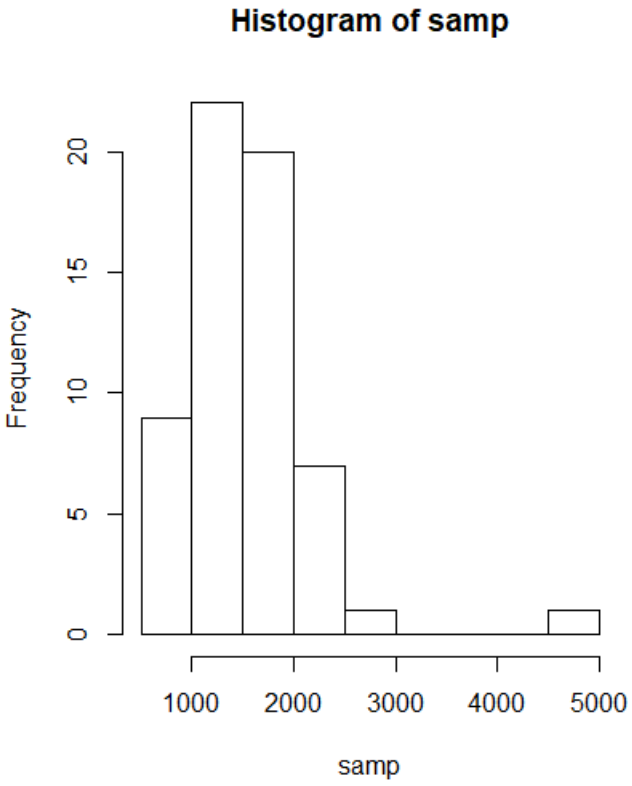
Name: Connor Lauerman Collaborators: (Any one you worked with)

R Lab 6: Confidence Intervals

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: Describe the distribution of your sample. What would you say is the "typical" size within your sample? Also state precisely what you interpreted "typical" to mean. (Include plot.)

The distribution is relatively symmetric with what appears to be an outlier at 5000. A typical size would be 1000-2000 since the majority of the data lies in between these values. I thought typical was where most of the values of the houses lied specifically the higher bars in the histograms.



Exercise 2: Would you expect another student's distribution to be identical to yours? Would you expect it to be similar? Why or why not?

I would no expect another students distribution to be identical since we are randomly grabbing 60 from the population there will be variability in our data. Although distribution would be similar since we are pulling from the same population that means the sample will have similar mean and standard deviation to the population but different which means our samples will also be similar.

Exercise 3: For the confidence interval to be valid, the sample mean must be normally distributed and have standard error $\frac{s}{\sqrt{n}}$. What conditions must be met for this to be true?
The sample size must be greater than 30

Exercise 4: What does “95% confidence” mean?
95% confidence means that we are 95 percent confident the mean of the population lies with in this range of data

Exercise 5: Does your confidence interval capture the true average size of houses in Ames?
Yes the mean is 1499.69 and my range is 1394.783 1699.150

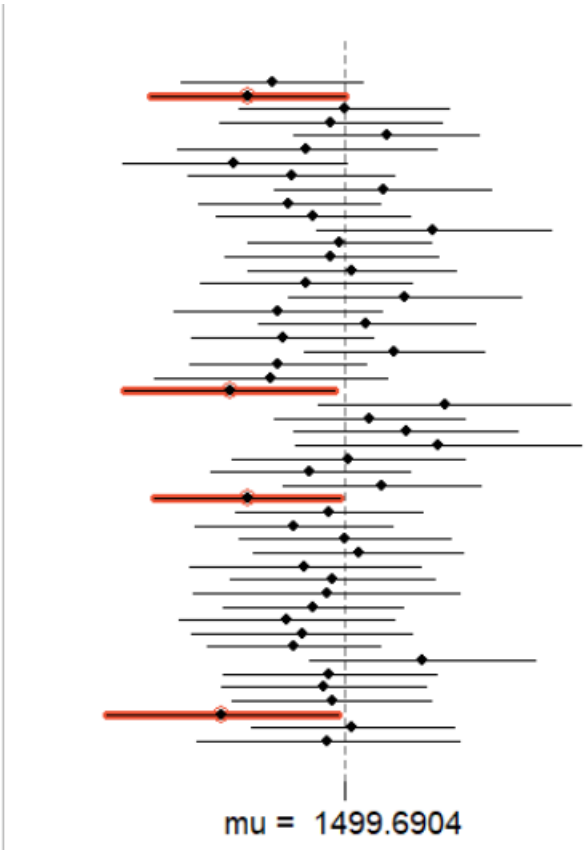
Exercise 6: Each student in your class should have gotten a slightly different confidence interval. What proportion of those intervals would you expect to capture the true population mean? Why?
I would expect 95% of the intervals capture the true population mean since we are using a confidence interval of 95% confident

On Your Own:

1) Using the following function (which was downloaded with the data set), plot all intervals. What proportion of your confidence intervals include the true population mean? Is this proportion exactly equal to the confidence level? If not, explain why. (Include plot.)

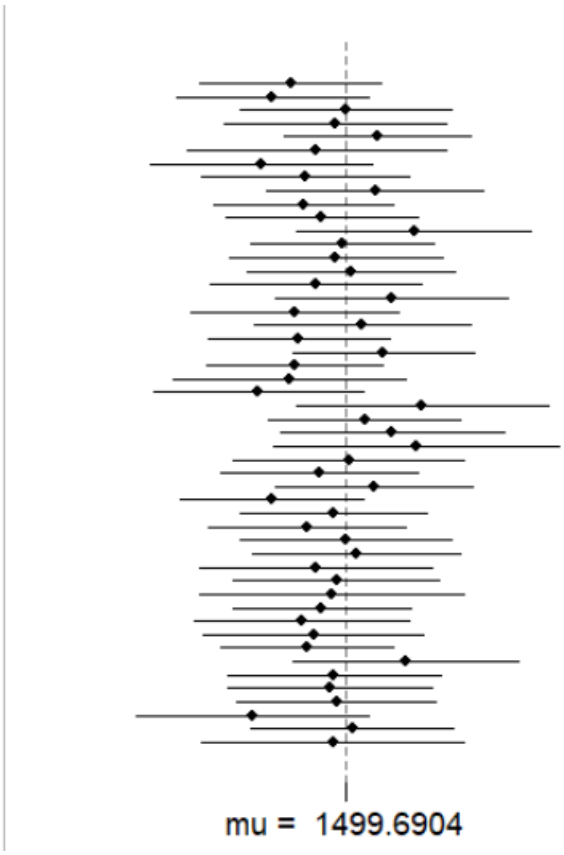
```
plot_ci(lower_vector, upper_vector, mean(population))
```

The proportion is not exactly 95% since we only took 50 samples if we took a larger number of samples we would have a proportion closer to the confidence interval.



2) Pick a confidence level of your choosing, provided it is not 95%. What is the appropriate critical value t?
Choosing 99% confidence would get a t value of 2.678

3) Calculate 50 confidence intervals at the confidence level you chose in the previous question. You do not need to obtain new samples, simply calculate new intervals based on the sample means and standard deviations you have already collected. Using the `plot_ci` function, plot all intervals and calculate the proportion of intervals that include the true population mean. How does this percentage compare to the confidence level selected for the intervals? (Include plot.)
This interval has 100 percent proportion that contain the true mean while the confidence interval I chose was 99% so I got a higher proportion then the confidence interval I chose.


mu = 1499.6904