

Sommaire

I	R et son fonctionnement	1
1	Concepts	3
1.1	Installation de R et RStudio	3
1.2	Environnement de travail	5
1.3	Introduction à RMarkdown	6
1.3.1	Création d'un document RMarkdown	6
1.3.2	Bases du langage	6
1.3.3	Insertion du code R	7
1.4	Les différentes aides	8
1.5	Les objets R	8
1.5.1	Création, affichage, suppression	9
1.5.2	Le mode d'un objet	10
1.5.3	La valeur manquante	11
1.5.4	Les vecteurs	11
1.5.5	Les matrices	17
1.5.6	Les facteurs	22
1.5.7	Les listes	23
1.5.8	Les data-frames	25
1.5.9	La classe d'un objet	26
1.6	Les fonctions	27
1.7	Les packages	28
1.7.1	Installation d'un package	29
1.7.2	Utilisation d'un package	29
1.7.3	Mise à jour des packages et de R	30
1.8	Exercices	31
2	Manipuler les données	35
2.1	Importer des données	35
2.1.1	Importation d'un fichier texte	35
2.1.2	Importation de données textes volumineuses	37

2.1.3	Importation d'autres formats de données	38
2.1.4	Importation via le menu de RStudio	38
2.2	Exporter des résultats	38
2.3	Manipuler les variables	39
2.3.1	Changer de type	39
2.3.2	Découpage en classes	41
2.3.3	Travail sur le niveau des facteurs	42
2.4	Manipuler les individus	45
2.4.1	Repérer les données manquantes	45
2.4.2	Repérer les individus aberrants univariés	47
2.4.3	Repérer et/ou éliminer des doublons	48
2.5	Concaténer des tableaux de données	49
2.6	Tableau croisé	52
2.7	Exercices	54
3	Visualiser les données	57
3.1	Les fonctions graphiques conventionnelles	57
3.1.1	La fonction plot	58
3.1.2	Représentation d'une distribution	62
3.1.3	Ajouts aux graphiques	64
3.1.4	Graphiques en plusieurs dimensions	66
3.1.5	Exportation de graphiques	68
3.1.6	Plusieurs graphiques	69
3.1.7	Amélioration et personnalisation des graphiques	71
3.2	Les fonctions graphiques avec ggplot2	74
3.2.1	Premiers graphes avec ggplot2	75
3.2.2	La grammaire ggplot	76
3.2.3	Group et facets	83
3.2.4	Compléments	86
3.3	Les graphiques interactifs	88
3.4	Construire des cartes	90
3.4.1	Carte statique dans R	90
3.4.2	Carte dans un navigateur	93
3.4.3	Carte avec contours : le format shapefile	97
3.5	Exercices	101
4	Programmer	109
4.1	Structures de contrôle	109
4.1.1	Commandes groupées	109
4.1.2	Les boucles (for ou while)	109
4.1.3	Les conditions (if, else)	111
4.2	Construire une fonction	112
4.3	La famille apply, des fonctions d'itération prédéfinies	114
4.4	Calcul parallèle	121

4.4.1	Introduction	121
4.4.2	Le package <code>parallel</code>	122
4.4.3	Le package <code>foreach</code>	124
4.4.4	Exemple avancé	125
4.5	Faire une application <code>shiny</code>	128
4.6	Exercices	137
5	Outils avancés pour la préparation des données	139
5.1	Le package <code>data.table</code>	139
5.1.1	Importation avec <code>fread</code>	140
5.1.2	Syntaxe	142
5.1.3	Sélection	142
5.1.4	Manipulation	145
5.1.5	Pour aller plus loin	149
5.2	Le package <code>dplyr</code> et le <code>tidyverse</code>	151
5.2.1	Le package <code>dplyr</code>	151
5.2.2	Manipulation de tables	155
5.2.3	Pour aller plus loin	158
5.3	Bases de données	162
5.3.1	SQL : Structured Query Language	162
5.3.2	JSON : JavaScript Object Notation	171
5.4	Web scraping	180
5.4.1	Introduction	180
5.4.2	Approche naïve	181
5.4.3	Le package <code>rvest</code>	182
5.4.4	Pour aller plus loin	187
5.5	Exercices	189
II	Fiches thématiques	193
6	Intervalles de confiance et tests d'hypothèses	195
6.1	Intervalle de confiance d'une moyenne	196
6.2	Test du χ^2 d'indépendance	199
6.3	Comparaison de deux moyennes	204
6.4	Tests sur les proportions	210
7	Analyses factorielles	213
7.1	Analyse en Composantes Principales	214
7.2	Analyse Factorielle des Correspondances	223
7.3	Analyse des Correspondances Multiples	228
7.4	Analyse Factorielle Multiple	237

8	Classification non supervisée	243
8.1	Classification Ascendante Hiérarchique	244
8.2	Méthode des K -means	252
8.3	Modèles de mélange	256
9	Méthodes usuelles de régression	263
9.1	Régression simple	264
9.2	Régression multiple	270
9.3	Analyse de la variance	276
9.4	Analyse de la covariance	282
9.5	Régression logistique	287
9.6	Analyse discriminante linéaire	294
9.7	Arbres	302
9.8	Régression Partial Least Square (PLS)	312
10	Machine learning	321
10.1	Calibration d'un algorithme avec <code>caret</code>	323
10.2	Forêts aléatoires	331
10.3	Régression sous contraintes	340
10.4	Gradient boosting	348
10.5	SVM	354
10.6	Réseaux de neurones et deep learning	362
10.7	Comparaison de méthodes	369
11	Divers	373
11.1	Gestion de données manquantes	374
11.2	Analyse de texte	383
11.3	Fouille de graphe	393
	Annexes	401
A.1	Écriture d'une formule pour les modèles	401
A.2	Environnement RStudio	402
A.3	Le package Rcmdr	403
	Bibliographie	405
	Index des fonctions	407
	Index	410