Homework 2

Wenxuan Zhou wz4388

This homework is due on Feb. 1, 2021 at 11:00pm. Please submit as a pdf file on Canvas.

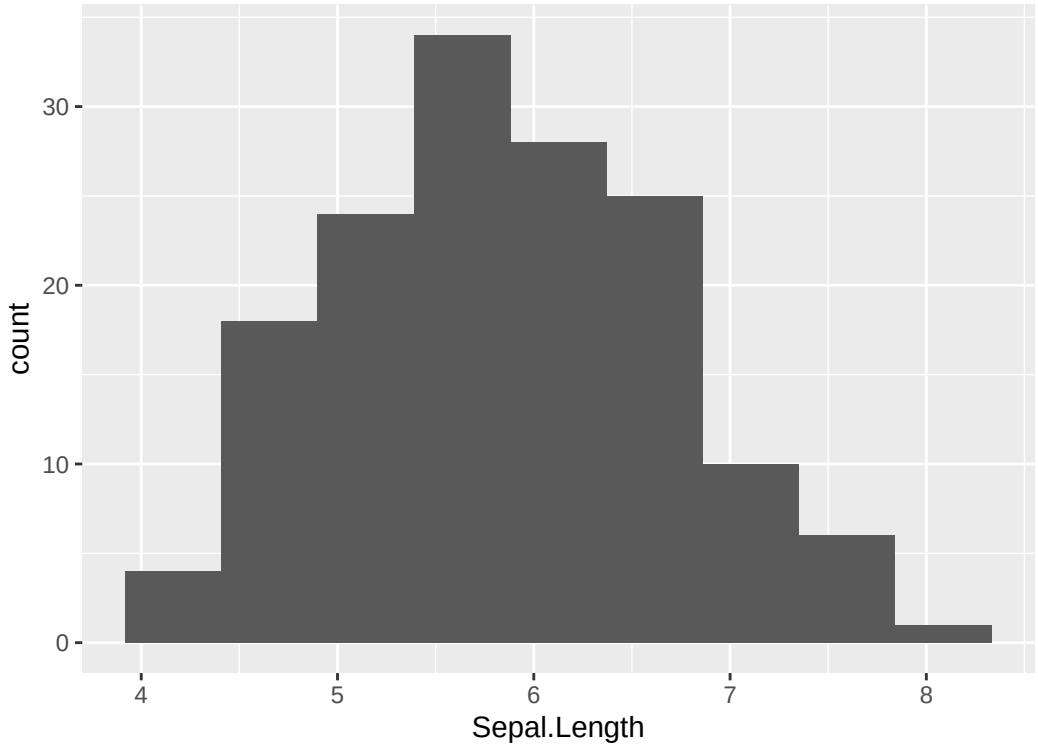
In this homework you will be working with the `iris` dataset built into R. This data set contains measurements of flowers (sepal length, sepal width, petal length, petal width) for three different *Iris* species (*I. setosa*, *I. versicolor*, *I. virginica*).

```
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Problem 1: (6 pts) Use ggplot to make a histogram of the `Sepal.Length` column. Manually choose appropriate values for `binwidth` and `center`. Explain your choice of values in 2-3 sentences.

```
ggplot(iris, aes(Sepal.Length)) + geom_histogram(binwidth = 0.49, center = 0.245)
```



I choose 0.49 as bin width because it would result in 9 bins in the histogram, and the pattern could be seen through the histogram. The center is 0.245, which is half of the bin width. This is to avoid the occurrence of negative sepal length (if applicable).

Problem 2: (4 pts) Modify the plot from Problem 1 to show one panel per species. Hint: Use `facet_wrap()`. See Slide 14 from Class 2.

```r
ggplot(iris, aes(Sepal.Length)) + geom_histogram(binwidth = 0.49, center = 0.245) +
  facet_wrap(vars(Species))
```