Homework Data Visualization

Napat Teekasuk

2023-07-04

Introduction

Use diamond data set to create 5 chart, also contain chart description. Submit PDF file

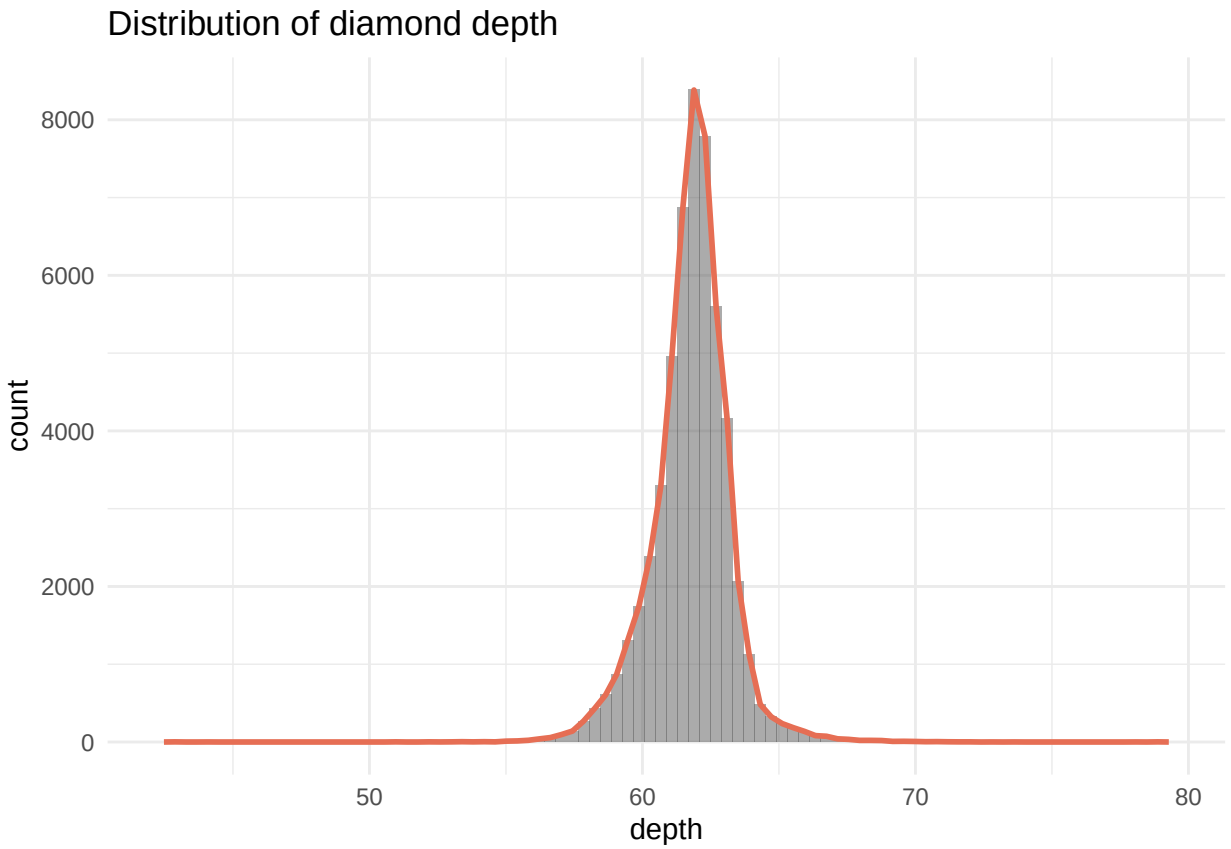
Load Library

```
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. The distributio

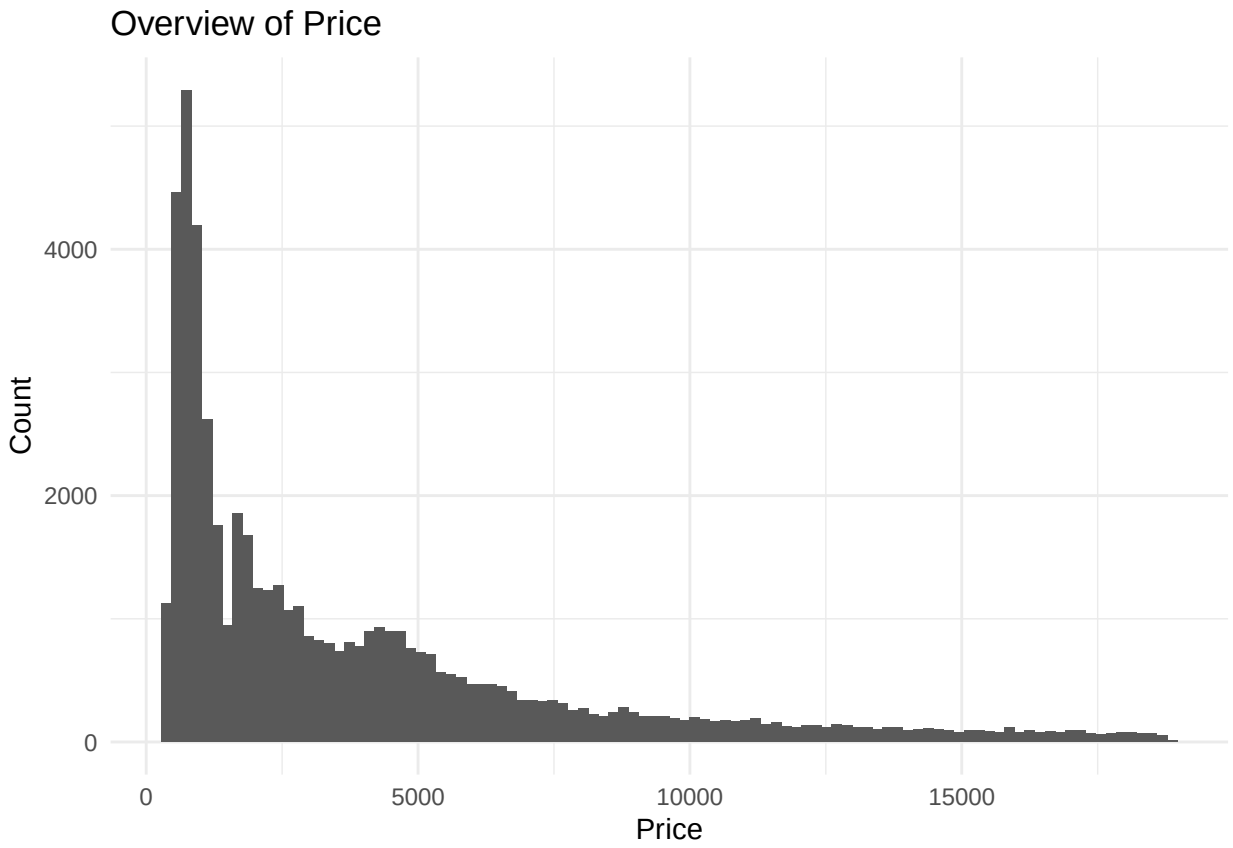
```
ggplot(diamonds, aes(depth)) +
  geom_histogram(bins = 90, alpha = 0.5) +
  geom_freqpoly(bins = 90, col = "#E66E54", linewidth = 1) +
  theme_minimal() +
  labs(title = "Distribution of diamond depth")
```



- This distribution is normal distribution

2. Overview of Price

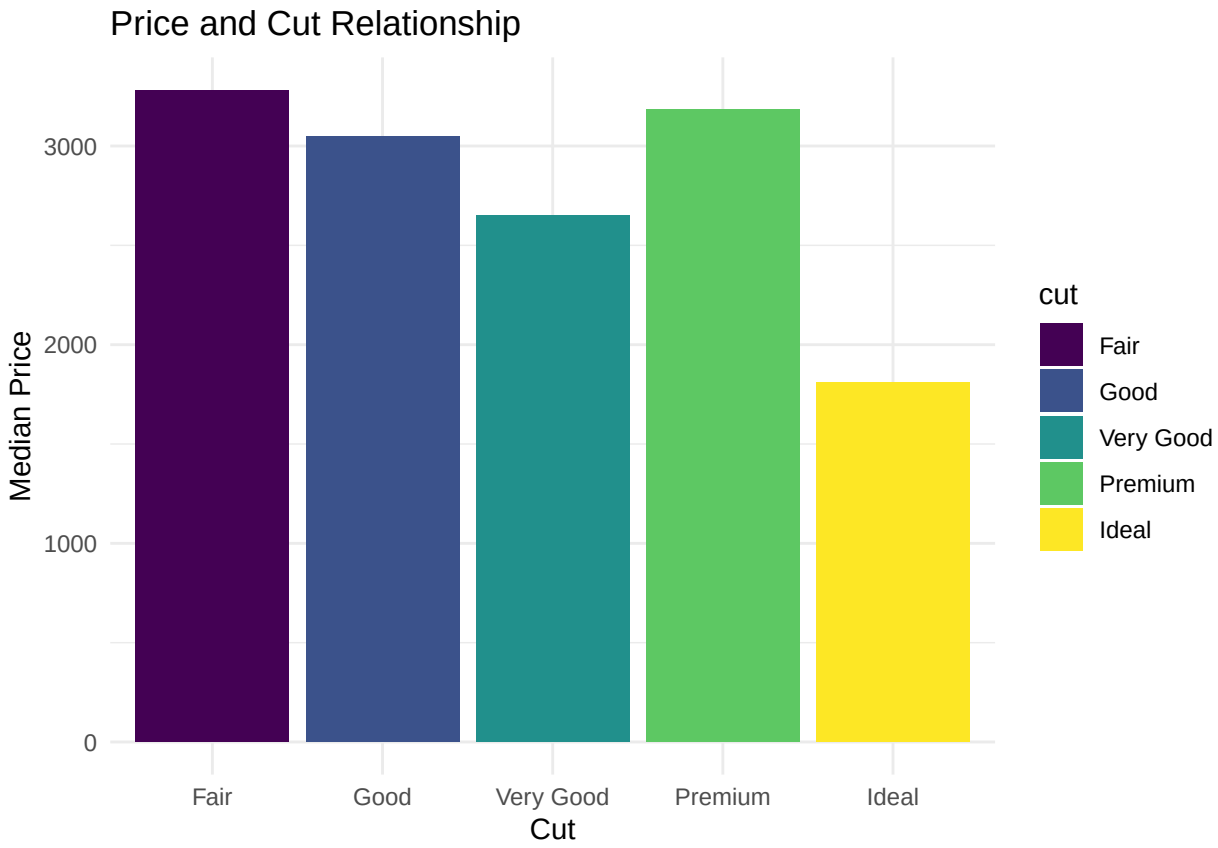
```
ggplot(diamonds, aes(price)) +
  geom_histogram(bins = 100) +
  theme_minimal() +
  labs(
    title = "Overview of Price",
    x = "Price",
    y = "Count"
  )
```



- Minimum Price is 326
- Maximun Price is 18823
- Average Price is 3933

3. Price and Cut Relationship

```
diamonds %>%
  group_by(cut) %>%
  summarise(
    med_price = median(price)
  ) %>%
  ggplot(aes(cut, med_price, fill = cut)) +
  geom_col() +
  theme_minimal() +
  labs(
    title = "Price and Cut Relationship",
    x = "Cut",
    y = "Median Price",
  )
```

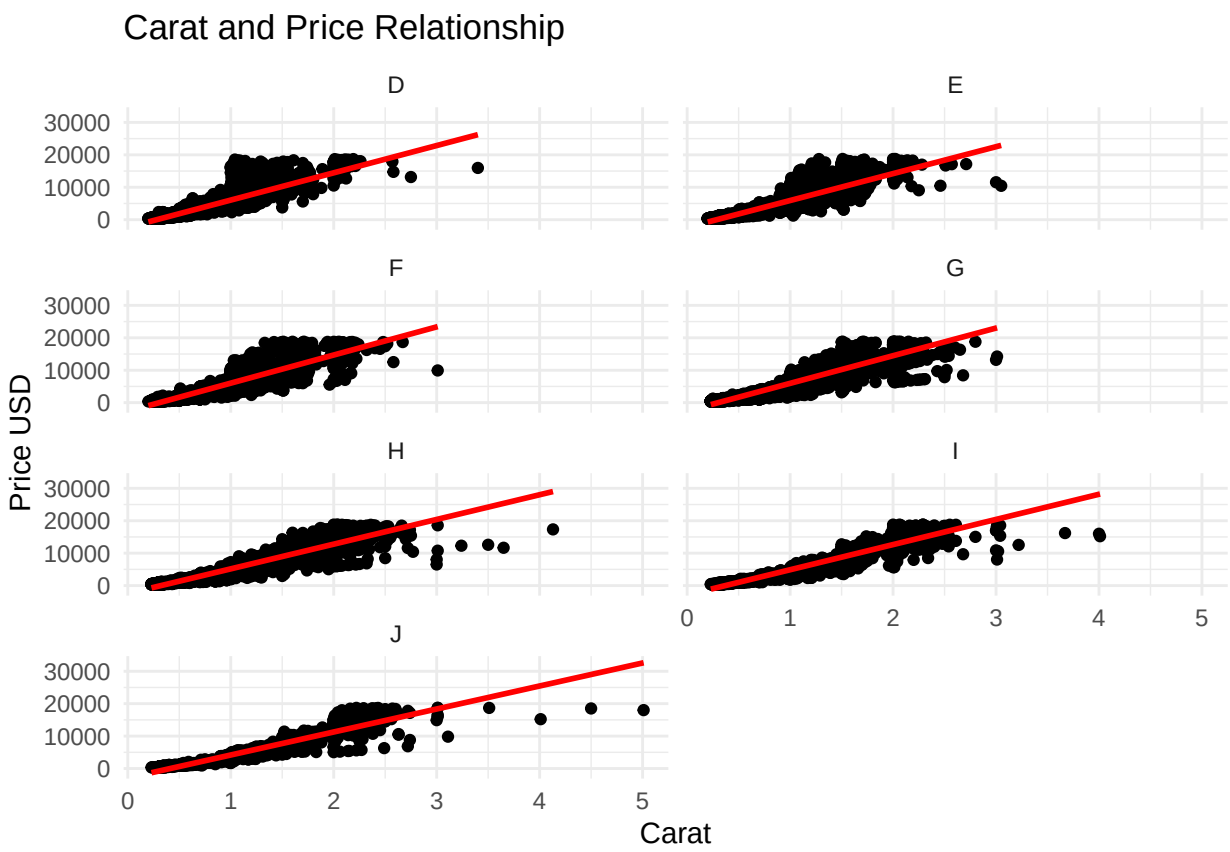


- Use median as middle value to represent the price of diamonds
- Fair » Premiun » Good » Very Good » Ideal (sort by median price)

4. Carat and Price Relationship

```
ggplot(diamonds, aes(carat, price)) +
  geom_point() +
  geom_smooth(method = "lm" , col = "red") +
  facet_wrap(~color, ncol = 2) +
  theme_minimal() +
  labs(title = "Carat and Price Relationship",
       x = "Carat",
       y = "Price USD")
```

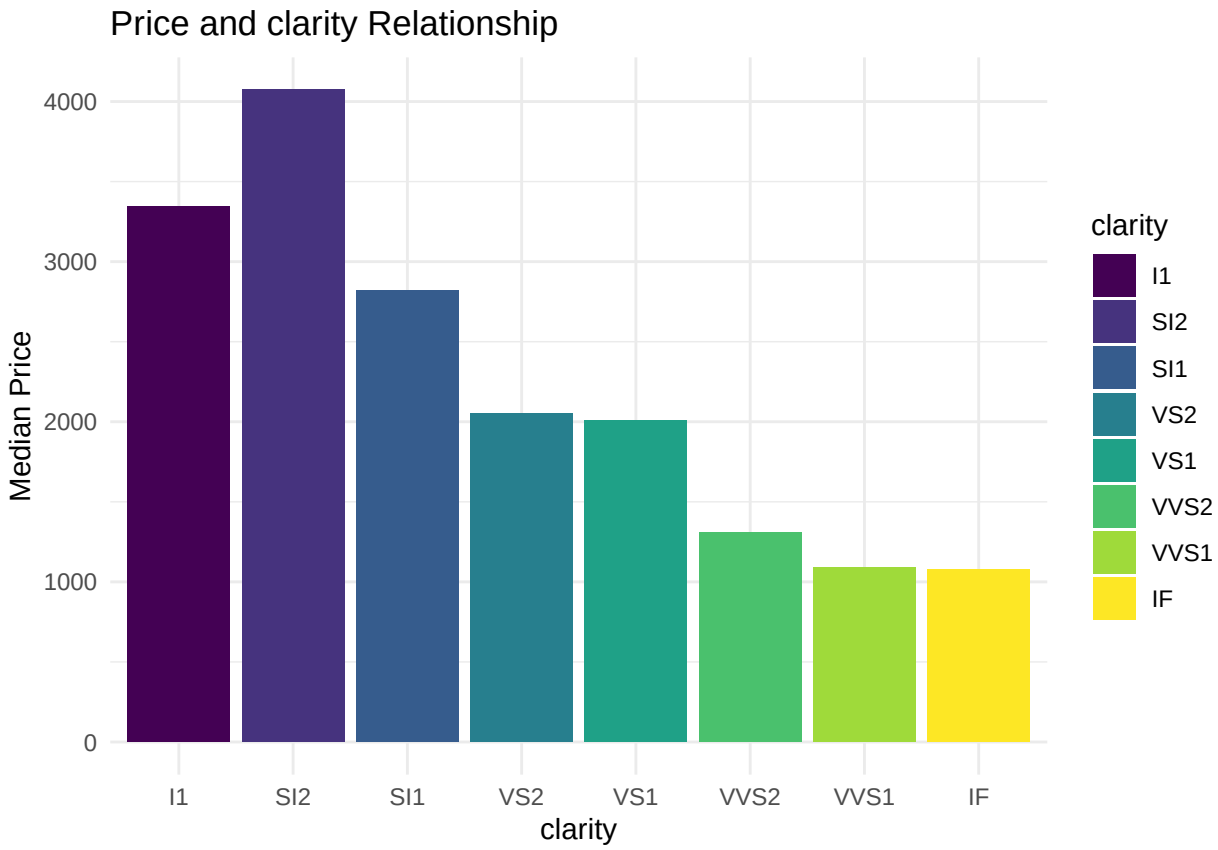
```
## `geom_smooth()` using formula = 'y ~ x'
```



- Correlation of Carat and Price are about 0.92

5. Price and clarity Relationship

```
diamonds %>%
  group_by(clarity) %>%
  summarise(median_price = median(price)) %>%
  ggplot(aes(clarity, median_price, fill=clarity)) +
  geom_col() +
  theme_minimal() +
  labs(
    title = "Price and clarity Relationship",
    x = "clarity",
    y = "Median Price",
  )
```



- Use median as middle value to represent the price of diamonds
- SI2 » I1 » SI1 » VS2 » VS1 » VVS2 » VVS1 » IF (sort by median price)