Name: Shreeya Chatterji
Roll: 102103447
Class: CO16

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 assumption of that analysis is that the data should be normally distributed.

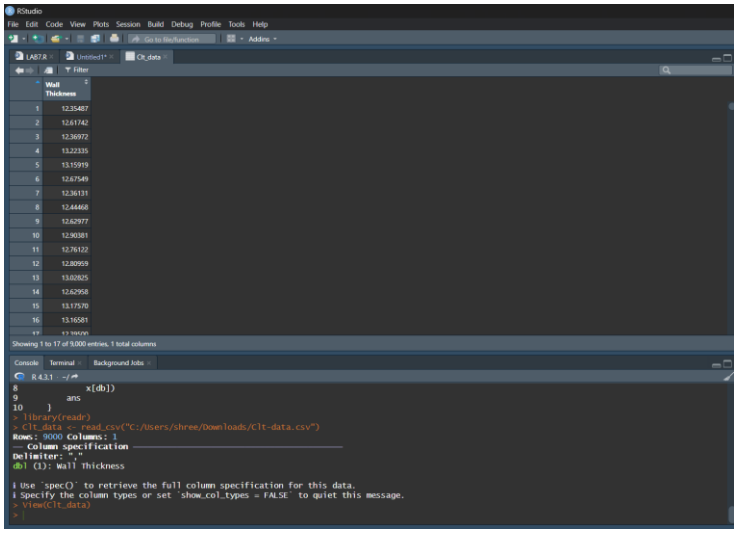
You have the following tasks to do:

(a) Import the csv data file in R.

Code:

```
# (a) Import the csv data file in R
library(readr)
Clt_data <- read_csv("C:/Users/shree/Downloads/Clt-data.csv")
View(Clt_data)
```

Output:



(b) Validate data for correctness by counting number of rows and viewing the top ten rows of the dataset.

Code:

```
# (b) Validate data for correctness
# Count number of rows
n_rows <- nrow(Clt_data)
print(paste("Number of rows: ", n_rows))
# View the top ten rows of the dataset
head(Clt_data, 10)
# Data about the column names
colnames(Clt_data)
str(Clt_data)
```

Output:

```
> n_rows <- nrow(Clt_data)
> print(paste("Number of rows: ", n_ro
[1] "Number of rows:  9000"
>
> # View the top ten rows of the datas
> head(Clt_data, 10)
# A tibble: 10 × 1
   `Wall Thickness`
              <dbl>
 1             12.4
 2             12.6
 3             12.4
 4             13.2
 5             13.2
 6             12.7
 7             12.4
 8             12.4
 9             12.6
10             12.9
```

```
> colnames(Clt_data)
[1] "Wall Thickness"
> str(Clt_data)
spc_tbl_ [9,000 × 1] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ Wall Thickness: num [1:9000] 12.4 12.6 12.4 13.2 13.2 ...
 - attr(*, "spec")=
  .. cols(
  ..   `Wall Thickness` = col_double()
  .. )
 - attr(*, "problems")=<externalptr>
>
```

Name: Shreeya Chatterji
Roll: 102103447
Class: CO16

(c) Calculate the population mean and plot the observations by making a histogram.

Code:

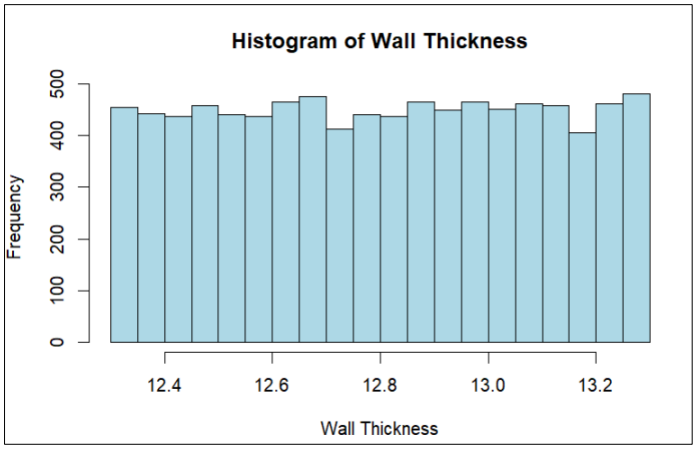
```
# (c) Calculate the population mean
population_mean <- mean(Clt_data$`Wall Thickness`)
print(population_mean)

# Plot the histogram of the observations
hist(Clt_data$`Wall Thickness`, breaks = 20, col = "lightblue", xlab = "Wall Thickness", main = "Histogram of Wall Thickness")
# Add a vertical line for population mean
abline(v = population_mean, col = "red", lwd = 2)
```

Output:

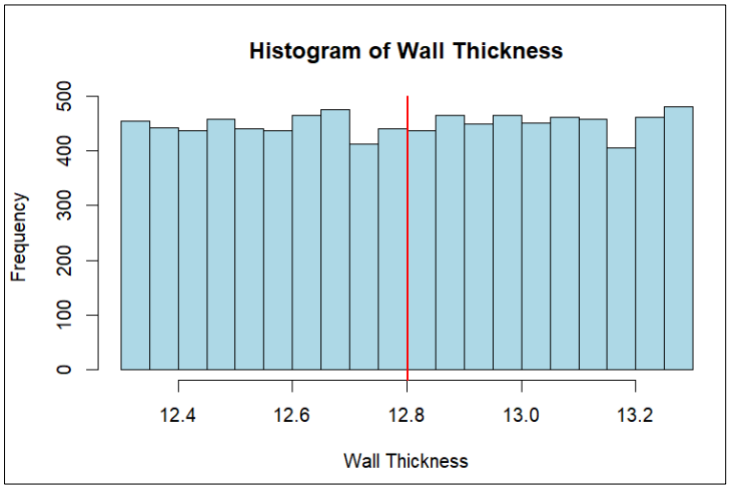
```
> # (c) Calculate the population mean
> population_mean <- mean(Clt_data$`Wall Thickness`)
> print(population_mean)
[1] 12.80205
> # Plot the histogram of the observations
> hist(Clt_data$`Wall Thickness`, breaks = 20, col = "lightblue
> # Add a vertical line for population mean
> abline(v = population_mean, col = "red", lwd = 2)
```

The Histogram:



(d) 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**.

Name: Shreeya Chatterji
Roll: 102103447
Class: CO16

Now perform the following tasks:

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

Code:

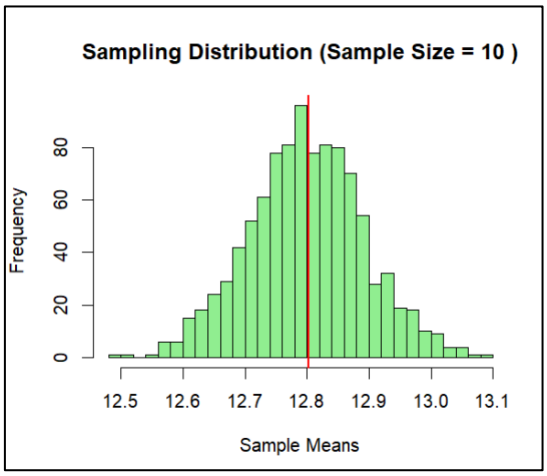
```
# Function to draw sufficient samples and plot histograms
draw_samples_and_plot <- function(sample_size) {
  # Generate 1000 samples of specified size
  sample_means <- replicate(1000, mean(sample(Clt_data$'Wall Thickness', size = sample_size, replace = TRUE)))

  # Plot histogram of sample means
  hist(sample_means, breaks = 30, col = 'lightgreen', xlab = "Sample Means", main = paste("Sampling Distribution (Samp

  # Add a vertical line for the population mean
  abline(v = population_mean, col = "red", lwd = 2)
}

# (a) Draw samples of size 10 and plot their means
draw_samples_and_plot(10)
```

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**.

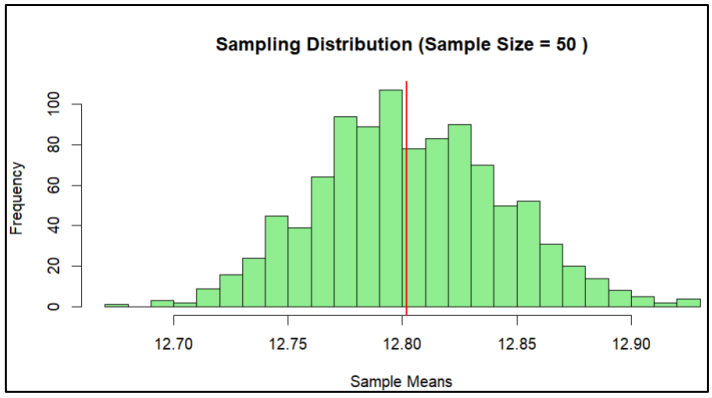
(b) Now repeat the same with sample size 50, 500 and 9000. Can you comment on what you observe.

Code:

```
# (b) Repeat for sample sizes 50, 500, and 9000
draw_samples_and_plot(50)
draw_samples_and_plot(500)
draw_samples_and_plot(9000)
```

50:

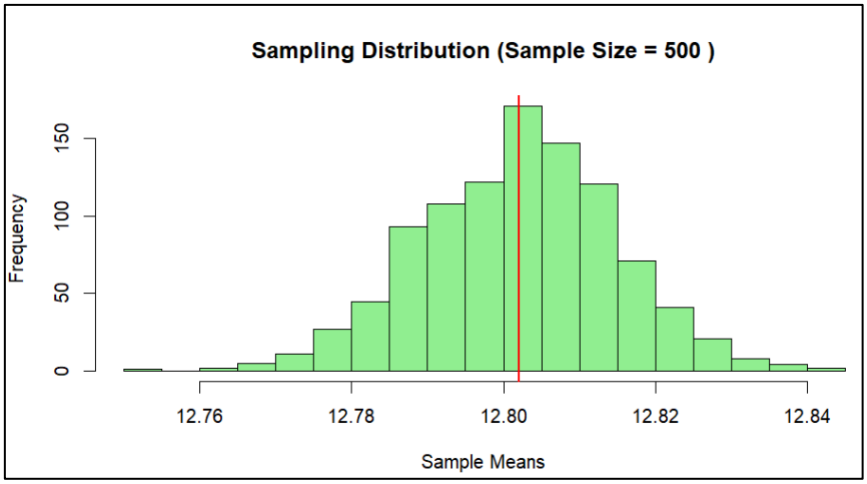
Histogram:



Name: Shreeya Chatterji
Roll: 102103447
Class: CO16

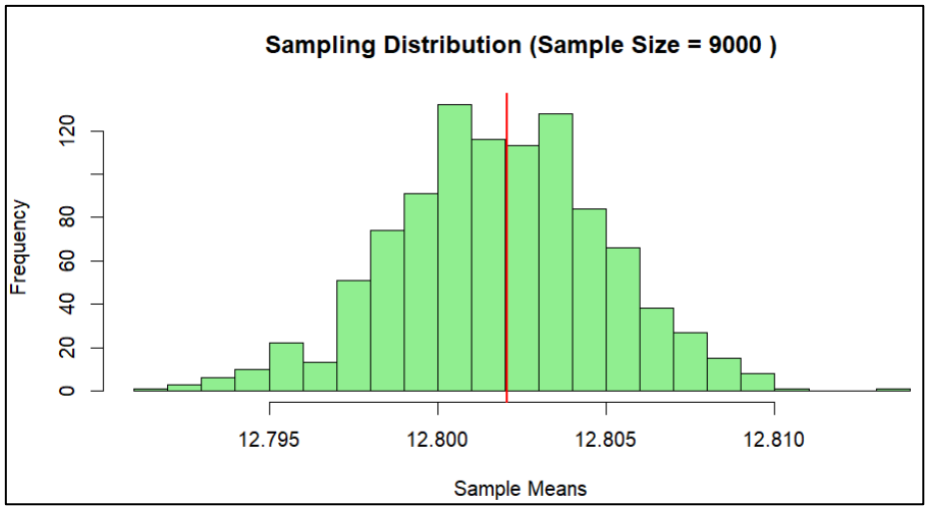
500:

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.