Before beginning the assignment I used the summary command in R to get and overview of the dataset.

The results show min, median and max values, quartiles and the mean for Petal/Sepal width and length. Also returned are the species names.

```
> summary (iris)
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
 Median :5.800   Median :3.000   Median :4.350   Median :1.300
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
       Species
 setosa    :50
 versicolor:50
 virginica :50
```

Q1.
To find the Mode in R the following command must be used
>my_mode <- table("Your Dataset here")
followed by the command to return results to the screen
>my_mode[which(my_mode ==max(my_mode))]

These commands will return the value that appears most frequently in the requested data set. For the Iris Data set the following values are represented most frequently. For the width and length of both Petal and Sepal, are represented above the frequency at which they appear.

```
> my_mode <- table(iris$Petal.Length)
> my_mode[which(my_mode ==max(my_mode))]

1.4 1.5
 13  13
>
> my_mode <- table(iris$Petal.Width)
> my_mode[which(my_mode ==max(my_mode))]
0.2
 29
> my_mode <- table(iris$Sepal.Length)
> my_mode[which(my_mode ==max(my_mode))]
 5
10
> my_mode <- table(iris$Sepal.Width)
> my_mode[which(my_mode ==max(my_mode))]
 3
26
>
```

The most common petal lengths are 1.4 and 1.5. They both appear 13 times. The petal width .2 appears 29 times. A sepal length of 5 appears 10 times and a sepal width of 3 appears 26 times.

The median is the middle value of the dataset when sorted in ascending order. To find the median in R we use the command shown below.
median("Your dataset here")

In the case of the iris dataset the median can be shown for both Petal and Sepal with and length.

```
> median(iris$Sepal.Length)
[1] 5.8
> median(iris$Sepal.Width)
[1] 3
> median(iris$Petal.Length)
[1] 4.35
> median(iris$Petal.Width)
[1] 1.3
>
```

If there are 15 values in a set the median will be the 8th value. In the iris set there are 50 of each species shown in the summary.

The mean or arithmetic mean is a value for the sum of all the numbers in the set divided by the count of all the numbers in the dataset. In R the command is shown below.

```
> mean(iris$Sepal.Width)
[1] 3.057333
> mean(iris$Sepal.Length)
[1] 5.843333
> mean(iris$Petal.Width)
[1] 1.199333
> mean(iris$Petal.Length)
[1] 3.758
>
```

The mean gives an average value for the data set but is not immune of being skewed by outliers. In the the iris data set it is clear to see that petal and sepal lengths are both greater than the relative widths. This is a common trait in plants and of course the three species of iris. The dataset allows for an understanding of average sizes in length and width. In iris petal length we had 2 returned values for the mode, which indicates that there are many petal lengths similar in size.

Q2.
The range is the difference between the lost value and the largest value in the set. In the iris data set it shows that the gap between the smallest petal length and largest is 5.9. This can cause skew in the results or perhaps it is normal for many different sizes to grow. The R command is shown below

range(iris$Sepal.Width)
min(iris$Sepal.Width)
max(iris$Sepal.Width)

```
> range(iris$Sepal.Width)
[1] 2.0 4.4
> min(iris$Sepal.Width)
[1] 2
> max(iris$Sepal.Width)
[1] 4.4
>
>
> range(iris$Sepal.Length)
[1] 4.3 7.9
> range(iris$Petal.Width)
[1] 0.1 2.5
> range(iris$Petal.Length)
[1] 1.0 6.9
>
```

The quartile shows values at 5 different places in the set on R. Normally just represented as 25% and 75% a quarter and three quarter points in the set. 0% and 100% are the values of the min and max in the data set.

```
> quantile(iris$Sepal.Length)
  0%  25%  50%  75% 100%
 4.3  5.1  5.8  6.4  7.9
> quantile(iris$Sepal.Width)
  0%  25%  50%  75% 100%
 2.0  2.8  3.0  3.3  4.4
> quantile(iris$Petal.Length)
  0%  25%  50%  75% 100%
1.00 1.60 4.35 5.10 6.90
> quantile(iris$Petal.Width)
  0%  25%  50%  75% 100%
 0.1  0.3  1.3  1.8  2.5
>
```

As suggested previously we can see that 75% for petal length is 5.10 which shows it is not uncommon that a some petal lengths can grow much larger than others. At 75% the difference is still 4.1 from the minimum, about right for the dataset.

The variance relates to the spread of data. The data set has many values but the variance describes the measurement between the values derived from the mean. The command in R is shown below

var(iris$Sepal.Length)

```
> var(iris$Sepal.Length)
[1] 0.6856935
> var(iris$Sepal.Width)
[1] 0.1899794
> var(iris$Petal.Length)
[1] 3.116278
> var(iris$Petal.Width)
[1] 0.5810063
```

The iris dataset shows that there is not a huge difference between Sepal and Petal width individually. However what can be seen is that the Petal length can vary greatly up to 3 times the length of the smallest petal is common.

The standard deviation is used to express the deviation from the mean. This can be positive or negative. A normal frequency distribution normally has observation with +/-1 of the mean. The command in R is shown below

sd(iris$Sepal.Length)

```
>
> sd(iris$Sepal.Length)
[1] 0.8280661
> sd(iris$Sepal.Width)
[1] 0.4358663
> sd(iris$Petal.Length)
[1] 1.765298
> sd(iris$Petal.Width)
[1] 0.7622377
>
```

The picture is becoming clearer regarding the Petal Length. It is without a doubt the value that deviates greatest in values with a large deviation of 1.7. With approximately 95% of all observations falling with 2 standard distributions of the mean it appears that the petal data may not be skewed by outliers.

The Z score shows how far from the mean a value is based on the number of standard deviations. The command in R is shown below.

x<-((iris$Sepal.Width)-mean(iris$Sepal.Width))/sd(iris$Sepal.Width)

```
> x<-((iris$Sepal.Width)-mean(iris$Sepal.Width))/sd(iris$Sepal.Width)
> x
  [1]  1.01560199 -0.13153881  0.32731751  0.09788935  1.24503015  1.93331463
  [7]  0.78617383  0.78617383 -0.36096697  0.09788935  1.47445831  0.78617383
 [13] -0.13153881 -0.13153881  2.16274279  3.08045544  1.93331463  1.01560199
 [19]  1.70388647  1.70388647  0.78617383  1.47445831  1.24503015  0.55674567
```

Q3.
The skew shows a numerical representation for the symmetry of the dataset. A value closer to 0 has a more symmetrical spread of values. A value far from 0 is a skew meaning outliers are present in the dataset. The command in R is shown below

skewness(iris$Sepal. Width)

```
> skewness(iris$Sepal.Width)
[1] 0.3157671
> skewness(iris$Sepal.Length)
[1] 0.3117531
> skewness(iris$Petal.Width)
[1] -0.1019342
> skewness(iris$Petal.Length)
[1] -0.2721277
>
```

After using the skew command in R on the dataset we can see that the skew is rather low, all numbers below .5 irrespective of their sign. Again width Petal Length it appears that there are no outliers and the data is spread evenly. It confirms that petal legth can vary greatly but there are many petals that grow as small as others grow large.

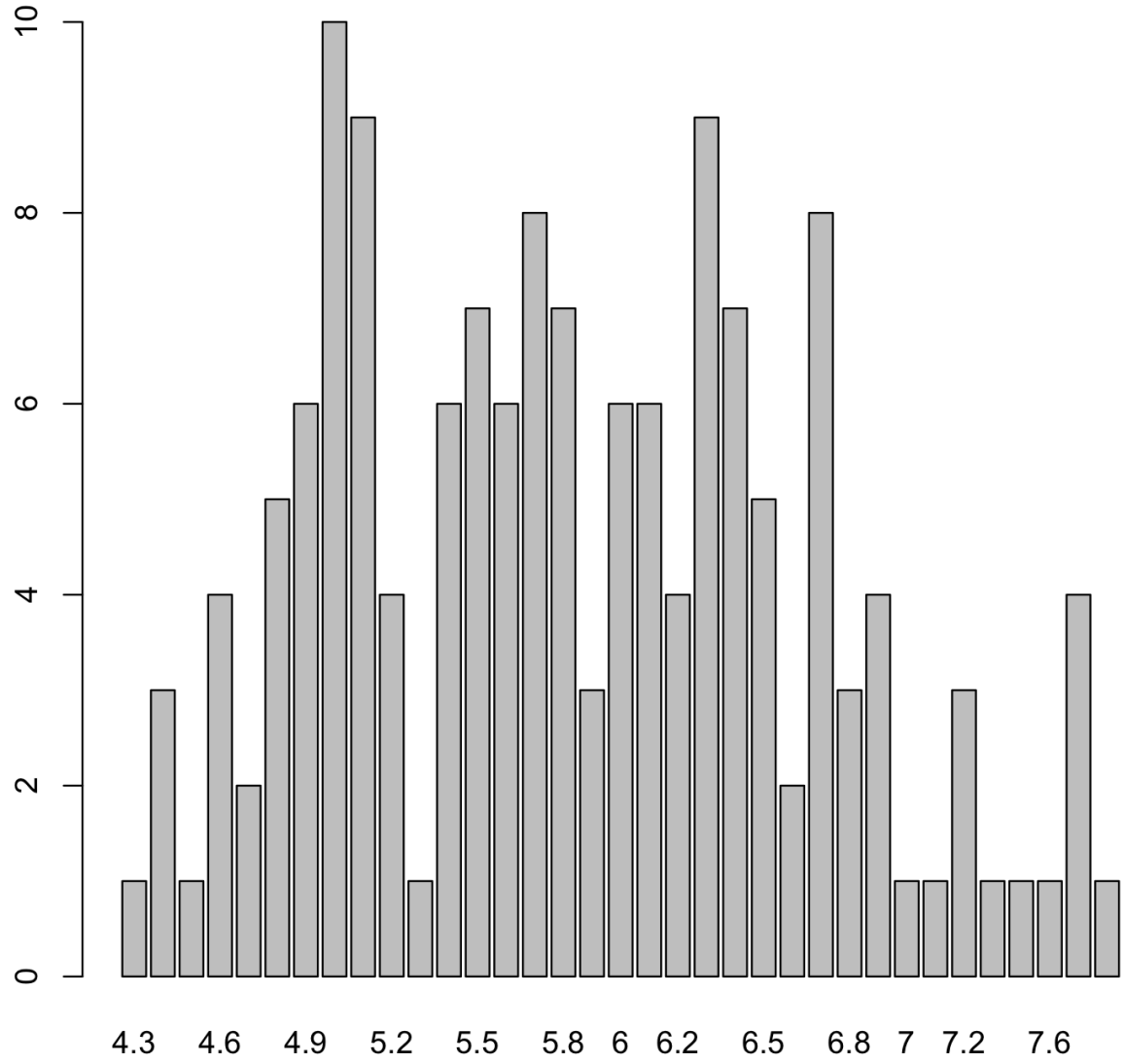
Kurtosis relates to the peak of distribution. Having a high Kurtosis score means that many of the dataset values are close to the mean and will cause a plot to have a pointed peak. A low kurtosis score will have many values spread out from the mean causing the peak to be flat. The command for the kurtosis score is shown below.

skewness(iris$Sepal. Width)

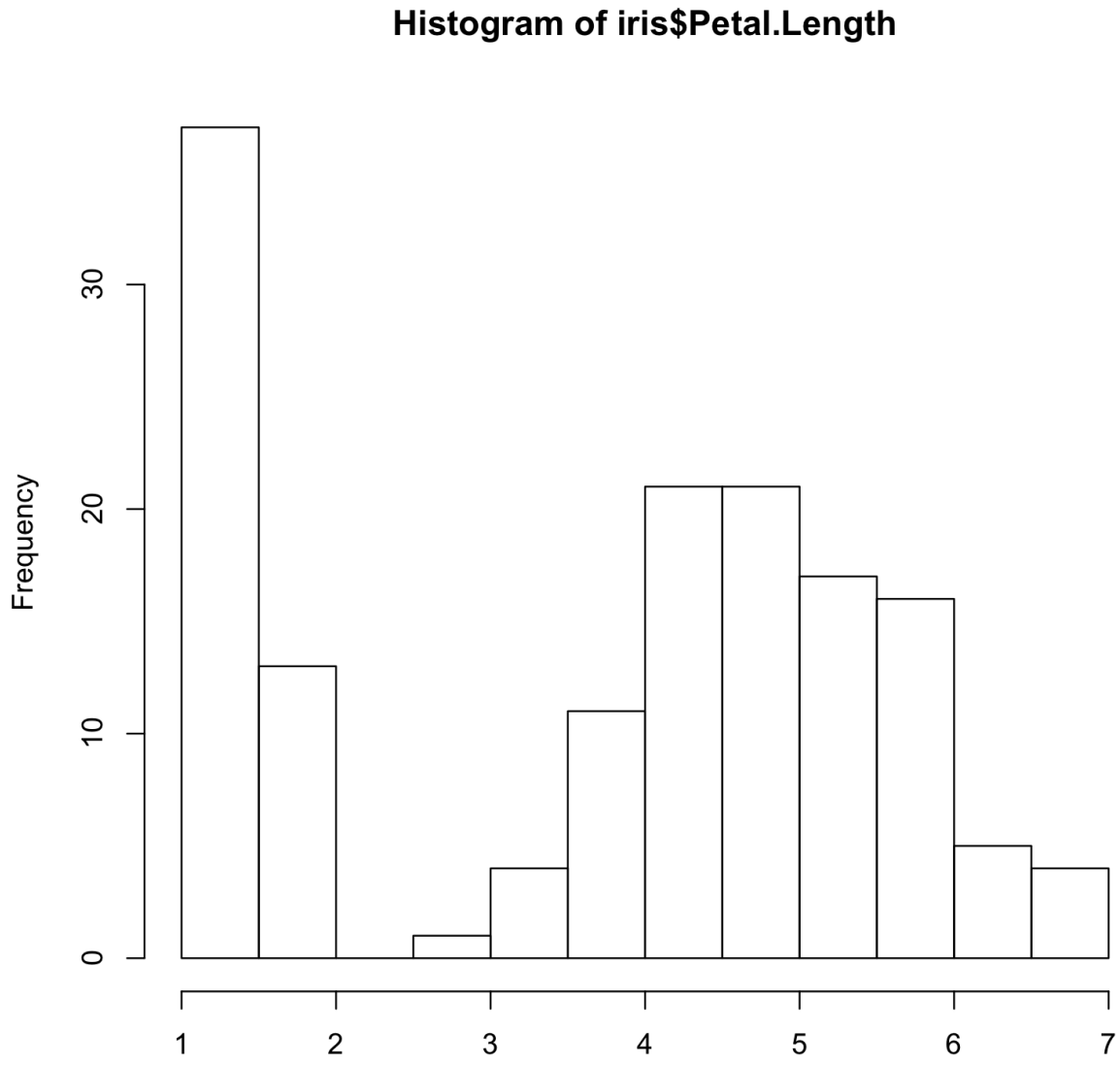
```
> kurtosis(iris$Sepal.Width)
[1] 3.180976
> kurtosis(iris$Sepal.Length)
[1] 2.426432
> kurtosis(iris$Petal.Width)
[1] 1.663933
> kurtosis(iris$Petal.Length)
[1] 1.604464
>
```

What is evident in the results of the kurtosis score for the dataset is that the Sepal width does not vary greatly from the mean. The size of the sepal is very close in all values. The rest of the data is spread out with a flatter peak but still the Sepal Length doesn't vary too much either. The real variable values occur in the petals.

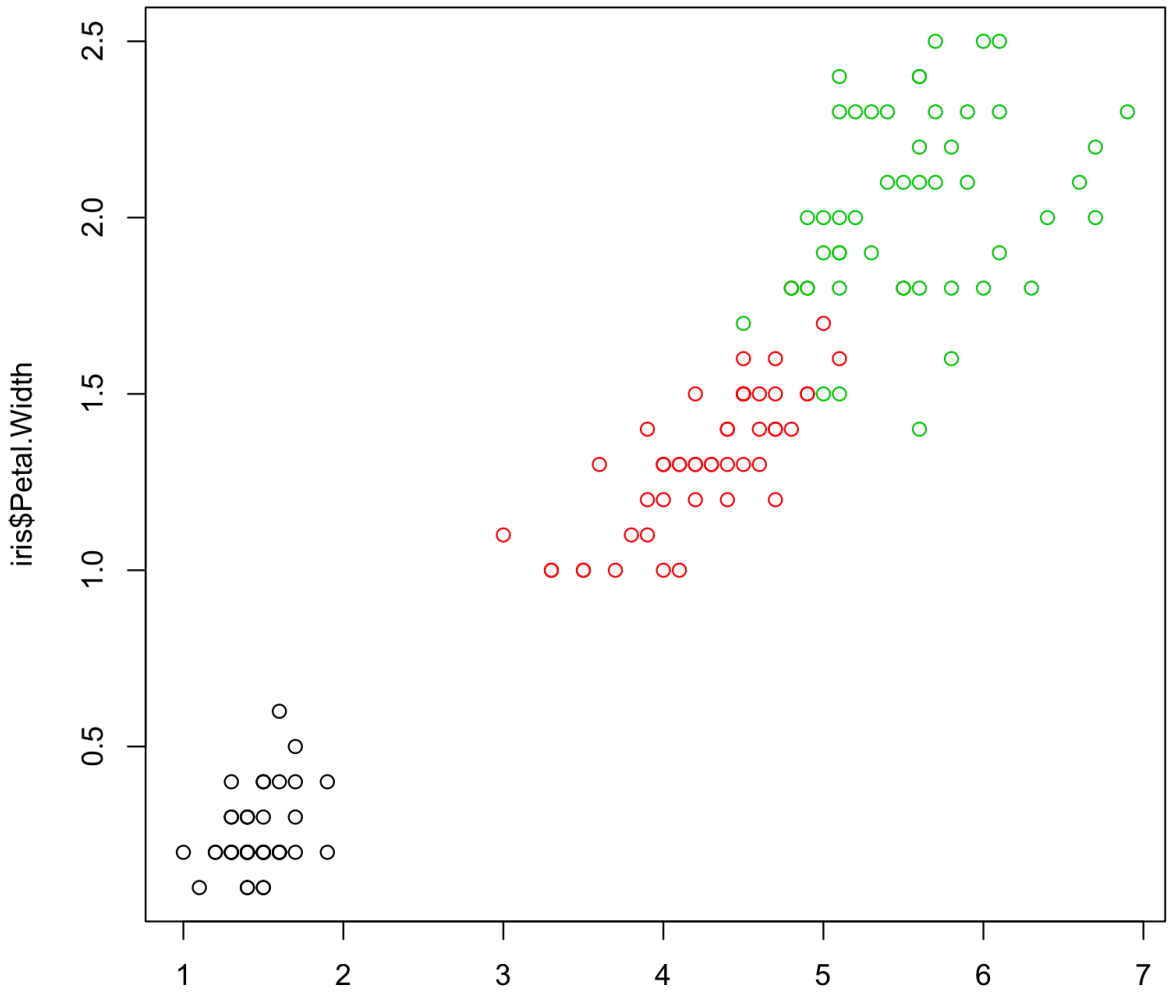
Q4.
Bar plot
barplot(table(iris$Sepal.Length))



Histogram
hist(iris$Petal.Length)



Scatter plot
plot(x=iris$Petal.Length, y=iris$Petal.Width, col=iris$Species)



Boxplot
boxplot(iris[,1:4])