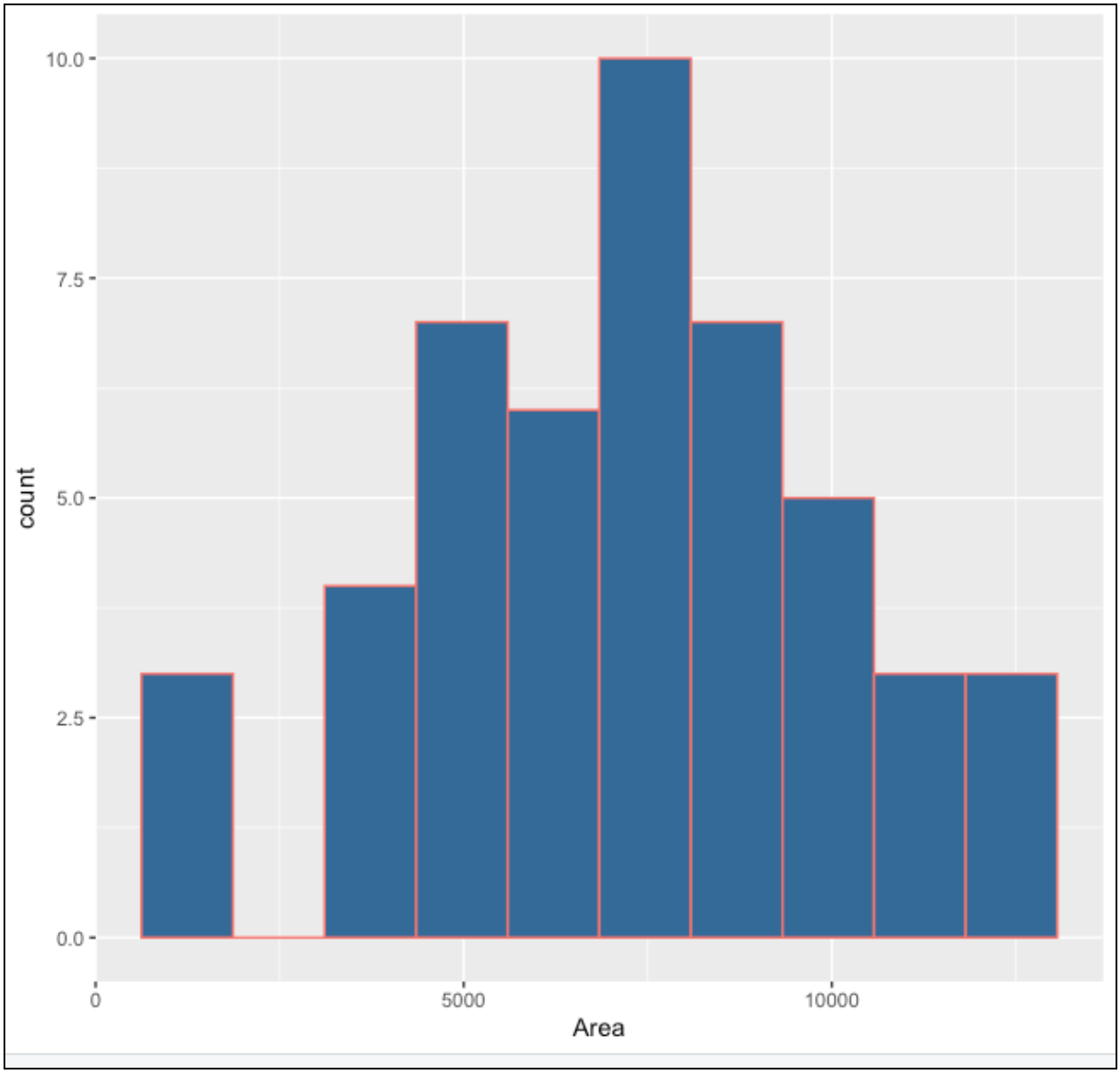
BONBUTTERY

Ggplot Exercises

Hands-on Output 1 (HOO1): Using the same dataset, create a histogram for the variable area, and set the bins to 10.



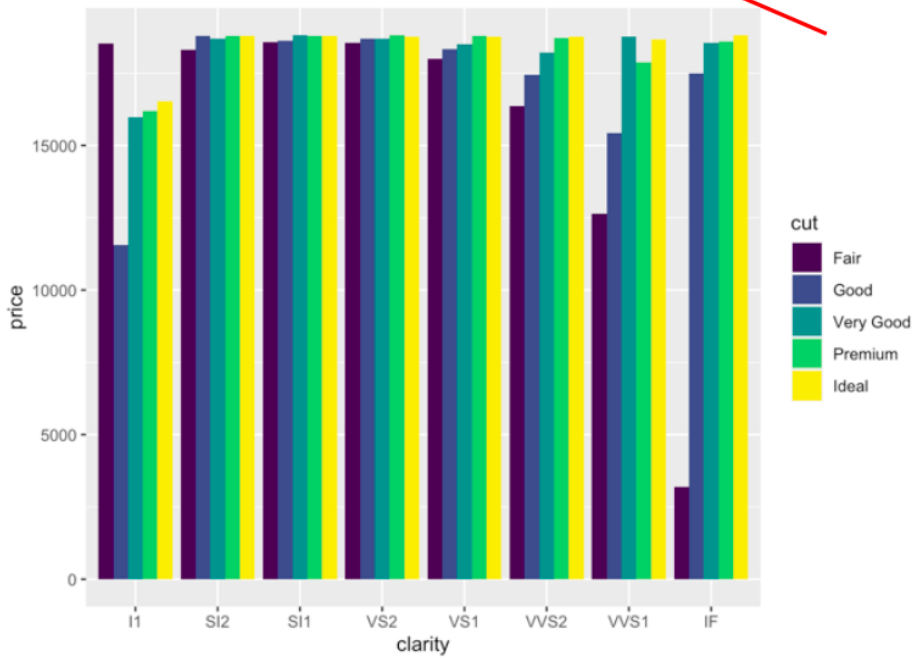
Code: > ggplot(data = rock, (aes(x = area, fill=area[as.factor("area")], color="red"))) + geom_histogram(bins = 10) +labs(x = "Area") + theme(legend.position = "none")

HOO2: Which variable was used as the basis of the grouping?

- The variable “cut” was used as the basis for grouping

```
data(diamonds)

ggplot(data = diamonds, aes(x = clarity, y = price, fill = cut))+ geom_bar(position="dodge", stat="identity")
```



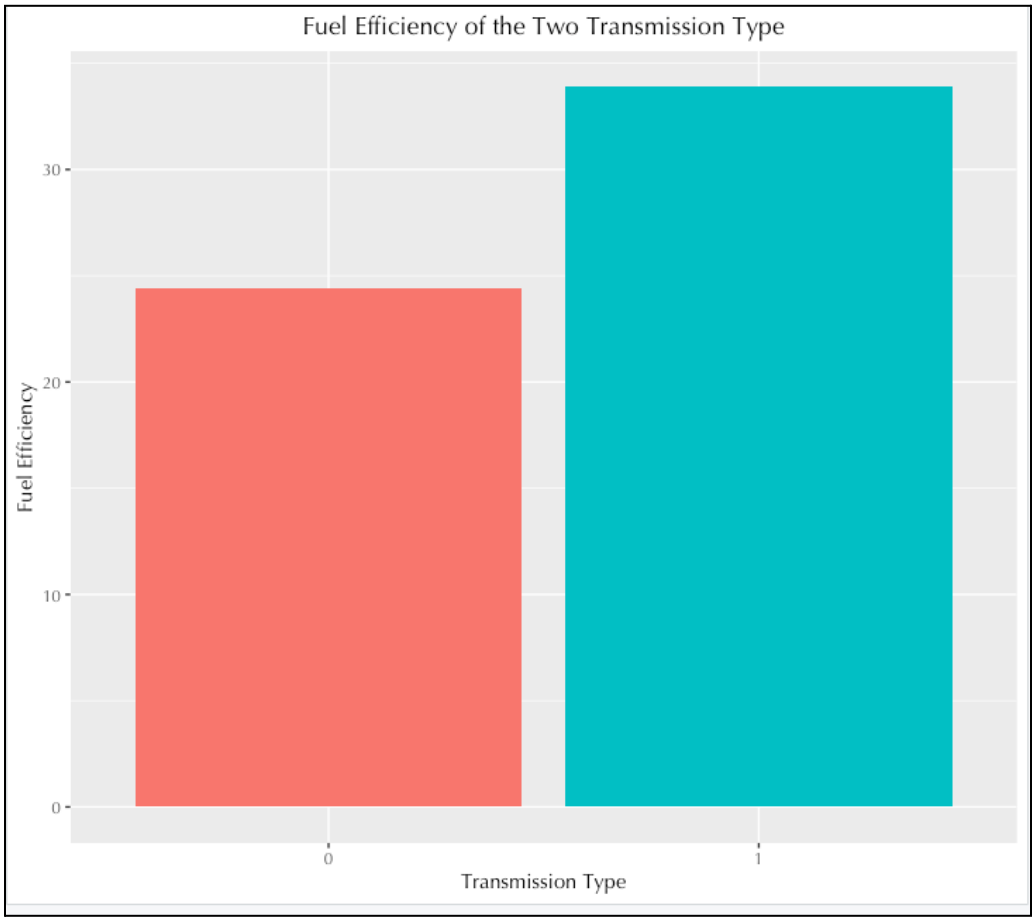
Exercises

Create the plots which will be able to answer the following questions. Make your plots as visually appealing as possible. Refer to the ggplot2 documentation for more information on plot customization.

1. Using the mtcars dataset, is there a significant difference in fuel efficiency based on transmission type?

- **Statistically, the graph cannot tell us if there is a significant difference in fuel efficiency. However, visually, we can see a huge difference in their fuel efficiencies. Therefore, yes, there is a significant difference in fuel efficiency.**

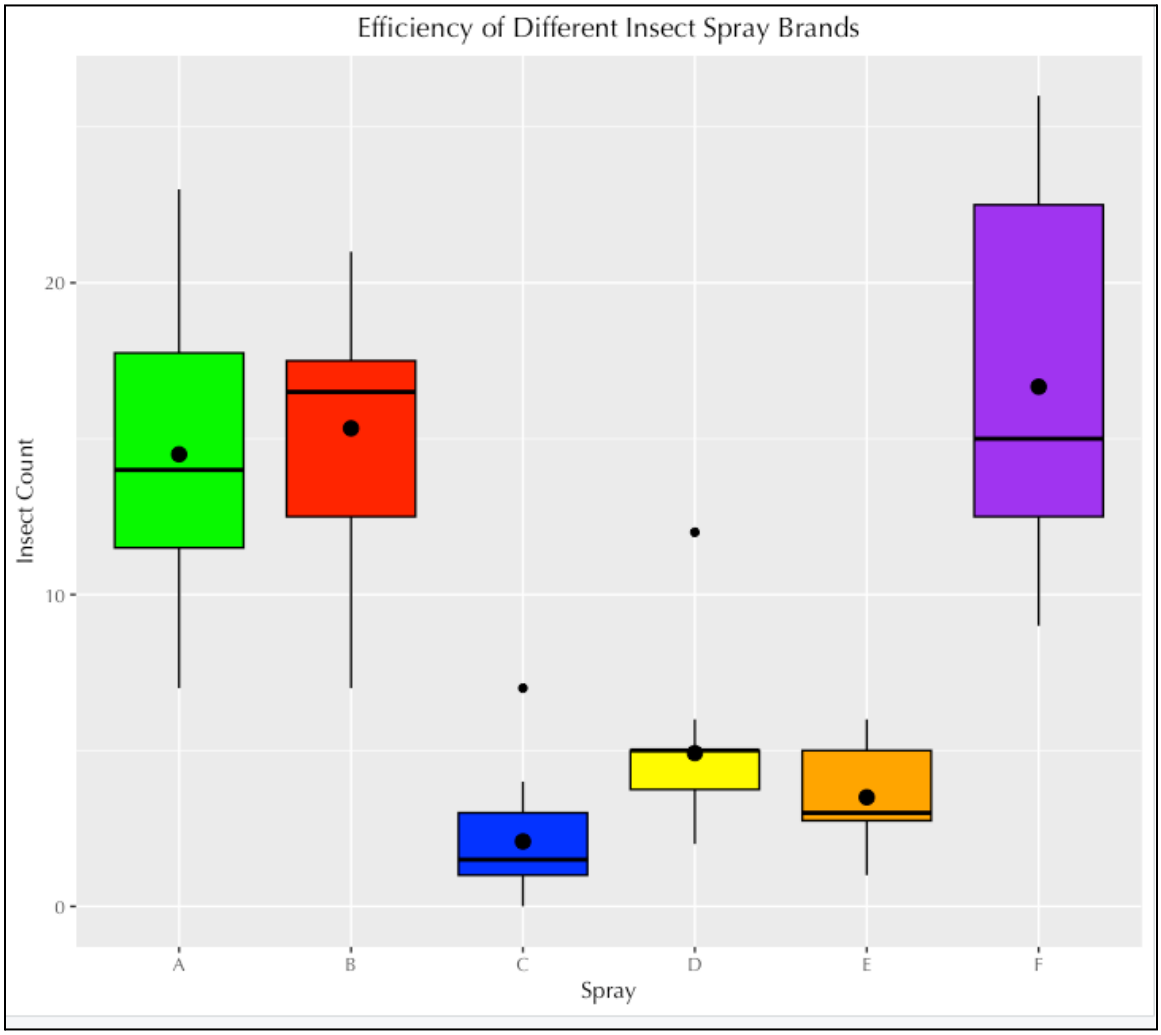
Variables: "mpg" = fuel efficiency and "am" = transmission type



Code: **> ggplot(data = mtcars, aes(x = as.factor(am), y = mpg, fill=as.factor(am))) + labs(x = "Transmission Type", y = "Fuel Efficiency") + ggtitle("Fuel Efficiency of the Two Transmission Type") + theme(text = element_text(family = "Optima")) + theme(plot.title = element_text(hjust = 0.5)) + geom_col(position="dodge") + theme(legend.position = "none")**

2. Using the InsectSprays dataset, which brand appears to be the most effective?

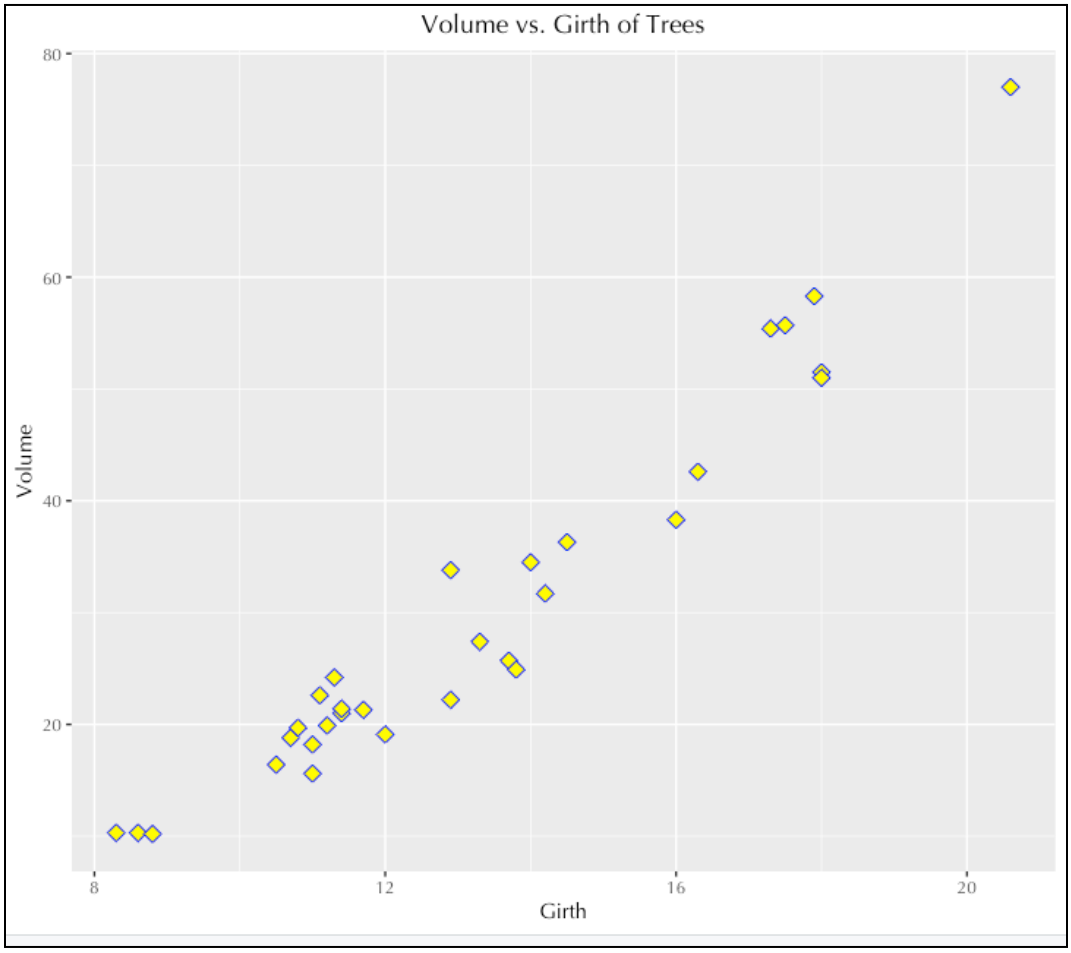
- **Brand C is the most effective Insect spray since it has the lowest mean insect count (black dot inside the boxplots).**



Code: > ggplot(data = InsectSprays, aes(x = spray, y=count)) + labs(x = "Spray", y = "Insect Count") + theme(legend.position = "none") + ggtitle("Efficiency of Different Insect Spray Brands") + theme(plot.title = element_text(hjust = 0.5)) + theme(text = element_text(family = "Optima")) + geom_boxplot(fill=c("Green", "red", "Blue", "Yellow", "Orange", "Purple"), col="Black") + stat_summary(fun=mean, geom='point', shape=19, size=3)

3. Using the trees dataset, which of the two variables exhibit a linear relationship?

- **The variables Girth and Volume exhibit a linear relationship**



Code: ggplot(data = trees, aes(x = Girth, y = Volume, color="Blue")) + labs(x = "Girth", y = "Volume") + theme(legend.position = "none") + ggtitle("Volume vs. Girth of Trees") + theme(plot.title = element_text(hjust = 0.5)) + theme(text = element_text(family = "Optima")) + geom_point(size = 3, shape = 23, fill = "Yellow", color="Blue")