

# Chapitre A

## Présentation du logiciel R

### Pré-requis et objectif

- La lecture du chapitre sur l'installation de R dans les Annexes peut se révéler utile.
- Ce chapitre présente les origines, l'objectif et les spécificités du logiciel R.

SECTION A.1

### Présentation du logiciel

#### A.1.1 Origines

Le logiciel **R** est un logiciel de statistique créé par Ross Ihaka & Robert Gentleman [21]. Il est à la fois un langage informatique et un environnement de travail : les commandes sont exécutées grâce à des instructions codées dans un langage relativement simple, les résultats sont affichés sous forme de texte et les graphiques sont visualisés directement dans une fenêtre qui leur est propre. C'est un clone du logiciel S-plus qui est fondé sur le langage de programmation orienté objet S, développé par AT&T Bell Laboratories en 1988 [4]. Ce logiciel sert à manipuler des données, à tracer des graphiques et à faire des analyses statistiques sur ces données.

#### A.1.2 Pourquoi utiliser R ?

Tout d'abord **R** est un logiciel **gratuit** et à **code source ouvert** (*open-source*). Il fonctionne sous UNIX (et Linux), Windows et Macintosh. C'est donc un logiciel **multi-plates-formes**. Il est développé dans la mouvance des logiciels libres par une communauté sans cesse plus vaste de bénévoles motivés.

Tout le monde peut d'ailleurs contribuer à son amélioration en y intégrant de nouvelles fonctionnalités ou méthodes d'analyse non encore implémentées. Cela en fait donc un logiciel en rapide et constante évolution.

C'est aussi un outil très puissant et très complet, particulièrement bien adapté pour la mise en œuvre informatique de **méthodes statistiques**. Il est plus difficile d'accès que certains autres logiciels du marché (comme SPSS ou Minitab par exemple), car il n'est pas conçu pour être utilisé à l'aide de « clics » de souris dans des menus. L'avantage en est toutefois double :

- l'approche est **pédagogique** puisqu'il faut maîtriser les méthodes statistiques pour parvenir à les mettre en œuvre ;
- l'outil est très **efficace** lorsque l'on domine le langage **R** puisque l'on devient alors capable de créer ses propres outils, ce qui permet ainsi d'opérer des analyses très sophistiquées sur les données.

#### Attention



**R** est plus difficile d'accès que certains autres logiciels du marché. Il faut passer du temps à en apprendre la syntaxe et les commandes.

Le logiciel **R** est particulièrement performant pour la manipulation de données, le calcul et l'affichage de graphiques. Il possède, entre autres choses :

- un système de documentation intégré très bien conçu (en anglais) ;
- des procédures efficaces de traitement des données et des capacités de stockage de ces données ;
- une suite d'opérateurs pour des calculs sur des tableaux et en particulier sur des matrices ;
- une vaste et cohérente collection de procédures statistiques pour l'analyse de données ;
- des capacités graphiques évoluées ;
- un langage de programmation simple et efficace intégrant les conditions, les boucles, la récursivité, et des possibilités d'entrée-sortie.

#### SECTION A.2

## R et les statistiques

**R** est un logiciel dans lequel de nombreuses techniques statistiques modernes et classiques ont été implémentées. Les méthodes les plus courantes permettant de réaliser une analyse statistique telles que :

- statistique descriptive ;
- tests d'hypothèses ;
- analyse de la variance ;

- méthodes de régression linéaire (simple et multiple) ;
- etc.

sont enchâssées directement dans le cœur du système. Notez également que la plupart des méthodes avancées de statistique sont aussi disponibles au travers de modules externes appelés *packages*. Ceux-ci sont faciles à installer directement à partir d'un menu du logiciel. Ils sont tous regroupés sur le site internet du *Comprehensive R Archive Network* (CRAN) (<http://cran.r-project.org>) sur lequel vous pouvez les consulter. Ce site fournit aussi, pour certains grands domaines d'étude, une liste commentée des *packages* associés à ces thèmes (appelée *Task View*), ce qui facilite ainsi la recherche d'une méthode statistique particulière. Par ailleurs, une documentation détaillée en anglais de chaque *package* est disponible sur le CRAN.

Il est par ailleurs utile de noter que les méthodes statistiques les plus récentes y sont régulièrement ajoutées par la communauté statistique elle-même.

#### Renvoi

Le lecteur pourra consulter avec profit la section C.2 détaillant les procédures à mettre en œuvre afin d'installer de nouveaux *packages*.



#### SECTION A.3

## R et les graphiques

Une des grandes forces de **R** réside dans ses capacités, bien supérieures à celles des autres logiciels courants du marché, à combiner un langage de programmation avec la possibilité de réaliser des graphiques de qualité. Les graphiques usuels s'obtiennent aisément au moyen de fonctions prédéfinies. Ces dernières possèdent de très nombreux paramètres permettant par exemple d'ajouter des titres, des légendes, des couleurs, etc. Mais il est également possible d'effectuer des graphiques plus sophistiqués permettant de représenter des données complexes telles que des courbes de surface ou de niveau, des volumes affichés avec un effet 3D, des courbes de densité, et bien d'autres choses encore. Il vous est également possible d'y ajouter des formules mathématiques. Vous pouvez aussi agencer ou superposer plusieurs graphiques sur une même fenêtre, et utiliser de nombreuses palettes de couleur.

Vous pouvez obtenir une démonstration des possibilités graphiques de **R** en tapant successivement les commande suivantes :

```
demo(image)
example(contour)
demo(graphics)
demo(persp)
```

```

demo(plotmath)
demo(Hershey)
require(lattice) # Charge le package que vous devez avoir
                  # préalablement installé en passant par le
                  # menu Packages/Installer le(s) package(s) .

demo(lattice)
exemple(wireframe)
require(rgl)      # Même remarque que ci-dessus.
demo(rgl)        # Possibilité d'interaction avec la souris.
exemple(persp3d)

```

La figure ci-dessous présente quelques-uns de ces graphiques.

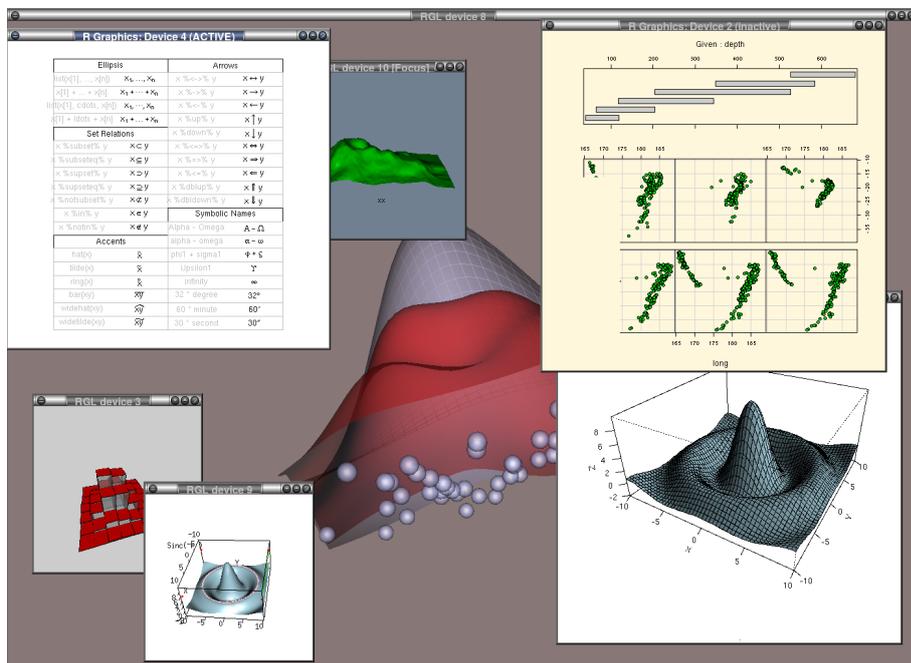


FIG. A.1: Quelques possibilités graphiques offertes par R.

#### SECTION A.4

## L'interface graphique de R

L'interface graphique de R (c'est-à-dire l'ensemble de ses menus) est très limitée, voire inexistante sur certaines plates-formes, en comparaison des autres logiciels standards. Cette minimalité peut dérouter les novices. Toutefois, nous pouvons nuancer cet inconvénient au travers des points suivants :

- cela offre l'avantage pédagogique d'inciter l'utilisateur à bien maîtriser la procédure statistique qu'il compte appliquer ;
- il existe des outils additionnels qui permettent d'étendre l'interface graphique.

Le *package Rcmdr*, à installer via le menu Packages et à charger (activer) à l'aide de la commande

`require(Rcmdr)`, permet d'effectuer des analyses graphiques et statistiques usuelles au moyen d'une interface plus conviviale à l'aide de menus déroulants. Par ailleurs, les instructions **R** permettant de réaliser l'analyse choisie dans les menus de *Rcmdr* s'affichent dans une portion de fenêtre dédiée. Cela peut par exemple être utile si l'on ne connaît pas ou que l'on a oublié une instruction **R** nécessaire à la réalisation d'une tâche particulière. En voici une petite illustration.

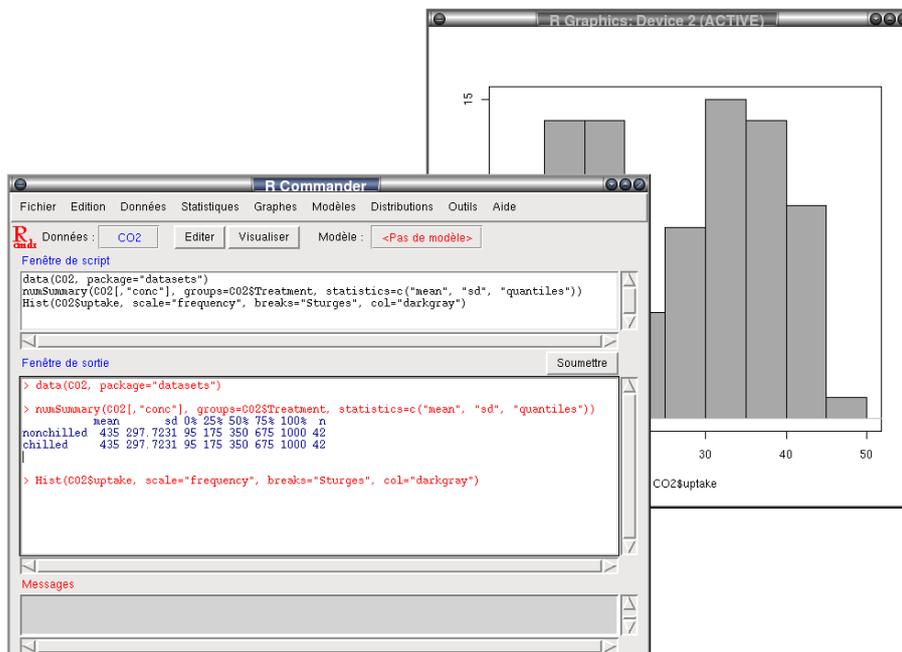


FIG. A.2: Illustration de l'interface graphique *Rcmdr*.

#### Astuce

Notez qu'il est possible, lorsque vous maîtriserez le **R**, de développer vous-même des outils similaires à *Rcmdr*, à destination d'un utilisateur final



qui ne souhaite pas apprendre le logiciel R mais seulement utiliser, de la façon la plus ergonomique possible, une procédure que vous aurez élaborée pour lui. Vous pourrez pour cela faire appel au *package tcltk*.

Mac



Les utilisateurs du système d'exploitation Macintosh, pourront consulter avec profit les instructions disponibles à l'adresse <http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/installation-notes.html>, après avoir installé le *package tcltk* disponible sur le CRAN.