

# Badanie jakości życia wybranych państw przy użyciu metod porządkowania liniowego oraz analizy skupień

## 1. Cel pracy:

Niewątpliwie na jakość życia ma wpływ mnóstwo czynników. Oczywiście różnica w jej ocenie może być subiektywna. Ja natomiast swojej pracy postaram się na podstawie kilku zmiennych, obejmujących kluczowe dziedziny życia w kontekście atrakcyjności danego kraju takie jak: zdrowie, środowisko, życie społeczne czy poziom zamożności ocenić jakość bytowania w każdym z nich. Państwa ujęte w analizie reprezentują każdy kontynent, są różnorodne kulturowo oraz etnicznie. Wszystkie dane obejmują 2022 rok.

## 2. Opis danych

Zmienne użyte w analizie:

- **life\_expectancy (stymulanta)** – jest to zmienna mówiąca o oczekiwanej długości życia. Jej wysoka wartość świadczy o wysokim poziomie opieki zdrowotnej, odpowiedniej jakości i dostępności jedzenia oraz o dobrym stanie środowiska
- **gdp\_per\_capita (stymulanta)** – Produkt krajowy brutto w przeliczeniu na jedną osobę (\$). W jego ustalaniu wzięto pod uwagę zmiany w sile nabywczej pieniądza w danym kraju. Jest to zmienna mówiąca o zamożności danego społeczeństwa. Im wyższa jej wartość tym kraj jest bogatszy
- **unemployment\_rate (destymulanta)** – stopa bezrobocia (%). Jej wysoka wartość jest niepożądanym zjawiskiem, świadczy o tym że na rynku brakuje wolnych stanowisk, co wiąże się z potencjalnym wzrostem ubóstwa w danym kraju
- **crime\_index (destymulanta)** – indeks przestępczości, to oszacowanie ogólnego poziomu przestępczości w danym kraju. Im niższa jego wartość tym potencjalnie w danym kraju jest bezpieczniej
- **air\_quality\_index (destymulanta)** - o wskaźnik opracowany przez agencje rządowe w celu informowania społeczeństwa o obecnym stopniu zanieczyszczenia powietrza lub przewidywanym poziomie jego zanieczyszczenia. Jego wysokie wartości bardzo niekorzystnie wpływają na

zdrowie człowieka, szczególnie na jego drogi oddechowe, mogąc w przyszłości prowadzić do wielu powikłań

- **gini\_index (destymulanta)** - wskaźnik Giniego (%), oparty na krzywej Lorenza, pokazuje nierówności w dochodach danego społeczeństwa – należy interpretować go w ten sposób, że im jest wyższy, tym nierówności w dochodach w danym państwie są większe

### 3. Wstępna analiza wykorzystywanych danych

#### Podstawowe statystyki:

	life_expectancy	gdp_per_capita	unemploymnet_rate	crime_index	air_quality_index	gini_index
count	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000
mean	77.650000	26670.125000	7.203750	46.479167	27.083333	40.641667
std	5.524019	22517.586397	6.297999	13.456079	19.978068	10.583659
min	65.500000	2184.000000	2.600000	22.100000	6.000000	25.600000
25%	74.600000	6739.500000	3.675000	38.800000	14.750000	32.300000
50%	78.350000	16374.000000	5.680000	46.300000	25.000000	37.650000
75%	82.300000	46033.750000	7.372500	53.725000	28.750000	51.200000
max	85.200000	75269.000000	33.500000	76.100000	96.000000	58.800000

Tabela przedstawia podstawowe statystyki opisowe odnoszące się do naszych zmiennych.

W przypadku średniej długości życia średnia równa 77,65 jest dość przybliżona do mediany która wynosi 78,35. Najkrótsze wartości oczekiwanej długości trwania życia można przypisać Nigerii, natomiast największe Japonii.

W przypadku PKB per capita średnia (26670,125) znacznie różni się od mediany (16374,00). Wpływ na to może mieć bardzo wysoka wartość PKB krajów najbogatszych w naszym zestawieniu. Najmniejszy dochód krajowy brutto per capita jest w Nigerii, a największy w USA.

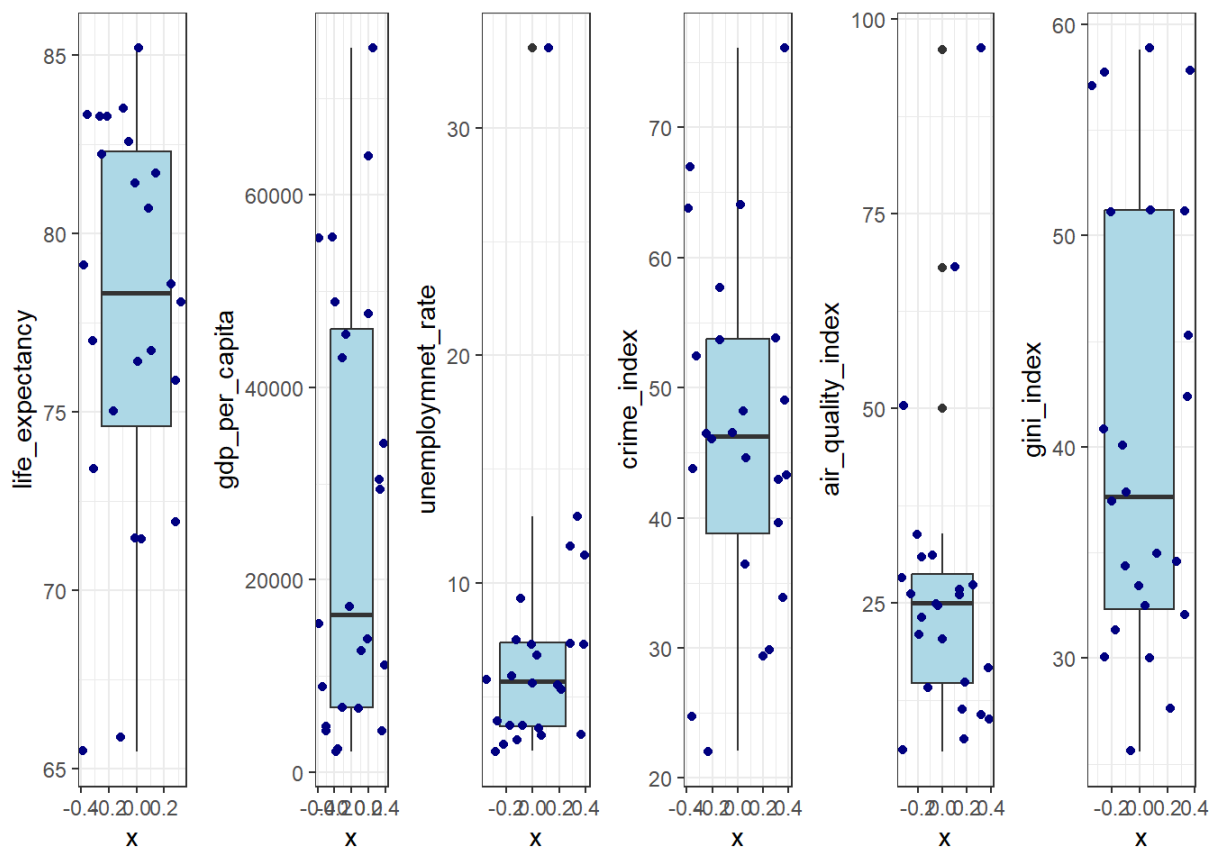
W przypadku stopy bezrobocia znów średnia przewyższa medianę. Ma to związek z istniejącymi wartościami odstającymi. Największe wartości przypisujemy Japonii równe 2,6%, a największe RPA równe 33,5%

Indeks przestępczości Ma stosunkowo równe wartości mediany i średniej. Największą przestępczości odnotowuje się w Południowej Afryce, a najmniejszą w Japonii

Indeks jakości powietrza również posiada przybliżone wartości średniej i mediany. Wyraźnie w oczy rzucają się państwa w których wartość indeksy jest największa np. Chiny (96)

Współczynnik Giniego również posiada przybliżoną wartość mediany i średniej. Do państw o największej nierówności społecznej należą: Nigeria (58,8) i Kolumbia (57,8), a najmniejsza nierówność szacowana jest w Niemczech (25,6)

### Wykresy pudełkowe dla cech:



Na powyższym wykresie zaprezentowane zostały wykresy pudełkowe wszystkich zmiennych.

Na pierwszym wykresie widzimy dane dotyczące przewidywanej długości życia. Jak łatwo można zauważyć różnica między medianą i pierwszym kwartylem a medianą i czwartym kwartylem jest dość podobna. Państwa z najmniejszą oczekiwaną długością trwania życia są dość mocno oddalone od pierwszego kwartyla, natomiast te z największą są dość blisko kwartyla czwartego.

W kolejnym wykresie zobrazowana została wartość PKB per capita. Tutaj już mediana jest przybliżona zdecydowanie w stronę pierwszego kwartyla, co oznacza że państw ze wskaźnikiem PKB stosunkowo niskim jest zdecydowanie więcej niż ze wskaźnikiem stosunkowo wysokim.

Trzeci wykres to zobrazowanie stopy bezrobocia w wybranych krajach. Bardzo mocno w tym przypadku rzucają się w oczy dwie wartości odstające (outlier), które są zdecydowanie poza zasięgiem górnego wąsa i mogą być problematyczne w dalszych etapach analizy

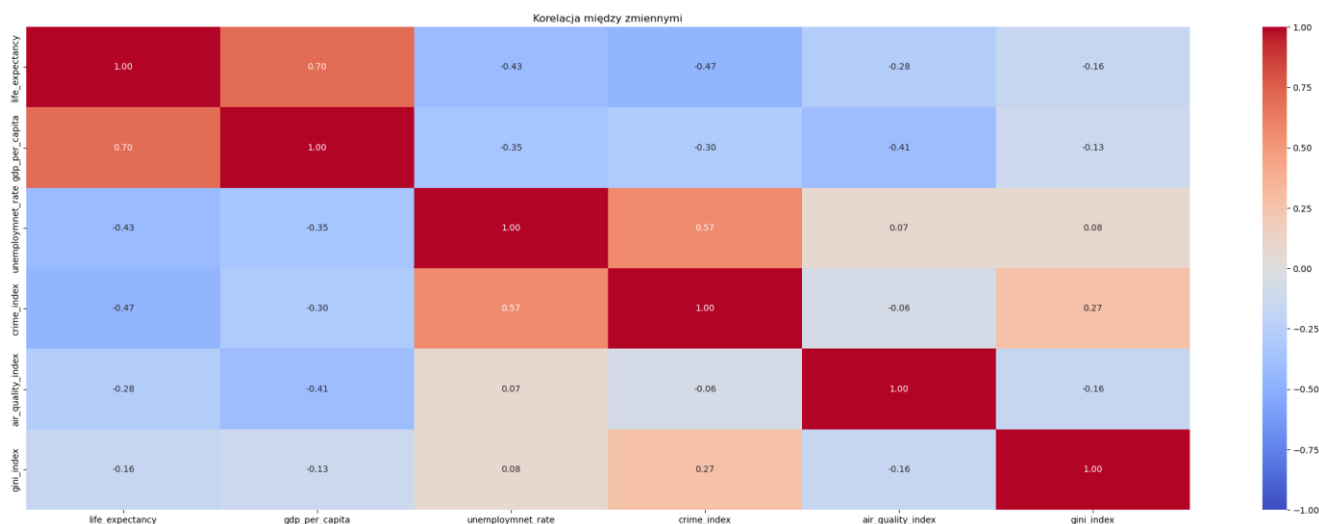
Czwarty wykres odnoszący się do indexu przestępczości jest stosunkowo symetryczny. Odległość między pierwszym i trzecim kwartylem jest równa a mediana znajdują się pomiędzy nimi.

W przypadku indeksu jakości powietrza mamy do czynienia z mocno odstającymi wartościami w przypadku 6 krajów które wyraźnie przekraczają długość naszego pudełka. Te wartości również mogą być kłopotliwe w dalszych analizach

W przypadku wykresu odnoszącego się do współczynnika Giniego możemy zauważyć stosunkowo dużą symetryczność danych. Mediana jest przesunięta w stronę pierwszego kwartyla co oznacza że więcej wartości jest do niższych wartości współczynnik niż wyższych.

## Macierz korelacji pomiędzy zmiennymi

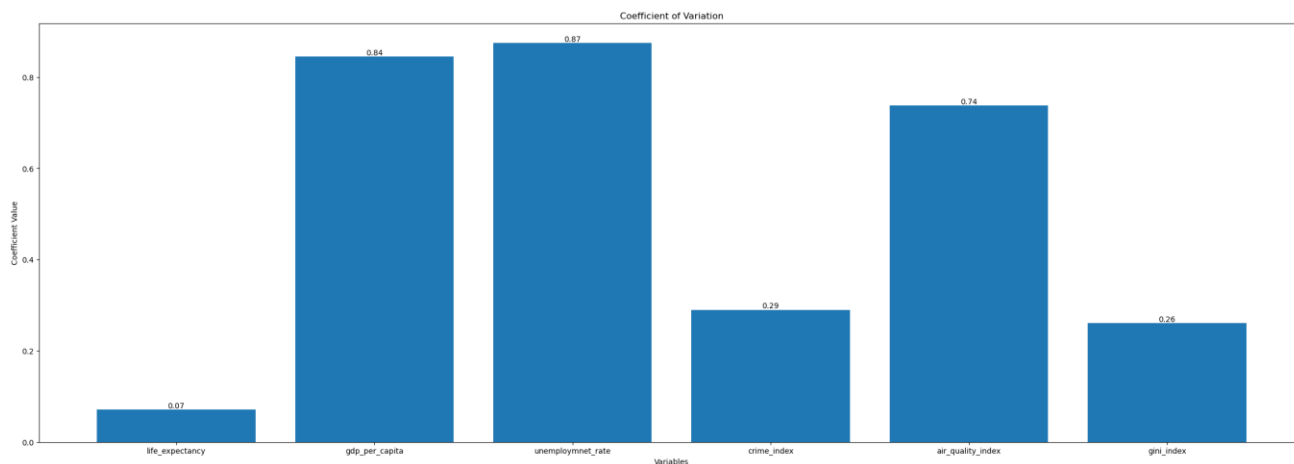
Jest to miara współzależności pomiędzy zmiennymi. Określa w jakim stopniu zmienna wartości jednej zmiana wpływa na zmianę wartości drugiej.



Korelacja w żadnym wypadku nie przekroczyła wartości bezwzględnej równej 0.9, co oznacza że nie ma potrzeby wykluczać żadnej zmiennej ze względu na stopień zależności.

## Współczynnik zmienności

Jest to miara mówiąca o stopniu zróżnicowania badanej cechy. Jego wartość powinna przekraczać 0.1



Wszystkie współczynniki oprócz oczekiwanej długości życia są większe od 0,1. W związku z tym że oczekiwaną długość życia uważam za niezwykle istotną w kontekście mojej analizy, pozwolę aby pozostała w dalszych rozważaniach mimo niskiej zmienności.

## 4. Porządkowanie liniowe

### Metoda Hellwiga

Przeprowadzę porządkowanie liniowe metodą Hellwiga. Polega ona na zmierzeniu odległości naszych obiektów od wzorca i na jej podstawie ułożeniu państw w ranking (od tych będących najbliżej wzorca do tych najdalszych ideałowi). Po dokonaniu niezbędnych obliczeń i przekształceń otrzymałem:

	country	life_expectancy	gdp_per_capita	unemploymnet_rate	crime_index	air_quality_index	gini_index	Hellwig
1	Germany	81.7	48845	3.10	36.4	23	25.6	0.739787
2	Australia	83.3	64003	3.70	43.8	11	34.4	0.735689
3	New Zealand	82.2	47680	3.30	43.3	15	27.6	0.709666
4	Japan	85.2	34135	2.60	22.1	27	34.6	0.689958
5	Canada	82.6	55646	5.30	42.9	17	40.1	0.673319
6	UK	81.4	45485	3.70	46.5	20	37.9	0.629319
7	France	83.3	43061	7.30	52.4	26	34.9	0.578065
8	Spain	83.5	29385	12.90	33.9	31	32.0	0.576826
9	Poland	78.6	17266	2.90	29.8	27	32.4	0.570539
10	Sweden	83.3	55543	7.50	49.0	11	51.2	0.539373
11	USA	79.1	75269	3.60	48.2	21	51.2	0.538959
12	Saudi Arabia	75.0	30436	5.60	24.7	6	51.2	0.498967
13	Russia	73.4	15482	3.90	39.6	8	40.8	0.464308
14	Indonesia	71.9	4788	5.90	46.1	25	30.1	0.387849
15	Mexico	75.9	11091	3.30	53.7	50	31.3	0.365596
16	Egypt	71.4	4295	7.30	46.6	34	33.4	0.352105
17	Argentina	77.0	13904	6.80	64.1	10	45.3	0.341182
18	Algeria	76.7	4274	11.60	53.8	28	42.4	0.336809
19	Colombia	80.7	6630	11.20	57.7	14	57.8	0.275088
20	China	78.1	12598	5.50	29.4	68	57.8	0.267987
21	Brazil	76.4	8918	9.30	67.0	26	57.1	0.212212
22	India	71.5	2389	7.33	44.6	96	30.1	0.132232
23	Nigeria	65.5	2184	5.76	63.8	25	58.8	0.096599
24	South Africa	65.9	6776	33.50	76.1	31	37.4	-0.095157

Ranking przedstawia kraje pod względem jakości życia ułożone malejąco. Wg. metody Hellwiga najlepiej żyje się w Niemczech i Australii. Podium zamyka Nowa Zelandia. Na ostatnich miejscach znajdują się kraje tzw. „trzeciego świata”, takie jak Indie, Nigeria czy RPA. Odległość od wzorca w ich przypadku jest bardzo daleka.

## Metoda standaryzowanych sum

W tej metodzie dla każdego obiektu sumuje się zestandaryzowane uprzednio wartości cech, a następnie konstruuje względny wskaźnik poziomu rozwoju, tzn. przekształca się go, tak aby przyjmował wartości z przedziału (0,1). Po uprzednich przekształceniach otrzymałem następujące wyniki:

	country	life_expectancy	gdp_per_capita	unemploymnet_rate	crime_index	air_quality_index	gini_index	MSS
1	Australia	83.3	64003	3.70	43.8	11	34.4	1.000000
2	Japan	85.2	34135	2.60	22.1	27	34.6	0.998941
3	Germany	81.7	48845	3.10	36.4	23	25.6	0.993848
4	New Zealand	82.2	47680	3.30	43.3	15	27.6	0.972988
5	Canada	82.6	55646	5.30	42.9	17	40.1	0.891982
6	UK	81.4	45485	3.70	46.5	20	37.9	0.847733
7	Poland	78.6	17266	2.90	29.8	27	32.4	0.831948
8	USA	79.1	75269	3.60	48.2	21	51.2	0.811524
9	France	83.3	43061	7.30	52.4	26	34.9	0.791690
10	Sweden	83.3	55543	7.50	49.0	11	51.2	0.790695
11	Spain	83.5	29385	12.90	33.9	31	32.0	0.787314
12	Saudi Arabia	75.0	30436	5.60	24.7	6	51.2	0.772287
13	Russia	73.4	15482	3.90	39.6	8	40.8	0.707884
14	Indonesia	71.9	4788	5.90	46.1	25	30.1	0.609570
15	Mexico	75.9	11091	3.30	53.7	50	31.3	0.573229
16	Argentina	77.0	13904	6.80	64.1	10	45.3	0.550181
17	Egypt	71.4	4295	7.30	46.6	34	33.4	0.529179
18	Algeria	76.7	4274	11.60	53.8	28	42.4	0.471810
19	China	78.1	12598	5.50	29.4	68	57.8	0.468103
20	Colombia	80.7	6630	11.20	57.7	14	57.8	0.461044
21	Brazil	76.4	8918	9.30	67.0	26	57.1	0.347577
22	India	71.5	2389	7.33	44.6	96	30.1	0.336785
23	Nigeria	65.5	2184	5.76	63.8	25	58.8	0.235376
24	South Africa	65.9	6776	33.50	76.1	31	37.4	0.000000

W wyniku przeprowadzenia porządkowania liniowego metodą standaryzowanych sum otrzymałem ranking państw pod względem jakości życia. Wyniki są porównywalne do metody Hellwiga. Na podium znów znalazła się Australia oraz Niemcy. Najgorszym miejscem do życia okazały się znów Indie, Nigeria i RPA.

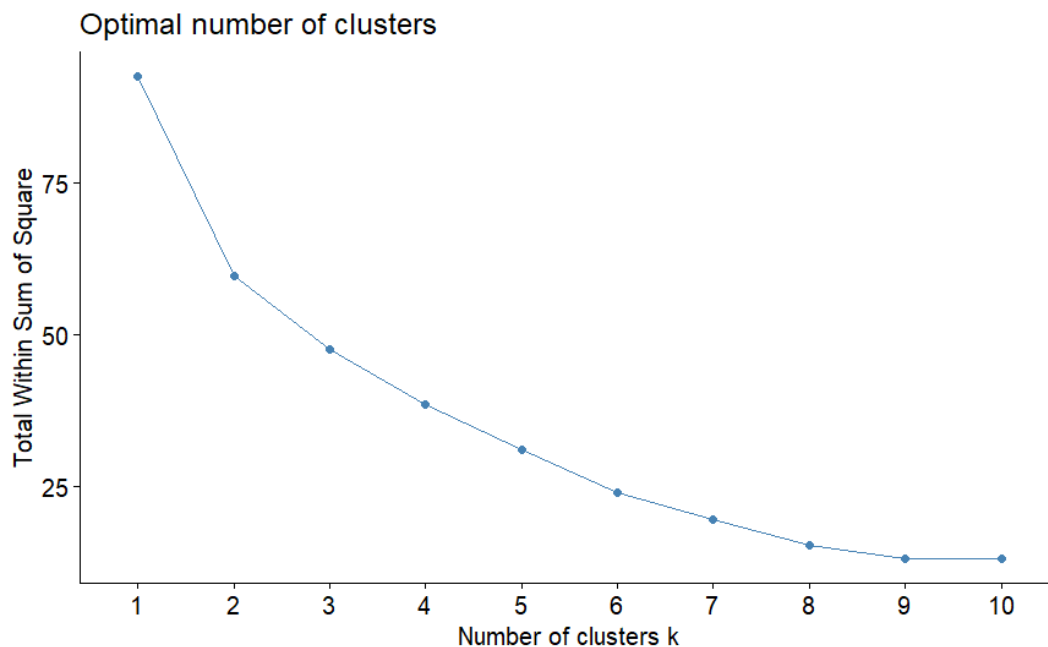
## 5. Analiza Skupień

### Metoda grupowania k-średnich

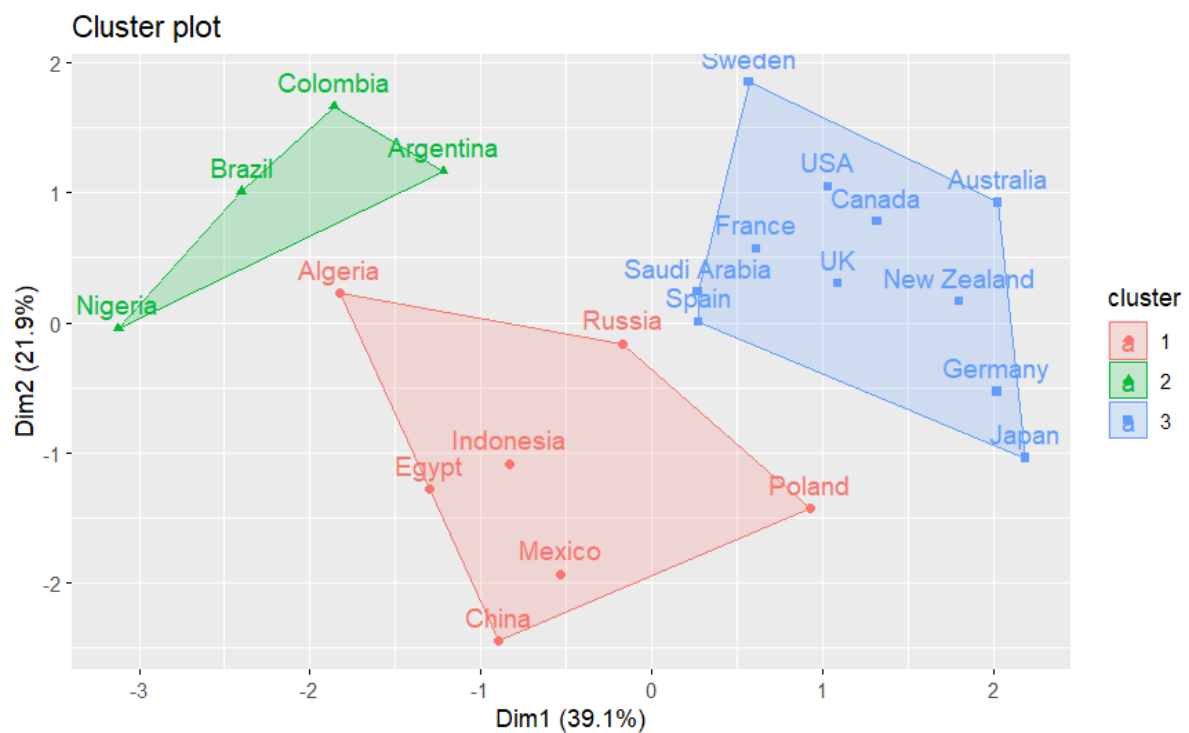
Jest to metoda polegająca na podziale zbioru danych na określoną przed początkiem badania  $k$  – rozłącznych skupień.

W trakcie przygotowań do badania należy sprawdzić czy nasze dane nie posiadają outlierów czyli obiektów znacznie odstających od reszty. W trakcie sprawdzania posłużyłem się metoda 3 sigm . W ich wyniku odkryłem, że obiektami nadającymi się do usunięcia i potencjalnie groźnymi dla wyników naszego badania są Indie i RPA.

Po ich usunięciu przechodzę do rozważań na temat liczby grup na jakie podzielę mój zbiór. W wyborze liczby skupień posłużę się metodą łokciową



Teoretycznie liczbą klastrow w której wykres zaczyna zwalniać jest 2 lecz wg mojej opinii odpowiedniejszy i bardziej oddający rzeczywistości będzie podział na 3 grupy i na taki się zdecyduje



Na podstawie podziałów policzyłem statystyki opisowe dla każdej z grup.

Statystyki dla grupy 1:



	index	life_expectancy	gdp_per_capita	unemploymnet_rate	crime_index	air_quality_index	gini_index
0	count	4.000000	4.000000	4.000000	4.00000	4.000000	4.000000
1	mean	74.900000	7909.000000	8.265000	63.15000	18.750000	54.750000
2	std	6.548791	4877.484734	2.456739	3.90939	7.973916	6.338507
3	min	65.500000	2184.000000	5.760000	57.70000	10.000000	45.300000
4	25%	73.675000	5518.500000	6.540000	62.27500	13.000000	54.150000
5	50%	76.700000	7774.000000	8.050000	63.95000	19.500000	57.450000
6	75%	77.925000	10164.500000	9.775000	64.82500	25.250000	58.050000
7	max	80.700000	13904.000000	11.200000	67.00000	26.000000	58.800000

Cechy które charakteryzują tą grupę to stosunkowo wysoka przestępczość oraz duży współczynnik Giniego, co oznacza duży stopień nierówności społecznej. Jest to charakterystyczna cecha krajów Ameryki południowej, gdzie powszechne jest zjawisko formowania się osiedli nędzy (faweli).

Statystyki dla grupy 2:

	index	life_expectancy	gdp_per_capita	unemploymnet_rate	crime_index	air_quality_index	gini_index
0	count	8.000000	8.000000	8.000000	8.000000	8.000000	8.000000
1	mean	74.687500	9022.875000	5.966250	42.950000	42.000000	37.287500
2	std	2.996873	5776.493559	2.838002	9.464218	28.248894	9.542452
3	min	71.400000	2389.000000	2.900000	29.400000	8.000000	30.100000
4	25%	71.800000	4289.750000	3.750000	37.150000	26.500000	31.000000
5	50%	74.650000	7939.500000	5.700000	45.350000	31.000000	32.900000
6	75%	77.050000	13319.000000	7.307500	48.375000	54.500000	41.200000
7	max	78.600000	17266.000000	11.600000	53.800000	96.000000	57.800000

Jest to grupa której cechami są wysoki poziom zanieczyszczania powietrza oraz niska oczekiwana długość życia. Są to kraje z średnio większym PKB od grupy 1. Cechą charakterystyczną jest również niski poziom bezrobocia

Statystyki dla grupy 3:

	index	life_expectancy	gdp_per_capita	unemploymnet_rate	crime_index	air_quality_index	gini_index
0	count	11.000000	11.000000	11.000000	11.000000	11.000000	11.000000
1	mean	81.872727	48135.272727	5.327273	40.290909	18.909091	38.245455
2	std	2.750306	14119.621688	3.013666	9.908224	7.713035	9.265675
3	min	75.000000	29385.000000	2.600000	22.100000	6.000000	25.600000
4	25%	81.550000	38598.000000	3.450000	35.150000	13.000000	33.200000
5	50%	82.600000	47680.000000	3.700000	43.300000	20.000000	34.900000
6	75%	83.300000	55594.500000	6.450000	47.350000	24.500000	45.650000
7	max	85.200000	75269.000000	12.900000	52.400000	31.000000	51.200000

Grupa trzecia jest grupą zdecydowanie najlepszą. Cechą rzucającą się w oczy jest bardzo długa oczekiwana długość życia oraz niezwykle wysokie PKB. W tej grupie zostały rozlokowane najbogatsze kraje na naszej liście. Charakteryzują się również wysoką dbałością o jakość środowiska i niską stopą bezrobocia

Podsumowując w naszym podziale można znaleźć pewne zależności.

Grupa pierwsza to grupa w której żyje się ciężko ze względu na wysoką przestępczość, niski poziom zarobków i wielkie nierówności społeczne. Zaliczają się do niej kraje Ameryki południowej i Afryki.

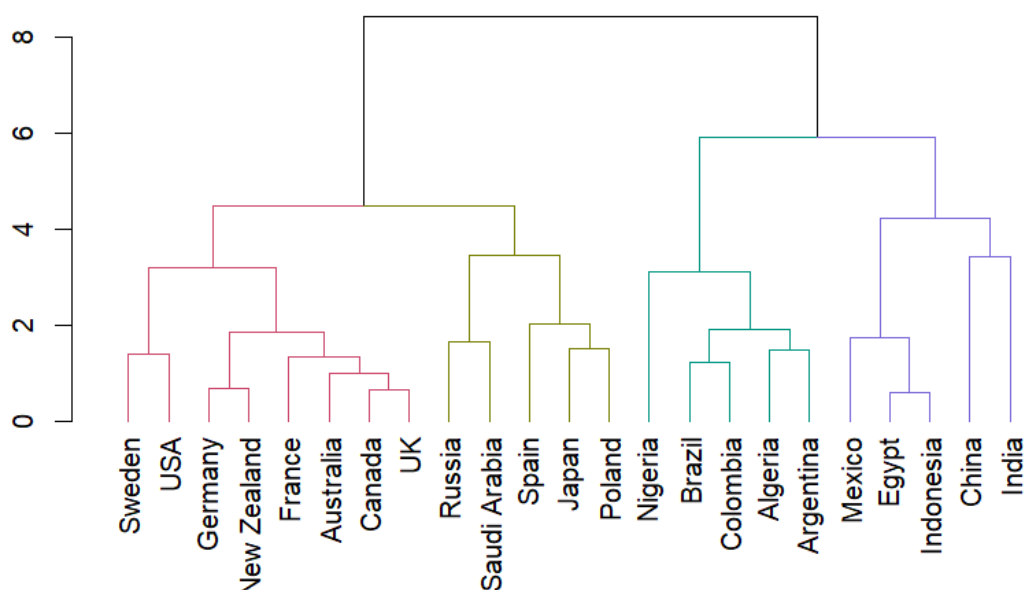
Grupa druga to kraje przemysłowe z bardzo niską stopą bezrobocia gdzie pod pewnymi względami jest lepiej niż w grupie 1, a pod pewnymi względami gorzej. Warto zauważyć że grupę tę reprezentują były kraje związku radzieckiego kraje Afryki północnej i Azji.

Grupa 3 to kraje najbogatsze z średnim PKB dużo większym od pozostałych grup. Jest to grupa krajów w których jakość życia powinna być największa ze względu na oczekiwaną długość życia, przestępczość i równość społeczną. Są to kraje tzw. „Zachodu”.

## Grupowanie hierarchiczne

W grupowaniu hierarchicznym wykorzystuje się funkcję odległości łączący każdy element zbioru w coraz większe grupy, aż do uzyskania jednego skupienia, które zawiera wszystkie obiekty (tzw. Metoda aglomeracyjna).

Po przeprowadzeniu odpowiednich obliczeń uzyskałem dendrogram:



Po przeprowadzeniu grupowania wyróżniły się nam 4 grupy. Spróbujemy je scharakteryzować

Grupa 1: Szwecja, USA, Niemcy, Nowa Zelandia, Francja, Australia, Kanada, Wielka Brytania

Statystyki dla grupy 1:

	index	life_expectancy	gdp_per_capita	unemployenet_rate	crime_index	air_quality_index	gini_index
0	count	8.00000	8.000000	8.000000	8.000000	8.000000	8.00000
1	mean	82.11250	54441.500000	4.687500	45.312500	18.000000	37.86250
2	std	1.42572	10787.853805	1.800347	4.857524	5.477226	9.54118
3	min	79.10000	43061.000000	3.100000	36.400000	11.000000	25.60000
4	25%	81.62500	47131.250000	3.525000	43.200000	14.000000	32.70000
5	50%	82.40000	52194.000000	3.700000	45.150000	18.500000	36.40000
6	75%	83.30000	57735.250000	5.800000	48.400000	21.500000	42.87500
7	max	83.30000	75269.000000	7.500000	52.400000	26.000000	51.20000

Grupa posiadająca bardzo wysoką oczekiwaną długość życia oraz bardzo duże PKB. Należą do niej najbogatsze kraje w naszym zestawieniu. Indeksy odnoszące się do jakości powietrza, równości społecznej i stopy bezrobocia na bardzo satysfakcjonujących poziomach

Grupa 2: Rosja, Arabia Saudyjska, Hiszpania, Japonia, Polska

Statystyki dla grupy 2:

	index	life_expectancy	gdp_per_capita	unemployenet_rate	crime_index	air_quality_index	gini_index
0	count	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000
1	mean	79.140000	25340.800000	5.580000	30.020000	19.800000	38.200000
2	std	5.150534	8397.215354	4.256407	7.033989	11.819475	8.074652
3	min	73.400000	15482.000000	2.600000	22.100000	6.000000	32.000000
4	25%	75.000000	17266.000000	2.900000	24.700000	8.000000	32.400000
5	50%	78.600000	29385.000000	3.900000	29.800000	27.000000	34.600000
6	75%	83.500000	30436.000000	5.600000	33.900000	27.000000	40.800000
7	max	85.200000	34135.000000	12.900000	39.600000	31.000000	51.200000

Grupa podobna do grupy pierwszej, państwa do niej należące mają średnio niższe PKB, ale jest tam znacznie bezpieczniej.

### Grupa 3: Nigeria, Brazylia, Columbia, Algieria, Argentyna

#### Statystyki dla grupy 3:

	index	life_expectancy	gdp_per_capita	unemploymnet_rate	crime_index	air_quality_index	gini_index
0	count	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000
1	mean	75.260000	7182.000000	8.932000	61.280000	20.600000	52.280000
2	std	5.728263	4526.039991	2.598292	5.380242	8.049845	7.786976
3	min	65.500000	2184.000000	5.760000	53.800000	10.000000	42.400000
4	25%	76.400000	4274.000000	6.800000	57.700000	14.000000	45.300000
5	50%	76.700000	6630.000000	9.300000	63.800000	25.000000	57.100000
6	75%	77.000000	8918.000000	11.200000	64.100000	26.000000	57.800000
7	max	80.700000	13904.000000	11.600000	67.000000	28.000000	58.800000

Grupa charakteryzujące się wysokim stopnie nierówności społecznych niskim PKB oraz wysokim współczynnikiem przestępstw. Posiada najwyższą stopę bezrobocia.

### Grupa 4: Meksyk, Egipt, Indonezja, Chiny, Indie

#### Statystyki dla grupy 4:

	index	life_expectancy	gdp_per_capita	unemploymnet_rate	crime_index	air_quality_index	gini_index
0	count	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000
1	mean	73.760000	7032.200000	5.866000	44.080000	54.600000	36.540000
2	std	3.063984	4514.974828	1.652205	8.926758	28.351367	11.960895
3	min	71.400000	2389.000000	3.300000	29.400000	25.000000	30.100000
4	25%	71.500000	4295.000000	5.500000	44.600000	34.000000	30.100000
5	50%	71.900000	4788.000000	5.900000	46.100000	50.000000	31.300000
6	75%	75.900000	11091.000000	7.300000	46.600000	68.000000	33.400000
7	max	78.100000	12598.000000	7.330000	53.700000	96.000000	57.800000

Grupa państw z bardzo zanieczyszczonym powietrzem, o porównywalnym PKB i oczekiwaną długością życia z grupą 3 za to z dużo niższym współczynnikiem nierówności społecznych

Można łatwo dojść do wniosku, że mamy tu podział na grupy gdzie jakość życia jest na wysokim poziomie( grupa 1 i grupa 2) i grupy gdzie żyje się bardzo źle( grupa 3 i 4). W zasadzie można powiedzieć w grupie 1 żyje się najlepiej, potem możemy uplasować grupę 2 ( tam mamy za to niższy indeks przestępstw. Grupy 3 i 4 nie różnią się znacząco ( w grupie 3 mamy wyższy poziom zanieczyszczeń, a w grupie 4 wyższy stopień nierówności)

## 6. Podsumowanie

Każda z przeprowadzonych analiz dawała różne wyniki, ale z łatwością można zauważyć że pewne państwa zawsze znajdowały się w grupach i klasyfikacjach przeznaczonych dla krajów z najlepszymi warunkami do życia. W większości były to państwa zachodnie, gdzie z uwagi na wysoki poziom rozwoju społecznego jakość

życia jest na bardzo wysokim poziomie. Warto również pamiętać że analiza ta zajmowała się tylko niektórymi aspektami wpływającymi na komfort funkcjonowania w wybranym miejscu świata i rzeczywistość może być odmienna. W sferze badań nad tym zagadnieniem pozostaje duże pole do rozwoju.

Autor:

Jakub Sornat