

Spis treści

Wstęp	9
Część I. Przegląd	21
1. Wizualizacja danych za pomocą pakietu ggplot2	23
Wstęp	23
Pierwsze kroki	24
Mapowanie estetyk	26
Typowe problemy	32
Panele	33
Obiekty geometryczne	35
Przekształcenia statystyczne	40
Dostosowanie położenia	46
Systemy współrzędnych	50
Warstwowa gramatyka graficzna	52
2. Organizacja pracy: podstawy	55
Podstawy kodowania	55
Co się kryje pod nazwą?	56
Wywoływanie funkcji	56
3. Przekształcanie danych za pomocą pakietu dplyr	59
Wprowadzenie	59
Filtrowanie wierszy za pomocą funkcji filter()	61
Organizowanie wierszy za pomocą funkcji arrange()	65
Wybieranie kolumn za pomocą funkcji select()	66
Dodawanie nowych zmiennych za pomocą funkcji mutate()	68
Zgrupowane wartości sumaryczne za pomocą funkcji summarize()	71
Grupowanie wyników mutowania (i filtrowania)	83

4. Organizacja pracy: skrypty	87
Uruchamianie kodu	88
Diagnostyka RStudio	88
5. Eksploracyjna analiza danych	91
Wstęp	91
Pytania	92
Odchylenie	93
Wartości brakujące	100
Kowariancja	102
Wzorce i modele	112
Wywołania ggplot2	115
Więcej informacji	115
6. Organizacja pracy: projekty	117
Co jest prawdziwe?	117
Gdzie przebywają nasze analizy?	118
Ścieżki i katalogi	119
Projekty RStudio	119
Podsumowanie	121
Część II. Przygotowywanie	123
7. Dane typu tibble z użyciem pakietu tibble	125
Wstęp	125
Tworzenie danych typu tibble	125
Typ tibble w porównaniu z typem data.frame	127
Interakcje ze starszym kodem	128
8. Importowanie danych za pomocą pakietu readr	131
Wstęp	131
Zaczynamy	131
Parsowanie wektora	134
Parsowanie pliku	140
Zapis do pliku	145
Inne typy danych	146
9. Czyszczenie danych z wykorzystaniem pakietu tidyr	149
Wstęp	149
Czyszczenie danych	150
Rozkład i gromadzenie	153

Rozdzielanie i łączenie	157
Brakujące wartości	160
Studium przypadku	162
Dane nieoczyszczone	166
10. Dane relacyjne z wykorzystaniem pakietu dplyr	167
Wstęp	167
nycflights13	168
Klucze	170
Złączenia mutujące	172
Złączenia filtrujące	180
Problemy ze złączeniami	183
Operacje na zbiorach	184
11. Przetwarzanie napisów za pomocą pakietu stringr	187
Wstęp	187
Podstawy napisów	187
Dopasowywanie wzorców do wyrażeń regularnych	191
Grupowanie i odwołania wsteczne	197
Narzędzia	198
Inne typy wzorców	207
Inne sposoby użycia wyrażeń regularnych	209
Pakiet stringi	210
12. Czynniki z użyciem pakietu forcats	211
Wstęp	211
Tworzenie czynników	211
Badania General Social Survey	213
Modyfikowanie kolejności czynnika	214
Modyfikowanie poziomów czynników	218
13. Przetwarzanie daty i czasu za pomocą pakietu lubridate	221
Wstęp	221
Tworzenie daty lub czasu	222
Komponenty danych typu data i czas	226
Odcinki czasu	230
Strefy czasowe	234

Część III. Program 237

14. Potoki z wykorzystaniem pakietu magrittr	239
Wstęp	239
Alternatywy potoków	239
Kiedy nie należy używać potoków?	243
Inne narzędzia z pakietu magrittr	243
15. Funkcje	247
Wstęp	247
Kiedy powinienem napisać funkcję?	248
Funkcje są dla ludzi i komputerów	250
Wykonywanie warunkowe	252
Argumenty funkcji	256
Zwracane wartości	260
Środowisko	262
16. Wektory	263
Wstęp	263
Podstawy wektorów	264
Ważne typy wektorów atomowych	265
Używanie wektorów atomowych	267
Wektory rekurencyjne (listy)	272
Wektory rozszerzone	278
17. Iteracje za pomocą pakietu purrr	281
Wstęp	281
Pętla for	282
Odmiany pętli for	284
Pętla for kontra programowanie funkcyjne	288
Funkcje mapujące	290
Obsługa niepowodzeń	293
Mapowanie na podstawie wielu argumentów	295
Funkcja walk	298
Inne wzorce pętli for	299

Część IV. Model303

18. Podstawy modelowania z wykorzystaniem pakietu modelr	307
Wstęp	307
Prosty model	308
Wizualizowanie modeli	315
Formuły i rodziny modeli	318
Wartości brakujące	329
Inne rodziny modeli	329
19. Budowanie modelu	331
Wstęp	331
Dlaczego diamenty niskiej jakości są droższe?	332
Co wpływa na liczbę lotów w ciągu dnia?	339
Więcej informacji o modelach	349
20. Wiele modeli z użyciem pakietów purrr i broom	351
Wstęp	351
gapminder	352
Kolumny w postaci list	361
Tworzenie kolumn w postaci list	363
Upraszczenie kolumn w postaci list	367
Czyszczenie danych za pomocą pakietu broom	369

Część V. Komunikowanie371

21. R Markdown	373
Wstęp	373
Podstawy R Markdown	374
Formatowanie tekstu za pomocą Markdown	376
Fragmenty kodu	377
Rozwiązywanie problemów	382
Nagłówki YAML	383
Więcej informacji	385
22. Grafika dla komunikacji z wykorzystaniem ggplot2	387
Wstęp	387
Etykieta	388
Adnotacje	390

Skale	396
Powiększanie	404
Szablony	405
Zapisywanie wykresów	407
Więcej informacji	410
23. Formaty R Markdown	411
Opcje wyjścia	411
Dokumenty	412
Notatniki	413
Prezentacje	413
Pulpity	414
Interaktywność	415
Serwisy WWW	417
Inne formaty	418
Więcej informacji	418
24. Sposób pracy z R Markdown	419
Skorowidz	421