Genetic gain training and data management: Pre-workshop

Setup and Intro - R

LAGAT Abraham | 2023-08-30/31

Acknowledgement
HAMADZIRIPI Esnath
SNELLER Clay
MUSUNDIRE Lennin

CGIAR

CIMMYT

Content: 30-31 August, 2023

- Software Installation and troubleshoot
- Concepts
- Interface
- R Language
- Basic codes, Functions, Packages
- Resource links

Installations: Java -> R -> RStudio

1. Java

1. Check if java is already installed in your computer. Launch the terminal and issue the command *java –version*,
2. If not installed download [Java]
3. After downloading, right click and select run as administrator
4. Follow the prompts to install it in your application
5. After installation, open windows terminal and issue the command java --version to check again

2. R Software

1. Visit Cran-Project website
2. Click Download R-x.y.z for Windows and wait for it to complete downloading
3. Locate the downloaded software in your machine, right-click and install it on your machine (preferably as Administrator).
4. Follow the prompts to complete the installation

3. RStudio

1. Visit the posit website download page and click DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS and wait for the download to complete
2. Locate the downloaded software in in your machine, right-click and install it on your machine (preferably as Administrator).
3. Follow the prompts to complete the installation

R Concepts

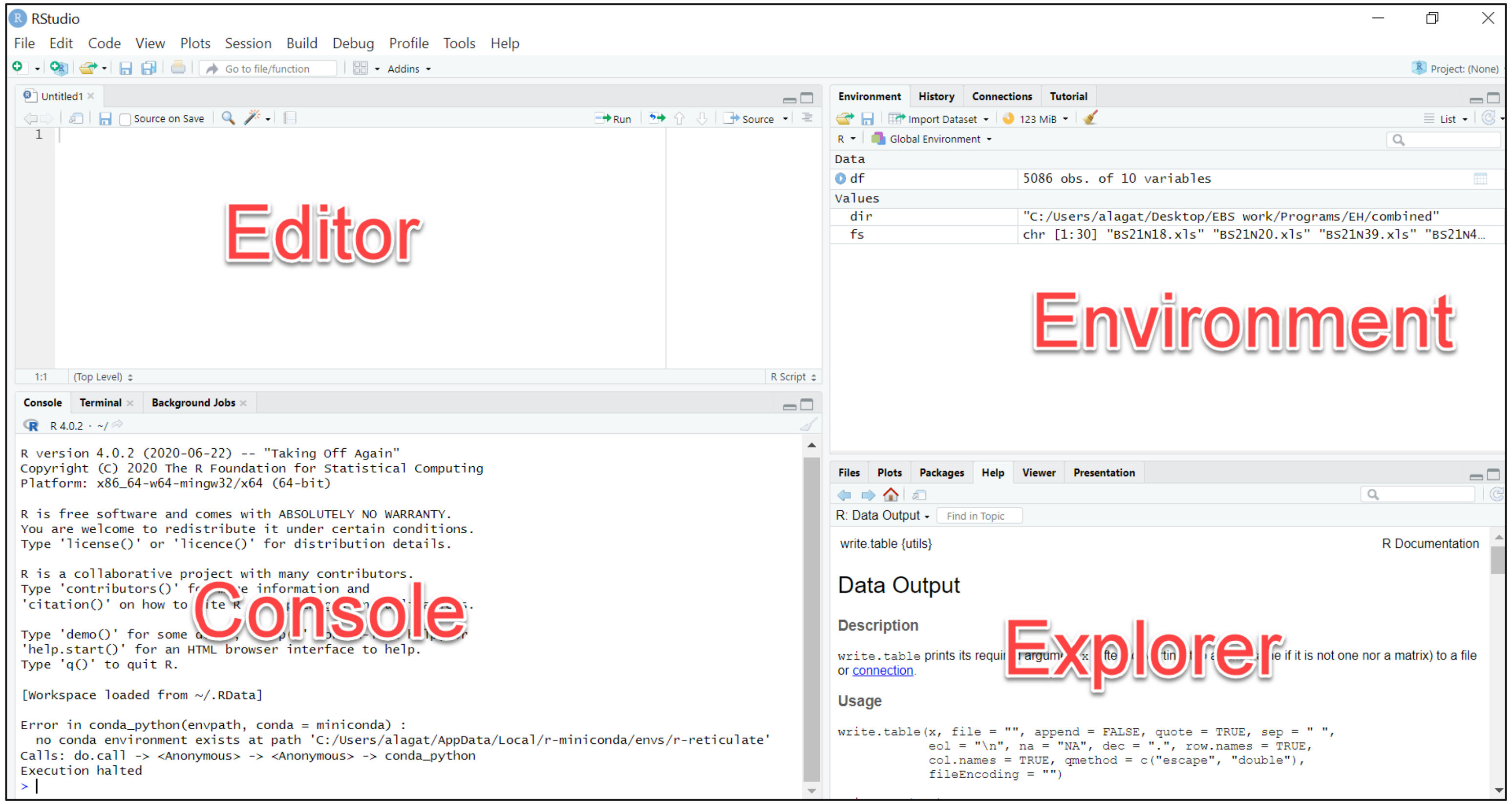
R is an **object-oriented** software.

Function – Instruction to do something. Can return results.
Parameters/arguments – inputs to a function
Package – compiled set of function(s). Installed packages become *libraries*.

Working directory vs environment vs project

RStudio – An integrated development environment (IDE)

RStudio Interface



Data

df	5086 obs. of 10 variables

Values

dir	"C:/Users/alagat/Desktop/EBS work/Programs/EH/combined"
fs	chr [1:30] "BS21N18.xls" "BS21N20.xls" "BS21N39.xls" "BS21N4…

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

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.

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.

[Workspace loaded from ~/.RData]

Error in conda_python(envpath, conda = miniconda) :
  no conda environment exists at path 'C:/Users/alagat/AppData/Local/r-miniconda/envs/r-reticulate'
Calls: do.call -> <Anonymous> -> <Anonymous> -> conda_python
Execution halted
> |
```

write.table {utils} R Documentation

Data Output

Description

write.table prints its required argument x (after converting it to a data frame if it is not one nor a matrix) to a file or connection.

Usage

```
write.table(x, file = "", append = FALSE, quote = TRUE, sep = " ",
            eol = "\n", na = "NA", dec = ".", row.names = TRUE,
            col.names = TRUE, qmethod = c("escape", "double"),
            fileEncoding = "")
```

1. Editor:

☑ Writing and running codes/scripts
☑ Show specific text-based files eg .txt, .r, .md, .rmd, etc
☑ View objects in the workspace

2. Console:

☑ Shows the codes run and results, any warnings and errors
☑ Codes can also be types directly on the console, useful for temporary short computations/calculator

3. Environment:

☑ Hold session information: objects, variables
☑ History of the codes run in the session
☑ Load/save environment

4. Explorer

☑ Files: Files and folders in the working directory
☑ Shows visualizations if any when running codes
☑ List of packages installed and loaded
☑ Provides support for using installed libraries and functions. Some packages have sample datasets.
☑ Setting working directory

Environment and Project

Working directory and environment

```
setwd("C:/Users/mimi/dataproject") # set working directory
getwd() # show working directory
ls() # show everything in the environment
remove(object) # remove the specified object in the environment also rm()
list.files() # list files and directories in the workspace
```

Project

Organized workspace for working on a project

Collaborative workspace

Configures it to be working directory

R Codes

- Instructions to R. Codes are case sensitive.
- Create objects by assignment: <-, ->, =
- Each instruction on its own line. Beware of \ symbol
- Anything within a line appearing after # symbol is omitted.
- Quotation marks " " or ' ' encloses literal strings. Back tick `   `

Operators

- Assignment operator: <-. Equivalent to =. Used for holding objects
- Arithmetic operators: +, -, /, *, ^
- Logical operators: ==, <, >, <=, >=, !, &, |

R Language

Data Types

- *integer*: whole numbers (integers). Represented with L eg -2L
- character: strings
- *logical*: Boolean values TRUE or FALSE (T or F)
- *factor*: categorize - default ascending alphanumeric
- *double/numeric*: decimal numbers
- *date*: time information

class(dataset)
as.character(dataset$age)

is.character(dataset$age)

Data objects

- Create new data in the workspace

```
a = 51;
b <- c(1, 5, 9)
```

- Loading data to workspace and held(assigned) to a variable

```
dataset <- read.csv("file.csv")
```

- Modify existing objects in the workspace

```
age = dataset$age
```

Q/A

Functions

```
# generate n number of fake ages between 0-120

fakeAges <- function(n = 1, yourname="mimi"){
        x = runif(n, 0, 120) # generate n values with range 0 to 120
        x = ceiling(x) # roundup to whole number
        return(x) # give the results of the function
}
```

Functions

- Repeatable, generalized set of codes instructing the program what to do
- Functions include a function name followed by brackets.
- Function can contain parameter inputs – added inside the bracket
- Most functions returns results

Distributed through: CRAN (19,198), Bioconductor, Github

https://github.com/lagvier/tempspace/blob/master/Introduction_codes.R

Basic codes

Data Access and explore

```
data<- read.csv("file.csv")

View(data)
str(data)
summary(data)

mean(data$age)
sd(data$age)
length(data$age)
```

Data Preparations

```
x = data$age
x = data[, 2]
ob1 = data[1, ]
obs_1 = data[-1, ]

x=subset(data, select=c("age"))
obs1 = subset(data, id == 1)

obs_1 = subset(data, id != 1)
obs_1 = subset(data, !is.na(age))
```

Analyze, Viz & Save

```
hist(data$age)
plot(data$age)

write.csv("output.csv")
```

Save images

Tips

- Help function help() or ? in a package: Check documentation of packages: input parameters, outputs objects, test data
- Read and add documentations in code with #
- Add blanks to lines to code blocks for readability
- Packages:

  Read excel: Can it read correctly? "readxl"
  Write Excel: Can it override sheets/data/format? "XLConnect"
  Data preparations: what preparations? "plyr", "dplyr", "reshape", "reshape2"
  Visualizations: what plot do I want? "ggplot2"
  Reports: can it author change? "officers"
- Function conflicts packagename::function()
- Be upto date with developments?

Packages

Distributed through: CRAN (19,198), Bioconductor, Github

- Install from CRAN: install.packages("readxl ", d = T)
- Load library to workspace: library("readxl")
- Remove from workspace: remove.package("readxl")
- Run function: read_excel()

Further resources

1. https://www.r-project.org/
2. https://www.stat.colostate.edu/~jah/talks_public_html/isec2020/installRStudio.html
3. R Beginner Videos:
   - https://www.youtube.com/watch?v=BvKETZ6kr9Q
   - https://www.youtube.com/watch?v=FY8BISK5DpM
   - https://www.youtube.com/watch?v=TQMAKGDIe_8 – RStudio interface
   - https://www.youtube.com/watch?v=Q5g6lYUn6Q4 – Mathematical functions
   - https://www.youtube.com/watch?v=HPJn1CMvtml – Data management and Viz

4. R Graphics
   - http://www.sthda.com/english/
   - https://r-graph-gallery.com
   - https://r-coder.com/r-graphs/

5. Codes equivalence with other programs - https://stats.oarc.ucla.edu/other/dae

Thank you for your interest!

Q/A