

Designing a Semi-automatic Map Construction Process for the Effective Visualisation of Business Geodata

Marion Simon[1(✉)] and Hartmut Asche[2(✉)]

[1] Institute of Geosciences, University of Potsdam, Potsdam, Germany
marion.simon@uni-potsdam.de
[2] Computer Graphics System Group, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
hartmut.asche@hpi.de

Abstract. This paper proposes a map construction process for the semi-automatic construction of thematic maps from business information data. Addressing a non-specialist user audience, the map construction process will allow a correct, at the same time effective cartographic visualisation for further (geo)visual analysis. Utilising the frequently disregarded geospatial component of existing business mass data, quality map representations facilitate the visual exploration, detection, and analysis of relevant spatial data distributions and structures hitherto unseen in the data. Presently, neither operational procedures nor appropriate software systems, such as BIS, DDS or GIS, are available in the industry for an effective map representation of the geocoded data. To put economic and business experts into a position to make full use of the geo coordinates present in the data, an easy-to-handle map construction process is required to exploit the full semantic and spatial potential of business data. Exemplified for an area diagram map, the map construction process discussed here provides the relevant tools and methods for the targeted audience.

Keywords: Geocoded business data · Geovisualisation · Map construction · Quality map representation · Geovisual analysis · Modular process design

1 Introduction

Most data in company databases possess a spatial reference. Hamilton [1] estimates a spatial reference in about 95 percent of all data; that is, 95 percent of all data is geospatial. Nevertheless, the use of spatial attributes is not fully exploited for a number of reasons. However, many connections, dependencies and interrelations are only expressed through the geo-component of the data. Potential further insights gained from the analysis of the spatial component of corporate (geo)data constitute a new and valuable contribution to corporate decision-making. Communication of spatial reference and their visual analysis is most effective through the visual channel, i.e. by visualisation, e.g. through map representations. Such maps also serve as a means of communication for these very decisions [2]. From a business perspective, it can be assumed that integration of the geo-component will result in all assessments for value

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12252, pp. 435–447, 2020.
https://doi.org/10.1007/978-3-030-58811-3_32

creation and employment being approximately two times higher [3]. Nevertheless, the expert effort and cost expenditure for the economic production of effective thematic maps need to be taken into account [4].

This article presents a solution approach that enables companies to visually analyse existing company (geo)data not only according to their semantic content but also with regard to their spatial reference. Up to now, there has been little research on the methods, processes and analyses that deals with both the expressive and effective graphic representation [5] as well as use of geocoded business data in the industry. Our focus in this presentation is on the effective representation of spatially related quantitative information in segmented proportional area symbols. Related issues of professional cartographic symbolization will be dealt with elsewhere.

2 Solution Approach

To empower corporate businesses to visualise their data expressively and effectively, it is necessary to make available the existing knowledge for the creation of professional, meaningful map graphics in the form of easy-to-use digital production processes. This enables even non-expert users to create effective map graphics from non-graphic company data. The basis of the solution approach is an executable visualisation process (Fig. 1), based on the classical visualisation pipeline [6]. The design and implementation of this principally scalable process is described below for the generation of small-scale map representations.

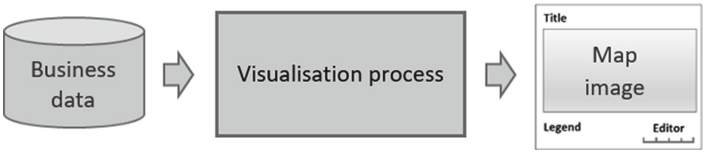

Fig. 1. Principle to convert non-graphic company data into graphic map representations

Explorative data analysis (EDA), also known as data mining, includes the special form of explorative spatial data analysis (ESDA) [7]. The process to be developed enables companies to extend the previously graphics-free data evaluation by analyses of the spatial distribution of the data based on their spatial attributes that have been left unused so far.

The way to design and develop an executable construction process for the production of effective cartographic visualisations from geodata consists of three essential stages:

a) formalisation of the map construction,
b) design a map construction process and
c) process automation.

In order to provide cartographic expertise for the visualisation of enterprise data, it is necessary to formalise this expertise. On the one hand this requires the identification of the relevant map construction steps, including the definition of the spatial reference, the selection of adequate map types, generalisation measures and the definition of the presentation form. On the other hand, it is necessary to define of the target group for which the map results are to be made available. In the course of this formalisation, the entire map production process is broken down into a finite number of content-defined modules that can be executed separately.

Initially, the design of such modular map construction process for the professional visualisation of spatial (mass) data is done on a semantic (external) and logical level. For this purpose, the classical visualisation pipeline with its modules is adapted. The map construction process is also modular, consisting of a fixed number of micro-modules. Depending on the application, these micro-modules can be combined flexibly to process chains via defined connectors. They are assembled semantically and logically into a dedicated card design process. The linking of the micro-modules by means of corresponding connectors depends on a company's analysis objectives defined in terms of data content. Map construction rules define the dependencies between the micro-modules. These interrelations are then graphically represented by process flow diagram elements.

To automate the cartographic expertise mentioned above, the process is first outlined with UML and SysML for a better understanding, and then graphically modeled with the jABC software [8]. This software system offers capabilities to generate Java code directly from graphic modelling. It can be determined which micro-modules do run automatically and for which modules user interaction is required or useful. Automation of the micro-modules will enable the user to generate alternative visualisation results with a higher degree of user control instead of default map graphics. It has to be noted that more user control is only possible to the extent that the quality criteria for the output of map visualisations implemented in the process modules are met. Finally, servicification of selected micro-modules is implemented. Existing open source tools and applications are into the map construction process integrated, such as the well-known ColorBrewer colour definition system [9].

The stages described provide a picture of the automation scope of the application.

3 Implementation Strategy

Taking the data analysis module as an example, the solution approach presented above will be dealt with in more detail (see Fig. 2). This module incorporates an automated process for the selection of the adequate map type, based on the analysis of the existing source data.

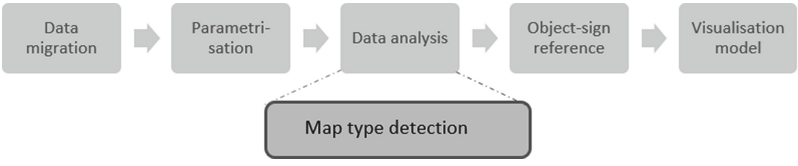

Fig. 2. Map construction pipeline (section, modified from [6])

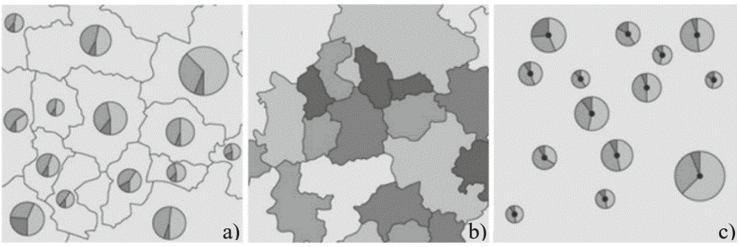

Fig. 3. Selection of map types relevant for business data: a) area diagram map, b) choropleth map, c) positional diagram map. Source: [10, 11]

Starting from the database, the possibility of a spatially related visualisation of the dataset in question is examined in a first step. If thre generation of a map graphic turns out to be feasible, the map type optimally representing the available geodata is identified. To do so, the selection process for choosing the appropriate map type is formalised. Map types are described based on their presentation characteristics. Selection of the appropriate map type is from about a dozen of different map types widely accepted in cartographic theory and applications (see Fig. 3). The optimal map type is one that is best fitting the geographic and semantic characteristics of the data to be mapped.

It has already been mentioned that an analysis of the source data is essential to determine the selection process for the dataset to be mapped. This is performed in a second component. To set up a selection process, an analysis of the data to be visualised in relation with the data characteristics is required for the creation of the adequate map type. Based on the definition of the relevant map types in cartographic literature, all map types representing quantitative data are identified. Suitable map types are characterised by the geometric and semantic components of the data that constitute the thematic content as well as the geographic reference of the map to be constructed. The focus of this paper is on the quantitative semantic content for thematic maps. The thematic content refers to the semantic information of a particular dataset expressed numerically [12]. Requirements of the data characteristics essential for visualisation (spatial reference, attribute value, value development and scaling level) are listed separately for each map type. Table 1 gives as an example for area diagram maps which are described below.

Table 1. Formalisation of map type characteristics of the area diagram map (modified after [10, 13])

Map type characteristics	Area diagram map
Spatial reference	area-related
Attribute value	absolute
Value development	continuous/classified
Scaling level	ratio oder interval

Because of its spatial reference, the quantitative content of the map presentation is always linked to specific geographic units. Areas (e.g. administrative units), lines (e.g. traffic routes) or points (e.g. settlements) are possible. For example, quantitative data of the consumption behaviour of a particular country's population has an area reference (because the population is related to a national territory). The representation of the reference unit (here: the delimitation of the corresponding country, the highest-level administrative unit) is in the base map. The corresponding diagram is then placed in the administrative unit to show the spatial reference of the respective data [14]. The semantic reference is further differentiated in the attribution as an attribute value. Thus, the absolute value and the relative or ratio value are assigned to the quantitative attribute data category. Options for further processing depend on the relations between the attributes, which are determined by the attribute value [11].

The value range can be separated into continuous (i.e. unstructured) attributes as well as discrete (i.e. structured) attributes. This describes a characteristic of the relationship between the attributes, which, in turn, determines their further processing options and the selection of the optimal map type [11].

The level of measurement characterises the information content of the data. It defines possible transformations of attribute values and describes the graphical relational properties of characters that make these transformations visible on the map. A total of four levels are distinguished to which the semantic attributes of the geodata can be assigned: nominal scale, ordinal scale and metric scale, which can be further subdivided into interval and ratio scales [11, 15].

In the following, the formal data characteristics relevant for the selection of the adequate map type are subsequently combined in a structured way to form a subprocess of the map type selection process source data analysis. The decision tree for the selection of the optimal map representation for the respective data is derived from this. A detail is shown in Fig. 4.

The sequence of the individual process steps of the map type selection component begins by checking the spatial reference of the existing business statistics to be mapped. Spatial reference provides "the geometrical information about the location and shape of the individual objects necessary for any digital modelling and cartographic representation […]" [12]. Examples are postal codes, telephone area codes or even place names.

Export datasets can contain, for example, country names or customer locations to which export quantities listed in the dataset are delivered. Country names or place names can be automatically extracted from the selected data lists (e.g. table schemata of a database). These geographic names are then compared with digital directories of geographic names, so-called gazetteers. This process is limited to fields of the 'TEXT' data type, so that not every field entry is checked for matching and thus affects the performance of the service. Finally, non-matches are displayed for manual error correction of the geographic names. This procedure can be used to check location-based (point-referenced place names) and area-related data (country names) for their spatial reference.

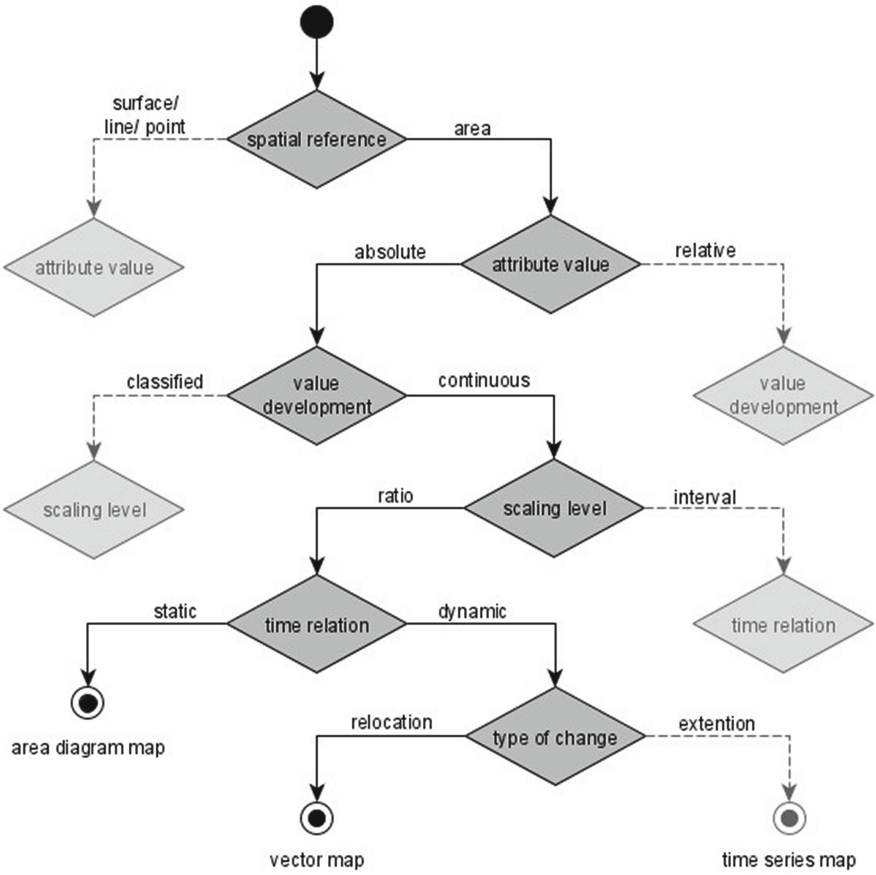

Fig. 4. Decision tree for selecting a map type (section)

Relative attribute values are generally expressed in percent or per thousand [14]. Absolute values characterise a quantity starting from a zero point. They are specified by value unit. It is possible to automate this sub-module that differentiates relative from absolute values by reading out field types (percentage or currency formats), field contents (% or currency symbol or weight unit addition) or column headings (entries in % or 'percent' or €/t or 'Euro'/'tons').

An unclassified value progression is assumed when specifying export volumes or export revenues. It is further assumed that the company's interest in data analysis is based on data that are as accurate as possible. Classified values are data already stored in a classified form.

In the case of quantitative data, options to select the scaling level are limited to the values of ratio and interval scale attributes. Weight data or sales revenues belong to the group of ratio-scaled data, which have a natural zero point. The system can use the results of attribute value determination for evaluation purposes.

To find out the time reference, interaction with the user is necessary. The system can make suggestions to the user on the basis of existing temporal attributes or time stamps read (date formats of various specifications).

Dynamics in the data can be expressed in two ways depending on the type of change (vector or time series map). Whether a representation of changing target regions of exported products is desired can be inquired. If the wish does not exist, a static temporal reference is assumed and the corresponding map type area diagram map is selected in the result.

4 Applying the Map Construction Process: A Use Case

To illustrate what has been presented, we take the use case a spread-producing company from a particular country. We assume this company exports its products Europe-wide. In order to be able to analyse its export data both in terms of content and region, an easy-to-comprehend graphic representation of the market penetration on the European market for spreads is required. The most effective way to do this is by a map graphic.

We also assume that the company aims at further market development by observing the European consumption behaviour of bread and cereals. For this purpose, sales managers extract the current figures from European databases and compare the numerical values with their own export figures using a so-called map construction assistant, which generates an effective map display from this geodata.

Our exemplary spread company keeps its own export data in table form under the file name 'ExportsSpreads.xls'. It contains annual export values (weight in tons-t and currency in euro-EUR) of the product category spreads to, say, the Scandinavian countries (Denmark, Finland, Norway, Sweden) (Fig. 5).

	A	B	C	D
1			export: weight	export: value
2	Year	NAME_ENGL	t	thsd. EUR
3	2018	Denmark	1390.8	4890
4	2018	Finland	328.5	787
5	2018	Sweden	927.6	2937
6	2018	Norway	265.0	749

(C)opyright Statistisches Bundesamt (Destatis), 2019

Fig. 5. Exemplary export figures of a product category of a manufacturing industry from Germany to Scandinavia

This export table forms the starting point for a data analysis to feed the described map type selection process. For the selection of the optimal map type, the decision tree presented in Sect. 3 is run through semi-automatically. For this data example the following can be determined: The available geodata

a) refer to state territories (selected Scandinavian countries are available in the digital GeographicNamesIndex),

b) are available in absolute values in the units t (tons) and Eur (euro) (no percentage signs in column headings),
c) are not broken down,
d) are ratio-scaled (identifiable by the system from the units t and Eur - conclusion: scale with natural zero = ratio-scaled),
e) refer to the year 2018 (static).

Thus, after having traversed the decision tree, the map type of the area diagram map to be implemented is selected. In order to go through this selection process on a technical level, the individual intermediate steps need to be described in more detail. As an example, the procedure for the first step, the detection of the spatial reference in the geodata set to be mapped, is presented here.

The tool-based determination of the geodata reference is carried out by searching for place or country names. First, the column must be identified, which holds a corresponding spatial reference. Since column designations are not titled in a standardised way and a certain range of names (e.g. country, city, name) exists, it is advisable to start with the selection procedure in the first or second line (depending on the table structure). It is also assumed at this point that export data is stored country- or location-wise and that these spatial data are listed at the beginning of the table for easier reading. For this reason, the search for country designations is started in column A.

```
/* Identification of the column with spatial reference data */
if field A2 formatted as text;
do check string compatibility to members of 'GeographicNamesIndex';
else if field B2 formatted as text;
     do check string compatibility to members of 'GeographicNamesIndex';
     else if field C2 formatted as text;
     do check string compatibility to members of 'GeographicNamesIndex';
     else break;
```

This loop is repeated over all contents contained in the table section (here: country names). A digital, internationally recognised index of names (Gazetteer Service) is fictitiously constructed here and shown as an example in Fig. 6. It represents the data basis for the spatial reference check within the source data analysis.

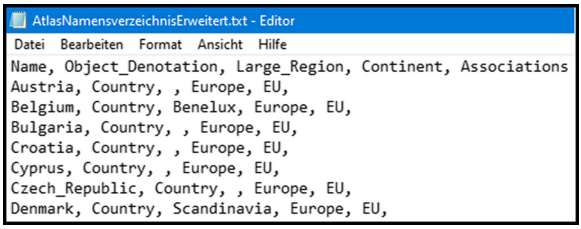

Fig. 6. Geographic names index (section)

The execution order (do) required in pseudocode is integrated into a separate micro-module and is described as follows:

```
/* Check for compatibility with the 'GeographicNamesIndex' */
Gazetteer Compatibility Check
{
    initialize x = read string;
    initialize y = read string from 'GeographicNamesIndex.txt' in (column
A, row 2);
    if (x unequal y);
    do initialize y = read string from 'GeographicNamesIndex.txt' in (col-
umn A, row 3);
    else check in 'GeographicNamesIndex.txt' if Object-
Denotation = country or city and load geometry data for base map via API
}
```

Based on this pseudo code description, the module description shown in Fig. 7 can be expected for the first segment of the bipolar decision tree "spatial reference". The coding of the other elements of the decision tree is also carried out according to this scheme. The decomposition and combination results in micro-modules such as the 'Gazetteer Checker'.

With the development and coding of the decision tree, the resulting micro-modules as well as their superordinate modules, shown here as roles in the SysML activity diagram, can now be described.

The coded and modularised decision tree will propose an area diagram map to represent the example data set of the selected export data. In the process module of object-sign reference, the design elements are assigned to the geodata to be displayed (Fig. 8).

After further execution of the map construction process (see Sect. 3), this map type, which has been identified as adequate, is implemented as standard output. In this example, the diagram type 'bar chart' is assigned as standard. To represent comparative export figures of the EU, the bar chart can be segmented at the users' request. The larger column represents the EU export figures for comparison with the embedded smaller column of the company's own export figures. The 'visualisation model' module displays the initial graphical design on a canvas, based on decisions made or predefined by the user, taking into account the output characteristics defined in the initial 'parameterization' module, i.e. the display of the base map and the embedding of the segmented bar charts, see Fig. 9.

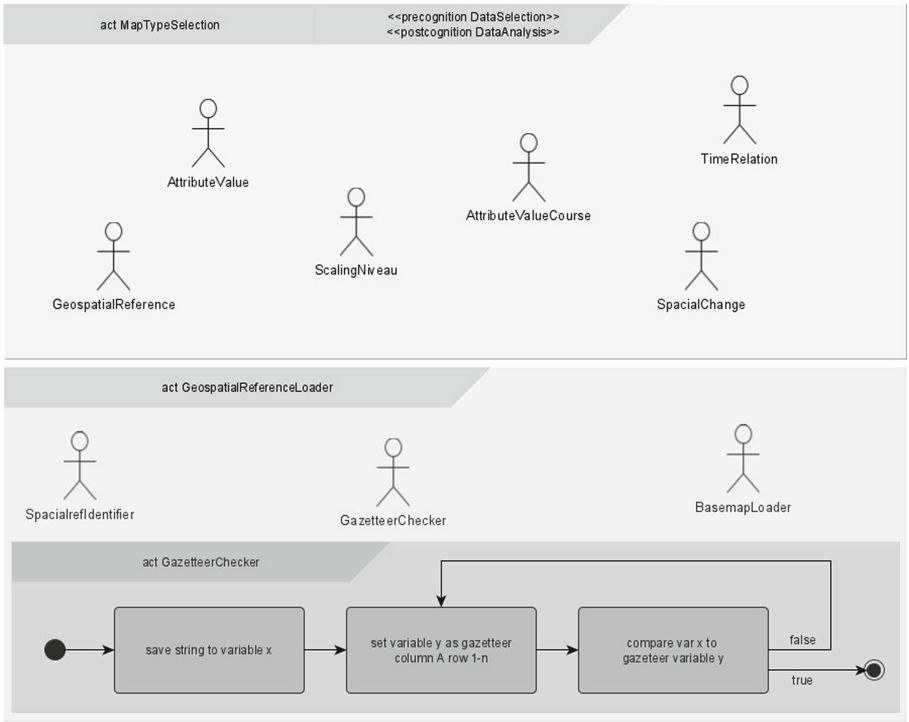

Fig. 7. Decision tree as SysML activity diagram

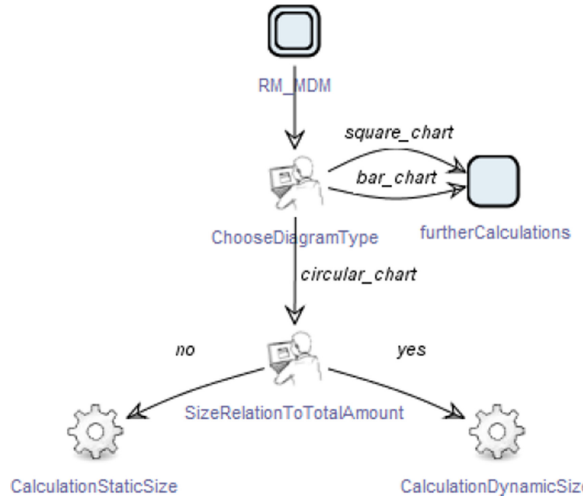

Fig. 8. Process model for diagram type selection in jABC (section, [16])

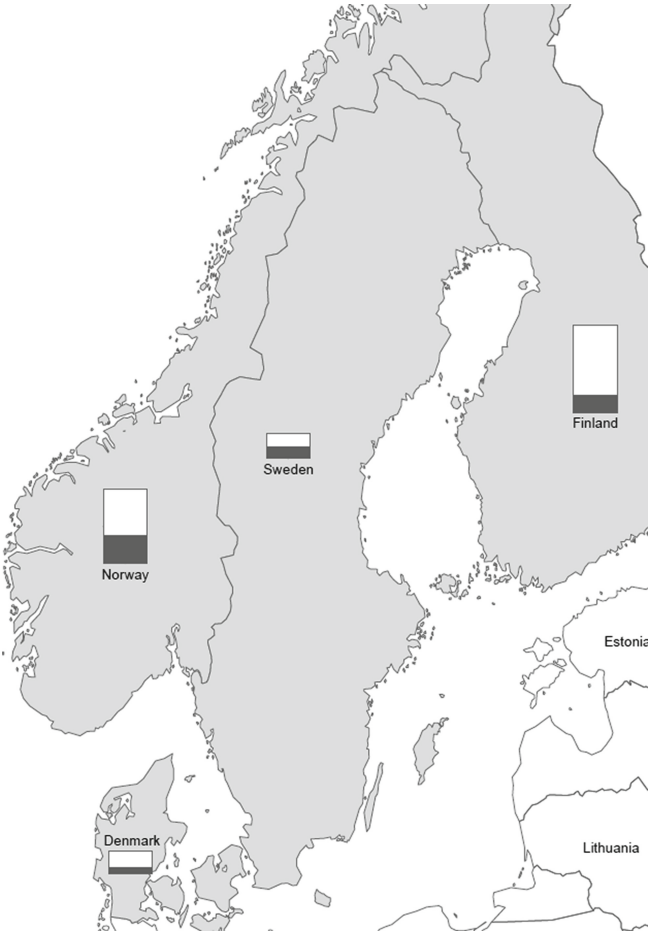

Fig. 9. Area diagram map for use case data set

5 Conclusion

This contribution presents a semi-automated design process for the generation of effective map visualisations from company databases with a spatial reference.

In order to incorporate the existing expertise of effective map visualisation into algorithms, the relevant map construction steps have been identified and broken down into executable steps that can be run separately. The semantic and logical linking of modules and micro-modules completes the step of formalisation. To facilitate the automated selection of the optimal map type a decision tree has been developed. By means of programming and servicification, users can access this expertise. The solution approach was explicitly exercised with an example data set. The data analysis module from the adapted visualisation pipeline has been used and enriched with systematically arranged expertise to identify the optimal map type for the example data. The modules

are described using activity components from SysML. A suitable map type for the use case has been identified and presented in a non-generalised map visualisation. It remains to be examined whether all process modules are mandatory for data analysis or whether, for example, the determination of the measurement scale of the processed dataset can be dispensed with. It is assumed that the micro-modules for examining the data for their value progression and scaling level will provide similar results when selecting the suitable map type. For the sake of completeness, they have been taken into account in this work in order not to prejudge a possible extension of the overall process.

The first visualisation result has not been free of graphical conflicts. Furthermore, essential components of the map layout such as title, legend and scale are missing. These deficits will be addressed in further visualisation process modules for georeferencing and map layout.

This contribution has demonstrated the importance and potential of spatial analysis of business geodata. These methods enable companies to use a completely new, previously neglected analysis method. The spatial analysis of their company data supports efficient entrepreneurial decisions and serves, at the same time, as an effective means of communication of these decisions.

References

1. Perkins (2010). http://computerworld.com/article/2549741/have-you-mapped-your-data-today-.html. Accessed 15 Mar 2020
2. Kraak, M.J., Ormeling, F.J.: Cartography. Visualization of Spatial Data. Addison Wesley Longman Ltd., Harlow (1996)
3. Fornefeld, M., Oefinger, P., Rausch, U.: Der Markt für Geoinformationen: Potentiale für Beschäftigung, Innovation und Wertschöpfung. Kurzfassung. Studie. MICUS-Management Consultig GmbH, Düsseldorf (2003)
4. Asche, H., Herrmann, C.: Desktop Mapping in der thematischen Kartographie. Stand der Technik und Marktübersicht. In: Deutsche Gesellschaft für Kartographie (DGfK) (edt.): Kartograpisches Taschenbuch 1994/95, 4th edn. Kirschbaum, Bonn (1994)
5. Mackinlay, J.: Automating the design of graphical presentations of relational information. ACM Trans. Graph. **5**, 110–141 (1986)
6. Haber, R.B., McNabb, D.A.: Visualization idioms: a conceptual model for scientific visualisation systems. In: Shriver, B., Nielsen, G.M., Rosenblum, M. (eds.) Visualization in Scientific Computing, Los Alamitos (1990)
7. Slocum, T.A., McMaster, R.B., Kessler, F.C., Howard, H.H.: Thematic cartography and geovisualization. In: Series in Geographic Information Science, 3rd edn. Prentice-Hall, Upper Saddle River, NJ. (2009)
8. Technische Universität (TU) Dortmund: jABC Link (2020). http://ls5-www.cs.tu-dortmund.de/projects/jabc/index.php. Accessed 13 May 2020
9. Harrower, M., Brewer, C.: ColorBrewer.org: an online tool for selecting color schemes for maps. Cartographic J. **40**(1), 27–37 (2003)
10. Bollmann, J., Koch, W.G. (eds.): Lexikon der Kartographie und Geomatik, vol. 1. Spektrum Akademischer Verlag, Heidelberg (2001)
11. Bollmann, J., Koch, W.G. (eds.): Lexikon der Kartographie und Geomatik, vol. 2. Spektrum Akademischer Verlag, Heidelberg (2002)

12. Hake, G.: Grünreich, D., Meng, L.: Kartographie. Visualisierung raum-zeitlicher Informationen, 8th edn. Walter de Gruyter & Co., Berlin, New York (2002)
13. Arnberger, E.: Thematische Kartographie. Westermann, Braunschweig (1997)
14. Witt, W.: Thematische Kartographie. Methoden und Probleme, Tendenzen und Aufgaben, 2nd edn. Gebr. Jänecke, Hannover (1970)
15. Kucharczyk, C.: Konzeption, Entwicklung und Implementierung eines regelbasierten Kartenkonstruktionsassistenten zur fachgerechten Visualisierung statistischer Massendaten. Masterthesis (2013). Unpublished
16. Simon, M., Asche, H.: Automated spatial data processing and refining. In: Lamprecht, A.-L. (ed.) ISoLA 2012/2014. CCIS, vol. 683, pp. 38–49. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-51641-7_3