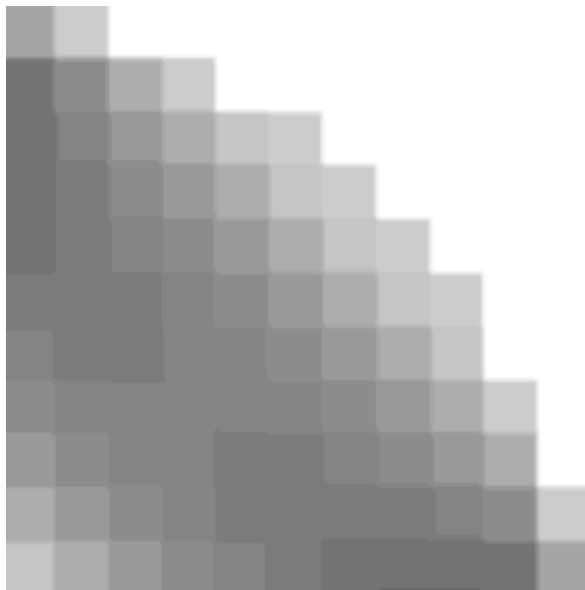


Gradacyjna analiza danych

Instytut Podstaw Informatyki PAN

Wiesław Szczesny

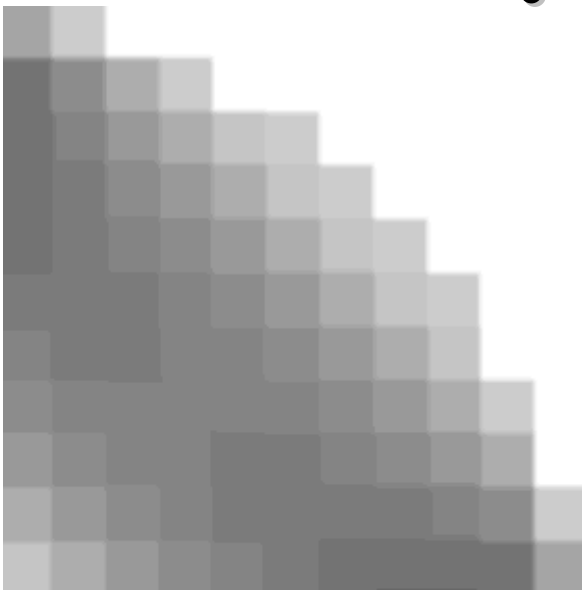
Emilia Jarochovska



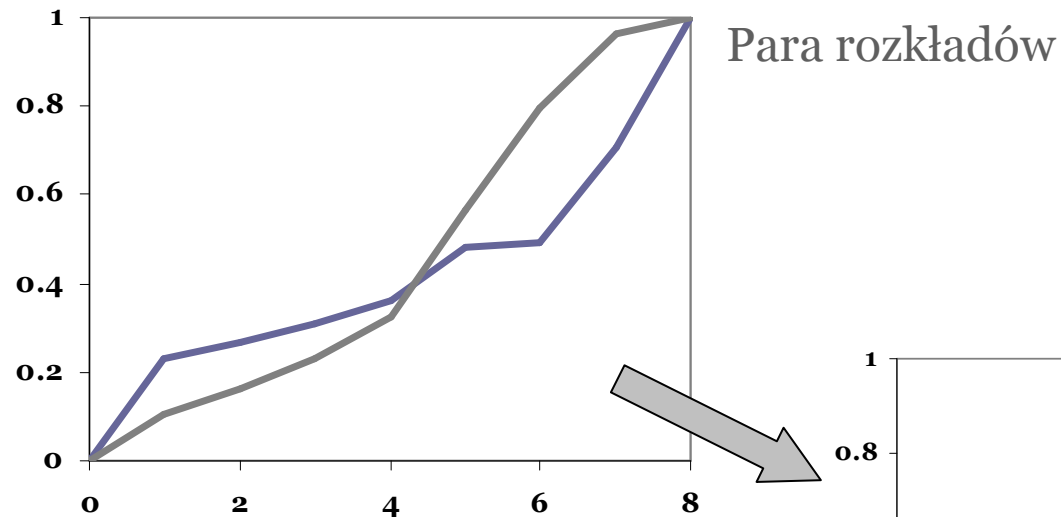
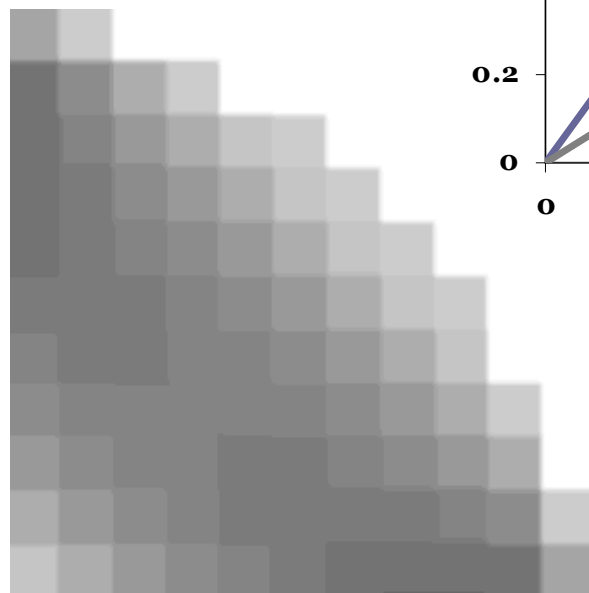
Gradacyjna analiza danych

Grade Correspondence Analysis

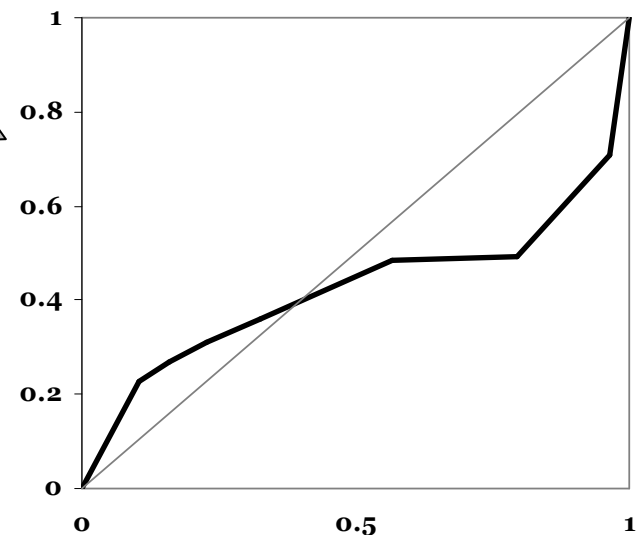
- Pomiar koncentracji, nadreprezentacja, GCA
- Przykład analizy
- Ku kompletnej infrastrukturze pojęć gradacyjnej analizy danych



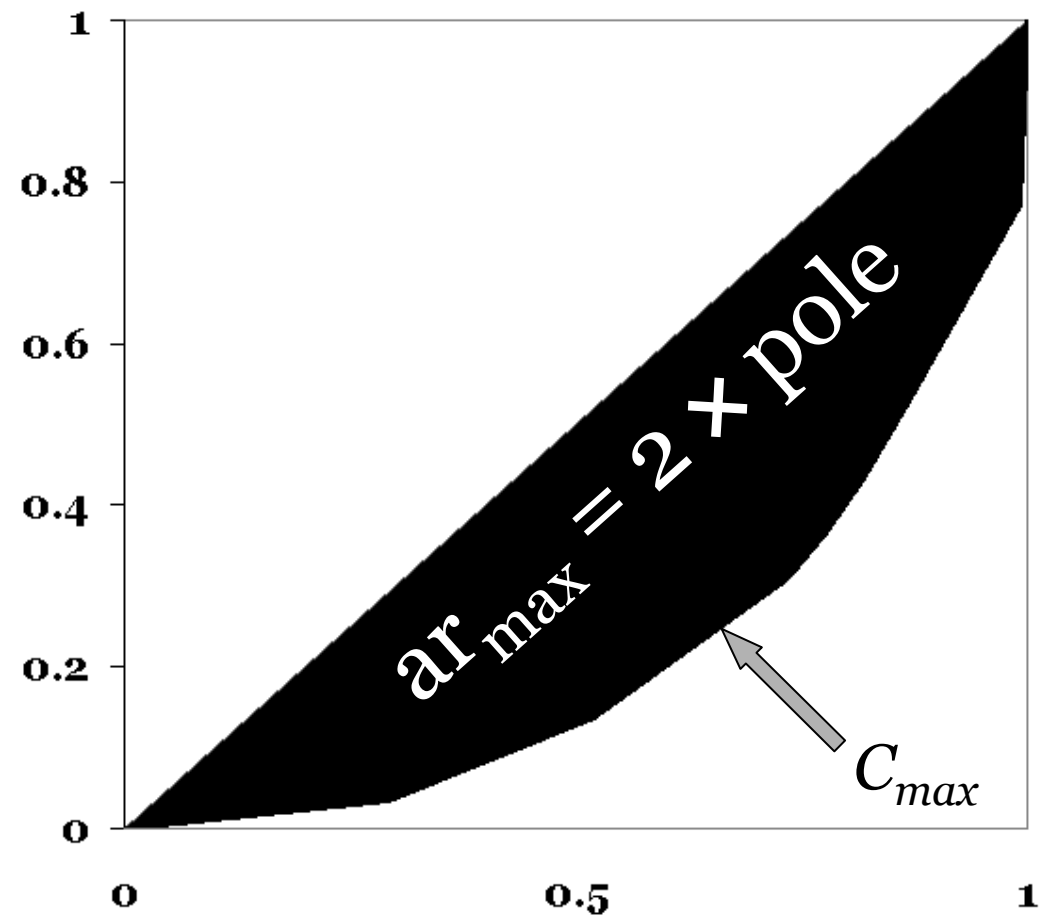
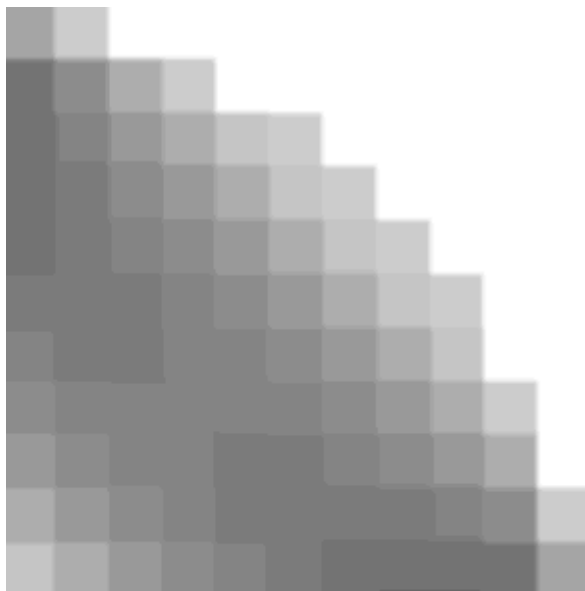
W gradacyjnej analizie danych dwóm jednowymiarowym rozkładom przyporządkowuje się krzywą w kwadracie jednostkowym, która reprezentuje koncentrację jednego rozkładu względem drugiego.



Krzywa koncentracji



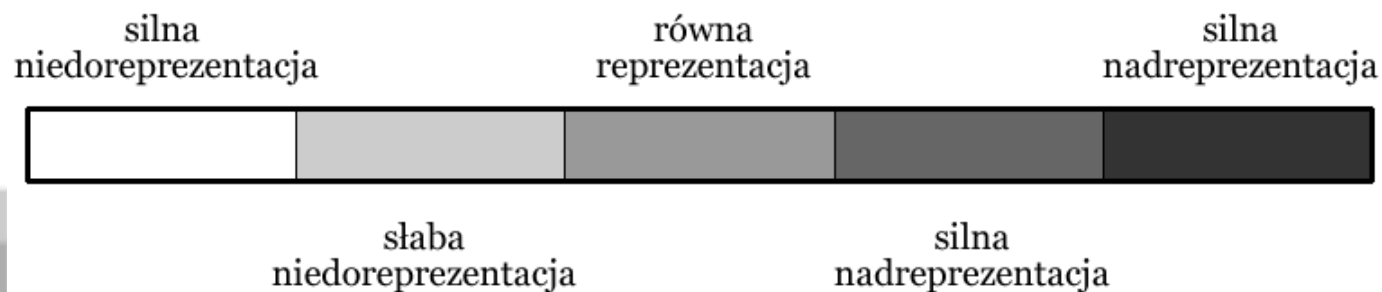
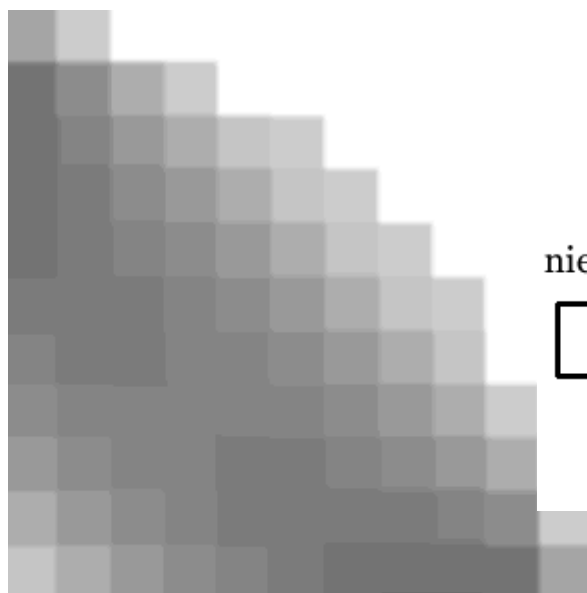
Pomiar maksymalnej koncentracji: krzywa C_{max} i wskaźnik ar_{max}

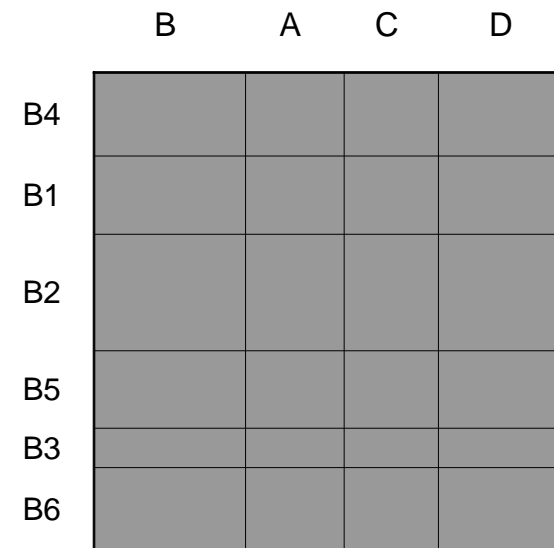
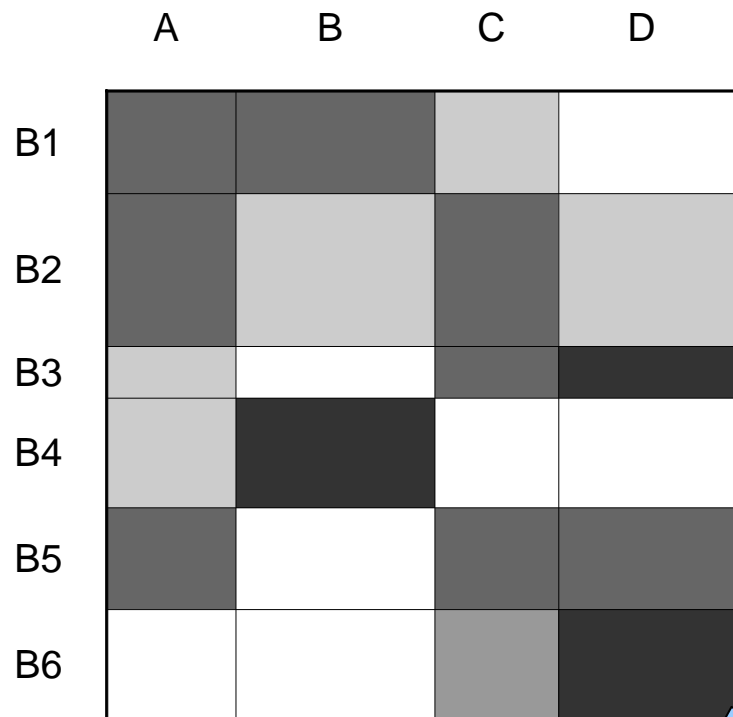


Nadreprezentacja

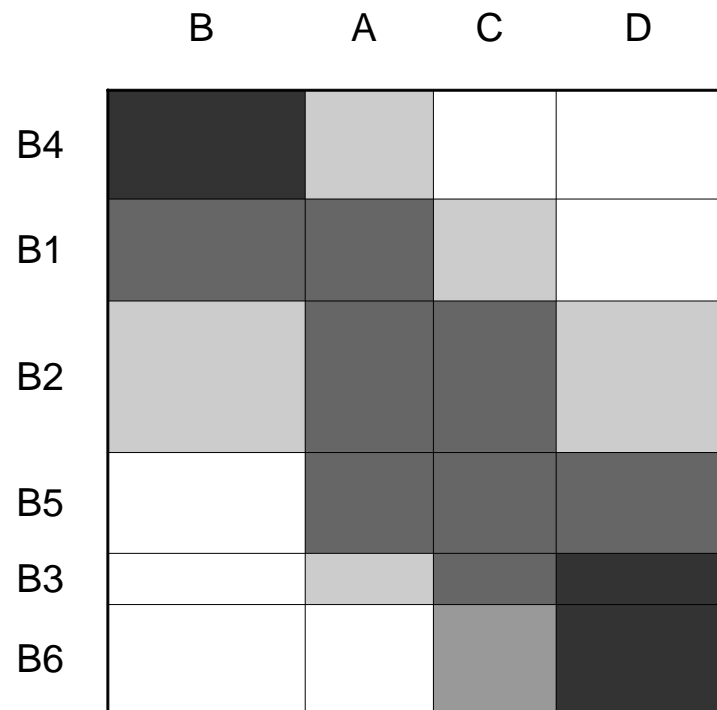
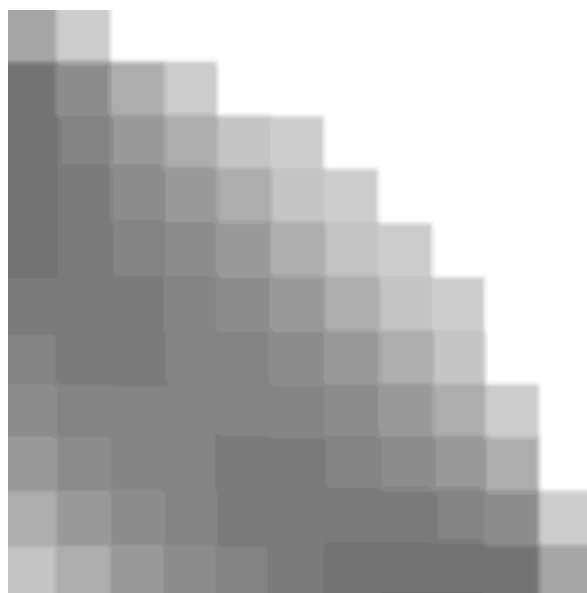
p_i	0.06	0.07	0.081	0.094	0.125	0.127	0.184	0.26
q_i	0.022	0.036	0.051	0.071	0.127	0.212	0.217	0.263
h_i	0.37	0.51	0.63	0.76	1.02	1.67	1.18	1.01

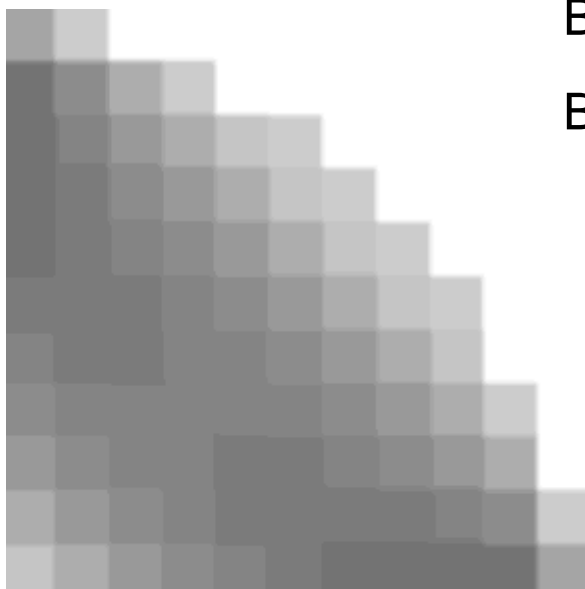
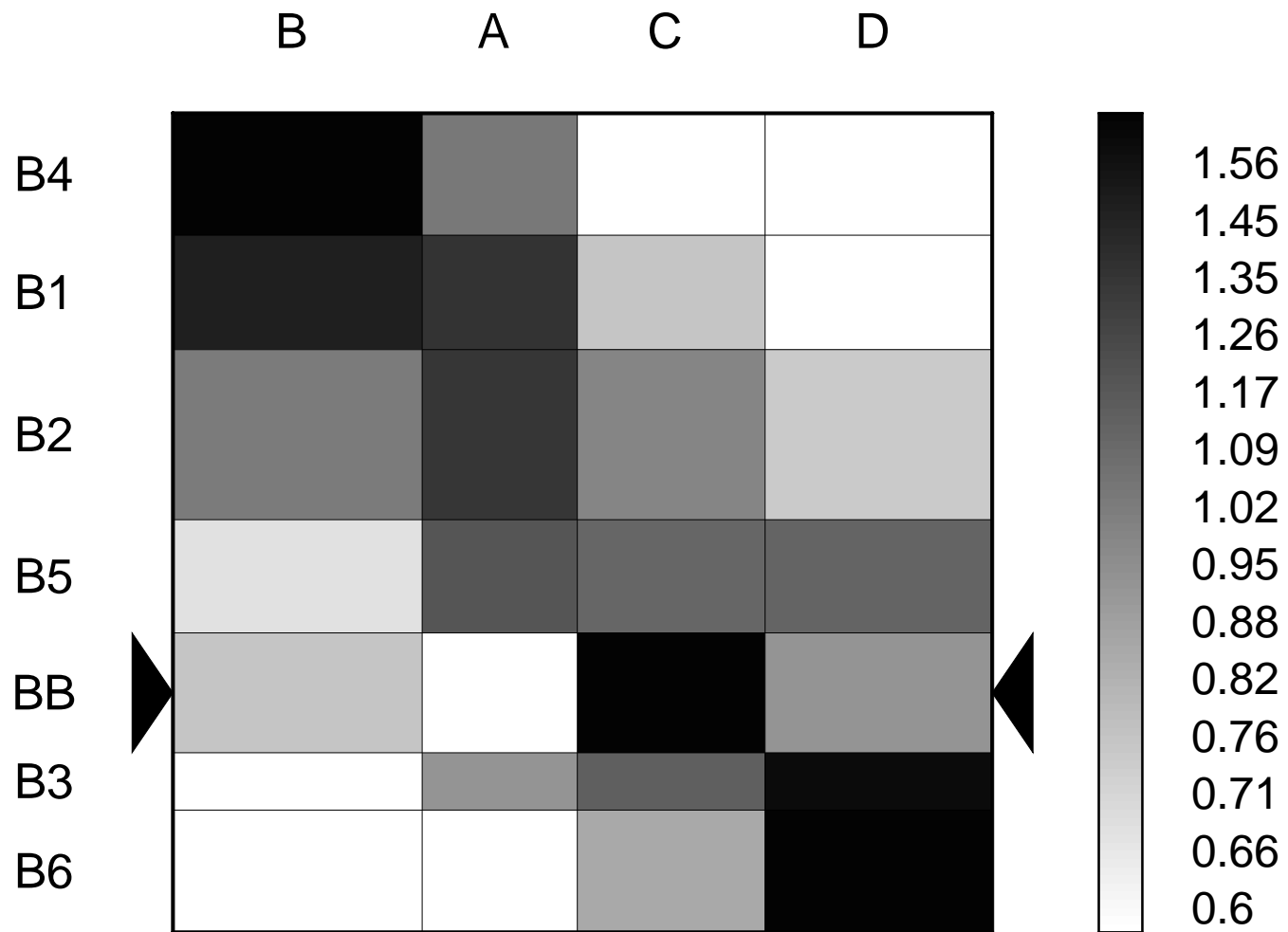
$$\text{Wskaźnik nadreprezentacji} = \frac{\text{„wielkość obserwowana”}}{\text{„wielkość wynikająca z modelu”}}$$





GCA





GradeStat - sss

Plik Edycja Wybór danych Normalizacja Przegląd danych Wykresy Obróbka wstępna Szeregowanie Analiza skupień Inne Okno Pomoc

GradeStat interface showing various data analysis windows and a central data table.

sss:1

Rho* =	0.19515	n	maks	Wyłącz.	0	0	0	0	0
Tau =	0.13168	n	maks	Grupa	1	4	30	13	8
				Waga	1	1	1	1	1
				Skupienie	1	1	1	2	2
Wyłącz.	Populacja	Skupienie	25x14	n./os 2007	/sr. w EU	st e-adm %	z. ciepł.04	ietn. 2004	atki
0	2	1	1	227	20	88.5	1.7		
0	2	1	1	113	67	121.5	1.8		
0	2	1	1	3.743	118	74	97.6	1.75	
0	2	2	2	123	72	116.6	1.42		
				125	32	100.8	1.73		
				122	58	106.3	1.78		
				138	50	125.2	1.99		
				110	50	98.1	1.9		
				119	35	100.6	1.64		
				103	53	111.6	1.33		
				117	59	86.7	1.74		
				84	25	152.8	1.49		
				109	47	81.5	1.37		
				79	45	98.1	1.22		
				69	40	129.1	1.37		
				82	32	123.2	1.29		
				72	40	136.7	1.42		
				98	55	140.6	1.32		
				71	30	75.7	1.23		
				60	15	68.1	1.28		

sss:2

Wystartuj GCA

Wykonaj iteracji: (puste=do przerwy)

Start

Rezultatów: 2000
Po filtrowaniu: 759 Wyczyść listę

Alg. Liczba Rho* Tau Populacja

RSS 1	0.195150	0.131680	Całość
RSS 2	0.195143	0.131675	F Całość
RSS 1	0.195140	0.131673	F Całość
RSS 1	0.195124	0.131663	F Całość
RSS 1	0.195124	0.131662	F Całość
RSS 1	0.195120	0.131660	F Całość
RSS 2	0.195120	0.131660	F Całość
RSS 1	0.195119	0.131659	F Całość
RSS 1	0.195117	0.131657	F Całość
RSS 1	0.195114	0.131655	F Całość
RSS 1	0.195113	0.131655	F Całość
RSS 1	0.195111	0.131653	F Całość
RSS 2	0.195111	0.131653	F Całość

sss:3

GradeStat

NUM

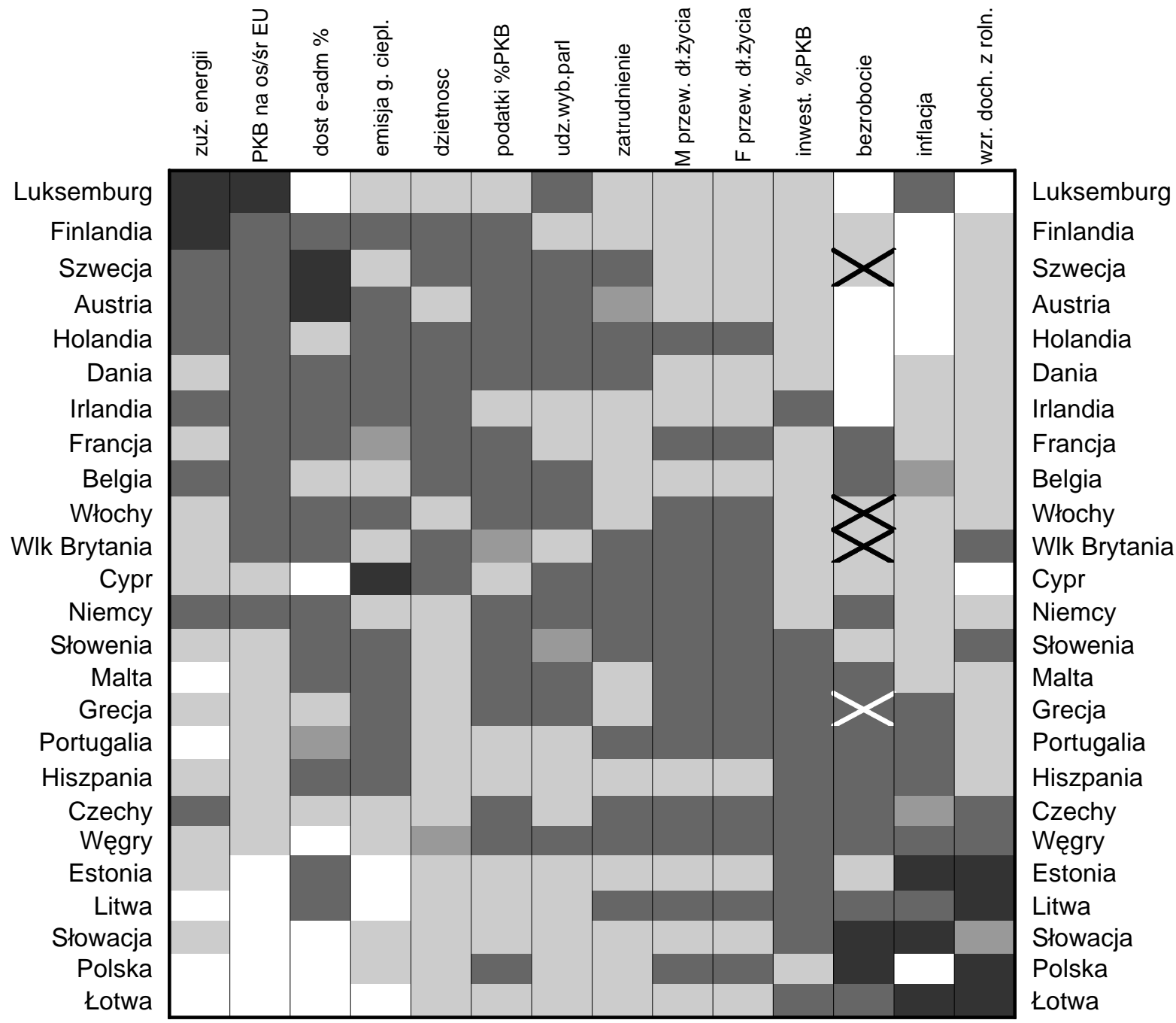
Przykład analizy Wskaźniki ekonomiczno-gospodarcze

	zuż. energii	PKB na os/śr EU	dost e- adm [%]	emisja g. ciepl.	dziatność	podatki %PKB	udz. wyb. parl	zatrudnienie	M przew. dl. życia	F przew. dl.życia	inwest. %PKB	bezrobocie	inflacja	wzr. doch. z roln.
Luksemburg	9.00	227.00	20.00	88.50	1.70	41.10	92.00	61.60	76.00	81.20	19.10	5.60	3.90	90.20
Finlandia	5.00	113.00	67.00	121.50	1.80	44.50	69.70	67.60	75.30	82.20	19.20	8.30	1.10	94.30
Szwecja	3.74	118.00	74.00	97.60	1.75	51.20	80.10	72.10	78.30	82.60	17.00		1.10	103.10
Austria	3.09	123.00	72.00	116.60	1.42	44.30	84.30	67.80	76.40	82.10	20.70	5.20	1.50	103.60
Holandia	3.16	125.00	32.00	100.80	1.73	38.80	80.00	73.10	76.90	81.40	19.50	4.60	1.40	86.20
Dania	2.76	122.00	58.00	106.30	1.78	49.90	87.10	75.70	75.40	80.10	20.50	4.40	2.10	94.70
Irlandia	3.00	138.00	50.00	125.20	1.99	31.70	62.60	66.30	76.40	81.20	25.50	4.30	2.50	107.70
Francja	2.59	110.00	50.00	98.10	1.90	45.30	60.30	63.10	76.70	83.80	19.50	9.20	2.00	86.60
Belgia	3.63	119.00	35.00	100.60	1.64	47.40	91.60	60.30	75.90	81.70	19.40	8.50	2.80	87.80
Włochy	2.22	103.00	53.00	111.60	1.33	42.10	81.40	57.60	76.80	82.50	20.60		2.20	86.30
Wlk Brytania	2.00	117.00	59.00	86.70	1.74	37.70	59.40	71.60	76.20	80.70	16.70		1.90	122.50
Cypr	2.32	84.00	25.00	152.80	1.49	33.70	91.80	68.90	76.60	81.70	19.20	4.90	2.30	75.20
Niemcy	2.80	109.00	47.00	81.50	1.37	40.00	79.10	65.00	76.50	82.10	17.10	9.50	2.10	121.70
Słowenia	2.36	79.00	45.00	98.10	1.22	39.90	70.40	65.30	73.50	80.70	25.00	6.40	2.30	129.10
Malta	1.14	69.00	40.00	129.10	1.37	36.70	95.70	54.00	76.70	80.70	21.30	7.50	2.30	88.40
Grecja	2.00	82.00	32.00	123.20	1.29	37.70	75.00	59.40	76.60	81.40	23.70		3.10	82.10
Portugalia	1.73	72.00	40.00	136.70	1.42	35.60	62.80	67.80	74.90	81.40	21.60	7.80	2.90	106.50
Hiszpania	2.00	98.00	55.00	140.60	1.32	35.40	68.70	64.10	75.20	80.20	22.10	8.50	1.10	100.90
Czechy	2.48	71.00	30.00	75.70	1.23	36.60	57.90	64.10	75.20	80.20	22.10	8.50	1.10	100.90
Węgry	1.74	60.00	15.00	68.10	1.28	39.20	73.50	64.10	75.20	80.20	22.10	8.50	1.10	100.90
Estonia	2.00	51.00	63.00	49.20	1.40	32.70	58.20	64.10	75.20	80.20	22.10	8.50	1.10	100.90
Litwa	1.21	48.00	40.00	33.80	1.26	28.70	58.20	64.10	75.20	80.20	22.10	8.50	1.10	100.90
Słowacja	1.98	52.00	15.00	71.80	1.25	30.60	70.10	64.10	75.20	80.20	22.10	8.50	1.10	100.90
Polska	1.46	49.00	10.00	67.90	1.23	34.90	46.20	51.70	70.60	79.20	18.30	17.20	0.90	205.70
Łotwa	1.62	43.00	5.00	41.50	1.24	29.10	71.20	62.30	65.90	76.20	28.90	8.10	7.00	243.10

Dane z Grzegorek, 2006
na podstawie
<http://epp.eurostat.cec.eu.int>

Przykład analizy

GCA – gradacyjna analiza odpowiedniości + imputacja



Przykład analizy

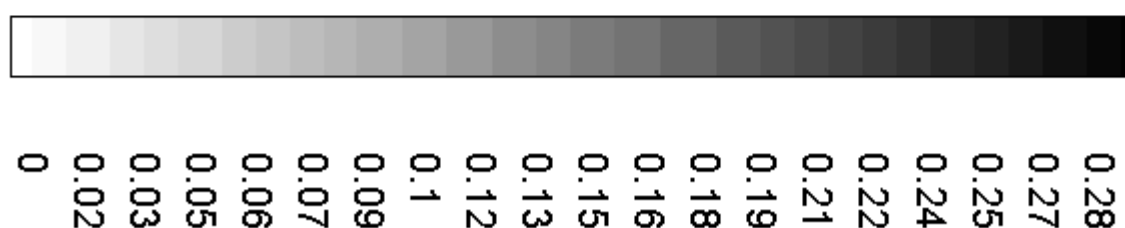
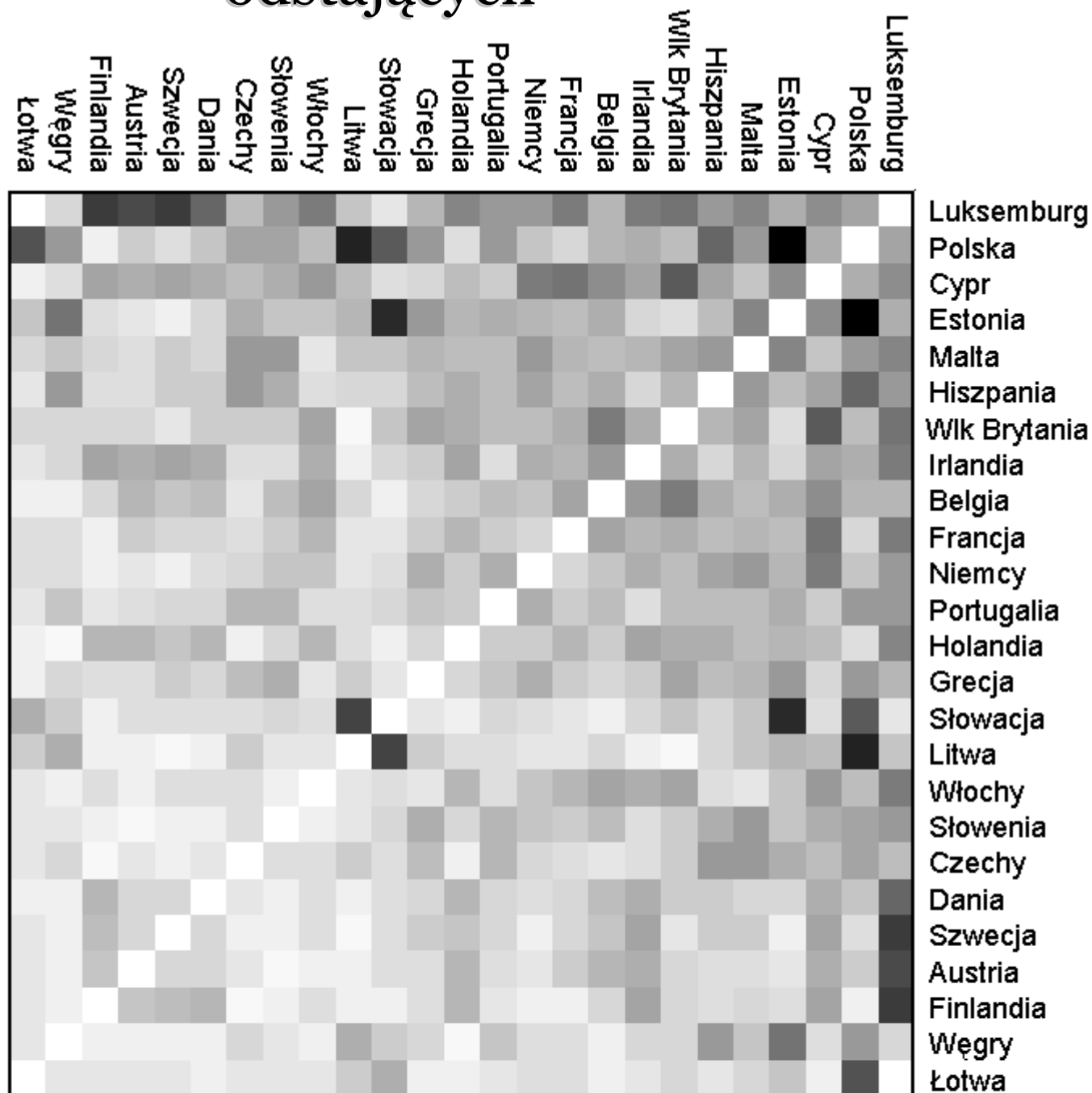
GCCA – gradacyjna analiza

skupień

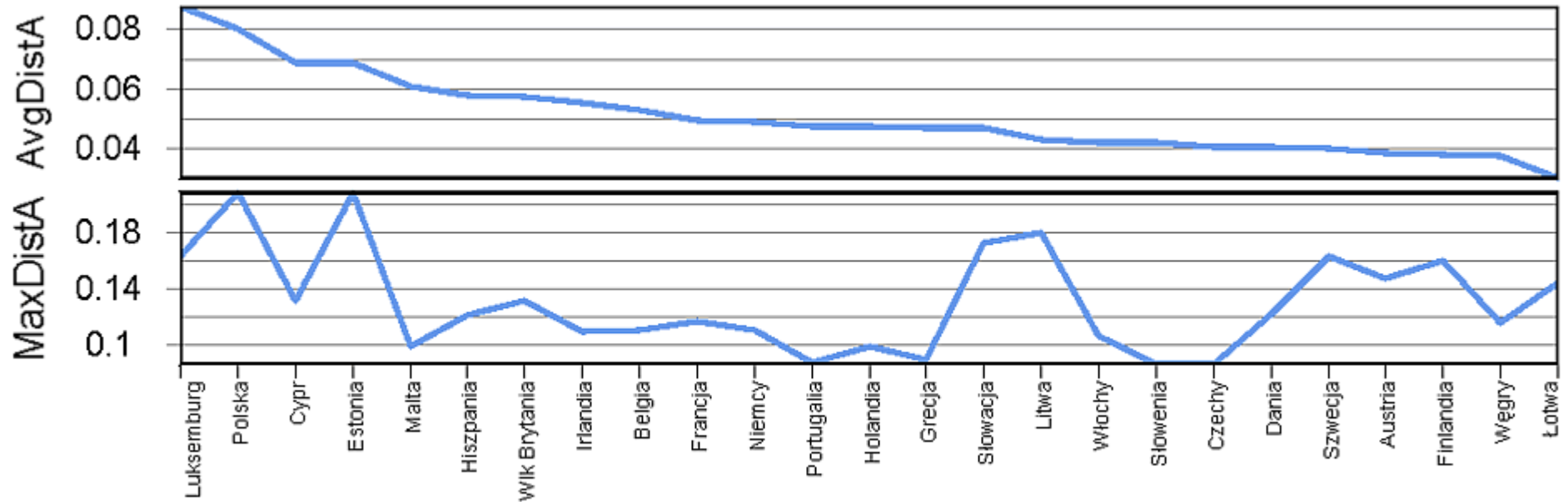
	zuż. energii	PKB na os/śr EU	dost e-adm %	emisja g. ciepl.	dzielnosc	podatki %PKB	udz.wyb.parl	zatrudnienie	M przew. dł.życia	F przew. dł.życia	inwest. %PKB	bezrobocie	inflacja	wzr. doch. z roln.	
Luksemburg	Dark	Dark	White	Light	Light	Light	Light	Light	Light	Light	Light	White	Light	Light	Luksemburg
Finlandia	Dark	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Light	Light	Light	Finlandia
Szwecja	Dark	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Light	Szwecja
Austria	Dark	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	White	Light	Light	Austria
Holandia	Dark	Dark	Light	Light	Light	Dark	Light	Light	Light	Light	Light	White	Light	Light	Holandia
Dania	Light	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Light	Light	Light	Dania
Irlandia	Dark	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Dark	White	Light	Light	Irlandia
Francja	Light	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Light	Francja
Belgia	Dark	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Light	Belgia
Włochy	Light	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Light	Włochy
Wlk Brytania	Light	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Dark	Wlk Brytania
Cypr	Light	Light	White	Dark	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	White	Cypr
Niemcy	Dark	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Light	Niemcy
Słowenia	Light	Light	Dark	Light	Light	Dark	Light	Light	Light	Light	Dark	Light	Light	Light	Słowenia
Malta	White	Light	Dark	Light	Light	Dark	Light	Light	Light	Light	Dark	Light	Light	Light	Malta
Grecja	Light	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Light	Grecja
Portugalia	White	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Light	Portugalia
Hiszpania	Light	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Light	Hiszpania
Czechy	Dark	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	Light	Light	Czechy
Węgry	Light	Light	Dark	Light	Light	Dark	Light	Light	Light	Light	Dark	Light	Light	Light	Węgry
Estonia	Light	White	Dark	Light	Light	Dark	Light	Light	Light	Light	Dark	Light	Light	Dark	Estonia
Litwa	White	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Dark	Light	Light	Dark	Litwa
Słowacja	Light	Dark	Dark	Light	Light	Dark	Light	Light	Light	Light	Dark	Light	Light	Dark	Słowacja
Polska	White	White	White	Light	Light	Dark	Light	Light	Light	Light	Light	Dark	White	Dark	Polska
Łotwa	White	White	White	Light	Light	Light	Light	Light	Light	Light	Dark	Light	Light	Dark	Łotwa

Przykład analizy

Znajdowanie elementów odstających



Przykład analizy



Znajdowanie elementów odstających

Strona główna

Książki

Zamawianie

Biblioteki

Twórcy

Publikacje

Recenzje i opinie

Zastosowania

Pobierz GradeStat

Licencja programu

Zrzuty ekranu

Instytut Podstaw

Informatyki PAN

Wybrane przykłady zastosowania GradeStata

Pełna lista jest w przygotowaniu i wkrótce pojawi się na tej stronie.

- [Wybory parlamentarne 2005](#)
- Wizualizacja spożycia produktów żywnościowych w Europie w 2001 roku według FAO [[pdf](#)]
- Dane antropometryczne z Centrum Zdrowia Dziecka - ulotka [[pdf](#)]

Wybrane przykłady zastosowania wskaźnika Giniego, krzywych Lorenza i pokrewnych narzędzi statystycznych

- [International Conference in Memory of Two Eminent Social Scientists: C. Gini and M.O. Lorenz, 23-26 marca 2005](#) w programie m.in. Magdalena Niewiadomska-Bugaj, Teresa Kowalczyk, Hend Auda: Statistical Tests Based on Gini Index and Gini Mean Difference [[abstract](#)]
- [Laboratorium Wizualizacji Informacji Chun-houh Chena](#), Instytut Nauk Statystycznych, Academia Sinica

<http://gradestat.ipipan.waw.pl>

Zapraszamy do przysyłania pytań, uwag, opinii:

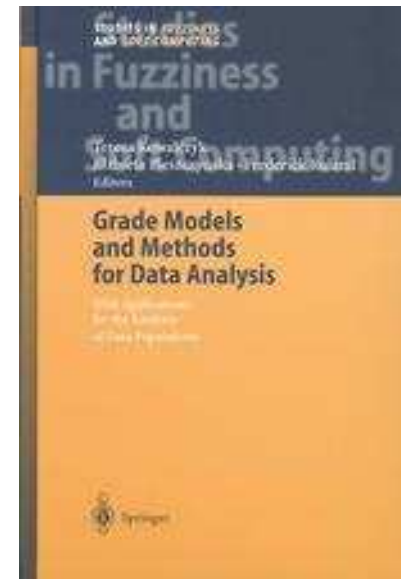
- dotyczących programu GradeStat - gradestat@ipipan.waw.pl
- dystrybucja książki z dołączonym programem - Biblioteka i Dział Wydawniczy IPI PAN biblioteka@ipipan.waw.pl
- niniejsza strona - [Emilia Jarochovska](#), [wersja angielska Marek Wiech](#)

Książki

Grade Models and Methods for Data Analysis

With applications for the analysis
of data populations

Kowalczyk T., Pleszczyńska E., Ruland F. (red.) 2004



Książki

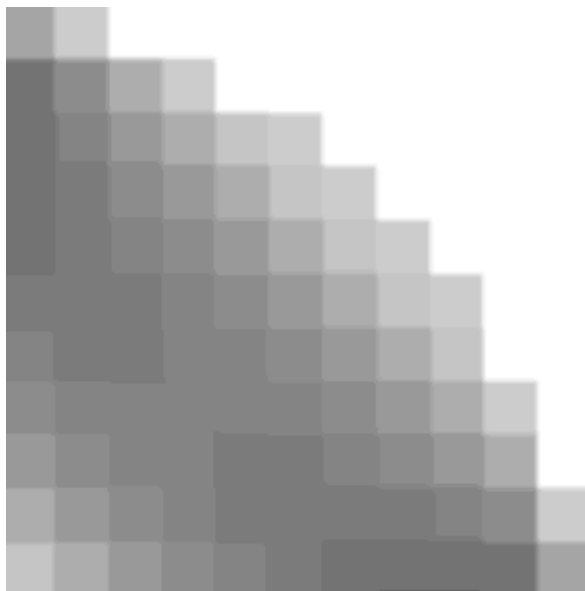
Analiza danych medycznych i demograficznych

Przy użyciu programu GradeStat

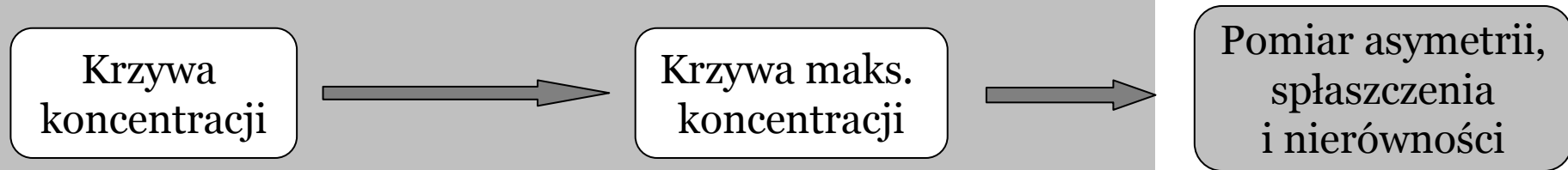
Książek J., Matyja O., Pleszczyńska E., Wiech M. (red.) 2005
książka wydana we współpracy
Instytutu Podstaw Informatyki
z Centrum Zdrowia Dziecka



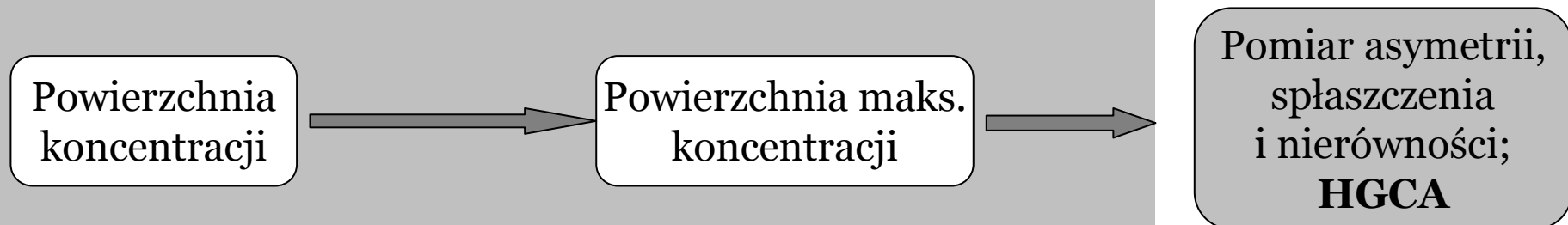
Ku kompletnej
infrastrukturze pojęć
gradacyjnej analizy danych



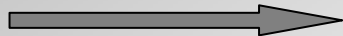
Dla pary zmiennych:



Dla tablicy $m \times k$:



Krzywa koncentracji



Krzywa maks. koncentracji



Pomiar asymetrii, spłaszczenia i nierówności



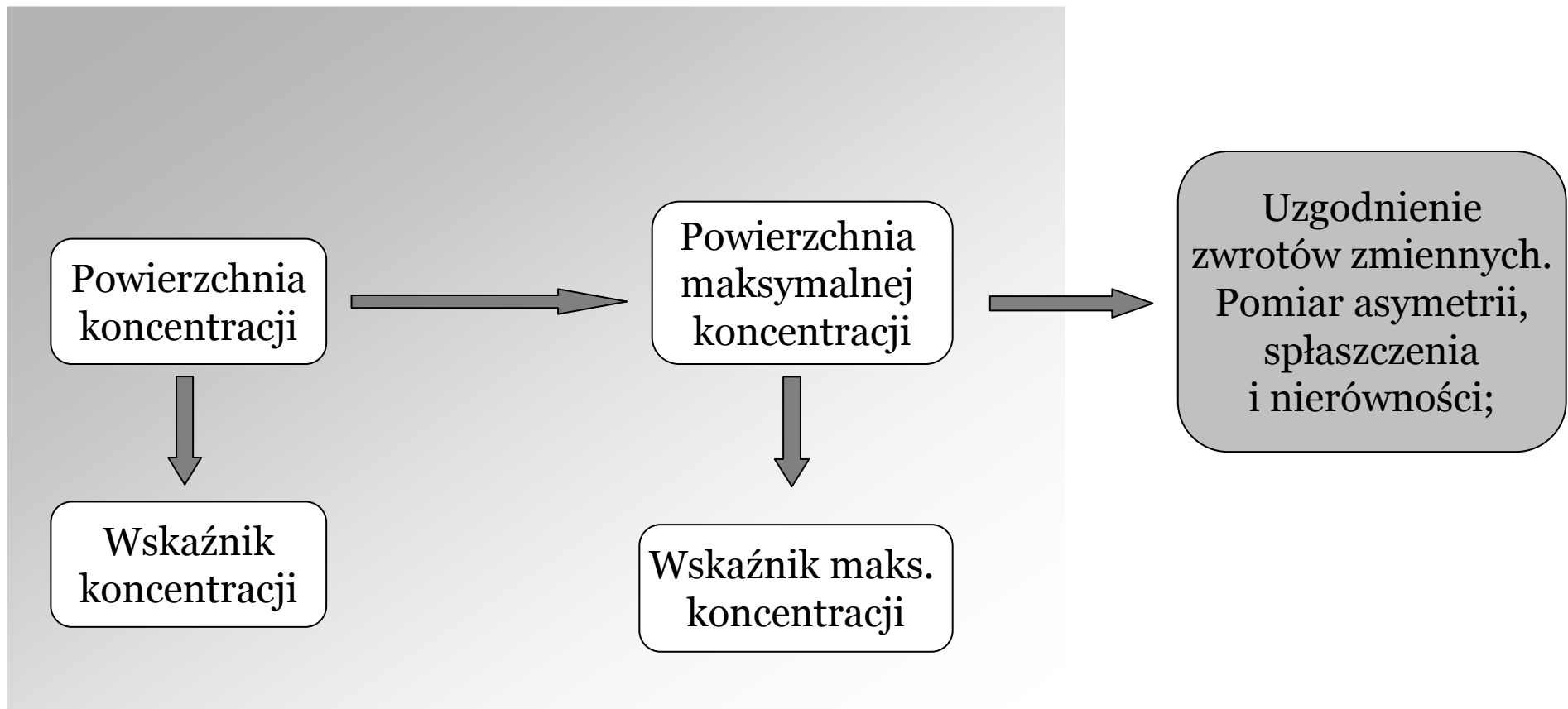
Wskaźnik koncentracji

Krzywa Lorenza



Wskaźnik maks. koncentracji

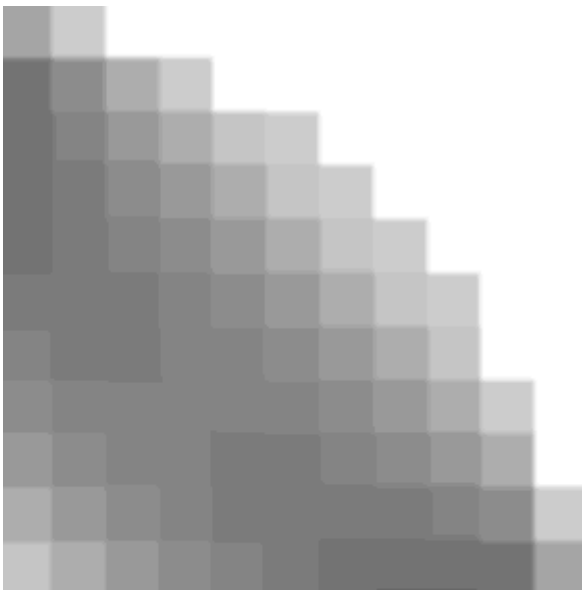
Dla pary zmiennych



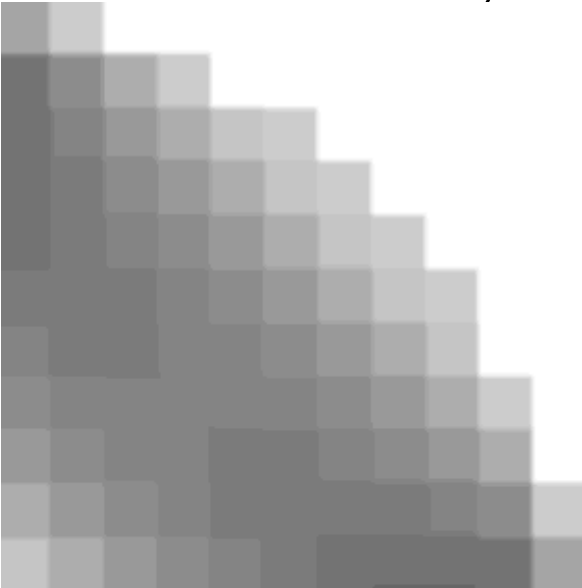
Dla tablicy $m \times k$

Tablica kontyngencji lub
macierz danych wielowymiarowych

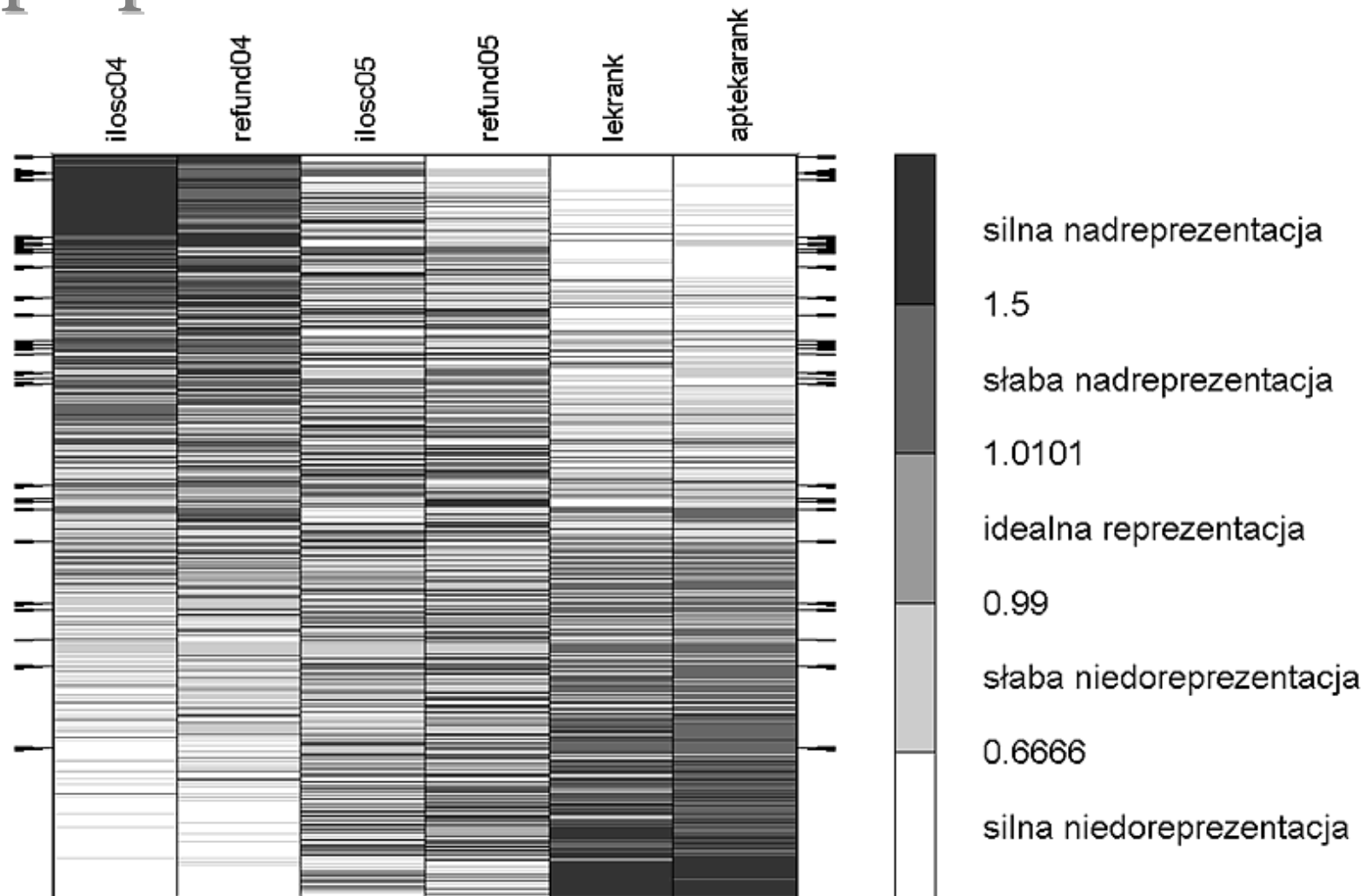
Ten schemat będzie w przyszłości rozwijany
w wielu pracach dotyczących
infrastruktury pojęciowej analizy danych
wielowymiarowych



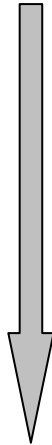
- Powiązania z innymi metodami wizualizacji, np. *generalized association plots* (Szczesny i Wiech, 2006)
- Dekompozycja obrazów medycznych, np. NMR (*Grzegorek, 2005*)
- Prace nad uzgadnianiem zwrotu zmiennych

- 
- European Economic Survey – zastosowanie GCA (m.in. praca magisterska)

Zastosowanie w monitorowaniu ordynacji lekarskiej współpraca z Łódzkim Oddziałem NFZ



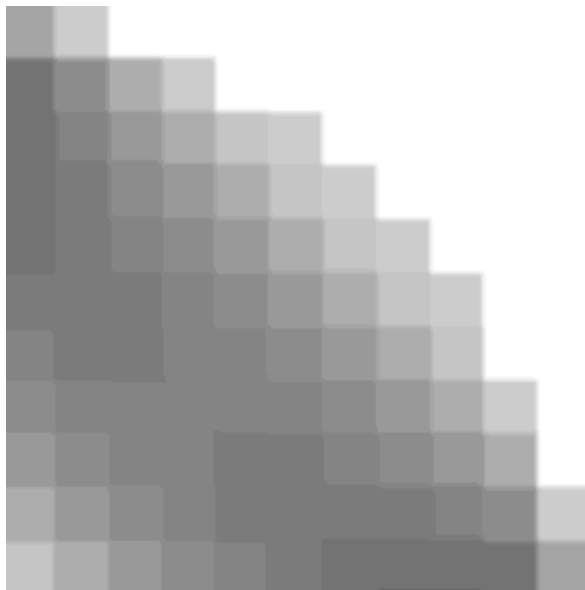
Przykłady zastosowań



Zapraszamy na stronę
<http://gradedstat.ipipan.waw.pl>

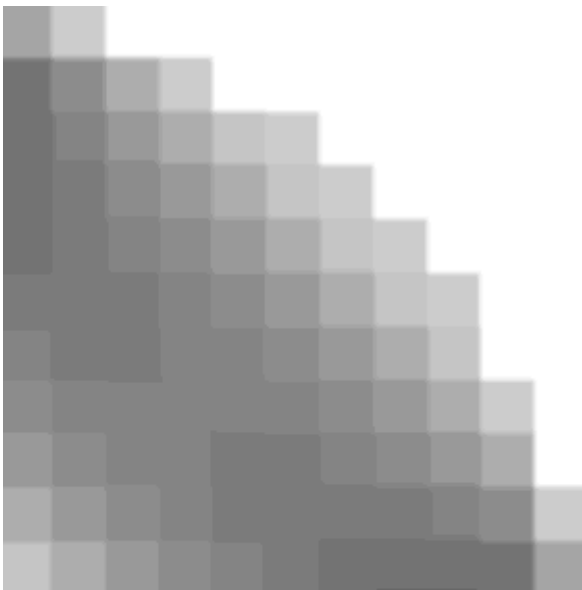


W przygotowaniu książka
Gradacyjna Analiza Danych
dla użytkowników na rozmaitych poziomach
zaawansowania

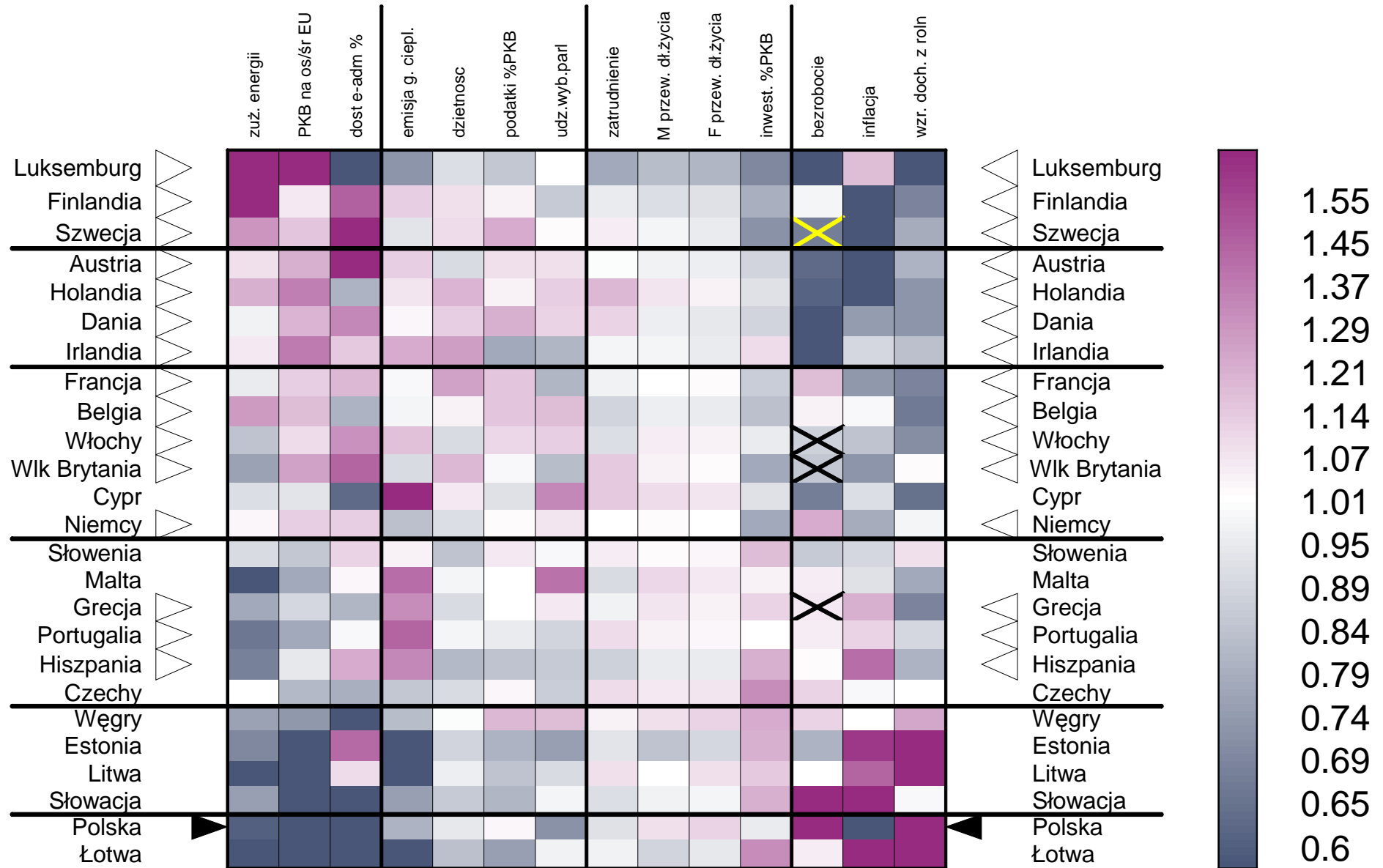


Dziękujemy za uwagę

Wiesław Szczesny wsz@ipipan.waw.pl
Emilia Jarochovska emkaj@hell.pl



Mapa nadreprezentacji w kolorze



Mapa współczynników korelacji rangowej Spearmana

