

# CourseOutline

Shonte

5/1/2019

Course Title: Data Validation in R

## Chapter 1: Introduction to Data Validation

- Lesson 1.1: Why is Clean Data Important?
  - A Learning Objective: Understand the importance of clean data
- Lesson 1.2: Working with dirty data
  - A Learning Objective: Summarize a dataset and identify the more obvious data quality errors causing friction in using the data for analytics.
- Lesson 1.3: Introduction to `validate`
  - A Learning Objective: Learn the basic concepts of the `validate` package including an overview of functions to evaluate data

## Chapter 2: Setting Great Expectations for Quality Data

- Lesson 2.1: Setting Basic Expectations to Validate
  - A Learning Objective: Create a basic rule set in order to validate individual columns, variable types (eg., numeric variable) and number of rows in the dataset.
- Lesson 2.2: Evaluating Expectations using `check_that`
  - A Learning Objective: Determine whether a basic rule set written in previous exercises meets the expectation using `check_that`
- Lesson 2.3: Validating Multiple Columns
  - A Learning Objective: Combine complex rules that apply to multiple columns using boolean operators and inequalities in a validation object.
- Lesson 2.4: Validating Aggregates (Sample Exercise Based on this Lesson)
  - A Learning Objective: Validate aggregates with arithmetic operations (mean/median) for checking against the data.

## Chapter 3: Creating Reproducible Rule Sets

- Lesson 3.1: Introducing the `validator` and `Indicator` nouns in more detail
  - A Learning Objective: Learn the primary purpose for creating a validator object.
- Lesson 3.2: Confronting a Dataset using `Validator`
  - A Learning Objective: Summarize and visualize the results after evaluating a validation rule set against data.
- Lesson 3.3: Confronting a Dataset using `Indicator`
  - A Learning Objective: Store indicator expressions and confront a dataset with these indicators, get values, confront, and summarise results

## Chapter 4: Defining a Validation Workflow for Reproducibility and Regular Monitoring

- Lesson 4.1: Bringing it all together

- A Learning Objective: Learn the structure of a basic workflow for evaluating data
- Lesson 4.2: Converting Validators/Indicators to a Data Frame.
  - A Learning Objective: Learn how to store validators / indicators as dataframes. Concepts of workflow are reinforced with a rule set already defined for students to step through proces.
- Lesson 4.3: Merging Results to a Data Frame
  - A Learning Objective: Learn how to work through the workflow and merge results back to data for monitoring
    - Challenge Exercise Idea: Students are given a new data set and must follow the basic workflow plan of defining rules, confront with data, analyze results and merging results back to data.