Under the aegis of Vijayam Educational Trust

# CATALYST COLLEGE
(A Unit of CIMAGE Group of Institutions)
Institution approved by Education Department, Government of Bihar, Affiliated to Patliputra University, Patna

Ref: CC\WRSP-MOT\19\15/20

Date: 26-May-2019

## NOTICE

This is to inform all the Students that a workshop on R Programming by IIT Bombay will be organized on 12.6.2019 from 9:30 AM to 5:30 PM in the auditorium of Catalyst College.

The workshop is completely free, and no money will be charged for the Training or Certification.

Interested students are instructed to meet the Activity In-Charge / Class Coordinator for more details and their registration.

By the order of

Principal

Date:12-06-2019

Workshop Title

# R Programming: Unlocking Data Science Potential

# Presented by IIT Bombay

Number of Students Participated: 52

---

Objective:

This workshop aims to introduce participants to R Programming, a powerful open-source language widely used for data analysis, statistical computing, and data visualization. Delivered by experts from IIT Bombay, the workshop will provide hands-on experience with R, teach key programming concepts, and explore its real-world applications in data science, machine learning, and statistical modeling.

By the end of this workshop, participants will gain the necessary skills to use R for analyzing complex data sets, building predictive models, and visualizing results, empowering them to make data-driven decisions in their respective fields.

---

Model 1. Welcome & Introduction to R Programming

- Welcome Remarks: Introduction to the workshop objectives, format, and key takeaways.

- About the Facilitators: Brief introduction to the speakers from IIT Bombay and their expertise in R programming, data science, and analytics.

- Overview of R:

  o Why R is one of the most powerful tools for data analysis and statistical computing.

  o Key benefits of R: open-source, extensive libraries, and strong community support.

Model 2. Getting Started with R

- Installing R & RStudio:

    o Step-by-step instructions on downloading and installing R and RStudio (the IDE for R).

    o Setting up RStudio and configuring your environment for efficient coding.

- Introduction to R Studio:

    o Navigating the RStudio interface: Console, Script, Environment, and Plots.

    o Writing and running your first R script: Basic commands and syntax.

- Basic R Syntax:

    o Variables, data types, and operators in R. o Working with vectors, matrices, and data frames.

    o Functions and how they are used to perform calculations and operations in R.

Model 3. Data Structures in R

- Vectors, Lists, Matrices, and Data Frames:

    o Understanding how to work with various data structures in R, and when to use them.

    o Hands-on practice: Creating and manipulating data structures.

- Importing and Exporting Data:

    o How to load and read datasets into R (e.g., CSV, Excel, and SQL databases). o Data wrangling techniques: Cleaning and preprocessing data (handling missing values, changing data types, etc.).

- Exploring Data with Basic Functions:

    o summary(), str(), head(), tail(), and dim() to explore datasets and inspect the structure of data frames.

    Sorting, filtering, and subsetting data.

Model 4. Data Manipulation and Cleaning

- dplyr Package:

    o Introduction to the dplyr package for data manipulation (filter, select, mutate, arrange, and summarize). o How to perform common data manipulation tasks such as aggregating data, summarizing, and grouping.

- Tidyr Package:

    o Introduction to tidyr for tidying and reshaping data.

    o Functions like spread(), gather(), and separate() to convert data into a usable format.

- Handling Missing Data:

    o Techniques for identifying and dealing with missing values in your datasets using tidyverse functions.

    o Imputation strategies and data cleaning workflows.

Model 5. Data Visualization with R

- Introduction to ggplot2:

    o Introduction to ggplot2, R's powerful and flexible visualization library.

    o Understanding the grammar of graphics and how to create plots in R.

- Creating Basic Plots:

    o How to create bar plots, scatter plots, line plots, histograms, and box plots.

    o Customizing plots: Titles, labels, colors, and themes.

- Advanced Visualizations:

    o Creating complex visualizations like heatmaps, geographic maps, and interactive plots using plotly and leaflet.

- Visualization Best Practices:

    How to communicate data effectively using visualizations.

o   Guidelines for creating clear, informative, and aesthetically pleasing charts.

---

Model 6. Statistical Analysis and Modeling in R

- Descriptive Statistics:

  o   Using R to compute basic statistical measures like mean, median, standard deviation, correlation, and variance.

  o   Understanding and applying hypothesis testing (t-tests, chi-squared tests) in R.

- Linear Regression:

  o   Introduction to simple and multiple linear regression models in R.

  o   How to interpret model coefficients, residuals, and performance metrics.

- Building Predictive Models:

  o   Introduction to predictive modeling with machine learning in R. o Overview of model training, testing, and validation using algorithms like decision trees, random forests, and k-nearest neighbors.

---

Model 7. R for Machine Learning and AI

- Overview of Machine Learning in R:

  o   The role of R in machine learning and AI applications.

  o   Popular machine learning libraries in R: caret, randomForest, xgboost.

- Supervised vs. Unsupervised Learning:

  o   Difference between supervised and unsupervised learning.

  o   Hands-on session with classification models (e.g., logistic regression, k-NN) and clustering models (e.g., k-means).

- Model Evaluation and Performance:

- o Model 8. Real-World Applications of R How to evaluate model performance using metrics like accuracy, precision, recall, and confusion matrix. Cross-validation and overfitting/underfitting in machine learning model

- R in Data Science:

  - o How data scientists use R for cleaning, analyzing, and visualizing large datasets. o Case studies of R in industries like healthcare, finance, marketing, and education.

- R in Business Analytics:

  - o How businesses use R for data-driven decision-making: Market analysis, customer segmentation, and sales forecasting.

- R in Research and Academia:

  - o Applications of R in academic research, including statistical analysis, survey data, and hypothesis testing.

- R in Big Data:

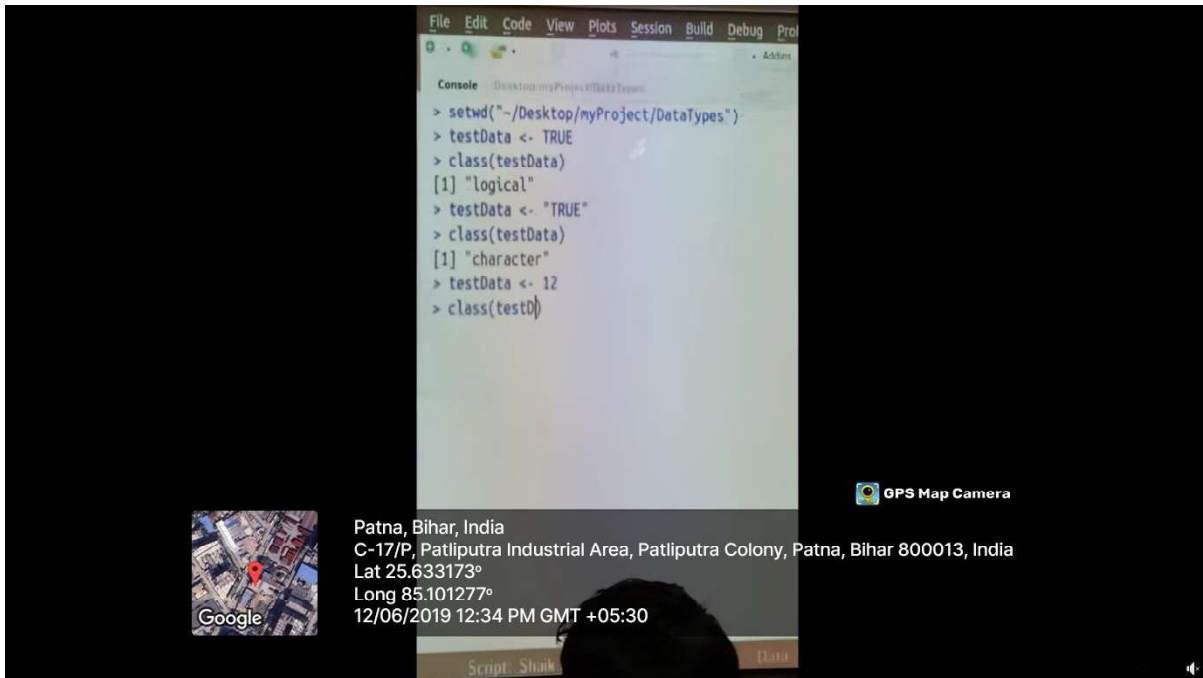  - o Using R to work with big data sets, including integration with Hadoop and Spark.

Key Takeaways:

- R Basics: Understanding R syntax, data types, and key libraries like dplyr, ggplot2, and caret.

- Data Analysis Skills: Learning how to manipulate, clean, and visualize data with R.

- Statistical and Predictive Modeling: Knowledge of regression analysis, hypothesis testing, and building machine learning models in R.

- Real-World Applications: Understanding how R is used in various industries for data-driven decision-making and research.

# Workshop on R Programming by IIT Bombay

## Date:-12/06/2019



Patna, Bihar, India
C-17/P, Patliputra Industrial Area, Patliputra Colony, Patna, Bihar 800013, India
Lat 25.633133°
Long 85.101247°
12/06/2019 12:14 PM GMT +05:30



Patna, Bihar, India
C-17/P, Patliputra Industrial Area, Patliputra Colony, Patna, Bihar 800013, India
Lat 25.633173°
Long 85.101277°
12/06/2019 12:34 PM GMT +05:30
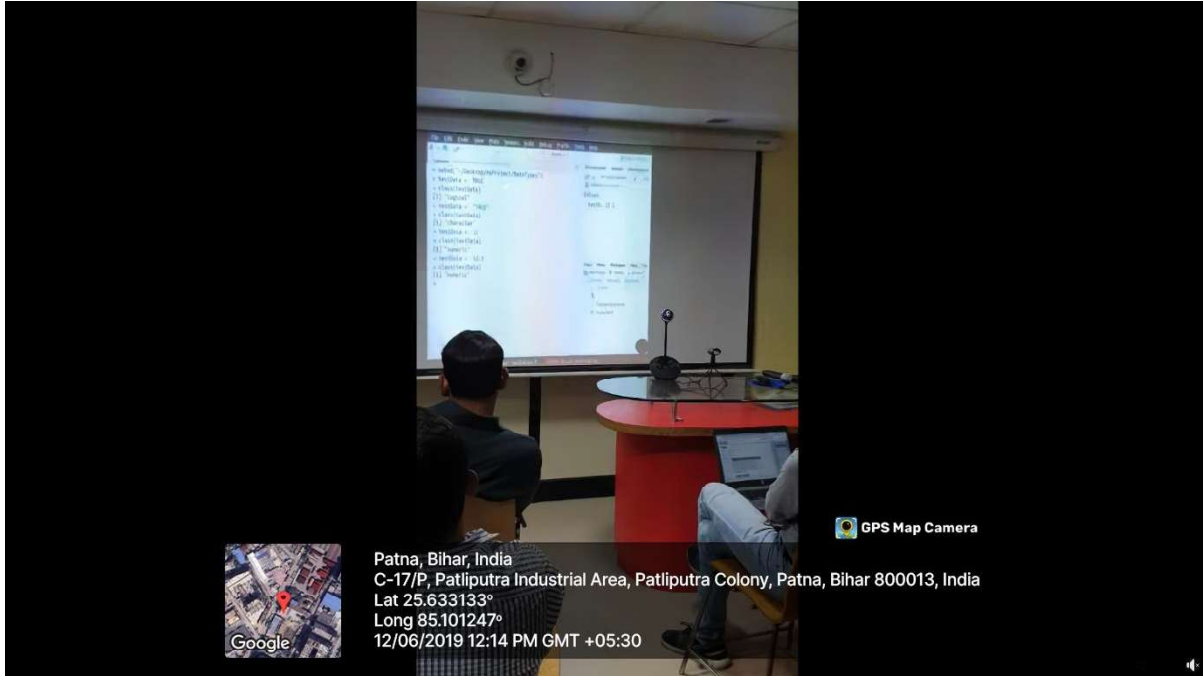
**Principal**
**CATALYST COLLEGE**
Plot No.- C-16(P) Patliputra Industrial Area
Patliputra, Patna-13

# Registration

## For Workshops/Seminars/Conferences during Academic Year 2019-2020

## Workshop on R Programming by IIT Bombay

## (12 June 2019)

| S. No. | ID | Name of the student | Student's Signature |
|--------|------|---------------------|---------------------|
| 1 | 445-7250 | Manish Kumar | |
| 2 | 445-6977 | Nur Alam | |
| 3 | 445-6862 | Prakash Raj | |
| 4 | 445-6853 | Prashant Kumar | |
| 5 | 445-6974 | Prince Kumar Singh | |
| 6 | 445-6730 | Raghav Raman Choudhary | |
| 7 | 445-6747 | Ranjeet Kumar Yadav | |
| 8 | 445-6733 | Raunak Rani | |
| 9 | 445-6854 | Sanjeev Kumar | |
| 10 | 445-7423 | Satish Kumar | |
| 11 | 445-6883 | Saurav Kumar | |
| 12 | 445-6761 | Shankar Kumar | |
| 13 | 445-6993 | Shiv Jee Kumar Yadav | |
| 14 | 445-6728 | Shivam Shekhr | |
| 15 | 445-7029 | Sonal Kumar Singh | |
| 16 | 445-6770 | Subham Kumar | |
| 17 | 445-6742 | Subham Shankar | |
| 18 | 445-7604 | Tanuja | |
| 19 | 445-6991 | Ujjval Kumar Verma | |
| 20 | 445-7001 | Vikash Kumar | |
| 21 | 445-7023 | Vikash Kumar | |
| 22 | 445-6739 | Vinayak Gupta | |
| 23 | 445-6759 | Vishal Pandey | |
| 24 | 45-7432 | Bolbam Kumar | |
| 25 | 445-6741 | Kanish Kumar | |
| 26 | 445-6948 | Manish Raj | |
| 27 | 445-6737 | Manisha Kumari | |
| 28 | 445-6933 | Ravnak Kumar | |
| 29 | 445-7275 | Aaseen Alam | |
| 30 | 445-7343 | Akshat Raj | |
| 31 | 445-7027 | Anish Raj | |
| 32 | 445-7345 | Avinash Kumar | |

| 33 | 445-7384 | Deeplal Ram | Deeplal Ram |
|----|----------|-------------|-------------|
| 34 | 445-7392 | Kajal Kumari | Kajal Kumari |
| 35 | 445-7033 | Kajal Kumari | R. Kumari |
| 36 | 445-6886 | Kamlesh Kumar Singh | Kamlesh Kr |
| 37 | 445-7377 | Kundan Kumar | R. Kumar |
| 38 | 445-7039 | Manish Kumar | Manish kumar |
| 39 | 445-7483 | Md Arbaz Ansari | Md. Arbaz. |
| 40 | 445-7252 | Md Faizan | Md. Faizan. |
| 41 | 445-7430 | Mukesh Kumar Jha | M. K. Jha. |
| 42 | 445-7469 | Nitish Kumar | Nitish k |
| 43 | 445-7379 | Pankaj Kumar | Pankaj Kuma |
| 44 | 445-7375 | Prashant Kumar | P. Kr. |
| 45 | 445-7041 | Rahul Kumar | Rahul Kr |
| 46 | 445-6979 | Ramesh Kumar | Ramesh. |
| 47 | 445-7363 | Ramesh Ranjan | R. Ranjan. |
| 48 | 445-7347 | Shakir Ansari | Shakir |
| 49 | 445-7438 | Sunny Kumar | Sunny Kumar |
| 50 | 445-7471 | Tannu Priya | Tannu Priya |
| 51 | 445-7485 | Deepankar Kumar | Deepankar |
| 52 | 445-7361 | Poonam Kumari | Poonam |

(Sign.)
Course Coordinator