

# Introduction to R for Data Science

## ABOUT THE COURSE

### INTRODUCTION

If you've been looking for an intense, methodological intro training class on R for data science, you've come to the right place. Our introduction to R for data science training class provides an in-depth look into the world of applied data science. Weaving complex methods with practical application, our training ensures the highest level of comprehension regarding basics of handling data from importing until decision making. Specifically, you'll learn how to use RStudio for data science related projects along with handling data in R. You'll also learn about tools and techniques for creating analytic graphics for data analysis. Furthermore, you will learn how to make a decision using various applied machine learning techniques. This training will provide you with all of the basic knowledge you need to know to dive into data science field.

### OBJECTIVES

At the end of this program participants will be able to achieve the following objectives:

1. How to use RStudio for data science related projects.
2. How to import, clean and manipulate data for analysis.
3. How to create analytic graphics for in-depth analysis of the data.
4. How to make a decision using applied and practical machine learning techniques.
5. Industry used tools and best practices for data science.

### METHODOLOGY

A variety of presentation methods will be used including instructor-led training, tutorial & discussion, laboratory sessions and exercises.

### TECHNOLOGIES LEARNT

Technologies that you will learn and develops throughout the course:

- R statistical programming language.
- RStudio Integrated Development Environment (IDE).
- Data science basics.
- Using R for data science.
- Concepts of machine learning.
- Applied and practical machine learning.

## COURSE SETTINGS

- Venue : FSKTM, UM
- Date : 31<sup>st</sup> October - 1st November 2017 (2 days) from 9.00am-5.00pm
- Level : Basic to Intermediate
- Inclusive : Morning Break, Lunch, Notes and Certificate of Completion
- Registration : <https://goo.gl/WtHVdS>
- Audience : Lecturers, instructors, IT Officers, Data enthusiast, Data analyst, Data Engineer, Undergraduates, Postgraduates



**COURSE SCHEDULE****DAY 1****Introduction to RStudio**

- Installation and overview of RStudio application.
- Introduction to various functions of RStudio along with basics of R language.

**Data Importing**

- Introduction to multiple types of data.
- Import various types of data into RStudio.
- Learn the parameters in the data import functions to further enhance importing efficiency.

**Data Importing**

- Introduction to basics of data cleaning using RStudio.
- Understand the concepts of raw data and tidy data.
- Learn how to handle missing values, ordering and sorting of data.
- Clean data using built-in functions and external packages in RStudio.

**Data Manipulation**

- Introduction to basics of data manipulation using RStudio.
- Perform subsetting, binding and sampling of the data.
- Learn how to manipulate data using various functions in RStudio.

**DAY 2****Graphing**

- Introduction to different types of graph
- Understand the principle of analytic graphics.
- Learn the parameter of different graphic functions in RStudio.

**Decision Making and Machine Learning using R**

- Introduction to the basic components of building and applying prediction functions with an emphasis on practical applications.
- Provide basic grounding in concepts such as training and tests sets, overfitting, and error rates.
- Introduction to a range of model-based and algorithmic machine learning methods
- Cover the complete process of building prediction functions including data collection, feature creation, algorithms, and evaluation.

**INSTRUCTOR BIOGRAPHY**

Nor Badrul Anuar obtained his Ph.D. in Information Security from Centre for Security, Communications and Network Research (CSCAN), Plymouth University, UK in 2012 and Master of Computer Science from University of Malaya, Malaysia in 2003. He is an Associate Professor at the Faculty of Computer Science and Information Technology in University of Malaya, Kuala Lumpur. He has published a number of conference and journal papers locally and internationally. His research interests include information security (i.e. intrusion detection systems), data sciences, artificial intelligence and library information systems.

**MORE INFORMATION**



**FSKTM TECHNOVATIONS**

**Email: [fsktm.technovations@um.edu.my](mailto:fsktm.technovations@um.edu.my) | Office: +603-79676435**

UM INNOVATIONS SDN. BHD. (772583-X)  
Suites 2A & 2B, Incubator 5, Level 5,  
Research Management & Innovation Complex  
University of Malaya, 50603 Kuala Lumpur  
Malaysia