

Homework 3 2022.05.10

Leo

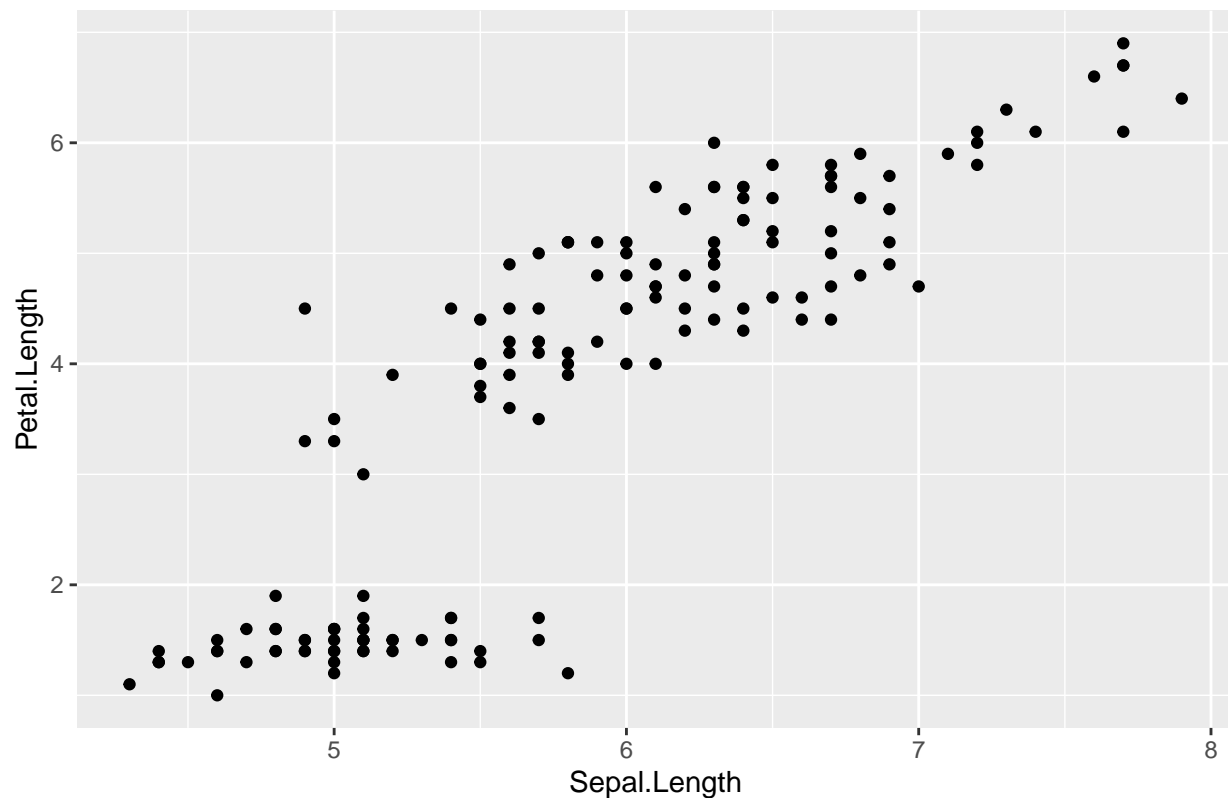
10/05/2022

First loading basic settings and related packages, and the code is not displayed.

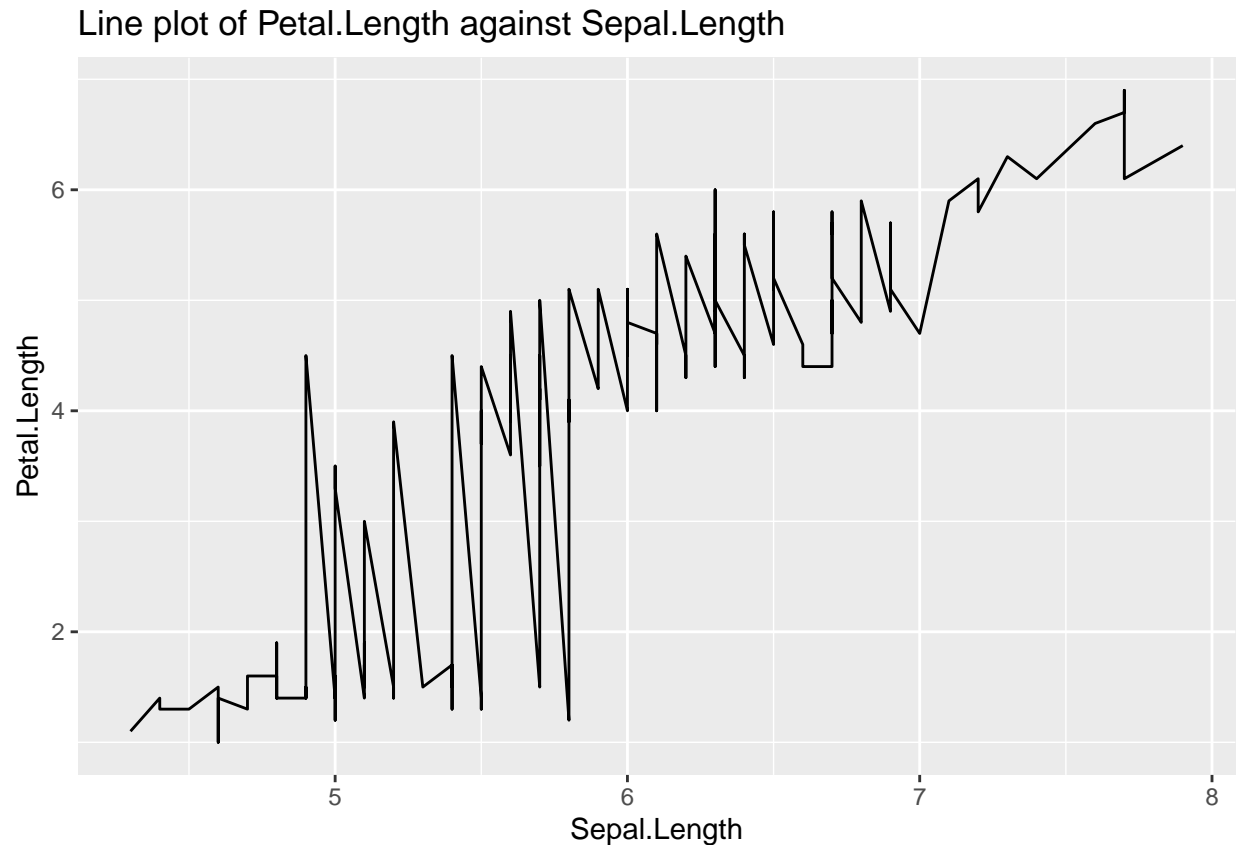
Scatter and Line plot of only Petal.Length against Sepal.Length

```
> ggplot(data = iris, mapping = aes(x = Sepal.Length, y = Petal.Length)) +  
+   geom_point() +  
+   ggtitle("Scatter plot of Petal.Length against Sepal.Length")
```

Scatter plot of Petal.Length against Sepal.Length



```
> ggplot(data = iris, mapping = aes(x = Sepal.Length, y = Petal.Length)) +  
+   geom_line() +  
+   ggtitle("Line plot of Petal.Length against Sepal.Length")
```



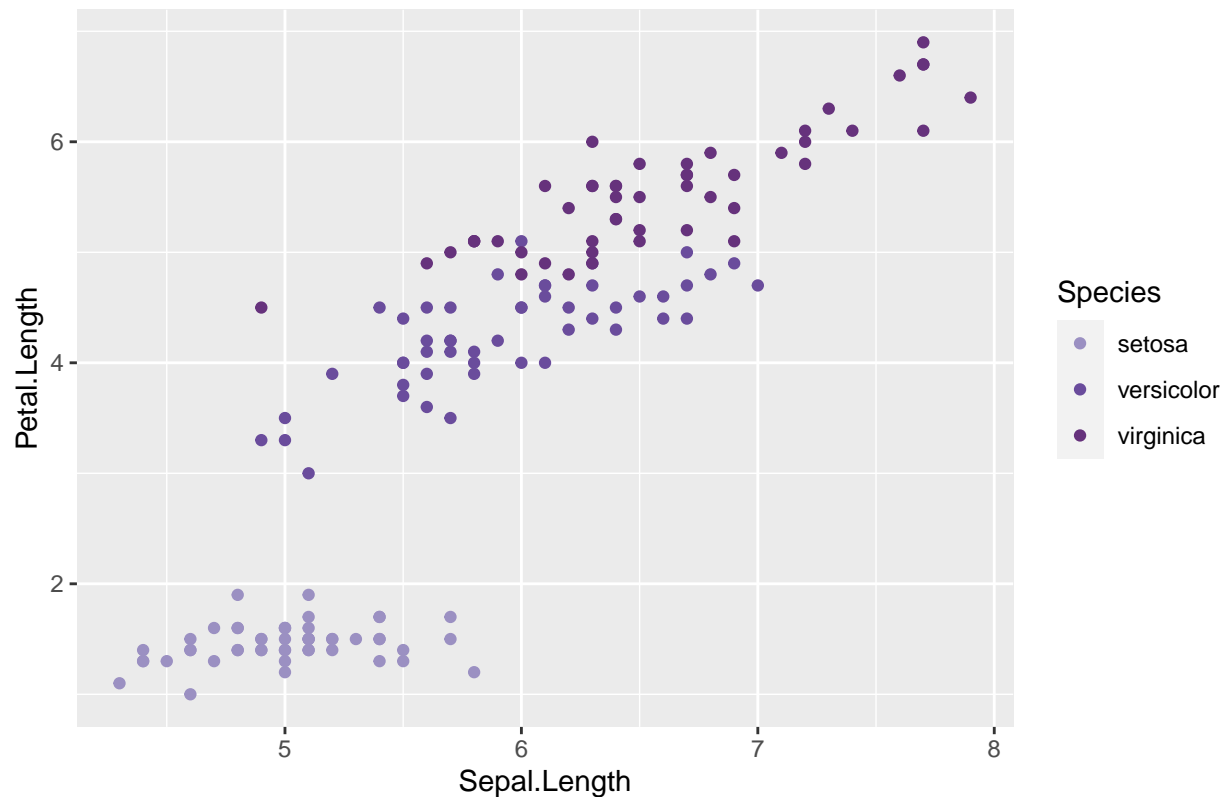
Enconter an analytic problem

As is shown in the Scatter and Line plot, we can see that statistics are widely distributed. In general the Petal.Length increases as the increase of Sepal.Length. But there's still some parts which don't follow this pattern. For example, when the Sepal.Length is around five, the Petal.Length stays still as the increase of Sepal.Length. So maybe we should take more parameters into account.

Scatter plot of Petal.Length against Sepal.Length(Taking Species into account)

```
> colorset1 <- c("#9B90C2", "#6A4C9C", "#66327C")
> iris%>%
+   ggplot(mapping = aes(x = Sepal.Length, y = Petal.Length, color = Species)) +
+   geom_point() +
+   ggtitle("Different Species' scatter plot of Petal.Length against Sepal.Length") +
+   scale_colour_manual(values = colorset1)
```

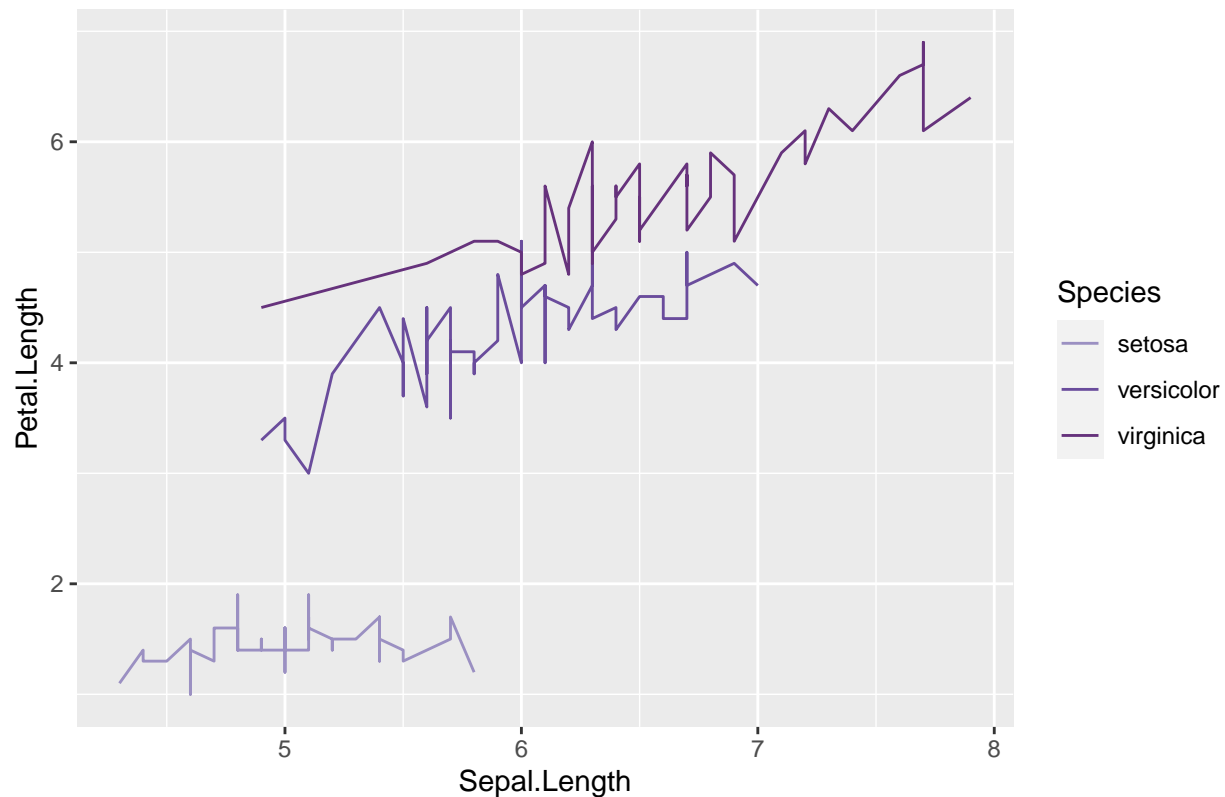
Different Species' scatter plot of Petal.Length against Sepal.Length



Line plot of Petal.Length against Sepal.Length(Taking Species into account)

```
> iris%>%
+   ggplot(mapping = aes(x = Sepal.Length,y = Petal.Length,color = Species)) +
+   geom_line() +
+   ggtitle("Different Species' line plot of Petal.Length against Sepal.Length") +
+   scale_colour_manual(values = colorset1)
```

Different Species' line plot of Petal.Length against Sepal.Length



Relationship between Sepal.Length and Petal.Length

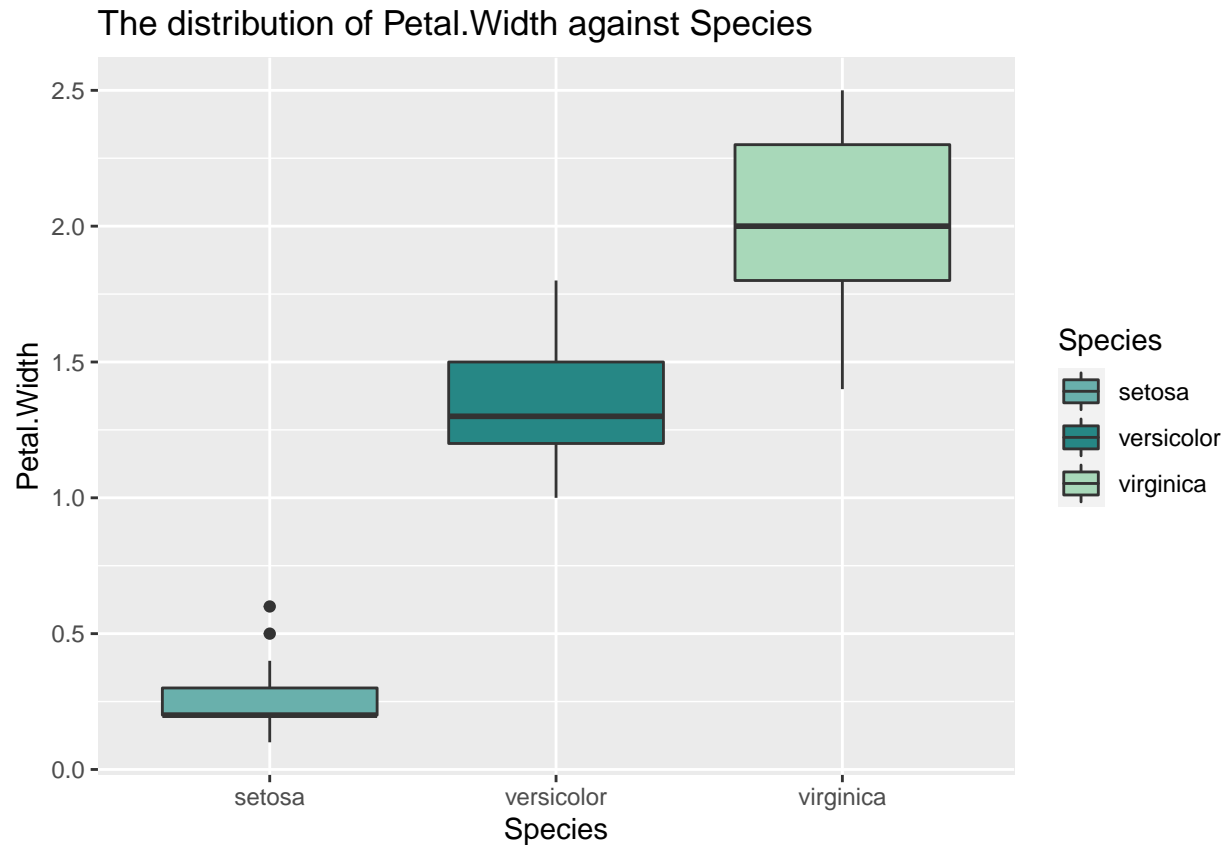
After taking species into account, we can apparently see different patterns of relationship between Sepal.Length and Petal.Length.

When their Sepal.Length are equal, generally Setosa has the shortest Petal.Length and Virginica has the longest.

As for the relationship pattern, with the increase of Sepal.Length, Setosa's Petal.Length would stay still, while the Petal.Length of Versicolor and Virginica would increase, and the latter would increase more pronounced.

Boxplot of Petal.Width against Species

```
> colorset2 <- c("#69B0AC", "#268785", "#A8D8B9")
> iris%>%
+   ggplot(mapping = aes(x = Species, y = Petal.Width, fill = Species)) +
+   geom_boxplot() +
+   ggtitle("The distribution of Petal.Width against Species") +
+   scale_fill_manual(values = colorset2)
```



The distribution of Petal.Width against Species

As is shown in the box plot, the distribution of Petal.Width between different species is significantly different. We can see that Setosa has the narrowest Petal.Width in general and Virginica has the widest.

Constructing new variables

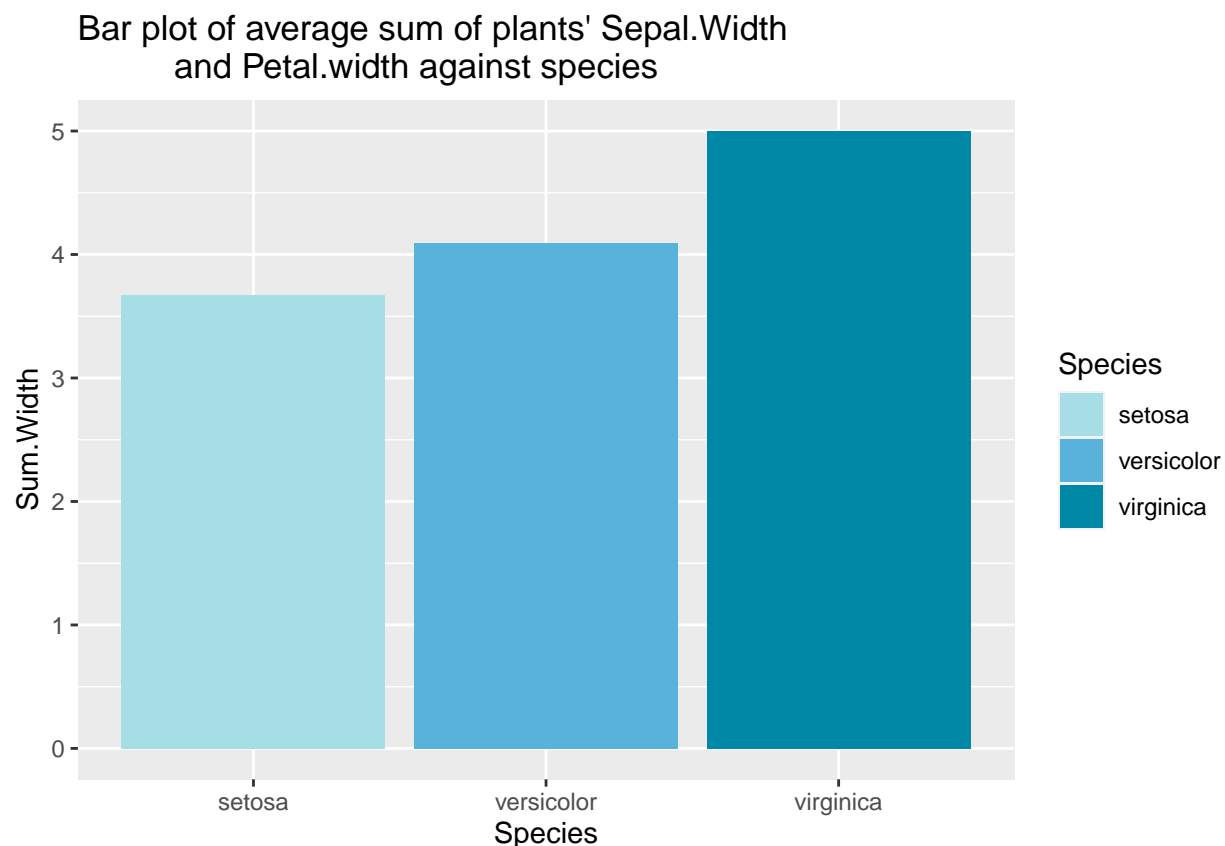
```
> iris$Sum.Width <- iris$Sepal.Width + iris$Petal.Width  
> iris$Sum.Length <- iris$Sepal.Length + iris$Petal.Length
```

Bar plot of Sum.Width against Species

```

> colorset3 <- c("#A5DEE4", "#58B2DC", "#0089A7")
> iris%>%
+   group_by(Species) %>%
+   summarise(n = n(), Sum.Width = mean(Sum.Width)) %>%
+   ggplot(aes(x = Species, y = Sum.Width, fill = Species)) + geom_bar(stat = "identity") +
+   ggtitle("Bar plot of average sum of plants' Sepal.Width
+           and Petal.width against species") +
+   scale_fill_manual(values = colorset3)

```

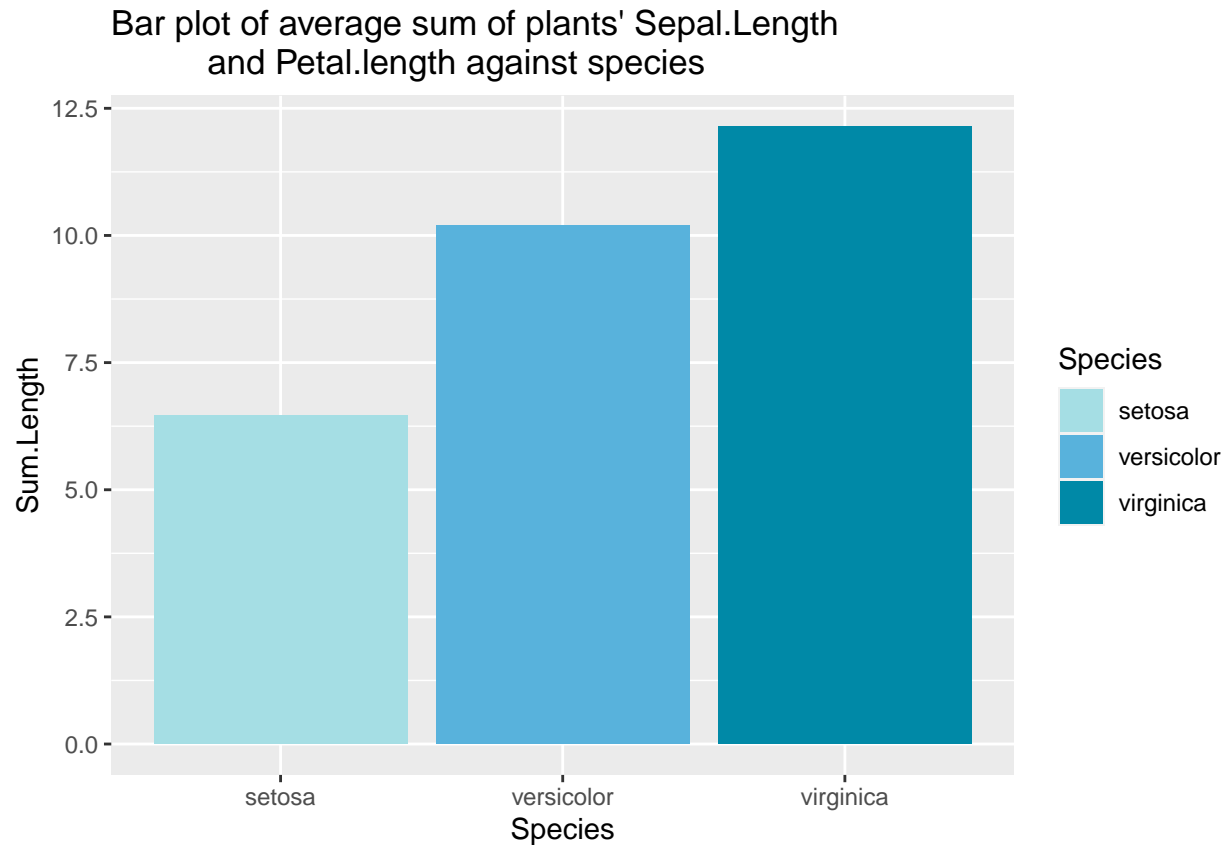


Bar plot of Sum.Length against Species

```

> iris%>%
+   group_by(Species) %>%
+   summarise(n = n(), Sum.Length = mean(Sum.Length)) %>%
+   ggplot(aes(x = Species, y = Sum.Length, fill = Species)) + geom_bar(stat = "identity") +
+   ggtitle("Bar plot of average sum of plants' Sepal.Length
+           and Petal.length against species") +
+   scale_fill_manual(values = colorset3)

```



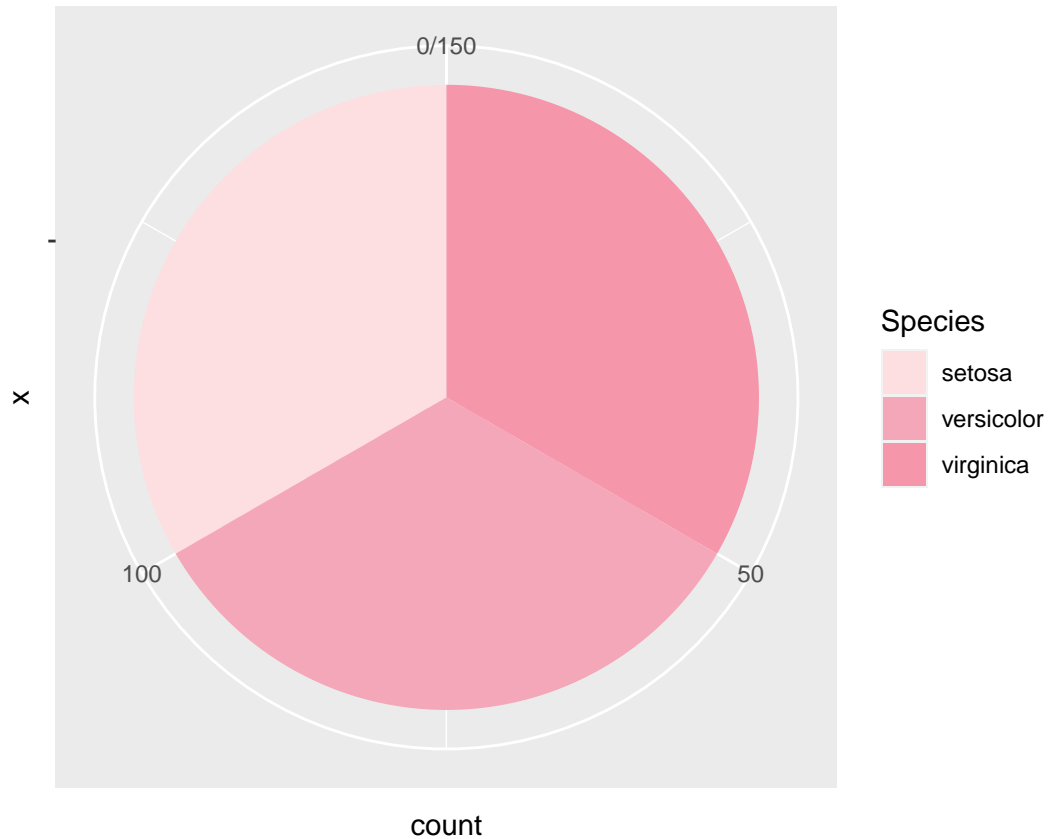
Pie plot of Species

I still don't know how to add percentage to pie plot without using ggpiestat()

```
> colorset4 <- c("#FEDFE1", "#F4A7B9", "#F596AA")
> options(digits = 4)
> iris.percent <- transform(as.data.frame(table(iris$Species)), percentage = Freq/nrow(iris)*100)
> iris.percent
```

```
##      Var1 Freq percentage
## 1  setosa   50      33.33
## 2 versicolor 50      33.33
## 3 virginica  50      33.33
```

```
> iris%>%
+   ggplot() +
+   geom_bar(aes(x = "", fill = Species), width = 1) +
+   coord_polar(theta = "y") +
+   scale_fill_manual(values = colorset4)
```



Merge two cols

```
> iris.new <- iris%>%
+   pivot_longer(c(Sepal.Length, Petal.Length), names_to = "Length.Type", values_to = "Length")
> iris.new <- iris.new%>%
+   pivot_longer(c(Sepal.Width, Petal.Width), names_to = "Width.Type", values_to = "Width")
> pander::pander(head(iris.new, n = 10))
```

Species	Sum.Width	Sum.Length	Length.Type	Length	Width.Type	Width
setosa	3.7	6.5	Sepal.Length	5.1	Sepal.Width	3.5
setosa	3.7	6.5	Sepal.Length	5.1	Petal.Width	0.2
setosa	3.7	6.5	Petal.Length	1.4	Sepal.Width	3.5
setosa	3.7	6.5	Petal.Length	1.4	Petal.Width	0.2
setosa	3.2	6.3	Sepal.Length	4.9	Sepal.Width	3
setosa	3.2	6.3	Sepal.Length	4.9	Petal.Width	0.2
setosa	3.2	6.3	Petal.Length	1.4	Sepal.Width	3
setosa	3.2	6.3	Petal.Length	1.4	Petal.Width	0.2
setosa	3.4	6	Sepal.Length	4.7	Sepal.Width	3.2
setosa	3.4	6	Sepal.Length	4.7	Petal.Width	0.2