

# Iris Analysis-Week-6-Riaz

Riaz Ahmed Tamim Ansari

2023-10-08

## Iris Analysis

```
df = data.frame(iris)
head(df)
```

```
##   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
```

## Average Sepal Length by Species

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
df %>% group_by(Species) %>% summarize(Meanlength_Sepal = mean(Sepal.Length))
```

```
## # A tibble: 3 x 2
##   Species    Meanlength_Sepal
##   <fct>         <dbl>
## 1 setosa         5.01
## 2 versicolor     5.94
## 3 virginica      6.59
```

## Visualizations

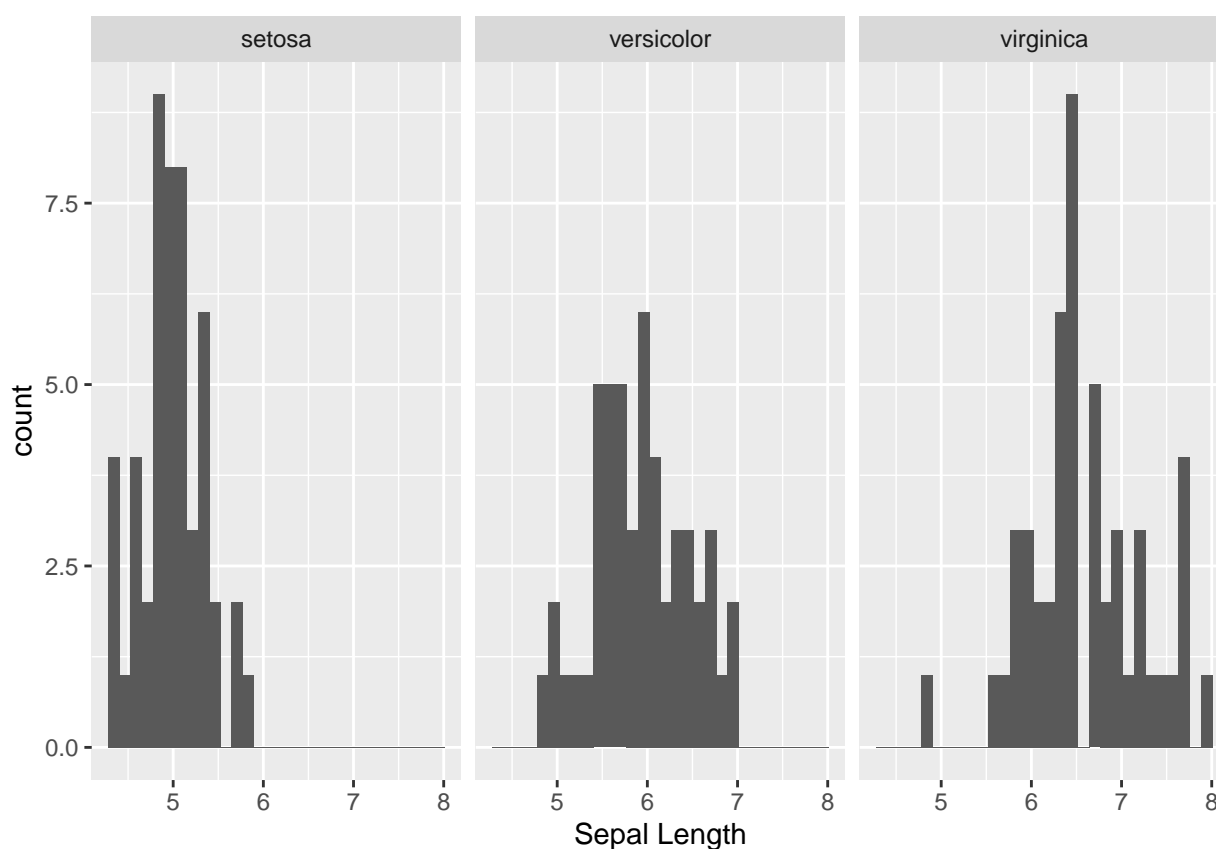
Looking through the Sepal length of the species, I can infer virginica has the longest with the average of 6.5 and some in the range above 7.5. For setosa, mean is around 5 and for versicolor it is less than 6. Almost all of the distributions are close to normal distribution. All the three are having negative kurtosis, meaning there are very less presence at the tail ends. Same information is inferred from box plot also, the box plot of virginica also says there is an outlier at the lower end, below 5

Looking at the Sepal width, I can infer setosa has the highest mean in between 3.25 and 3.5. I can also observe outliers on both the ends in setosa and virginica. From the histogram, I can see that setosa and versicolor are slightly positively skewed.

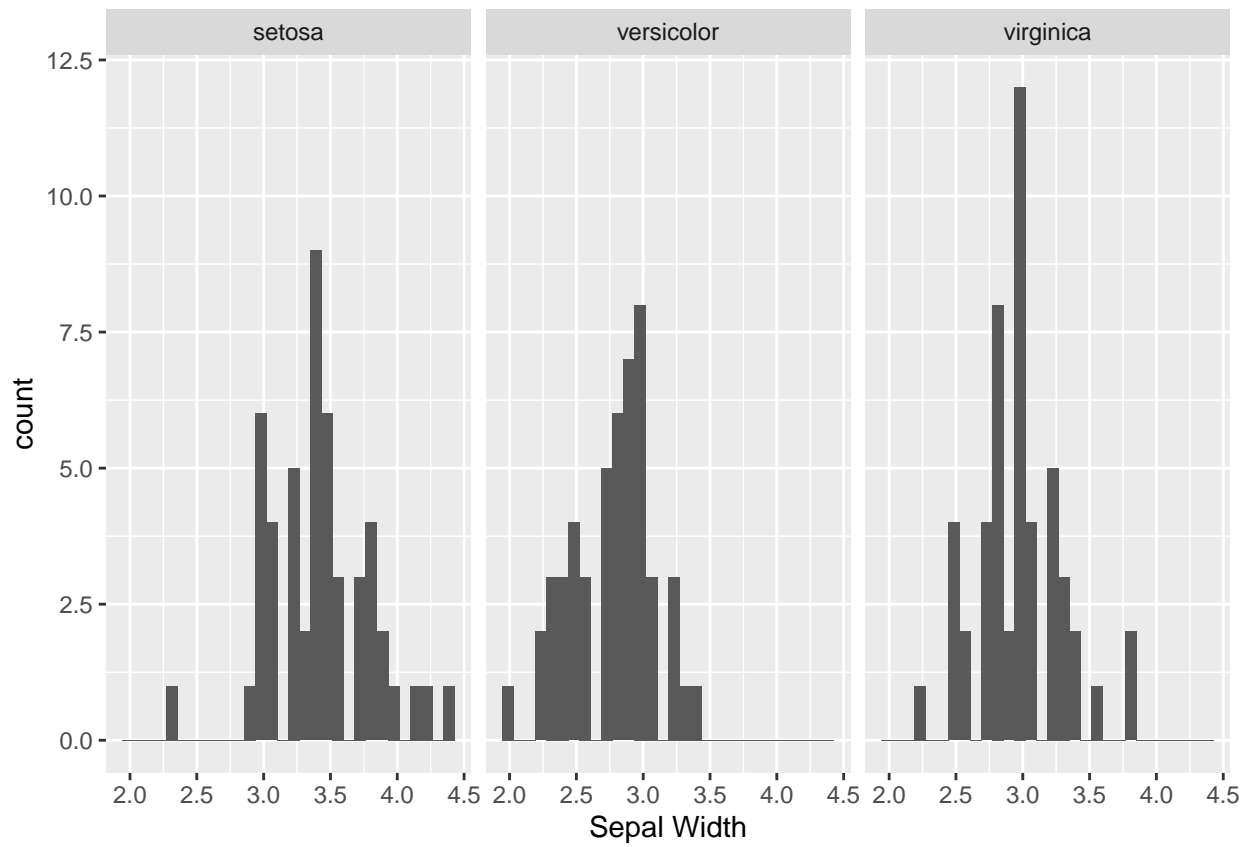
Petal length does not follow normal distribution. I can see that setosa is the shortest followed by versicolor and virginica. From boxplot, for versicolor and virginica, most of the observations lie below mean. There are also outliers observed in setosa and versicolor.

Petal width, setosa is the shortest followed by versicolor and virginica. Most of the values lie above mean for both versicolor and virginica. They do not follow normal distribution.

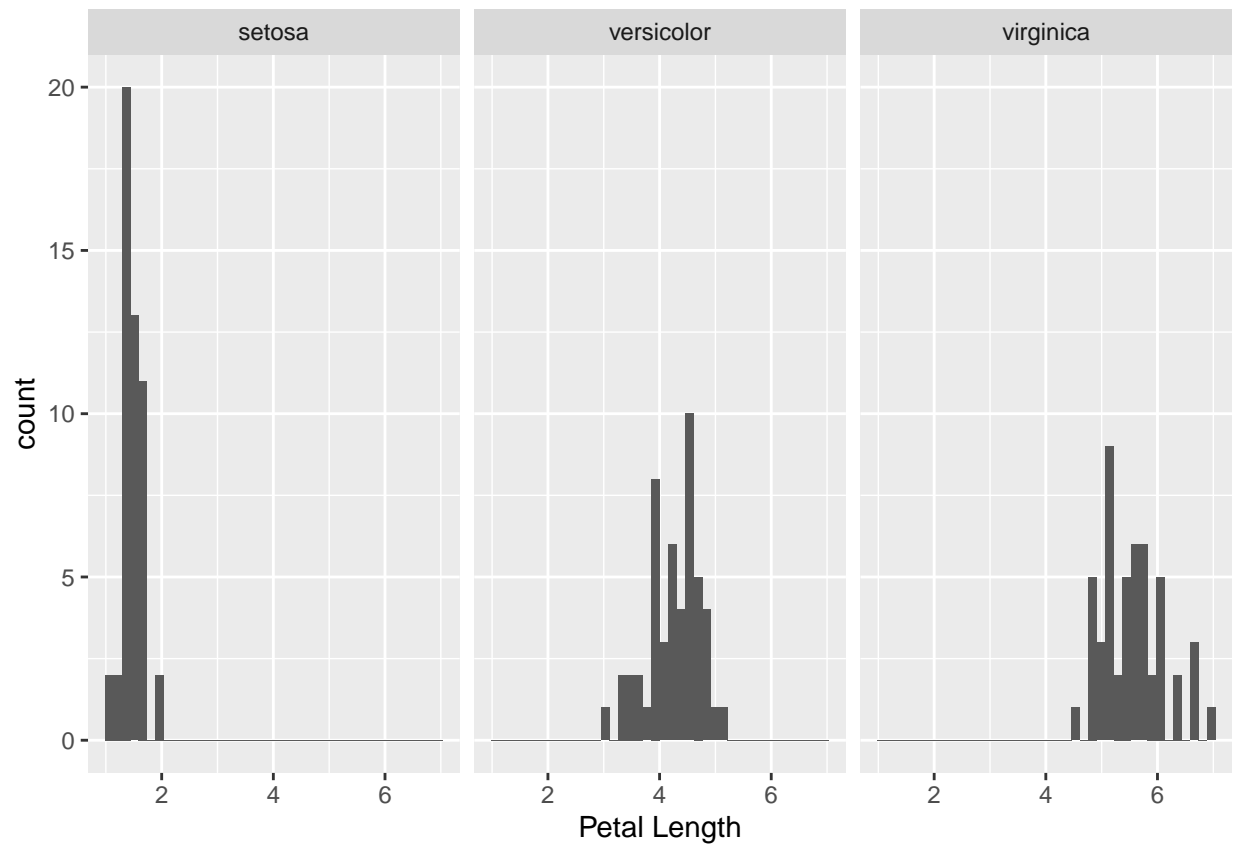
```
df_plot <- ggplot(df)
df_plot + geom_histogram(aes(x=Sepal.Length), bins = 30) + facet_wrap(~Species) + xlab("Sepal Length")
```



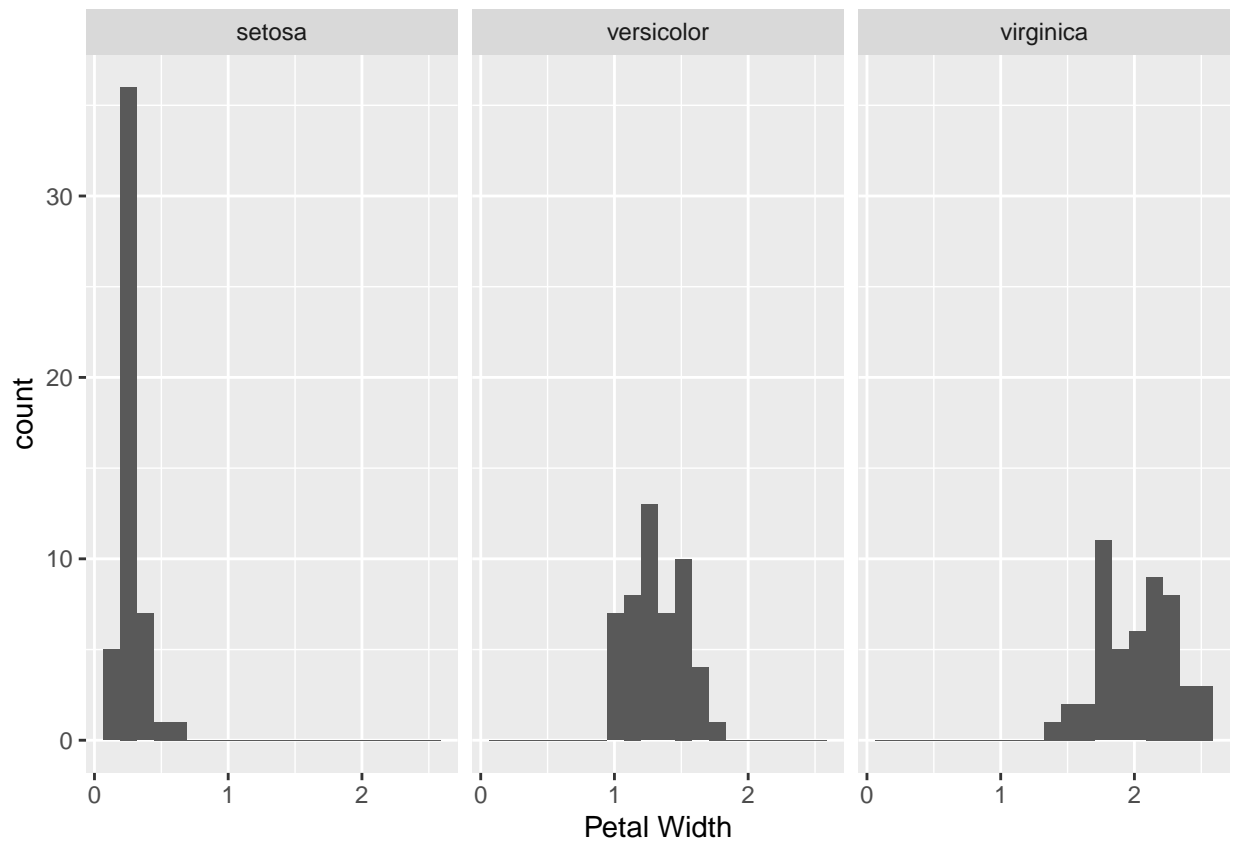
```
df_plot + geom_histogram(aes(x=Sepal.Width), bins = 30) + facet_wrap(~Species) + xlab("Sepal Width")
```



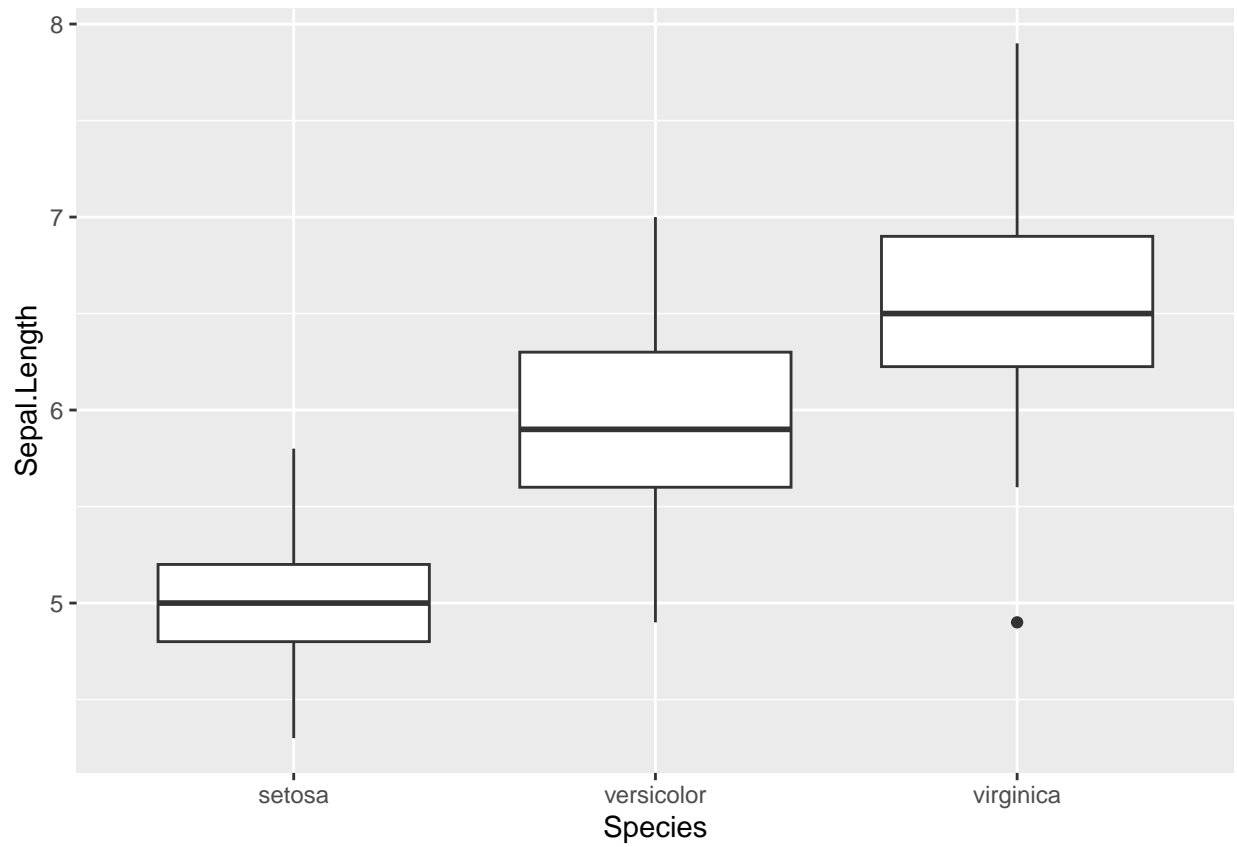
```
df_plot + geom_histogram(aes(x=Petal.Length),bins = 40) + facet_wrap(~Species) + xlab("Petal Length")
```



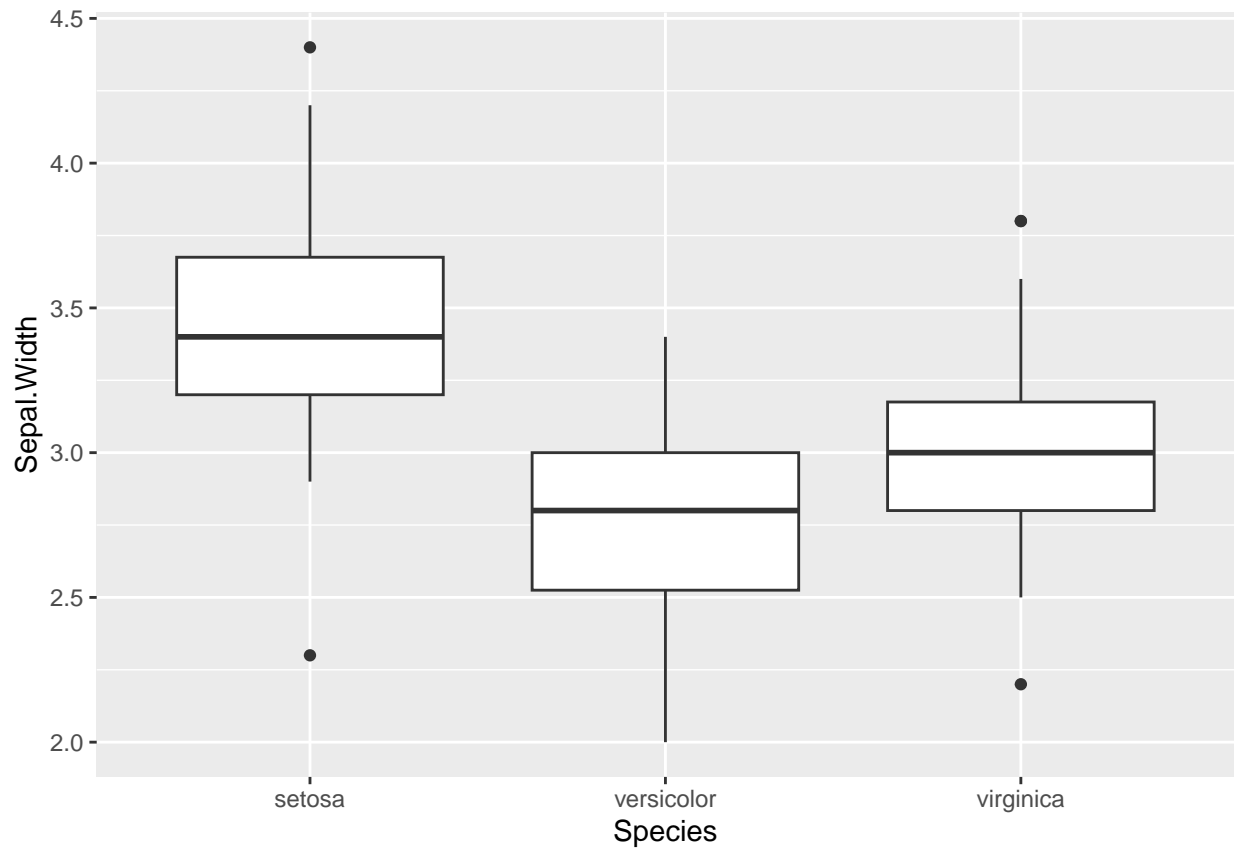
```
df_plot + geom_histogram(aes(x=Petal.Width),bins = 20) + facet_wrap(~Species) + xlab("Petal Width")
```



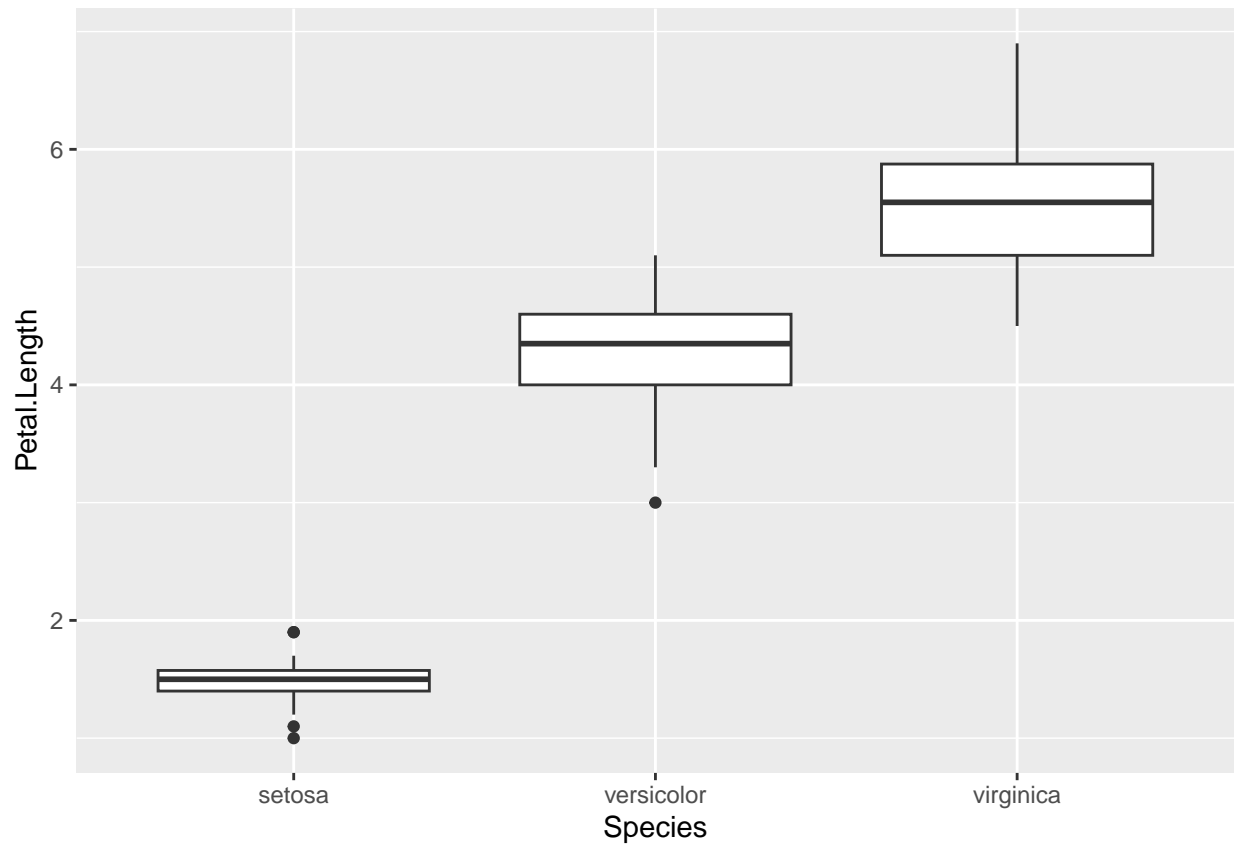
```
df_plot + geom_boxplot(aes(Species, Sepal.Length))
```



```
df_plot + geom_boxplot(aes(Species, Sepal.Width))
```



```
df_plot + geom_boxplot(aes(Species,Petal.Length))
```



```
df_plot + geom_boxplot(aes(Species,Petal.Width))
```



