



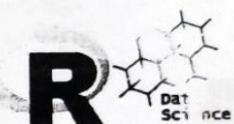
# R for Data Science

## Bài 1: *Tổng quan R*

Phòng LT & Mạng

[https://csc.edu.vn/lap-trinh-va-cSDL/R-Programming-Language-for-Data-Science\\_190](https://csc.edu.vn/lap-trinh-va-cSDL/R-Programming-Language-for-Data-Science_190)

2020



## Nội dung



1. Giới thiệu R
2. Môi trường làm việc
3. Tạo tập tin và thực thi



## Giới thiệu R

- R là một ngôn ngữ lập trình và là một môi trường dành cho tính toán thống kê, đồ họa máy tính, Data Science.
- R là một GNU project tương tự như S được phát triển tại Bell Laboratories (trước đây là AT&T, hiện tại là Lucent Technologies), do John Chambers và đồng nghiệp thực hiện
- R là một bộ phần mềm tích hợp với các thao tác xử lý dữ liệu, tính toán và hiển thị đồ họa.



R programming language for Data Science

3

## Giới thiệu R

- Lịch sử phát triển của R:

Phiên bản	Ngày công bố
0.49	1997-04-23
0.60	1997-12-05
0.65.1	1999-10-07
1.0	2000-02-29
1.4	2001-12-19
2.0	2004-10-04
2.1	2005-04-18
...	...
3.0	2013-04-03
...	...
3.4.3	2017-11-30
3.5.1	2018-07-02
...	...
3.6.3	2020-02-29



Nếu cần báo 3.5.1

4

## Giới thiệu R

### □ Tóm tắt về R

<b>Designed by</b>	Ross Ihaka and Robert Gentleman
<b>Developer</b>	R Core Team <sup>[1]</sup>
<b>First appeared</b>	August 1993; 26 years ago <sup>[2]</sup>
<b>Stable release</b>	3.6.3 ('Holding the Windsock') <sup>[3]</sup> / February 29, 2020; 0 days ago
<b>Typing discipline</b>	Dynamic
<b>License</b>	GNU GPL v2 <sup>[4]</sup>
<b>Filename extensions</b>	.r .rdata .rds .rda
<b>Website</b>	<a href="http://www.r-project.org">www.r-project.org</a> 

R programming language for Data Science

5

## Giới thiệu R

### □ Đặc điểm

- Chứa các kỹ thuật thống kê
- Có tính mở rộng cao
- Đồ họa chuyên nghiệp
- Giao diện dòng lệnh – command line
- Sử dụng R script
- Có nhiều gói mở rộng
- Chạy trên Windows, Unix, MacOS
- *có giao diện dòng lệnh, hỗ trợ deploy thành sản phẩm UI*

## Giới thiệu R

### □ Lý do chọn R cho Data Science

- Dễ học (nhưng khó trở thành “bậc thầy”)
- Mã nguồn mở, miễn phí
- Có cộng đồng người sử dụng lớn
- Do các chuyên gia thống kê phát triển
- Có số lượng thư viện hỗ trợ lớn cho Data Science và được cập nhật thường xuyên
- Có thể sử dụng để triển khai từng bước trong quy trình của Data Science



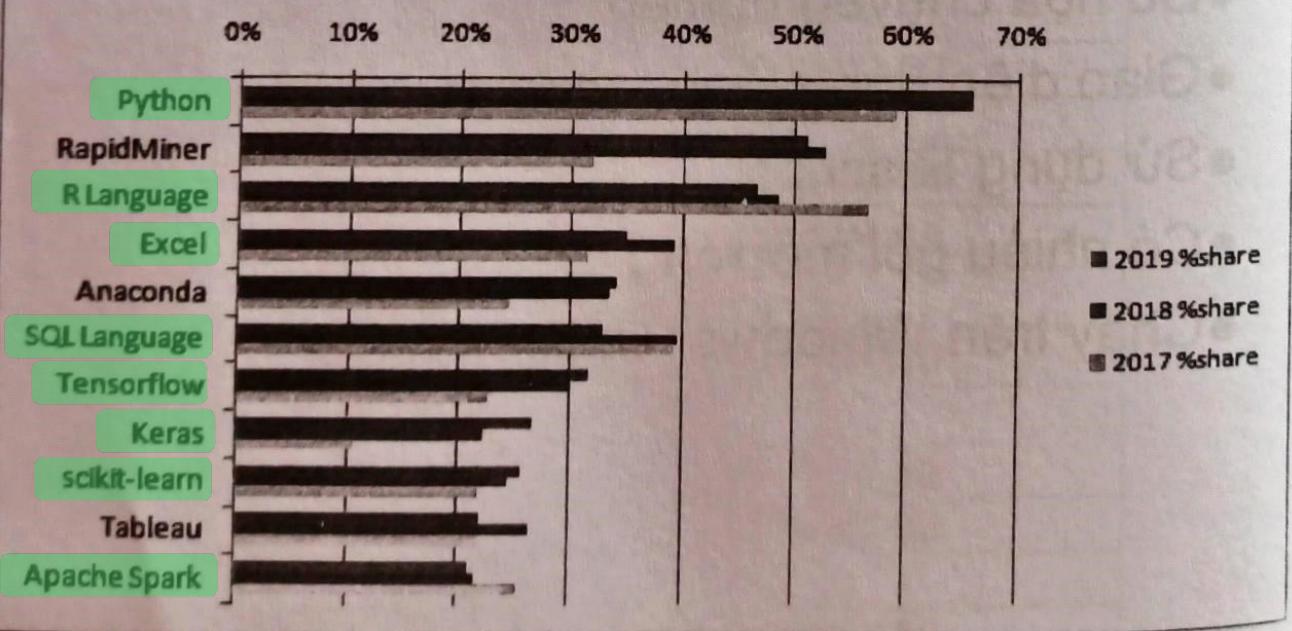
R programming language for Data Science

7

## Giới thiệu R

### □ Lý do chọn R cho Data Science

**Top Analytics, Data Science, Machine Learning Software 2017-2019, KDnuggets Poll**



R programming language for Data Science

8

# Nội dung

- 
1. Giới thiệu R
  2. Môi trường làm việc
  3. Tạo tập tin và thực thi



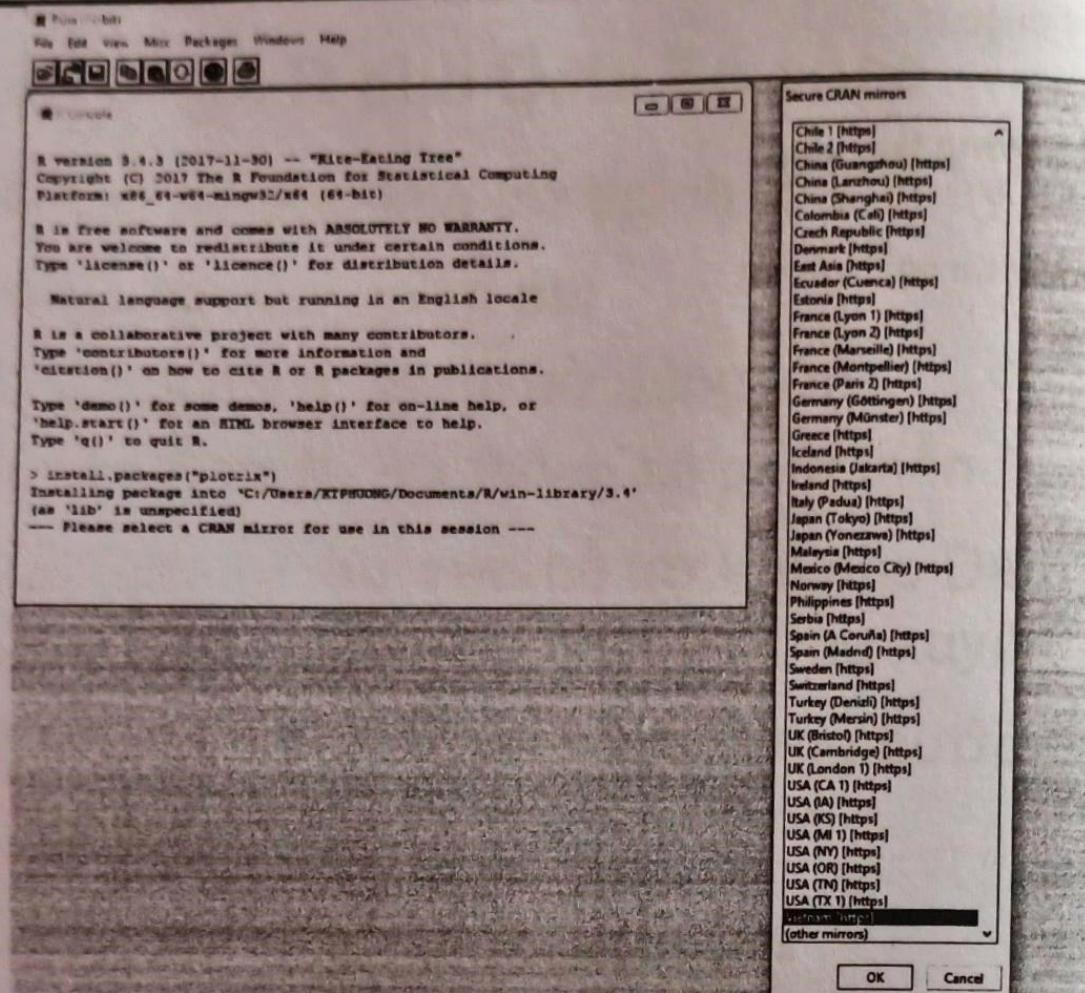
## Môi trường làm việc

### □ Có thể làm việc với:

- R-3.4.3 for Windows (32/64 bit) trở lên (download tại <https://cran.r-project.org/bin/windows/base/>)
- Cài đặt R
- Để sử dụng các thư viện đồ họa: vào command line của R gõ: `install.packages("plotrix")`



# Môi trường làm việc



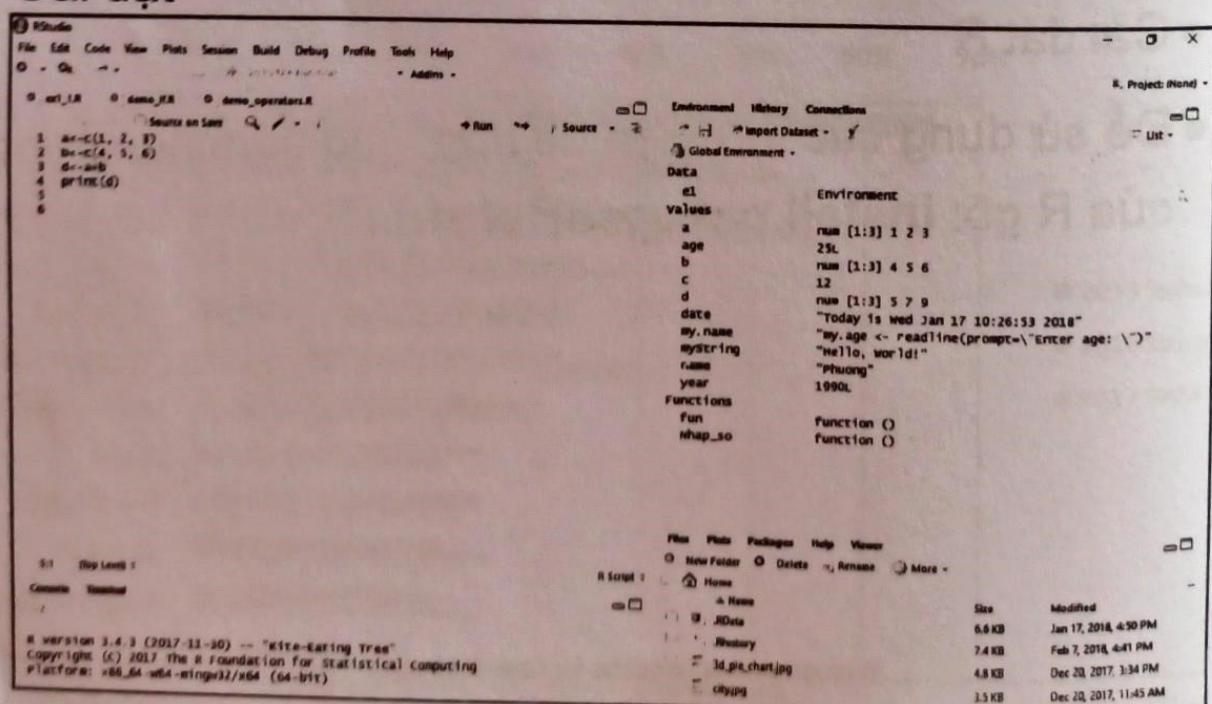
11

# Môi trường làm việc

## ☐ Rstudio (Editor dành cho R)

- Download tại <https://www.rstudio.com/>

- Cài đặt



12

## Môi trường làm việc

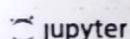
### ❑ Jupyter notebook

- Cài jupyter notebook (dùng chung với Python)
- Đưa R vào jupyter notebook (sau khi đã cài R)
  - pip install zmq
  - Ở command line của R console, gõ lệnh:
    - install.packages(c('repr', 'IRdisplay', 'evaluate', 'crayon', 'pbZIP4', 'devtools', 'uuid', 'digest'))
    - devtools::install\_github('IRkernel/IRkernel')
    - IRkernel::installspec()



## Môi trường làm việc

### ● Kết quả



Quit Logout

Files Running Clusters

Select items to perform actions on them.

- 0 / Chapter1
  - 
  - demo\_Organizing\_data.ipynb
  - demo\_preprocessing\_data\_1.ipynb
  - demo\_Visualization.ipynb
  - read\_spss\_file.ipynb

Name 

Notebook: Python 3

R

Other:

Text File

kB

Folder

kB

Terminal

kB

4 days ago 144 kB

Create a new notebook with R



## Nội dung

1. Giới thiệu R
2. Môi trường làm việc
3. Tạo tập tin và thực thi

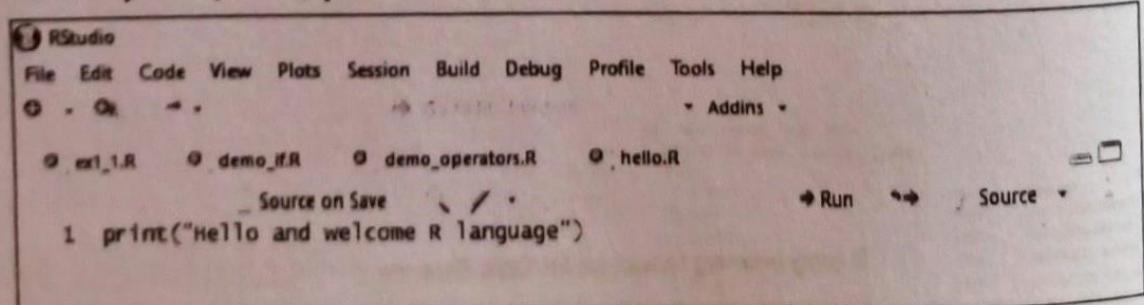


## Tạo tập tin và thực thi

### □ Dùng Rstudio

#### • Tạo tập tin

- Tạo thư mục chứa ứng dụng (thường trong Documents/R)
- Tạo tập tin .R: File > New File > R script
- Ctrl + S > chọn nơi lưu trữ > đặt tên file >
- Ví dụ: Tạo tập tin hello.R xuất ra màn hình câu chào.



```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
  - Ok   - Addins
  ex1_1.R  demo_if.R  demo_operators.R  hello.R
Source on Save  /  Run  Source
1 print("Hello and welcome R language")
  
```



## Tạo tập tin và thực thi

- Thực thi:

- Cách 1: Để chạy dòng code muốn thực thi > chọn các dòng > Nhấn Run
- Cách 2: Để chạy cả file .R > chọn Source
- Ví dụ: Thực thi chương trình hello.R

```
> print("Hello and welcome R language")
[1] "Hello and welcome R language"
>
```

```
> source('~/R/Chuong1/hello.R')
[1] "Hello and welcome R language"
>
```



## Tạo tập tin và thực thi

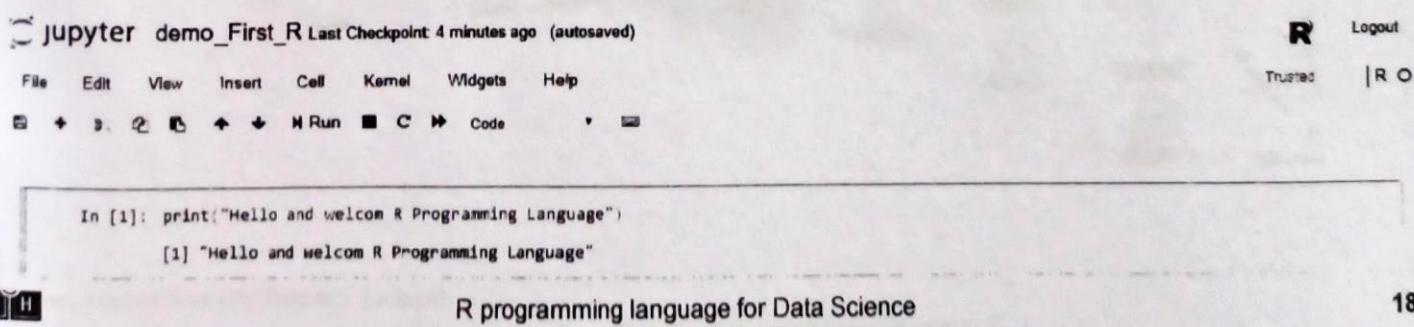
### □ Dùng Jupyter notebook

- Đến nơi sẽ lưu trữ tập tin > cmd > gõ jupyter notebook để vào màn hình làm việc.

- Tạo thư mục

- Tạo tập tin > đặt tên tập tin > Viết code

- Ví dụ: Tạo tập tin demo\_First\_R.ipynb xuất ra màn hình câu chào.



The screenshot shows a Jupyter Notebook interface. At the top, there's a toolbar with File, Edit, View, Insert, Cell, Kernel, Widgets, Help, and a Trusted button. Below the toolbar is a menu bar with File, Edit, View, Insert, Cell, Kernel, Widgets, Help, and a Trusted button. The main area contains a code cell with the following content:

```
In [1]: print("Hello and welcome R Programming Language")
[1] "Hello and welcome R Programming Language"
```

## Tạo tập tin và thực thi

### • Thực thi:

- Cách 1: Để chạy dòng code muốn thực thi > chọn các dòng > nhấn Run (hoặc nhấn Ctrl+Enter ngay dòng code)
- Cách 2: Để chạy cả file > chọn Kernel > Restart & Run All





# Chapter 1: Tổng quan R

## Exercise 1:

- Viết chương trình tính tổng, hiệu, tích, thương của 2 số a và b (với giá trị tùy chọn). In kết quả

```
In [1]: a = 5  
       b = 7  
       tong = a + b  
       hieu = a - b  
       tich = a * b  
       thuong = a/b
```

```
In [2]: print(paste("a=", a))  
       print(paste("b=", b))  
  
[1] "a= 5"  
[1] "b= 7"  
  
In [3]: print(paste("Tong = ", tong))  
       print(paste("Hieu = ", hieu))  
       print(paste("Tich = ", tich))  
       print(paste("Thuong = ", thuong))  
  
[1] "Tong = 12"  
[1] "Hieu = -2"  
[1] "Tich = 35"  
[1] "Thuong = 0.714285714285714"
```