

# Iris Dataset Analysis

Nick Blackford

2024-01-21

## Iris Dataset Analysis

```
library(datasets)
data(iris)
my_iris <- iris
```

### Average Sepal Length 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

avg_sepal_length <- my_iris %>%
  group_by(Species) %>%
  summarize(Average_Sepal_Length = mean(Sepal.Length))
avg_sepal_length

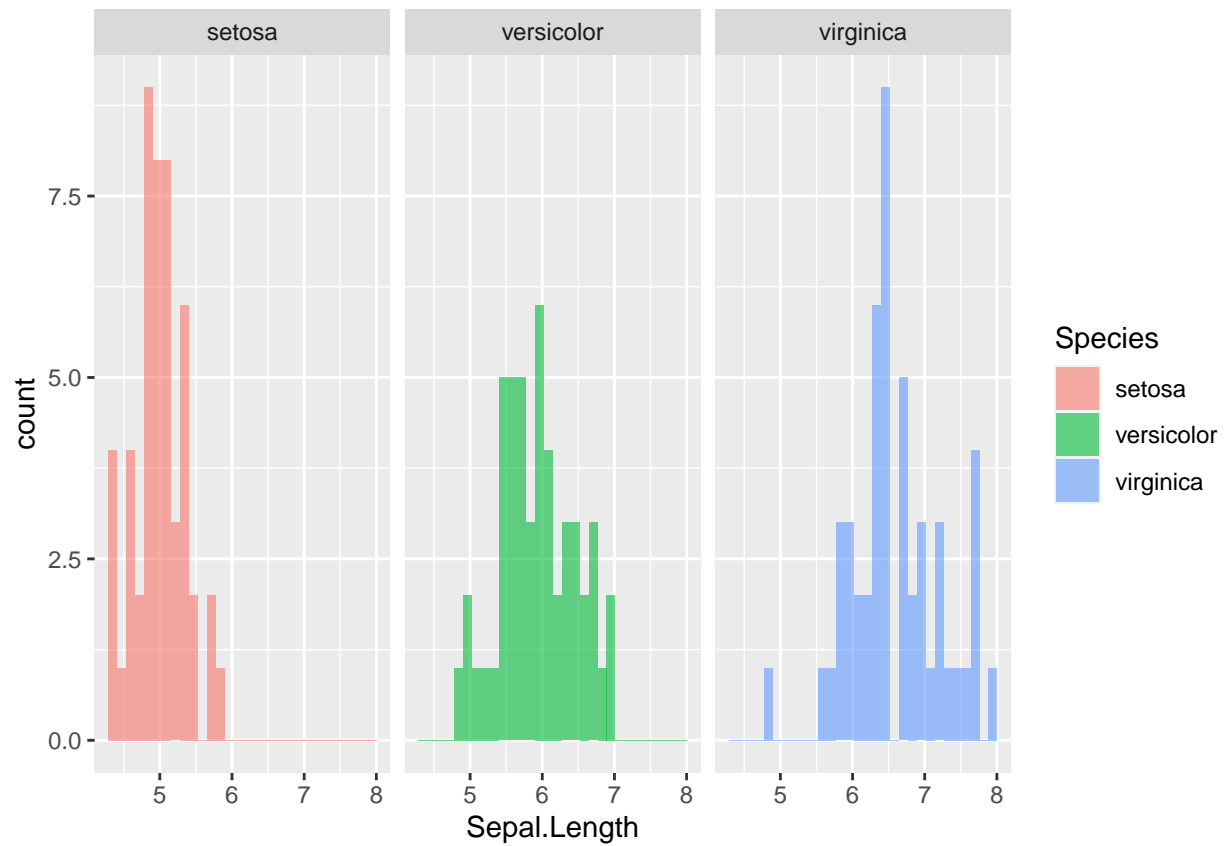
## # A tibble: 3 x 2
##   Species      Average_Sepal_Length
##   <fct>          <dbl>
## 1 setosa              5.01
## 2 versicolor         5.94
## 3 virginica          6.59
```

### Visualizations

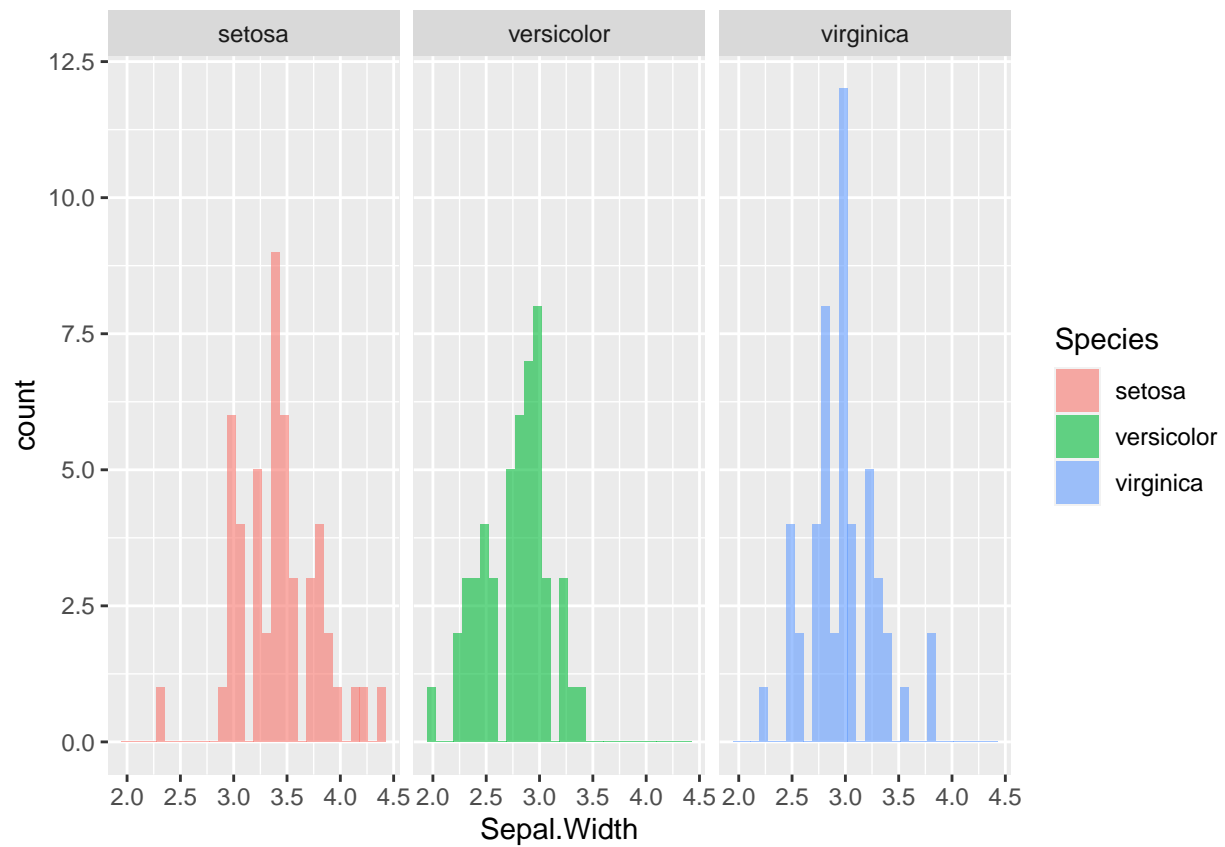
Below are visualizations showing the differences in sepal length, sepal width, petal length, and petal width between the different species in the Iris dataset.

```
# Sepal Length
ggplot(my_iris, aes(x = Sepal.Length, fill = Species)) +
```

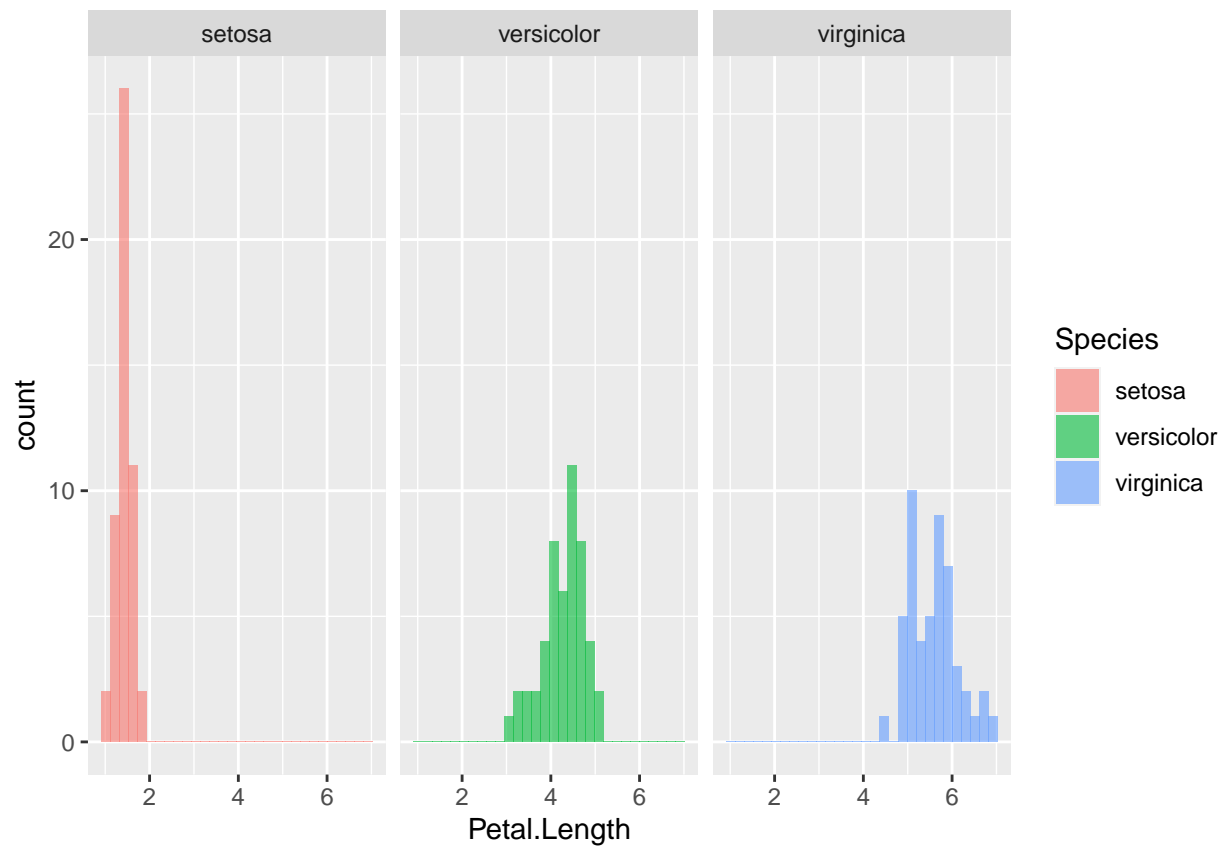
```
geom_histogram(bins = 30, alpha = 0.6) +  
facet_wrap(~Species)
```



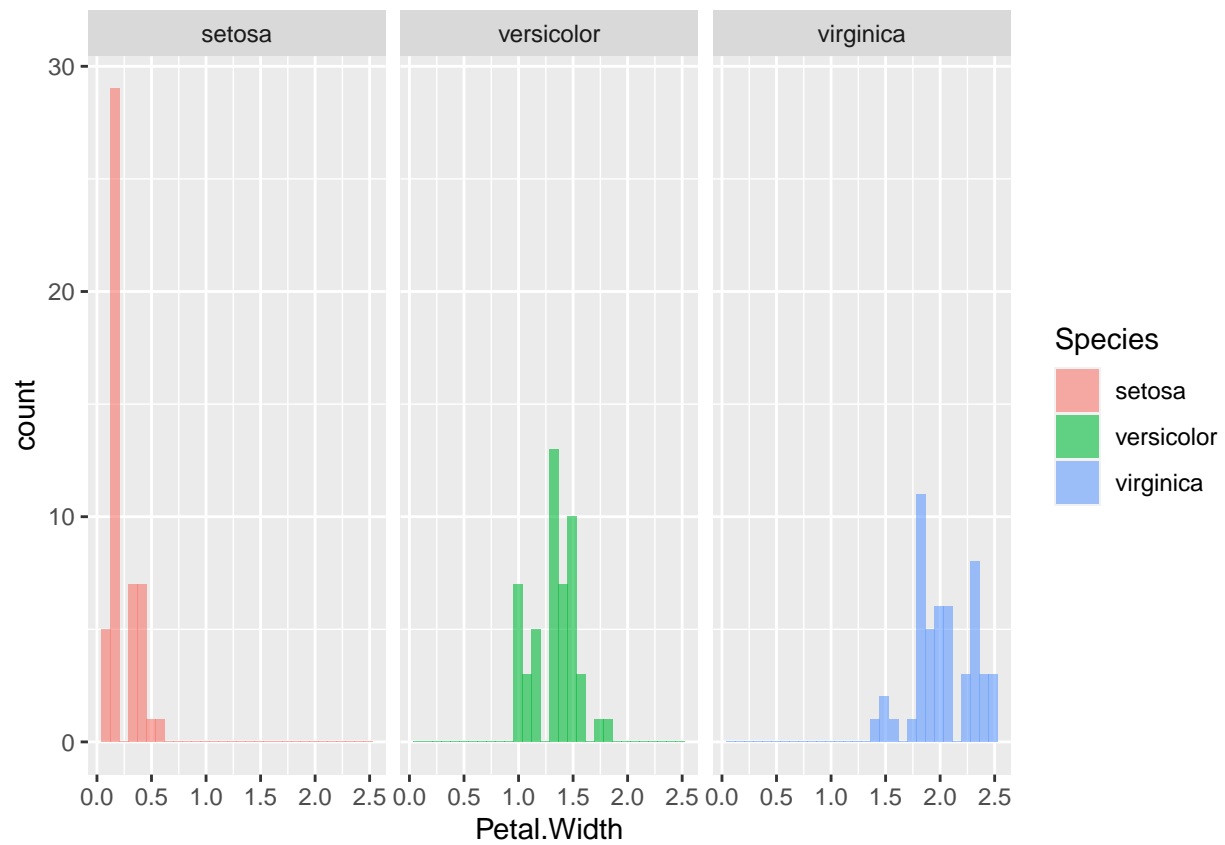
```
# Sepal Width  
ggplot(my_iris, aes(x = Sepal.Width, fill = Species)) +  
  geom_histogram(bins = 30, alpha = 0.6) +  
  facet_wrap(~Species)
```



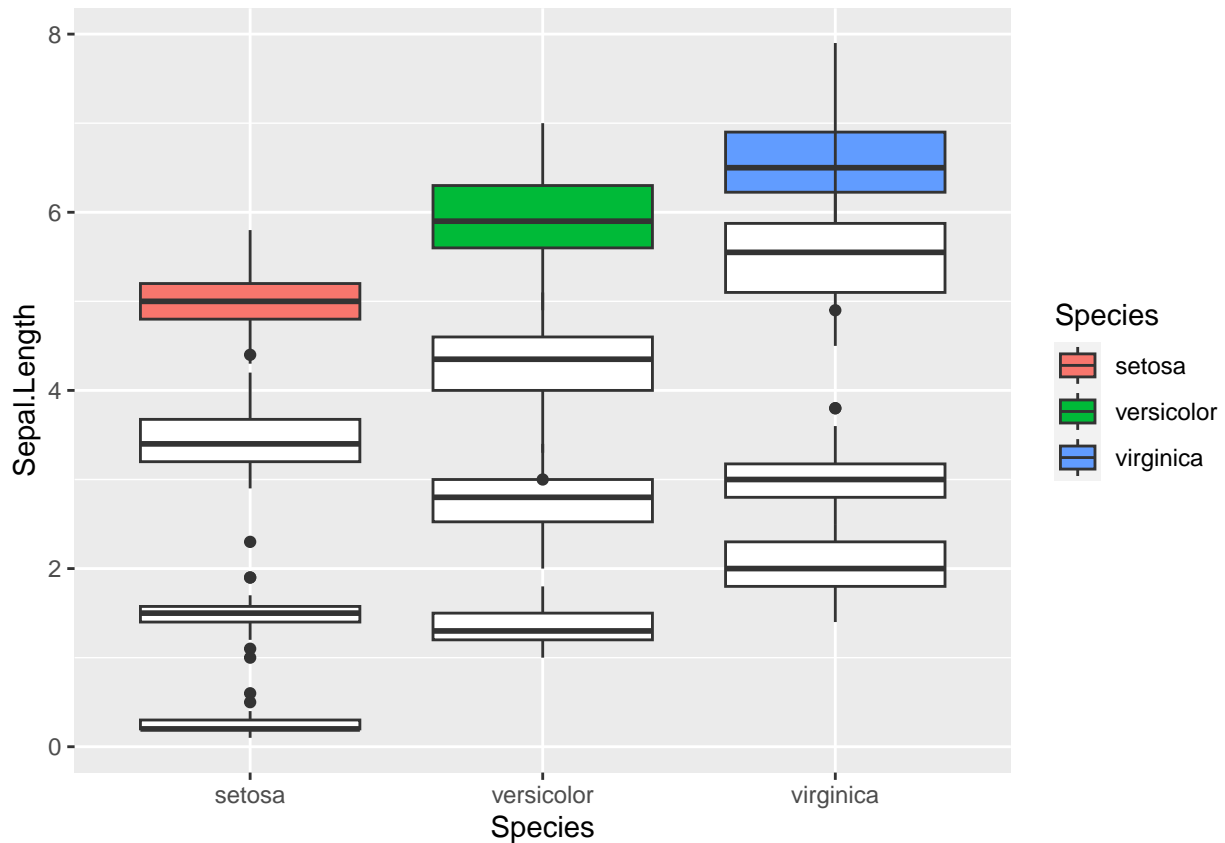
```
# Petal Length
ggplot(my_iris, aes(x = Petal.Length, fill = Species)) +
  geom_histogram(bins = 30, alpha = 0.6) +
  facet_wrap(~Species)
```



```
# Petal Width
ggplot(my_iris, aes(x = Petal.Width, fill = Species)) +
  geom_histogram(bins = 30, alpha = 0.6) +
  facet_wrap(~Species)
```



```
# Boxplot for all measurements
ggplot(my_iris, aes(x = Species, y = Sepal.Length)) +
  geom_boxplot(aes(fill = Species)) +
  geom_boxplot(aes(y = Sepal.Width)) +
  geom_boxplot(aes(y = Petal.Length)) +
  geom_boxplot(aes(y = Petal.Width))
```



The histograms and box plots above illustrate distinct differences in sepal length, sepal width, petal length, and petal width among the Iris species. In terms of sepal length, the sertosa is shortest, followed by the versicolor, then virginica respectively. The distributions for sepal length are approximately normal for all species, with virginica having the largest standard deviation. The largest in terms of sepal width are sertosa, virginica, then versicolor respectively. Petal length has much larger differences with sertosa averaging less than 2cm, versicolor at about 4cm, and virginica centered near 6cm.