

R言語

徹底解説

Hadley Wickham 著

石田 基広・市川 太祐・高柳 慎一・福島 真太郎 記

you can't combine factors
r("a"), factor("b"))

共立出版

R言語 徹底解説

Hadley Wickham 著

石田 基広・市川 太祐・高柳 慎一・福島 真太郎 訳

共立出版

R 言語徹底解説

(原題: *Advanced R*)

2016年2月15日 初版1刷発行

検印廃止

NDC007.64.417

ISBN 978-4-320-12393-9

著 者 Hadley Wickham

訳 者 石田 基広・市川 太祐 © 2016
高柳 慎一・福島真太郎

発行者 南條光章

発行所 共立出版株式会社

〒112-0006

東京都文京区小日向4-6-19

電話 03-3947-2511 (代表)

振替口座 00110-2-57035

URL <http://www.kyoritsu-pub.co.jp/>

印 刷 啓文堂

製 本 プロケード



一般社団法人
自然科学書協会
会員

Printed in Japan

JCOPY <出版者著作権管理機構委託出版物>

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつど事前に、出版者著作権管理機構 (TEL: 03-3513-6969, FAX: 03-3513-6979, e-mail: info@jcopy.or.jp) の許諾を得てください。

Advanced R

By Hadley Wickham

© 2015 by Taylor & Francis Group, LLC

CRC Press is an imprint of Taylor & Francis Group, an Informa business

All Rights Reserved.

Authorized translation from English language edition published by
Science, an imprint of Taylor & Francis Group LLC

No claim to original U. S. Government works

Japanese language edition published by KYORITSU SHUPPAN CO., LTD.

目 次

1	導入	1
1.1	本書が想定する読者層	4
1.2	本書から読者が得られるもの	4
1.3	メタテクニック	5
1.4	推奨される文献	6
1.5	助言を得る	7
1.6	謝辞	8
1.7	本書での表記	10
1.8	本書の公開にあたって	10
	第I部 基本編	11
2	データ構造	13
2.1	ベクトル	15
2.1.1	アトムックベクトル	15
2.1.2	リスト	18
2.1.3	エクササイズ	19
2.2	属性	20
2.2.1	因子	22
2.2.2	エクササイズ	25
2.3	行列および配列	25
2.3.1	エクササイズ	28
2.4	データフレーム	29

2.4.1	データフレームの生成	29
2.4.2	データフレームの判定および変換	30
2.4.3	データフレームの結合	31
2.4.4	特殊な列	32
2.4.5	エクササイズ	33
2.5	クイズの解答	33
3	データ抽出	35
3.1	データ抽出の型	36
3.1.1	アトムベクトルからのデータ抽出	37
3.1.2	リストからのデータ抽出	39
3.1.3	行列や配列からのデータ抽出	39
3.1.4	データフレームからのデータ抽出	41
3.1.5	S3 オブジェクトからのデータ抽出	42
3.1.6	S4 オブジェクトからのデータ抽出	42
3.1.7	エクササイズ	42
3.2	データ抽出演算子	43
3.2.1	データ抽出における簡易形式と構造保存形式の違い	44
3.2.2	\$ 演算子	46
3.2.3	欠損/範囲外の添字	47
3.2.4	エクササイズ	48
3.3	データ抽出と付値	48
3.4	応用例	50
3.4.1	ルックアップテーブル（文字列によるデータ抽出）	50
3.4.2	マッチングおよび結合（整数値によるデータ抽出）	50
3.4.3	ランダムサンプリングとブートストラップ（整数値によるデータ抽出）	52
3.4.4	並べ替え（整数値によるデータ抽出）	53
3.4.5	集約されたデータ行を復元する（整数値によるデータ抽出）	54

3.4.6	データフレームから列を削除する（文字列を用いたデータ抽出）	54
3.4.7	条件に応じて行を抽出する（論理値を用いたデータ抽出）	55
3.4.8	ブール代数と集合（論理値および整数値を用いたデータ抽出）	56
3.4.9	エクササイズ	58
3.5	クイズの解答	59
4	ボキャブラリ	61
4.1	基本的な関数群	61
4.2	よく使われるデータ構造	63
4.3	統計学関連	64
4.4	Rを制御する関数群	65
4.5	入出力関連	66
5	コーディングスタイルガイド	67
5.1	表記および命名	68
5.1.1	ファイル名	68
5.1.2	オブジェクト名	68
5.2	文法	69
5.2.1	スペースの入れ方	69
5.2.2	波カッコ	70
5.2.3	一行の長さ	71
5.2.4	インデント	71
5.2.5	付値	71
5.3	コードの構造化	72
5.3.1	コメントについて	72
6	関数	73
6.1	関数の構成要素	75
6.1.1	プリミティブ関数	76

6.1.2	エクササイズ	77
6.2	レキシカルスコープ	77
6.2.1	ネームマスキング	79
6.2.2	関数と変数	80
6.2.3	フレッシュスタート	81
6.2.4	ダイナミックルックアップ	82
6.2.5	エクササイズ	83
6.3	すべての操作は関数呼び出しである	84
6.4	関数の引数	86
6.4.1	関数呼び出し時の引数	87
6.4.2	引数リストによる関数呼び出し	88
6.4.3	デフォルト引数および未指定の引数	89
6.4.4	遅延評価	90
6.4.5	ドット引数(...)	93
6.4.6	エクササイズ	95
6.5	特殊な関数呼び出し	96
6.5.1	二項関数	96
6.5.2	置換関数	97
6.5.3	エクササイズ	100
6.6	返り値 [*]	100
6.6.1	処理を抜ける際の処理	103
6.6.2	エクササイズ	104
6.7	クイズの解答	105
7	オブジェクト指向実践ガイド	107
7.1	基本型	110
7.2	S3	111
7.2.1	オブジェクトと総称関数, メソッドの違い	112
7.2.2	クラス定義とオブジェクト生成	114
7.2.3	総称関数とメソッドの作成	117
7.2.4	メソッドディスパッチ	117

7.2.5	エクササイズ	120
7.3	S4	121
7.3.1	オブジェクトと総称関数, メソッドをそれぞれ見分ける方法	122
7.3.2	クラス定義とオブジェクト生成	123
7.3.3	新しいメソッドと総称関数の設定	126
7.3.4	メソッドディスパッチ	126
7.3.5	エクササイズ	127
7.4	RC	127
7.4.1	クラスの定義とオブジェクトの生成	128
7.4.2	オブジェクトとメソッドの確認	131
7.4.3	メソッドディスパッチ	131
7.4.4	エクササイズ	131
7.5	オブジェクト指向システムの選び方	132
7.6	クイズの解答	133
8	環境	135
8.1	環境の基礎	136
8.1.1	エクササイズ	142
8.2	環境の再帰	142
8.2.1	エクササイズ	145
8.3	関数の環境	145
8.3.1	エンクロージング環境	146
8.3.2	束縛環境	146
8.3.3	実行環境	149
8.3.4	呼び出し環境	151
8.3.5	エクササイズ	153
8.4	名前と値の束縛	154
8.4.1	エクササイズ	157
8.5	明示的環境	157
8.5.1	コピーを避ける	159

8.5.2	パッケージの状態管理	159
8.5.3	ハッシュマップとしての環境	160
8.6	クイズの解答	160
9	デバ깅、条件ハンドリング、防御的プログラミング	161
9.1	デバック技法	164
9.2	デバグのツール	165
9.2.1	呼び出しをたどる	166
9.2.2	エラーをブラウズする	168
9.2.3	任意のコードをブラウズする	170
9.2.4	コールスタック: <code>traceback()</code> , <code>where</code> , <code>recover()</code>	171
9.2.5	他のタイプのエラー	172
9.3	条件ハンドリング	173
9.3.1	<code>try()</code> でエラーを無視する	174
9.3.2	<code>tryCatch()</code> による条件ハンドリング	176
9.3.3	<code>withCallingHandlers()</code>	179
9.3.4	シグナルクラスをカスタマイズする	180
9.3.5	エクササイズ	182
9.4	防御的プログラミング	183
9.4.1	エクササイズ	184
9.5	クイズの解答	185
第II部	関数型プログラミング	187
10	関数型プログラミング	189
10.1	モチベーション	190
10.2	無名関数	196
10.2.1	エクササイズ	197
10.3	クロージャ	198
10.3.1	関数ファクトリ	201
10.3.2	可変な状態	201

10.3.3	エクササイズ	203
10.4	関数のリスト	204
10.4.1	関数のリストをグローバル環境へ移動させる	207
10.4.2	エクササイズ	209
10.5	ケーススタディ：数値積分	209
10.5.1	エクササイズ	213
11	汎関数	215
11.1	初めての汎関数：lapply()	217
11.1.1	ループのパターン	219
11.1.2	エクササイズ	220
11.2	For ループ汎関数：lapply() の仲間たち	221
11.2.1	ベクトル出力：sapply と vapply	222
11.2.2	複数の引数：Map (に加え, mapply)	224
11.2.3	ローリング計算	226
11.2.4	並列化	229
11.2.5	エクササイズ	230
11.3	行列やデータフレームの操作	231
11.3.1	行列と配列の操作	231
11.3.2	グループへの apply() 適用	233
11.3.3	plyr パッケージ	235
11.3.4	エクササイズ	236
11.4	リストの操作	236
11.4.1	Reduce()	236
11.4.2	叙述（プレディケート）汎関数	238
11.4.3	エクササイズ	239
11.5	数学的な汎関数	239
11.5.1	エクササイズ	242
11.6	ループを維持すべき場合	242
11.6.1	即時修正	242
11.6.2	再帰的な関係	243

11.6.3	while ループ	244
11.7	関数族	245
11.7.1	エクササイズ	251
12	関数演算子	253
12.1	挙動に関わる FO	255
12.1.1	メモ化	258
12.1.2	関数呼び出しの捕捉	260
12.1.3	遅延評価	263
12.1.4	エクササイズ	264
12.2	出力に関わる FO	266
12.2.1	軽微な修正	266
12.2.2	関数の動作を変更する	268
12.2.3	エクササイズ	269
12.3	入力に関わる FO	270
12.3.1	あらかじめ決められた関数の引数：部分関数適用	270
12.3.2	引数型の変更	271
12.3.3	エクササイズ	273
12.4	FO を結び付ける	274
12.4.1	関数の合成	274
12.4.2	論理型叙述関数とプール代数	277
12.4.3	エクササイズ	278
第 III 部	言語オブジェクトに対する計算	279
13	非標準評価	281
13.1	表現式の捕捉	283
13.1.1	エクササイズ	284
13.2	subset() における非標準評価	285
13.2.1	エクササイズ	289
13.3	変数のスコープに関する問題	290

13.3.1	エクササイズ	292
13.4	別の関数からの呼び出し	292
13.4.1	エクササイズ	296
13.5	substitute()	297
13.5.1	substitute() へのエスケープハッチの追加	299
13.5.2	未評価の三連ドット(...) の捕捉	300
13.5.3	エクササイズ	301
13.6	非標準評価の欠点	301
13.6.1	エクササイズ	303
14	表現式	305
14.1	表現式の構造	306
14.1.1	エクササイズ	310
14.2	名前	311
14.2.1	エクササイズ	312
14.3	呼び出し	313
14.3.1	呼び出しの修正	314
14.3.2	要素からの呼び出し生成	315
14.3.3	エクササイズ	316
14.4	現在の呼び出しの捕捉	317
14.4.1	エクササイズ	321
14.5	ペアリスト	322
14.5.1	エクササイズ	324
14.6	パーシングとデパーシング	325
14.6.1	エクササイズ	326
14.7	再帰関数を用いた抽象構文木の巡回	327
14.7.1	F と T の探索	328
14.7.2	代入で生成された変数すべての探索	329
14.7.3	呼び出し木の変更	334
14.7.4	エクササイズ	336

15	ドメイン特化言語	339
15.1	HTML	340
15.1.1	本節の目標	342
15.1.2	エスケープ	342
15.1.3	基本的なタグ関数	344
15.1.4	タグ関数	346
15.1.5	すべてのタグの処理	347
15.1.6	エクササイズ	349
15.2	LaTeX	349
15.2.1	LaTeX の数式	350
15.2.2	本節の目標	351
15.2.3	数式への変換	351
15.2.4	既知の記号	351
15.2.5	未知の記号	352
15.2.6	既知の関数	354
15.2.7	未知の関数	356
15.2.8	エクササイズ	358
第IV部	パフォーマンス	359
16	パフォーマンス	361
16.1	R はなぜ遅いか	362
16.2	マイクロベンチマーキング	363
16.2.1	エクササイズ	365
16.3	言語のパフォーマンス	365
16.3.1	究極的な動的特性	366
16.3.2	変更可能な環境を用いた名前の検索	368
16.3.3	遅延評価のオーバーヘッド	370
16.3.4	エクササイズ	371
16.4	実装のパフォーマンス	372
16.4.1	データフレームからの単一の値のデータ抽出	373

16.4.2	<code>ifelse()</code> , <code>pmin()</code> , <code>pmax()</code>	373
16.4.3	エクササイズ	376
16.5	代替の R の実装	376
17	コードの最適化	381
17.1	パフォーマンスの測定	383
17.1.1	制約	387
17.2	パフォーマンスの改善	388
17.3	コードの系統化	389
17.4	誰かがすでにその問題を解決していないか	391
17.4.1	エクササイズ	392
17.5	可能な限り処理を少なくする	392
17.5.1	エクササイズ	399
17.6	ベクトル化	401
17.6.1	エクササイズ	403
17.7	コピーの回避	404
17.8	バイトコードのコンパイル	405
17.9	ケーススタディ:t検定	406
17.10	並列化	409
17.11	その他のテクニック	411
18	メモリ	413
18.1	オブジェクトのサイズ	414
18.1.1	エクササイズ	420
18.2	メモリの使用量とガベージコレクション	420
18.3	lineprof パッケージを用いたメモリプロファイリング	423
18.3.1	エクササイズ	427
18.4	即時修正	428
18.4.1	ループ	431
18.4.2	エクササイズ	433

19 Rcpp パッケージを用いたハイパフォーマンスな関数	435
19.1 C++ を始めよう	438
19.1.1 引数がなく出力がスカラー	438
19.1.2 引数がスカラー・出力がスカラー	439
19.1.3 引数がベクトル・出力がスカラー	440
19.1.4 引数がベクトル・出力がベクトル	442
19.1.5 引数が行列・出力がベクトル	443
19.1.6 <code>sourceCpp()</code> を使用する	444
19.1.7 エクササイズ	446
19.2 属性とその他のクラス	448
19.2.1 リストとデータフレーム	449
19.2.2 関数	450
19.2.3 その他の型	451
19.3 欠損値	451
19.3.1 スカラー	451
19.3.2 文字列	454
19.3.3 論理値	454
19.3.4 ベクトル	454
19.3.5 エクササイズ	455
19.4 Rcpp パッケージのシュガー	456
19.4.1 算術・論理演算	456
19.4.2 論理値の集約関数	457
19.4.3 ベクトルのビュー	458
19.4.4 その他の有益な関数	458
19.5 STL	459
19.5.1 イテレータの使用	459
19.5.2 アルゴリズム	461
19.5.3 データ構造	462
19.5.4 ベクタ	463
19.5.5 セット	464
19.5.6 マップ	466

19.5.7	エクササイズ	466
19.6	ケーススタディ	467
19.6.1	ギブスサンプラー	467
19.6.2	R のベクトル化 vs. C++ のベクトル化	469
19.7	パッケージでの Rcpp パッケージの利用	472
19.8	さらに学ぶために	473
19.9	謝辞	475
20	R と C 言語のインターフェイス	477
20.1	R から C 言語の関数を呼び出す	479
20.2	C 言語でのデータ構造	480
20.3	ベクトルの生成と修正	482
20.3.1	ベクトルの生成とガベージコレクション	482
20.3.2	欠損値や不定値	484
20.3.3	ベクトルデータへのアクセス	486
20.3.4	文字ベクトルとリスト	487
20.3.5	引数の変更	488
20.3.6	スカラーへの変換	489
20.3.7	ロングベクトル	490
20.4	ペアリスト	490
20.5	引数の検証	494
20.6	関数の C 言語ソースを探す方法	495
	訳者あとがき	501
	索 引	505