Stat 135 Ch9 Q64 (Lab 2)

Jiying Zou

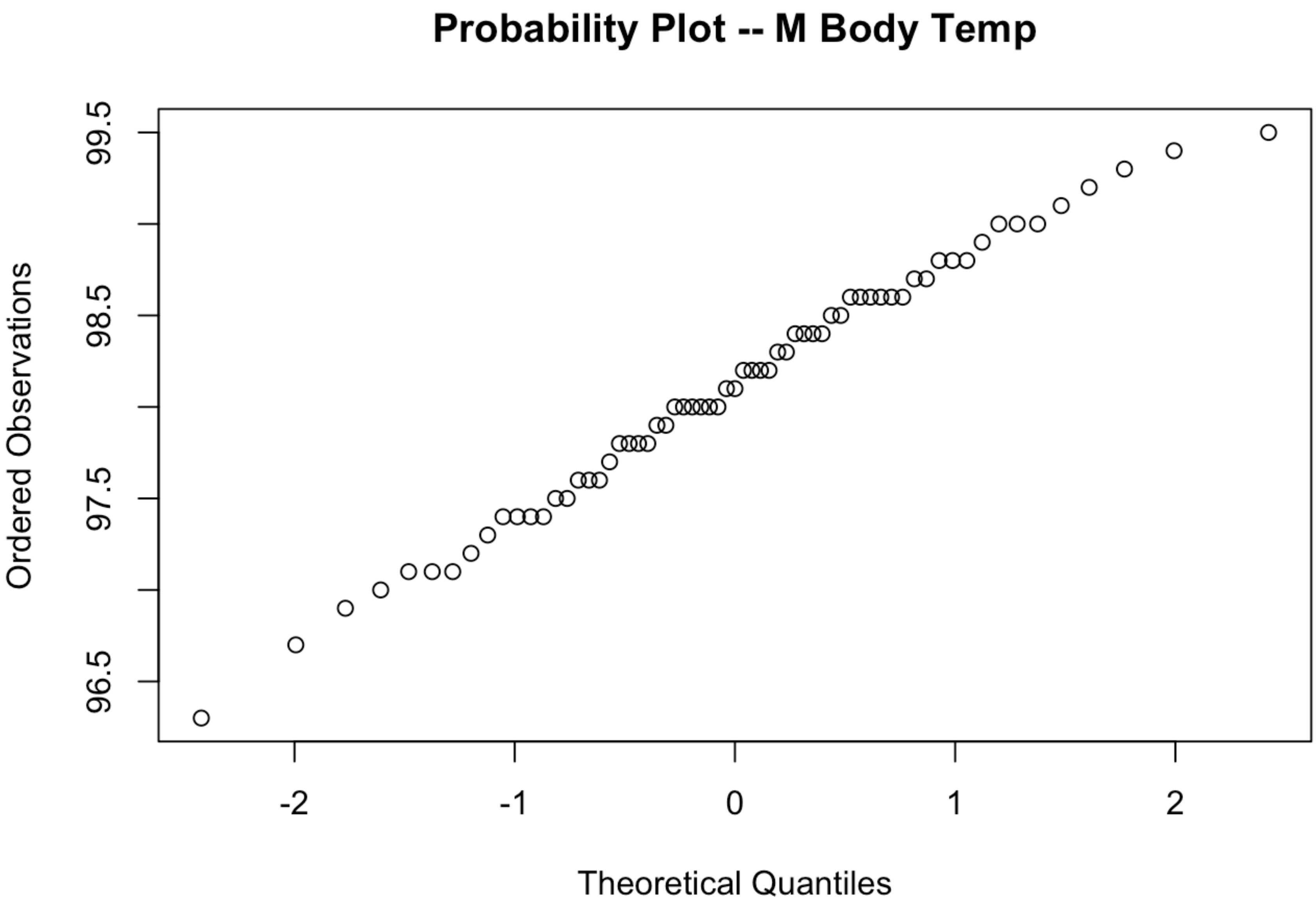
April 3, 2017

“The file `bodytemp` contains normal body temperature readings (degrees Fahrenheit) and heart rates (beats per minute) of 65 males (coded by 1) and 65 females (coded by 2) from Shoemaker (1996).”

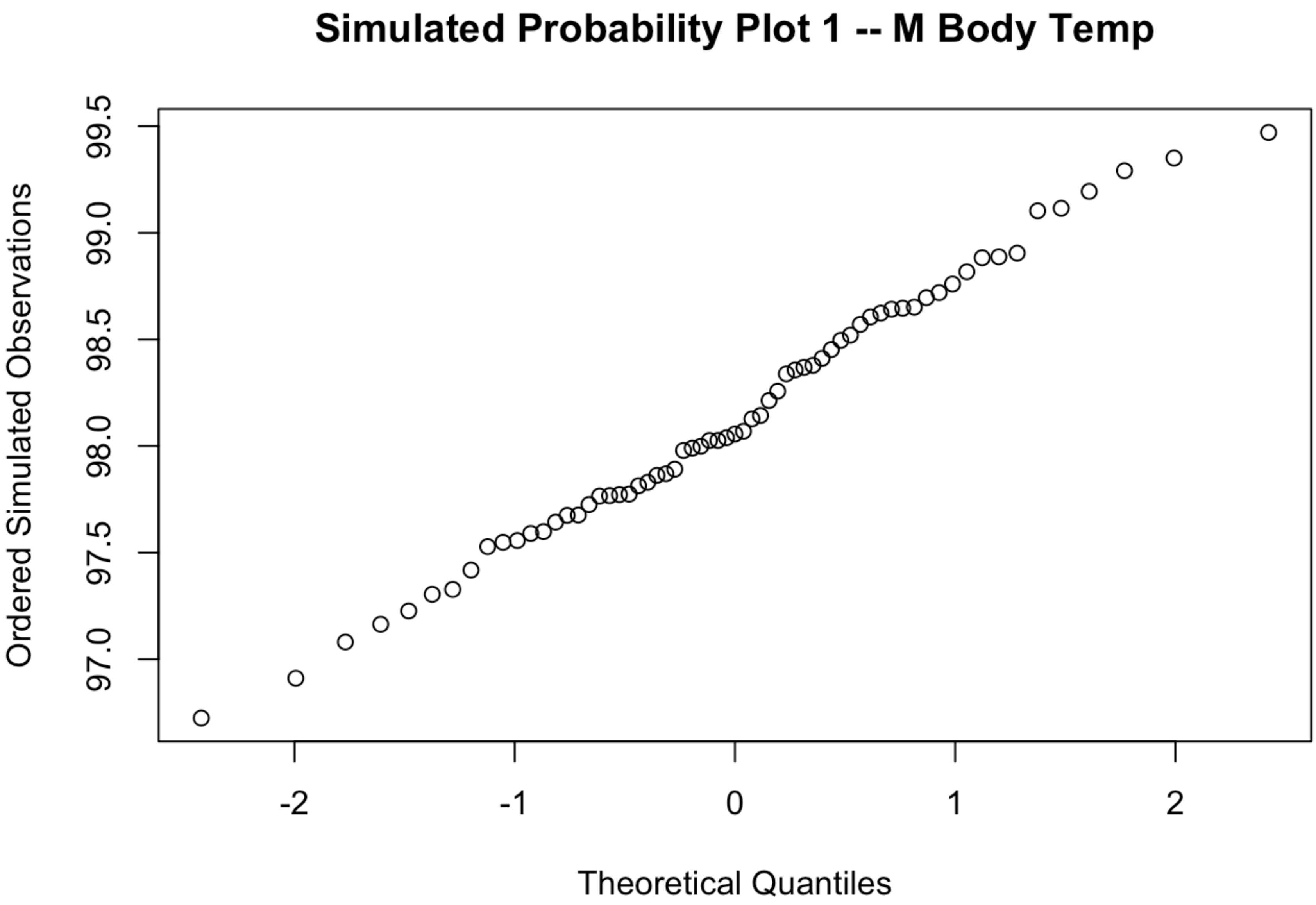
a)

“Assess the normality of the male and female body temperatures by making normal probability plots. In order to judge the inherent variability of these plots, simulate several samples from normal distributions with matching means and standard deviations, and make normal probability plots. What do you conclude?”

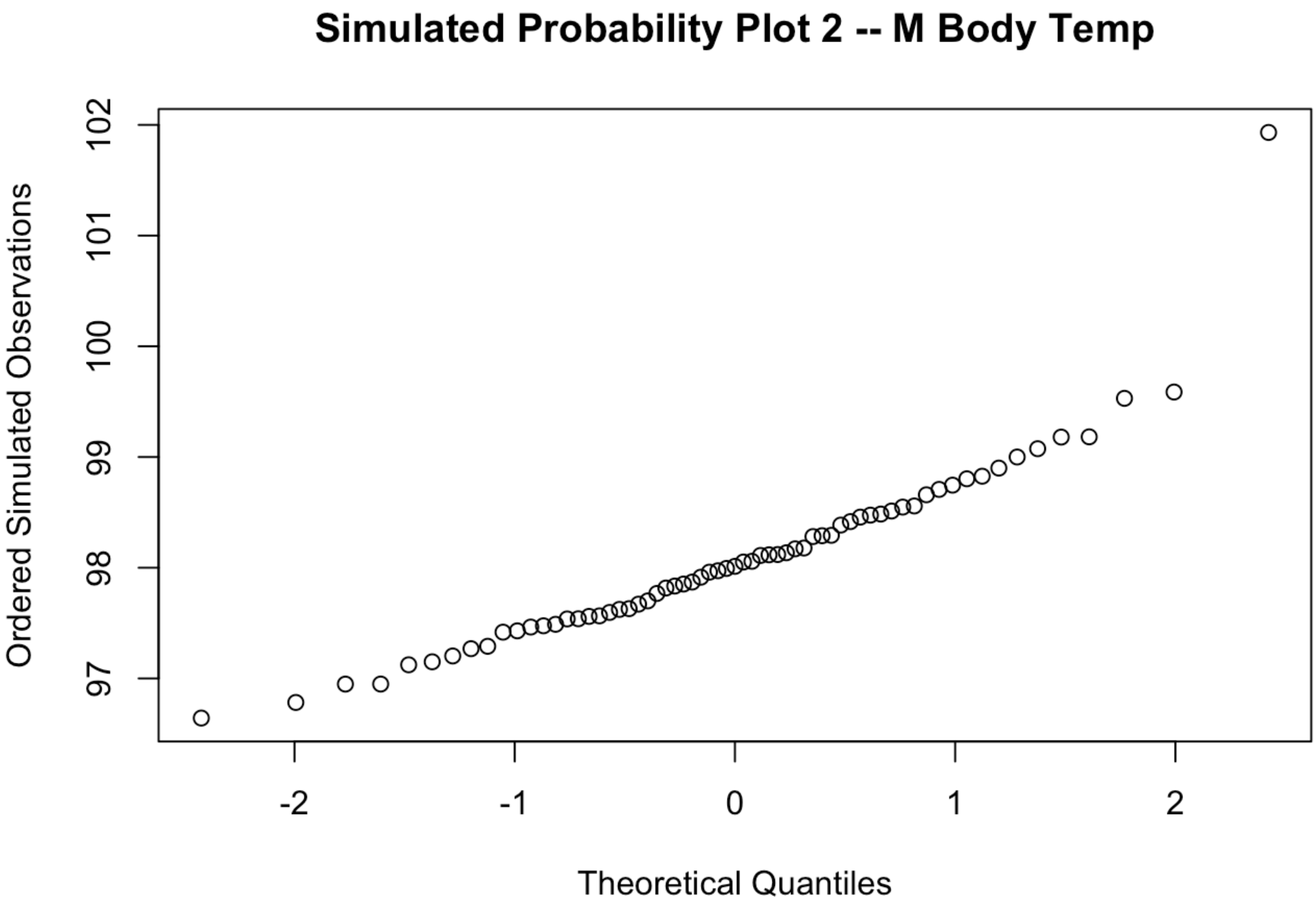
```
#Extract body temperatures of men and women separately
m_btemp <- df$temperature[df$gender == 1]
f_btemp <- df$temperature[df$gender == 2]

#Dataset is inherently ordered, no need to sort! :)

#Probability plot for male body temperatures
th_quantiles_m <- qnorm(ppoints(length(m_btemp))) #generate theoretical Normal quantiles
plot(th_quantiles_m, m_btemp, main = "Probability Plot -- M Body Temp", xlab = "Theoretica
l Quantiles", ylab = "Ordered Observations")
```



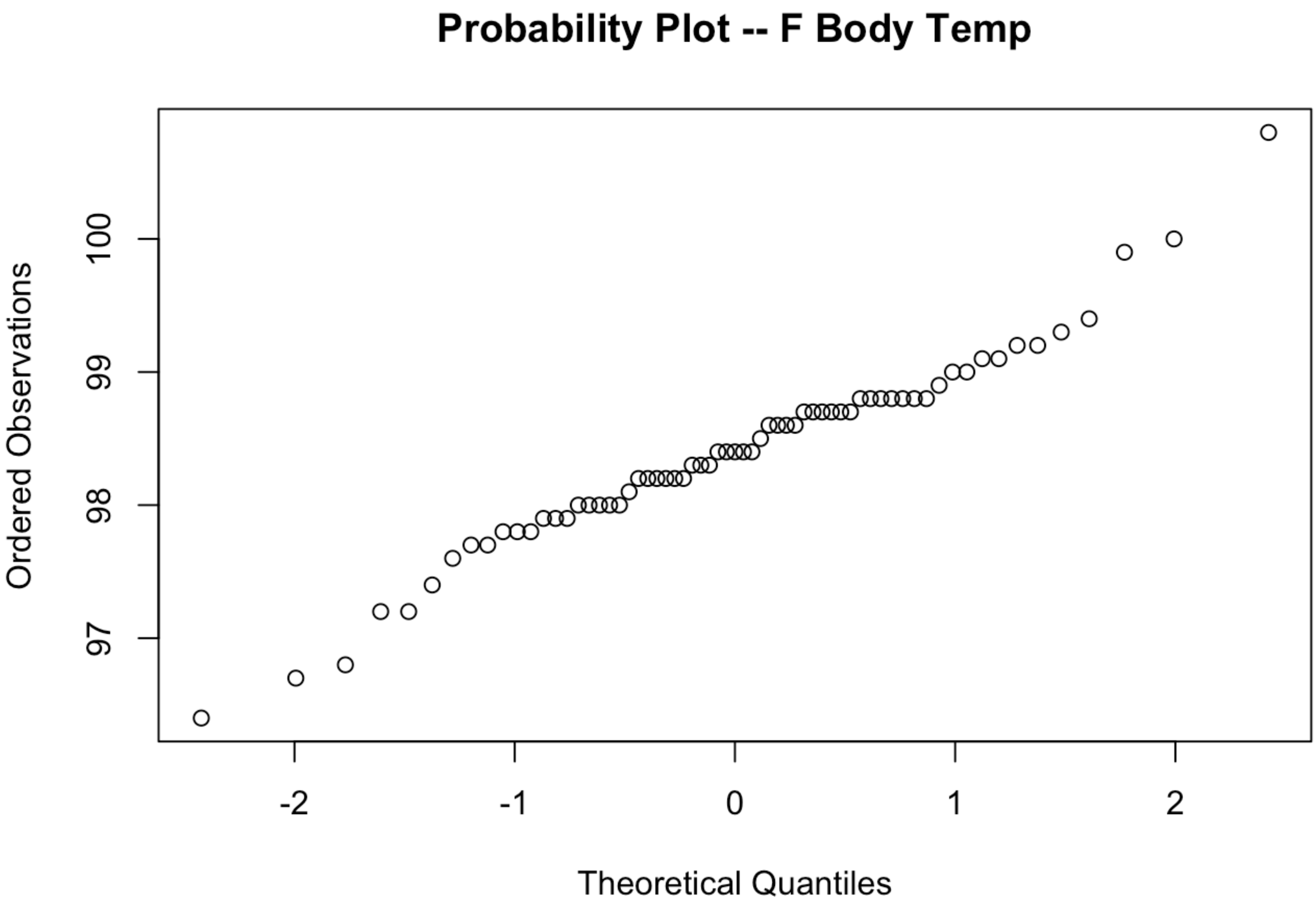

```r
#Two sample prob plots for male
m_btemp_avg <- mean(m_btemp) #find mean
m_btemp_sd <- sd(m_btemp) #find SD
m_btemp_sim1 <- sort(rnorm(length(m_btemp), m_btemp_avg, m_btemp_sd)) #simulate data, sort
m_btemp_sim2 <- sort(rnorm(length(m_btemp), m_btemp_avg, m_btemp_sd))

plot(th_quantiles_m, m_btemp_sim1, main = "Simulated Probability Plot 1 -- M Body Temp", x
lab = "Theoretical Quantiles", ylab = "Ordered Simulated Observations")
```



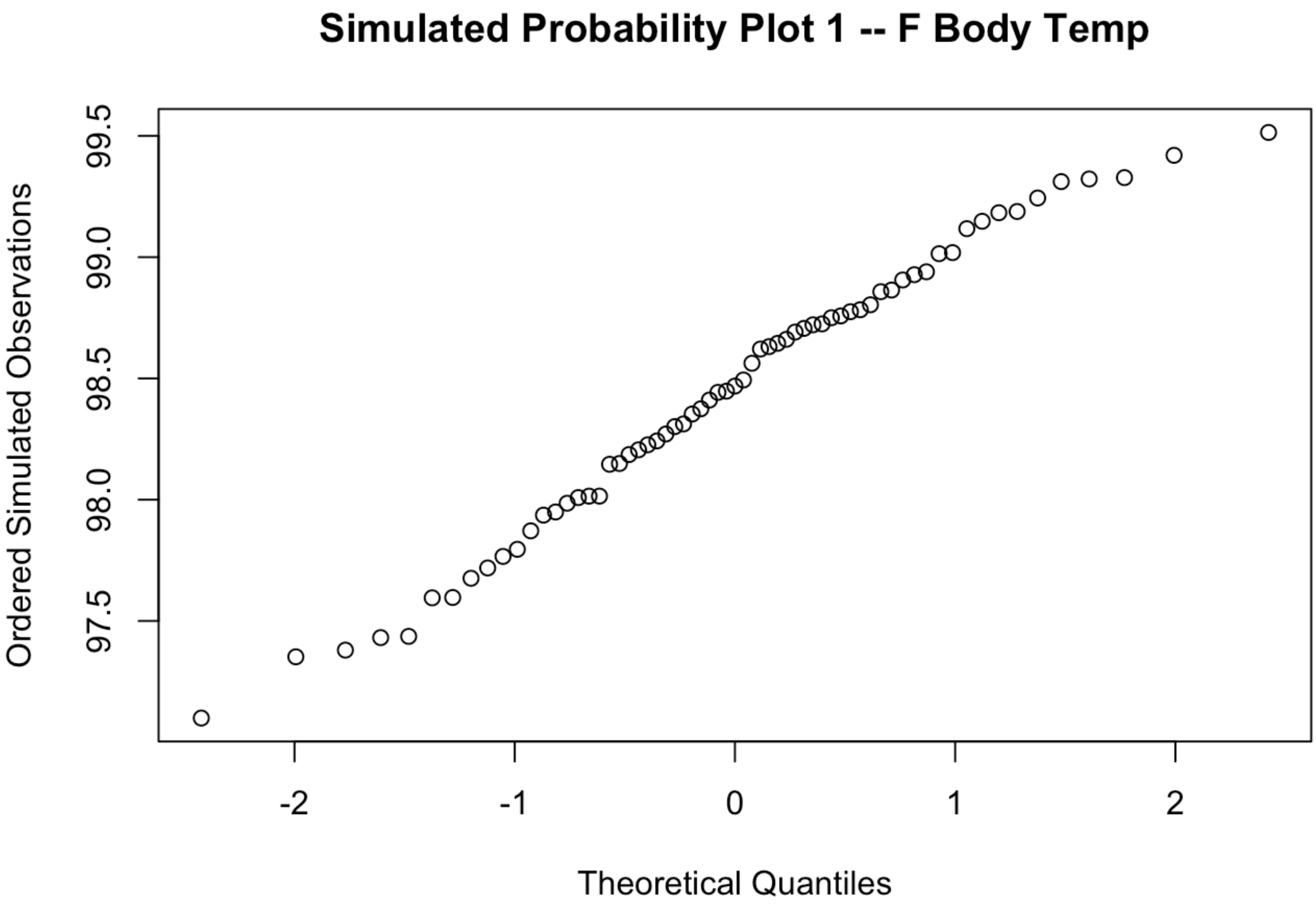

```
plot(th_quantiles_m, m_btemp_sim2, main = "Simulated Probability Plot 2 -- M Body Temp", x
lab = "Theoretical Quantiles", ylab = "Ordered Simulated Observations")
```



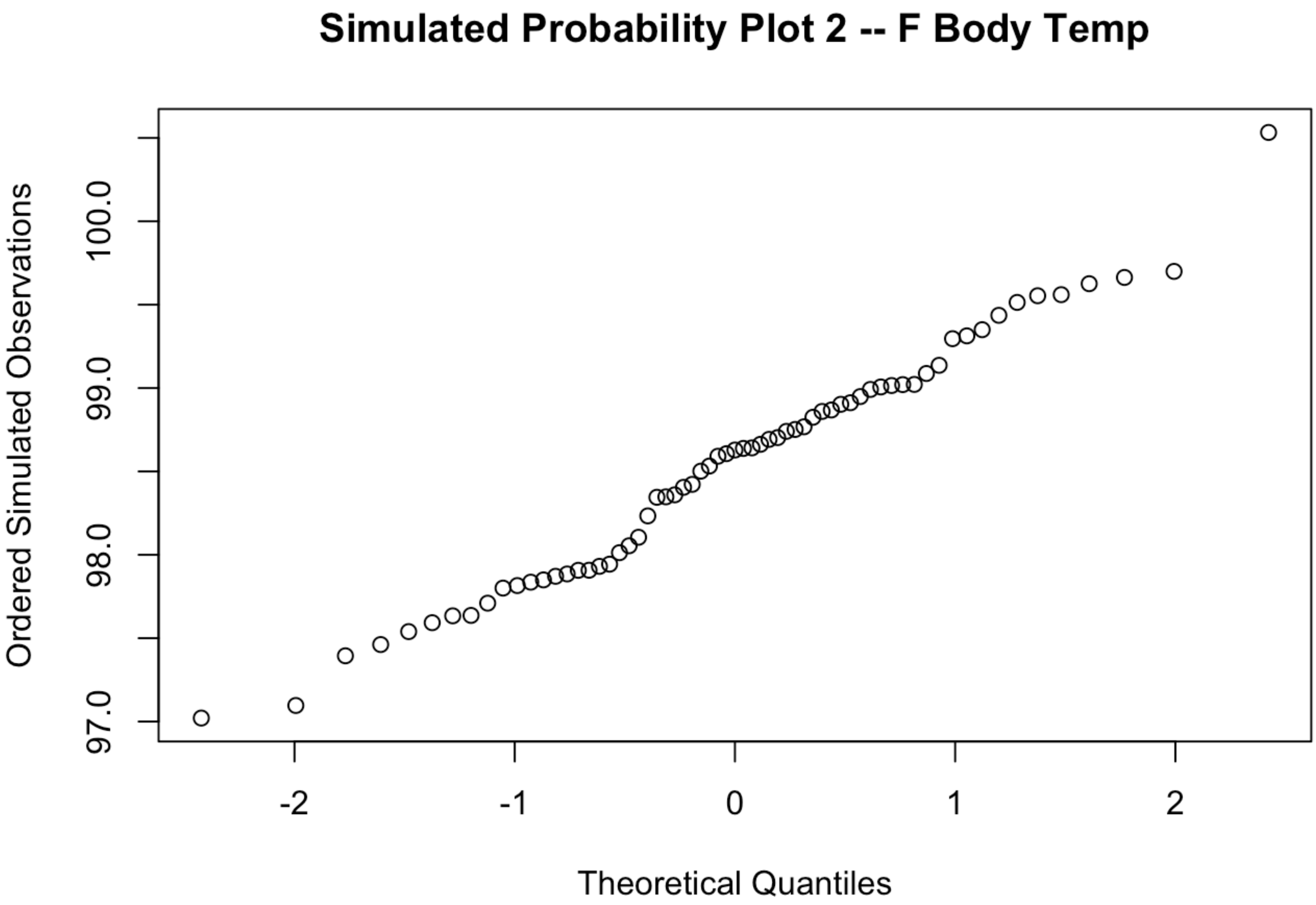

```
#Probability plot for female body temperatures
th_quantiles_f <- qnorm(ppoints(length(f_btemp))) #generate theoretical Normal quantiles
plot(th_quantiles_f, f_btemp, main = "Probability Plot -- F Body Temp", xlab = "Theoretica
l Quantiles", ylab = "Ordered Observations")
```




```
#Two sample prob plots for female
f_btemp_avg <- mean(f_btemp) #find mean
f_btemp_sd <- sd(f_btemp) #find SD
f_btemp_sim1 <- sort(rnorm(length(f_btemp), f_btemp_avg, f_btemp_sd)) #simulate data, sort
f_btemp_sim2 <- sort(rnorm(length(f_btemp), f_btemp_avg, f_btemp_sd))

plot(th_quantiles_f, f_btemp_sim1, main = "Simulated Probability Plot 1 -- F Body Temp", x
lab = "Theoretical Quantiles", ylab = "Ordered Simulated Observations")
```




```
plot(th_quantiles_f, f_btemp_sim2, main = "Simulated Probability Plot 2 -- F Body Temp", x
lab = "Theoretical Quantiles", ylab = "Ordered Simulated Observations")
```



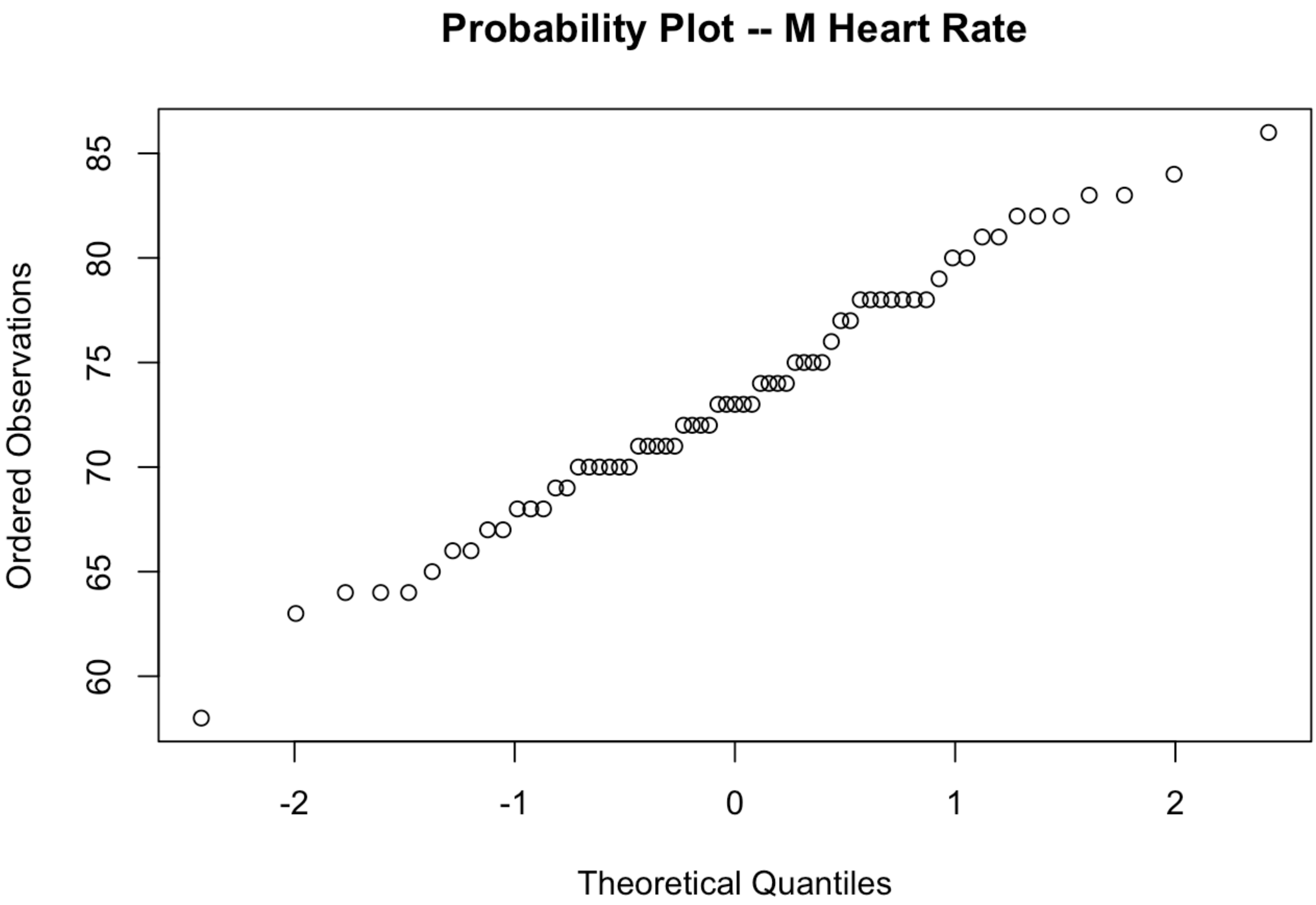
The probability plots for both male and female body temperature show strong linearity with little significant deviation, meaning that the Normal distribution seems to fit the data well for both genders.

Given that I consistently used the theoretical Normal quantiles I generated per gender, the simulated probability plots should demonstrate what natural variation looks like. The actual data is very in line with these simulated plots, with the small exception that the actual data looks a bit more staircase-like and less smoothly random. This could be due to body temperature measurements being taken in a less precise fashion (e.g. read off to the nearest degree).

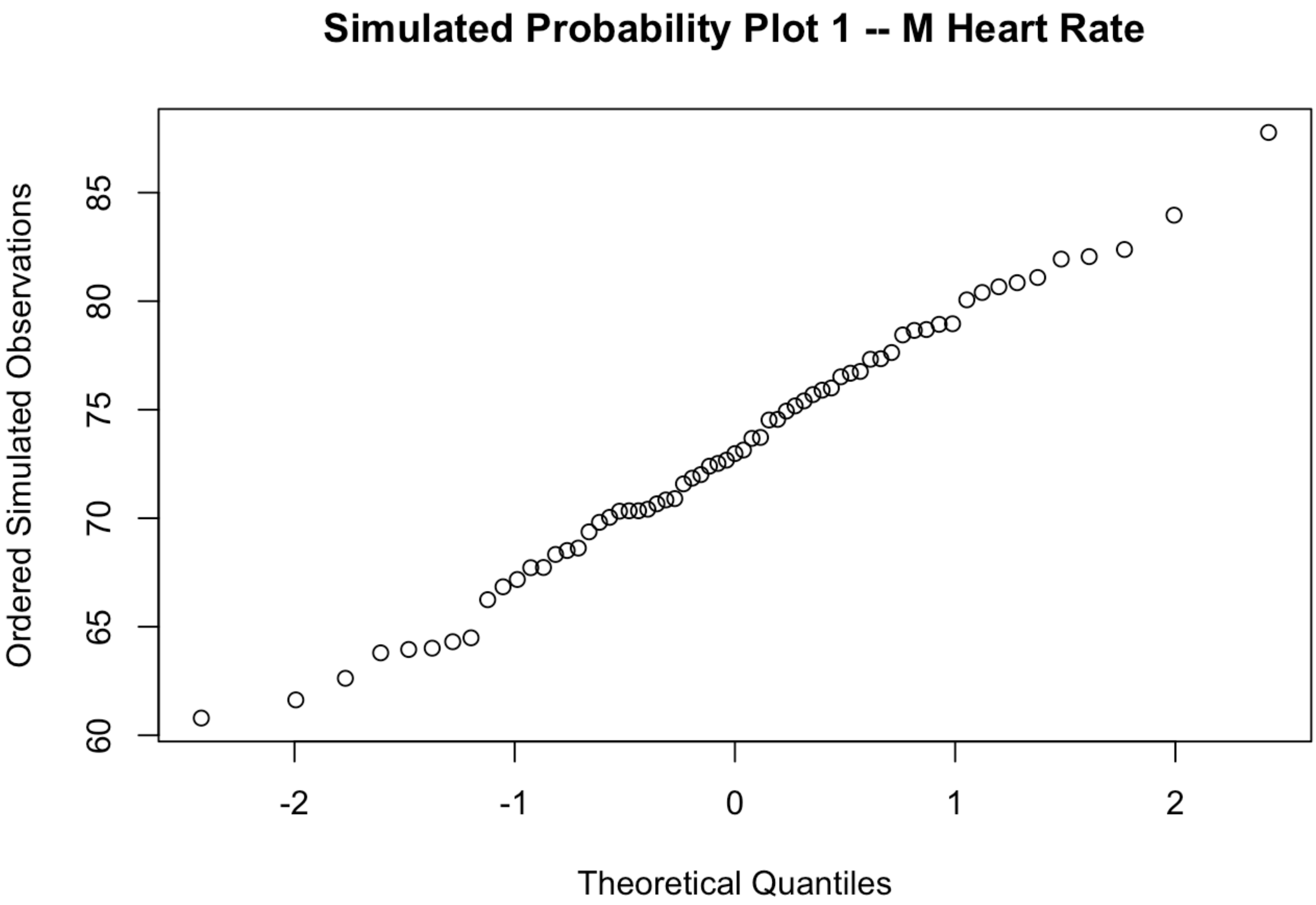
b)

“Repeat the preceding problem for heart rates.”

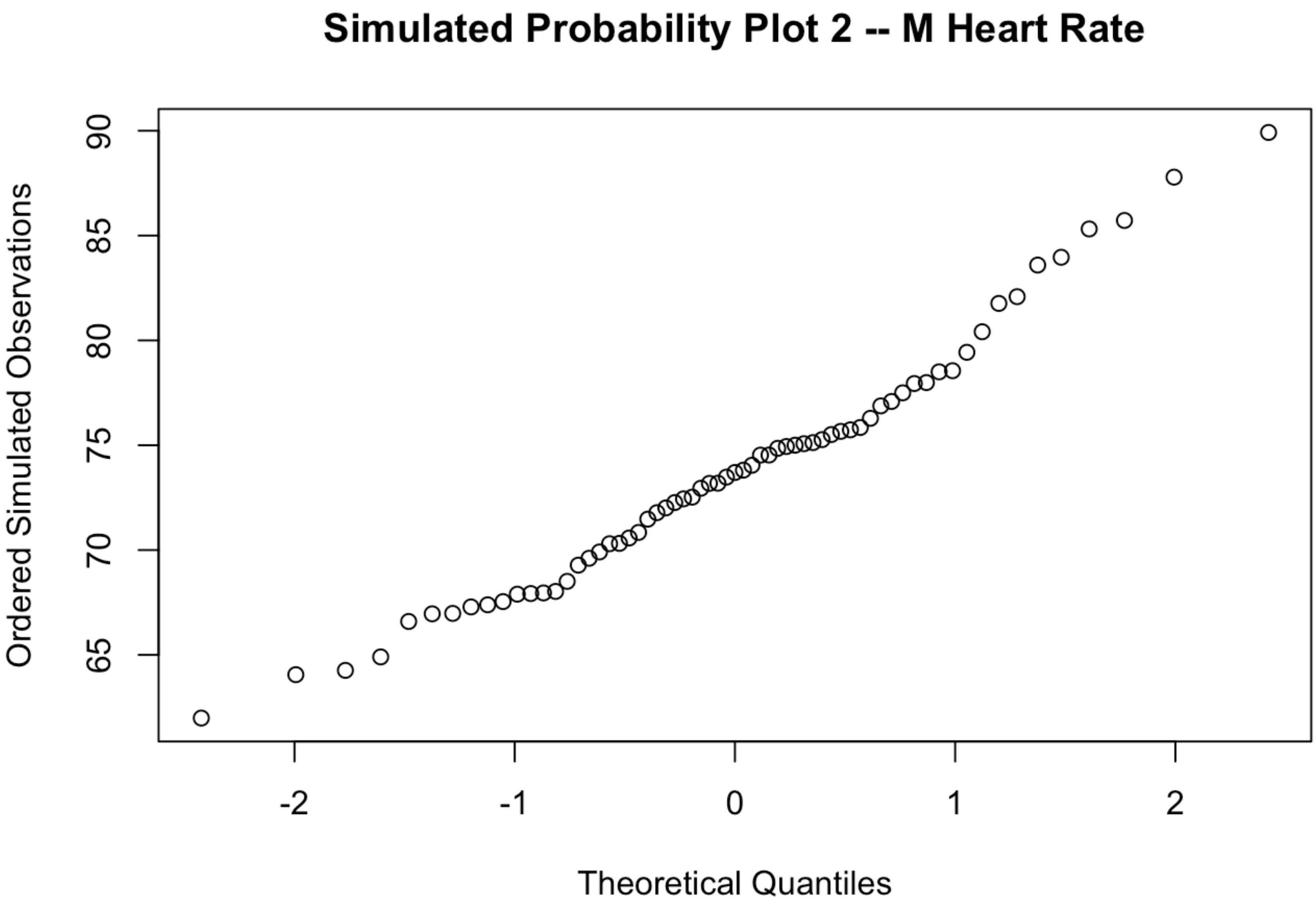
```
#Extract heart rates of men and women separately
m_rate <- sort(df$rate[df$gender == 1])
f_rate <- sort(df$rate[df$gender == 2])


#Probability plot for male body temperatures
th_quant_m <- qnorm(ppoints(length(m_rate))) #generate theoretical Normal quantiles
plot(th_quant_m, m_rate, main = "Probability Plot -- M Heart Rate", xlab = "Theoretical Qu
antiles", ylab = "Ordered Observations")
```



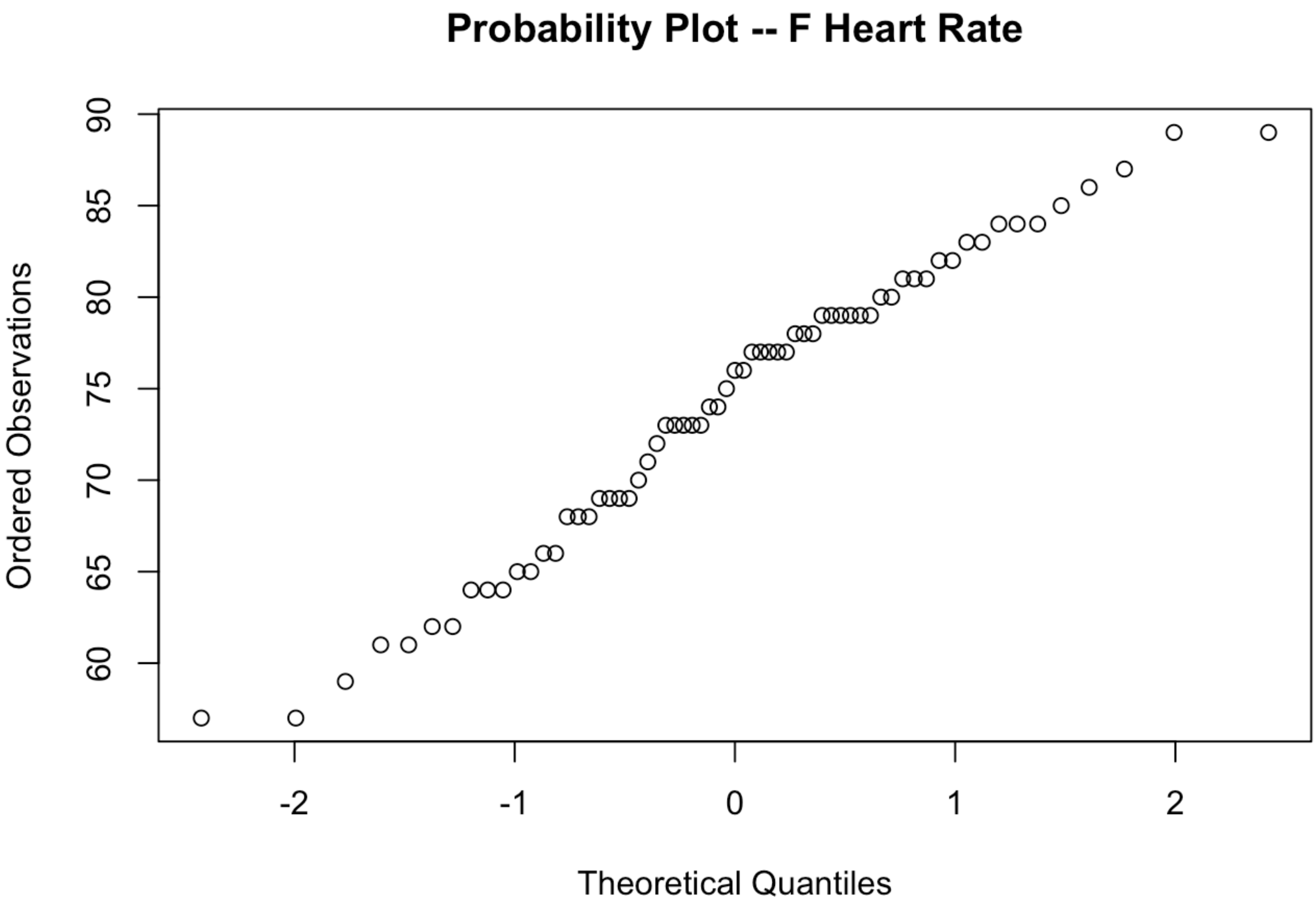

```
#Two sample prob plots for male
m_rate_avg <- mean(m_rate) #find mean
m_rate_sd <- sd(m_rate) #find SD
m_rate_sim1 <- sort(rnorm(length(m_rate), m_rate_avg, m_rate_sd)) #simulate data, sort
m_rate_sim2 <- sort(rnorm(length(m_rate), m_rate_avg, m_rate_sd))

plot(th_quant_m, m_rate_sim1, main = "Simulated Probability Plot 1 -- M Heart Rate", xlab
= "Theoretical Quantiles", ylab = "Ordered Simulated Observations")
```



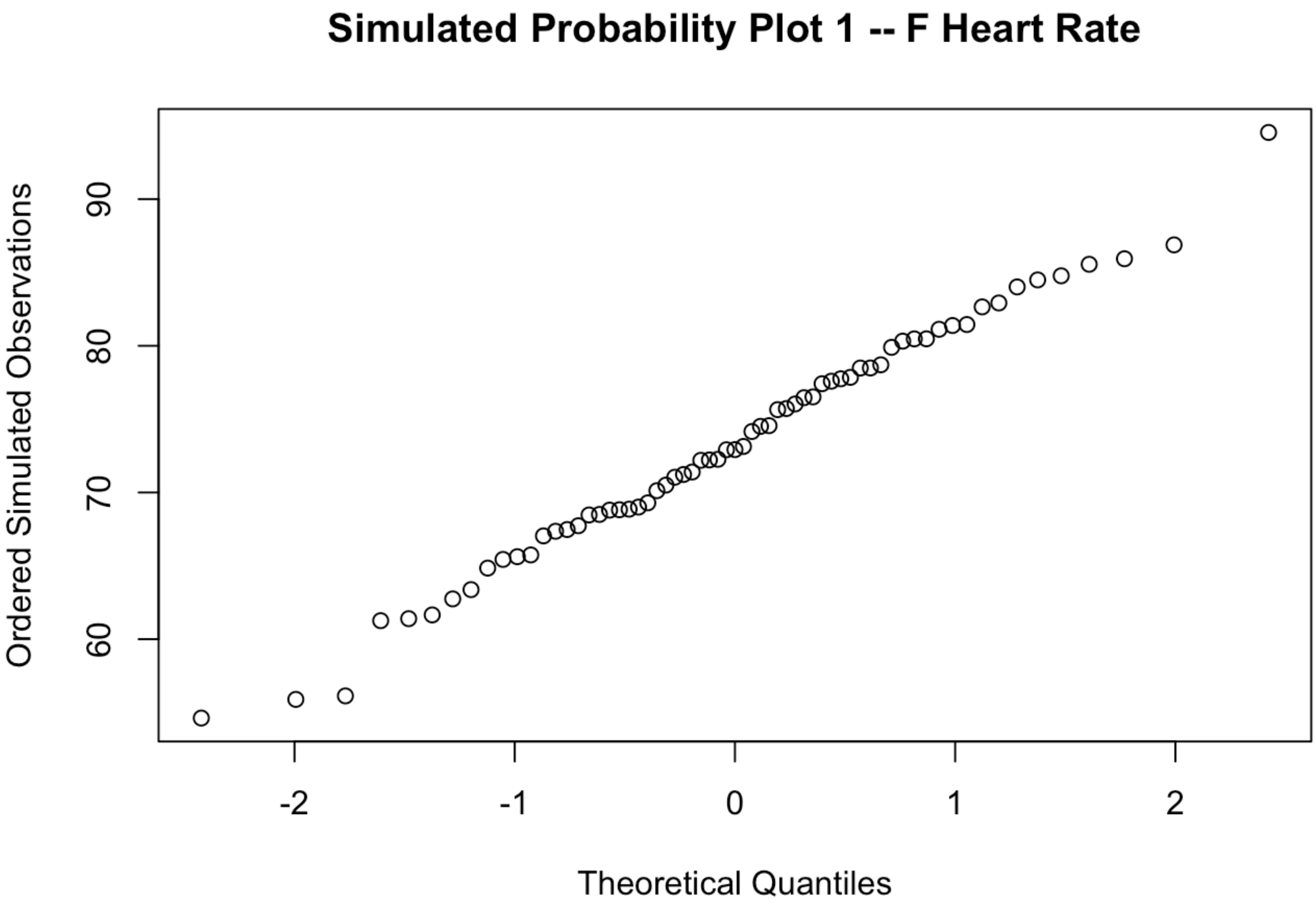

```
plot(th_quant_m, m_rate_sim2, main = "Simulated Probability Plot 2 -- M Heart Rate", xlab
= "Theoretical Quantiles", ylab = "Ordered Simulated Observations")
```



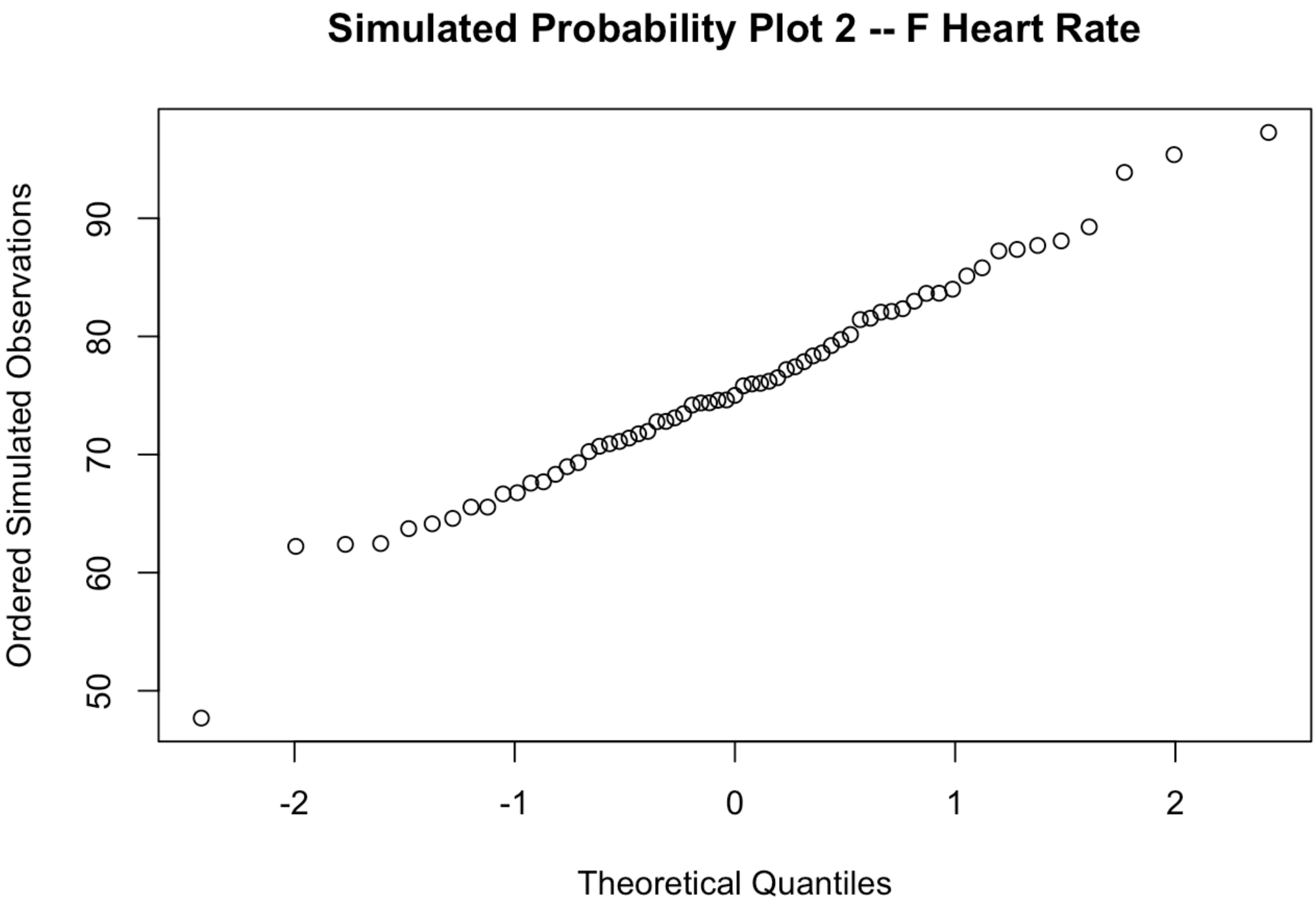

```
#Probability plot for female body temperatures
th_quant_f <- qnorm(ppoints(length(f_rate))) #generate theoretical Normal quantiles
plot(th_quant_f, f_rate, main = "Probability Plot -- F Heart Rate", xlab = "Theoretical Qu
antiles", ylab = "Ordered Observations")
```




```
#Two sample prob plots for female
f_rate_avg <- mean(f_rate) #find mean
f_rate_sd <- sd(f_rate) #find SD
f_rate_sim1 <- sort(rnorm(length(f_rate), f_rate_avg, f_rate_sd)) #simulate data, sort
f_rate_sim2 <- sort(rnorm(length(f_rate), f_rate_avg, f_rate_sd))

plot(th_quant_f, f_rate_sim1, main = "Simulated Probability Plot 1 -- F Heart Rate", xlab
= "Theoretical Quantiles", ylab = "Ordered Simulated Observations")
```




```
plot(th_quant_f, f_rate_sim2, main = "Simulated Probability Plot 2 -- F Heart Rate", xlab
= "Theoretical Quantiles", ylab = "Ordered Simulated Observations")
```



The probability plots for male and female heart rates also show strong linearity with variation easily attributable to chance variation. Once again we see the more step-wise shape of the data, indicating perhaps again the semi-categorical nature of how measurements were taken (e.g. rounded to the nearest beat).

c)

"For the males, test the null hypothesis that the mean body temperature is 98.6 degrees versus the alternative that the mean is not equal to 98.6 degrees. Do the same for the females. What do you conclude?"

We will perform the following two-sided t-tests (since population variance unknown) at the $\alpha = 0.05$ significance level.

For males:

$$H_o: \textit{The mean male body temperature is } 98.6^o \qquad H_A: \textit{The mean male body temperature is not } 98.6^o$$

```
print(paste("Average male body temperature is", m_btemp_avg, ", with an SD of", m_btemp_sd
,". There are", length(m_btemp), "observations."))
```

```
## [1] "Average male body temperature is 98.1046153846154 , with an SD of 0.69875576232659
1 . There are 65 observations."
```

$$test\ statistic: \frac{98.1046 - 98.6}{0.6987} = -0.7090 \qquad follows\ t_{65-1} = t_{64}\ distribution$$

$$p - value: \quad P(t_{64} < -0.7090) \approx 0.2404 = 24.04\% \quad (using\ calculator)$$

The p-value is insignificant at the 5% significance level. It is too easy to get results as extreme as ours due to chance variation. We fail to reject the null; there is not enough evidence against the claim that mean male body temperature is 98.6 degrees.

For females:

$$H_o: \ The\ mean\ female\ body\ temperature\ is\ 98.6^o \qquad H_A: \ The\ mean\ female\ body\ temperature\ is\ not\ 98.6^o$$

```
print(paste("Average female body temperature is", f_btemp_avg, ", with an SD of", f_btemp_
sd,". There are", length(f_btemp), "observations."))
```

```
## [1] "Average female body temperature is 98.3938461538462 , with an SD of 0.743487752731
366 . There are 65 observations."
```

$$test\ statistic: \frac{98.3938 - 98.6}{0.7434} = -0.2773 \qquad follows\ t_{65-1} = t_{64}\ distribution$$

$$p - value: \quad P(t_{64} < -0.2773) \approx 0.3912 = 39.12\% \quad (using\ calculator)$$

This p-value is also insignificant at the 5% significance level. It is too easy to get results as extreme as ours due to chance variation. We fail to reject the null here too; there is not enough evidence against the claim that mean female body temperature is 98.6 degrees either.