

R Programming for Data Scientists

Duration: 3 Days

Description

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.

Audience

- This is an intermediate and beyond level course, geared for experienced Data Analyst and Data Scientists who need to learn the details beyond the essentials of how to program in R. Incoming students should have prior data analytics background, and should have experience working with Excel and should know the basics of SQL.

Prerequisites

Students should have attended the course(s) below, or should have basic skills in these areas:

- Introduction to SQL
- Working with Excel

Outline

From Excel to R

- Common problems with Excel
- The R Environment
- Hello, R

R Basics

- Simple Math with R
- Working with Vectors
- Functions
- Comments and Code Structure

- Using Packages

Vectors

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

Reading and Writing

- Text Manipulation
- Factors

Dates

- Working with Dates
- Date Formats and formatting
- Time Manipulation and Operations

Multiple Dimensions

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

R in Data Science

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

R with MadLib

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

Data Visualization

- Powerful Data Through Visualization: Communicating the Message
- Techniques in Data Visualization
- Data Visualization Tools
- Examples

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

Business Rule Systems

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