

Guide R

Prof. Audrey Bürki, Samuel Arthers, Mégane Bollenrücher

2024-08-25

Contents

1	Introduction	5
2	Installation de R et RStudio	7
3	Packages et données	9
3.1	Installation et gestion des packages	9
3.2	Téléchargement des données	9
4	Gestion des données	11

Chapter 1

Introduction

Ceci est le guide R que nous proposons pour vous accompagner durant le cours de Statistiques I.

Chapter 2

Installation de R et RStudio

Chapter 3

Packages et données

Ce premier chapitre introduit deux concepts importants dans R. Le premier est les packages et le second concerne les données.

3.1 Installation et gestion des packages

Les packages sont des regroupements de fonctions et de jeux de données développés dans doivent se télécharger une seule fois, mais ils devront être importés à chaque utilisation. Le code ci-dessous permet d'installer les packages:

```
install.packages(ggplot2, dependencies = TRUE)
```

Avant chaque utilisation des packages, il est nécessaire d'importer le package grâce au code suivant:

```
library(ggplot2)
```

3.2 Téléchargement des données

Dans R, les bases de données se déclinent de plusieurs façons:

1. Les bases de données peuvent être directement incluses dans R ou dans les packages.
2. Les bases de données peuvent être créées dans l'environnement sauvees dans l'environnement R. Ces fichiers ont une extension .RData
3. Les bases de données peuvent être issues de fichiers externes. Ces fichiers ont une extension .csv

3.2.1 Bases de données issues de la base de R ou des packages

Lorsque les données sont issues de la base de R ou des packages, elles peuvent être importées en utilisant le code suivant:

```
data("iris")
```

Le code précédent vous présente également une fonction qui permet de nommer les variables contenues dans cette base de données. Si les données sont dans un package, le package doit être importé au préalable.

Chapter 4

Gestion des données

Plusieurs fonctions permettent de décrire la base de données. Le code suivant vous propose plusieurs fonction avec en commentaire leur utilités.

```
str(iris) #montre la structure de la base de données

## 'data.frame':    150 obs. of  5 variables:
## $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...

dim(iris) #donnes les dimensions de la base de données

## [1] 150    5

nrow(iris) #donne le nombre de lignes de la base de données

## [1] 150

ncol(iris) #donne le nombre de colonnes de la base de données

## [1] 5

head(iris) #affiche les premières lignes de la base de données, par défaut 5 lignes

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1          3.5          1.4          0.2  setosa
## 2          4.9          3.0          1.4          0.2  setosa
## 3          4.7          3.2          1.3          0.2  setosa
## 4          4.6          3.1          1.5          0.2  setosa
## 5          5.0          3.6          1.4          0.2  setosa
## 6          5.4          3.9          1.7          0.4  setosa
```

```
tail(iris, n = 10) #affiche les dix dernières lignes de la base de données
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 141	6.7	3.1	5.6	2.4	virginica
## 142	6.9	3.1	5.1	2.3	virginica
## 143	5.8	2.7	5.1	1.9	virginica
## 144	6.8	3.2	5.9	2.3	virginica
## 145	6.7	3.3	5.7	2.5	virginica
## 146	6.7	3.0	5.2	2.3	virginica
## 147	6.3	2.5	5.0	1.9	virginica
## 148	6.5	3.0	5.2	2.0	virginica
## 149	6.2	3.4	5.4	2.3	virginica
## 150	5.9	3.0	5.1	1.8	virginica