

---

## Introduction to R Programming for Data Science & Analytics (TTDS6683)

---

### Overview

R is a functional programming environment for business analysts and data scientists. It's a language that many non-programmers can easily work with, naturally extending a skill set that is common to high-end Excel users. It's the perfect tool for when the analyst has a statistical, numerical, or probabilities-based problem based on real data and they've pushed Excel past its limits. This course is a comprehensive hands-on look at the common scenarios encountered in analysis and presents practical solutions. In this course, special attention is paid to data science theory including AI grouping theory. A discussion of using R with AI libraries like Madlib are included.

---

### Prerequisite Comments

This course, geared for Data Analyst and Data Scientists who need to learn the essentials of how to program in R. Incoming students should have prior experience working with Excel or SAS, and should know the basics of SQL. Students should have intermediate-level experience in their field, and prior experience working with programming languages.

Follow On Courses: Our core R and Python programming, data science, analytics, AI and machine learning training courses provide students with a solid foundation for continued learning based on role, goals, or their areas of specialty. Please inquire for next step recommendations based on your goals

---

### Target Audience

This course, geared for Data Analyst and Data Scientists who need to learn the essentials of how to program in R. Incoming students should have prior experience working with Excel or SAS, and should know the basics of SQL. Students should have intermediate-level experience in their field, and prior experience working with programming languages.

---

### Course Objectives

This course provides indoctrination in the practical use of the umbrella of technologies that are on the leading edge of data science development focused on R and related tools. Working in a hands-on learning environment, led by our expert practitioner, students will learn R and its ecosystem, and where it's a better a tool than Excel.

This course is approximately 50% hands-on, combining expert lecture, real-world demonstrations and group discussions with machine-based practical labs and exercises. Our engaging instructors and mentors are highly experienced practitioners who bring years of current "on-the-job" experience into every classroom. Working in a hands-on learning environment, guided by our expert team, attendees will learn about and explore:

R Language and Mathematics

How to work with R Vectors

How to read and write data from files, and how to categorize data in factors

How to work with Dates and perform Date math

How to work with multiple dimensions and DataFrame essentials

Essential Data Science and how to use R with it

Visualization in R

How R can be used in Spark (Optional / Overview)

---

### Course Outline

## 1 - From Excel or SAS to R (Optional)

Common challenges with Excel / SAS  
The R Environment  
Hello, R

## 2 - Working with R Studio

Rshiny  
Rpresentations  
Rmarkdown

## 3 - R Basics

Simple Math with R  
Working with Vectors  
Functions  
Comments and Code Structure  
Using Packages

## 4 - Vectors

Vector Properties  
Creating, Combining, and Iterating  
Passing and Returning Vectors in Functions  
Logical Vectors

## 5 - Reading and Writing

Text Manipulation  
Factors

## 6 - Dates

Working with Dates  
Date Formats and formatting  
Time Manipulation and Operations

## 7 - Multiple Dimensions

Adding a second dimension  
Indices and named rows and columns in a Matrix  
Matrix calculation  
n-Dimensional Arrays  
Data Frames  
Lists

## 8 - R in Data Science

AI Grouping Theory  
K-means  
Linear Regression  
Logistic Regression  
Elastic Net

## 9 - R with MadLib

Importing and Exporting static Data (CSV, Excel)  
Using Libraries with CRAN  
K-means with Madlib  
Regression with Madlib  
Other libraries

## 10 - Data Visualization

Powerful Data through Visualization: Communicating the Message  
Techniques in Data Visualization  
Data Visualization Tools  
Examples

## 11 - Databases, Data lakes & Additional Topics

Building connections to Databases and Data lakes, for both Python and R (using Hive server)  
Methods to “query” data from database and data lakes, for both Python and R  
Creating and passing macro variables. Specifically, R sprint, paste, paste0, and paste3 (not sure of the equivalent in Python).

## 12 - R with Hadoop

Overview of Hadoop  
Overview of Distributed Databases  
Overview of Pig  
Overview of Mahout  
Exploiting Hadoop clusters with R  
Hadoop, Mahout, and R

## 13 - Business Rule Systems

Rule Systems in the Enterprise  
Enterprise Service Busses  
Drools & Using R with Drools

## 14 - R with AWS

Best practices for working with AWS (completely outside of R and Python)