

Article

# Where Maps Lie: Visualization of Perceptual Fallacy in Choropleth Maps at Different Levels of Aggregation

Giedrė Beconytė \*, Andrius Balčiūnas, Aurelija Šturaitė and Rita Viliuviene

Institute of Geosciences, Faculty of Chemistry and Geosciences, Vilnius University, M. K. Čiurlionio 21, LT-03101 Vilnius, Lithuania; andrius.balciunas@chgf.vu.lt (A.B.); a.sturaite@gis-centras.lt (A.Š.); rita.viliuviene@chgf.vu.lt (R.V.)

\* Correspondence: giedre.beconyte@gf.vu.lt

**Abstract:** This paper proposes a method for quantitative evaluation of perception deviations due to generalization in choropleth maps. The method proposed is based on comparison of class values assigned to different aggregation units chosen for representing the same dataset. It is illustrated by the results of application of the method to population density maps of Lithuania. Three spatial aggregation levels were chosen for comparison: the 1 × 1 km statistical grid, elderships (NUTS3), and municipalities (NUTS2). Differences in density class values between the reference grid map and the other two maps were calculated. It is demonstrated that a perceptual fallacy on the municipality level population map of Lithuania leads to a misinterpretation of data that makes such maps frankly useless. The eldership level map is, moreover, also largely misleading, especially in sparsely populated areas. The method proposed is easy to use and transferable to any other field where spatially aggregated data are mapped. It can be used for visual analysis of the degree to which a generalized choropleth map is liable to mislead the user in particular areas.

**Keywords:** choropleth maps; perception; fallacy; aggregation; mapping



**Citation:** Beconytė, G.; Balčiūnas, A.; Šturaitė, A.; Viliuviene, R. *Where Maps Lie: Visualization of Perceptual Fallacy in Choropleth Maps at Different Levels of Aggregation*. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 64. <https://doi.org/10.3390/ijgi11010064>

Academic Editor: Wolfgang Kainz

Received: 29 November 2021

Accepted: 12 January 2022

Published: 14 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Choropleth maps are compiled and used in order to comprehend the spatial patterns of distribution of different social and natural phenomena. The obvious motivator is the rapid and easy interpretation of a map image compared to statistical data tables intended more for the specialist. Moreover, special software and specialized knowledge is necessary for customized spatial analysis and the generalization of data. People today become used to choropleth maps at school; they trust cartographers and are accustomed to regarding the represented visual information as generally speaking to be authoritative. The presumed facts are well memorized. Spatial patterns and differences are quickly noticed. Maps are therefore powerful tools for forming public opinion and impacting decision making. However, cartographic information is in one way or another misleading [1]. Apart from instances of intentional manipulation, there are many ways in which representation of information can lead to possibly incorrect interpretation of the facts. Users are rarely aware of the risks that may result from cartographic overgeneralization. Descriptive statistics, such as dispersion of represented values, may or may not reveal the actual extent of the impact of these errors on users' perception of maps.

Different methods are used for the mapping of population and other point data, for example, locations of objects, crime or fire events, morbidity, and mortality [2]. If there is some uncertainty of location or the data are sensitive, the data are usually aggregated by spatial units. Depending on the characteristics of the represented phenomenon (level of uncertainty of location, spread, sensitivity, etc.) and on the purpose of the visualization, the size of the units can vary from very small to very large. Small units such as cells of a regular grid that are often used for aggregation of events allow for relatively precise

representation. However, the cells of a regular grid can only be referred to by abstract identifiers. Thus, such maps are inconvenient, especially for non-professional users who are accustomed to common choropleth maps that represent named administrative units. In addition, in some cases using small aggregation units results in possibilities for the divulging of sensitive personal data. On the other hand, aggregation to a smaller number of standard administrative units results in traditional choropleth maps that are easier to compile and more user-friendly.

However, such aggregated maps convey a distorted image of the phenomena represented. For example, in the case of population maps, fallacies are due to uneven distribution of population over the area—in each administrative unit there may be relatively highly populated settlements as well as large areas of arable land or forest with no residents. Average values in such cases do not represent the actual situation. The less even the spatial distribution, the larger the aggregation unit, and the greater the difference will be between data represented on maps at different aggregation levels. We term the phenomenon in which map users get incorrect information from maps perceptual fallacy (similar to the phenomenon of ecological fallacy described by Robinson [3]). It is illustrated by two pairs of map fragments representing the same five classes of population density at different levels of aggregation. In the first case, the patterns of the  $1 \times 1$ -km grid and of the same data aggregated by larger administrative units are approximately the same (Figure 1a). For the second fragment with more heterogeneous distribution, they emerge looking starkly different (Figure 1b) and the generalised visualization is misleading. Some territories with extreme class values at a detailed level are assigned intermediate class values on a generalized map. It is very difficult to predict where such distortions are apt to occur.



**Figure 1.** Population represented at different levels of detail:  $1 \times 1$  km cells and administrative units, similar pattern (a);  $1 \times 1$  km cells and administrative units, different pattern (b).

Many research papers are devoted to problems of spatial units in density mapping: modifiable spatial units [4–8]; disaggregation of spatial data [9,10]; aggregation of spatial data; and spatial uncertainty [11–13]. The impact of the data aggregation level on the accuracy of mapping has been extensively discussed in the field of health geography. It

is commonly agreed that the higher the aggregation level, the lower the statistical power of the analysis [9,10,14]. A similar approach is used in the geography of crime [15,16]. Gerell demonstrated that the smallest types of units are of the greatest importance for analysis of arson but found little difference between large and medium-sized geographical units (1.2 and 0.4 square kilometers, respectively) [17]. Shi [12] concluded that there is no effective way to estimate, represent, and present uncertainty resulting from data aggregation. It is recommended that individual data be used whenever possible. Various methods of density mapping such as dasymetric maps and cartograms [18–21] that have been proposed result in more accurate representation, but there is an obvious lack of studies on the effect of spatial aggregation error on map perception. Cartographic and visualization approaches to representing local spatial statistics are extensively described by Nelson and Brewer who also proposed a method that ‘provides guidance to researchers in selecting the most appropriate scales to aggregate, analyze, and represent data for problem-specific analyses’ [22]. We appreciate the analytical approach taken by these authors; however, we find their cartographic results to be sophisticated to a point that makes them difficult to interpret for inexperienced map makers.

Simple cartographic representations, such as choropleth maps, are not usually created by professional cartographers. Everyone with some knowledge of GIS can represent data on a choropleth map. A small qualitative survey conducted in the form of a semi-structured interview in 2019 revealed that 16 GIS specialists, teachers, data analysts, and a journalist tend to trust in the general correctness of choropleth maps they often create. They all said they have never thought about possible substantial deviations from spatial patterns of primary data due to aggregation of values in large units. It is paradoxical that when asked what is the main purpose of map communication, they all specifically indicated spatial pattern. Moreover, they were all informed about ecological inference fallacy but never related it with significant changes of original spatial pattern. A more extensive survey with 140 participants was conducted in 2021. It is described in Section 3.1. The results are compared to actual pattern of perceptual fallacy of population density map of Lithuania in the Section 3.3.

The pattern seen on a map depends on the chosen method of classification, but users, including professional geographers, generally did not care about the method used. They also expected that the choropleth maps display more or less the same patterns at different levels of aggregation (Figure 2).

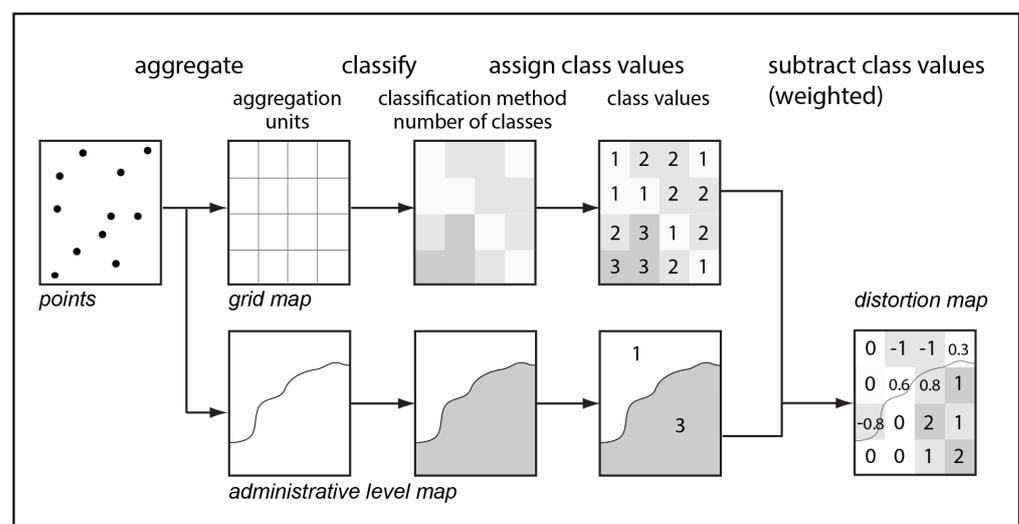


Figure 2. Calculation of perceptual fallacies.

The aim of this research was to find a way to evaluate the level of misinterpretation of data for particular areas of choropleth maps that are derived from known detailed data.

In this paper, we focus on the **problem** of quantitative evaluation of deviation that would somehow relate to *human perception*:

- (a) Where there are deviations from expected relative (class/rank) values, and;
- (b) How much perception of spatial distribution is distorted due to these deviations.

## 2. Materials and Methods

We propose a method of visualization that helps to determine to what extent perception of map information would be distorted due to the aggregation of individual data to larger units as compared to small standardized units. The method is based on quantitative evaluation of erroneous information on class values deduced from choropleth maps compared to those of a detailed grid map. The method is hereinafter also referred to as VPF—visualized perceptual fallacy. It allows for visual representation of areas where the user is effectively given wrong information on the class of the represented phenomenon (e.g., population density, crime, morbidity). Visualizations of fallacy very clearly show how misleading decisions can be when they are based on information, derived from points or small-scale choropleth maps, which represent data aggregated into administrative units. They also reveal the spatial distribution of fallacies over the territory of concern.

The variable input data used in the VPF method are:

1. The benchmark dataset (point data aggregated into small spatial units sufficient to adequate representation of the visualized phenomenon);
2. The dataset of the larger aggregation units;
3. The number of classes;
4. Break values that depend on the chosen method of classification for both datasets.

The process of calculation is represented in Figure 2 and is rather simple. It consists of the following steps:

1. Point data are aggregated into a benchmark grid and into larger aggregation units;
2. Both datasets are classified using the chosen method with the chosen number of classes;
3. The benchmark grid is intersected with larger aggregation units;
4. Class values of the larger aggregation units are compared to the benchmark grid cells. For the benchmark grid cells that intersect several larger aggregation units, class is calculated as weighted average and rounded to integer value;
5. The difference between larger aggregation units class value and benchmark grid class value is calculated:

$$PF_i = \sum_j a_j w_{ij} - c_i; C_i \cap A_j \neq O \quad (1)$$

where  $PF_i$  is a decimal measure of perceptual fallacy;  $c_i$  is the integer class value assigned to cell  $C_i$  from the benchmark dataset;  $a_j$  is the integer class value assigned to cell  $A_j$  from aggregated dataset;  $w_{ij}$  is percent of  $C_i$  overlapping with  $A_j$ .

There are several factors that are different for different spatial data and have to be considered.

1. The size and shape of a cell. The method works in the same way for cells of any shape, administrative units, or other polygons. The optimal size and shape of the cells depend on the nature of data. Any cartographic projection can be used as long as the maps used for comparison and the map of distortions are in the same projection;
2. The number of classes. While interpreting maps, people operate not the values, but the categories, such as ‘average’, ‘low’, ‘high’, or ‘normal’. The results of some pilot tests conducted as a part of the preparation for this research allow for the assertion that most users do well at perceiving and memorizing just three classes: ‘average’, ‘low’, and ‘high’ [23]. As stated in a study on visual perception of choropleth maps by Schiewe [24], ‘the number of classes should be kept as small as possible, [ ... ]. Defining a universal upper limit (e.g., 5 or 6 classes) is of course rather

- difficult [ . . . ].’ According to previous interviews and experiments with memorizing, seven classes appeared to be too many even for geography students. Five classes allow for representing the data with higher precision than just three, but further studies are needed to find out whether people really distinguish between the two ‘low’ and two ‘high’ categories. The method described can be applied for any number of classes;
3. The method of classification. Any method of classification can be used. The method operates with semantic class values. These values remain comparable across choropleth maps that use different number of classes ad different classification methods. Thus, different methods of classification may be applied for benchmark and for the aggregated dataset. An appropriate classification method must be selected depending on the nature of represented data;
  4. The color scheme. Both diverging and gradual color scales can be used. If they are applied correctly from the viewpoint of cartography and special users’ needs are considered, they would have no major impact on perception of the general spatial pattern.

The main characteristics of other groups of methods and the VPF method are summarized in Table 1. We did not include visual analytics [25] because (a) it is an approach rather than a method and (b) we did not investigate it in the context of usability (it is worth noting that we have no doubts concerning the utility of this approach).

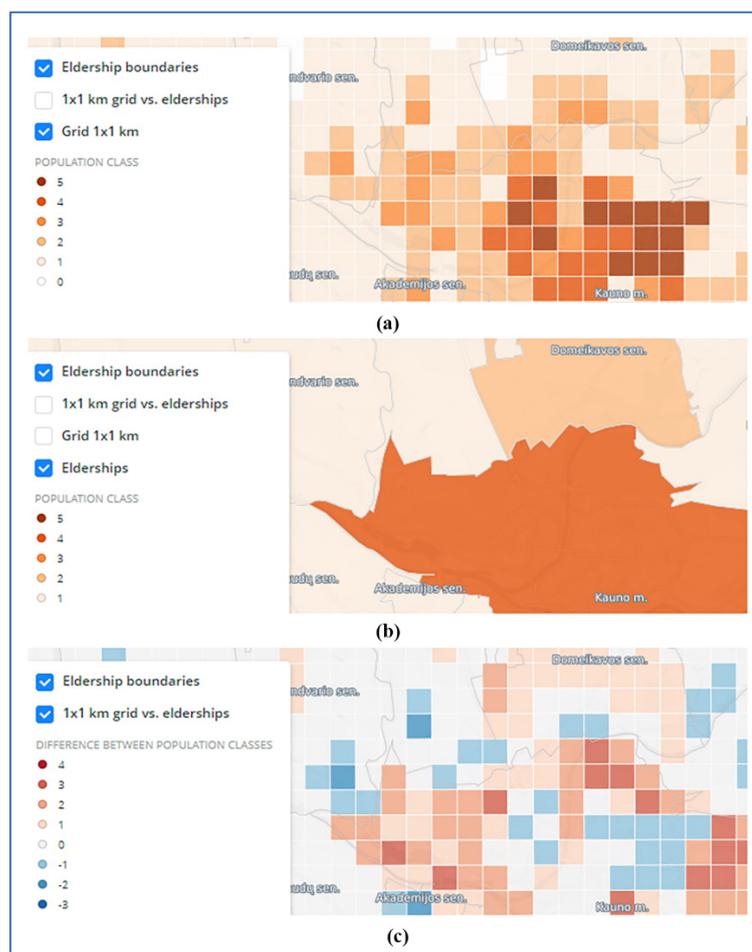
**Table 1.** Comparison of quantitative methods of assessment of problems resulting from spatial aggregation.

Methods	<i>Descriptive Statistics</i>	<i>Evaluation of Data Stability in Aggregation Structures [19]</i>	<i>VPF Method</i>
<b>Approach</b>	Statistical (standard deviation or dispersion index for an aggregation unit)	Statistical–visual (local indicators of spatial association)	Direct comparison of class values assigned to mapped units
<i>Cartographic visualization of fallacies</i>	No	Yes (requires additional interpretation)	Yes (direct)
<i>Takes map perception into account</i>	No	No	Yes
<i>Precision</i>	Precise values	Similarity values	Relative ordinal values
<i>Usability of results</i>	Requires interpretation	Requires additional analysis and interpretation	Usable without specialized knowledge
<i>Main disadvantages</i>	Does not represent spatial distribution of fallacies Evaluates possible impact of fallacy on data analysis, but not on perception	Complexity	Shows generalized picture (may depend on the method of classification)
<i>Main advantages</i>	Precise and commonly used	Educates users	Yields immediately usable maps of fallacies

The VPF method is based on a presumption that ordinary map users intuitively compare class values represented on maps and memorize relative characteristics, for example, ‘low’ values compared to ‘average’ or ‘very high’ values. This observation requires more extensive methodological reasoning that is outside the scope of this paper, but it is obvious that use of ordinal values is very common when population density issues

are discussed. The results of the above-mentioned qualitative survey also facilitate the assertion that people perceive and memorize classes of values represented on choropleth maps as a limited number of ranks from ‘lowest’ to ‘highest’.

The VPF method yields maps of differences between representations by relative class values (Figure 3). Negative values (in blue) show that the class of the cell in representation of aggregated data is lower than in the benchmark representation, for example; a cell that was assigned a class value of 3 (‘average’) is within larger area with class value of 1 (‘low’). Positive values (shown in orange color) show how much class values increase after aggregation and reclassification.



**Figure 3.** Fragment of map of (a) class values for benchmark 1 km<sup>2</sup> grid cells; (b) class values for elderships (NUTS3 units of ca.100 km<sup>2</sup>); (c) PF values; published on [carto.com](https://carto.com) (accessed on 28 November 2021).

### 3. Case Study

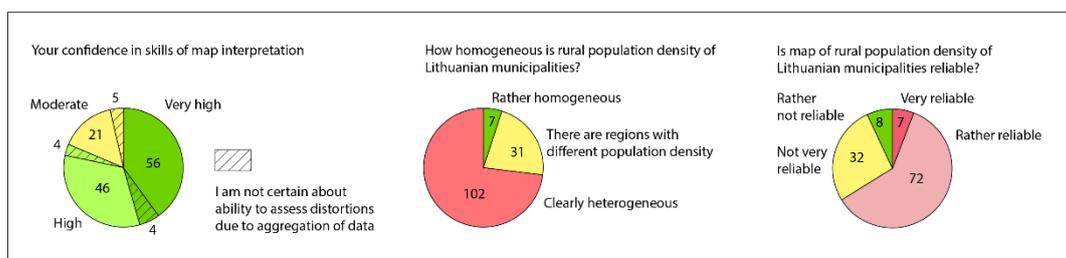
#### 3.1. Rationale for the Case Study

To provide proof for the concept, the proposed method was applied for evaluation and mapping of the perceptual fallacies in population density maps that are widely used in Lithuania. The choice of the population density of Lithuania to exemplify the method is rooted in the repeated misinterpretation of small-scale population density maps in recent years. Depopulation in Lithuania is due to growing emigration and a low fertility rate that does not ensure population replacement. Since 1990, Lithuania has had population loss indicators that are among the highest in Europe. The decline of population will continue in the coming decades [26]. In recent decades, researchers have observed and analyzed sparsely populated territories that are expanding [27–32]. The data of population

density in Lithuania are compiled from census data and from data of the state registers that adequately reflect the actual situation, but population density maps of Lithuania are usually compiled at small scales (1:1,000,000 or smaller). Censuses of the population and/or registries data normally refer to address points. In order to prevent the leaking of personal information, individual data are aggregated by administrative units and, since 2012, by statistical grid cells. Detailed population information has been used extensively in crime mapping for Lithuania [33,34]. Nevertheless, in 2019, most population density maps for Lithuania used for education and for evaluation of the demographic situation show information that was aggregated by second level administrative units (municipalities of 40 to 2216 square kilometers). Maps that represent administrative units are presented to politicians and the wider public and serve to form ‘the common view’ about territorial differences with respect to population.

Detailed grid maps can be found in the portal of *Statistics Lithuania*, but it will take time for public users to get used to this method of visualizations. As the majority of users of such maps are not cartography professionals, it is important to ensure that they interpret the maps correctly. This is a sensitive issue in Lithuania, where depopulation due to emigration and the decreasing birth rate has been gaining pace. Planning processes dedicated to preventing depopulation also make use of this information along with the precise individual data. The problem is that the majority of people interested in distribution of the population obtain their information not from databases, but from maps. Small-scale choropleth maps are quickly fathomed and memorized. All possible measures must therefore be taken in order to minimize potential misinterpretation of the information presented.

A pilot survey of 140 GIS/geography professionals in Lithuania was conducted prior to the experiment. The purpose was to understand the peculiarities of their mental maps of distribution of rural population. The respondents were employees of public and private organizations engaged in GIS and mapping: geography teachers (99; 71%), researchers in geosciences and social sciences (17; 12%), and GIS engineers (24; 17%). Out of them, 135 (96.4%) were confident about their ability to use and assess map information and 119 (85%) answered that they were informed about ecological inference fallacy. Then they were asked questions, two of which were relevant to our experiments: (a) about their opinion concerning homogeneity of rural population density in Lithuania and (b) whether they consider the demonstrated map (the same as presented in Figure 5) appropriate for representation of rural population density. There was a possibility provided to add comments as free text. Only 7 respondents (5.7%) stated that rural population density is more or less homogeneous across the country (Figure 4). The majority (102; 72.9%) stated that there are large regional differences and more than a half of them even provided various explanations why—whereas this was actually not the case in point of fact. Thus, the results of the survey supported our assumption that municipality level choropleth charts commonly seen in school atlases and in mass media may result in fallacious mental maps that persist despite availability of accurate data and detailed maps.



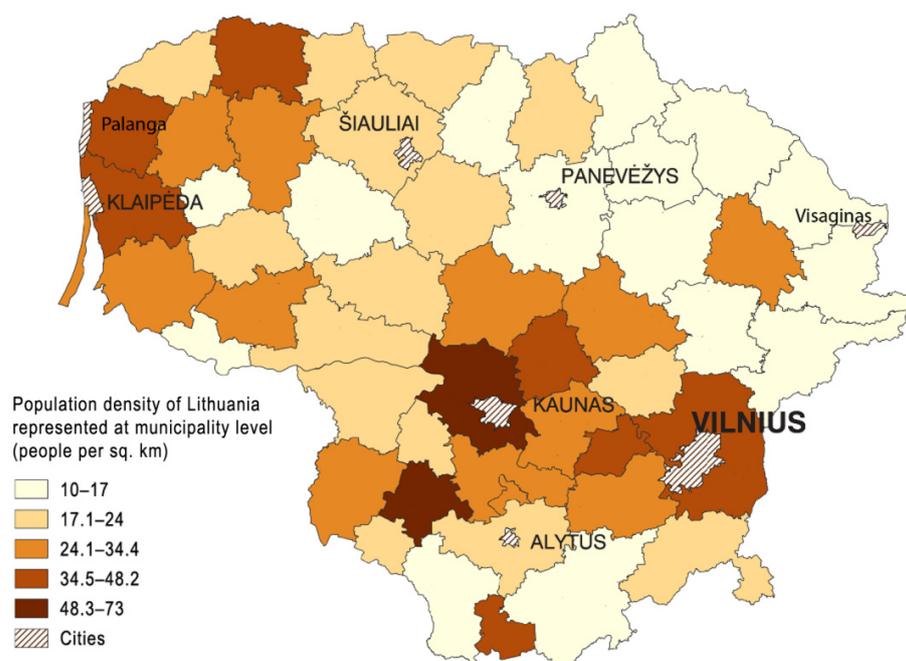
**Figure 4.** Results of survey concerning assumptions about population density ( $n = 140$ ).

### 3.2. Experiment Design

**Data and aggregation units.** The method was applied to two aggregation levels of estimated population data of the Republic of Lithuania in 2020.

Our benchmark dataset was a  $1 \times 1$  km rectangular grid. We assumed that people commonly move in the vicinity of their homes within a circle whose radius is at least 0.5 km. It corresponds to about a 5-min walking distance to the nearest facilities [35]. It would therefore not make sense to analyze population data with higher precision, even though the census data is available in  $100 \times 100$  m cells. For the same reason standard rectangular grid cells were used instead of more compact hexagonal cells. This dataset was used as a reference dataset to determine deviations in two population density choropleth maps.

1. The eldership level map (encompassing all elderships). Elderships are the smallest administrative divisions of Lithuania with an area from 0.7 to 594.8 km<sup>2</sup>; average 116.2 km<sup>2</sup>;
2. The municipality level map (rural municipalities with an area from 40.0 to 2213.9 km<sup>2</sup>; average 1080.1 km<sup>2</sup>). For this map, urban areas with extreme density values (the seven largest cities) were removed as outliers to the issues at hand (Figure 5). Inclusion of cities would only mean higher levels of distortion. On the other hand, people commonly understand that towns are by their nature populated many times more densely than rural areas (in Lithuania, correspondingly averaging 984 and 31 inhabitants per km<sup>2</sup>, respectively). In the discourse of depopulation, urban areas need to be analyzed and discussed as a separate issue.



**Figure 5.** Rural population density of Lithuania represented at municipality level.

The population dataset is publicly available online at the portal of official statistics in the form of an interactive atlas [36] and as a downloadable set of grids for the entire territory of the country. Due to personal data protection requirements, public point data have been aggregated by standard cells starting with the  $100 \times 100$  m.

Classification of data. As the purpose of the study was to demonstrate not the absolute but the perceived difference in population density values, we classified the density values into five classes using the Jenks natural breaks classification method implemented in ArcGIS 10.2.2. It should be noted that this method may not be the best option for population density mapping, but for our purpose the method selected is not of primary importance. It is also important to note that break values differ for different implementations of the same classification method.

The classes were assigned the numeric values: 0—unpopulated (in the reference grid only); 1—relatively lowest population density; 2—low, 3—average, 4—high, and

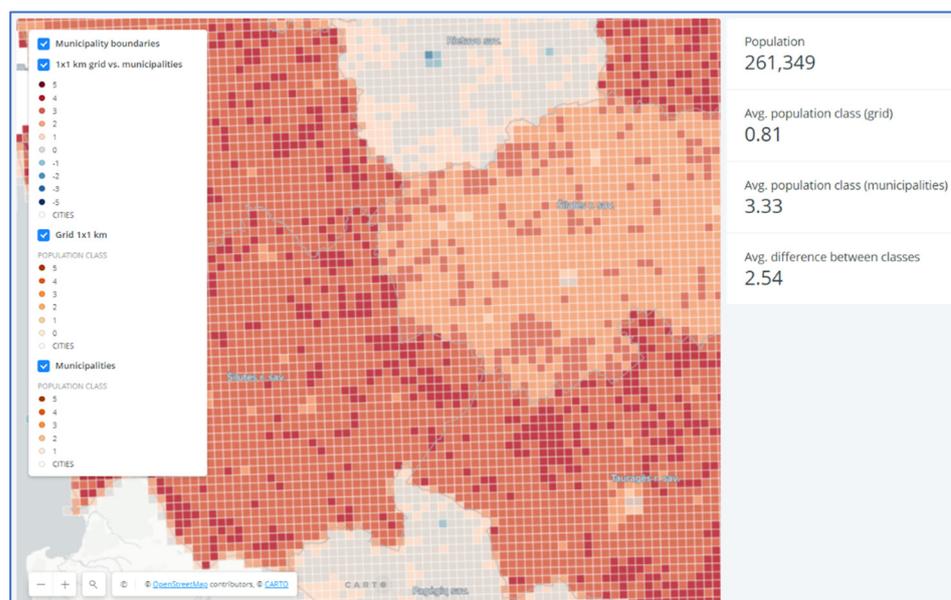
5—relatively highest density. The actual density value ranges for each class were different in all maps, but the five classes represent the way people actually read maps: the value is either ‘average’ or moderately/significantly lower or higher than the average. The class values of both administrative maps were then resampled to the  $1 \times 1$  km grid and compared to the values of the reference grid.

Visualization. In case of population density, a gradient scale is more popular and semantically justified—the values increase from zero to the highest without a change in any other quality. Pale to dark orange represent low to high density class values.

### 3.3. Outcomes

Two interactive maps were published on the *Carto* platform (<https://carto.com/>, accessed on 28 November 2021). They display PF values for  $1 \text{ km}^2$  grid cells against:

- Eldership level (<https://vucarto.carto.com/builder/e86912d2-c776-449d-8a27-4423b718ef55/embed>, accessed on 28 November 2021); and
- Municipality level (<https://vucarto.carto.com/builder/6f48cac4-5770-482f-aaf2-1b41561553d0/embed>, accessed on 28 November 2021) choropleth map layers and statistics of class values for visible area (Figure 6).



**Figure 6.** Interactive map of perceptual fallacies.

Interactive maps can be explored in order to understand peculiarities of deviations. The spatial patterns investigated in this case study are unexpected for people who are used to working with population data and maps.

The existing difference between the class value assigned to a unit on an administrative map and the smaller reference grid values (‘true’ values) means that people would attribute a wrong class value to a grid cell: For example, they might perceive a part of a particular territory as densely populated whereas it is not. Quantitative difference (in units) shows the size of the gap between the estimated and perceived values. Deviations from the reference grid class values were calculated for each reference grid cell. For the municipality level map, the average PF value was +2.27, which is surprisingly high. The maximum positive deviation was five—this was the case for unpopulated cells that were nested in aggregated units with the ‘highest’ population density. The maximum negative deviation was two. Negative deviations were rare and mostly appeared for small towns (mostly less than 10,000 inhabitants) covering up to 4 reference grid cells. Considering that urban areas were not analyzed in this case, they were ignored.

Deviations of the municipality level map clearly show in which areas the perception of density was mostly distorted:

- (a) The areas surrounding the five largest cities (Vilnius, Kaunas, Klaipėda, Šiauliai and Panevėžys, Lithuania) and, surprisingly, several municipalities which have several times smaller population and no bigger towns. It means that although there is a trend of higher positive deviations to be observed near cities and towns, this trend is far from universal. The deviations were determined more by the configuration of administrative districts and by a specific mosaic of land use in each district;
- (b) Central and Western Lithuania with prevailing deviations of +3 and +4 over the entire area. These distortions are responsible for the wrong assumption repeatedly used (and even explained by higher fertility of land) by Lithuanian researchers who contend that rural population density is relatively much higher in middle Lithuania.

Thus, this pattern of distortion may also be the explanation why 46% of respondents of the survey described in Section 3.1 indicated in their free text comments that the density of rural population is highest in precisely these regions (false).

The remainder of the municipalities contains about 97% of cells with positive deviation of 1 to 3 density classes.

There are only two municipalities—Rietavas and Pagėgiai, Lithuania—with prevailing zero distortion. This fact cannot be explained by specificity of population distribution as it is very typical for all rural municipalities.

Almost the entire rural territory of Lithuania in the municipality level map was assigned too high a class value, whereas too low a value was found in about just 0.1% of the territory. The fact that over 97% of the map represents semantically incorrect information qualifies this type of population density map as unusable.

The eldership map distorted population density data to a much lesser extent: the average PF value was only +0.27. Still, class values that are too high were assigned to about 27% of the territory. The distribution of distortions was more even. In the eldership level map, the fallacies depend less on proximity to cities and may also be observed in sparsely populated areas. Further analysis of the distribution of fallacies may lead to assumptions regarding the demographic situation in particular areas.

#### 4. Discussion

The described method appears to be applicable in other fields where point data are aggregated. It can be fully automated and used as an analytic tool by mapping experts. Creating a tool was not a purpose of this study, and appropriate automation of the task would require some efforts. Such a tool should provide a possibility to upload custom datasets for benchmark and for tested choropleth maps; to select the method of classification, ideally with a possibility of applying different methods for benchmark and test units. A graphical user interface also needs to be designed that simplifies the choice and use of the parameters and visualizes the results. Such a tool would make the evaluation of perceptual distortions rather easy for non-specialist users, especially when they have to visualize big data with unknown distribution. Upon implementation, it is intended that usability analysis and more case studies will follow.

In this study, we only demonstrated the results with one classification algorithm. It is worth mentioning that application of two other popular methods—equal interval and quartile—yielded slightly different cartographic results. Differences were due to specificity of concrete data distribution. The impact of the map's chosen classification method on perception can also be analyzed and evaluated, but it is beyond the scope of this paper.

The described method of comparison and fallacy mapping can be applied for any sets of maps where point data are aggregated to different units: unemployment, crime, health or other maps. Before making the decision to choose a larger aggregation unit we advise visualizing and comparing possible perceptual distortions. If the phenomenon is evenly spread, the differences can be minor and evenly distributed—which means, in effect, that more general maps can be used. This option is usually cheaper and more easily perceived

by inexperienced map users. If the resulting discrepancies are unacceptable, it may be a sign that smaller aggregation units should be used.

In this paper, we demonstrated the feasibility of the VPF method using particular spatial aggregation units and the population dataset of Lithuania. Evaluation of perceptual fallacies, perhaps with different initial assumptions and different variables, is applicable to any spatial areals and to any set of individual data that are aggregated to these areals. It is also applicable with any method of classification used for visualizations provided that the same classification method is used for the compared maps.

For any type of phenomenon or pattern of administrative boundaries, the spatial pattern of deviations that may lead to wrong perception cannot be understood and controlled using known parameters. For example, we cannot make any inference about what this pattern would look like for crime or mortality data. However, the pattern is quickly seen in the fallacy maps generated by the VPF method. A quick glance at the distortion map gives a fair idea about the extent and volume of misrepresentations. For deeper understanding, statistics of deviations can easily be calculated for each map area.

We do not expect that because of demonstrable distortions, choropleth population density maps will anytime soon be replaced by grid or interpolated density maps. But before disseminating small-scale choropleth maps, the extent of the distortions must be evaluated and suitability for purpose assessed. Obviously, the potential damage done by misrepresentation and corresponding quality requirements differs markedly depending on whether the maps are being used for research, decision making, education, or just providing information to the public.

## 5. Conclusions

In this paper, we proposed a cartographic method that allows for the assessing of impact of aggregation on perception of a spatial distribution pattern seen on a map.

The method is applicable to any data that are aggregated in variable areal units. Perceptual fallacy can be presented visually as a map of distortions, or described by statistics of distortions. The statistics of distortions can be used for correction of automatically assigned class values.

We demonstrated the method with data of rural population density in Lithuania. Based on the results of the test, a decision can be made regarding suitability of an aggregated map for correct perception of the spatial distribution seen on the benchmark map—in the given case, the choropleth map was suitable at eldership level but became misleading at the municipality level. Common use of the municipality level map may be the cause of widespread fallacious beliefs concerning spatial distribution of the rural population.

**Author Contributions:** Conceptualization, Giedrė Beconytė and Aurelija Šturaitė; methodology, Giedrė Beconytė and Andrius Balčiūnas; software and visualisation, Andrius Balčiūnas; validation, Rita Viliuvienė; investigation, Aurelija Šturaitė and Rita Viliuvienė; writing—original draft preparation, Giedrė Beconytė; writing—review and editing, Andrius Balčiūnas and Rita Viliuvienė. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Thanks are due to the managers of the national spatial information portal ([www.geoportal.lt](http://www.geoportal.lt), accessed on 28 November 2021) and to Statistics Lithuania ([www.stat.gov.lt](http://www.stat.gov.lt), accessed on 28 November 2021) for enabling unrestricted access to interoperable data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Monmonier, M. *How to Lie with Maps*, 2nd ed.; University of Chicago Press: Chicago, IL, USA, 2014.
2. Jovanovic, J.; Zivkovic, D. Cartographic Modeling of the Population Density in the Function of Research of Spatial-demographical Relations. *Geogr. Inst. Jovan Cvijin SASA Collect. Pap.* **2005**, *54*, 115–127. [[CrossRef](#)]
3. Robinson, W.S. Ecological Correlations and the Behavior of Individuals. *Am. Sociol. Rev.* **1950**, *15*, 351–357. [[CrossRef](#)]
4. Openshaw, S. *The Modifiable Areal Unit Problem*; Geo Books: Norwich, UK, 1984.
5. Root, E.D. Moving Neighborhoods and Health Research Forward: Using Geographic Methods to Examine the Role of Spatial Scale in Neighborhood Effects on Health. *Ann. Am. Assoc. Geogr.* **2012**, *102*, 986–995. [[CrossRef](#)] [[PubMed](#)]
6. Manley, D. Scale, Aggregation, and the Modifiable Areal Unit Problem. In *Handbook of Regional Science*; Springer: Berlin, Germany, 2014; pp. 1157–1171.
7. Xu, P.; Huang, H.; Dong, N.; Abdel-Aty, M. Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. *Accid. Anal. Prev.* **2014**, *70*, 110–120. [[CrossRef](#)] [[PubMed](#)]
8. Karsznia, I.; Gołębiowska, I.M.; Korycka-Skorupa, J.; Nowacki, T. Searching for an Optimal Hexagonal Shaped Enumeration Unit Size for Effective Spatial Pattern Recognition in Choropleth Maps. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 576. [[CrossRef](#)]
9. Luo, L.; McLafferty, S.; Wang, F. Analyzing Spatial Aggregation Error in Statistical Models of Late-stage Cancer Risk: A Monte Carlo Simulation Approach. *Int. J. Health Geogr.* **2010**, *9*, 51. [[CrossRef](#)]
10. Shi, X.; Miller, S.; Mwenda, K.; Onda, A.; Rees, J.; Onega, T.; Gui, J.; Karagas, M.; Demidenko, E.; Moeschler, J. Mapping Disease at an Approximated Individual Level Using Aggregate Data: A Case Study of Mapping New Hampshire Birth Defects. *Int. J. Environ. Res. Public Health* **2013**, *10*, 4161–4174. [[CrossRef](#)]
11. Shi, X. Evaluating the Uncertainty Caused by P.O. Box Addresses in Environmental Health Studies: A restricted Monte Carlo Approach. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 325–340. [[CrossRef](#)]
12. Shi, X. A GeoComputation Process for Characterizing the Spatial Pattern of Lung Cancer Incidence in New Hampshire. *Ann. Am. Assoc. Geographers.* **2009**, *99*, 521–533. [[CrossRef](#)]
13. Griffith, D.A. Uncertainty and Context in Geography and GIScience: Reflections on Spatial Autocorrelation, Spatial Sampling, and Health Data. *Ann. Am. Assoc. Geogr.* **2018**, *6*, 1499–1505. [[CrossRef](#)]
14. Jacquez, G.M.; Waller, L.A. The Effect of Uncertain Locations on Disease Cluster Statistics. In *Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing*; Mowrer, H.T., Congalton, R.G., Chelsea, M.I., Eds.; Sleeping Bear Press: Chelsea, MI, USA, 1999; pp. 53–64.
15. Andresen, M.A.; Malleson, N. Spatial Heterogeneity in Crime Analysis. In *Crime Modeling and Mapping Using Geospatial Technologies*; Leitner, M., Ed.; Springer: Dordrecht, The Netherlands, 2013; pp. 3–23.
16. Boessen, A.; Hipp, J.R. Close-ups and the Scale of Ecology: Land Uses and the Geography of Social Context and Crime. *Criminology* **2015**, *53*, 399–426. [[CrossRef](#)]
17. Gerell, M. Smallest is better? The Spatial Distribution of Arson and the Modifiable Areal Unit Problem. *J. Quant. Criminol.* **2017**, *33*, 293–318. [[CrossRef](#)]
18. Petrov, A. One Hundred Years of Dasymetric Mapping: Back to the Origin. *Cartogr. J.* **2012**, *49*, 256–264. [[CrossRef](#)]
19. Brewer, C.A.; Eicher, C.L. Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartogr. Geogr. Inf. Sci.* **2013**, *28*, 125–138. [[CrossRef](#)]
20. Zandbergen, P.A.; Ignizio, D.A. Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. *Cartogr. Geogr. Inf. Sci.* **2013**, *37*, 199–214. [[CrossRef](#)]
21. Ricker, B.; Kraak, M.-J.; Roth, R. The Promise of Dasymetric Maps for Monitoring Progress towards the United Nations Sustainable Development Goals. Abstracts of the International Cartographic Association. In Proceedings of the 30th International Cartographic Conference, Florence, Italy, 14–18 December 2021. [[CrossRef](#)]
22. Nelson, J.K.; Brewer, C.A. Evaluating Data Stability in Aggregation Structures across Spatial Scales: Revisiting the Modifiable Areal Unit Problem. *Cartogr. Geogr. Inf. Sci.* **2017**, *44*, 35–50. [[CrossRef](#)]
23. Šturaitė, A. Lietuvos Gyventojų Tankumo Kartografavimo Metodų Lyginamoji Analizė. Master's Thesis, Vilnius University, Vilnius, Lithuania, 2016. (Unpublished work).
24. Schiewe, J. Empirical Studies on the Visual Perception of Spatial Patterns in Choropleth Maps. *K J. Cartogr. Geogr. Inf.* **2019**, *69*, 217–228. [[CrossRef](#)]
25. Andrienko, G.; Andrienko, N.; Demšar, U.; Dransch, D.; Dykes, J.; Fabrikant, S.; Jern, M.; Kraak, M.-J.; Schumann, H.; Tominski, C. Space, Time and Visual Analytics. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 1577–1600. [[CrossRef](#)]
26. Berzins, A.; Zvidrins, P. Depopulation in the Baltic states. *Lith. J. Stat.* **2011**, *50*, 39–48. [[CrossRef](#)]
27. Vaitiekūnas, S. Lietuvos gyventojų tankumas ir koncentracija. *Tiltai* **2004**, *4*, 21–39.
28. Stanaitis, S. Social, Economic and Demographic Changes of Rural Areas in Lithuania. In *Changing Functions of Rural Areas in the Baltic Sea Region*; Polish Academy of Science: Warsaw, Poland, 2004; pp. 45–58. [[CrossRef](#)]
29. Vaitiekūnas, S. *Lietuvos Gyventojai per du Tūkstantmečius*; Mokslo ir Enciklopedijų Leidybos Institutas: Vilnius, Lithuania, 2006.
30. Daugirdas, V.; Baubinas, R. Retai apgyventos teritorijos Lietuvoje. 1. Teritorinės sklaidos aspektai. *Ann. Geogr.* **2008**, *40*, 28–37.
31. Stanaitis, S.; Stanaitis, A. Lietuvos miestų ir jų gyventojų skaičiaus bei struktūros kaita nepriklausomybės metais. *Tiltai* **2013**, *3*, 69–82.

32. Daugirdas, V.; Burneika, D.; Kriauciūnas, E.; Ribokas, G.; Stanaitis, S.; Ubarevičienė, R. *Lietuvos Retai Apgyventos Teritorijos*; Lietuvos Socialinių Tyrimų Centras: Vilnius, Lithuania, 2014.
33. Eismontaite, A.; Beconytė, G. Prognosis of Crime Rate Growth in Lithuania, Based on its Spatial Distribution and Relationship with Registered Unemployment Rates. *Filos. Sociol.* **2011**, *22*, 236–245.
34. Vasiliauskas, D.; Beconytė, G. Cartography of Crime: Portrait of Metropolitan Vilnius. *J. Maps* **2016**, *12*, 1236–1241. [[CrossRef](#)]
35. Levasseur, M.; Cohen, A.A.; Dubois, M.F.; Génereux, M. Environmental Factors Associated With Social Participation of Older Adults Living in Metropolitan, Urban, and Rural Areas: The NuAge Study. *Am. J. Public Health* **2015**, *105*, 1718–1725. [[CrossRef](#)] [[PubMed](#)]
36. Interaktyvus Atlasas. Oficialiosios Statistikos Portalas. 2017. Available online: <https://osp.stat.gov.lt/interaktyvus-atlasas> (accessed on 29 November 2021).