Homework Data Viz Batch 10

Wisaithat Tiiw

2024-08-13

TEST.H1

TEST.H2

TEST.H3

TEST TEXT

In-Depth Analysis of Diamond Data: Exploring Trends and Distributions of Price and Quality Using Five Graphs

1. Load the library (run this every time you open R)

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- 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
```

2. Prepare Data

```
set.seed(18)
prep.diamonds <- diamonds %>%
  sample_frac(0.25)
```

3. Visualization

1. Histogram of Diamond Prices

This histogram shows the distribution of diamond prices in US dollars.As we can see from the histogram,most diamonds are priced in the lower range, and only a small proportion of diamonds have very high prices.

```
ggplot(prep.diamonds, aes(x = price)) +
  geom_histogram(binwidth = 500, fill = "blue", col = "black") +
```

```
  theme_minimal() +
  labs(title = "Histogram of Diamond Prices",
       x = "Price ($USD)",
       y = "Count")
```

Histogram of Diamond Prices
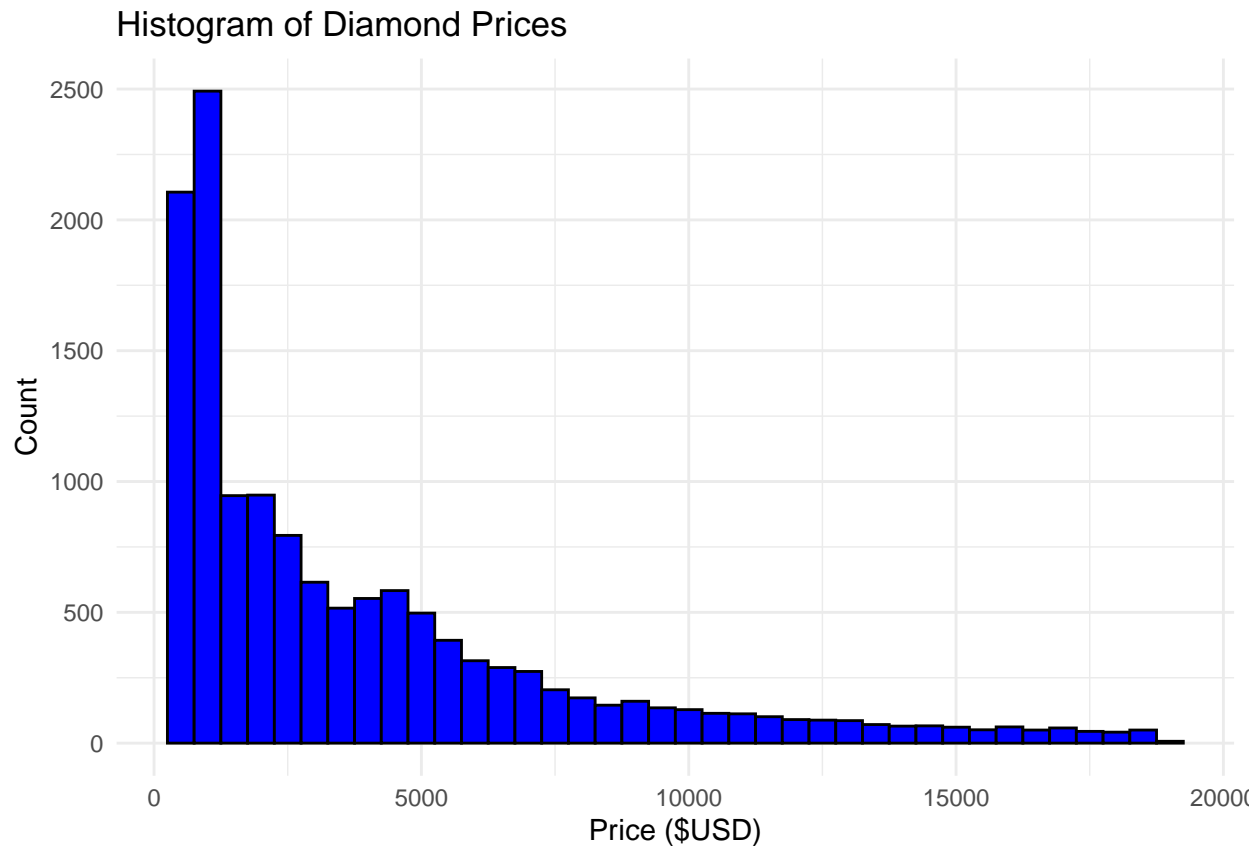
2. Bar Chart of Diamond Cut Quality

This bar chart displays the number of diamonds categorized into five different cut quality levels.And we can tell diamonds with an Ideal cut quality are the most numerous

```
ggplot(prep.diamonds, aes(x = cut)) +
  geom_bar(fill = "blue", col = "black") +
  theme_minimal() +
  labs(title = "Bar Chart of Diamond Cut Quality",
       x = "Cut Quality",
       y = "Count")
```

Bar Chart of Diamond Cut Quality
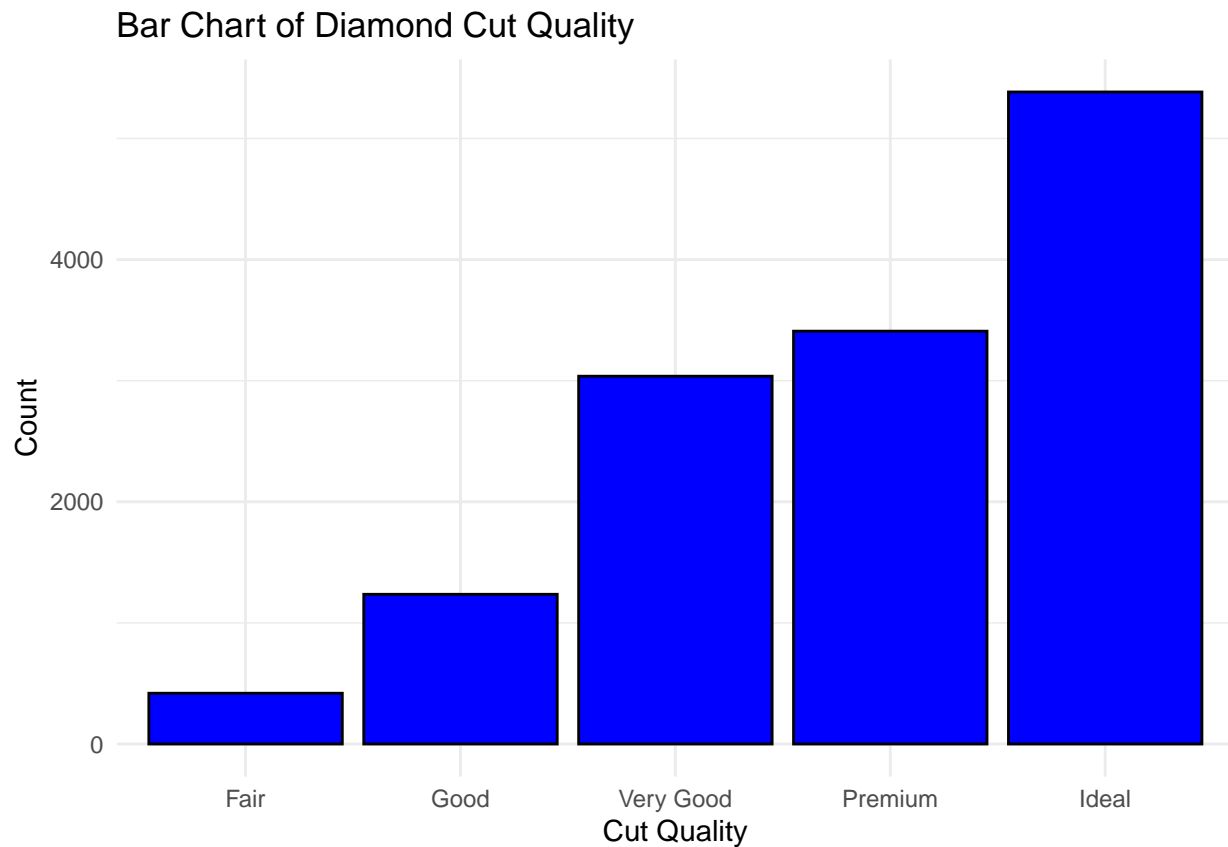
3. Stacked bar chart (Bar Chart + mapping Col)

Each bar represents a "cut" quality (Fair, Good, Very Good, Premium, Ideal), and the height of each bar shows the total count of diamonds within that cut category.

```
ggplot(prep.diamonds, aes(x = cut , fill = color)) +
  geom_bar() +
  theme_minimal() +
  labs(title = "Diamond Count by Cut and Color",
       x = "Cut Quality",
       y = "Count")
```

Diamond Count by Cut and Color
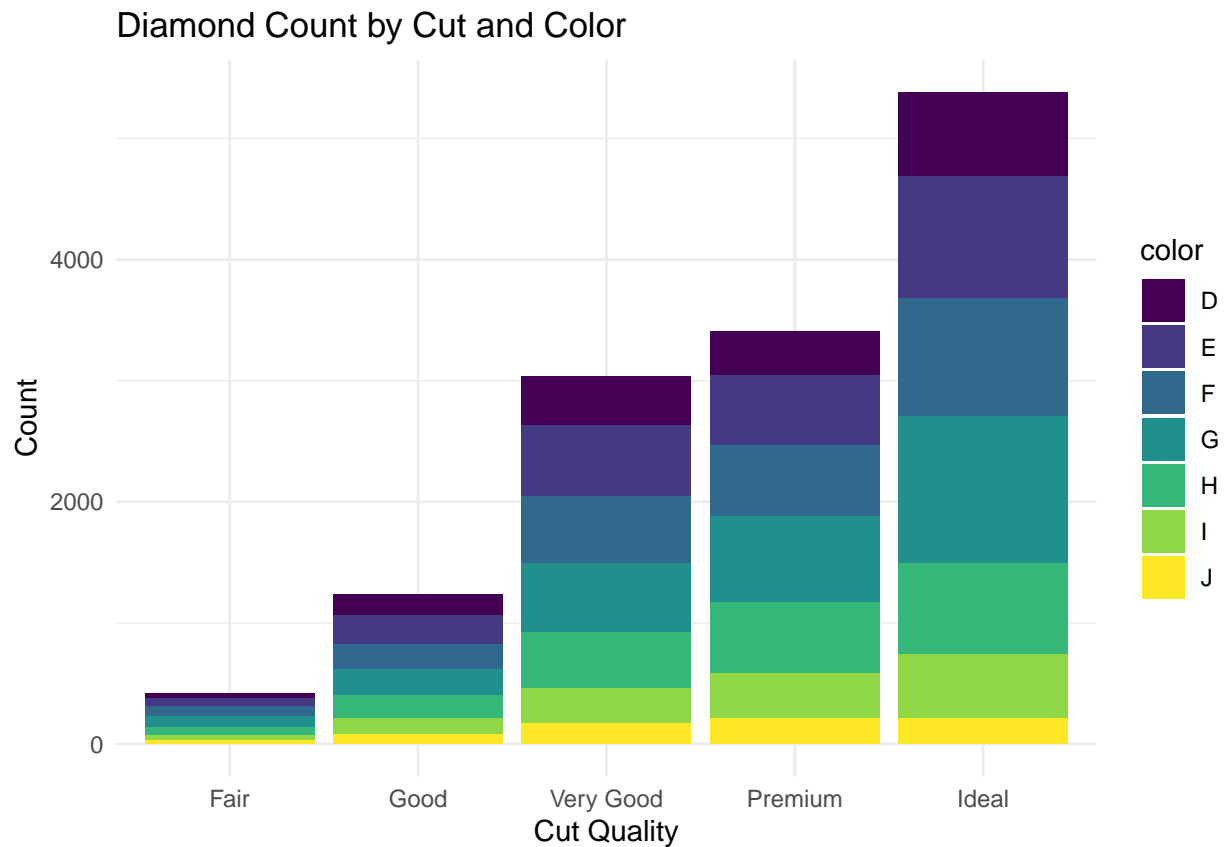
4. Average Diamond Price by Color

Diamonds with color grade D have the highest average price, while those with color grade J have the lowest.

```r
avg_price_by_col <- prep.diamonds %>%
  group_by(color) %>%
  summarise(avg_price = mean(price))
```

```r
ggplot(avg_price_by_col, aes(x = color, y = avg_price, fill = color)) +
  geom_col() +
  theme_minimal() +
  labs(title = "Average Diamond Price by Color",
       x = "Color of Diamond",
       y = "Average price")
```

Average Diamond Price by Color
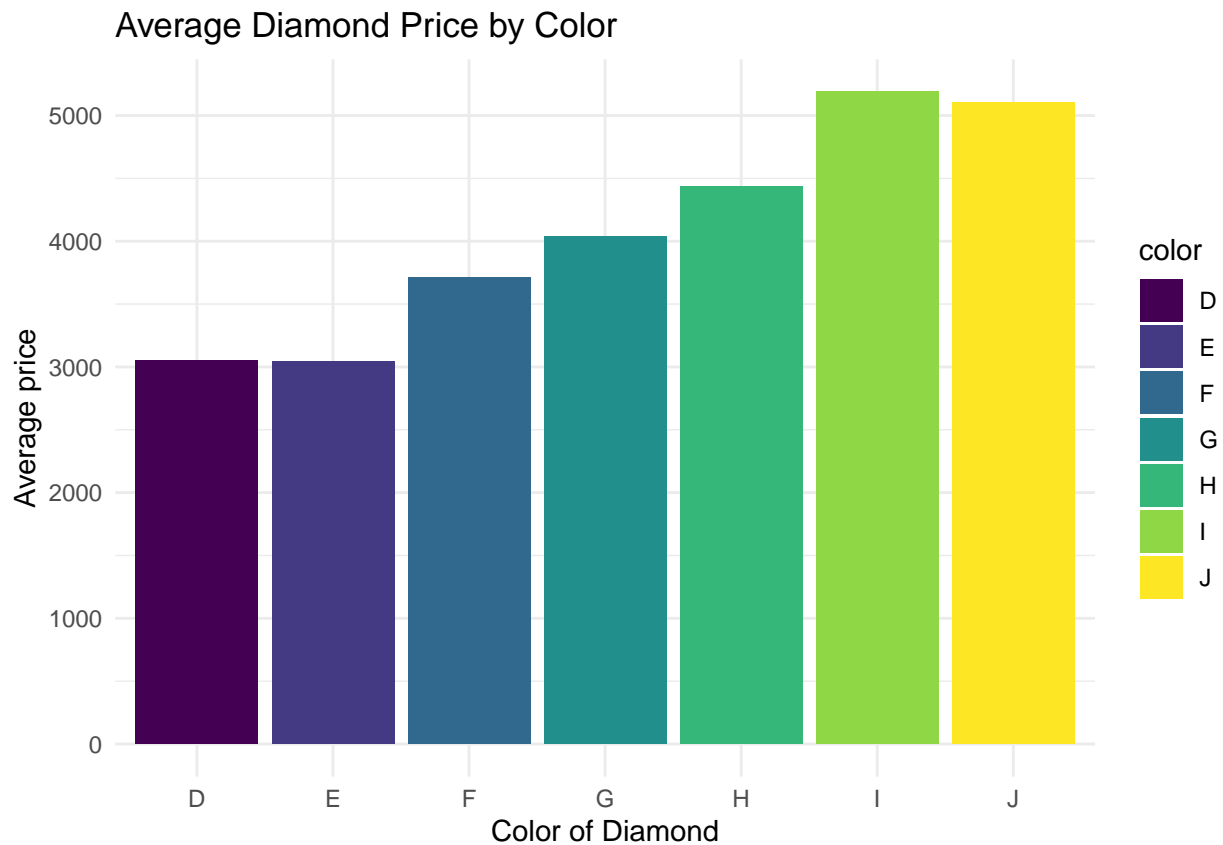
5. Facetted Scatter Plot: Price vs. Carat by Cut mapping with color

Each subplot shows the relationship between "carat" (x-axis) and "price" (y-axis) for diamonds of a specific cut quality. The color of each point represents the "color" of the diamond. The chart shows what makes a diamond more expensive. Bigger diamonds, diamonds with better cuts, and diamonds with clearer colors usually cost more.

```
ggplot(prep.diamonds, aes(x = carat, y = price, col = color)) +
  geom_point(size = 1, alpha = 0.1) +
  geom_smooth(method = "lm") +
  facet_wrap(~ cut) +
  labs(title = "Price vs. Carat by Cut",
       x = "Carat",
       y = "Price")
```

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

Price vs. Carat by Cut
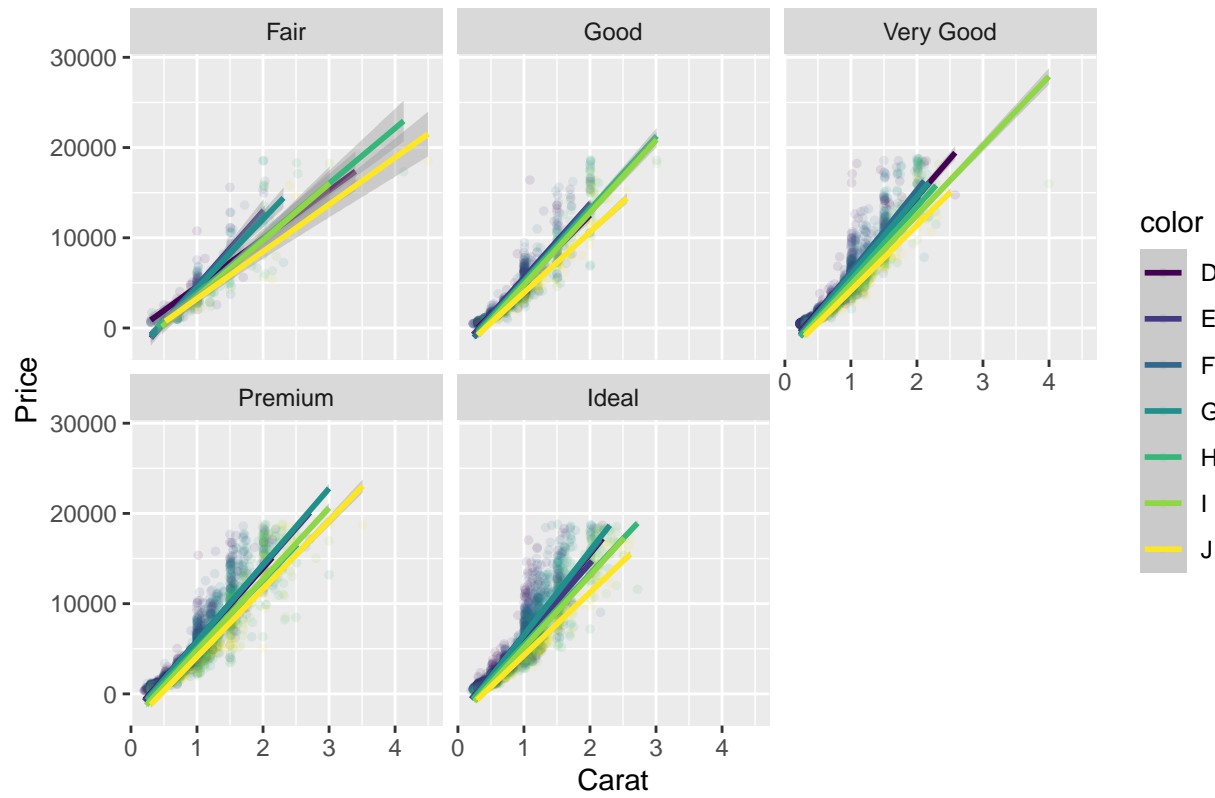
Studied to the point of begging for mercy