DGfK

Distortion Effects in Equal Area Unit Maps

Jochen Schiewe[1] ●

Abstract
Maps that correctly represent the geographic size and shape of regions, taking into account scaling and generalization, have the disadvantage that small regions can easily be overlooked or not seen at all. Hence, for some map use tasks where small regions are of importance, alternative map types are needed. One option is the so-called equal area unit maps (EAUMs), where every enumeration unit has the same area size, possibly also the same basic shape such as squares or hexagons. The geometrical distortion of EAUMs, however, leads to a more difficult search for regions as well as a falsification of topological relationships and spatial patterns. To describe these distortions, a set of analytical measures is proposed. But it turns out that the expressiveness of these measures is rather limited. To better understand and to model the influence of distortions, two user studies were conducted. The study on the search in EAUMs (also with the aim of reconstruct the search strategies of the users) revealed how important it is to consider the local topology (e.g. corner or border positions of regions) during the generation process. With regard to pattern identification, it could be shown that EAUMs significantly increase the detection rate of local extreme values. On the other hand, global lateral gradients or geostatistical hot spots often get blurred or even lost. As a consequence, a task-oriented selection of map types and further developments are recommended.

Keywords Equal area unit maps · Choropleth maps · Pattern identification · Usability

Kurzfassung
Karten, die unter Beachtung des Maßstabes und der Generalisierung die geographische Größe und Form von Regionen korrekt wiedergeben, haben den Nachteil, dass kleine Regionen leicht übersehen oder gar nicht gesehen werden. Daher werden für einige Kartennutzungsaufgaben, bei denen kleine Regionen von Bedeutung sind, alternative Kartentypen benötigt. Eine Option sind sogenannte Equal Area Unit Maps (EAUMs), bei denen jede Einheit dieselbe Flächengröße und zumeist auch dieselbe Grundform wie Quadrate oder Hexagone besitzt. Die geometrische Verzerrung von EAUMs führt jedoch zu einer schwierigeren Suche nach Regionen sowie zu einer Verfälschung topologischer Beziehungen und räumlicher Muster. Um diese Verzerrungen zu beschreiben, wird ein Satz von analytischen Maßzahlen vorgeschlagen. Es zeigt sich jedoch, dass die Aussagekraft dieser Maßzahlen eher begrenzt ist. Um den Einfluss von Verzerrungen besser verstehen und modellieren zu können, wurden zwei Anwenderstudien durchgeführt. Die Studie zur Suche in EAUMs (auch mit dem Ziel, die Suchstrategien der Benutzer nachzuvollziehen) ergab, wie wichtig es ist, die lokale Topologie (z. B. Eck- oder Randpositionen von Regionen) während des Generierungsprozesses zu berücksichtigen. In Bezug auf die Identifikation räumlicher Muster konnte gezeigt werden, dass EAUMs die Erkennungsrate lokaler Extremwerte signifikant erhöhen. Andererseits verschwimmen globale laterale Gradienten oder geostatistische Hot Spots häufig oder gehen sogar verloren. Infolgedessen werden eine aufgabenorientierte Auswahl von Kartentypen sowie weitere Entwicklungen empfohlen.

Schlüsselworte Equal Area Unit Maps · Choroplethenkarten · Mustererkennung · Gebrauchstauglichkeit

✉ Jochen Schiewe
jochen.schiewe@hcu-hamburg.de

1   Lab for Geoinformatics and Geovisualization (g2lab),
     HafenCity University Hamburg, Henning-Voscherau-Platz 1,
     20457 Hamburg, Germany

1 Introduction

Maps generally follow the principle of correctly representing the geographic size and shape of regions, of course, under consideration of scale and generalization. However, due to

Fig. 1 Example of equal area unit maps (EAUMs), composed of regular grid cells (middle) or hexagons (right), in comparison to original geographical regions (left)
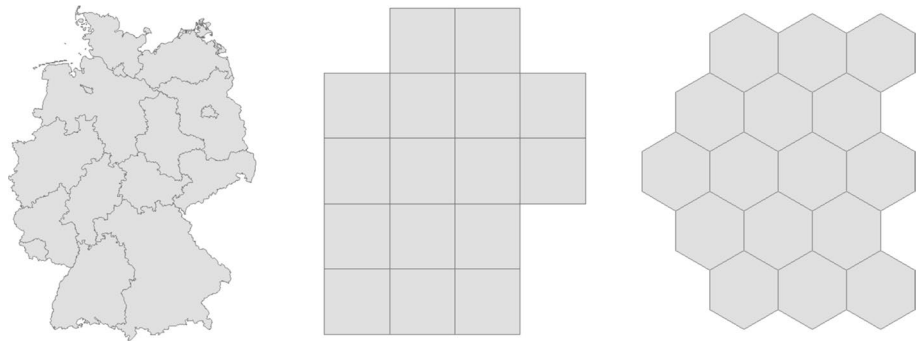
the heterogeneous differences in the size of the enumeration areas, this often leads to a very small regions appearing so small that they can be overlooked or not seen at all (e.g. Singapore is shown in a world map at a scale of 1: 90 million as an area of 0.3 mm by 0.5 mm). In choropleth maps, there are also the cognitively caused effect that larger regions are perceived as more dominant, although they have the same color value compared to the small regions, which leads to misinterpretations and wrong decisions (Slocum et al. 2009). Empirical studies have confirmed that the detection rates in the order of 60% for global extreme values in small areas within choropleth maps are significantly worse than those for large areas, which are in the order of 90% (Schiewe 2019).

As an alternative to the true-scale representation, Equal Area Unit Maps (EAUMs) have gained an increasing popularity in the recent past, especially in the news media.[1] EAUMs represent each enumeration area in the same size and possibly use uniform basic shapes (e.g. squares, rectangles, hexagons, or multi-hexagons—see Fig. 1). Chapter 2 gives a deeper insight into terminology related to EAUMs and methods for generating them.

The advantage of EAUMs is that for a variety of map use tasks all enumeration areas are shown with equal prominence. This makes it easy to look-up and compare values and to avoid the area size-bias. Assuming a certain size of the basic units, EAUMs can accommodate not only one color or hachure for every region, but also text, symbols or small diagrams, which enables the display of multivariate data. EAUMs are also motivating for use because they represent an unusual depiction of regions and produce a certain

surprise effect due to the false-scale representation. It can also be argued that the consistent use of clear shapes such as squares or hexagons is aesthetically pleasing.

Looking at the disadvantages, the generation of EAUMs leads to distortions concerning the distances between regional units, area sizes, shapes and topological relationships (such as loss and addition of neighborhoods and memberships) compared to the original geographical data. As a result, the search for specific regions and the interpretation of spatial distributions or patterns in the original data becomes more difficult and more error-prone. To describe the distortions in a more objective and quantitative manner, Chapter 3 presents selected geometrical and topological measures. From this, it can be concluded that these measures show a large variation within a map and do not produce a clear overall impression of distortions. Furthermore, the quite large set of analytical measures is not suitable for purposes such as comparing map designs or controlling an EAUM generation algorithm.

To better understand the relevance of distortions, this contribution also follows an empirical approach. While for some map types, such as value-by-area maps (Sun and Li 2010), studies on the effects of the distortions on the interpretability exist, there is still no empirical evidence for the cognitive effect of EAUMs.

A first user study (Chapter 4) deals with the effectiveness of the search in EAUMs but also with the derivation of typical search strategies that can be of interest for controlling the EAUM generation process. The second study (Chapter 5) focuses on the identification of patterns in EAUMs: firstly, the hypothesis is examined that patterns related to individual regions (such as local extreme values) are well recognized within EAUMs. The second hypothesis is related to the identification of other patterns such as North–South gradients or hot spots, which is much more difficult due to the inherent geometrical distortions. Chapter 6 summarizes the results of the two studies and gives recommendations for future developments in the EAUM environment.

---

2 Equal Area Unit Maps

2.1 Terminology

In the following I define an equal area unit map (EAUM) as a base map where every enumeration unit (geographic entity) is given the same size, possibly also the same shape (such as squares, rectangles or hexagons). Overlapping units are not allowed; however, gaps are possible. The arrangement of units is done according to one or more optimization criteria (e.g. minimizing the positional shift of centroids between geographic original and resulting EAUM).

It should to be noted that there is no common terminology in literature for this type of representation. One option is the term "equal area cartogram" (Braith 2015; Ordnance Survey 2020). Strictly speaking, a cartogram is not appropriate here, since a modification of shapes and sizes of units is not based on data values as defined by Slocum et al. (2009). On the other hand, the term "equal area cartogram" leaves out the term "map"—as Raisz (1934) does, who states that the "rectangular statistical cartograms" in his work are not maps.

If one takes the strict definition by Hake et al. (2002) into account, an EAUM does not fulfill the condition of a correct orthogonal projection; instead, it could be categorized as a map-alike product. On the other hand, taking into account the less strict definition of the International Cartographic Association (ICA 2003), it is applicable to use the term "map" since the primary relevance of spatial relationships is still given.

Sometimes, this map type is associated to the term "tile map" (e.g. McNeill and Hale 2017). However, there is also another (and more dominant) meaning that tile maps cut geographical space—without explicitly considering administrative regions—into regular elements to allow faster and individualized access (Peterson 2012). Other authors such as Eppstein et al. (2015) or Wongsuphasawat (2016) use the term "grid map". This suggests a regular raster, neglecting other unit representations such as hexagons.

Finally, EAUMs should not be confused with value-by-area maps (Dent 1999) or area cartograms (Slocum et al. 2009), in which the areas of each enumeration unit are scaled or sized proportionally to an attribute value (e.g. population).

2.2 Generation of EAUMs

The generation of EAUMs can be viewed as a so-called assignment problem, which is known from graph theory: given is a bipartite graph, which consists of two subsets with the same number of nodes, one partition for the original geographic input situation and the other for the output EAUM. While the nodes represent region centroids, the edges model the actual neighborhood of regions. The graph matching is now a combinatorial optimization problem that looks for a unique assignment between nodes of the two partitions, in which the overall costs of edge weights (e.g. the distances of centroids) must be minimized.

The conventional algorithm for weighted matchings in bipartite graphs is the Hungarian method, also called the Kuhn-Munkres algorithm (Kuhn 1955; Munkres 1957). It solves the assignment problem in $O(n^3)$ time (where n is the number of nodes). This algorithm was also used for generating map examples for the following study.[2] There are several specialized or extended forms of this algorithm as well as other approaches such as the simplex method (Barr et al. 1977). However, it is beyond the scope of this paper to give a comprehensive overview of the variety of these methods.

3 Distortions in EAUMs

To enable an objective and numerical description of distortions in EAUMs for absolute or relative comparison purposes, various measures are presented in the following (Sect. 3.1). These measures are then applied to three example maps in order to assess their expressiveness (Sect. 3.2).

3.1 Measures

The distortion effect between two maps can be described by several measures. For example, Nusrat and Kobourov (2016) use the categories statistical accuracy (preservation of thematic values), contiguity (presence of overlaps or gaps), geography (preservation of shape and size), and topology (especially preservation of neighborhoods). Alam et al. (2016) refer to distortions of geometry, topology and complexity. EAUMs are not dealt with in these publications. For our purposes, the following geometrical and topological measures have been defined:

Geometrical distortions:

- Distance distortion (G_DD): mean value of all distances between region centroids in geographical and EAUM map that correspond to each other's (normalized by distance of longest extension of total map, i.e. either in North–South or East–West direction).
- Single enumeration area size variance (G_SA): standard deviation of area sizes of geographical regions (normalized by division by mean value of area sizes). The larger the standard deviation is, the larger the distortion within an EAUM (with area standard deviation of Zero) will be.

Table 1 Geometrical and topological distortions of EAUMs for three example areas

Area	Australia		Germany		USA	
No. of regions	7		16		49	
EAUM basic shape	HEX	REG	HEX	REG	HEX	REG
Geometrical distortions						
G_DD	0.19	0.20	0.01	0.01	0.15	0.16
G_SA	0.91	0.91	0.81	0.81	0.77	0.77
G_TA	0.11	0.08	0.05	0.08	0.04	0.02
G_DP	0.37	0.43	0.27	0.25	0.18	0.18
Topological distortions						
T_ON	0.00	0.10	0.34	0.16	0.50	0.45
T_CN	0.19	0.48	0.48	0.60	0.54	0.81
T_TN	0.19	0.57	0.83	0.76	1.04	1.27
T_PI	0.00	0.00	0.13	0.13	0.00	0.00

- Total enumeration area size variance (G_TA): difference of total areas between geographical and EAUM representation (normalized by total area of geographical area).
- Difference Area of perimeter (G_DP): sum of protruding and clipped areas of EAUM in comparison to geographical area (normalized by total geographical area).
    Topographical distortions:
- Omission of neighbors (T_ON): number of neighborhood relationships that are missing after transformation to EAUM (normalized by total number of neighborhoods in map)
- Commission of neighbors (T_CN): number of neighborhood relations that are added after transformation to EAUM (normalized by total number of neighborhoods in map).
- Total number of neighborhood errors (T_TN): sum of T_ON and T_CN.
- Proportion of islands (T_PI): number of islands (which by definition cannot be preserved in EAUMs) in relation to total number of regions.

In all cases the measures are designed in such a way that the value range is limited to [0; 1] and the ideal case (i.e., no distortion) is reflected by the value 0.

3.2 Example

The aforementioned measures have been applied to three EAUMs that have been derived from given base maps with correct geographical appearance—showing the states of Australia, Germany, and the United States. Using the PyEAC software mentioned in chapter 2.2, two types of unit shapes have been created, namely regular squares (REG) and hexagons (HEX). Table 1 summarizes the results for all given measures.

Comparing the individual measures between the two basic EAUM shapes (REG and HEX) for one and the same area, no significant differences with regard to geometrical distortions can be observed (e.g. in the case of Australia: G_DD = 0.19 for HEX and G_DD = 0.20 for REG). On the other hand, there are clear differences between REG and HEX in topological measures, however, with no clear trends. The different maximum number of neighbors (REG: eight, HEX: six) together with the actual distribution of regions is probably responsible for these differences.

When looking at the differences between areas (with the same basic EAUM shape), Australia shows largest distortions, which is due to the small number of regions and the specific location of Tasmania. Compared to the German map, the US example shows (with one exception) slightly smaller geometrical distortions. Regarding topological changes, again no clear trends can be derived.

Figure 2 shows for an exemplary measure (D_GG) and a selected basic shape (REG) that there are large spatial variances within the measures listed in Table 1. The specified mean values, therefore, show a strong smoothing effect and do not create a clear overall impression of distortions for a specific data set.

All in all, due to the in many cases unclear trends, the regional variation of the measures and the rather large set of analytical measures, this analytical approach is not feasible for purposes such as comparing maps or even controlling an EAUM generation algorithm. As a consequence, this contribution will follow an empirical approach in the next chapters to better understand the importance and relevance of these measures.
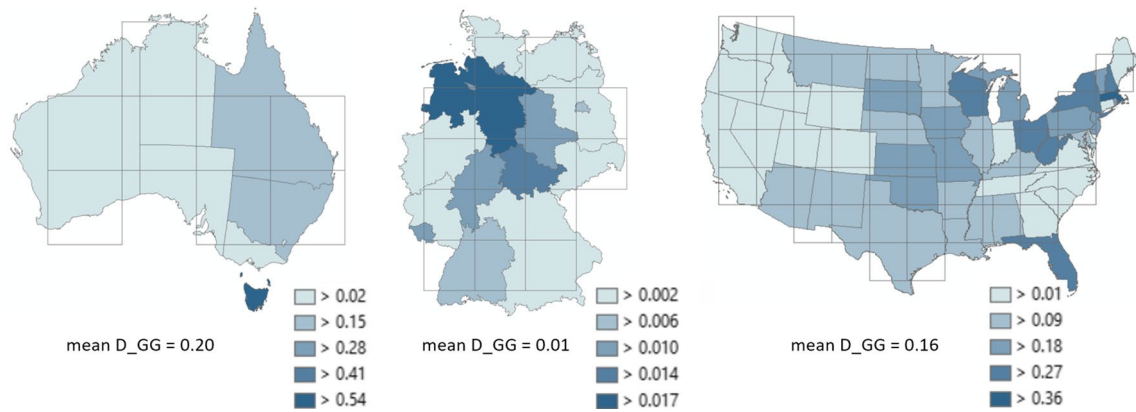
Fig. 2 Spatial distribution of measure D_GG for three EAUM data sets with basic REG shape (note: maps use different data classification schemes because absolute value ranges differ too much between examples)

4 Empirical Study: Search for Regions in EAUMs

4.1 Research Question

For many map usage tasks that are to be solved with the help of thematic maps with area reference (e.g. mosaic or choropleth maps), the search for regions is an essential operation. In this case either the name of the region or the location on another map for comparison purposes is given. Since geometrical distortions are inevitable with EAUMs, the question arises how effective the search can be. For further developments of EAUMs and other map types, it is also interesting to know which search strategies the users are using.

4.2 Hypotheses

Based on the knowledge of cognitive psychology and visual perception (e.g. top-down object recognition; Biederman 1981), it is assumed that topological relationships are key criteria for searching for specific regions in maps. This leads to the following hypotheses:

- [H1.1] Users apply different search strategies for regions depending on their topological situation (e.g. corner, border, center, and island).
- [H1.2] Algorithms for generating EAUMs can preserve the respective topological relationships only to a certain degree and thus influence the search effectiveness.

4.3 Study Design

Considering the various topological cases with distinct solutions that allow single or multiple choice questions, a quantitative, web-based study has been generated. Because no user group-specific characteristics are assumed, socio-demographic were not gathered. In order not to let the study processing time become too long, only map examples for Germany and the USA were shown.

The study used the two most common basic geometrical shapes for EAUMs, namely regular squares (REG) and hexagons (HEX). The region to be searched for was marked on a map with the "correct" geographical shape (GEO). Right to it an EAUM—either in REG or HEX layout—was presented and a single choice selection from four suggested target regions was requested; in all cases the question was: "Please locate the marked region in the right map. Choose the best solution for you" (Fig. 3).

"Please locate the marked region in the right map. Choose the best solution for you" (Fig. 3). With that approach, a possible lack of knowledge about the names of the regions was bypassed. A total of 11 cases were given—considering combinations of possible topologies (corner, border, center, and island) and different base maps (federal states of Germany and USA). When applying these cases to the two EAUM shapes (REG, HEX), a total of 22 tasks had to be solved.
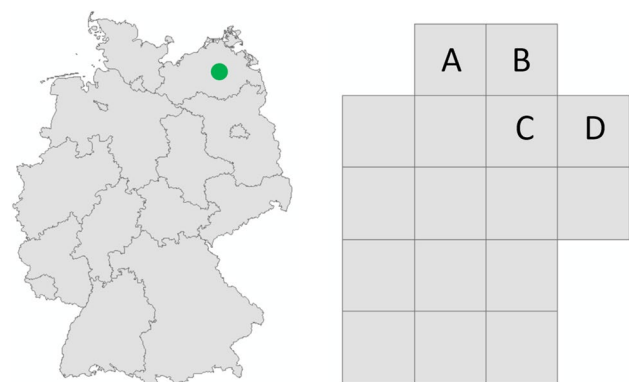

Fig. 3 Exemplary task for searching given region in map with geographical shape (GEO; left) in the EAUM map (here: REG; right)
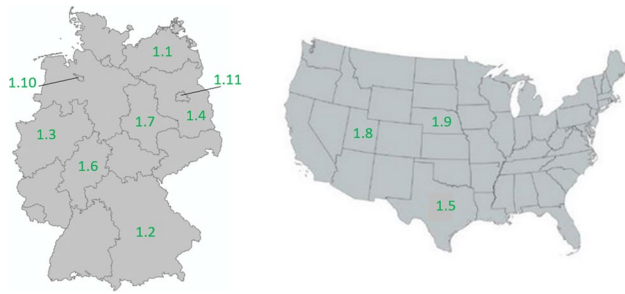
Fig. 4 Localization of cases 1 to 11 in study 1

The map examples were not presented in a systematic order in order to avoid learning effects. As efficiency is not in the focus of this survey, time measurements were not conducted.

4.4 Results

A total of 222 participants took part in the survey. While Fig. 4 shows the respective locations of the cases mentioned, Table 2 gives an overview of the results. The column "user preference" shows the percentage of users who selected the most common solution, regardless of whether this solution was identical with the algorithmic assignment of the respective state. This measure evaluates the pure

search strategy. The column "calculated state detected" shows the percentage at which users selected the region with a name match between geographical and EAUM representation. Only in 9 of 22 cases (marked with "Y" in the column "comparison") did user preference actually referred to the region with the calculated state.

No consistent strategies and results could be observed for the search for corner regions. In case 1.1 (region in an upper right corner; Fig. 5), 91% of the regions were assigned in the REG representation and 87% in the HEX representation. In contrast, the localization in case 1.2 (lower right corner; Fig. 6) shows significantly poorer values. The lower right corner was selected in REG only by 48% (though the majority) and in HEX by 41% of the participants. The search criterion for a corner (in this example the lowest and rightmost) was obviously not predominant, but the peripheral location in combination with the imaginary overlap of the EAUM region with the original region. Case 1.2 is also an example for a mismatch between solutions that were preferred by users and the one that was generated by the algorithm. It is therefore not surprising that only 4% (for REG and HEX) of the participants identified that algorithmic solution as Bavaria.

Also for border regions, detection results differed strongly between examples. North Rhine-Westphalia (case 1.3) shows very good detection rates of 98% (REG) and

Table 2 Results of study 1 concerning region search in EAUMs

Topological situation	Case	State	EAUM	User preference	Calculated state detected	Comparison
Corner	1.1	Mecklenburg-Western Pomerania	REG	91%	91%	Y
			HEX	87%	87%	Y
	1.2	Bavaria	REG	48%	48%	Y
			HEX	49%	4%	N
Border	2.3	North Rhine-Westphalia	REG	98%	98%	Y
			HEX	90%	90%	Y
	1.4	Brandenburg	REG	59%	59%	Y
			HEX	60%	38%	N
	1.5	Texas	REG	45%	–	–
			HEX	65%	1%	N
Center	1.6	Hesse	REG	63%	36%	N
			HEX	78%	78%	Y
	1.7	Saxony-Anhalt	REG	40%	32%	N
			HEX	55%	–	–
	1.8	Utah	REG	41%	41%	Y
			HEX	95%	–	–
	1.9	Nebraska	REG	70%	1%	N
			HEX	79%	6%	N
Island	1.10	Bremen	REG	52%	46%	N
			REG	50%	50%	Y
	1.11	Berlin	REG	62%	34%	N
			REG	79%	18%	N

Fig. 5 Case 1.1: Search for Mecklenburg-Western Pomerania (left, geographical shape)—user allocations (in percent) in REG (middle) and HEX (right) representations; green column indicates allocation of that region by algorithm
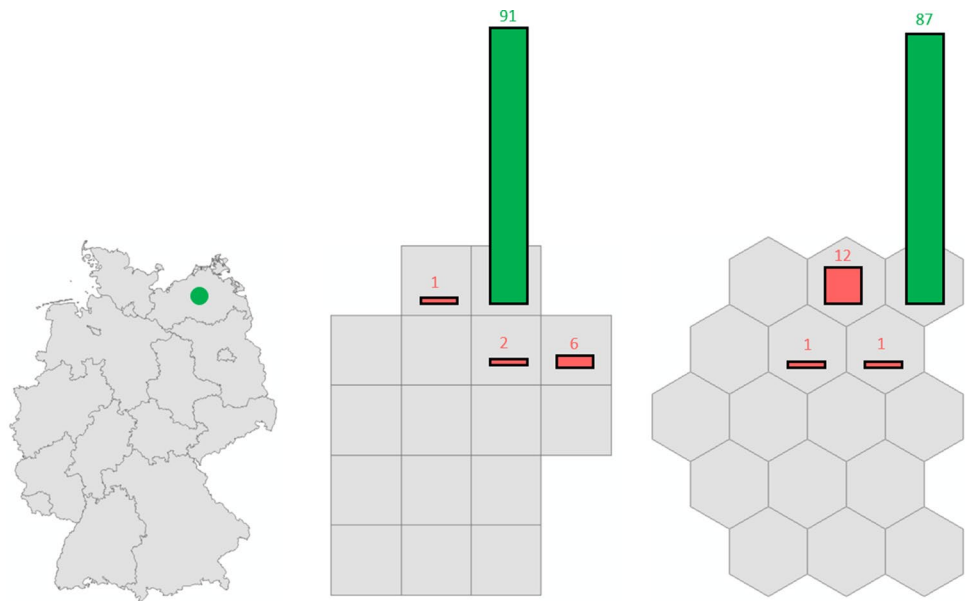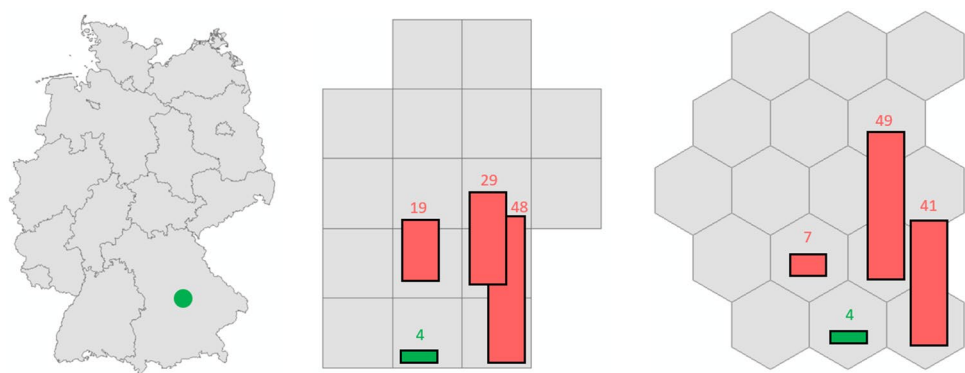
Fig. 6 Case 1.2: Search for Bavaria (left, geographical shape)—user allocations (in percent) in REG (middle) and HEX (right) representations; green column indicates allocation of that region by algorithm

Fig. 7 Case 1.5: Search for Texas (left, geographical shape)—user allocations (in percent) in HEX (right) representation; green column indicates allocation of that region by algorithm
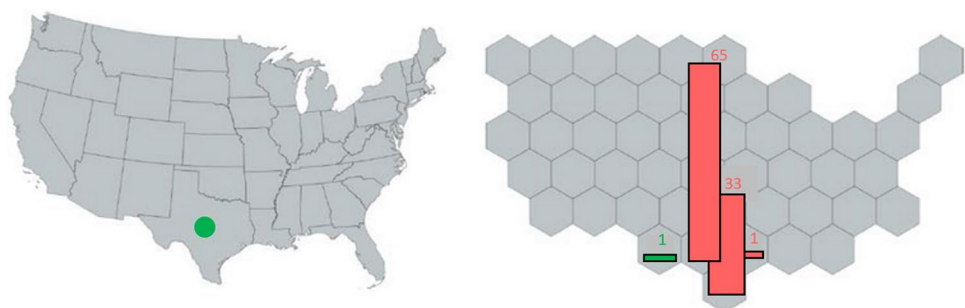
90% (HEX). The strategy here was obviously to either count along the edge from top to bottom, or to select the region at about the same height as in the original. In case 1.4 (Brandenburg), the result is significant worse (59%/38%), which is apparently due to the unclear position of the state Brandenburg relative to the island Berlin in the EAUMs. In the case of Texas (case 1.5; Fig. 7) it was not the bulge at the lower edge, but the first areal overlap with the original region that was selected most frequently

(65%)—using the reading order from the left to right. The HEX-algorithm has Texas placed to a rather distant location, which, unsurprisingly, was only recognized by 1% of users.

Also for center regions, a strong deviation between examples can be observed—the user preference values range between 40 and 70% (REG) and 55% and 95% (HEX). Obviously, most users solved these tasks by counting enumeration units. However, the counting as such was not always
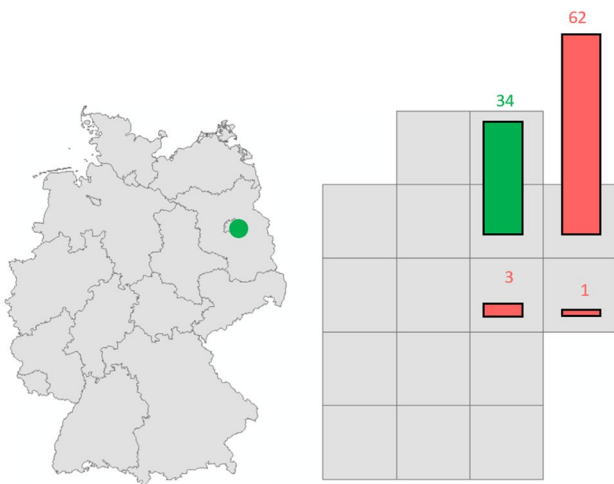
Fig. 8 Case 1.11: search for Berlin (left, geographical shape)—user allocations (in percent) in REG (right) representation; green column indicates allocation of that region by algorithm

correct—in particular, if more than three regions had to be considered like in the USA example. Furthermore, the misalignment of spatial units in the geographical and (to some extent) in the HEX-type hampers the counting. In general, HEX has higher recognition rates than REG in the given examples.

Island regions represent a problem with the representation in EAUMs as there is per se no perfect solution. Accordingly, the detection rates are also not satisfying. In 50–79% of the cases, the island and the surrounding polygon were arranged in the order in which they were read (usually from left to right; e.g. Berlin is assumed to be to the right of Brandenburg, Fig. 8). No HEX solution was offered for the search for island regions but one and the same REG variant had to be processed twice by the participants in the study with a certain time lag. While the detection rate remained the same in the case of Bremen, the rate for Berlin fluctuated very strongly (62% vs. 79%), which suggests low intuitiveness.

4.5 Interpretation

Hypothesis [H1.1]—assuming different search strategies for different topological situations—could be partially confirmed. In particular, corner and border positions were taken into consideration during the search. Essentially, the relative location was estimated or spatial units were counted—the left–right or top-down search sequences clearly dominated. In the case of larger geographic regions, those EAUM regions were often selected that show an imaginary overlap with the geographical region (e.g. in case 1.2).

Due to the scattered results and the unsatisfying mean of 66% for the user preference neither a clear nor a successful

strategy can be derived. Nevertheless, the identified basic topological search preferences should be taken into account when generating the EAUM. It can be concluded that the frequently used parameter "minimizing distances of centroids" (see Chapter 2.2) does not coincide with user's expectations.

Hypothesis [H1.2] could be confirmed: distortions in EACMs lead to significant deviations that users cannot easily compensate for. Compared to the aforementioned mean value of 66% for the user preference, the detection rate for the regions with a name match is only 43%. Furthermore, only in 9 out of 22 tasks did the user's major choice agree with the algorithmic solution.

5 Empirical Study: Pattern Identification in EAUMs

5.1 Research Question

Choropleth maps are primarily used to identify spatial patterns. In the following, regions with local extreme values (LE), global lateral gradients (GLG; i.e. North–South or East–West gradients) as well as hot spots (HS) are considered. The alternative use of EAUMs harbors the risk that the inherent spatial distortion lead to the fact that patterns are no longer recognized well or not at all. On the other hand, EAUMs should have the potential of a better recognition for local extremes of small enumeration areas, thus avoiding the area size-bias.

5.2 Hypotheses

It is assumed that due to geometrical distortions also the identification of spatial patterns in EAUMs is affected in a negative manner with the exception for the identification of local extreme value regions. This leads to the following hypotheses:

- [H2.1] The identification of LE takes place in EAUMs with greater effectiveness, because the area size bias is avoided. Since neighborhoods are irrelevant in this context, no difference between REG and HEX is to be expected.
- [H2.2] The identification of GLG and HS is best done in the original geographical layout (GEO) due to potential neighborhood distortions in REG and HEX.

5.3 Study Design

Considering the various pattern cases with distinct solutions that allow single or multiple choice questions, a

Fig. 9 Case 2.3: Representation of the same single LE in small enumeration area (here: Bremen in the North–West) in GEO, REG and HEX maps (from left; the darker the color the larger the value)
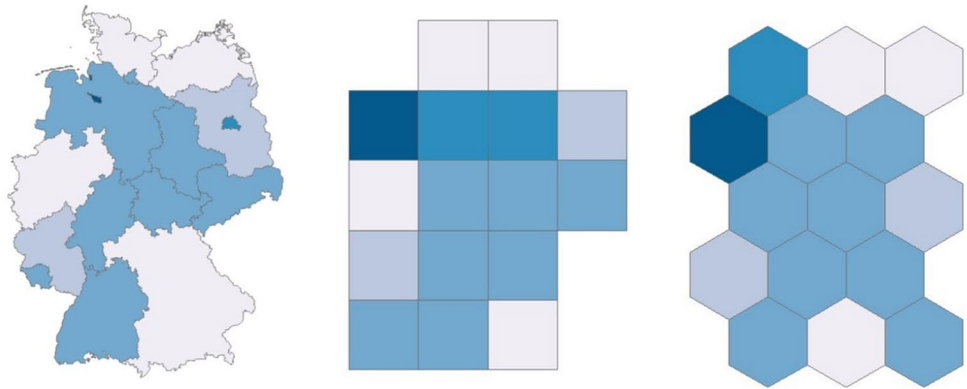
Fig. 10 Representation of the same East–West patterns in GEO, REG and HEX maps (from left; the darker the color the larger the value)
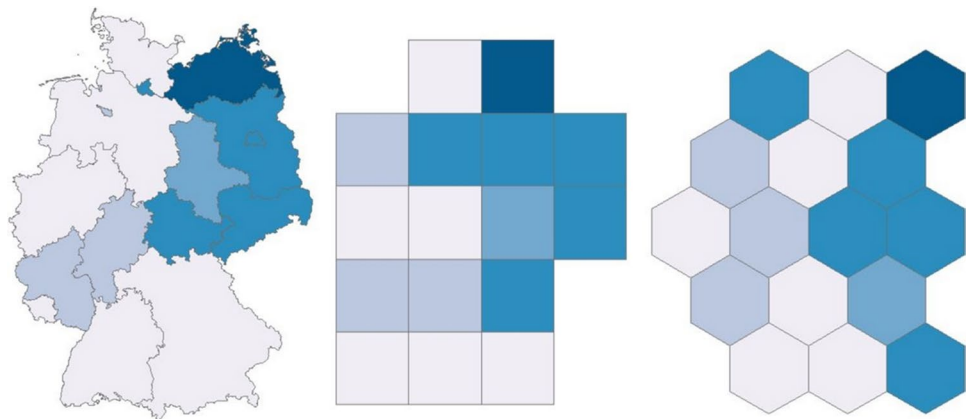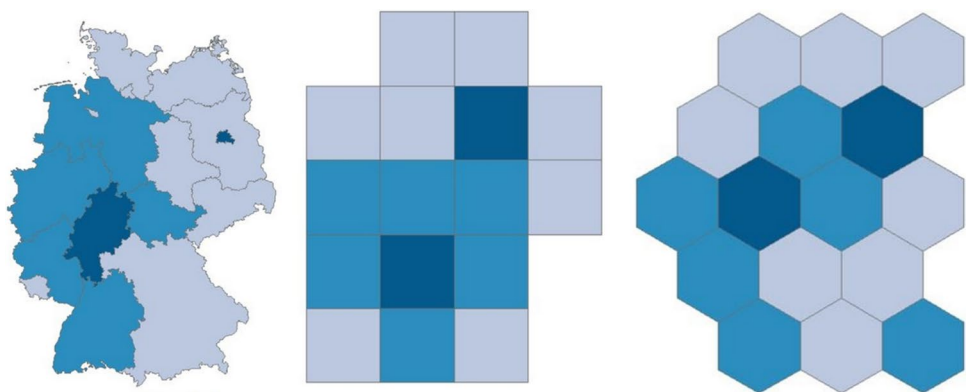
Fig. 11 Representation of the same hot spot in GEO, REG and HEX maps (from left; the darker the color the larger the value)

quantitative, web-based study has been generated. Because no user group-specific characteristics are assumed, socio-demographic were not gathered. In order not to let the study processing time become too long, only map examples for Germany and the USA were shown.

The task was to identify patterns in different map types (GEO/REG/HEX). For each task, one map with one or more patterns was presented; in total, eight maps were given (examples are given in Figs. 9, 10, 11). Participants were asked for a multiple choice-selection with the options of one

locale extreme value region (LE), two LE, North–South and East–West gradients, and HS (while the latter was described as, high value surround by other quite high values " for better understanding"). In all cases the question was: "Please name the spatial pattern that you view in this map. Multiple options are possible. Note that darker colors mean higher values". By purpose, the respective locations (i.e., the states) should not be named in order to avoid localization errors (see study 1).

Table 3 Results for study 2—correct pattern identification (with *p* values of 0.05, 0.01 and 0.001 for the definition of difference being "significant/*", "highly significant/**" or "highest significant/***")

Case	Base map	Pattern	Measure	GEO	REG	HEX	*p*(GEO-REG)	*p*(GEO-HEX)	*p*(REG-HEX)
2.1	Germany	One LE (large size)	M1	91%	87%	86%	0.16	0.10	0.48
			M2	68%	70%	66%	0.46	0.18	0.10
2.2	Germany	One LE (medium size)	M1	81%	87%	89%	0.10	0.03*	0.41
			M2	50%	58%	70%	0.04*	<0.001***	0.002**
2.3	USA	One LE (small size)	M1	58%	89%	90%	<0.001***	<0.001***	0.50
			M2	40%	73%	58%	<0.001***	<0.001***	<0.001***
2.4	Germany	Two LE (large and small)	M1	88%	82%	94%	0.08	0.03 *	<0.001***
			M2	62%	55%	69%	0.09	0.08	<0.001***
2.5	USA	Two LE (large and small)	M1	67%	82%	86%	<0.001***	<0.001***	0.19
			M2	51%	63%	63%	0.005**	0.003**	0.49
2.6	Germany	North–South	M1	93%	93%	89%	0.46	0.10	0.18
			M2	59%	63%	53%	0.21	0.12	0.006**
2.7	Germany	East–West	M1	78%	64%	64%	0.002**	<0.001***	0.50
			M2	64%	54%	76%	0.31	0.008**	0.001**
2.8	Germany	Hot Spot	M1	43%	43%	25%	0.50	<0.001***	<0.001***
			M2	22%	28%	17%	0.06	0.09	<0.001***

5.4 Results

Again, a total of 222 participants took part in the survey. Table 3 summarizes the results. Presenting a total of eight tasks for each map type (GEO, REG, and HEX) by using two base maps (either Germany or USA), the measures M1 and M2 were derived from the answers:

- M1: Ratio of users with correct pattern identification divided by all users (e.g. 90% of all users have recognized LE)—i.e., only omission errors are considered.
- M2: Ratio of correctly identified patterns divided by all recognized patterns (considering the possibility of multiple choice answers)—i.e., also commission errors are taken into account.

To evaluate the difference between the map types, the *p* value was derived from a pair-wise Chi-Square test.

Concerning local extreme values (LE), one can observe a clear decrease in the identification rate from large LE regions (case 2.1) over medium (case 2.2) to small ones (case 2.3; Fig. 9). The respective numbers for the GEO maps (M1 of 91%, 81% and 58%) correspond very well to the results of an earlier study (Schiewe 2019). If two LE were presented (cases 2.4 and 2.5), this effect could also be confirmed. In general, REG and HEX performed in a similar manner.

Considering global lateral gradients (GLG), the identification rates for the North–South example (case 2.6) are rather high and very similar between map types. On the other hand, for the East–West scenario (case 2.7) the GEO map was superior, although the absolute rate (78%) is not satisfying (Fig. 10).

The worst identification rates are observed for hot spots (HS; Fig. 11), although the example given shows a significant GI*score of 1.97. This result is due to the fact that the number of neighbors, which also have high values, has been reduced within the REG and HEX maps, so that the hot spot impression is also weakened.

5.5 Interpretation

With regard to hypothesis [H2.1], the area-size bias for the identification of LE in maps with geographical shape (GEO) could be clearly shown—the detection rate decreased significantly from large over medium to small enumeration areas. On the other hand, the identification of LE takes place in EAUMs with greater effectiveness. If the task is (only) to identify LE and smaller enumeration areas are affected, EAUMs should definitely be used. Nevertheless, an optimization concerning localization and preservation of neighborhoods is still meaningful (see study 1).

Hypothesis [H2.2] could also be confirmed: with regard to global lateral gradients (East–West or North–South), clear boundary lines are lost due distorted neighborhoods. Instead, other "artificial" patterns such as LE could dominate. With regard to hot spots, it was observed that the number of neighbors with large values is reduced so that hot spots are less noticeable. In summary, the actual geographical shape (GEO) should be preferred for these patterns, especially when small regions are not of interest.

6 Summary and conclusions

Maps that correctly reflect the geographic size and shape of regions have the disadvantage that small regions can easily be overlooked or not seen at all. This is not feasible for many map tasks in which small regions are of importance (e.g. the detection of local extreme values). An alternative approach is to use maps where every enumeration unit (geographic entity) is given the same size and possibly also the same shape (such as squares, rectangles or hexagons). Due to an inconsistent definition so far, the term "Equal Area Unit Map (EAUM)" is proposed for this map type.

EAUMs are being used more and more, in particular in news media. However, the geometrical distortion of EAUMs leads to a more difficult search and localization of regions as well as a falsification of topological relationships and spatial patterns. A set of analytical measures has been proposed to describe these distortions; however, the expressiveness of these measures is rather limited due to unclear trends, their strong spatial variation and the rather large set of parameters.

As a consequence, two user studies were conducted in order to better understand and model the influence of distortions. Concerning the search of regions in EAUMs, no clear trend in the search strategy could be observed. On the one hand, clear topological properties such as corner and border positions were taken into account; on the other hand, regions were often selected that simply have an imaginary overlap with the geographical region. Many users estimate the relative location by counting regions with the search sequences dominating from the left or from the top. In addition to this search problem, the actual placement regions by the algorithm, which does not take topological relationships into account, must be considered—only in 9 out of 22 tasks the user's main selection agreed with the algorithmic solution. This indirectly leads to the recommendation that labels should be used when quick identification is necessary and there is enough space to place them.

With respect to pattern identification it could be shown that EAUMs significantly increase the detection rate of local extreme values, thus avoiding the area-size bias in contrast to maps with "correct" geographical shapes and sizes. On the other hand, global lateral gradients (East–West or North–South) or geostatistical hot spots often get blurred or even lost. If certain spatial patterns are known and of importance, a task-oriented selection of map types is therefore recommended.

Our future research work will consider two aspects. Firstly, local topological properties will be included into the process of generating EAUMs. Although there is no clear search strategy, as many topological properties as possible (such as corner or border positions) should be taken into account in order to improve the search results.

In this context, the handling of island enumeration units is a very challenging topic; one solution to this could be the generation of semi-units[3] (e. g., putting Berlin and Brandenburg into one larger unit). Secondly, also the distortion of local and overall shape affects search and pattern identification in a negative manner. Hence, the approach of using approximate shapes instead of equal shapes will be taken but still maintaining equal enumeration area sizes. To do this, EAUMs will be composed of many smaller units (e.g. squares) that can be grouped in different ways to best reflect the actual geographical shape. In this context, also the consideration of local topology must be considered.

Data Availability Software PyEAC and example data available under https://gitlab.com/g2lab/pyeac.

Code Availability Software PyEAC available under https://gitlab.com/g2lab/pyeac.

Compliance with Ethical Standards

Conflict of interest None.

References

Alam J, Kobourov SG, Veeramoni S (2016) Quantitative measures for cartogram generation techniques. Proc Eurogr Conf Vis (EuroVis). https://doi.org/10.1111/cgf.12647

Barr RS, Glover F, Klingman D (1977) The alternating basis algorithm for assignment problems. Math Prog 13:1

Biederman I (1981) On the Semantics of a glance at a scene. In: Kubovy M, Pomerantz (eds) Perceptual organization. Erlbaum, Hillsdale

Brath R (2015) Equal area cartograms and multivariate labels. https://richardbrath.wordpress.com/2015/10/15/equal-area-cartograms-and-multivariate-labels/. Last Accesses 1 Dec 2020

---

[3] https://github.com/ebp-group/Switzerland_Tilemap

Dent BD (1999) Cartography: thematic map design, 5th edn. McGraw-Hill, Boston

Eppstein D, vanKreveld M, Speckmann B, Staals F (2015) Improved grid map layout by point set matching. Int J Comput Geom Appl 25(2):101–122

Hake G, Grünreich D, Meng L (2002) Kartographie, de Gruyter

ICA (2003) A Strategic Plan for the International Cartographic Association 2003–2011. https://icaci.org/files/documents/reference_docs/ICA_Strategic_Plan_2003-2011.pdf. Last Accessed 1 Dec 2020

Kuhn HW (1955) The hungarian method for the assignment problem. Naval Res Logist Q 2:83–97

McNeill G, Hale SA (2017) Generating tile maps. Proc Eurogr Conf Vis (EuroVis). https://doi.org/10.1111/cgf.13200

Munkres J (1957) Algorithms for the assignment and transportation problems. J Soc Ind Appl Math 5(1):32–38

Nusrat S, Kobourov S (2016) The state of art in cartograms. Proc Eurogr Conf Vis (EuroVis). https://doi.org/10.1111/cgf.12932

Peterson MP (2012) The tile-based mapping transition in cartography. In: Zentai L, Reyes Nunez J (eds) Maps for the future. Lecture notes in geoinformation and cartography, vol 5. Springer, Berlin

Raisz E (1934) The rectangular statistical cartogram. Geogr Rev 24(2):292–296

Schiewe J (2019) Empirical studies on the visual perception of spatial patterns in choropleth maps. KN J Cartogr Geogr Inf 69(3):217–228. https://doi.org/10.1007/s42489-019-00026-y

Slocum TA, McMaster RB, Kessler FC, Howard HH (2009) Thematic cartography and geovisualization, 3rd edn. Prentice Hall, Upper Saddle River, p 576

Sun H, Li Z (2010) Effectiveness of cartogram for the representation of spatial data. Cartogr J 47(1):12–21

Ordnance Survey (2020) https://github.com/OrdnanceSurvey/equal-area-cartogram. Last Accessed 1 Dec 2020

Wongsuphasawat K (2016) Whose grid map is better? Quality metrics for grid map layouts. https://kristw.medium.com/whose-grid-map-is-better-quality-metrics-for-grid-map-layouts-e3d6075d9e80. Last Accessed 1 Dec 2020