

Homework 2

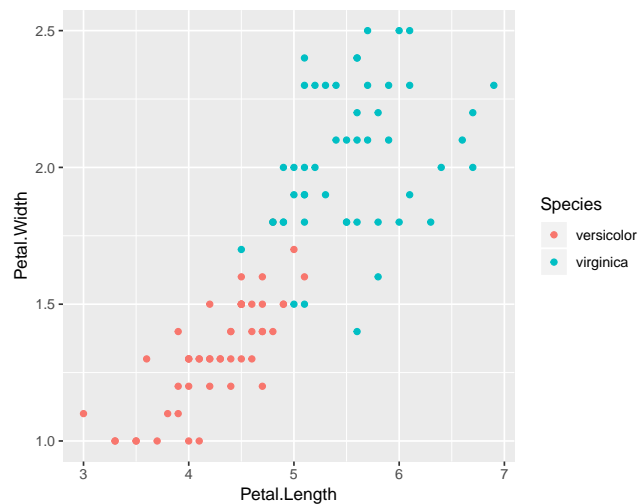
1. Recall `iris` data set. Find the maximum value of the sum `Sepal.Length + Sepal.Width`. To which species does the sample with the maximum value you found above belong? *Hint*: you might want to start with defining a new variable `Sepal.Sum` using `mutate()`, for example.

```
iris %>%  
  mutate(sum = Sepal.Length + Sepal.Width) %>%  
  arrange(desc(sum)) %>%  
  slice(1) %>%  
  pull(Species)
```

```
## [1] virginica  
## Levels: setosa versicolor virginica
```

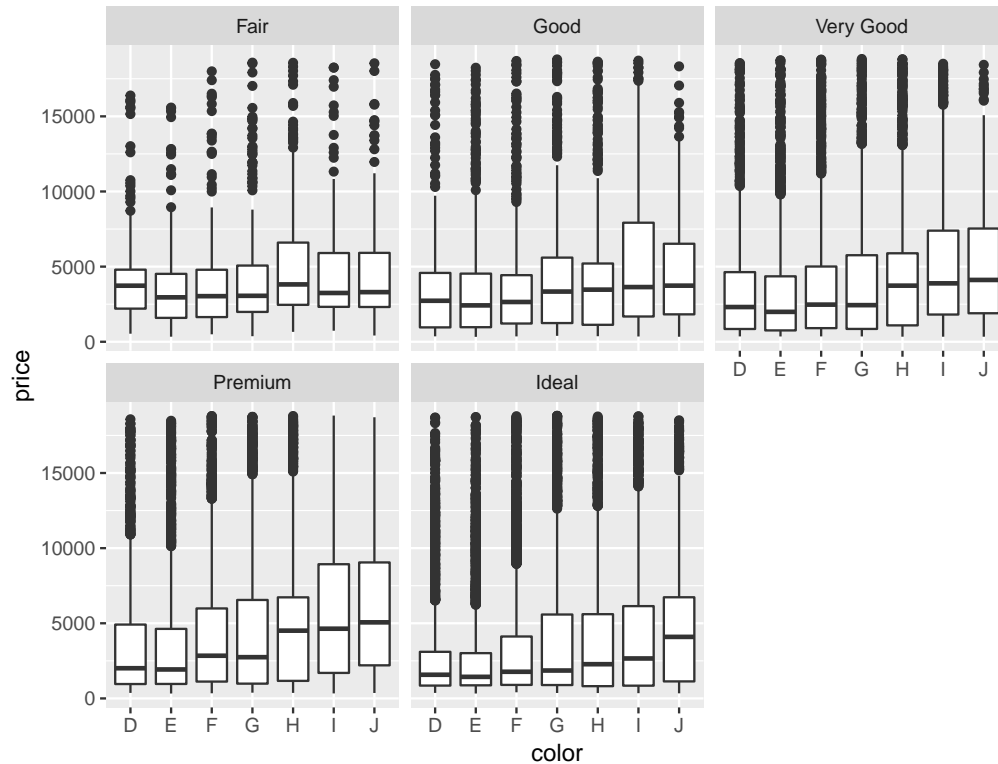
2. Create the following plot using `iris` data set. Note that `setosa` species do not show up. You do NOT need to adjust size or width/height ratio of the plot.

```
iris %>% filter(Species != "setosa") %>%  
  ggplot(aes(Petal.Length, Petal.Width, color = Species)) +  
  theme(aspect.ratio = 1) +  
  geom_point()
```



3. Create the following plot using `diamonds` data set.

```
ggplot(diamonds, aes(color, price)) +  
  geom_boxplot() +  
  facet_wrap(~cut)
```



4. Determine c so that the probability $0 \leq Z \leq c$ is 0.45, where Z follows the standard normal distribution $N(0, 1^2)$.

```
qnorm(0.5 + 0.45, mean = 0, sd = 1)
```

```
## [1] 1.644854
```

5. Recall the `heights.txt` data we used in class with 1375 mom/daughter height pairs:

```
ht <- read.table(file = "heights.txt", header = T, sep = " ")
```

Assuming that UK women's heights follow a normal distribution $N(\mu, \sigma^2)$, compute the 95% confidence interval of μ based on the mom's height data.

```
t.test(ht$Mheight)
```

```
##
## One Sample t-test
##
## data: ht$Mheight
## t = 983.32, df = 1374, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 62.32821 62.57739
## sample estimates:
## mean of x
## 62.4528
```