

Prise en main de R
M1MINT
d'après R for Data Science de G. Golemund et H. Wickham

Agathe Guilloux

Introduction

R et Rstudio

- ▶ R est un logiciel libre pour la traitement et l'analyse de données (développée depuis les années 90) disponible sur le CRAN
- ▶ Rstudio est une interface graphique ou, plutôt, un IDE (environnement de développement intégré ou integrated development environment) gratuitement disponible sur <http://rstudio.com>

R version 3.3.3 (2017-03-06) -- "Another Canoe"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

>

Environment History

Global Environment -

Environment is empty

Files Plots Packages Help Viewer

Install Update

Name	Description	Version
System Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-5
<input type="checkbox"/> acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
<input type="checkbox"/> AnnotationDbi	Annotation Database Interface	1.36.2
<input type="checkbox"/> asaur	Data Sets for "Applied Survival Analysis Using R"	0.50
<input type="checkbox"/> asserthat	Easy Pre and Post Assertions	0.2.0
<input type="checkbox"/> astsa	Applied Statistical Time Series Analysis	1.8
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.2
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayesm	Bayesian Inference for Marketing/Micro-Econometrics	3.1-0.1
<input type="checkbox"/> bestglm	Best Subset GLM and Regression Utilities	0.37
<input type="checkbox"/> BH	Boost C++ Header Files	1.66.0-1
<input type="checkbox"/> bindr	Parameterized Active Bindings	0.1.1
<input type="checkbox"/> bindrpp	An 'Rcpp' Interface to Active Bindings	0.2.2
<input type="checkbox"/> Biobase	Biobase: Base functions for Bioconductor	2.34.0
<input type="checkbox"/> BiorGenerics	S4 generic functions for Bioconductor	0.20.0
<input type="checkbox"/> BiocInstaller	Install/Update Bioconductor, CRAN, and github Packages	1.24.0
<input type="checkbox"/> bit	A Class for Vectors of 1-Bit Booleans	1.1-14
<input type="checkbox"/> bit64	A S3 Class for Vectors of 64bit Integers	0.9-7

Console, calculs, variables

Dans la console, exécuter

- ▶ Calculs

```
1+1
```

```
5/2
```

```
3*4
```

```
sqrt(8)
```

```
5^2
```

```
log(1)
```

- ▶ Variables numériques :

```
a <- 1
```

```
b <- 5
```

```
a+b
```

```
a-b
```

```
a < b
```

- ▶ Booléens

```
a <- TRUE
```

```
b <- FALSE
```

```
a & b
```

Vecteurs, fonctions

Dans la console, exécuter

- ▶ vecteurs

```
tailles <- c(156, 164, 197, 147, 173)
tailles
tailles_m <- tailles / 100
tailles_m
tailles + 10
tailles^2
poids <- c(45, 59, 110, 44, 88)
imc <- poids / (tailles / 100) ^ 2
imc
```

- ▶ fonctions

```
length(tailles)
min(tailles)
mean(tailles)
?sum
```

Script 1

The screenshot shows the RStudio interface. The main editor window contains the following R code:

```
(2017-03-06) -- "Another Canoe"  
17 The R Foundation for Statistical Computing  
apple-darwin13.4.0 (64-bit)  
  
1 libre livré sans AUCUNE GARANTIE.  
redistribuer sous certaines conditions.  
' ou 'licence()' pour plus de détails.  
  
collaboratif avec de nombreux contributeurs.  
prs()' pour plus d'information et  
'citation()' pour la façon de le citer dans les publications.  
  
Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide  
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.  
Tapez 'q()' pour quitter R.  
  
> |
```

The Environment pane on the right shows the current environment is empty:

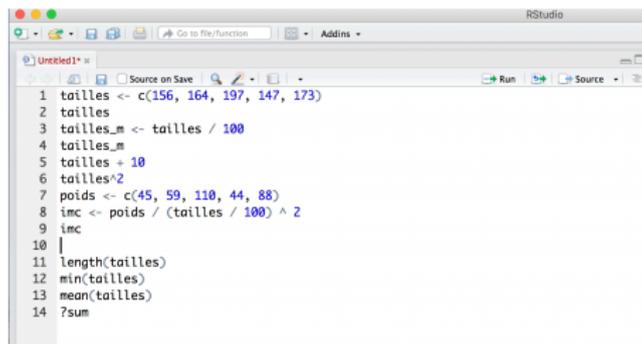
Environment History
Global Environment -
Environment is empty

The Packages pane on the right shows a list of installed and available packages:

Name	Description	Version
System Library		
<input type="checkbox"/> abind	Combine Multidimensional Arrays	1.4-5
<input type="checkbox"/> acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
<input type="checkbox"/> AnnotationDbi	Annotation Database Interface	1.36.2
<input type="checkbox"/> asaur	Data Sets for "Applied Survival Analysis Using R"	0.50
<input type="checkbox"/> assertthat	Easy Pre and Post Assertions	0.2.0
<input type="checkbox"/> astsa	Applied Statistical Time Series Analysis	1.8
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.2
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3
<input type="checkbox"/> bayesm	Bayesian Inference for Marketing/Micro-Econometrics	3.1-0.1
<input type="checkbox"/> bestglm	Best Subset GLM and Regression Utilities	0.37
<input type="checkbox"/> BH	Boost C++ Header Files	1.66.0-1
<input type="checkbox"/> bindr	Parameterized Active Bindings	0.1.1
<input type="checkbox"/> bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2.2
<input type="checkbox"/> Biobase	Biobase: Base functions for Bioconductor	2.34.0
<input type="checkbox"/> BiocGenerics	S4 generic functions for Bioconductor	0.20.0
<input type="checkbox"/> BiocInstaller	Install/Update Bioconductor, CRAN, and github Packages	1.24.0
<input type="checkbox"/> bit	A Class for Vectors of 1-Bit Booleans	1.1-14
<input type="checkbox"/> bit64	A S3 Class for Vectors of 64bit integers	0.9-7

Script II

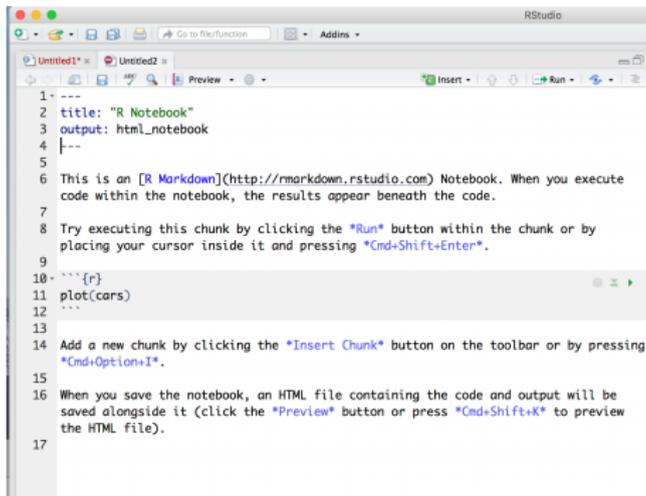
On peut regrouper son code dans un script



```
1 tailles <- c(156, 164, 197, 147, 173)
2 tailles
3 tailles_m <- tailles / 100
4 tailles_m
5 tailles + 10
6 tailles^2
7 poids <- c(45, 59, 110, 44, 88)
8 imc <- poids / (tailles / 100) ^ 2
9 imc
10 |
11 length(tailles)
12 min(tailles)
13 mean(tailles)
14 ?sum
```

Notebook I

ou dans un notebook

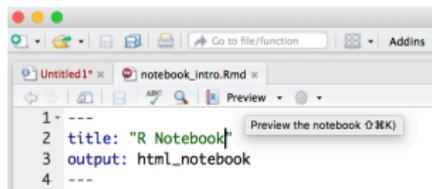


The screenshot shows the RStudio interface with a notebook open. The notebook content is as follows:

```
1 ---
2 title: "R Notebook"
3 output: html_notebook
4 |---
5
6 This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute
7 code within the notebook, the results appear beneath the code.
8 Try executing this chunk by clicking the *Run* button within the chunk or by
9 placing your cursor inside it and pressing *Cmd+Shift+Enter*.
10
11 ```{r}
12 plot(cars)
13 ```
14
15 Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing
16 *Cmd+Option+I*.
17
18 When you save the notebook, an HTML file containing the code and output will be
19 saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview
20 the HTML file).
```

que l'on peut visualiser avec

Notebook II



R Notebook

Code ▾

This is an [R Markdown](#) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

Hide

```
plot(cars)
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

On utilise, Rmarkdown voici un guide

1. Changer le titre
2. Enlever les commentaires inutiles
3. Ajouter le code déjà écrit

Notebook III

4. Ajouter (en markdown)

- ▶ un titre de section
- ▶ un titre de sous-section
- ▶ une liste
- ▶ un commentaire

On obtient

Notebook IV

~/Dropbox/Enseignements/ModelisationStatistique/TP1/notebook_intro.nb.html

notebook_intro.nb.html Open in browser Find Publish

Introduction à R et Rstudio

Code

Introduction à Rstudio notebook

Premier exemple

Voici des vecteurs

- de tailles
- de poids

Hide

```
tailles <- c(156, 164, 197, 147, 173)
tailles
```

```
[1] 156 164 197 147 173
```

Hide

```
poids <- c(45, 59, 110, 44, 88)
```

On calcule l'imc

Hide

```
imc <- poids / (tailles / 100) ^ 2
imc
```

```
[1] 18.49112 21.93635 28.34394 20.36189 29.40292
```

Voici des premières statistiques descriptives

Hide

```
length(imc)
```

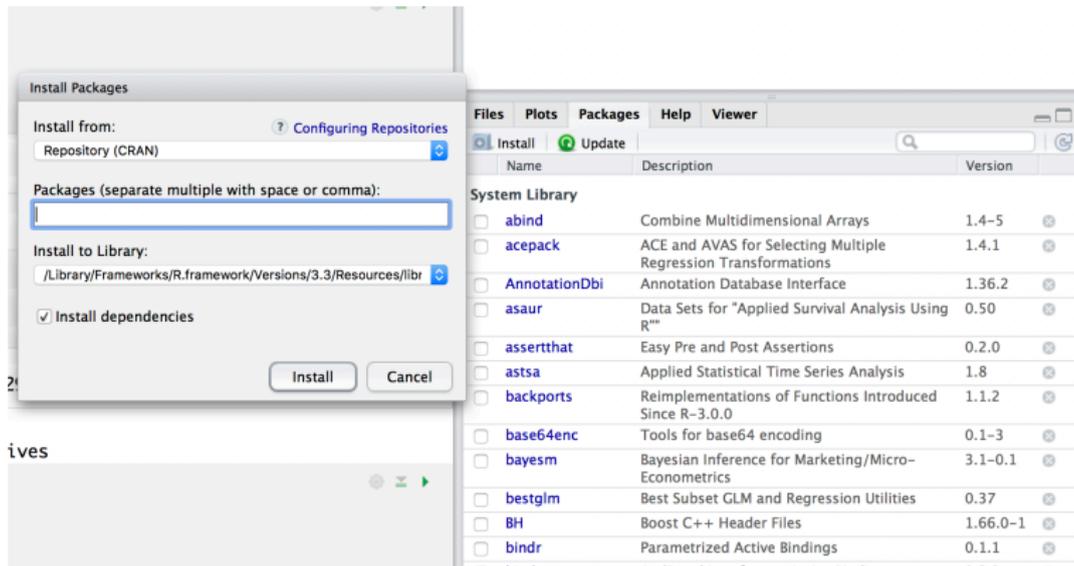
```
[1] 5
```

Hide

```
min(imc)
```

Les packages

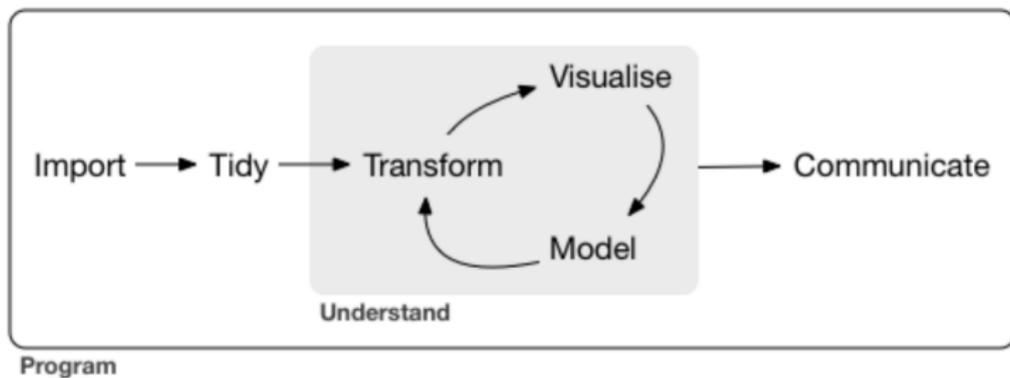
On peut ajouter des extensions, appelées “packages”, qui sont développées par la communauté.



que l'on charge ensuite avec la fonction `library`.

Data science avec R et Rstudio

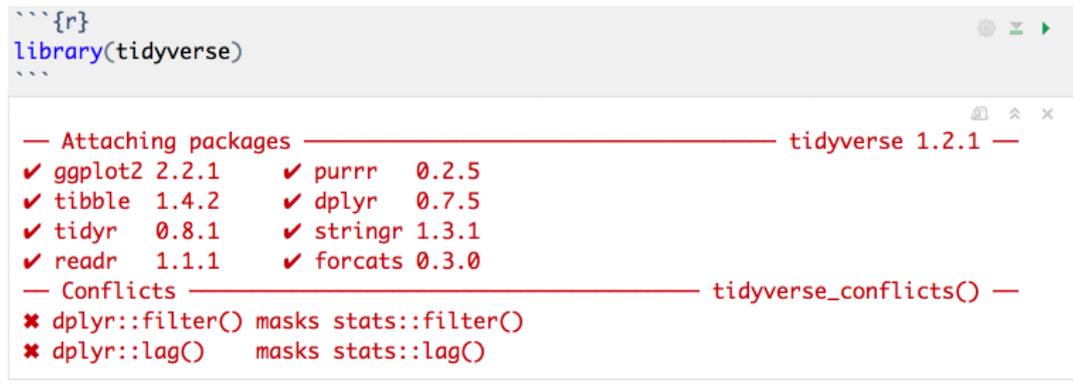
Data Science



Le package tidyverse

Nous avons utiliser le package (ou plutôt l'ensemble de package) tidyverse.
Installer et charger le.

```
```${r}``  
library(tidyverse)
```${r}``
```



```
— Attaching packages — tidyverse 1.2.1 —  
✓ ggplot2 2.2.1    ✓ purrr  0.2.5  
✓ tibble  1.4.2    ✓ dplyr  0.7.5  
✓ tidyr   0.8.1    ✓ stringr 1.3.1  
✓ readr   1.1.1    ✓ forcats 0.3.0  
— Conflicts — tidyverse_conflicts() —  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag()    masks stats::lag()
```

Importation

Les packages R contiennent souvent des données d'exemple

```
data(mpg)
glimpse(mpg)
```

Autrement on peut utiliser le package readr

```
mtcars <- read_csv(file = "mtcars.csv")
```

Data wrangling et visualisation

On utilisera principalement les packages

- ▶ `tidyr` et `dplyr` qui servent à gérer les données, voici une anti-sèche
- ▶ `ggplot2` pour les graphiques, voici une anti-sèche