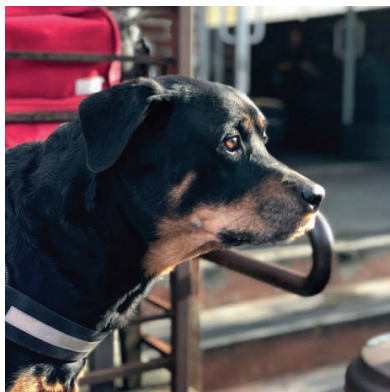


*Посвящается Мине,
лучшему напарнику в написании книги.
Мы будем по тебе скучать*



Содержание

От издательства	19
О переводчике	20
Предисловие	21
 Глава 1 Введение	 24
1.1 Почему R?	24
1.2 Для кого эта книга.....	26
1.3 Что вы узнаете из этой книги	27
1.4 Чего вы не узнаете из этой книги.....	28
1.5 Метатехники	28
1.6 Список рекомендуемой литературы	29
1.7 Помощь в разработке.....	29
1.8 Благодарности.....	30
1.9 Соглашения	34
1.10 Выходные данные	34
 Часть I ОСНОВЫ.....	 36
Введение	37
 Глава 2 Имена и значения	 39
2.1 Введение	39
2.2 Основы связывания	41
2.2.1 Синтаксически неправильные имена	42
2.2.2 Упражнения	43
2.3 Копирование при изменении	44
2.3.1 <code>tracemem()</code>	45
2.3.2 Вызовы функций	45
2.3.3 Списки.....	47
2.3.4 Датафреймы	48
2.3.5 Символьные векторы.....	49
2.3.6 Упражнения	50
2.4 Размеры объектов.....	51
2.4.1 Упражнения	52
2.5 Изменение на месте.....	53
2.5.1 Объекты с единственной привязкой	53
2.5.2 Окружения	56
2.5.3 Упражнения	57
2.6 Отвязывание и сборщик мусора	58
2.7 Ответы на контрольные вопросы	60

Глава 3 Векторы	61
3.1 Введение	61
3.2 Атомарные векторы	63
3.2.1 Скаляры	63
3.2.2 Создание длинных векторов с помощью функции <code>s()</code>	64
3.2.3 Пропущенные значения	65
3.2.4 Определение и приведение типов векторов	66
3.2.5 Упражнения	67
3.3 Атрибуты	67
3.3.1 Получение и установка	68
3.3.2 Имена	69
3.3.3 Размерности	70
3.3.4 Упражнения	71
3.4 Атомарные векторы <code>S3</code>	72
3.4.1 Факторы	73
3.4.2 Даты	75
3.4.3 Даты со временем	75
3.4.4 Длительности	76
3.4.5 Упражнения	77
3.5 Списки	77
3.5.1 Создание	77
3.5.2 Определение и приведение типов	79
3.5.3 Матрицы и массивы	80
3.5.4 Упражнения	80
3.6 Датафреймы и тибблы	80
3.6.1 Создание	82
3.6.2 Имена строк	85
3.6.3 Вывод на экран	86
3.6.4 Извлечение подмножеств	87
3.6.5 Определение и приведение табличных типов	88
3.6.6 Списки в колонках	89
3.6.7 Матрицы и датафреймы в колонках	90
3.6.8 Упражнения	91
3.7 <code>NULL</code>	91
3.8 Ответы на контрольные вопросы	92
Глава 4 Подмножества	93
4.1 Введение	93
4.2 Выбор нескольких элементов	94
4.2.1 Атомарные векторы	94
4.2.2 Списки	97
4.2.3 Матрицы и массивы	97
4.2.4 Датафреймы и тибблы	98
4.2.5 Сохранение размерностей	100
4.2.6 Упражнения	101
4.3 Выбор одного элемента	101
4.3.1 <code>[[</code>	102

4.3.2	\$	103
4.3.3	Отсутствующие и выходящие за границы индексы	104
4.3.4	@ и slot()	105
4.3.5	Упражнения	105
4.4	Извлечение множеств и присваивание	106
4.5	Применение	107
4.5.1	Таблицы поиска (символьное извлечение подмножеств)	107
4.5.2	Сопоставление и объединение в ручном режиме (целочисленное извлечение подмножеств)	107
4.5.3	Случайные выборки и бутстрэпы (целочисленное извлечение подмножеств)	108
4.5.4	Упорядочивание (целочисленное извлечение подмножеств)	109
4.5.5	Разворачивание агрегированных данных (целочисленное извлечение подмножеств)	110
4.5.6	Удаление столбцов из датафрейма (символьное извлечение подмножеств)	111
4.5.7	Выбор строк по условию (логическое извлечение подмножеств)	111
4.5.8	Булева алгебра против множеств (логическое и целочисленное извлечение подмножеств)	112
4.5.9	Упражнения	114
4.6	Ответы на контрольные вопросы	114
Глава 5	Управляющие структуры	115
5.1	Введение	115
5.2	Выбор	116
5.2.1	Некорректные входные значения	117
5.2.2	Векторизованный if	117
5.2.3	Оператор switch()	118
5.2.4	Упражнения	119
5.3	Циклы	120
5.3.1	Распространенные ловушки	121
5.3.2	Сопутствующие инструменты	122
5.3.3	Упражнения	123
5.4	Ответы на контрольные вопросы	123
Глава 6	Функции	124
6.1	Введение	124
6.2	Основы функций	125
6.2.1	Компоненты функций	126
6.2.2	Примитивные функции	127
6.2.3	Функции первого класса	127
6.2.4	Вызов функций	128
6.2.5	Упражнения	129
6.3	Комбинирование функций	130
6.4	Лексический поиск	131
6.4.1	Маскировка имен	132
6.4.2	Функции против переменных	133

6.4.3	С чистого листа.....	134
6.4.4	Динамический поиск.....	135
6.4.5	Упражнения.....	135
6.5	Ленивые вычисления.....	136
6.5.1	Промисы.....	136
6.5.2	Аргументы по умолчанию.....	137
6.5.3	Пропущенные аргументы.....	138
6.5.4	Упражнения.....	140
6.6	... (точка–точка–точка).....	141
6.6.1	Упражнения.....	143
6.7	Выход из функции.....	144
6.7.1	Явный и неявный возврат.....	144
6.7.2	Невидимые значения.....	145
6.7.3	Ошибки.....	146
6.7.4	Обработчики выхода.....	146
6.7.5	Упражнения.....	148
6.8	Формы записи функций.....	149
6.8.1	Преобразование в префиксную форму записи.....	149
6.8.2	Префиксная форма.....	151
6.8.3	Инфиксная форма.....	152
6.8.4	Замещающая форма.....	153
6.8.5	Особая форма.....	154
6.8.6	Упражнения.....	155
6.9	Ответы на контрольные вопросы.....	156
Глава 7	Окружения.....	157
7.1	Введение.....	157
7.2	Основы окружений.....	158
7.2.1	Основы.....	158
7.2.2	Важные окружения.....	160
7.2.3	Родители.....	161
7.2.4	Присваивание в родительском окружении <<-.....	163
7.2.5	Получение и установка значений.....	163
7.2.6	Продвинутые привязки.....	165
7.2.7	Упражнения.....	166
7.3	Рекурсия по окружениям.....	167
7.3.1	Упражнения.....	169
7.4	Особые окружения.....	169
7.4.1	Окружения пакетов и путь для поиска.....	170
7.4.2	Окружения функций.....	171
7.4.3	Пространства имен.....	172
7.4.4	Окружения выполнения.....	175
7.4.5	Упражнения.....	178
7.5	Стеки вызовов.....	178
7.5.1	Простые стеки вызовов.....	179
7.5.2	Ленивые вычисления.....	180
7.5.3	Фреймы.....	181
7.5.4	Динамический поиск.....	182
7.5.5	Упражнения.....	182

7.6	Окружения как структуры данных	182
7.7	Ответы на контрольные вопросы	183

Глава 8 Состояния 184

8.1	Введение	184
8.1.1	Требования	186
8.2	Сигнализирование о состояниях	186
8.2.1	Ошибки	187
8.2.2	Предупреждения	188
8.2.3	Сообщения	189
8.2.4	Упражнения	191
8.3	Игнорирование состояний	191
8.4	Обработка состояний	193
8.4.1	Объекты состояний	194
8.4.2	Обработчики выхода	195
8.4.3	Обработчики вызова	196
8.4.4	Стеки вызовов	199
8.4.5	Упражнения	200
8.5	Пользовательские состояния	201
8.5.1	Предпосылки	202
8.5.2	Сигнализирование	202
8.5.3	Обработка	204
8.5.4	Упражнения	205
8.6	Практическое применение	205
8.6.1	Значение ошибки	205
8.6.2	Значения успеха и ошибки	206
8.6.3	Повторное сигнализирование	208
8.6.4	Запись состояний	208
8.6.5	Отсутствие поведения по умолчанию	210
8.6.6	Упражнения	212
8.7	Ответы на контрольные вопросы	213

Часть II Функциональное программирование 214

Введение 215

Глава 9 Функционалы 218

9.1	Введение	218
9.2	Мой первый функционал: <code>map()</code>	220
9.2.1	Создание атомарных векторов	221
9.2.2	Анонимные функции и сокращенная запись	223
9.2.3	Передача дополнительных аргументов	225
9.2.4	Имена аргументов	227
9.2.5	Изменение другого аргумента	228
9.2.6	Упражнения	229
9.3	Стиль <code>pipe</code>	230
9.4	Разновидности функции <code>map</code>	232
9.4.1	Вход и выход одного типа: <code>modify()</code>	232

9.4.2	Два входа: функция <code>map2()</code> и другие	233
9.4.3	Никакого вывода: функция <code>walk()</code> и другие	236
9.4.4	Итерации по значениям и индексам	239
9.4.5	Любое количество входов: функция <code>map()</code> и другие	239
9.4.6	Упражнения	242
9.5	Семейство функций <code>reduce()</code>	243
9.5.1	Основы	243
9.5.2	<code>Accumulate</code>	245
9.5.3	Выходные типы	246
9.5.4	Множественные входы	247
9.5.5	<code>Map-reduce</code>	248
9.6	Предикаты-функционалы	249
9.6.1	Основы	249
9.6.2	Вариации функции <code>map()</code>	250
9.6.3	Упражнения	250
9.7	Функционалы базового R	251
9.7.1	Матрицы и массивы	251
9.7.2	Математическое применение	253
9.7.3	Упражнения	253
Глава 10	Фабрики функций	254
10.1	Введение	254
10.2	Основы фабрик функций	256
10.2.1	Окружения	256
10.2.2	Обозначения на диаграммах	257
10.2.3	Форсирование вычислений	258
10.2.4	Функции с отслеживанием состояния	259
10.2.5	Сборка мусора	261
10.2.6	Упражнения	261
10.3	Графические фабрики функций	263
10.3.1	Метки	263
10.3.2	Столбики на гистограммах	264
10.3.3	<code>ggsave()</code>	266
10.3.4	Упражнения	267
10.4	Статистические фабрики функций	267
10.4.1	Преобразование Бокса-Кокса	268
10.4.2	Генераторы повторных выборок по методу бутстрэпа	269
10.4.3	Оценка максимального правдоподобия	270
10.4.4	Упражнения	273
10.5	Фабрики функций + функционалы	274
10.5.1	Упражнения	275
Глава 11	Функциональные операторы	277
11.1	Введение	277
11.2	Существующие функциональные операторы	278
11.2.1	Перехват ошибок с помощью <code>purrr::safely()</code>	278
11.2.2	Кеширование вычислений с помощью функции <code>memoise::memoise()</code>	282

11.2.3	Упражнения	284
11.3	Пример для разбора: создание собственных функциональных операторов	284
11.3.1	Упражнения	286
Часть III Объектно ориентированное программирование.....		288
Введение		289
Глава 12 Базовые типы		294
12.1	Введение	294
12.2	Базовые объекты против объектов ООП	295
12.3	Базовые типы	295
12.3.1	Числовой тип	297
Глава 13 Система S3.....		299
13.1	Введение	299
13.2	Основы.....	300
13.2.1	Упражнения	303
13.3	Классы.....	304
13.3.1	Конструкторы	305
13.3.2	Валидаторы.....	307
13.3.3	Помощники	308
13.3.4	Упражнения	310
13.4	Обобщенные функции и методы	310
13.4.1	Диспетчеризация методов	311
13.4.2	Поиск методов.....	312
13.4.3	Создание методов	312
13.4.4	Упражнения	313
13.5	Стили объектов	314
13.5.1	Упражнения	315
13.6	Наследование	315
13.6.1	NextMethod().....	317
13.6.2	Разрешение создавать подклассы.....	318
13.6.3	Упражнения	320
13.7	Детали диспетчеризации методов	320
13.7.1	S3 и базовые типы	321
13.7.2	Внутренние обобщенные функции	322
13.7.3	Групповые обобщенные функции	322
13.7.4	Двойная диспетчеризация методов	323
13.7.5	Упражнения	324
Глава 14 Система R6		325
14.1	Введение	325
14.2	Классы и методы	326
14.2.1	Цепочки методов	327
14.2.2	Важные методы	328
14.2.3	Добавление методов после создания класса.....	330

14.2.4	Наследование	330
14.2.5	Интроспекция	331
14.2.6	Упражнения	331
14.3	Управление доступом	332
14.3.1	Приватность	332
14.3.2	Активные поля	333
14.3.3	Упражнения	335
14.4	Ссылочная семантика	335
14.4.1	Осмысление кода	336
14.4.2	Финализатор	337
14.4.3	Поля R6	338
14.4.4	Упражнения	339
14.5	Почему R6?	339
Глава 15	Система S4	341
15.1	Введение	341
15.2	Основы	343
15.2.1	Упражнения	344
15.3	Классы	344
15.3.1	Наследование	345
15.3.2	Интроспекция	346
15.3.3	Переопределение	346
15.3.4	Помощники	347
15.3.5	Валидаторы	347
15.3.6	Упражнения	349
15.4	Обобщенные функции и методы	349
15.4.1	Аргумент signature	350
15.4.2	Методы	350
15.4.3	Метод show	350
15.4.4	Функции доступа	351
15.4.5	Упражнения	352
15.5	Диспетчеризация методов	352
15.5.1	Простая диспетчеризация	353
15.5.2	Множественное наследование	354
15.5.3	Множественная диспетчеризация	356
15.5.4	Множественная диспетчеризация и множественное наследование	358
15.5.5	Упражнения	358
15.6	S4 и S3	359
15.6.1	Классы	359
15.6.2	Обобщенные функции	360
15.6.3	Упражнения	361
Глава 16	Компромиссы	362
16.1	Введение	362
16.2	S4 против S3	363
16.3	R6 против S3	364
16.3.1	Пространства имен	365

16.3.2	Режим протягивания	366
16.3.3	Сцепление методов.....	369
Часть IV МЕТАПРОГРАММИРОВАНИЕ.....		370
Введение		371
Глава 17 Общая картина.....		373
17.1	Введение	373
17.2	Код как данные.....	374
17.3	Код как дерево	375
17.4	Код может генерировать код.....	376
17.5	Исполнение запускает код	378
17.6	Вычисления с измененными функциями	379
17.7	Вычисления с измененными данными	380
17.8	Quosure	381
Глава 18 Выражения		382
18.1	Введение	382
18.2	Абстрактные синтаксические деревья	383
18.2.1	Внешнее отображение	384
18.2.2	Компоненты без кода.....	385
18.2.3	Инфиксные вызовы	386
18.2.4	Упражнения	387
18.3	Выражения	388
18.3.1	Константы.....	388
18.3.2	Символы	389
18.3.3	Вызовы.....	390
18.3.4	Резюме	393
18.3.5	Упражнения	393
18.4	Разбор и грамматика	394
18.4.1	Приоритет операций	394
18.4.2	Ассоциативность	396
18.4.3	Парсинг и депарсинг.....	397
18.4.4	Упражнения	398
18.5	Проход по AST с помощью рекурсивных функций	399
18.5.1	Находим F и T	401
18.5.2	Находим все переменные, созданные в результате присваивания	402
18.5.3	Упражнения	405
18.6	Специализированные структуры данных	406
18.6.1	Списки пар.....	406
18.6.2	Пропущенные аргументы.....	407
18.6.3	Векторы выражений	408
Глава 19 Квазицитирование		410
19.1	Введение	410
19.2	Предпосылки.....	411

19.2.1	Лексика	413
19.2.2	Упражнения	414
19.3	Цитирование	414
19.3.1	Захват выражений	414
19.3.2	Захват символов	415
19.3.3	Базовый R	416
19.3.4	Подстановка	417
19.3.5	Подведение итогов	417
19.3.6	Упражнения	418
19.4	Расцитирование	419
19.4.1	Расцитирование одного аргумента	419
19.4.2	Расцитирование функции	422
19.4.3	Расцитирование пропущенного аргумента	423
19.4.4	Расцитирование в особых случаях	423
19.4.5	Расцитирование нескольких аргументов	423
19.4.6	Учтливое притворство оператора !!	424
19.4.7	Нестандартные AST	426
19.4.8	Упражнения	427
19.5	Техники отмены цитирования	427
19.6	... (точка-точка-точка)	430
19.6.1	Пример	432
19.6.2	exec()	432
19.6.3	dots_list()	433
19.6.4	В базовом R	434
19.6.5	Упражнения	435
19.7	Практические примеры	436
19.7.1	lobstr::ast()	436
19.7.2	Map-reduce для генерации кода	436
19.7.3	Срезы массива	438
19.7.4	Создание функций	439
19.7.5	Упражнения	441
19.8	История	441
Глава 20	Вычисление	443
20.1	Введение	443
20.2	Основы вычислений	444
20.2.1	Применение: local()	445
20.2.2	Применение: source()	446
20.2.3	Ловушка: function()	447
20.2.4	Упражнения	448
20.3	Структура данных quosure	449
20.3.1	Создание	449
20.3.2	Вычисление	450
20.3.3	Точки	450
20.3.4	Под капотом	451
20.3.5	Вложенные quosure	452
20.3.6	Упражнения	453
20.4	Маски данных	453

20.4.1	Основы	454
20.4.2	Местоимения	454
20.4.3	Применение: subset()	455
20.4.4	Применение: transform()	456
20.4.5	Применение: select()	457
20.4.6	Упражнения	458
20.5	Использование tidy evaluation	459
20.5.1	Цитирование и расцитирование	459
20.5.2	Разрешение неопределенности	460
20.5.3	Цитирование и неопределенность	461
20.5.4	Упражнения	462
20.6	Вычисления в базовом R	462
20.6.1	substitute()	462
20.6.2	match.call()	465
20.6.3	Упражнения	469
Глава 21	Транслирование кода R	470
21.1	Введение	470
21.2	HTML	471
21.2.1	Цель	472
21.2.2	Экранирование	473
21.2.3	Базовые функции тегов	474
21.2.4	Функции тегов	475
21.2.5	Обработка всех тегов	477
21.2.6	Упражнения	479
21.3	LaTeX	480
21.3.1	Математика LaTeX	480
21.3.2	Цель	481
21.3.3	to_math()	481
21.3.4	Известные символы	482
21.3.5	Неизвестные символы	482
21.3.6	Известные функции	484
21.3.7	Неизвестные функции	486
21.3.8	Упражнения	487
Часть V	ТЕХНИКИ	488
Введение		489
Глава 22	Отладка	490
22.1	Введение	490
22.2	Общий подход	491
22.3	Обнаружение ошибок	492
22.3.1	Отложенные вычисления	494
22.4	Интерактивный отладчик	495
22.4.1	Команды browser()	495
22.4.2	Альтернативы	496
22.4.3	Скомпилированный код	498

22.5	Неинтерактивная отладка	498
22.5.1	dump.frames()	499
22.5.2	Вывод отладочной информации	499
22.5.3	RMarkdown	500
22.6	Проблемы, не связанные с ошибками	501
Глава 23	Измерение производительности	503
23.1	Введение	503
23.2	Профилирование	504
23.2.1	Визуализация профилирования	505
23.2.2	Профилирование памяти	507
23.2.3	Ограничения	509
23.2.4	Упражнения	510
23.3	Эталонное микротестирование	510
23.3.1	Результаты bench::mark()	511
23.3.2	Интерпретирование результатов	512
23.3.3	Упражнения	513
Глава 24	Повышение производительности	514
24.1	Введение	514
24.2	Организация кода	515
24.3	Поиск существующих решений	517
24.3.1	Упражнения	517
24.4	Не делайте больше, чем нужно	518
24.4.1	mean()	519
24.4.2	as.data.frame()	520
24.4.3	Упражнения	521
24.5	Векторизация	522
24.5.1	Упражнения	524
24.6	Избежание создания копий	524
24.7	Практический пример: расчет t-критерия Стьюдента	525
24.8	Другие техники	527
Глава 25	Переписывание кода R на C++	529
25.1	Введение	529
25.2	Начинаем работать с C++	531
25.2.1	Нет входа, скалярный выход	531
25.2.2	Скалярный вход, скалярный выход	532
25.2.3	Векторный вход, скалярный выход	533
25.2.4	Векторный вход, векторный выход	534
25.2.5	Использование sourceCpp	535
25.2.6	Упражнения	537
25.3	Другие классы	539
25.3.1	Списки и датафреймы	539
25.3.2	Функции	540
25.3.3	Атрибуты	540
25.4	Пропущенные значения	541

25.4.1	Скаляры	541
25.4.2	Строки	543
25.4.3	Булевы значения	543
25.4.4	Векторы	543
25.4.5	Упражнения	544
25.5	Стандартная библиотека шаблонов	544
25.5.1	Использование итераторов	544
25.5.2	Алгоритмы	546
25.5.3	Структуры данных	547
25.5.4	Векторы	548
25.5.5	Множества	549
25.5.6	Ассоциативные массивы	550
25.5.7	Упражнения	550
25.6	Практические примеры	551
25.6.1	Семплирование по Гиббсу	551
25.6.2	Векторизация в R против векторизации в C++	552
25.7	Использование Rcpp в пакете	554
25.8	Дополнительная литература для изучения	555
25.9	Благодарности	556

Решения и комментарии к упражнениям

Ответы на упражнения из главы 2	557
Ответы на упражнения из главы 3	567
Ответы на упражнения из главы 4	577
Ответы на упражнения из главы 5	581
Ответы на упражнения из главы 6	582
Ответы на упражнения из главы 7	595
Ответы на упражнения из главы 8	602
Ответы на упражнения из главы 9	610
Ответы на упражнения из главы 10	618
Ответы на упражнения из главы 11	629
Ответы на упражнения из главы 13	634
Ответы на упражнения из главы 14	651
Ответы на упражнения из главы 15	659
Ответы на упражнения из главы 18	667
Ответы на упражнения из главы 19	680
Ответы на упражнения из главы 20	688
Ответы на упражнения из главы 21	697
Ответы на упражнения из главы 23	714
Ответы на упражнения из главы 24	717
Ответы на упражнения из главы 25	729

Библиография

Предметный указатель

От издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

О переводчике



Александр Гинько, обладающий богатым опытом работы в сфере ИТ и более десяти лет посвятивший переводам книг и статей на самые разные темы, в последние годы специализируется на переводе книг в области бизнес-аналитики и программирования для издательства «ДМК Пресс» по направлениям Python, R, SQL, Power BI, DAX, Excel, Power Query, Tableau... На данный момент в активе Александра уже более 20 книг, включая одну авторскую, и он продолжает плодотворно работать над переводом

новых книг. В частности, читатели этой книги могут заинтересоваться его переводом книги «Изучаем Shiny» того же автора – Хэдли Уикема. Ознакомиться и приобрести ее можно на странице издательства по адресу <https://dmkpress.com/catalog/computer/statistics/978-5-97060-964-4>.

Помимо перевода книг, Александр ведет свой канал в Telegram (https://t.me/alexanderginko_books), где вы можете из первых уст получить ответы на все интересующие вас вопросы об уже переведенных книгах, находящихся в работе и запланированных на будущее. Также на канале можно найти промокоды на все книги Александра для покупки книг на сайте издательства «ДМК Пресс» с большими скидками.



Илья Шутов (руководитель направления прикладной аналитики данных ООО «Медиа-тел», PhD): «Книга **“R. К вершинам мастерства”** (Advanced R) является именно тем недостающим звеном мостика между аналитиком и аналитиком-программистом. Разница между ними бесконечно велика. Аналитик проводит локальные расчеты и использует R как мощный калькулятор, программист создает законченные приложения и сервисы, включающие в себя в том числе и аналитические блоки.

R, будучи языком общего применения, использует типы данных, которыми оперируют CPU, и опирается на общепринятые концепции по созданию программных конструкций, таких как управление памятью, stringpool, окружения, замыкания, цитирование и многие другие.

Книга Advanced R открывает занавес и позволяет каждому посмотреть на детали и устройство всех этих конструкций. Понять причины того или иного поведения программ, научиться писать эффективный код на R, понять магию non-standard evaluation (NSE) и метапрограммирования. Мечта о свободном полете “как птицы” сопровождает человека всю его историю. Знания, изложенные в этой книге, позволят вам взмыть ввысь, увидеть всю картину целиком и парить легко и непринужденно».

Предисловие

Добро пожаловать в книгу *R. К вершинам мастерства* (Advanced R, 2nd Edition). При написании этого издания (в оригинале второго. – Прим. пер.) я ставил перед собой три главные цели:

- улучшить освещение концепций, важность которых я осознал только после публикации первого издания книги;
- уменьшить количество тем, практичность которых, несмотря на их кажущуюся привлекательность, оказалась под вопросом или полезность которых не прошла проверку временем;
- улучшить читаемость книги за счет упрощения повествования, облегчения кода и добавления более сотни полезных диаграмм.

Одно из важнейших изменений в этом издании книги связано с переходом на новые пакеты, в частности на `rlang` (<http://rlang.r-lib.org>), обеспечивающий прозрачное взаимодействие с низкоуровневыми структурами данных и операциями. В первом издании мы практически полностью пользовались исключительно базовыми функциями языка R, что приводило к определенным сложностям, связанным с их постоянными изменениями в отношении имен и аргументов на протяжении многих лет. Я продолжу по необходимости указывать эквиваленты используемого функционала применительно к базовому пакету R, но если вам важна реализация описываемых концепций исключительно в базовом R, вам лучше будет обратиться к первому изданию этой книги, которое можно найти по адресу <http://adv-r.had.co.nz>.

С момента выхода первого издания базовые концепции языка R не претерпели серьезных изменений, чего не скажешь о моем их понимании. В связи с этим общая структура части под названием *Основы* осталась практически без изменений, в то время как отдельные главы были значительно расширены и улучшены:

- глава 2 – *Имена и значения* – представляет собой абсолютно новую главу, которая поможет вам лучше понять отличия между объектами и их именами. Это понимание позволит вам более точно прогнозировать ситуации, когда R выполняет копирование структур данных, и заложит основы представления о функциональном программировании;
- глава 3 – *Векторы* (ранее она называлась *Структуры данных*) – была полностью переписана с прицелом на векторные типы данных в R, в числе которых – целочисленные значения, факторы и датафреймы. Также в ней мы более детально обсудим важную тему, связанную с векторами S3 (такими как дата и дата со временем), и поговорим о разновидности датафреймов, представленной в пакете `tibble` [Кирилл Мюллер (Kirill Müller) и Хэдли Уикем (Hadley Wickham), 2018]. Кроме того, я поделюсь

своим претерпевшим изменения мнением об использовании векторных типов данных;

- в главе 4 – *Подмножества* – мы поговорим о различиях между функциями `[]` и `[]` по их назначению. Функция `[]` извлекает множество значений, тогда как `[]` – единственное значение (ранее эти операции упоминались как функция с *сохранением формата* (preserved) и *упрощенная* (simplified) соответственно). В разделе 4.3 вы увидите паровозик, который поможет понять, как функция `[]` работает со списками, а также познакомитесь с новыми функциями, обеспечивающими предсказуемое поведение объектов при выходе индексов за допустимые границы;
- глава 5 – *Управляющие структуры* – впервые появилась в этом издании. Как-то я забыл в первом издании упомянуть про `if` и `for`...;
- в главе 6 – *Функции* – был изменен порядок повествования: оператор создания конвейеров (`%>%`) теперь упоминается в качестве третьего варианта вызова последовательности функций (раздел 6.3). Кроме того, большое внимание уделяется формам функций (раздел 6.8);
- в главе 7 – *Окружения* – подвергся изменениям раздел 7.4, посвященный особым окружениям, а также была расширена дискуссия, касающаяся стека вызовов (раздел 7.5);
- в главе 8 – *Состояния* – содержится материал, ранее находившийся в главе *Исключения и отладка*, а также абсолютно новые темы, касающиеся того, как в R работает система состояний. Помимо этого, в разделе 8.5 будет показано, как можно создать собственные классы системы состояний.

Главы после первой части были серьезно реорганизованы вокруг трех важнейших парадигм языка R: функционального программирования, объектно ориентированного программирования и метапрограммирования:

- тему функционального программирования я разбил на три главы по основным техникам: функционалы (глава 9), фабрики функций (глава 10) и функциональные операторы (глава 11). При этом основной упор я сделал на темах, имеющих практическое применение в науке о данных, и сократил количество теоретических отступлений. Теперь в этих главах используются функции из пакета `purrr` [Лайонел Хенри (Lionel Henry) и Хэдли Уикем (Hadley Wickham), 2018], что позволило мне больше сосредоточиться на лежащих в основе описываемых подходов идеях, а не на малозначительных деталях. В результате мне удалось существенно упростить главу, посвященную функциональным операторам;
- во втором издании книги объектно ориентированному программированию (ООП) уделено довольно много внимания. В том числе были добавлены абсолютно новые главы, посвященные базовым типам (глава 12), классам `S3` (глава 13), `S4` (глава 15), `R6` (глава 14) и компромиссам при выборе подходящей системы ООП. В этих главах мы поговорим о том, как работают разные системы в R, а не о том, как использовать их максимально эффективно. Это вынужденная мера, связанная с тем, что многие технические подробности в других местах не описаны, а эффек-

тивному применению объектно ориентированного программирования в R можно посвятить целую книгу;

- метапрограммирование относится к набору инструментов, который может быть использован для создания кода из самого кода. В сравнении с первым изданием материал в этих главах был значительно расширен, теперь он целиком построен на базе фреймворка *tidy evaluation*, представляющего совокупность концепций, облегчающих использование принципов метапрограммирования и делающих его более безопасным. В главе 17 – *Общая картина* – описывается концепция метапрограммирования в широком смысле. В главе 18 – *Выражения* – мы взглянем на структуры данных, лежащие в основе концепции. Глава 19 будет посвящена квазичитированию, а глава 20 – выполнению кода в особых окружениях. В главе 21 – *Транслирование кода R* – мы соберем все полученные знания воедино и узнаем, как можно выполнять транслирование кода из одного языка (программирования) в другой;
- в заключительной части книги я собрал главы, посвященные важным техникам программирования, включая профилирование, измерение и повышение производительности, а также использование пакета Rcpp, обеспечивающего взаимодействие между R и C++. Содержание этих глав практически не изменилось по сравнению с первым изданием книги, но их организация была немного скорректирована. В частности, я поправил текст под использование более новых пакетов (`microbenchmark -> bench`, `lineprof -> profvis`), но по большей части материал остался прежним.

Несмотря на более подробное освещение многих тем во втором издании книги, с пятью главами мне пришлось распрощаться:

- глава со словарем была исключена из книги, поскольку она всегда казалась мне немного странной. Существуют более подходящие способы для демонстрации словарных сведений, не обязательно отводить для них отдельную главу;
- глава, посвященная стилям, перекочевала в онлайн по адресу <https://style.tidyverse.org>. Руководство по стилям также содержит информацию о новом пакете `styler` [Кирилл Мюллер (Kirill Müller) и Лоренц Вальтхерт (Lorenz Walthert), 2018], с помощью которого можно применять различные правила;
- глава, посвященная C, переехала на сайт <https://github.com/hadley/r-internals>, на котором со временем будет собрано полное руководство по написанию кода на C, способного работать со структурами данных R;
- глава, в которой обсуждались вопросы, связанные с памятью, также была исключена из книги. Большинство имеющегося материала было включено в главу 2, а с оставшейся частью, более технического характера, было решено расстаться;
- из книги пропала глава, посвященная производительности языка R. В ней было не так много полезных сведений, и она утратила свою актуальность с развитием R.

1

Введение



Я программирую на языке R вот уже более 15 лет, а последние пять лет делаю это на постоянной основе. Это позволило мне довольно хорошо изучить, как работает данный язык. И книга, которую вы держите в руках, – это моя попытка поделиться опытом и сделать путь постижения R моих читателей максимально легким и безболезненным, насколько это возможно. Чтение ее поможет вам избежать ловушек и тупиков, в которых я провел немало времени, и позволит изучить полезные инструменты и техники для решения большинства возникающих задач. Вместе с тем я попытаюсь показать вам, что, несмотря на некоторые причуды этого языка, идеально подходящего для применения в области науки о данных, он обладает присущей ему одному красотой и элегантностью.

1.1 Почему R?

Если вы только краем уха слышали об R, то можете задаться вопросом, зачем вам стоит изучать столь причудливый язык. Вот лишь несколько ответов на него:

- R – бесплатный язык с открытым исходным кодом, и он доступен на всех известных платформах. Таким образом, при выполнении анализа в языке R вы можете быть уверены, что любой сможет воспроизвести полученные вами результаты вне зависимости от места обитания и уровня достатка;
- язык R славится своим многогранным и дружелюбным сообществом как в онлайне (например, #rstats в твиттере – <https://twitter.com/search?q=%23rstats>), так и в офлайне (планирование митапов по адресу <https://www.meetup.com/topics/r-programming-language>). Также стоит упомянуть об интереснейшей еженедельной новостной рассылке <https://rweekly.org> и отдельном сообществе R-Ladies по адресу <https://rladies.org>, в котором представительницы прекрасного пола могут с легкостью найти своих единомышленниц;
- для языка R написано множество пакетов в области статистического моделирования, машинного обучения, визуализации, импорта и манипулирования данными. Какую бы модель или график вы ни задумали

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru