

Data Science : programowanie, analiza i wizualizacja danych z wykorzystaniem języka R / Michael Freeman, Joel Ross. – Gliwice, copyright © 2020

Spis treści

Przedmowa	11
Wprowadzenie	13
O autorach	19
CZĘŚĆ I Wprowadzenie	21
Rozdział 1 Przygotowywanie komputera	23
1.1. Przygotowywanie narzędzi używanych w wierszu poleceń	24
1.1.1. Wiersz poleceń w systemie macOS	24
1.1.2. Wiersz poleceń w systemie Windows	25
1.1.3. Wiersz poleceń w systemie Linux	25
1.2. Instalowanie systemu git	25
1.3. Tworzenie konta w serwisie GitHub	26
1.4. Wybieranie edytora tekstu	26
1.4.1. Atom	26
1.4.2. Visual Studio Code	27
1.4.3. Sublime Text	27
1.5. Pobieranie języka R	28
1.6. Pobieranie środowiska RStudio	28
Rozdział 2 Używanie wiersza poleceń	31
2.1. Uruchamianie wiersza poleceń	31
2.2. Poruszanie się w systemie plików	32
2.2.1. Przechodzenie do innego katalogu	33
2.2.2. Wyświetlanie listy plików	35
2.2.3. Ścieżki	35
2.3. Zarządzanie plikami	37
2.3.1. Uczenie się nowych instrukcji	37
2.3.2. Symbole wieloznaczne	39
2.4. Radzenie sobie z błędami	40
2.5. Przekierowywanie danych wyjściowych	41
2.6. Polecenia związane z siecią	42
CZĘŚĆ II Zarządzanie projektami	45
Rozdział 3 Kontrola wersji z użyciem systemu git i serwisu GitHub	47
3.1. Czym jest git?	47

3.1.1. Podstawowe zagadnienia związane z systemem git	48
3.1.2. Czym jest GitHub?	49
3.2. Konfigurowanie narzędzi i tworzenie projektu	50
3.2.1. Tworzenie repozytorium	51
3.2.2. Sprawdzanie stanu	51
3.3. Śledzenie zmian w projekcie	52
3.3.1. Dodawanie plików	53
3.3.2. Zatwierdzanie	54
3.3.3. Proces używania systemu git	55
3.4. Zapisywanie projektów w witrynie GitHub	56
3.4.1. Forki i klonowanie	57
3.4.2. Wysyłanie i pobieranie	59
3.5. Dostęp do historii projektu	60
3.5.1. Historia rewizji	60
3.5.2. Powrót do starszych wersji	61
3.6. Ignorowanie plików w projekcie	62
Rozdział 4 Tworzenie dokumentacji za pomocą języka Markdown	65
4.1. Pisanie kodu w języku Markdown	65
4.1.1. Formatowanie tekstu	66
4.1.2. Bloki tekstu	66
4.1.3. Hiperłącza	67
4.1.4. Rysunki	67
4.1.5. Tabele	68
4.2. Wyświetlanie dokumentów w języku Markdown	68
CZĘŚĆ III Podstawowe umiejętności z zakresu języka R	71
Rozdział 5 Wprowadzenie do języka R	73
5.1. Programowanie z użyciem języka R	73
5.2. Uruchamianie kodu w języku R	74
5.2.1. Używanie środowiska RStudio	74
5.2.2. Używanie języka R w wierszu poleceń	76
5.3. Dodawanie komentarzy	78
5.4. Definiowanie zmiennych	78
5.4.1. Podstawowe typy danych	80
5.5. Szukanie pomocy	83
5.5.1. Nauka uczenia się języka R	84
Rozdział 6 Funkcje	89
6.1. Czym jest funkcja?	89
6.1.1. Składnia funkcji w języku R	90
6.2. Wbudowane funkcje języka R	91
6.2.1. Argumenty nazwane	92
6.3. Wczytywanie funkcji	93
6.4. Pisanie funkcji	95
6.4.1. Debugowanie funkcji	97

6.5. Instrukcje warunkowe	98
Rozdział 7 Wektory	101
7.1. Czym jest wektor?	101
7.1.1. Tworzenie wektorów	101
7.2. Operacje wektorowe	102
7.2.1. Ponowne używanie elementów	104
7.2.2. Prawie wszystko jest wektorem	105
7.2.3. Funkcje wektorowe	105
7.3. Indeksy w wektorach	107
7.3.1. Listy indeksów	108
7.4. Filtrowanie wektorów	109
7.5. Modyfikowanie wektorów	110
Rozdział 8 Listy	113
8.1. Czym jest lista?	113
8.2. Tworzenie list	114
8.3. Dostęp do elementów listy	115
8.4. Modyfikowanie list	117
8.4.1. Pojedyncze i podwójne nawiasy kwadratowe	118
8.5. Stosowanie funkcji do list za pomocą wywołania lapply()	119
CZĘŚĆ IV Przekształcanie danych	121
Rozdział 9 Jak zrozumieć dane?	123
9.1. Proces generowania danych	123
9.2. Wyszukiwanie danych	124
9.3. Rodzaje danych	126
9.3.1. Skale pomiarowe	126
9.3.2. Struktury danych	127
9.4. Interpretowanie danych	129
9.4.1. Zdobywanie wiedzy w danej dziedzinie	129
9.4.2. Jak zrozumieć schematy danych?	131
9.5. Odpowiadanie na pytania na podstawie danych	133
Rozdział 10 Ramki danych	135
10.1. Czym jest ramka danych?	135
10.2. Praca z ramkami danych	136
10.2.1. Tworzenie ramek danych	136
10.2.2. Opisywanie struktury ramek danych	137
10.2.3. Dostęp do ramek danych	138
10.3. Praca z danymi CSV	139
10.3.1. Katalog roboczy	140
10.3.2. Zmienne w postaci faktorów	142
Rozdział 11 Operowanie danymi za pomocą pakietu dplyr	145
11.1. Gramatyka operowania danymi	145

11.2. Podstawowe funkcje pakietu dplyr	146
11.2.1. Pobieranie (funkcja selekt())	147
11.2.2. Filtrowanie (funkcja filter())	149
11.2.3. Dodawanie kolumn (funkcja mutate())	150
11.2.4. Sortowanie danych (funkcja arrange())	151
11.2.5. Tworzenie podsumowań (funkcja summarize())	152
11.3. Wykonywanie operacji sekwencyjnych	153
11.3.1. Operator potoku	154
11.4. Analizowanie ramek danych z wykorzystaniem grupowania	155
11.5. Złączanie ramek danych	157
11.6. Pakiet dplyr w praktyce — analizowanie danych na temat lotów	162
Rozdział 12 Porządkowanie danych za pomocą pakietu tidyr	169
12.1. Czym jest porządkowanie danych?	169
12.2. Od kolumn do wierszy — gather()	171
12.3. Z wierszy na kolumny — spread()	172
12.4. Pakiet tidyr w praktyce — eksplorowanie statystyk na temat edukacji	174
Rozdział 13 Dostęp do bazy danych	181
13.1. Przegląd relacyjnych baz danych	181
13.1.1. Czym jest relacyjna baza danych?	182
13.1.2. Tworzenie relacyjnej bazy danych	184
13.2. Wstęp do języka SQL	185
13.3. Dostęp do bazy danych w języku R	189
Rozdział 14 Używanie internetowych interfejsów API	193
14.1. Czym jest internetowy interfejs API?	193
14.2. Żądania REST	194
14.2.1. Identyfikatory URI	194
14.2.2. Operacje (czasowniki) z protokołu HTTP	201
14.3. Dostęp do internetowych interfejsów API w R	201
14.4. Przetwarzanie danych w formacie JSON	203
14.4.1. Przetwarzanie danych w formacie JSON	205
14.4.2. Spłaszczanie danych	207
14.5. Interfejsy API w praktyce — znajdowanie kubańskiego jedzenia w Seattle	209
CZĘŚĆ V Wizualizacje danych	215
Rozdział 15 Projektowanie wizualizacji danych	217
15.1. Cel wizualizacji	217
15.2. Wybieranie układu graficznego	219
15.2.1. Wizualizowanie jednej zmiennej	220
15.2.2. Wizualizowanie wielu zmiennych	223
15.2.3. Wizualizowanie danych hierarchicznych	227
15.3. Wybieranie skutecznego kodowania graficznego	229

15.3.1. Skuteczne kolory	231
15.3.2. Wykorzystanie atrybutów przed u wagowych	234
15.4. Ekspresywne prezentacje danych	236
15.5. Zwiększanie estetyki	238
Rozdział 16 Tworzenie wizualizacji za pomocą pakietu ggplot2	241
16.1. Gramatyka grafiki	241
16.2. Tworzenie podstawowych wykresów za pomocą ggplot2	242
16.2.1. Określanie obiektów geometrycznych	245
16.2.2. Odzworowania aspektów estetycznych	247
16.3. Złożone układy i dostosowywanie opcji	248
16.3.1. Dostosowywanie pozycji	248
16.3.2. Zmienianie stylu za pomocą skal	250
16.3.3. Układ współrzędnych	253
16.3.4. Fasety	254
16.3.5. Etykiety i uwagi	255
16.4. Tworzenie map	257
16.4.1. Kartogramy	258
16.4.2. Mapy punktowe	261
16.5. Pakiet ggplot2 w praktyce — mapa eksmisji w San Francisco	262
Rozdział 17 Interaktywne wizualizacje w języku R	267
17.1. Pakiet plotly	269
17.2. Pakiet rbokeh	271
17.3. Pakiet leaflet	273
17.4. Interaktywne wizualizacje w praktyce — analizowanie zmian w Seattle	276
CZĘŚĆ VI Tworzenie i udostępnianie aplikacji	281
Rozdział 18 Tworzenie dynamicznych raportów za pomocą platformy R Markdown	283
18.1. Konfigurowanie raportu	283
18.1.1. Tworzenie plików .rmd	284
18.1.2. Kompilowanie dokumentów	286
18.2. Integrowanie tekstu w formacie Markdown i kodu w języku R	287
18.2.1. Wykonywalne fragmenty kodu w języku R	287
18.2.2. Kod wewnątrzwerszowy	288
18.3. Wyświetlanie danych i wizualizacji w raportach	289
18.3.1. Wyświetlanie łańcuchów znaków	289
18.3.2. Wyświetlanie list w formacie Markdown	290
18.3.3. Wyświetlanie tabel	291
18.3.4. Wyświetlanie wykresów	292
18.4. Udostępnianie raportów jako stron internetowych	293
18.5. Platforma R Markdown w praktyce — raport na temat oczekiwanej długości życia	295

Rozdział 19 Tworzenie interaktywnych aplikacji internetowych za pomocą platformy Shiny	301
19.1. Platforma Shiny	301
19.1.1. Podstawowe zagadnienia dotyczące platformy Shiny	302
19.1.2. Struktura aplikacji	303
19.2. Projektowanie interfejsów użytkownika	307
19.2.1. Treści statyczne	308
19.2.2. Dynamiczne dane wejściowe	310
19.2.3. Dynamiczne dane wyjściowe	311
19.2.4. Układy	312
19.3. Tworzenie serwerów aplikacji	315
19.4. Publikowanie aplikacji na platformę Shiny	318
19.5. Platforma Shiny w praktyce — wizualizacja śmiertelnych postrzeleń przez policję	320
 Rozdział 20 Praca zespołowa	 327
20.1. Śledzenie różnych wersji kodu za pomocą gałęzi	327
20.1.1. Praca z różnymi gałęziami	329
20.1.2. Scalanie gałęzi	332
20.1.3. Scalanie a konflikty	333
20.1.4. Scalanie w serwisie GitHub	335
20.2. Prowadzenie projektów z użyciem gałęzi funkcji	337
20.3. Współpraca w ramach scentralizowanego procesu pracy	338
20.3.1. Tworzenie centralnego repozytorium	339
20.3.2. Używanie gałęzi funkcji w scentralizowanym procesie pracy	341
20.4. Współpraca w procesie pracy z użyciem forków	342
 Rozdział 21 Dalsza nauka	 347
21.1. Uczenie statystyczne	347
21.1.1. Ocena zależności	347
21.1.2. Prognozowanie	348
21.2. Inne języki programowania	348
21.3. Odpowiedzialność etyczna	349
 Skorowidz	 351