

Data Visualization Bootcamp Homework

Ananya Abdulla

2023-08-10

Instruction

Use the diamonds dataset to create five visualizations. Generate a PDF of the visualizations and submit it to Discord.

```
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 ggplot2    3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
## -- 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
```

```
library(ggplot2)
```

1.) Scatter Plot of Carat vs. Price

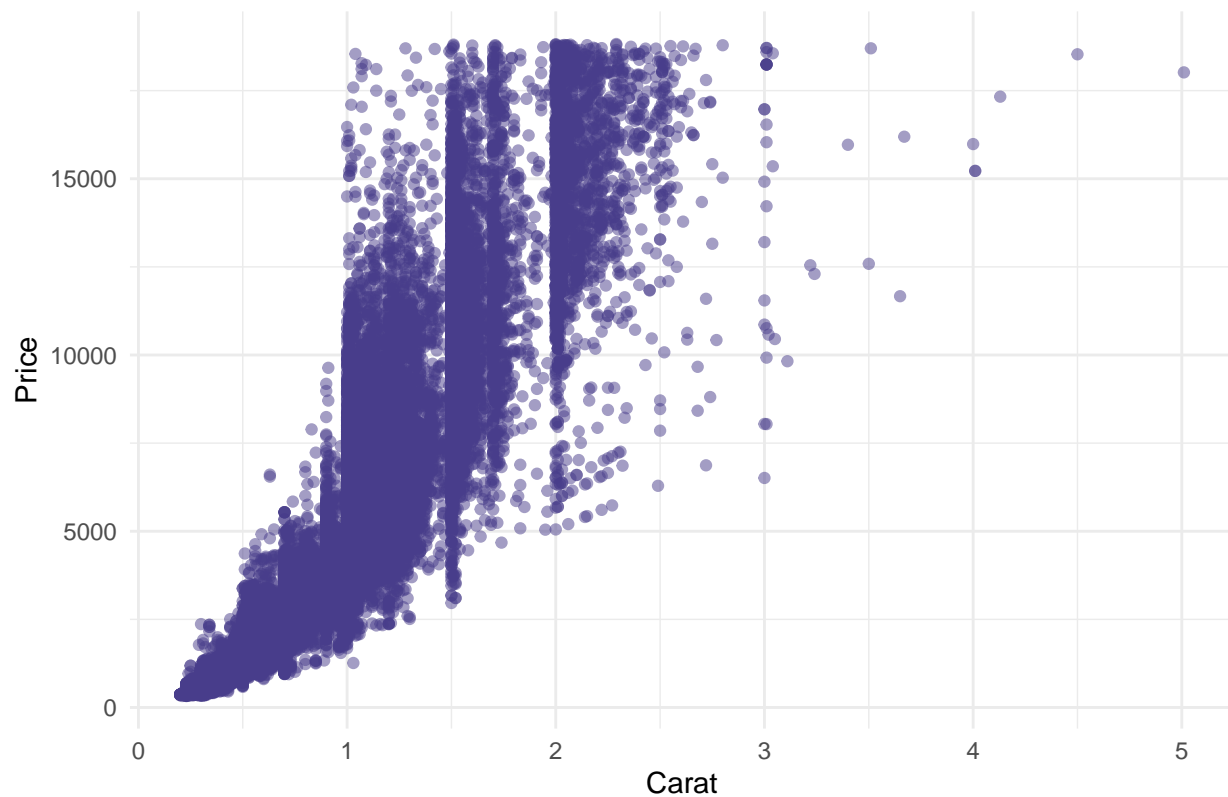
This chart visualizes the relationship between diamond carat and price using a scatter plot.

```
library(ggplot2)
```

```
scatter_plot <- ggplot(diamonds, aes(x = carat, y = price)) +
  geom_point(alpha = 0.5, color = "darkslateblue") +
  labs(title = "Scatter Plot of Carat vs. Price",
       x = "Carat", y = "Price") +
  theme_minimal()

print(scatter_plot)
```

Scatter Plot of Carat vs. Price



The scatter plot shows the relationship between the carat weight of diamonds and their prices. Each diamond is represented by a point, and the x-axis shows the carat weight while the y-axis shows the price. The transparency of the points is adjusted to prevent them from overlapping.

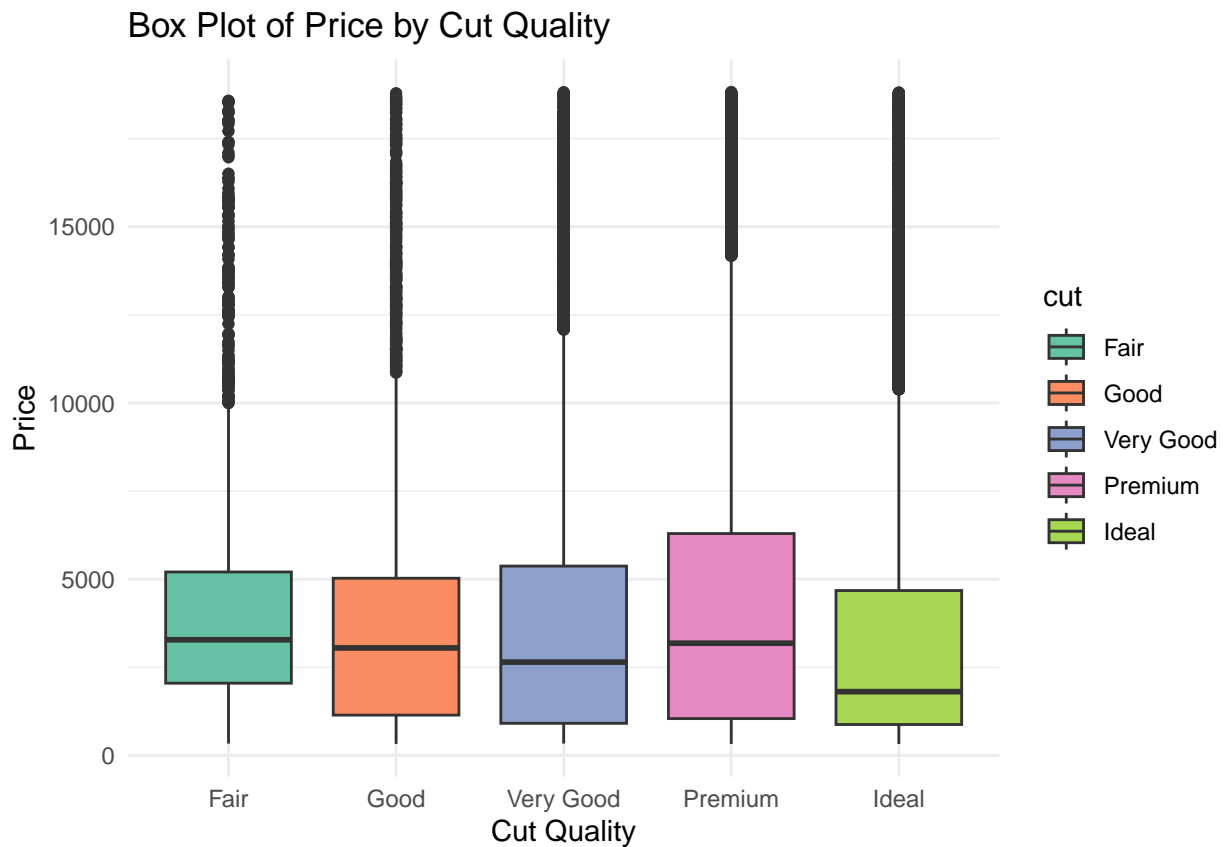
2.) Box Plot of Price by Cut Quality

This chart visualizes the spread of diamond prices for different cut qualities.

```
library(ggplot2)

box_plot <- ggplot(diamonds, aes(x = cut, y = price, fill = cut)) +
  geom_boxplot() +
  labs(title = "Box Plot of Price by Cut Quality",
       x = "Cut Quality", y = "Price") +
  scale_fill_brewer(palette = "Set2") +
  theme_minimal()

print(box_plot)
```



The box plot shows the distribution of diamond prices by cut quality. Each box represents a cut quality category, and the y-axis shows the price. This visualization helps to analyze the variation in prices across different cut qualities.

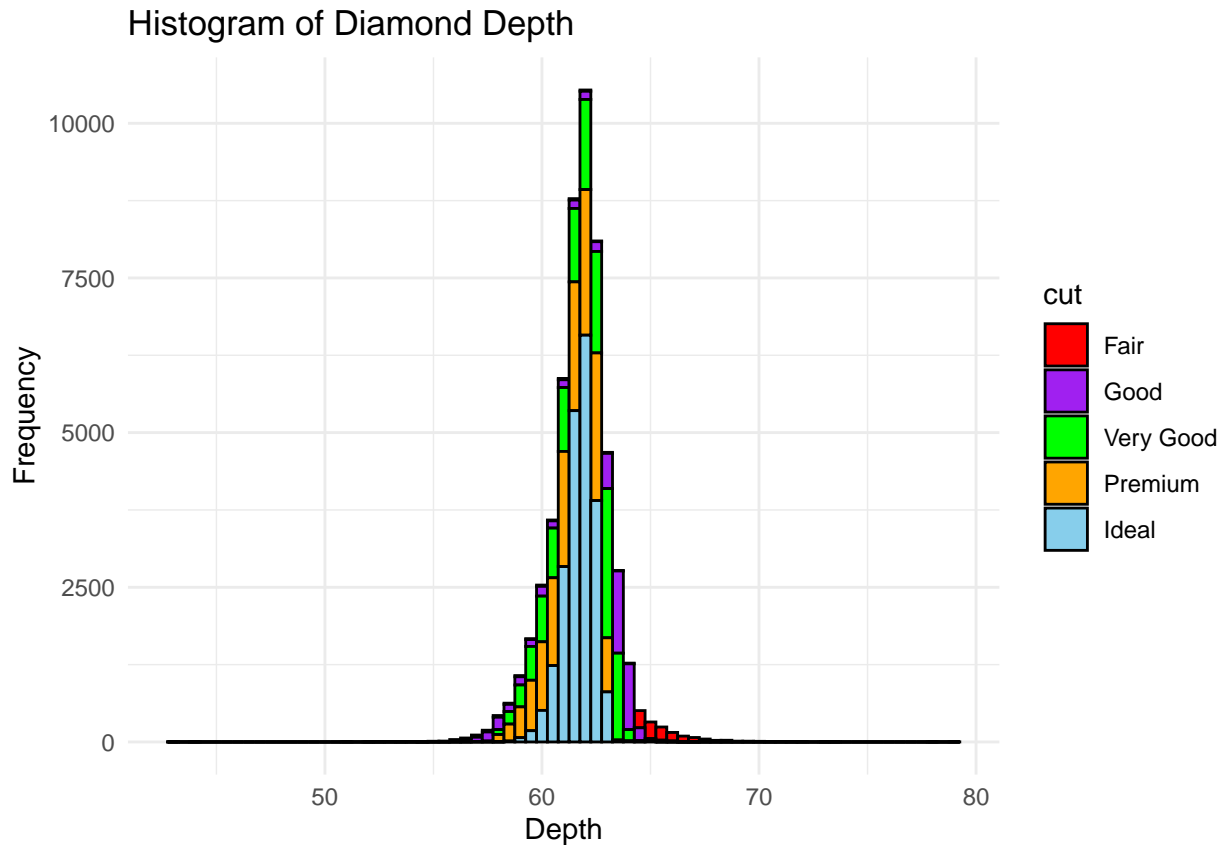
3.) Histogram of Diamond Depth

This chart visualizes the distribution of diamond depths.

```
library(ggplot2)

histogram_plot <- ggplot(diamonds, aes(x = depth, fill = cut)) +
  geom_histogram(binwidth = 0.5, color = "black") +
  labs(title = "Histogram of Diamond Depth",
       x = "Depth", y = "Frequency") +
  scale_fill_manual(values = c("Ideal" = "skyblue", "Premium" = "orange",
                              "Very Good" = "green", "Good" = "purple",
                              "Fair" = "red")) +
  theme_minimal()

print(histogram_plot)
```



The histogram shows the distribution of diamond depths. The x-axis shows the depth values, and the y-axis shows the number of diamonds with that depth. The title of the plot indicates that it is exploring the patterns of diamond depths.

4.) Average Price of Diamonds by Cut Quality

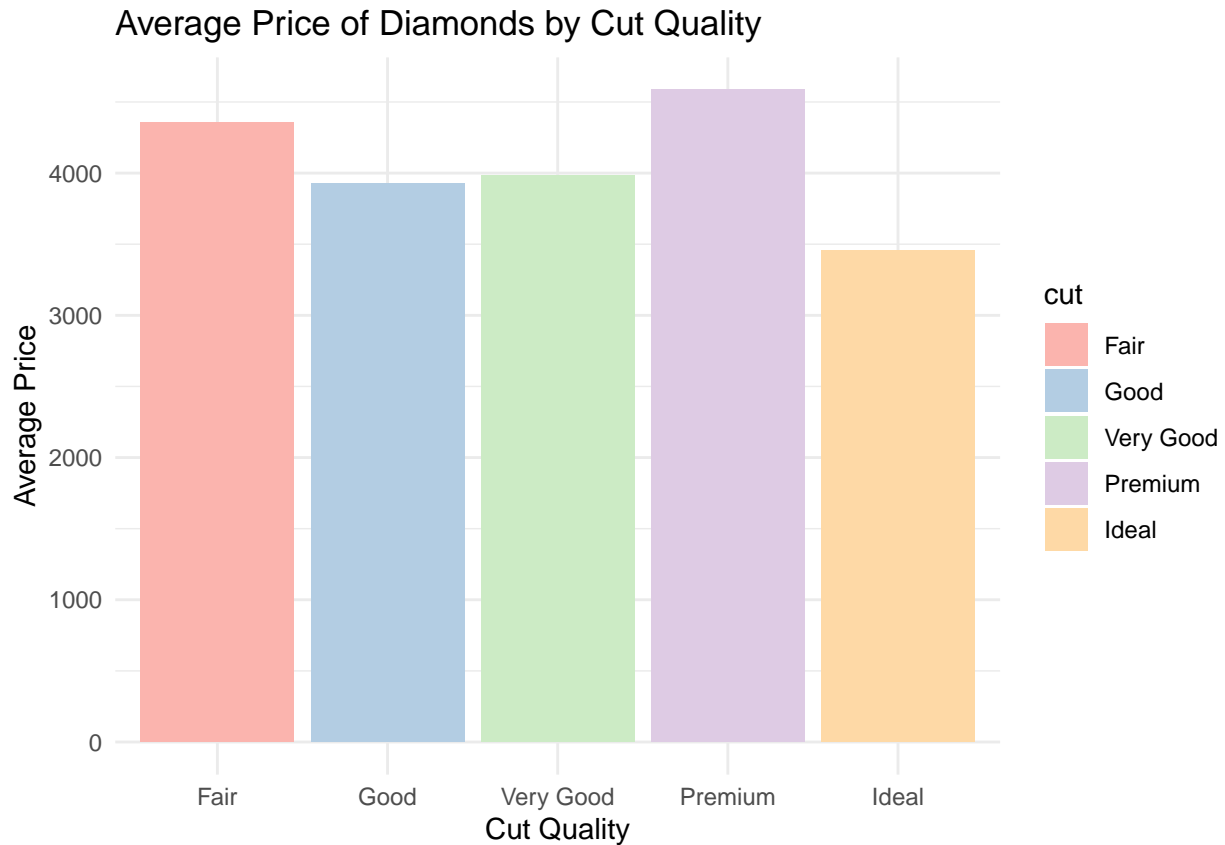
This bar chart shows the average price of diamonds by diamond cut.

```
library(ggplot2)
library(dplyr)

avg_price_by_cut <- diamonds %>%
  group_by(cut) %>%
  summarize(avg_price = mean(price))

bar_plot <- ggplot(avg_price_by_cut, aes(x = cut, y = avg_price, fill = cut)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Price of Diamonds by Cut Quality",
       x = "Cut Quality", y = "Average Price") +
  scale_fill_brewer(palette = "Pastel1") +
  theme_minimal()

print(bar_plot)
```



The bar chart shows the average price of diamonds by cut quality, with each bar representing a cut quality category and the height of the bar corresponding to the average price. The title of the plot indicates that it is exploring how average prices vary across different cut qualities.

5.) Carat-Price Relationship by Diamond Cut

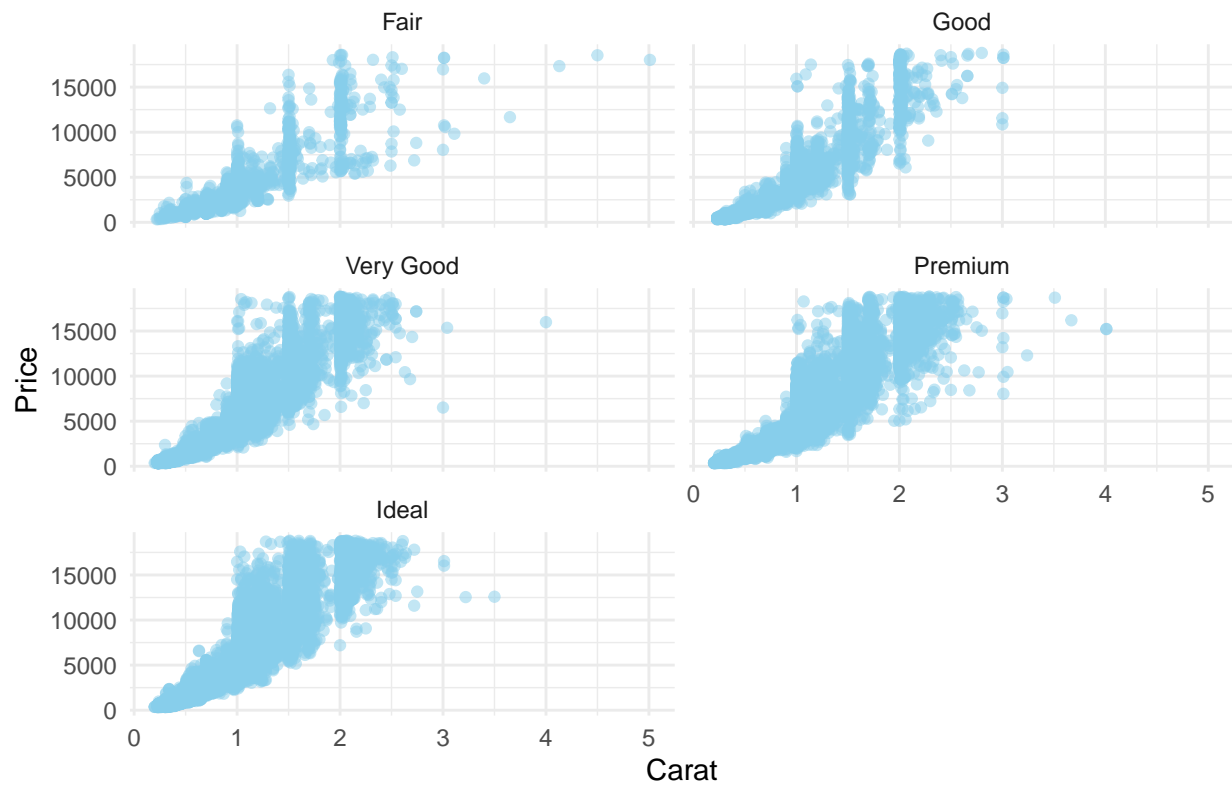
This scatter plot shows the relationship between carat and price for different diamond cuts, with facets used to differentiate the data by cut.

```
library(ggplot2)

faceted_scatter_plot <- ggplot(diamonds, aes(x = carat, y = price)) +
  geom_point(alpha = 0.5, color = "skyblue") +
  facet_wrap(~ cut, ncol = 2) +
  labs(title = "Carat-Price Relationship by Diamond Cut",
       x = "Carat", y = "Price") +
  theme_minimal()

print(faceted_scatter_plot)
```

Carat–Price Relationship by Diamond Cut



The scatter plot with facets shows the carat-price relationship for different diamond cuts. Each facet represents a different cut quality, and the points within each facet show the distribution of carat and price for that cut quality. The title of the plot indicates that it is exploring how the carat-price relationship varies across different cut qualities.