## **Course duration**

• 3 days

## **Course Benefits**

- Learn the basics of R and Rstudio.
- Import and manipulate tabular data with R.
- Conduct exploratory analysis.
- Generate rich graphics with GGPlot2.
- Test for group differences using inferential methods.
- Build statistical regression models using R.

## **Course Outline**

- 1. Course Overview
  - 1. Why R? Advantages and disadvantages
  - 2. Downloading and installing
  - 3. How to find documentation
- 2. Introduction
  - 1. Using the R console/RStudio
  - 2. Getting help
  - 3. Learning about the environment
  - 4. Writing and executing scripts
  - 5. Introduction to vectorized calculations
  - 6. Introduction to data frames
  - 7. Installing packages
  - 8. Working directory
  - 9. Saving your work
- 3. Variable types and data structures
  - 1. Variables and assignment
  - 2. Data types
  - 3. Numeric, character, boolean, and factors
  - 4. Data structures
  - 5. Vectors, matrices, arrays, dataframes, lists
  - 6. Indexing, subsetting
  - 7. Assigning new values
  - 8. Viewing data and summaries
  - 9. Naming conventions
  - 10. Objects
- 4. Manipulating Data with R
  - 1. Getting data into the R environment and understanding dataframes

- 2. Built-in data
- 3. Overview of dataframes
- 4. Reading data from structured text files
- 5. Reading data using ODBC
- 6. Dataframe manipulation with dplyr
- 7. Renaming columns
- 8. Adding new columns
- 9. Managing data types
- 10. Binning data (continuous to categorical)
- 11. Combining categorical values
- 12. Transforming variables
- 13. Handling missing data
- 14. Long to wide and back
- 15. Merging datasets together
- 16. Stacking datasets together (concatenation)
- 17. Handling dates in R
- 18. Date and date-time classes in R
- 19. Formatting dates for modeling
- 5. Exploratory data analysis (descriptive statistics) including base graphics
  - 1. Continuous data
  - 2. Distributions
  - 3. Quantiles, mean
  - 4. Bi-modal distributions
  - 5. Histograms, box-plots
  - 6. Categorical data
  - 7. Tables
  - 8. Barplots
  - 9. Group by calculations with dplyr
  - 10. Split-apply-combine
  - 11. Long to wide and back, tidy data structures
- 6. Advanced graphics in R: using GGPlot
  - 1. Understanding the grammar of graphics
  - 2. Quick plots (qplot function)
  - 3. Building graphics by pieces (ggplot function)
  - 4. Understanding geoms (geometries)
  - 5. Linking chart elements to variable values
  - 6. Controlling legends and axes
  - 7. Exporting graphics
- 7. Testing for Group differences
  - 1. Traditional Inferential Statistics, A/B testing
  - 2. Null hypothesis testing and p-values
  - 3. Comparing Groups
  - 4. P-Values, summary statistics, sufficient statistics, inferential targets
  - 5. T-Tests (equal and unequal variances)
  - 6. ANOVA
  - 7. Chi-Square Tests
  - 8. Correlation

- 8. Modeling with R
  - 1. Frequentist Approaches to multivariable Statistics:
  - 2. Linear Regression
  - 3. Multivariate linear regression
  - 4. Capturing Non-linear Relationships
  - 5. Comparing Model Fits
  - 6. Scoring new data

## **Class Materials**

Each student will receive a comprehensive set of materials, including course notes and all the class examples.

**Class Prerequisites** 

Experience in the following *is required* for this R Programming class:

• Basic programming background.