Week 6 (6.2 Exercise)

Maxim Bilenkin

2025-01-14

Iris Dataset Analysis.

```
# Loading datasets and necessary Iris packages.
library(datasets)
data(iris)
```

```
# Creating and assigning iris dataset to the variable.
iris_dataset <- iris
head(iris_dataset)
```

```
##   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
```

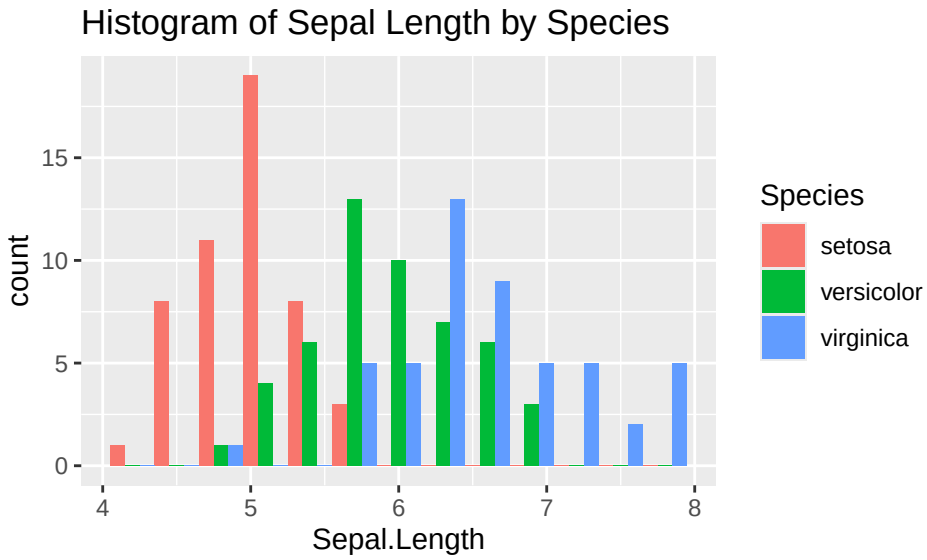
Average Sepal Length by Species.

```
# Loading dplyr package and suppressing all the startup messages.
suppressPackageStartupMessages(library(dplyr))
library(knitr)

# Grouping all the data by species and calculating the average sepal length.
avg_length_of_sepal <- iris_dataset %>%
    group_by(Species) %>%
    summarize(Average_Sepal_Length = mean(Sepal.Length))

# Printing all the calculated results.
kable(avg_length_of_sepal, col.names = c("Species", "Average Sepal Length"))
```

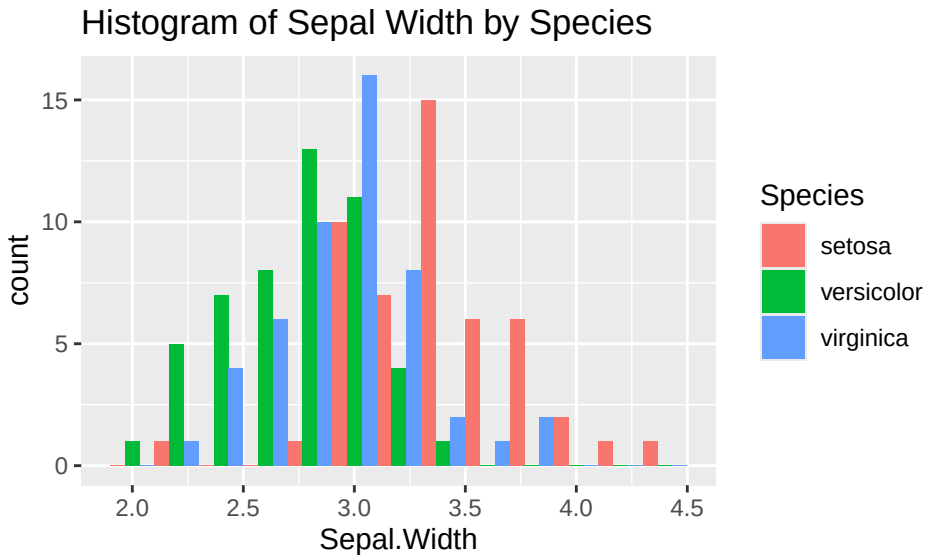
Species	Average Sepal Length
setosa	5.006
versicolor	5.936
virginica	6.588

Visualizations

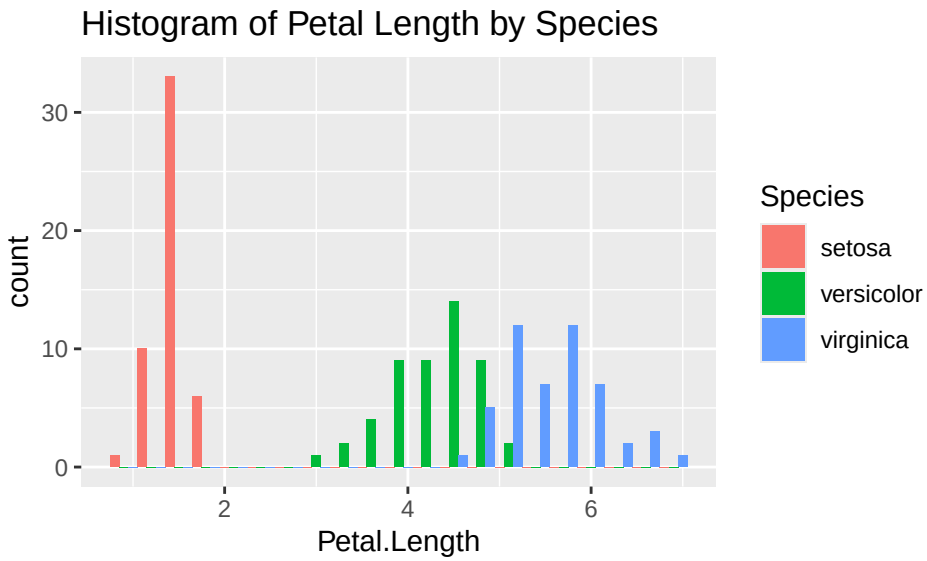
```
# Creating Histograms for Sepal Length, Sepal Width, Petal Length,
# and Petal Width by Species.
ggplot(iris_dataset, aes(x=Sepal.Length, fill=Species)) +
  geom_histogram(binwidth=0.3, position="dodge") +
  ggtitle("Histogram of Sepal Length by Species")
```



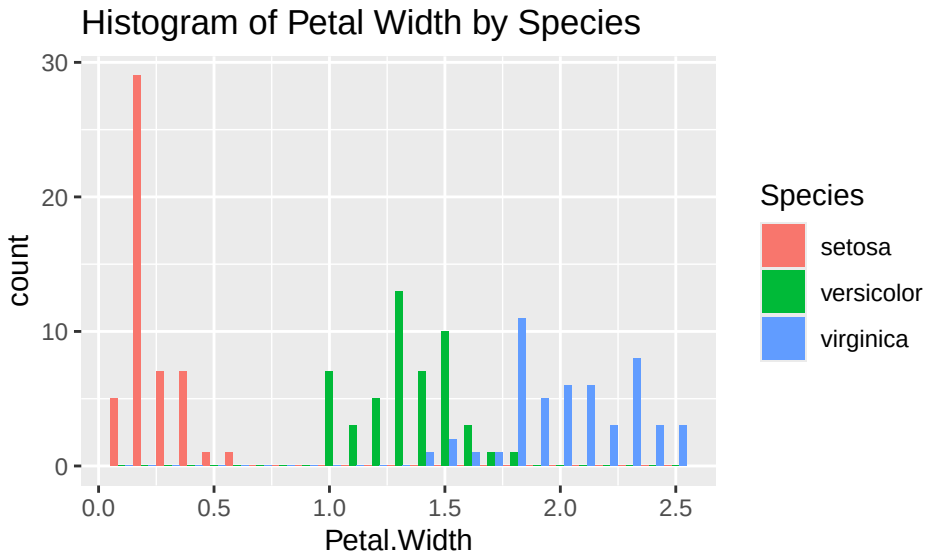

```
ggplot(iris_dataset, aes(x=Sepal.Width, fill=Species)) +
  geom_histogram(binwidth=0.2, position="dodge") +
  ggtitle("Histogram of Sepal Width by Species")
```



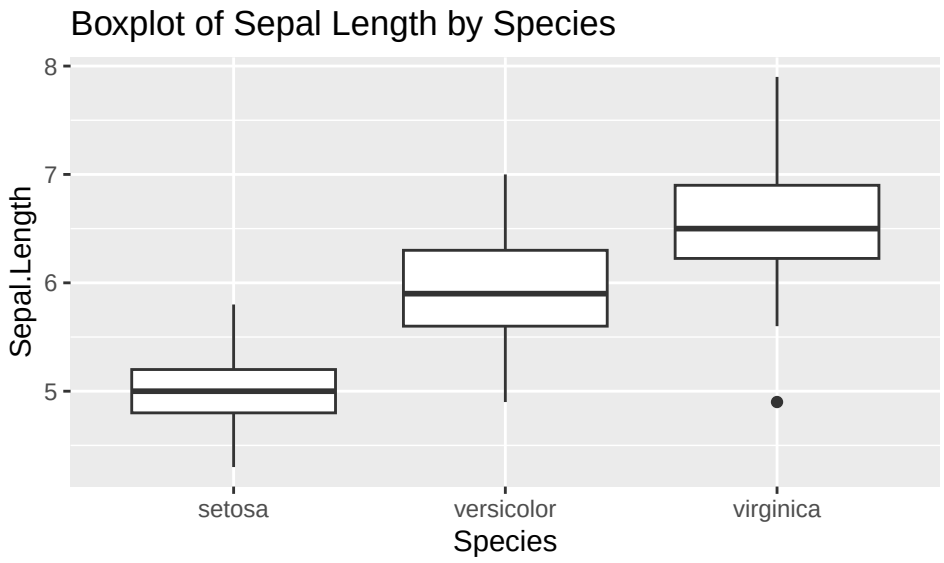

```
ggplot(iris_dataset, aes(x=Petal.Length, fill=Species)) +
  geom_histogram(binwidth=0.3, position="dodge") +
  ggtitle("Histogram of Petal Length by Species")
```



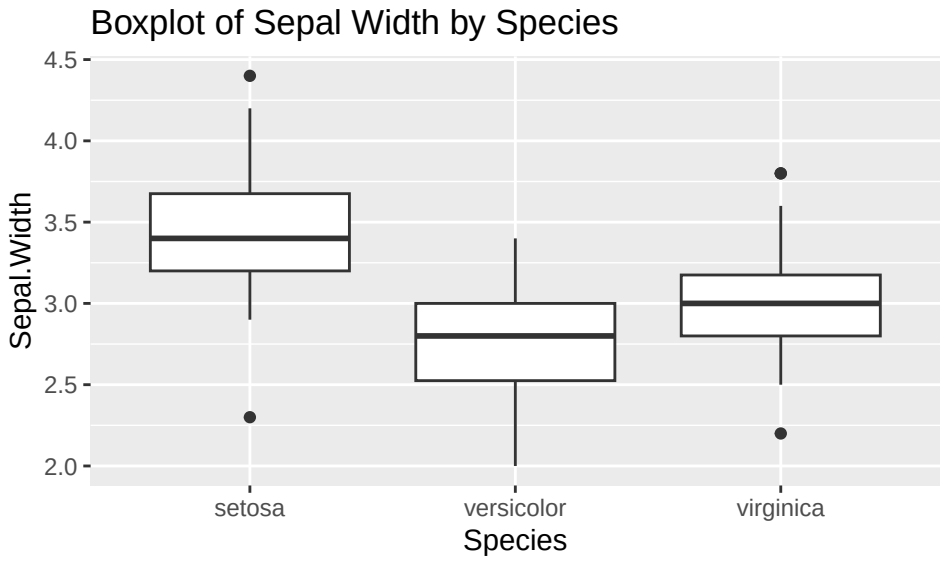

```
ggplot(iris_dataset, aes(x=Petal.Width, fill=Species)) +
  geom_histogram(binwidth=0.1, position="dodge") +
  ggtitle("Histogram of Petal Width by Species")
```




```
# Creating Boxplots for Sepal Length, Sepal Width, Petal Length, and Petal Width
# by Species
ggplot(iris_dataset, aes(x=Species, y=Sepal.Length)) +
  geom_boxplot() +
  ggtitle("Boxplot of Sepal Length by Species")
```




```
ggplot(iris_dataset, aes(x=Species, y=Sepal.Width)) +
  geom_boxplot() +
  ggtitle("Boxplot of Sepal Width by Species")
```




```
ggplot(iris_dataset, aes(x=Species, y=Petal.Length)) +
  geom_boxplot() +
  ggtitle("Boxplot of Petal Length by Species")
```


```
ggplot(iris_dataset, aes(x=Species, y=Petal.Width)) +
  geom_boxplot() +
  ggtitle("Boxplot of Petal Width by Species")
```

By looking at all histograms and plots, we can quickly notice that all three species differ significantly from each other. For example, Iris species such as setosa, versicolor, and virginica exhibit substantial differences in their sepal and petal length, width, and size. By examining the calculated average sepal length, we see that Iris setosa has the shortest average sepal length among the three species, while Iris virginica has the longest. Iris versicolor stands in between the two. Additionally, the boxplots reveal that Iris virginica has both a larger average sepal length and petal length compared to the other two species. Although Iris setosa is generally smaller in most dimensions, it has the largest sepal width.