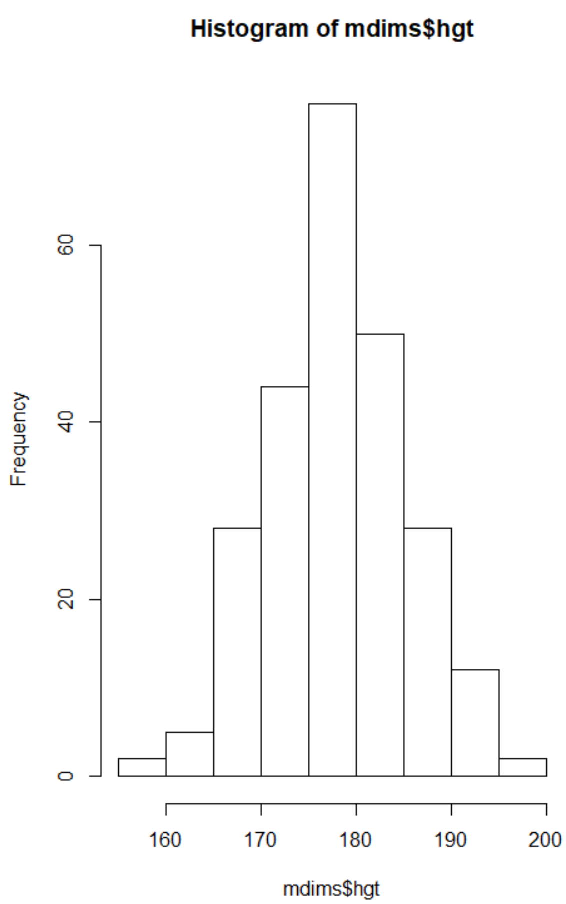
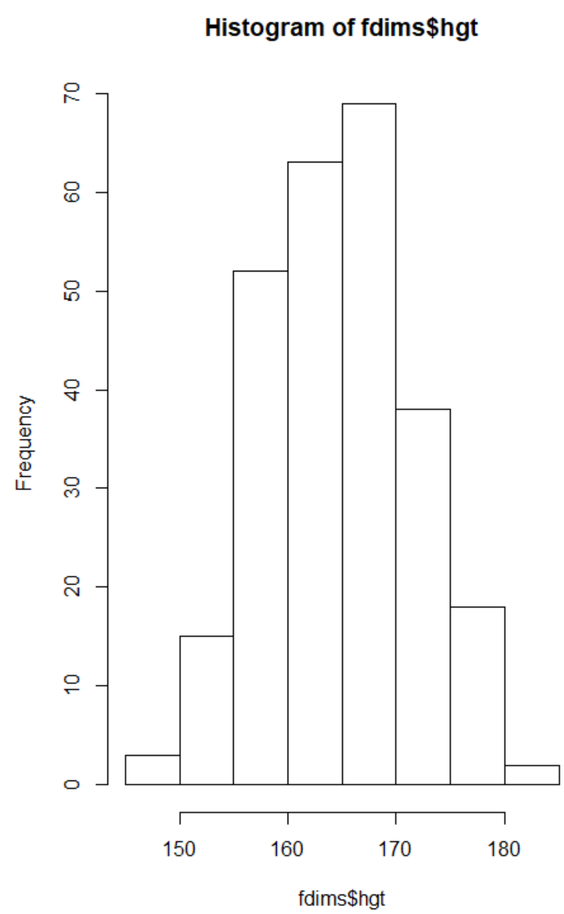
Name: Samuel L. Peoples	Collaborators: (Any one you worked with)

R Lab 3: The Normal Distribution

Exercise 1: Make a histogram of men's heights and a histogram of women's heights. How would you compare the various aspects of the two distributions?

```
> hist(mdims$hgt)
> hist(fdims$hgt)
```

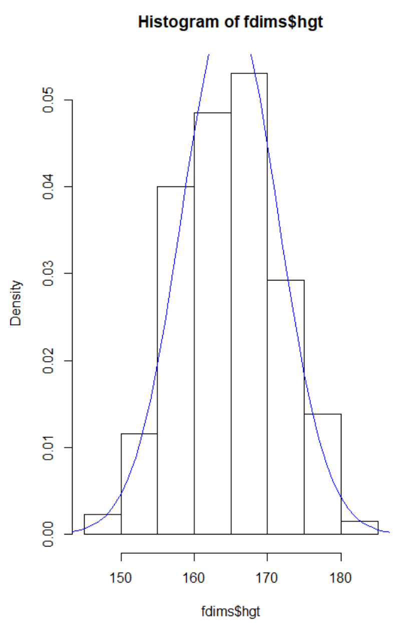




In the above histograms, the mean for the men is roughly 178cm, while the women have a mean around 168cm. The men have a smaller standard deviation than the women, which implies that women may have a wider variety of heights, where men cluster near the mean.

Exercise 2: Based on the this plot, does it appear that the data 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")
```

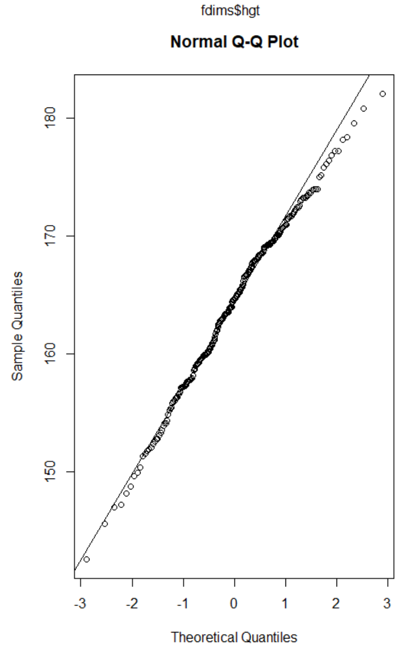


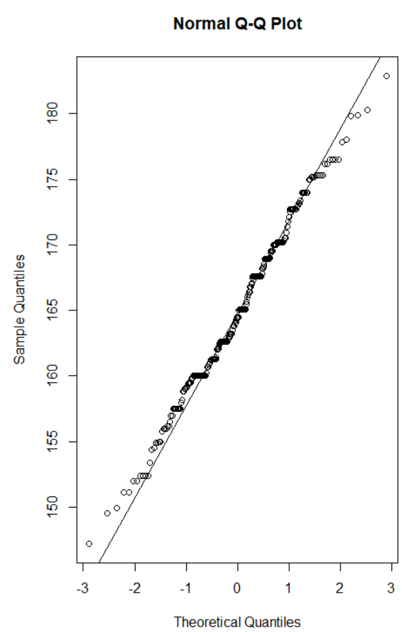
The data roughly falls within the shape of the normal distribution.

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 real data?

```
> sim_norm <- rnorm(n = length(fdims$hgt), mean +=
fhgtmean, sd = fhgtsd)
> qqnorm(sim_norm)
> qqline(sim_norm)
> qqnorm(fdims$hgt)
> qqline(fdims$hgt)
```

`In the normal Q-Q plot for sim_norm, below, a majority of the points are on the line; compared to our data on the right, it generally follows the same trends, and points closely follow the line.`

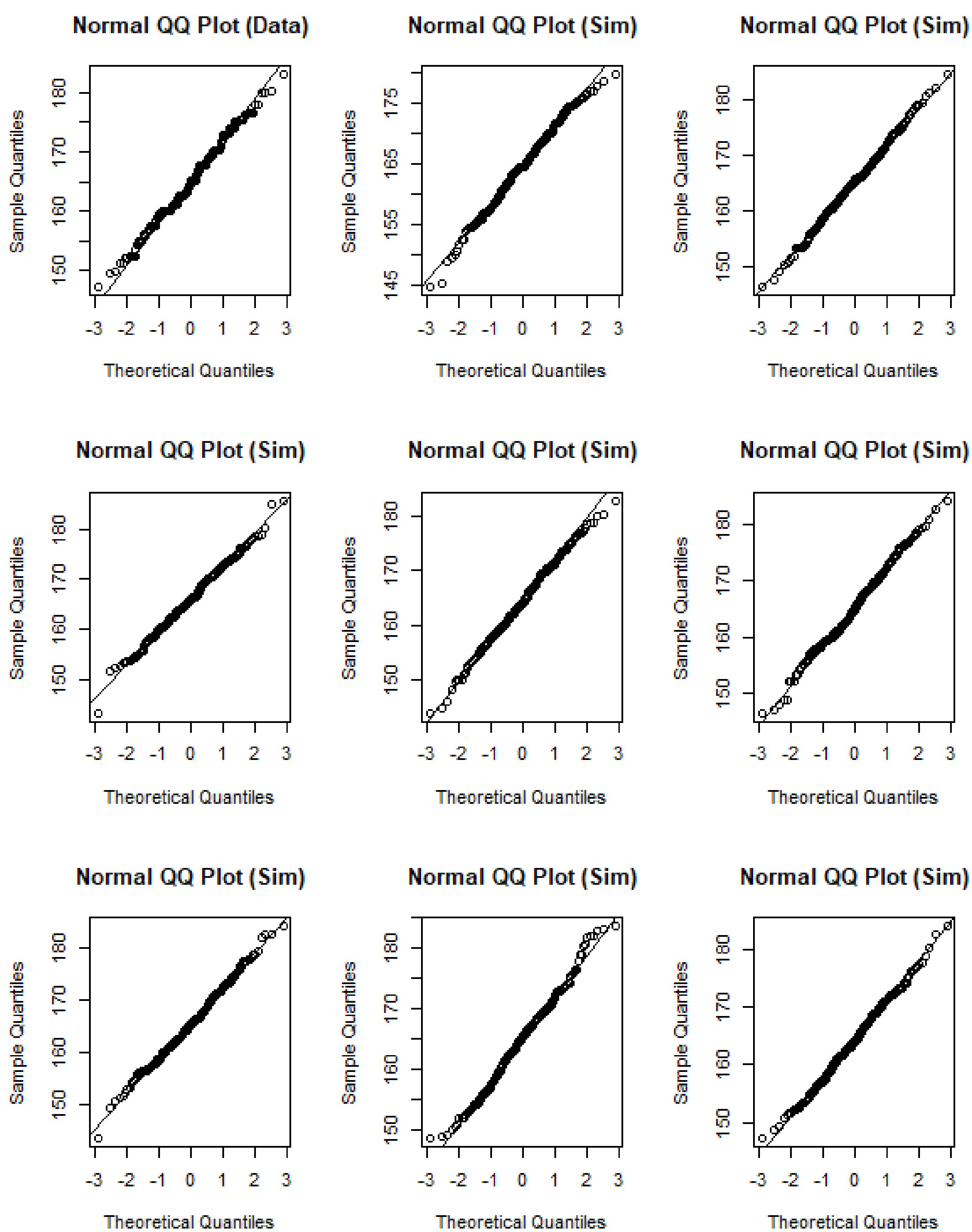




Exercise 4: Does the normal probability plot for fdims$hgt look similar to the plots created for the simulated data? That is, do plots provide evidence that the female heights are nearly normal?

```
>qqnormsim(fdim$hgt)
```

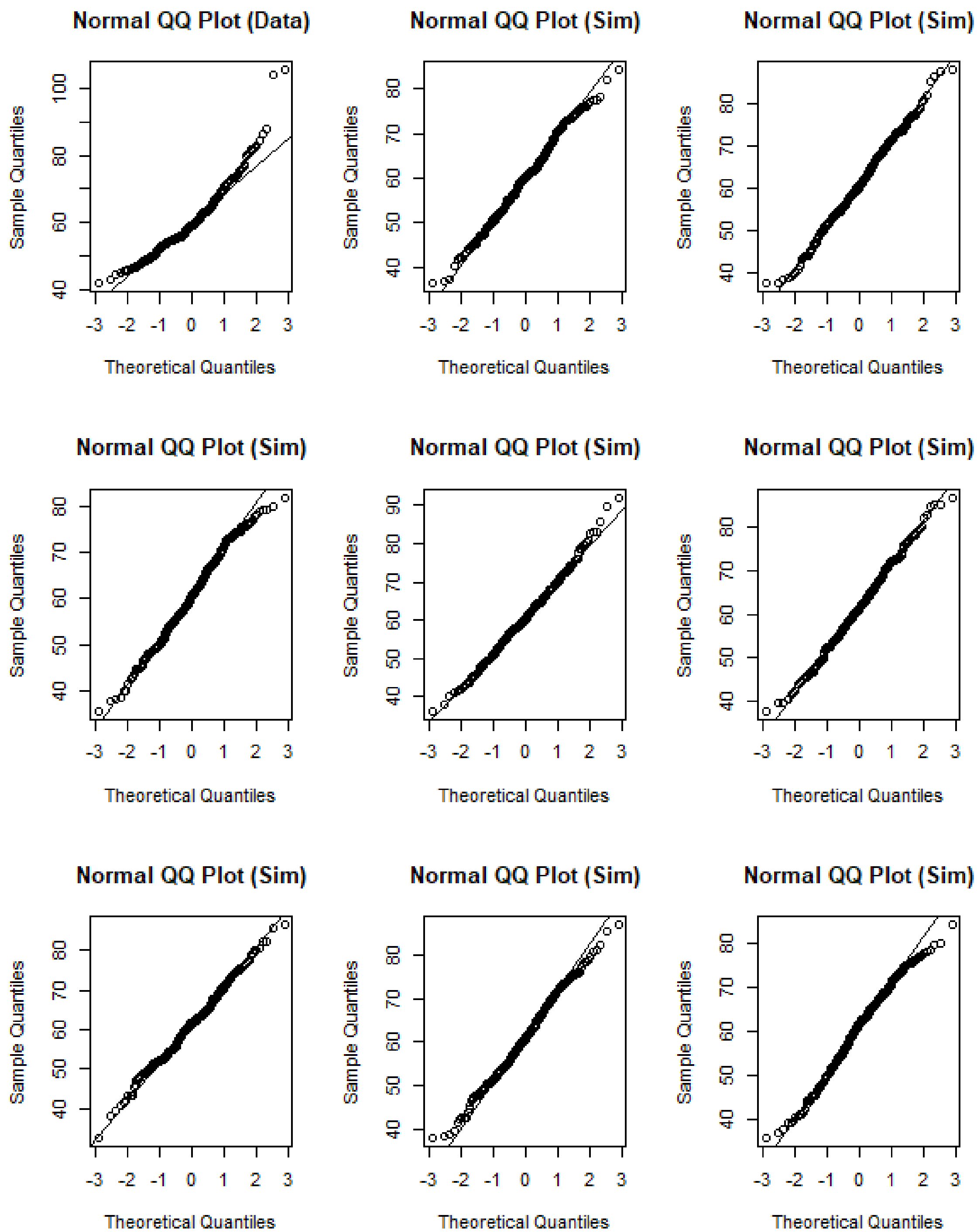
When compared to eight other normal simulations, it would be difficult to pick our data out of a line-up, and our confidence in women's height being distributed normally grows further.



Exercise 5: Using the same technique, determine whether or not female weights appear to come from a normal distribution.

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

In the below figure, the women's weights do not seem to be distributed normally. As the theoretical quantities leave the mean, they seem to follow a nonlinear curve, with many points off of the line.



Exercise 6: Write out two probability questions that you would like to answer; one regarding female heights and one regarding female weights. Calculate the 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 two methods?

What is the probability a randomly chosen young adult female is shorter than 155cm?

```
> pnorm(q = 155, mean = fhgtmean, sd = fhgtsd)
[1] 0.06571769
```

What is the probability a randomly chosen young adult female weighs more than 70kg?

```
> 1-pnorm(q = 70, mean = mean(fdims$wgt), sd = mean(fdims$wgt))
[1] 0.438368
```

On Your Own:

1: Now let's consider some of the other variables in the body dimensions data set. Using the figures at the end of the exercises, match the histogram to its normal probability plot. All of the variables have been standardized (first subtract the mean, then divide by the standard deviation), so the units won't be of any help. If you are uncertain based on these figures, generate the plots in R to check.

a. The histogram for female biiliac diameter (`bii.di`) belongs to normal probability plot letter B.

b. The histogram for female elbow diameter (`elb.di`) belongs to normal probability plot letter C.

c. The histogram for general age (`age`) belongs to normal probability plot letter D.

d. The histogram for female chest depth (`che.de`) belongs to normal probability plot letter A.

```
> qqnorm(fdims$bii.di)

> qqnorm(fdims$elb.di)

> qqnorm(bdims$age)

> qqnorm(fdims$che.de)
```

2: Note that normal probability plots C and D have a slight stepwise pattern.
Why do you think this is the case?

```
>fdims$elb.di

>bdims$age
```

The elbow data is rounded to one decimal place, while age is rounded to the nearest whole number, so discrete values will cause a slightly stepped pattern which diminishes as the sample is increased.

3: As you can see, normal probability plots can be used both to assess normality and visualize skewness. Make 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? Use a histogram to confirm your findings.

```
> qqnorm(fdims$kne.di)

> qqline(fdims$kne.di)
```

In the plot below, there is clustering towards the mean, with a trail to large quantities in the sample, which implies that the histogram is skewed right.