

Wizualizacja spożycia produktów żywnościowych w Europie przy użyciu programu GradeStat

1. Wprowadzenie

Poniżej przedstawiamy tabelę zaczerpniętą z danych FAO dla roku 2001, zawierającą spożycie grup produktów żywnościowych w kilogramach na osobę dla 24 krajów europejskich. Tablica ta jest mało czytelna bez stosownej wizualizacji. Przedstawimy wyniki takiej wizualizacji dokonanej za pomocą programu GradeStat (dostępnego pod adresem <http://gradestat.ipipan.waw.pl>). Zastosowane będą metody, których opis znajduje się w książce „Analiza danych medycznych i demograficznych przy użyciu programu GradeStat”, wydanej w 2005 roku przez Instytut Podstaw Informatyki PAN przy współfinansowaniu otrzymanym z Instytutu-Pomnika „Centrum Zdrowia Dziecka”.

| | Strącz- kowe | Warzy- wa | Ryby | Oleje roślinne | Zbożo- we | Mięso | Owoce | Jaja | Cukier , miód | Ziem- niaki | Tłuszcze zwierzęce | Mleko | Używki |
|----|-----------------|--------------|------|-------------------|--------------|-------|--------|------|------------------|----------------|-----------------------|-------|--------|
| FI | 1,4 | 70,2 | 32,1 | 11,2 | 112,8 | 66,4 | 93,6 | 8,8 | 36,7 | 70,6 | 11,3 | 355,5 | 12,4 |
| CH | 1,5 | 74,3 | 30,9 | 17 | 103,4 | 70,6 | 101,5 | 12,6 | 47,5 | 52,6 | 16,7 | 357,7 | 11,6 |
| NT | 1,9 | 89,7 | 20,5 | 14,7 | 77 | 91,4 | 129,3 | 17,1 | 46,5 | 86,8 | 10 | 342,7 | 11,3 |
| IE | 2,7 | 80,8 | 15,9 | 15,3 | 129,1 | 88,3 | 99,7 | 7,8 | 44,1 | 119,8 | 13,9 | 288,5 | 5,7 |
| SL | 1,1 | 55,2 | 7,7 | 9 | 136,2 | 94,2 | 108,3 | 10,3 | 14,5 | 59,1 | 18,6 | 249,6 | 11,2 |
| AT | 0,9 | 97,9 | 14,5 | 17,3 | 117,9 | 110,1 | 125,6 | 12,7 | 50,1 | 69 | 17,8 | 290,8 | 9,9 |
| EE | 0,6 | 76,3 | 22,1 | 11,1 | 138,1 | 58,6 | 77,8 | 12,5 | 32,8 | 131,6 | 4,8 | 218,5 | 12,1 |
| DE | 1,6 | 92,7 | 14,5 | 22,8 | 101,7 | 83,1 | 119,6 | 12,5 | 46,6 | 77,8 | 21,8 | 240,8 | 9,4 |
| DK | 1,1 | 99,8 | 26,5 | 6,9 | 127,1 | 116,8 | 95,6 | 16,3 | 51,4 | 81,5 | 27,5 | 257,8 | 14,3 |
| GB | 6,7 | 89,5 | 21,6 | 18,2 | 102,5 | 77,8 | 91,4 | 10,5 | 38,9 | 110,9 | 6,9 | 227,2 | 7,2 |
| LV | 0,1 | 86,2 | 20 | 13,6 | 113,1 | 38,7 | 59 | 10,4 | 31,4 | 99,6 | 15,6 | 194,8 | 6,2 |
| CZ | 1,5 | 75,1 | 12,7 | 17,5 | 112,6 | 75,9 | 66,9 | 15,8 | 38,5 | 81,3 | 9,5 | 201,7 | 7,1 |
| FR | 2 | 129,9 | 31,1 | 17,7 | 117,1 | 102,4 | 97,3 | 15,6 | 40,9 | 120 | 19 | 272,3 | 9,6 |
| LT | 0,5 | 94,9 | 20,3 | 12,3 | 167,6 | 46,6 | 69,8 | 10,4 | 43 | 130,8 | 10,8 | 228,1 | 5,1 |
| BE | 2,4 | 140,4 | 21,6 | 24,4 | 107,7 | 89,7 | 84,4 | 12,8 | 52,2 | 113 | 24,9 | 228,1 | 3,6 |
| HU | 3 | 110,8 | 4,4 | 15,9 | 125,6 | 81,6 | 62,3 | 16,8 | 50,4 | 67,8 | 26,8 | 176,1 | 6,8 |
| PL | 2 | 119,2 | 13,1 | 14,1 | 155,5 | 70,5 | 54,9 | 11 | 43,1 | 132,2 | 14 | 176,8 | 5 |
| SK | 3,5 | 80,5 | 7,4 | 17,4 | 128,4 | 55,4 | 53,9 | 13,1 | 40,3 | 67,9 | 13,5 | 119,8 | 3,8 |
| MT | 3,2 | 151,5 | 37,6 | 5,5 | 185,7 | 68,2 | 63,9 | 17,6 | 56,5 | 86,1 | 11,6 | 197,7 | 7,1 |
| IT | 5,5 | 177,9 | 24,7 | 27,5 | 162,1 | 91,2 | 139,9 | 12 | 31,6 | 39,3 | 16,1 | 247,1 | 7,2 |
| PT | 3,9 | 187,5 | 76,1 | 16,8 | 132,2 | 88,6 | 132 | 10,7 | 34,1 | 123,7 | 12,9 | 226,7 | 6,3 |
| CY | 4,36 | 163,76 | 25,0 | 9 | 15,88 | 111,1 | 103,97 | 11,6 | 4 | 42,46 | 42,9 | 179,2 | 6,48 |
| ES | 5,7 | 154,2 | 44,7 | 28,1 | 101,3 | 118,1 | 122,6 | 12 | 32,6 | 82,4 | 4,7 | 164,7 | 6,8 |
| GR | 4,5 | 271,9 | 25,1 | 27,9 | 153,7 | 91,5 | 145,4 | 9,2 | 35,8 | 73,3 | 3,1 | 242,4 | 5,4 |

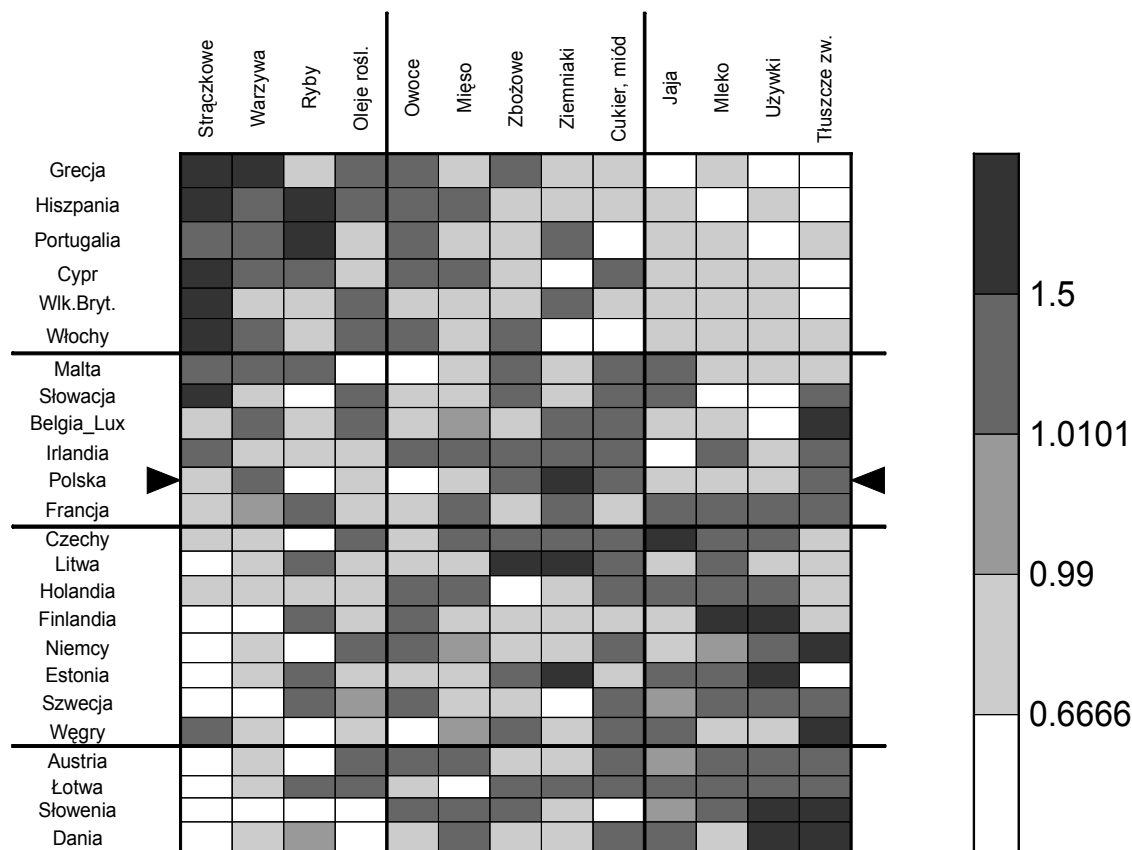
Objaśnienia mniej znanych skrótów:

CH Szwajcaria
 NT Holandia
 IE Irlandia
 AT Austria
 EE Estonia
 DK Dania
 LV Łotwa

LT Litwa
 SK Słowacja
 MT Malta
 IT Włochy
 PT Portugalia
 CY Cypr
 ES Hiszpania

2. Optymalnie uporządkowana macierz nadreprezentacji

Pierwszy etap wizualizacji tabeli danych polega na utworzeniu tak zwanej *mapy nadreprezentacji*, widocznej poniżej. Kolejność krajów i kolejność produktów jest na mapie inna niż w tabeli.



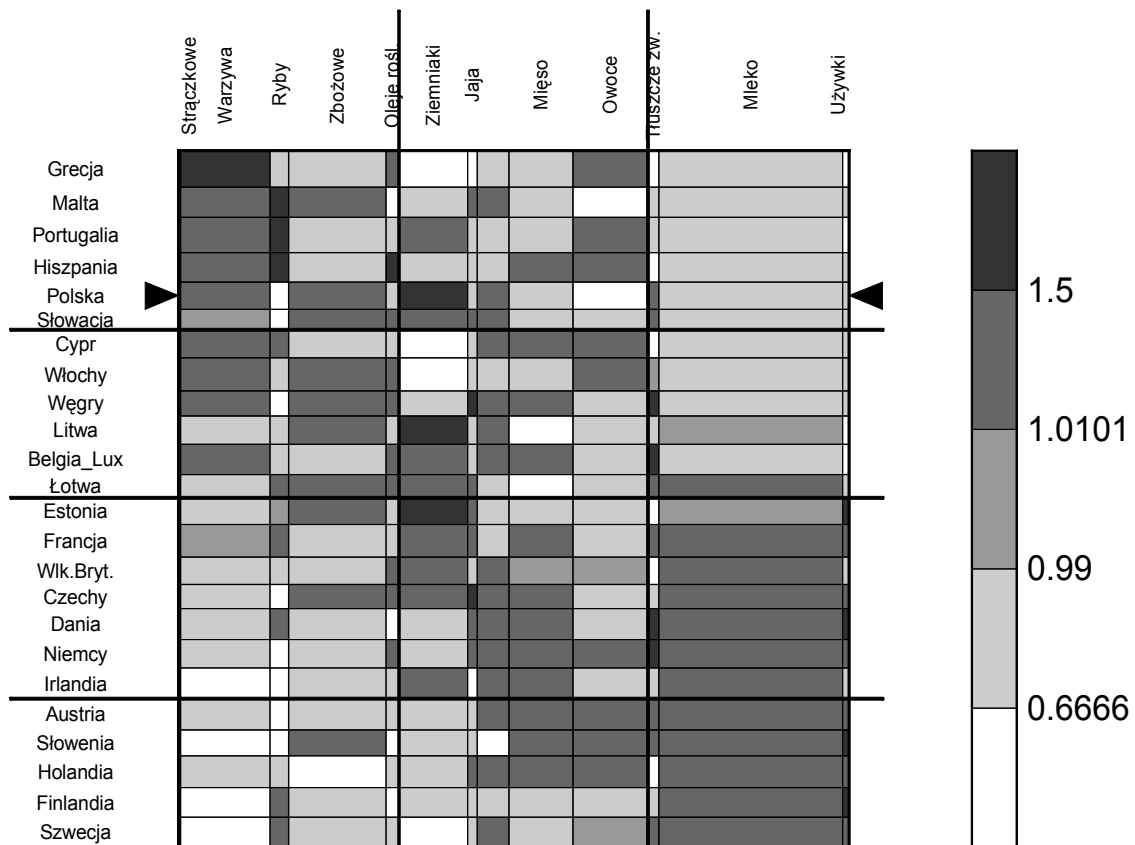
Uporządkowanie wierszy i kolumn są dobrane tak, aby tworzyły cztery spójne skupienia krajów, których mieszkańcy spożywają podobne produkty, oraz trzy skupienia produktów. Im ciemniejsza jest barwa prostokąta na przecięciu kraju i produktu, tym silniejsza jest nadreprezentacja spożycia tego produktu w danym kraju. Jak widać, ciemne prostokąty skupiają się głównie w rogach lewym dolnym i prawym górnym, a jasne w dwóch pozostałych rogach.

Małymi czarnymi trójkątami została oznaczona Polska.

Czarne, pogrubione linie poziome oddzielają grupy (zwane dalej skupieniami) krajów, których mieszkańcy spożywają podobne ilości poszczególnych produktów. Pogrubionymi pionowymi liniami oddzielone są skupienia produktów, które są podobnie preferowane we wszystkich rozpatrywanych krajach.

Liczba skupień dla krajów (4) i dla produktów (3) została ustalona przez analityka po rozpatrzeniu kilku możliwych podziałów.

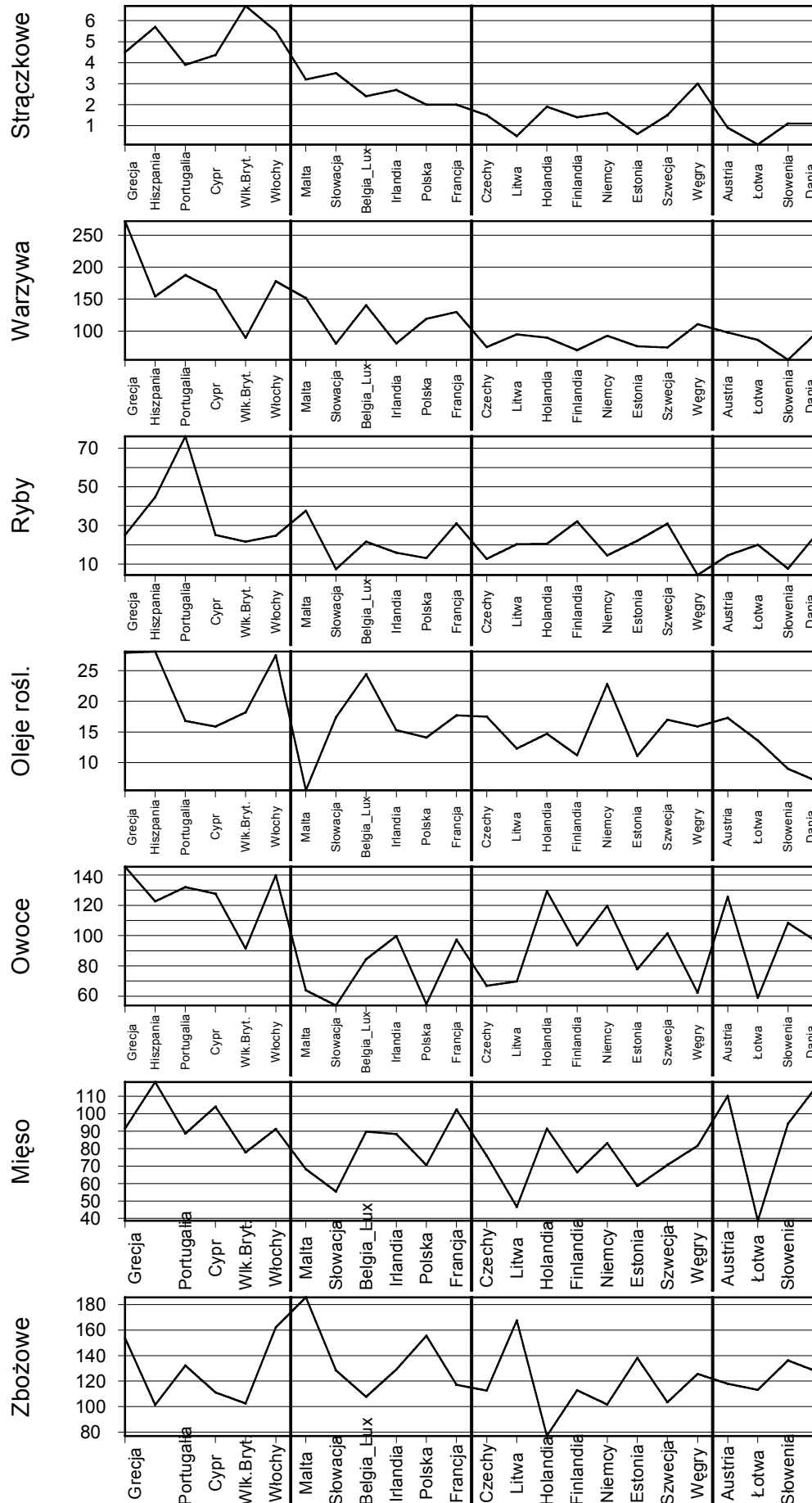
Przy tworzeniu mapy dokonano wstępnej normalizacji danych, polegającej na podzieleniu każdej liczby w tabeli przez sumę komórek z kolumny, w której ta liczba się znajduje. Z tego właśnie powodu szerokości wszystkich kolumn są takie same. Dla porównania przedstawiamy mapę nadreprezentacji otrzymaną wtedy, gdy każdą liczbę w tabeli dzieli się przez sumę komórek z całej tabeli.

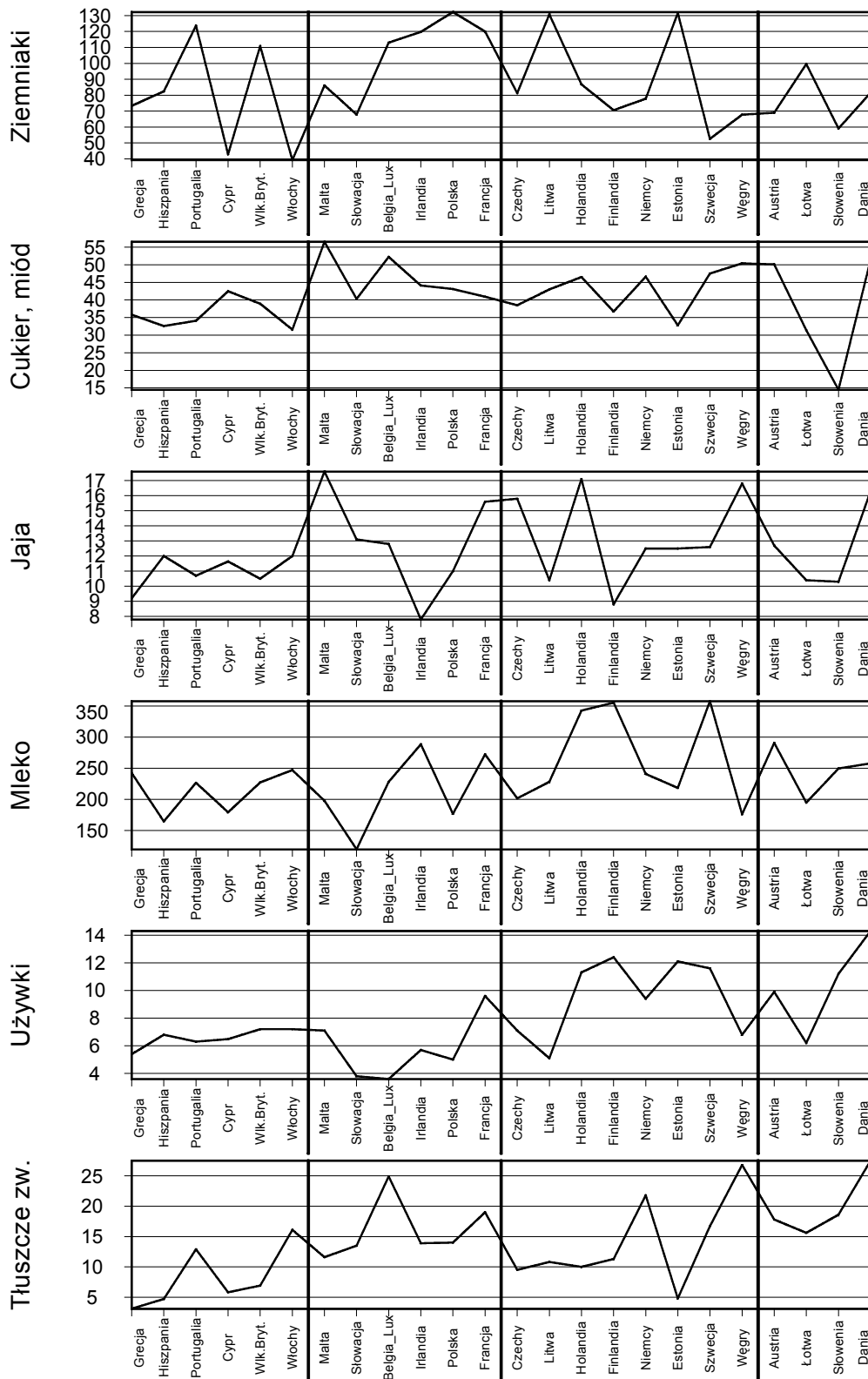


Na mapie widać ogromne różnice w wielkości spożycia poszczególnych produktów, co wydaje się zaciemniać faktyczne znaczenie danych grup produktów w diecie.

Dalej posługujemy się danymi znormalizowanymi w pierwszy sposób.

3. Tabela danych w postaci wykresów liniowych dla wszystkich krajów i produktów

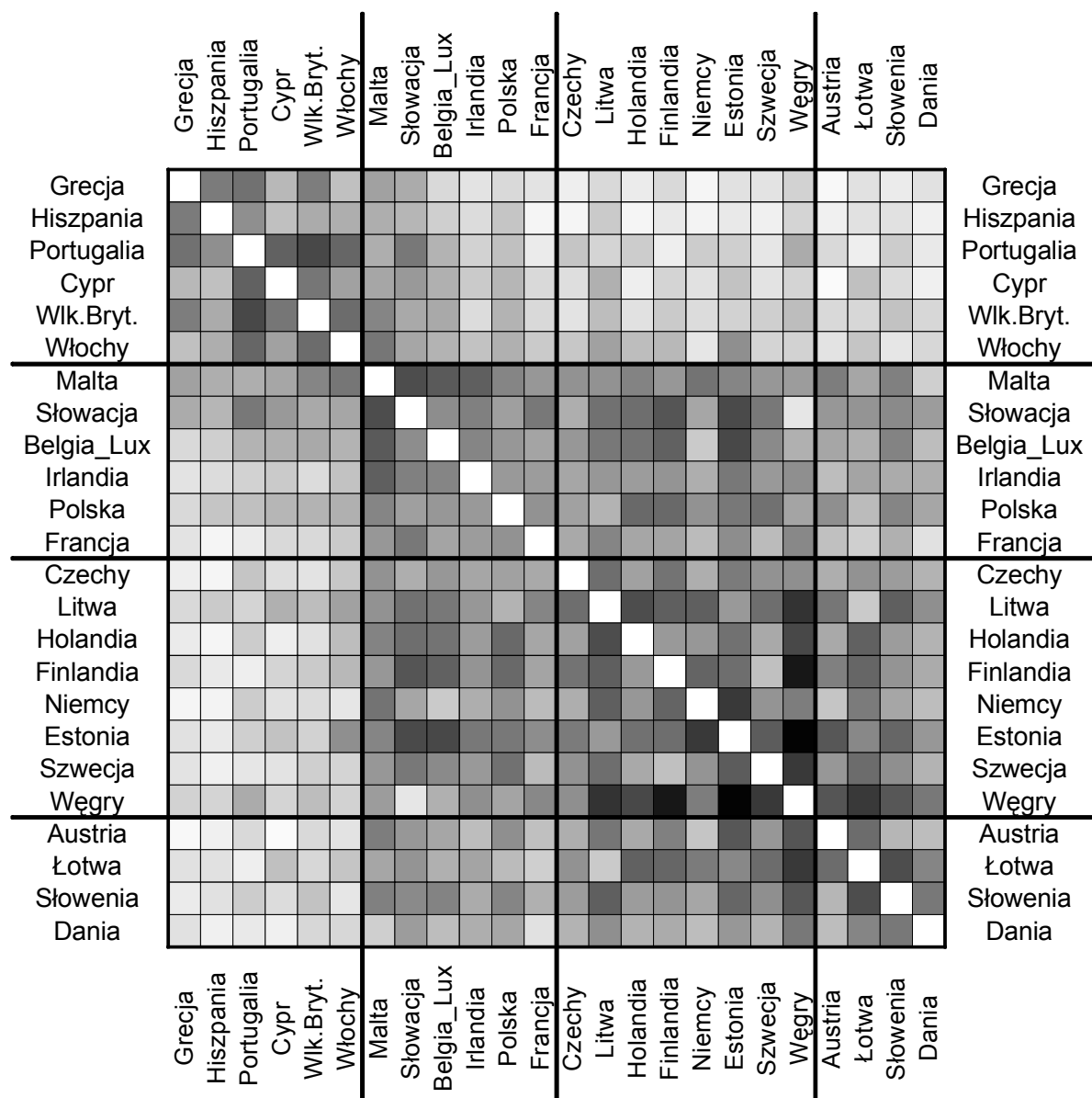




4. Wyznaczenie elementów odstających dla produktów i dla krajów

| | Strączkowe | Warzywa | Ryby | Oleje roślin. | Owoce | Mięso | Zbożowe | Ziemniaki | Cukier, miód | Jaja | Mleko | Używki | Tłuszcze zw. | |
|---------------|------------|---------|------|---------------|-------|-------|---------|-----------|--------------|------|-------|--------|--------------|---------------|
| Strączkowe | | | | | | | | | | | | | | Strączkowe |
| Warzywa | | | | | | | | | | | | | | Warzywa |
| Ryby | | | | | | | | | | | | | | Ryby |
| Oleje roślin. | | | | | | | | | | | | | | Oleje roślin. |
| Owoce | | | | | | | | | | | | | | Owoce |
| Mięso | | | | | | | | | | | | | | Mięso |
| Zbożowe | | | | | | | | | | | | | | Zbożowe |
| Ziemniaki | | | | | | | | | | | | | | Ziemniaki |
| Cukier, miód | | | | | | | | | | | | | | Cukier, miód |
| Jaja | | | | | | | | | | | | | | Jaja |
| Mleko | | | | | | | | | | | | | | Mleko |
| Używki | | | | | | | | | | | | | | Używki |
| Tłuszcze zw. | | | | | | | | | | | | | | Tłuszcze zw. |

Na mapie są wyniki szukania odstępstwa od regularności dla każdego produktu. Im ciemniejsza jest kolumna (wiersz) odpowiadająca danemu produktowi (krajowi), tym bardziej jego spożycie odstaje od tendencji dominującej w całym zbiorze produktów. Po rybach, ziemniakach i tłuszczach zwierzęcych (w których występują czarne kwadraty) najbardziej odstają produkty zbożowe – jednak nie zostały one usunięte z dalszej analizy.



Spośród krajów najbardziej wyróżniają się Węgry, Estonia, Litwa, Malta, a lokalnie Portugalia i Wielka Brytania. Natomiast Polska dobrze przystaje do pozostałych krajów (z wyj. Holandii i Finlandii).

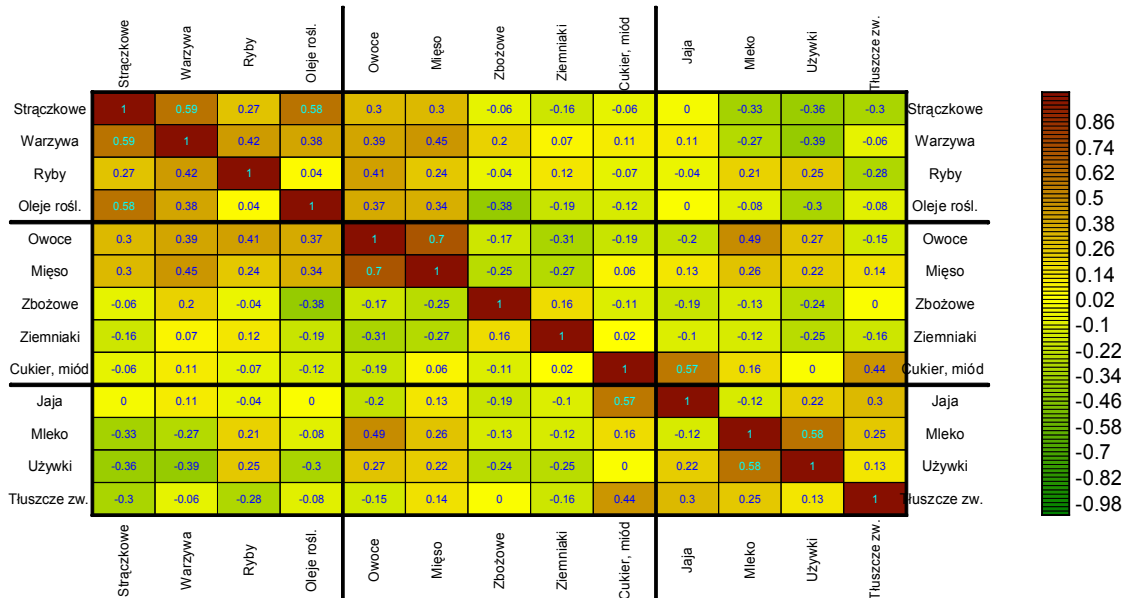
W dalszej analizie rozpatrywane są wszystkie kraje.

5. Agregacja w skupieniach dla krajów po usunięciu produktów odstających od regularności

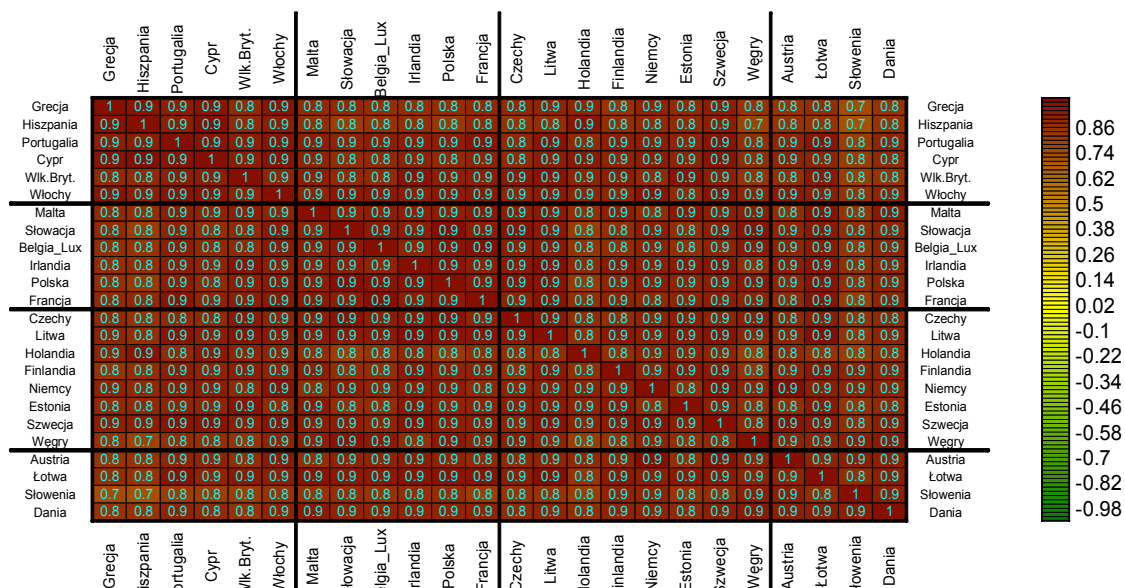
| | Strączkowe | Warzywa | Oleje roślinne | Owoce | Mięso | Zbożowe | Cukier, miód | Jaja | Mleko | Użytki |
|-------------|------------|---------|----------------|-------|-------|---------|--------------|------|-------|--------|
| skupienie 1 | 5.11 | 174 | 22.3 | 126 | 95.1 | 127 | 35.9 | 11 | 214 | 6.56 |
| skupienie 2 | 2.63 | 117 | 15.7 | 75.6 | 79 | 137 | 46.1 | 12.9 | 213 | 5.8 |
| skupienie 3 | 1.5 | 85.5 | 15.3 | 90.1 | 71.7 | 117 | 42.7 | 13.3 | 265 | 9.47 |
| skupienie 4 | 0.8 | 84.7 | 11.7 | 97.1 | 89.9 | 123 | 36.8 | 12.4 | 248 | 10.4 |

Skupienia 1 i 4 są wyraźnie przeciwstawne, z wyjątkiem owoców, których średnie spożycie w obu tych skupieniach jest wysokie. Środkowe skupienia są pośrednie, a z tabeli można odczytać średnie wartości spożycia dla poszczególnych produktów. Przypomnijmy, że w pierwszym skupieniu znalazły się kraje śródziemnomorskie i Wielka Brytania – ich dietę charakteryzuje największe spożycie roślin strączkowych, warzyw i olejów roślinnych. Skrajnym przeciwieństwem w tym zbiorze okazały się Austria, Łotwa, Słowenia i Dania, z predylekcją zwłaszcza do jaj, mleka i używek, ale również do owoców.

6. Macierz podobieństw dla produktów i dla krajów



Podobieństwami są tu współczynniki korelacji Spearmana. Zamieszczona obok skala pokazuje, że czerwonym kolorem zaznaczono pary produktów o wysokiej dodatniej wartości tego współczynnika, np. strączkowe i oleje roślinne lub jaja i cukier. Zielonym kolorem zaznaczono najsilniej ujemnie skorelowane pary produktów, np. zbożowe i oleje roślinne. Widać, że rozmieszczenie czerwieni i zieleni odpowiada rozmieszczeniu czerni i bieli na pokazanej na początku mapie nadreprezentacji dla całej tabeli.





Podobna analiza dla krajów pokazuje, że przy zachowaniu tej samej skali dla wsp. korelacji cała tabela jest prawie jednolicie czerwona – różnice w podobieństwie między krajami są dużo mniejsze. Większe zróżnicowanie można by uzyskać zmieniając skalę np. na 0.6-1.