

# Analiza ekonomiczna państw

Gabriela Kuczyńska, Klaudia Grabowska i Martyna Stalmach

2023-03-30

## Spis treści

Wprowadzenie .....	1
Opis wybranych zmiennych oraz ich źródło .....	2
Podstawowe statystyki zmiennych .....	3
Sprawdzenie czy PCA jest możliwe.....	3
Korelacje między zmiennymi .....	3
Test sferyczności Bartletta .....	5
Kryterium KMO .....	5
PCA .....	6
Wnioski z PCA.....	7
Grupowanie .....	7
Wyświetlenie dendrogramu - plot na obiekcie hclust. ....	7
Grupa wraz ze zmiennymi.....	9
Graficzne przedstawienie grupy .....	10
Wnioski z grupowania .....	10
Wnioski ogólne .....	10

## Wprowadzenie

W naszym projekcie dokonaliśmy analizy wpływu zmiennych na PKB per capita wzajemnie na siebie. PKB per capita to jeden z najważniejszych wskaźników gospodarczych, który odzwierciedla wartość produkcji krajowej brutto podzielonej przez liczbę mieszkańców kraju. Jest to wskaźnik, który pokazuje poziom życia i potencjał konsumpcyjny mieszkańców kraju. Naszym celem było przeprowadzenie analizy wpływu różnych zmiennych na poziom PKB per capita, takich jak produkcja energii, wydajność pracowników, zarobki roczne, czy wskaźnik konsumpcji. W ramach projektu przeprowadziliśmy analizy statystyczne, które umożliwiły nam zidentyfikowanie zmiennych, które najbardziej wpływają na poziom PKB per capita oraz pomogły nam określić relacje między nimi. Analiza została przeprowadzana w celu zredukowania rozmiaru danych i znalezieniu między nimi zależności.

## Opis wybranych zmiennych oraz ich źródło

Przedstawienie danych w postaci tabeli razem z ich opisami i linkami do strony Eurostat:

Zmienna	Opis wraz z linkiem
<b><i>Real GDP per capita (PKB)</i></b>	Wskaźnik jest obliczany jako stosunek realnego PKB do średniej populacji w danym roku. ( <a href="https://tiny.pl/wslzh">https://tiny.pl/wslzh</a> )
<b><i>Discharges from hospitals (Zwolnienia ze szpitala)</i></b>	Zwolnienie pacjenta ze szpitala po zabiegu lub przebiegu leczenia. (liczone na 100tys. mieszkańców) ( <a href="https://tiny.pl/wslz7">https://tiny.pl/wslz7</a> )
<b><i>Internet use by individuals (Zużycie Internetu)</i></b>	Ostatnie korzystanie z Internetu. (procent) ( <a href="https://tiny.pl/wslzf">https://tiny.pl/wslzf</a> )
<b><i>Share of government budget appropriations or outlays on research and development (Budżet na badania i rozwój)</i></b>	Mierzy wsparcie rządowe na rzecz badań i rozwoju na podstawie danych zgromadzonych z budżetów (procent) ( <a href="https://tiny.pl/wslzk">https://tiny.pl/wslzk</a> )
<b><i>Energy productivity (Produkcja energii)</i></b>	Wskaźnik mierzy wielkość produkcji gospodarczej, która jest wytwarzana na jednostkę dostępnej energii brutto. ( <a href="https://tiny.pl/wslz2">https://tiny.pl/wslz2</a> )
<b><i>Labour productivity per person employed and hour worked (Wydajność pracowników)</i></b>	Obliczana jako rzeczywista wydajność na jednostkę nakładu pracy - mierzona całkowitą liczbą przepracowanych godzin. (procent) ( <a href="https://tiny.pl/wsl3h">https://tiny.pl/wsl3h</a> )
<b><i>Annual net earnings (Zarobki roczne)</i></b>	Roczne zarobki netto osoby zatrudnionej w pełnym wymiarze czasu pracy, bez dzieci, zarabiających przeciętnie (PPS) ( <a href="https://tiny.pl/wsl3q">https://tiny.pl/wsl3q</a> )
<b><i>GDP per capita in PPS (Wskaźnik konsumpcji)</i></b>	Jest wyrażony w stosunku do średniej Unii Europejskiej ustalonej na poziomie 100. (PPS) ( <a href="https://tiny.pl/wsl3g">https://tiny.pl/wsl3g</a> )

Dane wczytałyśmy z własnej bazy, którą utworzyłyśmy na podstawie danych ze strony Eurostat. Wszystkie dane są na rok 2019.

## Podstawowe statystyki zmiennych

Kraj	PKB na osobę	Zwolnienia ze szpitala	Zużycie Internetu
Length:31	Min. : 6630	Min. : 691.8	Min. :67.95
Class :character	1st Qu.:15685	1st Qu.:1628.0	1st Qu.:81.01
Mode :character	Median :25500	Median :2061.6	Median :87.75
	Mean :30717	Mean :2291.2	Mean :87.23
	3rd Qu.:38505	3rd Qu.:2950.7	3rd Qu.:95.43
	Max. :83590	Max. :4509.1	Max. :99.03

Budżet na badania i rozwój	Produkcja energii	Wydajność pracowników
Min. :0.530	Min. : 2.190	Min. : 49.0
1st Qu.:0.920	1st Qu.: 4.890	1st Qu.: 77.1
Median :1.330	Median : 7.310	Median : 98.5
Mean :1.310	Mean : 7.658	Mean : 98.6
3rd Qu.:1.585	3rd Qu.: 9.275	3rd Qu.:113.5
Max. :2.900	Max. :19.400	Max. :188.2

Zarobki roczne	Wskaźnik konsumpcji
Min. :10641	Min. : 53.0
1st Qu.:15052	1st Qu.: 76.0
Median :21985	Median : 97.0
Mean :21547	Mean :105.6
3rd Qu.:27219	3rd Qu.:123.5
Max. :40873	Max. :251.0

## Sprawdzenie czy PCA jest możliwe

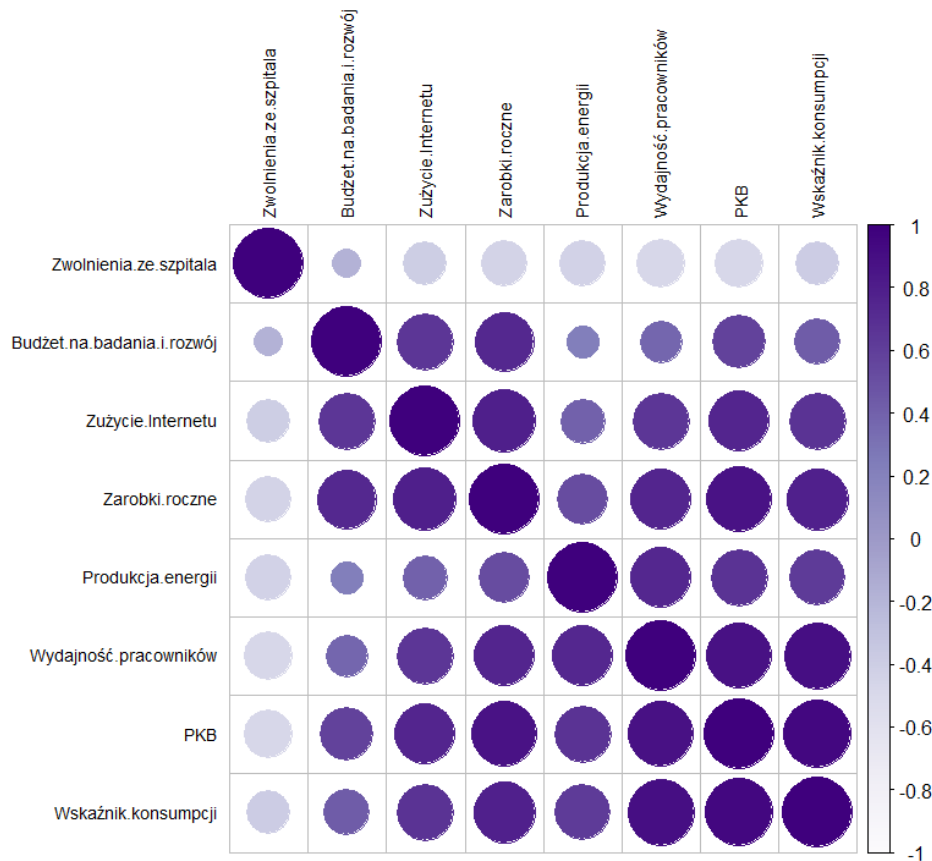
### Korelacje między zmiennymi

Poniższa tabela przedstawia korelacje zawarte między naszymi zmiennymi oraz tym jak oddziałują na wybraną zmienną y (PKP na osobę). Korelacja między zmiennymi może być mierzona za pomocą współczynnika korelacji Pearsona, który przyjmuje wartości między -1 a 1. Współczynnik korelacji Pearsona 0 oznacza brak korelacji, a wartość -1 lub 1 oznacza idealną korelację odpowiednio negatywną lub pozytywną. Można zauważyć, że korelacje y ze zmiennymi x są silne. Osiągają wartości ponad 0,50, co daje nam duże prawdopodobieństwo, że zmienne x mają znaczący wpływ na zmienną y. Korelacje między zmiennymi przedstawione zostały w Tabeli 1.

	Budżet na badania i rozwój	Zużycie Internetu	PKB na osobę	Produkcja energii	Wskaźnik konsumpcji	Wydajność pracowników	Zarobki roczne	Zwolnienia ze szpitala
PKB na osobę	0,581	0,751	1,000	0,673	0,943	0,882	0,872	-0,472
Budżet na badania i rozwój	1,000	0,659	0,581	0,220	0,437	0,378	0,734	-0,179
Użycie Internetu	0,659	1,000	0,751	0,402	0,671	0,651	0,796	-0,397
Produkcja energii	0,220	0,402	0,673	1,000	0,623	0,748	0,521	-0,432
Wskaźnik konsumpcji	0,437	0,671	0,943	0,623	1,000	0,907	0,784	-0,387
Wydajność pracowników	0,378	0,651	0,882	0,748	0,907	1,000	0,759	-0,474
Zarobki roczne	0,734	0,796	0,872	0,521	0,784	0,759	1,000	-0,442
Zwolnienia ze szpitala	-0,179	-0,397	-0,472	-0,432	-0,387	-0,474	-0,442	1,000

Tabela 1. Korelacje między zmiennymi.

PCA wykonuje się tylko wtedy, gdy między zmiennymi występują korelacje.



Wykres 1. Wykres korelacji między zmiennymi.

**WNIOSEK:** Korelacje są wyraźne, więc PCA jest możliwa do wykonania.

## Test sferyczności Bartletta

- **chisq:** 240.9
- **p.value:** 9.661e-36
- **df:** 28

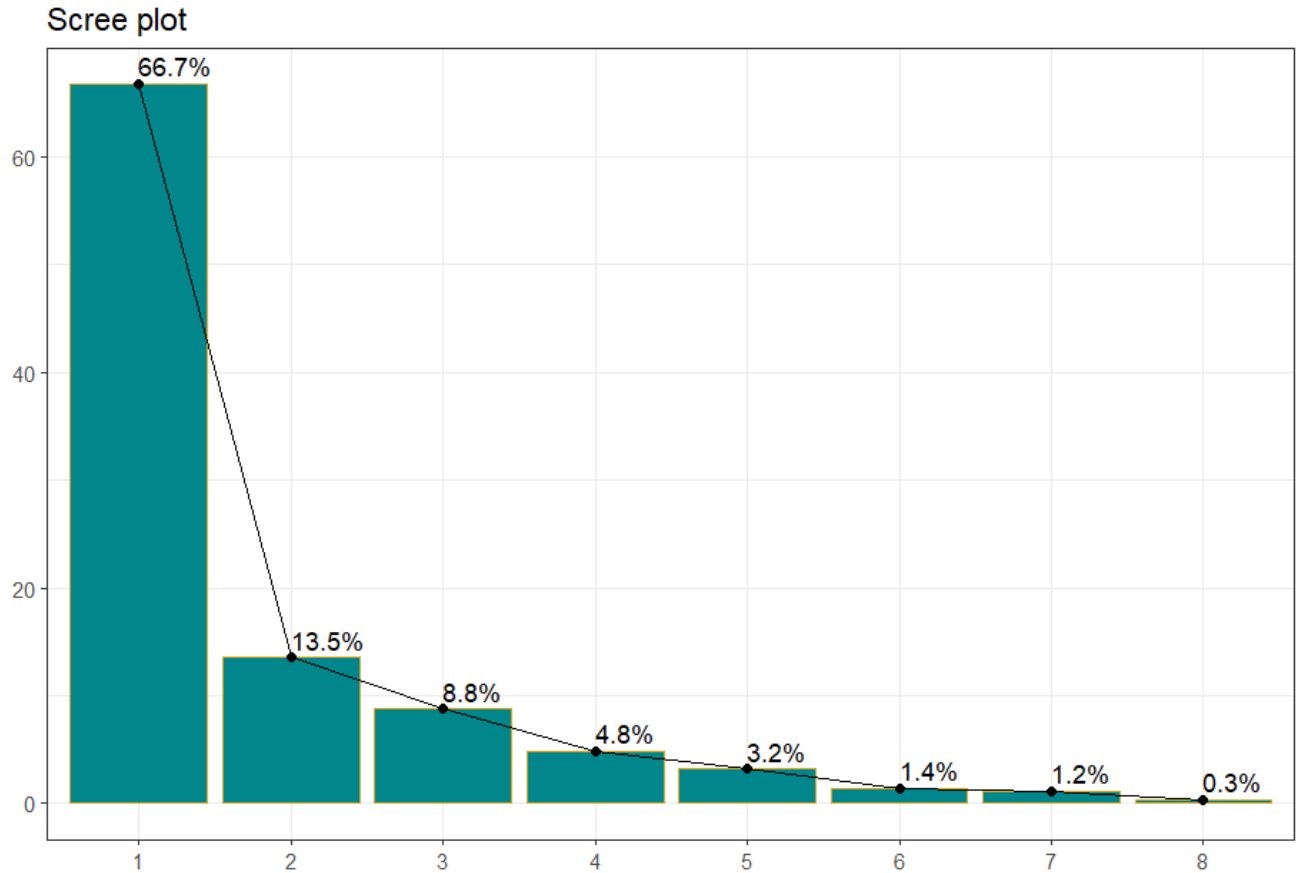
**WNIOSEK:** P-value bardzo małe, więc można odrzucić hipotezę, że macierz korelacji jest macierzą jednostkową.

## Kryterium KMO

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = cor(proj[, 2:9]))
## Overall MSA = 0.82
## MSA for each item =
##          PKB   Zwolnienia.ze.szpitala
##          0.79         0.81
##   Zużycie Internetu Budżet.na.badania.i.rozwój
##          0.94         0.77
##   Produkcja.energii   Wydajność.pracowników
##          0.76         0.83
##   Zarobki.roczne     Wskaźnik.konsumpcji
##          0.89         0.74
```

**WNIOSEK:** Powyższe mocne korelacje oznaczają dużą szansę na powodzenie PCA.  $KMO > 0.5$ , więc PCA jest jak najbardziej dopuszczalna.

## PCA



Wykres 2. Wykres zawartości informacji w składowych.

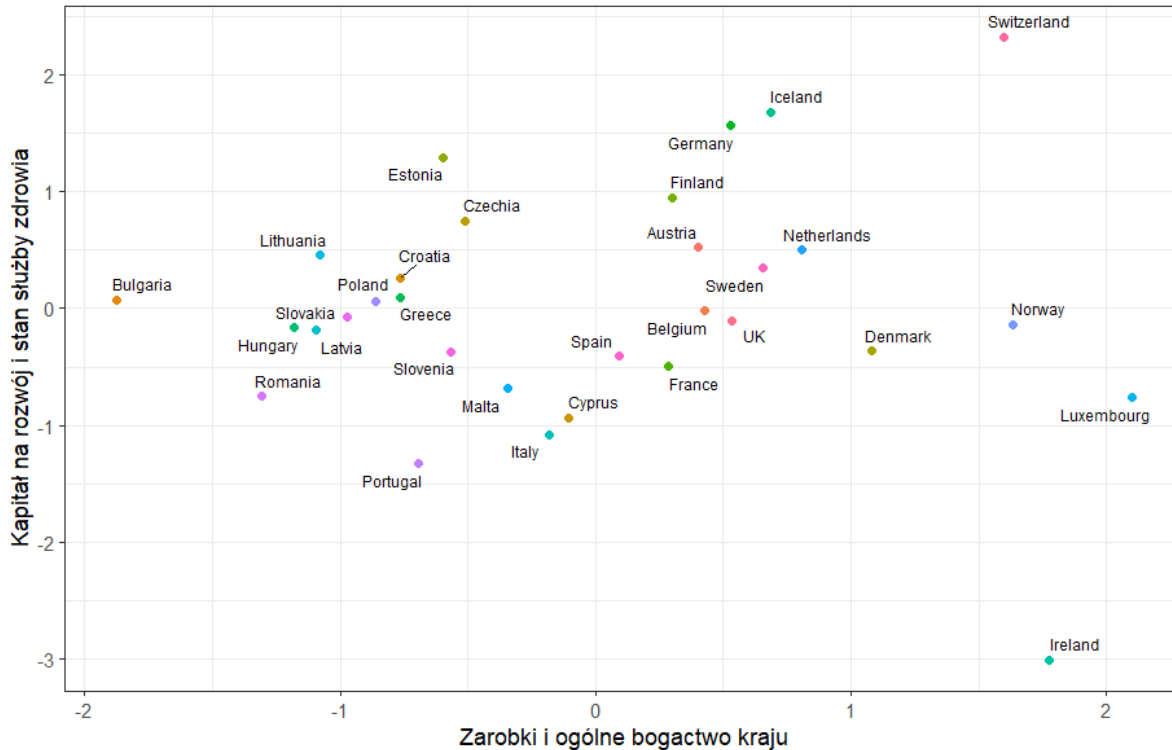
Wybieramy 2 składowe, ponieważ łącznie wyjaśniają ponad 80% zmienności – można to łatwo odczytać z wykresu słupkowego – dwie pierwsze składowe wyjaśniają ponad 80% ( $66,7\% + 13,5\% = 80,2\%$ ).

Określamy nasze wymiary w następujący sposób:

PC1 = zarobki i ogólne bogactwo kraju

PC2 = kapitał na rozwój i stan służby zdrowia

Definicje naszych wymiarów są poglądowe, zostały stworzone aby ułatwić analizę.



Wykres 3. Wykres krajów według składowych PCA.

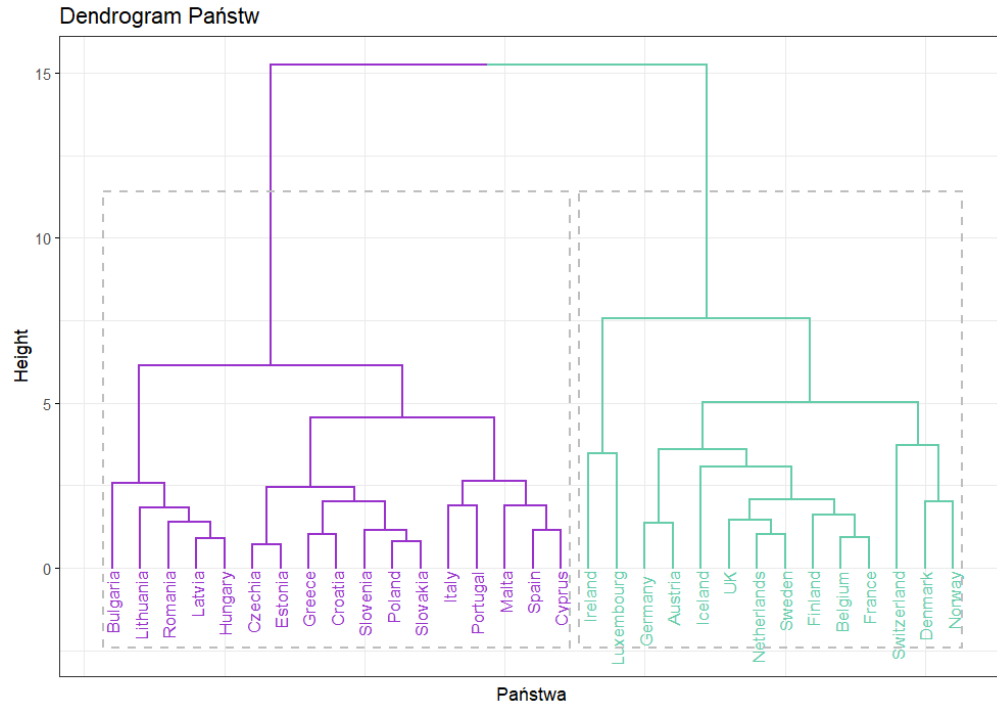
## Wnioski z PCA

Największy budżet na rozwój oraz badania posiada Szwajcaria i w tej kwestii znacząco się wyróżnia. Najniższe PKB w badanej grupie ma Bułgaria. Irlandia wyróżnia się prawdopodobnie tym, że produkuje dużą ilość energii i ma wydajnych pracowników. Francja również może się pochwalić swoją produkcją energii. Na Islandii budżet na rozwój również jest spory i zarobki są zadowalające, podobnie jest w Holandii. Kraje biedniejsze, mniej rozwinięte usytuowały się po lewej stronie wykresu. Kraje bogatsze, lepiej rozwinięte znajdują się po prawej. Zwolnienia ze szpitala wliczają w siebie również śmierci, dlatego odwrotnie korelują z PKB. Tam gdzie opieka zdrowotna jest na wyższym poziomie pacjenci kardiologiczni częściej przeżywają np. zawały, lub otrzymują stałą opiekę medyczną. Portugalia nie słynie z kolei z posiadania sporego budżetu na rozwój. Polska nie wyróżnia się specjalnie od innych krajów słowiańskich.

## Grupowanie

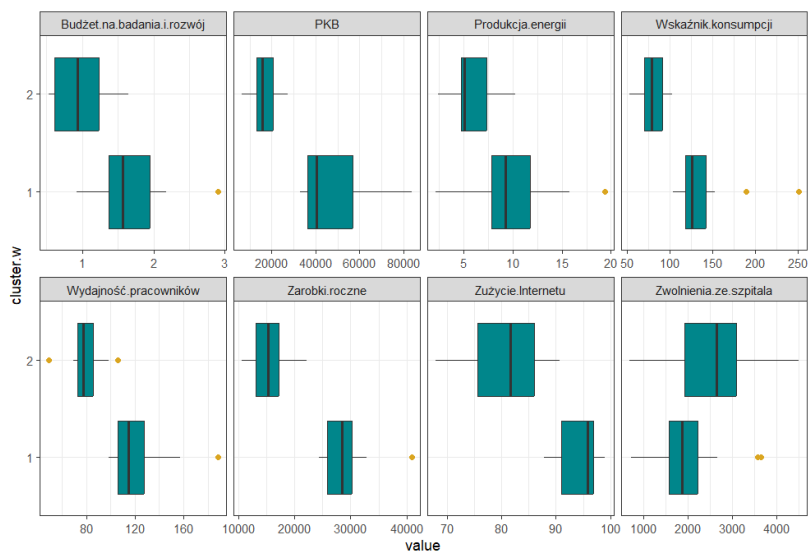
### Wyświetlenie dendrogramu - plot na obiekcie hclust.

Wybieramy grupowanie hierarchiczne po przeprowadzeniu badania funkcją clusterboot. Stabilność tej metody oscyluje między 92-94%. Badany zbiór danych jest dość mały, więc ta metoda wydaje się najbardziej odpowiednia.



Wykres 4. Dendrogram Państw.

Najbliżej siebie są Czechy i Estonia. Odległość na wykresie tych dwóch zmiennych od siebie jest najmniejsza. Odczytujemy to na podstawie dendrogramu i najniższej położonej kreski łączącej parę. Dendrogram wyraźnie wyznacza dwie grupy. Odczytujemy go metodą "bottom up", czyli od dołu. Dzielimy nasz zbiór na dwie grupy i odczytujemy, która grupa ma przewagę dla poszczególnych zmiennych.

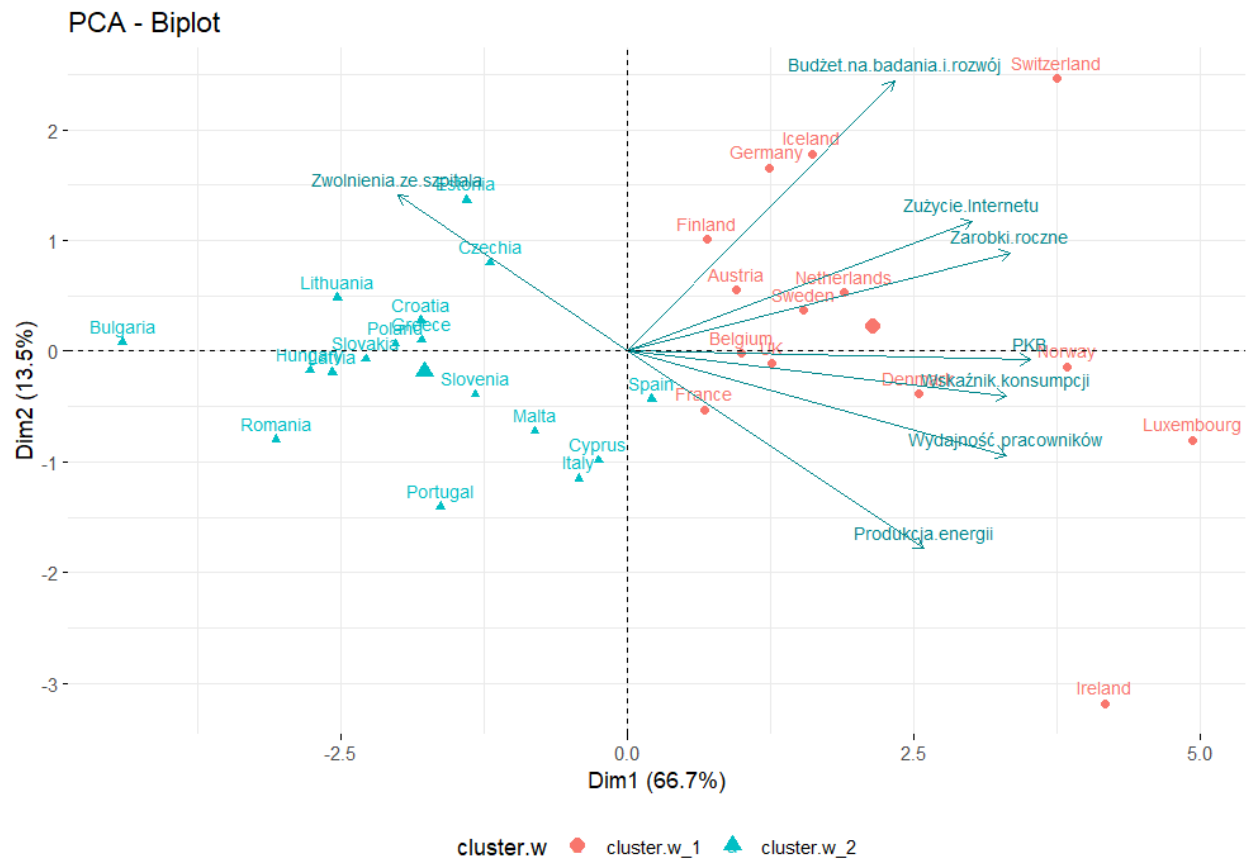


Wykres 5. Boxploty poszczególnych zmiennych.



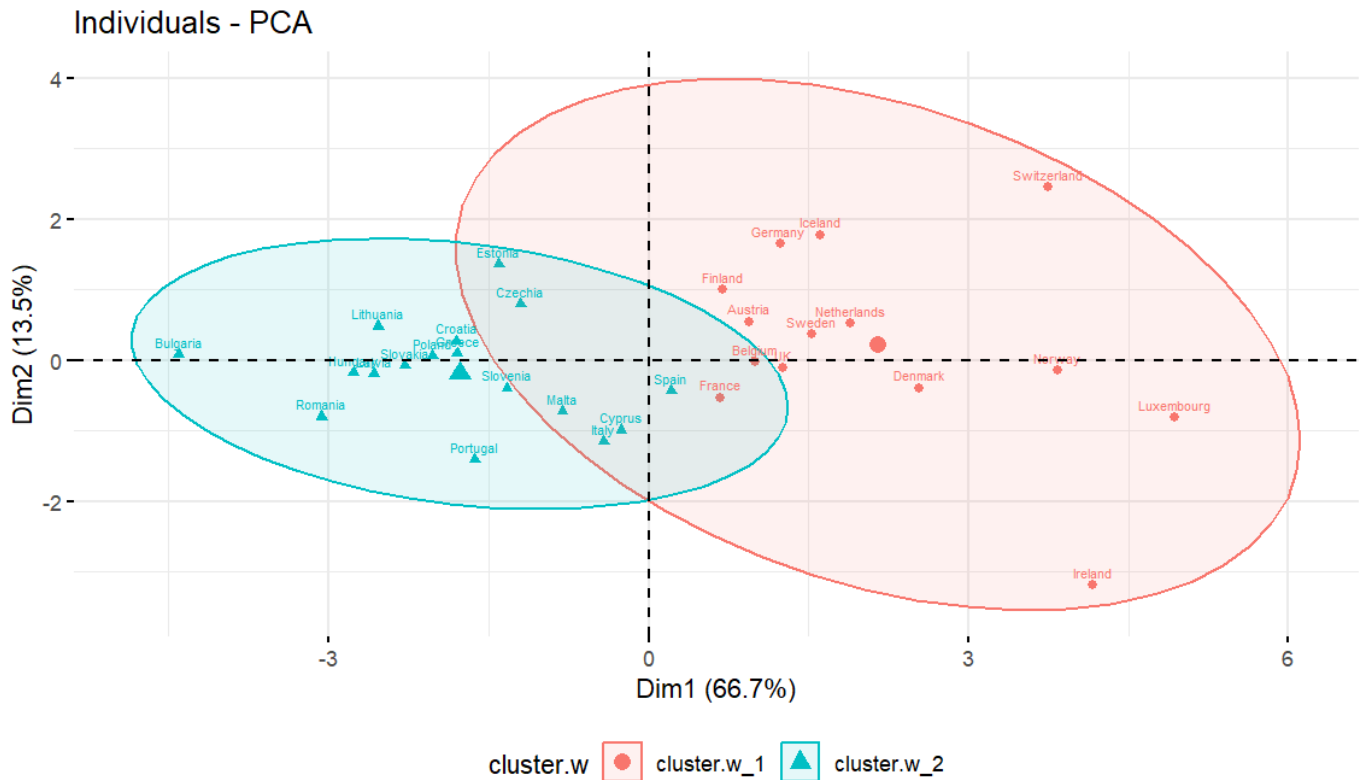
W 1 grupie znajdują się kraje, które cechuje duża liczba zwolnień ze szpitala pacjentów z oddziałów kardiologicznych (w tym śmierci), z niskim wskaźnikiem konsumpcji i mało wydajnymi pracownikami. W 2 grupie przeważają państwa wysoko rozwinięte, z dużą ilością użytkowników internetowych i dużym budżetem na rozwój.

## Grupa wraz ze zmiennymi



Wykres 6. Wykres krajów na Biplocie.

## Graficzne przedstawienie grupy



Wykres 7. Wykres według grupowania.

## Wnioski z grupowania

W 1 grupie przeważają państwa wysoko rozwinięte, z dużą ilością użytkowników internetowych, dużym budżetem na rozwój.

W 2 grupie znajdują się kraje, które cechuje duża liczba zwolnień ze szpitala pacjentów z oddziałów kardiologicznych (w tym śmierci), z niskim wskaźnikiem konsumpcji i mało wydajnymi pracownikami.

## Wnioski ogólne

Grupowanie i PCA pomogła nam określić, w jakiej sytuacji ekonomicznej były badane państwa oraz co je łączyło na rok 2019. Określiła jak można je podzielić, które z nich odstają od reszty, czym się wyróżniają i które są bardzo do siebie zbliżone dla podanych parametrów. Ewidentnym wnioskiem jest to, że kraje byłego bloku radzieckiego cechuje opóźnienie względem krajów zachodnich i Skandynawii. Dzięki PCA mogliśmy łatwo zmniejszyć nasze dane do dwóch wymiarów – dzięki temu można łatwo odczytać w czym nasze kraje są do siebie podobne. Początkowo nie sądziłyśmy, że niektóre zmienne przez nas wybrane będą miały między sobą tak silne korelacje, ale dzięki temu nasze PCA wyszło wzorcowe. Podsumowując całą naszą analizę uważamy, że wybrałyśmy bardzo dobre zmienne do naszej bazy danych.