Iris Dataset Analysis

Miles Peña

2024-01-21

Iris Analysis

```
# load iris

library(datasets)
data(iris)
```

```
# new iris variable

newIrisVar <- iris
```

Average Sepal Length by Species

```
# load dplyr and group by species

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
newIrisVar %>% group_by(Species) %>%
summarise(Avg_Sepal_Length = mean(Sepal.Length))
```

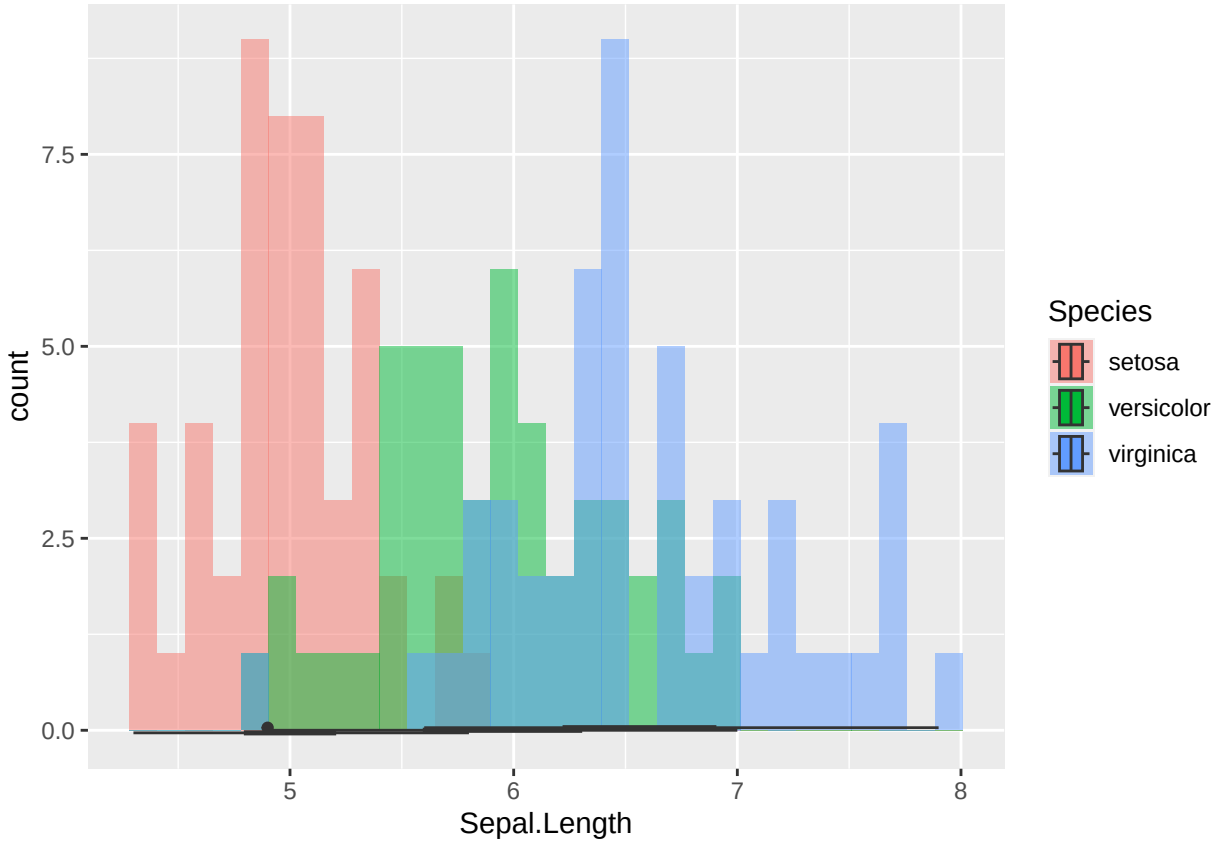
```
## # A tibble: 3 x 2
##   Species    Avg_Sepal_Length
##   <fct>                 <dbl>
## 1 setosa                 5.01
## 2 versicolor             5.94
## 3 virginica              6.59
```

Visualizations

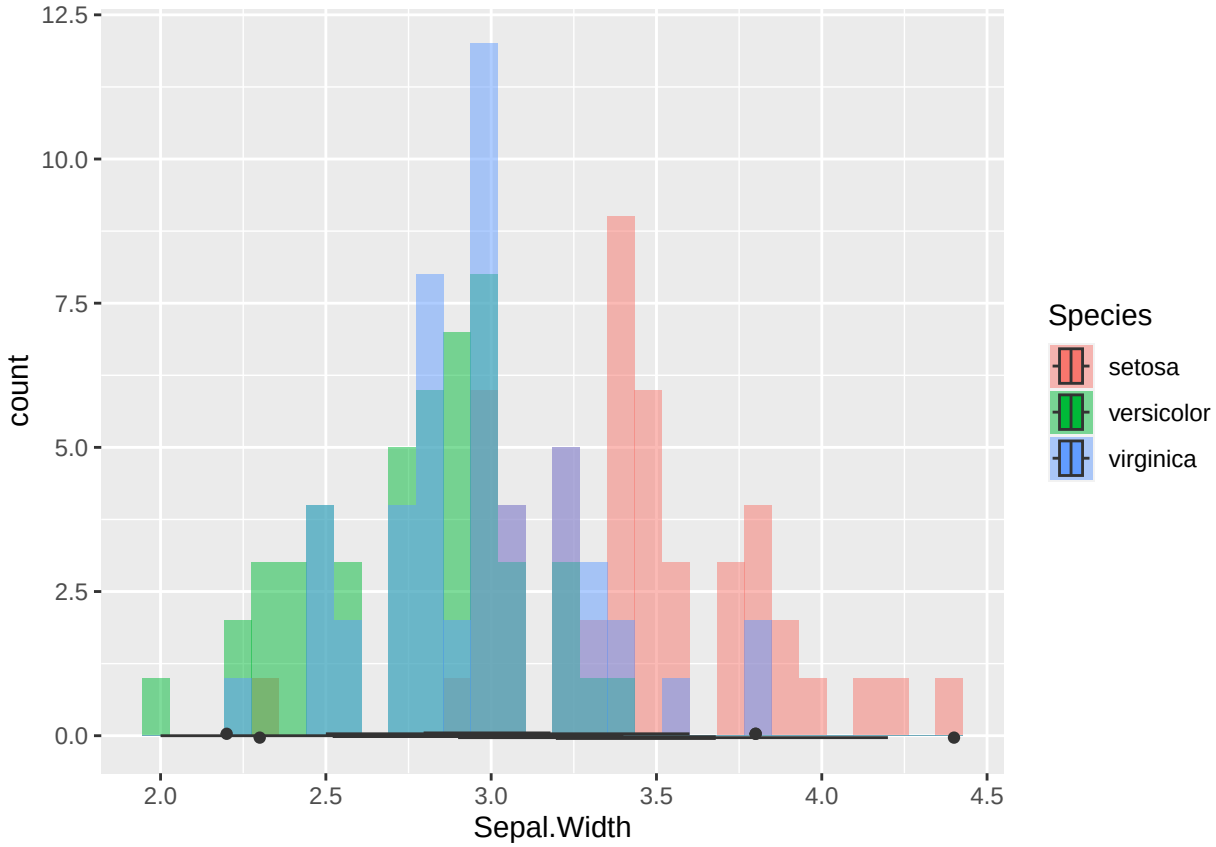
```
# histograms and boxplots for sepal length, sepal width, petal length and petal width

# Sepal Length
```

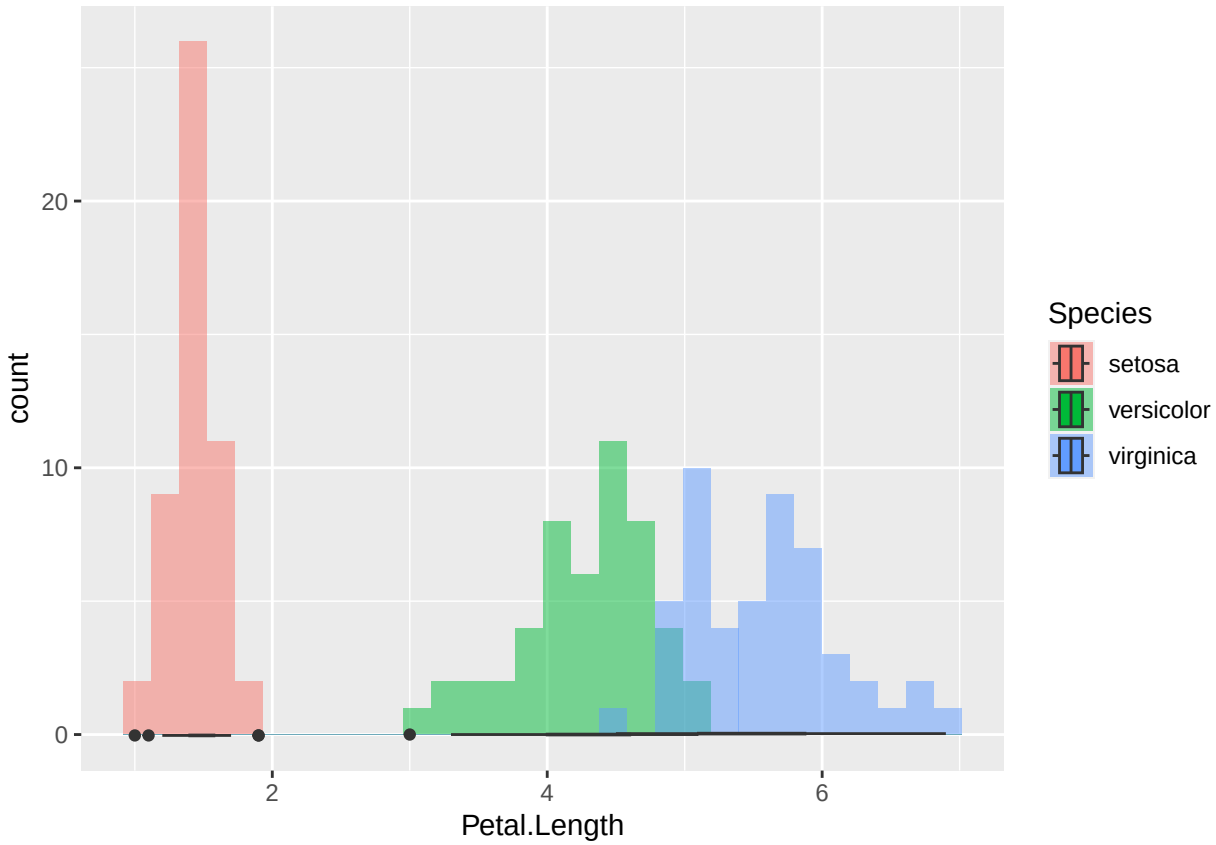
```
sepal_length <- ggplot(newIrisVar, aes(x = Sepal.Length, fill = Species)) +
geom_histogram(position = "identity", alpha = 0.5, bins = 30) +
geom_boxplot(width = 0.1)
sepal_length
```



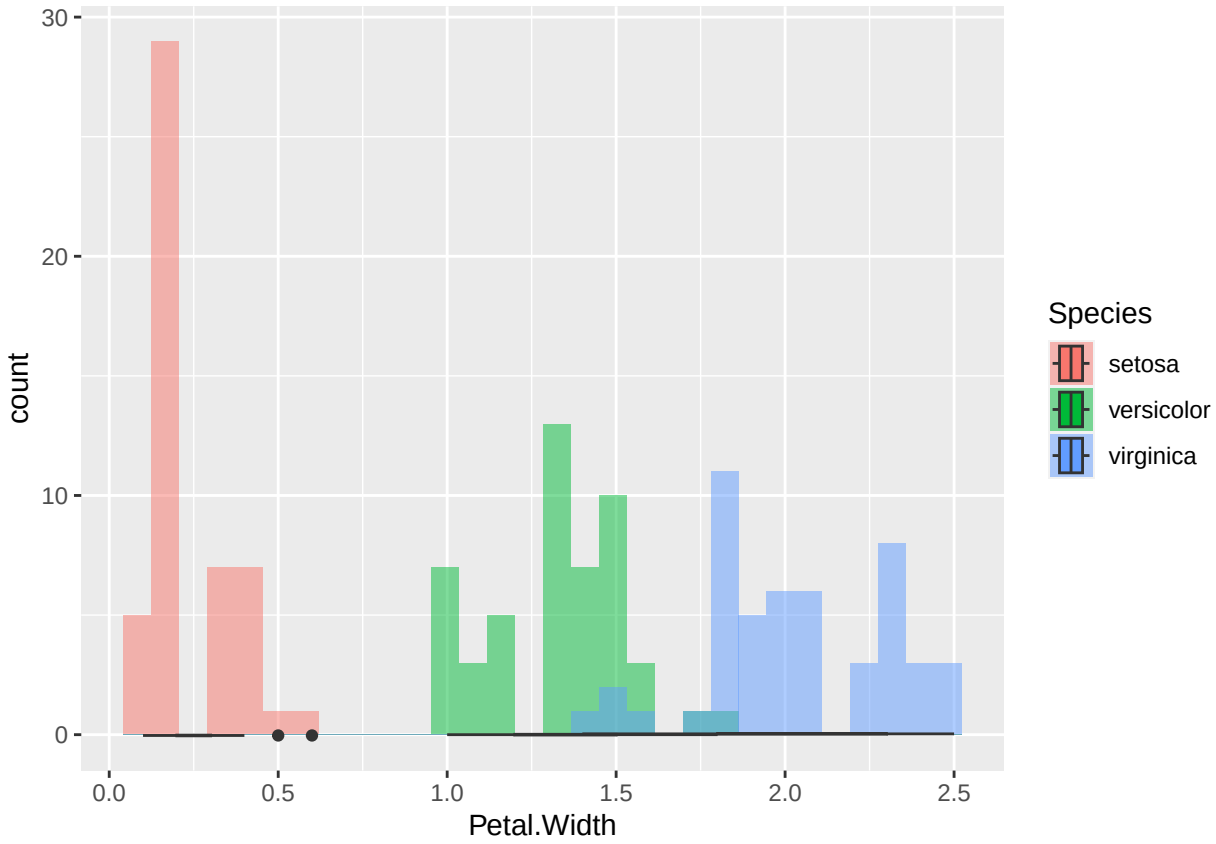

```
# Sepal Width
sepal_width <- ggplot(newIrisVar, aes(x = Sepal.Width, fill = Species)) +
geom_histogram(position = "identity", alpha = 0.5, bins = 30) +
geom_boxplot(width = 0.1)
sepal_width
```




```
# Petal Length
petal_length <- ggplot(newIrisVar, aes(x = Petal.Length, fill = Species)) +
geom_histogram(position = "identity", alpha = 0.5, bins = 30) +
geom_boxplot(width = 0.1)
petal_length
```




```
# Petal Width
petal_width <- ggplot(newIrisVar, aes(x = Petal.Width, fill = Species)) +
geom_histogram(position = "identity", alpha = 0.5, bins = 30) +
geom_boxplot(width = 0.1)
petal_width
```




```
# Based on the histograms/boxplots, the species setosa has the smallest lengths of sepals but has the w
```