



## INTRODUCTION

R is a flexible, powerful and free software application for statistics and data analysis. It is used in a wide variety of research disciplines, and has become increasingly popular in recent years. This workshop aims to provide the participants with essential skills for analyzing big data with R. The workshop will cover the basics of data visualization, data management, programming, the implementation of commonly used data analytics methods. The workshop will focus on the practical implementation of the methods covered. No prior experience with R is required.

## OBJECTIVES

Participants in this workshop will:

- ◆ Learn the basics of R, data handling in R and the various ways to create scripts and programs.
- ◆ Learn to explore and visualize data in R through several packages including dplyr and ggplot2.
- ◆ Apply predictive analytics methods on big data to generate meaningful results.
- ◆ Get familiar with different big data platforms including Spark and how to integrate them with R.
- ◆ Explore fast, streaming, and scalable data analysis with the most cutting-edge technologies in the market.

## DAY 1

### Introduction to R Programming

- ~ Brief R background, why R?
- ~ RStudio basics, panes, snippets and projects
- ~ The R system: workspace, working directory, packages
- ~ Getting help
- ~ The R language basics, function arguments
- ~ R objects
- ~ Working with R objects: subsetting, changing data types, names, missing values, reading and writing data
- ~ The “new R workflow”: dplyr, magrittr and tidyr packages for data manipulation
- ~ Traditional (base) plotting in R

## DAY 3

### Inferential Statistics and Sharing Results

- ~ Basic analyses: general linear models
- ~ Sharing R results: markdown, Rdata files, rds files
- ~ Introducing supervised and unsupervised learning methods

## DAY 5

### Unsupervised Learning Methods and Big Data

- ~ Unsupervised learning methods: K-means and hierarchical clustering methods
- ~ Introducing big data analytics and Spark
- ~ Big data in R through the sprklyr package
- ~ Putting it all together: workshop project

## DAY 2

### Intermediate and advanced R programming topics

- ~ Data visualization with ggplot2 package
- ~ Advanced R: control structures (for, if else, while)
- ~ Advanced R: writing more useful functions
- ~ A brief hands-on demonstration of other relevant technologies (R markdown, Shiny)

## DAY 4

### Supervised Learning Methods

- ~ Decision Trees
- ~ Neural Networks
- ~ Support vector machines
- ~ Model fitting and validation
- ~ Introducing the caret package

## DR. HUSSAM ALSHRAIDEH...

Dr. Alshraideh is an Associate Professor of Operations Research and Statistics at the Mechanical and Industrial Engineering Department at Rochester Institute of Technology (RIT-Dubai). He holds a dual Ph.D degree in Industrial Engineering and Operations Research with a minor in Statistics from The Pennsylvania State University.

He also holds a Masters degree in Quality Engineering from Arizona State University. His current research interests include statistical process optimization and smart data analytics applications in healthcare related fields.

## HOW TO REGISTER?

[www.rit.edu/dubai/big-data-r](http://www.rit.edu/dubai/big-data-r)

## NEED MORE INFORMATION?

Speak to Mr. Mustafa Hariri at +971 4 371 2xxx