# RELATIONSHIP BETWEEN CITY'S PHYSICAL STRUCTURE AND ECONOMIC PERFORMANCE

Presented to the Faculty

ISM University of Management and Economics

in Fulfilment of the Requirements for the Degree of

Master in Financial Economics

by

Karolina Štreimikytė

May 2018

## Table of content

**Abstract**

An average European urban population is 70% and cities are shown to be the driving force of economic growth. Cities themselves are complex functional and physical structures. Although there have been studies that correlate particular features of cities with their economic performance for decades, only recent availability of small scale multi-layered geographical data has opened up an opportunity to dive deeper into the overall complexity of cities.

This study constructs and uses structural, functional and geographical features of cities, clusters them into 8 groups with unsupervised learning k-means algorithm and briefly evaluates the differences between the economic indicators of clusters. The results show that clusters perform differently among themselves from the economic perspective. They differ in their employment composition, GVA per worker output, education levels and composition, numbers of patents produced.

The findings lead to conclusions that economic performance and structural characters of cities are connected and that agglomeration economies play an important role in the overall map of city competitiveness.

## Introduction

Urbanization is a dominant trend, it influences societies, cultures, economies and the environment worldwide (OECD, 2012a). 55% of the world population is now living in cities and the average of European urban population around 70% of country population. As economies across the world have become increasingly integrated, 'global cities' – better connected cities - have been seen more and more as dominant economically competitive nodes in the networked global economy (Pain et al., 2015). 600 largest cities in the world are projected to generate 60 percent of global GDP by 2025 (Dobbs et al., 2011). Also, it has been recognized that analysis at the National level is not sufficient anymore and that intra-national scale economic analysis is needed to truly capture the essence of economic growth (Porter, 2003).

The European Union (EU) forms a transnationally interlinked economic and institutional space that is being considered as an economic block which is put out to the global competition with other large economic blocks in North America and Asia. It is aiming to become a worldwide leading economic territory of the knowledge-intensive economy ("Lisbon process") (Kratke, 2007). In the meantime, European Union unification process is diminishing the importance of national differences through increasingly unifying legislations and due to the convergence of economic performance and therefore increases the importance of the competitiveness among cities within the EU. European urban areas and cities vary in their geographic situations, climate, heritage and trajectory, activities, governance, population, etc. (Aksoy et al., 2016).

Mobility, is one of the cornerstones of European integration: the free movement of people, capital, goods and services. Catching migrating human and economic capital can mean a sustainable economic future for a city. Due to the decreasing costs of transportation, movement of physical goods is becoming less expensive and movement of information is practically free.

Nonetheless, frequent movement of people is becoming increasingly more expensive and cities are taking the stage as important facilitators of contact of people (Glaeser & Kohlhase, 2004). Therefore, competition for relocating inhabitants is recognized, especially when taking into account the importance of demographic changes: it is vital for the city to maintain a healthy demographical structure and competition among cities is only increasing (Bessis, 2016).

Maximizing and rebalancing economic performance of urban areas is at the top of the national and regional agendas (OECD, 2012a). Agglomeration, industrial clustering and innovation are recognized as having positive externalities and appear frequently in dense urban areas, but at this point, it is widely accepted, that geographical shape is as important or sometimes defines the performance better than simply the scale of agglomeration.

The development of the EU territory reveals considerable inconsistencies among cities: connectivity, concentration, activity employment rates, living conditions and the amount of human capital. Dias Jordão et al. (2017) have found that Intellectual Capital has a positive influence on corporate profitability and return of companies either individually, globally or by industry and that IC helps increase financial performance, systematically, over time. Cities are recognized as platforms for agglomerations of such companies and their workers. A city cannot base the sole development of Creative Capital (Kratke, 2011), but the regions' capacities in the field of innovation activity, research and technological development (Cooke et al., 2000) are considered as a key factor of regional economic success.

The analysis of the cities has been taking place for many years, but only recently the extensive collection of data and availability of quantitative analysis tools have allowed researchers to evaluate the city as a complex system and calculate some features that could not be evaluated in a simplified manner. The interdisciplinary approach is becoming more applicable and available and economic fields such as urban economics, economic geography or developmental economics have been an increasing field of interest. Newly available data is scaling into smaller modes and gives the opportunity to examine those differences while comparing the performance of urban areas based on economic function rather than administrative boundaries of countries or greater regions (OECD, 2012a). Unified European estimations are now available for cities, greater cities and urban centers with the number of inhabitants of at least 50 000 for the EU, Switzerland, Croatia, Iceland and Norway.

Can the sustainable economic and financial performance of a city be predicted by its physical structure? Are there common physical features of more competitive cities? This thesis will aim to recognize the unifying structural factors of the most competitive cities. The study will base research on data at different areas: at level of a city and its administrative bounds and the level of functional urban areas (FUA).

# Literature Review

## Cities and regions

Cities are organized hierarchically into areas, neighborhoods, blocks, whose scale and geometrical characteristics depend on a number of factors, including their economic purpose and function (Buhl et al., 2003). Cities tend to have one or multiple centers (mono and multi-centered cities), the structures are divided into areas (the old town, industrial areas, etc.) which hold one or multiple functions. Infrastructure connections are vital for a city: electricity, water, central heating, gas and road networks. Since road networks are the most physically demanding of the engineering infrastructures, it plays a vital role in shaping the geometry of cities. The larger areas of cities with similar characteristics tend to be divided and also joined by the roads of different intensity and classes. Roads act as joining elements when an object has to be transferred between remote points, but they act as a division when analyzing neighboring areas. Both of these functions increase with the size (width) and the intensity of the streets. As a tendency, one of the defining characteristics of an urban area is the density and it's gradient: urban and suburban characters.

Often, a wider net of cities for a country or an interconnected structure follows a Zipf's law, which states that the ordinal number *n* of the city by size contains *1/(n+1)* of the structures population (the biggest city in the network contains ½ of the reachable population), however this only holds if the administrative boundaries are ignored and cities are redefined by their functional area (Veneri, 2013) which is done by the new definition of metropolitan areas (OECD, 2012a). Redefining urban "functional economic units", which since then have been adapted as functional urban areas – FUAs. The definition calculations were proposed by OECD in 2012 and applied by the European Union later. This new approach gives the possibility to compare cities in a completely new manner and scale. Combining the FUA data with city administrative data can provide us with fresh insights.

**What predicts the well-being of a city or a region? Logics of cities**

**Human Capital**

Human capital tends to hold a decisive role in studies that analyze the attractiveness or economic sustainability of a city (Bessis, 2016). Higher stock of human capital is positively correlated with urban productivity as shown by Abel, Dey & Gabe (2012). Glaeser & Saiz (2004) argue that the best predictor for the growth of a city is their geographical position, exclusively – the mean yearly temperature, which positively correlates with the attractiveness and performance of the city. The second best estimator – skill composition of a city. Caragliu et al. (2011) adds to the argument by finding that city's endowment of hard infrastructure (physical capital) is not as decisive in predicting urban competitiveness as the human and social capital (access to quality information and communication technologies). But the competitiveness is highly related to scale. The tendency to host universities, R&D facilities, and other knowledge-generating institutions correlates with the city size (Van Oort, 2004; Meijers et al., 2015), bigger cities also tend to provide a good environment for information exchange and consumption (Meijers et al. 2015).

The ability of urban concentration to connect people and exchange ideas is closely associated with the concept of human capital. Human capital predicts population and productivity growth at the city and metropolitan area level as surely as it predicts income growth at the country level (Glaeser & Saiz, 2004). Graham (2009) finds that urban concentration has large effects on the productivity of the sectors that are disproportionately common in highly urbanized area: transport services, business and management consultancy, financial services and public services.

Frick & Rodríguez-Pose (2018) add that "it is well known that different sectors benefit from agglomeration economies to a varying degree with high-tech industries and the professional services sectors among those that benefit the most". They also find results that prove the importance of urban concentration for developed countries, which as a rule are highly dependent on information-based industries. As already stated – cities are complex and vary greatly with their ability to nurture different industries and functions. Therefor it is no surprise that studies such as Van Winden et al. (2007) show structural differences among European cities endowment of human capital and knowledge – based industries and the consequential variety of potential to thrive in a knowledge - based society.

**Agglomeration Benefits**

Links between human capital and agglomeration are shown in various research. It is clearly understood that economical units with more possibilities to obtain assets have a clear advantage. Meijers et al. (2015) look at the agglomeration from the view of a firm and claim that costs of maintaining a firm are smaller in "[l]arger, denser and more diverse cities". These firms have an advantage which leads to higher outputs and utility gains for firms and households as well. "They profit from larger input markets, larger labor pools, the presence of better infrastructure, public facilities and more specialized business services, all facilitating better matches between supply and demand" (Meijers et al., 2015).

Due to the agglomeration effects, high-income, high-technology and intense human capital clusters - concentration can boost the overall productivity of the economy (Ozyurt & Dees, 2015). The concentration of employment is found higher than the concentration of population, which suggests that employment agglomeration produces more economic benefits than the housing agglomeration (Tsai, 2005).

Latest studies by Frick & Rodriguez-Pose (2018) have also found that "urban concentration is beneficial for economic growth in high-income countries", although this will not hold true for developing countries. Furthermore, some studies shift from a mere concentration analysis. Distribution among cities and within cities are topics explored. The European-level city distribution and growth differences have shown that Europe has already surpassed the mega growth period, where economic growth is concentrated in the first-tier cities. Studies argue that oversimplified applications of agglomeration and city size are a thing of the past. Europe is a connected polycentric structure, which can enable cities to "borrow" populations, jobs and thus be more efficient as an overall structure (Dijkstra et al., 2013). Proximity and connection to other regions and cities also show a spillover effect: "business investment and human capital of the neighbouring regions have a positive impact — both direct and indirect — on economic performance of a given region" (Ozyurt & Dees, 2015).

**Urbanization**

The world as a whole is still under great wave of urbanization, but developed countries and Europe as a continent is recognized as a region that also experiences reverse effects (Dijkstra et al., 2013, Meijers et al., 2015). Urbanization trends in some developed European countries have slowed down or even reversed. This adds to the understanding that agglomeration is not the sole predictor of performance and that the structure of functional layouts and connection has great effect. It is important to notice, that, as Henderson (2003) states, the process of urbanization is merely transitional and is not a universal measure to estimate growth.

Henderson (2003) also argues that there is a degree of urban concentration that results in the highest output growth numbers. He also calculates that there are costs involved in under or over-concentrated regions. The degree of change of population concentration (which implies urbanization or reverse processes) itself varies across the sizes and development of countries. These findings have led to a proposal of the existence of a turnaround point – a threshold that is most appropriate for the concentration of the city. While Frick & Rodríguez-Pose (2018) acknowledge this point, they also state that the developed high-income countries tend to experience positive growth correlation with increasing concentration levels, while this may not be true to developing countries. For European countries the positive effect on growth of urban concentration is shown my Photo-Bala (2009). These studies contradict Henderson (2003), where conclusions state that the concentration is reversely correlated with the size and development of the country: small and underdeveloped countries tend to gain most from physical expansion and population growth, which will not necessarily hold true for city structures of greater size.

### Measures of city performance

As noted in Stiglitz & Fitoussi (2009): "It has become a commonplace to say that it is important to monitor and assess performance, whether of firms or countries or individuals. This is especially the case as our societies have become more performance-oriented." Nevertheless, assessing performance of a complex structure is always a challenging task and there is no agreed upon benchmark for cities that is accepted by the global community. Due to these reasons and the fact that cities are at the spotlight of modern economic development, came an era of performance measures and attempts to compile a unified measurement. Private and public organizations have proposed their ways of measuring performance and competitiveness of cities: Urban Development Indicators developed by the World Bank, European Commision as a project SUD-LAB EC, EU-TISSUE programme, European Common Indicators (ECI), by research institute Ambiente Italia, the Smart City index, and many others. Nevertheless, there is still no one unified measure (Mavrič & Bobek, 2015). GDP is used and financed heavily as one common and recognized benchmark, but active discussions are taking place if it is still relevant and able to fully represent the performance of cities.

Performance measures for cities are often comprised of several topics: quality of life and environment (health, air quality, inhabitant commuting habits, satisfaction with the environment, green areas, service accessibility, etc.), sustainability in development (resource allocation, green infrastructure, ctizen's participation (local involvement), etc.), socioeconomic factors (population, economic output, wage and job growth, education, development of industries, etc.) (Mavrič & Bobek, 2015, Dobbs et al., 2011). Netherlands cities performance compiled by Bureau Louter is constructed of 41 variables which include: floor space, employment, employment dispersion in different sectors, value added and development of labor productivity (Pellenbarg & Groote, 2016). In any of existing cases the measurements are compiled together while physical determinants and economic determinants are often included in the compiling of performance measure sets. This research will be focusing of separating physical and economic measures and looking at the ability to predict economic performance from the measures of physical structure.

**General economic measures (GDP, GVA, GMP)**

Main measurement of economic performance is still defined as pre-constructed indicators such as GDP per capita as used in Dobbs et al. (2011). This measurement It is strongly influenced by business investment and human capital (Ozyurt & Dees, 2015). Alternative measures of GVA (Gross Value added) or GMP (Gross Metropolitan Product) are used as well depending on the data availability and scale. Gross Metropolitan Product (GMP) is used as a measurement by Strano & Sood (2016), but it can be hard to obtain a unified valuation in an intra-country environment of the EU. GVA is a good measure of the economic output of a geographical region, as proposed by Dunnel (2009). In contrast, it is not advised to measure GVA per head, which divides output of those working in a region by all local residents. It does not represent neither the income of the residents, nor the actual value added per working inhabitant. Dunnel (2009) also suggests that "GVA per hour worked and GVA per job as productivity promotes the use of productivity, income and labour market indicators to give a more complete picture of regional and subregional economic performance".

### Employment and Employment Composition

Other measures that show economic stability of a city are employment spatial homogeneity, income equality, age-dependency ratios. Employment and employment growth are metrics adapted in city-performance studies (Porter, 2003). Individuals with high-tech skill sets tend to receive higher wages (Moretti, 2003), which are directly related to economic growth. Regional economic performance is strongly influenced by the clusters which appear to shape wages in industries. Employment composition is used in the studies that examine competitiveness and economic performance at an intra-national level (Porter, 2003, Pain et al., 2015, Bessis, 2016). General unemployment level measures are doubtful, since high unemployment can suggest a highly attractive point, where people migrate to use the possibilities of the job market (Yap, 1977).

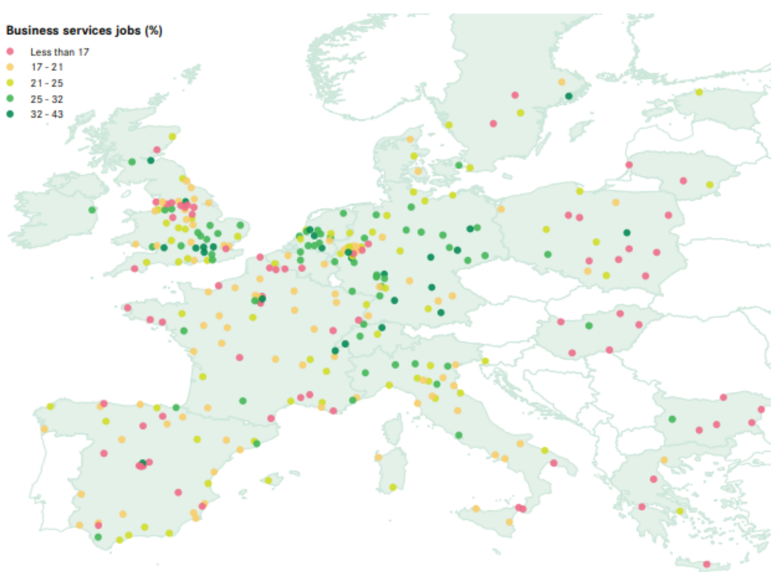



*Figure 1*. Cities by share of business services jobs, 2011. Eurostat data, constructed by Bessis (2016)

**Innovation**

Innovation is one of the main indicators that push long-term economic growth. The impact of innovation on economic performance is difficult to measure, but research estimates that innovation contributes to about a third of economic growth (Cameron, 1996). Patents and R&D investments are considered to be a representation of human capital and innovation. This figure is used often in researches concerning geographical distribution of human capital (Porter, 2003). Analyzing cities and their economic performance through numbers of patents is highly common and is found to be a significant factor affecting performance (Porter, 2003, Ozyurt & Dees, 2015, Bessis, 2016).

Nevertheless, patent registration might have a distorted view on the R&D investment and overall representation of human capital volumes in the region or city. Since patents are non-physical items, it is hard to recognize the actual geographical location of the R&D development for a specific patent and easy to change the registration country. As shown in Alstadsæter et al. (2018), a business can be tempted to develop products and maintain research and development activities in cheaper location with lower wages and then register the patent in countries which provide large financial benefits to patent owners through corporate taxation deductions.

**Education**

Education is a more direct and easily available representation of human capital. Photo-Bala (2009) examines appropriate variables used in other studies that compare city structure and economic performance and he finds that besides urban concentration, and education of the labor force (which represents human capital), capital and output growth rates are also good representations. In the spirit of defining the "smart city" highlight the importance of Information and Communication Technologies (ICTs) in the last 20 years has been shown to reflect the competitiveness of a city by Caragliu et al. (2011). In addition, it is important to note that education can strongly predict productivity of labor measures, which are direct representatives of economic performance (Van Winden et al., 2007).

Education is a highly common measurement of economic success for a nation, region or a city as observed in other studies (Pain et al., 2015). A proper incorporation of several factors can result in even more promising numbers: agglomeration and high levels of education tend to give higher economic growth. For example, Rosenthal & Strange (2008) have found that college-educated workers increase lower-educated workers' wages and the overall level of the economy.

**Alternative economic performance measures**

Some usual general economic sustainability measurements will not be considered. For example: poverty levels in cities are found to appear due to people migrating towards better opportunities (Yap, 1977, Glaeser, Kahn & Rappaport, 2000) and are better representations of economic performance and stability at a national level.

There are a few alternative approaches of how to recognize cities economic performance and competitiveness. The size of the population directly shows the demand for being able to use the functions of a particular city. Nevertheless, since the supply of city land is not elastic, the number of people might not fully represent the demand. City economic success can be measured in second best approach through increased prices for that city amenities: land, properties and services. This relationship can be recognized by land or residential property values and rents for living areas (Song & Knaap, 2003; Matthews & Turnbull, 2007). Discounted cash flow valuation is performed for commercial real estate by summing the expected future returns of the future cash flows and expected land value increases (Wendt, 1957). These expectations capitalize into current commercial land values and property rents as well. Nevertheless, the land price data is difficult to obtain for large data sets as found in the Eurostat data base and will not be the topic of this research.

### Physical Measures

Physical measures of cities are diverse and various researches approach this topic from different perspectives. These perspectives include ratios and more complex analyses of geometrical characteristics and morphological components. These approaches include the density measurements, street network analysis. The overview of the possible methods will be given in this section and should provide a good perspective on the complexity of city structures, textures and the implications of the physical structure influence on the behavior and choices of local residents and business units.

Variables that will be discussed onward will mostly look at the inner-city structure, but the positioning of the city is extremely important as well. Taking variables that compensate for pre-existing conditions are important and can influence the later changes. These variables are suggested as following: tier of the city (Capital city dummy variable), poly-centric or mono-centric character of the city network or the area that the city is located in (Cardoso & Meijers, 2016), temperature of the cities (Glaeser & Saiz, 2004), precipitation (Aksoy et al., 2016).

### Morphology

"Morphological analysis, which refers to the geometric characteristics of urban sites, illustrates its usefulness in determining the analogies between patterns of cities and their "physical" characters providing indicators of the aspect of settlement form and structure" (Colaninno et al., 2009). Urban morphology "implies 'form,' 'land use,' and 'density,' and has connotations with the shape, structure, pattern and organization of land use, and the system of relation between them" (Donnay, Barnsley, and Longley, 2001 as cited in Colaninno et al., 2009).

Road length and intersection number are closely related measures, although their pattern of fractals can differ, while population density, employment density and building gross area can explain each other since they correlate at a high degree (Peiravian, de Lapparent & Derrible, 2015).

This is also a result of the increased availability of geographical data in smaller scales and the efficiency of today’s tools, which has let the researchers to dive into smaller structures and make more complex analyses as found in the studies of Aksoy et al., 2016 (Europe) and Nasri & Zhang, 2018 (USA) which take as many city-defining parameters as available and use machine learning algorithms to group cities by their defining characteristics.

**Compactness and sprawl**

Compactness is a widely used term to describe concentration. *Compact City Policies: A Comparative Assessment* published by OECD (2012b) describes this property as a positive characteristic for a city, which enables local inhabitants to efficiently invest in infrastructure, lessens automobile dependency by shortening intra-urban distances. Compact cities have dense development and transport system patterns. “They play a part in the economy by giving residents easier access to services, jobs, and social networking.” The report values cities economic performance in relation to other cities and their features of compactness: density, proximity, public transportation systems, job and service supply and accessibility. Most common measures to OECD (2012b) suggests variables to measure compactness of a city, that include population and urban land growth, population density measured as an average density over a 24-hour period, urban land cover, average trip distances (as a measure of proximity to various functions), means of transport used, proximity to public transport nodes (measuring by the ratio of inhabitants within a certain radius of public transport stops), share of trips using public transport.

The urban primacy measure is sometimes used to calculate the concentration of the city: it measures the share of urban population living in the central area. For a country-level analysis, this variable is measured between the total population of the country and the population living in the biggest city. The variable is widely used and has extensive historical data available. It is very simple and data is easily available. It is also correlated to other country agglomeration measures (Henderson, 2003). Bertinelli and Strobl (2003) argues that this measure is not the best as it differs widely among the size of the country – the bigger the country the higher probability of having multiple urban centers, which decreases the total share of population in one biggest urban unit (Photo-Bala, 2009).

Sprawl does not have an agreed-upon definition, but is often defined by four land use characteristics: "low density; scattered development (i.e. decentralised sprawl); commercial strip development; and, leapfrog development" (Ewing, 1997 as cited in Tsai, 2005). Tsai (2005) develops four quantitative measurements for dimensions of metropolitan regions: "metropolitan size, activity intensity, the degree that activities are evenly distributed, and the extent that high-density sub-areas are clustered" and uses Moran coefficient to distinguish compactness from sprawl. Moran coefficient approaches zero for very compact monocentric regions, while the sprawl characteristics increase the values of this coefficient.

While Jaeger et al. (2010) uses different vocabulary, the study provides almost identical measurements to identify the dimensions in identifying sprawling areas:

1. the expansion of urban areas;
2. the scattering of settlement areas, that is how densely clumped or widely dispersed the buildings and patches of built-up areas are within the landscape (area-intensive growth);
3. low-density development (i.e. high land uptake per person).

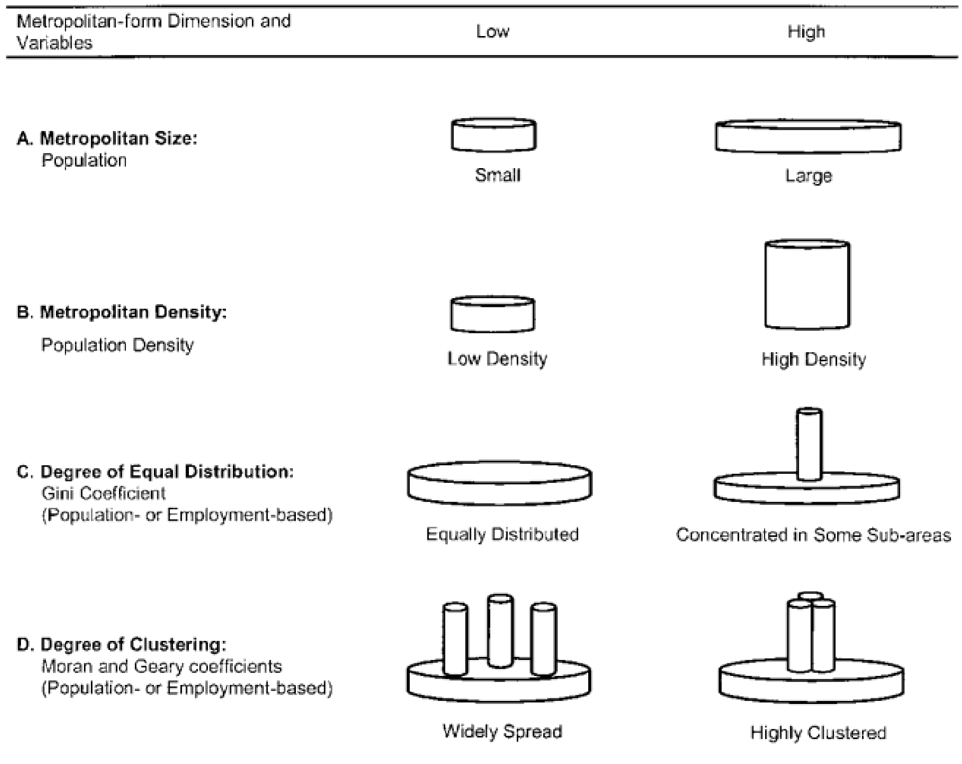



*Figure 2. Four dimensions of metropolitan form (Tsai, 2005)*

**Connectivity and transportation**

Transportation are important in 2 scales: in a network of cities and within the city itself. These are 2 different scales provide different functions. Cities connectivity with the rest of the world or other cities (which is usually measured by land, water and air transportation links) gives the ability for a metropolitan region to borrow labor and human capital from surrounding areas which positively enhances metropolitan functions and competitiveness of a region, but most significant is the local size of the urban unit. The optimal size of local labor and human capital pool and network connectivity differs across metropolitan functions and cities (Van Winden et al., 2007, Meijers et al., 2015).

The position in the network is also found to be important: Cardoso & Meijers (2016) find that the first tier cities (mostly capitals) have a significant surplus (Paris), while the second tier cities are better off when a dominant first-tier city does not exist in the network (German polycentric structure). Nonetheless, there is no linear relationship found between cities economic performance and connectivity (Pain et al., 2015) which is not surprising since a single measure cannot represent the whole complexity of the city structure and competitiveness.

Interior transportation systems and road networks determine the daily transition of the labor force within the city and the comfort and types of transport used by the local inhabitants. These physical structures have geometric properties, which can be quantified and measured. Peiraviana & Derriblea (2015) have managed to observe geometric characteristics of cities in the USA that could predict the location (Eastern versus Westers or coastal versus inland) or car driven versus pedestrian friendly cities. The last feature also correlates with the cities age – in the USA the younger cities tend to have bigger blocks and be more oriented towards an automobile society. Nevertheless, the amount of roads is not the sole important factor.

Transportation structure in the city has been found to correlate with socio-economic factors. Higher concentration of people and walking time per capita is observed in smaller scale areas, where intersections are denser and which is also considered as a more pedestrian-friendly environment (Peiravian & Derriblea, 2015). Walkability also correlates with higher property values (Pivo & Fisher, 2009 & 2011). The configuration of the streets of the city is the generator of the movement within the city (Kostakos, 2010). Although small scale walkable cities add to the attractiveness, large urban vessels tend to divide the overall area into segregated plots (Nasri & Zhang, 2018). Walkscore is a new derivative of real-time interactive data that is constructed and adapted in the USA research and business applications as an additive. Nazri & Zhang (2018) incorporates this variable in their study and the possibility to include this particular variable in our study was explored even prior to their input. Nevertheless, the index is calculated by a business unit and is not yet available in Europe.

Space syntax is an emerging field of analysis that concentrates on the notion of connectedness and quantifying the layout of the streets. This field has shown that street network can be characterized as having bad or good syntax because it can cause social segregation and antisocial behavior (Kostakos, 2010). Nevertheless, the data sets significant in volume, constructs of these indicators are under development. The processes use a lot of time and computational power and there is no pre-made data set for street syntax indicators for the wide data set European cities yet. Therefore, unfortunately, this method of analysis will have to be postponed to further research in cities physical performance.

### Form and geometry

Previously mentioned influences and building policies shape the geometry of the cities. Differences in urban geometry translate into differences among residential property values (Song & Knaap, 2003). Matthews & Turnbull (2007) find that this difference is only significant in pedestrian oriented structures, where automobile-oriented neighborhoods see no effect. As Pivo & Fisher (2009) have overviewed:

> "Contributing attributes to a walkable neighbourhood include urban density, land use mixing, street connectivity (i.e. the directness of links and the density of connections), traffic volume, distance to destinations, sidewalk width and continuity, city block size, topographic slope, perceived safety, and aesthetics (Frank and Pivo 1994, Hoehner et al. 2005, Cao et al. 2006, Lee and Moudon 2006, Parks and Schofer 2006)."

The cities are usually comprised of many different types of neighborhoods, but given the recent Geographical Information System tools it is possible to quantify forms and generalize mean geometrical caharacteristics (Peiravian & Derriblea, 2015).

### City Classification Attempts

Each city has a unique typology, but city classification is recognized as an important possible tool in the political economy, that could help make key decisions in setting city development priorities. There have been several attempts to put such complex systems as cities into groups based on various factors with their numbers increasing in later studies.

One of the first researchers to analyze smaller than National scale in the USA was Porter (2003), who has classified the USA Economic Areas (and equivalent to European Functional Urban Areas) into 3 categories: traded, local and resource-oriented. The process of classification was based on the employment between industries and geographical distribution.

Van Winden et al. (2007) points out that the knowledge-based industries will have different effects on different cities. An appropriate concentration and mix of knowledge-based education and business functions and their links correlate with lower unemployment levels and higher property prices, nevertheless one without the other is much less effective. They propose six types of cities: stars, metropoles in transition, knowledge pearls, star nicheplayers, nicheplayers in transition, intellectuals. The classification is based on the composition of existing industries, education base and the scale and connectivity of the city.

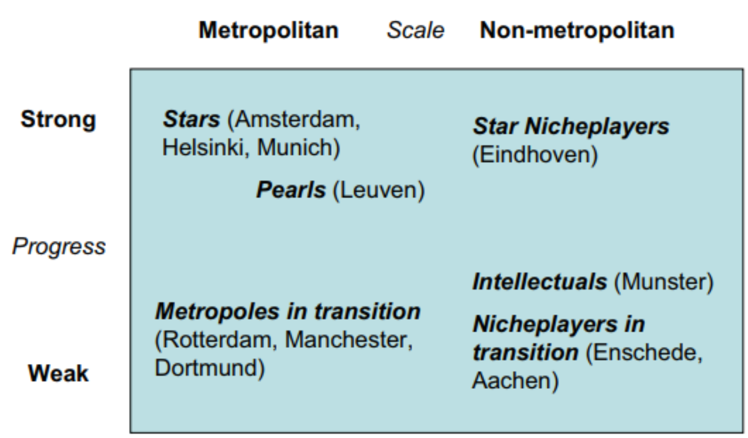



**Figure 3.** Typology of cities.

*Figure 3.*(Van Winden et al., 2007)

As mentioned in the beginning of this research, increased availability of geographical data in smaller scales and the efficiency of today's tools, have let the researchers to dive into smaller structures and make more complex analyses as found in the studies of Aksoy et al. (2016) (and Nasri & Zhang (2018) for European and USA cities respectively. They take as many city-defining parameters as available and use machine learning clustering algorithms to group cities by their defining characteristics and analyze them in and among separate groups.

Aksoy et al. (2016) recently have attempted to classify 385 European cities in a working paper. The classes are based on the evaluation of 59 indicators, which are derived from various recent data sources. This classification process is done by K-means methodology creating 10 artificial groups. This early clustering methodology can predict the determinants for the differences of cities. They combine the dataset from environmental, governance, structural and socio-economic factors and their changes. The variable set is large and highly versatile. Eight out of ten clusters are defined by unemployment levels, size of administrative areas, share of green urban areas, tourism activities, urban sprawl indicators ("Urban permeation", "built-up area" and "dispersion"), economic significance, proportion of national population, World Governance Index, old-age dependency with different combinations. Paris and London are defined as 2 separate clusters.

The most recent study of Nasri & Zhang (2018) presented slightly ahead of the research you are reading looks at clustering from another perspective and gives a fresh insight on physical structure parameters. They extract the most defining characteristics and cluster cities into 3 groups:

A. Compact, well mixed and highly accessible cities
B. Cities defined by moderate density, average accessibility and random clustering patterns

C. Sprawled, low density suburban settings.

They have managed to identify the most influential physical structure parameters to the clustering algorithm and after dividing the dataset of 50 biggest cities into 3 groups, and as Aksoy et al. (2016) analyze the results with descriptive manner.

**Final remarks**

As we can see from the literature review, the city form cannot be identified easily with one figure, since different relationships between city pattern additives result in different outcomes. For example: many roads can lead to a very small – gradient, walkable, human-scale friendly neighborhood, similar to the Old Town areas of cities. But if those roads are 4 lanes wide, they will not act as a pedestrian friendly instance and will most probably suggest an outskirt-type pattern. Population living in an urban settlement of first type of mentioned character will most probably be closely involved in common activities and the agglomeration effect, human capital sharing will have lots of potential area. The second character of the mentioned urban pattern, will most probably show segregation, long commuting hours – economic and time resources put into relocating daily from point A to point B.

Considering all this, it is no wonder that quantifying measures of cities and connecting them with the economic performance factors literature varies greatly in the topics, variables and approaches. Urban economic research is still an area of touch-and-go, but it is changing rapidly and evolving with ever so increasing amount of information and possibility to make complex analyses.

## Research methodology

The research consists of two major parts. First part involves gathering and clustering physical data of European cities and the second part consists of gathering and evaluating the economic performance differences between clusters of cities. This research will use K-Means algorithm and follow the most recent research papers that were done for city analyses, mentioned in the previous section.

> Cluster analysis has a wide range of applications in many research fields such as marketing, insurance, biology and psychiatry (…). In land-use planning and policy-making, cluster analysis can be very useful to identify areas with similar land-use pattern to propose, implement and evaluate land-use policies more efficiently (Smith & Saito, 2001, as cited in Nasri & Zhang, 2018).

There are several approaches to cluster testing and the algorithms can be identified as partitioning, hierarchical, density-based, grid-based or model-based methods. K-means algorithm is a form of partitioning approach. The closest method is k-median method. Both of the processes work in similar manner by clustering data instances into mean or median bundles. The logical steps of the algorithm will be described in further sections. K-means and k-median algorithms require the researcher to set the initial number of clusters that the data set will be grouped to. Hierarchical algorithms, on the other hand, can set the number of clusters on their own merits and the cluster-setting is optional. For more detailed comparison of partitioning and hierarchical clustering methods one can refer to Nasri & Zhang (2018).

### Data gathering

The data gathering and selection process was a challenging task. It involved several applications (ArcGIS, QGIS, excel, R) and substantial amounts of time and attention to prepare several variables. The data is gathered from sources that are not perfectly coherent in time, but it is important to understand that the city structure does not react to changing conditions in a very speedy way, especially the overall ratios of land structure. The processes of urbanization and land-uptake are slow due to regulatory environments and time taken to essentially change the city structure. A usual urban development change of course consists of several steps: public policy setting, urban assessment, urban planning proposals, public discussion, approval processes, urban design project proposals, approval processes, financing, building and paving large engineering structures, architectural and building engineering design, financing, building. Each of the steps might take a year or several. For a relatively developed city to essentially change its course, it needs time counted in decades. Therefor the differences in data timing will not be considered as essential.

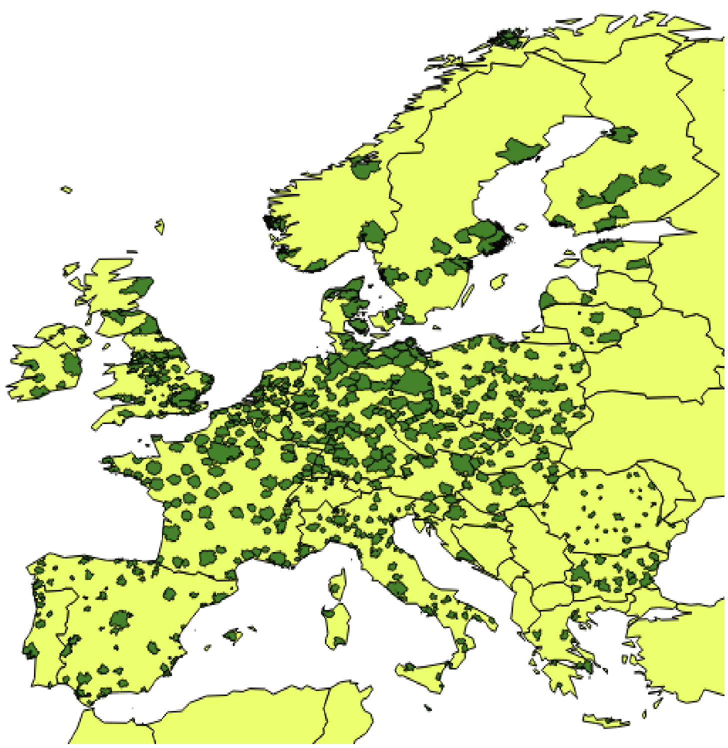


*Figure 4.* European country shapefile extract with and overlay of Functional urban areas layer

The geographical data has been obtained from the European Commission data portal Eurostat, the European Environmental Agency (EEA - an agency of the European Commission) and the Copernicus project in the form of data tables, GEOrasters and shapefiles – geospatial vector format. Urban Audit (URAU) 2011-2014 shapefile was used as the basis of determining the physical shapes and characters of cities and their functional areas as provided by the Eurostat data portal and approved by the European Commission. It is important to note that the data span 2011-2014 marks the preparation period and does not represent the time dimension, since this detail data set is highly complex and therefore is a sophisticated estimate of the distribution of cities and their functions at a point in time. This data set contains around more than 800 cities and more than 600 Functional Urban Areas. Cities of 50 000 inhabitant or more are market in the data set. Data manipulations and calculations have been preceded in QGIS – an open source Geographical Information System program, MS Excel and R. Economic variables have been obtained from the Centre for Cities – an independent urban policy research unit and a charity registered in England. The main aim of this organization is researching cities and their economic performance. The collection of data includes more than 300 cities in Europe. The data is based on the year 2011 and has been checked to have matching names with the URAU city naming, nothing else in the data set has been manipulated. Calculations and later implementation into research was conducted using MS Excel and R.

**Physical measures**

Physical measures included in this research are based on previous studies in the attempt to recognize the most characteristic measurements of city structure that influence the economic choices of people and companies associated with that geographical location. Most of the variables were based on the working paper of Aksoy et al. (2016). Some additional variables were added according to other related researches, discussed earlier in this paper. Nasri & Zhang (2018) have proposed a fresh approach on the topic, but their data frame is constructed from variables that are only available in the USA (Walkscore) or have a similar overview. An adapted measure of the urban primacy measure will be used as a ratio of the central area and the functional urban area size and population.

**Density and sprawl**

As mentioned previously in this research, there are no agreed-upon sprawl quantitative measures (EEA, 2016) and often this poses a challenge. Many of the early attempts to evaluate the spread of people included visual representation of city layout or ill-informing means and ratios. Nevertheless, recent studies have gained the advantage of using powerful computational tools and detailed data sets. Still, studies in this field use different calculations for density and sprawl measurements. Nasri et al. (2018) proposes Global Moran's I and high/low clustering (Getis Org General-G) for the analysis of the surface parameters. Weighted urban proliferation and its additives are used by Aksoy et. Al (2016). Following this research, we will adapt the Jaeger et al. (2010) proposal of sprawl measurement as it is a highly sophisticated complex variable that manages to construct a single number for the completely different features: density, concentration and layout that have been discussed in the research methodology part on compactness and sprawl. The research and adapted definition is based on the concept of sprawl described by Jaeger et al. (2010), later improved by Jaeger & Schwick (2014) and detailed for the European continent in the EEA Urban Sprawl report (2016):

> Urban sprawl is a phenomenon that can be visually perceived in the landscape. A landscape [is affected by urban sprawl] if it is permeated by urban development or solitary buildings and when land uptake per inhabitant or job is high. The more area built over in a given landscape (amount of built-up area) and the more dispersed this built-up area in the landscape (spatial configuration), and the higher the uptake of built-up area per inhabitant or job (lower utilization intensity in the built-up area), the higher the degree of urban sprawl. The term 'urban sprawl' can be used to describe both a state (the degree of sprawl in a landscape) as well as a process (increasing sprawl in a landscape).
> The causes and consequences of urban sprawl are distinguished from the phenomenon of urban sprawl itself, and therefore are not a part of this definition.

In this research, adaptation of EEA report 11/2016 grid shapefile dataset is repurposed for the Urban Audit spatial units of 2011-2014: City areas and/or Functional Urban Areas. The authors argue that there have been many attempts to derive an index or a set of indexes to quantify sprawl and propose a *WUP* (Weighted Urban Proliferation) sprawl measurement that is comprised of three other measures and two functions.

*PBA* – percentage of built-up area, corresponds to the relationship between built-up area in the measured geographical unit and the total area of the unit. *PBA* can be directly compared to any other area or geographical unit, since it does not depend on the size of the geographical instance. It is the most common dimension of land-uptake and the ratio is a fairly straightforward concept. Nevertheless, the HRL mapping in small scales and ability to choose the density of imperviousness in a shapefiles, from which the ratio is extracted provides us with flexibility and possibilities for experimenting. That being said, this research will not be testing the impact of different imperviousness levels on PBA and will be following earlier works.

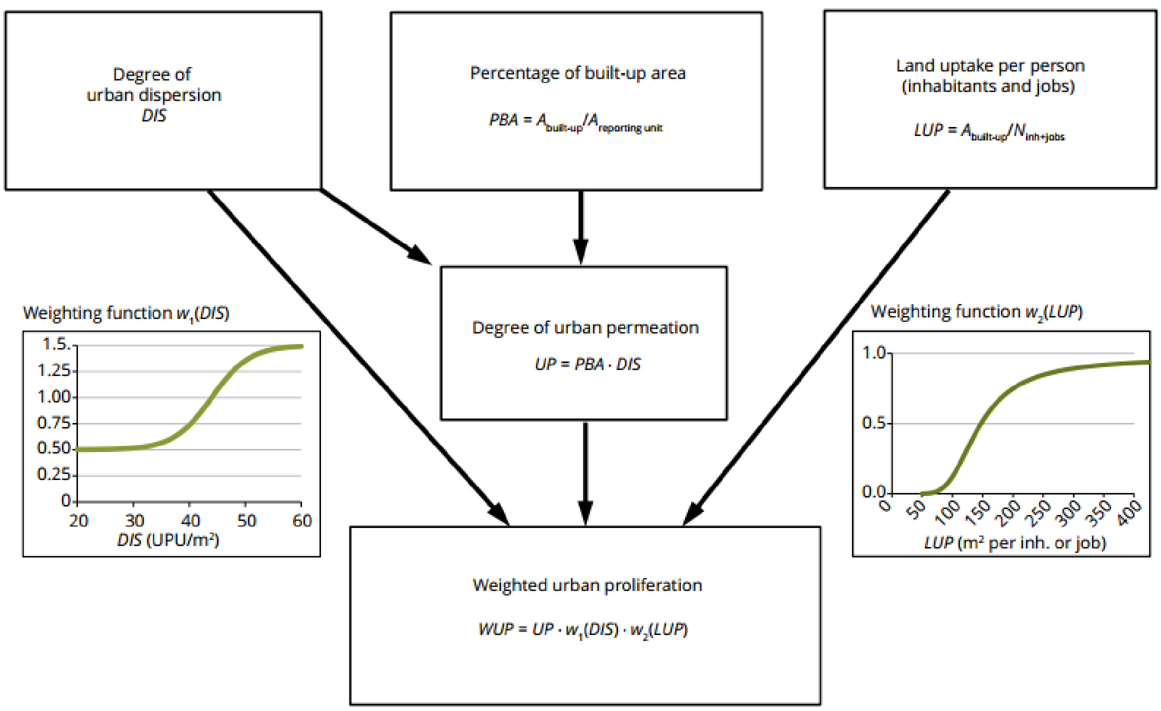



*Figure 5*. Weighted Urban Proliferation (WUP) composition showing the relationship with composing measurements: Dispersion (DIS), Percentage of built-up area (PBA) and LUP (EEA, 2016, p. 41)

*DIS* – dispersion - is a metric of settlement geometry characteristic, which is calculated from the distances between any two points withing the built-up area up to a maximum distance defined as *HP* - horizon of perception, which is used as 2km in this dataset and proposed by the study of EEA (2016) and introduces the reasoning of this choice further in the text. The longer distance between two points contributes to a higher *DIS* value. *DIS* is calculated as *UPU/(area of buil-up surface in square meters)*.

*Equation 1* Degree of urban dispersion calculation as provided by Jaeger et al. (2010), on which the automated calculation algorithms are based and used in this research

$$\text{Degree of urban dispersion} = DIS = \frac{1}{A_{\text{urban}}} \int_{\vec{x} \in \text{urban areas}} \frac{1}{\int_{\vec{y} \in \text{urban areas and } |\vec{x}-\vec{y}| < HP} d\vec{y}} \int_{\vec{y} \in \text{urban areas and } |\vec{x}-\vec{y}| < HP} f(|\vec{x}-\vec{y}|)\, d\vec{y}\, d\vec{x},$$

Then the metrics are transformed by a weighted function to have more distinct observation of sprawl size, where $w_1$ *(DIS)* = 1 is a mean dispersion, which is set at the 1960's Swiss average of 43.986 $UPU/m^2$ (*UPU* - Urban Permeation Unit). $w_1(DIS)$ ranges from 0.5 to 1.5, where a higher value indicates higher dispersion. (Jaeger and Schwick, 2014 as cited in EEA anexes, 2016).

*Equation 2.* Weighing of dispersion (DIS) for further Weighted Urban Proliferation calculations (EEA, 2016, annexes):

$$w_1(DIS) = 0.5 + \frac{e^{\left(0.294432 \frac{m^2}{UPU} \times DIS - 12.955\right)}}{1 + e^{\left(0.294432 \frac{m^2}{UPU} \times DIS - 12.955\right)}}$$

*UPU* – Urban Permeation Unit is not defined as a separate variable, but it is a measurement unit derived from mathematical approaches. It is extracted from a calculation that is based on two-dimensional data from the map GIS raster files. This calculation is an additive in the calculations of *DIS* and *WUP* that will be explained in greater detail further on. *UPU* units are calculated for grid squares (which are the size of 100m x 100m and are at least 30%

built-up – has 30% or more imperviousness). 30% is chosen as a point, recognized by Orlitová et al. (2012) to be the best measure used for greatest *DIS* and *WUP* value ranges: this measure gives the biggest possible differences. Impervious surfaces, are surfaces built with buildings, sidewalks, roads or any other cover that prevents the surface from developing naturally. *UPU* was calculated using a cross-boundary connections procedure (CBC), which is considered to be a more appropriate and precise way than the alternative cutting-out procedure. A cutting-out procedure takes into account only the impervious squares that are within the calculated unit and does not consider any built-up surfaces outside of the City or functional urban area. This simplifies the data set needs, but does not represent the full interaction between the closest surrounding areas. CBC gives the possibility to be more precise in the evaluation process of real urban land interactions and positioning. An example of CBC and cutting-out procedure calculations can be better understood in the following figure:

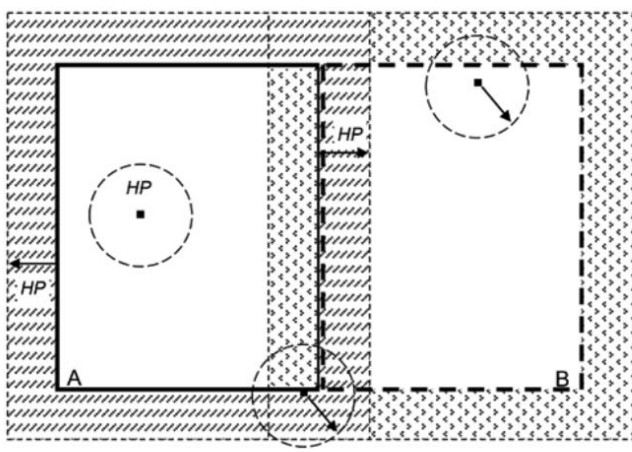



*Figure 6.* Areas A and B both contain very small urban patches. HP is the horizon of perception radius. In area A the UPU calculations are identical for cutting-out procedures and CBC, but in area B the distance from the urban point to the edge of the reporting is smaller than the horizon of perception. In this case anything outside the boundary will be not considered in in the calculations with cutting-out procedure implementation, even if the exterior patches are built-up. CBC takes into account anything within the radius of HP (modified after Jaeger et al., 2010, as cited in EEA 2016).

Urban sprawl can be measured in different scales. *HP* – horizon of perception – is the component of this function that defines the scale of the measurement. In theory, any HP value can be correct to some degree and a good choice can vary somewhere between 1 km and 5 km (EEA, 2016):

1. Visual perception. A medium-height person is able to see up to 5 km into distance because of the curvature of the earth.
2. Values lower than 1 km are too small, because this scale does not represent the real relationships of the cities. For example, it is perfectly normal that two urban areas share same functions: schools, medical centers, etc., even if they are more than 1 km apart.

*LUP* – Land Uptake per Person - describes the intensity of the use of the built-up area. It assumes that the land is used more efficiently, if more people live and work in a smaller built-up area. The reciprocal of *LUP* is *UD* – utilization density (*1/LUP*). *LUP* value varies between 0 and 1, where 0 indicates dense areas and 1 indicates sprawl (EEA, 2016).

> If the *LUP* is higher than 250 $m^2$/inhabitant or job, the $w_2(LUP)$ is close to 1. If it is less than 100 $m^2$/ inhabitant or job (e.g. in city centre areas), the $w_2(LUP)$ is close to 0 because such areas are not considered to be sprawled. Accordingly, if the *UD* is less than 4 000 inhabitants and jobs per km2, the weighting factor is close to 1, and if it is more than 10 000 inhabitants and jobs per $km^2$, the weighting factor is nearly 0. A value of 4 500 inhabitants and jobs per $km^2$ corresponds to the limit of 400 $m^2$ of urban area per inhabitant (without taking jobs into consideration) suggested by the Swiss Federal Council in 2002 as a maximum acceptable average value (Swiss Federal Council, Schweizerischer Bundesrat, 2008, p. 27 as cited in EEA 2016).

For calculations without jobs using only inhabitant numbers, we can recalculate the weighing function for the

*Equation 3.* Weighting function for Land Uptake per Person or job (LUP) for further Weighted Urban Proliferation calculation (EEA, 2016, annexes):

$$w_2(LUP) = \frac{e^{\frac{4.159-613.125\frac{inh.+jobs}{m^2}}{LUP}}}{1+e^{\frac{4.159-613.125\frac{inh.+jobs}{m^2}}{LUP}}}$$

*UP* – Urban Permeation - is a product of *PBA* and *DIS*. It describes to what degree the landscape is permeated by built-up patches. *UP* is a metric that can be compared directly between geometries since it is an intensive value that does not depend on the size of the landscape (Jaeger et al., 2010, as cited in EEA 2016).

*WUP* – Weighted Urban Proliferation - is a product of *UP*, $w_1(DIS)$ and $w_2(LUP)$ which indicates higher sprawl values with a higher *WUP* since it is a product of three measurements that represent sprawl characteristics in 3 different dimensions: part of used-up land (*PBA* and *UP*), the spread of built-up land (DIS) and the intensity of it's use (*LUP*).

**K-means Clustering**

The K-means clustering algorithm is an "unsupervised" learning (machine learning) algorithm that assigns data points to a specific number of clusters and, as mentioned, is observed in recent research applications that study cities (Aksoy et al., 2016, Nasri & Zhang, 2018).

The algorithm starts with an initial number of clusters given by the researcher (in our research it is 8). The clusters are placed randomly by the algorithm and assigns all the data points to one of the clusters. After – a procedure of reassigning data points and recalculating cluster centers takes place. The movement of cluster centers is called iterations. These steps are repeated until an optimal layout emerges and usually up to 10 iterations are needed to get an optimal data point clustering. The clustering is called optimal, when there is no within-clusters sum of squares (WSS) reduction while moving a point between any clusters – the sum of squares of distances from points to the assigned cluster centers is minimized (Hartigan & Wong, 1979). No minimum number of points in any cluster is set.

### Data Standardization

As in Nasri & Zhang (2018), the K-means algorithm uses Euclidean distances to calculate the coordinates in the data space. With a number of variables *n*, an n-space Euclidean distance is used in the K-means algorithm and the distance between two points is described with a Pythagorean theorem.

*Equation 4* The Euclidean distance calculation mathematical expression, based on the Pythagorean theorem:

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-p_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

The K-means algorithm does not identify between different scales of the variables, therefore the variables have to be transformed for an efficient clustering. If this is not performed, the algorithm and the clusters will be based on the variables with significantly larger values. For example: if one variable represents population (which will have five to seven digits in natural number values) and another represents employment rate (in percentage or ratio), then there is a tendency to see the clusters being more influenced by the population numbers rather than the employment rate. In addition, the algorithm will be more affected by the variables that contain values of greatest ranges. The values need to be as "dimensionless" as possible.

There is no universal standardization method and a number of alternative standardization transformations are applicable. For example: scaling can be used to make the overall estimations converge to a similar range. In this process all the values are divided by a specific identified number which depends on the selection of the researcher. Decimal scaling is very common and simple: is this method the decimal point is moved. This generally means that each variable is multiplied by $10^n$, where $n$ is separately and manually defined for each variable and can be any integer ($n \in \mathbf{Z}$). A more advanced normalization process is a normalization process. This transformation uses a linear function to reprocess the data into a specific range (Mohamad & Usman, 2013).

A common normalization range is 0 to 1, where each element is divided by its minimum variable value and then divided by the row range. This transformation does not distort the existing variance and generally improves the performance of clustering mechanisms (Mohamad & Usman, 2013).

This research will be using Z-score standardization. This transformation is useful where the actual minimum or maximum is not defined or unknown. It has been found that z-score standardization is more effective for K-means clustering preprocessing of data than min-max and decimal transformations. The process is performed with centering and scaling data with a function *center_scale* form an R package "ClusterR". When the data is scaled and centered, each value is transformed in the following order: a mean of the variable is subtracted from each value of a variable and divided by the standard deviation. This provides the data in the form of Z-scores with a mean of 0 and a standard deviation of 1 (Mohamad & Usman, 2013).

**Selecting the number of clusters**

A major question is the number of clusters that has to be set by the researcher since it is not automated by the method itself (Tibshirani, Walther & Hastie, 2001). There are more than

30 methods to decide on the number of clusters while general logics is not exempted (Kassambara, 2017). Many early researches that include K-means clustering method do not give formal reasoning for the selection of number of clusters and often the number is set as an outcome of a trial-and-error process with a visual identification of an acceptable number (Pham et al., 2005). In current days there are developed methods that make this selection process more transparent and justified. These methods are compiled in R packages and are ready-to-use. The methods used to determine the number of clusters in this study are:

- Explained variance or the *F-test* gives an estimation as percentage of total variance that is explained by the clustering. It is calculated by dividing the sum of the within-cluster-sum-of-squares-of-all-clusters (previously mentioned as WSS) by the total sum of squares (TSS) of all data points. Function criterion *variance_explained* in a function *Optimal_Clusters_KMeans* will be set for this performance measure estimation.
- WCSS - the sum of the within-cluster-sum-of-squares-of-all-clusters. This particular function criterion is programmed as *WCSSE* in the function *Optimal_Clusters_KMeans* and provides a suggested estimate.
- Dissimilarity: the average intra-cluster-dissimilarity of all clusters. A function *Optimal_Clusters_KMeans* has this criterion incorporated as *dissimilarity*.
- Average Silhouette method. This method measures the quality of the clustering by measuring the average distance between clusters. For each point $i$ in a cluster $C$ an average dissimilarity distance $a_i$ is calculated. Then the average dissimilarity of point $i$ for all the elements in clusters $C$ that do not contain $i$ are calculated s $d(i,C)$. The minimum observation $b_i = min_C\, d(C,i)$ represents the dissimilarity between the closes neighboring cluster and point $i$. The silhouette width of instance $i$ is defined as $S_i = (a_i – b_i) / max(a_i, b_i)$ and can take a value

between -1 and 1. The higher the value – the better the clustering. 0 indicates that the point is between clusters, while a value of -1 suggests that the point is in a wrong cluster (Kassambara, 2017). Function criterion *silhouette* in a function *Optimal_Clusters_KMeans* will be set for this performance measure estimation.

- *f(K) distortion* is a criterion based on Pham et al. (2005). It measures the irregularities of the data set by evaluating and weighing the sum of distortions of clusters. This method performs well with an unknown distribution, and therefore can be considered to be more proficient than the Akaike information criterion and the Bayesian information criterion which are both constructed to perform well with data sets with Gaussian distribution. Method by Pham et al. (2005) bases their prediction of clusters with an algorithm closely associated with K-means clusteringFunction criterion *distribution_fK* in a function *Optimal_Clusters_KMeans* will be set for this performance measure estimation.
- Adjusted_Rsquared : the adjusted R^2 statistic

An elbow method is a method of validation and interpretation on clustering analysis that is based on visual exploration and ratios. Plotting variance the previously mentioned indicators for a range of the possible number of clusters from 1 to *n*, where *n* is selected by the researcher. This method is estimated by a visual identification of an „elbow" – a point on the plot, where an increase in the number of clusters does not provide significantly better explained variance. This is due to the observation that adding an additional cluster when the number is low gives a high increase in explanation. But at a certain point the marginal gain of explained variance will fall with an additional cluster added. This is the elbow point. Nevertheless, often there is no clearly defined point and the fall of explanation perentage is gradual.

A major draw-back of the F-test, Elbow, WCSS or other previously discussed methods is that these tests depend highly on the judgement of the researchers performing this computation (Mohamad & Usman, 2013). With gradual increase in the performance, where an increase in the number of clusters does not provide the researchers with a clear „elbow" point, the decision power lies in the hands of the research team. In this case it is important to choose an appropriate number based on pure logics and implications (Aksoy et al., 2016). On the one hand, the number of clusters has to be small enough, so that the clusters can be explored and valuable information can be extracted, but big enough to give reasonable explanation value.

**Economic performance measures**

As discussed in the previous parts, economic performance can be estimated taking pre-constructed economic performance variables as GVA or assesed as an inderect outcome of socioeconomical parameters as level of education, employment and employment in high-tech industries. The data obtained for this research is constructed for 17 European countries where the analysed cities are located: Belgium, Bulgaria, Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Italy, Lithuania, Hungary, the Netherlands, Poland, Sweden, Switzerland and the United Kingdom (Bessis, 2016). The selection was made due to the data availability at a certain scale of the analysis. The data was constructed from Eurostat and Urban Atlas information with additional gap-filling from the national statistical bureaus. The data set is cross-sectional for the year 2011 with several instances taken from the year 2012. The economic data is constructed for medium-sized and big cities for every city that has more than 125 000. Soft economic data is harder to obtain for smaller cities and although the clustering will be performed for more than 750 cities, the economic evaluation can only be fully possible for the greater ones. The data is collected for the administrative bounds of cities and not FUAs.

The variables used in the comparison of the economic valuation iclude:

GVA – the GVA values are expressed in pound sterling, as this was the most direct provided measurement. Nonetheless, the currency will not be a problem in the overall comparison. GVA measures are pre-adjusted for purchasing power parities (PPP). This removes the pre-existing differences between countries in the levels of economic development and accounts better for the true value created by the local population. For example: if the same amount of work will be priced lower in Lithuania than in the UK, but it can provide the same amount of goods to be purchased. This would overestimate the amount of value created in the UK, but underestimate Lithuanian GVA. These monetary differences disappear in the PPP adjustment process. GVA per worker – is the value of GVA divided by the number of filled working spots.

Education and education distribution measures – these measures are taken as population portion with high, medium or low levels on education. The levels are determined by the International Standard Classification of Education (ISCED). Under the 1997 classification.

Labor composition measures can show the development of an area and the ability to generate value. The employment composition is defined by NACE classification that is available in the European statistical level. "The industrial structure of cities, which reflects the type of investment they have attracted over time, varies across Europe." (Bessis, 2016). It will be assessed with the following variables:

- Business services –working population portion in the areas of business-type, knowledge-based service jobs. Due to data availability, the variable will incorporate some other professions. The variable is supposed to represent the employment in areas that human capital provides higher additional value. The variable refers to jobs in financial and insurance activities, information and communication, real estate activities, professional, scientific and technical activities, administrative and support service activities.

- Mining, manufacturing and utilities – this indicator refers to the industrial activities of the cities. Data availability constrains us from a deeper dive into the distribution, but is a good overall marker. The term accounts for jobs in manufacturing, mining and quarrying, electricity, gas, water, steam and air conditioning supply, sewerage, waste management and remediation activities.
- Public services – employment proportion of the workers in public institutions.
- Other private services – employment in private service areas, not defined in other way by NACE.

Other measures will be assessed. For more reasoning and advantages of these particular measures one can refer to Bessis (2016). These measurements are: business stock, employment rate, patent applications to the EPO per 100,000 of population, total jobs, unemployment rate.

# Research

## Urban Audit. Cities and Functional Urban Areas

City areas and Functional urban areas are defined by the European Commission Eurostat/GISCO Urban Audit 2011-2014 GIS datasets, last updated in 2015 December. City areas represent the administrative areas of cities. These shapes are important, since these areas are administered directly by the municipalities and the local governance unit can actively interfere in the process of the development. On the other hand, Functional urban areas (FUAs) are constructed to better represent the extent of the city and activities that it influences. These boundaries do not represent the administrative city area, but are defined by the dispersion of inhabitants and functional relationships. Functional urban areas provide us with the true understanding of how widely the city influences surrounding areas, but the local city administration often does not have the means to directly influence the development of FUAs. This provides us with some contradictory insights and Vilnius area can be a good example to briefly describe some basic observations.

Vilnius city municipality sets direct standards for the development of the interior areas. Nevertheless, from the URAU provided extent of the functional urban area, we see that the economic and commuting activity extents far from its administrative bounds. Vilnius city FUA covers the whole Vilnius city municipality, but also a part of Vilnius regional municipality. In addition, it extends into regional municipality areas of Trakai and Elektrėnai. This simple example provides us with some insight of the usefulness of good management of central cities as well as the extent of their influence on surrounding governance units. For a more distance thought – it also gives a reason to doubt local administrative unit autonomy and prefer a more centralized strategic planning approach.

*Figure 7.* European City areas and Functional Urban Areas (FUAs) as represented in the Urban Audit shapefiles 2011-2014.

**Urban Morphological Zones**

Urban Morphological Zones (UMZ) have been constructed from Corine Land Cover database updates of 2000 and 2006 and made available in 2013 by the EEA and the Joint Research Center population density grid of 2001. UMZ change 2000-2006 and UMZ coverage of 2006 are used in this research in GIS shapefile formats provided by the EEA (EEA, 2014).

UMZs are defined by the CORINE Land Cover classes as built-up areas that are not further than 200 m apart. The definitions of classes that construct the urban fabric are defined by the EEA (2014) as:

- 'Continuous urban fabric' comprises buildings, roads and artificially surfaced area covering almost all ground; non-linear areas of vegetation and bare soil are exceptional.

- 'Discontinuous urban fabric' comprises buildings, roads and artificially surfaced areas with vegetation and bare soil occupying discontinuous but significant surfaces
- 'Industrial or commercial units' primarily comprise artificial surfaces (concrete, asphalt) devoid of vegetation but also contain buildings and/or vegetated areas.
- 'Green urban areas' are patches of vegetation within urban fabric including parks and cemeteries with vegetation.

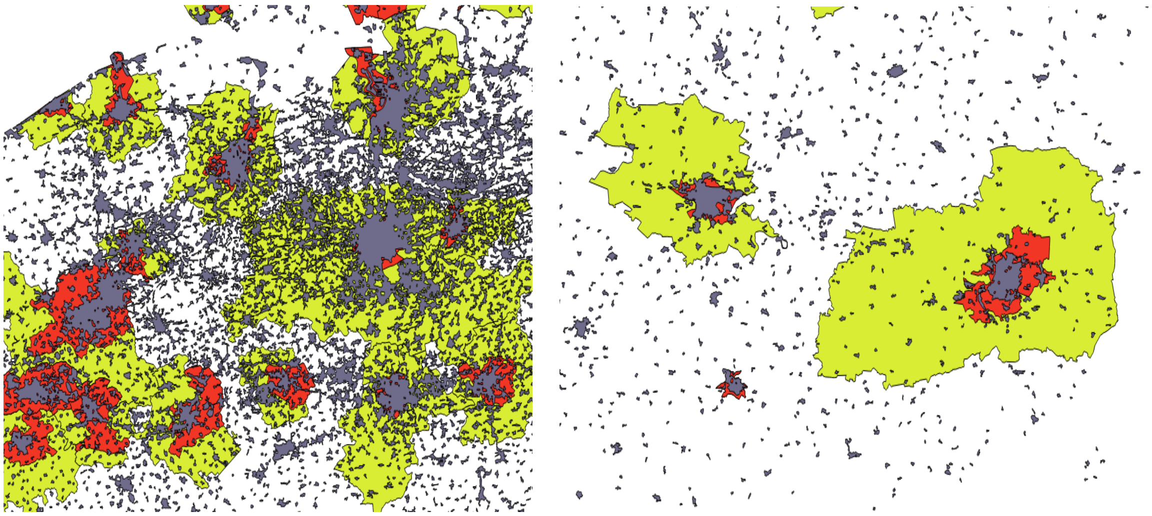


*Figure 8.* Left: Paris and Brussels. Right: Kaunas and Vilnius. Representation in the shapefile dataset. Yellow – FUA, red – City areas (URAU 2011-2014), blue – UMZ (EEA 2006), black – UMZ changes (2000-2006).

UMZs are useful to observe the continuity and the outreach of the fabric. Especially when that fabric can extend between cities. The differences in the size of continuous fabric and the sizes of UMZs vary extensively.

For example, Lithuanian cities contain dispersed patches of UMZ coverage while a single UMZ surrounding Brussels can extend far beyond the boundaries of an administrative city, the FUA and even cover other surrounding cities.

*Table 1*. Comparison table of Urban morphological units as described by the European commission and it's daughter organizations

| Abbr. | Name | Description | Source |
|---|---|---|---|
| **UMZ** | Urban Morphological Zone | Continously built-up areas, with a maximum spacing of 200m (Milego 2007). UMZ delineations can also be adjusted to LAU2 geometry by the means of a UMZ-LAU2 dictionary. | EEA |
| **City (C)** | City | Administrative city area taken as base (Eurostat, 2012) | Eurostat |
| **Kernel (K)** | Greater city | Greater city is constructed of several cities that function as one urban unit (Eurostat, 2012) | Eurostat |
| **FUA** | Functional Urban Area | These urban areas are based on the analysis of commuting patterns around morphological urban cores (Peeters, 2011). | IGEAT |
| **LUZ** | Larger Urban Zone | Larger Urban Zones harmonized correspond to the new definition of cities for the latest Urban Audit and were created by the consortium Eurostat, Urban Audit and OECD (Djikstra, Poelman, 2012). Later reclassified as Functional Urban Areas (FUAs). | Eurostat |

Source : ESPON 2013 Database Dictionary of Spatial Units. Abbreviations of organizations : IGEAT - Institut de Gestion de l'Environnement et d'Aménagement du Territoire, Unversité Libre de Bruxelles ; EEA – European Environmental Agency ; ESPON – European Spatial Planning Observation Network

*Figure 9.* European City areas, Functional Urban Areas (FUAs) and Urban Morphological Zones (UMZs) as represented in the Urban Audit shapefiles 2011-2014.

**Population and jobs grids**

The population and number of jobs cummulative grid is taken from the a data set provided by the EEA Urban Sprawl project (EEA, 2016) and from the official web page and consists of a of 1x1 km pieces that covers the inhabited areas of the European Union countries. The grid consists of 2006 and 2009 population (including jobs) data and also calculates the Sprawl variables for each squared kilometer grid cell. Data premaration and adjustment processes can be found in the report by the EEA (2016). Due to the complexity, sophistication and time consuming processes, population data is not re-made, since the grid is already available and is officially provided by the European Environmental agency. The population data for year 2012 is constructed from the cumulative data for cities from the Eurostat web page. The processes used in the preparation and adjustments of the data are described in detail in the annexes of the research, but the general data gathering methodology included the following steps:

- gathering population data from Eurostat population grid files for years 2006 and 2011, which contained most of the needed data points,
- interpolating population data for the year 2009, for which the imperviousness coverage is provided as a processed HRL (High resolution layers) by the Copernicus project (the European Environmental agency) described further in this research,
- gathering and work place based number of jobs data from Eurostat web page,
- adding missing data from the National statistical bureaus,
- if the number of jobs is not available from the work place based data, a step of adding number of jobs data from employment numbers and taking into account the commuting population ratios is performed and calculated,
- interpolating missing values, if still existing.

The report and the data sets were made for NUTS2 regions and look at bigger geometrical units, not taking into account Cities or FUAs. These smaller units (Cities and Functional urban areas) are used in our research. Instead of new data gathering, we reappropriate the data of the report for new geographical units with the help of shapefile overlays and assigning each grid of the data set to a FUA or city area by location functions. In this way, we are fairly certain to use similar methodology as the creators of the methodology itself and changing scales should not interfere with the correctness of the results.

The link to the shapefiles and additional information can be found in the appendix „Data sources". Nevertheless, it is important to recognise essential limitations: the data is prone to be flawed since it is constructed and adjusted in relatively complex ways. The data set is also hard to check point-by-point and it is almost impossible to recognize if small mismatches appear.

Population grid is overlaid with the Urban Audit 2011-2014 shapefile that contains the geometrical vectoral outlines of administrative cities and Functional urban areas as described previously in the research in the Urban Audit section. After that every grid is given an identity of the Functional urban area and the city that it overlaps with. The values of population count are added by their location – FUA or City. These numbers are used as the base of our calculations for perviousness and density and sprawl measures.

The algorithm assigns all the overlaping squares of 1x1 square km to a shape of FUA or City. This can lead to certain degree of precision errors. Nevertheless, since most of the population is gathered in the centers of the shapes, these fluctuations should not drastically affect the results.

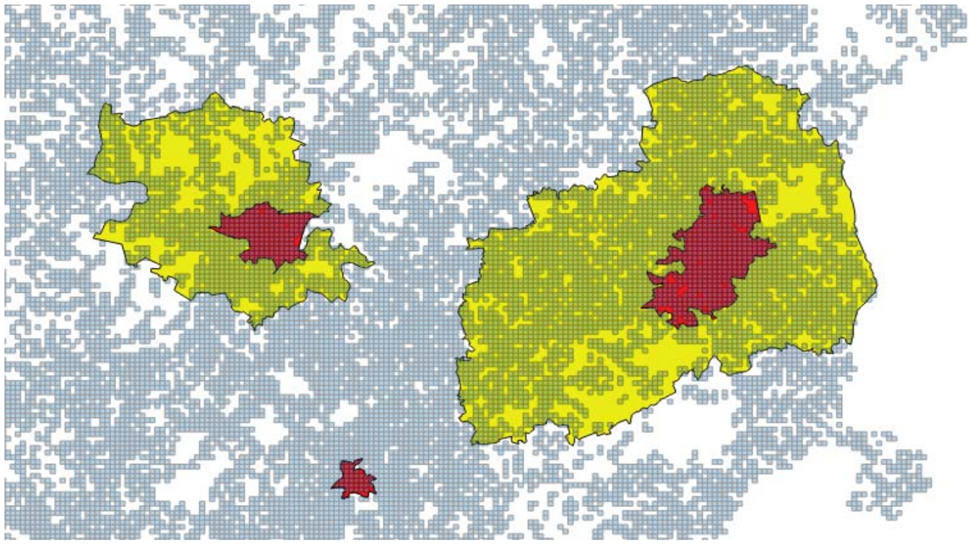


*Figure 10.* An extract of the European shapefile used in research with Vilnius and Kaunas. City areas and FUA areas covered with population grids.

**Imperviousness**

Imperviousness information is taken from the Copernicus project. This project comprices satellite-borne earth maps preproduced for research purposes. The project was created for monitoring the environment and providing sound data for the European Commission and the European Environmental Agency (Copernicus, 2017). Imperviousness layers are High Resolution Layers (HRLs) created through automated processing and interactive classifications. The layers are available in the years of 2006, 2009 and a recent addition of 2012 and 2015. In addition change layers are provided as well. At the moment or the research the change layers involving 2015 are not yet available and being processed. Imperviousness degree (IMD) layers are available in 20x20 m and 100x100 m data raster formats. The research will be using 100x100 m data formats as this has been used in the previous studies (EEA, 2016) while proving to provide suitable information. The research will be using 2012 data as the data set for was not yet available.

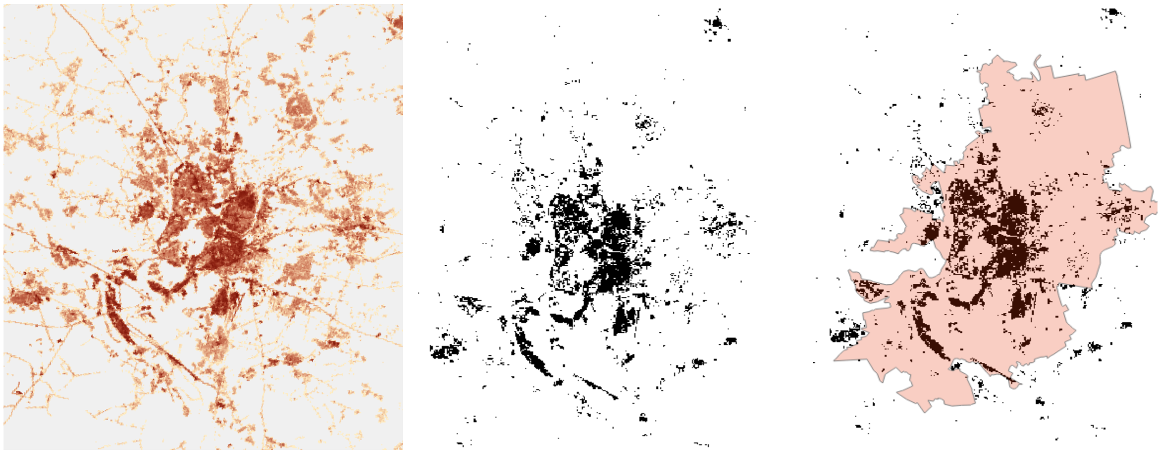


*Figure 11.* LEFT: Vilnius imperviousness 2011 extract of the original file. Grey - no values, Red - 100% imperviousness with decreasing color intensity at lover imperviousness levels. MIDDLE: Vilnius imperviousness 2011 extract of a reproduced binary file. Black – cells that contain a minimum value of 30% imperviousness. RIGHT: A binary file extract of Vilnius with a City shapefile overlay from Urban Audit, Eurostat.

The change files have been added to the official web page in April, 2018. Due to time-consing processes of preparation and nalysis, these data sets will not be involved in our research.

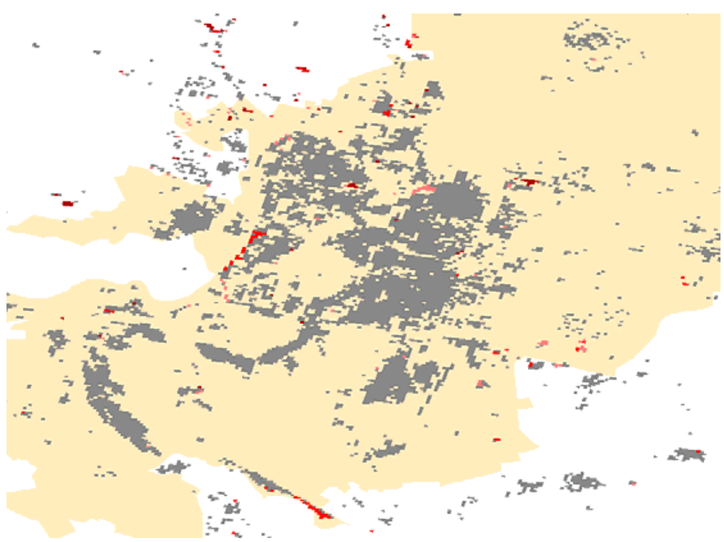


*Figure 12.* Vilnius imperviousness coverage in 100 m grid with 30 % or more imperviousness coverage. Yellow color covers land bounded by the administrative outline of the city of Vilnius, grey represents the 30 % or more imperviousness coverage of the year 2006, pink - 2009, red - 2012, dark red – 2015.

**Corine Land Cover Data**

Corine land cover (CLC) started off as a project in 1994, financed by the European Environmental agency and was based on analogue methods of examining land imaging data and classifying it into main categories. Hence, the accepted common methodology for CORINE Land Cover consists of computer assisted photo-interpretation of satellite images, with the simultaneous consultation of ancillary data, classifying data into classes of the CORINE Land Cover nomenclature (Bossard et al., 2010). The data sets are small-scale and contain detailed information on all Europe. For example, CORINE land cover data set for 2012 has 2370829 data instances as vector shapes.

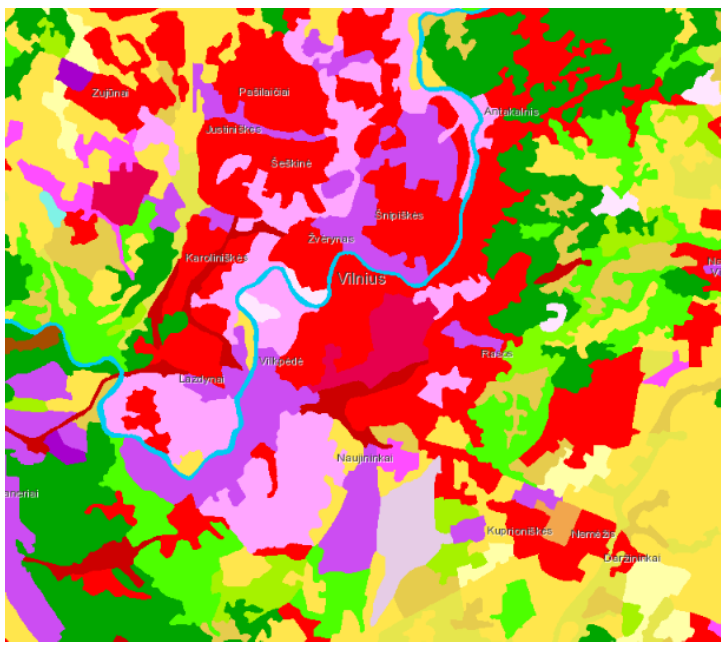



*Figure 13* Vilnius in CORINE land cover data set with original covering with all data classes and legend coloring. Red color represents the urban fabric.

CORINE land cover layers, descriptions and codes can be found in APPENDIX 2. The datasets are available in ESRI shapefile formants for years 2006 and 2012. Our research uses 2012 data coverage with particular interest in urban classes. The layer information is crossed with FUA and CITY areas defined by the URAU shapefile geographical forms. Ratios in percentage are taken into account which are constructed by crossing the shapefile data for geographical urban units and the CORINE layers. Layer information is crossed with city and (or) Functional urban area bounds according to general logics of the influence on the territory.

*Table 2.* Corine land cover layers used in this research and calculated as a ratio to a geographical unit's area. Reasoning

| Layer name | Calculated for FUA | Calculated for CITY | Reasoning |
|---|---|---|---|
| **Agriculture** | Yes | No | Agricultural land is mostly found outside of the central city area therefore it is not necessary to calculate the land for the central area as it is fragmental and would not give a reasonable estimate of the actual urban functions. Calculating agricultural area as a part of functional urban area gives an insight on how much the urban and non-urban functions are mixed in the city-influenced geographical area. |
| **Continuous urban fabric** | Yes | Yes | Continuous urban fabric represents efficient land use and is mostly common for central areas of cities. Calculating the proportion of city area that is occupied with continuous urban fabric gives an understanding of the efficiency of the management of the city. Nevertheless, since the central city territory is artificially defined, FUA ratio of urban land provides us with the full understanding of efficient resource allocation for the actual whole economic unit. |
| **Discontinuous urban fabric** | Yes | Yes | Discontinuous urban fabric represents less efficient land allocation. It is common as a surrounding texture of the central area, but also plays a role in identifying the efficiency of management of the city. Discontinuous urban fabric is closely related to the term "sprawl". |
| **Industrial land and commercial units** | Yes | Yes | The land is primary dedicated to industrial and economic purposes. These patches are not commonly located in the central areas of the cities. These areas mostly contain large-scale structures and are in great contrast with the small-scale textures of the central mixed-use areas of the cities. If found in the central area, they can help predict the overall physical character of the city, which differs greatly from the usual human-scale built environment. Locating these areas in the Functional zone suggests dedication to industry and economic development. |

| | | | |
|---|---|---|---|
| **Green Urban Areas** | Yes | No | Green urban areas are defined as green plots that are constructed and took care of by people and also incorporate leisure, sports and leisure facilities (stadiums and parks). These artificially created green land patches are common for the central parts of the city and take on a relatively small part of functional urban areas. |
| **Marine** | No | Yes | Marine connections are influential for trade economy and attractiveness of an urban area. It also gives insight on the climatic conditions, which influence physical structures of cities. Marine variable is a dummy variable that identifies all functional urban areas that have a direct connection with the shoreline of marine water faces. |
| **Forests (Trees)** | No | Yes | Forest areas that are covered with trees are not common in the central areas of the cities. As with agricultural land. Locating these geographical patches in the functional zone provides us with an insight to the physical connection to nature of the inhabitants. Ratios of considerable size also might suggest low demand for built-up land, concentration on forestry activities and geographical location. |
| **Vegetation** | Yes | Yes | Vegetation layer consists of shrubs and herbaceous vegetation areas, which include low-height shrubs, natural grass-lands, moors, heathlands, transitional woodlands and are more common for surrounding areas, not the central city points. It will be calculated as a proportion to the FUA and city areas. |
| **Wetlands** | No | Yes | Wetlands represent the land that is expensive to built-up and is a limiting factor for economic activities and construction. It is an important physical factor for the development of the city and can influence the density, concentration and geometrical layout of the city and its functions. |
| **Bare land** | Yes | Yes | Natural areas with open spaces that contain sands, beaches, rock surfaces, scarcely vegetated areas, glaciers and burnt areas. These territories can be associated with low agricultural value and land-use intensity. Often they are found in mountainous regions. In city areas they can imply economic constrains since construction work in these areas is resource-consuming. They also act as physical barriers. |

**Summary of variables**

Variable used in clustering is a merged set that represents the physical character of the city in a form of coverage and density representations. Population and jobs count (demographical units), number of households and conventional dwellings (excluding institutional dwellings) with geographical area variables are set to represent the hierarchical character and overall size of the structure. These variables are best estimates of total influence on potential agglomeration effects. Ratio variables provide us with information on the physical conditions of the area (the surfaces of the teritory) and the texture of the city. These ratios also give the understanding of the land-use intensity. Density variables are constructed by the EEA (2016) report and represent the urban patterns of human settlement. These parameters show the physical distances and are important to understand the level of sustainable development policy incorporation and inter-city economic connections. Economic, employment and human capital variables will be used to measure the economic performance of the cities. An overview and summary of the variables can be used in the tabe below.

*Table 3.* An overview of the variables used in this research.

| VARIABLE | Name | Unit | Variable analysis area | Description | Year | Obtained from |
|---|---|---|---|---|---|---|
| **Variables used for Physical Structure Analysis** | | | | | | |
| **TOTpop** | Total population in the administrative city bounds | Number of inhabitants + jobs | Demography | Total population in the geographical area of administrative city and functional urban areas | 2006, 2009, 2012 | Constructed Constructed by this research from URAU and EEA data |
| **Area** | Total area of the analyzed geographical unit | $km^2$ | Structural | Total areas of the shape of the geographical unit as agreed upon by the European Commission. Provided by the Urban audit. | | Constructed from URAU shapefiles by this research |
| **Poly** | Polycentricity of country | Number | Structural, Geography | The index constructed by the ESPON project 1.1.1 | | ESDP |
| **LUP** | Land Uptake Per Person | $m^2$ / inh.+jobs | Density | Built up area / number of users. Computed for city areas | 2009 | Constructed by this research from URAU and EEA data |
| **PBA** | Percentage of Built-Up Area | Percentage | Density | Built-up area / total geographical area. Computed for city areas | 2009 | Constructed by this research from URAU and EEA data |
| **$W_1$ DIS** | Weighted Dispersion | UPU / $m^2$ | Density | A metric of settlement geometric characteristics representing the level of two-dimensional sprawl (UPU / built up area). Computed for city areas | 2009 | Constructed by this research from URAU and EEA data |
| **UP** | Urban Proliferation | UPU / $m^2$ | Density | PBA * DIS. Computed for city areas | 2009 | Constructed by this research from URAU and CORINE data |
| **WUP** | Weighted Urban Proliferation | Construct | Density | A metric derived from UP, DIS and LUP representing urban sprawl. Computed for city areas | 2009 | Constructed by this research from URAU and EEA data |
| **AVEwTemp** | Mean temperature of the warmest month | Number | Geography | Average temperature of the means of warmest month data. If data for a specific city does not exist, it was replaced by the countries mean data | 2000 - 2016 | Eurostat and national statistics |
| **AVEcTemp** | Mean temperature of the coldest month | Number | Geography | Average temperature of the means of coldest month data. If data for a specific city does not exist, it was replaced by the countries mean data | 2000 - 2016 | Eurostat and national statistics |
| **Rain** | Precipitation | mm / $m^2$ | Geography | Average yearly precipitation of 2000 - 2015 | 2000 - 2015 | Eurostat and national statistics |
| **Coastal** | A Coastal FUA | Dummy | Geography | 1 if by coast | 2011 | Constructed by this research |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | from URAU shapefile |
| **Capital** | Capital City | Dummy | Structural / Geography | 1 if is a capital city, else 0 | 2011 | URAU data |
| **Kernel** | City in Kernel | Dummy | Structural / Geography | 1 if belongs to a kernel, else 0 | 2011 | URAU data |
| **Soviet** | Belonging To Communist Regime | Dummy | Historical / Geography | 1 if belonged to the soviet regime area, else 0 | | Constructed by this research from URAU and historical map data |
| **UMZinCity** | UMZ area as proportion of City area | Percentage | Structural | The sum of all overlapping Urban Morphological Zones areas / City area | 2006 | Eurostat, URAU data |
| **UMZinFUA** | UMZ area as proportion of FUA area | Percentage | Structural | The sum of all overlapping Urban Morphological Zones areas / FUA area | 2006 | Eurostat, URAU |
| **CITYinFUA** | Administrative city area in FUA | Percentage | Structural | The sum of city areas in the FUA, in which the city is located. | 2011 - 2014 | Constructed from URAU shapefile data |
| **Cont** | Continuous urban fabric | Percentage | Structural | Continuous urban fabric as percentage of urban unit of the administrative city and functional urban areas. (Corine code 111) | 2012 | Copernicus, EEA, URAU |
| **Discont** | Discontinuous urban fabric | Percentage | Structural | Discontinuous urban fabric as percentage of urban unit of the administrative city and functional urban areas. (Corine code 112) | 2012 | Copernicus, EEA, URAU |
| **Industrial** | Industrial and commercial areas | Percentage | Structural | Sum of industrial areas as percentage of urban unit of the administrative city and functional urban areas. (Corine code 121) | | Copernicus, EEA, URAU |
| **GreenUrban** | Green urban areas, including sporting facilities and leisure outdoor green areas | Percentage | Structural | Sum of green urban, sports and leisure areas as a percentage of city area. (Corine codes 141, 142) | | Copernicus, EEA, URAU |
| **Vegetation** | Low-height vegetation | Percentage | Structural | Natural vegetation area as a percentage of urban unit areas. (Corine codes 321, 322, 323, 324) | | Copernicus, EEA, URAU |
| **Forest** | Areas covered with forests and full grown trees | Percentage | Structural | Forest areas covered with various kinds of full-grown trees as a percentage of FUA area. (Corine codes 311, 312, 313). | | Copernicus, EEA, URAU |
| **Bare** | Bare land | Percentage | Structural | Open spaces with little or no vegetation (rocks, glaciers, etc.) as a percentage of FUA area. (Corine codes 331 – 335). | | Copernicus, EEA, URAU |
| **Wetlands** | Wetland area | Percentage | Structural | Wetland area as a percentage of the area of FUA. (Corine codes 411, 412, 421, 422, 423). | | Copernicus, EEA, URAU |
| **NoHH** | No. of Households | Number | Demography | Number of households (excluding institutional households) | 2011 | Eurostat |

| | | | | | | |
|---|---|---|---|---|---|---|
| **NoDWELL** | Number of dwellings | Number | Structural | Number of conventional dwellings | 2011 | Eurostat |
| **DWELLtoHH** | Dwellings per household | Ratio | Structural | Number of conventional dwellings per household | 2011 | Constructed from Eurostat |
| **CAR** | Car Ownership | Number | Structural | Number per 100 inhabitants | 2011 | Constructed from Eurostat data |
| **ch** | Changes | Dependent on variable | Dependent on variable | Additional variables with respect to observed changes over time was added. Change in population and several ratios | | Constructed by this research |
| **Variables used for Economic Performance Analysis** | | | | | | |
| **GVAtot** | GVA | Monetary Value | Economic | Gross Value Added adjusted for purchasing power parities | 2011 | CFC |
| **GVApw** | GVA per worker | Monetary Value | Economic | Gross Value Added per worker adjusted for purchasing power parities | 2011 | CFC |
| **BsnStock** | Business stock | Number / 100k of population | Economic | Number of business units per 100 000 population | 2011 | CFC, Eurostat |
| **HighPOP** | High skilled population | Percentage | Economic / Development | The percentage of population with high-skill population determined by the ISCED under the 1997 classification | 2011 | CFC, Eurostat |
| **MedPOP** | Medium skilled population | Percentage | Economic / Development | The percentage of population with high-skill population determined by the ISCED under the 1997 classification | 2011 | CFC, Eurostat |
| **LowPOP** | Low skilled population | Percentage | Economic / Development | The percentage of population with high-skill population determined by the ISCED under the 1997 classification | 2011 | CFC, Eurostat |
| **MMU** | Manufacturing, mining and utilities | Percentage | Economic / Development | Employment (jobs) in mining, manufacturing, energy (NACE Rev. 2, B-E) | 2011 | CFC, Eurostat |
| **Patents** | Patent applications to the EPO | Number / 100k of population | Economic | Patent registration as defined European Patent Office, whereas the UK tool shows UK registrations. | 2011 | CFC, Eurostat |
| **BsnServ** | Business services | Percentage | Economic | Percentage of employment in Business service industries. Employment (jobs) in business service activities (NACE Rev. 2, G to L) | 2011 | CFC, Eurostat |
| **PubServ** | Public services | Percentage | Economic | Employment (jobs) in public administration, defense, education, human health and social work activities (NACE Rev. 2, O to Q) | 2011 | CFC, Eurostat |
| **OtherServ** | Other private services | Percentage | Economic / Development | | 2011 | CFC, Eurostat |
| **Emp** | Employment | Percentage | Economic | Employment rate | 2011 | CFC, Eurostat |
| **Unemp** | Unemployment rate | Percentage | Economic | | 2011 | CFC, Eurostat |

Abbreviations: CFC – center of cities, URAU – urban audit, EEA – European Environmental agency (an agency of the European Commission), Copernicus – land monitoring services project by the EEA, CORINE – Corine land cover monitoring sercive by the EEA, FUA – functional urban area, ESDP – ESPON data base portal, financed by the EU European Regional Development Fund, ISCED - International Standard Classification of Education, EPO – European patent office.

**Defining the number of clusters**

The number of clusters was chosen after several tests described in the methodology part. The tests have shown that there is no clearly defined elbow point, when running the test for standardized data with a Z-score transformation. Number of clusters test was also made for data with subjective decimal scaling transformations. This transformation gives a much better predictive power for the data set, but will not be used further in the research. The decimal scaling data predicts 3 – 10 clusters and the Z-score transformation data shows that the clustering will not fully explain the data differences. Nevertheless, as it was mentioned before, this is common in complex data sets. Aksoy et al. (2016) and Nasri & Zhang (2018) have performed the clustering for 10 and 3 clusters respectively, suggesting that the intuitive number for the physical typology types of cities lies somewhere in between. A test for 8 clusters was chosen since it is a number that can provide us with some general insights, although probably not perfect.

**Cluster testing**

The clustering was performed on 759 after accounting for inconsistent or missing physical data and outliers. Clustering is performed using KMeans_prcc cluster testing from an R package "ClusterR", which uses Armadillo library functions and provides a much wider oversight on the clustered data than an alternative process KMeans_arma. KMeans_prcc was built on KMeans_arma, but the results are better described and the function provides more flexibility. It allows for multiple initializations, that means it can run several clustering algorithms consequently by selecting the starting centroids in different point. The centroid initialization is set at "optimal_init": this algorithm adds new data point one by one while checking that they don't already exist. Initiation can also be set on a random mode, the cluster centers can be pre-set, using KMeans++. Nevertheless, these initialization parameters usually do not provide with generally different results. The number of initializations was set at 5, which is arbitrary. The centroids of the clusters are provided in the annexes.

The main descriptive results of the are provided below:

- Total SSE: 32594;
- WCC per cluster in the order of clusters: 2560.177, 705.434, 3435.021, 3918.105, 1152.852, 4452., 1818.31, 3437.964;
- Between SS divided by TSS (total sum of squares) – the ratio that gives us the explanatory power of the clustering algorithm was quite low: 0.3409819.

**Clustering results**

8 clusters were grouped. The physical parameters in the different groups are inconsistent and might suggest, that even the descriptive value of the clustering is not very big, it is reasonable to assume that the results might have some explanatory power on the city structure, its composition, attractiveness and later on – economic performance. The map with assigned clusters provides with an insight that even though the cities were not defined by their geographical location, the structure and environmental conditions (which largely influence the structural patterns of urban units) are enough to guess which geographical area the city belongs to. The description of the clusters with the implications and short insights on the results are given below.

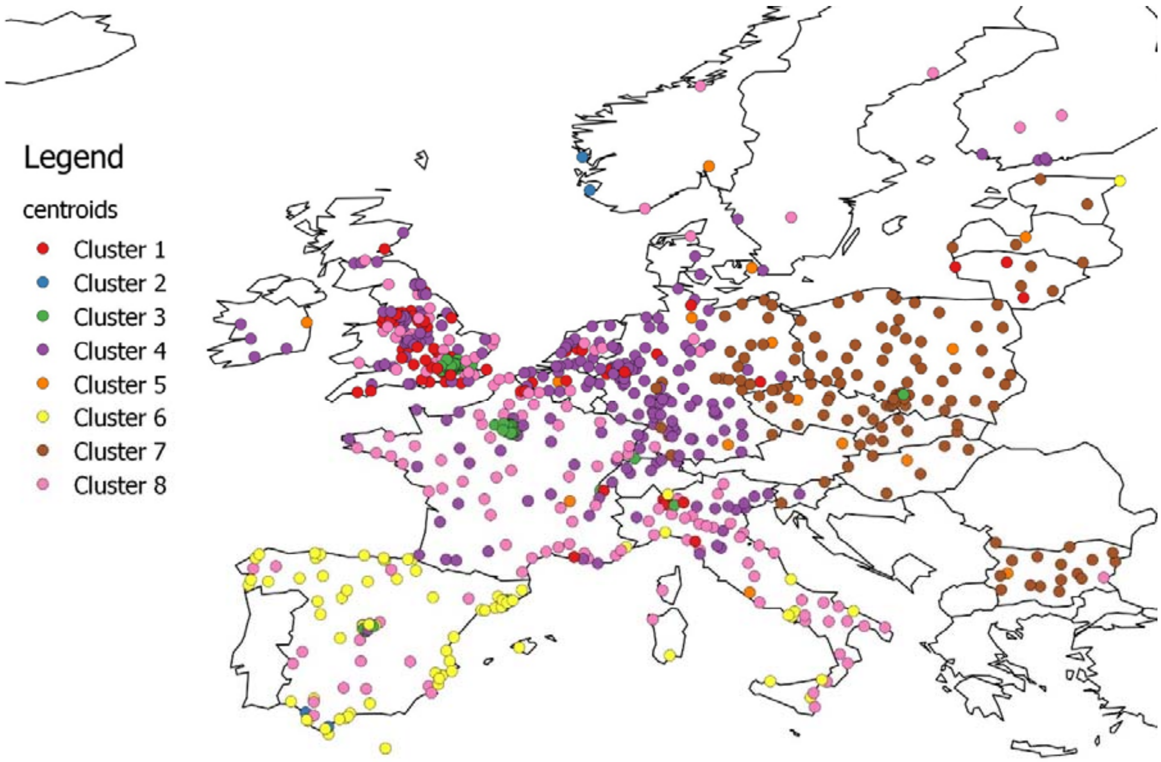



*Figure 14* The map of the European continent with overlaid cities colored by cluster. Source: this research. Visualization tool: QGIS.

***Cluster 1: the masses.*** The 66 cities in cluster 1 have the biggest Z-score mean for the discontinuous urban fabric, industrial land area uptake in city as in FUA area. They also have comparatively small FUA areas and the biggest city to FUA land area ratio. These cities appear to be medium density and dispersion. Most probably they have surrounding barriers that limit their expansion or they are within a cluster of cities and a part of an inter-connected net. Urban characteristics could identify with second-tier cities that act as a buffer for greater urban structures.

***Cluster 2: the attractions.*** These cities have the largest city area mean and a relatively big city to FUA area ratio. They have very low dispersion and big city areas. Greatest population and the most positive population change. They appear to be highly attractive and well used, with high levels of continuous urban fabric. Highly monocentric, have high levels of precipitation and close connection to the bare land as well as the marine shoreline. These are only 5 city areas: 3 Norwegian West coast cities and two Spanish coastal cities. They have a large reach of bare land which is highly characteristic for marine beach resorts and mountainous regions.

***Cluster 3. the suburbs.*** 82 cities were assigned to this subset. The cluster has the biggest proportion of kernel cities. The cities themselves are small in area, but their Functional urban area is the most spread. High percentage of built-up area in the city and a very small proportion of the FUA population in this particular city. These cities have a medium WUP and a second-biggest positive population change and comparatively the biggest proportion of green urban areas. It would be safe to guess that the cities are walking-friendly and would probably have a good public transport infrastructure while maintaining good connections to the central area.

***Cluster 4: the sprawled.*** Cities in this cluster are defined by the highest WUP index, and are often incorporated in a polycentric structure. They are sprawled and have no other characteristic that stands out. Most probably this is due to the fact that the cluster is composed of highly diverse cities. This can be explained by the high number of cities that were assigned to the subset – 230.

***Cluster 5: the giants.*** The cluster is small and contains only 19 cities. 12 of the cities are capital cities: Berlin, Warsaw, Paris, Oslo, Prague, Wien, Madrid, etc. Other four are Munich, Milano, Barcelona and Hamburg. These cities have distinctively bigger administrative city population than the mean, low urban dispersion in the administrative city bounds and relatively low sprawl measures (WUP). Highly concentrated central city areas are surrounded by big FUA areas, this means that the administrative city has a large reach in the surrounding areas. The population ratio between administrative city area and the FUA is not far from the mean. This fact, in combination with the small city/FUA land area ratio imply that the city in cluster 1 unit does not fit in its administrative bounds and acts as a catalyst for the lands around. This city or FUA area has little or no natural vegetation patches, on the other hand, these cities tend to have large amounts of urban green area. This might be due to the fact that most of the green areas are surrounded by urban development and are simply efficiently used by the inhabitants.

***Cluster 6: the mediteranian.*** The cities are small in area with population numbers close to the data set mean. The lowest sprawl indicator WUP is also assigned here. These cities have a population that is leaving the central area and transferring to the FUA surroundings. The subset distinctly has high mean temperatures. They have high values of continuous urban fabric and quite high ratios of low-rise vegetation land. The cities have a monocentric character and act as stand-alone units of functional clusters. It is no wonder that 66 out of 79 cities assigned to this cluster are Spanish cities and another 10 are Italian.

***Cluster 7: the eastern identity.*** 133 cities belong to the subset. Out of them: 51 is Polish, 25 are German, 17 – Czech, 16 – Bulgarian, 8 are Slovakian and 6 come from Hungary. 7 cities are from the Baltic region, including Vilnius, Kaunas, Tallinn. This cluster appears to have the highest negative population change mean and lowest cold temperatures. This fact, in combination with close to the mean warm temperatures, can suggest an in-land climate and the low coastal Z-score confirms our assumption. The urban units belong to monocentric areas, with large surrounding FUA zones with close-to-mean amounts of densely built-up land patches in the central areas. The cluster contains the highest concentration of post-communist cities.

***Cluster 8: the simple.*** A mix of French (49), Italian (37), UK (26) and Spanish (18) cities occupy most of the cluster. 7 cities come from the Nordic region. These cities have high administrative city area mean and low percentage of built-up area. Small amounts of industrial land patches and low amount of green urban infrastructure areas. High number of car ownership numbers. It would appear, that the cities are very similar to cluster 6, but have smaller dispersion indicators, lower temperatures and higher, above-the-mean administrative city area.

*Table 4.* Definitions of clusters and their main physical characteristics

| Cluster No. | Cities in Cluster | Code Name | Defining Structural Characteristics |
|---|---|---|---|
| 1 | 66 | Masses | Large amounts of discontinuous urban fabric and city to FUA ratios. Marine. |
| 2 | 5 | Attractions | Marine. High use intensity. High positive population change. |
| 3 | 82 | Suburbs | Dense urban areas with green character |
| 4 | 230 | Sprawled | High sprawl indicators |
| 5 | 19 | Giants | Capitals, big populations. |
| 6 | 79 | Mediteranian | Lowest WUP values. |
| 7 | 133 | Eastern identity | Large FUA areas. High post-communist score. |
| 8 | 145 | Simple | High car ownership. Little urban green areas |

After the initial clustering algorithm, we assign a cluster number to each instance in the economic data set. The assigning is dome through the names of the cities with an automated merging R function. Cluster 2 has been omitted from our research and will not be analyzed further. This might be due to the limited information on the urban unit or a relatively small size.

*Table 5* Changes in the number of data points for economic analysis.

| Cluster No. | Cities in Physical Cluster | Cities in Econ. cluster | Code Name | Decrease in cluster size for economic analysis |
|---|---|---|---|---|
| 1 | 66 | 28 | Masses | 58% |
| 2 | 5 | 0 | Attractions | 100% |
| 3 | 82 | 11 | Suburbs | 87% |
| 4 | 230 | 104 | Sprawled | 55% |
| 5 | 19 | 15 | Giants | 21% |
| 6 | 79 | 29 | Mediteranian | 63% |
| 7 | 133 | 30 | Eastern identity | 77% |
| 8 | 145 | 48 | Simple | 67% |

**Economic Performance**

As we can see, the cities tend to show some clustering in economic activity. Table of mean and median of economic performance characteristics by clusters are provided in the chart below:

*Table 6* The mean and median parameters of economic performance measures by cluster.

| | **cluster 1** | | **cluster 3** | | **cluster 4** | | **cluster 5** | |
|---|---|---|---|---|---|---|---|---|
| | mean | median | mean | median | mean | median | mean | median |
| **Business services 2011 (%)** | 22.74 | 21.41 | 23.16 | 21.31 | 23.72 | 23.27 | 31.30 | 31.07 |
| **Business stock 2011 (per 100k of population)** | 315.48 | 260.61 | 427.76 | 423.17 | 484.27 | 463.73 | 734.04 | 729.26 |
| **Employment rate 2011 (%)** | 72.65 | 74.20 | 69.53 | 71.15 | 70.36 | 70.33 | 68.52 | 69.96 |
| **GVA 2011 (£ billion)** | 6.23 | 5.54 | 9.57 | 5.16 | 10.13 | 7.77 | 81.52 | 66.82 |
| **GVA per worker 2011 (£)** | 48100.00 | 47100.00 | 61463.64 | 62700.00 | 53121.15 | 52100.00 | 62360.00 | 60700.00 |
| **High skilled population 2011 (%)** | 29.54 | 27.53 | 33.04 | 30.67 | 32.92 | 33.26 | 37.10 | 40.43 |
| **Low skilled population 2011 (%)** | 27.98 | 29.42 | 31.15 | 30.74 | 22.60 | 20.69 | 20.83 | 17.63 |
| **Manufacturing, mining and utilities 2011 (%)** | 13.00 | 12.02 | 13.94 | 14.22 | 12.30 | 10.99 | 8.94 | 9.15 |
| **Medium skilled population 2011 (%)** | 38.29 | 34.06 | 32.43 | 30.07 | 41.91 | 42.24 | 38.03 | 41.15 |
| **Other private services 2011 (%)** | 31.08 | 29.77 | 33.49 | 30.80 | 28.81 | 28.11 | 31.23 | 30.24 |
| **Other sectors 2011 (%)** | 4.60 | 4.24 | 6.44 | 6.40 | 4.93 | 4.90 | 5.12 | 5.00 |
| **Patent applications to the EPO 2011 (per 100K of population)** | 11.73 | 8.20 | 23.05 | 16.62 | 24.20 | 15.85 | 16.28 | 11.21 |
| **Population 2011** | 276862.21 | 239162.00 | 380847.64 | 181861.00 | 343158.38 | 252545.50 | 2345592.33 | 1729040.00 |
| **Public services 2011 (%)** | 28.52 | 29.27 | 22.96 | 23.14 | 29.48 | 30.25 | 21.76 | 22.45 |
| **Total jobs 2011** | 129717.86 | 112850.00 | 178118.18 | 79500.00 | 190087.50 | 142600.00 | 1264706.67 | 940600.00 |
| **Unemployment rate 2011 (%)** | 8.47 | 7.51 | 12.64 | 10.31 | 9.06 | 8.39 | 10.70 | 10.41 |

| | **cluster 6** | | **cluster 7** | | **cluster 8** | | **TOTAL** | |
|---|---|---|---|---|---|---|---|---|
| | **mean** | **median** | **mean** | **median** | **mean** | **median** | **mean** | **median** |
| **Business services 2011 (%)** | 20.25 | 20.56 | 18.97 | **17.26** | **18.36** | 18.06 | 22.13 | 21.23 |
| **Business stock 2011 (per 100k of population)** | 770.11 | 762.72 | 575.89 | 604.54 | 556.66 | 498.10 | 502.12 | 468.79 |
| **Employment rate 2011 (%)** | 60.68 | 60.63 | 62.92 | 59.79 | 65.91 | 65.28 | 67.76 | 68.29 |
| **GVA 2011 (£ billion)** | 6.74 | 4.77 | 4.69 | 4.19 | 4.77 | 4.26 | 11.78 | 5.64 |
| **GVA per worker 2011 (£)** | 52806.90 | 52500.00 | 35413.33 | 34900.00 | 50779.17 | 49900.00 | 50996.60 | 50400.00 |
| **High skilled population 2011 (%)** | 35.01 | 37.60 | 21.69 | 16.19 | 29.75 | 30.80 | 31.16 | 31.93 |
| **Low skilled population 2011 (%)** | 38.76 | 38.71 | 21.62 | 27.64 | 28.22 | 24.94 | 26.12 | 24.96 |
| **Manufacturing, mining and utilities 2011 (%)** | 9.13 | 7.31 | 17.80 | 17.32 | 12.29 | 10.90 | 12.53 | 11.08 |
| **Medium skilled population 2011 (%)** | 25.82 | 24.02 | 53.57 | 53.84 | 39.02 | 40.52 | 39.96 | 40.68 |
| **Other private services 2011 (%)** | 35.40 | 35.43 | 28.85 | 27.09 | 32.06 | 31.67 | 30.70 | 29.30 |
| **Other sectors 2011 (%)** | 8.14 | 7.87 | 6.77 | 6.77 | 7.85 | 7.24 | 6.06 | 5.61 |
| **Patent applications to the EPO 2011 (per 100K of population)** | 2.79 | 2.06 | 3.25 | 1.27 | 8.84 | 5.84 | 15.05 | 9.22 |
| **Population 2011** | 328625.97 | 242503.00 | 317644.40 | 234380.50 | 218137.06 | 186618.50 | 423939.20 | 235076.00 |
| **Public services 2011 (%)** | 22.41 | 23.69 | 26.75 | 27.65 | 24.91 | 31.04 | 26.76 | 28.67 |
| **Total jobs 2011** | 128300.00 | 92100.00 | 127513.33 | 112750.00 | 94087.50 | 84000.00 | 212805.28 | 115900.00 |
| **Unemployment rate 2011 (%)** | 20.85 | 20.58 | 10.07 | 10.46 | 14.56 | 13.83 | 11.64 | 10.48 |

***Cluster 1: the masses.*** Low business stock might be due to the fact the cities in this cluster. High employment numbers go along with this assumption. Balanced proportions of the levels of education within the population. High level of employment in public sector

***Cluster 3. the suburbs.*** Relatively low business stock. It might be safe to assume that is due mostly provide the labor force for the central capital region, which is found in cluster 5 as they surround and accompany the central regional giant. High GVA per worker ratios. Low level of employment in public sector. High numbers of patent applications per capita.

***Cluster 4: the sprawled.*** Cities in this cluster are defined by High level of employment in public sector. These cities have high levels of patent applications. This might be due to the cheaper labor force as the cities are mostly second – tier, located in the central cluster of European economy.

***Cluster 5: the giants.*** Highest employment in business services. High number of business stock and extreme levels of GVA and Relatively high GVA per worker ratio. High numbers of highly skilled population. Low level of employment in public sector, probably due to the fact that overall output of the city is much greater than the public infrastructure needed to sustain this level. This assumption also extends levels of employment in mining, manufacturing and utilities workplaces. Reasonable numbers of patent applications per capita. This might be due to the fact that the research prices are expensive in the central region and can be outsourced.

***Cluster 6: the mediteranian.*** High number of business stock. Low employment, GVA and GVA per worker levels. High numbers of highly skilled and low-skilled population with a small number and high unemployment levels. Low levels of employment in mining, manufacturing and utilities workplaces.

***Cluster 7: the eastern identity.*** Decent business stock numbers. Low employment. High levels of employment in the manufacturing facilities. Most of the results are relatively medium and the cities in this cluster most probably are quite versatile. Nevertheless, this is the post-communist cluster and there is no doubt that the region still has some developmental catching-up to do.

***Cluster 8: the simple.*** Relatively high level of employment in the public sector. Relatively high overall unemployment levels. Reasonable numbers of patent applications per capita. European medium cities, second – tier and located in developed countries. It is no doubt, that the economic conditions of the surroundings boost the output.

## Discussion, implications and Limitations

### Limitations

Data collection process is highly complex and is prone to human error. The data sets have been pre-produced by a number of institutions that differ in their methodology of collection. This holds true for data in various levels. For example: CORINE land cover data is collected from national institutions and does involve some manual adjustments that can be very objective and depend on personal perspectives. Missing data was interpolated or replaced by a value that is closest to the missing value. This can cause discourses from true results. In missing points, where appropriate, the data point was adapted from a different geographical scale. For example: the mean temperature data that was absent for a city scale was assigned from the country mean. This will not distort the results for small countries. Nevertheless, if a country is big and has varying landscapes (mountainous regions, coastal areas and inland dry lands like Spain), the temperature might have somewhat substantial differences.

Choosing data adjustments and number of clusters for K-Means clustering is arbitraty. The Z-score standardization was not previously assesed for this particular type of data and therefore might not be the most appropriate approach. This might be due to the lack of adjusting the weights of each variable, since the algorithm calculates Euclidean distances and evaluates each point with equal proportion, not adjusting for the distribution or size differences between the variables. In further data processing it would be reasonable to analyze the best standardization methods (including Z-score) and research the appropriate weights for each variable with the careful selection of the number of best representative variables, since some contain closely related information.

The plots of the F-score, WCSS and other methods discussed in the research methodology part could not predict a clearly defined number of clusters and the increase in explanatory power by clustering increased gradually with the number of clusters. This might suggest a more linear relationship between standardized variables which could be explored.

In addition, we did not account for the street network syntax, which could be added to the exploration. The Walkscore is another variable that has shown to have a predictive power, but is not yet easily accessible for the European geographical region.

**Discussion and Implications**

From the results we can see that clustering with physical parameters can provide some basic insight into the functional and economical structure or the city. In addition, we can extract some reasonable overviews in the geographical distribution. This might be due to the fact that different countries tend to have various traditions of building and react to the climatic conditions. It is possible to assume, that the physical structure represents the culture of the place, not just the economic performance.

Nevertheless, some general insights can be concluded:

- Big capital cities tend to cluster together and are easily defined as having exceptional structural parameters: density, spread, land-use parameters.

- The surrounding cities of the central capitals work as "donors" and provide the living area for the central territories, which contain most of the business stock of the surrounding functional pattern. The livable suburban-type city can be observed in the geographical characteristics such as high levels of coverage by urban green infrastructure that provide an attractive area to live in.
- No single economical unit will work better than an area densely interconnected and sharing the public infrastructure. This includes the public sector employment measures as well as high density and land-uptake.

Cluster analysis is a valuable tool in the assessment of complex organisms as cities are and the possibilities for fine-tuning this process are endless. It is a great time of technological improvement and available information to crack open even the most head-turning structures. This research adds to the geographical economics assessment for cities in the European region and acknowledges the importance of integrated fields.

**Bibliography**

Abel, J. R., Dey, I., & Gabe, T. M. (2012). Productivity and the density of human capital. *Journal of Regional Science*, *52*(4), 562-586.

Aksoy, E., Gregor, M., Fons, J., Garzillo, C., Cugny-Seguin, M., Löhnertz, M., Schröder, C. (2016). City typologies of Europe: a tool to support urban sustainability studies and practices. *WIT Transactions on Ecology and The Environment, 204,* p. 199-210

Alstadsæter, A., Barrios, S., Nicodème, G., Skonieczna, A. M., & Vezzani, A. (2018). Patent boxes design, patents location, and local R&D. *Economic Policy*, *33*(93), 131-177.

Bessis, H. (2016). Competing with the continent: How UK cities compare with their European counterparts. *Centre for Cities, September. Available at: http://www. centreforcities. org/wp-content/uploads/2016/09/16-09-21-Competing-with-the-continent. pdf (accessed 17 April 2017). Google Scholar*.

Bossard, M., Feranec, J., & Otahel, J. (2000). CORINE land cover technical guide: Addendum 2000.

Buhl, J. et al. (2006). Topological patterns in street networks of self-organized urban settlements. *The European Physical Journal, 49,* 513–522.

Cameron, G. (1996). *Innovation and economic growth*. Centre for Economic Performance, London School of Economics and Political Science.

Camagni, R., Capello, R., & Caragliu, A. (2015). The rise of second-rank cities: what role for agglomeration economies?. *European Planning Studies*, *23*(6), 1069-1089.

Caragliu, A., Del Bo, C., Nijkamp, P. (2011). Smart Cities in Europe, *Journal of Urban Technology, 18(2), 65-82.*

Cardoso, R., Meijers, E. (2016). Contrasts between first-tier and second-tier cities: a functional perspective. *European Planning Studies, 24(5),* 996-1015.

Colaninno, N., Alhaddad, B., & Cladera, J. R. (2009). The effectiveness of morphology and street networks in determining models of urban growth at different spatial scales analysis. In *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology IX* (Vol. 7478, p. 747813). International Society for Optics and Photonics.

Cooke, Ph., Boekholt, P. & Todtling, F. (2000). The Governance of Innovation in Europe. Regional Perspectives on Global Competitiveness (London: Pinter).

Copernicus, (2017). Copernicus Project Partners. Retrieved March 16, 2018, from https://land.copernicus.eu/user-corner/partners

Dias Jordão, R. V., Almeida, V. R., (2017). Performance measurement, intellectual capital and financial sustainability", *Journal of Intellectual Capital, 18(3), 643-666.*

Dijkstra, L., Garcilazo, E., & Mccann, P. (2013). The Economic Performance of European Cities and City Regions: Myths and Realities. *European Planning Studies, 21*(3), 334-354.

Dijkstra, L., & Poelman, H. (2012). Cities in Europe: the new OECD-EC definition. *Regional Focus*, *1*(2012), 1-13.

Dobbs, R., Smit, S., Remes, J., Manyika, J., Roxburgh, C., & Restrepo, A. (2011). Urban world: Mapping the economic power of cities. *McKinsey Global Institute.*

Dunnell, K. (2009). National Statistician's article: measuring regional economic performance. *Economic & Labour Market Review*, *3*(1), 18-30.

EEA. (2014, June 27). Urban morphological zones 2006. Retrieved February 20, 2018, from https://www.eea.europa.eu/data-and-maps/data/urban-morphological-zones-2006-1#tab-methodology

EEA, E. (2016). Urban sprawl in Europe-joint EEA-FOEN report.

ESPON (2013). Database Dictionary of Spatial Units. Obtained March 30, 2018, from: http://database.espon.eu/db2/jsf/DicoSpatialUnits/DicoSpatialUnits_onehtml/index.html#N10F7D

ESPON Database Portal. Retrieved March 30, 2018, from http://database.espon.eu/db2/resource?idCat=72

Eurostat (2012). Annual national accounts. Reference metadata in Euro SDMX metadata structure (ESMS).

Frick, S. A., Rodriguez-Pose, A. (2018). Change in urban concentration and economic growth, Discussion Paper DP12566, Published 05 January 2018.

Graham, D. J. (2009). Identifying urbanisation and localisation externalities in manufacturing and service industries. *Papers in Regional Science, 88(1),* 63-84. Obtained January 21, 2018 from: http://onlinelibrary.wiley.com.ezproxy.ulb .ac.be/doi/10.1111/j.14355957.2008.00166.x/full

Gardiner, B., Martin, R., & Tyler, P. (2011). Does spatial agglomeration increase national growth? Some evidence from Europe. Journal of Economic Geography, 11(6), 979- 1006. doi:10.1093/jeg/lbq047

Glaeser, E. L., & Kohlhase, J. E. (2004). Cities, regions and the decline of transport costs. In *Fifty Years of Regional Science* (pp. 197-228). Springer, Berlin, Heidelberg.

Glaeser, E. L., Kahn, M. E., & Rappaport, J. (2000). *Why do the poor live in cities?* (No. w7636). National Bureau of Economic Research.

Glaeser, E., & Saiz, A. (2004). The Rise of the Skilled City. *Brookings-Wharton Papers on Urban Affairs, 2004*(1), 47-105.

Guérois, M., Bretagnolle, A., Mathian, H., & Pavard, A. (2014). Functional Urban Areas (FUA) and European Harmonisation. A Feasibility Study from the Comparison of Two Approaches: Commuting Flows and Accessibility Isochrones. Paris: ESPON Database. In *Technical report, ESPON 2013 database*.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100-108.

Henderson, J. V. (2003). The urbanization process and economic growth: The so-what question. Journal of Economic Growth, 8(1), 47-71. Obtained January 11, 2018 from: https://link-springer-com.ezproxy.ulb.ac.be/article/10.1023/A%3A1022860800744

IGEAT, I. LATTS et TSAC, 2007. *ESPON project 1.4. 3 Study on Urban Functions-Final report*.

Jaeger, J. A., Bertiller, R., Schwick, C., Cavens, D., & Kienast, F. (2010). Urban permeation of landscapes and sprawl per capita: New measures of urban sprawl. *Ecological Indicators*, *10*(2), 427-441.

Jaeger, J. A., & Schwick, C. (2014). Improving the measurement of urban sprawl: Weighted Urban Proliferation (WUP) and its application to Switzerland. *Ecological Indicators*, *38*, 294-308.

Jiang, B. (2015). Geospatial analysis requires a different way of thinking: The problem of spatial heterogeneity. *GeoJournal*, *80*(1), 1-13.

Jiang, B. (2016). Head/tail breaks for visualization of city structure and dynamics. *European Handbook of Crowdsourced Geographic Information*, *169*.

Jiang, B., & Yin, J. (2014). Ht-index for quantifying the fractal or scaling structure of geographic features. *Annals of the Association of American Geographers*, *104*(3), 530-540.

Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning* (Vol. 1). STHDA.

Kostakos, V. (2010). Space syntax and pervasive systems. In *Geospatial analysis and modelling of urban structure and dynamics* (pp. 31-52). Springer, Dordrecht.

Krätke, S. (2007). Metropolisation of the European Economic Territory as a Consequence of Increasing Specialisation of Urban Agglomerations in the Knowledge Economy, European Planning Studies.

Krätke, S. (2011). The creative capital of cities: Interactive knowledge creation and the urbanization economies of innovation. Chichester: Wiley-Blackwell.

Matthews, J. W., & Turnbull, G. K. (2007). Neighborhood street layout and property value: The interaction of accessibility and land use mix. *The journal of real estate finance and economics*, *35*(2), 111-141.

Mavrič, J., & Bobek, V. (2015). Measuring Urban Development and City Performance. In *Perspectives on Business and Management*. InTech.

Mccann, P. (1995). Rethinking the Economics of Location and Agglomeration. *Urban Studies (Routledge)*, *32*(3), 563-578. doi:10.1080/00420989550012979

Meijers, E. J., Burger, M. J., Hoogerbrugge, M. M. (2015). Borrowing size in networks of cities: City size, network connectivity and metropolitan functions in Europe, *Papers in Regional Science, 95(1),* p. 183-184. Obtained November 14, 2017 from: http://evertmeijers.com/wp-content/uploads/2017/07/Meijers-E-Burger-M-Hoogerbrugge-M-2016-PiRS-Borrowing-size-in-networks-of-cities.pdf

Mesev, T. V., Longley, P. A., Batty, M., & Xie, Y. (1995). Morphology from imagery: detecting and measuring the density of urban land use. *Environment and Planning A*, *27*(5), 759-780.

Milego, R. (2007). Urban morphological zones 2000, definition and procedural steps. Final Report, Copenhagen: European Environment Agency, ETC Terrestrial Environment (2007).

Milego, R., & Ramos, M. J. (2013). Disaggregation of socioeconomic data into a regular grid and combination with other types of data. *Technical Report, ESPON*.

Mohamad, I. B., & Usman, D. (2013). Standardization and its effects on K-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, *6*(17), 3299-3303. Obtained November 05, 2017 from https://pdfs.semanticscholar.org/1d35/2dd5f030589ecfe8910ab1cc0dd320bf600d.pdf

Moretti, E. (2004). Human capital externalities in cities. In *Handbook of regional and urban economics* (Vol. 4, pp. 2243-2291). Elsevier.

Nasri, A., & Zhang, L. (2018). A multi-dimensional multi-level approach to measuring the spatial structure of US metropolitan areas. *Journal of Transport and Land Use*, *11*(1).

OECD. Publishing, & Organisation for Economic Co-operation and Development Staff. (2012a). *Redefining Urban: A New Way to Measure Metropolitan Areas*. Organisation for Economic Cooperation and Development (OECD).

OECD (2012b), *Compact City Policies: A Comparative Assessment*, OECD Publishing, Paris. Obtained September 04, 2017 from: http://dx.doi.org/10.1787/9789264167865-en

Orlitová, E., Soukup, T., Kopecky, M., Vobora, V., Gregor, M., Schwick, C., & Jaeger, J. (2012). Urban sprawl typology: Critical analysis of input datasets, indicators calculated and dataflows proposal.

Ozyurt, S., & Dees, S. (2015). Regional dynamics of economic performance in the EU: To what extent spatial spillovers matter?

Pain, K., Hamme, G. V., Vinciguerra, S., David, Q. (2015). Global networks, cities and economic performance: Observations from an analysis of cities in Europe and the USA.

Peeters, D. (2011). The functional urban areas database. *Paris: ESPON, final rapport of the 2013 ESPON Data Base project*

Peiravian, F., de Lapparent, M., & Derrible, S. (2015). *Exploration of the Complex Similarity of Urban System Components* (No. EPFL-REPORT-207468).

Peiraviana, F., Derriblea, S. (2015), Multi-Dimensional Geometric Complexity in Urban Transportation Systems. Obtained November 10, 2017 from: https://arxiv.org/ftp/arxiv/papers/1507/1507.03607.pdf 2017 01 30

Pellenbarg, P. H., & Groote, P. D. (2016). Economic Performance of Cities in the Netherlands. *Tijdschrift voor economische en sociale geografie*, *107*(2), 254-256.

Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *219*(1), 103-119. Obtained November 24, 2017 from https://www.ee.columbia.edu/~dpwe/papers/PhamDN05-kmeans.pdf

Pivo, G., & Fisher, J. (2009). Effects of walkability on property values and investment returns. *Responsible Property Investing Center Working Paper. Boston College and University of Arizona Benecki Center for Real Estate Studies, Indiana University*.

Pivo, G., & Fisher, J. D. (2011). The walkability premium in commercial real estate investments. *Real Estate Economics*, *39*(2), 185-219.

Pholo-Bala, A. (2009). Urban concentration and economic growth: checking for specific regional effects. Center for Operations Research and Econometrics (CORE). Université Catholique de Louvain. Discussion Paper 2009/38. Retrieved November 22, 2017 from https://www.uclouvain.be/cps/ucl/doc/core/documents /Chapter2COREDP.pdf

Porter, M. E. (2003). The economic performance of regions. In *Regional Competitiveness* (pp. 137-166). Routledge.

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, *64*(2), 373-389.

Song, Y., & Knaap, G. J. (2003). New urbanism and housing values: a disaggregate assessment. *Journal of Urban Economics*, *54*(2), 218-238.

Srinivasan, S. (2002). Quantifying spatial characteristics of cities. *Urban Studies*, *39*(11), 2005-2028.

Strano, E., & Sood, V. (2016). Rich and poor cities in Europe. An urban scaling approach to mapping the European economic transition. *PloS one*, *11*(8), e0159465.

Stiglitz, J., Sen, A. K., & Fitoussi, J. P. (2009). The measurement of economic performance and social progress revisited: reflections and overview.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411-423.

Tsai, Y. H. (2005). Quantifying urban form: compactness versus' sprawl'. *Urban studies*, *42*(1), 141-161.

Van Winden, W., Van den Berg, L., & Pol, P. (2007). European cities in the knowledge economy: towards a typology. *Urban Studies*, *44*(3), 525-549.

Wendt, P. F. (1957). Theory of urban land values. *Land economics*, *33*(3), 228-240.

Veneri, P. (2013). On city size distribution: evidence from OECD functional urban areas. *OECD Regional Development Working Papers*, *2013*(27).

**APPENDIX 1**
**Geographical data and data sources**

| Name | Source | Publication date | Year | Link |
|---|---|---|---|---|
| **Population grid, 1 sq.km, Census 2011** | Eurostat | 01 Feb 2016 | 2011 | http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat |
| **Population grid, 1 sq.km, Census 2006** | Eurostat | 23 Jan 2012 | 2006 | http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat |
| **Imperviousness grid 100m, Europe 2015** | EEA | 22 Mar 2018 | 2015 | https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/status-maps/2015 |
| **Imperviousness grid 100m, Europe 2012** | EEA | 06 Apr 2018 | 2012 | https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/status-maps/2012?tab=metadata |
| **Imperviousness grid 100m, Europe 2009** | EEA | 06 Apr 2018 | 2009 | https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/status-maps/2009?tab=download |
| **CORINE land cover vector data sets** | EEA, Copernicus | 12 Dec 2017 | 2012 | https://www.eea.europa.eu/data-and-maps/data#c0=5&c11=&c5=all&b_start=0 |
| **URAU shapefiles** | | 07 Dec 2015 | 2011-2014 | http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/urban-audit |
| **Selected data** | Eurostat | Various additional variables | 2000 - 2015 | http://ec.europa.eu/eurostat/data/database |

## APPENDIX 2
## CORINE land cover classification

Source: shapefile technical data by CORINE

| CLC CODE | LABEL1 | LABEL2 | LABEL3 | USED |
|---|---|---|---|---|
| **111** | Artificial surfaces | Urban fabric | Continuous urban fabric | Yes |
| **112** | Artificial surfaces | Urban fabric | Discontinuous urban fabric | Yes |
| **121** | Artificial surfaces | Industrial, commercial and transport units | Industrial or commercial units | Yes |
| **122** | Artificial surfaces | Industrial, commercial and transport units | Road and rail networks and associated land | Yes |
| **123** | Artificial surfaces | Industrial, commercial and transport units | Port areas | No |
| **124** | Artificial surfaces | Industrial, commercial and transport units | Airports | No |
| **131** | Artificial surfaces | Mine, dump and construction sites | Mineral extraction sites | No |
| **132** | Artificial surfaces | Mine, dump and construction sites | Dump sites | No |
| **133** | Artificial surfaces | Mine, dump and construction sites | Construction sites | No |
| **141** | Artificial surfaces | Artificial, non-agricultural vegetated areas | Green urban areas | No |
| **142** | Artificial surfaces | Artificial, non-agricultural vegetated areas | Sport and leisure facilities | No |
| **211** | Agricultural areas | Arable land | Non-irrigated arable land | Yes |
| **212** | Agricultural areas | Arable land | Permanently irrigated land | Yes |
| **213** | Agricultural areas | Arable land | Rice fields | Yes |
| **221** | Agricultural areas | Permanent crops | Vineyards | Yes |
| **222** | Agricultural areas | Permanent crops | Fruit trees and berry plantations | Yes |
| **223** | Agricultural areas | Permanent crops | Olive groves | Yes |
| **231** | Agricultural areas | Pastures | Pastures | Yes |
| **241** | Agricultural areas | Heterogeneous agricultural areas | Annual crops associated with permanent crops | Yes |
| **242** | Agricultural areas | Heterogeneous agricultural areas | Complex cultivation patterns | Yes |
| **243** | Agricultural areas | Heterogeneous agricultural areas | Land principally occupied by agriculture, with significant areas of natural vegetation | Yes |
| **244** | Agricultural areas | Heterogeneous agricultural areas | Agro-forestry areas | Yes |
| **311** | Forest and semi natural areas | Forests | Broad-leaved forest | No |
| **312** | Forest and semi natural areas | Forests | Coniferous forest | No |
| **313** | Forest and semi natural areas | Forests | Mixed forest | No |

| | | | | |
|---|---|---|---|---|
| **321** | Forest and semi natural areas | Scrub and/or herbaceous vegetation associations | Natural grasslands | No |
| **322** | Forest and semi natural areas | Scrub and/or herbaceous vegetation associations | Moors and heathland | No |
| **323** | Forest and semi natural areas | Scrub and/or herbaceous vegetation associations | Sclerophyllous vegetation | No |
| **324** | Forest and semi natural areas | Scrub and/or herbaceous vegetation associations | Transitional woodland-shrub | No |
| **331** | Forest and semi natural areas | Open spaces with little or no vegetation | Beaches, dunes, sands | No |
| **332** | Forest and semi natural areas | Open spaces with little or no vegetation | Bare rocks | No |
| **333** | Forest and semi natural areas | Open spaces with little or no vegetation | Sparsely vegetated areas | No |
| **334** | Forest and semi natural areas | Open spaces with little or no vegetation | Burnt areas | No |
| **335** | Forest and semi natural areas | Open spaces with little or no vegetation | Glaciers and perpetual snow | No |
| **411** | Wetlands | Inland wetlands | Inland marshes | No |
| **412** | Wetlands | Inland wetlands | Peat bogs | No |
| **421** | Wetlands | Maritime wetlands | Salt marshes | No |
| **422** | Wetlands | Maritime wetlands | Salines | No |
| **423** | Wetlands | Maritime wetlands | Intertidal flats | No |
| **511** | Water bodies | Inland waters | Water courses | Yes |
| **512** | Water bodies | Inland waters | Water bodies | Yes |
| **521** | Water bodies | Marine waters | Coastal lagoons | Yes |
| **522** | Water bodies | Marine waters | Estuaries | Yes |
| **523** | Water bodies | Marine waters | Sea and ocean | Yes |

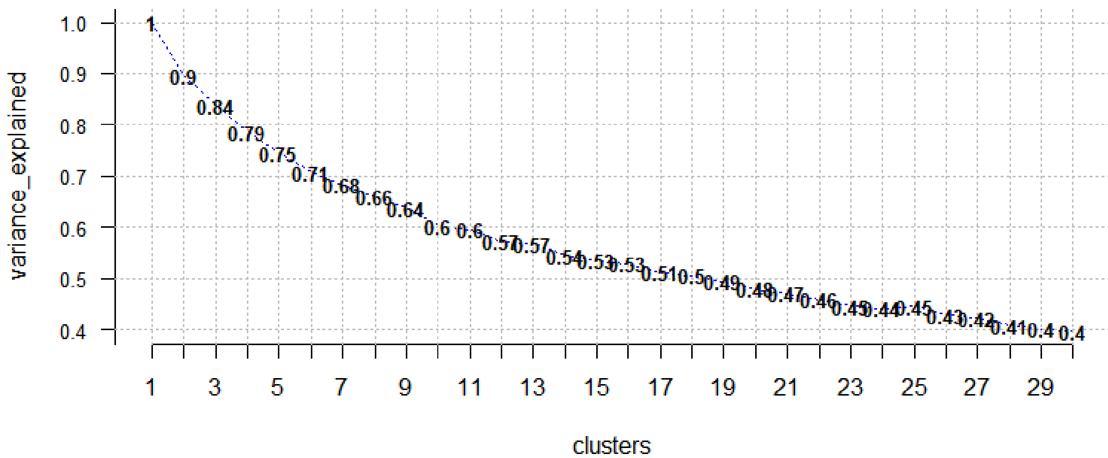
## APPENDIX 3
## "Center For Cities" collection of economic data sources

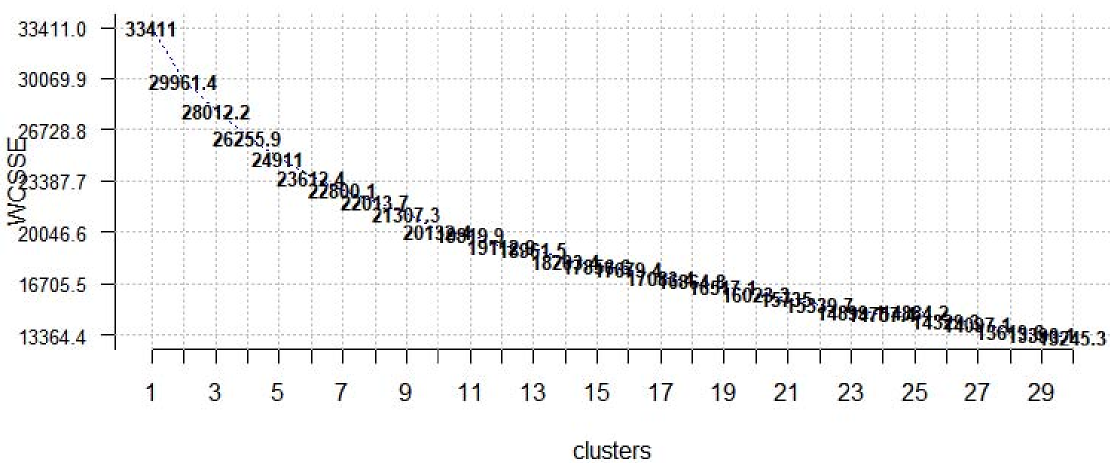
| Variable | Sources and missing data |
|---|---|
| **Business services** | Eurostat, Labour market - cities and greater cities. Data unavailable for Belfast |
| **Business stock** | Eurostat, Economy and finance - cities and greater cities; FSO, Statent; CSO, Business Demography; INSEE, répertoire des établissements et des entreprises; Istat, Census 2011; CBS, Vestigingen van bedrijven. Centre for Cities calculation. Data unavailable for Polish, Spanish and Swedish cities. |
| **Employment rate** | Eurostat, Labour market - cities and greater cities; Statistical Office of Poland, employment rate by municipality; DST, register-based labour force survey; ISTAT, Census 2011. Data for Paris (FR) is for 2012 |
| **GVA** | Eurostat, Gross value added at basic prices by NUTS 3 regions, Employment (thousand persons) by NUTS 3 regions, Labour market - cities and greater cities; ONS, BRES; INSEE, Recensements de la population 2012; ISTAT, Census 2011; Statistical Office of Poland, Labour market; FSO, Statent, GVA per canton. Data for some cities is for 2012. Centre for Cities calculation. |
| **GVA per worker** | Eurostat, Gross value added at basic prices by NUTS 3 regions, Employment (thousand persons) by NUTS 3 regions, Labour market - cities and greater cities; ONS, BRES; INSEE, Recensements de la population 2012; ISTAT, Census 2011; Statistical Office of Poland, Labour market; FSO, Statent, GVA per canton. Data for some cities is for 2012. Centre for Cities calculation. |
| **High skilled population** | Eurostat, Education - cities and greater cities; Statistics Belgium, Census 2011; FSO, Urban Audit statistics; DST, Educational attainment; CSO, Census 2011; Hellenic Statistical Authority; INE, Urban Audit statistics; INSEE, RP 2012; CBS, Municipal populaton; SCB, level of education by municipality. Data for Paris is for 2012. Data unavailable for Milan (IT) and Naples (IT) |
| **Low skilled population** | Eurostat, Education - cities and greater cities; Statistics Belgium, Census 2011; FSO, Urban Audit statistics; DST, Educational attainment; CSO, Census 2011; Hellenic Statistical Authority; INE, Urban Audit statistics; INSEE, RP 2012; CBS, Municipal populaton; SCB, level of education by municipality. Data for Paris is for 2012. Data unavailable for Milan (IT) and Naples (IT) |
| **Manufacturing, mining and utilities** | Eurostat, Labour market - cities and greater cities. Data unavailable for Belfast |
| **Medium skilled population** | Eurostat, Education - cities and greater cities; Statistics Belgium, Census 2011; FSO, Urban Audit statistics; DST, Educational attainment; CSO, Census 2011; Hellenic Statistical Authority; INE, Urban Audit statistics; INSEE, RP 2012; CBS, Municipal populaton; SCB, level of education by municipality. Data for Paris is for 2012. Data unavailable for Milan (IT) and Naples (IT) |
| **Other private services** | Eurostat, Labour market - cities and greater cities. Data unavailable for Belfast |
| **Other sectors** | Eurostat, Labour market - cities and greater cities. Data unavailable for Belfast |
| **Patent applications to the EPO** | Eurostat, Patent applications to the EPO by priority year by NUTS 3 regions, Gross value added at basic prices by NUTS 3 regions, Employment (thousand persons) by NUTS 3 regions, Labour market - cities and greater cities; ONS, BRES; INSEE, RP 2012; ISTAT, Census 2011; Statistical Office of Poland, Labour market; FSO, Statent, GVA per canton. Centre for Cities calculation. Data unavailable for some cities. |

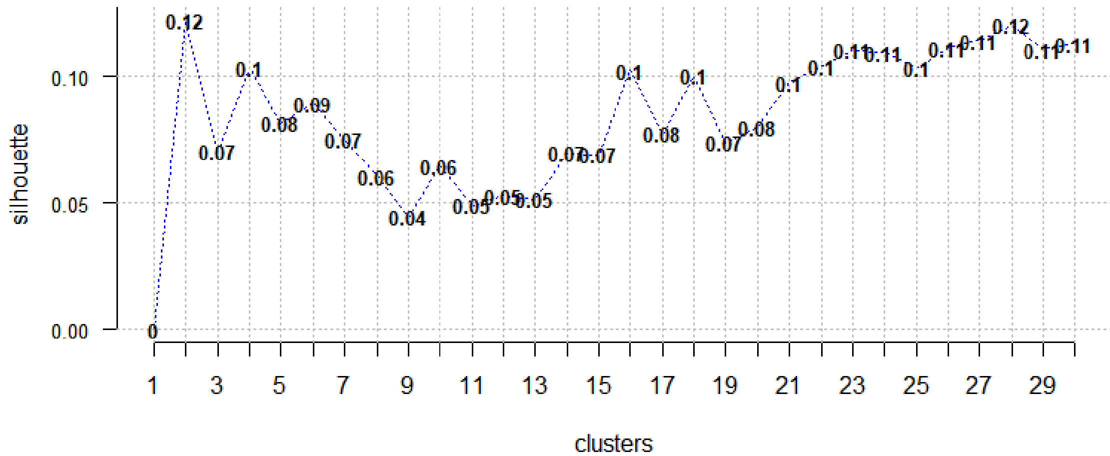
| | |
|---|---|
| **Population** | Source: Eurostat, Population on 1 January by age groups and sex - cities and greater cities; Population on 1 January by broad age group, sex and NUTS 3 region |
| **Public services** | Source: Eurostat, Labour market - cities and greater cities. Data unavailable for Belfast |
| **Total jobs** | Source: Eurostat, Labour market - cities and greater cities. Data unavailable for Belfast |
| **Total jobs** | Source: NOMIS, Business Register and Employment Survey. |
| **Total jobs** | Source: NOMIS, Business Register and Employment Survey |
| **Unemployment rate** | Source: Eurostat, Labour market - cities and greater cities; Statistical Office of Poland, labour market. Data for Paris (FR) is for 2012 |

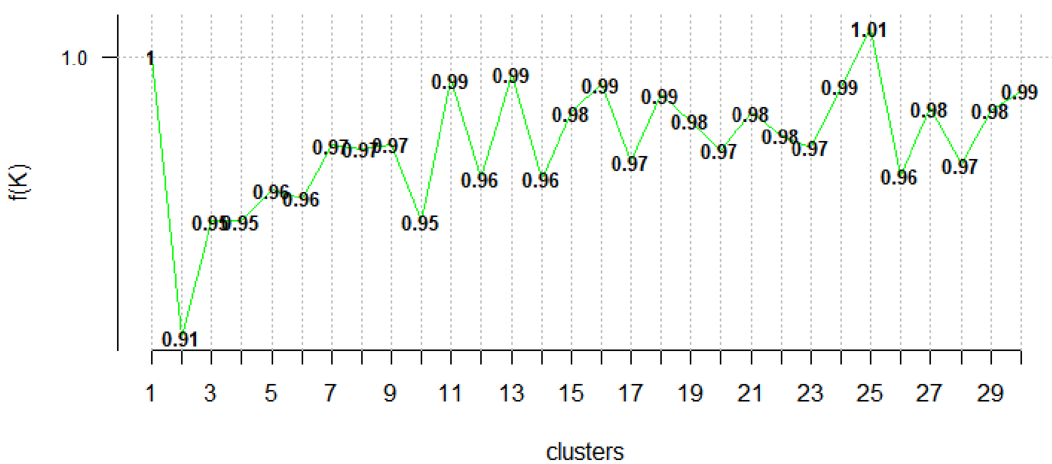
**APPENDIX 4**
**Number of cluster testing graphs**

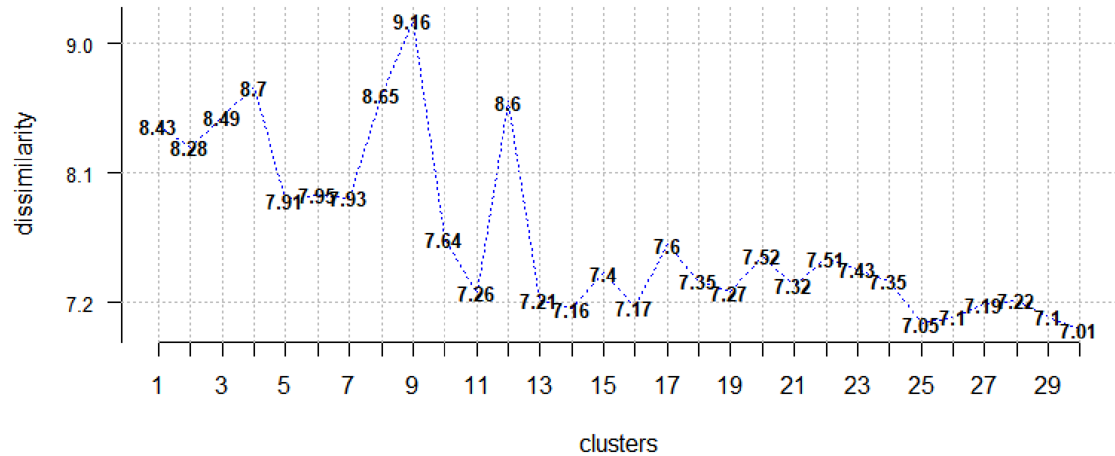
Z-score transformation results for testing the number of clusters:

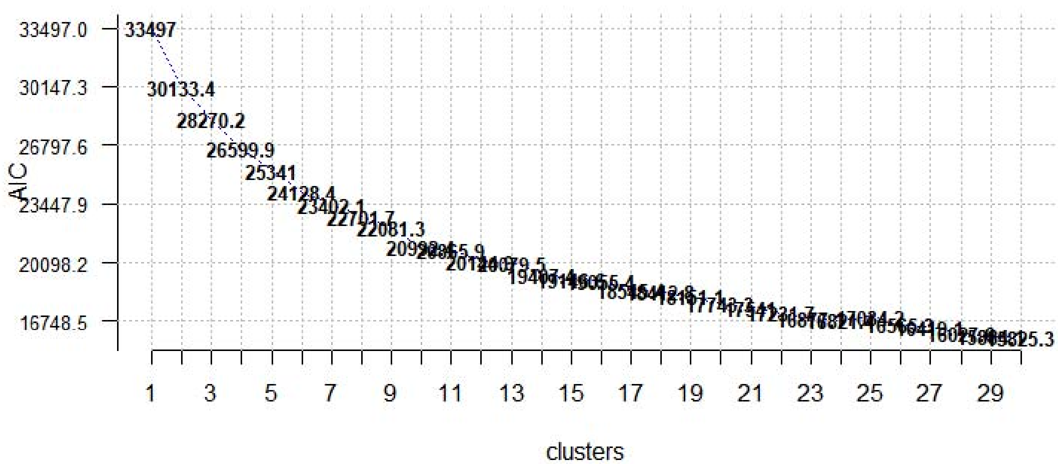


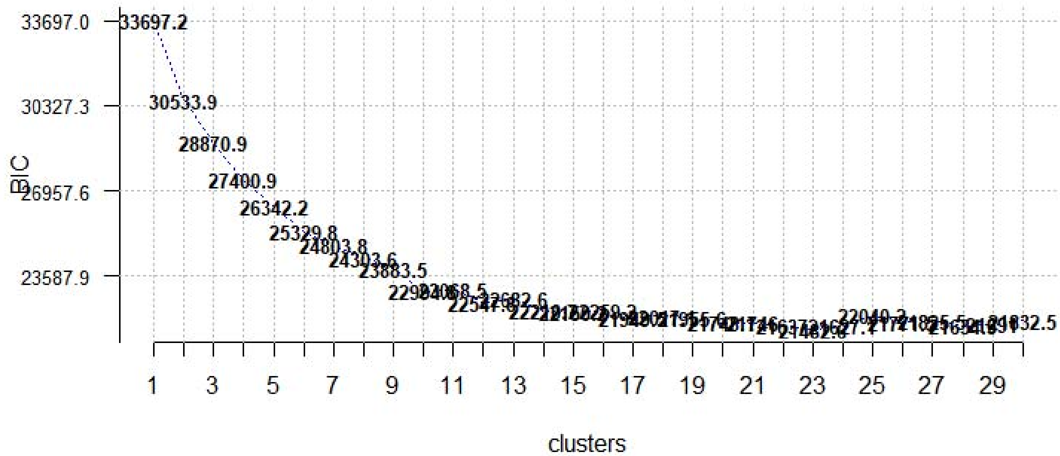



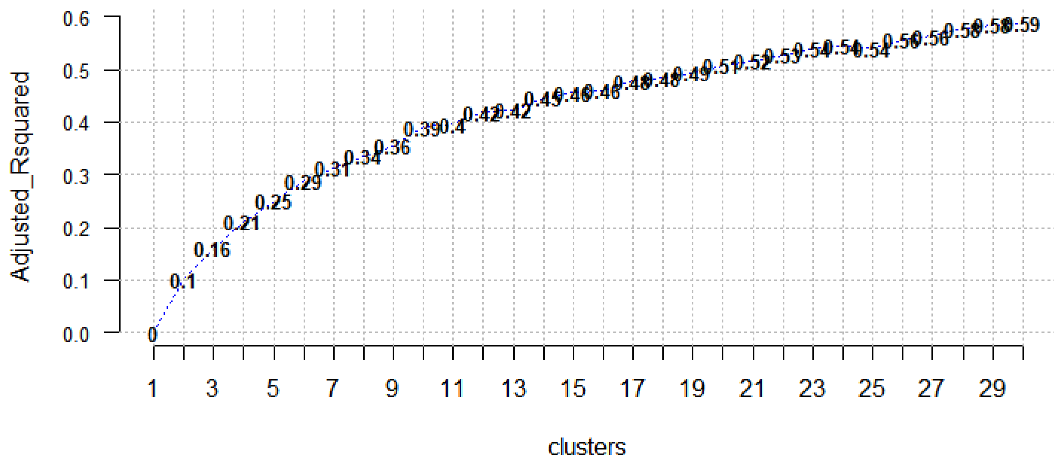

silhouette
clusters


f(K)
clusters

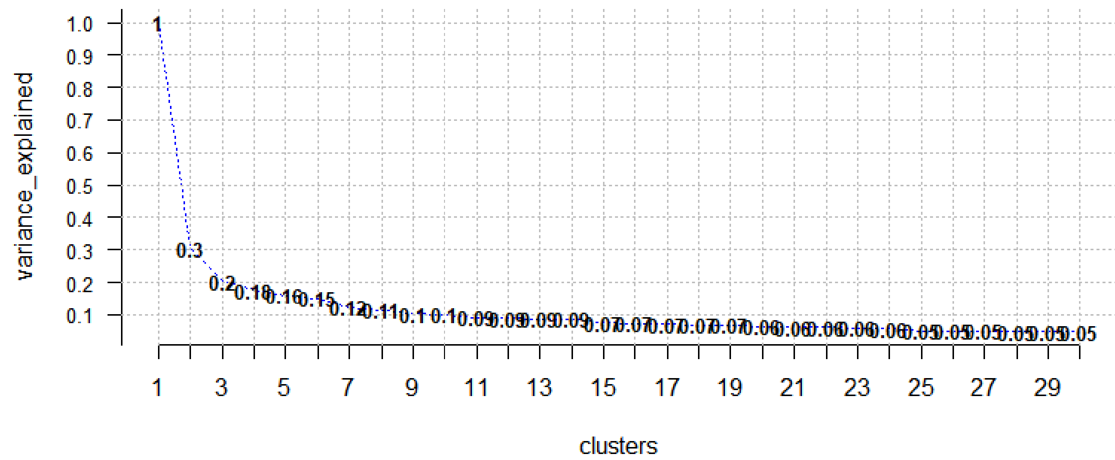

dissimilarity
9.0
8.1
7.2
clusters
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
8.43
8.28
8.49
8.7
7.91
7.95
7.93
8.65
9.16
7.64
7.26
8.6
7.21
7.16
7.4
7.17
7.6
7.35
7.27
7.52
7.32
7.51
7.43
7.35
7.05
7.1
7.19
7.22
7.1
7.01

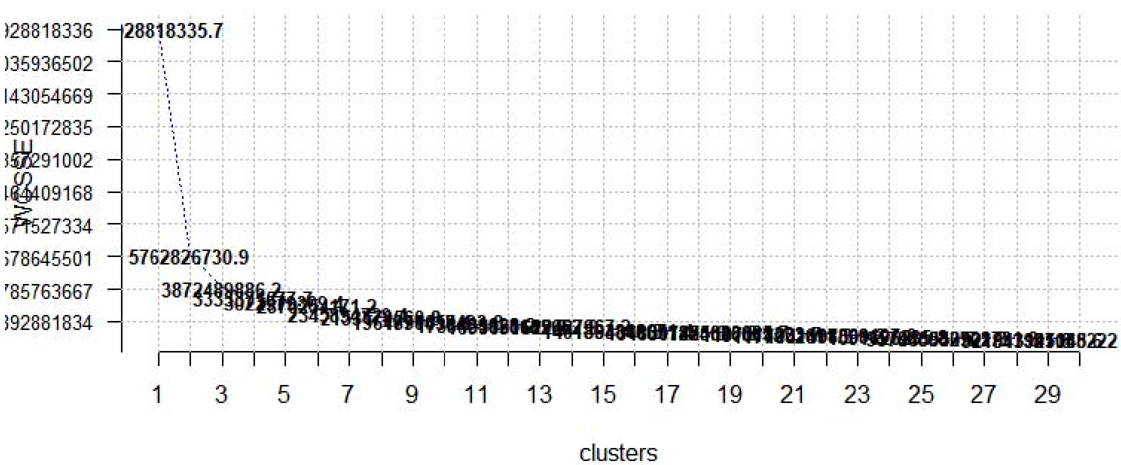

AIC
33497.0
30147.3
26797.6
23447.9
20098.2
16748.5
clusters
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
33497
30133.4
28270.2
26599.9
25341
24128.4
22081.3

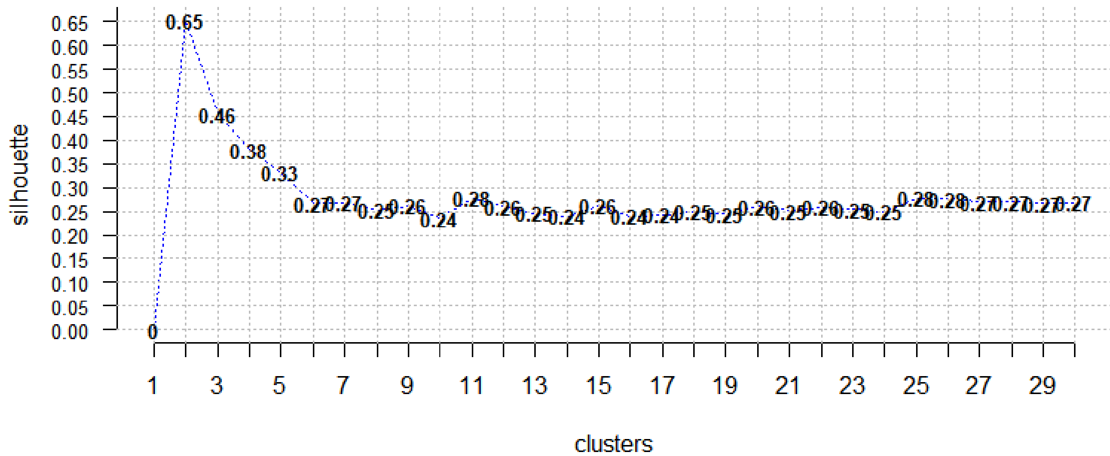

BIC
clusters
33697.0
30327.3
26957.6
23587.9
33697.2
30533.9
28870.9
27400.9
26342.2
25329.8
24803.8
24303.6
23883.5
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29

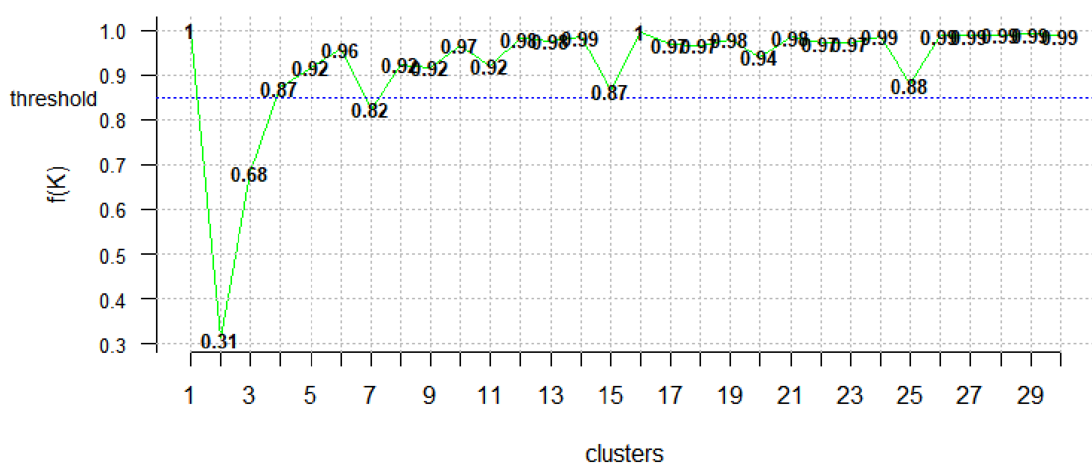

Adjusted_Rsquared
clusters
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
0.1
0.16
0.21
0.25
0.29
0.31
0.34
0.36
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29

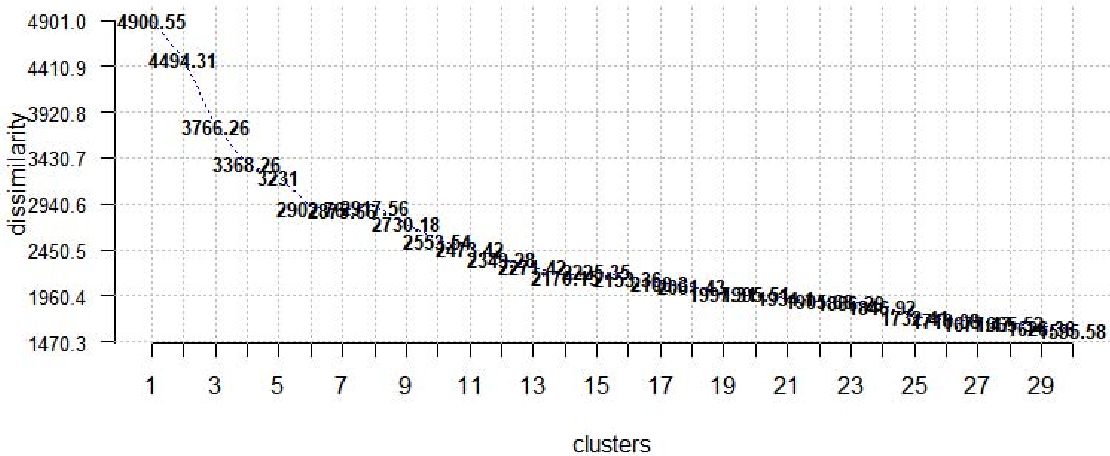
Number of cluster testing after subjective decimal transformation:

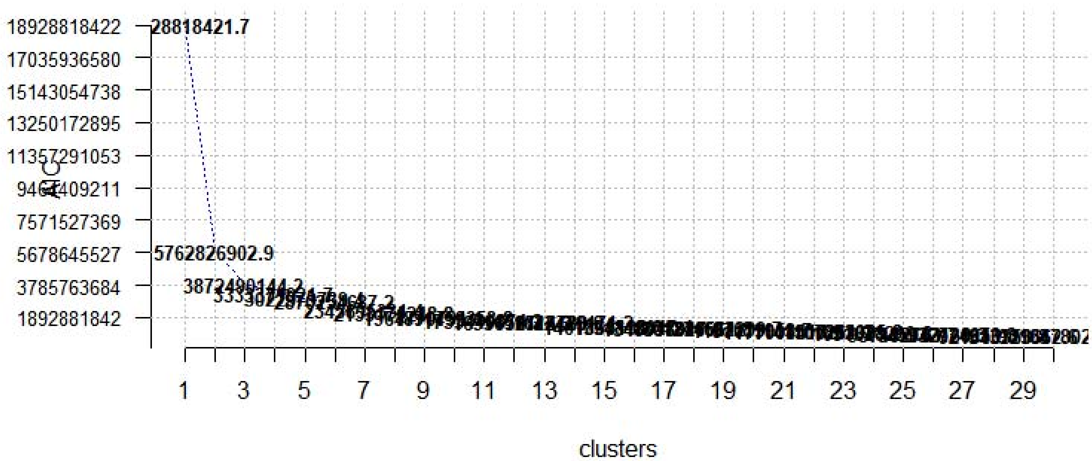


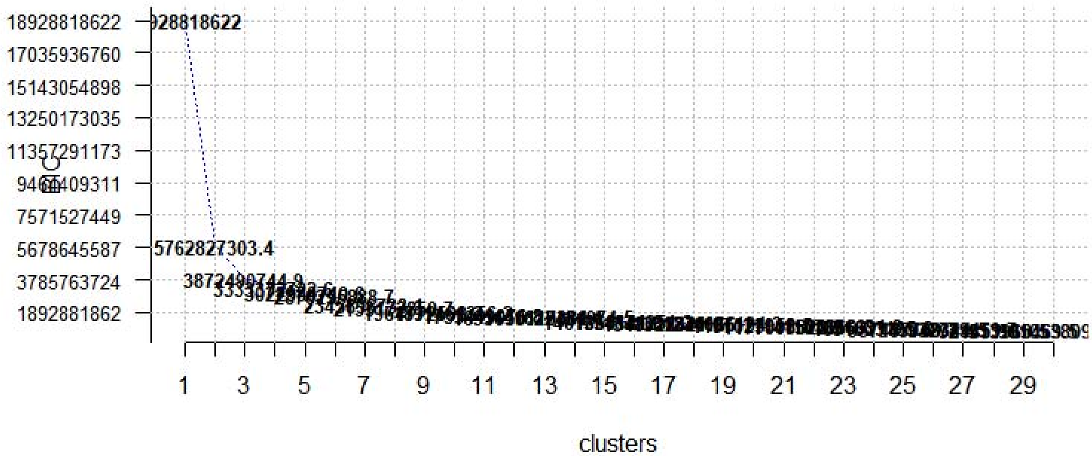



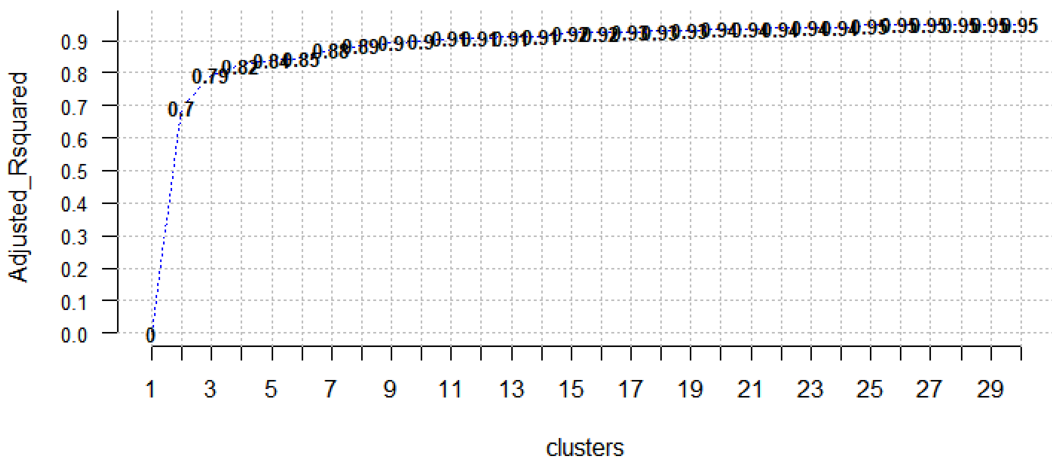

silhouette
clusters


threshold
f(K)
clusters


dissimilarity
4901.0
4410.9
3920.8
3430.7
2940.6
2450.5
1960.4
1470.3
4900.55
4494.31
3766.26
3368.26
3231
2730.18
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29
clusters


AIC
18928818422
17035936580
15143054738
13250172895
11357291053
7571527369
5678645527
3785763684
1892881842
28818421.7
5762826902.9
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29
clusters


18928818622
17035936760
15143054898
13250173035
11357291173
7571527449
5678645587
3785763724
1892881862
5762827303.4
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29
clusters


Adjusted_Rsquared
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
0.7
0.79
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29
clusters

# APPENDIX 5
# Clustered Cities. Complete list

**C. CITIES**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **1** | Remscheid | Krefeld | Mönchengladbach | Chemnitz | Solingen | Neumünster | Bielefeld | Wuppertal |
| | Šiauliai | Alytus | Klaipeda | Varese | Bergamo | Busto Arsizio | Como | Massa |
| | Kortrijk | Leuven | Alphen aan den Rijn | Hilversum | Amersfoort | Martigues | Lens - Liévin | Hénin - Carvin |
| | Annemasse | Bradford | Warrington | North East Lincolnshire | Reading | Sefton | Thanet | Warwick |
| | St. Helens | Torbay | Southampton | Wirral | Crawley | Dundee City | Mansfield | Kirklees |
| | Portsmouth | Bracknell Forest | Liverpool | Sheffield | Hyndburn | Hartlepool | Sunderland | Milton Keynes |
| | Rotherham | Luton | Plymouth | Chesterfield | Eastbourne | Guildford | Stevenage | Gloucester |
| | Burnley | Worcester | Halton | Medway | Redditch | Worthing | Swindon | Northampton |
| | Oxford | Blackburn with Darwen | | | | | | |
| **2** | Sanlúcar de Barrameda | Línea de la Concepción, La | Tromsø | Bergen | Stavanger | | | |
| **3** | Chorzów | Monza | Leganés | Coslada | Alcorcón | Prat de Llobregat, El | Sabadell | Fuenlabrada |
| | Torrejón de Ardoz | L'Hospitalet de Llobregat | Mollet del Vallès | Parla | Cornellà de Llobregat | Badalona | Saint-Quentin en Yvelines | CA Europ' Essonne |
| | Mantes en Yvelines | CA du Val d'Irge | CC des Coteaux de la Seine | Argenteuil - Bezons | CA Brie Francilienne | CC de l'Ouest de la Plaine de France | CA de la Vallée de Montmorency | CA Val de France |
| | CA des Lacs de l'Essonne | CA de Seine Essonne | CA les Portes de l'Essonne | CA Val et Forêt | Cergy-Pontoise | CA du Plateau de Saclay | CA le Parisis | CA Marne et Chantereine |
| | Marne la Vallée | CA Sénart - Val de Seine | CA du Val d'Yerres | Versailles | Sénart en Essonne | CA des deux Rives de la Seine | Evry | CC de la Boucle de la Seine |
| | Islington | Hackney | Haringey | Croydon | Hounslow | Woking | Basildon | Ealing |
| | Bromley | Slough | Tower Hamlets | Newham | Greenwich | Lewisham | Kensington and Chelsea | Wandsworth |
| | Redbridge | Southend-on-Sea | Sutton | Barnet | Hammersmith and Fulham | Lambeth | Kingston upon Thames | Bexley |
| | Enfield | Richmond upon Thames | Hillingdon | Manchester | Barking and Dagenham | Gravesham | Westminster | Merton |
| | Waltham Forest | Brent | Southwark | Harrow | Harlow | Havering | Camden | St Albans |
| | Basel | Genève | | | | | | |
| **4** | Leipzig | Reutlingen | Mülheim a.d.Ruhr | Recklinghausen | Bochum | Flensburg | Lübeck | Braunschweig |
| | Wolfsburg | Sankt Augustin | Bonn | Kiel | Ludwigsburg | Göttingen | Saarbrücken | Siegen |
| | Rosenheim | Erlangen | Oldenburg (Oldenburg) | Aschaffenburg | Speyer | Karlsruhe | Ludwigshafen am Rhein | Essen |
| | Köln | Herne | Bottrop | Leverkusen | Bergisch Gladbach | Kempten (Allgäu) | Augsburg | Neu-Ulm |
| | Oberhausen | Düsseldorf | Gelsenkirchen | Trier | Hannover | Dortmund | Darmstadt | Stuttgart |
| | Wilhelmshaven | Kassel | Münster | Wiesbaden | Mannheim | Frankenthal (Pfalz) | Nürnberg | Bremerhaven |
| | Dresden | Bamberg | Bremen | Lüneburg | Osnabrück | Friedrichshafen | Witten | Neuss |
| | Esslingen am Neckar | Offenburg | Fürth | Heilbronn | Hildesheim | Hamm | Koblenz | Frankfurt am Main |
| | Freiburg im Breisgau | Fulda | Regensburg | Mainz | Landshut | Ingolstadt | Hagen | Paderborn |
| | Ulm | Heidelberg | Hanau | Würzburg | Schweinfurt | Bayreuth | Kaiserslautern | Gießen |
| | Iserlohn | Duisburg | Offenbach am Main | Moers | Aachen | Sindelfingen | Bologna | Venezia |
| | Padova | Pordenone | Treviso | Vicenza | Brescia | Torino | Verona | Udine |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Firenze | Livorno | Prato | Antwerpen | Charleroi | Brugge | Gent | Liège |
| | Oostende | Mons | Zoetermeer | Breda | Hengelo | Hoorn | Purmerend | Almere |
| | Haarlemmermeer | Enschede | Zaanstad | Nijmegen | Zwolle | Tilburg | Venlo | Almelo |
| | Maastricht | Leeuwarden | Roosendaal | Helmond | Groningen | 's-Hertogenbosch | Bergen op Zoom | Odense |
| | Århus | Getafe | Pozuelo de Alarcón | Caen | Grenoble | Nantes | Reims | Angoulème |
| | Bayonne | Toulouse | Roanne | Brest | Melun | Pau | Bordeaux | Nancy |
| | Troyes | Tours | Cherbourg | Meaux | Annecy | Dijon | Creil | Orléans |
| | Marseille | Strasbourg | Lille | Toulon | Le Mans | Tarbes | Clermont-Ferrand | Montbelliard |
| | Le Havre | Bolton | Preston | Glasgow | Trafford | Hastings | Blackpool | Newcastle upon Tyne |
| | Coventry | Oldham | Nuneaton and Bedworth | North Lanarkshire | Middlesbrough | Bury | Rochdale | Aberdeen |
| | Bournemouth | Exeter | Gateshead | Dudley | North Tyneside | Newport | Norwich | Leicester |
| | Stockport | Kingston-upon-Hull | Ipswich | Brighton and Hove | Wakefield | Leeds | Stoke-on-trent | Cheltenham |
| | Wolverhampton | Tamworth | Poole | Thurrock | Cambridge | Cardiff | Stockton-on-Tees | Lincoln |
| | Tameside | Edinburgh | Derby | Bristol | Wigan | South Tyneside | Birmingham | Solihull |
| | Salford | Sandwell | Nottingham | Walsall | Göteborg | Malmö | Cork | Waterford |
| | Galway | Limerick | Turku / Åbo | Espoo / Esbo | Vantaa / Vanda | Helsinki / Helsingfors | Ljubljana | Lausanne |
| | Luzern | St. Gallen | Biel/Bienne | Bern | Winterthur | Zürich | | |
| | | | | | | | | |
| **5** | München | Berlin | Hamburg | Riga | Warszawa | Sofia | Roma | Milano |
| | Bruxelles / Brussel | København | Barcelona | Madrid | Paris | Lyon | Oslo | Dublin |
| | Budapest | Praha | Wien | | | | | |
| | | | | | | | | |
| **6** | Narva | Genova | Palermo | Napoli | Bari | Pescara | Messina | Cagliari |
| | Giugliano in Campania | Catania | Caserta | Elche/Elx | Alicante/Alacant | Viladecans | Avilés | Gijón |
| | Burgos | Vigo | Santa Coloma de Gramenet | Alcalá de Henares | Pontevedra | Granollers | Valencia | Ponferrada |
| | Mataró | Reus | Torrevieja | Ferrol | Zamora | Barakaldo | Sant Cugat del Vallès | Marbella |
| | Tarragona | Palma de Mallorca | Granada | Torremolinos | Bilbao | Majadahonda | Pamplona/Iruña | Palencia |
| | Benidorm | Rubí | San Fernando | Manresa | Fuengirola | Girona | Melilla | Irun |
| | Terrassa | Castelldefels | Logroño | Santander | Alcobendas | Elda | Zaragoza | Castellón de la Plana/Castelló de la Plana |
| | Ceuta | Ourense | Sevilla | Cádiz | Móstoles | Getxo | Algeciras | Málaga |
| | Sant Boi de Llobregat | Puerto de Santa María, El | Oviedo | Salamanca | Valladolid | Talavera de la Reina | A Coruña | San Sebastián/Donostia |
| | Gandia | Huelva | León | Vilanova i la Geltrú | Cerdanyola del Vallès | Nice | Lugano | |
| | | | | | | | | |
| **7** | Rostock | Weimar | Stralsund | Greifswald | Dessau-Roßlau | Passau | Marburg | Jena |
| | Magdeburg | Tübingen | Brandenburg an der Havel | Schwerin | Pforzheim | Görlitz | Zwickau | Halle an der Saale |
| | Cottbus | Neubrandenburg | Gera | Plauen | Erfurt | Wetzlar | Konstanz | Frankfurt (Oder) |
| | Potsdam | Tartu | Tallinn | Daugavpils | Liepaja | Jelgava | Sosnowiec | Zory |
| | Konin | Czestochowa | Poznan | Radom | Kielce | Szczecin | Grudziadz | Leszno |
| | Katowice | Bytom | Zabrze | Gdansk | Nowy Sacz | Lublin | Bialystok | Walbrzych |
| | Zamosc | Piotrków Trybunalski | Suwalki | Koszalin | Ruda Slaska | Gdynia | Tychy | Wloclawek |
| | Przemysl | Plock | Chelm | Kraków | Gliwice | Rybnik | Lódz | Elblag |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Siedlce | Bydgoszcz | Olsztyn | Tarnów | Kalisz | Wroclaw | Bielsko-Biala | Zielona Góra |
| | Jelenia Góra | Lomza | Slupsk | Jastrzebie-Zdrój | Gorzów Wielkopolski | Opole | Torun | Legnica |
| | Rzeszów | Sliven | Pleven | Stara Zagora | Veliko Tarnovo | Shumen | Varna | Vidin |
| | Pazardzhik | Blagoevgrad | Pernik | Haskovo | Yambol | Plovdiv | Ruse | Dobrich |
| | Vratsa | Kaunas | Panevežys | Vilnius | Trieste | Miskolc | Gyor | Nyíregyháza |
| | Szombathely | Kecskemét | Pécs | Hradec Králové | Liberec | Havírov | Ostrava | Zlín |
| | Plzen | Chomutov-Jirkov | Kladno | Ústí nad Labem | Brno | Most | Karviná | Jihlava |
| | Ceské Budejovice | Olomouc | Karlovy Vary | Pardubice | Maribor | Trencín | Nitra | Prešov |
| | Košice | Žilina | Trnava | Banská Bystrica | Bratislava | | | |
| | | | | | | | | |
| **8** | Celle | Villingen-Schwenningen | Salzgitter | Burgas | Foggia | Reggio di Calabria | Modena | Sassari |
| | Pesaro | Piacenza | Latina | Cremona | Savona | Campobasso | Reggio nell'Emilia | Asti |
| | Ancona | Bolzano | Catanzaro | Pavia | Taranto | Lecce | Pisa | Cosenza |
| | Trento | Perugia | Matera | Potenza | Barletta | Ravenna | Terni | Ferrara |
| | La Spezia | Parma | Siracusa | Forlì | Lecco | Avellino | Novara | Salerno |
| | Acireale | Namur | Deventer | Apeldoorn | Lelystad | Aalborg | Santiago de Compostela | Cáceres |
| | Guadalajara | Lugo | Albacete | Badajoz | Ciudad Real | Jaén | Murcia | San Sebastián de los Reyes |
| | Córdoba | Lleida | Dos Hermanas | Cartagena | Toledo | Jerez de la Frontera | Vitoria/Gasteiz | Rozas de Madrid, Las |
| | Montpellier | Aubagne | Angers | Saint-Quentin | Beauvais | Arras | Evreux | Bourges |
| | Metz | Béziers | Niort | Aix-en-Provence | Rennes | Rouen | Nimes | Châteauroux |
| | Boulogne-sur-mer | CA de Sophia-Antipolis | Chambery | Saint-Brieuc | Quimper | Valence | Valenciennes | Chartres |
| | La Rochelle | Chalon-sur-Saône | Besançon | Mulhouse | Belfort | Compiègne | Albi | Perpignan |
| | Brive-la-Gaillarde | Lorient | Charleville-Mézières | Limoges | Vannes | Dunkerque | Avignon | Fréjus |
| | Saint-Etienne | Saint-Nazaire | Châlons-en-Champagne | Colmar | Poitiers | Calais | Douai | Ajaccio |
| | Amiens | Trondheim | Kristiansand | Wrexham | Tunbridge Wells | Ashford | East Staffordshire | Waveney |
| | Darlington | Maidstone | Peterborough | Basingstoke and Deane | Swansea | Falkirk | Wycombe | Carlisle |
| | Barnsley | Telford and Wrekin | York | Doncaster | Dacorum | Great Yarmouth | Chelmsford | Bedford |
| | Bath and North East Somerset | Colchester | Newcastle-under-Lyme | Cheshire West and Chester | Cannock Chase | Umeå | Jönköping | Jyväskylä |
| | Tampere / Tammefors | | | | | | | |

## APPENDIX 6
## Cluster Z-score Means for Physical Parameters

| Clst | Capital | KERNEL | AREAc | AREAf | cPOP | PBAc | DIScity | w1DISc | Upc |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.15575 | -0.16225 | -0.24247 | -0.76692 | -0.14109 | 0.157002 | 0.180363 | 0.214432 | 0.14916 |
| 2 | -0.15575 | -0.38704 | 1.997119 | -0.17577 | -0.35756 | -0.55374 | -1.0886 | -1.08651 | -0.5724 |
| 3 | -0.15575 | 1.168996 | -0.55664 | 1.993389 | -0.12035 | 1.337536 | 0.965602 | 0.878962 | 1.354547 |
| 4 | -0.09864 | -0.0129 | -0.1572 | -0.13358 | 0.020746 | 0.255462 | 0.451282 | 0.462202 | 0.255647 |
| 5 | 4.683568 | 0.393833 | 1.09739 | 1.303331 | 4.649257 | 0.993423 | 1.199643 | 1.097689 | 1.022098 |
| 6 | -0.15575 | -0.01143 | -0.30684 | -0.41029 | -0.04825 | 0.063288 | -0.24003 | -0.22844 | 0.042569 |
| 7 | -0.05699 | -0.23087 | -0.24919 | -0.18427 | -0.22828 | -0.44739 | -0.22979 | -0.19059 | -0.45043 |
| 8 | -0.15575 | -0.38704 | 0.85759 | -0.3385 | -0.26183 | -0.96827 | -1.1221 | -1.13491 | -0.96366 |

| | w2UDc | LUPc | WUPc | POPf | AREA CinF | POPpropCtoF | POPch | PROPch | UMZtoC |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.267357 | 0.12023 | 0.4201 | -0.47104 | 1.68138 | 1.079502 | 0.008188 | 0.243852 | -0.11906 |
| 2 | 0.190702 | -0.05487 | -0.50777 | -0.51303 | 0.840529 | 0.999298 | 1.379853 | -0.55096 | -0.34441 |
| 3 | -0.88966 | -0.76009 | 0.024356 | 2.453128 | -0.25215 | -1.41328 | 0.653726 | 0.291144 | 2.135239 |
| 4 | 0.532063 | 0.419286 | 0.737066 | -0.19115 | -0.34923 | -0.38495 | -0.06587 | 0.177801 | -0.22991 |
| 5 | -1.08142 | -0.94889 | -0.30672 | 0.421571 | -0.54644 | 0.155775 | 0.17031 | -0.17733 | -0.19749 |
| 6 | -1.34416 | -1.11737 | -0.89282 | -0.15819 | 0.254612 | 0.274854 | 0.147125 | -0.70943 | -0.19163 |
| 7 | -0.02255 | -0.04231 | -0.2711 | -0.42062 | -0.59767 | 0.144937 | -0.71648 | 0.127673 | -0.329 |
| 8 | 0.425601 | 0.48386 | -0.58133 | -0.43523 | 0.383346 | 0.58092 | 0.238187 | -0.24602 | -0.3447 |

| | UMZtoF | AVEwTEMP | AVEcTEMP | Rain | Agriculturef | ContUrbc | ContUrbf | DiscontUrbf | DiscontUrbc |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.428554 | -0.56617 | 0.173295 | 0.373596 | 0.432925 | -0.14822 | 0.028102 | 1.86994 | 0.221262 |
| 2 | -0.18855 | -0.53362 | 0.220168 | 2.636145 | 0.802107 | 0.150357 | 1.943737 | -0.18839 | -0.12466 |
| 3 | -0.08753 | -0.33382 | 0.572719 | 0.159436 | -0.17782 | 0.335457 | 0.146545 | 0.382605 | 0.439954 |
| 4 | -0.08309 | -0.27627 | -0.10855 | 0.288623 | -0.08234 | -0.15207 | -0.27959 | -0.08084 | 0.044883 |
| 5 | -0.13783 | 0.331663 | -0.37627 | -0.3835 | -0.12656 | 0.43514 | -0.03501 | -0.08266 | 0.399281 |
| 6 | -0.09946 | 0.944163 | 0.905476 | -0.02905 | 0.179481 | 0.744496 | 1.469919 | -0.26365 | -0.17728 |
| 7 | -0.22554 | -0.21319 | -1.23443 | -0.72366 | -0.08491 | -0.20584 | -0.35155 | -0.39034 | -0.10303 |
| 8 | -0.18331 | 0.540784 | 0.450074 | -0.07908 | 0.003124 | -0.16004 | -0.19302 | -0.42028 | -0.27763 |

| | Industrialc | Industrialf | wetlandsf | Vegetationf | Vegetationc | GreenUrbc | BAREc | BAREf | Forestf |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.07568 | 1.424698 | 0.766426 | -0.23047 | -0.15647 | 0.163615 | 0.03672 | -0.0366 | -0.17713 |
| 2 | -0.3398 | 0.320505 | 0.433718 | 0.971929 | 0.412325 | -0.48481 | 7.688586 | 9.853515 | -0.08495 |
| 3 | 0.478626 | 0.03099 | -0.10048 | -0.27404 | -0.25104 | 1.185344 | -0.18414 | -0.15167 | -0.1597 |
| | 0.132228 | -0.0206 | 0.050691 | -0.23587 | -0.1859 | 0.04304 | -0.11583 | -0.10539 | -0.11883 |
| 4 | 0.345848 | 0.114239 | -0.1197 | -0.17277 | -0.15994 | 0.891734 | -0.11376 | -0.12222 | 0.077915 |
| 5 | 0.382366 | 0.622527 | -0.05473 | 1.558315 | 1.239683 | -0.34942 | 0.251632 | 0.176702 | 0.240658 |
| 6 | -0.0532 | -0.548 | -0.25132 | -0.23275 | -0.15726 | -0.12286 | -0.15299 | -0.15915 | 0.09374 |
| 7 | -0.6391 | -0.49588 | -0.11137 | -0.01238 | -0.01637 | -0.61014 | 0.024171 | -0.00445 | 0.135047 |
| 8 | -0.07568 | 1.424698 | 0.766426 | -0.23047 | -0.15647 | 0.163615 | 0.03672 | -0.0366 | -0.17713 |

| | Soviet | Coastal | NoHH | NoDwell | DWELLperHH | CarsPerpop | POLY |
|---|---|---|---|---|---|---|---|
| 1 | -0.32661 | -0.09906 | -0.15328 | -0.16533 | -0.22673 | -0.09609 | 0.130897 |
| 2 | -0.48149 | 1.409102 | -0.32578 | -0.30651 | 0.720625 | -0.21058 | -2.74912 |
| 3 | -0.45033 | 0.453492 | -0.19471 | -0.20471 | -0.1363 | -0.59744 | -0.10961 |
| 4 | -0.44816 | -0.11942 | 0.064117 | 0.046081 | -0.24483 | -0.22941 | 0.44628 |
| 5 | 0.325545 | 0.07152 | 4.655339 | 4.69113 | -0.06286 | -0.23181 | -0.20645 |
| 6 | -0.44914 | 0.712092 | -0.13312 | -0.0646 | 1.233348 | 0.344789 | -1.51057 |
| 7 | 1.958839 | -0.51765 | -0.21852 | -0.2328 | -0.20694 | -0.16831 | 0.431812 |
| 8 | -0.46387 | 0.006947 | -0.24764 | -0.23747 | 0.069878 | 0.749665 | -0.15671 |