Data Visualization Bootcamp Homework

Nut Pakorn

2023-07-06

Homework

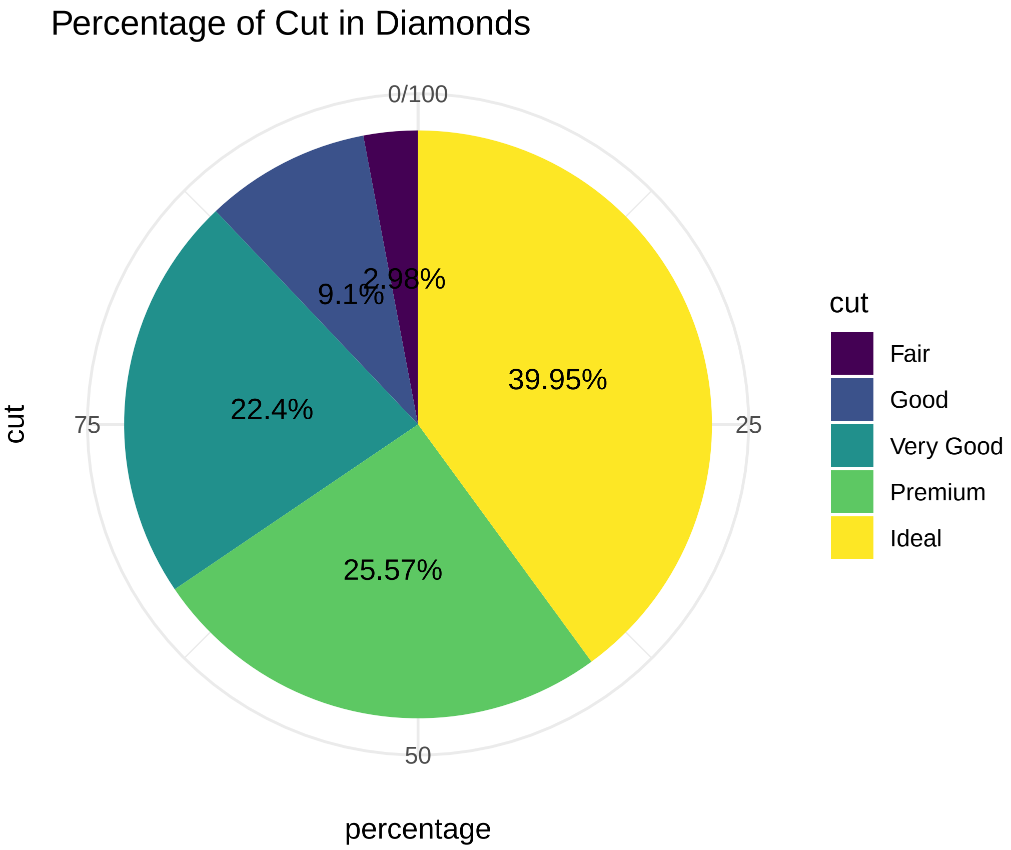
Use diamonds dataset to create 5 Charts

1.Percentage of Cut in Diamonds

```
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 error
```

```
HW1_Chart <- diamonds %>%
  group_by(cut) %>%
  summarise(percentage = n() / nrow(diamonds) * 100) %>%
  ggplot(aes("", percentage, fill = cut)) +
  geom_bar(stat = "identity") +
  coord_polar(theta = "y") +
  geom_text(aes(label = paste0(round(percentage, 2), "%")),
            position = position_stack(vjust = 0.5)) +
  labs(title = "Percentage of Cut in Diamonds",
       x = "cut") +
  theme_minimal()

HW1_Chart
```



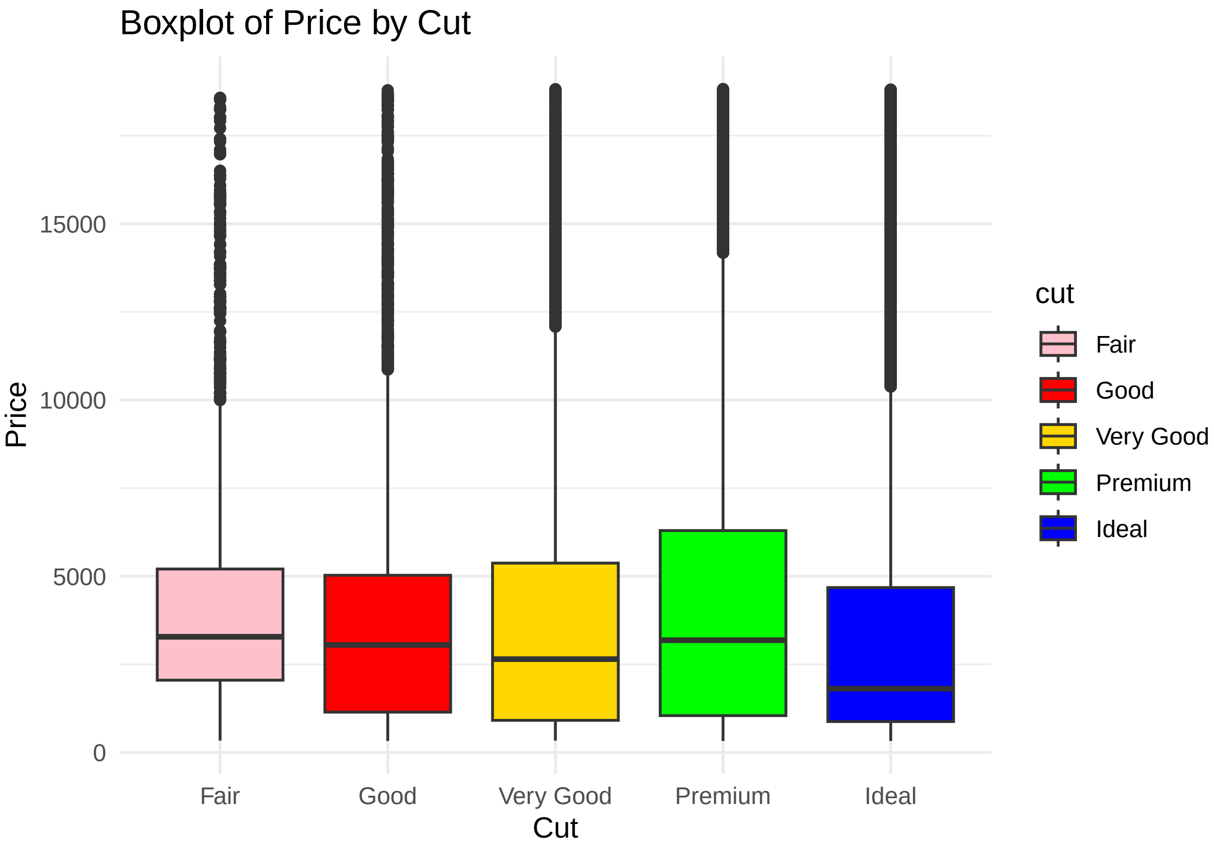
```
HW1_Ans <- diamonds %>%
  group_by(cut) %>%
  summarise(percentage = paste(round(n() / nrow(diamonds) * 100,2),"%"))

HW1_Ans
```

```
## # A tibble: 5 x 2
##   cut       percentage
##   <ord>     <chr>
## 1 Fair      2.98 %
## 2 Good      9.1 %
## 3 Very Good 22.4 %
## 4 Premium   25.57 %
## 5 Ideal     39.95 %
```

2.Average price of each Cut

```
HW2_Chart <- diamonds %>%
  ggplot(aes(x = cut, y = price, fill = cut)) +
  geom_boxplot() +
  labs(title = "Boxplot of Price by Cut",
       x = "Cut",
       y = "Price") +
  scale_fill_manual(values = c("pink","red","gold","green","blue","purple","light blue")) +
  theme_minimal()
```

```
HW2_Chart
```



```
HW2_Ans <- diamonds %>%
  group_by(cut) %>%
  summarise(Average_price = round(mean(price),2))

HW2_Ans
```

```
## # A tibble: 5 x 2
##   cut       Average_price
##   <ord>             <dbl>
## 1 Fair              4359.
## 2 Good              3929.
## 3 Very Good         3982.
## 4 Premium           4584.
## 5 Ideal             3458.
```

3.Max price of each clarity in cut

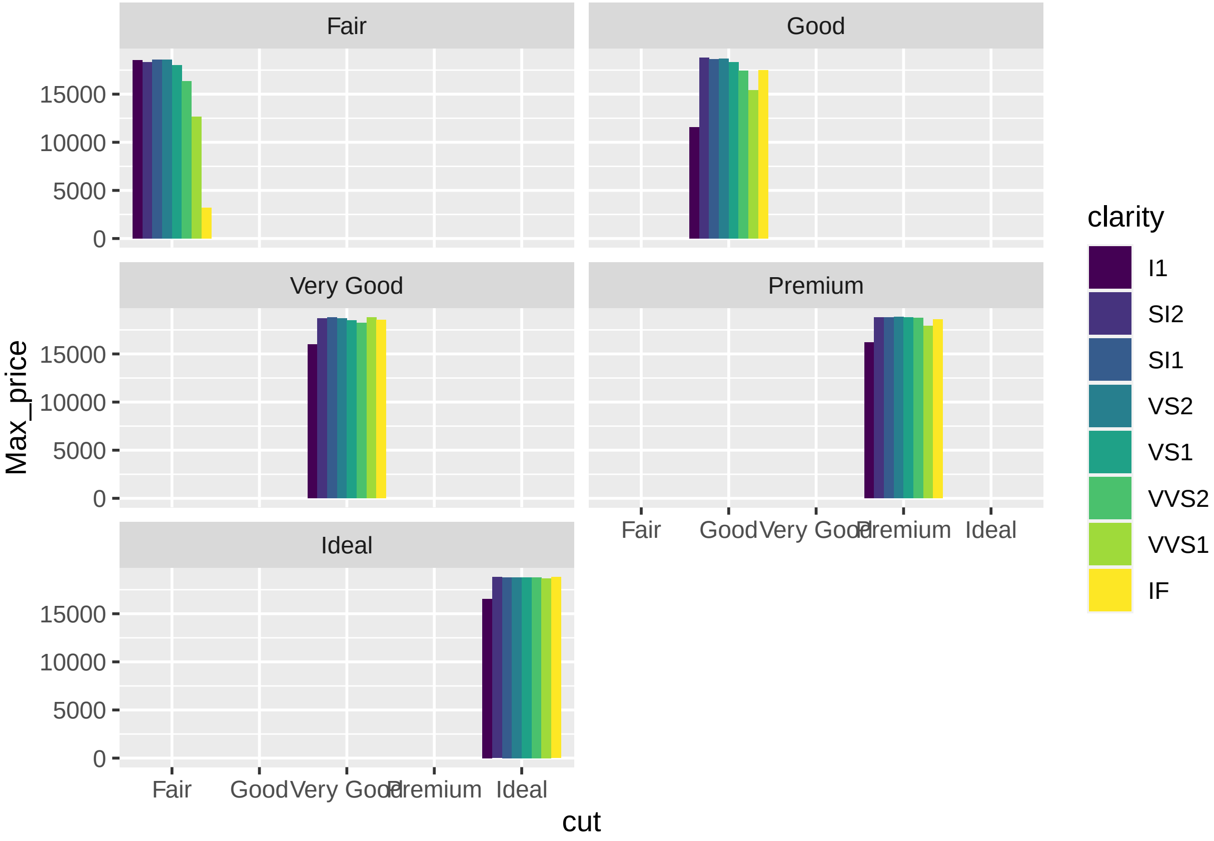
```
HW3_Chart <- diamonds %>%
  group_by(cut, clarity) %>%
  summarise(Max_price = max(price)) %>%
  ggplot(aes(cut, Max_price, fill = clarity)) +
  geom_col(position = "dodge") +
  facet_wrap(~cut , ncol = 2)
```

```
## `summarise()` has grouped output by 'cut'. You can override using the `.groups`
## argument.
```

```
labs(title = "Maximum Price by Clarity",
     x = "Clarity",
     y = "Maximum Price") +
  theme_minimal()
```

```
## NULL
```

```
HW3_Chart
```



```
HW3_Ans <- diamonds %>%
  group_by(cut,clarity) %>%
  summarise(Max_price = max(price))
```

```
## `summarise()` has grouped output by 'cut'. You can override using the `.groups`
## argument.
```

```
HW3_Ans
```

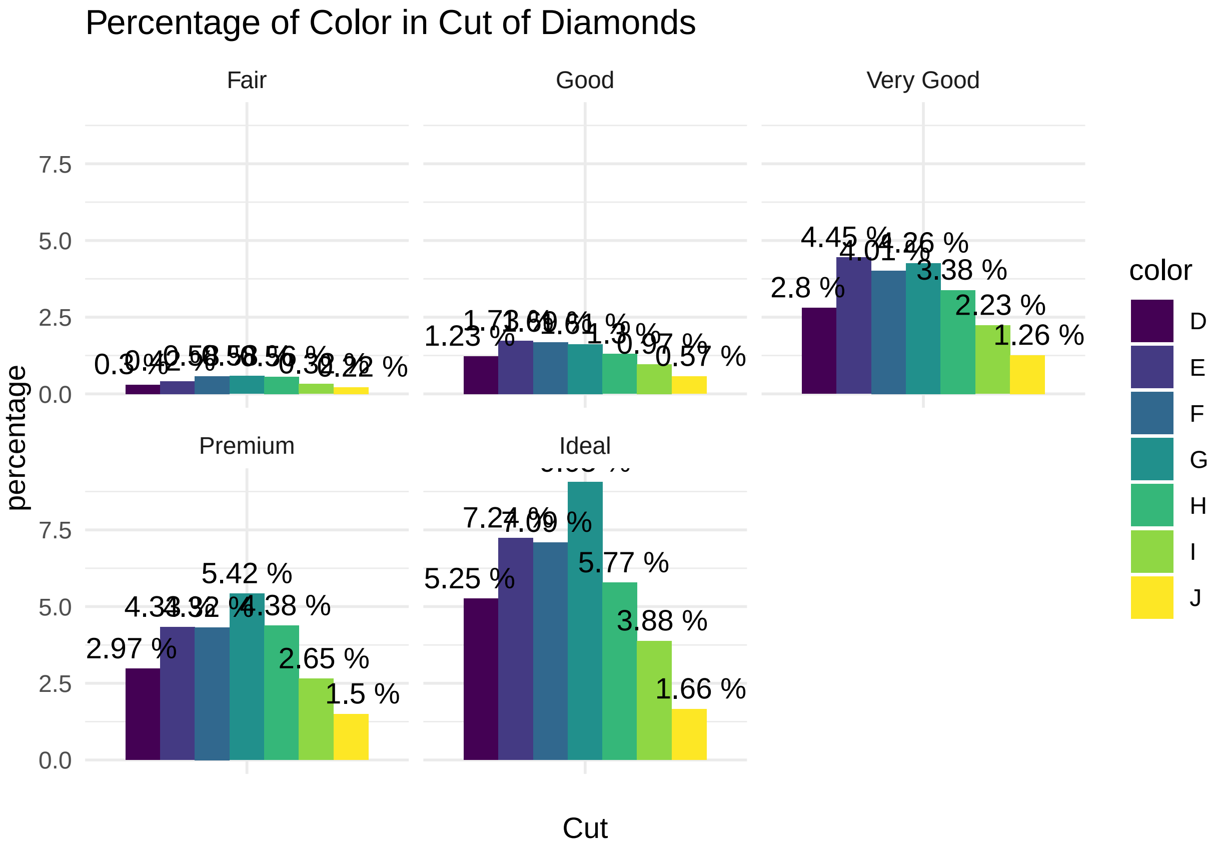
```
## # A tibble: 40 x 3
## # Groups:   cut [5]
##    cut   clarity Max_price
##    <ord> <ord>       <int>
##  1 Fair  I1          18531
##  2 Fair  SI2         18308
##  3 Fair  SI1         18574
##  4 Fair  VS2         18565
##  5 Fair  VS1         17995
##  6 Fair  VVS2        16364
##   7 Fair  VVS1        12648
##   8 Fair  IF           3205
##   9 Good  I1          11548
## 10 Good  SI2         18788
## # i 30 more rows
```

4.Percentage of Color in Cut of Diamonds

```
HW4_Chart <- diamonds %>%
  group_by(cut,color) %>%
  summarise(percentage = n() / nrow(diamonds) * 100) %>%
  ggplot(aes("", percentage , fill = color)) +
  geom_bar(position = "dodge",stat = "identity") +
  facet_wrap(~cut, ncol = 3 ) +
  geom_text(aes(label = paste(round(percentage,2),"%")),
            position = position_dodge(width = 1),vjust = -0.5) +
  labs(title = "Percentage of Color in Cut of Diamonds",
       x = "Cut") +
  theme_minimal()
```

```
## `summarise()` has grouped output by 'cut'. You can override using the `.groups`
## argument.
```

```
HW4_Chart
```



```
HW4 <- diamonds %>%
  group_by(cut,color) %>%
  summarise(percentage = paste(round(n() / nrow(diamonds) * 100,2),"%"))
```

```
## `summarise()` has grouped output by 'cut'. You can override using the `.groups`
## argument.
```

```
total_percentage <- HW4 %>%
  summarise(total_percentage =
            paste(sum(as.numeric(sub("%","",percentage))),"%"))

HW4_Ans <- list(HW4,total_percentage)

HW4_Ans
```

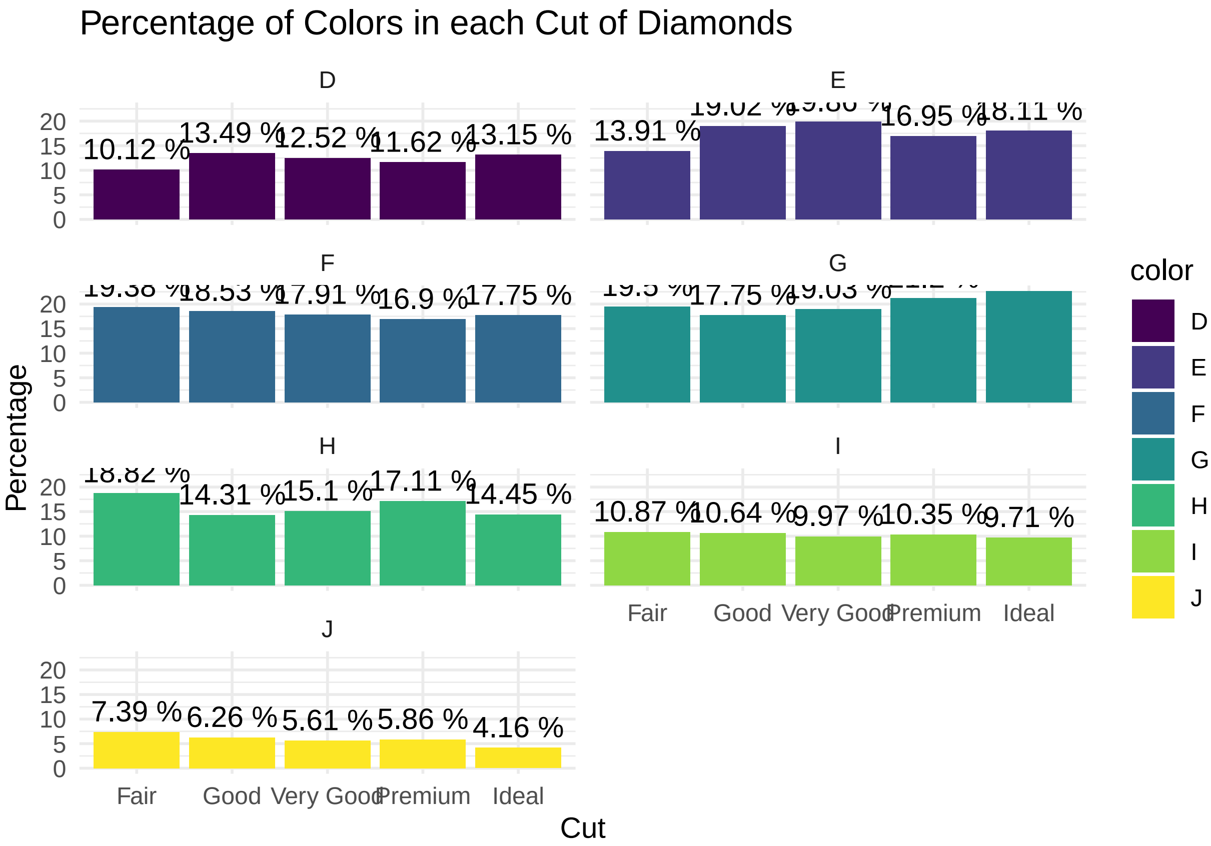
```
## [[1]]
## # A tibble: 35 x 3
## # Groups:   cut [5]
##    cut   color percentage
##    <ord> <ord> <chr>
##  1 Fair  D     0.3 %
##  2 Fair  E     0.42 %
##  3 Fair  F     0.58 %
##  4 Fair  G     0.58 %
##  5 Fair  H     0.56 %
##  6 Fair  I     0.32 %
##  7 Fair  J     0.22 %
##  8 Good  D     1.23 %
##  9 Good  E     1.73 %
## 10 Good  F     1.69 %
## # i 25 more rows
##
## [[2]]
## # A tibble: 5 x 2
##   cut       total_percentage
##   <ord>     <chr>
## 1 Fair      2.98 %
## 2 Good      9.1 %
## 3 Very Good 22.39 %
## 4 Premium   25.57 %
## 5 Ideal     39.94 %
```

5.Percentage of Colors in each Cut of Diamonds

```
HW5_Chart <- diamonds %>%
  group_by(cut,color) %>%
  summarise(count = n()) %>%
  group_by(cut) %>%
  mutate(percentage = round(count / sum(count) * 100,2 )) %>%
  ggplot(aes(cut , percentage , fill = color)) +
  geom_col(position = "dodge") +
  geom_text(aes(label = paste(round(percentage,2),"%")) ,
            position = position_dodge(width = 1) , vjust = -0.5) +
  facet_wrap(~color , ncol = 2) +
  labs(title = "Percentage of Colors in each Cut of Diamonds",
       x = "Cut",
       y = "Percentage") +
  theme_minimal()
```

```
## `summarise()` has grouped output by 'cut'. You can override using the `.groups`
## argument.
```

```
HW5_Chart
```



```
HW5_Ans <- diamonds %>%
  group_by(cut,color) %>%
  summarise(count = n()) %>%
  mutate(percentage = paste(round(count / sum(count) * 100,2),"%" )) %>%
  print(n = 35)
```

```
## `summarise()` has grouped output by 'cut'. You can override using the `.groups`
## argument.
```

```
## # A tibble: 35 x 4
## # Groups:   cut [5]
##    cut       color count percentage
##    <ord>     <ord> <int> <chr>
##  1 Fair      D       163 10.12 %
##  2 Fair      E       224 13.91 %
##  3 Fair      F       312 19.38 %
##  4 Fair      G       314 19.5 %
##  5 Fair      H       303 18.82 %
##  6 Fair      I       175 10.87 %
##  7 Fair      J       119 7.39 %
##  8 Good      D       662 13.49 %
##  9 Good      E       933 19.02 %
## 10 Good      F       909 18.53 %
## 11 Good      G       871 17.75 %
## 12 Good      H       702 14.31 %
## 13 Good      I       522 10.64 %
## 14 Good      J       307 6.26 %
## 15 Very Good D      1513 12.52 %
## 16 Very Good E      2400 19.86 %
## 17 Very Good F      2164 17.91 %
## 18 Very Good G      2299 19.03 %
## 19 Very Good H      1824 15.1 %
## 20 Very Good I      1204 9.97 %
## 21 Very Good J       678 5.61 %
## 22 Premium   D      1603 11.62 %
## 23 Premium   E      2337 16.95 %
## 24 Premium   F      2331 16.9 %
## 25 Premium   G      2924 21.2 %
## 26 Premium   H      2360 17.11 %
## 27 Premium   I      1428 10.35 %
## 28 Premium   J       808 5.86 %
## 29 Ideal     D      2834 13.15 %
## 30 Ideal     E      3903 18.11 %
## 31 Ideal     F      3826 17.75 %
## 32 Ideal     G      4884 22.66 %
## 33 Ideal     H      3115 14.45 %
## 34 Ideal     I      2093 9.71 %
## 35 Ideal     J       896 4.16 %
```

```
HW5_Ans
```

```
## # A tibble: 35 x 4
## # Groups:   cut [5]
##    cut   color count percentage
##    <ord> <ord> <int> <chr>
##  1 Fair  D       163 10.12 %
##  2 Fair  E       224 13.91 %
##  3 Fair  F       312 19.38 %
##  4 Fair  G       314 19.5 %
##  5 Fair  H       303 18.82 %
##  6 Fair  I       175 10.87 %
##  7 Fair  J       119 7.39 %
##  8 Good  D       662 13.49 %
##  9 Good  E       933 19.02 %
## 10 Good  F       909 18.53 %
## # i 25 more rows
```