

DTPrac3

secary

2024-06-05

R Markdown

```
library(tidyverse)
library(inspectdf)
```

```
data(diamonds)
diamond_df <- diamonds
diamonds
```

```
## # A tibble: 53,940 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
## 7  0.24 Very Good I      VVS1    62.3    57    336  3.95  3.98  2.47
## 8  0.26 Very Good H      SI1     61.9    55    337  4.07  4.11  2.53
## 9  0.22 Fair    E      VS2     65.1    61    337  3.87  3.78  2.49
## 10 0.23 Very Good H      VS1     59.4    61    338  4     4.05  2.39
## # i 53,930 more rows
```

```
inspect_na(diamonds)
```

```
## # A tibble: 10 x 3
##   col_name  cnt pcnt
##   <chr>    <int> <dbl>
## 1 carat      0    0
## 2 cut        0    0
## 3 color      0    0
## 4 clarity    0    0
## 5 depth      0    0
## 6 table      0    0
## 7 price      0    0
## 8 x          0    0
## 9 y          0    0
## 10 z         0    0
```

```
inspect_num(diamonds)
```

```
## # A tibble: 7 x 10
##   col_name   min    q1  median    mean    q3    max    sd pcnt_na hist
##   <chr>    <dbl> <dbl>  <dbl>   <dbl> <dbl> <dbl>  <dbl> <dbl> <named >
## 1 carat     0.2   0.4    0.7    0.798   1.04 5.01e0 4.74e-1    0 <tibble>
## 2 depth     43    61   61.8   61.7    62.5 7.9 e1 1.43e+0    0 <tibble>
## 3 table     43    56    57    57.5    59   9.5 e1 2.23e+0    0 <tibble>
## 4 price    326   950  2401  3933.   5324. 1.88e4 3.99e+3    0 <tibble>
## 5 x         0    4.71   5.7    5.73    6.54 1.07e1 1.12e+0    0 <tibble>
## 6 y         0    4.72   5.71   5.73    6.54 5.89e1 1.14e+0    0 <tibble>
## 7 z         0    2.91   3.53   3.54    4.04 3.18e1 7.06e-1    0 <tibble>
```

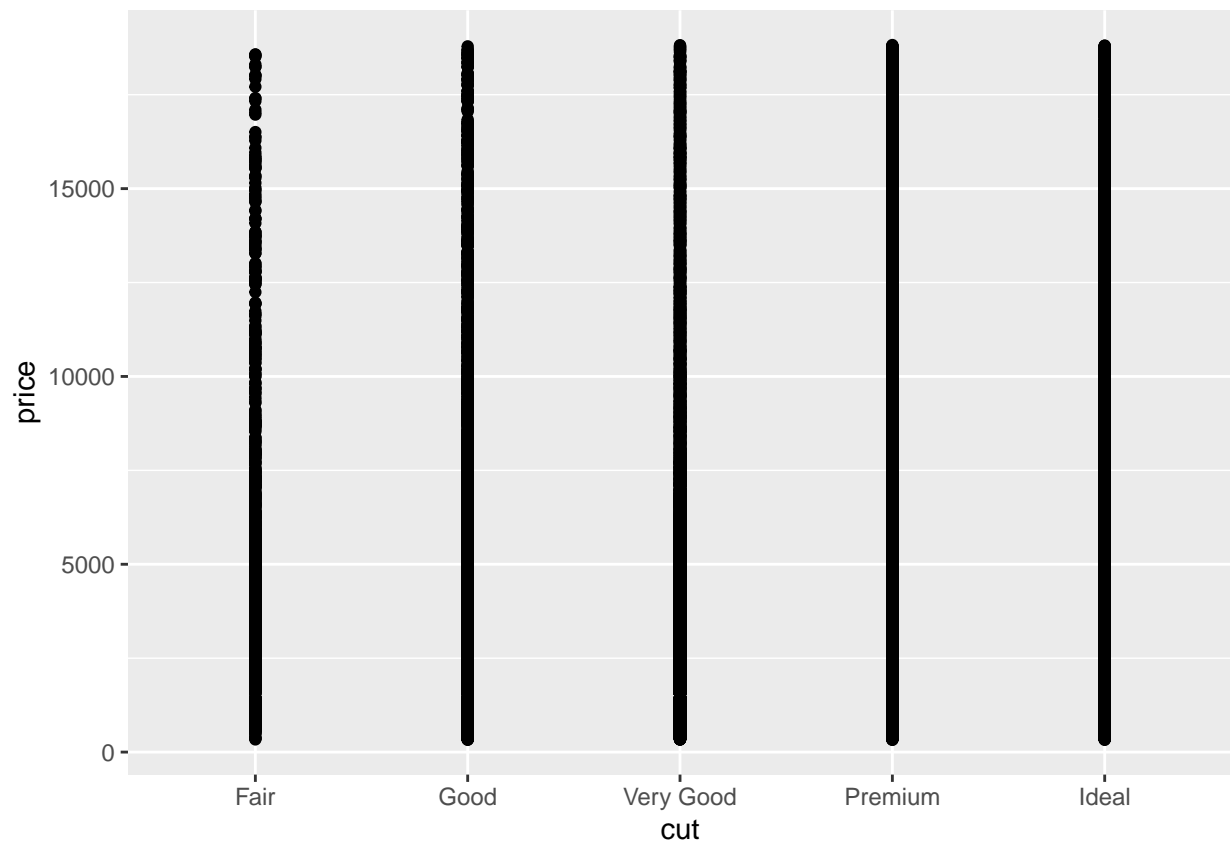
```
unique(diamonds$cut)
```

```
## [1] Ideal      Premium    Good       Very Good Fair
## Levels: Fair < Good < Very Good < Premium < Ideal
```

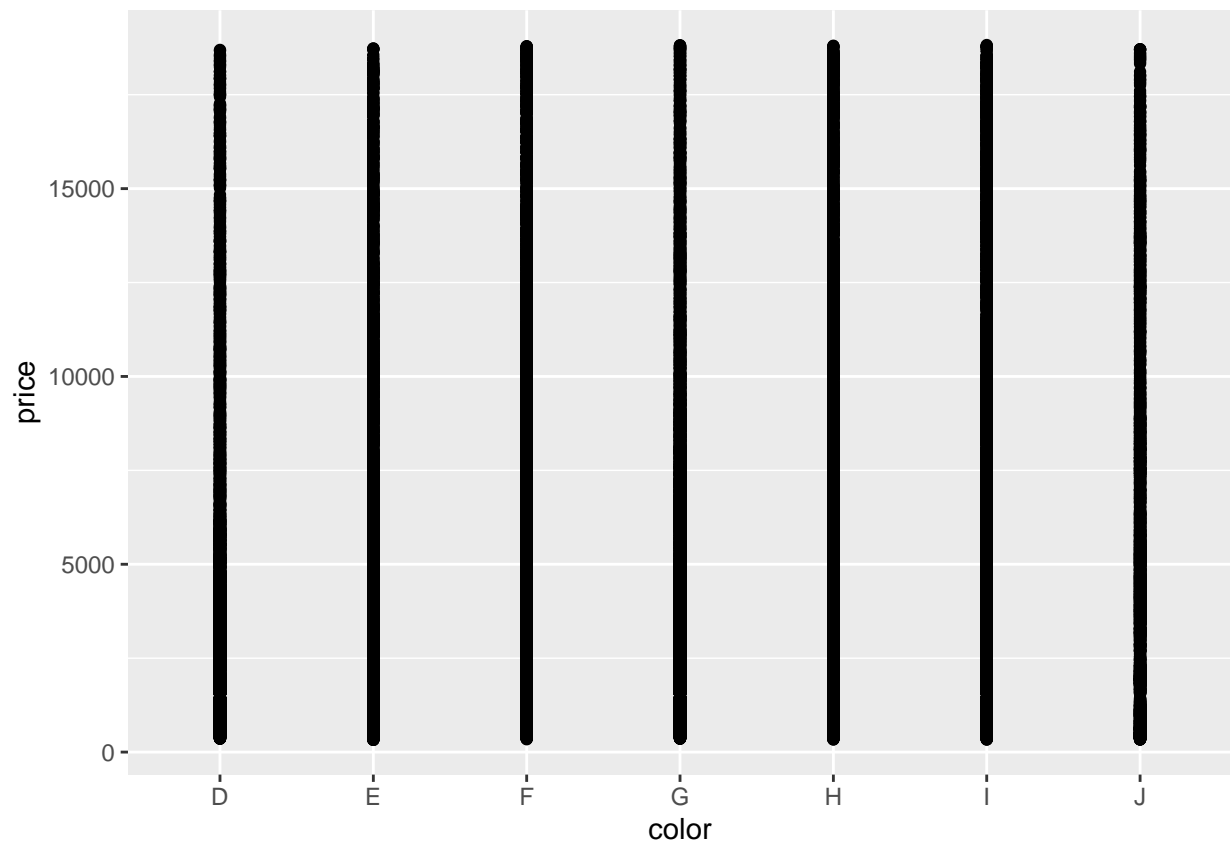
```
count(diamonds,cut)
```

```
## # A tibble: 5 x 2
##   cut      n
##   <ord>  <int>
## 1 Fair    1610
## 2 Good    4906
## 3 Very Good 12082
## 4 Premium 13791
## 5 Ideal   21551
```

```
ggplot(diamonds,aes(x=cut,y=price))+
  geom_point()
```



```
ggplot(diamonds,aes(x=color,y=price))+  
  geom_point()
```



```
library(ggplot2)
ggplot(diamonds, aes(y=price, fill=cut)) +
  geom_boxplot()
```

