BeVera

COURSE OUTLINE

EDA II – Data Wrangling

Course Overview

The EDA II course is an introduction to data wrangling which is the process of getting data into a useful form for visualization and modeling. Data Analysis can be divided into three parts: Extraction, Transform, and Visualize. During this course, attendees will learn to visualize and summarize their data in preparation for more advanced analysis.

Intended Audience

This Exploratory Data Analysis in R course intended for individuals, data analysis, data scientists and researchers that have a working knowledge of the R programming language.

Prerequisites

Attendees must download the R and the R Studio IDE onto their personal laptops.

Learning Objective

- Learn to work with the most useful data manipulation package in R
- Learn the basic vocabulary of dplyr
- Determining important features and patterns in data and unusual nonconformities
- Learn to invoke basic EDA methods to understand datasets in preparation for more advanced analysis
- Discuss what the data tells us? (In contrast to "confirmatory" data analysis)
- Learn 5-Number summary, Box plots, Stem and Leaf Diagrams
- Learn Extremes (min and max), Median & quartiles
- Identify more robust to skewed & long tailed distributions