Introduction to R

Outline

- Introduction to R

- Introduction to RStudio

- R objects

  1. Variables and Function Calls

  2. Data Types

  3. Vectors

  4. Data Frames

- Control statements

  1. Conditionals

  2. Loops

Why is R so popular in statistics, data science, data analytic fields?

R is a scripting language for statistical data manipulation and analysis, widely used in both academia and industry.

Advantages of R:

- Public-domain and free
- Available for the Windows, Mac, and Linux operating systems
- Superior to provide statistical operations and modeling
- Powerful in data visualization
- Save data sets and command history

RStudio
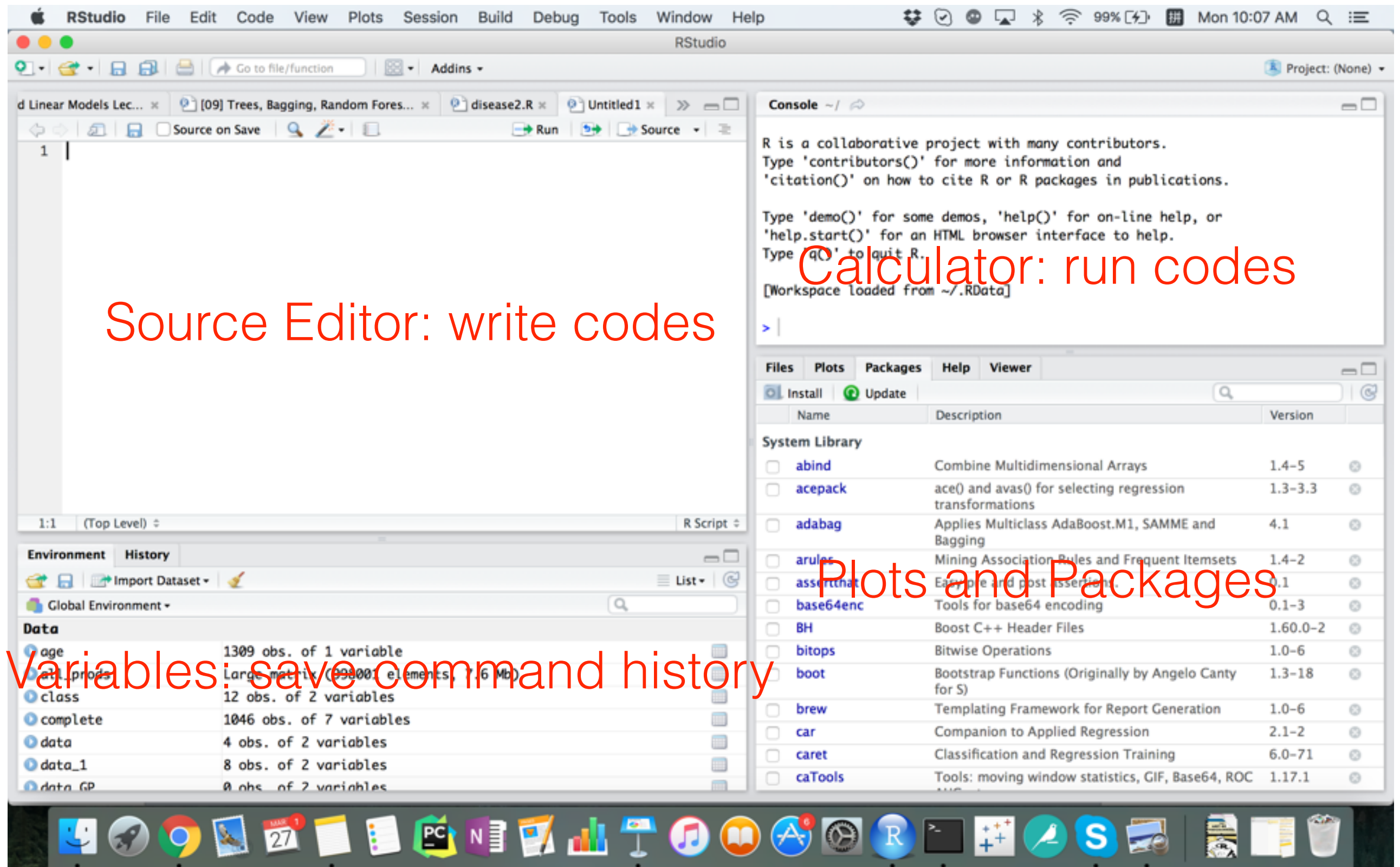
Variables and Function Calls

A variable: a storage location to save a value
Assign a value to a variable: either "=" or "<-"

Variable names rules:
• Letters, i.g. a,B
• Letters+Numbers, i.g. x1, y2
• Letters+_+letters/numbers, i.g. x_a, x_1
• Can not start with number

Functions takes argument and return values, like you write any expression
• Math, i.g. +(addition),-(subtraction),*(multiplication),/(division),%/% (quotient), %%(reminder), ^(power)
• Statistical analysis, i.g. mean, standard deviation

Data types

Primitive data types:

- numeric: number

- integer: integer

- character: string

- logical: TRUE/FALSE

- factor: numeric/string

Change data types:

- from numeric/integer to character:

  as. character()

- from character to numeric/integer

  as.numeric()/ as.integer()

Vectors

Vector: a sequence of primitive data values of the same data type

Create a vector:
- c(): created a vector
- seq(): a sequence of integers
- rnorm(): pick random values from a normal distribution with mean equal to 0 and standard deviation equal to 1
- runif(): pick random values from 0 to 1
- sample(): random sample
- rep(): a set of repeated values/vectors

Slice a vector:
- a single element based on its index
- a vector based on several indexes
- negative index: delete several indexes
- conditional index

Modify a vector:
- re-assignment a value to a element based on its index

Vector computation:
- math operations: +, -, *, /
- aggregation functions:sum, maximum, minimum, mean, standard deviation

Data Frames

Data frames: database tables, include columns and rows

Create a data frame: combination of vectors by column

add a new column

Explore a data frame:
- colnames()
- rownames()
- dim() : dimension
- str(): data type of each column
- summary(): statistical aggregation of numeric columns

Data Frames

Selecting by row/column index:
- By row
- By column
- By row/column

Selecting by column name

Slicing:
- One condition
- Multiple conditions

Print the data:
- head()
- tail()
- View()

Data frame manipulation:
order(): sort by one column

Exporting:
write.csv(): save at the current working directory

Importing
read.csv(): read from the working directory of a folder that contains the file

Control Statements

Conditionals:
if/else loop
if (conditions){
statement1}
else{
statement2}

Loops: computer programming
- for loop
for (test_expressions)
{
    statement
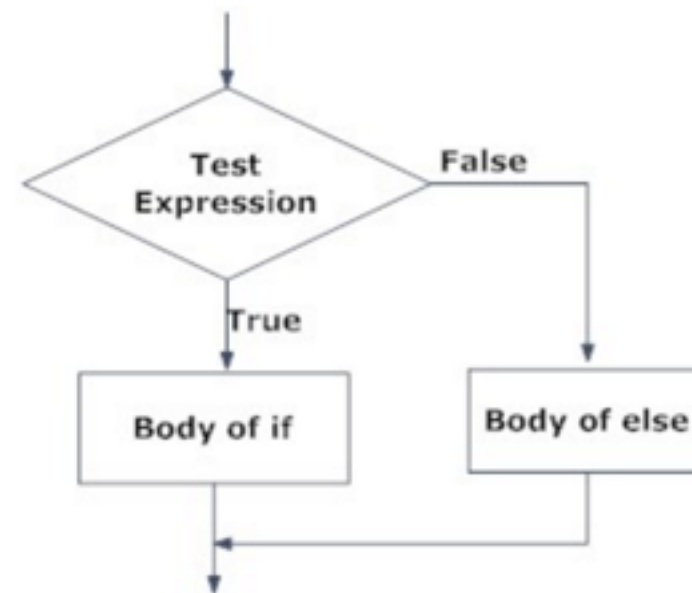}

- while loop
initial assignment
while (test_expressions)
{
    statement
}

Fig: Operation of if...else statement