



Application – Site Analysis Furniture Store

13

Nicolai Moos

Abstract

This chapter deals with the virtual situation that a furniture company is searching for a site for a new furniture store. The user is facing several different datasets (e.g. streets, landuse, existing sites, etc.) with the task to process them with a variety of tools and methods. The decisions on when to use which tool with which parameters are subjected to distinct requirements and restrictions given by chosen conditions on the new site.

The workflow illustrates a simple site analysis with data that can be obtained freely from the web and thereby shows the application and combination of few of the most common tools. By revealing a practical approach on how to use geodata to end with a result layer that only contains features that fulfil all requirements as well as they obey all restrictions, this case study is a vivid example of the calculation of both economic and spatial data.

Keywords

Spatial analysis · Georeferencing · Geoprocessing · Logic operators · Table joins

13.1 Introduction

This chapter is not a classical case study in the narrower sense but gives an applicable and practical example on how the two scientific fields of spatial and economic data acquisition and processing can each be extended and improved by combining them with each other. The simulated workflow is kept basic to focus on the general idea and illustrates a possible routine that can be applied practically to distinct real situations.

It is conceivable that a successful company that produces and sells furniture is looking for a new location for another store in North-Rhine Westphalia (NRW). Several requirements on the new site must be fulfilled, while other parameters need to be excluded. The whole process of combining these inclusions and exclusions is based on spatial data while the thresholds are defined by economical approaches.

For the analysis, the requirements and restrictions for the new site have to be separated from each other, as they have to be processed individually.

Requirements

- maximum distance of 500 m to federal highways
- maximum distance of 1000 m to freeways
- population density of at least 800 inhabitants per square kilometer in respective municipality

N. Moos (✉)
Geography, Geomatics Group, Ruhr-University Bochum,
Bochum, Germany
e-mail: nicolai.moos@rub.de

Restrictions

- must not be in range of 40 km to existing sites
- must not be in settlement areas

The number and structure of requirements and restrictions are only limited by available data. No matter if it is annual income per capita, the prices of certain premises or perfect accessibility to water or electricity – as long as there is the possibility of obtaining data for an issue, it is possible to include the parameters into the analysis.

Given datasets in this exemplary workflow are vector datasets (shapefiles) with municipalities, settlement areas and the street network of NRW, an unreferenced raster map in GIF format and a csv-table with the population data of NRW (for more information on data structure and acquisition, see Chap. 1. Data Sources).

Using this data, the analysis workflow starts with finding the suitable areas that will match the requirements. It then proceeds by finding the unsuitable areas and finally subtracts these areas from the suitable ones to get the result layer that will display all areas matching both, requirements and restrictions.

This analysis can be done in any GIS (ArcGIS, QGIS, MapInfo, etc.). To provide the possibility of a proper replicability, the whole workflow will be processed in QGIS for this case study, as it is open source.

13.2 Project Setup and Data Preparation

At first all datasets that already have a coordinate system need to be uploaded into an empty project with a proper visualization (see Fig. 13.1). The project properties should be checked to make sure that the overall project coordinate system is correct - in this case it is the UTM projection of zone 32 N with the ETRS89 reference frame (EPSG: 25832).

13.2.1 Georeferencing

Once the layer structure is set up the next step is to perform a georeferencing on the raster map that will contain the information about where in NRW

the existing sites of the company are located. A proper georeferencing connects points of the image coordinate system with their geographical coordinates of the given reference system. Either the geographical coordinates are picked out of already referenced material (shapefile, raster image, etc.) which is then called point-to-point-referencing or your unreferenced image data is overlaid by a cartographical grid that defines the coordinates of certain points in the map (point-to-coordinate-referencing). As the unreferenced raster map here covers the area of whole NRW as well as the shapefile of the municipalities the georeferencing can be processed via the point-to-point referencing method, using the georeferencer in QGIS. It is necessary to distribute the connection points homogenously over the whole map to make sure that the it is georeferenced equally accurate, as an algorithm is interpolating the coordinates for the rest of the map according to the chosen points. Clustering these points in a certain area will affect results in a negative way. As image coordinates are transferred to geographical coordinates, the order of setting the points is always first point on the unreferenced map, second point on the referenced material (Fig. 13.2).

After at least four well-distributed control points are set and the mean error is in an acceptable range, the software interpolates the coordinates for the rest of the referenced raster image and saves it as a new and overall referenced dataset (for more information on coordinate systems and projections see Part I, Sect. 1.1.2. Spatial Data Models).

13.2.2 Digitizing

After the raster map has geographic coordinates, the locations need to be extracted out of the map by digitizing the coordinates in a point shapefile as until now the locations are only pixels in a continuous dataset (see Part I, Chap. 1. Data Sources). To do this, a new point layer has to be created that has both the corresponding coordinate system and the necessary attribute table field to store the name information of the city where the existing sites are located (Fig. 13.3).

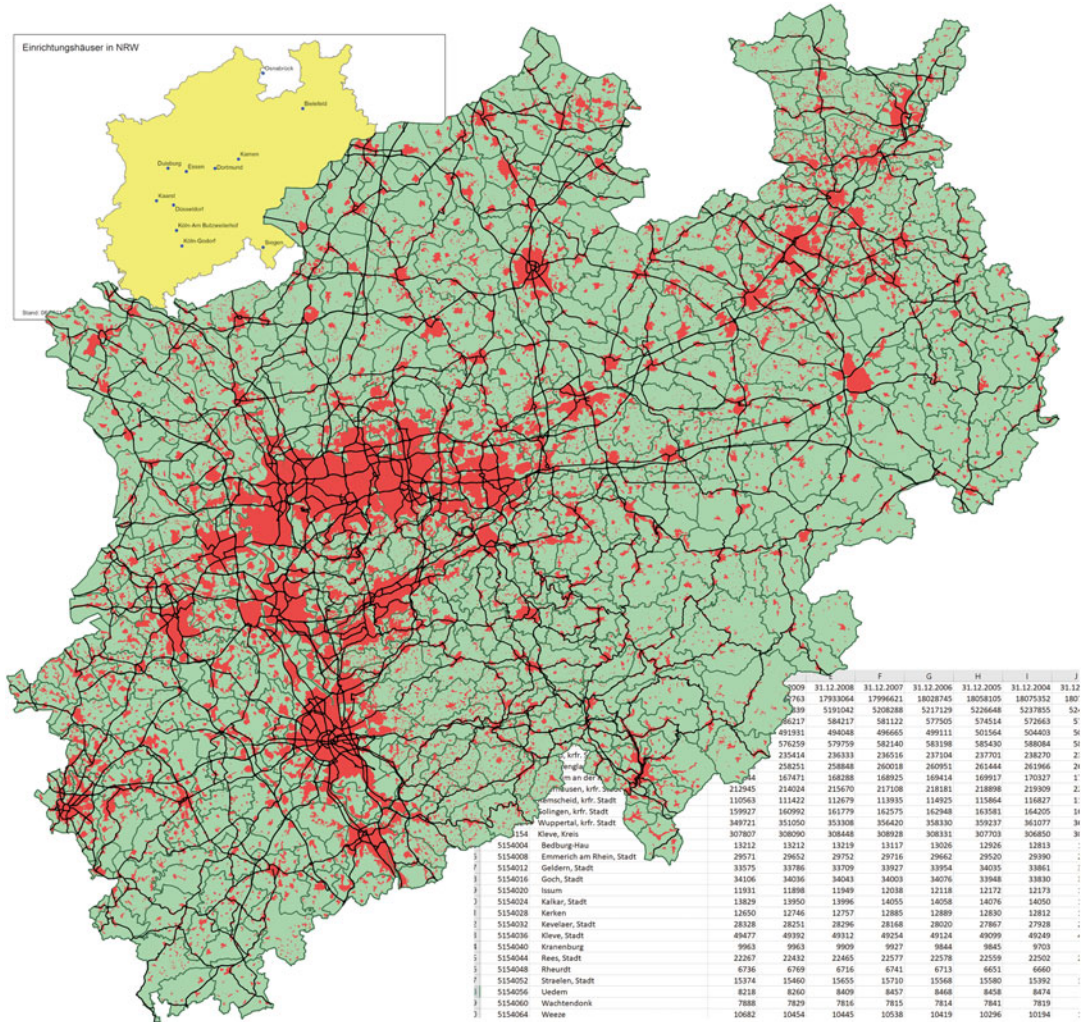


Fig. 13.1 Feature layer municipalities NRW (green), settlement areas (red), streets (black), raster map (top left), csv-table (bottom right)

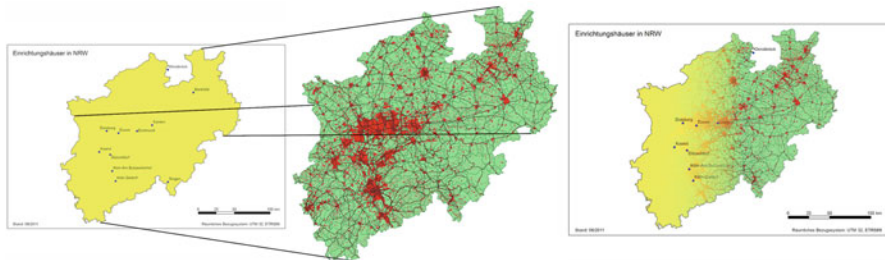
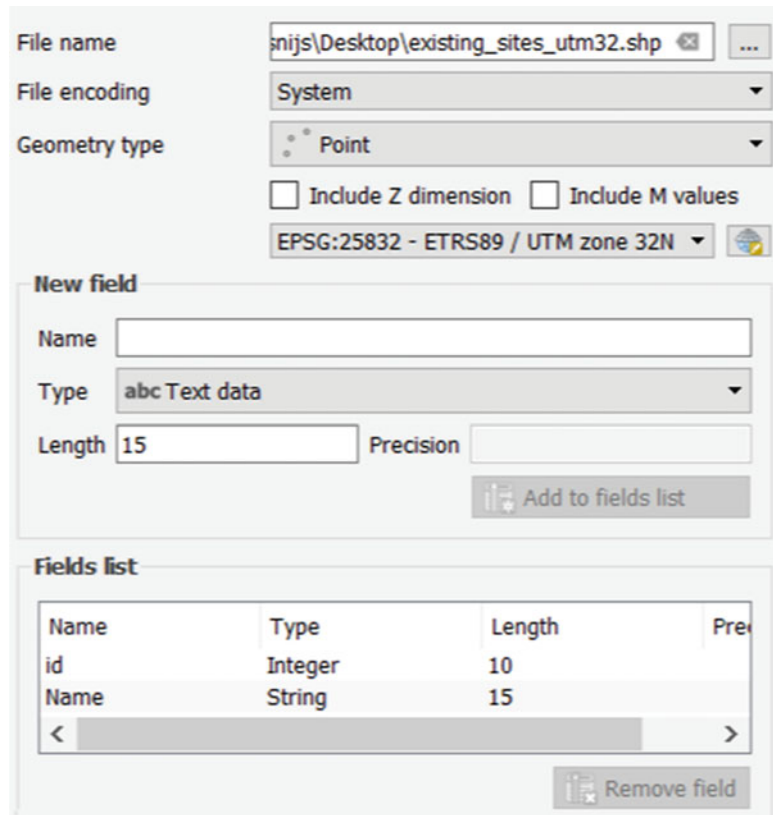


Fig. 13.2 Georeferencing from unreferenced to referenced map and result

Fig. 13.3 Creating a new point shapefile with a proper attribute field (QGIS screenshot)



Once there is a new and empty shapefile that has the coordinate system and the table field, the edit session needs to be initiated for this particular layer. Only then features and their attributes can be added to it (here: city names). To digitize the features as well as the linked attributes it is necessary to add the point features via the ‘add features’-tool and then extend them by the information of the city name in the respective table field (Fig. 13.4). When this is done the edit session needs to be saved and toggled off.

13.2.3 Table Join

One of the given requirements for the analysis is a population density of at least 800 inhabitants per

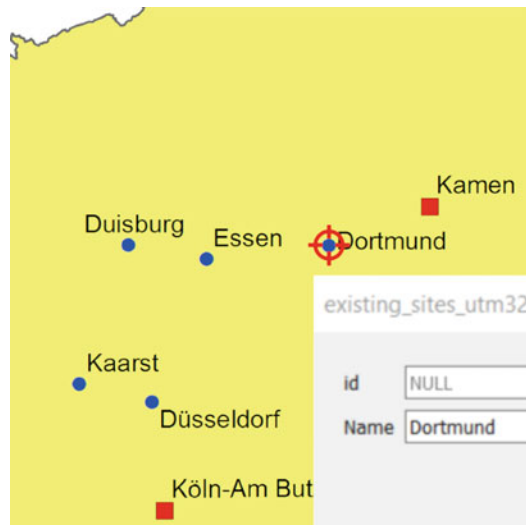


Fig. 13.4 Digitizing existing sites (red crosshairs)

square kilometer in each municipality. As the attribute table of the shapefile that contains the geometries of the municipalities stores neither the information of the relative nor the absolute number of inhabitants per municipality, these values have to be added.

They can be found in the csv-table that contains the (absolute) population data of NRW and needs to be imported into your GIS beforehand so that all the columns and lines are displayed properly.

For the import of the csv-file it is necessary to set several parameters. First there is the delimiter of the different fields (comma, tab, space, etc.), which ensures a proper conversion and puts each value into an individual cell. Second the encoding of the table content has to be set to make certain that all values are displayed correctly. Finally, one has to check if there are coordinates in the table

that could be converted into geometries (Fig. 13.5) which isn't the case here. After importing the table, it needs to be saved as either dbf- or spatial lite-format to provide that it's editable.

To then add this imported data to the attribute table of the shapefile, it is necessary to perform a table join. This requires the identification of a key field (unique ID) that connects both tables – attribute and imported table – via identical values of each feature (Fig. 13.6).

Since a table join can only be done successfully if both key fields are of the same data type (text, number [integer, double, etc.], date, etc.), it can be necessary to copy the values of the key field of the imported table into a new field, setting a data type that matches the key field in the attribute table.

In this case, the key field is the field that stores the code of the particular municipality consisting

Fig. 13.5 Parametrizing the import of a csv-table (QGIS-screenshot)

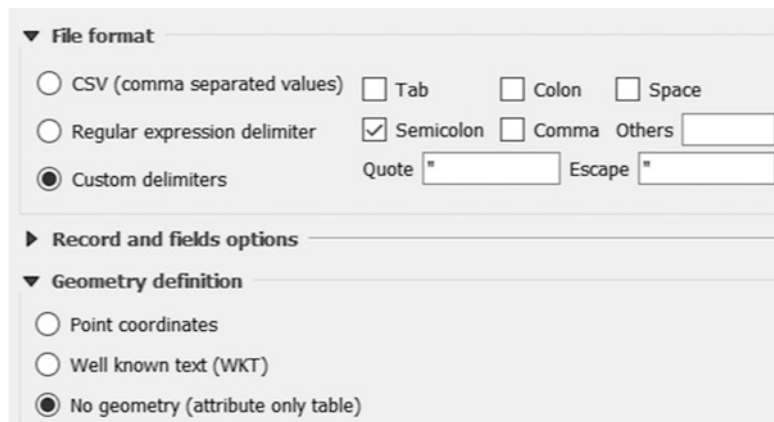


Fig. 13.6 Example for a table join via a key field

area	name	code	code	avr_income	population
250,7	City 1	1	1	1.003,00 €	76.000
134,8	City 2	2	2	1.090,70 €	85.000
42,9	City 3	3	3	985,70 €	107.000
1.000,2	City 4	4	4	1.307,30 €	40.000
782,6	City 5	5	5	790,80 €	95.000

area	name	code	avr_income	population
250,7	City 1	1	1.003,00 €	76.000
134,8	City 2	2	1.090,70 €	85.000
42,9	City 3	3	985,70 €	107.000
1.000,2	City 4	4	1.307,30 €	40.000
782,6	City 5	5	790,80 €	95.000

of seven digits specifying the data type as a number (Fig. 13.7). As the data type of the field in the imported table is a text and there are missing several digits in some cells that prohibit an overall join, it is necessary to create a new field and copy the code values of the text field as numbers to this new field. Therefore the table was modified to dbf or spatial lite in the first place to make it editable.

To now add the missing digits to the respective values, all cells that need to be updated are selected by an SQL expression (see Part I, Sect. 3.1. Simple Spatial Analysis). Looking at the values in the cells it can be stated that either

they are complete and ready for a join or they are missing three zeros on their right to match the ones in the attribute table of the spatial layer. The task now is to find an expression that filters the table and only selects the specific values that need an update and then multiply them by one thousand which will add three zeros. As all affected values are in between a range of five thousand and six thousand the right expression for the selection here is "*<field name>*" > 5000 AND "*<field name>*" < 6000. Via the field calculator the selection is multiplied by one thousand as mentioned above (Fig. 13.8).

After that the table join can be done by navigating to the properties of the municipality shapefile and add a join, choosing the key field in the attribute table, then choosing the table whose values should be added and then choosing the key field in that particular table.

To prevent the result dataset from being changed it is then exported into a new single and persistent shapefile that contains the geometries of the municipalities as well as the absolute population data.

categorie	data type	storable range
Numbers	short integer	-32.768 to 32.767
	long integer	-2.147.483.648 to 2.147.483.647
	float	-3,4E38 to 1,2E38
	double	-2,2E308 to 1,8E308
Text		
Date		

Fig. 13.7 Data types in table fields

The screenshot shows a software interface for updating data. On the left, a table lists 'Code' values: 5158008, 5954024, 5954032, 5515, 5158012, 5116, 5513, and 5370028. The rows for 5515, 5116, and 5513 are highlighted in blue. In the center, a 'Field Calculator' window is open. It has 'Only update 23 selected features' checked. Under 'Update existing field', the 'Output field type' is set to 'Whole number (integer)' and the 'Output field length' is 10. The 'Expression' field contains the text `"Code" * 1000`. Below the expression field is a 'Function Editor' with a search bar and a list of categories: row_number, Aggregates, Arrays, Color, Conditionals, Conversions, Date and Time, Fields and Values, Fuzzy Matching, General, Geometry, Map Layers, Maps, and Math. On the right, another table shows the same 'Code' values, but the updated values are 5158008, 5954024, 5954032, 5515000, 5158012, 5116000, 5513000, and 5370028. The rows for 5515000, 5116000, and 5513000 are highlighted in blue. A large black arrow on the left points from the original table to the updated table, and a large black arrow on the right points from the updated table back to the original table.

Fig. 13.8 Updating selected features (values in blue) via field calculator

13.3 Suitable Areas

To obtain suitable areas it is necessary to extract particular layers out of the prepared data. For this purpose it is helpful to arrange all the relevant parameters in a table (Fig. 13.9). The suitable areas will be exported into *positive layers* and the unsuitable areas will be exported into *negative layers*.

13.3.1 Freeways

The first positive layer for the analysis contains all areas that are in range of 1000m within freeways. That is a valuable approach to provide an easy access to the new store avoiding small roads as most customers use to go shopping with their own car. To create this layer as a shapefile, all freeway features of the street network layer need to be selected, buffered by 1000 m and then exported to a new shapefile.

The selection is done most accurately and effectively via an expression in structured query language (SQL, see Part I, Sect. 3.1. Simple Spatial Analysis). The relevant information for this selection is stored in a certain field that implies abbreviations of the street classes combined with the individual numeric value of the street. All freeways in NRW are tagged with an A (freeway in German: Autobahn) and an individual number. As the expression should manage to select all highways at once, the individual number of the freeways is replaced by ‘%’, changing the ‘=’ to a ‘LIKE’, as now the value does not equal a certain value but is implying all the different freeways at

once, what then leads to the expression “<field name>” LIKE ‘A%’.

The result layers of all upcoming calculations are saved in two different directories – one called *positive* (for the requirement-layers) and one called *negative* (for the restrictive layers).

To then buffer all selected freeways by the value of 1000 m and save it in the positive-folder as an independent shapefile, one has to open the buffer tool, select the input layer, set the beforehand selected features as the only ones that should be used and define a buffer distance with a value in the unit of the used coordinate system.

If the result features should be accumulated to have one mutual border in case of overlaps, it is necessary to dissolve them (Fig. 13.10). In some versions of the tool (depending on the GIS-software and its version), it is as well necessary to define the number of segments that build the outline of the buffered area (the more segments, the rounder the shape).

13.3.2 Federal Highways

As there are not only freeways but also federal highways that should be in direct range of the new store to assure a good accessibility, the whole process can be repeated in almost the same manner. The only thing that needs an adjustment is the expression for the selection of the particular features. In this case to “<field name>” LIKE ‘B %’ OR “<field name>” LIKE ‘N%’ since the federal highways are tagged with two different letters. To not select the features that contain both – B and N – the logical operator OR connects the

Selection Criteria				
Criteria	Dataset	Attribute	Value(s)	Buffer
<i>include:</i>				
population density at least 800 inh./km ²	municipalities_utm32.shp	pop_dens	>=800	-
maximum distance to federal highways 500m	streets_utm32.shp	SHIELDNUM	‘B%’, ‘N%’	500m
maximum distance to freeways 1000m	streets_utm32.shp	SHIELDNUM	‘A%’	1000m
<i>exclude:</i>				
distance to existing sites	furniture_stores.gif	-	-	40km
outside of settlement areas	built_up_areas_wgs84.shp	-	-	-

Fig. 13.9 Table with the selection criteria of included and excluded layers

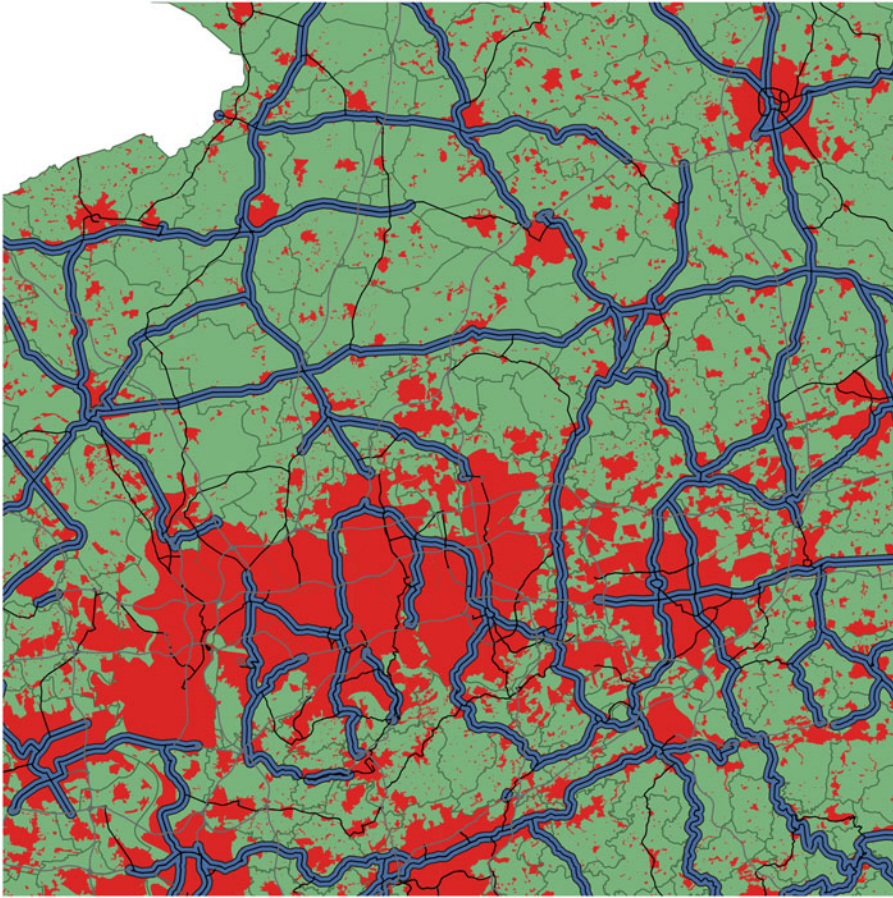


Fig. 13.10 Buffered federal highways (blue)

two expressions, for only one of the two commands has to be true to select the corresponding feature.

Afterwards the selection is buffered by 500 m (Fig. 13.11) and then saved into the positive-folder.

13.3.3 Population Density

For the new site of the store it is desirable to only include regions into the consideration that have a high population density to reach as many customers as possible. To calculate the population density of each municipality in NRW it is necessary to look into the attribute table of the joined layer and detect the table field that contains the absolute number of inhabitants per municipality.

In this case, the field with the population data of 2008 is used for the relative population per square kilometer. In order to perform the calculation for all features in one step, again the field calculator is the tool of choice. The new information is stored in a new table field that has to be created with a decimal number field type and a sufficient length. The population density is then calculated dividing the absolute numbers of inhabitants by the area of the certain municipality e.g. "*<field name population>/area [sqkm]*".

The UTM coordinate system is in the metric unit *meter*, which can require another important step (depending on the used GIS) for the calculation of the population density per square kilometer, as the area has to be multiplied by 1.000.000 ($1 \text{ km}^2 = 1 \text{ m}^2 \times 1.000.000$) in case the unit of the area can't be defined before the calculation.

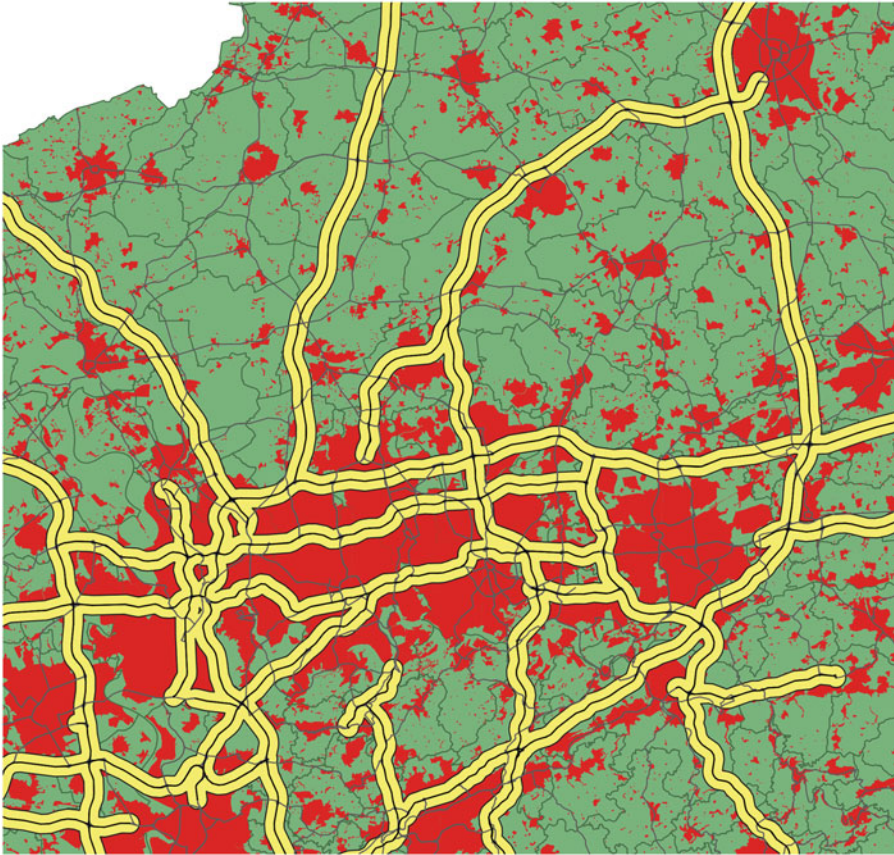


Fig. 13.11 Buffered freeways (yellow)

After the population density is calculated properly, all municipalities that are greater or equal than 800 inhabitants per square kilometer need to be selected by another expression.

As the only reasonable operator here is the greater than or equal to operator (\geq) the expression for this selection is “*<field name population density>*” \geq 800. The selection then is exported to a separate and persistent layer into the positive-folder (Fig. 13.12).

13.4 Unsuitable Areas

To subtract all restricted areas from the result layer a negative-folder is created to store the two layers which define the areas that need to be avoided for the new location. These are built-up areas, as it is too expensive to build a huge new

warehouse in an area that is more valuable for living than for commercial use, and areas that are in between a radius of 40 km of already existing warehouses. This buffer around the existing sites is reasonable as the company rather wants to approach new customers than give already nearby ones a second opportunity where to buy their new piece of furniture. In a more detailed analysis it could be considered to obtain the buffer values not from a simple linear buffer but from a service area that is processed using the street network instead (see Part 1, Sect. 3.3. Network Analysis).

As the built-up areas layer already contains only features that define the built-up areas themselves this layer is used directly and completely as a negative layer for one restriction by copying it into the corresponding directory.

However, the distance to existing sites needs to be processed before the layer is useable for the

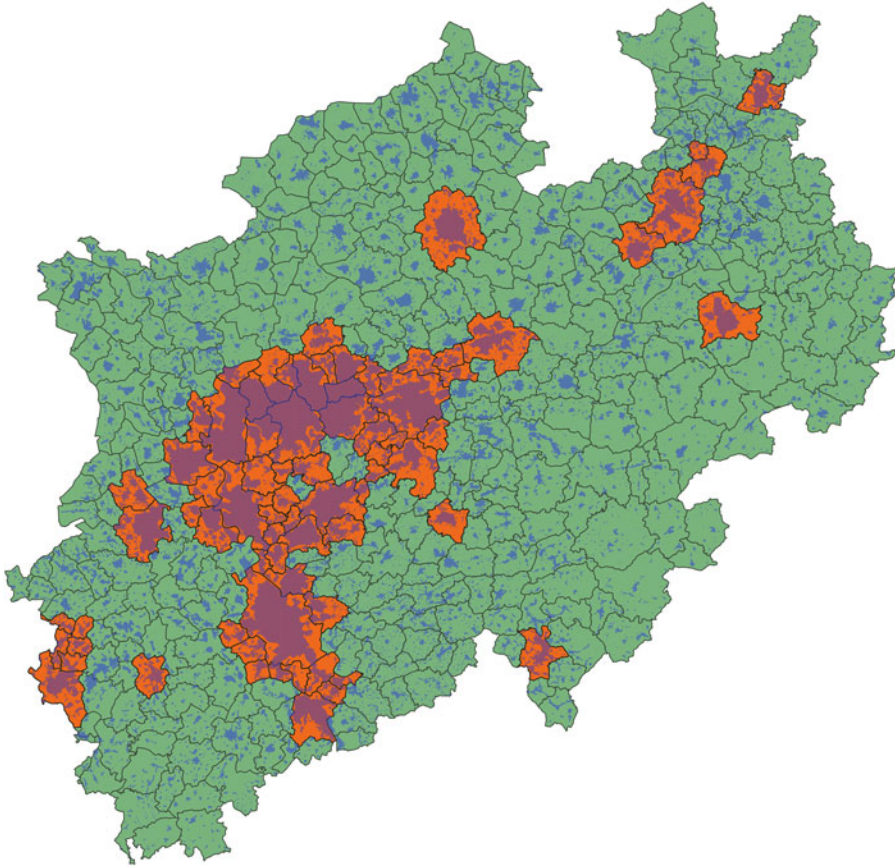


Fig. 13.12 Municipalities with over 800 inhabitants per square kilometer (orange) and built-up sites (light blue) to compare

following calculations and therefore added to the folder containing the two negative layers. The processing here only consists of a single buffering by 40 km that is again done via the buffering tool, setting the beforehand digitized layer of the existing sites as input layer and buffer it by 40.000 m to then save it into the negative-folder (Fig. 13.13).

13.5 Combining Layers

Now that the two folders *positive* and *negative* each contain all the files that are either a requirement or a restriction, all files in each folder need to be combined to one overall file (example of logical operators in Fig. 13.14).

13.5.1 Positive Layer

The combination of all three files from the positive-folder to a single layer is done via an intersection, which is equal to a logical AND ('&'). After simply setting the three layers as input, the output of the intersect-tool only contains areas where all requirements are fulfilled concurrently. Areas where only one or two requirements are accomplished are excluded from the positive result layer (Fig. 13.15).

13.5.2 Negative Layer

For the overall negative layer the single layers have to be combined to an overarching layer via

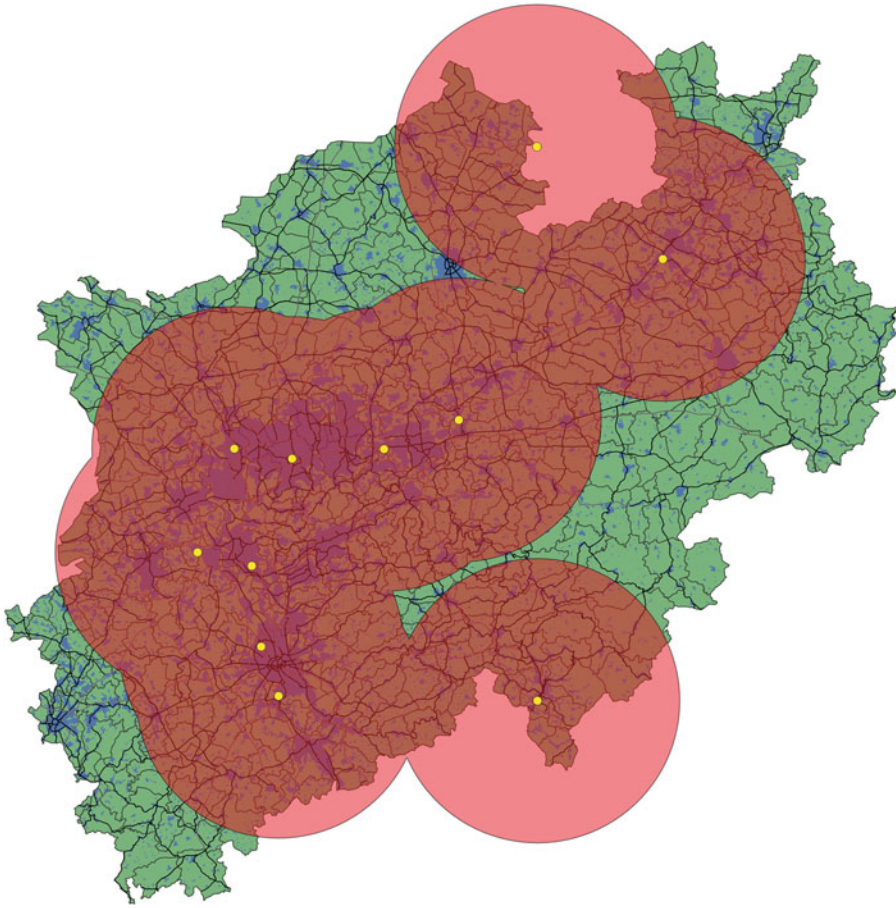


Fig. 13.13 Existing sites (yellow dots) buffered by 40 km (light red)

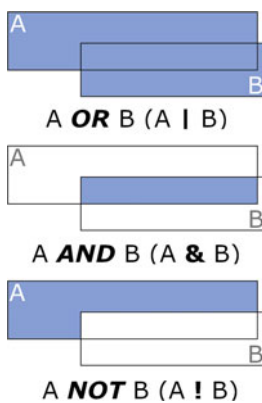


Fig. 13.14 Logical operators

the merge or union tool, which is equal to a logical OR (' \cup '). No matter if one or all restrictions are regarded – there is no way the new site will be

built in an area that is located within a single negative layer (Fig. 13.16).

13.6 Final Result

For the final result the two layers (overall negative and overall positive) need to be combined somehow. The tool of choice here is the difference tool, as it is the same as a logical NOT (' \setminus ') and subtracts all unsuitable areas from the suitable ones (Fig. 13.17).

The final result now can be readout in area size and precise location and then put into a map for an overview of what is left and where. Following this, new investigations can be done on the suitable areas, initializing accurately the upcoming search on a smaller scale with further approaches

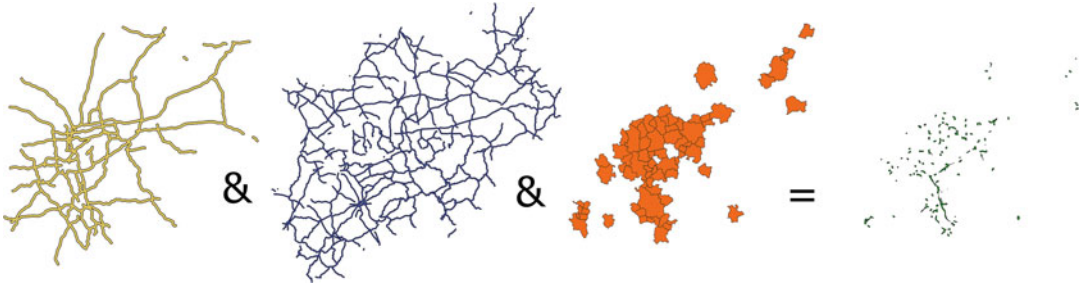


Fig. 13.15 Intersecting all positive layers via a logical AND ('&')

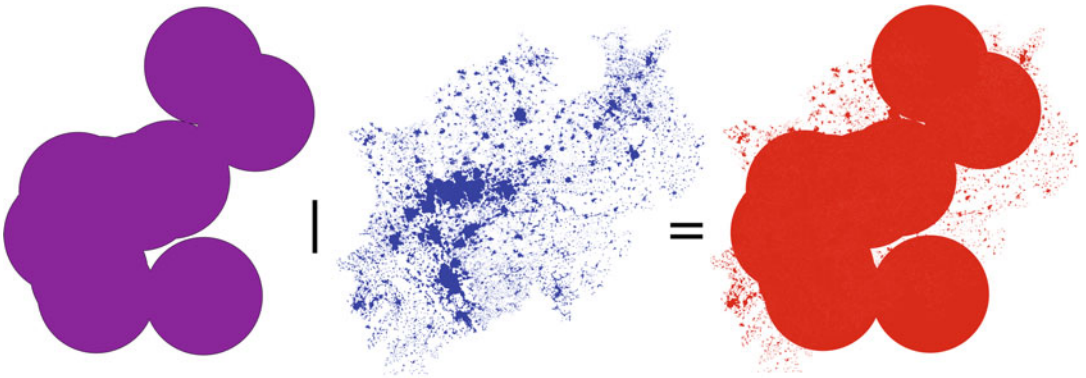


Fig. 13.16 Merging the two negative layers via a logical OR ('|')

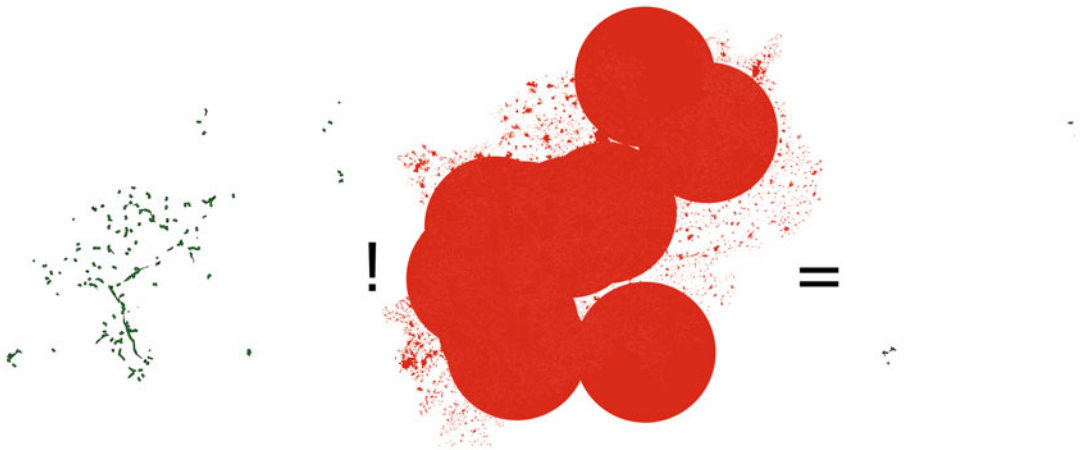


Fig. 13.17 Final result: the difference or logical NOT ('!') of the positive and negative layer

for a property to build a new furniture store (Fig. 13.18).

This short workflow could be one possible scenario for a valuable combination of both spatial and economic approaches. There are certainly

several other factors that can be included to the whole process like income per capita, land cost, availability of water and electricity supply etc. To do this, corresponding data needs to be acquired, combined and processed to then set the particular

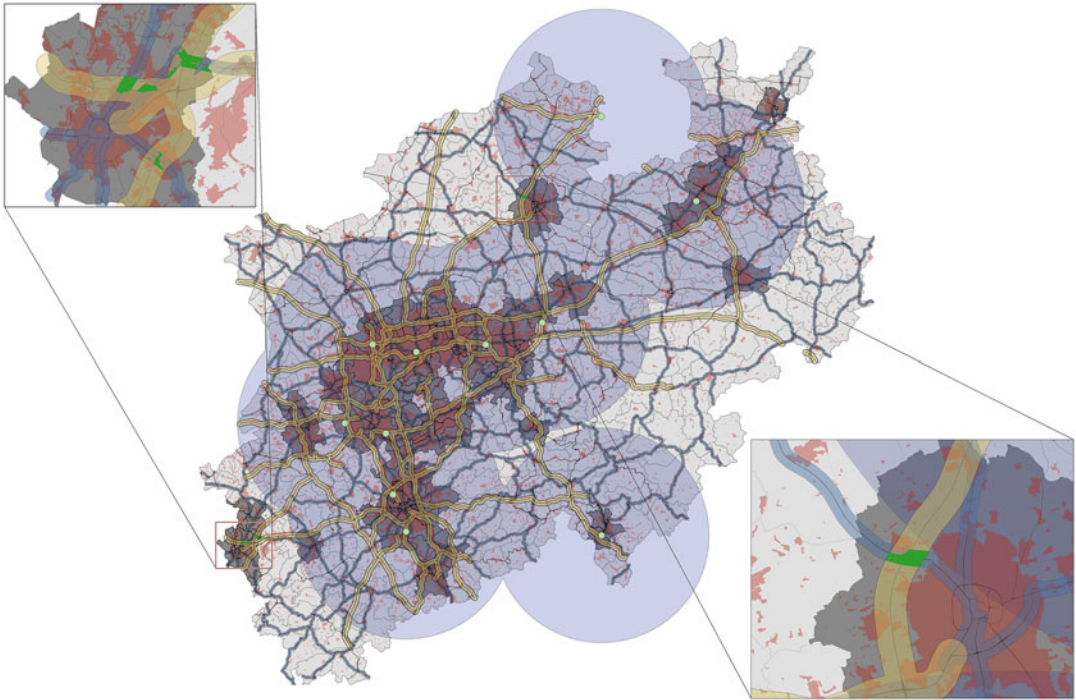


Fig. 13.18 Final result, suitable areas in green

thresholds and extract the layers that should affect the final result.

The two fields of GIS-science and economy require and complement each other in almost every step. This becomes explicit while

parametrizing the different calculations and tools, dealing with values that underly business related decision making while the whole processing is done spatially with layers and ends up in a characteristic and solving map.

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.

