Statistical Analysis using R

IITA
Transforming African Agriculture

Introduction to R and RStudio

Ibnou Dieng
Kayode Fowobaje
Sam Ofodile
Moshood Bakare
Trushar Shah
Margaret Karanja

IITA Biometrics Unit

01-02 & 07-09 December 2021

Course overview

1. Short Introduction to R and RStudio
2. Preparation of Data for Statistical Analysis
3. Data wrangling
4. Experimental Designs for Plant Breeding
5. ANOVA and MET analysis
6. Multivariate analysis
7. Graphics in R with ggplot2

What's R

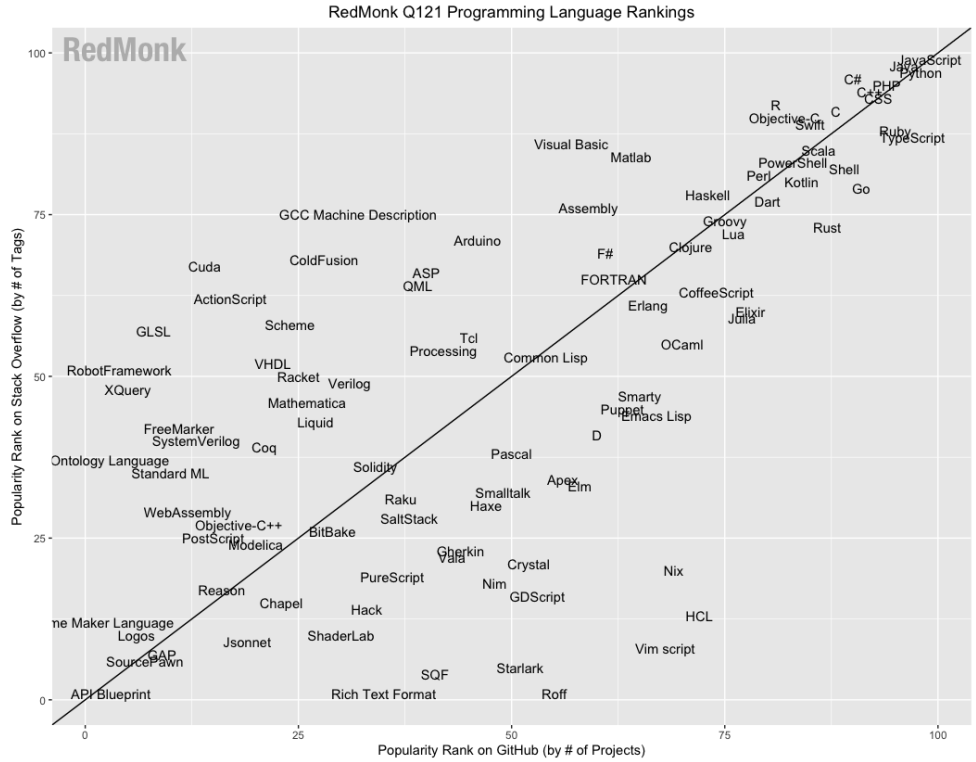
R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. See https://www.r-project.org/

R provides a wide variety of statistical and graphical techniques: linear and nonlinear modeling, statistical tests, time series analysis, classification, clustering, etc.

Many users make their data, functions and documentation freely available, grouped together in the form of packages.

What's R

R is one of the most popular language for data scientists — and it's been around for almost 20 years



https://redmonk.com/sogrady/2021/03/01/language-rankings-1-21/

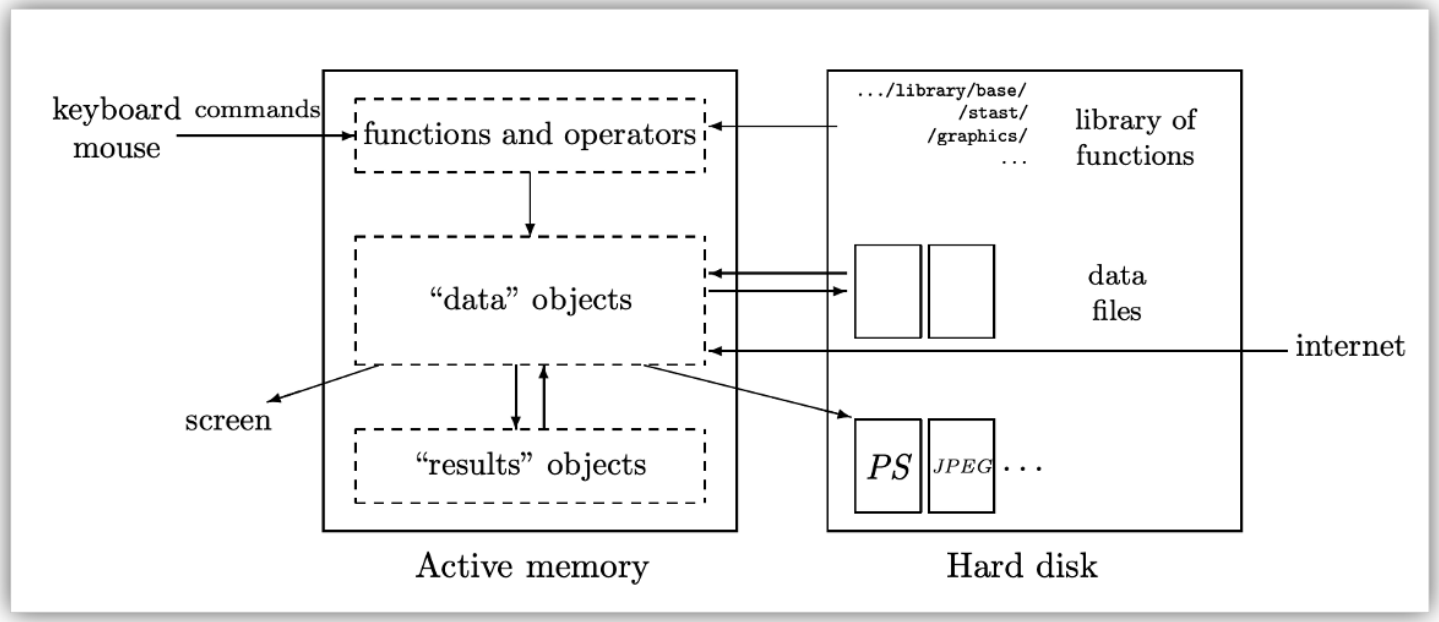
What's R

R key features:

- R provides robust facilities for data management, analysis and visualization
- R makes it easier to debug errors in code
- R facilitates complex operations with vectors, arrays, data frames, and other data objects of various sizes.
- R can integrate with other programming languages such as C, C ++, Python, Java, FORTRAN and JavaScript.
- R has extensive community support that provides technical assistance
- R is cross-platform compatible. R packages can be installed and used on any OS in any software environment without any modification
- R is at the cutting edge of new methodologies: as of May 2021, there were 19,000+ available packages, mostly stored in CRAN, the global repository of open-source packages that extend the capabilities of R
- **R is free**

What's R

- A classical software displays immediately the results of an analysis,
- R stores these results in an **“object”**, so that an analysis can be done with no result displayed



https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

What's RStudio

We are using R via RStudio.

R is like a car's **engine** while **RStudio** is like a car's **dashboard**

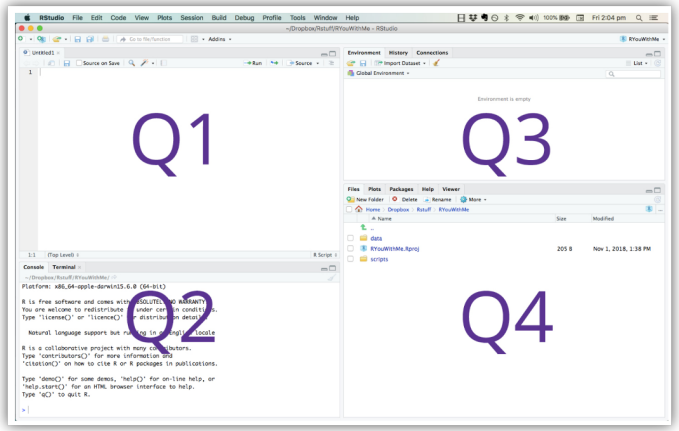
R: Engine	RStudio: Dashboard
	

https://moderndive.netlify.app/index.html

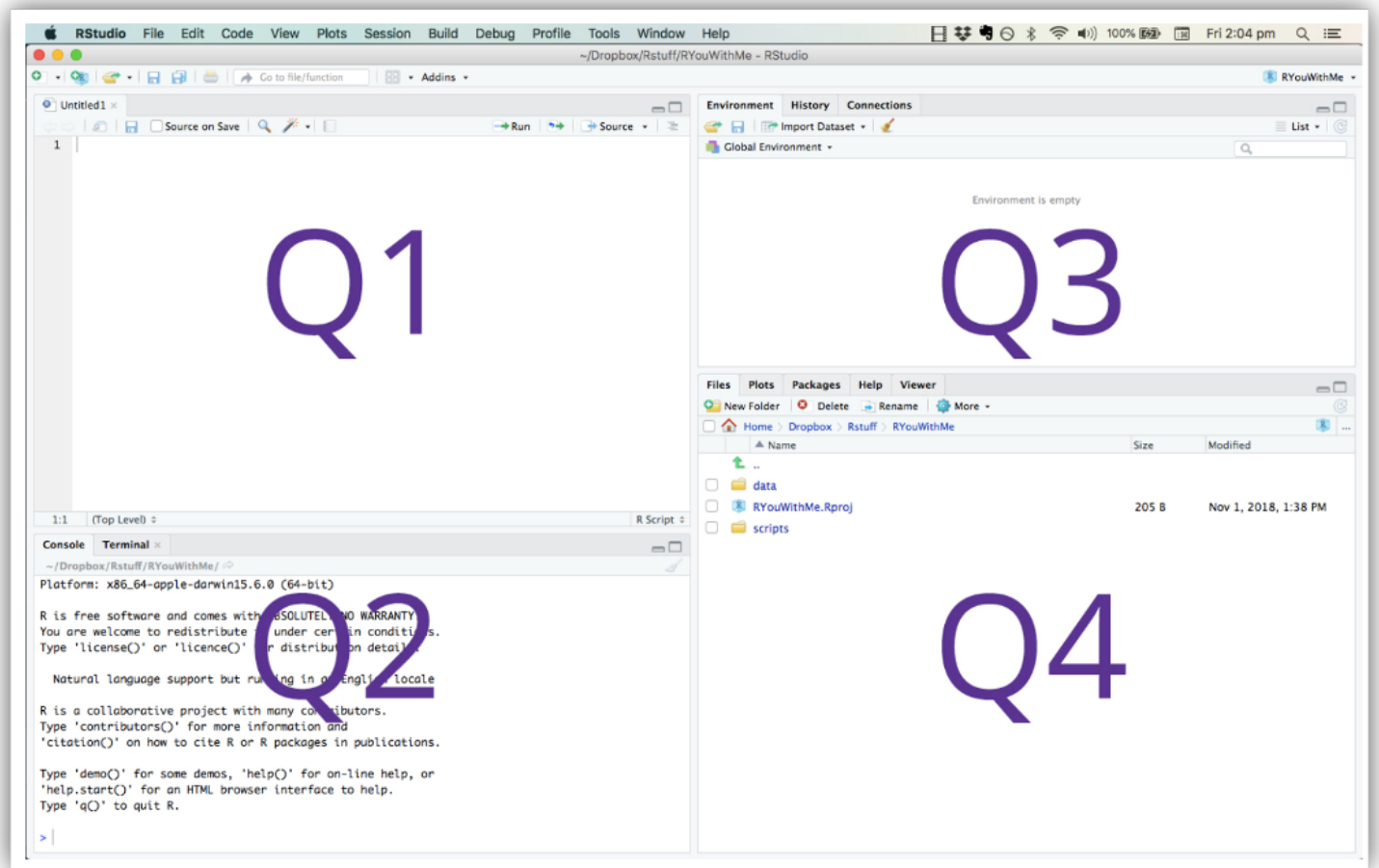
What's RStudio

There are four quadrants in **RStudio**

- **Q1**: script, data, command to run script
- **Q2**: console
- **Q3**: environment
- **Q4**: files, plots, packages, help



What's RStudio



R and RStudio together

R is a programming language that runs computations, while **RStudio** is an Integrated Development Environment (IDE) that provides an interface by adding many convenient features and tools.

- Install R from CRAN, the Comprehensive R Archive Network: https://cran.r-project.org/
- Install RStudio: https://www.rstudio.com/
- Install Rtools: https://cran.r-project.org/bin/windows/Rtools/
- If you already have R and/or RStudio installed, we highly recommend that you re-install both and get as current as possible
- For this course, we'll be using:
    - R version 4.1.2 -- "Bird Hippie"
    - RStudio Version 2021.09.1-372

Interactive mode

- R defaults to an interactive mode (Q2): a prompt > is presented to users
- Each input expression is evaluated… and a result returned

```
> # Simple arithmetic
> 8 + 5
```

```
## [1] 13
```

```
> # Operator precedence
> 2 + 3 * 5
```

```
## [1] 17
```

```
> # Exponentiation
> 3^2
```

```
## [1] 9
```

```
> # Basic mathematical functions
> sqrt(10)
```

```
## [1] 3.162278
```

```
> # Some constants are predefine
> pi
```

```
## [1] 3.141593
```

```
> # Circomference of earth at Eq
> 2 * pi * 6378
```

```
## [1] 40074.16
```

The R help

- The R help provides useful information on how to use functions.
- Help is available directly for a given function, for instance:
  - typing `?mean` or `help(mean)` or `help("mean")` in the console (Q2)
  - will display, the help page for the function `mean` in Q4

Selected Resources

- R for Data Science
- Advanced R online
- STAT 545
- Tidyverse
- ggplot2 quick reference