```
#LIBRARY'S
library(tidyverse)
library(plotly)
library(data.table)
library(ggplot2)
library(maps)
library(dplyr)
library(tidyr)
library(lubridate)
```

Iris Dataset

> "The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by the British statistician and biologist Ronald Fisher in his 1936 paper The use of multiple measurements in taxonomic problems as an example of linear discriminant analysis" https://en.wikipedia.org/wiki/Iris_flower_data_set (https://en.wikipedia.org/wiki/Iris_flower_data_set)

```
#  Read the iris.csv file  (2 points)
data_iris = fread("iris.csv")
```

```
# Show some values from data frame (2 points)
head(data_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
```
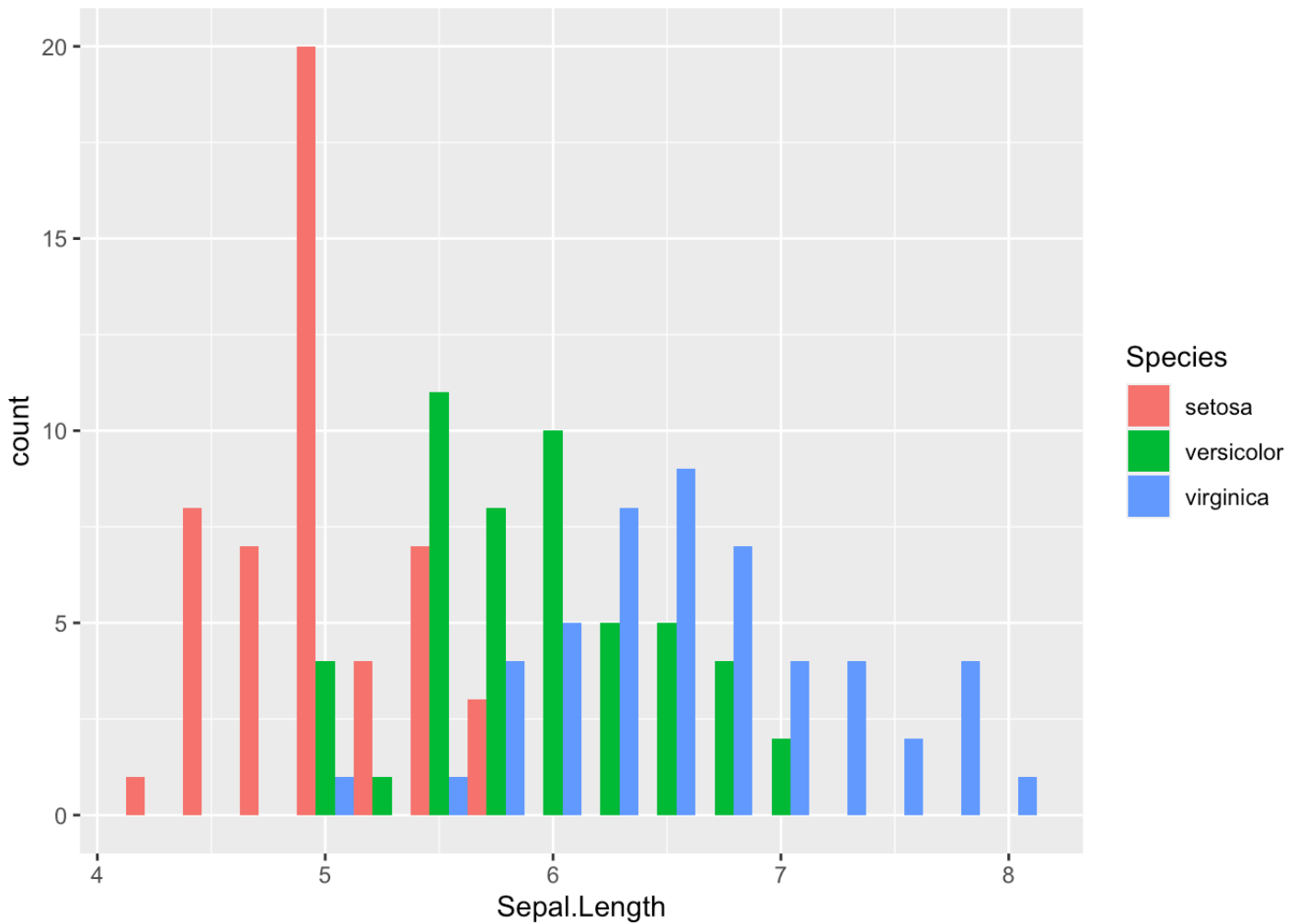
```
hist_iris.1 = plot_ly(data_iris, x = ~Sepal.Length, color = ~Species, type = "histogr
am")
hist_iris.1
```

25

setosa
versicolor
virginica
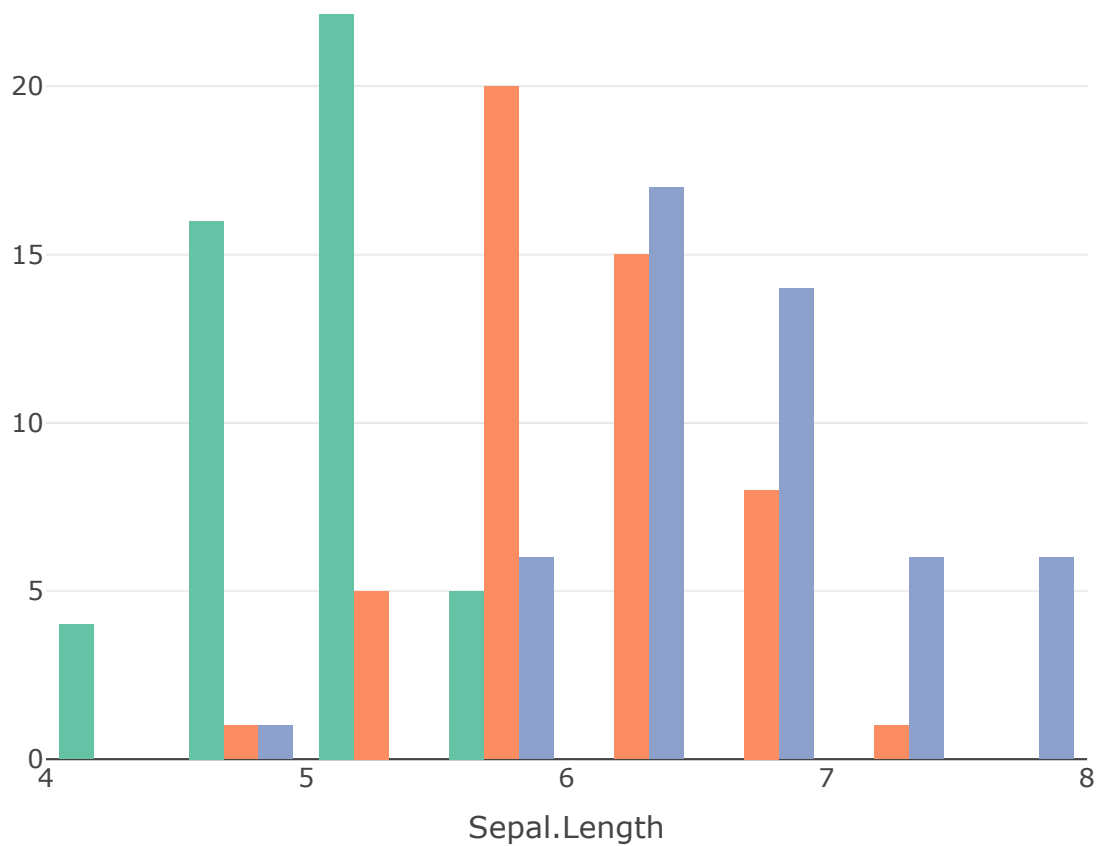
```
ggplot_hist_iris = ggplot(data_iris, aes(x = Sepal.Length, fill = Species)) +
  geom_histogram(position = "dodge",binwidth = 0.25)

ggplot_hist_iris
```
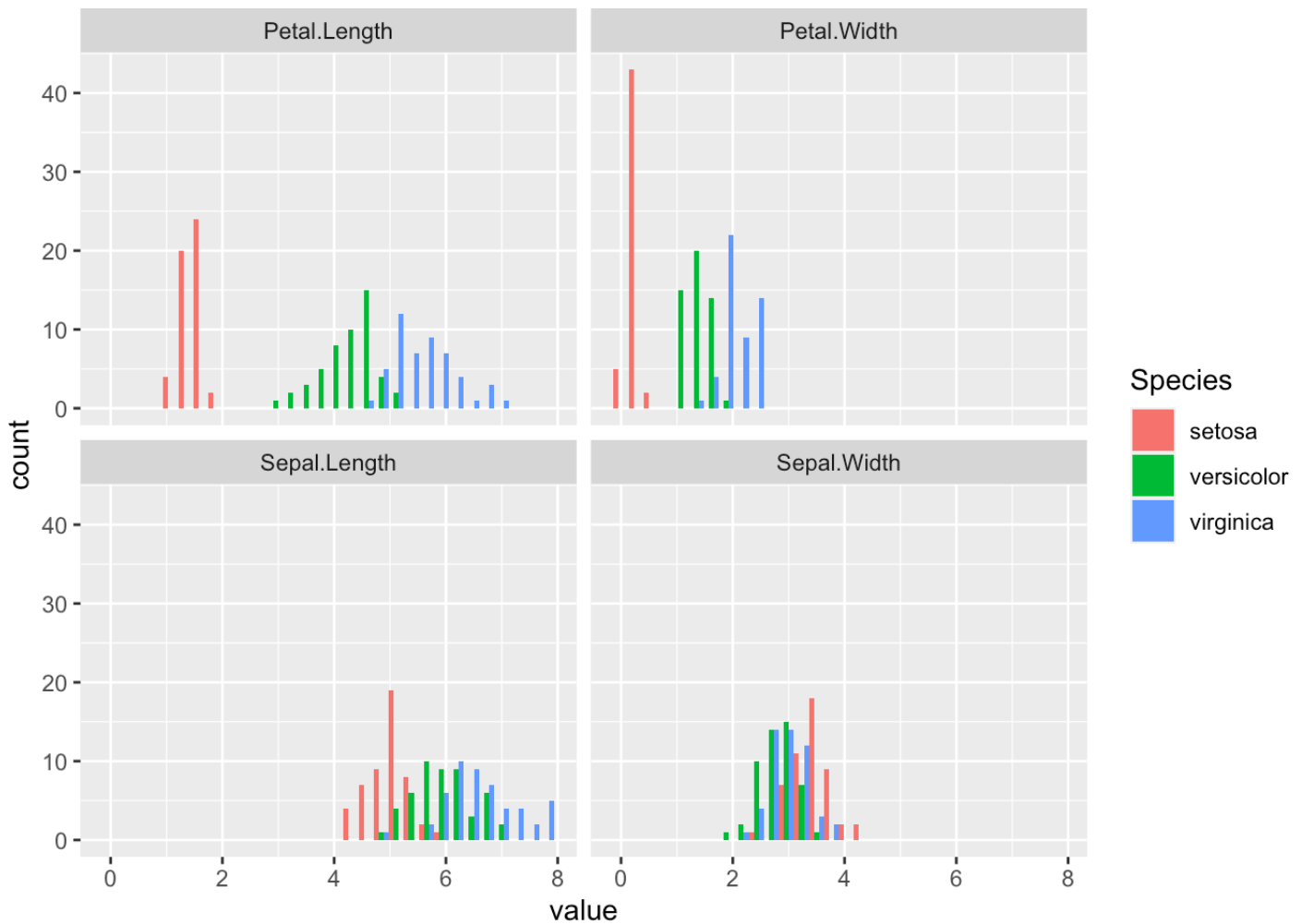
```
#Formating the dataframe
data_iris.new_format = data_iris %>%
  pivot_longer(-Species, names_to = "Metric", values_to = "value")

#Plotting
ggplot(data_iris.new_format, aes(x=value, fill = Species)) + facet_wrap(factor(data_i
ris.new_format$Metric)) +
  geom_histogram(position = "dodge")
```
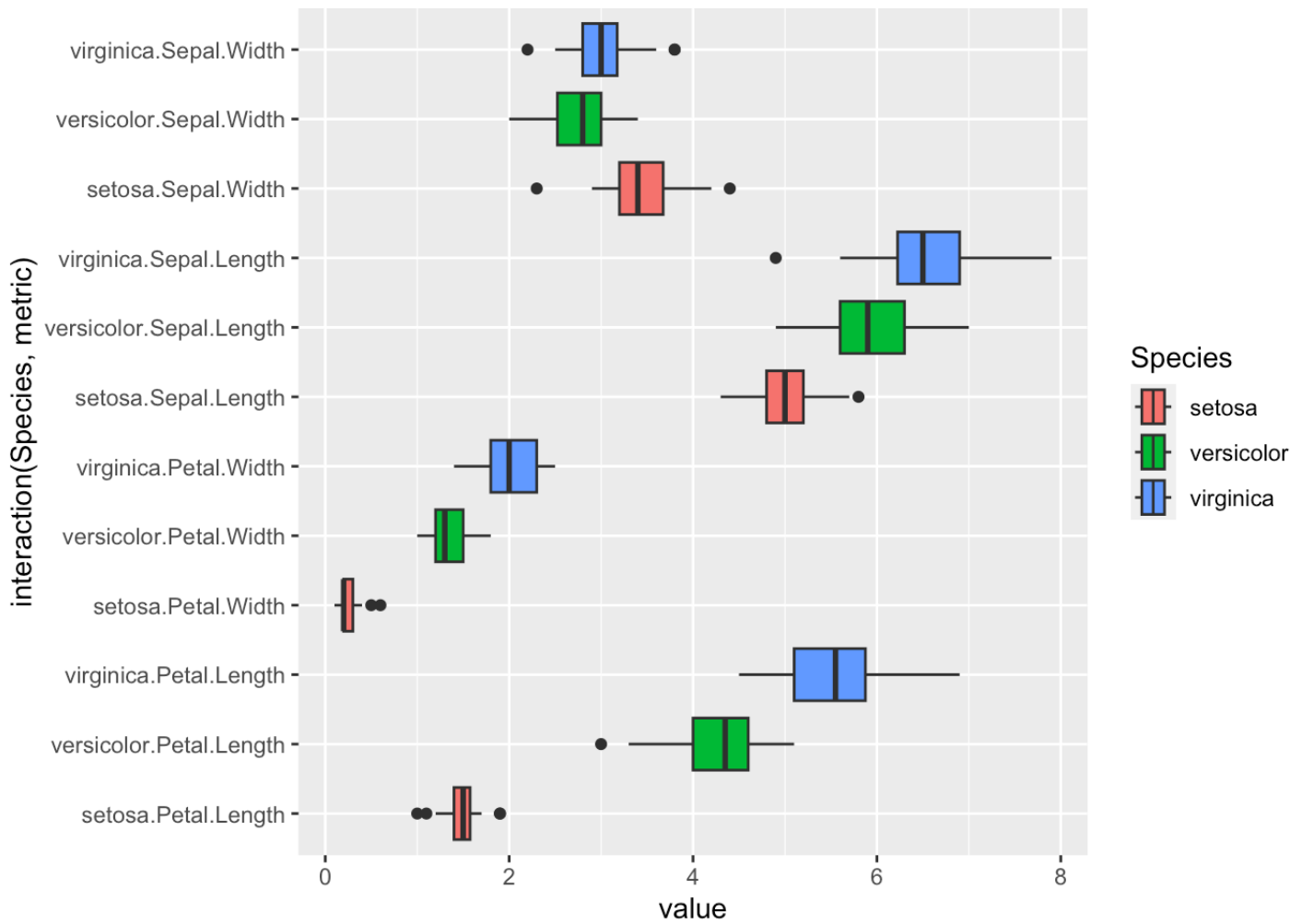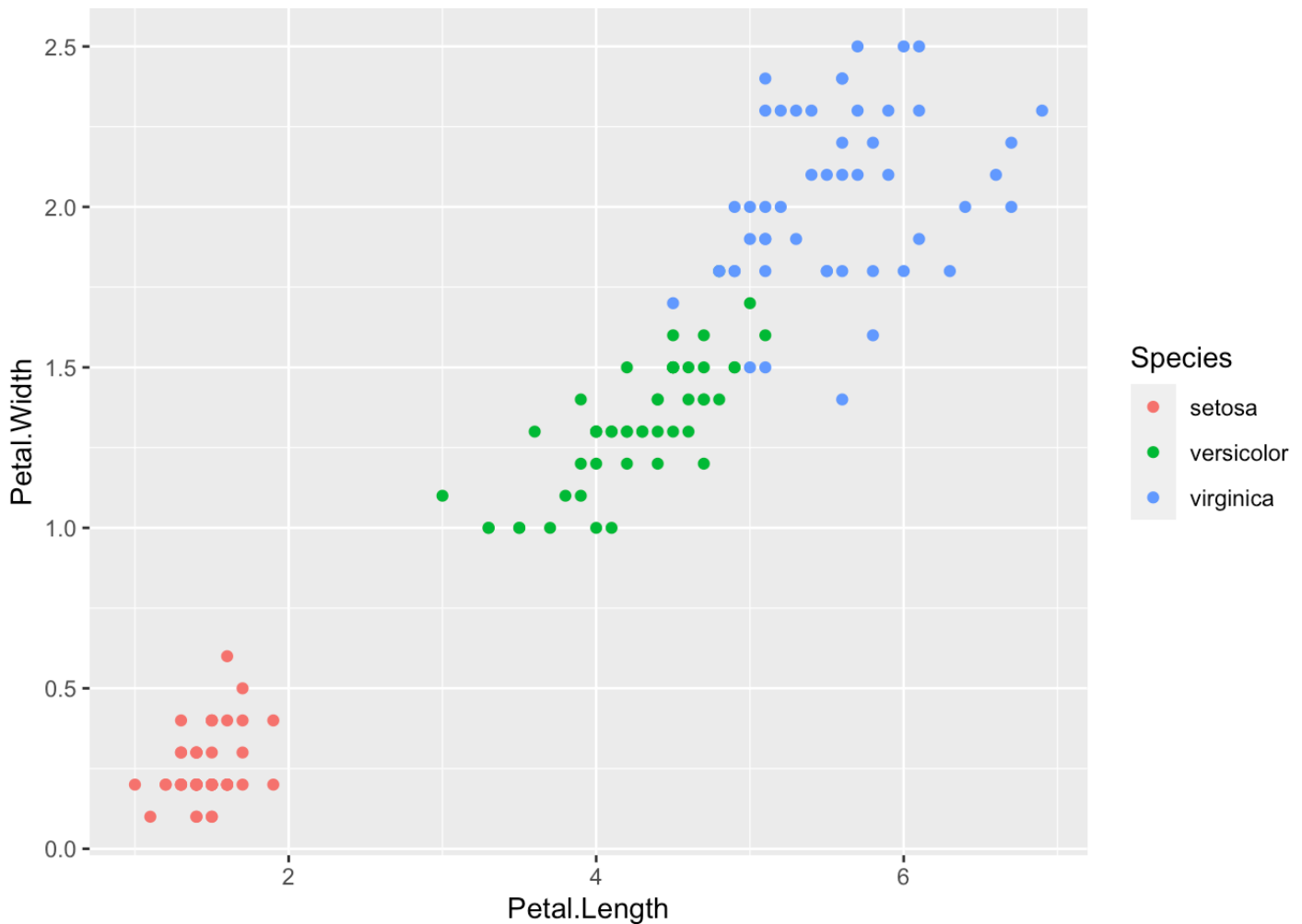
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
data_iris.new_format$interact_feild = interaction(data_iris.new_format$Species, data_
iris.new_format$Metric, sep="." )
ggplot(data_iris.new_format, aes(x = value, fill = Species, y = interact_feild))+ geo
m_boxplot()+ labs(y = "interaction(Species, metric)", x = "value")
```

```
ggplot(data_iris, aes(x = Petal.Length, color = Species, y = Petal.Width))+
    geom_point()
```

```
plot_ly(data_iris, x = ~Petal.Length,y = ~Petal.Width, z = ~Sepal.Length, color = ~Sp
ecies)
```

```
## No trace type specified:
##    Based on info supplied, a 'scatter3d' trace seems appropriate.
##    Read more about this trace type -> https://plotly.com/r/reference/#scatter3d
```
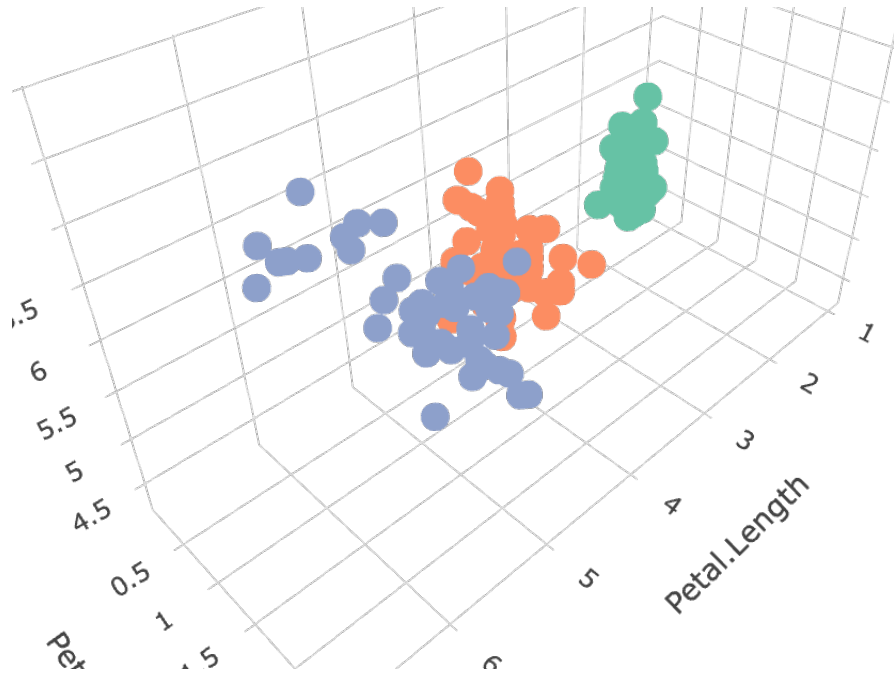
```
## No scatter3d mode specifed:
##    Setting the mode to markers
##    Read more about this attribute -> https://plotly.com/r/reference/#scatter-mode
```

The Iris data set helps to distinguish between the species by its features like Petal.Length,Petal.Width,Sepal.Length,Sepal.width. By above procedure we can clearly see that petal length classifies the data set well and sepal width classify the least. Even if we check the co-effients of each features it confirms the same. To conclude, the 4 features of flower helps us the classify the dataset into setosa, versicolor and virginca clearly.