



Introduction to Spatial Exploration of Economic Data

11

Vít Pászto

Abstract

In the introductory chapter, firstly, the summary of how this chapter evolved is provided as well as the organisation of the whole Part III of this book is described. Then, the main body serves as an overview of how the spatial exploration of economic data and respective methods of interdisciplinary analytics can be approached. Based on the authors experiences up to this date, five levels or stages of an analytical approach to spatially analyse economic data are defined. At the lowest level, the author focuses on a “simple” visualisation of data, following by the level of merging (multivariate) statistics of economic data with their spatial component. In the middle, as a third level, spatial statistics – as an implicit use of statistics in the spatial analysis – is mentioned. As the fourth stage, a workflow of the previous ones is depicted, which leads to the final and most advanced level of spatial-economic modelling. This chapter strives to define a universal workflow in economic data analysis. The conceptual framework introduced in this chapter is based on 3 years of interdisciplinary cooperation within the Spationomy project.

Keywords

Geovisual analysis · Spatial statistics · Exploratory analysis · Spatial modelling

11.1 Introduction

This chapter has been formerly work-titled as a “methods of interdisciplinary analytics”, which turned out to be a rather ambitious plan as it might take a whole book to write about methods. In this book, we talk about a fusion of several distinct fields – geoinformatics/geomatics, geography, spatial analysis, geovisual tools and (geo) visualisation on one side, on the other, we refer to economy, business, business informatics, economic data and quantitative methods to work with them, and also about a management. In the Spationomy project, to simplify the mixture of disciplines, we call the former disciplines and people (staff and students) simply as “geo” part; keeping the same logic, the latter (disciplines and people) had a label as “eco”. As the “eco” label may be confused with “ecology”, we lately re-branded the label to “business” part. Anyway, each one above mentioned has a broad theoretical framing, old concepts, methodologies, and contemporary issues to deal with. It will be almost impossible to capture every aspect of these disciplines in on coherent text, and it was not intended at all. However, in previous parts of the book (Parts I and II), we provide a comprehensive overview of the subjects’ bases. The Parts

V. Pászto (✉)

Department of Informatics and Applied Mathematics,
Moravian Business College Olomouc, Olomouc, Czech
Republic

Department of Geoinformatics, Palacký University
Olomouc, Olomouc, Czech Republic
e-mail: vit.paszto@gmail.com

© The Author(s) 2020

V. Pászto et al. (eds.), *Spationomy*, https://doi.org/10.1007/978-3-030-26626-4_11

235

I and II are meant to be an optimal start for those interested in the spatial economic topics.

Part III is dedicated to examples and case studies on how the “geo” and “business” part can be used jointly. Following chapters illustrate a few application of how common knowledge from geomatics, geography, economy and business informatics could be used in practice and research. Firstly, in Chap. 12, an interesting fusion of geospatial tools (mainly for presenting data) with purely business and managerial needs in a water management company is described in detail. This fusion is an example of how both parts (“geo” and “economic”) can be utilised in real-life situations. On the other hand, the case study in Chap. 13, shows an artificial site selection of a fictitious furniture store. This example is a typical process task for a new branch, store, or any facility allocation. In Chap. 14, the issue of demographical development was merged with spatial planning in cities. It represents a standard research paper approach to study given phenomenon, with a unique combination of the main, very actual topic (population ageing), and its spatial pattern in the studied region. Lastly, in Chap. 15, another example of a scientific study is provided. The study aims to capture the spatial implications of a European CO₂ emission trading system over the ten years. Cartographic methods and geovisual analysis are used to (spatially) explore basic environmental and economic data referring to pollution allowances market. The set of overviewing introduction and four different case studies demonstrates how methods of interdisciplinary analytics can be deployed in both real-life situation and research. It brings new knowledge and opens possibilities for novel approaches in the new joint field of the “spationomy”.

Now let’s get back to the “methods of interdisciplinary analytics” or “(spatial) exploration of economic data”. During the 3 years of the Spationomy project, we learnt how beneficial it is to combine ideas, methods, approaches, and topics among (and from) already mentioned disciplines. From various experiences – learning and teaching material creation, joint scientific paper preparations, brainstorming and

discussions – that we had a chance to have in the Spationomy team, and also from the interaction with other stakeholders, we feel the need to propose an optimal workflow of the spatial and economic data analytics. It is also a practice of the author of this chapter to advise students to follow the next five steps towards a successful bachelor or master thesis. In the following pages, a way to approach the data analysis is presented from the simple to more advanced and complex application of spatial and statistical methods. Introduced five levels reflect the authors’ experiences that were proven in practice during the project. However, it is not a dogma that could not be changed, modified or adjusted to the reader’s needs.

11.1.1 Level 1 – (Geo)Visual Analysis

The geovisual analysis represents the first step in the exploration of the data (spatial, economic, business and any other types). The general objective of data visualisation is to transform textual or numerical information into the form of its graphical representation. Whether it is the picture, scheme, chart, graph, workflow, infographics, map, interactive application, 3D graphics or something else, it focuses on a transfer of information to the reader. The visualisation also serves as a tool for data exploration. We can perform simple (and effective at the same time) exploratory analysis, e.g. to find extreme values, outliers, in a data. By depicting a boxplot, scatterplot, or just linear chart (see Part II, Chap. 8), we can immediately see such outliers, which could be hardly detected when “looking” at the numbers. Indeed, the experienced data analyst can find out anomalies in a raw dataset, or when we have a small data sample, it is easy to capture outliers. However, in the case of big data or other highly heterogeneous data, the hidden pattern could be revealed with considerable difficulties, if at all. With the use of visualisation techniques, we can analyse such messy data, we can describe data patterns inside the dataset, uncover and show extreme values, find relationships in data and compare them, and most importantly to communicate the information much more clearly.

Visualisation as the seemingly the simplest level of proposed ways of economic data exploration, however, must also follow some rules and recommendations (for further reference, see Part II). Otherwise, visualisation tools might be misused, which would consequently lead to possible misinterpretation of the graphical representation of data. A great example of the strength and appropriateness of data visualisation is the Anscombe quartet. It is the unique dataset containing four subsets of two-variable data. By calculating mean, variance, the correlation between two variables in each dataset, the same values are returned. Thus, statistically, all four datasets are the same (share identical statistical properties). But if we visualise Anscombe

quartet, we achieve something completely different (Fig. 11.1). Of course, it is vital to be familiar with basic statistical properties of analysed data, but we strongly recommend complementing it with any form of a visualisation.

The first step of data analysis should always include visualisation of the data (besides basic statistics). In other words, level one in the data exploration should be that we take the data and visualise them. At this level, we do not modify, filter, select, or aggregate the data before we visualise it. That is why it is the first and rather straightforward way for data analysis. When we talk about geovisualisation, it is nothing more than using maps as a means of the medium to visualise data. In the case of geovisualisation, we

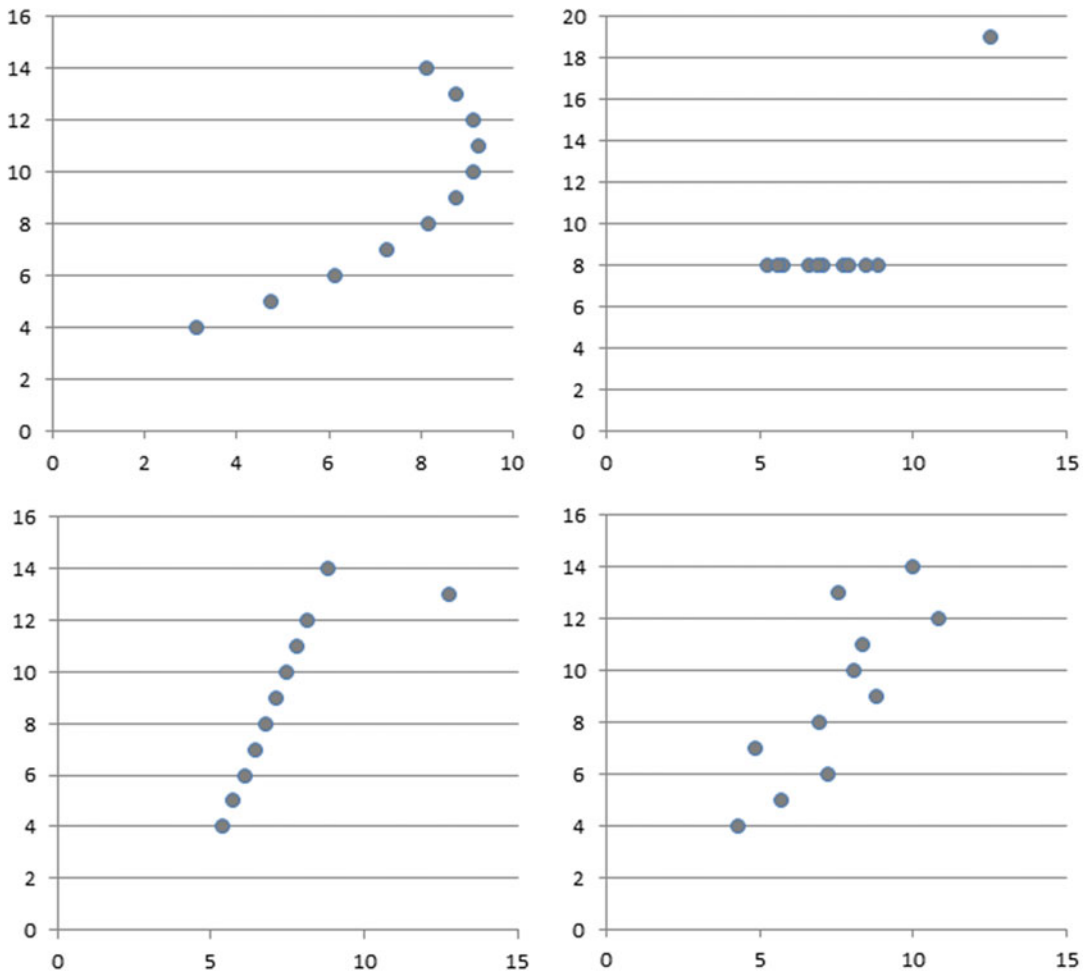


Fig. 11.1 Scatterplots of the Anscombe quartet. (Source: Authors)

have cartography and the rules that we need to follow. At the same time, we should still keep in mind that the information transfer is the primary goal of (geo)visualisation, not blind conformation to the rules. Figure 11.2 serves as an example of geovisualisation, with sample points representing economic subjects displayed on the left (with no attributes reflected), and population density map within administrative units on the right. From both examples in Fig. 11.2, we can easily see otherwise hidden spatial pattern of data.

11.1.2 Level 2 – Statistics, Exploratory Data Analysis and Its (Geo)Visualisation

At this level, all the analytical processes take place outside the GIS environment, or better to say without a spatial component implicitly included in the data. We understand this level to be covering advanced techniques of statistics (e.g. testing hypothesis, multivariate statistics, regression or correlation analysis and such), mathematics. For example, one of the most frequently applied techniques is multivariate statistics. In the field of spatial exploration of economic

data, we commonly work with multiple attributes of geographical units (e.g. regions' GDP, income, unemployment rates, demographic structure and others). Unfortunately, there is a somewhat limited number of implemented tools of multivariate statistics directly in the GIS environment. In other words, we can use more advanced settings, or several more variations of such statistical tools outside the GIS environment (or in the environment directly working with a spatial component of data). That is why it is often better to “run away” from GIS to “normal” statistics or mathematics – and multivariate data analysis is the case. In (spatial) data analysis, we commonly use instruments like Correlation analysis, Factorial analysis, Analysis of Variance (ANOVA), Principle Component Analysis, or clustering methods (see Fig. 11.3). These statistical tools help us to reveal relationships in the data variables, reduce the dimension of data that can be then easier handled in GIS or find groups with similar variable values (characteristics). Non-spatial visualisation of such analyses can help us to understand a data better, and it also serves as the support for data interpretation. Since we usually work with data that are geographically referenced (e.g. to the country level, particular geographic

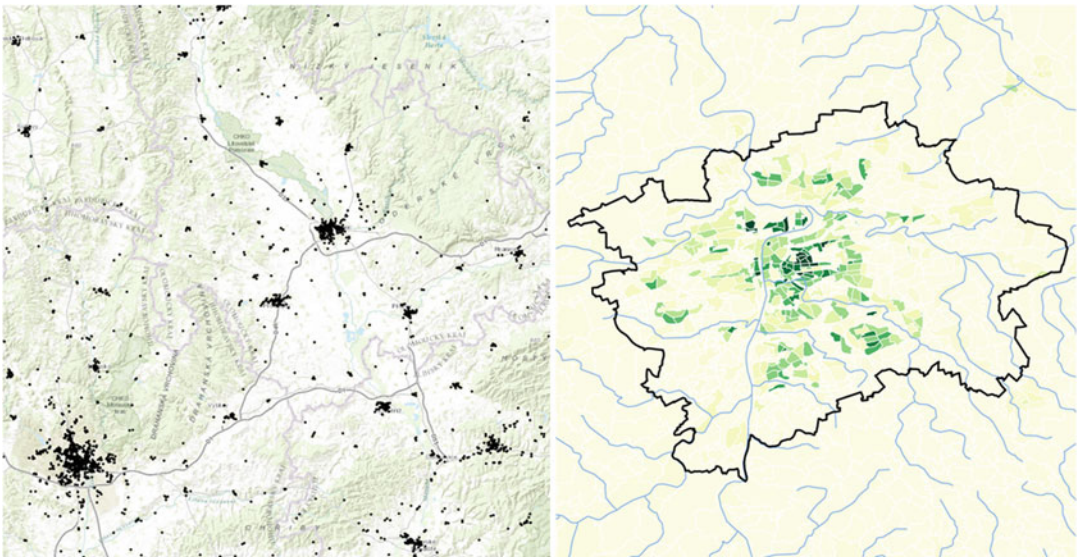


Fig. 11.2 Distribution of economic subjects – sample data (left), population density visualisation in administrative units (right). (Source: Authors)

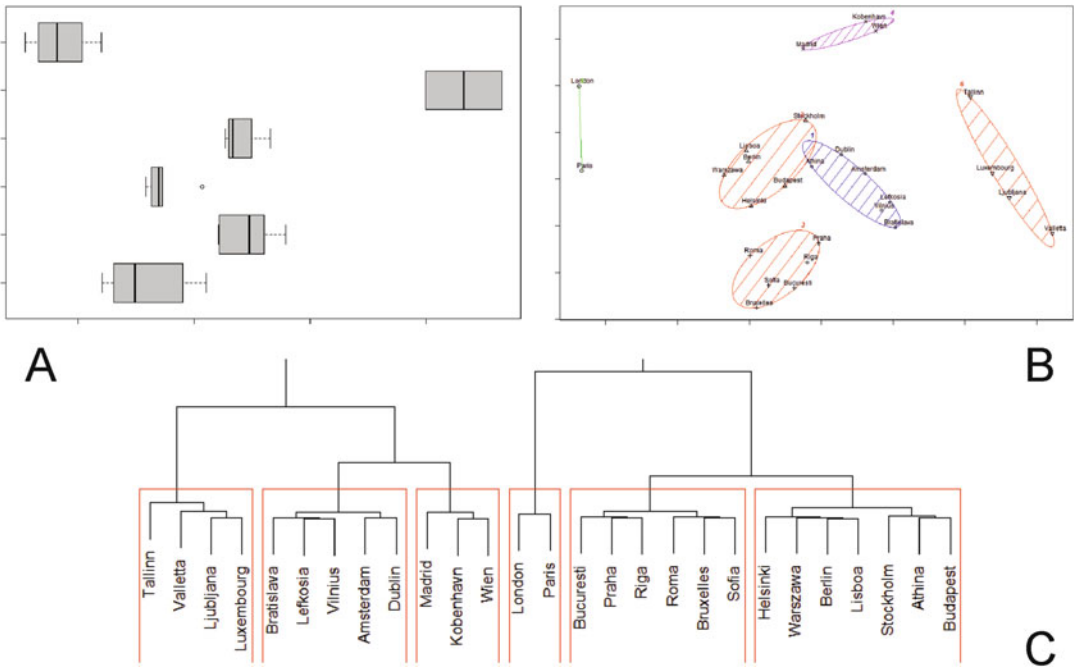


Fig. 11.3 Examples of statistical exploration of data visualised in the form of boxplots (a), hierarchical clustering tree (b), and in-hierarchical clustering (c). (Source: Authors)

region or area, or even to a concrete position – XY geographical coordinates), it is possible to display a data in the form of a map. In these kinds of geovisualisation, however, we need to bear in mind that we depict results of non-spatial analyses spatially. Therefore, no spatial relationships are taken into account during the analysis, and we can “only” observe if data relationships and patterns are also in correspondence in the geographical/spatial context.

11.1.3 Level 3 – Spatial Statistics, Exploratory Spatial Data Analysis and Its (Geo)Visualisation

The level three is referring about types of analyses that use some techniques from the previous part but this time with the spatial component inherently included. Analogically to level two, the methods used for (proper) spatial exploration of data are then labelled as spatial statistics or exploratory spatial data analyses. Examples of

such tools include for instance spatial autocorrelation, Morans’ I, Local Indicators of Spatial Associations, area local or global statistics, geographically weighted regression and others. Spatial statistics and exploratory data analysis help us to examine and measure the geographic distribution of your data, look for global and local outliers, search for global trends, examine local variation or spatial autocorrelation, analyse geographical patterns, mapping clusters, or find a spatial relationship in data. The main advantage of such methods is that they implicitly include the spatial component in the analysis. For instances, if we deal with a point data, the XY coordinates are taken into account. So in the case of a grouping analysis (Fig. 11.4a), points are grouped based on both their attributes and their position. Another example could be a use of neighbouring characteristics of data, meaning that when performing cluster analysis (Fig. 11.4b), a predefined number surrounding polygons/countries are included in clustering to capture also the spatial configuration of data. Figure 11.4 provides examples of (geo)visualisation of some

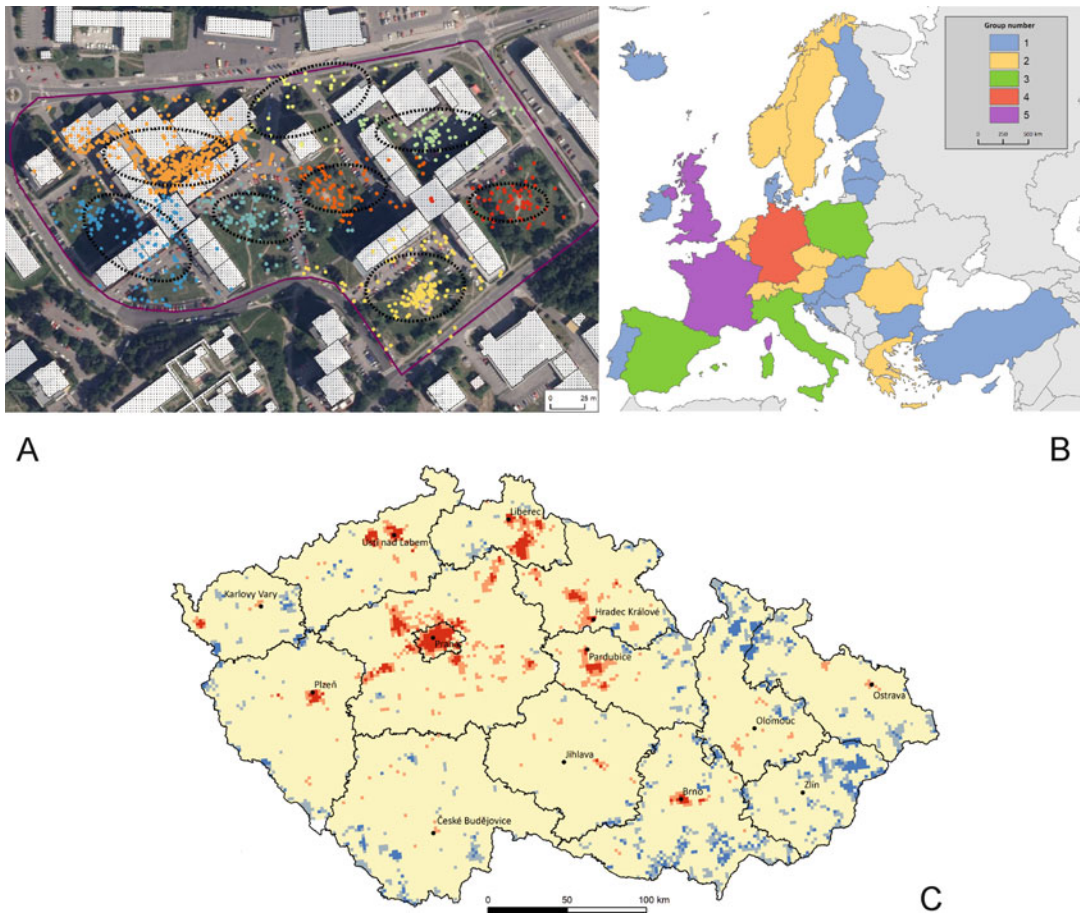


Fig. 11.4 Examples of results from the exploratory spatial data analysis – grouping analysis and standard deviation ellipses (a), spatial clustering (b), and hot-spot analysis (Getis-Ord G_i^*) (c). (Source: Authors)

of the methods – grouping analysis of point patterns with evaluation of their dispersion around their geographical centre using standard deviations ellipses (a), spatial cluster analyses taking neighbourhood proximity measures as a spatial constraint (b), or (c) hot-spot analysis identifying statistically significant spatial clusters of high values (hot spots) and low values (cold spots).

11.1.4 Level 4 – A Combination of Analytical Methods (Level 2 and 3)

Level four represents a fusion of level two and three. Ideally, we should analyse (spatial) data

concerning both of their characteristics – non-spatial and spatial. Therefore, we advise to start with a basic and advanced statistics followed by their non-spatial and spatial visualisation (levels one and two), and then to move on to the solely spatial methods (level three). By a combination of such methods, we can grasp the most important properties of (spatial) data and deliver results, interpretations, and practical implications. Sometimes, it is required to repeat particular steps in the joint analysis since some preliminary or intermediate results might influence the further investigation of the data. For example, we perform correlation analysis to identify redundant variables. Remaining data variables are then inputs for a Principal Component Analysis that is later used for a (spatial) cluster analysis and

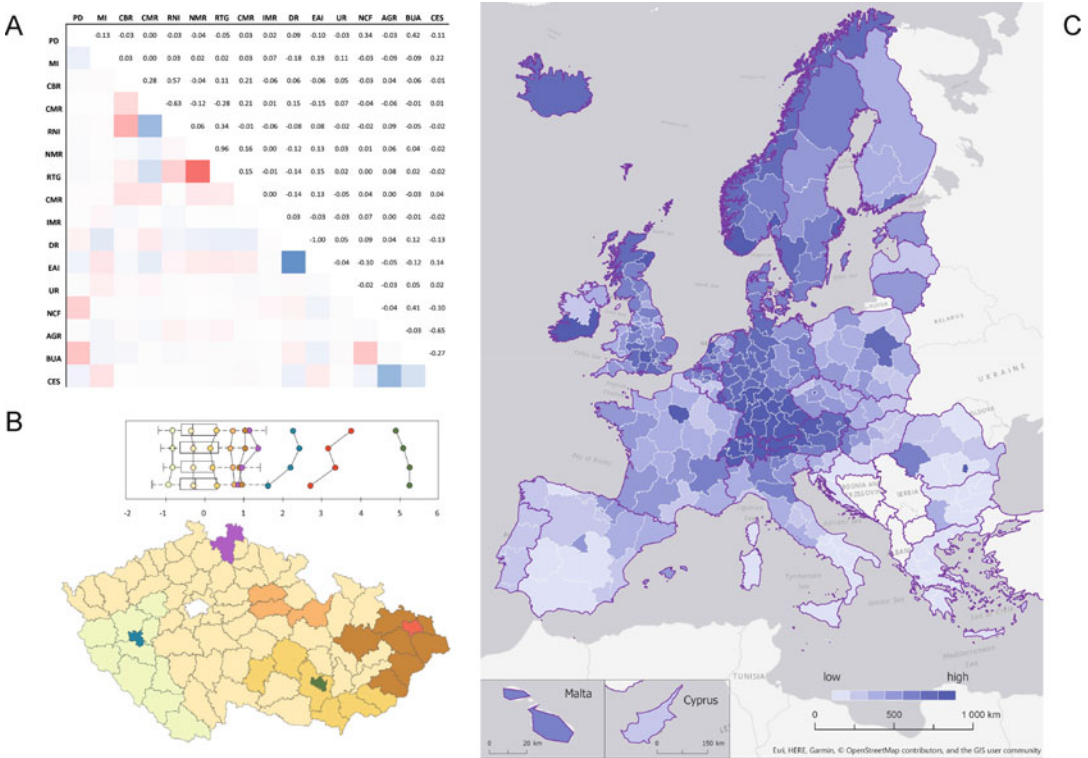


Fig. 11.5 Examples of the combination of analytical methods – correlation matrix (a), boxplots with (spatial) clustering (b), and visualisation of the first component from PCA (Principal Component Analysis) (c). (Source: Authors)

(geo)visualisation. In the end, we might end up with results that need to be validated, modified or adjusted. In this case, we go through the whole cycle again to find the best fit of methods to given data which leads to finer exploratory analysis and understanding of results. This complex process of (spatial) data analysis encompasses a great variety of techniques that are time-consuming, and also demanding as regards the researcher expertise. Thus, interdisciplinary cooperation in forming an expert team is often inevitable. There is no need that one person masters all the skills to perform a complex analytical workflow presented here. To be more specific, examples of individual techniques for (spatial) data analysis is given in Fig. 11.5, where a correlation matrix highlighting variables with low/high values is depicted (a). In Fig. 11.5b, a joint visualisation of spatial clustering in a chart together with boxplot (upper part) and map output (lower) part illustrates an excellent showcase of the combination of methods. Finally, geovisualisation of the first component

from a Principal Component Analysis is depicted in Fig. 11.5c.

11.1.5 Level 5 – (Spatial) Modelling

The highest level of data analysis involves techniques connected with modelling approaches. Again, the modelling part of data analysis can take place outside a “geographical” domain, e.g. mathematical, statistical, machine learning, or another computation is performed first. Then, if possible, such modelling results can be visualised by charts or on a map. The second approach includes modelling directly counting with a spatial component of data. However, this could also be done with no use of a geographical information system. But most of the GIS software offers ways to run modelling within its environment. It must be noted, that such modelling within GIS is more or less a sequence of separate analytical tools connected into a model workflow.

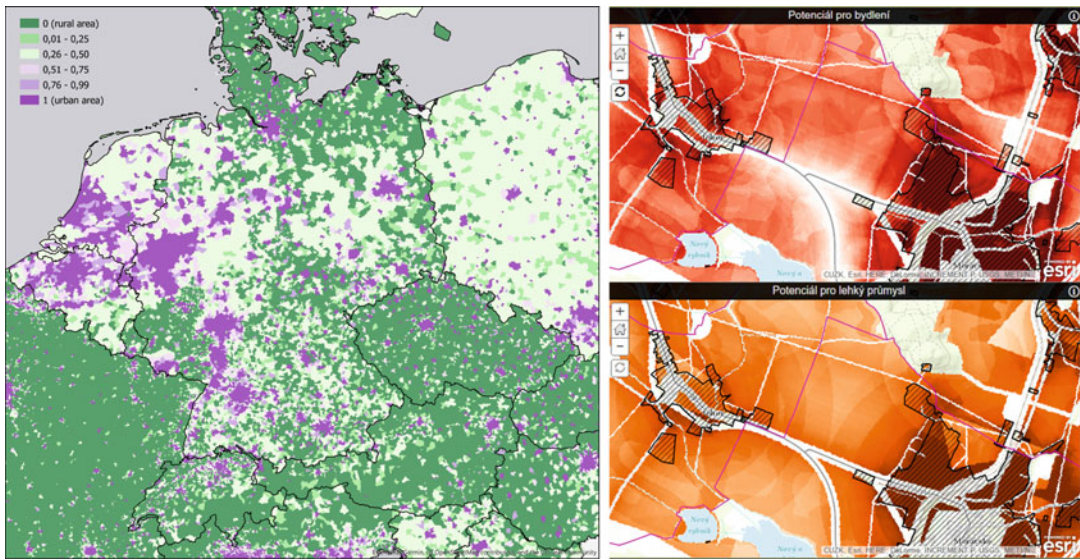


Fig. 11.6 Visualisation made from modelling using fuzzy sets and logic (left), and results from the Urban Planner modelling of a land-use potential (right). (Source: authors)

Complex spatial modelling often requires expert's programming skills and ready-to-go modelling tools are available in the form of plugins or specialised extensions. In Fig. 11.6 (left), there is a geovisualisation of the non-spatial modelling that took place outside GIS by using special fuzzy inference system. On the contrary, Fig. 11.6 (right) shows results from the Urban Planner extension to GIS software used for a land suitability modelling.

11.2 Summary

This chapter addressed an ideal workflow of “methods of interdisciplinary analytics”, but we need to note that it is delivered from the author's practical experiences. Nevertheless, when analysing data, it is advised to proceed from the

simplest to more advanced procedures. That is why, the first step to understand a data is to use proposed level one techniques, i.e. simple (geo) visualisation. Then, level two and three can be applied if we aim to explore specific characteristics of data or to conduct comprehensive statistical or spatial analysis. Ideally, after this stage, a combination of both should follow – again, with respect to the research goals. Finally, a modelling phase can be the concluding step in the whole workflow. As mentioned in a previous text, to follow such workflow, it usually requires cooperation among several experts. Therefore, as synopsis to the very first sentence in this introductory chapter – “methods of interdisciplinary analytics” for (spatial) data exploration is indeed a good label. In the next chapter of this part of the book, one synthetic/artificial and four real-life examples are presented.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

