

# Data Analytics Using R

By Dr. Viswanath Rao

## Overview

- History of R
- Advantages and disadvantages
- Downloading and installing
- How to find documentation

## Introduction

- Using the R console
- Getting help
- Learning about the environment
- Writing and executing scripts
- Object oriented programming
- Introduction to vectorized calculations
- Introduction to data frames
- Installing packages
- Working directory
- Saving your work

## Variable types and data structures

- 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

- Built-in data
- Reading data from structured text files
- Reading data using ODBC

## Dataframe manipulation with dplyr

- Renaming columns
- Adding new columns
- Binning data (continuous to categorical)
- Combining categorical values
- Transforming variables
- Handling missing data
- Long to wide and back
- Merging datasets together
- Stacking datasets together (concatenation)

#### Handling dates in R

- Date and date-time classes in R
- Formatting dates for modeling

#### Control flow

- Truth testing
- Branching
- Looping

#### Functions in depth

- Parameters
- Return values
- Variable scope
- Exception handling

#### Applying functions across dimensions

- Sapply, lapply, apply

#### 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
- Melting and casting data

#### Inferential statistics

- Bivariate correlation

- T-test and non-parametric equivalents
- Chi-squared test

#### Base graphics

- 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)

#### General linear regression

- Linear and logistic models
- Regression plots
- Confounding / interaction in regression
- Scoring new data from models (prediction)

#### Conclusion