



APPLIED STATISTICS PROJECT

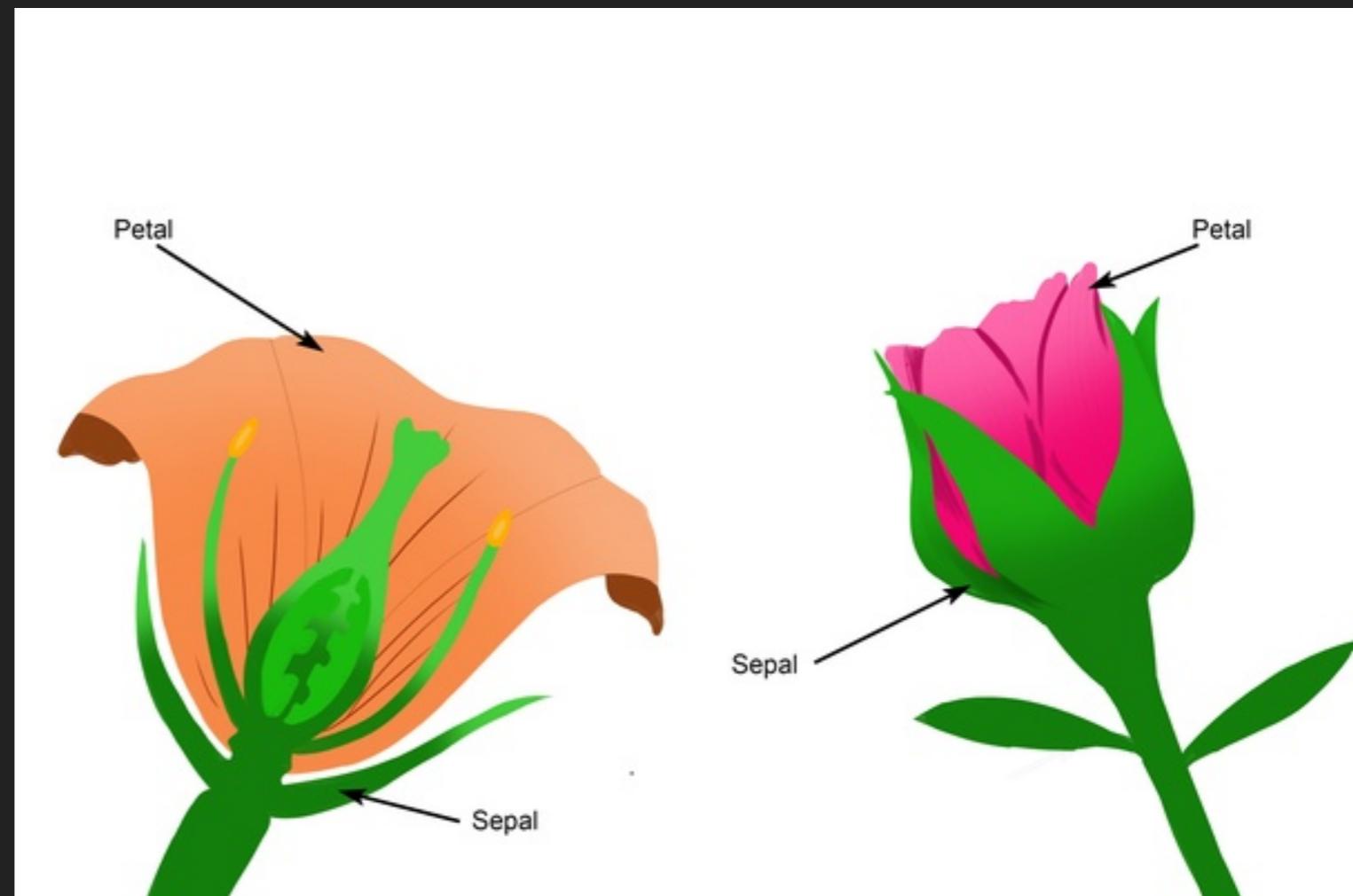
DE COMBARIEU MARTIN

GUEGNOLLE MAXENCE

IRIS DATASET ANALYSIS

FEATURES PRESENTATION

- ▶ Sepal.Length
- ▶ Sepal.Width
- ▶ Petal.Length
- ▶ Petal.Width



FEATURES PRESENTATION

- ▶ Sepal.Length
- ▶ Sepal.Width
- ▶ Petal.Length
- ▶ Petal.Width

Centimeters

0.1 → 7.9

FEATURES PRESENTATION

Versicolor



Setosa



Virginica



50

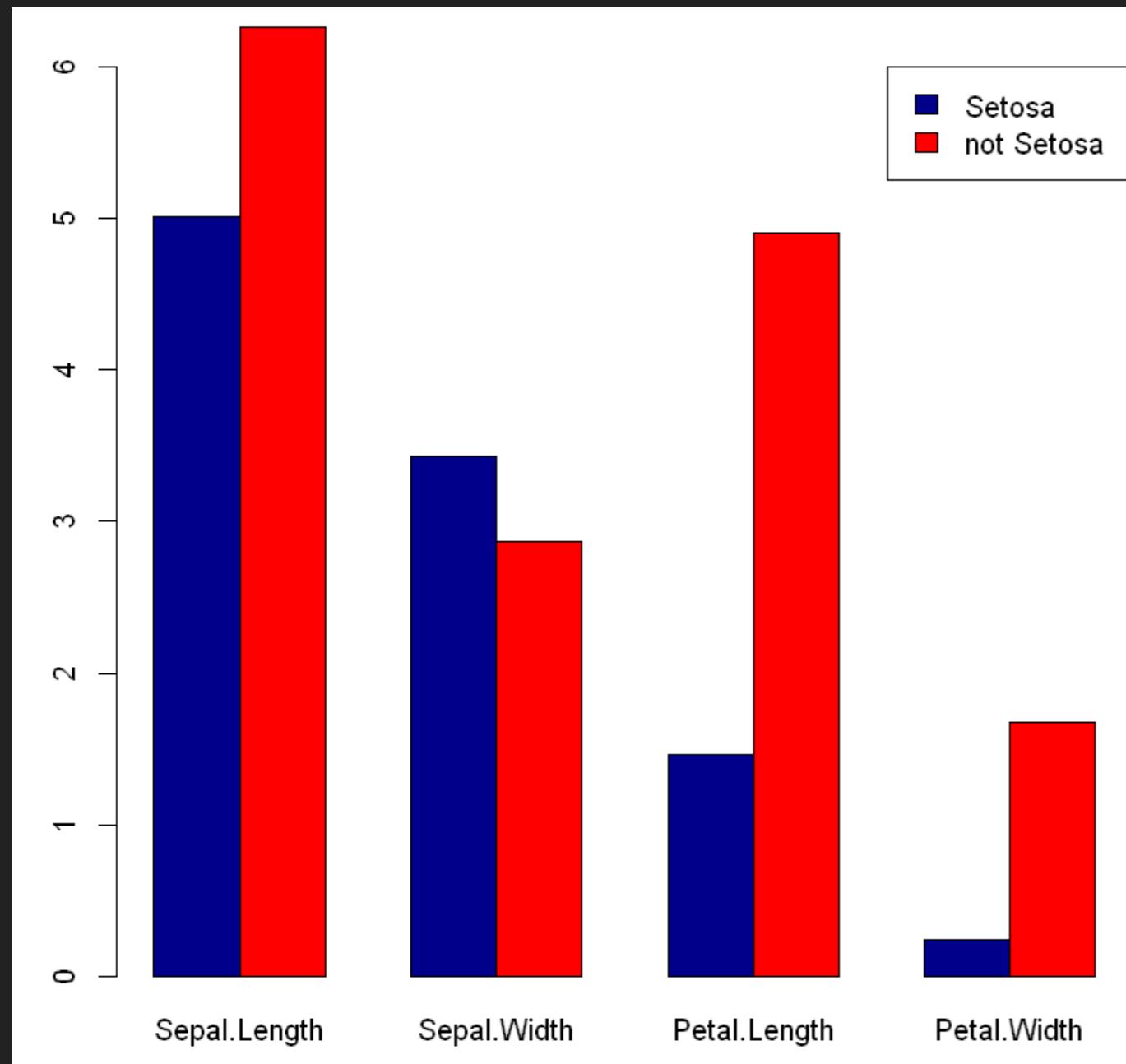
50

50



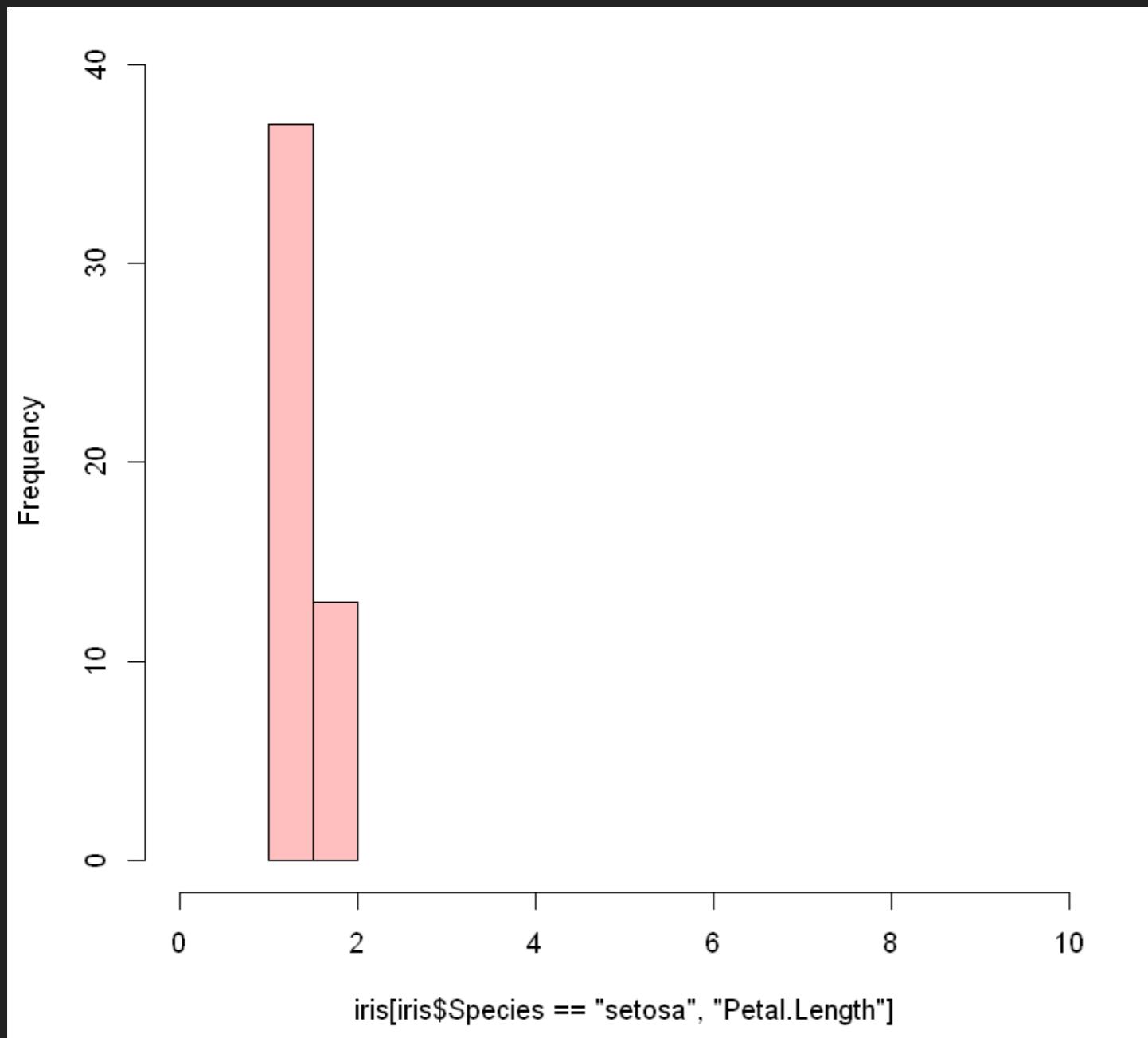
CAN WE PREDICT
THE SPECIE OF AN
IRIS BASED ON
THOSE
FEATURES ?

SETOSA ANALYSIS



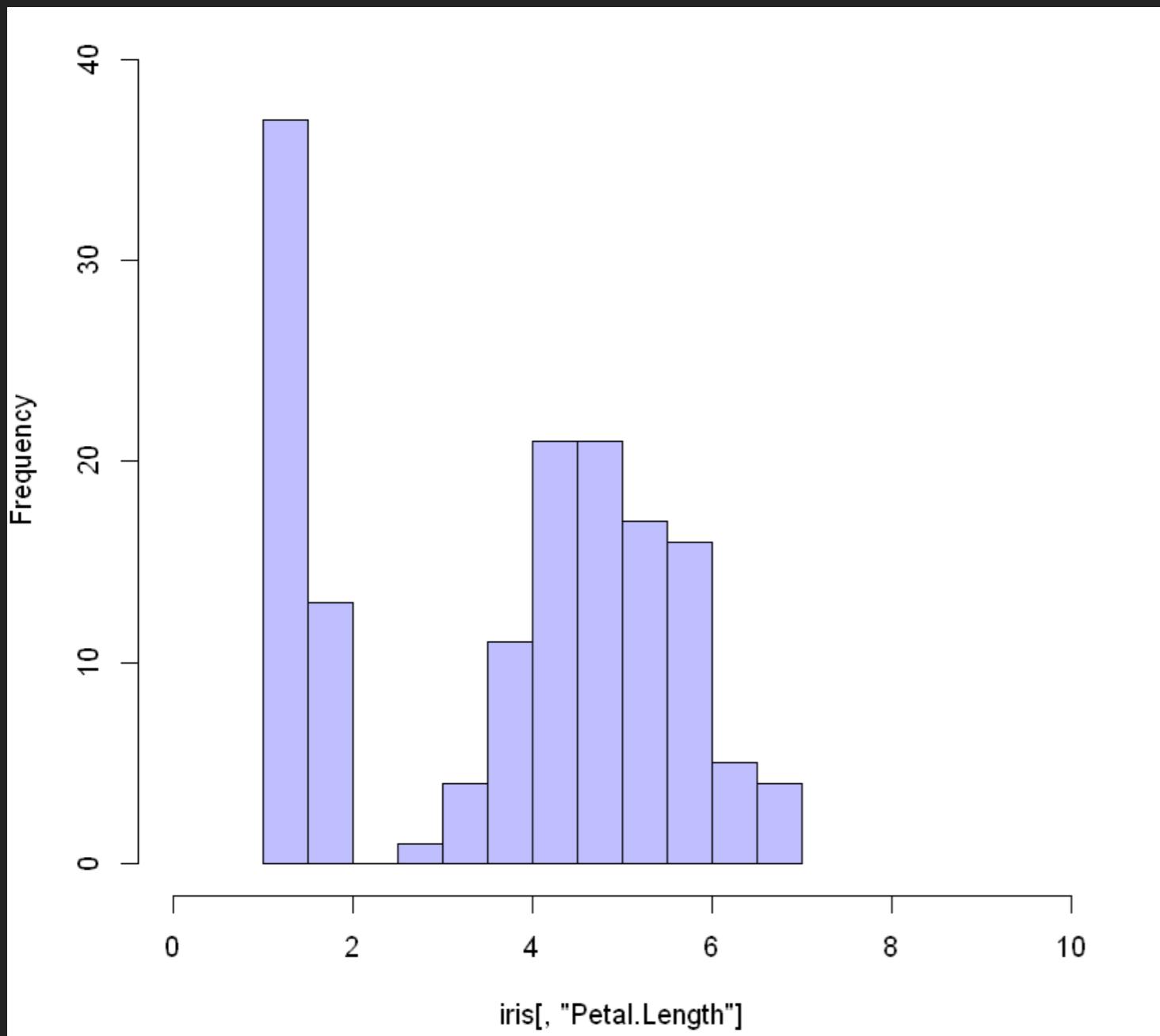
Setosa's Petal lenght

SETOSA ANALYSIS



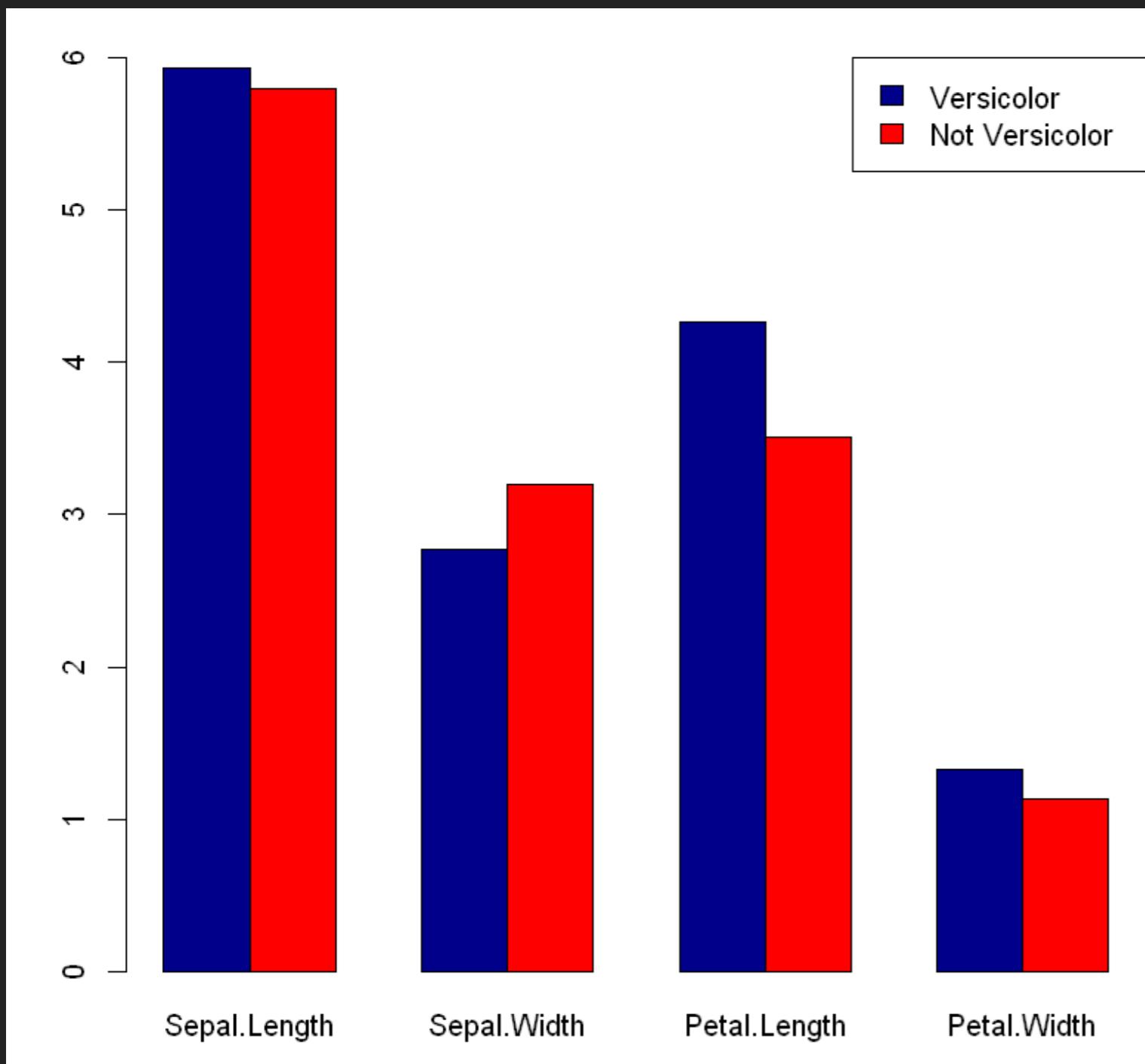
Setosa petals
are the smallest

SETOSA ANALYSIS



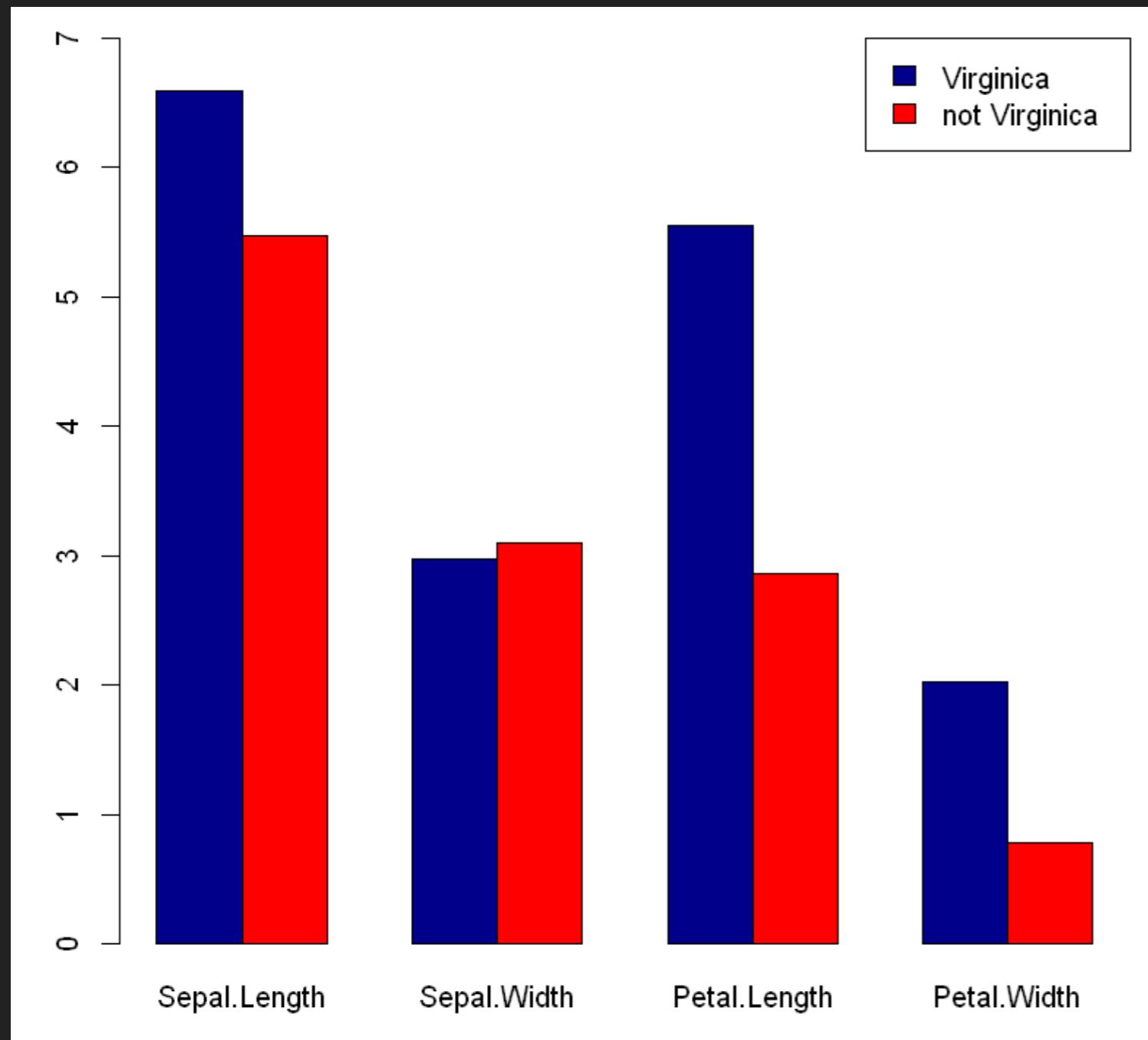
Setosa petals
are the smallest

VERSICOLOR ANALYSIS



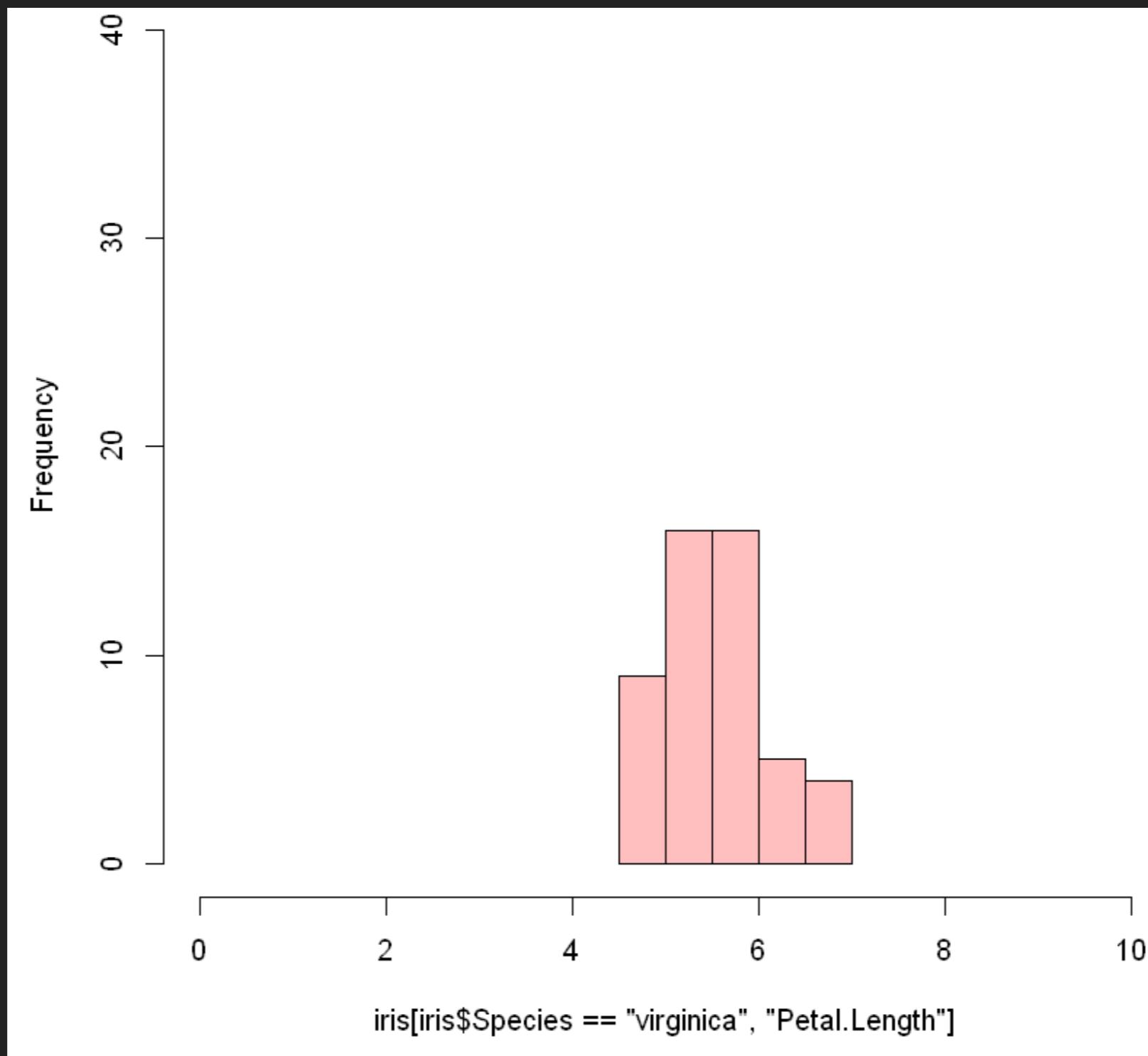
Seem in the average

VIRGINICA ANALYSIS

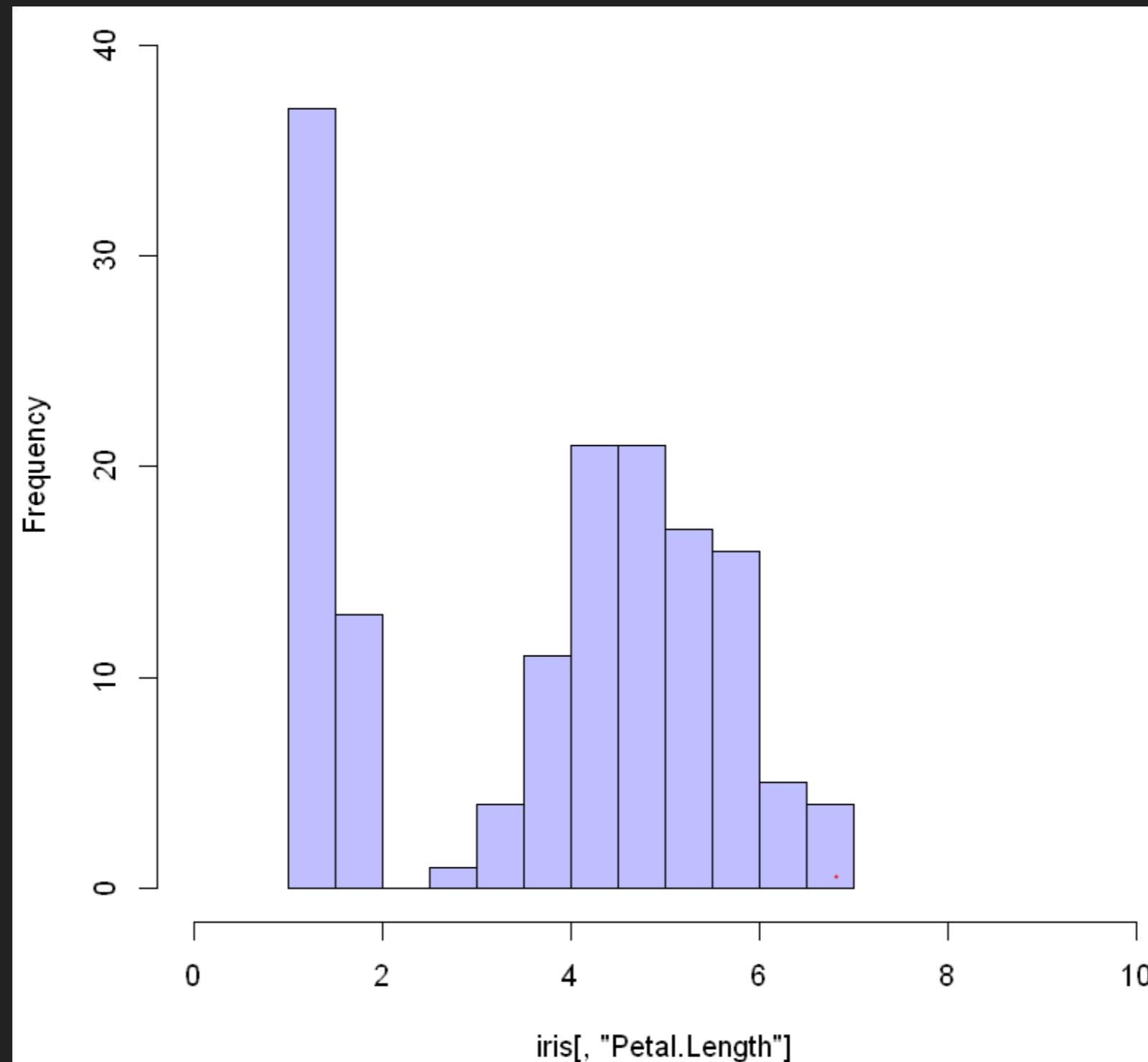


Virginica's petal lenght

VIRGINICA ANALYSIS



VIRGINICA ANALYSIS



VIRGINICA ANALYSIS



```
virginicagt5 <- iris[(iris$Species=="virginica") & (iris$Petal.Length >= 5.0), ]  
globalgt5 <- iris[(iris$Petal.Length >= 5.0), ]  
nrow(virginicagt5)/nrow(globalgt5)*100
```

95% !!

COMPARAISON



```
versicolorTest <- iris[(iris$Species=="versicolor") & (iris$Petal.Length <= 5.0) & (iris$Petal.Length >= 2.0), ]  
globalTest <- iris[(iris$Petal.Length <= 5.0) & (iris$Petal.Length >= 2.0), ]  
nrow(versicolorTest)/nrow(globalTest)*100
```

84% !!

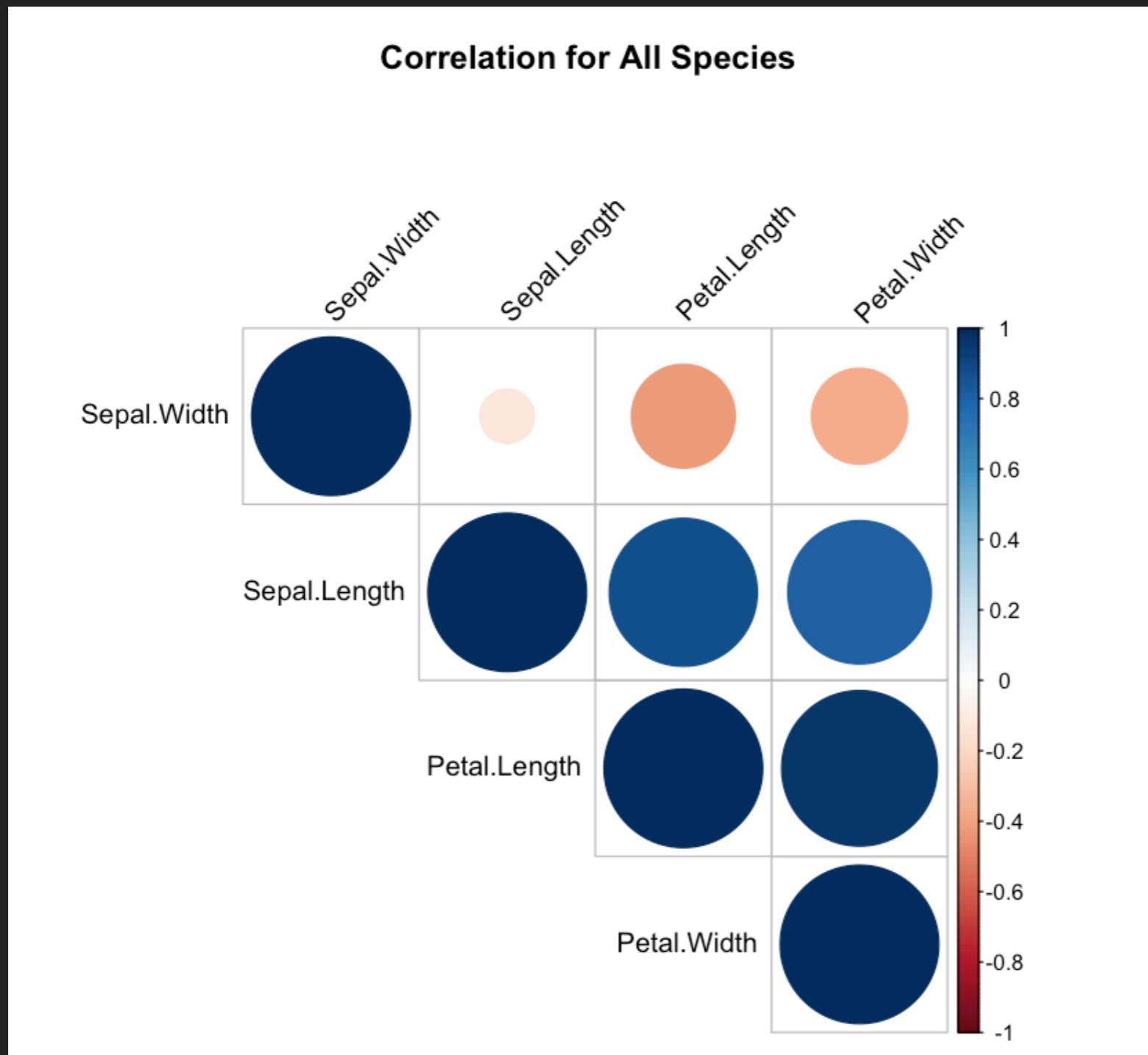
COMPARAISON



```
virginicaTest <- iris[(iris$Species=="virginica") & (iris$Petal.Length <= 5.0) & (iris$Petal.Length >= 2.0), ]
```

16% !!

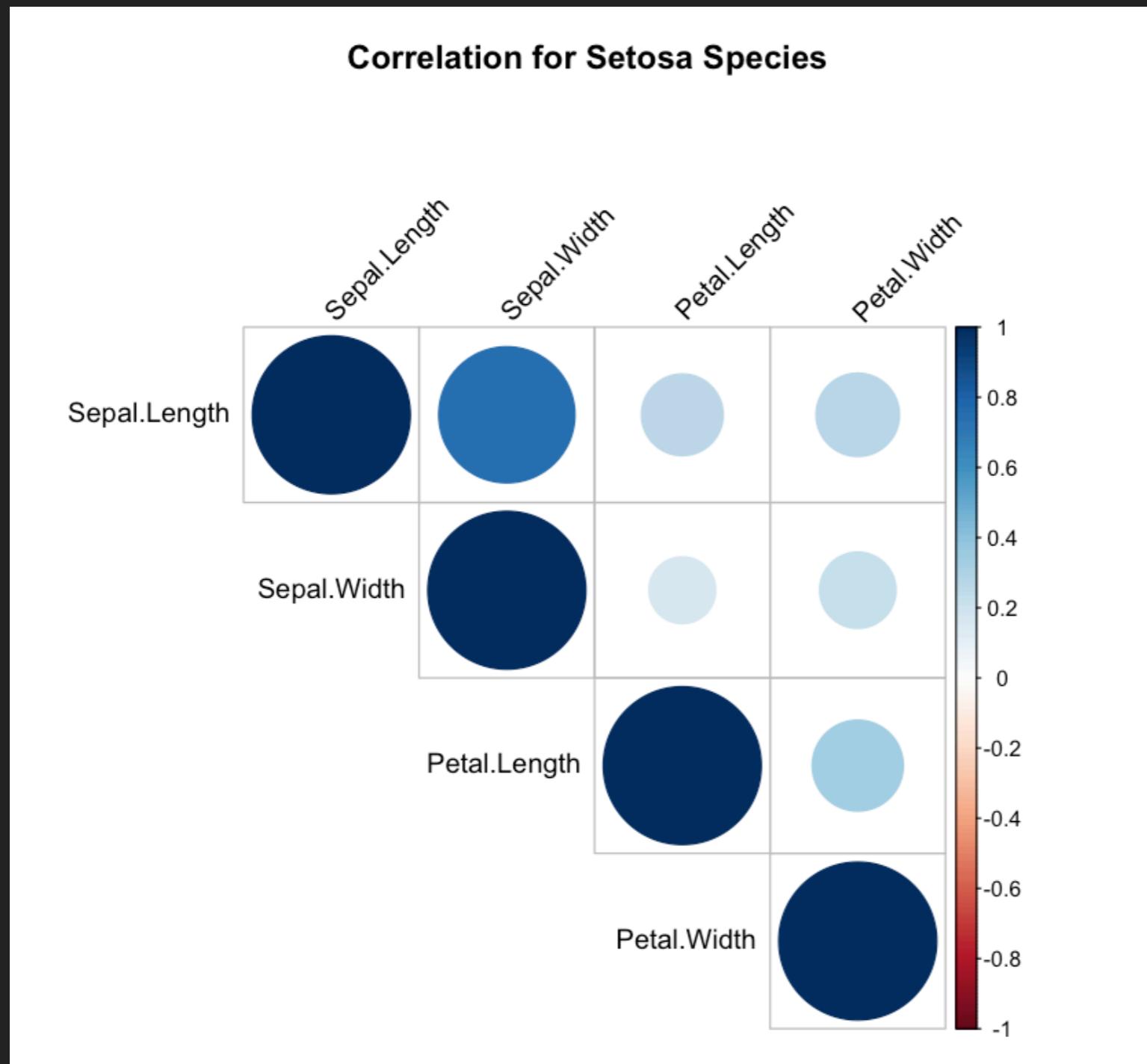
CORRELATION FOR ALL SPECIES



Petals - Sepals

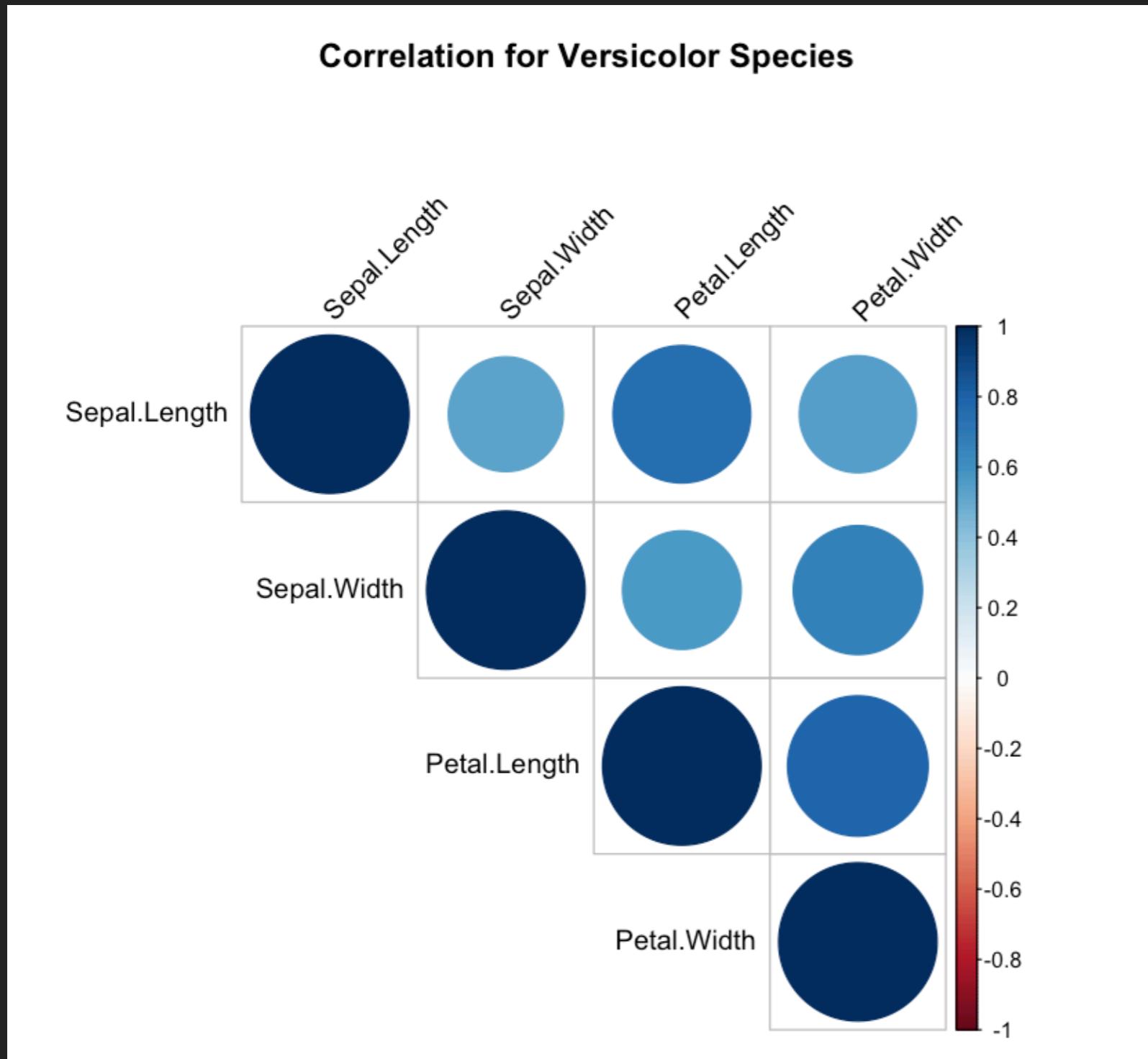
Petals - Petals

CORRELATION FOR SETOSA SPECIES



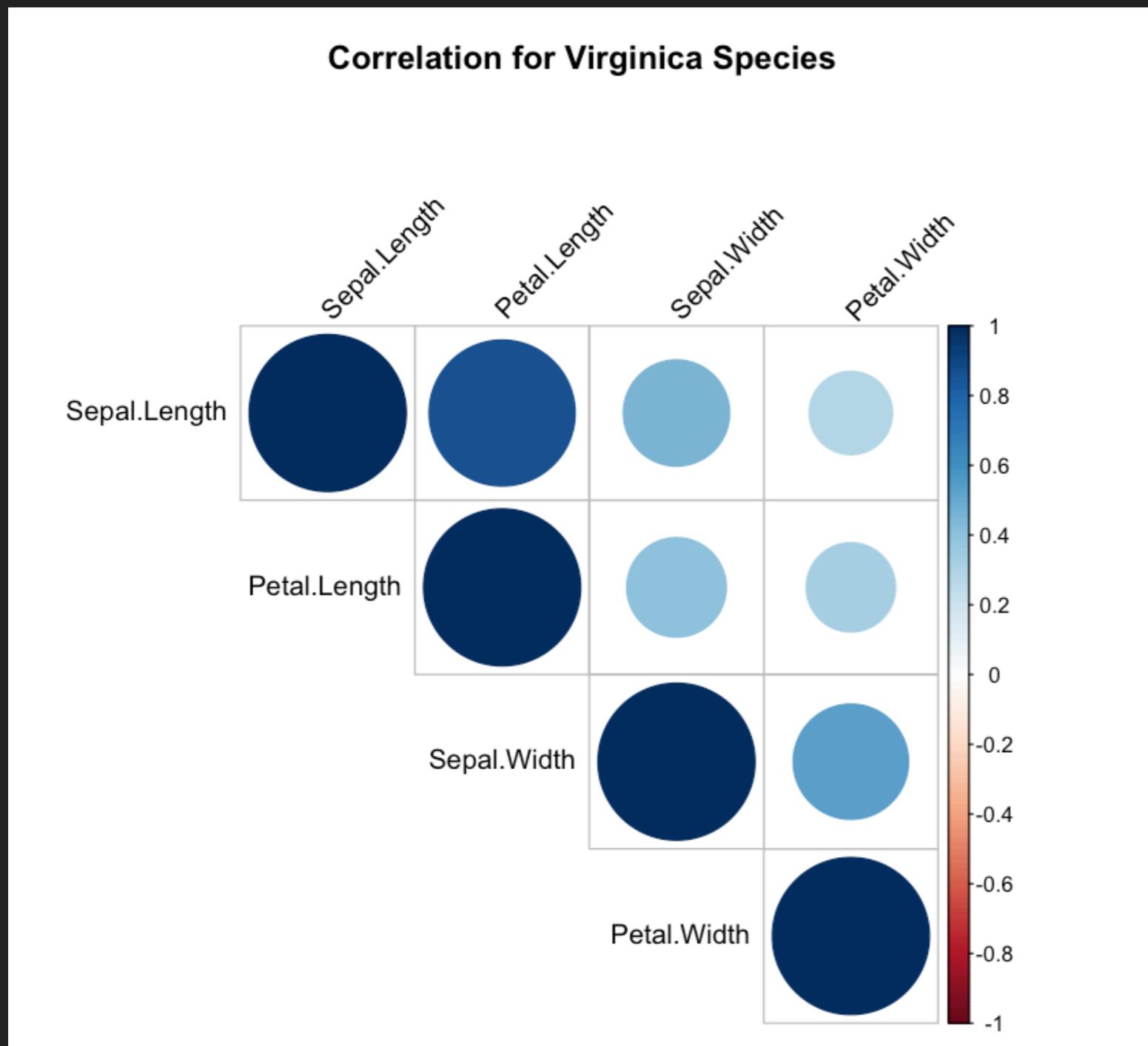
Speal L - Sepal W

CORRELATION FOR VERSICOLOR SPECIES



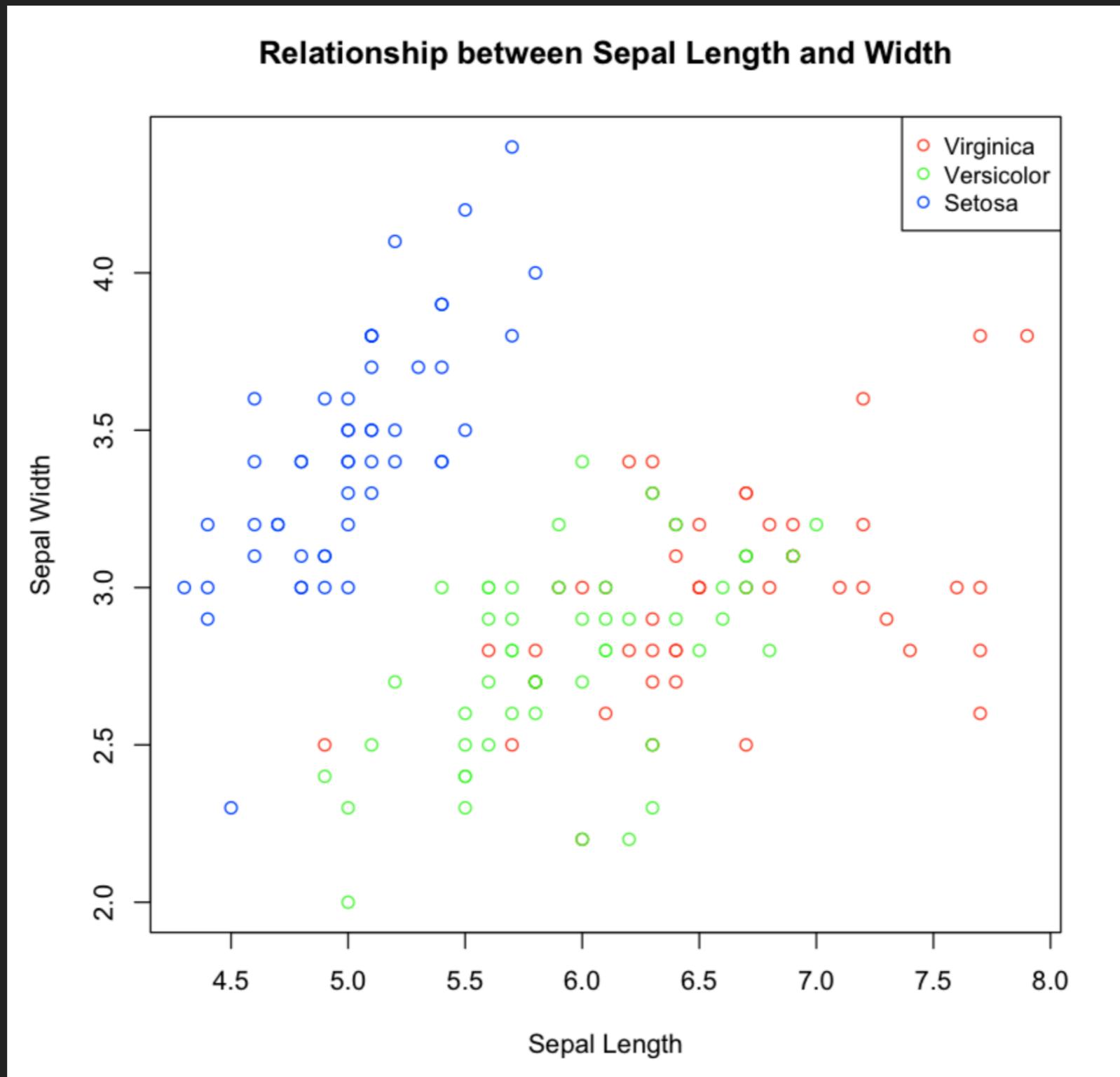
Global correlation
Petal L - Petal W
Sepal L - Petal L

CORRELATION FOR VIRGINICA SPECIES

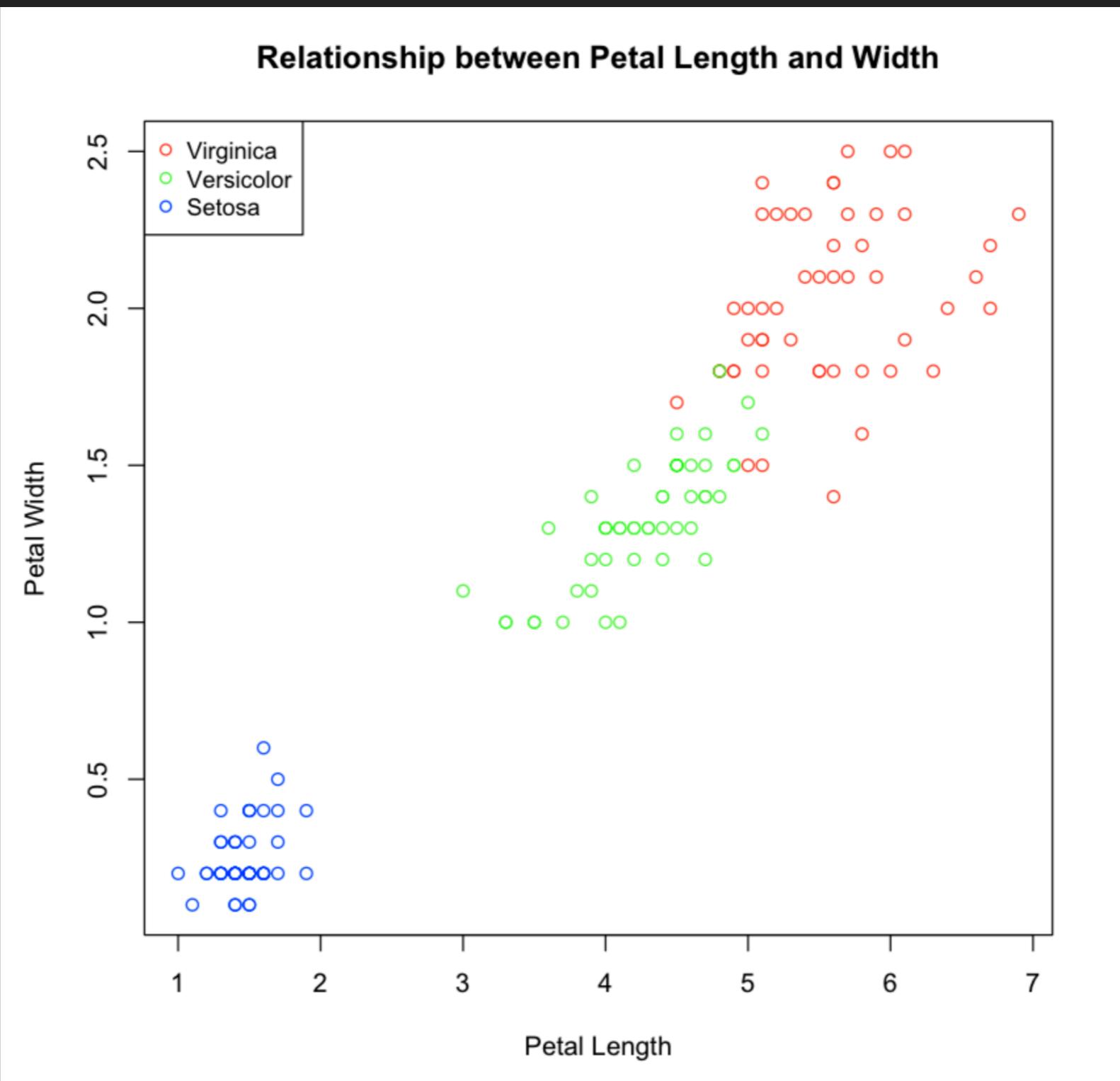


Sepal L - Petal L

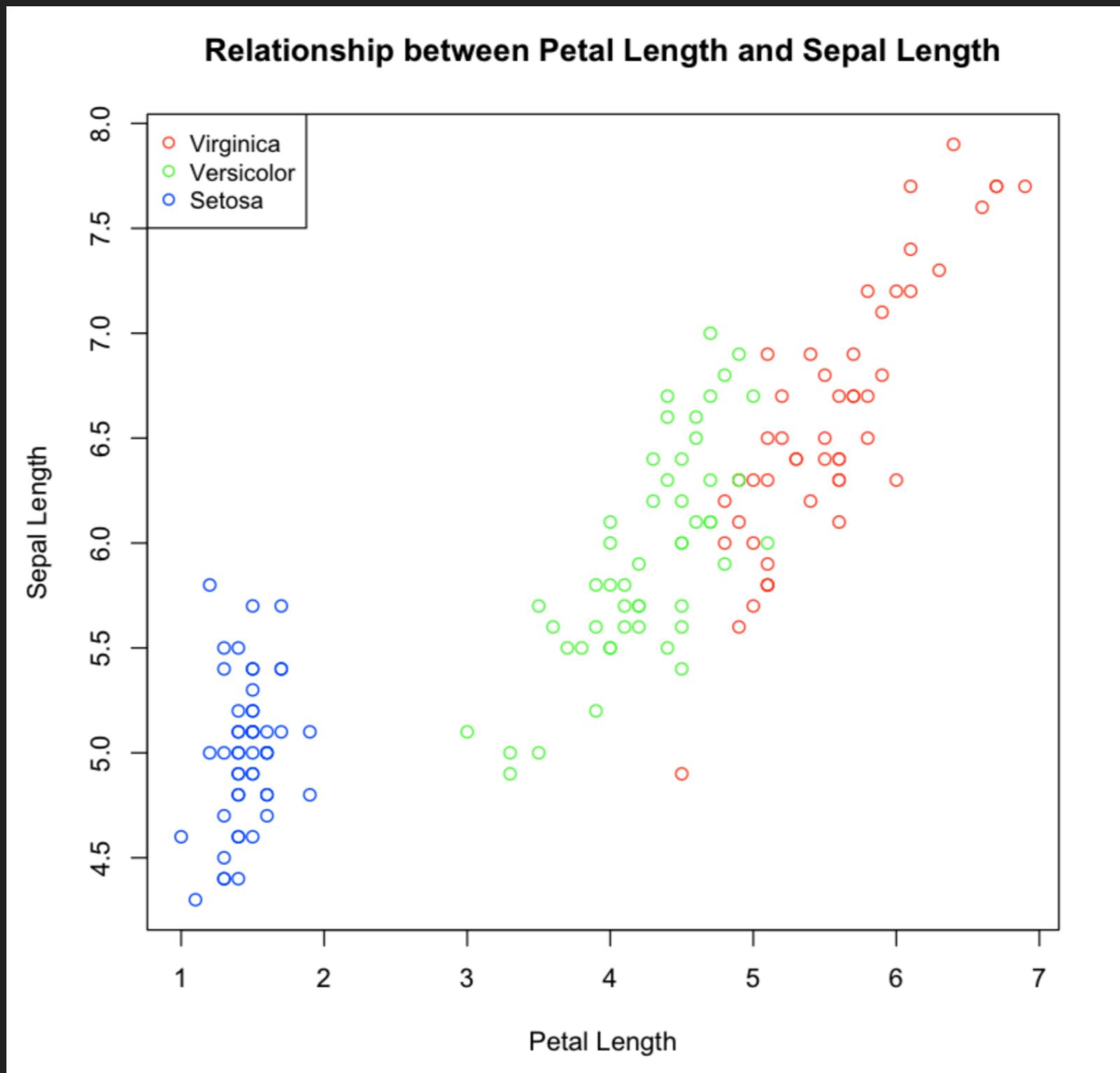
VISUALIZING THE CORRELATION - SEPAL LENGTH AND WIDTH



RELATIONSHIP BETWEEN PETAL LENGTH AND WIDTH



RELATIONSHIP BETWEEN PETAL LENGTH AND SEPAL LENGTH



CORRELATION

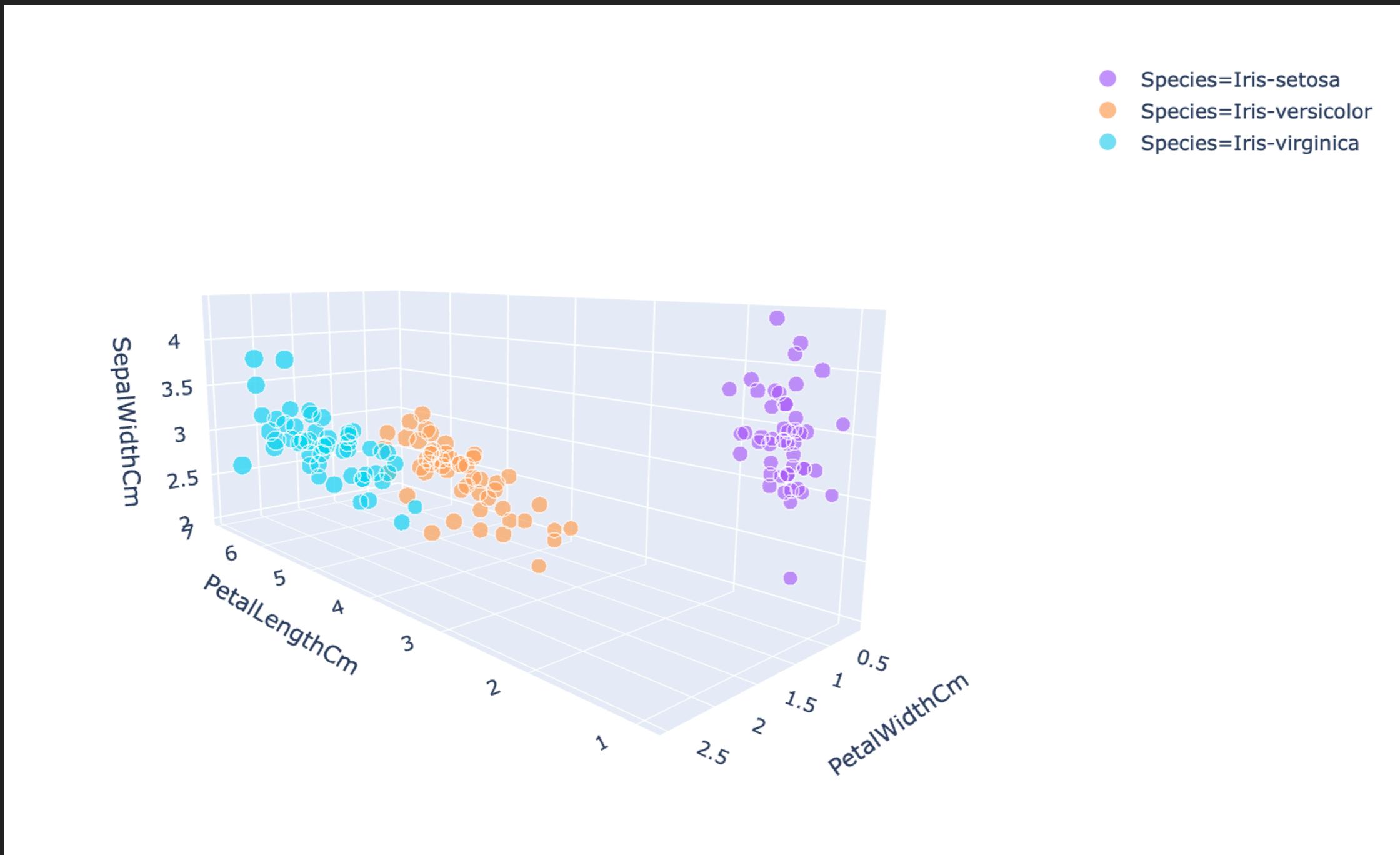
DISCRIMINATIVE CLASSES



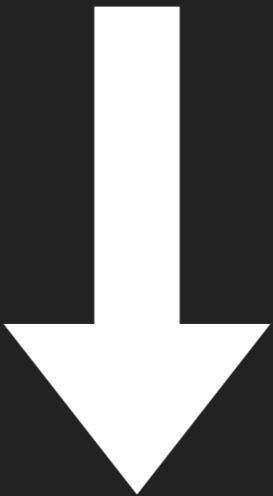
REGRESSION FOR PREDICTION

APPLIED STATISTICS PROJECT

VISUALIZATION



SCALING K NEIGHBORS CLASSIFIER



THE ACCURACY IS AT 95.24% ON TRAINING DATA

THE ACCURACY IS AT 100.00% ON TEST DATA