Introduction to Data Science

Daniel Gutierrez, Data Scientist

Los Angeles, Calif.

Course Outcomes

- Introduction to R Programming Part II

Lesson Objectives

- Learning R language constructs for manipulating data

- These techniques will be useful for data munging

- Extracting parts of vectors, matrices, lists (subsetting)

- Managing NA values found in data sets

R Fundamentals – Part 2

- A very important part of the R language is called "subsetting" or the ability to extract parts of an object

- Subset a vector

- Subset a matrix

- Subset a list

- The `subset()` function

R Fundamentals – Part 2

- Handling missing values (NAs) is very important in R

- Removing NA values

- NA versus NULL versus NaN versus empty string " "

- NA and NULL in statistical functions

- NAs with the `subset()` function

Code module

- WEEK 2-1 Code module – subsetting vectors and matrices
- WEEK 2-2 Code module – subsetting lists
- WEEK 2-3 Code module – more subsetting lists
- WEEK 2-4 Code module – `subset()` function
- WEEK 2-5 Code module – removing NA values
- WEEK 2-6 Code module – NA vs. NULL vs. NaN vs. empty string
- WEEK 2-7 Code module – NAs in statistical functions

Summary

- In WEEK 2 we continued building our data science toolbox by adding a number of tools used for data munging:

- Subsetting

- Missing value handling