

Содержание

Об авторах	19
Об изображении на обложке	19
Предисловие	21
О чем эта книга	21
Как организована книга	23
О чем не будет говориться в книге	24
Большие данные	25
Python, Julia и прочее	26
Непрямоугольные данные	26
Подтверждение гипотез	27
Что необходимо для работы с книгой	28
R	28
RStudio	28
Библиотека tidyverse	29
Другие пакеты	30
Выполнение R-кода	31
Получение справочной и дополнительной информации	32
Электронная версия книги	34
Соглашения, принятые в книге	35
Использование примеров кода	36
Ждем ваших отзывов!	37
Часть I. Предварительный анализ	39
Глава 1. Визуализация данных с помощью ggplot2	41
Введение	41
Используемые ресурсы	41
Первые шаги	42
Фрейм данных mpg	43
Создание графика	43

Шаблон команды создания графика	45
Упражнения	45
Эстетика визуализации	46
Упражнения	52
Распространенные ошибки	53
Панели	54
Упражнения	56
Геометрические объекты	57
Упражнения	62
Статистические преобразования	64
Упражнения	69
Позиционные настройки	70
Упражнения	76
Системы координат	76
Упражнения	78
Многослойная грамматика графики	79
Глава 2. Рабочий процесс: основы	83
Основы написания кода	83
Что представляют собой имена	84
Вызов функций	85
Упражнения	87
Глава 3. Преобразование данных с помощью пакета dplyr	89
Введение	89
Используемые ресурсы	89
Пакет nycflights13	90
Основные сведения о пакете dplyr	91
Фильтрация строк с помощью функции filter ()	92
Сравнения	93
Логические операторы	93
Отсутствующие значения	95
Упражнения	96
Перестановка строк с помощью функции arrange ()	97
Упражнения	98
Выбор столбцов с помощью функции select ()	99
Упражнения	101
Добавление новых переменных с помощью функции mutate ()	102

Полезные функции создания объектов	103
Упражнения	106
Получение групповых итогов с помощью функции <code>summarize()</code>	107
Объединение нескольких операций с помощью канала	108
Отсутствующие значения	110
Счетчики	111
Полезные итоговые функции	116
Группирование по нескольким переменным	121
Разгруппирование	122
Упражнения	123
Групповое видоизменение (и фильтрация)	124
Упражнения	125
Глава 4. Организация рабочего процесса: скрипты	127
Выполнение кода	128
Диагностические средства RStudio	129
Упражнения	130
Глава 5. Предварительный анализ данных	131
Введение	131
Необходимые ресурсы	131
Вопросы	132
Вариация	133
Визуализация распределений	134
Типичные значения	137
Необычные значения	140
Упражнения	142
Отсутствующие значения	143
Упражнения	145
Ковариация	145
Категориальные и непрерывные переменные	146
Упражнения	152
Две категориальные переменные	153
Упражнения	155
Две непрерывные переменные	155
Упражнения	159
Шаблоны поведения переменных и модели	159
Вызовы функций <code>ggplot2</code>	163
Дополнительная информация	164

Глава 6. Организация рабочего процесса: проекты	165
Что является вашим “достоянием”	165
Где находится ваш анализ	167
Пути и каталоги	168
Проекты RStudio	168
Резюме	171
Часть II. Подготовка данных	173
Глава 7. Создание tibble-фреймов с помощью пакета tibble	175
Введение	175
Необходимые ресурсы	175
Создание tibble-фреймов	176
Сравнение tibble-фреймов с фреймами <code>data.frame</code>	178
Вывод на печать	178
Извлечение поднаборов	179
Взаимодействие с разработанным ранее кодом	180
Упражнения	180
Глава 8. Импорт данных с помощью пакета readr	183
Введение	183
Необходимые ресурсы	183
Приступаем к работе	183
Сравнение с базовыми возможностями R	186
Упражнения	187
Синтаксический анализ векторов	188
Числа	190
Строки	192
Факторы	193
Календарные даты, календарные дата и время, время суток	194
Упражнения	196
Синтаксический анализ файлов	197
Стратегия	197
Проблемы	198
Другие стратегии	201
Запись в файл	203
Другие типы данных	205

Глава 9. Аккуратизация данных с помощью пакета <code>tidyr</code>	207
Введение	207
Необходимые ресурсы	208
Аккуратные данные	208
Упражнения	211
Распределение и сведение столбцов	212
Сведение столбцов	213
Распределение столбцов	215
Упражнения	216
Разделение и объединение столбцов	217
Разделение столбцов	218
Объединение столбцов	220
Упражнения	222
Отсутствующие значения	222
Упражнения	224
Учебный пример	225
Упражнения	230
Неаккуратные данные	230
Глава 10. Работа с реляционными данными с помощью пакета <code>dplyr</code>	233
Введение	233
Необходимые ресурсы	234
Пакет <code>nycflights13</code>	234
Упражнения	237
Ключи	237
Упражнения	240
Мутирующие соединения	240
Что представляют собой соединения	242
Внутреннее соединение	243
Внешние соединения	244
Неуникальные ключи	246
Определение ключевых столбцов	248
Упражнения	250
Другие реализации	251
Фильтрующие соединения	252
Упражнения	255
Проблемы соединений	255
Операции над множествами	257

Глава 11. Работа со строками с помощью пакета <code>stringr</code>	259
Введение	259
Необходимые ресурсы	259
Основы работы со строками	259
Длина строки	261
Объединение строк	261
Извлечение подстрок	262
Локали	263
Упражнения	264
Поиск соответствия шаблонам с помощью регулярных выражений	265
Базовые соответствия	265
Упражнения	267
Якоря	267
Упражнения	269
Символьные классы и чередования	269
Упражнения	270
Повторения	270
Упражнения	272
Группирование и обратные ссылки	273
Упражнения	273
Инструментарий	274
Обнаружение совпадений	276
Упражнения	278
Извлечение совпадений	279
Упражнения	281
Разбиение совпадений на группы	281
Упражнения	282
Замена совпадений	283
Упражнения	283
Разбиение строк	283
Упражнения	285
Поиск совпадений	285
Другие типы шаблонов	286
Упражнения	289
Другие применения регулярных выражений	289
Пакет <code>stringi</code>	290
Упражнения	290

Глава 12. Работа с факторами с помощью пакета forcats	291
Введение	291
Необходимые ресурсы	291
Создание факторов	292
Опрос GSS	293
Упражнения	296
Изменение порядка следования факторов	296
Упражнения	302
Изменение уровней факторов	302
Упражнения	305
Глава 13. Работа с датами и временем с помощью пакета lubridate	307
Введение	307
Необходимые ресурсы	308
Создание переменных, описывающих дату и время	308
Создание значений даты/времени на основе строк	309
Создание значений даты/времени на основе отдельных компонентов	310
Создание значений даты/времени на основе других типов	313
Упражнения	314
Компоненты даты/времени	314
Получение компонентов	314
Округление	318
Задание значений компонентов	318
Упражнения	319
Временные промежутки	320
Длительности	321
Периоды	322
Интервалы	324
Резюме	325
Упражнения	326
Часовые пояса	326
Часть III. Программирование	329
Глава 14. Работа с каналами с помощью пакета magrittr	333
Введение	333
Необходимые ресурсы	333
Альтернативы каналам	333
Промежуточные объекты	334

Замена исходного объекта	336
Композиция функций	336
Использование канала	337
Когда не следует использовать канал	338
Другие инструменты, доступные в пакете <code>magrittr</code>	339
Глава 15. Функции	343
Введение	343
Необходимые ресурсы	344
Когда следует писать функции	344
Упражнения	347
Функции создаются для компьютеров и людей	348
Упражнения	351
Условное выполнение	351
Условия	352
Множественные условия	353
Стиль кода	354
Упражнение	355
Аргументы функций	356
Выбор имен	358
Проверка значений	358
Точка-точка-точка (...)	360
Ленивые вычисления	361
Упражнения	361
Возвращаемые значения	362
Явные команды возврата	362
Написание функций, пригодных для включения в канал	363
Окружение	364
Глава 16. Векторы	367
Введение	367
Необходимые ресурсы	367
Основные сведения о векторах	368
Важные типы атомарных векторов	370
Логический тип	370
Числовой тип	370
Символьный тип	371
Отсутствующие значения	372
Упражнения	372

Использование атомарных векторов	373
Приведение типов	373
Функции проверки типов	375
Скаляры и правила заикливания	375
Именованые векторы	377
Извлечение элементов	377
Упражнения	379
Рекурсивные векторы (списки)	380
Визуализация списков	381
Извлечение элементов	382
Список специй	383
Упражнения	385
Атрибуты	385
Расширенные векторы	387
Факторы	388
Дата и дата/время	388
Tibble-фреймы	389
Упражнения	390
Глава 17. Итерирование с помощью пакета purrr	391
Введение	391
Необходимые ресурсы	392
Циклы for	392
Упражнения	394
Варианты цикла for	395
Изменение существующего объекта	396
Шаблоны формирования циклов	397
Выходные данные неизвестной длины	398
Последовательность неизвестной длины	399
Упражнения	400
Циклы for и функционалы	401
Упражнения	404
Функции семейства map	404
Сокращения	406
Базовые возможности R	407
Упражнения	408
Обработка ошибок	409
Функции семейства map с несколькими входными переменными	412
Вызов различных функций	415

Функции семейства <code>walk</code>	416
Другие шаблоны циклов <code>for</code>	417
Функции-предикаты	417
Функции <code>reduce</code> и <code>accumulate</code>	418
Упражнения	420
Часть IV. Моделирование	421
Глава 18. Базовое моделирование с помощью пакета <code>modelr</code>	425
Введение	425
Необходимые ресурсы	426
Простая модель	427
Упражнения	435
Визуализация моделей	436
Предсказания	437
Остатки	439
Упражнения	441
Формулы и семейства моделей	441
Категориальные переменные	442
Взаимодействия (непрерывная и категориальная переменные)	445
Взаимодействия (две непрерывные переменные)	449
Преобразования	453
Упражнения	456
Отсутствующие значения	457
Другие семейства моделей	457
Глава 19. Создание модели	459
Введение	459
Необходимые ресурсы	460
Почему бриллианты низкого качества оказались более дорогими	460
Цена и вес	462
Более сложная модель	467
Упражнения	470
Что влияет на количество ежедневных авиарейсов	470
День недели	471
Сезонный субботний эффект	475
Вычисляемые переменные	481
Времена года: альтернативный подход	482
Упражнения	483
Дополнительная информация о моделях	484

Глава 20. Обработка множества моделей с помощью пакетов purrr и broom	487
Введение	487
Необходимые ресурсы	488
Набор данных <code>garminder</code>	489
Вложенные данные	491
Столбец-список	493
Отмена вложения данных	494
Качество модели	497
Упражнения	499
Столбцы-списки	500
Создание столбцов-списков	502
Создание столбцов-списков на основе вложений	503
Создание столбцов-списков на основе векторизованных функций	504
Создание столбцов-списков на основе многозначных итогов	505
Создание столбцов-списков на основе именованных списков	506
Упражнения	507
Упрощение столбцов-списков	508
Преобразование списка в вектор	509
Отмена вложения	510
Упражнения	511
Приведение данных к аккуратной форме с помощью пакета <code>broom</code>	511
Часть V. Коммуникация	513
Глава 21. Язык разметки R Markdown	515
Введение	515
Необходимые ресурсы	516
Основы R Markdown	516
Упражнения	519
Форматирование текста с помощью Markdown	519
Упражнение	521
Фрагменты кода	521
Имя фрагмента	522
Опции фрагмента	523
Таблица	524
Кеширование	525
Глобальные опции	526

Встроенный код	527
Упражнения	528
Устранение неполадок	528
Заголовок YAML	529
Параметры	529
Библиографические ссылки и цитаты	531
Дополнительная информация	532
Глава 22. Обмен графикой с помощью пакета ggplot2	533
Введение	533
Необходимые ресурсы	534
Подписи	534
Упражнения	537
Аннотации	538
Упражнения	545
Шкалы	545
Метки на осях и ключи легенды	546
Компоновка легенды	549
Замена шкалы	550
Упражнения	556
Изменение масштаба изображения	556
Темы	558
Сохранение графиков	560
Изменение размера рисунка	561
Другие важные опции	564
Дополнительная информация	565
Глава 23. Выходные форматы R Markdown	567
Введение	567
Опции выходных форматов	568
Документы	569
Блокноты	570
Презентации	571
Демонстрационные панели	571
Интерактивность	573
Виджеты	573
Пакет Shiny	575

Веб-сайты	576
Другие форматы	577
Дополнительная информация	577
Глава 24. Организация работы с R Markdown	579
Предметный указатель	583