IDENTIFICATION OF SPECIFIC AREAS WITHIN PROVINCIAL CAPITAL CITIES AND THEIR FUNCTIONAL AREAS IN TERMS OF THE DEMOGRAPHIC AND ECONOMIC SITUATION OF THEIR INHABITANTS USING GIS-BASED SPATIAL ANALYSIS

Name (s) of author(s):

Adam Dąbrowski, Sylwia Filas-Przybył, Dawid Pawlikowski

Statistical Office in Poznan (POLAND)

A.Dabrowski@stat.gov.pl

Organization: **Central Statistical Office in Poland**

Introduction

Over the last few years Polish cities have witnessed rapid changes in their demographic structure, particularly as regards aging and suburban migrations. Until now there has been no data on a sufficient level to analyze those processes within the city boundaries. Fortunately, the latest census in Poland (2011) has produced detailed information, including geographic coordinates, which can be used to perform more in-depth spatial analysis and to identify specific areas in terms of the demographic and economic situation of the population. We have selected and calculated indicators necessary to analyze socio-economic variation within selected cities. On this basis we have been able to distinguish specific, homogenous areas. This work has been prepared as a part of the project funded from the European Union funds from “Technical Assistance” Operational Programme 2014-2020: “Identification of specific areas within provincial capital cities and their functional areas in terms of the demographic and economic situation of their inhabitants using GIS-based spatial analysis”.

Methods

The study was conducted based on administrative data collected during census from 2011, which characterize the population‘s socio-economic structure. The geographical scope of the research was limited to provincial capitals. The results of the study presented in this paper have been shown on the example of Poznań – the provincial capital of Greater Poland region. It is populated by 545 680 people and has the area of 262 km^2. About 235.6 thousand people are being employed and the registered unemployment rate is equal to 3.1% which places Poznań as the city with the least unemployed rate among the provincial capitals.

All of the indicators were aggregated to a grid of 500m x 500m squares. One advantage of the grid approach over administrative boundaries is its stability over time, making it possible to monitor changes and analyze their dynamics, which was one of the key points of the methodology used in the project. Administrative boundaries present the data in a generalized form often blurring the real structure and concealing any spatial patterns. To remain privacy it was necessary to exclude from the analysis squares populated by less than 10 people. Other sizes of grids have also been tested but, while preserving the above mentioned privacy rule, smaller squares have excluded too much population and bigger ones presented the data in a generalized form often resulting in too few spatial objects to properly calculate some statistics.

The target indicators include 12 demographic (Tab.1) and 7 economic indices (Tab.2). In addition to those indices, Local Indicators of Spatial Association (LISA) have been computed for five demographic indexes to determine the presence of elderly people and identify specific areas of advanced age (Tab.3). Those five indexes were: the median age, the index of elderly people by Cyrus Chu, the support for the eldest people index, the dependency ratio of the elderly and the index of the elderly.

LISA allows to show the presence of spatial autocorrelation of analysed data. It is useful to determine the existence of clusters of high or low value. It indicates as well the existence of hot or cold spots, meaning high values surrounded by low values or vice versa. By calculating local Moran's I statistic (one of the LISA indicators) for five demographic indexes that characterise the influence of the elderly people on the demographic structure we were able to pin point the areas with the most probable problems with aging population.

Tab. 1 Demographic indicators used in the study

ID.	NAME	EQUATION
1	Total population	L
2	Feminization	$\frac{L_f}{L_m} * 100$
3	The percentage of people aged 0-14	$\frac{L_{(0-14)}}{L} * 100$
4	The percentage of people aged 15-64	$\frac{L_{(15-64)}}{L} * 100$
5	The percentage of people aged 65+	$\frac{L_{(65+)}}{L} * 100$
6	Median age	-

7	The index of elderly people by Cyrus Chu[1]	$I = \frac{1}{\omega - z} \sum_{p_j = p_z}^{p_\omega} (j - z) * p_j$ where: *I* – index of elderly people, p_j – percentage of people in range j, *z* – the threshold of old age, *ω* – the upper limit of the oldest age range
8	The index of elderly people by United Nations (UN)	$\frac{L_{(65+)}}{L} * 100$
9	The support for the eldest people index	$\frac{L_{(85+)}}{L_{(50-64)}} * 100$
10	The dependency ratio of the children	$\frac{L_{(0-14)}}{L_{(15-64)}} * 100$
11	The dependency ratio of the elderly	$\frac{L_{(65+)}}{L_{(15-64)}} * 100$
12	The index of elderly people	$\frac{L_{(65+)}}{L_{(0-14)}} * 100$

Tab. 2 Economic indicators used in the study.

ID.	NAME
13	The share of people for which the main source of income is the wage labor
14	The share of people for which the main source of income is the retirement pay/pension
15	The share of elderly people for which the main source of income is the wage labor in the total population of elderly people
16	Dominant source of income
17	The share of people receiving a pension in the total population in working age
18	The share of people receiving unemployment benefits in the total population in working age
19	The share of people receiving social pension or social assistance benefits in the total population

The local Moran's I statistic has the advantage over other statistical analysis by incorporating in spatial dimension the values around the region of interest. When it is calculated for areas that

[1] Kurek S. (2008), *Typologia starzenia się ludności Polski w ujęciu przestrzennym*, Wydawnictwo Naukowe Akademii Pedagogicznej, Kraków

have diverse boundaries (e.g. administrative boundaries) there is often a problem of big differences in number of neighbours. However calculating it for the grid of squares has lessen this problem. By defining the neighbourhood as the point or edge contact most of the squares had around 8 surrounding values (apart from those surrounded by grids populated by less than 10 people). The local Moran's I statistic is being calculated using the following equation:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^{n} w_{i,j}(x_j - \bar{X})$$

where:

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^{n} (x_j - \bar{X})^2}{n-1} - \bar{X}^2,$$

x_i is the value of analyzed unit (single square), $\bar{x}$ is the mean value from the city and w_{ij} is the weight matrix of the surrounding squares[2].

Most of the indexes have been calculated as a simple percentage or ratio of 2 values. However for calculating the dominant source of income another method had to be used. Our goal was to present the residents' economic structure. Since most of the residents work as employees (wage labor), treating the main source of income as the one with the highest percent in the area would result in an almost monochromatic map. The solution was to use the location quotient. It shows how the value in a single grid differs from the corresponding value calculated for the whole region (in this case: the city of Poznań). It is being calculated using the following equation:

$$LQ_i^r = \frac{x_{ri}/x_{r.}}{x_{.i}/x_{..}},$$

where:

x_{ri}	-	The number of people in a square *r*, for which the main source of income is the *i* source of income
$x_{r.} = \sum_i x_{ri}$	-	The total population in square *r*
$x_{.i} = \sum x_{ri}$	-	The number of people in the city, for which the main source of income is the *i* source of income
$x_{..} = \sum_r^T x_{r.} = \sum_i x_{.i} = \sum_T \sum_i x_{ri}$	-	The total population of the city

[2] Anselin L. (1995), *Local Indicators of Spatial Association – LISA*, Geographical Analysis 27(2), p. 93-115

Results

The outcome of the study is a grid of squares in a shapefile format and two datasets in the form of Excel databases and maps. Excel file includes all of the indices and unique identification numbers, which can be used to perform a spatial join with the grid shapefile. The maps contain 25 cartograms illustrating spatial variation of all the indices for each city. For the purpose of the paper Fig. 1 - 4 present the following indexes: total population (Fig.1), the local Moran's I statistic calculated for the median age (Fig.2), the synthesis of all high values calculated via local Moran's I statistic (Fig.3) and the dominant source of income calculated using location quotient (Fig.4). Other maps are publicly available on the Polish Central Statistical Office website (www.stat.gov.pl/en/) .

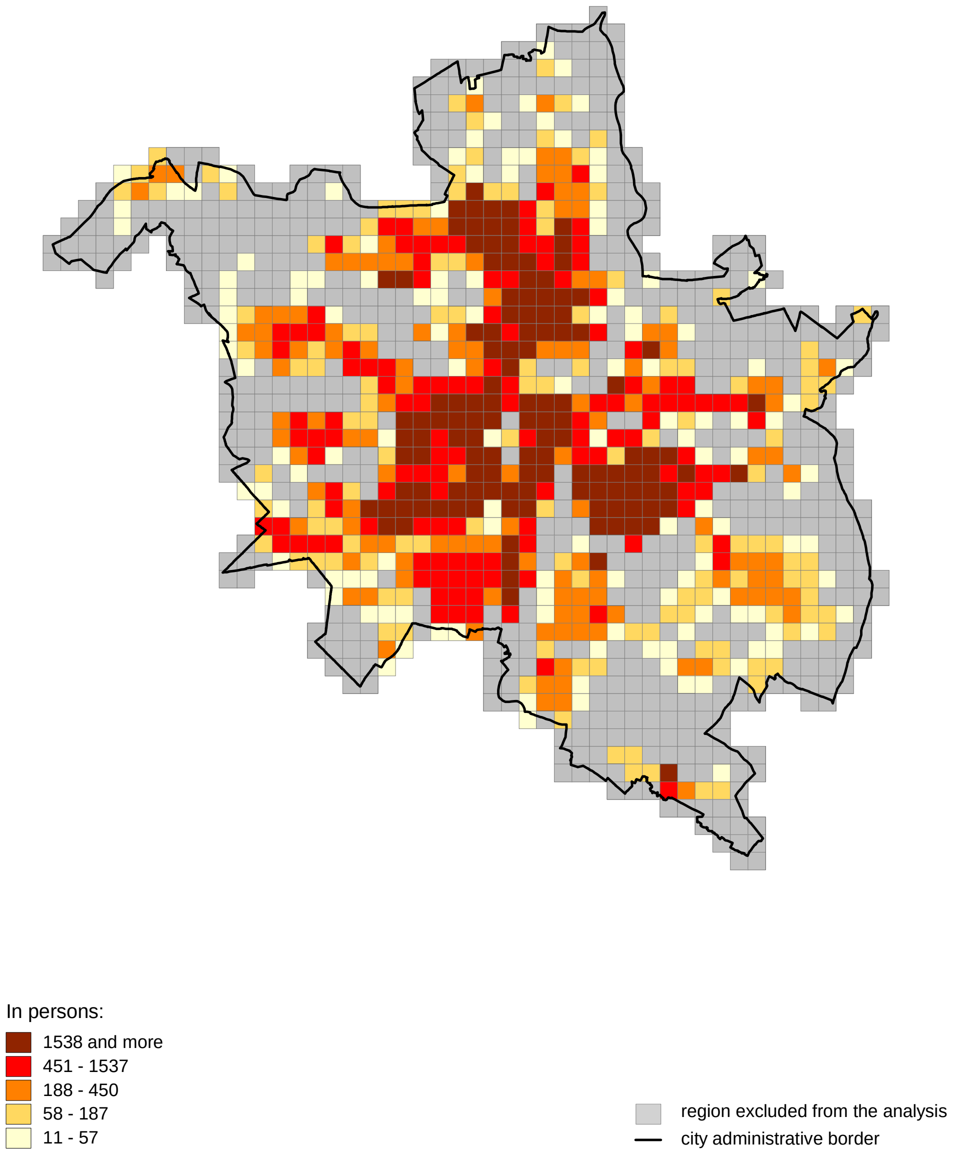

Fig. 1 Total population.

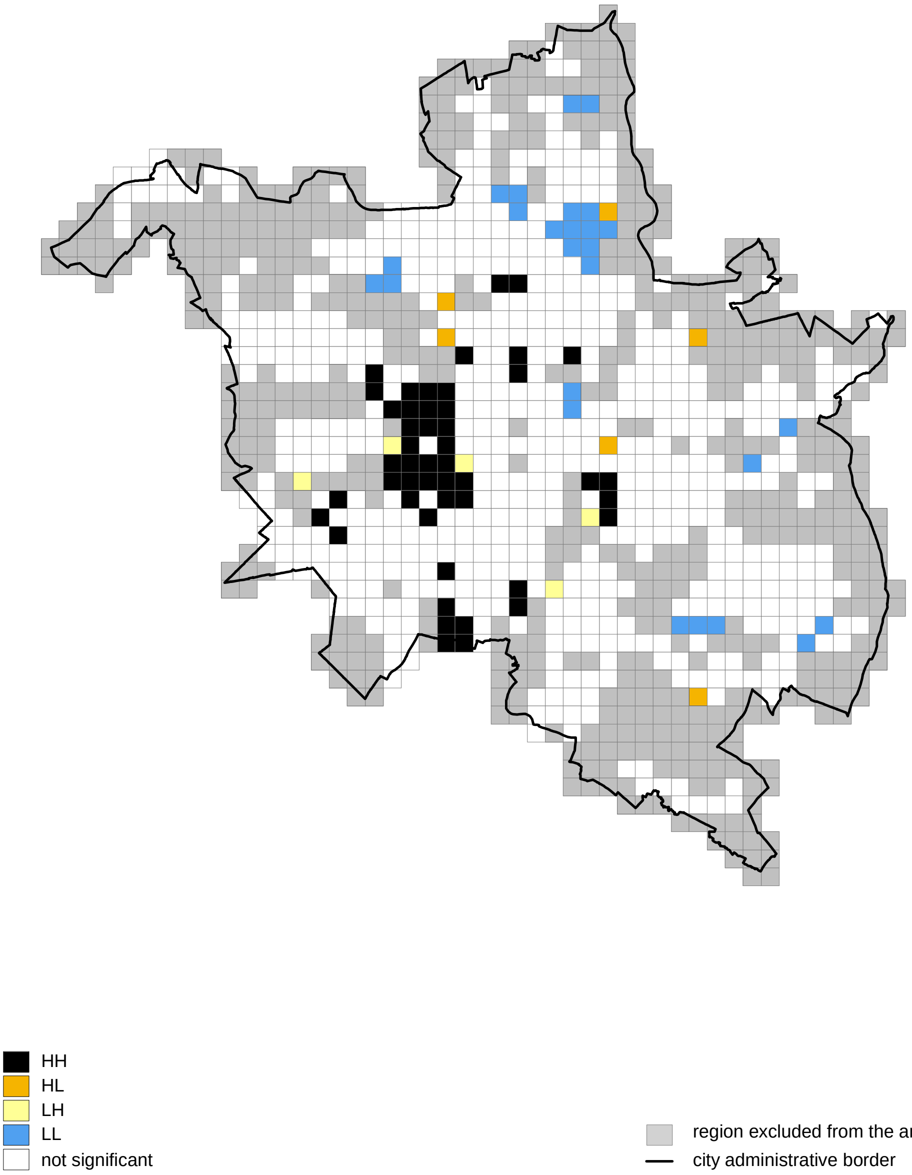

Fig. 2 Local Moran's I statistic calculated for the median age. HH are the clusters of high value surrounded by other high values; HL are the hot spots: high values surrounded by low values; LH are the cold spots: low values surrounded by high values; LL are the clusters of low values surrounded by other low values.

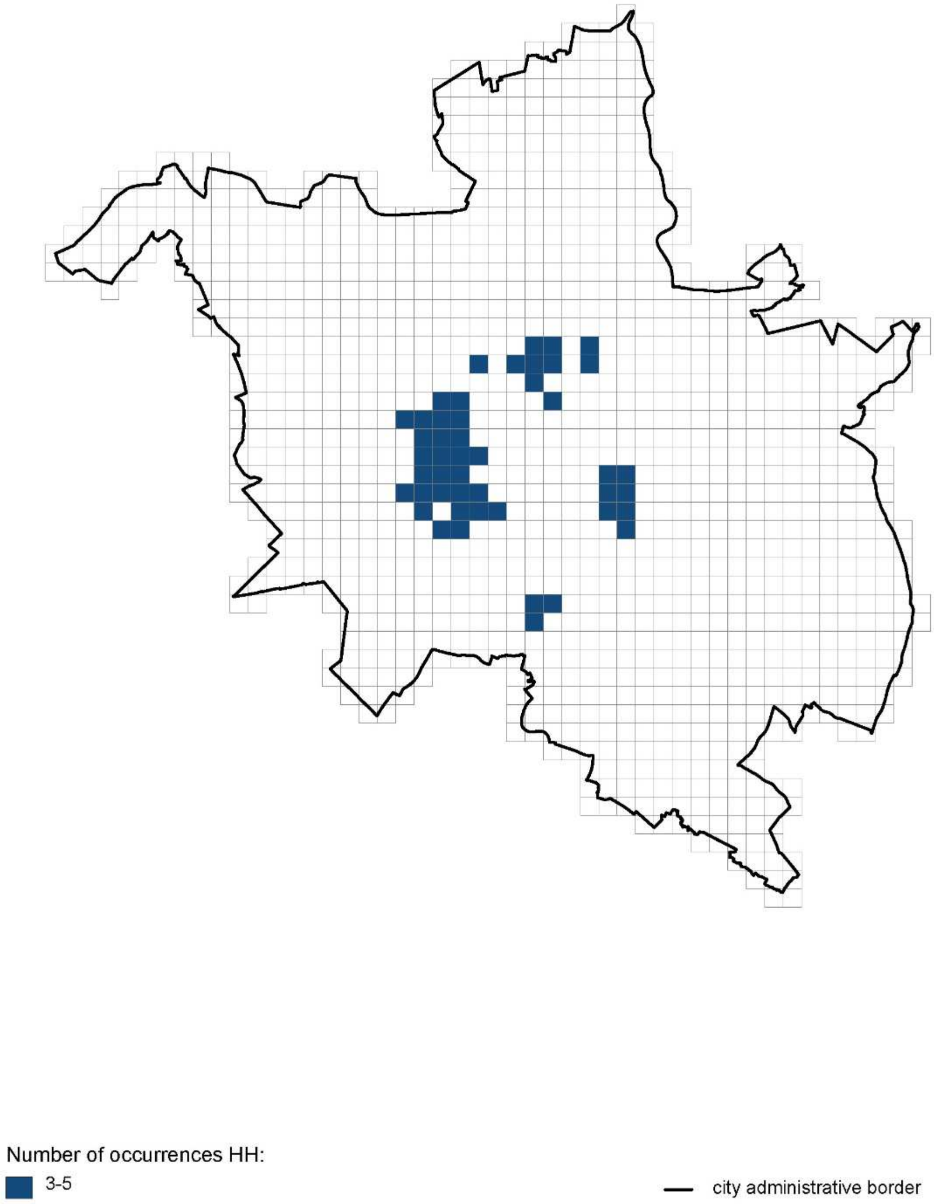

Fig. 3 The synthesis of all HH values determining the areas with the most probable problems with aging population.

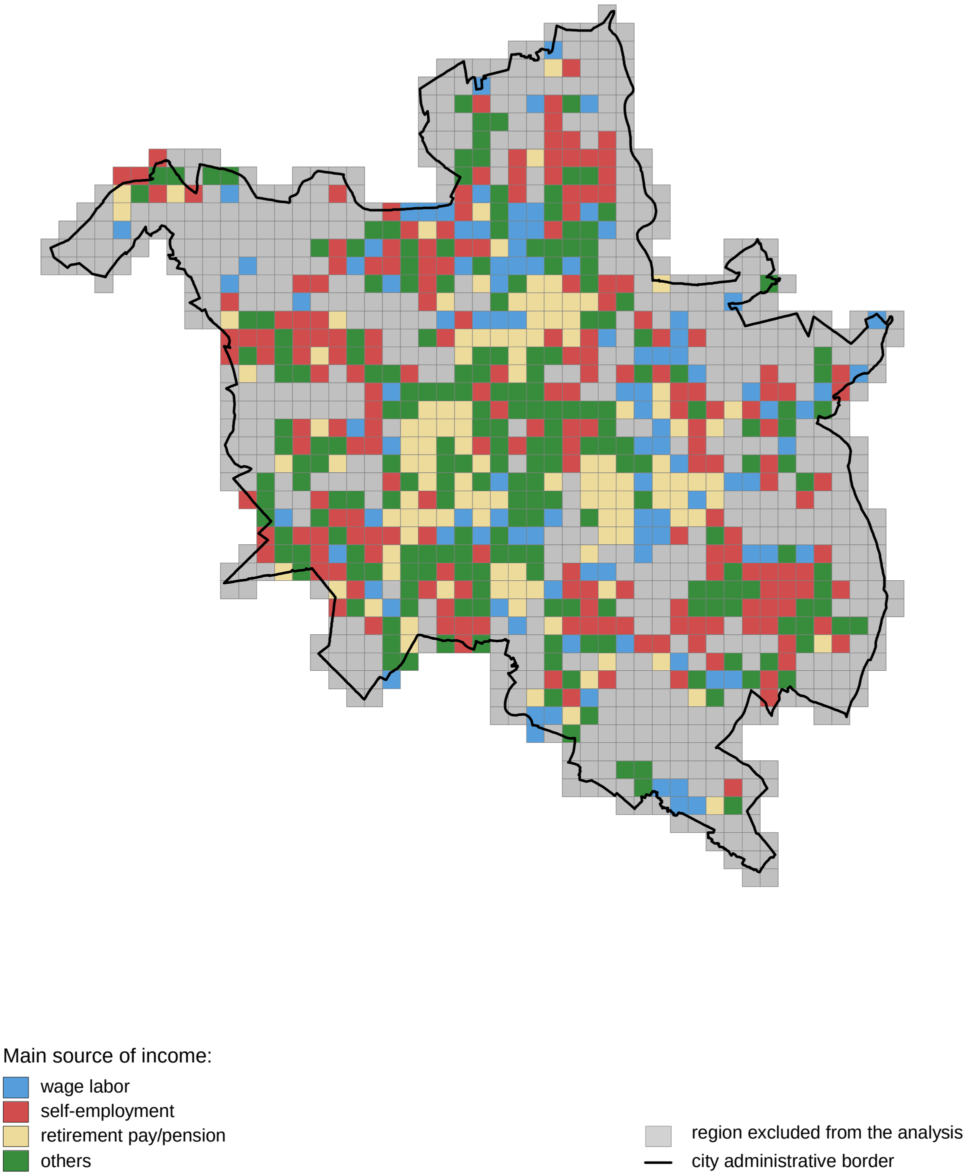

Fig. 4 The main source of income calculated by the location quotient.

Discussion

All the results are now publicly available and can be used for further analysis to enhance decision making processes at different administrative levels. It provides information on spatial distribution of elderly people as well as the economic structure such as the percentage of unemployed and the main source of income. By combining this knowledge with other information it may prove to be a useful tool for city planning or health care management.

The grid of 500m x 500m squares has proved to be a reliable spatial division that preserves its boundaries in time and is small enough to observe spatial patterns in local scale. It could be of great value to perform further analysis in this division by including different data sources e.g. income, crime or level of education.