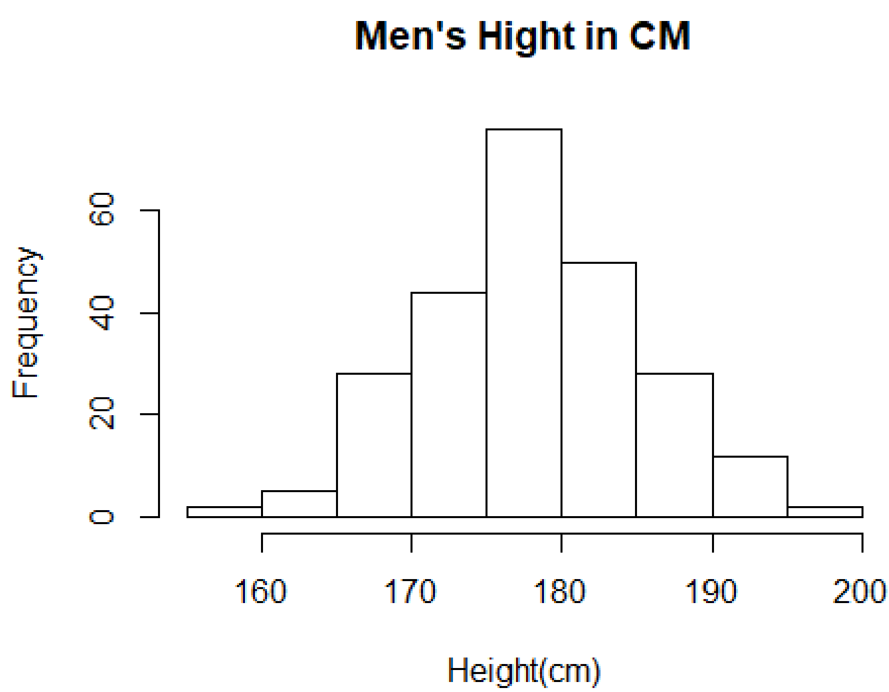
Lab4

```r
download.file("http://www.openintro.org/stat/data/bdims.RData", destfile =
"bdims.RData")
load("bdims.RData")
mdims <- subset(bdims, bdims$sex == 1)
fdims <- subset(bdims, bdims$sex == 0)
mhgtmean <- mean(mdims$hgt)
mhgtsd <- sd(mdims$hgt)
```

Exercise 1: Generate separate histograms of the men's and women's heights. Then, compare and contrast the center, spread, and shape of these two height distributions.

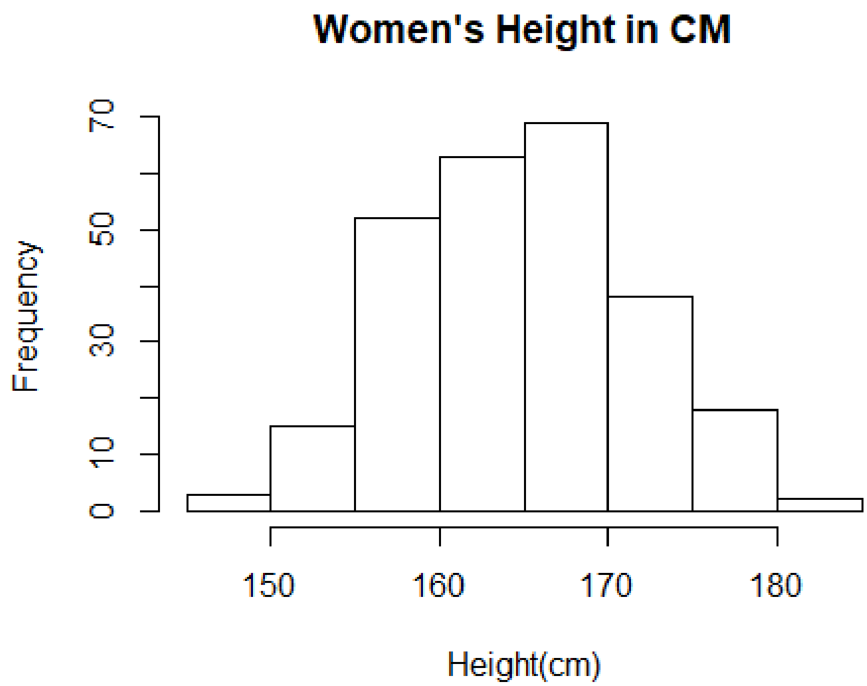
```r
#Male Height
hist(mdims$hgt, main = "Men's Hight in CM", xlab = "Height(cm)")
```




```r
summary(mdims$hgt) # 5 Number Summary of Men's Height in cM.

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   157.2   172.9   177.8   177.7   182.7   198.1

#Female Height
hist(fdims$hgt, main = "Women's Height in CM", xlab = "Height(cm)")
```



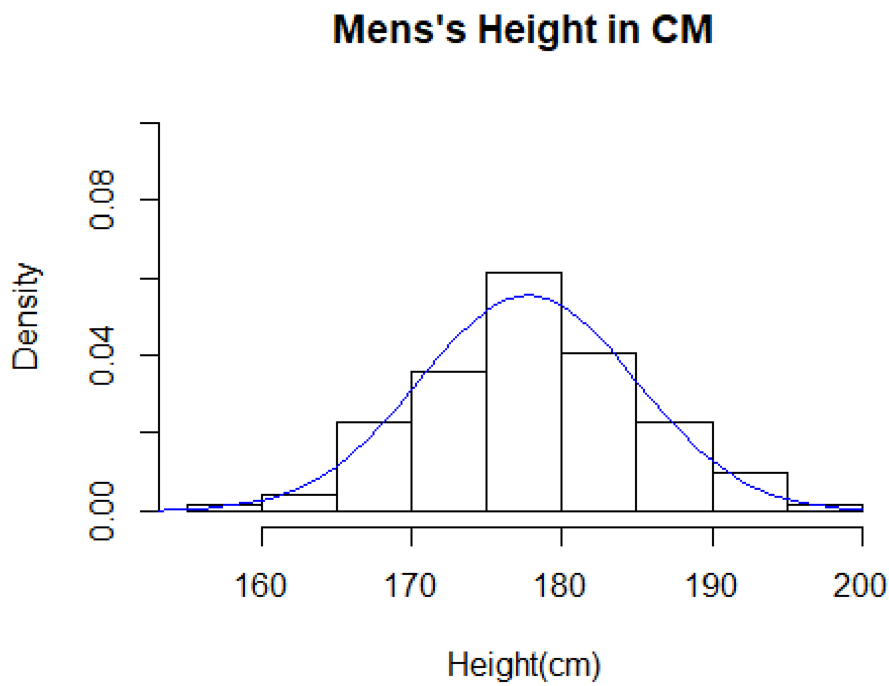

```r
summary(fdims$hgt) # 5 Number Summary of Women's Height in cM.

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   147.2   160.0   164.5   164.9   169.5   182.9
```

The men's distribution appears to be symmetical while the womens distribution is skewed left. The Men's median is 177.8 while the Women's is 164.5.

Exercise 2: Based on this plot, does it appear that the men's height data follow a nearly normal distribution? Explain.

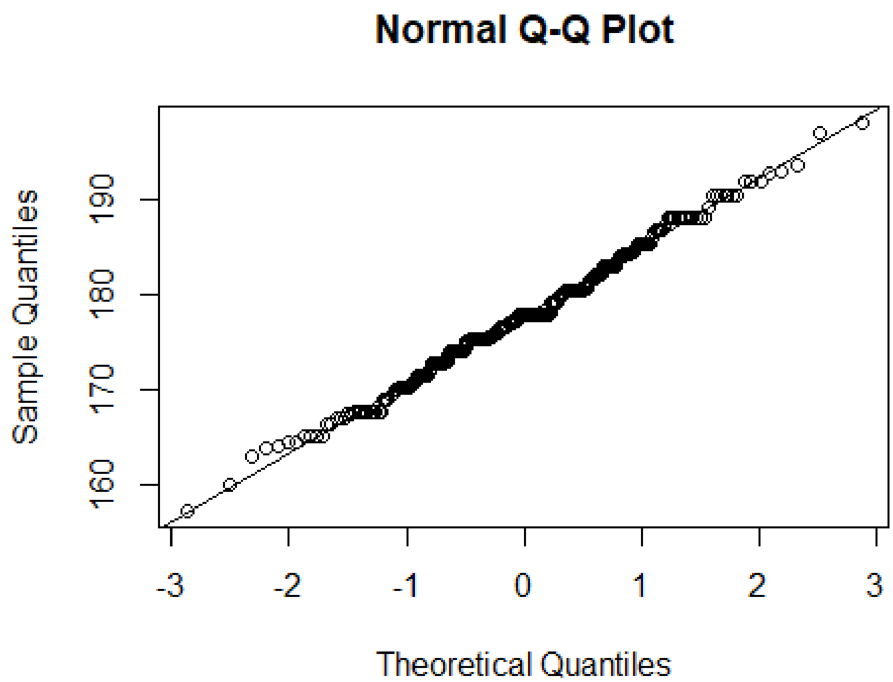
```r
hist(mdims$hgt, main = "Mens's Height in CM", probability = TRUE,
xlab="Height(cm)", ylim = c(0,0.1))
x <- 150:200
y <- dnorm(x = x, mean = mhgtmean, sd = mhgtsd)
lines(x = x, y = y, col = "blue")
```



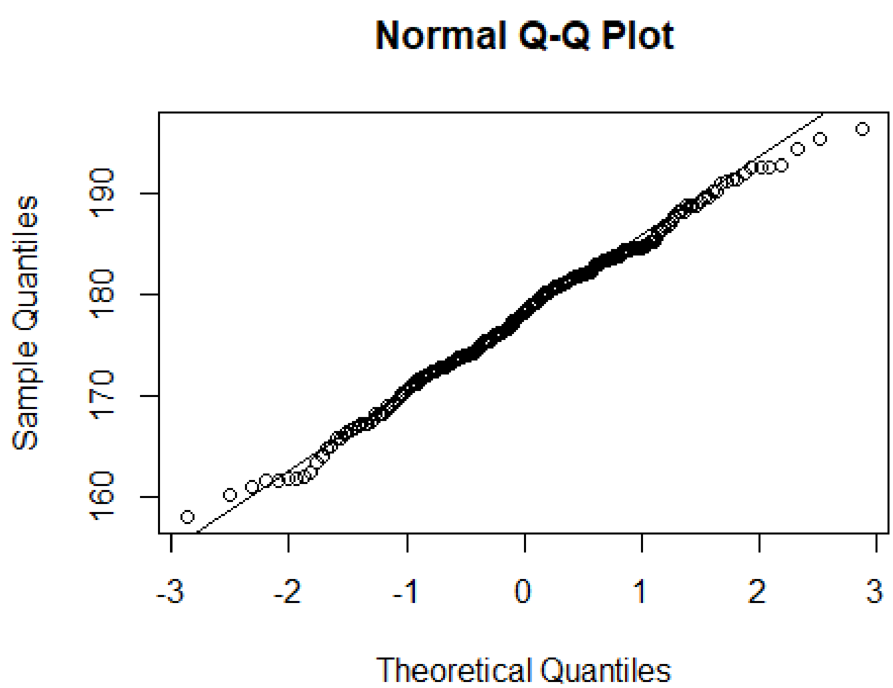
Based on this plot, the data follows a normal distribution. It follows the curve that has been imprinted from the dnorm command

Exercise 3: Make a normal probability plot of sim_norm. Do all of the points fall on the line? How does this plot compare to the probability plot for the actual data?

```
qqnorm(mdims$hgt)
qqline(mdims$hgt)
```



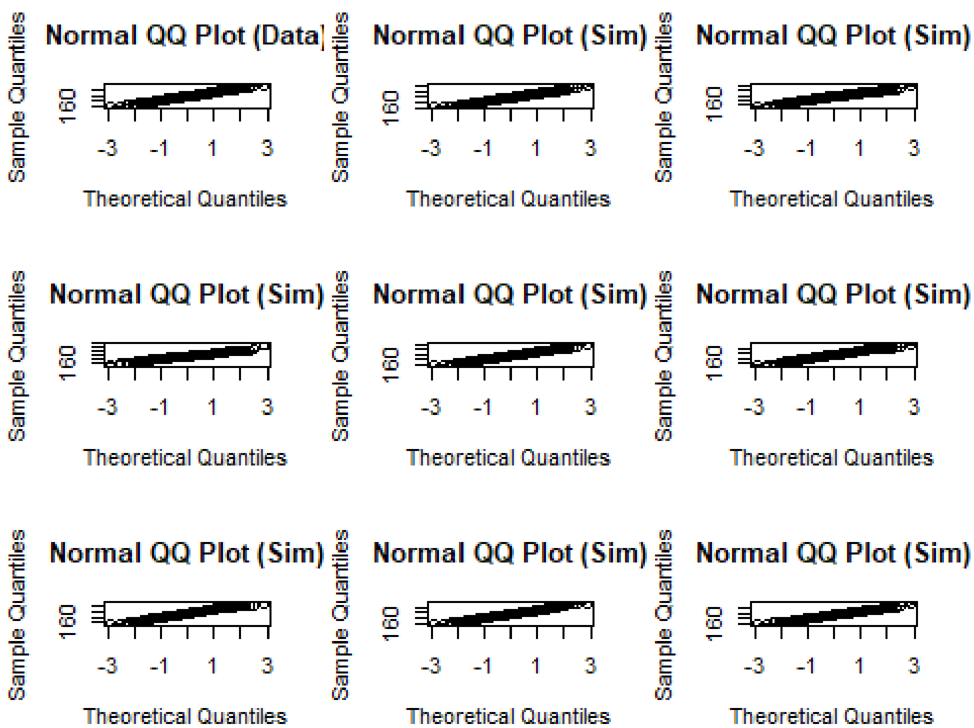

```
sim_norm <- rnorm(n = length(mdims$hgt), mean = mhgtmean, sd = mhgtsd)
qqnorm(sim_norm)
qqline(sim_norm)
```



When comparing the simulated data most of the points fall either near or close o the line. The simulated data is very close to the actual data.

Exercise 4: Does the normal probability plot for mdims$hgt look similar to the plots created for the simulated data? That is, do the plots provide evidence that the male heights are nearly normal? Explain.

```
qqnormsim(mdims$hgt)
```

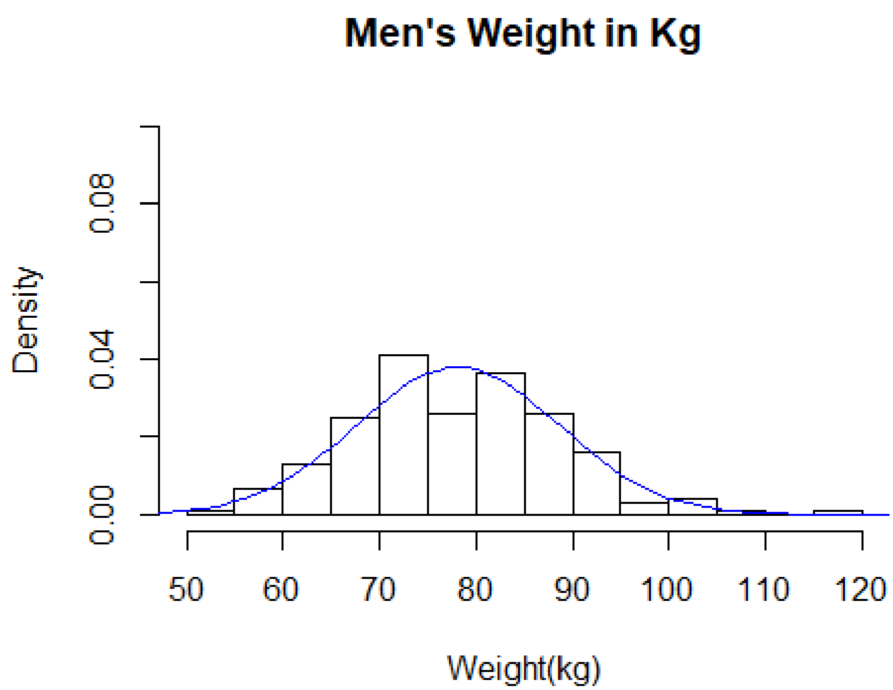


###The normal probability plot of mdins$hgt does look similar to the plots that were created for the simulated data. The plots do provide evidentce that the males heights are normal. Most of the points fall either on the line on very close to it.

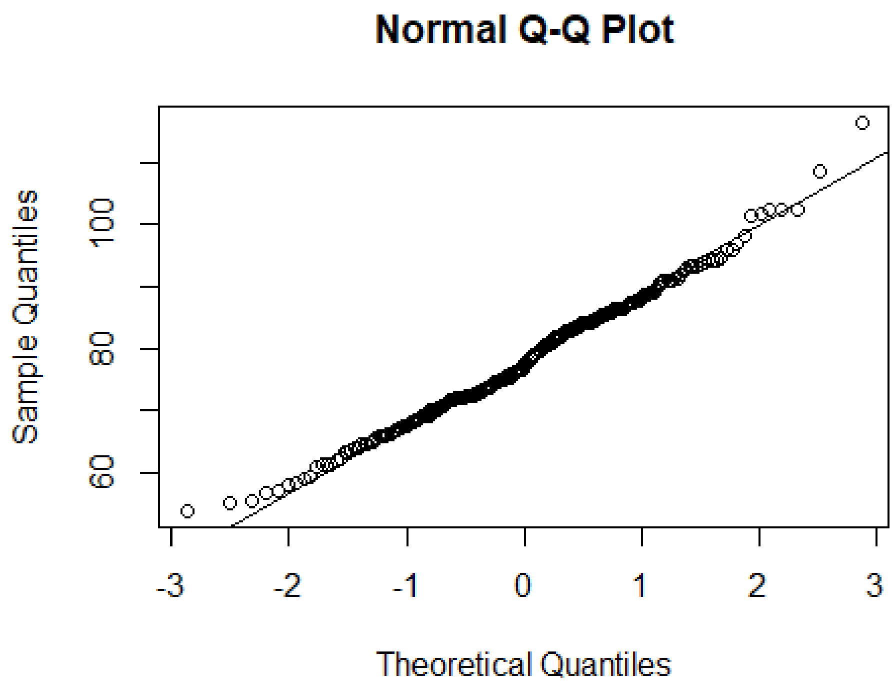
Exercise 5: Using the same procedure you used to judge the normality of the male height data in Exercises 2 through 4, explain your judgment as to whether or not the male weights appear to come from a normal distribution.

```
mwgtmean <- mean(mdims$wgt)
mwgtsd <- sd(mdims$wgt)
hist(mdims$wgt, main="Men's Weight in Kg", probability = TRUE, xlab =
"Weight(kg)", ylim = c(0,0.1))
x <- 0:1000
```

```
y <- dnorm(x = x, mean = mwgtmean, sd = mwgtsd)
lines(x = x, y = y, col = "blue")
```




```
qqnorm(mdims$wgt)
qqline(mdims$wgt)
```




```
summary(mdims$wgt)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   53.90   70.95   77.30   78.14   85.50  116.40
```

The male's weights seem to have a scewed right distribution.

Write out two probability questions that you would like to answer - one regarding male heights and one regarding male weights. Calculate those probabilities using both the theoretical normal distribution as well as the empirical distribution (four probabilities in all). Which variable, height or weight, had a closer agreement between the theoretical and empirical probabilities? Explain your reasoning.

```
#What is the probability of a male's height being greater than 5 foot 9
inches (175.26 cm)
1 - pnorm(q = 175.26, mean = mhgtmean, sd = mhgtsd)

## [1] 0.6353186

sum(mdims$hgt > 175.26)/length(mdims$hgt)

## [1] 0.6801619
```

```r
#What is the probability of a female's height being greater than 5 foot 9
inches (175.26 cm)
1 - pnorm(q = 175.26, mean = mean(fdims$hgt), sd = sd(fdims$hgt))

## [1] 0.05623195

sum(fdims$hgt > 175.26)/length(fdims$hgt)

## [1] 0.06538462
```

The two questions I choose was what was the probabilty of either a male's or female's height being greater than 5 foot 9 inches. Both the therorial and empirical probabilities were reasonibly close. For males, the empirical distrubution number was .635 while the empirical distribution number was .68. However with the female's empirical distribution number of .056 and the empirical distribution number was .065. Males are more likely to be taller than 5 foot 9 inches but females are not as likely.

Homework Assignment

1.1. D

1.2. A

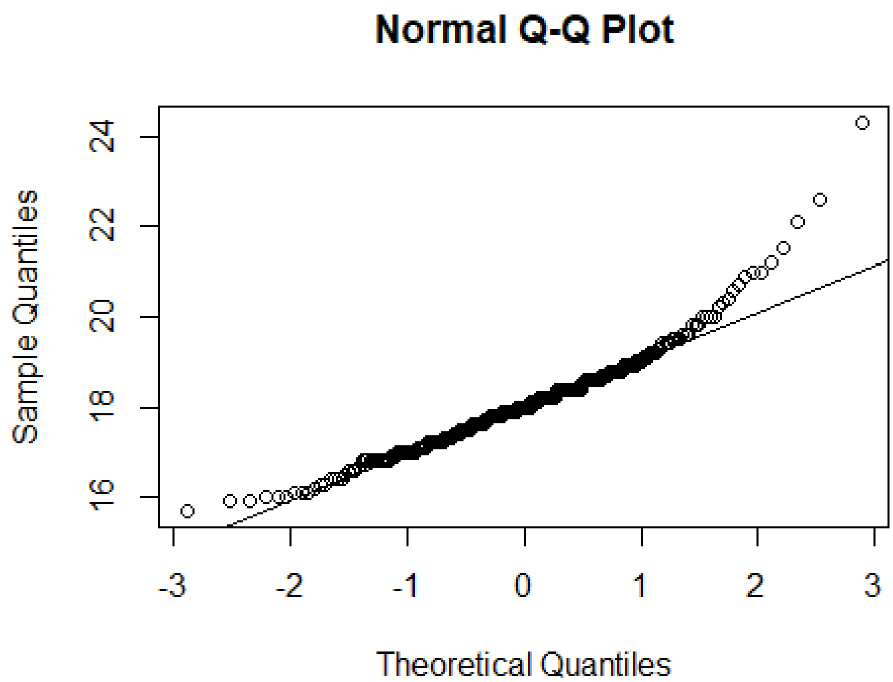
1.3. B

1.4. C

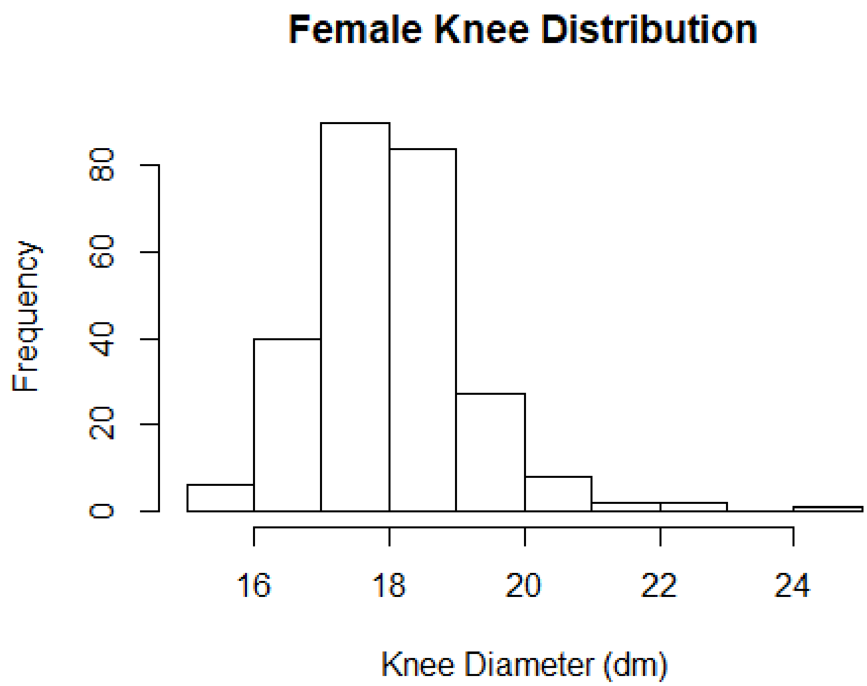
2. Age is a discrete variable and will produce a stepwise pattern.

As you can see, normal probability plots can be used both to assess normality and visualize skewness.Produce a normal probability plot for female knee diameter (kne.di). Based on this normal probability plot, is this variable left skewed, symmetric, or right skewed? Explain your reasoning. Use a histogram to confirm your findings.

```r
qqnorm(fdims$kne.di)
qqline(fdims$kne.di)
```




```r
hist(fdims$kne.di, main = "Female Knee Distribution", xlab = "Knee Diameter (dm)")
```



###With the nomral distribution and using the histogram, we can conclude tht the female knee diameter is skewed right.