**Course Name: Big Data and Analytics Lab (Code: BCSE0183)**

**Semester: 6th Section: 3-D/G (Batches) Academic Session: Feb 2022 – May 2022**

Lab Assignment – 07

Problems on Visualizing Given Dataset using R Programming

Download Dataset of **Students Survey** Record at (<https://drive.google.com/file/d/1dsJHUiSk4aYXyQ8ZbQUkfNXCyayLAmEa/view?usp=sharing>)

**Queries**:

1) How many students are in given dataset?

2) How many of the first 10 students in the dataset had names longer than 5 letters?

3) How long is the name of the first student in the dataset who is happy less than 40% of the time?

4) Which variable tells us how many letters are in each student’s first name? The name of this variable in the dataset is?

5) What type of variable is this in Q4 --categorical or quantitative?

**Determine the population parameters:**6) Visualize the shape of the population data by making a histogram.    
7) Calculate the “true” mean and standard deviation of the population.

**Compare the sample statistics:**8) Draw 1,000 samples of size n=5 from the population data.  Calculate the mean of each sample.  
9) Graph these 1,000 sample means in a histogram and examine the shape.  
10) Calculate the mean and standard deviation of the sampling distribution.  
11) Repeat this process for samples of size n=15 and n=25.  
12) Compare the results you get to the predictions of the Central Limit Theorem.

**NOTE**: Please visit the link for more examples on CSV file operation in R Programming at (<https://www.geeksforgeeks.org/working-with-csv-files-in-r-programming/> )

Course Faculty: **Dr. Robin Singh Bhadoria**, Assistant Professor Department of Computer Engineering & Applications

GLA University, Mathura (U.P.)