b-it Bonn-Aachen International Center for Information Technology

universitätbonn

Introduction to R-Programming

Paurush Praveen
Algorithmic Bioinformatics
BIT Research School

praveen@bit.uni-bonn.de

Outline

- R as a new language
- R as a calculator
- Variables and Operators
- Data Structure
- Control Structure
- Functions
- String handle and regex

R-Introduction

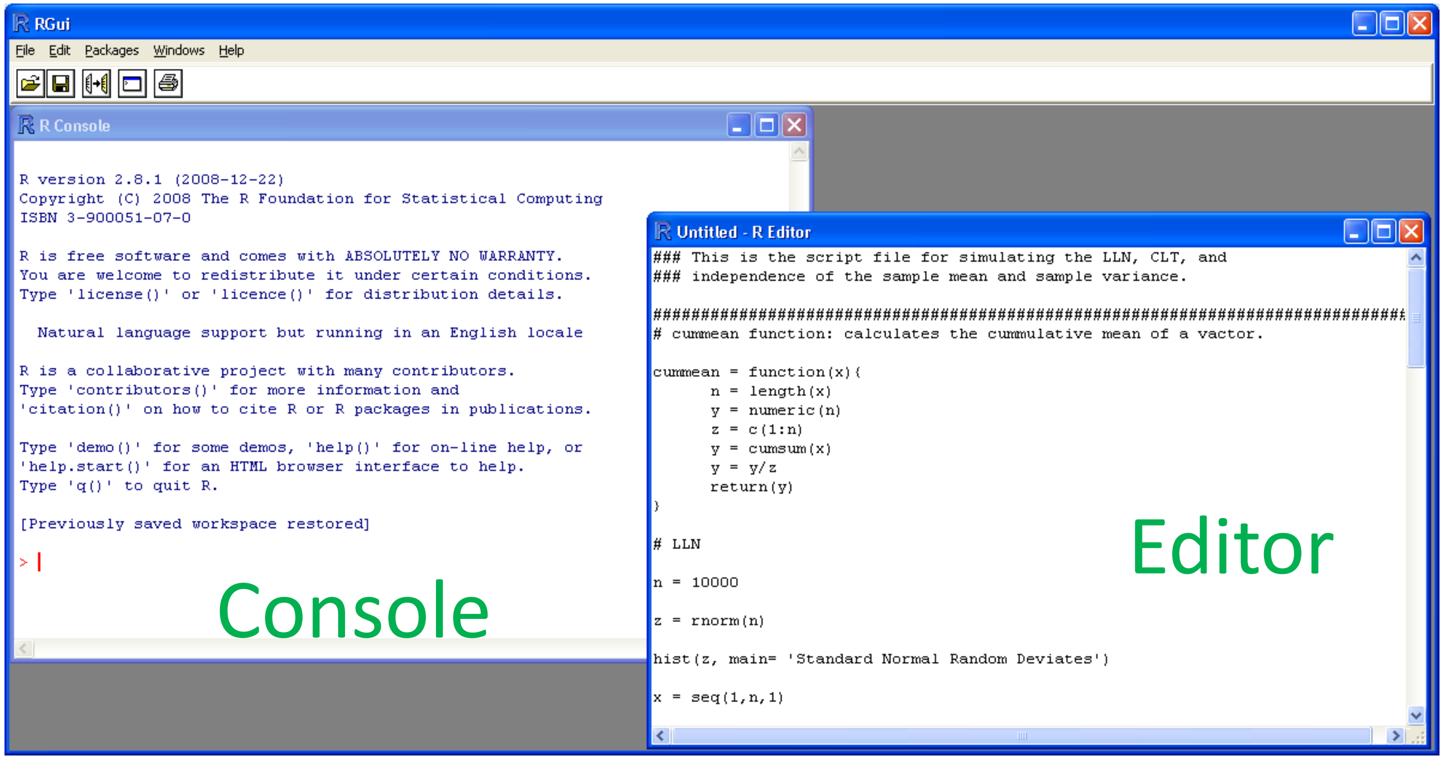
- A programming/scripting language widely used in statistical works
- An interpreted language, not a compiled one
- Free | Easy | Support | Compatibility
- Object oriented

How does it work

- variables, data, functions, results, etc, are stored in the active memory of the computer in the form of objects which have a name.
- The user can do actions on these objects with operators (arithmetic, logical, comparison, ...) and functions (which are themselves objects).
- The use of operators is relatively intuitive

How does it look

- command prompt symbol ">"



```
R version 2.8.1 (2008-12-22)
Copyright (C) 2008 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> |
```

```
### This is the script file for simulating the LLN, CLT, and
### independence of the sample mean and sample variance.

###############################################################################
# cummean function: calculates the cummulative mean of a vactor.

cummean = function(x){
      n = length(x)
      y = numeric(n)
      z = c(1:n)
      y = cumsum(x)
      y = y/z
      return(y)
}

# LLN

n = 10000

z = rnorm(n)

hist(z, main= 'Standard Normal Random Deviates')

x = seq(1,n,1)
```

ALGORITHMIC BIOINFORMATICS

Input / Output

- Interactive session
- Input at the console itself
- input a script
- source(“codefile”)
- Usually R gives output on the command line.
- To save as file use

```
sink("file.txt") <Not for graphical output>
##return to the normal mode
sink()
```

Running R-programs

- Save your commands in a file (*viz*. commands.R)
- Call R on the command line:

  R commands.R

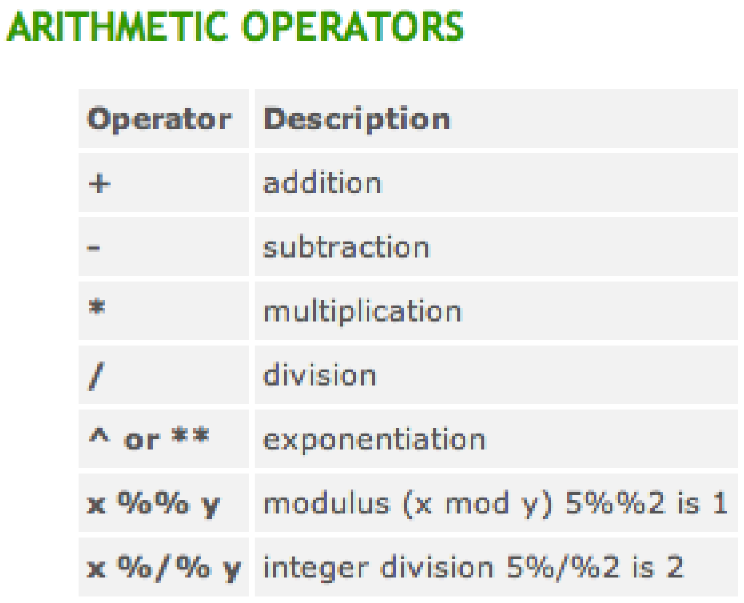
- Call the script from within R:

  > commands.R

R as a calculator

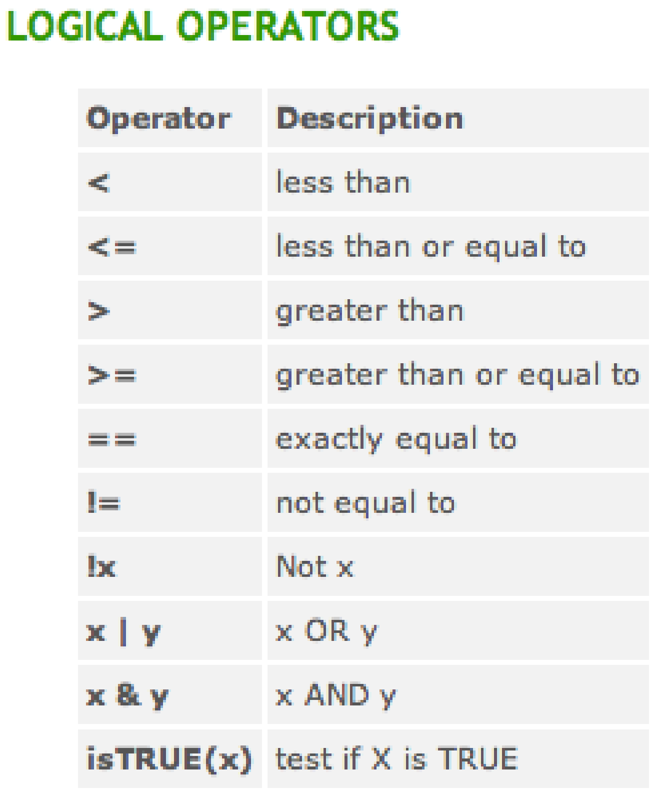
- + - * / work as usual
- pi etc. exist as Functions with brackets: sqrt(pi)
- a call withoutparetheses gives the sourcecode
- # denote comments
- assignment by <-
- library()
- data()

Operators



ARITHMETIC OPERATORS

Operator	Description
+	addition
-	subtraction
*	multiplication
/	division
^ or **	exponentiation
x %% y	modulus (x mod y) 5%%2 is 1
x %/% y	integer division 5%/%2 is 2



LOGICAL OPERATORS

Operator	Description
<	less than
<=	less than or equal to
>	greater than
>=	greater than or equal to
==	exactly equal to
!=	not equal to
!x	Not x
x | y	x OR y
x & y	x AND y
isTRUE(x)	test if X is TRUE

Binary operators work on vectors, matrices and scalars

Source: http://www.statmethods.net/management/operators.html

Data Types

- A vector contains an indexed set of values that are all of the same type:
    - logical
    - numeric
    - complex
    - Character

  – The numeric type can be further broken down into integer, single, and double types (but this is only important when making calls to foreign functions, eg. C or Fortran.)

Data Structures

- Data Structures
  - vector - arrays of the same type
  - factor - categorical
  - list - can contain objects of different types
  - matrix - table of numbers
  - data.frame - table of numbers and/or characters
  - environment
  - hashtable
  - function

Control Structures

for and *while* Loops

The syntax for writing for loops in R is

```
for(i in 1:N){
 Loop Code*
}
```

The syntax for while loops in R is

```
while(logical argument){
Loop Code*
}
```

```
Repeat {expression}
```

*where Loop Code should be substituted with the code you want to run in the loop.

Control structures

Conditional statements

- **if (condition) true_expression else false_expression**
- **if (condition) expression**
- **if(condition 1) { statement1**

```
if( condition 1) { statement1
     } else if(condition2) { statement2
          } else if(condition3) { statement3
     }
```

Writing Functions

- Writing R functions provides a means of adding new functionality to the language
- Functions that a user writes have the same status as those which are provided with R.
- Reading the functions provided with the R system is a good way to learn how to write functions.

Writing Functions

- You can declare your own function in R using the function() command. The syntax is

```
    myfunction() <- function(input1,input2,..)
  {
       Function code
  }
```

- Functions can be called with the arguments

```
    myfunction(argument1, argument2,...)
```

ALGORITHMIC BIOINFORMATICS

String Handle

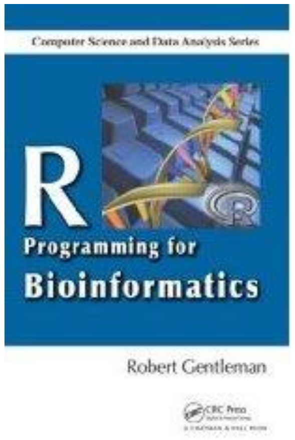
- Character strings are one of the basic types in R
- A character vector is a vector of character strings, not characters
- "" is an empty string
- Special characters -> a backslash followed by a symbol for the special character
  - **toupper(x) & tolower(x)** *: To convert case*
  - **chartr(old,new,x)** Replaces any occurrence in x
  - **sub and gsub** for substitution

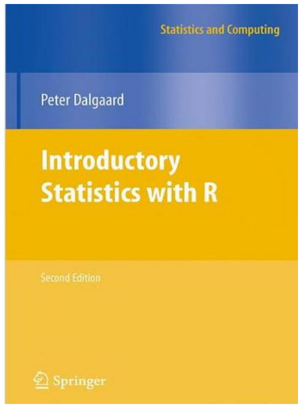
Regular Expression

- Regular Expression (Regex)
- Properties
  - Can be combined
  - Can be used with metacharacters to define the pattern
- Using regular expressions
  - **grep (pattern,x)**
  - **grep (pattern, x, value=TRUE)**

Further References

- Use **?<command>** in console to seek help or **help(<command>)**
- http://cran.r-project.org/doc/manuals/R-intro.html
- http://www.cyclismo.org/tutorial/R/
- Books


Computer Science and Data Analysis Series
R Programming for Bioinformatics
Robert Gentleman
CRC Press



- Web helps can be more useful
- Google to search for specific problems
- http://www.rseek.org/ is a search engine for R specific topics

Another Book "R in a Nutshell", Adler J, Oreilly