lab04

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(knitr)
library(broom)
```

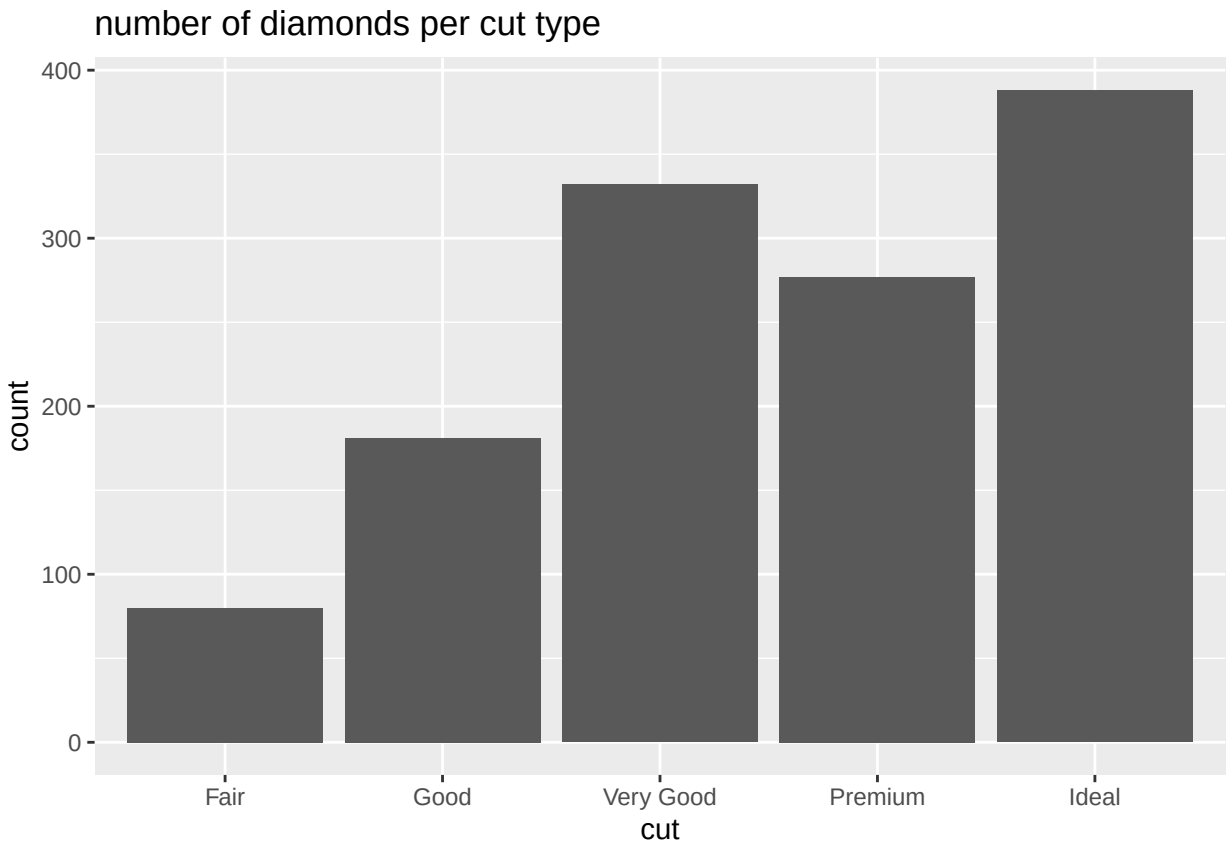
```
data <- diamonds
glimpse(diamonds)
```

```
## Rows: 53,940
## Columns: 10
## $ carat   <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, 0.23, 0.~
## $ cut     <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very Good, Ver~
## $ color   <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, J, J, I,~
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, VS1, SI1, VS1, ~
## $ depth   <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, 65.1, 59.4, 64~
## $ table   <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, 61, 54, 62, 58~
## $ price   <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338, 339, 340, 34~
## $ x       <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, 3.87, 4.00, 4.~
## $ y       <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, 3.78, 4.05, 4.~
## $ z       <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, 2.49, 2.39, 2.~
```

```
d_sub <- subset(data,  carat == .5)
```

There are 1258 observations in the dataset where carats are .5

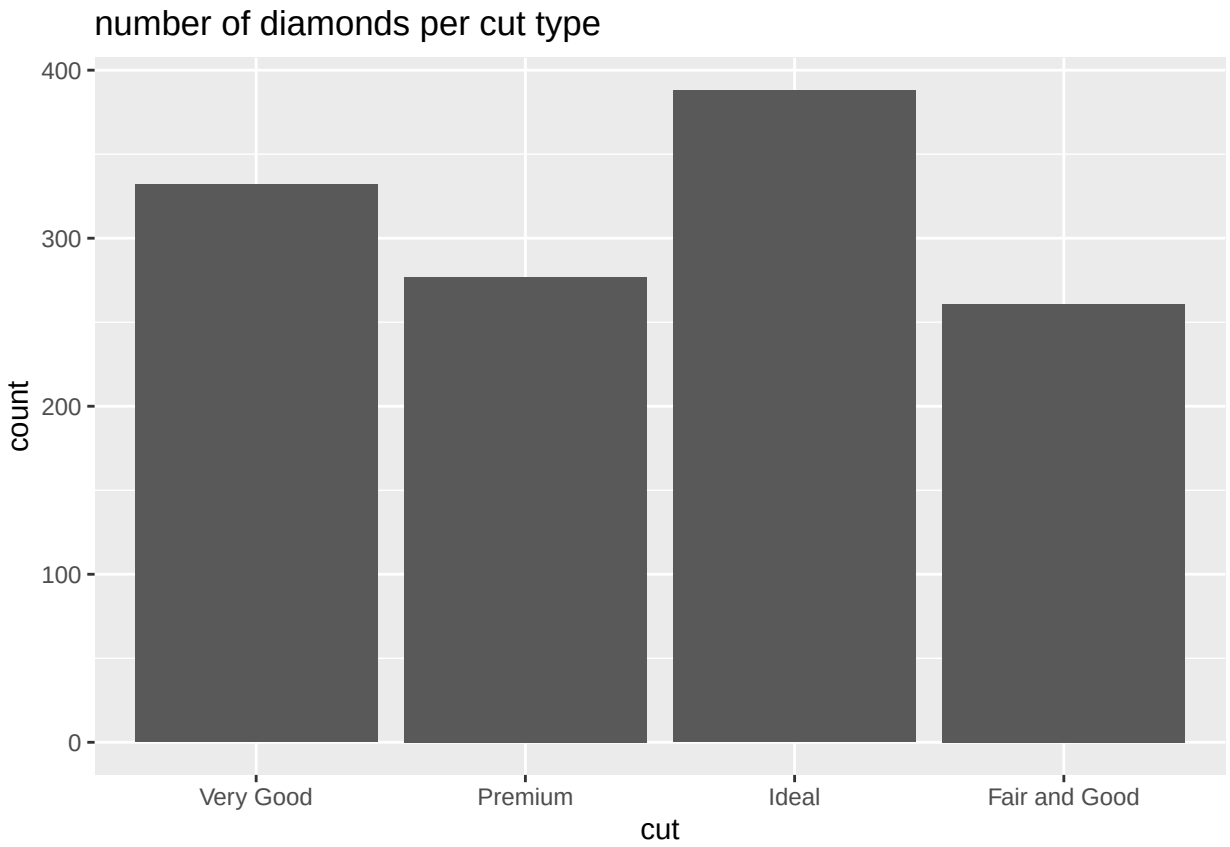
```
ggplot(data = d_sub, aes(x=cut)) +
         geom_bar() + labs(title = "number of diamonds per cut type")
```



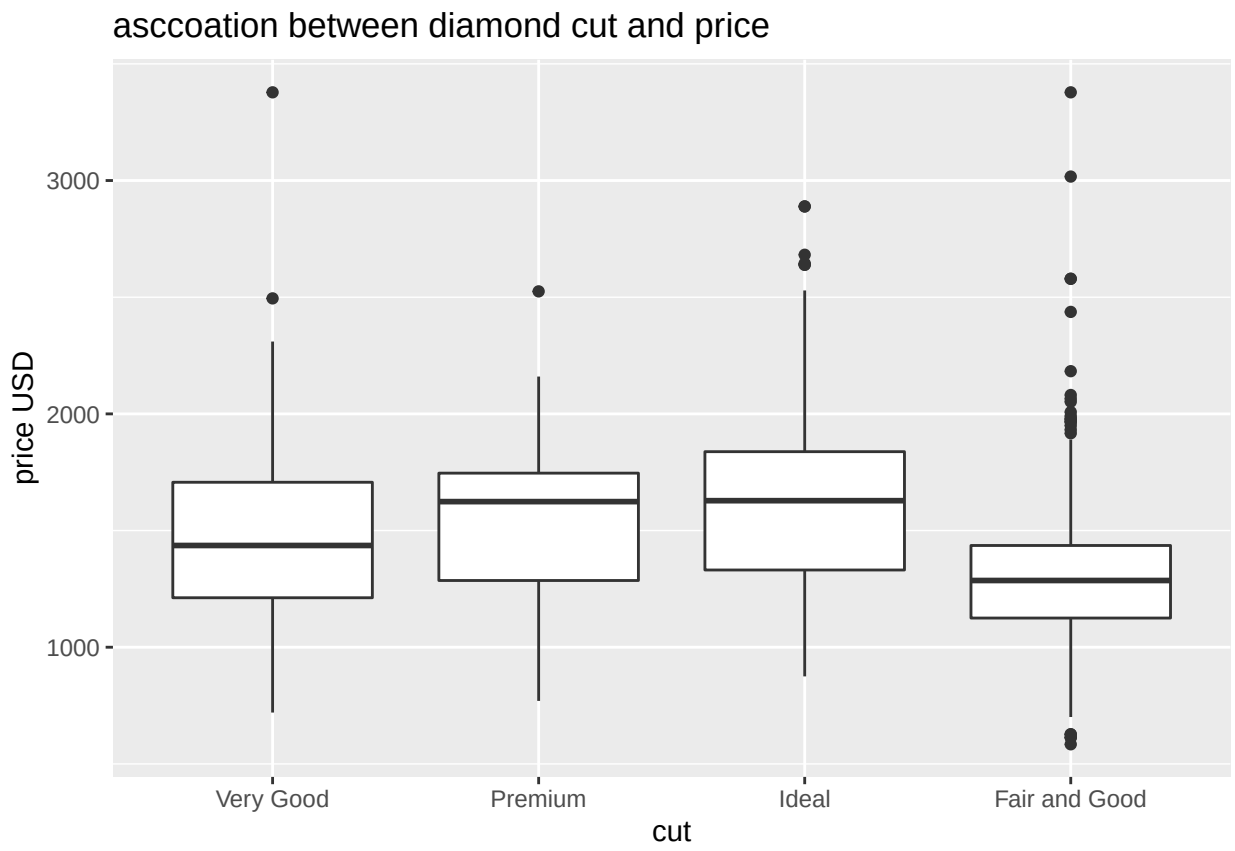
Cuts of fair and good have the fewest observations.

```
combine <- d_sub %>%
  mutate(cut = fct_lump(cut, n = 3, other_level = "Fair and Good"))
```

```
ggplot(data = combine, aes(x=cut)) +
         geom_bar() + labs(title = "number of diamonds per cut type")
```




```
ggplot(data = combine, aes(x=cut, y = price)) + geom_boxplot() + labs(title = "asccoation between diamo
```




```
vg <- combine %>% filter(cut == "Very Good")
p <- combine %>% filter(cut == "Premium")
i <- combine %>% filter(cut == "Ideal")
fg <- combine %>% filter(cut == "Fair and Good")
```

Summary Stats:

Very Good

```
mean(vg$price)
```

```
## [1] 1488.663
```

```
sd(vg$price)
```

```
## [1] 339.363
```

```
nrow(vg)
```

```
## [1] 332
```

Premium

```
mean(p$price)
```

```
## [1] 1531.776
```

```
sd(p$price)
```

```
## [1] 304.1443
```

```
nrow(p)
```

```
## [1] 277
```

Ideal

```
mean(i$price)
```

```
## [1] 1608.668
```

```
sd(i$price)
```

```
## [1] 368.3448
```

```
nrow(i)
```

```
## [1] 388
```

Fair

```
mean(fg$price)
```

```
## [1] 1340.644
```

```
sd(fg$price)
```

```
## [1] 364.5216
```

```
nrow(fg)
```

```
## [1] 261
```