Week 6 - Rmarkdown Assignment

Surenther

2024-10-05

Iris Dataset Analysis

```
#Assign Iris dataset
iris_data <- iris
```

Average Sepal Length by Species

```
# Load the dplyr package
library(dplyr,warn.conflicts = FALSE)

# Group the iris dataset by species and calculate the average sepal length
average_sepal_length <- iris_data %>%
  group_by(Species) %>%
  summarise(avg_sepal_length = mean(Sepal.Length))

# View the results
knitr::kable(average_sepal_length, caption='Average Sepal Length.')
```

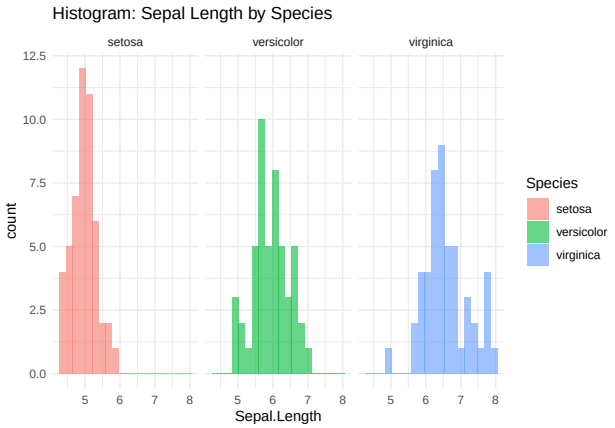
Table 1: Average Sepal Length.

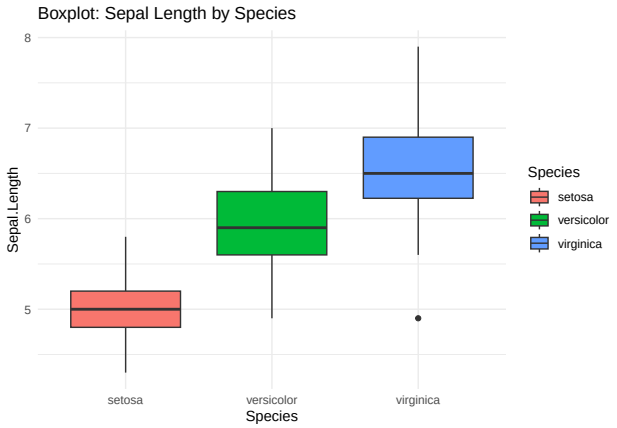
Species	avg_sepal_length
setosa	5.006
versicolor	5.936
virginica	6.588

Visualizations

```
# Histogram and Box plot for Sepal Length
  ggplot(iris, aes(x = Sepal.Length, fill = Species)) +
  geom_histogram(alpha = 0.6, position = "identity", bins = 20) +
  facet_wrap(~ Species) +
  theme_minimal() +
  ggtitle("Histogram: Sepal Length by Species")
```

```r
ggplot(iris, aes(x = Species, y = Sepal.Length, fill = Species)) +
geom_boxplot() +
theme_minimal() +
ggtitle("Boxplot: Sepal Length by Species")
```



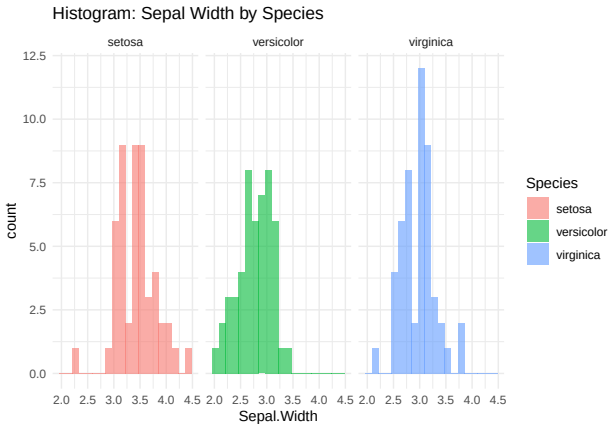


Findings

The Sepal Length shows noticeable differences between the species in the iris dataset.Iris virginica tends to have the longest sepals, with both the highest median and overall range, as shown by the box plot and histogram. Iris setosa has the shortest sepal lengths, with a much narrower range and a lower median compared to the other species. Iris versicolor falls between the two, with moderate sepal lengths, showing overlap with both setosa and virginica.

```r
# Histogram and Box plot for Sepal Width
ggplot(iris, aes(x = Sepal.Width, fill = Species)) +
geom_histogram(alpha = 0.6, position = "identity", bins = 20) +
facet_wrap(~ Species) +
theme_minimal() +
ggtitle("Histogram: Sepal Width by Species")

ggplot(iris, aes(x = Species, y = Sepal.Width, fill = Species)) +
geom_boxplot() +
theme_minimal() +
ggtitle("Boxplot: Sepal Width by Species")
```



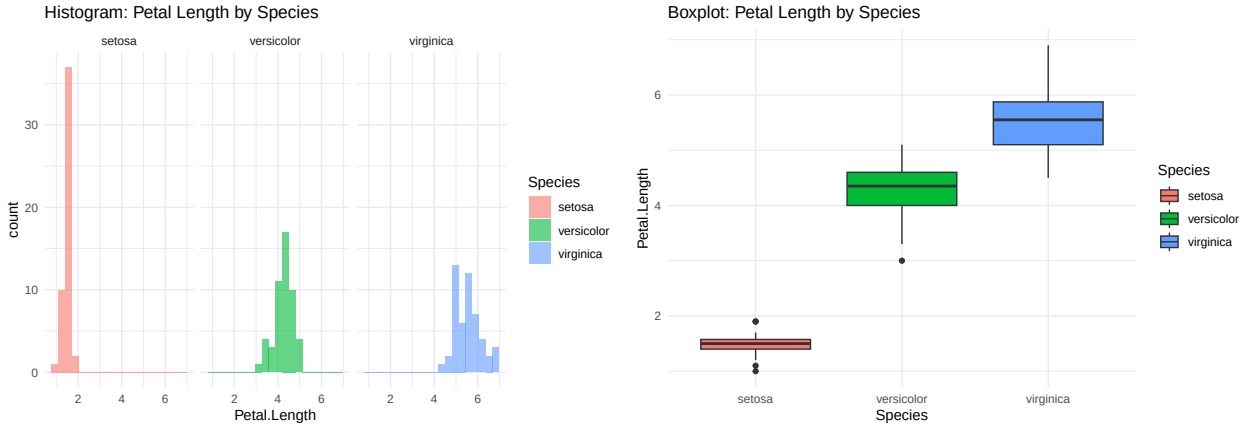
Findings

The species setosa tends to have the widest sepals, with most values clustered around 3.5 cm, as shown

by both the histogram and the box plot. In contrast, versicolor and virginica have narrower sepals, with versicolor displaying a smaller range of values centered around 2.8 cm and virginica around 3.0 cm.

```
# Histogram and Box plot for Petal Length
  ggplot(iris, aes(x = Petal.Length, fill = Species)) +
  geom_histogram(alpha = 0.6, position = "identity", bins = 20) +
  facet_wrap(~ Species) +
  theme_minimal() +
  ggtitle("Histogram: Petal Length by Species")

  ggplot(iris, aes(x = Species, y = Petal.Length, fill = Species)) +
  geom_boxplot() +
  theme_minimal() +
  ggtitle("Boxplot: Petal Length by Species")
```

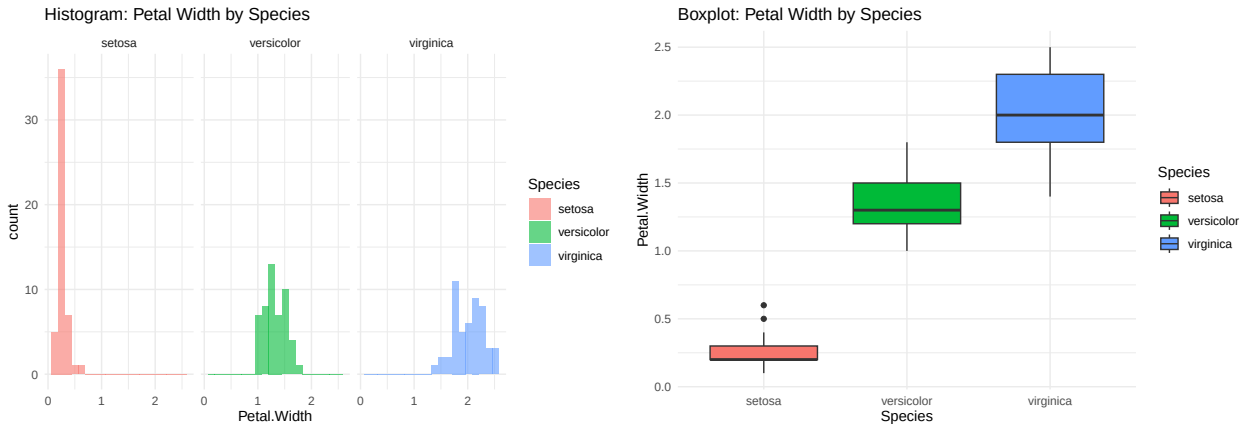


Findings

Setosa has the shortest petal lengths, with a relatively small range and little variation, typically falling below 2 cm. In contrast, Versicolor has moderate petal lengths, generally ranging between 3 and 5 cm, showing a wider spread compared to Setosa. Virginica exhibits the longest petal lengths, mostly above 5 cm, with the highest variability among the species

```
# Histogram and Box plot for Petal Width
  ggplot(iris, aes(x = Petal.Width, fill = Species)) +
  geom_histogram(alpha = 0.6, position = "identity", bins = 20) +
  facet_wrap(~ Species) +
  theme_minimal() +
  ggtitle("Histogram: Petal Width by Species")

 ggplot(iris, aes(x = Species, y = Petal.Width, fill = Species)) +
  geom_boxplot() +
  theme_minimal() +
  ggtitle("Boxplot: Petal Width by Species")
```



Findings

Setosa exhibits the smallest petal widths, with values tightly clustered around 0.2 cm, showing little variation. In contrast, Versicolor has petal widths that are moderately larger, ranging between 1.0 and 1.8 cm, with a wider spread. Virginica shows the greatest petal widths, ranging from approximately 1.8 to 2.5 cm, indicating both a higher average and greater variability compared to the other two species

Overall Findings

The visualization comparing the three species of the iris dataset setosa, versicolor, and virginica shows clear distinctions in their sepal and petal dimensions. Setosa consistently has shorter petal lengths and petal widths, as well as wider sepals compared to the other species. Versicolor and virginica overlap in sepal dimensions, but virginica tends to have larger petal lengths and petal widths overall. The boxplots reveal that virginica has the greatest variation in both petal length and width, while setosa shows the most consistency with fewer outliers.