Module - 4:

Statistical Analysis Using R:

In [ ]:

In [1]:

```
install.packages("e1071")
```

```
  There is a binary version available but the source version is later:
       binary source needs_compilation
e1071  1.7-6 1.7-12              TRUE

  Binaries will be installed
package 'e1071' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\nanth\AppData\Local\Temp\RtmpwNZvjN\downloaded_packages
```
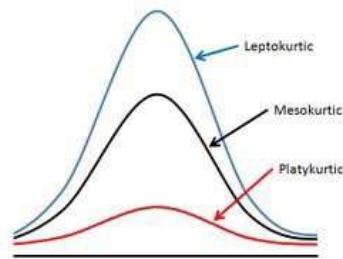
Kurtosis:

```
            * Kurtosis measure the shape of the frequency curve
            * It will measure the degree of the peakness of a frequency distribution

     Types of Kurtosis:
            * Leptokurtic (value > 0 )
            * Platykurtic (value < 0 )
            * Mesokurtic  (value == 0)
```


In [2]:

```
# install.packages("e1071")
```

In [9]:

```
library(e1071)
```

```
Warning message:
"package 'e1071' was built under R version 3.6.3"
```
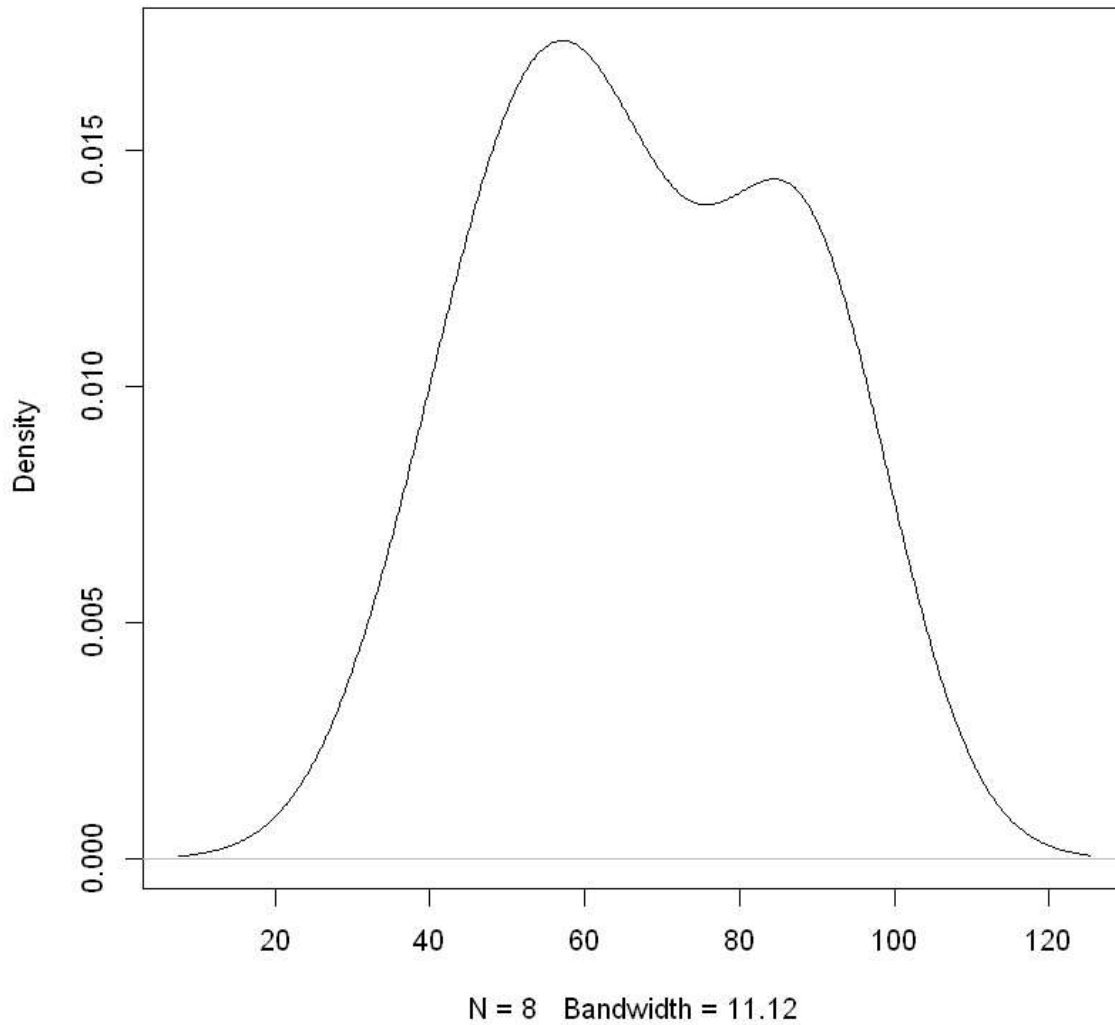
In [18]:

```r
x=c(41, 52, 53, 64, 65, 86, 86, 92)
a=kurtosis(x)
print(a)

plot(density(x))
```

[1] -1.790467

density.default(x = x)

N = 8   Bandwidth = 11.12

In [23]:

```r
# using if and else if condition:

print(paste("The Kurtosis value: ", a))

if(a<0){
    print("Platykurtic")
} else if(a>0){
    print("Leptokurtic")
} else if(a==0){
    print("Mesokurtic")
}
```

[1] "The Kurtosis value: -1.79046683946946"
[1] "Platykurtic"

HYPOTHESIS:

Hypothesis is a educated guess about something in the world around you. It should be testable, either by experiment or observation.

        Types of Hypothesis:
            * Alternative Hypothesis (negative form)
            * Null Hypothesis

    CASE 1: A new medicine you think might work
        Alternative = A new medicine will not work
        Null =

        * If the calculated value is greater than the critical value, we have to reject the null hypothesis.
        * If the calculated value is lesser than the critical value, then we accept the null hypothesis.

        Notation:
            * Alternative Hypothesis = Ha
            * Null Hypothesis = Ho

Problem:

It is believed that a candy making machine makes chocolate bars that are 5g in weight. A worker claims that the machine after maintenance no longer makes 5g.
Write down Ho and Ha :

    -> Ho (NULL) = Machine makes chocolate that are 5g in weight.
    -> Ha (ALTERNATIVE) = Machine doesn't make chocolate that are 5g in weight.

Level of Significance:

        * Default significance level = 0.05
        * A 5% risk of concluding that a difference exists whent here is no actual difference
        * Significance level also denoted as alpha or (), is the probability of rejecting the null
          hypothesis when it is true.

In [ ]:

Choosing Statistical Test:

        The Type of data user have:
                * Categorical Data
                * Quantitative Data

        Number of Samples we have:
                * One Sample - Making comparison against global value   --|
                * Two Sample - Comparing one group to other group   --|-- (do T Test)
                * Two Sample Special - One sample with different measurements   --|
                * Three or More Sample - (do Anova Test)

T- Test:

```
T-Test will tell the significance difference between the groups

When to use T-Test:
        * Data are Independent
        * Datas are noramlly distributed (check distribution by applying skewness)

Types of T-Test:
        * Paired T-Test (If the groups come from a single population)
        * Two-Sample T-Test (If the groups come from two different)
        * One-Sample T-Test (If there is one group being compared against a standard value)

T-Test:
    * One-tailed T-Test
    * Two-tailed T-Test

    j
```

In [ ]: