Data Visualization With ggplot2

Adejumo Ridwan Suleiman

2022-09-25

Installing and Loading ggplot2

```
#install.packages("ggplot2")
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.1
```

Diamonds dataset

```
head(diamonds)
```

```
## # A tibble: 6 x 10
##   carat cut       color clarity depth table price     x     y     z
##   <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1  0.23 Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
## 2  0.21 Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
## 3  0.23 Good      E     VS1      56.9    65   327  4.05  4.07  2.31
## 4  0.29 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63
## 5  0.31 Good      J     SI2      63.3    58   335  4.34  4.35  2.75
## 6  0.24 Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48
```

Economics dataset

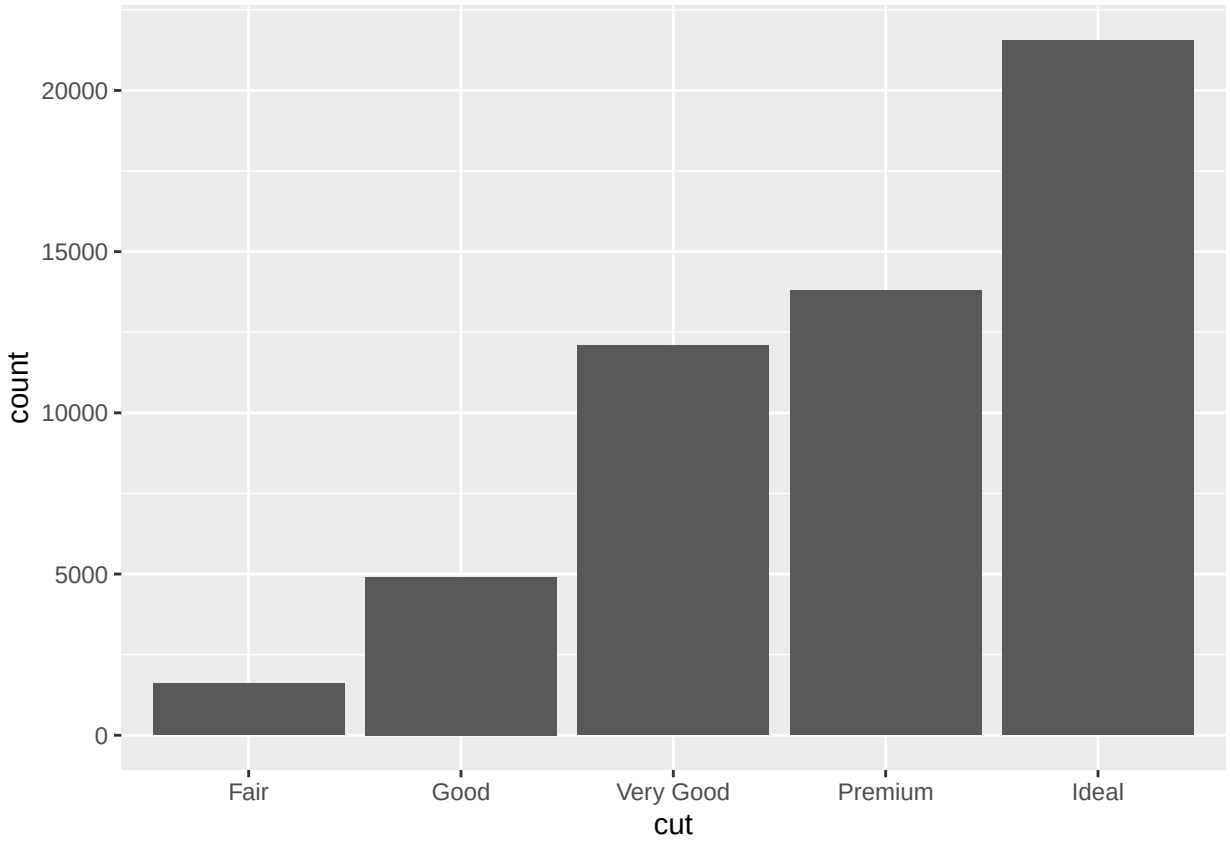
```
head(economics)
```

```
## # A tibble: 6 x 6
##   date         pce    pop psavert uempmed unemploy
##   <date>     <dbl>  <dbl>   <dbl>   <dbl>    <dbl>
## 1 1967-07-01  507. 198712    12.6     4.5     2944
## 2 1967-08-01  510. 198911    12.6     4.7     2945
## 3 1967-09-01  516. 199113    11.9     4.6     2958
## 4 1967-10-01  512. 199311    12.9     4.9     3143
## 5 1967-11-01  517. 199498    12.8     4.7     3066
## 6 1967-12-01  525. 199657    11.8     4.8     3018
```

Barplots

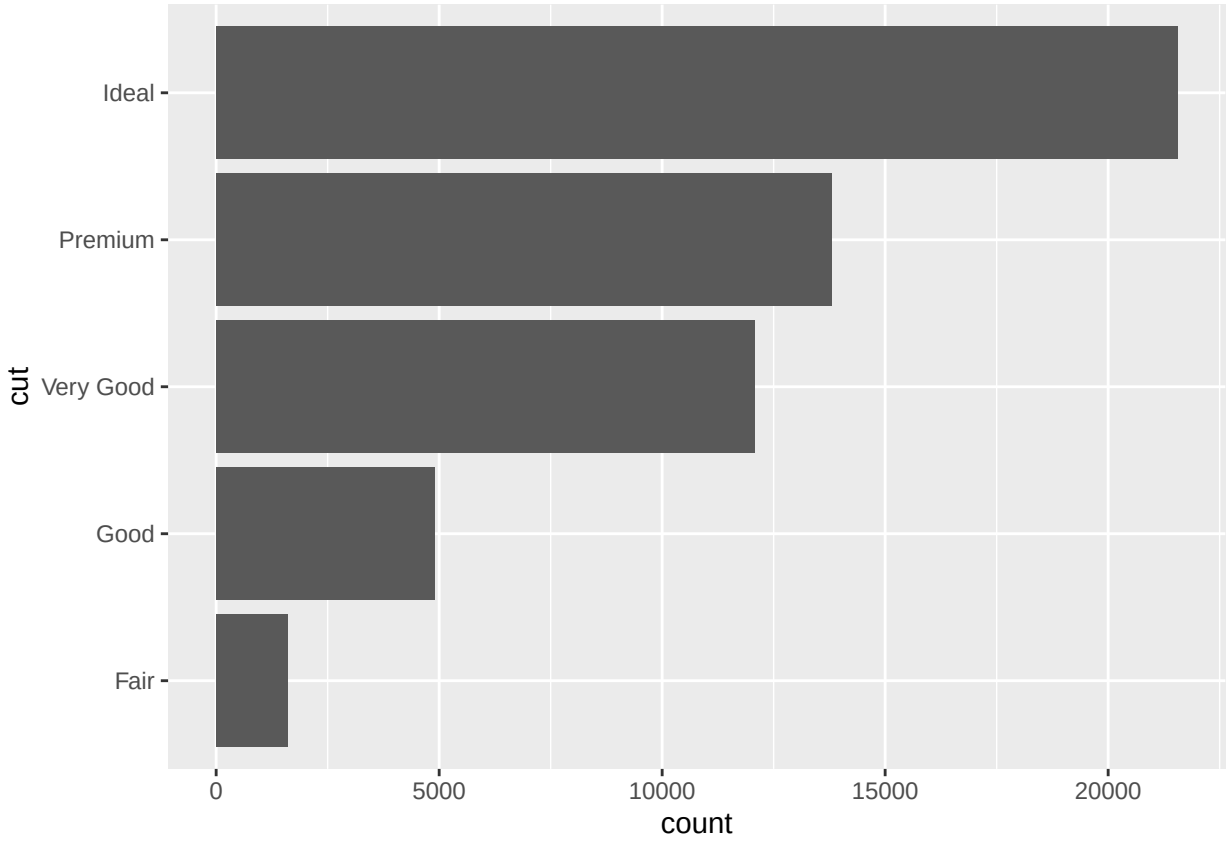
Vertical Bar Plot

```
# Most occurring diamonds (change fill with "fill = ", change line colour with "colour = ")
diamonds |>
  ggplot(mapping = aes(x = cut)) +
  geom_bar()
```



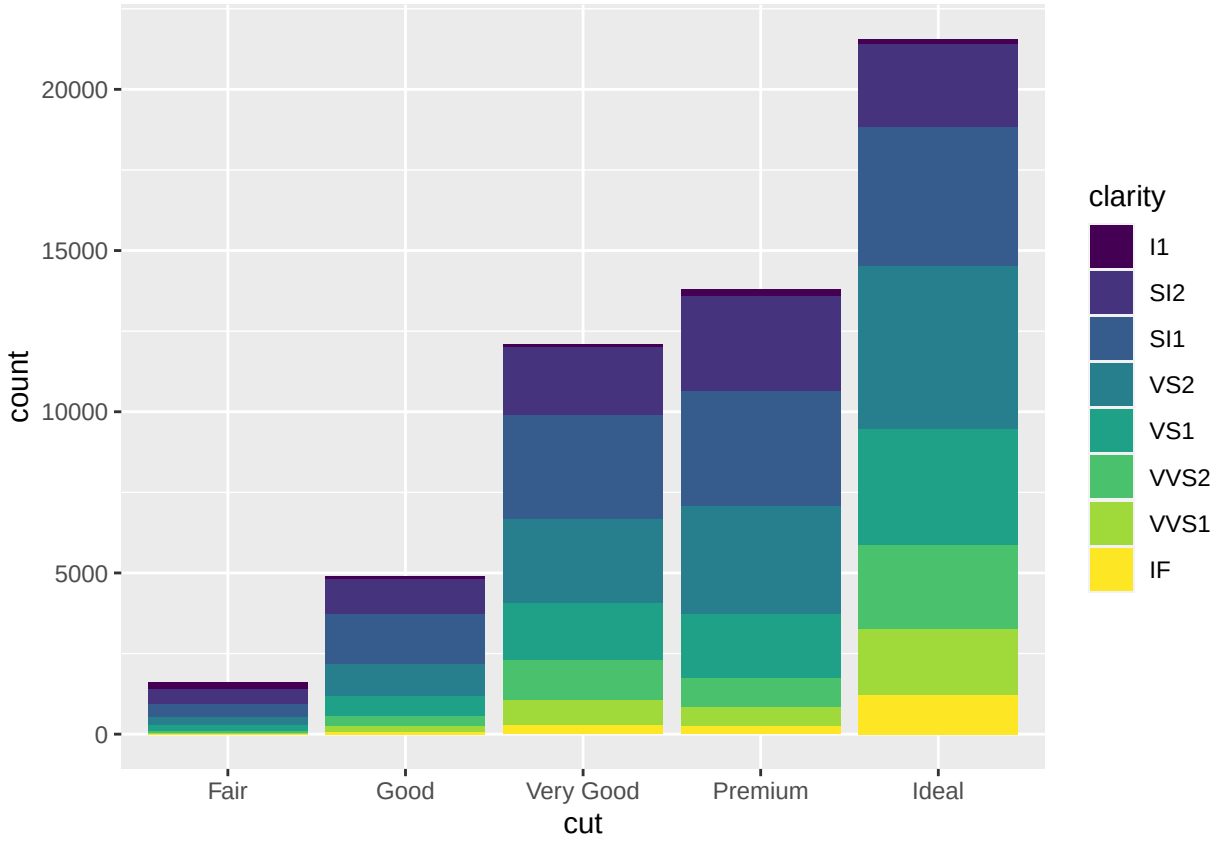
Horizontal Bar Plot

```
diamonds |>
  ggplot(mapping = aes(y = cut)) +
  geom_bar()
```



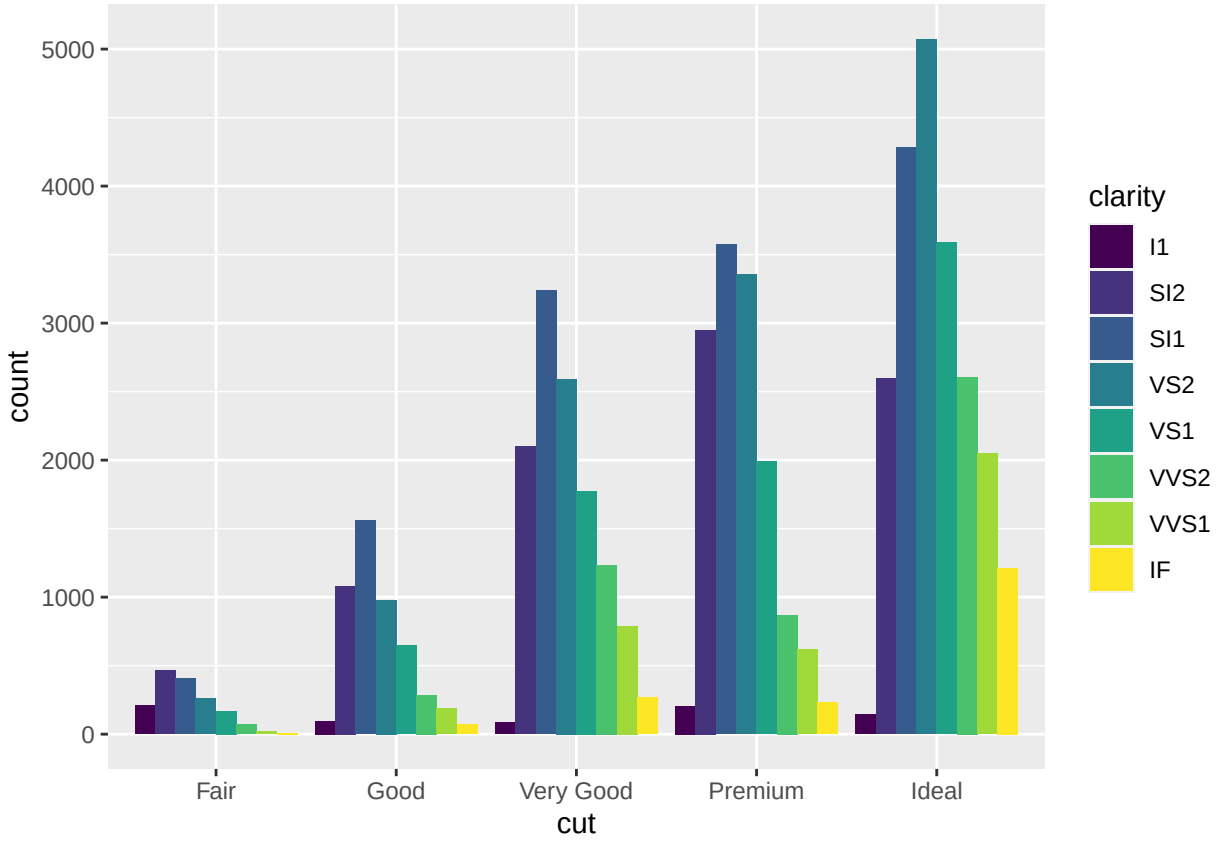
Stacked BarPlot

```
diamonds |>
  ggplot(mapping = aes(x = cut, fill = clarity), position = "fill") +
  geom_bar()
```



Grouped Barplot

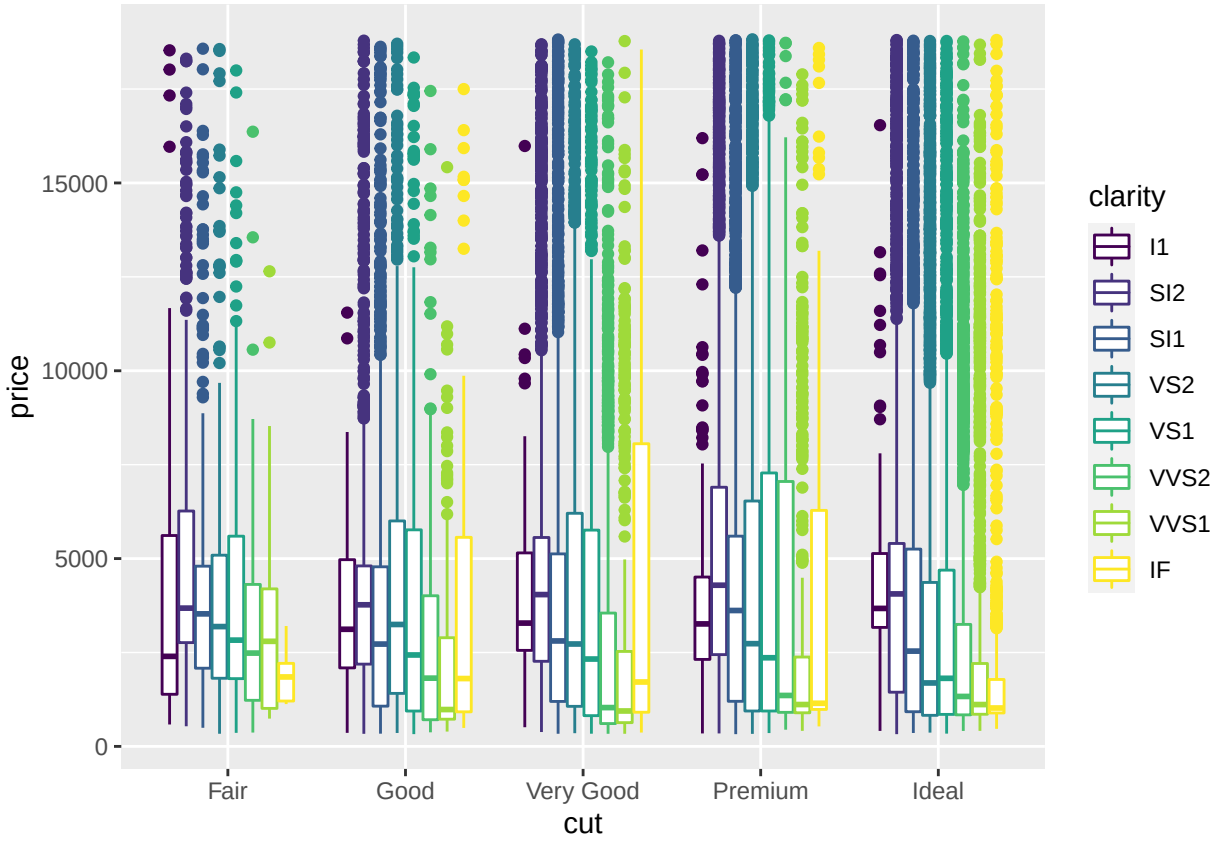
```
ggplot(data = diamonds) +
  geom_bar(mapping = aes(x = cut, fill = clarity), position = "dodge")
```



Boxplots

Comparing diamond cut and price

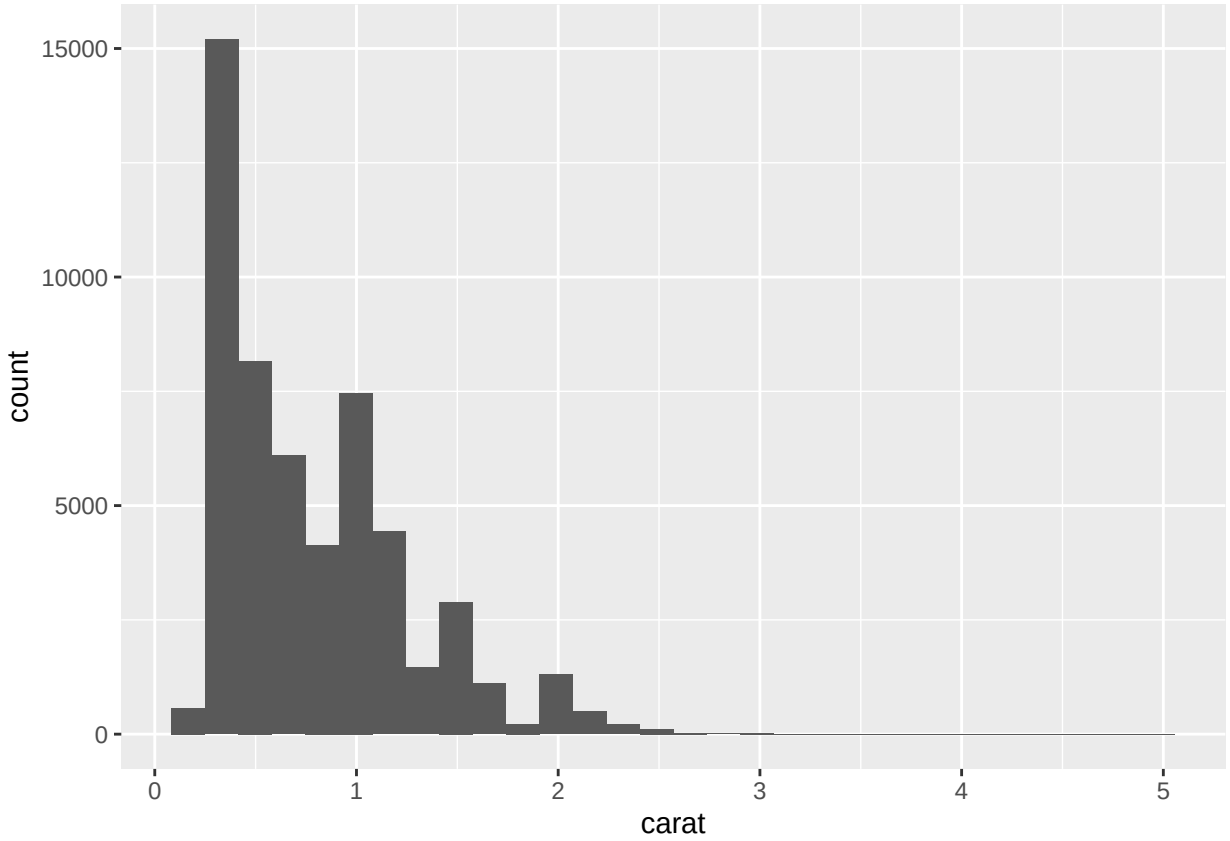
```
diamonds |>
  ggplot(mapping = aes(x = cut, y = price, colour = clarity)) +
  geom_boxplot()
```



Histogram

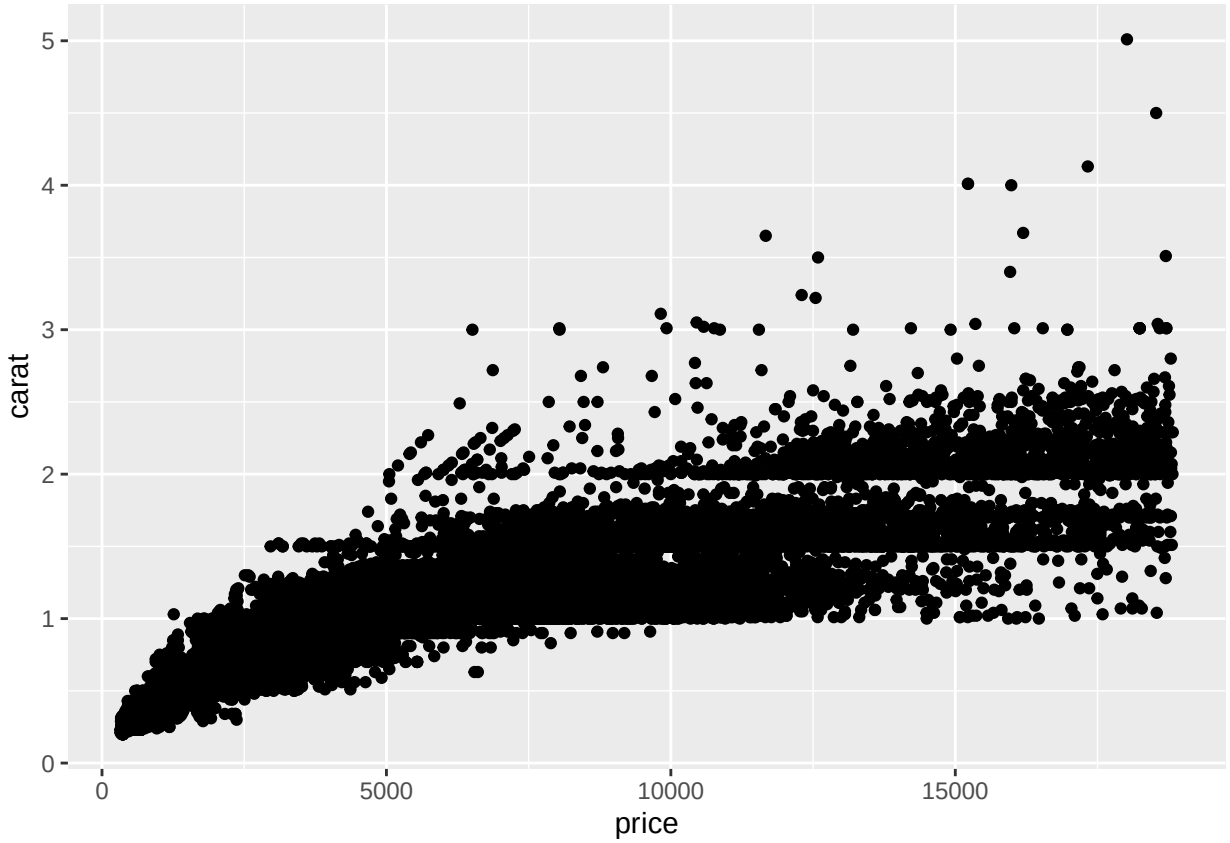
```
#Check for the distribution of diamond carat
diamonds |>
  ggplot(aes(x = carat)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



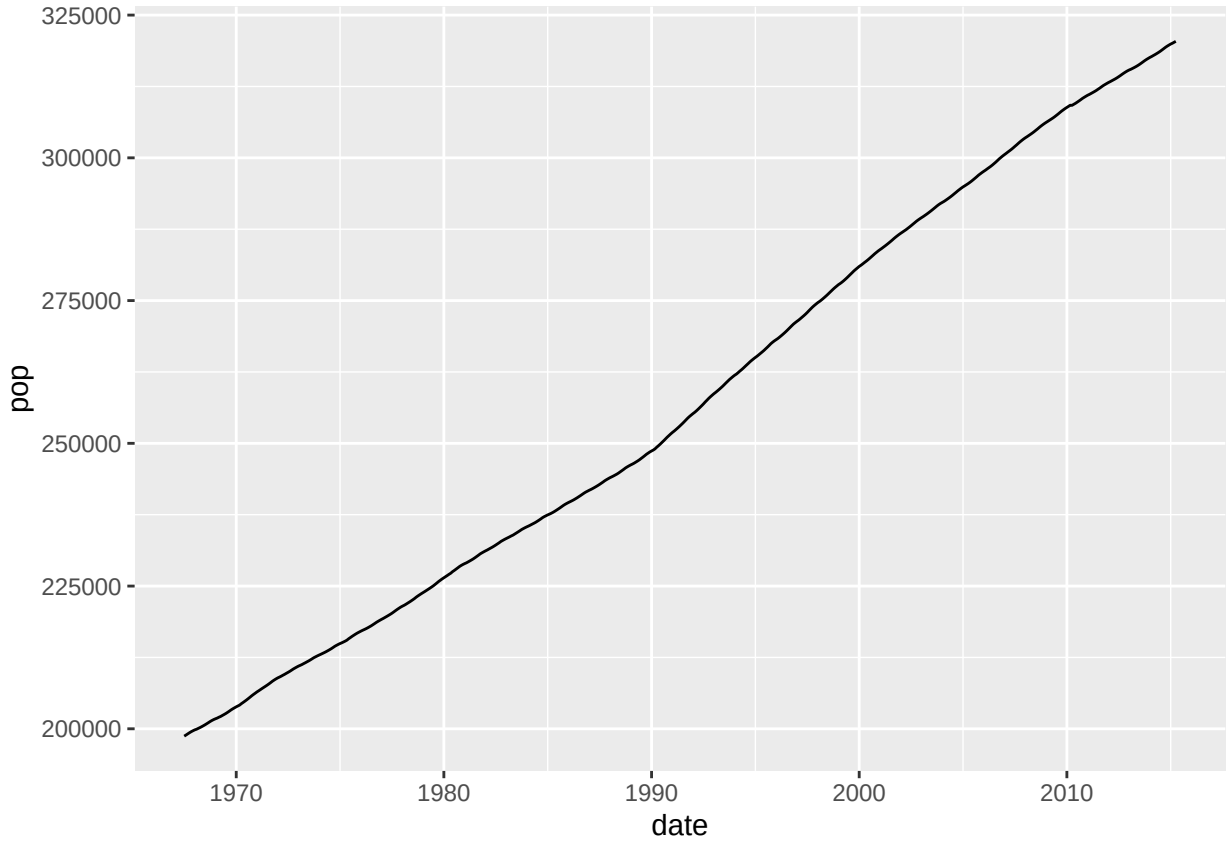
Scatter Plot

```
# Check for the relationship between diamond price and carat
diamonds |>
  ggplot(aes(x = price, y = carat)) +
  geom_point()
```



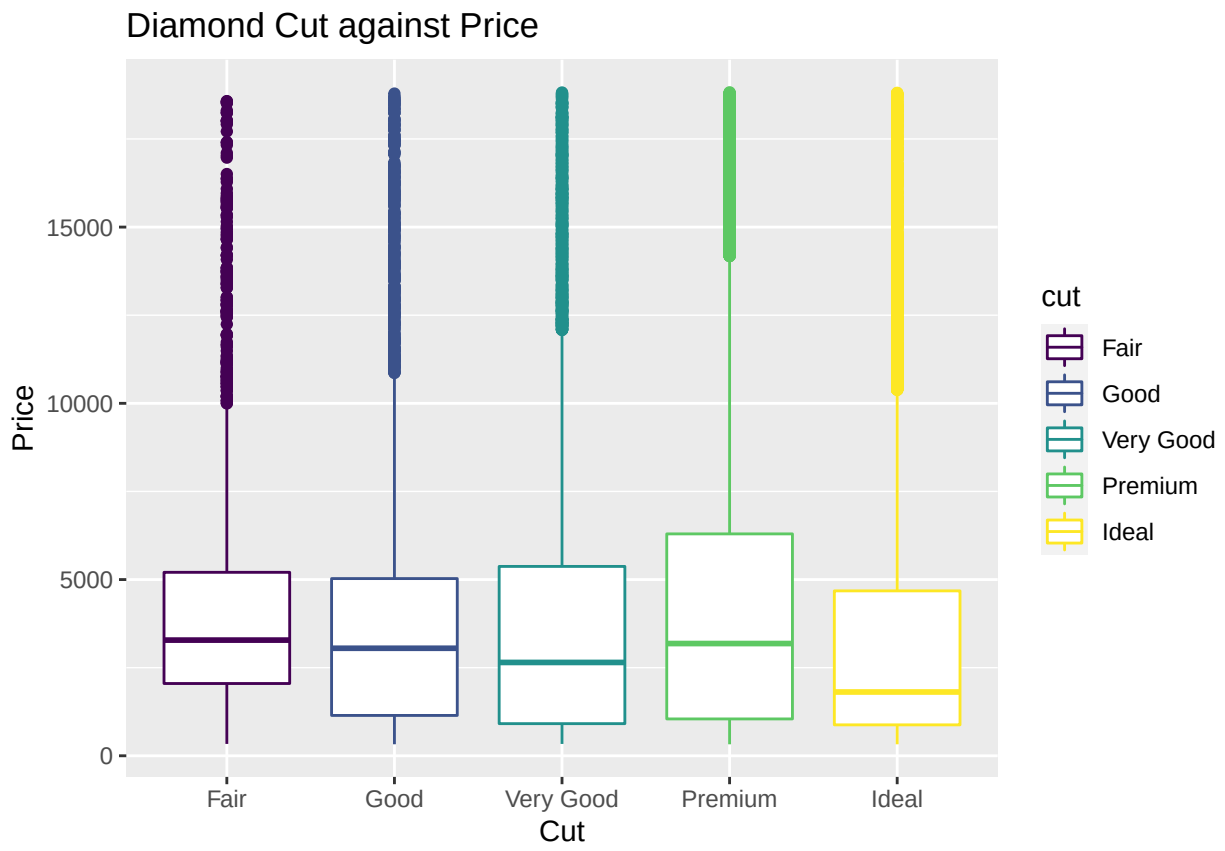
Line Graph

```
economics |>
  ggplot(aes(x = date, y = pop)) +
  geom_line()
```



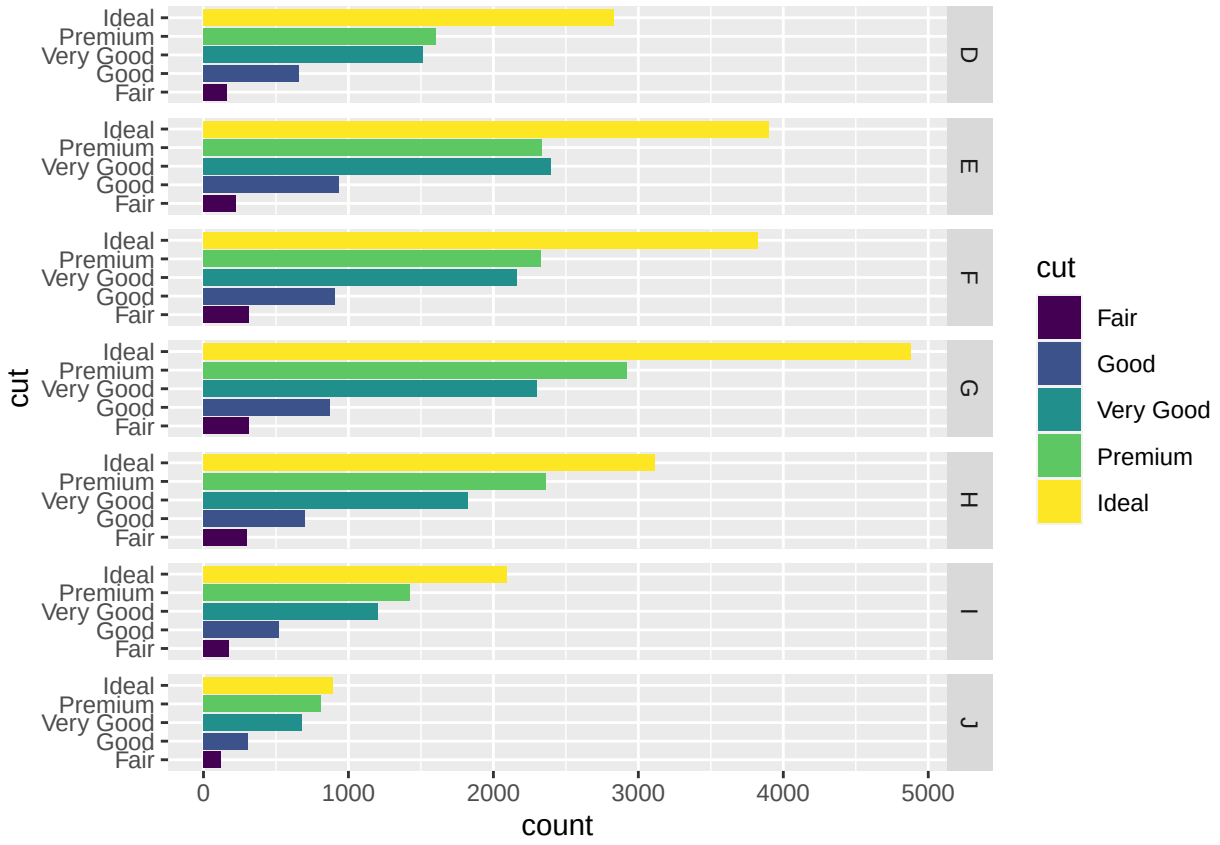
Labelling

```
diamonds |>
  ggplot(mapping = aes(x = cut, y = price, colour = cut)) +
  geom_boxplot() +
  labs(x = "Cut", y = "Price", title = "Diamond Cut against Price", )
```



Facetting

```
diamonds |>
  ggplot(mapping = aes(y = cut, fill = cut)) +
  geom_bar() +
  facet_grid(color ~ .)
```

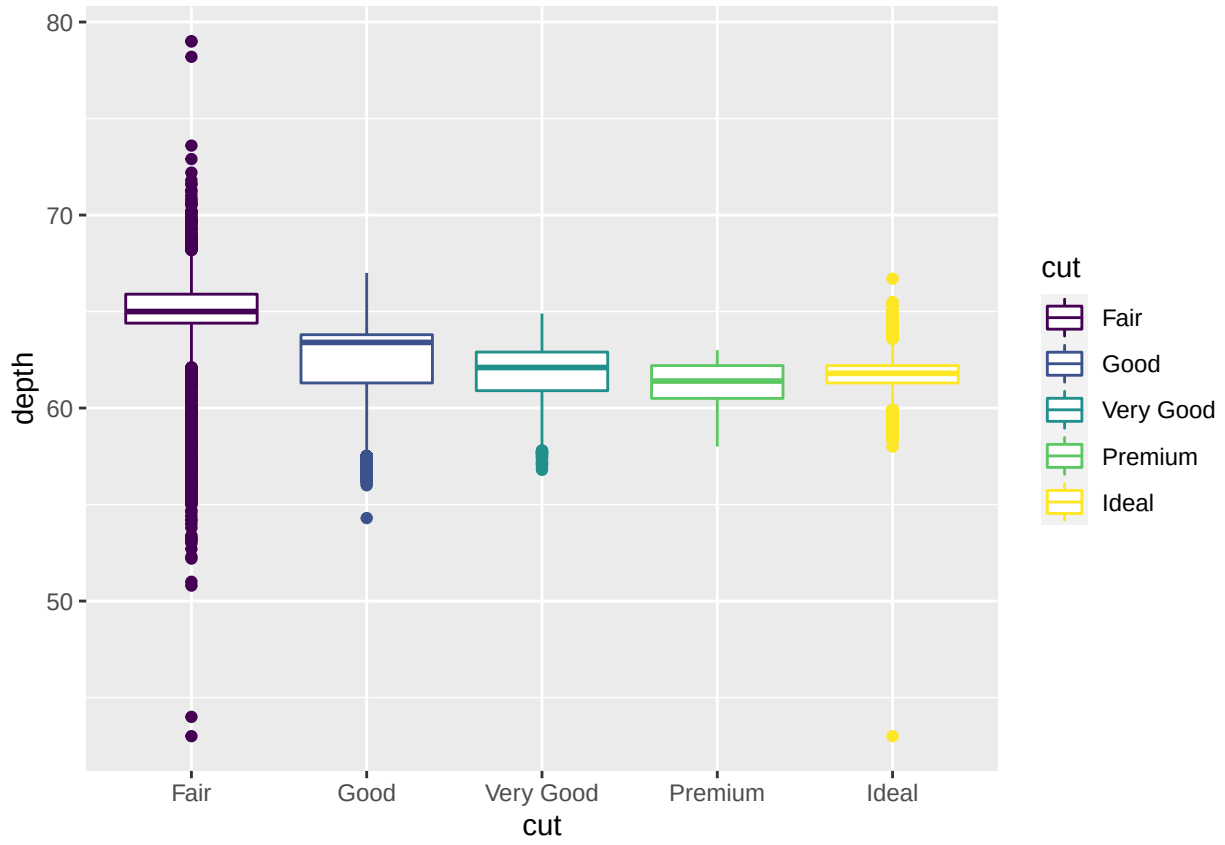


Exercises and Solutions

Question 1

Using a boxplot to visualize diamond cut against depth percentage. Which diamond cut have the highest median depth percentage. 1. Fair - ANSWER 2. Good 3. Very Good 4. Premium 5. Ideal

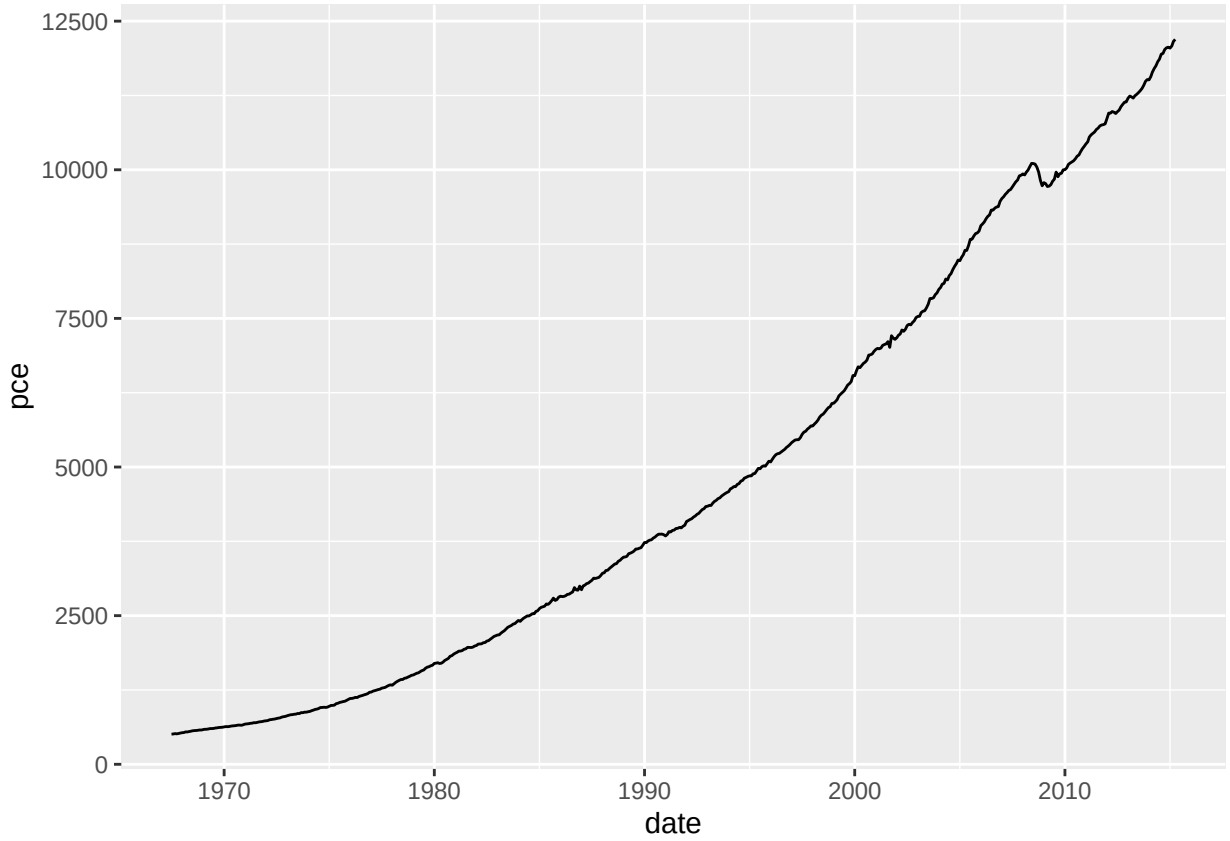
```
diamonds |>
  ggplot(mapping = aes(x = cut, y = depth, color = cut)) +
  geom_boxplot()
```



Question 2

Visualize using a line graph the change in United States Personal consumption expenditures in the economics data. In which year did the United States have the low personal consumption expenditure. 1. 1967 - ANSWER 2. 2000 3. 1999 4. 2015 5. 1990

```
economics |>
  ggplot(mapping = aes(x = date, y = pce)) +
  geom_line()
```

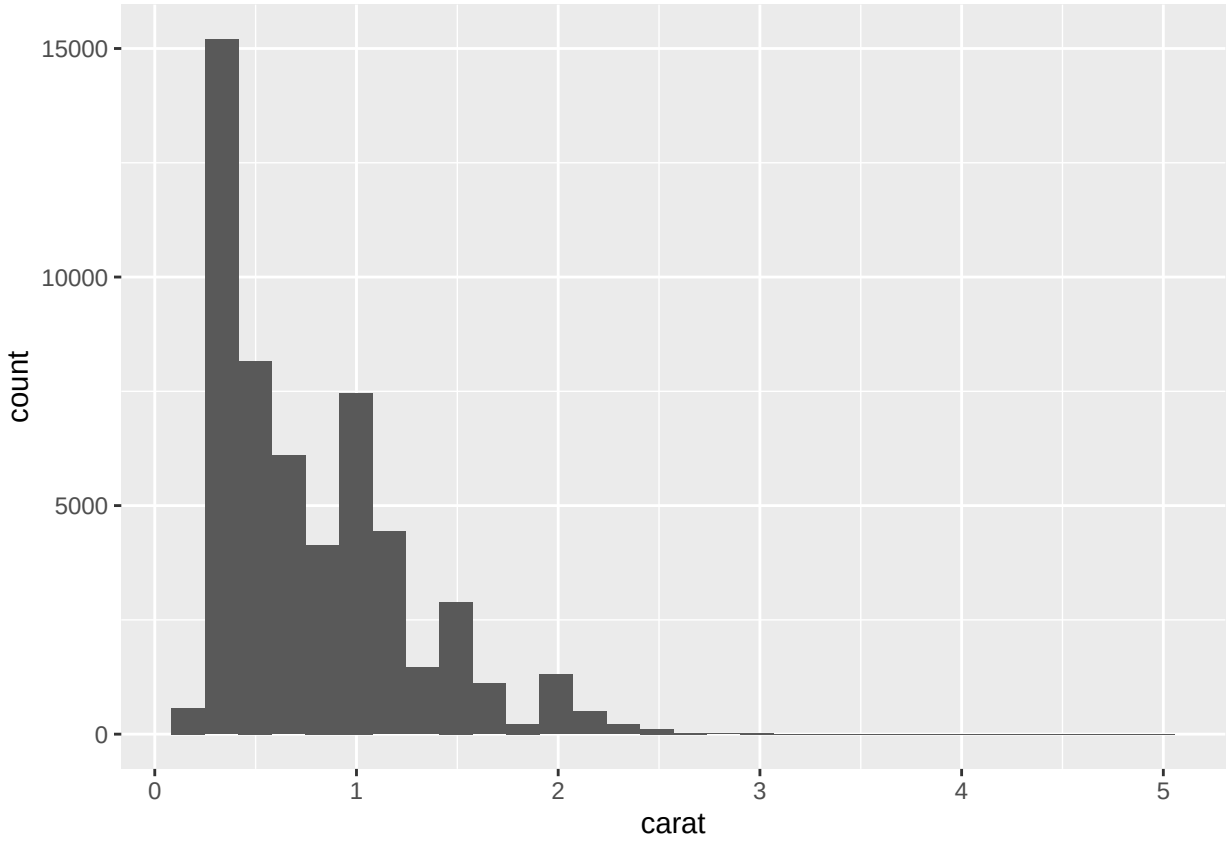


Question 3

Using a suitable bin width, visualize the distribution of diamond carat in the diamonds data set. In which range of carat value do most diamonds fall in. 1. 0.5 - 1.5 - ANSWER 2. 4.0 - 5.0 3. 2.0 - 3.5 4. 0.0 - 0.5

```
diamonds |>
  ggplot(mapping = aes(x = carat)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



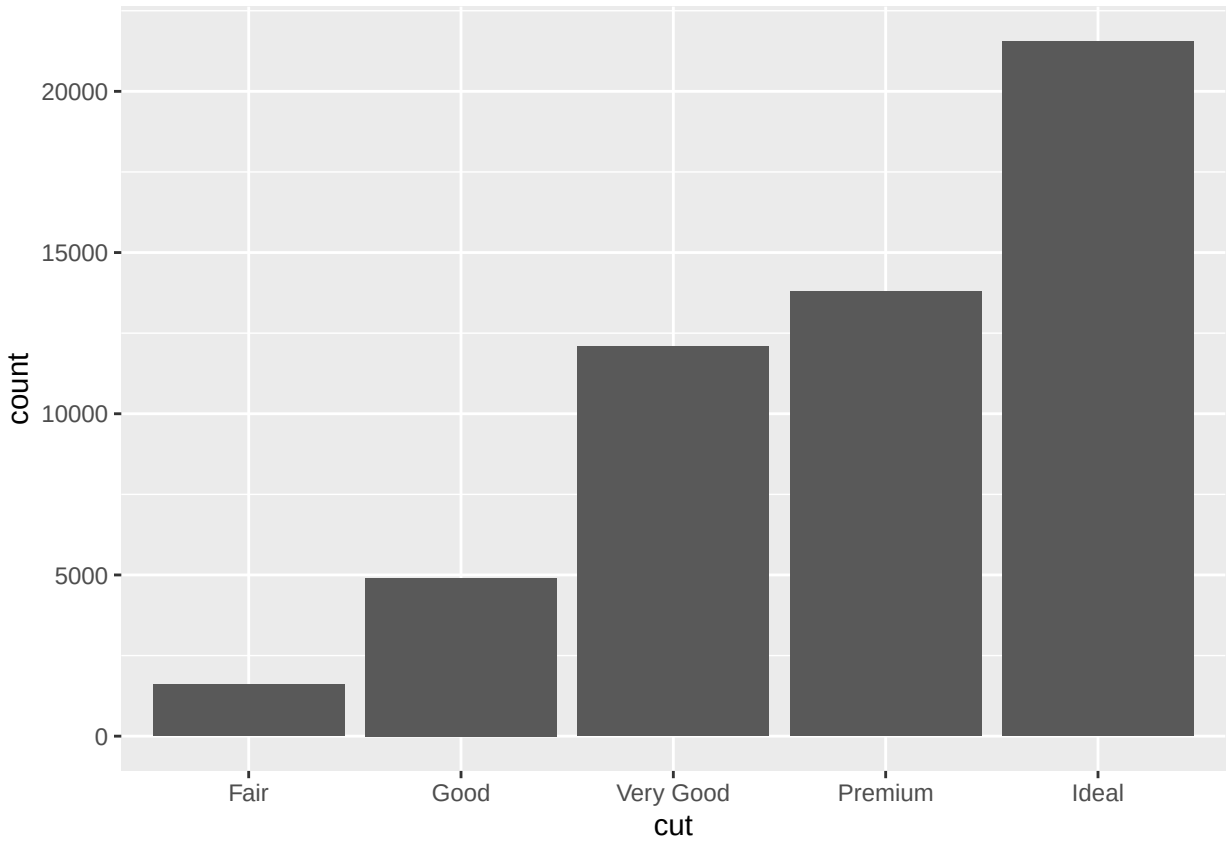
Question 4

Which of the following can we use to plot a numerical variable against a numerical variable 1. Box Plot 2. Scatter Plot - ANSWER 3. Bar Plot 4. Line Plot 5. None of the Above

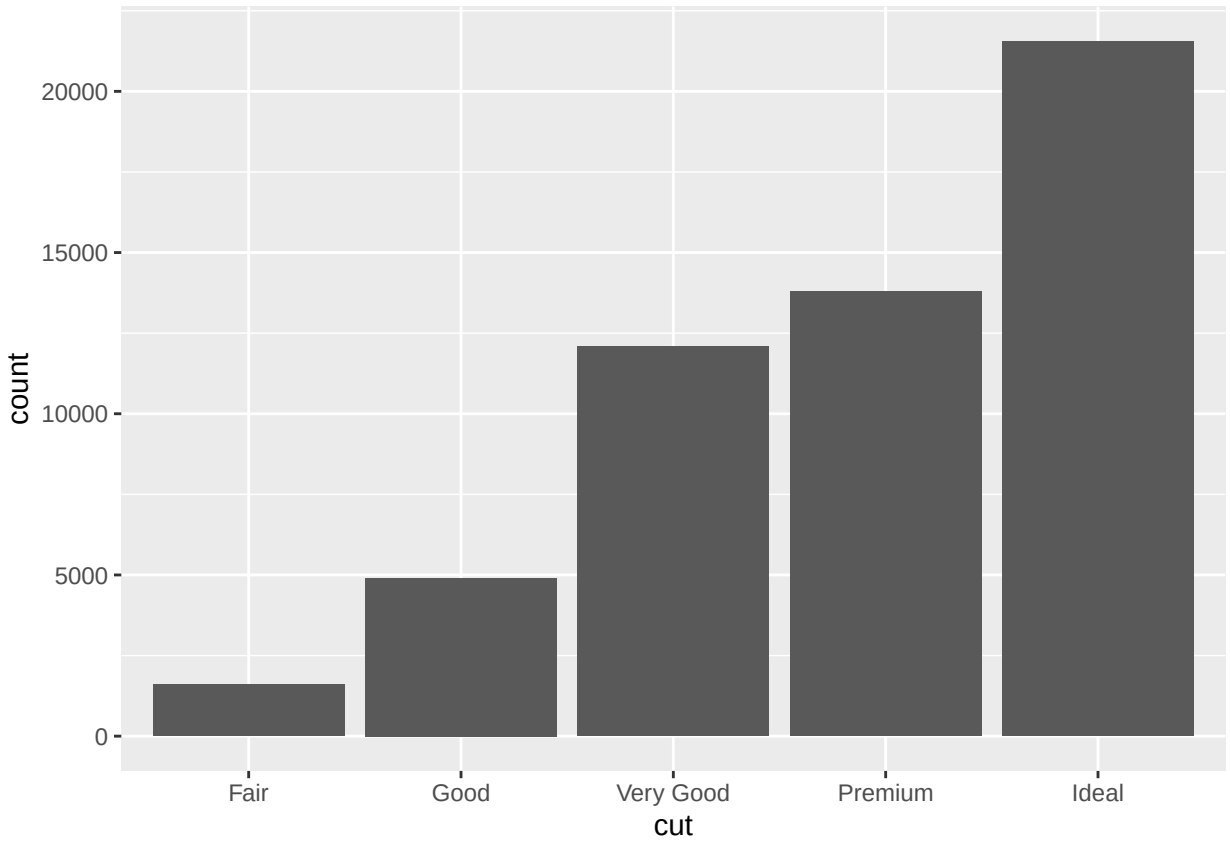
Question 5

Which of the following codes in the options, will give the graph below. SOLUTION - ALL THE GIVEN CODES WILL GIVE THE SAME GRAPH

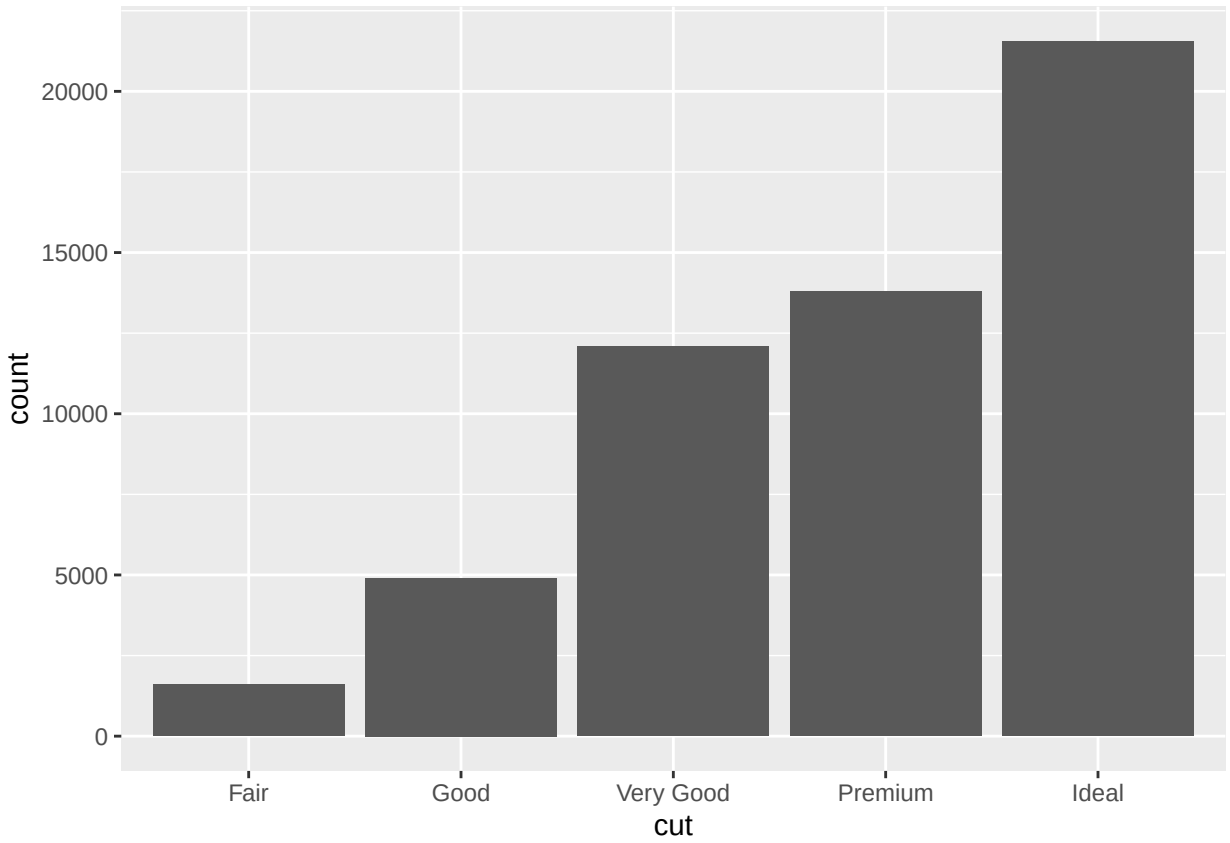
```
diamonds |>
  ggplot(mapping = aes(x = cut)) +
  geom_bar()
```



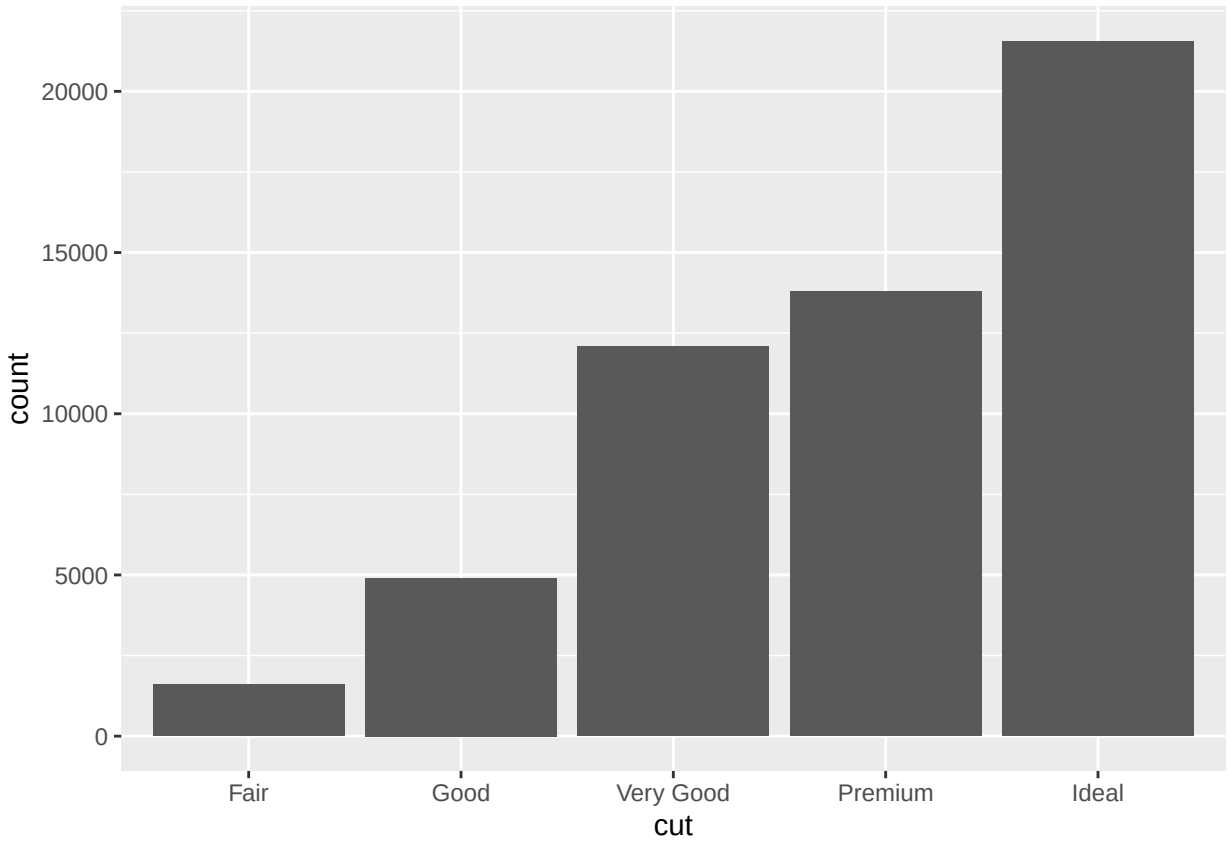

```
diamonds |>
  ggplot() +
  geom_bar(mapping = aes(x = cut))
```




```
ggplot(data = diamonds) +
  geom_bar(mapping = aes(x = cut))
```




```
ggplot(data = diamonds, mapping = aes(x = cut)) +
  geom_bar()
```


20000
15000
10000
5000
0
count
Fair
Good
Very Good
Premium
Ideal
cut