Lab for Week 2 - Normal Distribution

Name

Date of lab session

Lab report

```
download.file("http://www.openintro.org/stat/data/bdims.RData", destfile = "bdims.RData")
load("bdims.RData")
```

Load data:

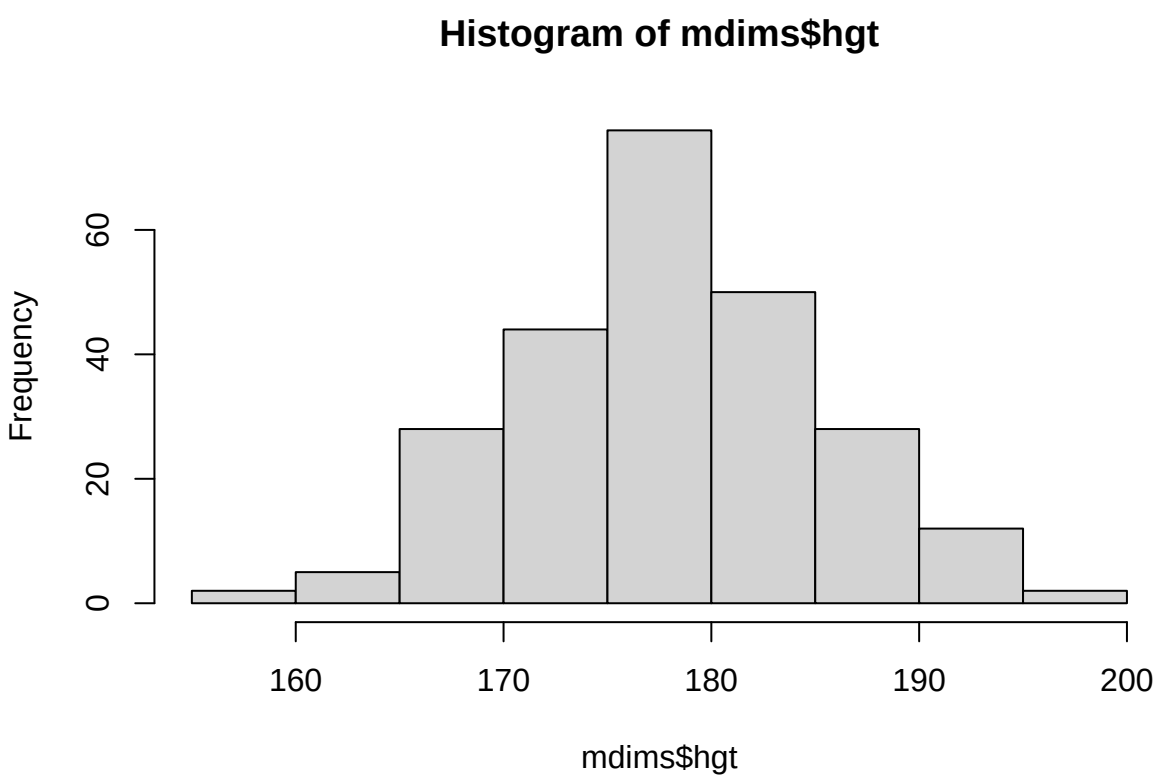
```
set.seed(14531)
```

Set a seed:

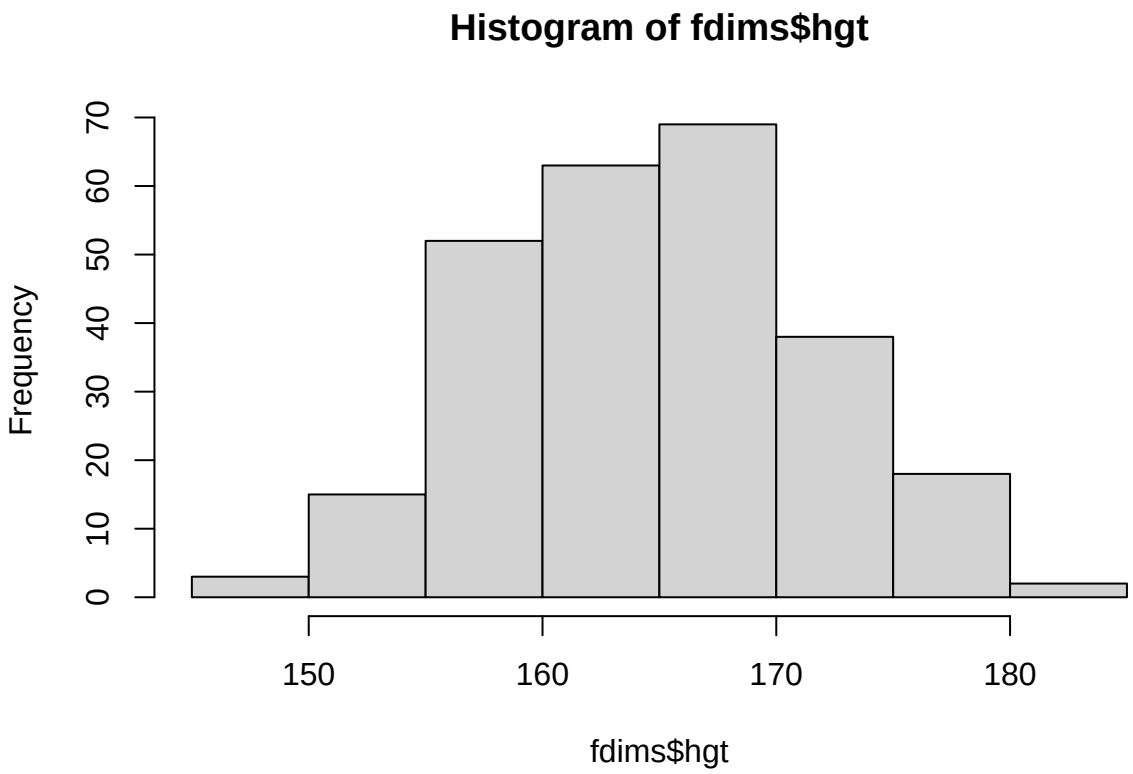
Exercises:

Exercise 1: The height for men seem to be more evenly distributed than the height of women. We can also see that the women's heights were concentrated while it was more distributed in the mens.

```
mdims <- subset(bdims, sex == 1)
fdims <- subset(bdims, sex == 0)
hist(mdims$hgt)
```

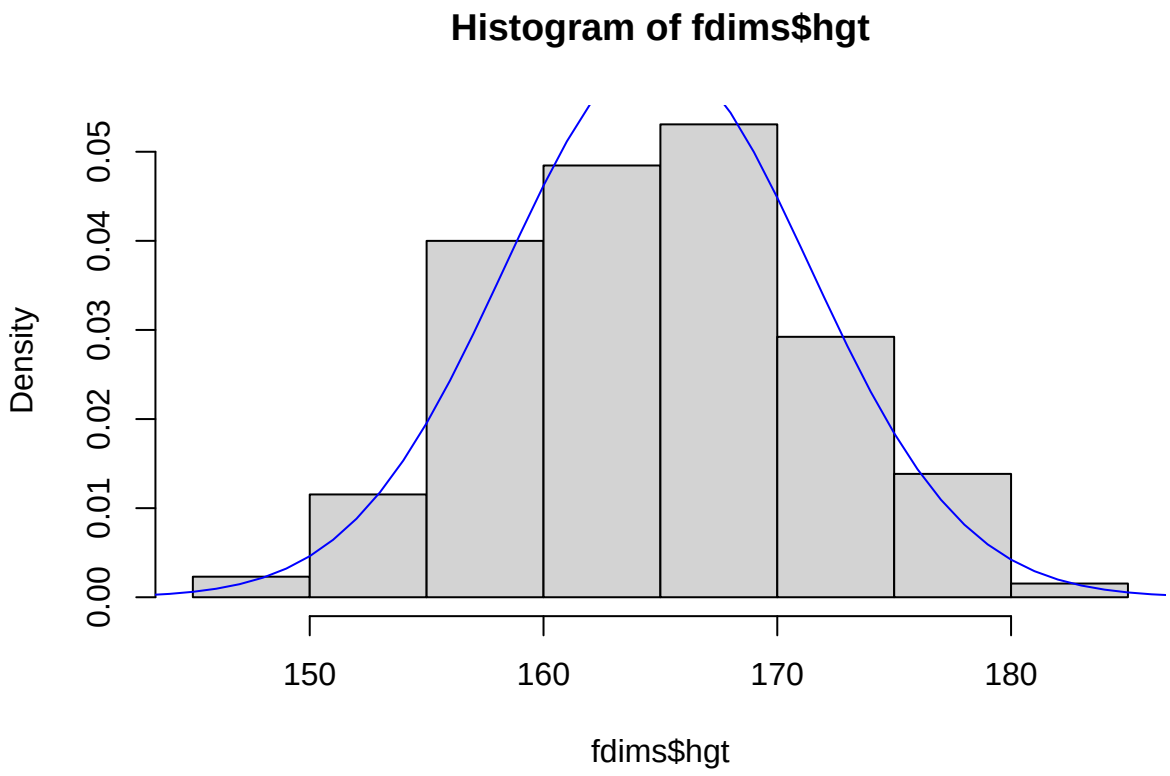



```
hist(fdims$hgt)
```



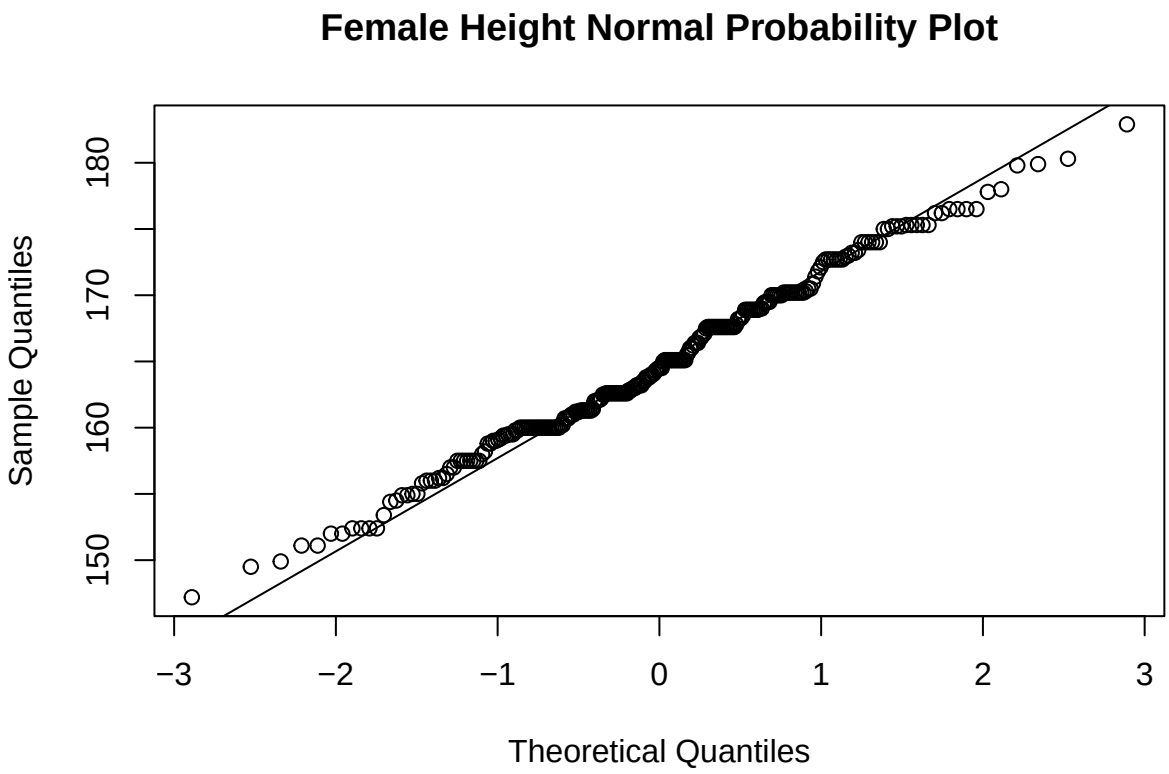
Exercise 2: The plot shows that the distribution does follow a nearly normal distribution.

```
fhgtmean <- mean(fdims$hgt)
fhgtsd   <- sd(fdims$hgt)
hist(fdims$hgt, probability = TRUE)
x <- 140:190
y <- dnorm(x = x, mean = fhgtmean, sd = fhgtsd)
lines(x = x, y = y, col = "blue")
```

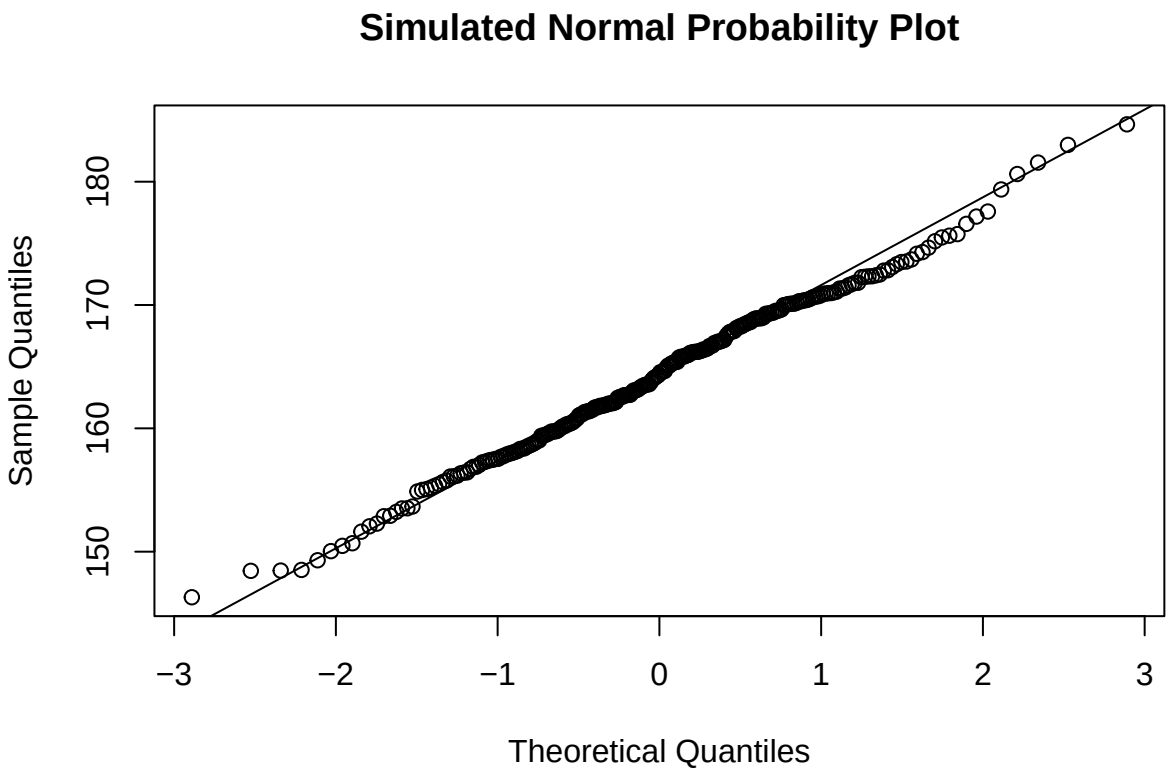


Exercise 3: Not all points are on the line, however, it does allign with the line much better than the real data.

```
qqnorm(fdims$hgt, main = "Female Height Normal Probability Plot")
qqline(fdims$hgt)
```

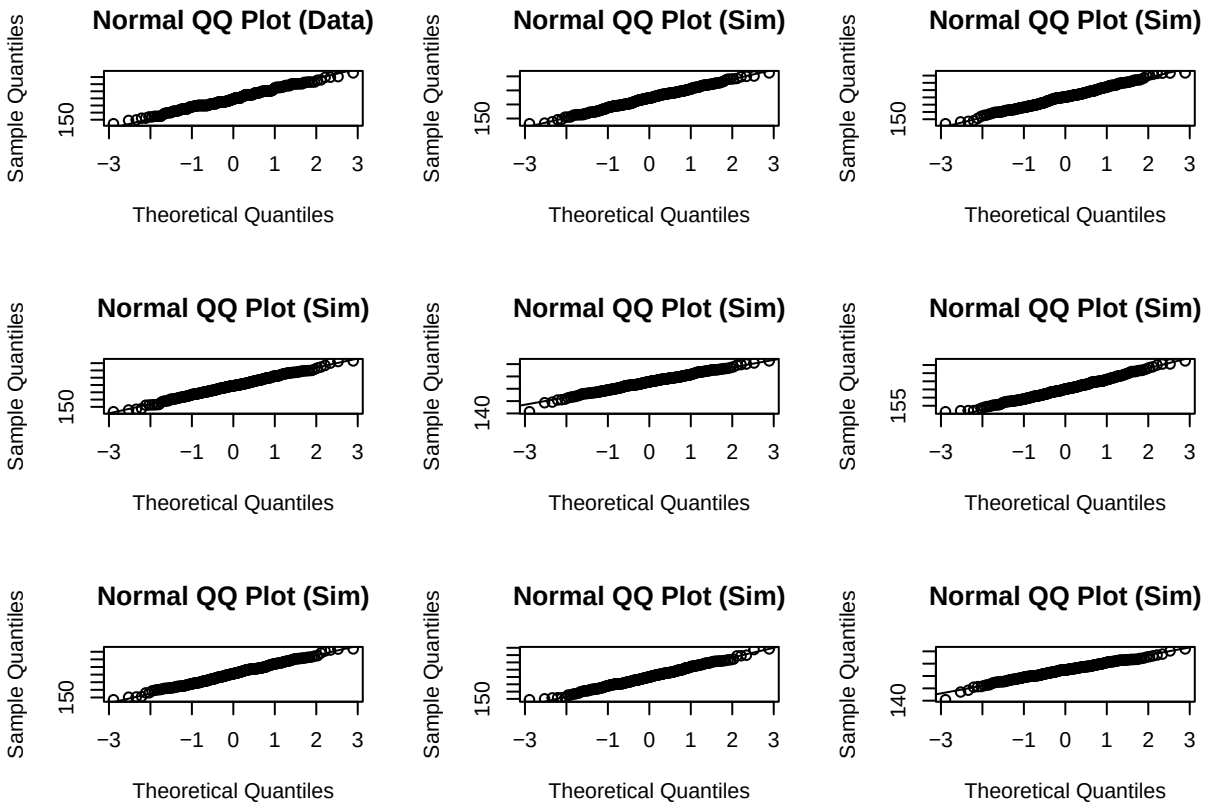



```r
sim_norm <- rnorm(n = length(fdims$hgt), mean = fhgtmean, sd = fhgtsd)
qqnorm(sim_norm, main = "Simulated Normal Probability Plot")
qqline(sim_norm)
```



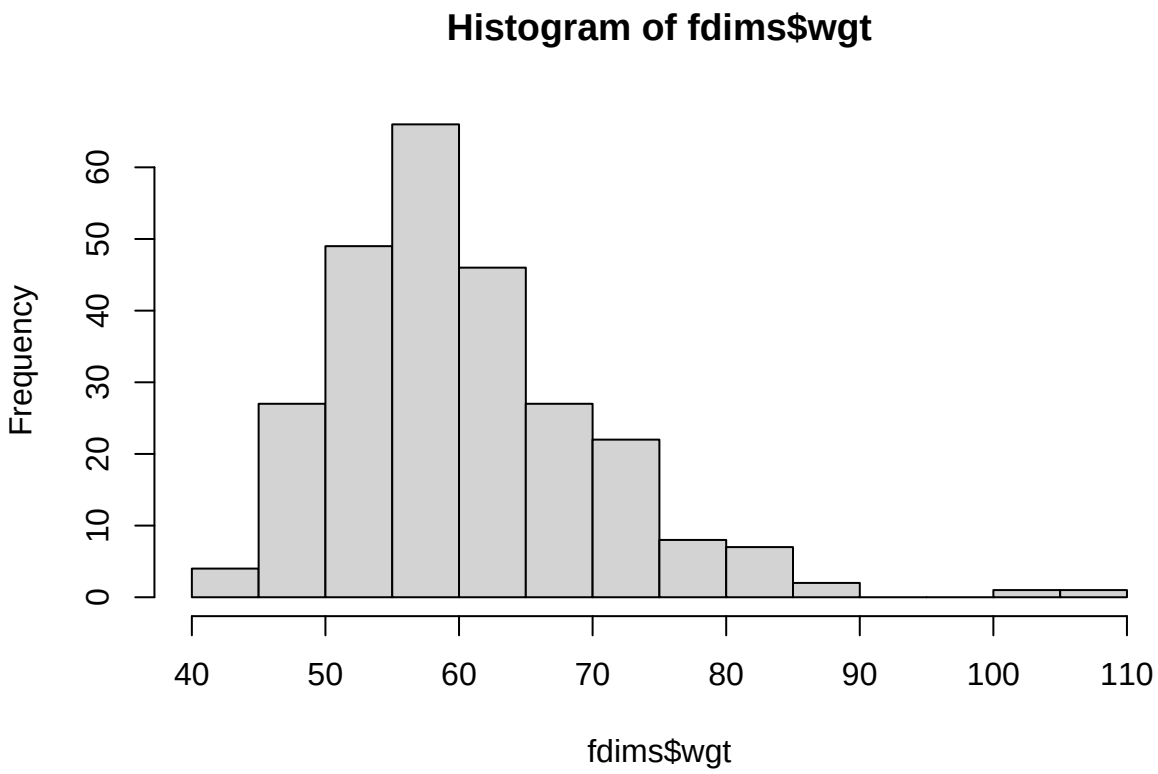
Exercise 4: The simulated plots do follow a similar line as the actual data. Therefore, plots do provide that the female height roughly follow a nearly normal distribution.

```
qqnormsim(fdims$hgt)
```

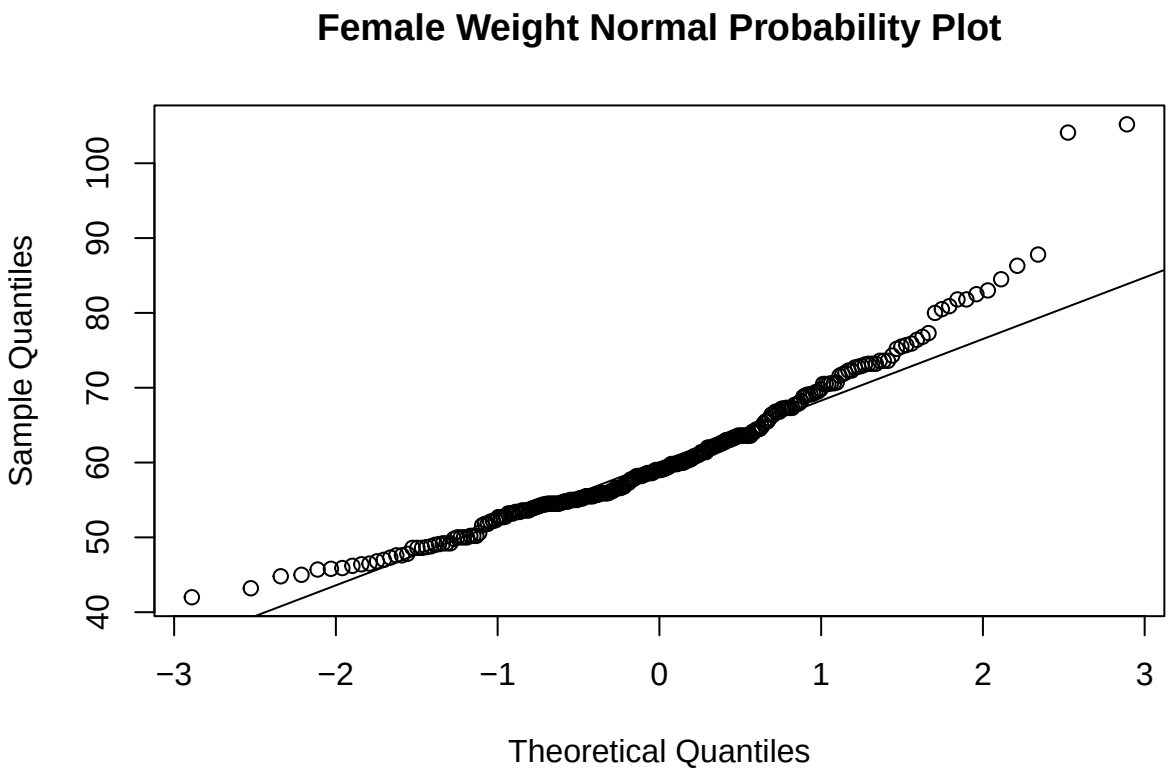


Exercise 5: Female weight does seem to follow a near normal distribution. The histogram isskewed right and the normal probability plot does not follow the line, but rather curves a slight bit. However, the sample has more than 30 data points, with no extreme outliers, so we can say that it is nearly a normal distribution.

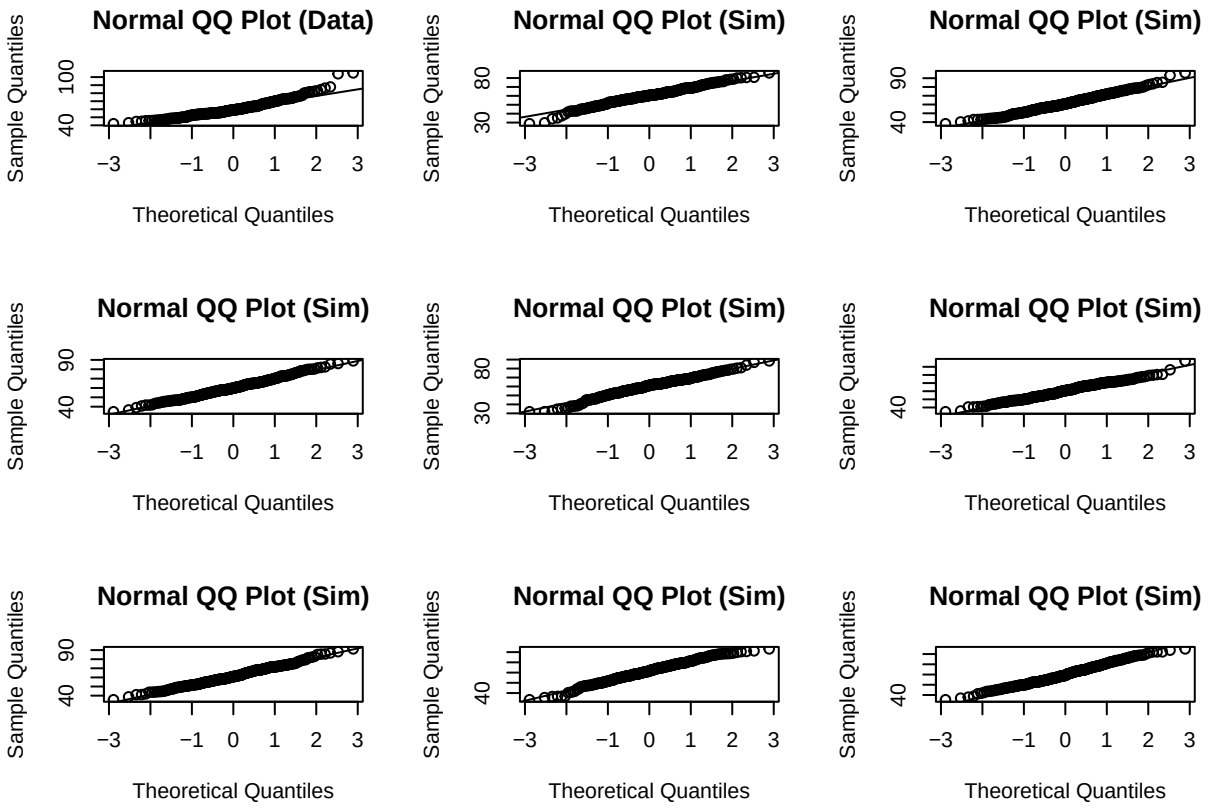
```
hist(fdims$wgt)
```




```
qqnorm(fdims$wgt, main = "Female Weight Normal Probability Plot")
qqline(fdims$wgt)
```




```
qqnormsim(fdims$wgt)
```



Exercise 6: What is the probability of a woman being 180 cm or taller? Around 1.04% using Normal distribution method and around 0.77% using empirical distribution. What is the probability of a woman weighing under 65 kg? Around 67.64% using Normal distribution method and 73.85% using empirical distribution.

The difference between the two methods is more apparent for weight. (Difference in two methods for height was 0.002 and for weight was 0.06) This can be because the height deviates less from the normal distribution, so it makes sense that the two are similar.

```
fwgtmean <- mean(fdims$wgt)
fwgtsd   <- sd(fdims$wgt)
hNormal <- 1 - pnorm(q = 180, mean = fhgtmean, sd = fhgtsd)
hNormal
```

```
## [1] 0.01040328
```

```
hEmpirical <- sum(fdims$hgt > 180) / length(fdims$hgt)
hEmpirical
```

```
## [1] 0.007692308
```

```
hDiff <- hNormal - hEmpirical
hDiff
```

```
## [1] 0.002710974
```

```r
wNormal <- 1 - pnorm(q = 65, mean = fwgtmean, sd = fwgtsd, lower.tail = F)
wNormal
```

```
## [1] 0.6763603
```

```r
wEmpirical <- sum(fdims$wgt < 65) / length(fdims$wgt)
wEmpirical
```

```
## [1] 0.7384615
```

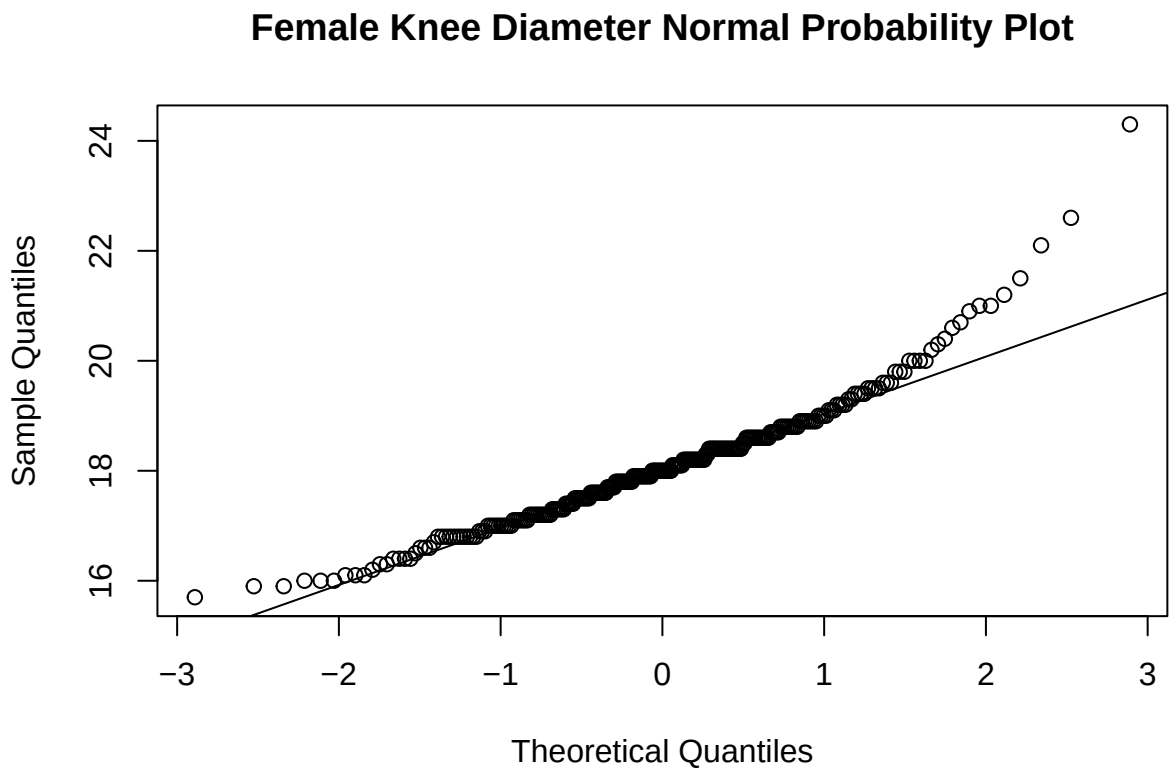
```r
wDiff <- wNormal - wEmpirical
wDiff
```

```
## [1] -0.0621012
```

On Your Own:

1: The distribution of female knee diameters skews left.

```r
qqnorm(fdims$kne.di, main = "Female Knee Diameter Normal Probability Plot")
qqline(fdims$kne.di)
```



```r
hist(fdims$kne.di)
```