George Tryfiates

10/6/2020

R Intro

1. Open source lang for stats and graphics
   1. Data handling and storage
   2. Op for matrices
   3. Intermediate data analysis
   4. Graphic for data analysis
   5. ML
2. Who uses R?
   1. Academic but increasingly in industry
   2. FED uses R
3. Why?
   1. Pro
      1. Most used in Research (Py most used in Industry)
      2. Created w/ stats and data in mind
      3. Open source
      4. Wide range of high quality packages
      5. Great for learning ML
      6. Easy to use and understand
   2. Con
      1. Performance
      2. Domain specificity
      3. Few unusual quirks
4. R Syntax
   1. Comment
   2. Multiline Hack:
      1. If(FALSE){“multiline comment”}
   3. Dtypes
      1. Vectors
         1. Apple = c(‘red’, ‘green’)
         2. Atomic
            1. Int
            2. Numeric
            3. Complex
            4. Raw
            5. Logical
            6. One more
      2. Lists
         1. List(c())
         2. List(lists, vectors, matrix…)
         3. One-based indexing (first index is one, not zero)
      3. Matrices
         1. 2D data set
      4. Arrays
         1. Matrix of any dimension
         2. Array(c(‘geen’, ‘yellow’), dim = c(3,3,2)
      5. Factors
         1. Created using a vector
         2. Creating a class basically
      6. Data Frame
         1. Ea. Col can contain different dtypes
         2. Table or 2D array struct
         3. Vectors of different length
         4. Data.frame()
   4. Variable = objects
      1. Assignment
         1. =
         2. ->
         3. <-
      2. Printing
         1. Print()
         2. Cat() continuous
      3. Finding
         1. Print(ls()) : prints all variables in env
         2. Print(ls(pattern = “var”)) : pattern matching
   5. Operators
      1. If
      2. If else
      3. Switch
   6. Loop / Decision Making
      1. Repeat
      2. While
      3. For
      4. Break / next
   7. Fx
      1. Seq(start, end) : returns sequence of num from start to end