

EDA - Analiza danych z zestawu ‘crabs’

2024-12-17

Contents

0. Wprowadzenie	1
0.1 Pakiety	2
0.2 Zmienne	2
1. Dane	2
1.1 Tabela z danymi	2
1.2 Struktura zestawu danych	6
1.3 Podstawowe statystyki	6
1.4 Braki danych w zestawie	7
2. Różnice międzygatunkowe wśród krabów	8
2.1 Jak różnią się pod względem rozmiaru płata czołowego? (zmienna FL)	8
2.2 Jak różnią się pod względem szerokości tylnej części kraba? (zmienna RW)	10
2.3 Jak różnią się pod względem długości karapaksu? (zmienna CL)	11
2.4 Jak różnią się pod względem szerokości karapaksu (zmienna CW)	13
2.5 Jak różnią się pod względem głębokości ciała? (zmienna BD)	14
3. Zależności między zmiennymi	16
3.1 Jaka jest zależność pomiędzy wielkością płata czołowego (FL) a szerokością karapaksu (CW)? .	16
3.2 Jaka jest zależność pomiędzy szerokością tylnej części kraba (RW) a głębokością ciała (BD)? .	17
3.3 Jaka jest zależność pomiędzy długością ciała (FL) a głębokością ciała kraba (BD)?	19
3.4 Jaka jest zależność pomiędzy wielkością płata czołowego (FL) a długością karapaksu (CL)? . .	20
3.5 Jaka jest zależność pomiędzy długością karapaksu (CL) a głębokością ciała kraba (BD)?	21

0. Wprowadzenie

Zestaw danych `crabs` składa się z 200 obserwacji dotyczących krabów, w tym 100 obserwacji dla każdego z dwóch gatunków: Blue (B) i Orange (O). W danych zawarte są cechy fizyczne krabów, które mogą być wykorzystane do analizy statystycznej, w tym do klasyfikacji gatunków. Dla zainteresowanych osób podrzucam badanie, które było inspiracją do stworzenia tego zestawu danych:

Campbell, N.A. and Mahon, R.J. (1974) A multivariate study of variation in two species of rock crab of genus *Leptograpsus*. *Australian Journal of Zoology* 22, 417–425.

0.1 Pakiety

Będę korzystał z następujących bibliotek:

```
library(psych)
library(MASS)
library(knitr)
library(ggplot2)
library(mice)
library(dplyr)
library(tinytex)
```

0.2 Zmienne

Zestaw danych `crabs` z pakietu `MASS` zawiera 8 zmiennych, z czego 3 stanowią zmienne katgoryczne a pozostałe 5 zmienne numeryczne, których wartości zostały wyrażone w milimetrach (mm). Poniżej przedstawiłem informacje dotyczące danych zawartych w poszczególnych zmiennych.

Kategoryczne:

- `sp`: Gatunek kraba. Odpowiednio “B” oznacza odmianę niebieską a “O” pomarańczową.
- `sex`: Płeć kraba. Odpowiednio “F” oznacza płć żeńską a “M” płć męską.

Numeryczne:

- `FL`: Rozmiar płata czołowego u kraba
- `RW`: Szerokość tylnej części kraba czyli w najwęższym miejscu ciała kraba
- `CL`: Długość karapaksu czyli twardej, zewnętrznej powłoki ciała kraba, która osłania jego głowę oraz część klatki piersiowej.
- `CW`: Szerokość karapaksu
- `BD`: Głębokość ciała kraba.

1. Dane

1.1 Tabela z danymi

`crabs`

	sp	sex	index	FL	RW	CL	CW	BD
1	B	M	1	8.1	6.7	16.1	19.0	7.0
2	B	M	2	8.8	7.7	18.1	20.8	7.4
3	B	M	3	9.2	7.8	19.0	22.4	7.7
4	B	M	4	9.6	7.9	20.1	23.1	8.2
5	B	M	5	9.8	8.0	20.3	23.0	8.2
6	B	M	6	10.8	9.0	23.0	26.5	9.8
7	B	M	7	11.1	9.9	23.8	27.1	9.8
8	B	M	8	11.6	9.1	24.5	28.4	10.4
9	B	M	9	11.8	9.6	24.2	27.8	9.7
10	B	M	10	11.8	10.5	25.2	29.3	10.3

11	B	M	11	12.2	10.8	27.3	31.6	10.9
12	B	M	12	12.3	11.0	26.8	31.5	11.4
13	B	M	13	12.6	10.0	27.7	31.7	11.4
14	B	M	14	12.8	10.2	27.2	31.8	10.9
15	B	M	15	12.8	10.9	27.4	31.5	11.0
16	B	M	16	12.9	11.0	26.8	30.9	11.4
17	B	M	17	13.1	10.6	28.2	32.3	11.0
18	B	M	18	13.1	10.9	28.3	32.4	11.2
19	B	M	19	13.3	11.1	27.8	32.3	11.3
20	B	M	20	13.9	11.1	29.2	33.3	12.1
21	B	M	21	14.3	11.6	31.3	35.5	12.7
22	B	M	22	14.6	11.3	31.9	36.4	13.7
23	B	M	23	15.0	10.9	31.4	36.4	13.2
24	B	M	24	15.0	11.5	32.4	37.0	13.4
25	B	M	25	15.0	11.9	32.5	37.2	13.6
26	B	M	26	15.2	12.1	32.3	36.7	13.6
27	B	M	27	15.4	11.8	33.0	37.5	13.6
28	B	M	28	15.7	12.6	35.8	40.3	14.5
29	B	M	29	15.9	12.7	34.0	38.9	14.2
30	B	M	30	16.1	11.6	33.8	39.0	14.4
31	B	M	31	16.1	12.8	34.9	40.7	15.7
32	B	M	32	16.2	13.3	36.0	41.7	15.4
33	B	M	33	16.3	12.7	35.6	40.9	14.9
34	B	M	34	16.4	13.0	35.7	41.8	15.2
35	B	M	35	16.6	13.5	38.1	43.4	14.9
36	B	M	36	16.8	12.8	36.2	41.8	14.9
37	B	M	37	16.9	13.2	37.3	42.7	15.6
38	B	M	38	17.1	12.6	36.4	42.0	15.1
39	B	M	39	17.1	12.7	36.7	41.9	15.6
40	B	M	40	17.2	13.5	37.6	43.9	16.1
41	B	M	41	17.7	13.6	38.7	44.5	16.0
42	B	M	42	17.9	14.1	39.7	44.6	16.8
43	B	M	43	18.0	13.7	39.2	44.4	16.2
44	B	M	44	18.8	15.8	42.1	49.0	17.8
45	B	M	45	19.3	13.5	41.6	47.4	17.8
46	B	M	46	19.3	13.8	40.9	46.5	16.8
47	B	M	47	19.7	15.3	41.9	48.5	17.8
48	B	M	48	19.8	14.2	43.2	49.7	18.6
49	B	M	49	19.8	14.3	42.4	48.9	18.3
50	B	M	50	21.3	15.7	47.1	54.6	20.0
51	B	F	1	7.2	6.5	14.7	17.1	6.1
52	B	F	2	9.0	8.5	19.3	22.7	7.7
53	B	F	3	9.1	8.1	18.5	21.6	7.7
54	B	F	4	9.1	8.2	19.2	22.2	7.7
55	B	F	5	9.5	8.2	19.6	22.4	7.8
56	B	F	6	9.8	8.9	20.4	23.9	8.8
57	B	F	7	10.1	9.3	20.9	24.4	8.4
58	B	F	8	10.3	9.5	21.3	24.7	8.9
59	B	F	9	10.4	9.7	21.7	25.4	8.3
60	B	F	10	10.8	9.5	22.5	26.3	9.1
61	B	F	11	11.0	9.8	22.5	25.7	8.2
62	B	F	12	11.2	10.0	22.8	26.9	9.4
63	B	F	13	11.5	11.0	24.7	29.2	10.1
64	B	F	14	11.6	11.0	24.6	28.5	10.4

65	B	F	15	11.6	11.4	23.7	27.7	10.0
66	B	F	16	11.7	10.6	24.9	28.5	10.4
67	B	F	17	11.9	11.4	26.0	30.1	10.9
68	B	F	18	12.0	10.7	24.6	28.9	10.5
69	B	F	19	12.0	11.1	25.4	29.2	11.0
70	B	F	20	12.6	12.2	26.1	31.6	11.2
71	B	F	21	12.8	11.7	27.1	31.2	11.9
72	B	F	22	12.8	12.2	26.7	31.1	11.1
73	B	F	23	12.8	12.2	27.9	31.9	11.5
74	B	F	24	13.0	11.4	27.3	31.8	11.3
75	B	F	25	13.1	11.5	27.6	32.6	11.1
76	B	F	26	13.2	12.2	27.9	32.1	11.5
77	B	F	27	13.4	11.8	28.4	32.7	11.7
78	B	F	28	13.7	12.5	28.6	33.8	11.9
79	B	F	29	13.9	13.0	30.0	34.9	13.1
80	B	F	30	14.7	12.5	30.1	34.7	12.5
81	B	F	31	14.9	13.2	30.1	35.6	12.0
82	B	F	32	15.0	13.8	31.7	36.9	14.0
83	B	F	33	15.0	14.2	32.8	37.4	14.0
84	B	F	34	15.1	13.3	31.8	36.3	13.5
85	B	F	35	15.1	13.5	31.9	37.0	13.8
86	B	F	36	15.1	13.8	31.7	36.6	13.0
87	B	F	37	15.2	14.3	33.9	38.5	14.7
88	B	F	38	15.3	14.2	32.6	38.3	13.8
89	B	F	39	15.4	13.3	32.4	37.6	13.8
90	B	F	40	15.5	13.8	33.4	38.7	14.7
91	B	F	41	15.6	13.9	32.8	37.9	13.4
92	B	F	42	15.6	14.7	33.9	39.5	14.3
93	B	F	43	15.7	13.9	33.6	38.5	14.1
94	B	F	44	15.8	15.0	34.5	40.3	15.3
95	B	F	45	16.2	15.2	34.5	40.1	13.9
96	B	F	46	16.4	14.0	34.2	39.8	15.2
97	B	F	47	16.7	16.1	36.6	41.9	15.4
98	B	F	48	17.4	16.9	38.2	44.1	16.6
99	B	F	49	17.5	16.7	38.6	44.5	17.0
100	B	F	50	19.2	16.5	40.9	47.9	18.1
101	O	M	1	9.1	6.9	16.7	18.6	7.4
102	O	M	2	10.2	8.2	20.2	22.2	9.0
103	O	M	3	10.7	8.6	20.7	22.7	9.2
104	O	M	4	11.4	9.0	22.7	24.8	10.1
105	O	M	5	12.5	9.4	23.2	26.0	10.8
106	O	M	6	12.5	9.4	24.2	27.0	11.2
107	O	M	7	12.7	10.4	26.0	28.8	12.1
108	O	M	8	13.2	11.0	27.1	30.4	12.2
109	O	M	9	13.4	10.1	26.6	29.6	12.0
110	O	M	10	13.7	11.0	27.5	30.5	12.2
111	O	M	11	14.0	11.5	29.2	32.2	13.1
112	O	M	12	14.1	10.4	28.9	31.8	13.5
113	O	M	13	14.1	10.5	29.1	31.6	13.1
114	O	M	14	14.1	10.7	28.7	31.9	13.3
115	O	M	15	14.2	10.6	28.7	31.7	12.9
116	O	M	16	14.2	10.7	27.8	30.9	12.7
117	O	M	17	14.2	11.3	29.2	32.2	13.5
118	O	M	18	14.6	11.3	29.9	33.5	12.8

119	0	M	19	14.7	11.1	29.0	32.1	13.1
120	0	M	20	15.1	11.4	30.2	33.3	14.0
121	0	M	21	15.1	11.5	30.9	34.0	13.9
122	0	M	22	15.4	11.1	30.2	33.6	13.5
123	0	M	23	15.7	12.2	31.7	34.2	14.2
124	0	M	24	16.2	11.8	32.3	35.3	14.7
125	0	M	25	16.3	11.6	31.6	34.2	14.5
126	0	M	26	17.1	12.6	35.0	38.9	15.7
127	0	M	27	17.4	12.8	36.1	39.5	16.2
128	0	M	28	17.5	12.0	34.4	37.3	15.3
129	0	M	29	17.5	12.7	34.6	38.4	16.1
130	0	M	30	17.8	12.5	36.0	39.8	16.7
131	0	M	31	17.9	12.9	36.9	40.9	16.5
132	0	M	32	18.0	13.4	36.7	41.3	17.1
133	0	M	33	18.2	13.7	38.8	42.7	17.2
134	0	M	34	18.4	13.4	37.9	42.2	17.7
135	0	M	35	18.6	13.4	37.8	41.9	17.3
136	0	M	36	18.6	13.5	36.9	40.2	17.0
137	0	M	37	18.8	13.4	37.2	41.1	17.5
138	0	M	38	18.8	13.8	39.2	43.3	17.9
139	0	M	39	19.4	14.1	39.1	43.2	17.8
140	0	M	40	19.4	14.4	39.8	44.3	17.9
141	0	M	41	20.1	13.7	40.6	44.5	18.0
142	0	M	42	20.6	14.4	42.8	46.5	19.6
143	0	M	43	21.0	15.0	42.9	47.2	19.4
144	0	M	44	21.5	15.5	45.5	49.7	20.9
145	0	M	45	21.6	15.4	45.7	49.7	20.6
146	0	M	46	21.6	14.8	43.4	48.2	20.1
147	0	M	47	21.9	15.7	45.4	51.0	21.1
148	0	M	48	22.1	15.8	44.6	49.6	20.5
149	0	M	49	23.0	16.8	47.2	52.1	21.5
150	0	M	50	23.1	15.7	47.6	52.8	21.6
151	0	F	1	10.7	9.7	21.4	24.0	9.8
152	0	F	2	11.4	9.2	21.7	24.1	9.7
153	0	F	3	12.5	10.0	24.1	27.0	10.9
154	0	F	4	12.6	11.5	25.0	28.1	11.5
155	0	F	5	12.9	11.2	25.8	29.1	11.9
156	0	F	6	14.0	11.9	27.0	31.4	12.6
157	0	F	7	14.0	12.8	28.8	32.4	12.7
158	0	F	8	14.3	12.2	28.1	31.8	12.5
159	0	F	9	14.7	13.2	29.6	33.4	12.9
160	0	F	10	14.9	13.0	30.0	33.7	13.3
161	0	F	11	15.0	12.3	30.1	33.3	14.0
162	0	F	12	15.6	13.5	31.2	35.1	14.1
163	0	F	13	15.6	14.0	31.6	35.3	13.8
164	0	F	14	15.6	14.1	31.0	34.5	13.8
165	0	F	15	15.7	13.6	31.0	34.8	13.8
166	0	F	16	16.1	13.6	31.6	36.0	14.0
167	0	F	17	16.1	13.7	31.4	36.1	13.9
168	0	F	18	16.2	14.0	31.6	35.6	13.7
169	0	F	19	16.7	14.3	32.3	37.0	14.7
170	0	F	20	17.1	14.5	33.1	37.2	14.6
171	0	F	21	17.5	14.3	34.5	39.6	15.6
172	0	F	22	17.5	14.4	34.5	39.0	16.0

```

173 0 F 23 17.5 14.7 33.3 37.6 14.6
174 0 F 24 17.6 14.0 34.0 38.6 15.5
175 0 F 25 18.0 14.9 34.7 39.5 15.7
176 0 F 26 18.0 16.3 37.9 43.0 17.2
177 0 F 27 18.3 15.7 35.1 40.5 16.1
178 0 F 28 18.4 15.5 35.6 40.0 15.9
179 0 F 29 18.4 15.7 36.5 41.6 16.4
180 0 F 30 18.5 14.6 37.0 42.0 16.6
181 0 F 31 18.6 14.5 34.7 39.4 15.0
182 0 F 32 18.8 15.2 35.8 40.5 16.6
183 0 F 33 18.9 16.7 36.3 41.7 15.3
184 0 F 34 19.1 16.0 37.8 42.3 16.8
185 0 F 35 19.1 16.3 37.9 42.6 17.2
186 0 F 36 19.7 16.7 39.9 43.6 18.2
187 0 F 37 19.9 16.6 39.4 43.9 17.9
188 0 F 38 19.9 17.9 40.1 46.4 17.9
189 0 F 39 20.0 16.7 40.4 45.1 17.7
190 0 F 40 20.1 17.2 39.8 44.1 18.6
191 0 F 41 20.3 16.0 39.4 44.1 18.0
192 0 F 42 20.5 17.5 40.0 45.5 19.2
193 0 F 43 20.6 17.5 41.5 46.2 19.2
194 0 F 44 20.9 16.5 39.9 44.7 17.5
195 0 F 45 21.3 18.4 43.8 48.4 20.0
196 0 F 46 21.4 18.0 41.2 46.2 18.7
197 0 F 47 21.7 17.1 41.7 47.2 19.6
198 0 F 48 21.9 17.2 42.6 47.4 19.5
199 0 F 49 22.5 17.2 43.0 48.7 19.8
200 0 F 50 23.1 20.2 46.2 52.5 21.1

```

1.2 Struktura zestawu danych

```
str(crabs)
```

```

'data.frame': 200 obs. of 8 variables:
 $ sp : Factor w/ 2 levels "B","O": 1 1 1 1 1 1 1 1 1 1 ...
 $ sex : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...
 $ index: int 1 2 3 4 5 6 7 8 9 10 ...
 $ FL : num 8.1 8.8 9.2 9.6 9.8 10.8 11.1 11.6 11.8 11.8 ...
 $ RW : num 6.7 7.7 7.8 7.9 8 9 9.9 9.1 9.6 10.5 ...
 $ CL : num 16.1 18.1 19 20.1 20.3 23 23.8 24.5 24.2 25.2 ...
 $ CW : num 19 20.8 22.4 23.1 23 26.5 27.1 28.4 27.8 29.3 ...
 $ BD : num 7 7.4 7.7 8.2 8.2 9.8 9.8 10.4 9.7 10.3 ...

```

1.3 Podstawowe statystyki

Ze względu na to że zestaw danych `crabs` zawiera zmienne `sex` oraz `sp` (reprezentuje ona dwa gatunki krabów), postanowiłem sprawdzić podstawowe statystyki dla każdej grupy w zależności od płci oraz rasy.

```
describeBy(crabs[, 4:8], group = interaction(crabs$sp, crabs$sex))
```

Descriptive statistics by group

group: B.F

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
FL	1	50	13.27	2.63	13.15	13.31	2.97	7.2	19.2	12.0	-0.12	-0.72	0.37
RW	2	50	12.14	2.44	12.20	12.15	2.45	6.5	16.9	10.4	-0.10	-0.64	0.34
CL	3	50	28.10	5.92	27.90	28.12	7.12	14.7	40.9	26.2	-0.05	-0.75	0.84
CW	4	50	32.62	6.80	32.35	32.66	7.93	17.1	47.9	30.8	-0.06	-0.69	0.96
BD	5	50	11.82	2.75	11.60	11.79	3.26	6.1	18.1	12.0	0.06	-0.73	0.39

group: O.F

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
FL	1	50	17.59	2.97	18.00	17.73	3.26	10.7	23.1	12.4	-0.33	-0.64	0.42
RW	2	50	14.84	2.35	14.65	14.96	2.45	9.2	20.2	11.0	-0.33	-0.22	0.33
CL	3	50	34.62	5.84	34.70	34.89	6.89	21.4	46.2	24.8	-0.30	-0.57	0.83
CW	4	50	39.04	6.54	39.55	39.38	6.75	24.0	52.5	28.5	-0.36	-0.46	0.93
BD	5	50	15.63	2.75	15.65	15.70	2.82	9.7	21.1	11.4	-0.13	-0.74	0.39

group: B.M

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
FL	1	50	14.84	3.20	15.10	14.92	3.34	8.1	21.3	13.2	-0.15	-0.78	0.45
RW	2	50	11.72	2.11	11.70	11.81	2.08	6.7	15.8	9.1	-0.31	-0.39	0.30
CL	3	50	32.01	7.31	32.45	32.26	7.64	16.1	47.1	31.0	-0.20	-0.75	1.03
CW	4	50	36.81	8.35	37.10	37.04	8.30	19.0	54.6	35.6	-0.17	-0.73	1.18
BD	5	50	13.35	3.20	13.60	13.41	3.56	7.0	20.0	13.0	-0.10	-0.85	0.45

group: O.M

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
FL	1	50	16.63	3.51	16.70	16.64	3.85	9.1	23.1	14.0	0.00	-0.88	0.50
RW	2	50	12.26	2.20	12.10	12.29	2.15	6.9	16.8	9.9	-0.07	-0.54	0.31
CL	3	50	33.69	7.61	33.35	33.74	7.49	16.7	47.6	30.9	0.00	-0.80	1.08
CW	4	50	37.19	8.39	36.30	37.24	8.45	18.6	52.8	34.2	0.01	-0.80	1.19
BD	5	50	15.32	3.53	15.00	15.35	3.56	7.4	21.6	14.2	-0.01	-0.79	0.50

1.4 Braki danych w zestawie

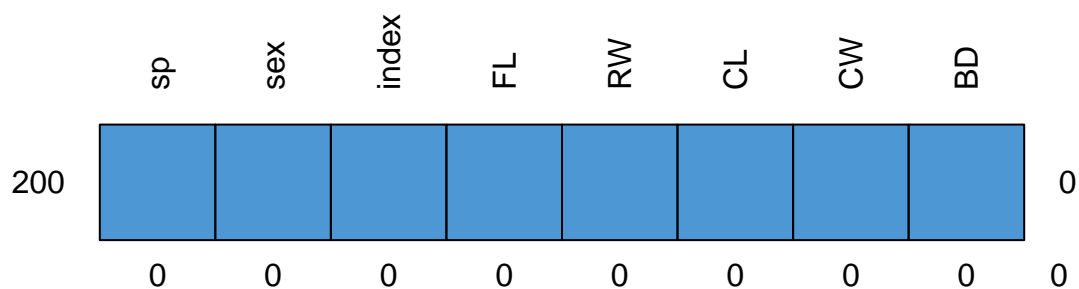
W tym przypadku wykorzystałem funkcję `md.pattern()` z pakietu `mice`. Rozwiązanie to umożliwia mi szybkie sprawdzenie czy w danym zbiorze danych znajdują się braki danych `NA`. Nie ukrywam że funkcję tę poznałem dzięki jednej z przykładowych prac podanych na platformie UPEL. Poniższy wynik wskazuje że baza danych nie zawiera braków dlatego można przejść już do właściwego etapu analizy danych.

```
md.pattern(crabs, rotate.names = TRUE)
```

```

/\      /\
{ '---' }
{ 0  0 }
==> V <== No need for mice. This data set is completely observed.
\  \||/ /
'-----'

```



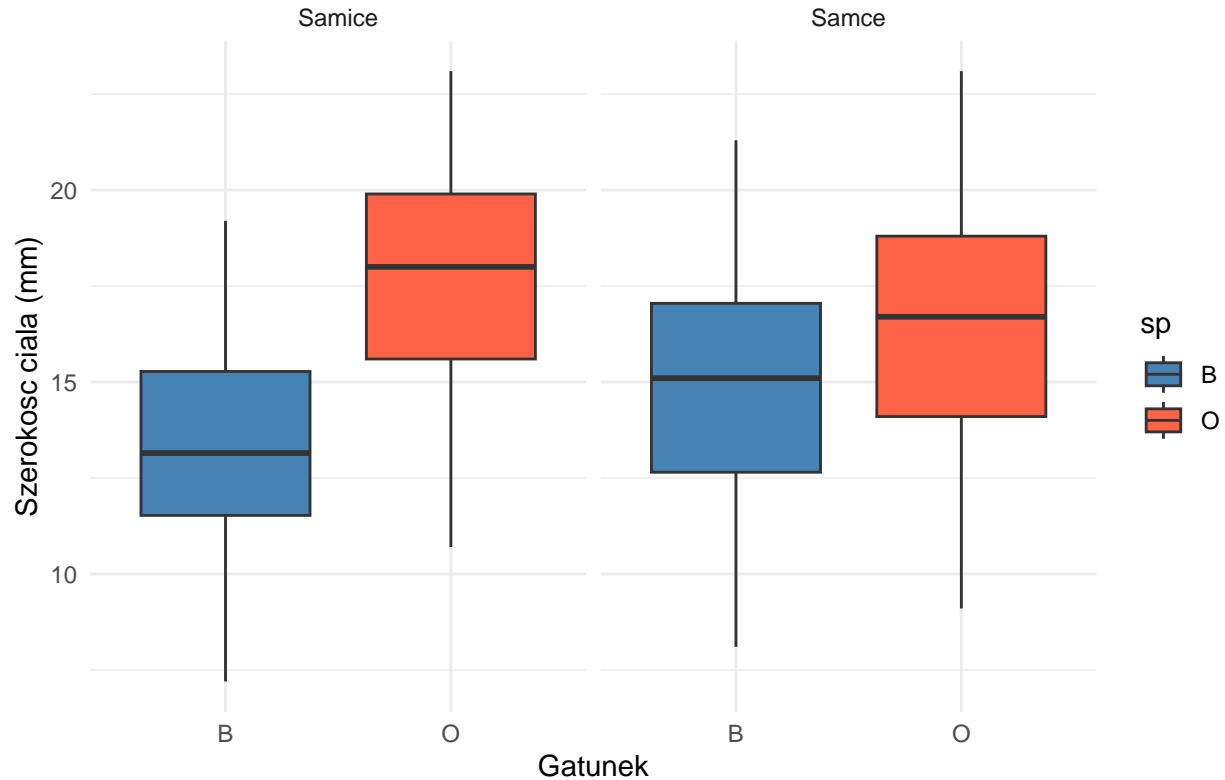
	sp	sex	index	FL	RW	CL	CW	BD
200	1	1	1	1	1	1	1	0
	0	0	0	0	0	0	0	0

2. Różnice międzygatunkowe wśród krabów

2.1 Jak różnią się pod względem rozmiaru płata czołowego? (zmienna FL)

```
ggplot(crabs, aes(x = sp, y = FL, fill = sp)) +
  geom_boxplot() +
  scale_fill_manual(values = c("#4682B4", "#FF6347")) +
  labs(title = "Długość płata czołowego krabów",
       x = "Gatunek",
       y = "Szerokość ciała (mm)") +
  theme_minimal() +
  facet_wrap(~sex, labeller = labeller(sex = c("F" = "Samice", "M" = "Samce")))
```


Długość pła czołowego krabów



```
stats_crabs_FL <- crabs %>%
  group_by(sex, sp) %>%
  summarise(
    Min = min(FL),
    Q1 = quantile(FL, 0.25),
    Median = median(FL),
    Q3 = quantile(FL, 0.75),
    Max = max(FL),
    .groups = "drop"
  )
```

```
stats_crabs_FL
```

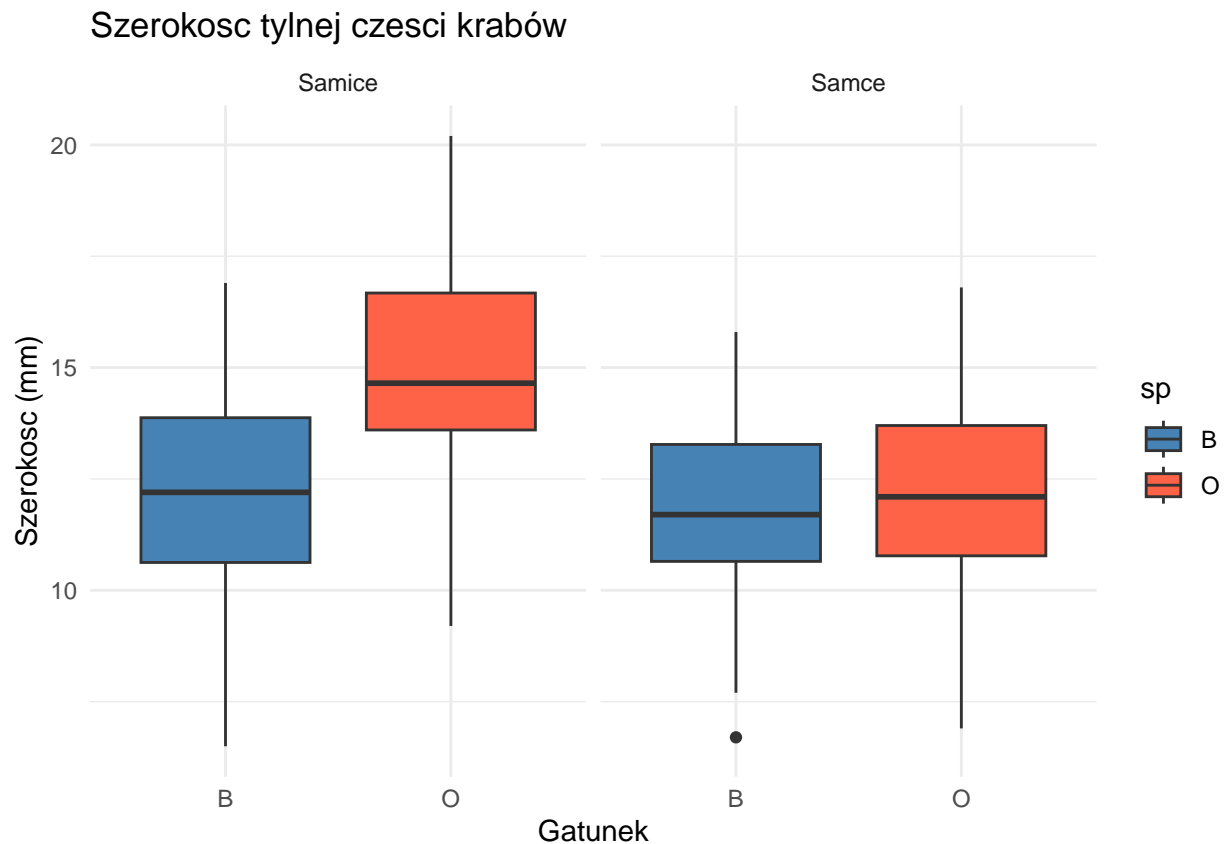
```
## # A tibble: 4 x 7
##   sex  sp    Min    Q1 Median    Q3    Max
##   <fct> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 F    B      7.2  11.5  13.2  15.3  19.2
## 2 F    O     10.7  15.6  18    19.9  23.1
## 3 M    B      8.1  12.6  15.1  17.0  21.3
## 4 M    O      9.1  14.1  16.7  18.8  23.1
```

Dla samic krabów, widać, że płaty czołowe samic Gatunku B są najmniejsze, z minimalnym rozmiarem wynoszącym 7.2, a maksymalnym 19.2. Mediana (13.15) oraz pozostałe kwartyłe wskazują na względnie mały rozrzut rozmiarów w tej grupie. Natomiast samice typu Orange (O) mają wyraźnie większe płaty czołowe, z minimalnym rozmiarem 10.7 i maksymalnym 23.1. Mediana wynosi 18.0, a kwartyła 3. (19.9) wskazuje na

szerszy zakres wielkości, co sugeruje większą zmienność w tej grupie. Z kolei dla samców, zauważamy podobny trend. Samce gatunku B mają płyty czołowe o rozmiarze od 8.1 do 21.3, z medianą wynoszącą 15.1, co także pasuje im w podobnym zakresie jak samice tego typu. Samce gatunku O mają płyty czołowe o rozmiarze od 9.1 do 23.1, z medianą 16.7. Wartości te są nieco wyższe niż u samic typu Orange, co wskazuje, że w tej grupie samce mają nieco większe płyty czołowe niż samice.

2.2 Jak różnią się pod względem szerokości tylnej części kraba? (zmienna RW)

```
ggplot(crabs, aes(x = sp, y = RW, fill = sp)) +
  geom_boxplot() +
  scale_fill_manual(values = c("#4682B4", "#FF6347")) +
  labs(title = "Szerokość tylnej części krabów",
       x = "Gatunek",
       y = "Szerokość (mm)") +
  theme_minimal() +
  facet_wrap(~sex, labeller = labeller(sex = c("F" = "Samice", "M" = "Samce")))
```



```
stats_crabs_RW <- crabs %>%
  group_by(sex, sp) %>%
  summarise(
    Min = min(RW),
    Q1 = quantile(RW, 0.25),
    Median = median(RW),
```

```

    Q3 = quantile(RW, 0.75),
    Max = max(RW),
    .groups = "drop"
)

stats_crabs_RW

```

```

## # A tibble: 4 x 7
##   sex   sp      Min    Q1 Median    Q3    Max
##   <fct> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 F     B      6.5  10.6  12.2  13.9  16.9
## 2 F     O      9.2  13.6  14.6  16.7  20.2
## 3 M     B      6.7  10.6  11.7  13.3  15.8
## 4 M     O      6.9  10.8  12.1  13.7  16.8

```

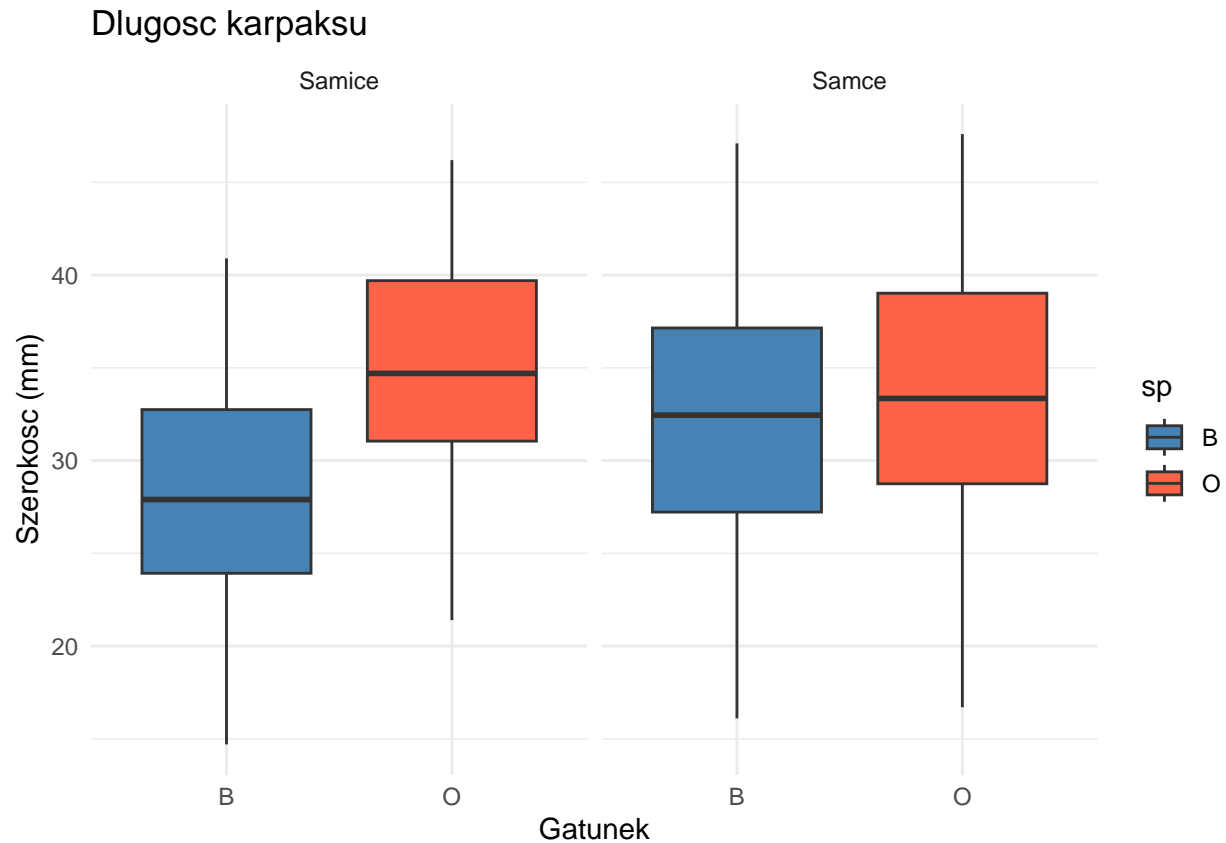
Gatunek krabów O charakteryzuje się wyraźnie większą średnią i bardziej jednorodną tylną szerokością pancerza (RW) niż gatunek B, zarówno u samic, jak i samców. Kraby gatunku O mają większe rozmiary w tej cesze, a ich rozrzut wielkości jest mniejszy, co sugeruje mniejszą zmienność wewnątrz grupy. Z kolei gatunek B wykazuje większą zmienność, co może wskazywać na większą różnorodność rozmiarów w tej populacji.

2.3 Jak różnią się pod względem długości karapaksu? (zmienna CL)

```

ggplot(crabs, aes(x = sp, y = CL, fill = sp)) +
  geom_boxplot() +
  scale_fill_manual(values = c("#4682B4", "#FF6347")) +
  labs(title = "Długość karpaksu",
       x = "Gatunek",
       y = "Szerokość (mm)") +
  theme_minimal() +
  facet_wrap(~sex, labeller = labeller(sex = c("F" = "Samice", "M" = "Samce")))

```



```
stats_crabs_CL <- crabs %>%
  group_by(sex, sp) %>%
  summarise(
    Min = min(CL),
    Q1 = quantile(CL, 0.25),
    Median = median(CL),
    Q3 = quantile(CL, 0.75),
    Max = max(CL),
    .groups = "drop"
  )
```

```
stats_crabs_CL
```

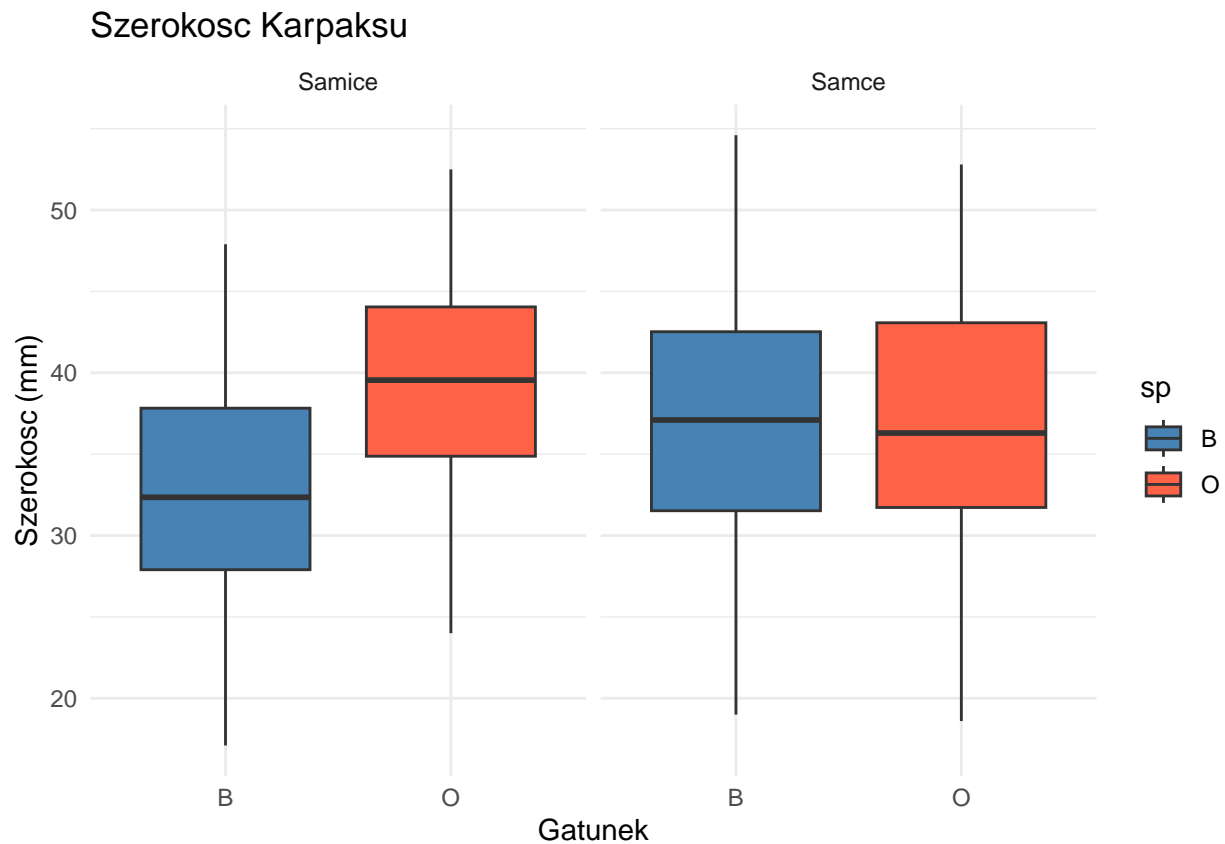
```
## # A tibble: 4 x 7
##   sex  sp      Min    Q1 Median    Q3    Max
##   <fct> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 F    B      14.7  23.9  27.9  32.8  40.9
## 2 F    O      21.4  31.0  34.7  39.7  46.2
## 3 M    B      16.1  27.2  32.4  37.2  47.1
## 4 M    O      16.7  28.8  33.3  39.0  47.6
```

Samice gatunku B mają długość karapaksu w zakresie od 14,7 mm do 40,9 mm, podczas gdy samice gatunku O mają mniejszy zakres – od 21,4 mm do 46,2 mm. U samców gatunku B długość karapaksu waha się od 16,1 mm do 47,1 mm, natomiast w gatunku O mieści się w przedziale od 16,7 mm do 47,6 mm. Choć gatunek B ma szerszy zakres długości, kraby gatunku O osiągają większe rozmiary. Mediana długości karapaksu dla

samic gatunku O to 34,7 mm, a dla samic gatunku B wynosi 27,9 mm. U samców mediana dla gatunku O jest średnio większy to 33,35 mm, a dla gatunku B 32,45 mm. Oznacza to, że karapaks krabów gatunku O jest średnio większy i bardziej jednorodny, podczas gdy gatunek B wykazuje większą zmienność w rozmiarach.

2.4 Jak różnią się pod względem szerokości karapaksu (zmienna CW)

```
ggplot(crabs, aes(x = sp, y = CW, fill = sp)) +
  geom_boxplot() +
  scale_fill_manual(values = c("#4682B4", "#FF6347")) +
  labs(title = "Szerokość Karpaksu",
       x = "Gatunek",
       y = "Szerokość (mm)") +
  theme_minimal() +
  facet_wrap(~sex, labeller = labeller(sex = c("F" = "Samice", "M" = "Samce")))
```



```
stats_crabs_CW <- crabs %>%
  group_by(sex, sp) %>%
  summarise(
    Min = min(CW),
    Q1 = quantile(CW, 0.25),
    Median = median(CW),
    Q3 = quantile(CW, 0.75),
    Max = max(CW),
```

```

    .groups = "drop"
  )

stats_crabs_CW

```

```

## # A tibble: 4 x 7
##   sex    sp      Min      Q1 Median      Q3      Max
##   <fct> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 F      B     17.1  27.9  32.4  37.8  47.9
## 2 F      O     24    34.9  39.6  44.1  52.5
## 3 M      B     19    31.5  37.1  42.5  54.6
## 4 M      O     18.6  31.7  36.3  43.1  52.8

```

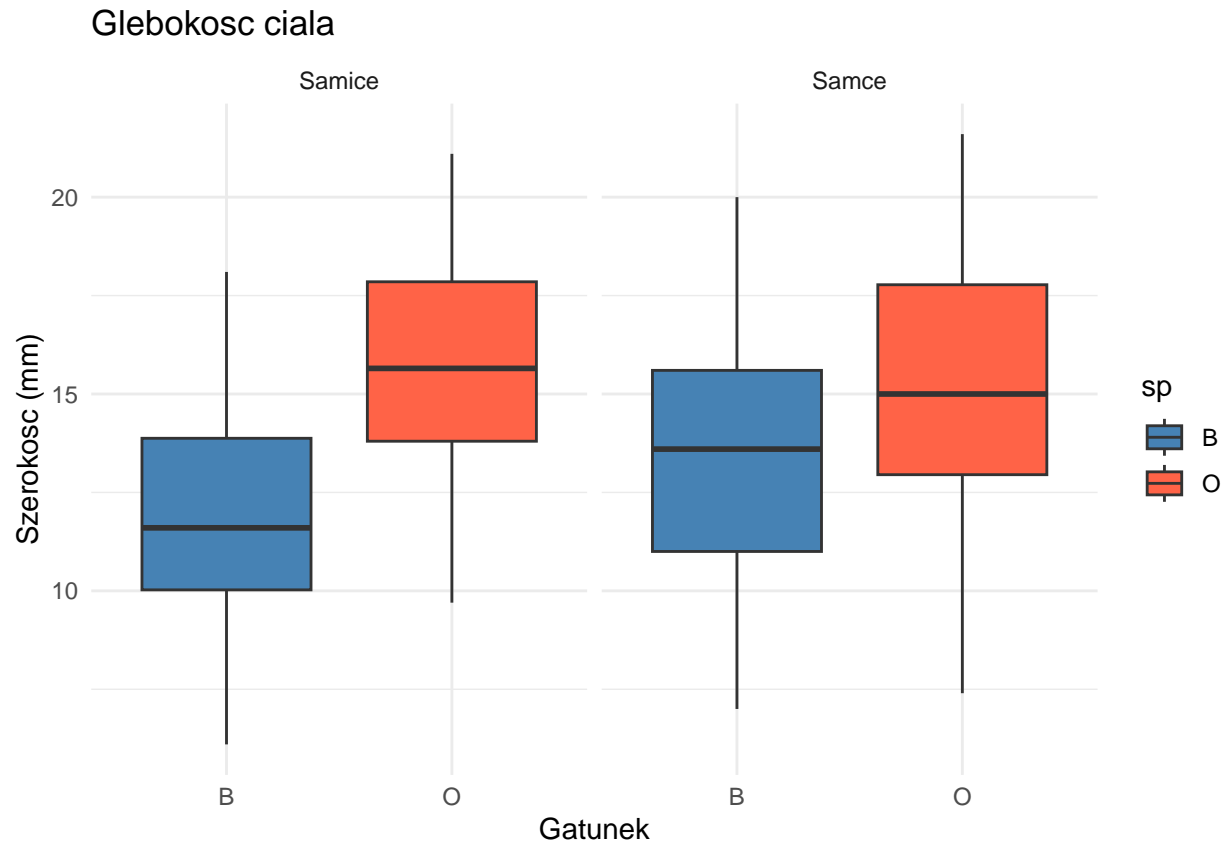
W przypadku szerokości karapaksu (CW) największe różnice między gatunkami O i B występują u samic, ale także u samców można zauważyć pewne różnice. U samic gatunku O mediana szerokości karapaksu wynosi 39.55 mm, co jest znacznie wyższe niż 32.35 mm u samic gatunku B. Z kolei wśród samców gatunku O mediana wynosi 36.30 mm, natomiast u samców gatunku B jest wyższa i wynosi 37.10 mm. Gatunek O cechuje się również szerszym zakresem szerokości karapaksu w porównaniu do gatunku B. U samic O wartości te mieszczą się w przedziale od 24.0 mm do 52.5 mm, podczas gdy samice B mają mniejszy zakres – od 17.1 mm do 47.9 mm. Podobnie u samców, szerokość karapaksu w gatunku O wynosi od 18.6 mm do 52.8 mm, a w gatunku B od 19.0 mm do 54.6 mm.

2.5 Jak różnią się pod względem głębokości ciała? (zmienna BD)

```

ggplot(crabs, aes(x = sp, y = BD, fill = sp)) +
  geom_boxplot() +
  scale_fill_manual(values = c("#4682B4", "#FF6347")) +
  labs(title = "Głębokość ciała",
       x = "Gatunek",
       y = "Szerokość (mm)") +
  theme_minimal() +
  facet_wrap(~sex, labeller = labeller(sex = c("F" = "Samice", "M" = "Samce")))

```



```
stats_crabs_BD <- crabs %>%
  group_by(sex, sp) %>%
  summarise(
    Min = min(BD),
    Q1 = quantile(BD, 0.25),
    Median = median(BD),
    Q3 = quantile(BD, 0.75),
    Max = max(BD),
    .groups = "drop"
  )
```

```
stats_crabs_BD
```

```
## # A tibble: 4 x 7
##   sex  sp      Min    Q1 Median    Q3    Max
##   <fct> <fct> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 F    B      6.1  10.0  11.6  13.9  18.1
## 2 F    O      9.7  13.8  15.6  17.8  21.1
## 3 M    B       7    11    13.6  15.6  20
## 4 M    O      7.4  13.0  15    17.8  21.6
```

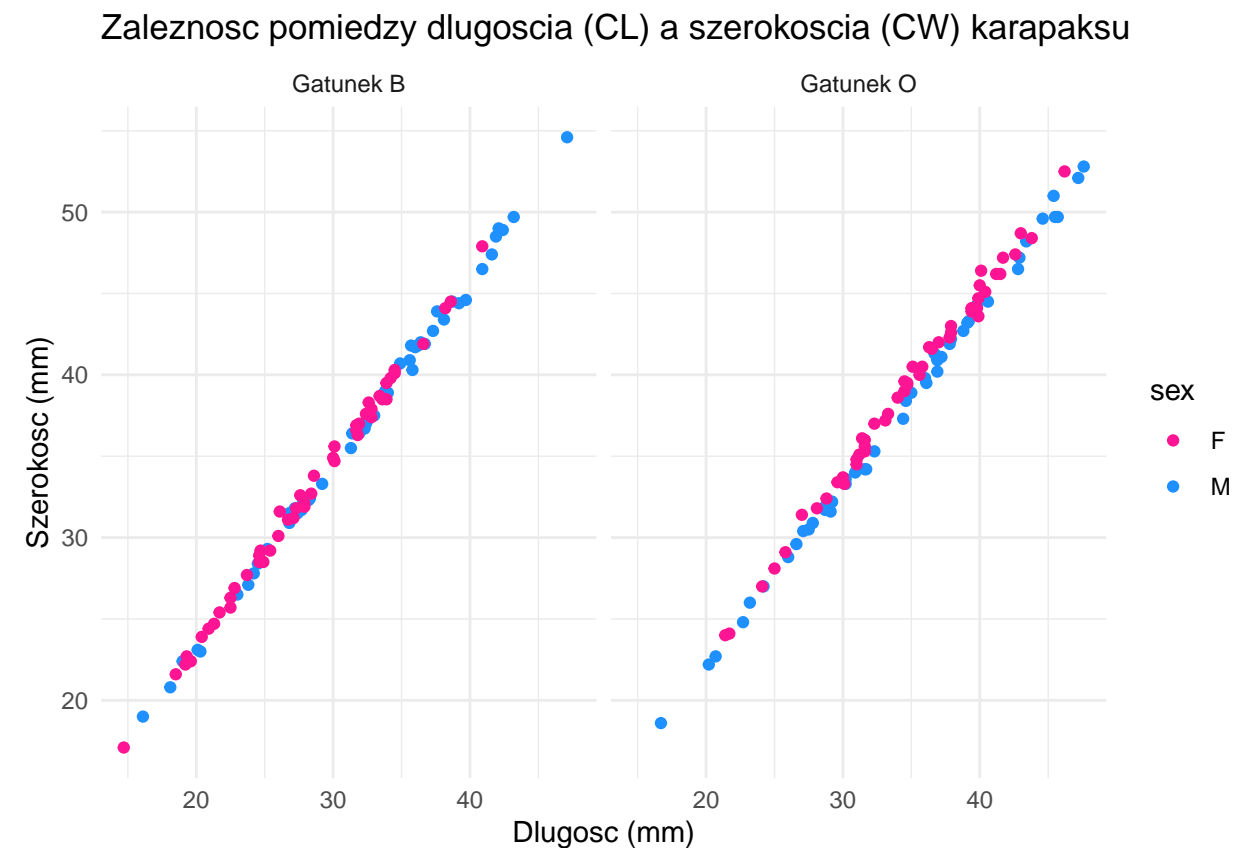
Głębokość ciała (BD) w przypadku badanych krabów różni się pomiędzy gatunkami, przy czym gatunek O wykazuje większe wartości niż gatunek B, zarówno u samic, jak i samców. Dla samic gatunku O mediana głębokości ciała wynosi 15,65 mm, podczas gdy dla samic gatunku B jest to 11,60 mm. U samców głębokość ciała w gatunku O ma medianę równą 15,00 mm, a w gatunku B 13,60 mm. Zakres głębokości ciała u samic

gatunku O waha się od 9,7 mm do 21,1 mm, natomiast w gatunku B rozciąga się od 6,1 mm do 18,1 mm. W przypadku samców, głębokość ciała w gatunku O mieści się w przedziale od 7,4 mm do 21,6 mm, podczas gdy w gatunku B wynosi od 7,0 mm do 20,0 mm. Te różnice wskazują na to, że kraby gatunku O mają większą i bardziej jednorodną głębokość ciała, w porównaniu do krabów gatunku B, które wykazują szerszy zakres wartości.

3. Zależności między zmiennymi

3.1 Jaka jest zależność pomiędzy wielkością płata czołowego (FL) a szerokością karapaksu (CW)?

```
ggplot(crabs, aes(x = CL, y = CW, color = sex)) +
  geom_point() +
  scale_color_manual(values = c("F" = "#FF1493", "M" = "#1E90FF")) +
  labs(title = "Zależność pomiędzy długością (CL) a szerokością (CW) karapaksu",
       x = "Długość (mm)",
       y = "Szerokość (mm)") +
  theme_minimal() +
  facet_wrap(~sp, labeller = labeller(sp = c("B" = "Gatunek B", "O" = "Gatunek O")))
```



```
crabs %>%
  filter(sp %in% c("O", "B"), sex %in% c("F", "M")) %>%
```



```
group_by(sp, sex) %>%
summarise(correlation = cor(CL, CW, method = "pearson", use = "complete.obs"))
```

```
## 'summarise()' has grouped output by 'sp'. You can override using the '.groups'
## argument.
```

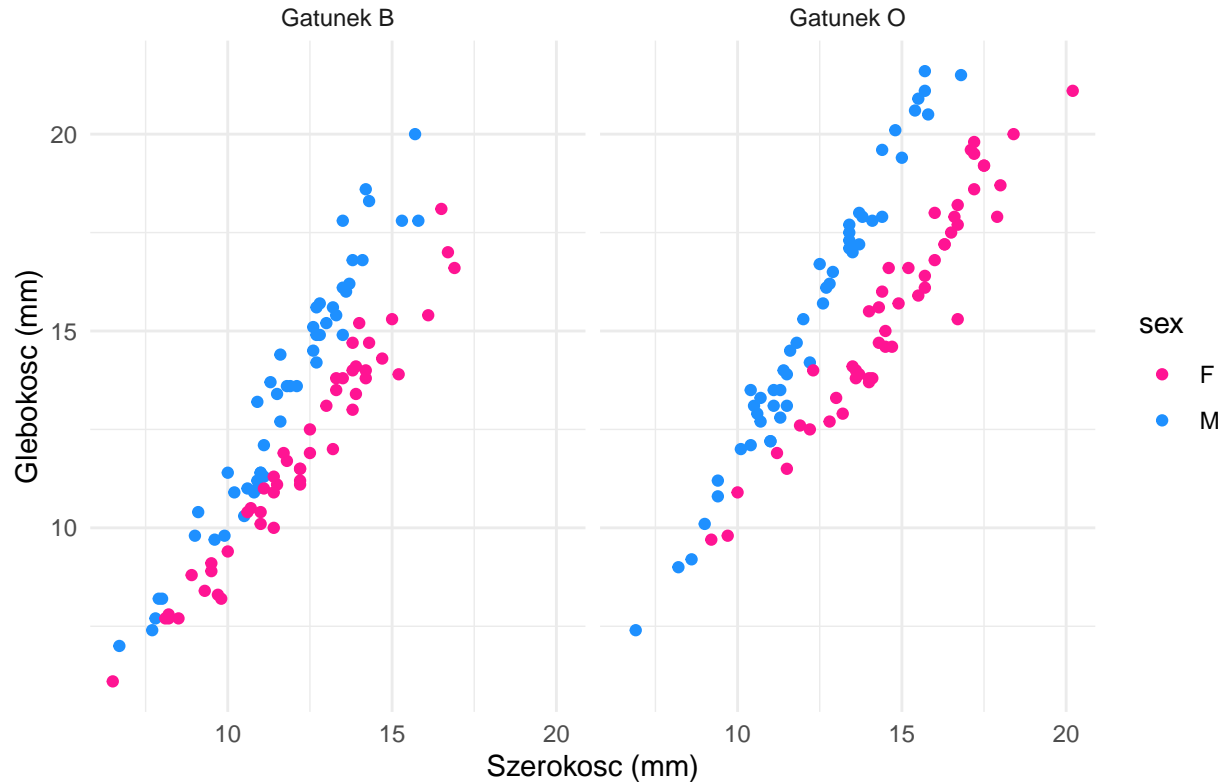
```
## # A tibble: 4 x 3
## # Groups:   sp [2]
##   sp    sex correlation
##   <fct> <fct>      <dbl>
## 1 B     F         0.998
## 2 B     M         0.999
## 3 O     F         0.997
## 4 O     M         0.999
```

Korelacje wśród samic i samców oraz pomiędzy gatunkami B i O są bardzo podobne co wskazuje na mniej więcej tak samo mocną zależność pomiędzy długością a szerokością karapaksu. Na podstawie jednej z tych zmiennych (np. CL) możemy bardzo dokładnie przewidzieć drugą zmienną (CW), ponieważ zmienne są silnie skorelowane w obrębie wszystkich grup.

3.2 Jaka jest zależność pomiędzy szerokością tylnej części kraba (RW) a głębokością ciała (BD)?

```
ggplot(crabs, aes(x = RW, y = BD, color = sex)) +
  geom_point() +
  scale_color_manual(values = c("F" = "#FF1493", "M" = "#1E90FF")) +
  labs(title = "Zależność pomiędzy szerokością tylniej części kraba (RW) a głębokością ciała (BD) ",
        x = "Szerokość (mm)",
        y = "Głębokość (mm)") +
  theme_minimal() +
  facet_wrap(~sp, labeller = labeller(sp = c("B" = "Gatunek B", "O" = "Gatunek O")))
```

Zależność pomiędzy szerokością tylnej części kraba (RW) a głębokością ciała



```
crabs %>%
  filter(sp %in% c("O", "B"), sex %in% c("F", "M")) %>%
  group_by(sp, sex) %>%
  summarise(correlation = cor(RW, BD, method = "pearson", use = "complete.obs"))
```

```
## 'summarise()' has grouped output by 'sp'. You can override using the '.groups'
## argument.
```

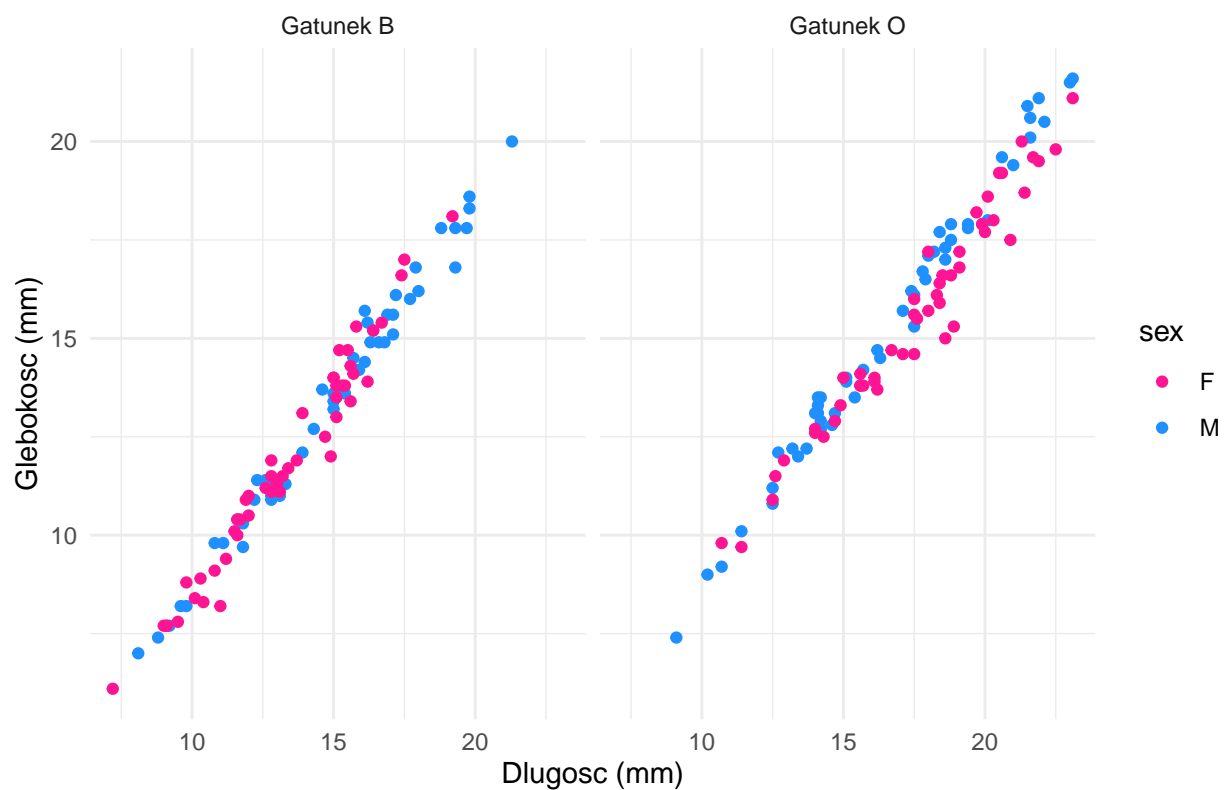
```
## # A tibble: 4 x 3
## # Groups:   sp [2]
##   sp    sex correlation
##   <fct> <fct>      <dbl>
## 1 B     F         0.977
## 2 B     M         0.969
## 3 O     F         0.961
## 4 O     M         0.987
```

Wszystkie wartości korelacji między tylną szerokością pancerza (RW) a głębokością ciała (BD) w różnych grupach (gatunki i płcie) wskazują na silną dodatnią korelację w zakresie od 0.96 do 0.99. Oznacza to, że zmienne te rosną razem, tzn. w miarę wzrostu jednej z nich (np. RW), druga (BD) również rośnie w sposób proporcjonalny. Najsilniejsza korelacja została zaobserwowana u samców gatunku O, gdzie wartość wynosi 0.9874, co wskazuje na niemal idealną liniową zależność między tymi dwoma zmiennymi. Choć korelacje są silne we wszystkich grupach, samice gatunku O wykazują najniższą wartość korelacji (0.9609)

3.3 Jaka jest zależność pomiędzy długością ciała (FL) a głębokością ciała kraba (BD)?

```
ggplot(crabs, aes(x = FL, y = BD, color = sex)) +  
  geom_point() +  
  scale_color_manual(values = c("F" = "#FF1493", "M" = "#1E90FF")) +  
  labs(title = "Zależność pomiędzy rozmiarem płata czołowego (FL) a głębokością ciała kraba (BD)",  
        x = "Długość (mm)",  
        y = "Głębokość (mm)") +  
  theme_minimal() +  
  facet_wrap(~sp, labeller = labeller(sp = c("B" = "Gatunek B", "O" = "Gatunek O")))
```

Zależność pomiędzy rozmiarem płata czołowego (FL) a głębokością ciała krab



```
crabs %>%  
  filter(sp %in% c("O", "B"), sex %in% c("F", "M")) %>%  
  group_by(sp, sex) %>%  
  summarise(correlation = cor(FL, BD, method = "pearson", use = "complete.obs"))
```

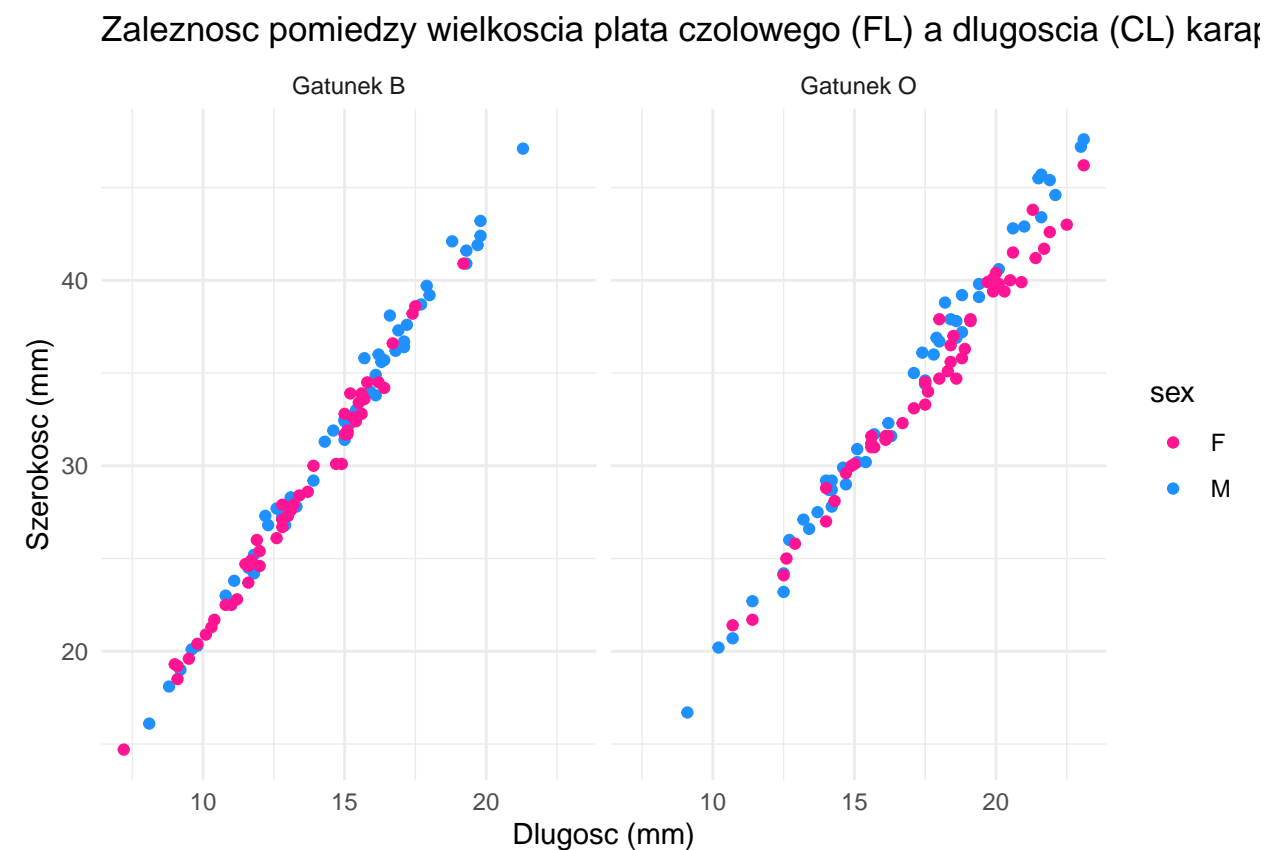
```
## 'summarise()' has grouped output by 'sp'. You can override using the '.groups'  
## argument.
```

```
## # A tibble: 4 x 3  
## # Groups:   sp [2]  
##   sp    sex correlation
```

```
##   <fct> <fct>      <dbl>
## 1 B     F         0.982
## 2 B     M         0.992
## 3 O     F         0.979
## 4 O     M         0.994
```

3.4 Jaka jest zależność pomiędzy wielkością płata czołowego (FL) a długością karapaksu (CL)?

```
ggplot(crabs, aes(x = FL, y = CL, color = sex)) +
  geom_point() +
  scale_color_manual(values = c("F" = "#FF1493", "M" = "#1E90FF")) +
  labs(title = "Zależność pomiędzy wielkością płata czołowego (FL) a długością (CL) karapaksu",
       x = "Długość (mm)",
       y = "Szerokość (mm)") +
  theme_minimal() +
  facet_wrap(~sp, labeller = labeller(sp = c("B" = "Gatunek B", "O" = "Gatunek O")))
```



```
crabs %>%
  filter(sp %in% c("O", "B"), sex %in% c("F", "M")) %>%
  group_by(sp, sex) %>%
  summarise(correlation = cor(FL, CL, method = "pearson", use = "complete.obs"))
```

```
## 'summarise()' has grouped output by 'sp'. You can override using the '.groups'
## argument.
```

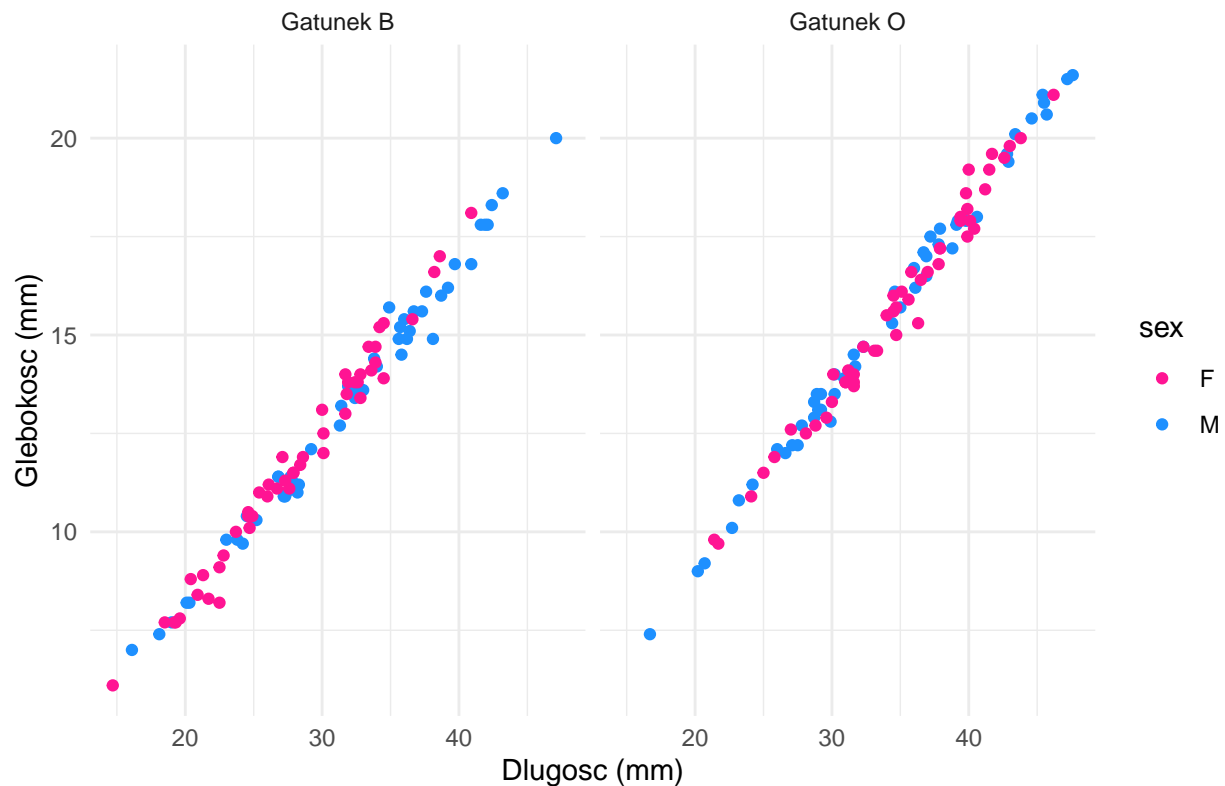
```
## # A tibble: 4 x 3
## # Groups:   sp [2]
##   sp    sex correlation
##   <fct> <fct>         <dbl>
## 1 B     F           0.995
## 2 B     M           0.995
## 3 O     F           0.990
## 4 O     M           0.994
```

Korelacja pomiędzy wielkością płata czołowego (FL) a długością ciała (CL) w różnych grupach jest bardzo silna i dodatnia w każdym przypadku. Najwyższą korelację zaobserwowano u samic gatunku B, gdzie wynosi ona 0,9955, co sugeruje niemal idealną liniową zależność między FL a CL w tej grupie. Również u samców gatunku B (0,9948) oraz samców gatunku O (0,9943) korelacja jest wyjątkowo wysoka. Dla samic gatunku O korelacja wynosi 0,9896, co jest nieco niższą wartością, ale wciąż wskazuje na bardzo silną liniową zależność między tymi dwoma cechami.

3.5 Jaka jest zależność pomiędzy długością karapaksu (CL) a głębokością ciała kraba (BD)?

```
ggplot(crabs, aes(x = CL, y = BD, color = sex)) +
  geom_point() +
  scale_color_manual(values = c("F" = "#FF1493", "M" = "#1E90FF")) +
  labs(title = "Zależność długością karapaksu (CL) a głębokością ciała kraba (BD)",
       x = "Długość (mm)",
       y = "Głębokość (mm)") +
  theme_minimal() +
  facet_wrap(~sp, labeller = labeller(sp = c("B" = "Gatunek B", "O" = "Gatunek O")))
```

Zależność długości karapaksu (CL) a głębokości ciała kraba (BD)



```
crabs %>%
  filter(sp %in% c("O", "B"), sex %in% c("F", "M")) %>%
  group_by(sp, sex) %>%
  summarise(correlation = cor(CL, BD, method = "pearson", use = "complete.obs"))
```

```
## 'summarise()' has grouped output by 'sp'. You can override using the '.groups'
## argument.
```

```
## # A tibble: 4 x 3
## # Groups:   sp [2]
##   sp    sex correlation
##   <fct> <fct>      <dbl>
## 1 B     F         0.990
## 2 B     M         0.994
## 3 O     F         0.990
## 4 O     M         0.997
```

Podobnie jak wyżej Wartości korelacji pomiędzy długością ciała (CL) a głębokością ciała (BD) w różnych grupach wskazują na silną dodatnią korelację we wszystkich przypadkach. Wartości korelacji wynoszą od 0.9901 do 0.9969, co oznacza, że zmienne te mają bardzo silną liniową zależność,