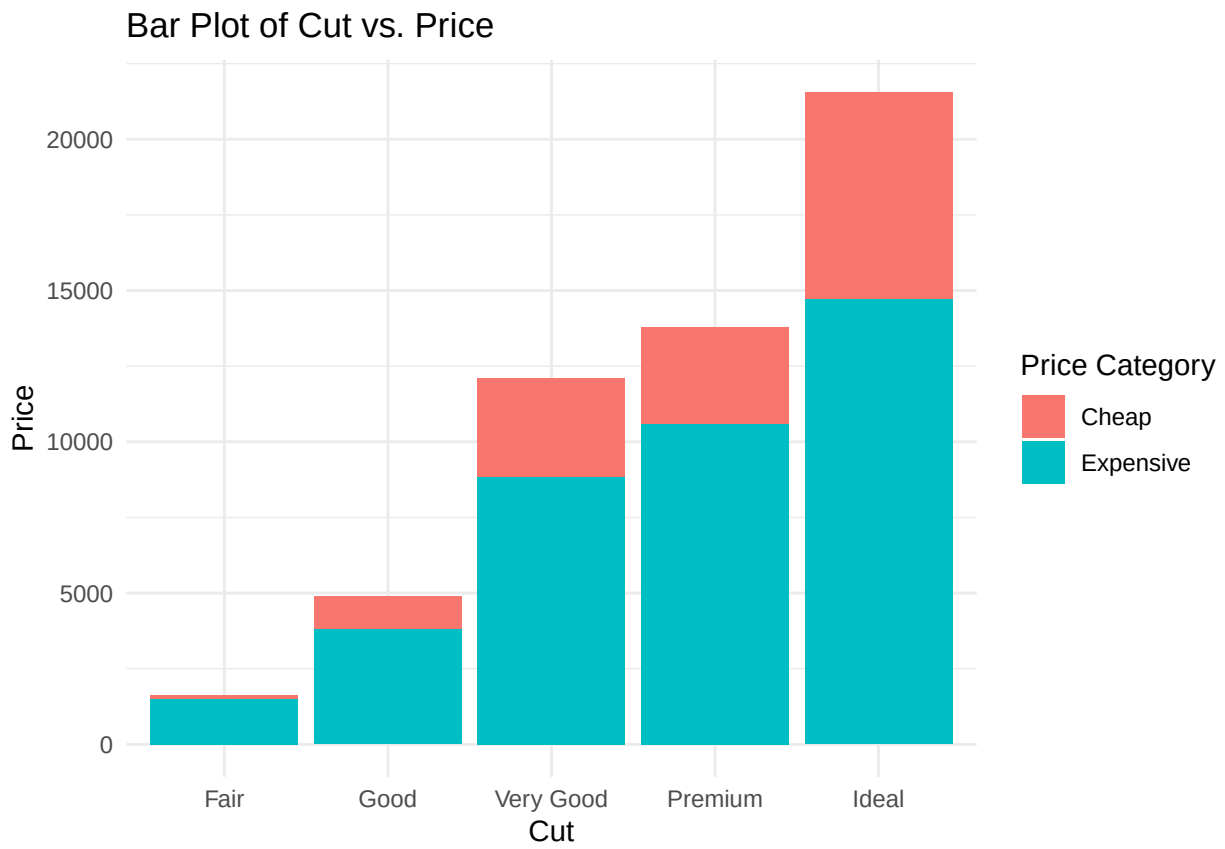
```
diamonds %>%
  group_by(cut) %>%
  summarise(meanprice = mean(price), maxprice = max(price)) %>%

  ggplot(mapping = aes(cut)) +
  geom_point(aes(y = maxprice), col = "salmon", size = 3, shape = 17) +
  geom_point(aes(y = meanprice), col = "green3", size = 3, shape = 16) +
  labs(x = "Cut", y = "Price") +
  theme_minimal()
```



This graph shows the highest and average prices of diamonds for each cut diamond, The pink point shows the highest price. The green point shows the average price.

4. The price of each diamond cut, divided into expensive and cheap categories

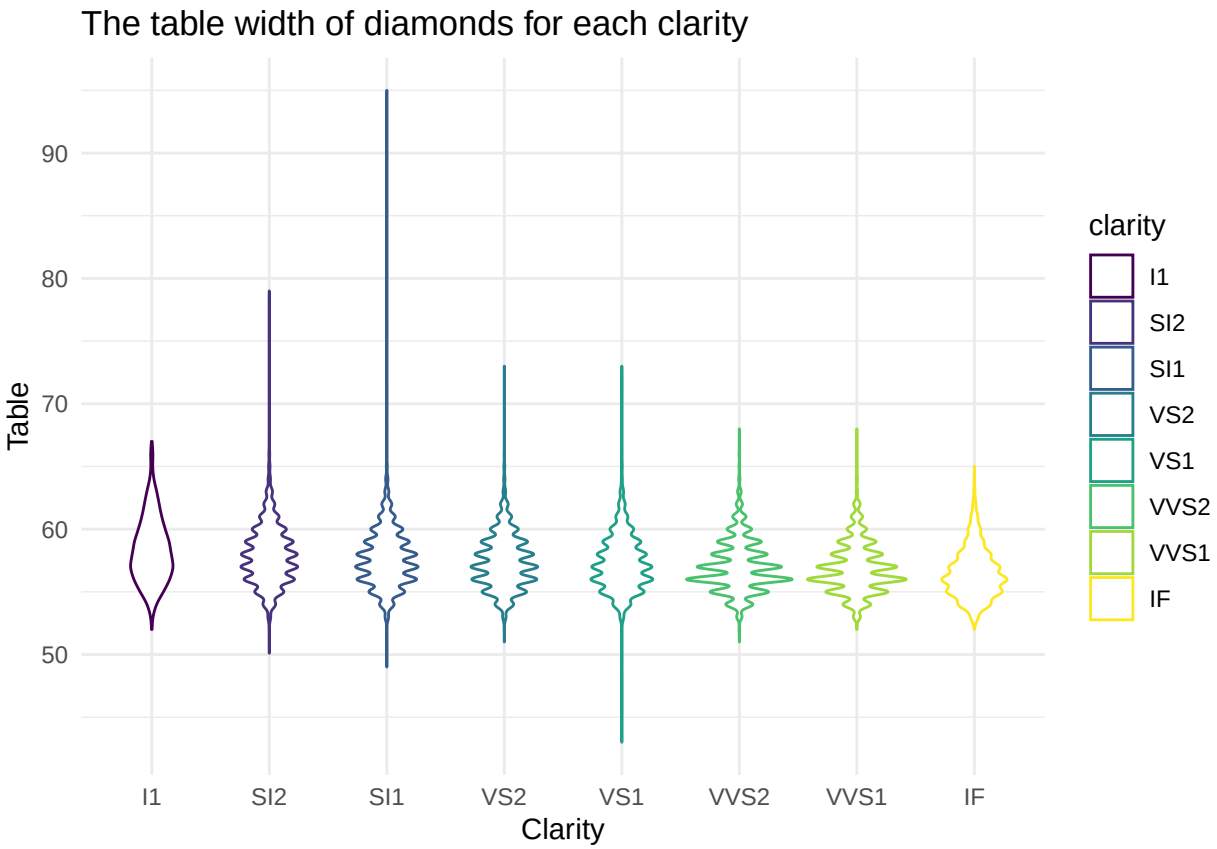
```
ggplot(diamonds %>%
         mutate(price = ifelse(price <= 1000, "Cheap","Expensive")),
       aes(x = cut, fill=price)) +
  geom_bar() +
  labs(title = "Bar Plot of Cut vs. Price",
       x = "Cut",
       y = "Price",
       fill = "Price Category") +
  theme_minimal()
```



The price of each diamond cut, divided into expensive and cheap categories. Based on the majority of the graphs, each cut is typically expensive.

5. The table width of diamonds for each clarity

```
ggplot(diamonds, aes(x = clarity, y = table, col = clarity)) +
  geom_violin() +
  labs(title = "The table width of diamonds for each clarity",x = "Clarity", y = "Table") +
  theme_minimal()
```



Each clarity has a table value ranging from approximately 55 to 65, with the SI1 clarity having the highest table value.