

# R

# dla każdego

---

Zaawansowane analizy  
i grafika statystyczna

Jared P. Lander

Przekład  
Marek Włodarz

APN Promise  
Warszawa 2018

# Spis treści

	<b>Słowo wstępne</b> .....	<b>xi</b>
	<b>Wprowadzenie</b> .....	<b>xiii</b>
<b>1</b>	<b>Poznajemy R</b> .....	<b>1</b>
	1.1 Pobieranie R.....	1
	1.2 Wersja R.....	2
	1.3 Wersja 32- czy 64-bitowa.....	2
	1.4 Instalowanie.....	3
	1.5 Microsoft R Open.....	14
	1.6 Podsumowanie.....	15
<b>2</b>	<b>Środowisko R</b> .....	<b>17</b>
	2.1 Interfejs wiersza polecenia.....	18
	2.2 RStudio.....	19
	2.3 Microsoft Visual Studio.....	35
	2.4 Podsumowanie.....	36
<b>3</b>	<b>Pakiety R</b> .....	<b>37</b>
	3.1 Instalowanie pakietów.....	38
	3.2 Ładowanie pakietów.....	40
	3.3 Budowanie pakietu.....	41
	3.4 Podsumowanie.....	42
<b>4</b>	<b>Podstawy R</b> .....	<b>43</b>
	4.1 Podstawowe działania arytmetyczne.....	43
	4.2 Zmienne.....	44
	4.3 Typy danych.....	46
	4.4 Wektory.....	52
	4.5 Wywoływanie funkcji.....	58
	4.6 Dokumentacja funkcji.....	58
	4.7 Brakujące dane.....	59
	4.8 Potoki.....	61
	4.9 Podsumowanie.....	62
<b>5</b>	<b>Zaawansowane struktury danych</b> .....	<b>63</b>
	5.1 Ramki danych.....	63

**vi** Spis treści

5.2	Listy .....	71
5.3	Macierze .....	77
5.4	Tablice .....	80
5.5	Podsumowanie .....	81
<b>6</b>	<b>Wczytywanie danych do R .....</b>	<b>83</b>
6.1	Czytanie plików CSV .....	83
6.2	Dane Excela .....	88
6.3	Wczytywanie z baz danych .....	91
6.4	Dane z innych narzędzi statystycznych .....	94
6.5	Pliki binarne języka R .....	95
6.6	Dane dołączone do R .....	97
6.7	Wydobywanie danych z witryn sieci Web .....	98
6.8	Wczytywanie danych JSON .....	101
6.9	Podsumowanie .....	103
<b>7</b>	<b>Grafika statystyczna .....</b>	<b>105</b>
7.1	Podstawowe funkcje graficzne .....	105
7.2	ggplot2 .....	109
7.3	Podsumowanie .....	123
<b>8</b>	<b>Tworzenie funkcji w języku R .....</b>	<b>125</b>
8.1	Hello, World! .....	125
8.2	Argumenty funkcji .....	126
8.3	Zwracane wartości .....	129
8.4	Funkcja do.call .....	130
8.5	Podsumowanie .....	131
<b>9</b>	<b>Wyrażenia sterujące .....</b>	<b>133</b>
9.1	if oraz else .....	133
9.2	switch .....	136
9.3	ifelse .....	138
9.4	Złożone testy .....	139
9.5	Podsumowanie .....	140
<b>10</b>	<b>Pętle – nie-R metoda iteracji .....</b>	<b>141</b>
10.1	Pętla for .....	141
10.2	Pętla while .....	143
10.3	Sterowanie pętlami .....	144
10.4	Podsumowanie .....	144
<b>11</b>	<b>Manipulacje grupowe .....</b>	<b>145</b>

11.1	Rodzina apply .....	145
11.2	aggregate .....	149
11.3	plyr .....	152
11.4	data.table .....	157
11.5	Podsumowanie .....	166
<b>12</b>	<b>Szybsze manipulacje grupowe przy użyciu dplyr. ....</b>	<b>167</b>
12.1	Potoki .....	167
12.2	tbl .....	168
12.3	select .....	170
12.4	filter .....	176
12.5	slice .....	180
12.6	mutate .....	181
12.7	summarize .....	184
12.8	group_by .....	185
12.9	arrange .....	186
12.10	do .....	187
12.11	Stosowanie dplyr dla baz danych .....	189
12.12	Podsumowanie .....	192
<b>13</b>	<b>Iteracje przy użyciu purrr .....</b>	<b>193</b>
13.1	map .....	193
13.2	Funkcja map ze wskazanymi typami .....	195
13.3	Iteracje przez obiekt data.frame .....	201
13.4	Funkcja map z wieloma wejściami .....	202
13.5	Podsumowanie .....	203
<b>14</b>	<b>Kształtowanie danych .....</b>	<b>205</b>
14.1	Funkcje cbind oraz rbind .....	205
14.2	Złączenia .....	206
14.3	reshape2 .....	212
14.4	Podsumowanie .....	216
<b>15</b>	<b>Kształtowanie danych w Tidyverse .....</b>	<b>217</b>
15.1	Sklejanie wierszy i kolumn .....	217
15.2	Złączenia przy użyciu dplyr .....	218
15.3	Konwertowanie formatów danych .....	223
15.4	Podsumowanie .....	227
<b>16</b>	<b>Manipulowanie ciągami znaków .....</b>	<b>229</b>
16.1	paste .....	229

16.2	sprintf. ....	230
16.3	Wyodrębnianie tekstu . . . . .	231
16.4	Wyrażenia regularne . . . . .	235
16.5	Podsumowanie . . . . .	242
<b>17</b>	<b>Rozkłady prawdopodobieństwa. ....</b>	<b>243</b>
17.1	Rozkład normalny . . . . .	243
17.2	Rozkład dwumianowy. ....	249
17.3	Rozkład Poissona . . . . .	254
17.4	Inne rozkłady. ....	257
17.5	Podsumowanie . . . . .	260
<b>18</b>	<b>Podstawowe statystyki . . . . .</b>	<b>261</b>
18.1	Statystyki podsumowujące . . . . .	261
18.2	Korelacja i kowariancja . . . . .	265
18.3	Test t-Studenta . . . . .	274
18.4	ANOVA. ....	283
18.5	Podsumowanie . . . . .	286
<b>19</b>	<b>Modele liniowe . . . . .</b>	<b>287</b>
19.1	Prosta regresja liniowa. ....	287
19.2	Regresja wieloraka . . . . .	293
19.3	Podsumowanie . . . . .	310
<b>20</b>	<b>Uogólnione modele liniowe . . . . .</b>	<b>311</b>
20.1	Regresja logistyczna. ....	311
20.2	Regresja Poissona. ....	315
20.3	Inne uogólnione modele liniowe . . . . .	319
20.4	Analiza przeżycia . . . . .	319
20.5	Podsumowanie . . . . .	325
<b>21</b>	<b>Diagnostyka modelu . . . . .</b>	<b>327</b>
21.1	Reszty. ....	327
21.2	Porównywanie modeli. ....	333
21.3	Sprawdzian krzyżowy . . . . .	337
21.4	Bootstrap . . . . .	342
21.5	Krokowe wybieranie zmiennych . . . . .	346
21.6	Podsumowanie . . . . .	350
<b>22</b>	<b>Regularyzacja i ściąganie . . . . .</b>	<b>351</b>
22.1	Elastic Net. ....	351
22.2	Ściąganie bayesowskie. ....	370

22.3	Podsumowanie .....	375
<b>23</b>	<b>Modele nieliniowe .....</b>	<b>377</b>
23.1	Modele nieliniowe najmniejszych kwadratów .....	377
23.2	Interpolacja funkcjami sklejanymi .....	380
23.3	Uogólnione modele addytywne .....	384
23.4	Drzewa decyzyjne .....	391
23.5	Wzmocnione drzewa decyzyjne .....	394
23.6	Lasy losowe .....	397
23.7	Podsumowanie .....	399
<b>24</b>	<b>Serie czasowe i autokorelacja .....</b>	<b>401</b>
24.1	Autoregresywne średnie ruchome .....	401
24.2	VAR .....	409
24.3	GARCH .....	415
24.4	Podsumowanie .....	422
<b>25</b>	<b>Grupowanie .....</b>	<b>423</b>
25.1	K-średnie .....	423
25.2	PAM .....	432
25.3	Grupowanie hierarchiczne .....	439
25.4	Podsumowanie .....	444
<b>26</b>	<b>Dopasowywanie modelu przy użyciu Caret .....</b>	<b>445</b>
26.1	Podstawy Caret .....	445
26.2	Opcje Caret .....	446
26.3	Dostrajanie wzmocnionego drzewa .....	448
26.4	Podsumowanie .....	452
<b>27</b>	<b>Reprodukowalność i raporty przy użyciu knitr .....</b>	<b>453</b>
27.1	Instalowanie programu L <sup>A</sup> T <sub>E</sub> X .....	453
27.2	Elementarz L <sup>A</sup> T <sub>E</sub> X .....	454
27.3	Korzystanie z knitr w połączeniu z L <sup>A</sup> T <sub>E</sub> X .....	457
27.4	Podsumowanie .....	463
<b>28</b>	<b>Tworzenie bogatych dokumentów przy użyciu RMarkdown .....</b>	<b>465</b>
28.1	Kompilowanie dokumentu .....	465
28.2	Nagłówek dokumentu .....	466
28.3	Elementarz języka Markdown .....	467
28.4	Wstawki kodu w Markdown .....	469
28.5	htmlwidgets .....	471

**x** Spis treści

28.6	Pokazy slajdów RMarkdown.....	485
28.7	Podsumowanie .....	486
<b>29</b>	<b>Tworzenie interaktywnych tablic kontrolnych przy użyciu Shiny .....</b>	<b>487</b>
29.1	Shiny w RMarkdown.....	488
29.2	Wyrażenia reaktywne w Shiny .....	493
29.3	Serwer i UI.....	496
29.4	Podsumowanie .....	505
<b>30</b>	<b>Tworzenie pakietów R.....</b>	<b>507</b>
30.1	Struktura folderów.....	507
30.2	Pliki pakietu.....	508
30.3	Dokumentacja pakietu .....	515
30.4	Testy.....	518
30.5	Sprawdzanie, kompilacja i instalowanie.....	521
30.6	Wysyłanie pakietu do CRAN .....	523
30.7	Kod C++.....	523
30.8	Podsumowanie .....	530
<b>A</b>	<b>Praktyczne zasoby .....</b>	<b>531</b>
A.1	Meetupy .....	531
A.2	Stack Overflow .....	533
A.3	Twitter.....	533
A.4	Konferencje .....	533
A.5	Witryny Web.....	534
A.6	Dokumenty .....	534
A.7	Książki.....	535
A.8	Podsumowanie .....	536
<b>B</b>	<b>Słownik.....</b>	<b>537</b>
	<b>Spis ilustracji .....</b>	<b>553</b>
	<b>Spis tabel.....</b>	<b>561</b>
	<b>Indeks ogólny .....</b>	<b>563</b>
	<b>Indeks funkcji .....</b>	<b>571</b>
	<b>Indeks osób .....</b>	<b>581</b>
	<b>Indeks pakietów .....</b>	<b>582</b>
	<b>Indeks zbiorów danych.....</b>	<b>584</b>