

## R Beginner: Data Prep & Descriptive Analytics



**Length:** 3 days / 6 half days

**Overview:** This course teaches students the basics of programming with R by focusing on reading in, manipulating, and visualizing data. It is for students who are new to programming with R or have limited experience with the language. The course will cover the basics of R, including installation, data structures, and essential functions. Students will also explore data manipulation techniques and perform descriptive analytics to derive insights from their datasets. By the end of the course, participants will be equipped to confidently analyze data and create informative visualizations, laying a strong foundation for further exploration in data science.

**Objectives:**

- Understand the fundamentals of R programming and its applications in data analysis.
- Set up and navigate the R and RStudio environments effectively.
- Create and manipulate various data structures, including vectors, matrices, and data frames.
- Import and export data from different file formats and databases using R.
- Perform basic calculations and control statements to analyze data.
- Utilize the dplyr package for data manipulation, including filtering, summarizing, and reshaping datasets.
- Conduct exploratory data analysis and compute descriptive statistics to summarize data.
- Create visualizations using ggplot and base R to effectively communicate insights.

### Intro to R and Data Prep

#### Overview of R

- History
- Uses
- Advantages/Disadvantages of R
- Download and installation of R and RStudio

#### Introduction to RStudio

- Console
- Scripts
- Environment
- Changing RStudio options
- Installing packages
- Setting/changing directories
- Setting up projects
- Save work

### Variables and Data Structures

- Creating variables
  - Numeric
  - Strings
  - Dates
  - Factor
  - Booleans
- Different data structures
  - Vector
  - Matrices
  - Data Frame
  - List
  - Tibble
  - Array?
  - Function?
- Overwriting/copying variables
- Index/subsetting data structures

### **Read in Data to R**

- Use of readr package
- Base R approach to reading in files
  - csv files
  - .txt files
  - Any delimiter file
- Read in data from database??
- Data from packages

### **Save a File From R**

### **Calculations in R**

- Simple calculator
- Vector calculations
- Matrix multiplication

### **Control Statements**

- If else
- Ifelse

### **Loops**

- For
- While

### **Creation of Functions**

- Parameters
- Arguments
- Return values
- Error/exception control

### **Iteration with purrr/apply**

### **Data Manipulation with dplyr**

- Rename columns
- Create new columns
- Create discrete column from continuous
- Filtering data
- Ordering data
- Group and summarize data
- Create variables with dates/extract information from dates

- Working/manipulating factors
- Create variables with strings
  - Matching patterns
  - Basic regular expressions
  - Creating string variables

### **Reshaping Data**

- Concatenate datasets
- Merging datasets
- Wide to long
- Long to wide

### **Descriptive Statistics**

### **Exploratory Data Analysis**

- Descriptive stats
  - Summary
  - Mean
  - Median
  - Mode
  - 5 Number Summary
  - Skewness?
  - Tables
  - Correlation
  - Probabilities with distributions
- Exploratory Plots (ggplot and baseR)
  - Correlation
  - Histograms
  - Density Plots
  - Bargraphs
  - Scatterplot
  - Box plots
  - Line graph
- Improving Plots
  - Legends
  - Color
  - Manipulate axis
  - Combine plots

## **Prerequisites**

None

## **Materials**

All students will receive slides with lecture material, data and labs.

## **Software needed on each student PC:**

- Microsoft Excel 2010 or later (2013 or later recommended)
- Internet access
- IR (installed and fully operational)
- RStudio (installed and fully operational)

## **Related data and lab files will be provided**

**Other R Courses Available:** R Intermediate: Diagnostic Analytics;  
R Advanced: Predictive Analytics