**EXPERIMENT 8**

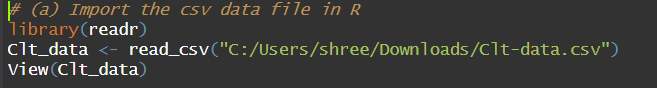
**A pipe manufacturing organization produces different kinds of pipes. We are given the monthly data of the wall thickness of certain types of pipes (data is available on LMS Clt-data.csv).**

**The organization has an analysis to perform and one of the basic assumptions of that analysis is that the data should be normally distributed.**

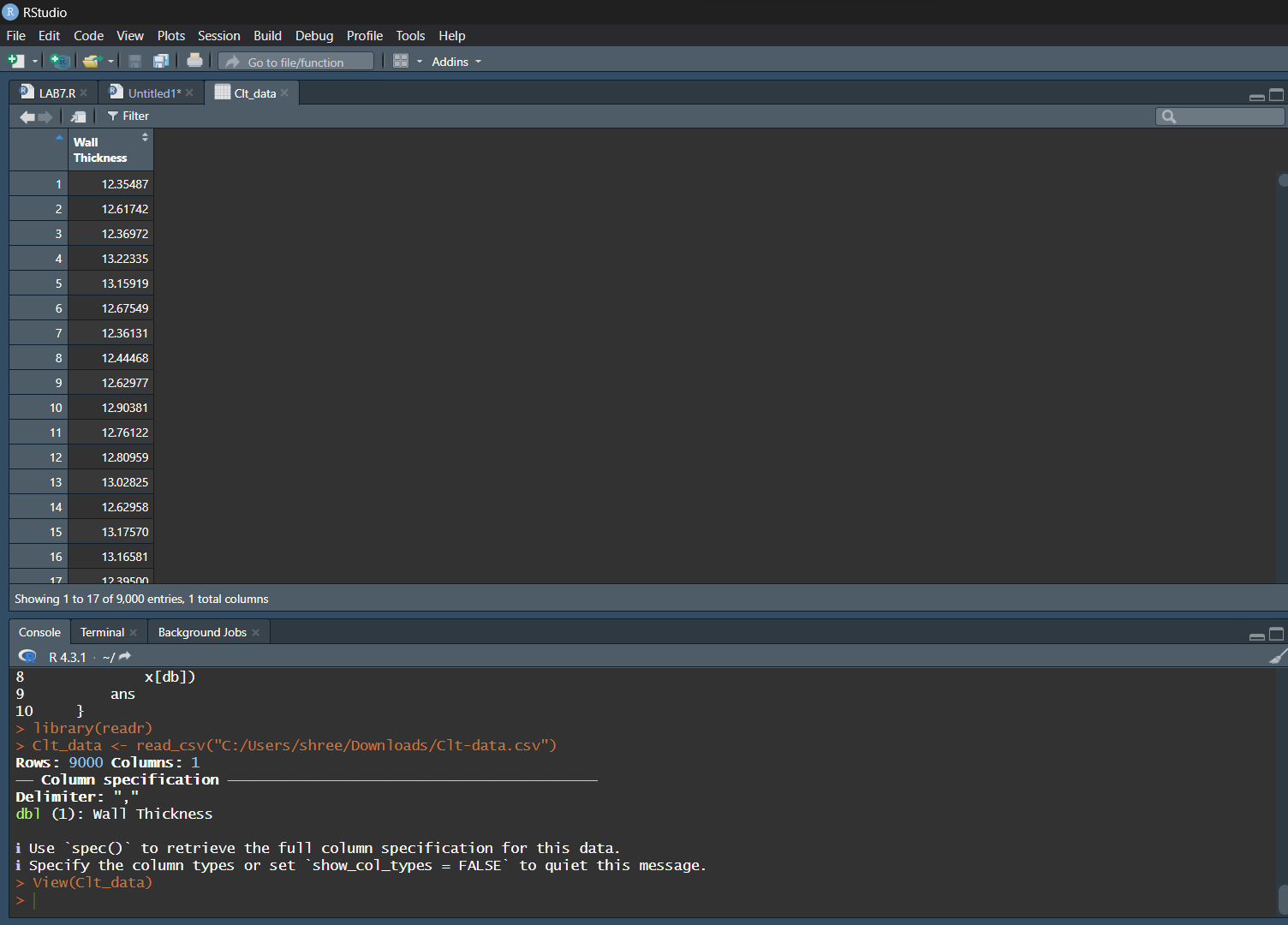
**You have the following tasks to do:**

1. **Import the csv data file in R.**

Code:

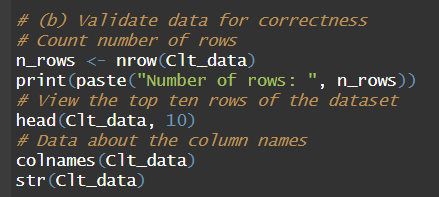


Output:

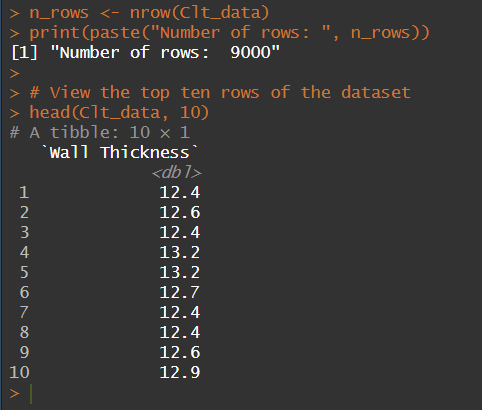
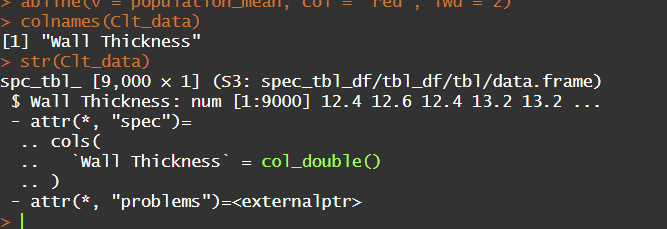


1. **Validate data for correctness by counting number of rows and viewing the top ten rows of the dataset.**

Code:

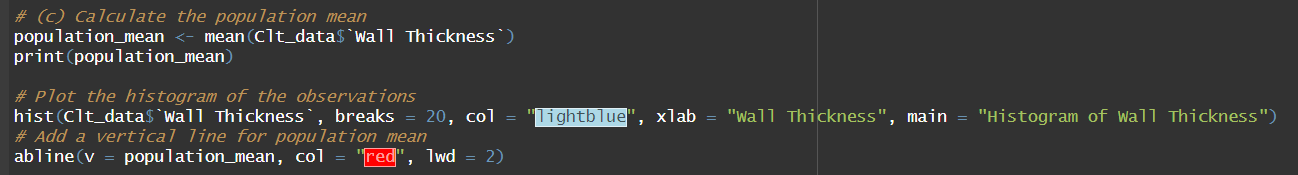


Output:

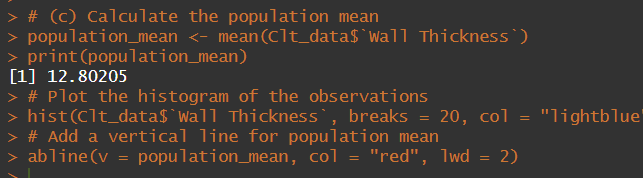
 

1. **Calculate the population mean and plot the observations by making a histogram.**

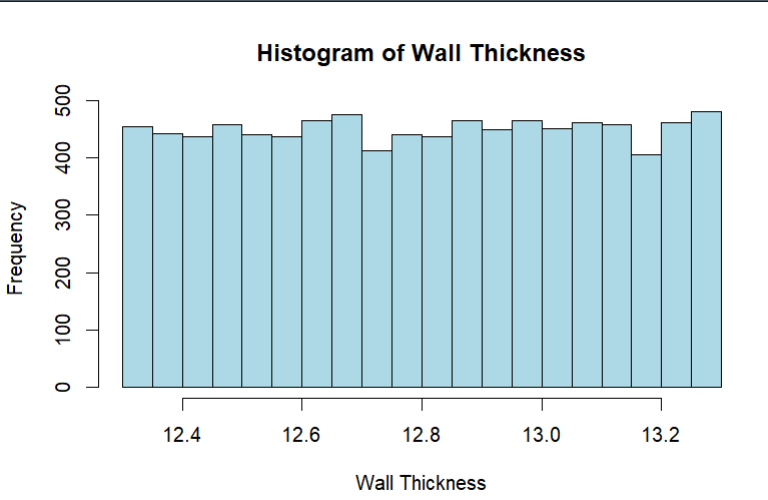
Code:



Output:

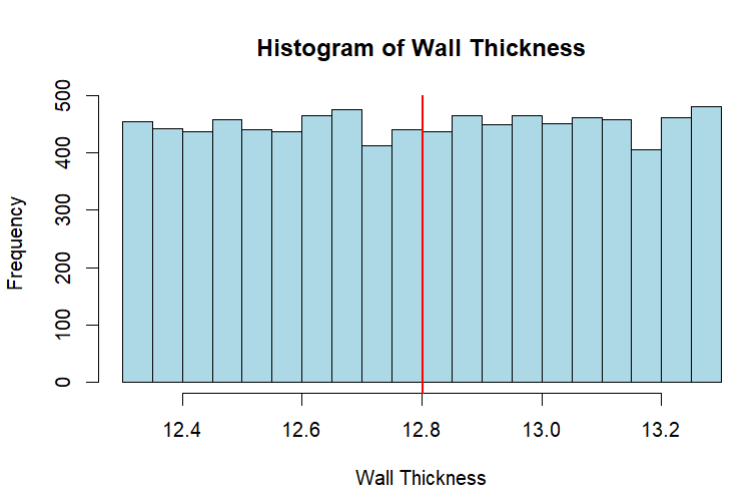


The Histogram:



1. **Mark the mean computed in last step by using the function abline.**

Histogram with abline:



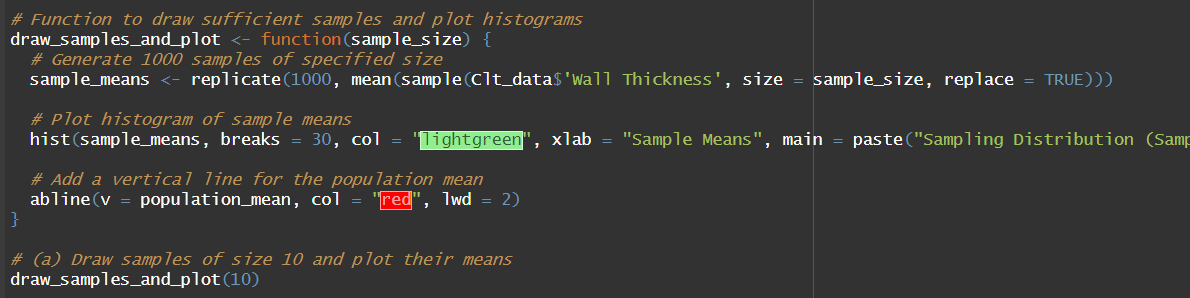
See the red vertical line in the histogram? That’s the population mean. Comment on whether the data is normally distributed or not?

Ans: Although the **abline** is right in the middle of the histogram still it does not confirm it is a normal distribution. After studying the histogram, we can clearly say that the histogram does NOT resemble a BELL-SHAPED CURVE so we can say that the data is **NOT NORMALLY DISTRIBUTED**.

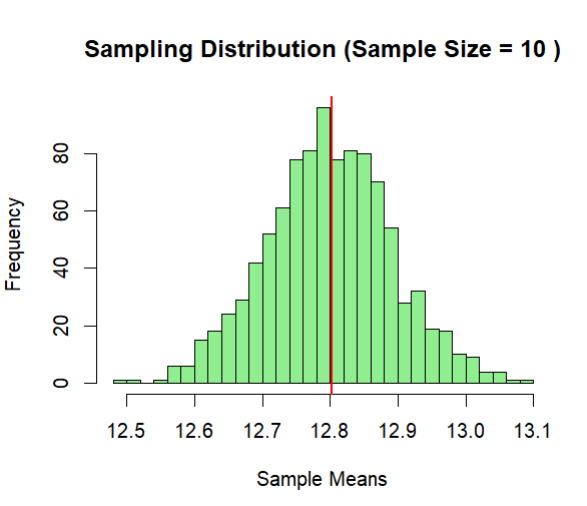
Now perform the following tasks:

1. **Draw sufficient samples of size 10, calculate their means, and plot them in R by making histogram. Do you get a normal distribution.**

Code:



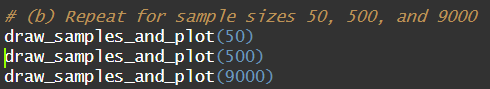
Histogram:



The mean is coming around 12.8. The histogram clearly represents a Bell-Shaped curve. So we can conclude that the sample is **NORMALLY DISTRIBUTED**.

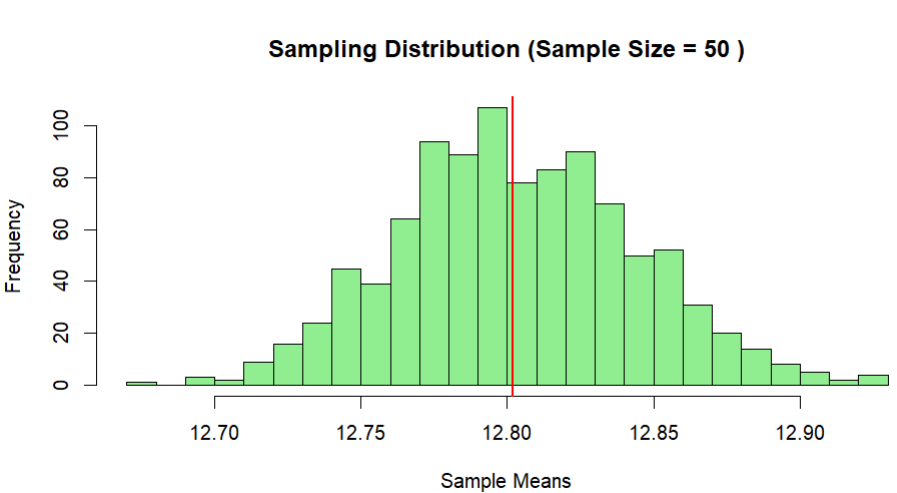
1. **Now repeat the same with sample size 50, 500 and 9000. Can you comment on what you observe.**

Code:



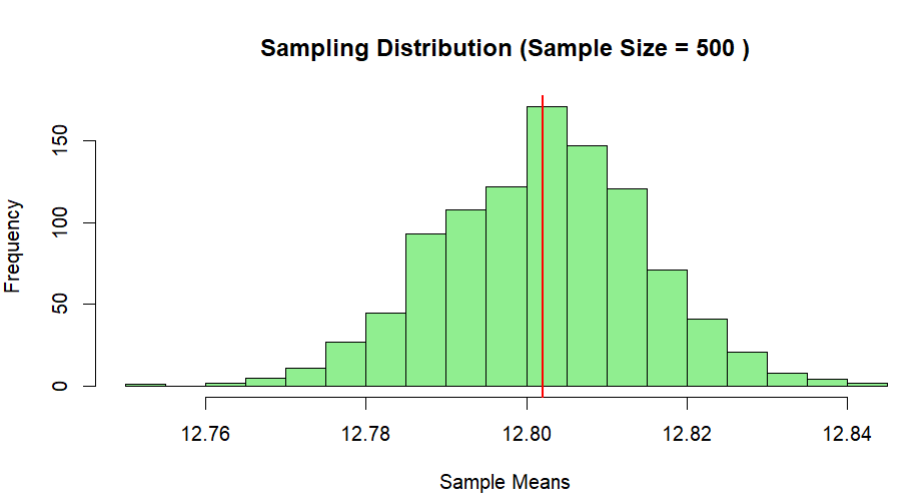
**50:**

Histogram:

****

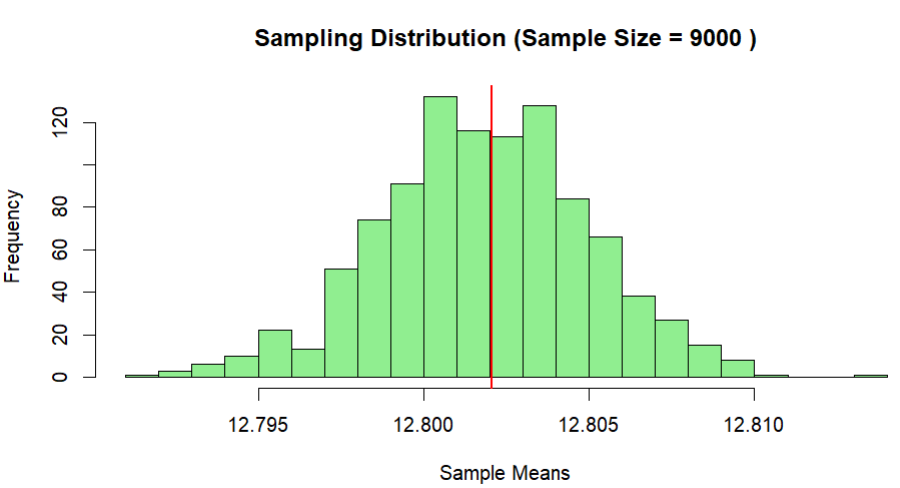
**500:**

Histogram:

****

**9000:**

Histogram:

****

Here, we get a good bell-shaped curve and the sampling distribution approaches normal distribution as the sample sizes increase. Therefore, we can recommend the organization to use sampling distributions of mean for further analysis.