Graphics with R

To test plot functions, we'll use the iris data.set

```
head(iris)
```

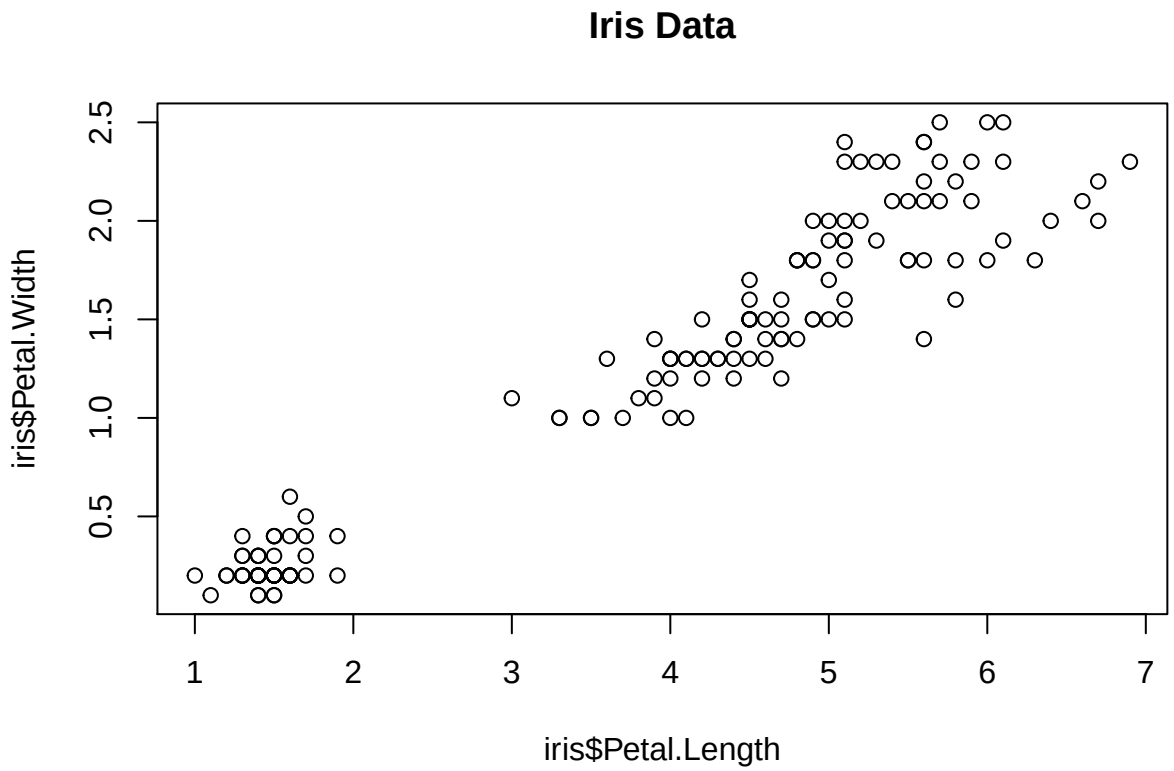
```
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
```

Basic plots

Scatter plots

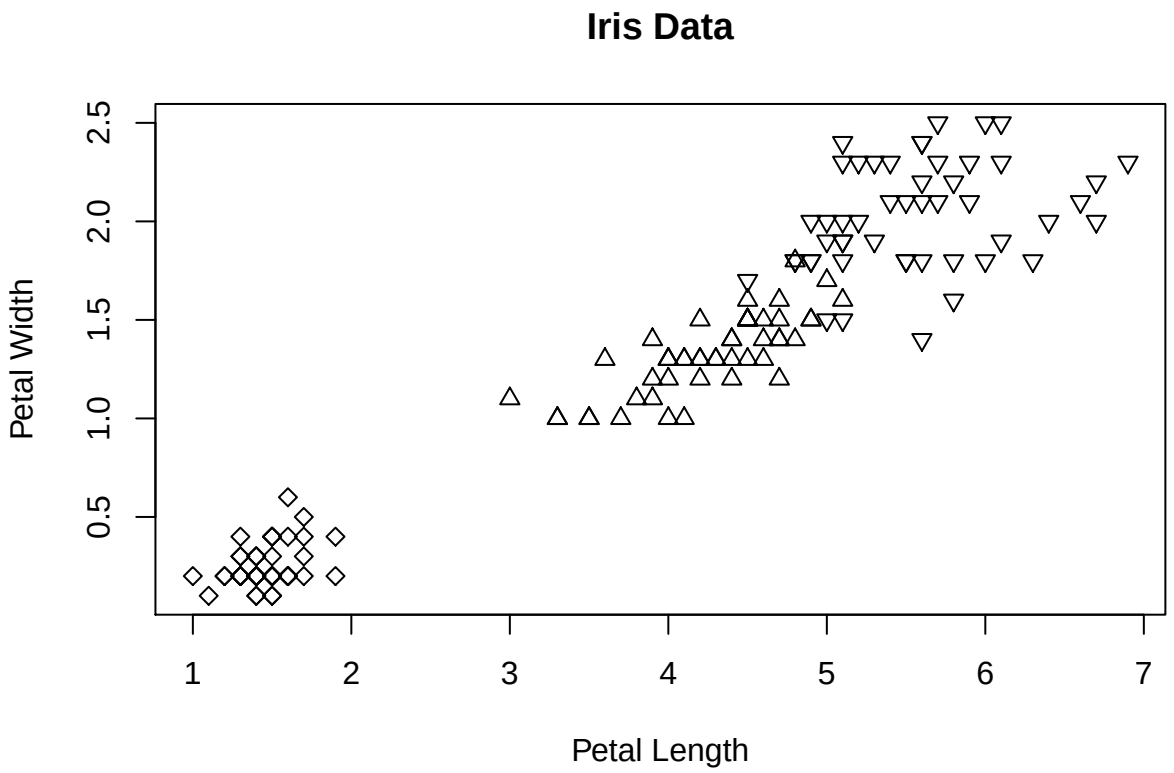
The simplest plots in R are scatter plots. A scatter plot pairs up values of two quantitative variables in a data set and display them as geometric points inside a Cartesian diagram. They are obtained by using the function `plot(x,y)`.

```
plot(iris$Petal.Length, iris$Petal.Width, main="Iris Data")
```



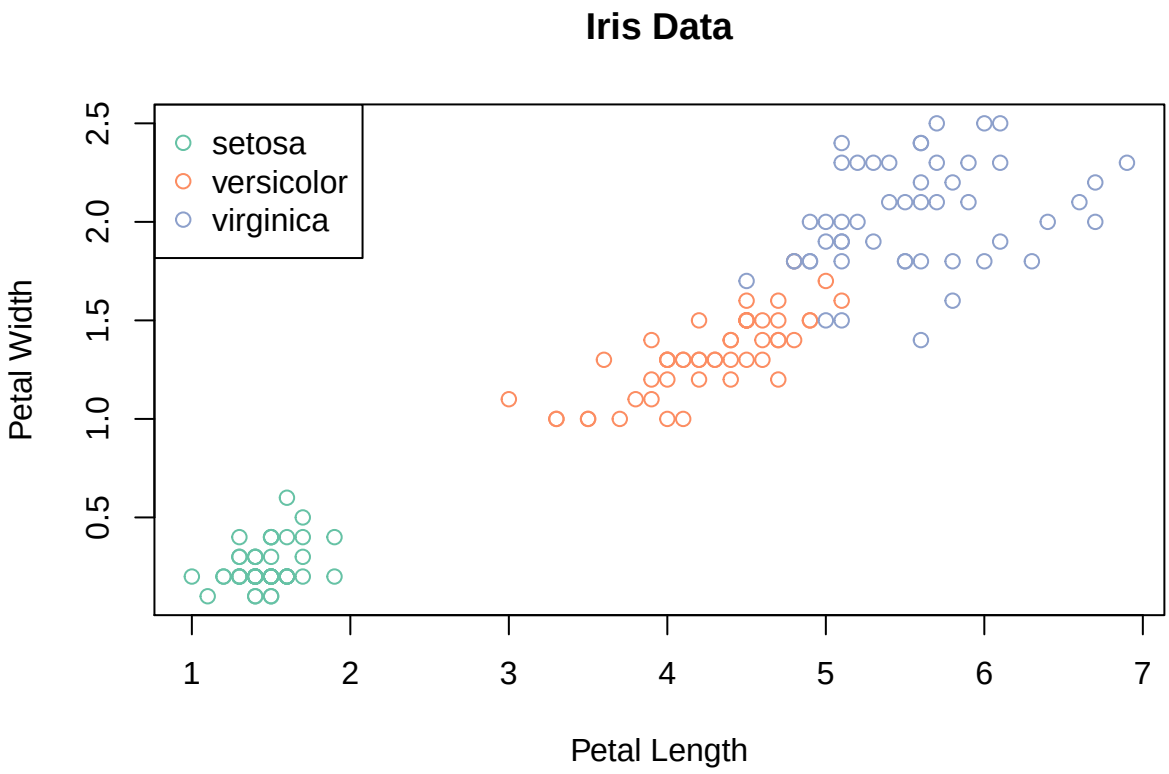
The `plot()` function has many arguments * type: the type of plot - "p" for points - "l" for lines - "b" for both - "o" for overplotted - "h" for histogram - "s" for stair steps * main: an overall tit;e for the plot * sub: a subtitle for the plot * xlab: a title for the x-axis * ylab: a title for the y-axis * pch: plot character - pch=21 for filled circles - pch=22 for filled squares - pch=23 for filled diamonds - pch=24 or pch=25 for up/down triangles.

```
plot(iris$Petal.Length, iris$Petal.Width, pch=c(23,24,25)[unclass(iris$Species)], main="Iris Data", xlab
```



We can also add colors and a legend

```
library(RColorBrewer)
plot(iris$Petal.Length, iris$Petal.Width, pch=21, col=brewer.pal(3, "Set2")[iris$Species], main="Iris Da
legend("topleft", legend=levels(iris$Species), col=brewer.pal(3, "Set2"), pch=1)
```

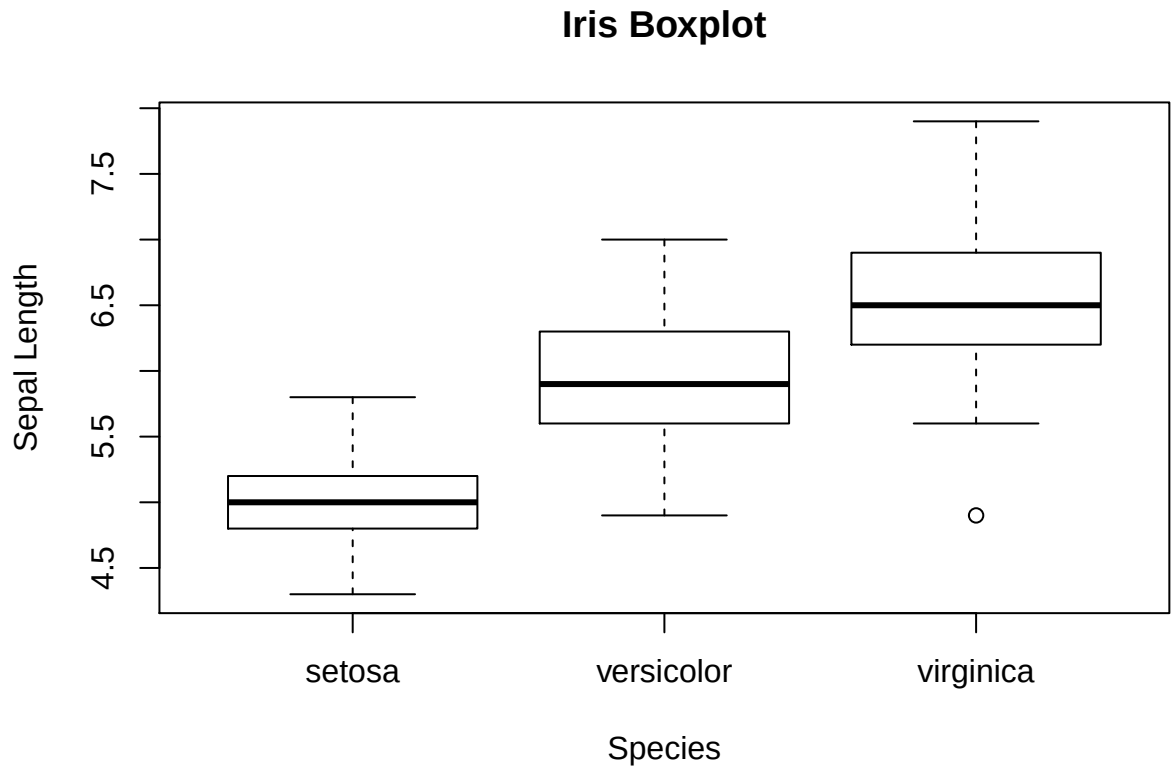


Box-plots

Boxplots are useful to represent simple variables. The box plot (a.k.a. box and whisker diagram) is a standardized way of displaying the distribution of data based on the five number summary: minimum, first quartile, median, third quartile, and maximum.

For example a box-plot of the petal-length can be obtained as follows:

```
boxplot(Sepal.Length~Species,data=iris,
        xlab="Species", ylab="Sepal Length", main="Iris Boxplot")
```

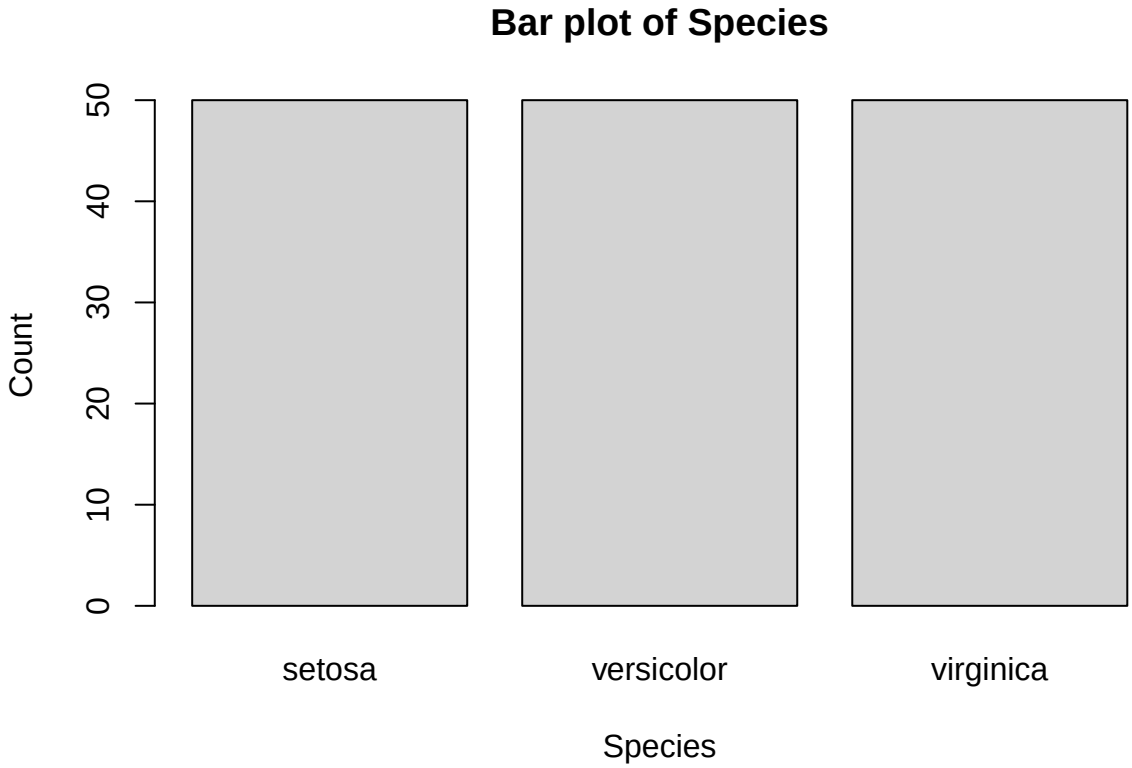


##

Bar plots A bar chart or bar graph is a chart or graph that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally.

For example a bar plot of the Iris Species can be obtained as follows:

```
barplot(table(iris$Species), col="lightgrey", xlab="Species", ylab="Count", main="Bar plot of Species")
```



As the iris

contains 50 samples of each species, lets create a sub-dataset to get a smoother histogram

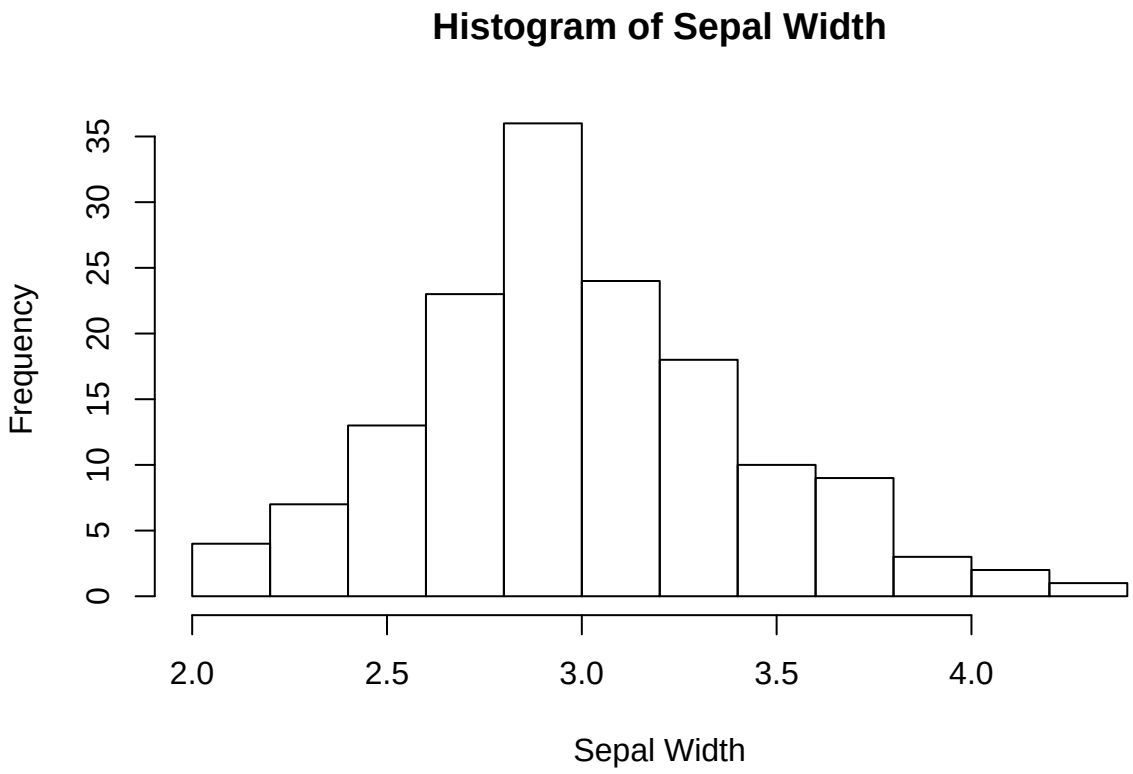
```
iris1 <- iris[sample(1:nrow(iris), 70), ]
barplot(table(iris1$Species), col="lightgrey", xlab="Species", ylab="Count", main="Bar plot of Species")
```

Histograms

Histograms are widely use in statistics to visually estimate the stastistical distribution of variables. A histogram is a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width is equal to the class interval.

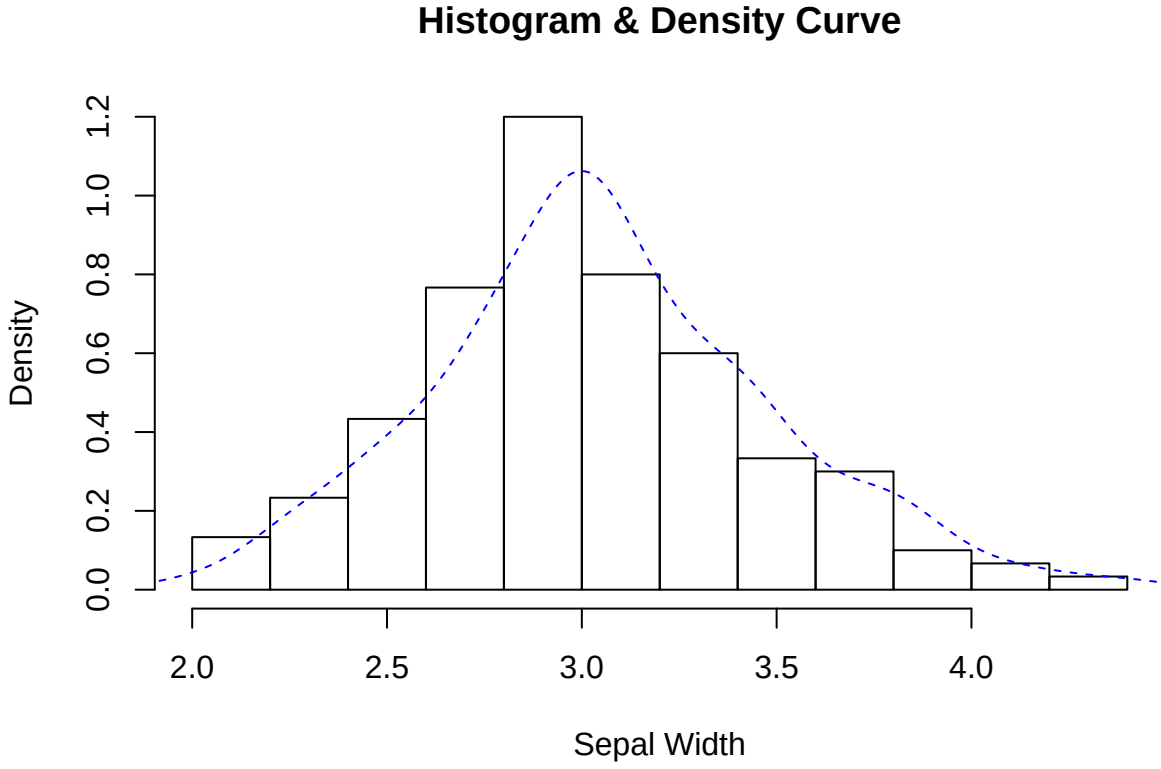
For example the histogram of petal-width is obrain as follows:

```
hist(iris$Sepal.Width, freq=NULL, density=NULL, breaks=12,
     xlab="Sepal Width", ylab="Frequency", main="Histogram of Sepal Width")
```



It is also possible to add the density curve:

```
 d <- density(iris$Sepal.Width)
hist(iris$Sepal.Width, breaks=12, prob=TRUE, xlab="Sepal Width", main="Histogram & Density Curve")
lines(d, lty=2, col="blue")
```



Advanced plots with ggplot2

ggplot2 is a plotting system for R, based on the grammar of graphics, which tries to take the good parts of base and lattice graphics and none of the bad parts. It takes care of many of the fiddly details that make plotting a hassle (like drawing legends) as well as providing a powerful model of graphics that makes it easy to produce complex multi-layered graphics.

Scatter plot

As a starting point, let's replicate the scatter plot previously obtained.

```
library(ggplot2)
scatter <- ggplot(data=iris, aes(x = Sepal.Length, y = Sepal.Width))
scatter + geom_point(aes(color=Species, shape=Species)) +
  xlab("Sepal Length") +  ylab("Sepal Width") +
  ggtitle("Iris data")
```

Another way of doing thta is using grdaient colors (temperature-like)

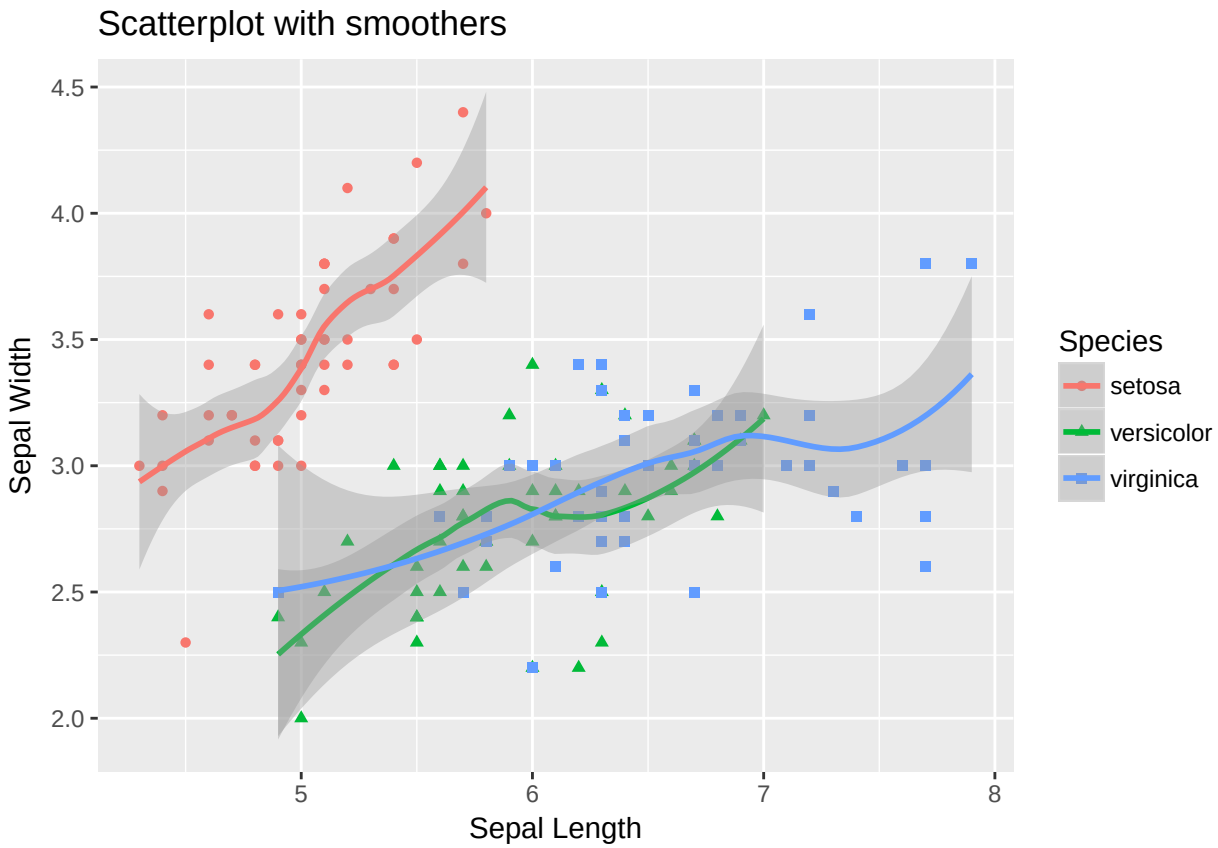
```
scatter + geom_point(aes(color = Petal.Width, shape = Species), size = 2, alpha = I(1/2)) +
          geom_vline(aes(xintercept = mean(Sepal.Length)), color = "red", linetype = "dashed") +
          geom_hline(aes(yintercept = mean(Sepal.Width)), color = "red", linetype = "dashed") +
          scale_color_gradient(low = "yellow", high = "red") +
          xlab("Sepal Length") +  ylab("Sepal Width") +
          ggtitle("Iris data")
```

ggplot2 allows adding smoothers to the scatterplot

```
smooth <- ggplot(data=iris, aes(x=Sepal.Length, y=Sepal.Width, color=Species)) +
  geom_point(aes(shape=Species), size=1.5) + xlab("Sepal Length") + ylab("Sepal Width") +
  ggtitle("Scatterplot with smoothers")

# Linear model
smooth + geom_smooth(method="lm")
```


```
# Local polynomial regression
smooth + geom_smooth(method="loess")
```

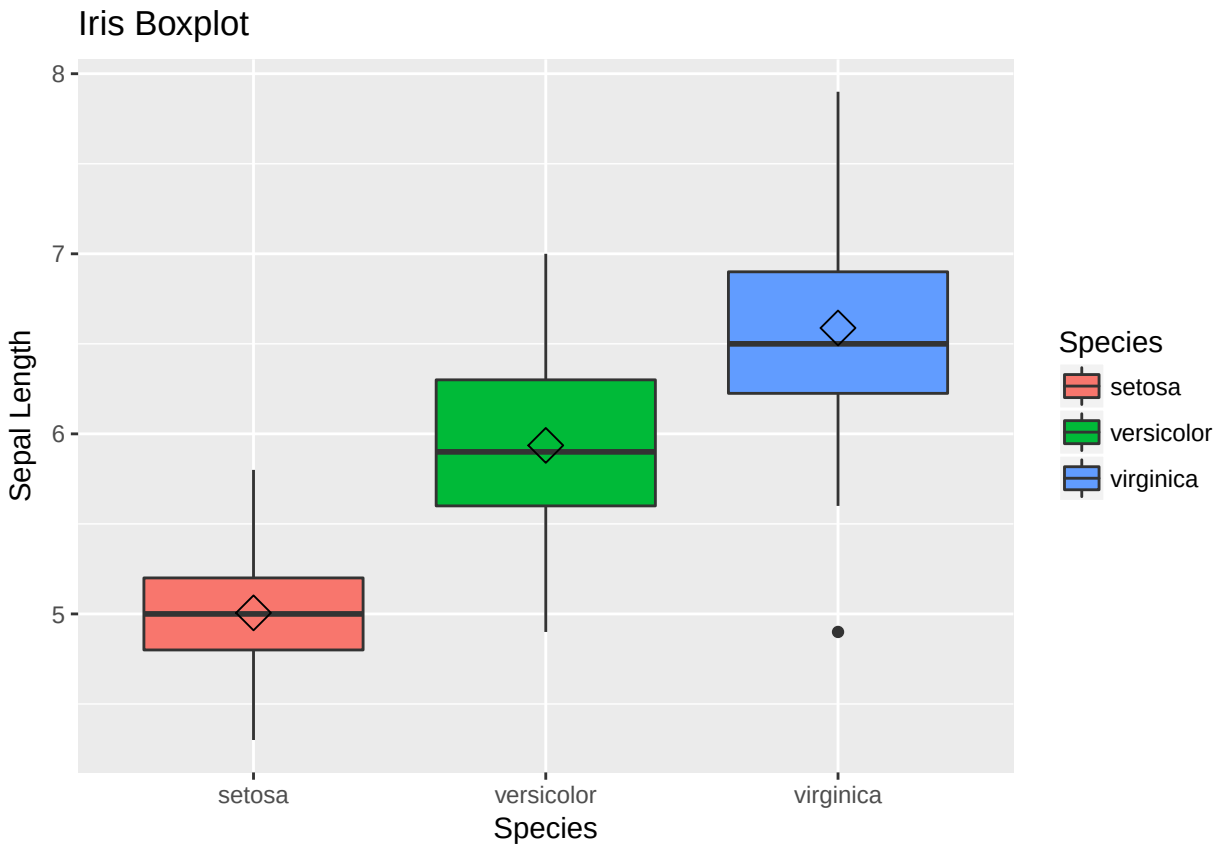



```
# generalised additive model
smooth + geom_smooth(method="gam", formula= y~s(x, bs="cs"))
```

Boxplot

With ggplot2, boxplots can be obtained as follows:

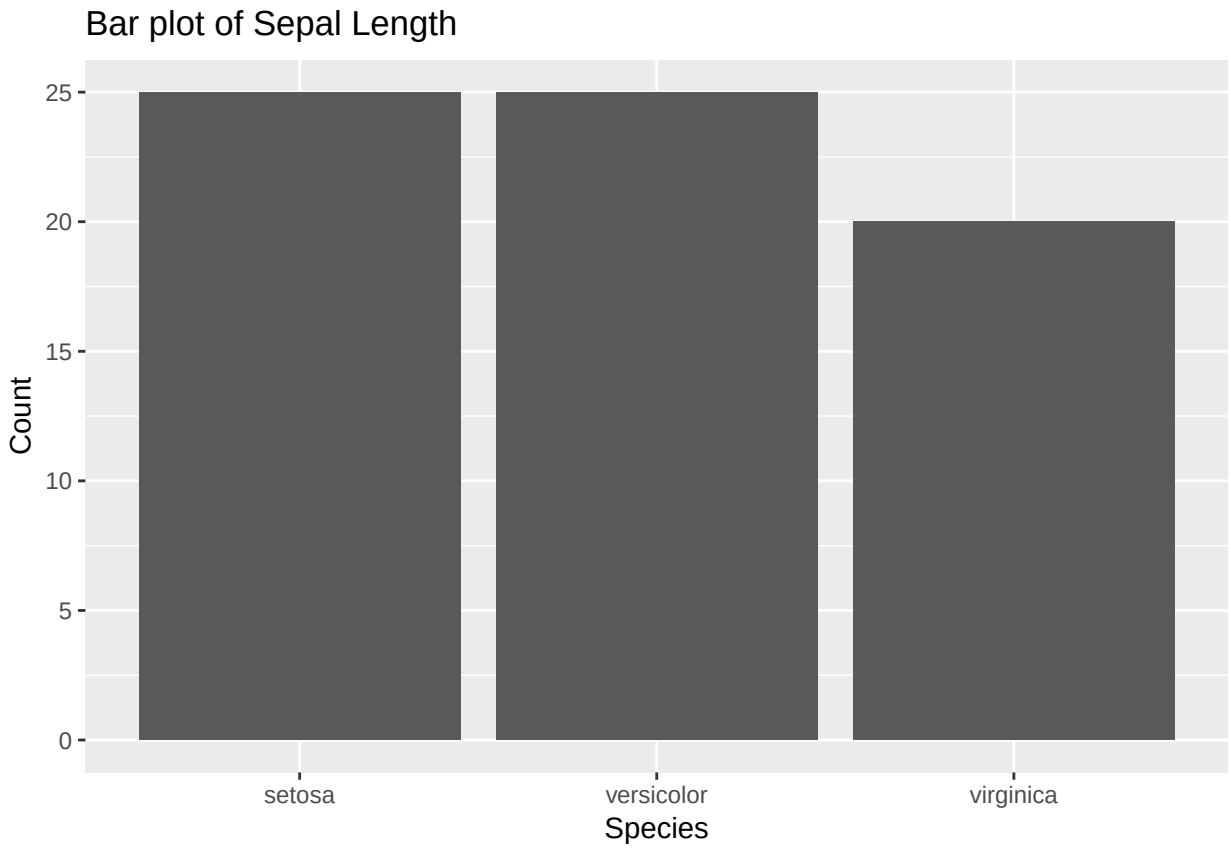
```
box <- ggplot(data=iris, aes(x=Species, y=Sepal.Length))
box + geom_boxplot(aes(fill=Species)) +
  ylab("Sepal Length") + ggtitle("Iris Boxplot") +
  stat_summary(fun.y=mean, geom="point", shape=5, size=4)
```



Bar plot

A bar-plot with ggplot2:

```
bar <- ggplot(data=iris1, aes(x=Species))
bar + geom_bar() +
  xlab("Species") +  ylab("Count") + ggtitle("Bar plot of Sepal Length")
```

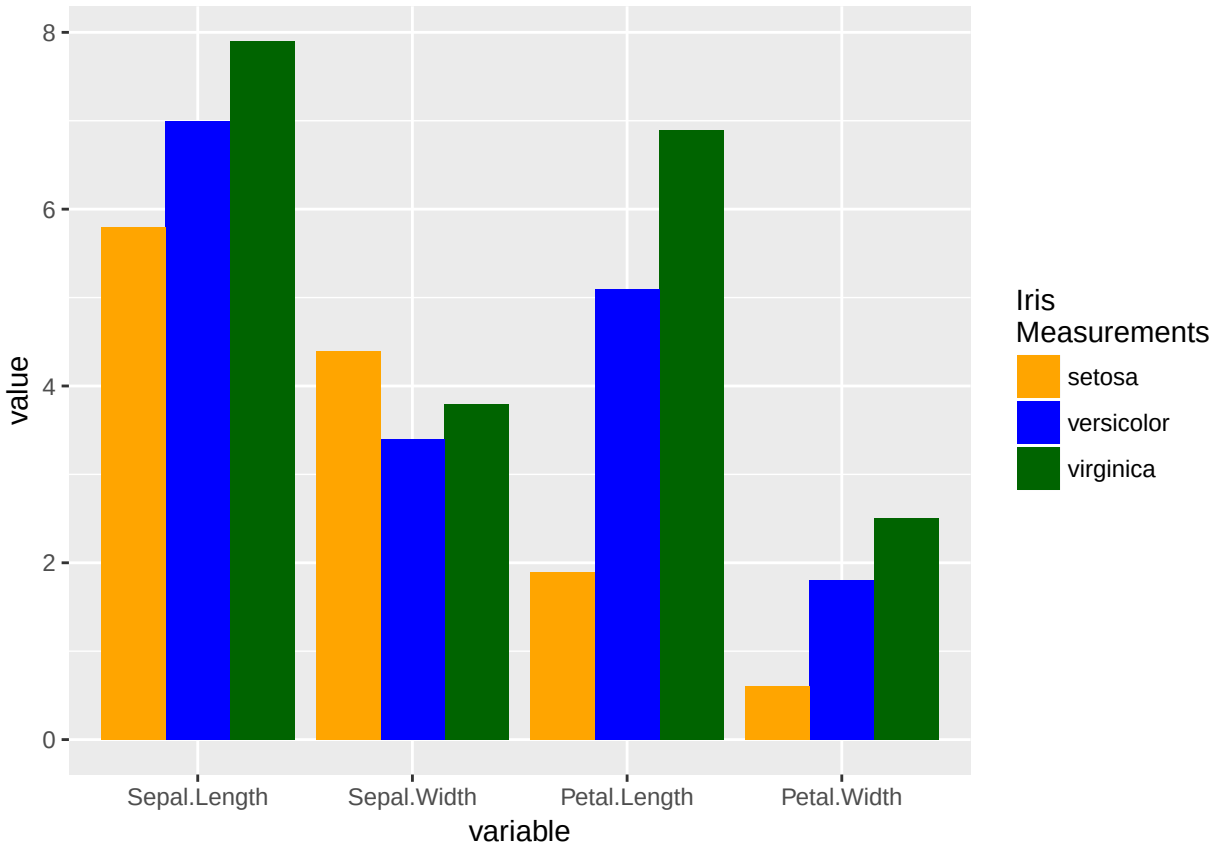


Advanced bar plots can be obtained by melting hte data as follows:

```
library(reshape2)
iris2 <- melt(iris, id.vars="Species")
head(iris2)
```

```
  Species     variable value
1  setosa Sepal.Length   5.1
2  setosa Sepal.Length   4.9
3  setosa Sepal.Length   4.7
4  setosa Sepal.Length   4.6
5  setosa Sepal.Length   5.0
6  setosa Sepal.Length   5.4
```

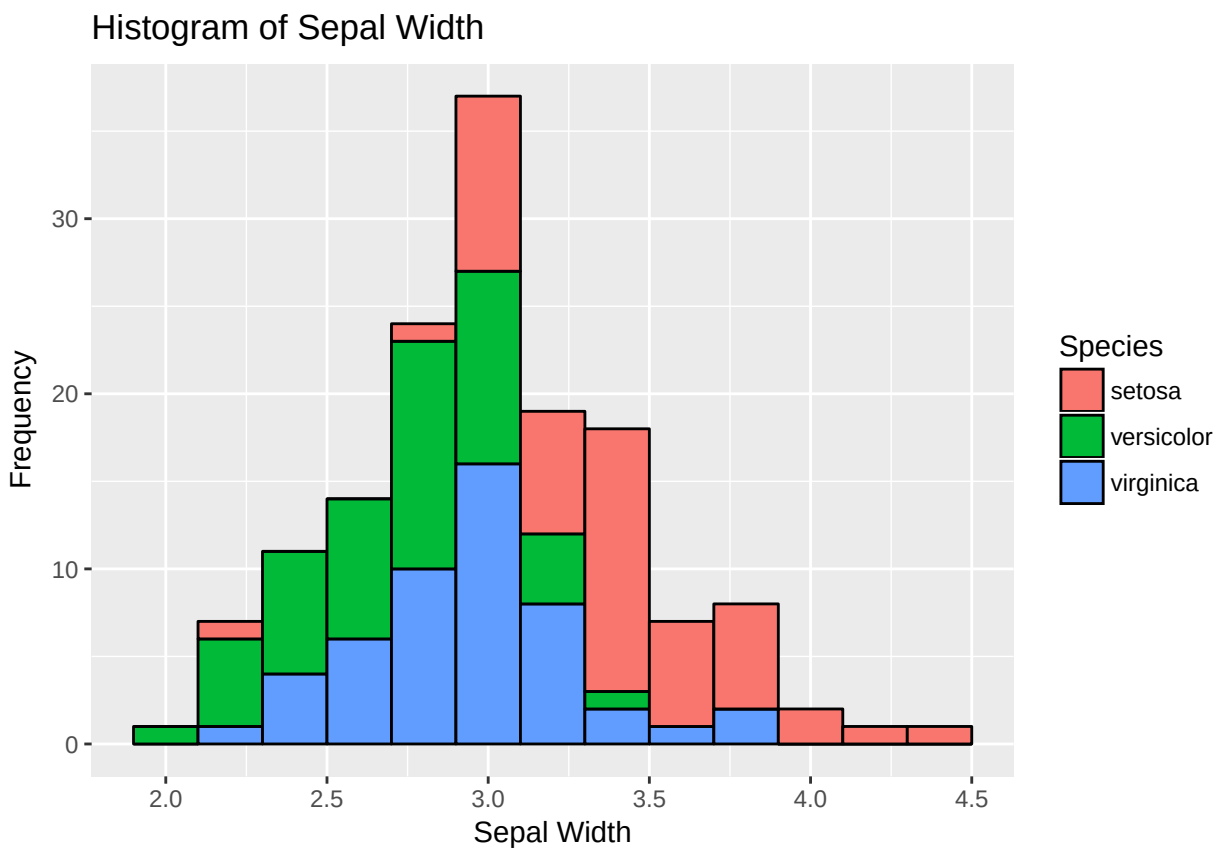
```
bar1 <- ggplot(data=iris2, aes(x=variable, y=value, fill=Species))
bar1 + geom_bar(stat="identity", position="dodge") +
  scale_fill_manual(values=c("orange", "blue", "darkgreen", "purple"),
                    name="Iris\nMeasurements",
                    breaks=unique(iris2$Species),
                    labels=unique(iris2$Species))
```



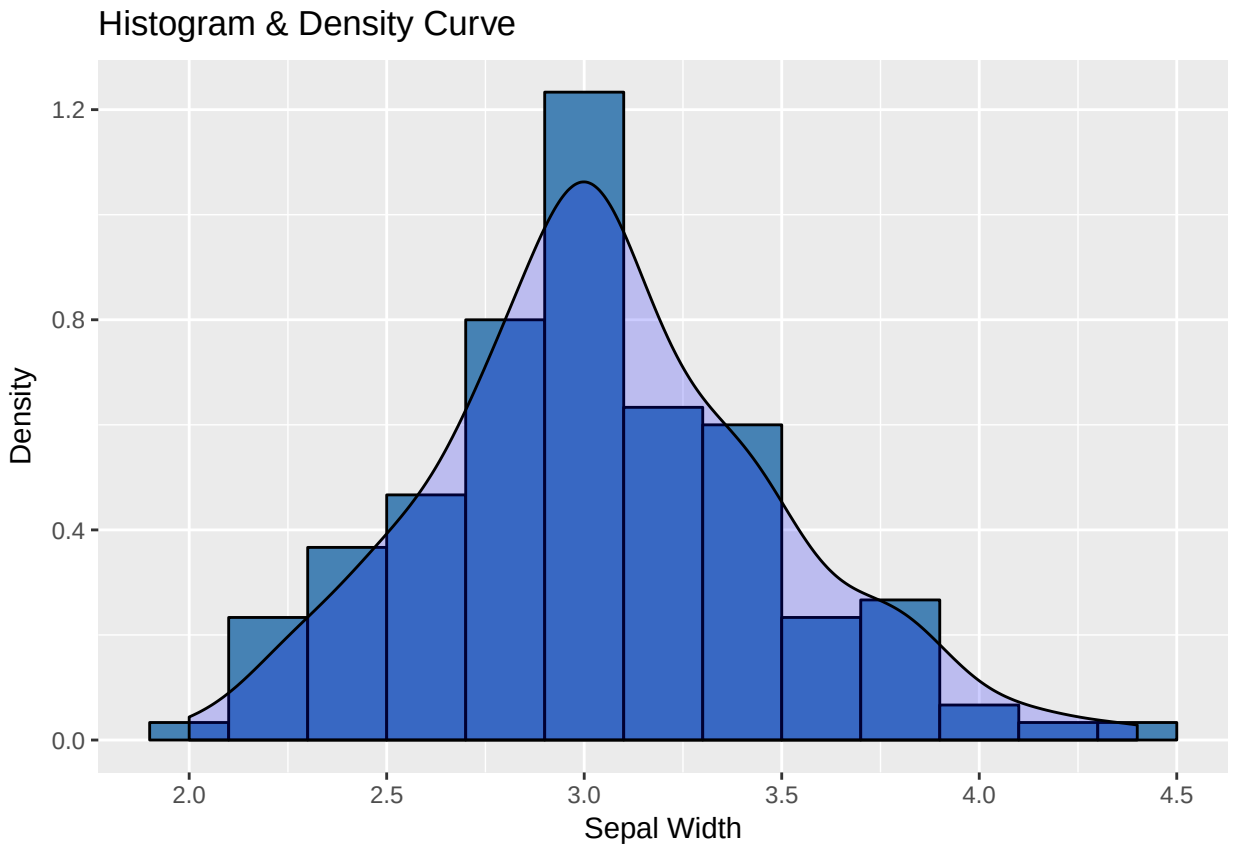
Histograms

Histograms can also be obtained with ggplot2:

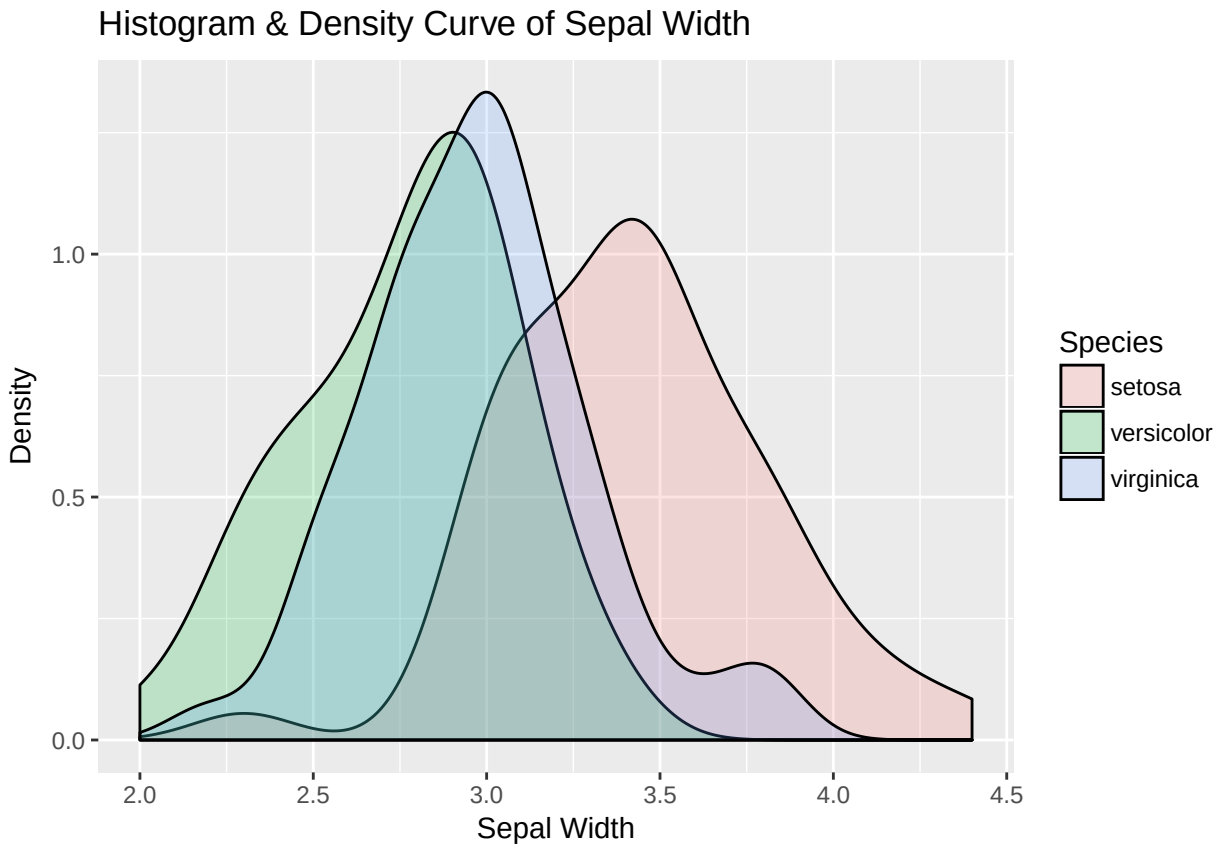
```
library(ggplot2)
histogram <- ggplot(data=iris, aes(x=Sepal.Width))
histogram + geom_histogram(binwidth=0.2, color="black", aes(fill=Species)) +
  xlab("Sepal Width") +  ylab("Frequency") + ggtitle("Histogram of Sepal Width")
```




```
density <- ggplot(data=iris, aes(x=Sepal.Width))
density + geom_histogram(binwidth=0.2, color="black", fill="steelblue", aes(y=..density..)) +
  geom_density(stat="density", alpha=I(0.2), fill="blue") +
  xlab("Sepal Width") +  ylab("Density") + ggtitle("Histogram & Density Curve")
```




```
density2 <- ggplot(data=iris, aes(x=Sepal.Width, fill=Species))
density2 + geom_density(stat="density", alpha=I(0.2)) +
  xlab("Sepal Width") +  ylab("Density") + ggtitle("Histogram & Density Curve of Sepal Width")
```


Histogram & Density Curve of Sepal Width
Density
1.0
0.5
0.0
2.0
2.5
3.0
3.5
4.0
4.5
Sepal Width
Species
setosa
versicolor
virginica