

**Introduction to R Programming**

**Course Number:** RPROG-100
**Duration:** 4 days

**Overview**

Accelebrate's Introduction to R Programming training course teaches attendees how to use R programming to explore data from a variety of sources by building inferential models and generating charts, graphs, and other data representations.

**Prerequisites**

Students should have knowledge of basic statistics (t-test, chi-square-test, regression) and know the difference between descriptive and inferential statistics. No programming experience is needed.

**Materials**

All attendees receive comprehensive courseware and a textbook.

**Software Needed on Each Student PC**

* A recent release of R 4.x
* IDE or text editor of your choice (RStudio recommended)

**Objectives**

* Master the use of the R and RStudio interactive environment
* Expand R by installing R packages
* Explore and understand how to use the R documentation
* Read Structured Data into R from various sources
* Understand the different data types in R
* Understand the different data structures in R
* Understand how to create and manipulate dates in R
* Use the tidyverse collection of packages to manipulate dataframes
* Write user-defined R functions
* Use control statements
* Write Loop constructs in R
* Use the apply family of functions to iterate functions across data
* Expand iteration and programming through the Purrr package
* Reshape data from long to wide and back to support different analyses
* Perform merge operations with R
* Understand split-apply-combine (group-wise operations) in R
* Identify and deal with missing data
* Manipulate strings in R
* Understand basic regular expressions in R
* Understand base R graphics
* Focus on GGplot2 graphics for R for generating charts
* Use RMarkdown to programmatically generate reproducible reports
* Use R for descriptive statistics
* Use R for inferential statistics
* Write multivariate models in R (general linear models)
* Understand confounding and adjustment in multivariate models
* Understand interaction in multivariate models
* Predict/Score new data using models
* Understand basic non-linear functions in models
* Understand how to link data, statistical methods, and actionable questions

**Outline**

* Overview
	+ History of R
	+ Advantages and disadvantages
	+ Downloading and installing
	+ How to find documentation
* Introduction
	+ Using the R console and RStudio
	+ Getting help
	+ Learning about the environment
	+ Writing and executing scripts
	+ Object oriented programming
	+ Introduction to vectorized calculations
	+ Introduction to data frames
	+ Installing and loading packages
	+ Working directory
	+ Saving your work
* Variable types and data structures in base R
	+ Variables and assignment
	+ Data types
		- Numeric, character, boolean, and factors
	+ Data structures
		- Vectors, matrices, arrays, dataframes, lists
	+ Indexing, subsetting
	+ Assigning new values
	+ Viewing data and summaries
	+ Naming conventions
	+ Objects
* Getting data into the R environment with readr
	+ Built-in data
	+ Reading data from structured text files
	+ Reading data using ODBC
* Dataframe manipulation with dplyr
	+ Introduction to tibbles, enhanced data frames
	+ Renaming columns
	+ Adding new columns
	+ Binning data (continuous to categorical)
	+ Combining categorical values
	+ Transforming variables
	+ Handling missing data
	+ Merging datasets together
	+ Stacking datasets together (concatenation)
* Handling dates in R using lubridate
	+ Date and date-time classes in R
	+ Formatting dates for modeling
* Exploratory data analysis (descriptive statistics)
	+ Continuous data
		- Distributions
		- Quantiles, mean
		- Bi-modal distributions
		- Histograms, box-plots
	+ Categorical data
		- Tables
		- Barplots
	+ Group by calculations with dplyr
		- Split-apply-combine
	+ Reshaping and pivoting data in R (long to wide with aggregation)
		- pivot\_wider and \_longer with tidyr
* Working with text data
	+ Finding and matching patterns in text
	+ Stringr package for text manipulation
	+ Introduction to regular expressions in R
	+ Categorical data wrangling with forcats
* Control flow
	+ Truth testing
	+ Branching
	+ Looping
* Functions in depth
	+ Parameters
	+ Return values
	+ Variable scope
	+ Exception handling
* Applying functions across dimensions
	+ Sapply, lapply, apply
	+ Programming with map and purrr
* Graphics in R Overview
	+ Base graphics system in R
	+ Scatterplots, histograms, barcharts, box and whiskers, dotplots
	+ Labels, legends, titles, axes
	+ Exporting graphics to different formats
* Advanced R graphics: ggplot2
	+ Understanding the grammar of graphics
	+ Quick plots (qplot function)
	+ Building graphics by pieces (ggplot function)
	+ Understanding geoms (geometries)
	+ Linking chart elements to variable values
	+ Controlling legends and axes
	+ Exporting graphics
* Inferential Statistics
	+ Bivariate correlation
	+ T-test and non-parametric equivalents
	+ Chi-squared test
* General Linear Regression Models in R
	+ Understanding formulas
	+ Linear and logistic regression models
	+ Regression plots
	+ Confounding / interaction in regression
	+ Evaluating residuals
	+ Scoring new data from models (prediction)
	+ Useful plots from regression models
* Conclusion