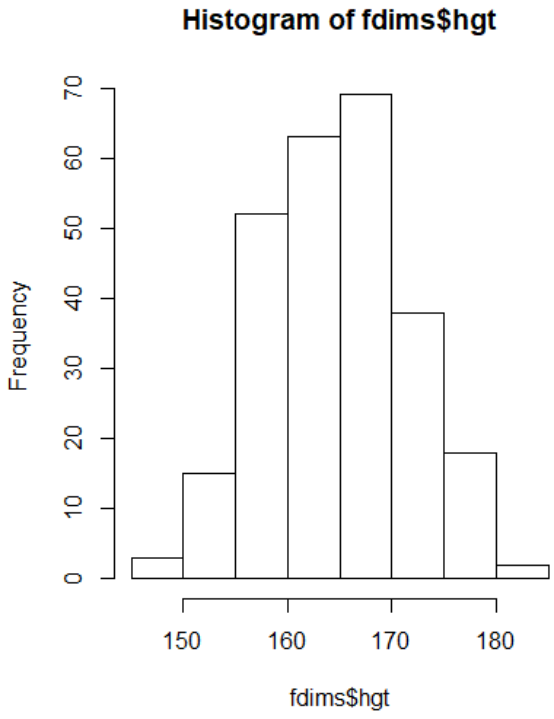
Name: Connor Lauerman Collaborators: (Any one you worked with)

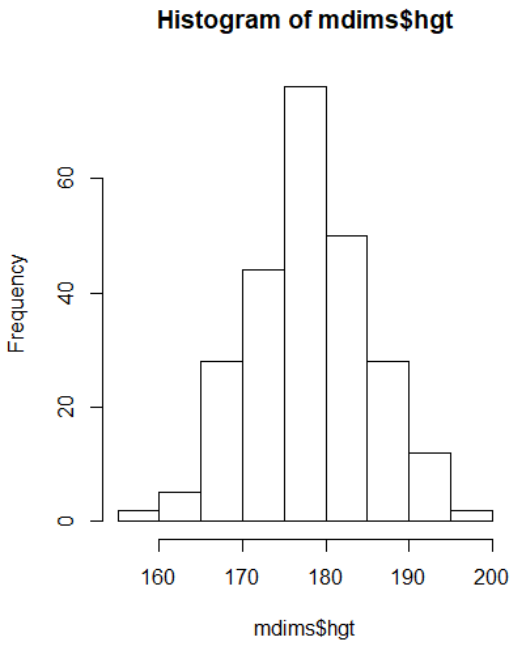
R Lab 4: Distributions

Please answer all the Exercises and the questions from the "On Your Own" section. If you use any graphs or charts to justify your answer, please include them.

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? (Include plot.)

The male heigts seem to be more symmetric than the female ones with more in the middle frequency. While the female heights are less symmetric and mean shifted right compared to the male heights



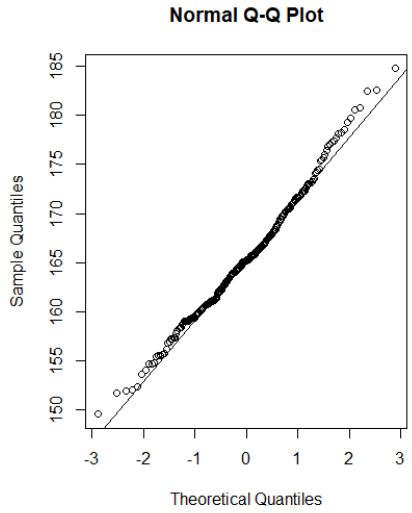


Exercise 2: Based on this plot, does it appear that the data follow a nearly normal distribution?

The data is nearly normal but isn't a perfect bell shape specifically the mean is not at the top of the bell

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? (Include plot.)

All the points don't fall on the line but they are very close. This plot is much better than the original at following the lines but in both graphs the tails are troublesome

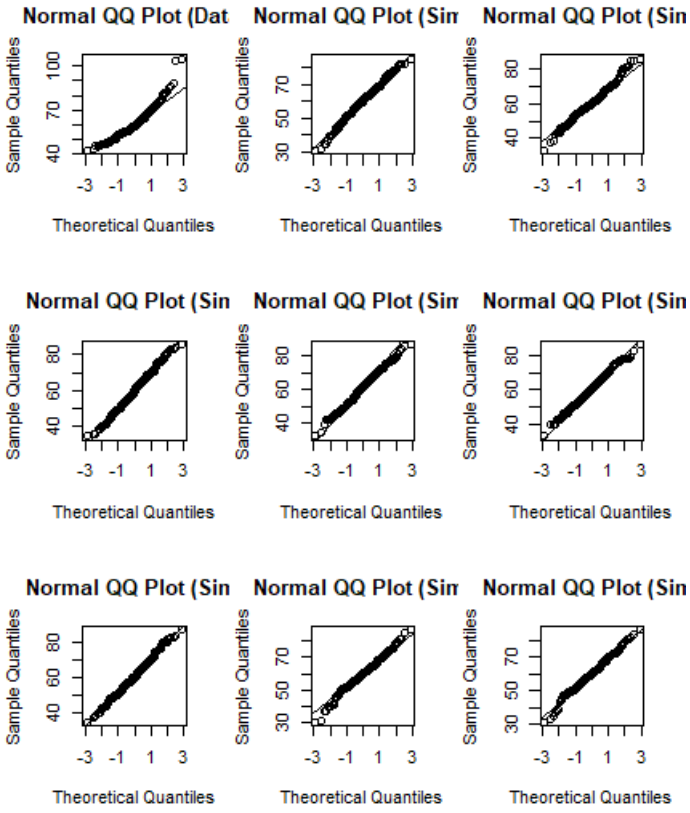


Exercise 4: Does the normal probability plot for `fdims$hgt` look similar to the plots created for the simulated data? That is, do plots provide evidence that the female heights are nearly normal?

They do provide evidence since our data has characteristics of the other plots behavior such as tails missing the line and oscillation over the normal line

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

Female weights do not appear to be normal because the data bends off the normal line shortly after the middle which I think is too much data to be called normal



Exercise 6: Write out two probability questions that you would like to answer; one regarding female heights and one regarding female weights. Calculate the answers to 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 that a random selected adult female is smaller than 152cm (5 feet)?
Pnorm = .0246
Prob = .0192
What is the probability that a random selected adult female weight is between 70 and 80?
Pnorm = .142
Prob = .112

On Your Own:

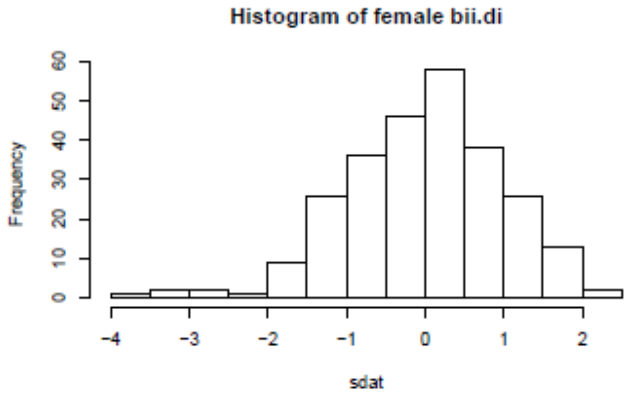
1) Now let's consider some of the other variables in the body dimensions data set. Using the figures at the end of the exercises, match the histogram to its normal probability plot. All of the variables have been standardized (first subtract the mean, then divide by the standard deviation), so the units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

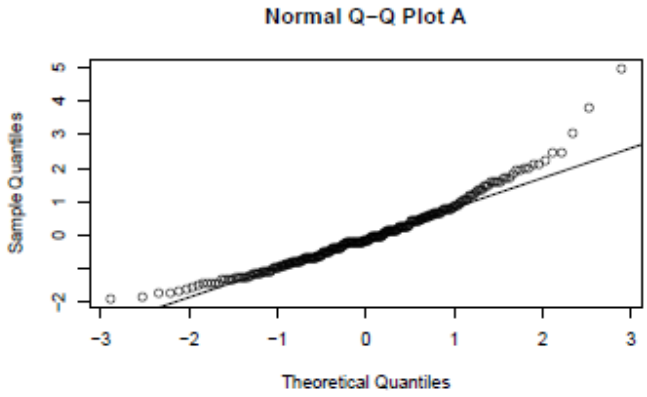
a. The histogram for female biiliac (pelvic) diameter (`bii.di`) belongs to normal probability plot letter _B_.

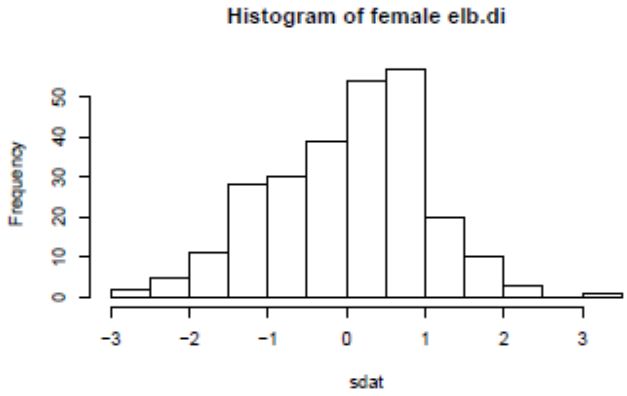
b. The histogram for female elbow diameter (`elb.di`) belongs to normal probability plot letter _C___.

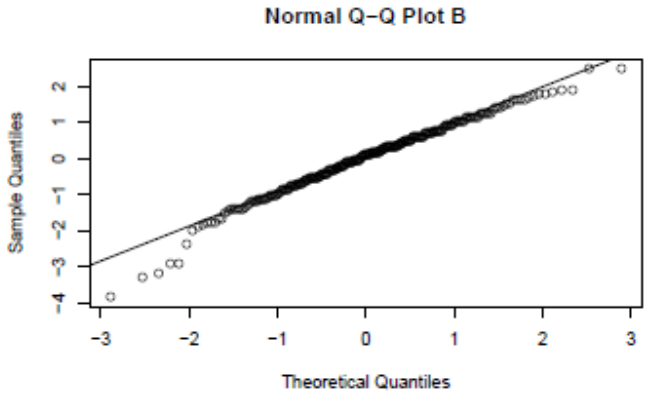
c. The histogram for general age (`age`) belongs to normal probability plot letter __D__.

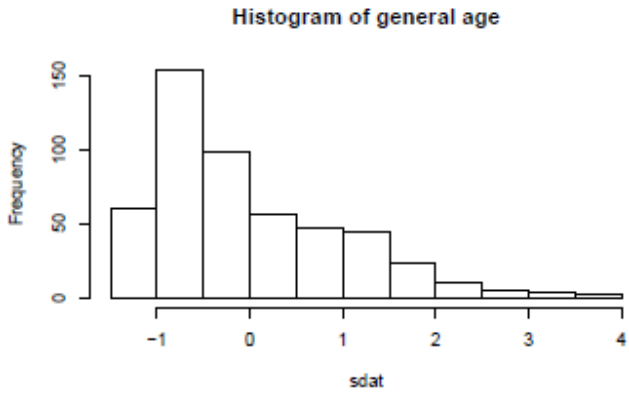
d. The histogram for female chest depth (`che.de`) belongs to normal probability plot letter __A__.

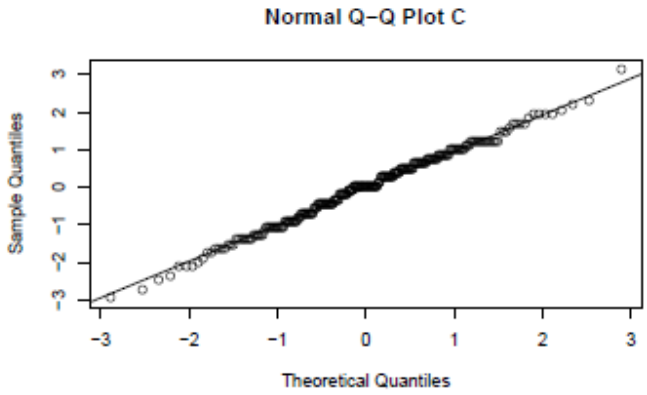


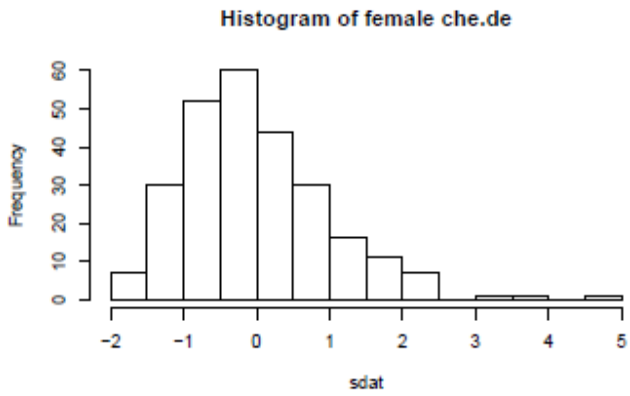


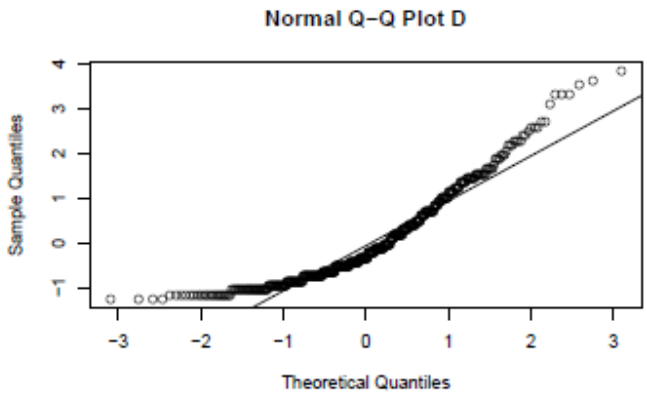












2) Note that normal probability plots C and D have a slight stepwise pattern.
Why do you think this is the case?

Since the graphs have similar frequencies next on some bars and then jumps up to normal on the following bar it creates a step like pattern

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. (Include plots.)

The data seems left skewed since more there is a denser amount of data in the negative x axis than there is positive