
Spis treści

1. Pierwsze kroki i uzyskiwanie pomocy	17
1.1. Pobranie i instalacja R	18
1.2. Instalacja środowiska RStudio	20
1.3. Uruchamianie środowiska RStudio	21
1.4. Wprowadzanie poleceń	23
1.5. Wyjście ze środowiska RStudio	24
1.6. Przerywanie realizacji kodu R	26
1.7. Przeglądanie dołączonej dokumentacji	27
1.8. Uzyskiwanie pomocy na temat funkcji	28
1.9. Wyszukiwanie dodatkowej dokumentacji	30
1.10. Uzyskiwanie pomocy na temat pakietu	31
1.11. Wyszukiwanie pomocy w internecie	32
1.12. Wyszukiwanie przydatnych funkcji i pakietów	35
1.13. Przeszukiwanie list dyskusyjnych	36
1.14. Przesyłanie pytań do serwisu Stack Overflow lub innego	37
2. Garść podstaw	41
2.1. Wyświetlanie interesujących nas danych na ekranie	41
2.2. Wyznaczanie zmiennych	43
2.3. Tworzenie listy zmiennych	44
2.4. Usuwanie zmiennych	46
2.5. Tworzenie wektorów	47
2.6. Obliczanie podstawowych statystyk	49
2.7. Tworzenie sekwencji	51
2.8. Porównywanie wektorów	52
2.9. Wybieranie elementów wektora	54
2.10. Wykonywanie obliczeń wektorowych	57
2.11. Ustalanie pierwszeństwa operatorów	59
2.12. Osiąganie więcej przy mniejszej liczbie znaków	61
2.13. Tworzenie strumienia wywołań funkcji	62
2.14. Unikanie najpowszechniejszych pomyłek	65

3. Korzystanie z oprogramowania	71
3.1. Sprawdzanie i wyznaczanie katalogu roboczego	71
3.2. Tworzenie nowego projektu RStudio	72
3.3. Zapisywanie przestrzeni roboczej	74
3.4. Przeglądanie historii wpisanych poleceń	75
3.5. Zapisywanie wyniku wcześniejszego polecenia	76
3.6. Wyświetlanie załadowanych pakietów poprzez ścieżkę wyszukiwania	77
3.7. Przeglądanie listy zainstalowanych pakietów	79
3.8. Uzyskiwanie dostępu do funkcji zawartych w pakiecie	80
3.9. Uzyskiwanie dostępu do wbudowanych zestawów danych	81
3.10. Instalowanie pakietów z repozytorium CRAN	82
3.11. Instalowanie pakietu z serwisu GitHub	84
3.12. Wyznaczanie lub zmiana domyślnego serwera CRAN	85
3.13. Uruchamianie skryptu	86
3.14. Uruchamianie skryptu wsadowego	87
3.15. Wyszukiwanie katalogu domowego R	89
3.16. Personalizowanie rozruchu R	91
3.17. Korzystanie z R i RStudio w chmurze	94
 4. Dane wejściowe i wyjściowe	 97
4.1. Wprowadzanie danych za pomocą klawiatury	97
4.2. Wyświetlanie mniejszej (lub większej) liczby znaków	98
4.3. Przekierowywanie wyników do pliku	100
4.4. Wyświetlanie listy plików	101
4.5. Problem z otwieraniem pliku w systemie Windows	103
4.6. Odczytywanie rekordów o stałej szerokości	104
4.7. Odczytywanie plików danych tabelarycznych	107
4.8. Odczytywanie plików CSV	110
4.9. Zapisywanie danych w pliku CSV	112
4.10. Odczytywanie danych tabelarycznych lub CSV z internetu	113
4.11. Odczytywanie danych z arkusza Excel	114
4.12. Zapisywanie ramki danych w pliku Excel	116
4.13. Odczytywanie danych z pliku SAS	118
4.14. Odczytywanie danych z tabel HTML	120
4.15. Odczytywanie plików o skomplikowanej strukturze	122
4.16. Odczyt baz danych MySQL	126
4.17. Uzyskiwanie dostępu do bazy danych za pomocą pakietu dbplyr	129
4.18. Zapisywanie i transportowanie obiektów	131

5. Struktury danych	135
5.1. Dodawanie danych do wektora	142
5.2. Wstawianie danych do wektora	144
5.3. Reguła zawijania	144
5.4. Tworzenie wektora czynnikowego (zmiennej kategorialnej)	146
5.5. Łączenie wielu wektorów w jeden wektor i wektor czynnikowy	147
5.6. Tworzenie listy	149
5.7. Wybieranie elementów listy za pomocą ich pozycji	150
5.8. Wybieranie elementów listy po nazwie	152
5.9. Tworzenie listy asocjacyjnej nazwa/wartość	153
5.10. Usuwanie elementu z listy	155
5.11. Spłaszczanie listy do postaci wektora	156
5.12. Usuwanie elementów o wartości NULL z listy	157
5.13. Warunkowe usuwanie elementów listy	158
5.14. Inicjowanie macierzy	159
5.15. Wykonywanie operacji macierzowych	161
5.16. Nadawanie nazw opisowych rzędom i kolumnom macierzy	162
5.17. Wybór jednego rzędu/kolumny macierzy	163
5.18. Inicjowanie ramki danych z danymi kolumny	164
5.19. Inicjowanie ramki danych z danymi rzędu	165
5.20. Dołączanie rzędów do ramki danych	168
5.21. Wybór kolumn ramki danych za pomocą pozycji	170
5.22. Wybór kolumn ramki danych za pomocą nazwy	174
5.23. Zmienianie nazw kolumn w ramce danych	175
5.24. Usuwanie wartości NA z ramki danych	176
5.25. Wykluczanie kolumn za pomocą nazwy	177
5.26. Łączenie dwóch ramek danych	178
5.27. Scalanie kolumn dwóch ramek danych	179
5.28. Przekształcanie jednej wartości atomowej w inną	181
5.29. Przekształcanie jednego ustrukturyzowanego typu danych w inny	183
6. Przekształcenia danych	187
6.1. Stosowanie funkcji wobec każdego elementu listy	187
6.2. Stosowanie funkcji wobec każdego rzędu ramki danych	190
6.3. Stosowanie funkcji wobec każdego rzędu macierzy	191
6.4. Stosowanie funkcji wobec każdej kolumny	192
6.5. Stosowanie funkcji wobec wektorów równoległych lub list	194
6.6. Stosowanie funkcji wobec grup danych	196
6.7. Tworzenie nowej kolumny na podstawie jakiegoś warunku	197

7. Łańcuchy znaków i daty	199
7.1. Uzyskiwanie długości łańcucha znaków	201
7.2. Łączenie łańcuchów znaków	202
7.3. Wydobywanie fragmentów łańcuchów znaków	203
7.4. Rozdzielanie łańcucha znaków zgodnie z rozgranicznikiem	204
7.5. Zastępowanie fragmentów łańcuchów znaków	205
7.6. Tworzenie wszystkich kombinacji par łańcuchów znaków	206
7.7. Uzyskiwanie bieżącej daty	208
7.8. Przekształcanie łańcucha znaków w obiekt Date	208
7.9. Przekształcanie obiektu Date w łańcuch znaków	209
7.10. Przekształcanie roku, miesiąca i dnia w obiekt Date	210
7.11. Uzyskiwanie daty juliańskiej	211
7.12. Wydobywanie elementów składowych daty	212
7.13. Tworzenie sekwencji dat	213
8. Prawdopodobieństwo	215
8.1. Wyznaczanie liczby kombinacji	217
8.2. Generowanie kombinacji	218
8.3. Generowanie liczb losowych	219
8.4. Generowanie odtwarzalnych liczb losowych	220
8.5. Generowanie próby losowej	222
8.6. Generowanie sekwencji losowych	223
8.7. Losowe permutacje wektora	224
8.8. Obliczanie prawdopodobieństwa rozkładów dyskretnych	225
8.9. Obliczanie prawdopodobieństwa rozkładów ciągłych	226
8.10. Przekształcanie prawdopodobieństw w kwantyle	228
8.11. Tworzenie wykresu funkcji gęstości	229
9. Statystyka ogólna	235
9.1. Podsumowywanie danych	237
9.2. Obliczanie częstości względnych	239
9.3. Zestawianie wektorów czynnikowych w tabeli i tworzenie tablic wielozmiennych	240
9.4. Sprawdzanie niezależności zmiennych kategoryjnych	241
9.5. Obliczanie kwantyli (i kwartyli) zestawu danych	242
9.6. Uzyskiwanie odwrotności kwantyli	243
9.7. Normalizowanie danych	244
9.8. Testowanie średniej próby (test t)	244
9.9. Kształtowanie przedziału ufności dla średniej	246
9.10. Kształtowanie przedziału ufności dla mediany	247
9.11. Testowanie proporcji próby	248
9.12. Kształtowanie przedziału ufności dla proporcji	249

9.13. Testowanie pod względem rozkładu normalnego	250
9.14. Testowanie przebiegów	251
9.15. Porównywanie średnich dwóch prób	252
9.16. Nieparametryczne porównywanie położenia dwóch prób	254
9.17. Testowanie korelacji pod względem istotności	256
9.18. Testowanie grup pod względem równych proporcji	257
9.19. Porównywanie parami średnich poszczególnych grup	259
9.20. Testowanie dwóch prób w kontekście tego samego rozkładu	260
10. Grafika	263
10.1. Tworzenie wykresu punktowego	267
10.2. Wstawianie tytułu i etykiet	267
10.3. Dodawanie (lub usuwanie) siatki	269
10.4. Stosowanie motywu wobec wykresu ggplot	272
10.5. Tworzenie wielogrupowego wykresu punktowego	277
10.6. Dodawanie (lub usuwanie) legendy	278
10.7. Rysowanie linii regresji na wykresie punktowym	282
10.8. Tworzenie wykresów par zmiennych	285
10.9. Tworzenie wykresów punktowych dla poszczególnych grup danych	287
10.10. Tworzenie wykresu kolumnowego	289
10.11. Umieszczanie przedziałów ufności na wykresie kolumnowym	292
10.12. Wprowadzanie kolorów na wykresie kolumnowym	295
10.13. Rysowanie linii łączącej pary punktów x i y	297
10.14. Zmiana rodzaju, szerokości i koloru linii	297
10.15. Tworzenie wykresu zawierającego wiele zestawów danych	301
10.16. Dodawanie linii pionowych lub poziomych	302
10.17. Tworzenie wykresu pudełkowego	304
10.18. Tworzenie po jednym wykresie pudełkowym na każdy poziom wektora czynnikowego	306
10.19. Tworzenie histogramu	308
10.20. Dodawanie oszacowania gęstości do histogramu	310
10.21. Tworzenie standardowego wykresu kwantyl-kwantyl	311
10.22. Tworzenie innych wykresów kwantyl-kwantyl	314
10.23. Rysowanie zmiennej w różnych kolorach	316
10.24. Tworzenie wykresu funkcji	319
10.25. Wyświetlanie wielu wykresów na jednej stronie	321
10.26. Zapisywanie wykresu do pliku	324
11. Regresja liniowa i analiza ANOVA	327
11.1. Przeprowadzanie prostej analizy liniowej	329
11.2. Przeprowadzanie wielorakiej regresji liniowej	331
11.3. Uzyskiwanie statystyk regresji	332

11.4. Omówienie podsumowania regresji	335
11.5. Przeprowadzanie regresji liniowej bez użycia punktu przecięcia z osią współrzędnych	338
11.6. Przeprowadzanie regresji wyłącznie przy użyciu zmiennych ściśle skorelowanych ze zmienną objaśnianą	339
11.7. Przeprowadzanie regresji liniowej z członami interakcyjnymi	342
11.8. Wybór najlepszych zmiennych regresji	344
11.9. Przeprowadzanie regresji na podzbiorze danych	349
11.10. Korzystanie ze wzorów w równaniu regresji	350
11.11. Przeprowadzanie regresji względem wielomianu	351
11.12. Regresja względem przekształconych danych	353
11.13. Wyszukiwanie najlepszego przekształcenia potęgowego (procedura Boxa-Coxa)	355
11.14. Kształtowanie przedziałów ufności dla współczynników regresji	359
11.15. Tworzenie wykresu elementów resztowych regresji	360
11.16. Diagnozowanie regresji liniowej	361
11.17. Wykrywanie najbardziej znaczących obserwacji	364
11.18. Testowanie wartości resztowych pod względem autokorelacji (test Durbina-Watsona)	366
11.19. Przewidywanie nowych wartości	367
11.20. Kształtowanie przedziałów predykcji	368
11.21. Przeprowadzanie jednoczynnikowej analizy ANOVA	369
11.22. Tworzenie wykresu interakcji	371
11.23. Wyszukiwanie różnic pomiędzy średnimi grup	372
11.24. Przeprowadzanie odpornej analizy ANOVA (test Kruskala-Wallisa)	375
11.25. Porównywanie modeli za pomocą analizy ANOVA	376
12. Przydatne sztuczki	379
12.1. Zagłębienie do danych	379
12.2. Wyświetlanie rezultatu przypisania	380
12.3. Sumowanie rzędów lub kolumn	382
12.4. Wyświetlanie danych w kolumnach	383
12.5. Grupowanie danych w przedziały	384
12.6. Określanie położenia danej wartości	385
12.7. Wybieranie co n-tego elementu wektora	385
12.8. Określanie minimów i maksimów	386
12.9. Tworzenie wszystkich kombinacji kilku zmiennych	388
12.10. Spłaszczanie ramki danych	389
12.11. Sortowanie ramki danych	390
12.12. Usuwanie atrybutów ze zmiennej	391
12.13. Odkrywanie struktury obiektu	392

12.14. Obliczanie czasu potrzebnego na realizację kodu	395
12.15. Wstrzymywanie ostrzeżeń i komunikatów o błędach	396
12.16. Pobieranie argumentów funkcji z listy	397
12.17. Definiowanie własnych operatorów binarnych	399
12.18. Blokowanie komunikatu rozruchowego	401
12.19. Przeglądanie i wyznaczanie zmiennych środowiskowych	401
12.20. Używanie sekcji kodu	402
12.21. Równoległe przetwarzanie kodu R na komputerze lokalnym	403
12.22. Równoległe przetwarzanie kodu R w sposób zdalny	406
13. Zaawansowane obliczenia numeryczne i statystyczne	411
13.1. Minimalizowanie lub maksymalizowanie funkcji jednoparametrowej	411
13.2. Minimalizowanie lub maksymalizowanie funkcji wieloparametrowej	412
13.3. Obliczanie wartości własnych i wektorów własnych	414
13.4. Przeprowadzanie analizy głównych składowych	415
13.5. Przeprowadzanie prostej regresji ortogonalnej	416
13.6. Wyszukiwanie skupień w danych	418
13.7. Przewidywanie zmiennej binarnej (regresja logistyczna)	421
13.8. Metody samowsporne	423
13.9. Analiza czynnikowa	425
14. Analiza szeregów czasowych	431
14.1. Reprezentowanie danych szeregów czasowych	433
14.2. Tworzenie wykresów danych szeregów czasowych	436
14.3. Wydobywanie najstarszych lub najnowszych obserwacji	437
14.4. Tworzenie podzbiorów z szeregów czasowych	439
14.5. Scalanie kilku szeregów czasowych	441
14.6. Uzupełnianie brakujących obserwacji w szeregach czasowych	443
14.7. Opóźnianie lub przyspieszanie szeregu czasowego	446
14.8. Obliczanie kolejnych różnic	447
14.9. Wykonywanie obliczeń na szeregu czasowym	449
14.10. Obliczanie średniej kroczącej	450
14.11. Stosowanie funkcji przy uwzględnieniu okresu kalendarzowego	451
14.12. Stosowanie funkcji rozwijającej	453
14.13. Tworzenie wykresu funkcji autokorelacji	455
14.14. Testowanie szeregów czasowych pod kątem autokorelacji	456
14.15. Tworzenie wykresu funkcji autokorelacji cząstkowej	457
14.16. Wyszukiwanie korelacji opóźnionych pomiędzy dwoma szeregami czasowymi	459
14.17. Usuwanie trendów z szeregów czasowych	461
14.18. Dopasowywanie modelu ARIMA	463
14.19. Usuwanie nieistotnych współczynników z modelu ARIMA	466

14.20. Diagnozowanie modelu ARIMA	468
14.21. Uzyskiwanie prognoz z modelu ARIMA	470
14.22. Tworzenie wykresu prognoz	471
14.23. Sprawdzanie występowania zjawiska równania do średniej w szeregu czasowym	472
14.24. Wygładzanie szeregu czasowego	475
15. Elementy prostego programowania	479
15.1. Wybór pomiędzy dwiema alternatywnymi opcjami: if/else	480
15.2. Przetwarzanie w pętli	482
15.3. Definiowanie funkcji	483
15.4. Tworzenie zmiennej lokalnej	485
15.5. Wybór pomiędzy wieloma alternatywnymi ścieżkami: funkcja switch	485
15.6. Definiowanie wartości domyślnych parametrów	487
15.7. Sygnalizowanie błędów	488
15.8. Ochrona przed błędami	489
15.9. Tworzenie funkcji anonimowej	490
15.10. Tworzenie zbioru funkcji wielokrotnego użytku	491
15.11. Automatyczne formatowanie kodu	492
16. Środowisko R Markdown i publikowanie	495
16.1. Tworzenie nowego dokumentu	496
16.2. Dodawanie tytułu, danych autora i daty	498
16.3. Formatowanie dokumentu tekstowego	499
16.4. Wstawianie nagłówków dokumentu	500
16.5. Wstawianie listy	500
16.6. Prezentowanie wyników kodu R	502
16.7. Kontrolowanie wyświetlania kodu i wyników	503
16.8. Wstawianie wykresu	505
16.9. Wstawianie tabeli	507
16.10. Wstawianie wygenerowanej tabeli	509
16.11. Wstawianie równań matematycznych	512
16.12. Generowanie wyniku w formacie HTML	513
16.13. Generowanie wyniku w formacie PDF	514
16.14. Generowanie wyników w formacie Microsoft Word	516
16.15. Generowanie pliku prezentacji	522
16.16. Tworzenie parametryzowanego raportu	524
16.17. Organizowanie pracy z dokumentami R Markdown	527