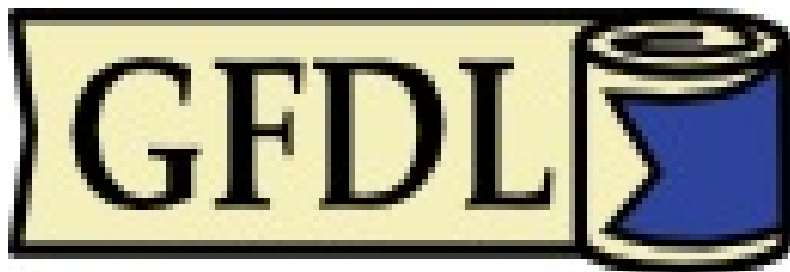


© 2012 Vincent Goulet



Il est permis de copier, distribuer ou modifier ce document selon les termes de la *GNU Free Documentation License*, version 1.3 ou toute version subséquente publiée par la Free Software Foundation ; avec aucune section inaltérable (*Invariant Sections*), aucun texte de couverture avant (*Front-Cover Texts*), et aucun texte de couverture arrière (*Back-Cover Texts*). Une copie de la licence est incluse à l'annexe D.

Code source

Le code source ~~LaTeX~~ et R de ce document est disponible dans le dépôt Subversion

https://svn.fsg.ulaval.ca/svn-pub/vgoulet/documents/intro_r/

ou en communiquant directement avec l'auteur.

Historique de publication

Avril 2012 : Troisième édition

Janvier 2007 Seconde édition

Janvier 2006 Première édition

ISBN 978-2-9809136-7-9

Dépôt légal – Bibliothèque nationale du Québec, 2012

Dépôt légal – Bibliothèque et Archives Canada, 2012

Couverture

Tornade de catégorie F5 à l'approche de la ville de Elie, Manitoba, le 22 juin 2007.

Crédits photo : Justin1569 via Wikimedia Commons

Table des matières

Introduction	v
1 Présentation du langage R	1
1.1 Bref historique	1
1.2 Description sommaire de R	2
1.3 Interfaces	3
1.4 Stratégies de travail	4
1.5 Éditeurs de texte	5
1.6 Anatomie d'une session de travail	8
1.7 Répertoire de travail	9
1.8 Consulter l'aide en ligne	9
1.9 Où trouver de la documentation	9
1.10 Exemples	10
1.11 Exercices	11
2 Bases du langage R	13
2.1 Commandes R	13
2.2 Conventions pour les noms d'objets	15
2.3 Les objets R	16
2.4 Vecteurs	20
2.5 Matrices et tableaux	21
2.6 Listes	24
2.7 <i>Data frames</i>	26
2.8 Indixage	26
2.9 Exemples	28
2.10 Exercices	39
3 Opérateurs et fonctions	41
3.1 Opérations arithmétiques	41

3.2	Opérateurs	42
3.3	Appels de fonctions	43
3.4	Quelques fonctions utiles	44
3.5	Structures de contrôle	50
3.6	Fonctions additionnelles	51
3.7	Exemples	52
3.8	Exercices	60
4	Exemples résolus	63
4.1	Calcul de valeurs présentes	63
4.2	Fonctions de masse de probabilité	64
4.3	Fonction de répartition de la loi gamma	66
4.4	Algorithme du point fixe	68
4.5	Suite de Fibonacci	69
4.6	Exercices	70
5	Fonctions définies par l'utilisateur	73
5.1	Définition d'une fonction	73
5.2	Retourner des résultats	74
5.3	Variables locales et globales	74
5.4	Exemple de fonction	75
5.5	Fonctions anonymes	76
5.6	Débogage de fonctions	76
5.7	Styles de codage	77
5.8	Exemples	78
5.9	Exercices	82
6	Concepts avancés	87
6.1	L'argument '...'	87
6.2	Fonction apply	88
6.3	Fonctions lapply et sapply	90
6.4	Fonction mapply	92
6.5	Fonction replicate	93
6.6	Classes et fonctions génériques	94
6.7	Exemples	95
6.8	Exercices	102
7	Fonctions d'optimisation	107
7.1	Contexte	107
7.2	Fonctions d'optimisation et de calcul de racines	108

7.3	Pour en savoir plus	109
7.4	Exemples	109
7.5	Exercices	113
8	Générateurs de nombres aléatoires	115
8.1	Générateurs de nombres aléatoires	115
8.2	Fonctions de simulation de variables aléatoires non uniformes . .	116
8.3	Exemples	117
8.4	Exercices	119
A	GNU Emacs et ESS : la base	121
A.1	Mise en contexte	121
A.2	Installation	122
A.3	Description sommaire	122
A.4	<i>Emacs-ismes</i> et <i>Unix-ismes</i>	123
A.5	Commandes de base	124
A.6	Anatomie d'une session de travail (bis)	127
A.7	Configuration de l'éditeur	128
A.8	Aide et documentation	128
B	Planification d'une simulation en R	129
B.1	Introduction	129
B.2	Première approche : avec une boucle	130
B.3	Seconde approche : avec <code>sapply</code>	130
B.4	Variante de la seconde approche	134
B.5	Gestion des fichiers	134
B.6	Exécution en lot	135
B.7	Conclusion	136
C	Installation de packages dans R	137
D	GNU Free Documentation License	139
D.1	APPLICABILITY AND DEFINITIONS	139
D.2	VERBATIM COPYING	141
D.3	COPYING IN QUANTITY	142
D.4	MODIFICATIONS	142
D.5	COMBINING DOCUMENTS	144
D.6	COLLECTIONS OF DOCUMENTS	145
D.7	AGGREGATION WITH INDEPENDENT WORKS	145
D.8	TRANSLATION	146

D.9 TERMINATION	146
D.10 FUTURE REVISIONS OF THIS LICENSE	146
ADDENDUM: How to use this License for your documents	147
Réponses des exercices	149
Chapitre 2	149
Chapitre 3	150
Chapitre 4	151
Chapitre 5	152
Chapitre 6	156
Bibliographie	161
Index	163

Bibliographie

- Abelson, H., G. J. Sussman et J. Sussman. 1996, *Structure and Interpretation of Computer Programs*, 2^e éd., MIT Press, ISBN 0-26201153-0.
- Becker, R. A. 1994, «A brief history of S», cahier de recherche, AT&T Bell Laboratories. URL <http://cm.bell-labs.com/cm/ms/departments/sia/doc/94.11.ps>.
- Becker, R. A. et J. M. Chambers. 1984, *S: An Interactive Environment for Data Analysis and Graphics*, Wadsworth, ISBN 0-53403313-X.
- Becker, R. A., J. M. Chambers et A. R. Wilks. 1988, *The New S Language: A Programming Environment for Data Analysis and Graphics*, Wadsworth & Brooks/Cole, ISBN 0-53409192-X.
- Braun, W. J. et D. J. Murdoch. 2007, *A First Course in Statistical Programming with R*, Cambridge University Press, ISBN 978-0-52169424-7.
- Cameron, D., J. Elliott, M. Loy, E. S. Raymond et B. Rosenblatt. 2004, *Learning GNU Emacs*, 3^e éd., O'Reilly, Sebastopol, CA, ISBN 0-59600648-9.
- Chambers, J. M. 1998, *Programming with Data: A Guide to the S Language*, Springer, ISBN 0-38798503-4.
- Chambers, J. M. 2000, «Stages in the evolution of S», URL <http://cm.bell-labs.com/cm/ms/departments/sia/S/history.html>.
- Chambers, J. M. 2008, *Software for Data Analysis: Programming with R*, Springer, ISBN 978-0-38775935-7.
- Chambers, J. M. et T. J. Hastie. 1992, *Statistical Models in S*, Wadsworth & Brooks/Cole, ISBN 0-53416765-9.
- Hornik, K. 2011, «The R FAQ», URL <http://cran.r-project.org/doc/FAQ/R-FAQ.html>, ISBN 3-90005108-9.

- Iacus, S. M., S. Urbanek et R. J. Goedman. 2011, «R for Mac OS X FAQ», URL <http://cran.r-project.org/bin/macosx/RMacOSX-FAQ.html>.
- IEEE. 2003, *754-1985 IEEE Standard for Binary Floating-Point Arithmetic*, IEEE, Piscataway, NJ.
- Ihaka, R. et R. Gentleman. 1996, «R: A language for data analysis and graphics», *Journal of Computational and Graphical Statistics*, vol. 5, n° 3, p. 299–314.
- Ligges, U. 2003, «R-winedt», dans *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, édité par K. Hornik, F. Leisch et A. Zeileis, TU Wien, Vienna, Austria, ISSN 1609-395X. URL <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/>.
- Redd, A. 2010, «Introducing NppToR: R interaction for Notepad++», *R Journal*, vol. 2, n° 1, p. 62–63. URL http://journal.r-project.org/archive/2010-1/RJournal_2010-1.pdf.
- Ripley, B. D. et D. J. Murdoch. 2011, «R for Windows FAQ», URL <http://cran.r-project.org/bin/windows/base/rw-FAQ.html>.
- Venables, W. N. et B. D. Ripley. 2000, *S Programming*, Springer, New York, ISBN 0-38798966-8.
- Venables, W. N. et B. D. Ripley. 2002, *Modern Applied Statistics with S*, 4^e éd., Springer, New York, ISBN 0-38795457-0.
- Venables, W. N., D. M. Smith et R Development Core Team. 2011, *An Introduction to R*, R Foundation for Statistical Computing. URL <http://cran.r-project.org/doc/manuals/R-intro.html>.
- Wheeler, B. 2009, **SuppDists**: *Supplementary Distributions*. URL <http://www.bobwheeler.com/stat,r> package version 1.1-8.

Index

Les numéros de page en caractères gras indiquent les pages où les concepts sont introduits, définis ou expliqués.

!, **43**
!=, **43**
*, **43**
+, **43**
-, **43**
->, **14**, **43**
->>, **43**
-Inf, **19**
..., **87**, **132**
/, **43**
:, **43**, **60**
;, **14**
<, **43**
<-, **14**, **42**, **43**
<<-, **43**, **74**
<=, **43**
=, **14**
==, **43**
>, **43**
>=, **43**
[, **27**
[<-, **27**
[[, **157**
[[], **25**, **25**
[], **21**, **23**, **25**, **27**
\$, **26**, **27**, **43**
\$<-, **27**
%*%, **43**, **83**
%/%, **43**
%%, **43**
%in%, **46**, **56**
%O%, **50**, **58**
&, **43**
&&, **43**
^, **42**, **43**
{ }, **15**

abs, **79**, **80**, **99**, **100**
add, **100**
affectation, **13**
apply, **51**, **58**, **61**, **87**, **88**, **88**, **89**, **90**, **95**,
 96
array, **22**, **34**, **95**, **96**
array (classe), **22**
arrondi, **46**
as.data.frame, **26**
attach, **26**, **37**
attr, **19**, **31**, **101**, **102**
attribut, **19**
attributes, **19**, **31**, **32**

boucle, **51**, **70**, **130**
break, **51**, **60**, **79**, **80**, **99**
by, **10**, **55**

- byrow, 22, 44
- c, 20
- cat, 99
- cbind, 24, 26, 34, 39, 57
- ceiling, 46, 56
- character, 21, 33
- character (mode), 17, 21
- choose, 65
- class, 32–34, 36, 94, 99
- class (attribut), 20
- colMeans, 49, 61, 88, 95
- colSums, 48, 58, 61, 68, 88
- compilé (langage), 2
- complex, 33
- complex (mode), 17
- cos, 28
- cummax, 48, 57
- cummin, 48, 57
- cumprod, 48, 57, 64
- cumsum, 48, 57
- curve, 100
- data, 31, 44, 55
- data frame, 26
- data.frame, 26
- data.frame (classe), 26
- dbeta, 110
- dbinom, 65
- density, 10
- det, 49
- detach, 26, 37
- dgamma, 111
- diag, 39, 49, 57
- diff, 47, 57
- différences, 47
- dim, 32–34, 36, 38, 57, 95
- dim (attribut), 20, 21, 22
- dimension, 20, 39
- dimnames, 32, 44
- dimnames (attribut), 20
- distribution
 - bêta, 117
 - binomiale, 64, 117
 - binomiale négative, 117
 - Cauchy, 117
 - exponentielle, 117
 - F, 117
 - gamma, 66, 84, 117
 - géométrique, 117
 - hypergéométrique, 117
 - khi carré, 117
 - log-normale, 117, 119
 - logistique, 117
 - mélange discret, 119
 - mélange Poisson/gamma, 119
 - normale, 82–84, 117
 - Pareto, 84, 103
 - Poisson, 65, 71, 117
 - t, 117
 - uniforme, 117
 - Weibull, 117
 - Wilcoxon, 117
- dnorm, 83
- dossier de travail, voir répertoire de travail
- dpois, 66
- écart type, 47
- ecdf, 158
- else, 50, 58–60, 80, 98
- Emacs, 7, 78
 - C-_, 125
 - C-g, 125
 - C-r, 125
 - C-s, 125
 - C-SPC, 125
 - C-w, 125
 - C-x 0, 126
 - C-x 1, 126

- C-x 2, 126
- C-x b, 126
- C-x C-f, 125
- C-x C-s, 125, 128
- C-x C-w, 125
- C-x k, 125
- C-x o, 126
- C-x o , 127
- C-x u, 125
- C-y, 125
- configuration, 128
- M-%, 125
- M-w, 125
- M-x, 125
- M-y, 125
- nouveau fichier, 125
- rechercher et remplacer, 125
- sélection, 125
- sauvegarder, 125
- sauvegarder sous, 125
- ESS, 7
 - C-c C-e, 126
 - C-c C-e , 127
 - C-c C-f, 126
 - C-c C-l, 126
 - C-c C-n, 126
 - C-c C-n , 127
 - C-c C-o, 126
 - C-c C-q, 126, 128
 - C-c C-r, 126
 - C-c C-v, 126
 - h, 126
 - l, 127
 - M-h, 126
 - M-n, 126
 - M-p, 126
 - n, 126
 - p, 126
 - q, 127
 - r, 127
 - x, 127
- étiquette, 20, 39
- eval, 80, 98
- exists, 37
- exp, 28, 65, 67, 111, 112
- expression, 13
- expression, 30, 80, 98
- expression (mode), 17
- extraction, voir aussi indiçage
 - derniers éléments, 45
 - éléments différents, 45
 - premiers éléments, 45
- F, voir FALSE
- factorial, 61, 65
- FALSE, 16, 77
- floor, 46, 56
- fonction
 - anonyme, 76
 - appel, 43
 - débogage, 76
 - définie par l'utilisateur, 73
 - générique, 94
 - maximum local, 108
 - minimum, 108
 - minimum local, 108
 - optimisation, 108
 - racine, 108
 - résultat, 74
- for, 51, 53, 54, 58, 59, 81, 92, 130
- function, 73, 79–81, 96–102, 109–111
- function (mode), 17
- gamma, 28, 61, 65
- head, 45, 56
- hist, 98, 104
- if, 50, 53, 54, 58–60, 77, 79, 80, 98, 99
- ifelse, 50
- Im, 113

- indichage
 - liste, **25**, **39**
 - matrice, **23**, **26**, **40**
 - vecteur, **26**, **39**
- Inf, **19**
- install.packages, **52**
- interprété (langage), **2**
- is.finite, **19**
- is.infinite, **19**
- is.na, **19**, **31**, **38**, **59**
- is.nan, **19**
- is.null, **18**
- lapply, **51**, **87**, **90**, **90**, **91**, **92**, **96**, **97**,
101, 130, 132
- length, **10**, **17**, 30–36, **38**, 55–57
- lfactorial, **61**
- lgamma, **61**
- library, **52**, **60**
- list, **25**, **30**, **32**, **35**, **36**, **96**, **97**, **99**, **101**
- list (mode), **17**, **24**
- liste, **24**
- lm, **102**
- log, **111**
- logical, **21**, **33**
- logical (mode), **17**, **19**, **21**
- longueur, **18**, **39**
- lower, **109**, **110**, **112**
- ls, **11**, **29**, **101**
- mapply, **51**, **92**, **97**
- match, **46**, **56**
- matrice, **61**, **82**, **83**, **88**
 - diagonale, **49**
 - identité, **49**
 - inverse, **49**
 - moyennes par colonne, **49**
 - moyennes par ligne, **49**
 - somme par colonne, **48**
 - sommes par ligne, **48**
 - transposée, **49**
- matrix, **11**, **22**, **28**, **33**, **34**, **36**, **53**, **55**,
82, 95, 96
- matrix (classe), **21**
- max, **10**, **11**, **47**, **57**, **95**
- maximum
 - cumulatif, **48**
 - d'un vecteur, **47**
 - local, **108**
 - parallèle, **48**
 - position dans un vecteur, **46**
- mean, **19**, **31**, **47**, **57**, **95**, **98**
- median, **47**, **57**
- médiane, **47**
- methods, **94**
- min, **10**, **11**, **47**, **57**
- minimum
 - cumulatif, **48**
 - d'un vecteur, **47**
 - fonction non linéaire, **108**
 - local, **108**
 - parallèle, **48**
 - position dans un vecteur, **46**
- mode, **17**, **39**
- mode, **16**, **30**, **31**, **35**, **36**
- moyenne
 - arithmétique, **47**
 - harmonique, **71**
 - pondérée, **70**, **102**
 - tronquée, **47**
- NA, **19**, **77**
- na.rm, **19**, **31**, **95**
- names, **32**, 36–38
- names (attribut), **20**
- NaN, **19**
- nchar, **18**, **30**
- ncol, **11**, **33**, **34**, **44**, **48**, **55**, **58**, **95**
- next, **51**
- nlm, **108**, **108**, **111**, **112**

- nlminb, **108**
- noms d'objets
 - conventions, **15**
 - réservés, **16**
- Notepad++, **8**
- nrow, **11, 33, 44, 48, 55, 57, 95**
- NULL, **18, 20**
- NULL (mode), **18**
- numeric, **21, 31, 33, 38, 58, 59, 81**
- numeric (mode), **17, 21**
- optim, **108, 112**
- optimize, **108, 110**
- order, **45, 56**
- outer, **49, 51, 58, 67, 76, 101**
- package, **51**
- paste, **103**
- pgamma, **67**
- plot, **10, 32, 94, 100, 119**
- pmax, **48, 57, 58**
- pmin, **48, 57**
- pnorm, **83**
- point fixe, **68, 75**
- points, **100**
- polyroot, **109, 113**
- print, **53, 54, 58–60, 77, 79, 80, 94, 95, 98, 99**
- prod, **47, 50, 57, 58, 95**
- produit, **47**
 - cumulatif, **48**
 - extérieur, **49**
- q, **8, 102**
- quantile, **47**
- quantile, **47, 57**
- répertoire de travail, **9**
- racine
 - d'un polynôme, **109**
 - d'une fonction, **108**
- rang, **45**
- range, **47, 57**
- rank, **45, 56**
- rbind, **24, 26, 34, 39**
- rbinom, **118**
- Re, **113**
- renverser un vecteur, **45**
- rep, **10, 44, 55, 58, 60, 96, 97**
- repeat, **51, 59, 68, 79, 80, 98**
- répétition de valeurs, **44**
- replace, **37, 57, 119**
- replicate, **93, 97, 98, 134**
- return, **74**
- rev, **45, 56, 57, 65**
- rgamma, **110, 118**
- rm, **11, 101**
- RNGkind, **118**
- rnorm, **10, 98, 102, 118**
- round, **11, 46, 56**
- row.names, **36**
- rowMeans, **49, 61, 88**
- rowSums, **48, 58, 61, 88, 95**
- rpois, **118**
- runif, **10, 11, 118**
- S, **1, 2**
- S+, **1**
- S-PLUS, **1**
- sample, **32, 37, 57, 61, 91, 95–97, 101, 117, 119**
- sapply, **51, 87, 90, 91–93, 96, 97, 101, 130, 132, 155**
- save.image, **4, 8, 128**
- Scheme, **2**
- sd, **47, 57, 98**
- search, **51, 60**
- seq, **10, 30, 36, 44, 55, 60, 96**
- simulation
 - nombres uniformes, **115**
 - planification, **129–136**

- variables aléatoires, 116
- sin, 28
- solve, 11, 49, 57
- somme, 47
 - cumulative, 48
- sort, 45, 55
- source, 134
- start, 79, 80, 98, 100
- stop, 99
- structure, 99
- style, 77
- suite de nombres, 44
- sum, 19, 47, 57, 58, 95, 96, 111
- summary, 47, 57, 94, 100
- switch, 51
- system.time, 81

- T, voir TRUE
- t, 11, 49, 57
- table, 119
- tableau, 61, 88
- tail, 45, 56
- tri, 45
- TRUE, 16, 77
- trunc, 46, 57
- typeof, 17

- unique, 45, 56
- uniroot, 108, 109
- unlist, 26, 36, 96
- upper, 109, 110, 112

- valeur présente, 63, 70–72
- var, 47, 57, 82
- variable
 - globale, 74
 - locale, 74
- variance, 47
- vecteur, 20, 41
- vector, 33, 35
- vide, voir NULL

- which, 45, 56
- which.max, 46, 56
- which.min, 46, 56
- while, 51, 59, 82
- WinEdt, 8

