

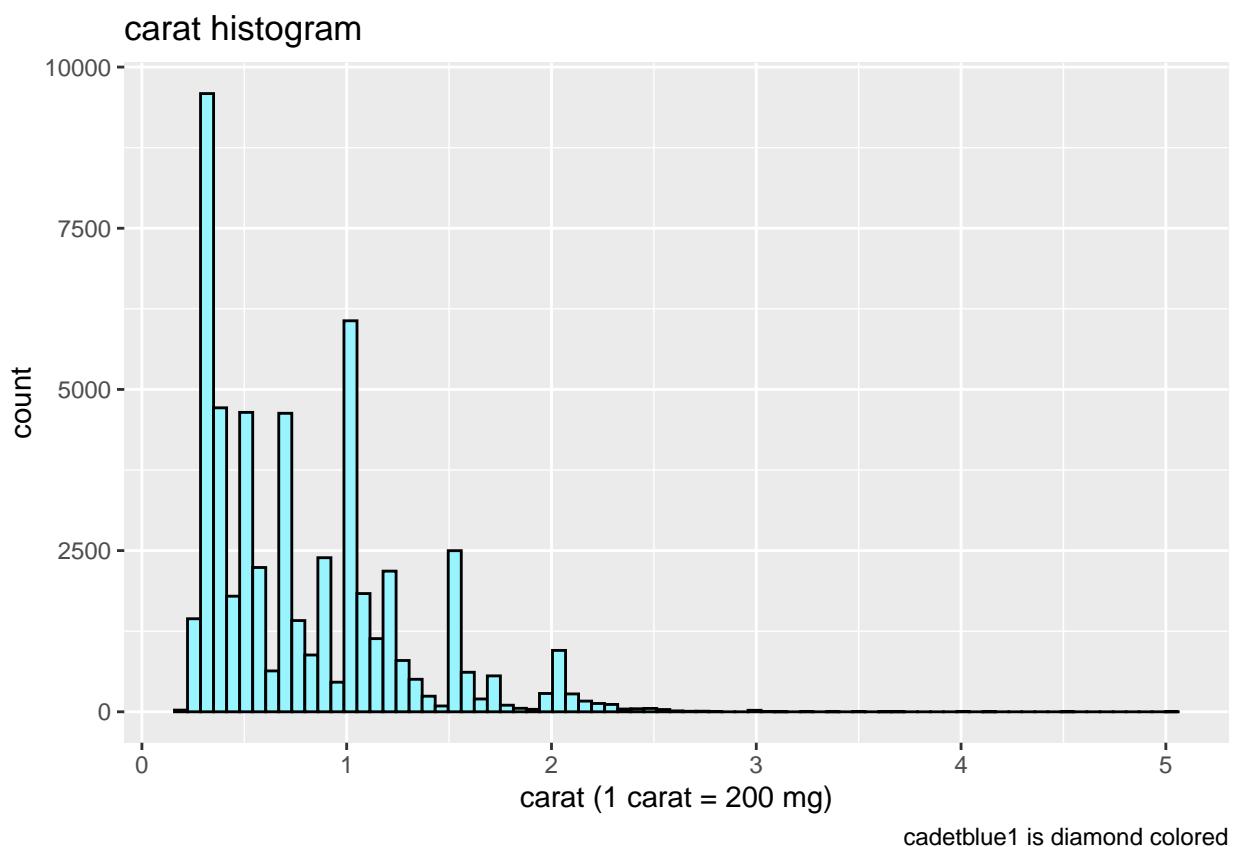
# Homework 5

Luke Bennett

10/19/2021

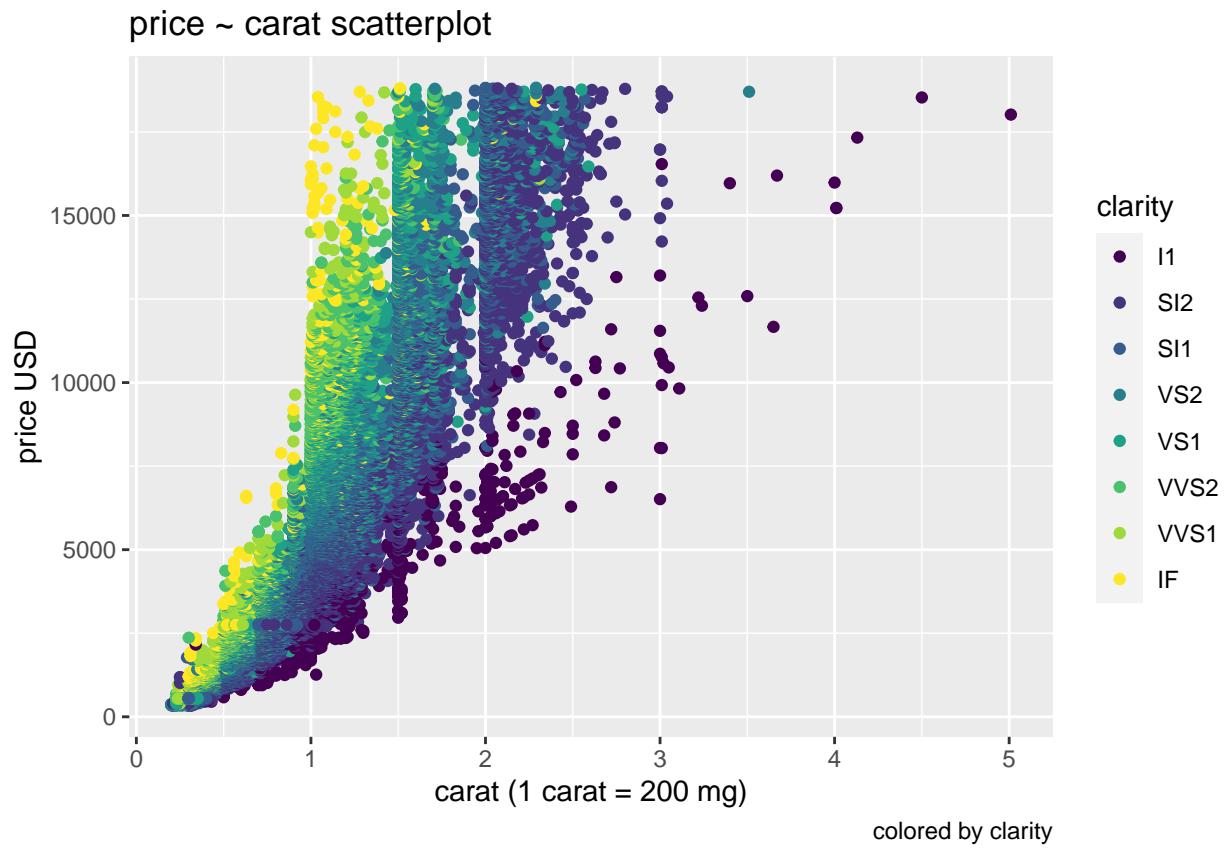
## Problem 1 a)

```
library(ggplot2)
data("diamonds")
x <- diamonds$carat
n <- length(x)
k <- 2 * (n)^(1/3)
h <- (max(x) - min(x)) / k
ggplot(data = diamonds) +
  geom_histogram(binwidth = h, mapping = aes(carat),
                 colour = "black", fill = "cadetblue1") +
  labs(title = "carat histogram", x = "carat (1 carat = 200 mg)",
       caption = "cadetblue1 is diamond colored")
```



```
# Problem 1 b)
```

```
ggplot(diamonds) +  
  geom_point(mapping = aes(x = carat, y = price, color = clarity)) +  
  labs(title = "price ~ carat scatterplot", y = "price USD",  
       x = "carat (1 carat = 200 mg)", caption = "colored by clarity")
```

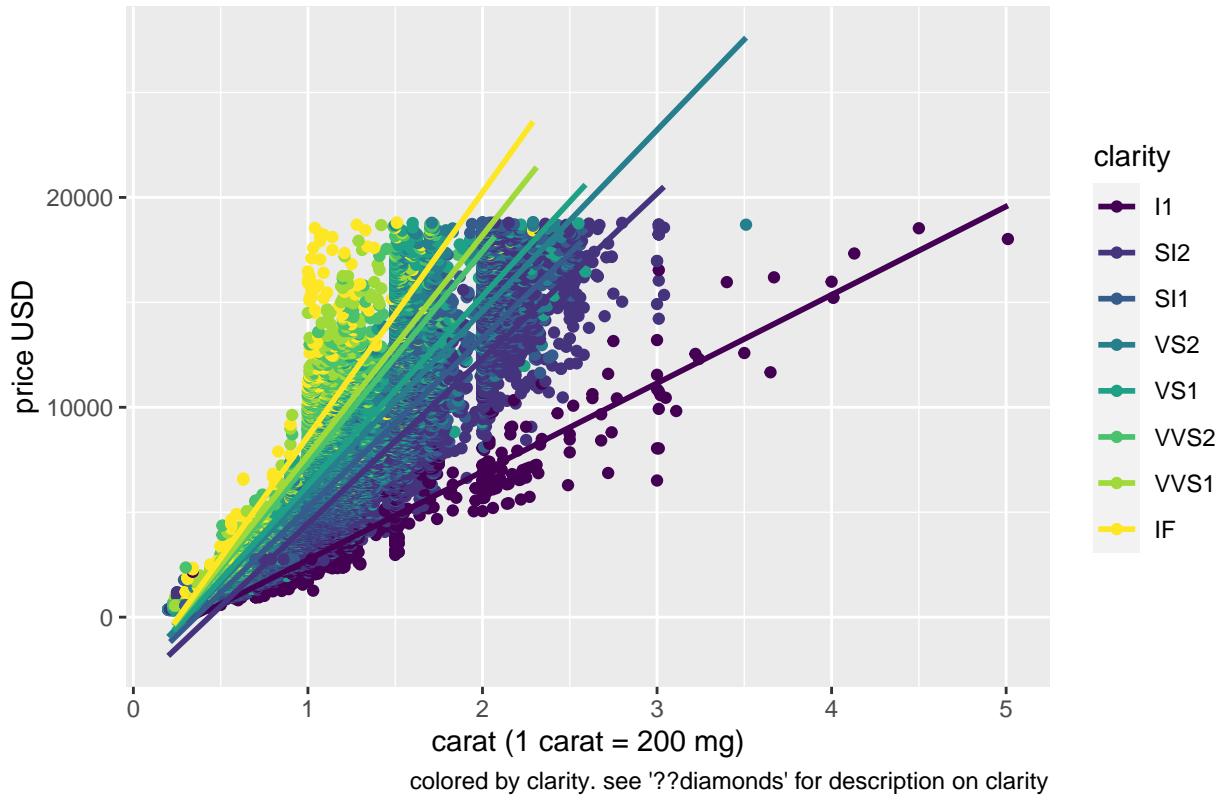


```
# Problem 1 c)
```

```
ggplot(diamonds) +  
  geom_point(mapping = aes(x = carat, y = price, color = clarity)) +  
  geom_smooth(method = lm, se = FALSE, mapping = aes(x = carat, y = price, color = clarity)) +  
  labs(title = "price ~ carat scatterplot w/ trendlines by clarity", y = "price USD",  
       x = "carat (1 carat = 200 mg)",  
       caption = "colored by clarity. see '??diamonds' for description on clarity")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

price ~ carat scatterplot w/ trendlines by clarity

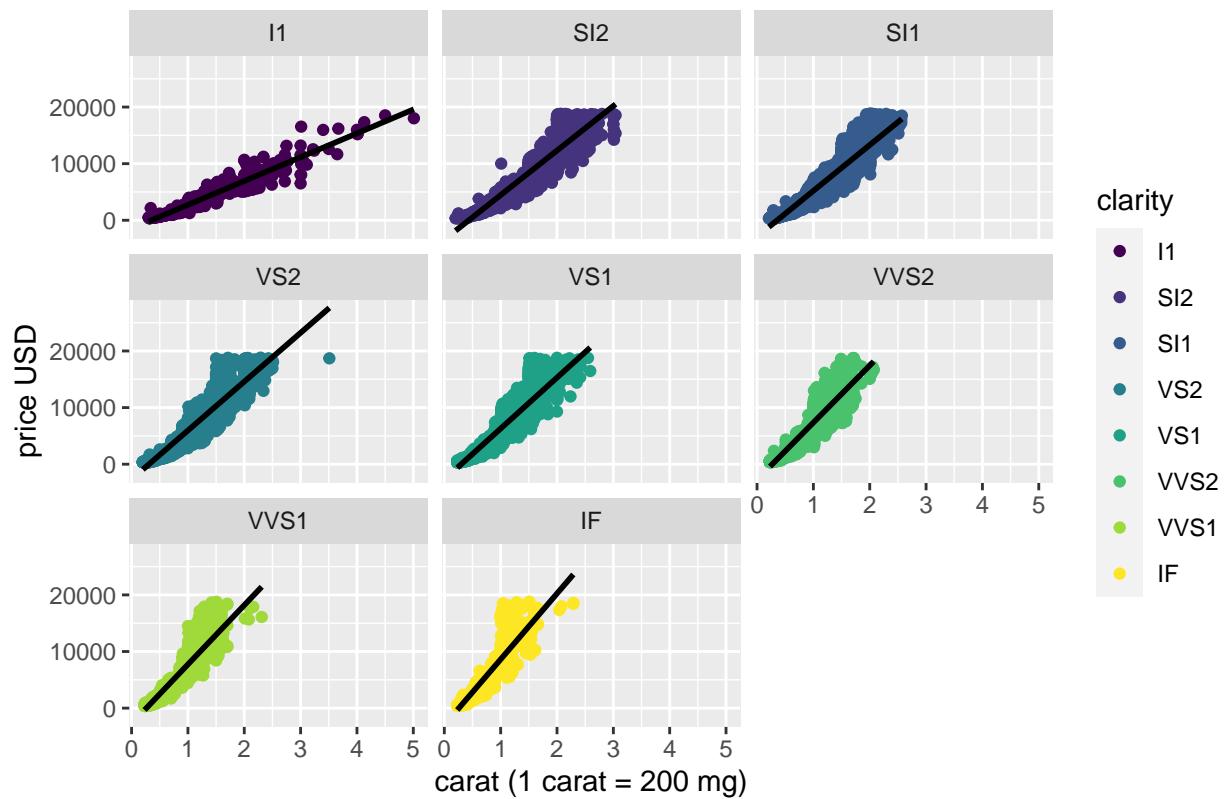


# Problem 1 d)

```
ggplot(diamonds, aes(x = carat, y = price)) +
  geom_point(mapping = aes(color = clarity)) +
  geom_smooth(method = lm, se = FALSE, color = "black") +
  facet_wrap(~clarity, nrow = 3) +
  labs(title = "price ~ carat by clarity", x = "carat (1 carat = 200 mg)",
       y = "price USD")
```

## `geom\_smooth()` using formula 'y ~ x'

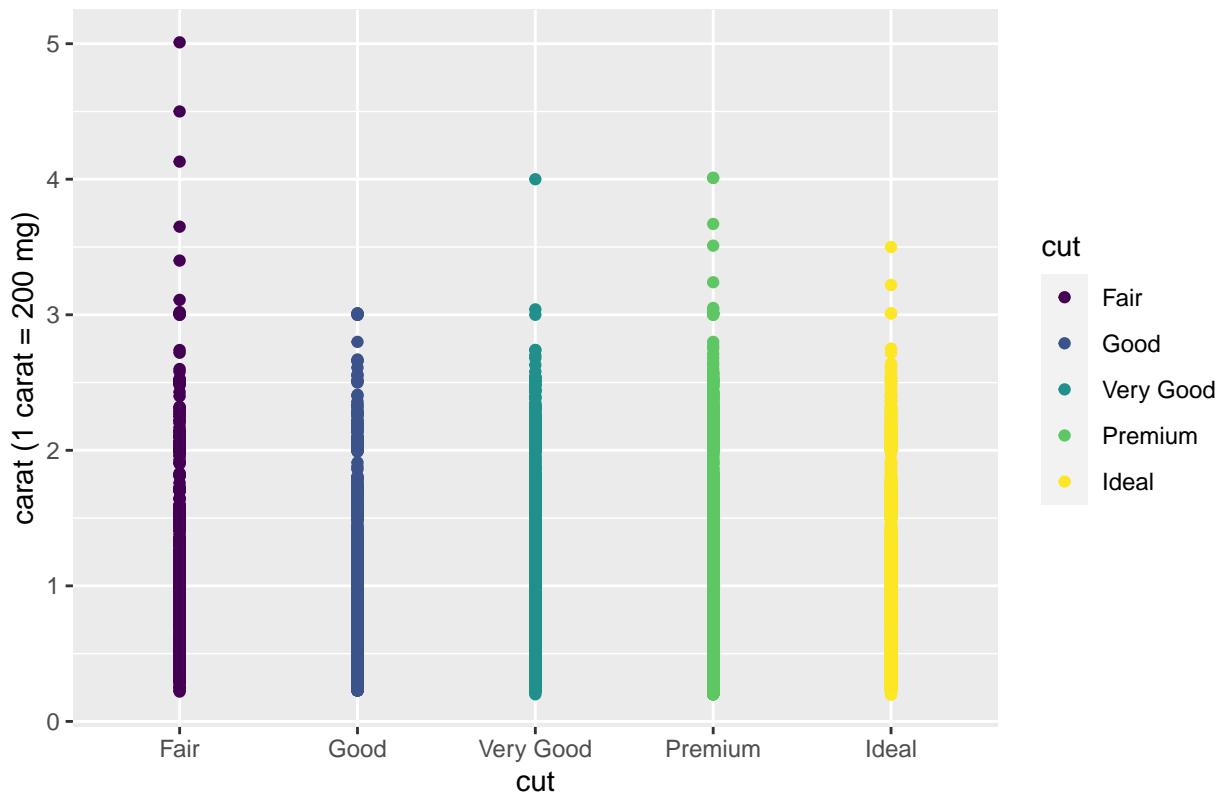
price ~ carat by clarity



```
# Problem 1 e.1)
```

```
ggplot(diamonds, mapping = aes(x = cut, y = carat)) +
  geom_point(mapping = aes(color = cut)) +
  labs(title = "carat by cut point plot", y = "carat (1 carat = 200 mg)")
```

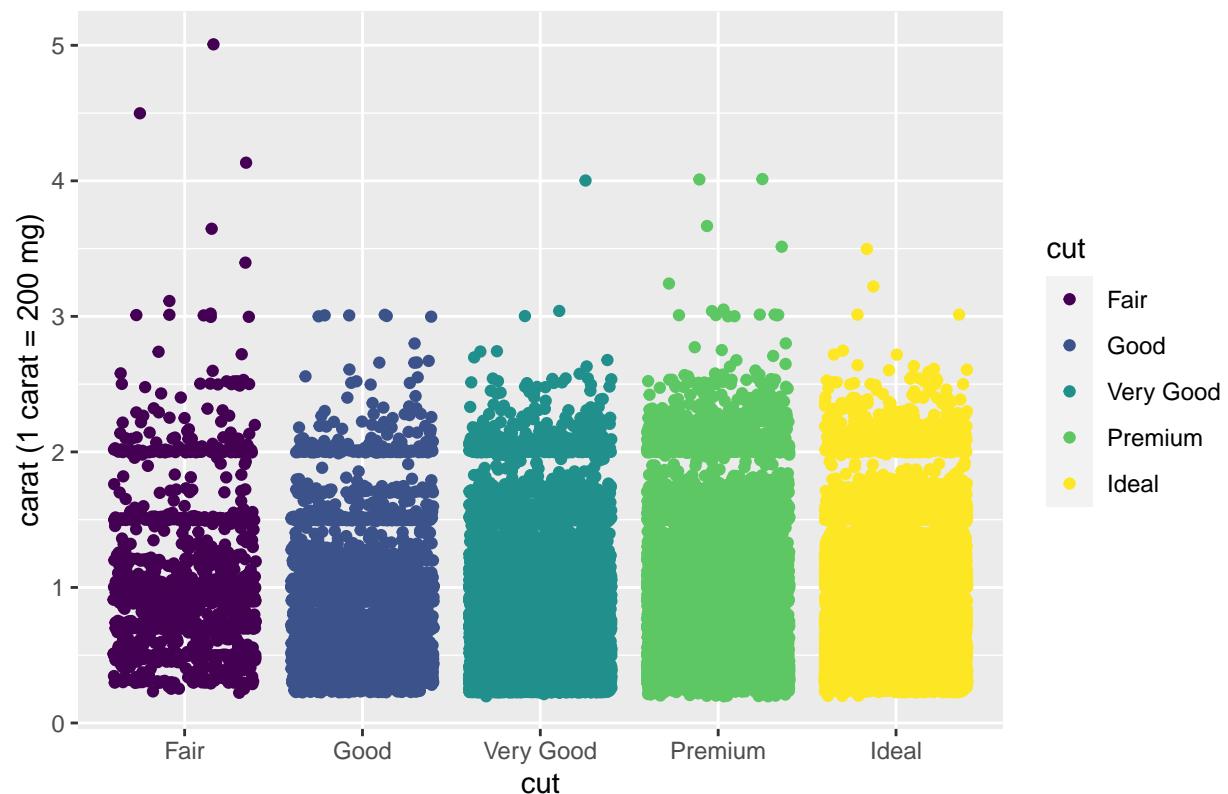
## carat by cut point plot



# Problem 1 e.2)

```
ggplot(diamonds, mapping = aes(x = cut, y = carat)) +  
  geom_jitter(mapping = aes(color = cut)) +  
  labs(title = "carat by cut jitter plot", y = "carat (1 carat = 200 mg)")
```

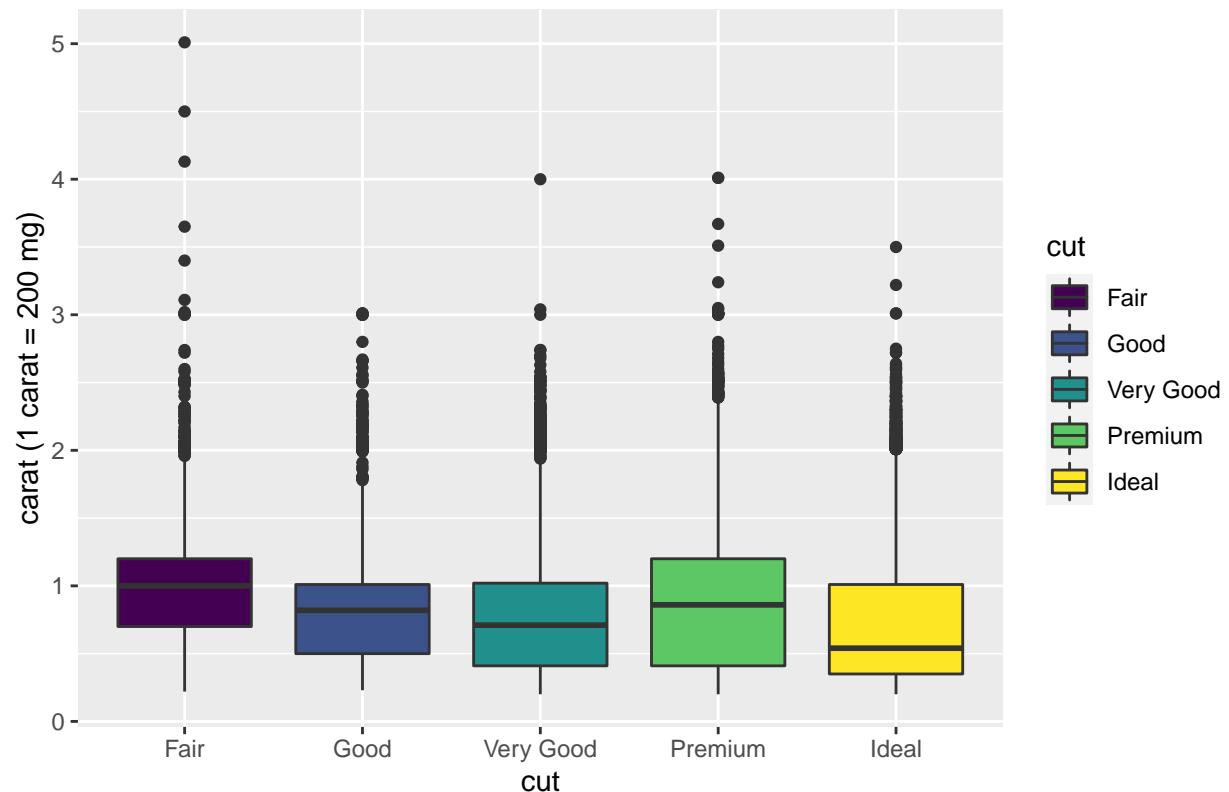
carat by cut jitter plot



```
# Problem 1 e.3)
```

```
ggplot(diamonds, mapping = aes(x = cut, y = carat)) +  
  geom_boxplot(mapping = aes(fill = cut)) +  
  labs(title = "carat boxplot by cut", y = "carat (1 carat = 200 mg)")
```

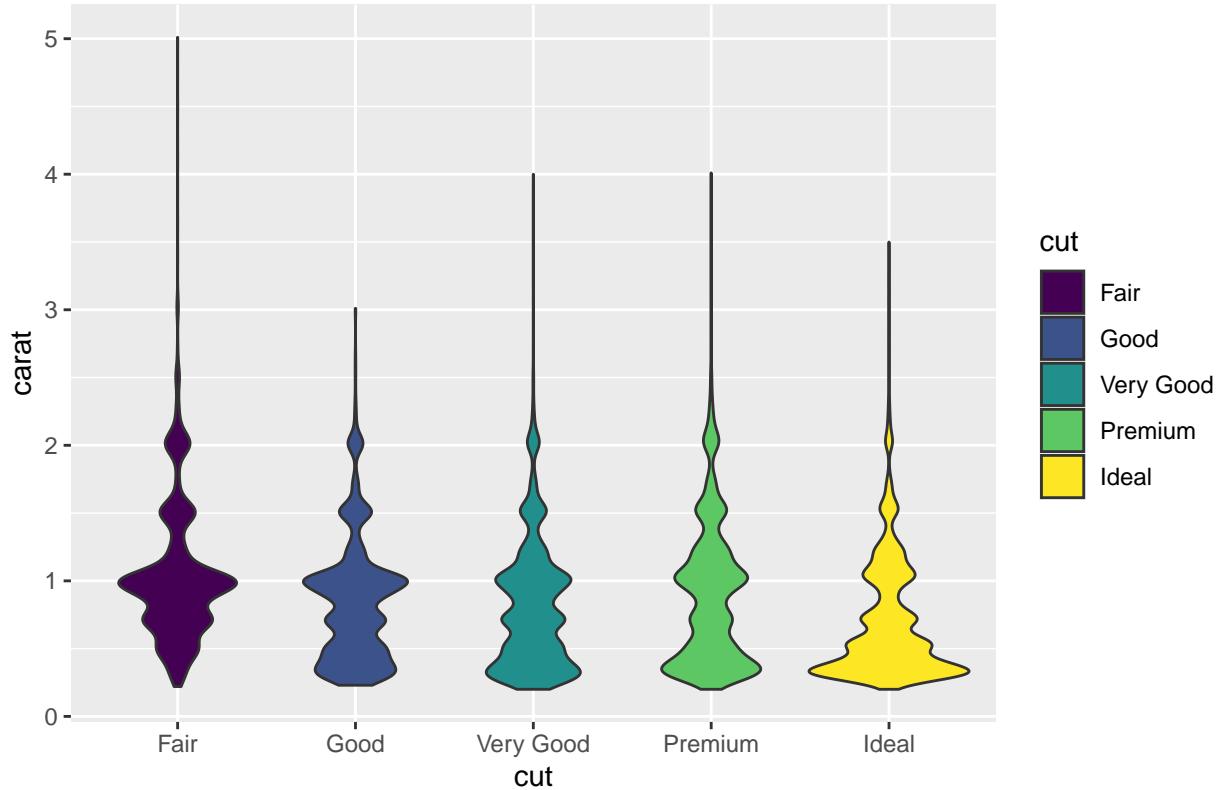
carat boxplot by cut



# Problem 1 e.4)

```
ggplot(diamonds, mapping = aes(x = cut, y = carat)) +  
  geom_violin(aes(fill = cut)) +  
  labs(title = "carat by cut violin plot", aes(title = "topright"))
```

carat by cut violin plot

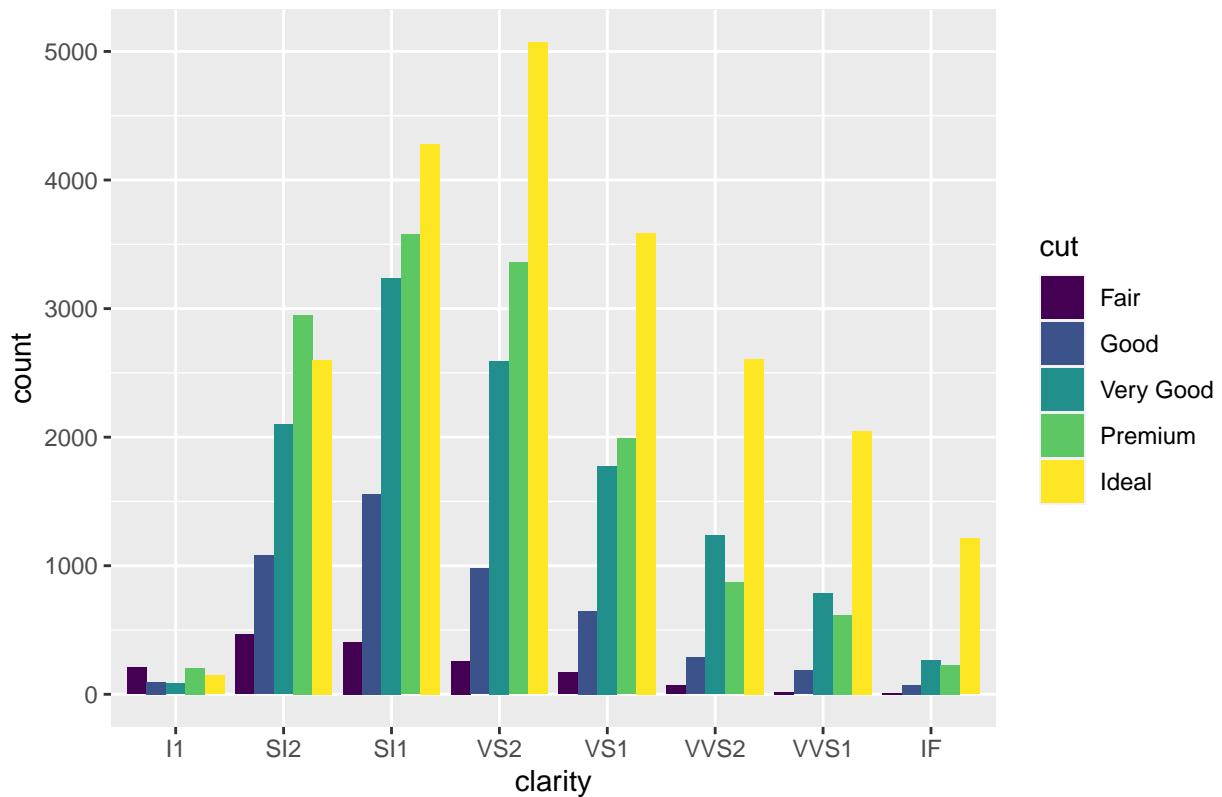


Out of point, jitter, box, and violin plots, violin or box seem to be the best for data visualization while jitter and point seem to be the worst.

### Problem 2 a.1)

```
ggplot(diamonds) +  
  geom_bar(mapping = aes(x = clarity, fill = cut), position = "dodge") +  
  labs(title = "barplot by clarity and cut")
```

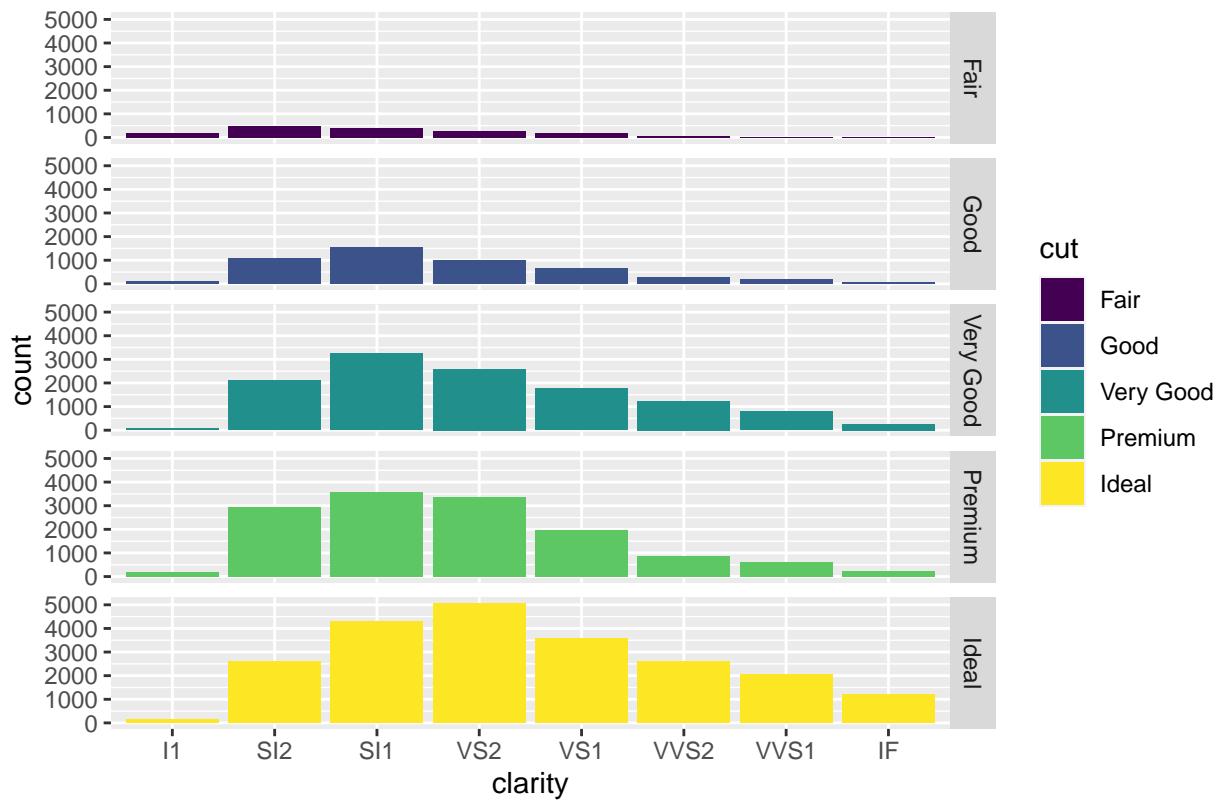
barplot by clarity and cut



## Problem 2 a.2)

```
ggplot(diamonds) +  
  geom_bar(mapping = aes(clarity, fill = cut)) +  
  facet_grid(rows = vars(cut)) +  
  labs(title = "barplot by clarity faceted by cut", x = "clarity")
```

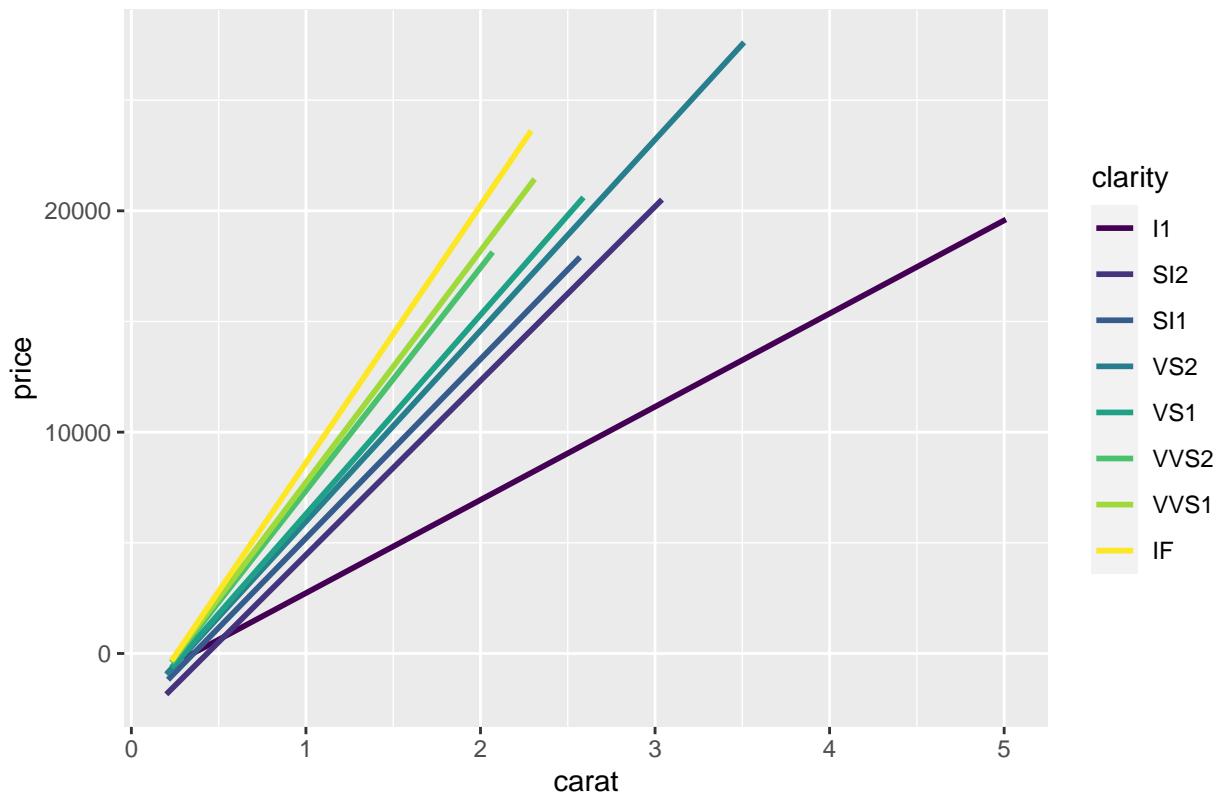
barplot by clarity faceted by cut



#Problem 2 b)

```
ggplot(diamonds, aes(x = carat, y = price, color = clarity)) +
  geom_smooth(formula = y ~ x, se = FALSE, method = lm) +
  labs(title = "price ~ carat linear regression by clarity")
```

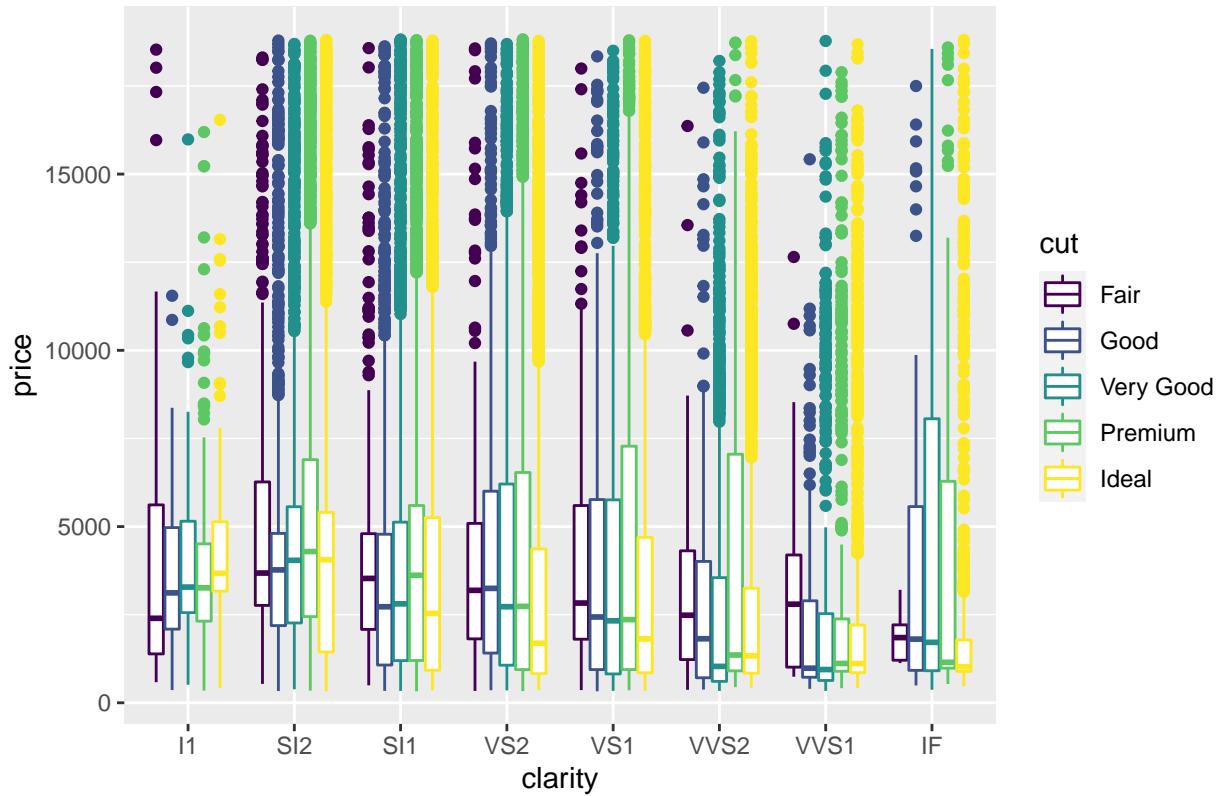
price ~ carat linear regression by clarity



### Problem 2 c)

```
ggplot(diamonds) +  
  geom_boxplot(mapping = aes(y = price, x = clarity, color = cut)) +  
  labs(title = "boxplot price by clarity and cut")
```

boxplot price by clarity and cut



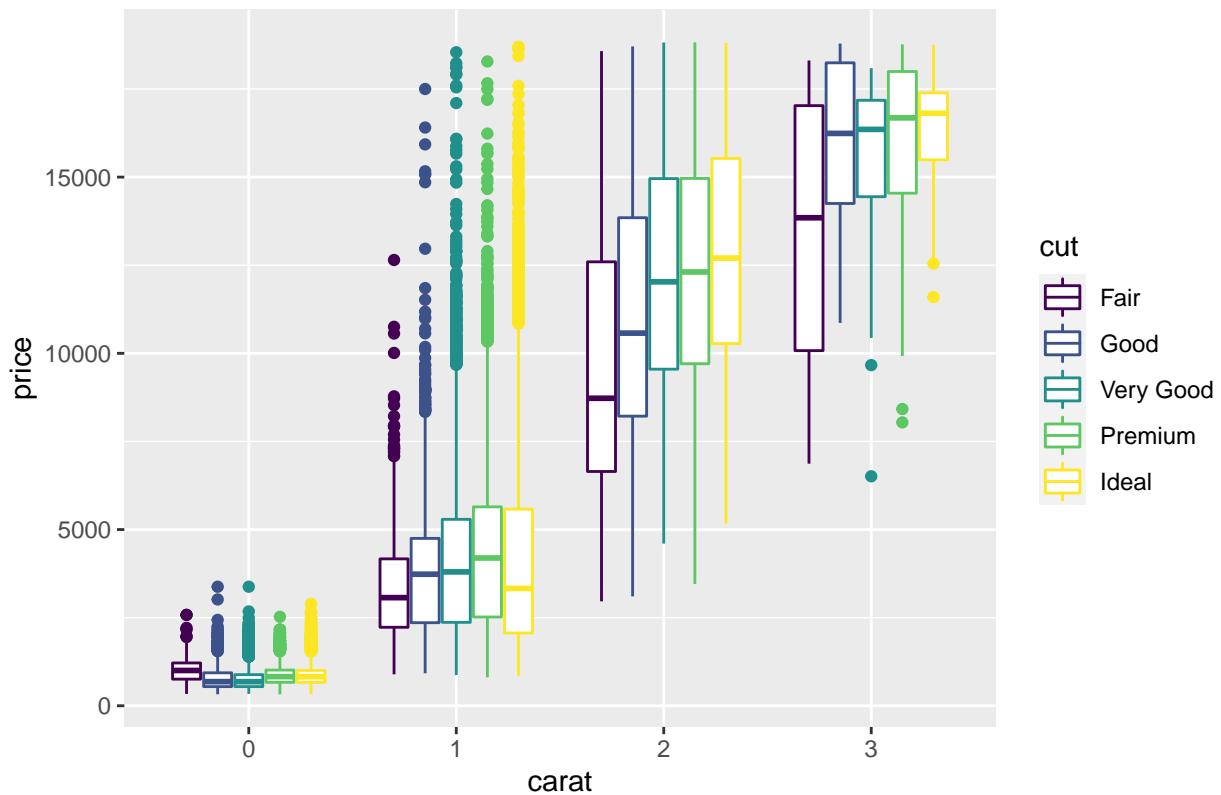
# Problem 2 d)

```

# d1 <- which(diamonds$carat <= 1 & diamonds$carat >= 0)
# d2 <- which(diamonds$carat > 1 & diamonds$carat <= 2)
# d3 <- which(diamonds$carat > 2 & diamonds$carat <= 3)
# d4 <- which(diamonds$carat > 3)
# diamonds$fcarat <- diamonds$carat
# diamonds$fcarat[d1] <- "0"
# diamonds$fcarat[d2] <- "1"
# diamonds$fcarat[d3] <- "2"
# diamonds$fcarat[d4] <- "3"
# factor(diamonds$fcarat, levels = 0:3)
# diamonds$fcarat <- factor(diamonds$carat, levels = 0:3)
ggplot(data = subset(diamonds, subset = round(carat) < 4)) +
  geom_boxplot(mapping = aes(y = price, x = as.factor(round(carat)), color = cut)) +
  labs(title = "boxplot price by carat and cut", x = "carat")

```

boxplot price by carat and cut



## Problem 2 e)

```
ggplot(diamonds) +  
  geom_histogram(binwidth = 0.1, mapping = aes(x = depth, y = ..density..)) +  
  facet_grid(rows = vars(cut)) +  
  labs(title = "density of depth variable facet by cut")
```

density of depth variable facet by cut

