

Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska

Uniwersytet Ekonomiczny we Wrocławiu

ANALIZA PRZESTRZENNA PROFILU SPOŁECZNEGO WOJEWÓDZTW Z WYKORZYSTANIEM PROGRAMU R

1. Wstęp

Celem analizy przestrzennej jest znalezienie informacji o przestrzennej zależności regionów oraz interakcjach pomiędzy wartościami badanych zmiennych w różnych lokalizacjach. Analiza przestrzenna danych umożliwia określenie podobieństwa i różnic pomiędzy regionami w ujęciu zarówno ogólnym, jak i indywidualnym. Dzięki tym metodom można np. wyodrębnić grupy regionów podobnych do siebie.

W artykule zaprezentowano wyniki analizy, której celem było m.in. wyodrębnienie czynników mających decydujący wpływ na sytuację społeczną w Polsce. Natomiast celem głównym opracowania było porównanie województw pod względem tych czynników, czyli identyfikacja różnic między województwami.

Wizualizacji danych w przestrzeni dokonano w programie **R** za pomocą map z podziałem administracyjnym (mapy typu *Choropleth map*).

2. Wyodrębnianie czynników różnicujących województwa

W badaniu wykorzystano 50 wskaźników określających profil społeczny województw, opisujących m.in.: ludność, rynek pracy, mieszkalnictwo, edukację, kulturę i zdrowie. Do dekompozycji informacji zawartych w zbiorze wskaźników (zmiennych obserwowalnych) wykorzystano analizę czynnikową. Celem analizy czynnikowej jest wykrycie pewnych wspólnych, nieznanych czynników powodujących istnienie zależności między zmiennymi. To znaczy, że obserwowalne zmienne można przedstawić w postaci liniowej funkcji mniejszej liczby nieobserwowalnych (ukrytych) zmiennych, zwanych czynnikami. Czynniki te zachowują znaczną część informacji zawartych w zmiennych pierwotnych i mają inną interpretację merytoryczną [Crawford, Lomas 1980; Sztemberg-Lewandowska 2008].

Z pierwotnego zbioru wskaźników za pomocą analizy czynnikowej wyodrębni-
no podzbiór, który tworzy czynniki różnicujące województwa. Wyodrębni-
ono trzy czynniki: rynek pracy, przedszkola, opieka zdrowotna.

Czynnik rynek pracy jest opisany przez siedem wskaźników zamieszczonych w
tab. 1. Wskaźnik Kaisera-Meyera-Olkin (KMO) dla tych zmiennych wynosi 0,76.

Tabela 1. Ładunki czynnikowe wskaźników opisujących czynnik rynek pracy uzyskane metodą
największej wiarygodności

Zmienne	Ładunki czynnikowe
Ogólne saldo migracji na 1000 ludności	0,634
Wskaźnik zatrudnienia w %	0,988
Wskaźnik zatrudnienia kobiet w %	0,988
Ogólny wskaźnik pracujących zawodowo w %	0,997
Ogólny wskaźnik czynnych zawodowo kobiet w %	0,985
Udział pracujących w %	0,995
Udział pracujących w sektorze prywatnym w %	0,661
Udział wariancji wyjaśnionej przez czynnik w %	82

Źródło: opracowanie własne z wykorzystaniem programu **R**.

W tabeli 2 zamieszczono zmienne opisujące czynnik przedszkola (KMO dla tych
zmiennych wynosi 0,75), natomiast w tab. 3 wskaźniki związane z czynnikiem opie-
ka zdrowotna (KMO = 0,69).

Tabela 2. Ładunki czynnikowe wskaźników opisujących czynnik przedszkola uzyskane metodą
największej wiarygodności

Zmienne	Ładunki czynnikowe
Liczba miejsc w przedszkolach na 1000 ludności	0,915
Liczba dzieci w oddziałach przedszkolnych przy szkołach na 1000 ludności	-0,830
Liczba dzieci w przedszkolach na 100 w wieku 3–6 lat	0,929
Udział wariancji wyjaśnionej przez czynnik w %	80

Źródło: opracowanie własne z wykorzystaniem programu **R**.

Warto zauważyć, że ładunek czynnika przedszkola ze zmienną liczbą dzieci w
oddziałach przedszkolnych przy szkołach jest ujemny, czyli wysoka wartość tej
zmiennej jest niekorzystna. Ujemny jest także ładunek czynnika opieka zdrowotna
ze wskaźnikiem liczba mieszkańców na przychodnię, tzn. im większa liczba osób
przypada na jedną przychodnię, tym gorsza sytuacja tam występuje pod względem
opieki zdrowotnej.

Następnie dla wszystkich województw wyznaczono wartości każdego z czynników (tab. 4).

Tabela 3. Ładunki czynnikowe wskaźników opisujących czynnik opieka zdrowotna uzyskane metodą największej wiarygodności

Zmienne	Ładunki czynnikowe
Łóżka w szpitalach na 10 000 ludności	0,931
Osobodni leczenia na 1000 mieszkańców	0,997
Liczba mieszkańców na przychodnię	-0,765
Udział wariacji wyjaśnionej przez czynnik w %	82

Źródło: opracowanie własne z wykorzystaniem programu **R**.

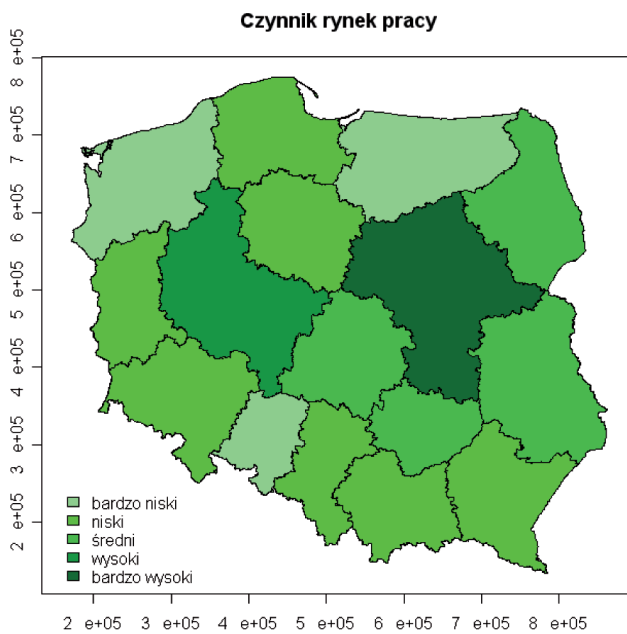
Tabela 4. Poziomy czynników

Województwo	Rynek pracy	Przedszkola	Opieka zdrowotna
Dolnośląskie	-0,292	0,183	-0,421
Kujawsko-pomorskie	-0,113	0,247	-0,989
Lubelskie	0,444	-1,001	1,859
Lubuskie	-0,871	0,514	-0,638
Łódzkie	0,996	0,158	1,314
Małopolskie	-0,104	-0,064	-0,125
Mazowieckie	2,505	0,323	-0,060
Opolskie	-1,305	2,451	-0,619
Podkarpackie	-0,402	-1,032	-0,060
Podlaskie	0,176	-0,791	0,617
Pomorskie	-0,394	-0,760	-1,791
Śląskie	-0,165	1,648	1,884
Świętokrzyskie	0,500	-1,299	0,085
Warmińsko-mazurskie	-1,267	-0,987	-0,856
Wielkopolskie	1,315	0,866	0,012
Zachodniopomorskie	-1,023	-0,457	-0,211

Źródło: opracowanie własne z wykorzystaniem programu **R**.

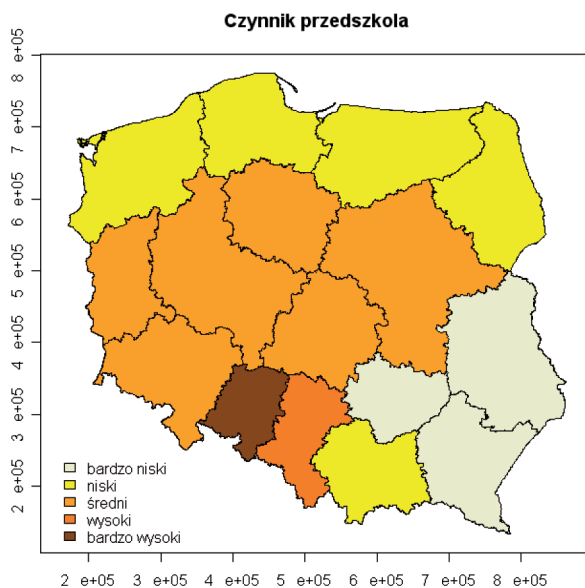
Poziomy czynników zamieszczone w tab. 4 posłużyły do przedstawienia na mapach sytuacji województw pod względem wyodrębnionych czynników.

Mapę z podziałem administracyjnym, grupującą województwa według wartości czynnika rynek pracy, przedstawia rys. 1.



Rys. 1. Mapa z podziałem administracyjnym (województwa) – rynek pracy

Źródło: opracowanie własne z wykorzystaniem programu R.

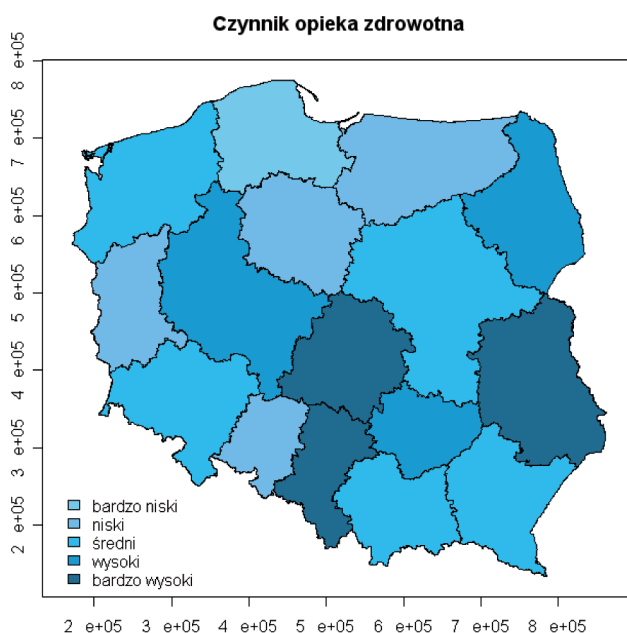


Rys. 2. Mapa z podziałem administracyjnym (województwa) – przedszkola

Źródło: opracowanie własne z wykorzystaniem programu R.

Rynek pracy jest na najwyższym poziomie w woj. mazowieckim, w woj. wielkopolskim jest nieznacznie gorzej. Najniższy poziom tego czynnika występuje w woj. opolskim, warmińsko-mazurskim i zachodniopomorskim.

Sytuacja przedszkoli jest na najwyższym poziomie w woj. opolskim. Natomiast w najgorszym położeniu są woj. podkarpackie, świętokrzyskie, lubelskie.



Rys. 3. Mapa z podziałem administracyjnym (województwa) – opieka zdrowotna

Źródło: opracowanie własne z wykorzystaniem programu **R**.

Opieka zdrowotna jest na najwyższym poziomie w woj. śląskim, łódzkim i lubelskim. Najniższy poziom tego czynnika występuje w woj. pomorskim.

3. Generowanie map

W przypadku analiz przestrzennych podstawą jest zgromadzenie danych o obiektach świata rzeczywistego będących przedmiotem zainteresowania. Dane te, opisujące cechy poszczególnych obiektów, nazywane są atrybutami. Ze względu na rodzaj przechowywanej w atrybutach informacji dzielimy je na:

- atrybuty opisowe – określające właściwości (nieprzestrzenne) i relacje obiektów,
- atrybuty przestrzenne – określające położenie, wielkość i geometryczny kształt obiektów oraz ich przestrzenne (topologiczne) relacje,
- atrybuty czasowe, określające zmienność – dynamikę w czasie.

Dane przestrzenne, w formie cyfrowej, mogą występować jako dane rastrowe lub dane wektorowe. Obraz tych danych składa się na treść mapy numerycznej. Mapa numeryczna (cyfrowa) w modelu wektorowym zbudowana jest z trzech głównych rodzajów obiektów [Sej-Kolasa 2002]:

- punktów – które można umiejscowić dzięki współrzędnym; uporządkowane pary współrzędnych w układzie kartezjańskim: (x, y) .
- linii, odcinków lub linii łamanych – szeregu połączonych ze sobą punktów tworzących linie na mapie; ciągi współrzędnych punktów tworzących linię,
- powierzchni (obszarów, wieloboków) – ciągi współrzędnych punktów tworzących linie ograniczające wielobok, charakteryzujących się pewną szerokością i długością, których granica i wszystkie punkty w obrębie mogą być dokładnie zlokalizowane.

Danym przestrzennym przypisane są dane liczbowe, mające zwykle różne wartości dla poszczególnych obserwacji.

Do modelowania wieloboków (województw) wykorzystano plik typu *shapefile*. Plik *shapefile* został przygotowany w układzie 1992, ponieważ taki układ jest rekomendowanym odwzorowaniem kartograficznym dla Polski¹.

Do wygenerowania takiego pliku wykorzystano program ArcInfo (oprogramowanie typu GIS oparte na oprogramowaniu firmy ESRI), w efekcie otrzymując plik *shapefile*, który:

- definiuje obszary (*.shp),
- przypisuje im identyfikator i kojarzy sąsiadów (*.shx),
- zawiera bazę danych z listą zmiennych (*.dbf).

Wprowadzenie danych przestrzennych wymaga doładowania do bibliotek oraz wczytania pakietów „przestrzennych” (komendą `library()`). Podstawowym pakietem wykorzystywanym do analiz przestrzennych na obszarach jest pakiet `spdep`. Dodatkowo wykorzystywany jest pakiet `maptools` oraz ewentualnie (w zależności od zastosowanej metody wprowadzania danych) pakiet `shapefiles`.

Program **R** wczytuje wszystkie trzy składowe pliku *shapefile* jedną komendą – `read.shapefile()` z pakietu `shapefiles` lub `read.shape()` z pakietu `maptools`².

Pliki muszą spełniać konwencję nazewniczą systemu MS-DOS, jak również pliki tego samego zbioru muszą się znajdować w tym samym folderze.

Efektem wczytania pliku typu *shapefile* komendą `read.shape()` jest obiekt klasy `Map` (`w.shape`). Komenda wczytuje bezpośrednio plik kształtu (*.shp) oraz domyślnie atrybuty pliku, czyli plik bazy danych (*.dbf). Obiekt danych należy utworzyć z atrybutu obiektu klasy `map`. Kolejno należy przekonwertować

¹ Dokonano konwersji z oryginalnego układu WGS-84 do układu 1992. Układ geodezyjny 1992 stanowi podstawę do sporządzania map w skalach 1:10 000 i mniejszych, ze względu na duże zniekształcenia nie jest polecany do opracowań w większych skalach.

² W zależności od przyjętego sposobu postępowania.

uzyskany obiekt (`w.shape`) na obiekt klasy `poly` (komenda `Map2poly()` z pakietu `maptools`), a następnie na obiekt klasy `nb` oraz `listw` (komenda `poly2nb()` oraz `nb2listw()` z pakietu `spdep`) [Kopczewska 2007].

```
> w.shape <- read.shape(„c:/w/w92.shp”) #wczytanie danych (pakiet
  maptools)
> w.dane <- w.shape$att.data
> w.poly <- Map2poly(w.shape) #konwersja obiektu klasy map na obiekt
  klasy poly (pakiet maptools)
> w.nb <- poly2nb(w.poly) #konwersja obiektu klasy poly na obiekt
  klasy nb (pakiet spdep)
> w.listw <- nb2listw(w.nb) #konwersja obiektu klasy nb na obiekt
  klasy listw (pakiet spdep)
```

Plik z danymi opisowymi najwygodniej jest wczytać osobno z formatu `*.csv`. Warunkiem takiego połączenia plików jest zmienna będąca identyfikatorem obszaru. W tym przypadku dane opisowe stanowiły poziomy danego czynnika dla województw (zob. tab. 4).

```
> w.dane <- read.csv2(„c:/w/w92.csv”) #wczytanie danych liczbowych:
  poziomy czynników.
```

Po wczytaniu danych przestrzennych i opisowych można przystąpić do tworzenia mapy z nałożonymi wartościami zmiennej, co wymaga:

- narysowania mapy konturowej; wykorzystana została mapa z podziałem administracyjnym (tzw. *choropleth map*),
- zdefiniowania przedziałów wartości dla grup kolorystycznych (dolne granice przedziałów; przedziały grupowania wynikają z wartości ładunków czynnikowych – zob. tab. 4),
- zdefiniowania palety kolorów, jakimi będzie wypełniona mapa konturowa,
- dodania legendy oraz tytułu do rysunku.

Poniżej zamieszczono fragment kodu w języku R (mapa dla czynnika rynek pracy):

```
> plot(w.poly, forcefill=FALSE) #rysunek mapy konturowej
> range <- c(-2,-1,0,1,2) # definicja przedziałów wartości dla grup
  kolorystycznych
> cols <- c(„lightgreen”, „green”, „limegreen”, „green4”, „dark-
  green”) # definicja palety kolorów
> plot(w.poly, col=cols[findInterval(w.dane$cz1, range)],
  forcefill=FALSE) #generowanie mapy
> # dodanie legendy i tytułu
> legend(x=c(150000,150000),y=c(250000,250000), fill=cols, legen-
  d=c(„bardzo niski”, „niski”, „średni”, „wysoki”, „bardzo wyso-
  ki”), bty=„n”)
> title(„Czynnik rynek pracy”).
```

4. Podsumowanie

Przeprowadzone badanie pozwala na sformułowanie wniosków dotyczących sytuacji społecznej województw. Do klasyfikacji województw pod względem sytuacji społecznej posłużyły wartości średnie z unormowanych poziomów wyodrębnionych czynników dla poszczególnych województw (tab. 5).

Tabela 5. Średnie z poziomów wyodrębnionych czynników

Województwo	Średnia z poziomów czynników
Śląskie	0,695
Mazowieckie	0,634
Łódzkie	0,613
Wielkopolskie	0,585
Lubelskie	0,511
Opolskie	0,440
Podlaskie	0,393
Małopolskie	0,366
Dolnośląskie	0,345
Świętokrzyskie	0,328
Kujawsko-pomorskie	0,314
Lubuskie	0,304
Podkarpackie	0,260
Zachodniopomorskie	0,243
Pomorskie	0,128
Warmińsko-mazurskie	0,116

Źródło: opracowanie własne.

Biorąc pod uwagę wszystkie czynniki, można stwierdzić, że najlepsza sytuacja pod względem społecznym jest w województwach: śląskim, mazowieckim i łódzkim. W Śląskiem czynnik rynek pracy jest na niskim, przedszkola na wysokim, a opieka zdrowotna na bardzo wysokim poziomie. W województwie mazowieckim czynnik rynek pracy jest na bardzo wysokim poziomie, natomiast czynniki opieka zdrowotna i przedszkola na średnim poziomie. Z kolei w Łódzkim czynniki rynek pracy i przedszkola są na średnim poziomie, opieka zdrowotna na bardzo wysokim poziomie.

Najgorsza sytuacja jest w województwach pomorskim i warmińsko-mazurskim. W województwie pomorskim czynniki rynek pracy i przedszkola są na niskim poziomie, opieka zdrowotna na bardzo niskim. W woj. warmińsko-mazurskim czynnik rynek pracy jest na bardzo niskim poziomie, a czynniki przedszkola i opieka zdrowotna na niskim.

Przeprowadzone analizy pozwoliły także na sformułowanie wniosków dotyczących metod wykorzystanych w badaniu. W przypadku danych mających atrybut przestrzenny konieczne jest przedstawienie ich na mapie. Dobrym narzędziem do takiej wizualizacji danych przestrzennych jest środowisko **R**, musimy dysponować jedynie plikiem typu *shapefile*. **R** jest pewną alternatywą dla programów typu GIS, umożliwia mapowanie wartości i wyników obliczeń, dając równocześnie o wiele lepsze narzędzia niż GIS do modelowania procesów przestrzennych (oferuje pakiety do analiz zarówno punktowych, jak i obszarowych). Wizualizacja danych znacznie ułatwia analizę i pozwala odkryć zależności, niekiedy niewidoczne, jeśli dysponujemy tylko danymi liczbowymi.

Literatura

- Crawford I.M., Lomas R.A., *Factor Analysis – a Tool for Data Reduction*, „European Journal of Marketing” 1980 vol. 14, no 7, s. 414-421.
- Kopczewska K., *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R*, CeDeWu, Warszawa 2007.
- Sej-Kolasa M., *Podstawy zarządzania informacją o środowisku*, AE, Wrocław 2002.
- Sztemberg-Lewandowska M., *Analiza czynnikowa w badaniach marketingowych*, UE, Wrocław 2008.

SPATIAL ANALYSIS OF THE SOCIAL PROFILES OF PROVINCES WITH THE USE OF THE R SOFTWARE

Summary

A spatial analysis of data makes it possible to define the similarities and differences among regions. The aim of this study was to identify the determiners of the social situation in Poland. Another important aspect was a comparative analysis of the determiner strengths in each region. The analysis covered 50 indices describing the social profile of every Polish province, including the sizes and levels of population, labour market, housing, education, culture and healthcare. Decomposition of the data covered by the set of indices (random variables) was performed through factor analysis. The spatial visualisation of data was performed with the **R** software. The data was plotted against choropleth maps of the administrative division of Poland. The basis of map creation was a shapefile generated with GIS-type software.