Alternatives for automatic rendering of Brazil and other geographic regions' territory thematic cartograms

Alternativas para a renderização automática de cartogramas temáticos do território do Brasil e de outras regiões geográficas

Luiz Nascimento, Regina Reznik, Tarsus Pinheiro, Cláudio Esperança, Ricardo Farias

cartograms, thematic maps, carto-centric, carto-step, data visualization, IBGE

This article presents alternative concepts for rendering thematic cartograms, using primitive shapes arranged under different concepts of spatial distribution, with the aim of meeting, in a more intuitive way, the criteria explained at the initially defended proposal by Barreto, Kosminsky, and Esperança (2018). Graphic rendering algorithms, developed for the automatic generation of thematic cartograms according to the demonstrated concepts, and following the methodology of Speckmann and Verbeek (2010)'s work, are also presented.

cartogramas, mapas temáticos, carto-centric, carto-step, visualização de dados, IBGE

Apresentam-se neste artigo conceitos alternativos para a renderização de cartogramas temáticos, utilizando formas primitivas dispostas sob diferentes conceitos de distribuição espacial, visando atender de forma mais intuitiva aos critérios explicitados na proposta inicialmente defendida por Barreto, Kosminsky e Esperança (2018). São também apresentados algoritmos de renderização gráfica, desenvolvidos para a geração automática de cartogramas temáticos conforme os conceitos aqui apresentados, e seguindo a metodologia do trabalho de Speckmann e Verbeek (2010).

1 Context and motivation

Data visualization is a relevant research problem; when the visualized data is georeferenced, that is, associated to a known location in geographic space, the result is a map.

Maps can simplify and regularize information, as seen in Tversky (2000), and remove unimportant information that could harm the cognitive load, as seen in Barkowsky, Latecki, and Richter (2000). For example, a thematic map can help in the visualization of data by federation units from the Population Census, which is an operation carried out every decade by the IBGE (Brazilian Institute of Geography and Statistics), thus avoiding great differences in area or population.

A data map allows many different levels of analysis, ranging from an overview, that is, a view of general global patterns, to tiny details, as seen in Tufte (2001).

Many ways of solving these visualization problems have already been discussed. We can cite the example of Jacques Bertin (1983), who studied the problem of this type of representation and listed a historical sequence of maps that allow us to understand their evolution in parallel with the evolution of technology.

Currently, map services on the Internet for mobile devices are very common applications, having become popular in our daily lives; therefore, the development of adequate solutions for the visualization of georeferenced data is a relevant research topic.

In the initial proposal by Barreto et al. (2018), when drawing a thematic map for the average grades of the National Secondary Education Exam (ENEM) by exam and by state, five requirements are presented for the visual representation of geographical units: an initial requirement was established to allow a comparison between the average grades of each of the five exams, to identify patterns and correlations, without disregarding the general situation of the country. Since such a comparison requires maps that roughly present the same shape, the use of cartograms should not be considered for this purpose.

A second requirement was to preserve the shape of the country, which suggested a choropleth map of Brazil as an appropriate subject of study. Choropleth and chorochromatic maps work by associating the represented regions with a list of colors whose distinction lies in their hue or saturation characteristics, which makes it possible to identify data in space and, therefore, associate the represented regions with the information that relates to them. Andrienko studied this relationship in 'Choropleth and chorochromatic maps work by associating the represented regions with a list of colors, capable of identifying data in space, and thus relating the represented regions to the information that relates to them' (Andrienko & Savinov, 2001).

Despite being very popular, this type of cartogram has limitations, such as the difficulty in visualizing small areas that may be very important. Sergipe (SE) and the Federal District (DF) are so small in relation to the rest of the units that they are not clearly differentiated when colors are used for comparison. States such as Amazonas and Pará, despite their low population density, would immediately be overemphasized, regardless of the variable being compared; the problem can be aggravated if the values are similar. To better compare values between Brazilian states of very different sizes, a requirement was established: that the states be rendered in a size as uniform as possible; thus, avoiding the bias associated with differences in the size of areas is a central issue in this study, and its third requirement. To achieve this goal, it was decided that each of the 26 states and the Federal District would be represented by basic forms with similar areas.

There are possible solutions for these visualizations, such as the statistical grid (reference), whose application can harmonize with contemporary technological resources, as seen in Pinheiro and Esperança (2018); the statistical grid, however, also has limitations, in the sense of decomposing the space into arbitrary units which are not precisely related to the geopolitical units that may need representation.

The fourth requirement was the use of monochromatic color palettes for representing the average grade magnitude for each subject by state. Although many studies – Yau (2013), Munzner (2014) – have emphasized the limits of perception when comparing values with colors, it was observed that using a monochromatic color palette for each school subject could personalize the subject maps. A color scale is not merely an aesthetic choice, but a useful and intuitive tool for visualizing magnitude. Furthermore, colors can be applied to represent ordered attributes, in this case, the different school subjects.

Finally, the fifth requirement was to preserve neighboring relations between states, facilitating the visual search for a specific state.

After many attempts to represent Brazil's Federation Units with different regular polygons, it was observed that, regardless of applied shapes, preserving the neighborhood between the states was still a problem. Thus, an algorithm was created to exhaustively test the possibilities of changing the unit's positions yet maintaining the topological neighborhood.

The final result met four of five established requirements, namely: normalizing state areas, allowing comparisons between figures by state, preserving the country's shape, and using monochromatic color palettes for magnitude representation. After all the visual experiments were carried out, it was observed that preserving neighboring relations between states – the fifth requirement – could only be fulfilled at the risk of losing the country's shape.

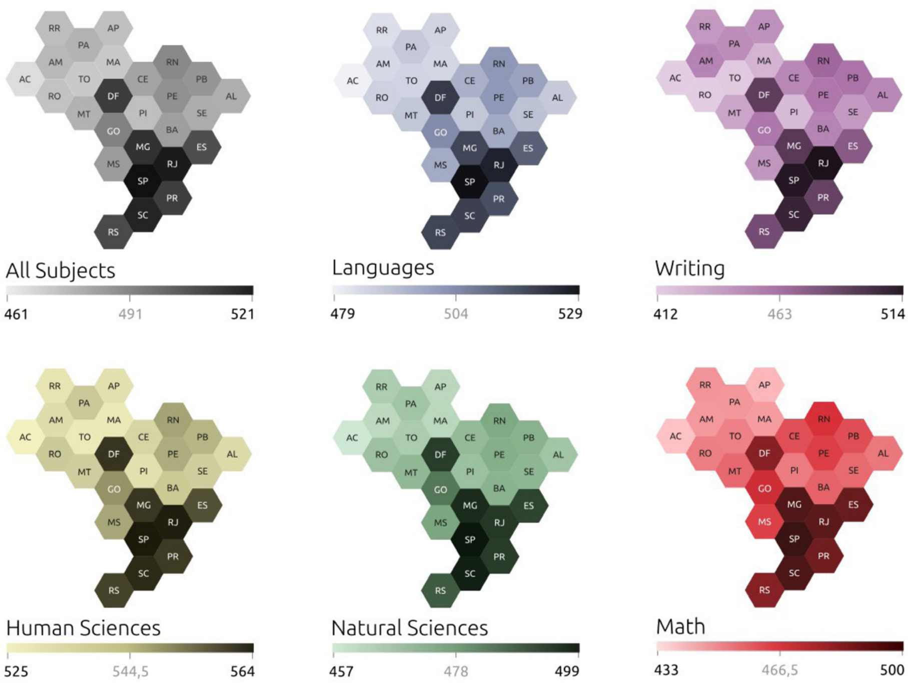

Figure 1 Regular shape (hexagon) cartograms representing Brazilian Federation Units (extracted from Barreto et al., 2018).

For the most part, this work confirmed the importance of developing more studies in information design focusing on the creation of maps that consider regional needs.

It is worth mentioning the work by Speckmann and Verbeek (2010), which develops the proposal of "necklace maps", consisting of "necklaces" that surround an underlying map, and are composed of symbols associated to the political-administrative divisions of the map; this work also presents a methodology for the automated creation of such maps.

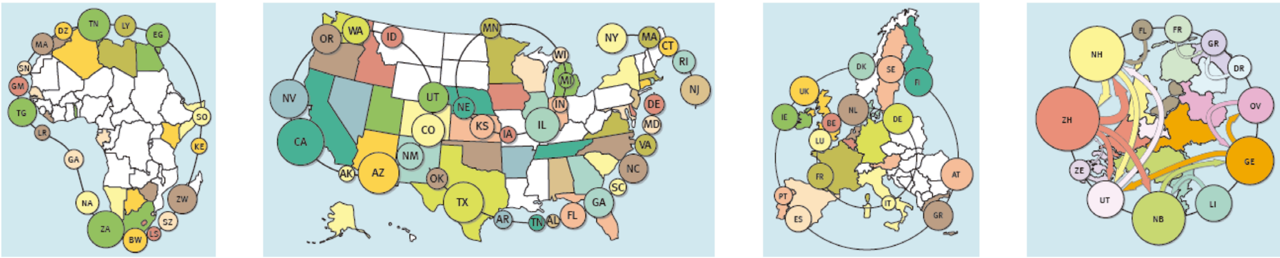

Figure 2 Examples of "necklace maps". Extracted from Speckmann and Verbeek (2010).

Therefore, this article departs from the need to foster studies to solve and meet the set-out requirements, to present two proposals for rendering cartograms using basic geometric shapes, explaining their methodologies and offering several implementation suggestions, with practical applications that use data collected by the IBGE in the 2023 Population Census.

2 Concepts and methods

Due to specific characteristics of its representation in terms of neighborhood and disparity in size of the federation units, the starting point in the search for new forms of representation was the map of the Brazilian territory, with the aim of formulating solutions to this unique representational problem. Subsequently, we sought to apply the proposed solutions to other territories and their respective political-administrative divisions, in order to validate their application under different conditions, taking into account that, due to the limitations imposed by each proposal, there is no way to present a single solution that simultaneously offers all territories a cartogram of perfect representation.

2.1 Carto-centric

The first proposal is based on a concept of geographic concentricity: a geographic unit, suitable for occupying the central position of the cartogram, is chosen and, based on neighborhood and distance, the remaining units are distributed over the perimeter of one or more circles, concentric to this central reference unit.

Under this proposal, keeping in mind a proper adaptation to the Brazilian territory and the concept of geographic centrality expressed in the design and construction of the new capital city, Brasília, founded in 1960, the Federal District was almost naturally identified as the most proper federation unit for occupying cartogram's central position.

Then, the following concentric layers were arbitrated:

- Geographical units closer to the central unit, but without coastline or borders with other countries. This layer is made up of three units: Goiás, Minas Gerais and Tocantins;
- Geographical units that border the previous units. This second layer is made up of fifteen units: Alagoas, Bahia, Ceará, Espírito Santo, Maranhão, Mato Grosso, Mato Grosso do Sul, Pará, Paraíba, Pernambuco, Piauí, Rio de Janeiro, Rio Grande do Norte, São Paulo, Sergipe;
- Subsequent layers, made up of geographical units bordering the previous units. Thus, the third layer will be composed of Amapá, Amazonas, Paraná and Rondônia; the fourth layer will be composed of Acre, Roraima and Santa Catarina, and the fifth layer will be made up of a single federation unit, Rio Grande do Sul.

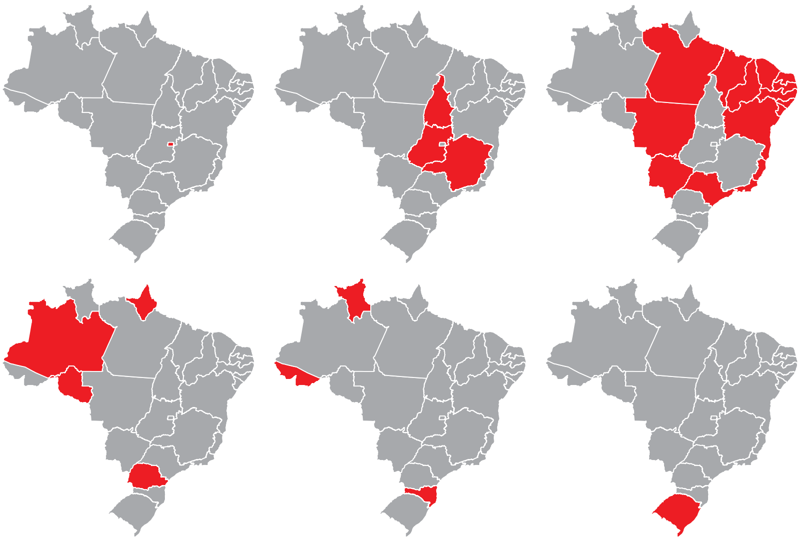

Figure 3 Concentricity levels with their respective federation units.

By collecting and establishing these aspects, the elements' spatial distribution is carried out within each layer of the cartogram, in order to bring them closer to their relative positions to neighboring elements.

Thus, for computational purposes, some general variables need to be established: the basic shape and size of each element, the cartogram's central element position, and each subsequently external level's distance, related to the central element.

To computationally represent the aspects related to each element, an array of tuples was used, each one composed of the following numerical values: a unique identifier (which can be an integer, a text, or both), the element's level, the element's angular position at that level, and optionally the data to be represented by the cartogram. The purpose of using an array of tuples is to make the representation more economical in terms of memory space when compared to traditional vector map formats. This feature tends to be interesting when we think about field devices for data collecting tasks, with limited data storage and processing capability.

```
const ufs =[[11,"RO",3,180],
            [12,"AC",4,180],
            [13,"AM",3,150],
            [14,"RR",4,150],
            [15,"PA",2,150],
            [16,"AP",3,120],
            [17,"TO",1,120],
            [21,"MA",2,120],
            [22,"PI",2,100],
            [23,"CE",2,80],
            [24,"RN",2,60],
            [25,"PB",2,40],
            [26,"PE",2,20],
            [27,"AL",2,0],
            [28,"SE",2,340],
            [29,"BA",2,320],
            [31,"MG",1,240],
            [32,"ES",2,290],
            [33,"RJ",2,265],
            [34,"SP",2,240],
            [41,"PR",3,240],
            [42,"SC",4,240],
            [43,"RS",5,240],
            [50,"MS",2,210],
            [51,"MT",2,180],
            [52,"GO",1,180],
            [53,"DF",0,0]
           ];
```

Figure 4 Tuple array for concentricity cartogram related to Brazilian Federation Units.

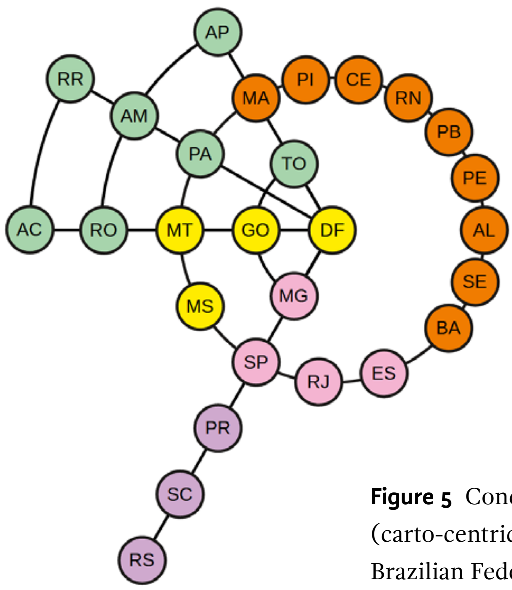

Figure 5 Concentricity cartogram (carto-centric) representing Brazilian Federation Units.

From this point onwards, we will name this first proposal as "carto-centric", in view of its main visual characteristic, concentricity.

To verify its validity on cartograms relating to other 'geographical areas', the same concept of concentricity was experimented with Rio de Janeiro State map. In addition to the greater number of geographical units (92 municipalities) compared to Brazilian territory (27 federation units), there is a very frequent fact, in which the central administrative unit (the capital) does not correspond to a territorial central position. In this case, two different cartograms can be created, in order to evaluate which one is more visually appropriate: one of them concentric to the capital, and the other one concentric to some arbitrary unit closer to the territory's centroid, and using some additional visual resource to distinguish the capital; at the presented examples, a different outline color was applied to that specific capital unit.

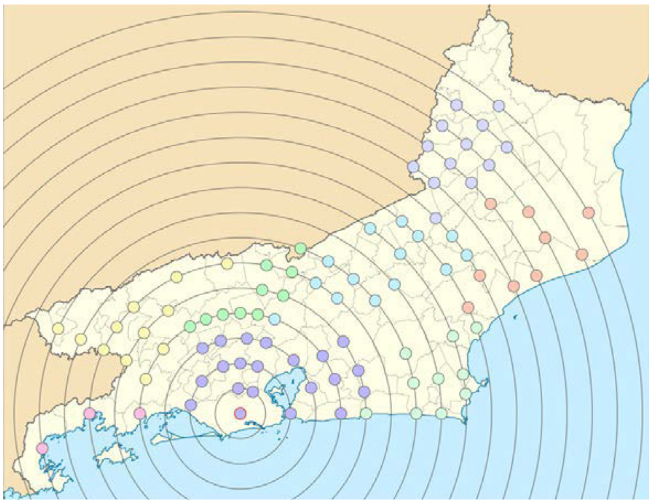

Figure 6 Concentricity cartogram (carto-centric) representing Rio de Janeiro state municipalities, capital-centered.

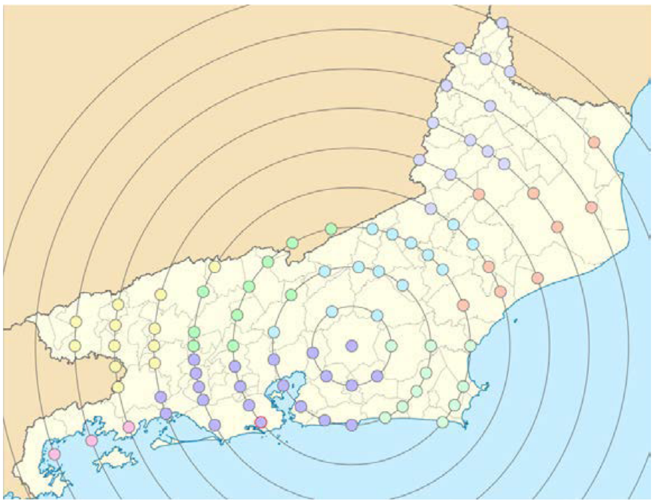

Figure 7 Concentricity cartogram (carto-centric) representing Rio de Janeiro state municipalities, centered at Cachoeiras de Macacu.

It is observed that the capital-centered concentricity produced a cartogram represented by more levels, since that the capital tends to aggregate units at its surroundings that are smaller in size and larger in amount.

2.2 Carto-step

The previous proposal brings the benefit of representing all units through symbols of uniform shape and size, but it is still possible to think about solutions that limit dimensions' differences only to some restricted scale of sizes, in order to make all units visible simultaneously and still protect some proportional character and represented territory's general shape.

Thus, the second proposal suggests the representation of each geographic entity through two similar and concentric basic shapes: one which is solid and with a single and constant area for every entity; and another presented in contour, and which area is proportionally related to the real geographical area, constrained to the arbitrarily chosen restricted sizes' scale.

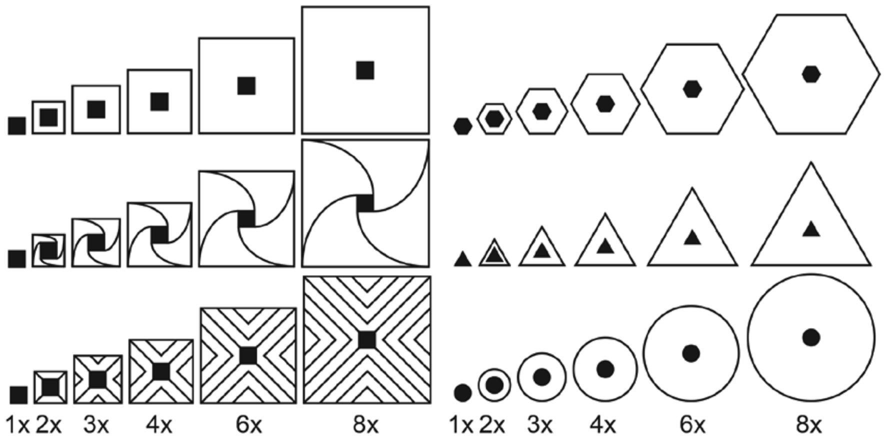

Figure 8 Suggestions for basic concentric shapes' scales.

Under this proposal, we initially seek to identify, for each geographic entity, the following key aspects:

- its centroid, in order to define its global positioning on the cartogram.
- its area, in order to define, together with all other units' areas, the global scale and the number of discrete sizes to be applied at the distinct units' representation on the cartogram.

With these aspects collected and established, the elements' spatial distribution on the cartogram is then carried out, which can follow the original position of the centroids, or approximate them to a regular grid. There must be no superposition of the solid internal forms, being permitted element's outlined external shapes superposition and interlacing.

Likewise, for computational purposes, some general variables need to be established; these are, in this proposal, the basic shape and the internal dimension of each element, in addition to the scale that compounds the discrete values' established set for external outlined shapes, relative to internal solid shapes.

To computationally represent the aspects related to each element, an array of tuples was used, each one made up of the following numerical values: a unique identifier (which can be an integer, a text, or both); the central position shared by the solid internal form and the outlined external form; the discrete value that multiplies the internal dimension for obtaining combined element's external dimension; and optionally the data to be represented by the cartogram.

```
const ufs =[
            [11,"RO",15.5,10,4],
            [12,"AC",11.5,9,4],
            [13,"AM",15,5.5,8],
            [14,"RR",16.5,1.5,4],
            [15,"PA",22,5.5,6],
            [16,"AP",22.5,2,3],
            [17,"TO",25,9,4],
            [21,"MA",27,6.5,4],
            [22,"PI",29.5,7,3],
            [23,"CE",31,7.5,2],
            [24,"RN",32.5,7,1],
            [25,"PB",33.5,8,1],
            [26,"PE",32.5,9,1],
            [27,"AL",33.5,10,1],
            [28,"SE",32.5,11,1],
            [29,"BA",29,11.5,6],
            [31,"MG",27,14.5,4],
            [32,"ES",30,15.5,2],
            [33,"RJ",28,17.5,2],
            [34,"SP",25,17,4],
            [41,"PR",22,18.5,2],
            [42,"SC",23.5,20,2],
            [43,"RS",21.5,22,3],
            [50,"MS",20.5,16,4],
            [51,"MT",20.5,11.5,6],
            [52,"GO",23.5,13,4],
            [53,"DF",25,13,1]
           ];
```

Figure 9 Tuple array for the step cartogram related to Brazilian Federation Units.

From this point onwards, we will name this second proposal as "carto-step", in view of its fundamental characteristic of combining two similar and concentric shapes, which can be understood as a "step" or base of a regular pyramid; thus favoring an additional variable representation through the use of three-dimensionality, suggested by the constructed solid.

In this representation, the Great Regions' general shapes are maintained, according to the officially proposed division.

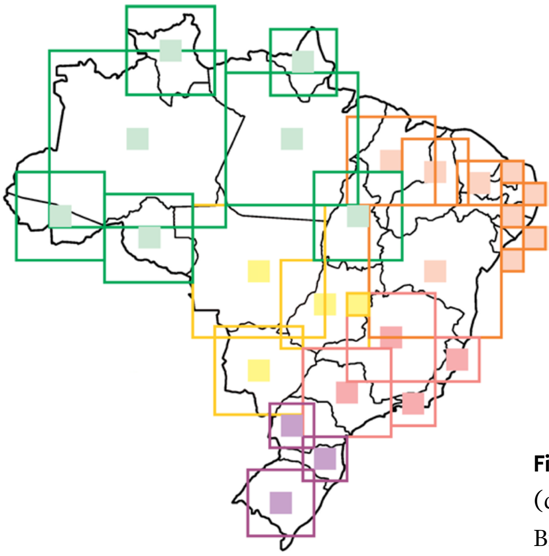

Figure 10 Step cartogram (carto-step) applied to Brazilian Federation Units.

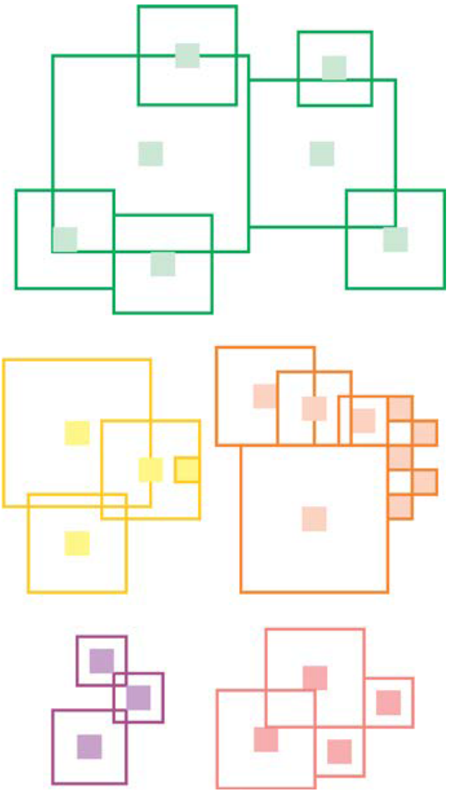

Figure 11 Step cartogram (carto-step) applied separately to Brazilian Great Regions.

To verify its validity on cartograms relating to other 'geographical areas', the same concept of composing shapes was experimented with Rio de Janeiro State map. Here the same discrete sizes' scale used for the Brazilian territory could be applied, containing values from 1x to 8x, and still preserving the visualization of all units; other scales can be applied, depending on each represented territory's characteristics.

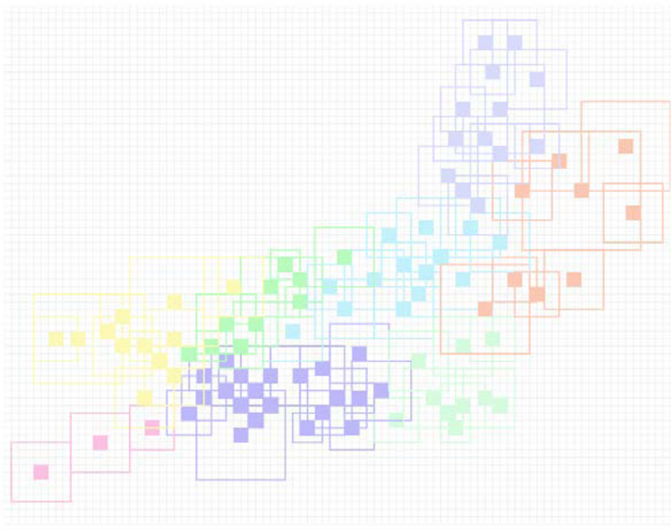

Figure 12 Step cartogram (carto-step) applied to Rio de Janeiro State municipalities.

2.3 Step-centric

It is worth noting that the two proposals presented here are not exclusionary; thus, its characteristic basic concepts – concentricity and steps – could be combined into one cartogram that displays both aspects simultaneously, producing what could be termed a "step-centric" cartogram.

```
const ufs =[[11,"RO",3,180,4],
            [12,"AC",4,180,4],
            [13,"AM",3,150,8],
            [14,"RR",4,150,4],
            [15,"PA",2,150,6],
            [16,"AP",3,120,3],
            [17,"TO",1,120,4],
            [21,"MA",2,120,4],
            [22,"PI",2,100,3],
            [23,"CE",2,80,2],
            [24,"RN",2,60,1],
            [25,"PB",2,40,1],
            [26,"PE",2,20,1],
            [27,"AL",2,0,1],
            [28,"SE",2,340,1],
            [29,"BA",2,320,6],
            [31,"MG",1,240,4],
            [32,"ES",2,290,2],
            [33,"RJ",2,265,2],
            [34,"SP",2,240,4],
            [41,"PR",3,240,2],
            [42,"SC",4,240,2],
            [43,"RS",5,240,3],
            [50,"MS",2,210,4],
            [51,"MT",2,180,6],
            [52,"GO",1,180,4],
            [53,"DF",0,0,1]
           ];
```

Figure 13 Tuple array for the "step-centric" cartogram relative to Brazilian Federation Units.

It should be noted that a simple application of an array of tuples containing the same numerical values used in the two separate proposals must be subjected to review and adjustment, in order to provide to this representation a better spatial adaptation.

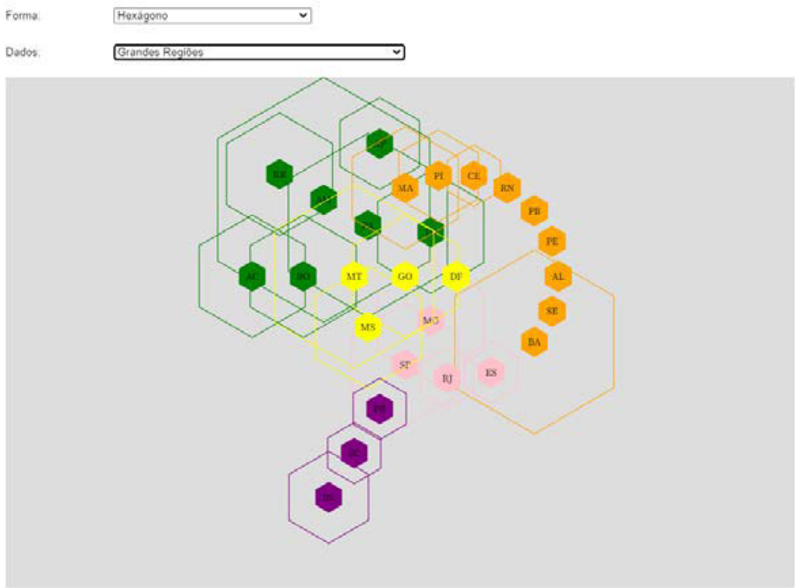

Figure 14 "Step-centric" cartogram applied to Brazilian Federation Units.

3 Results and discussion

The proposals presented above were applied to the first released data by IBGE after the data collection carried out during the 2023 Population Census. The cartographic representations suggested below were initially developed in Observable (2024), an online platform for data visualization, available for consultation through the URL https://observablehq.com/@lp; at this interface, it is possible to directly switch between data sets via a selection box.

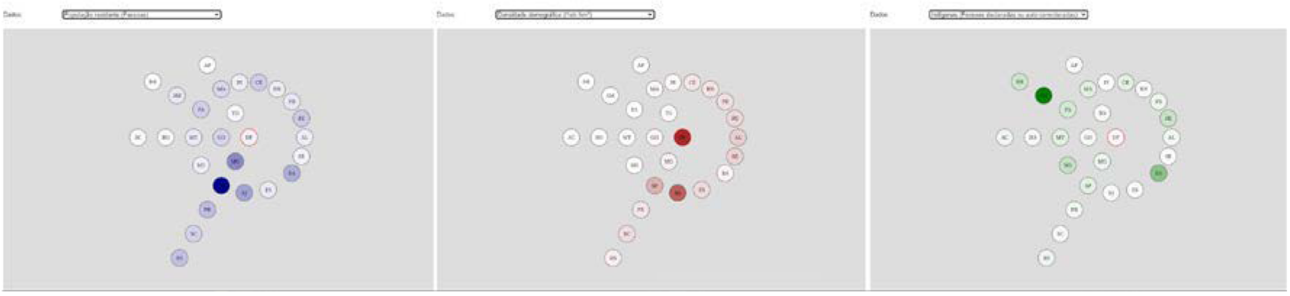

Figure 15 Carto-centric presenting data from the 2023 Census for Brazilian Federation Units.

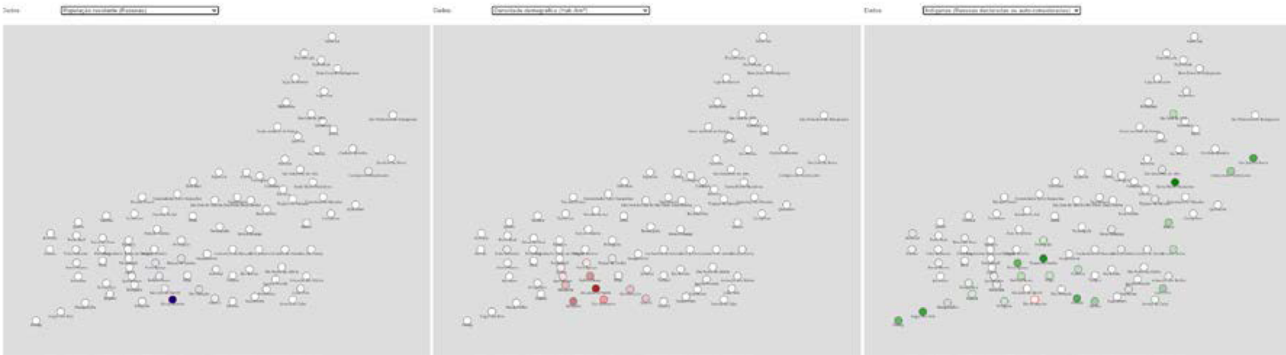

Figure 16 Carto-centric presenting data from the 2023 Census for Rio de Janeiro state municipalities, centered at Cachoeiras de Macacu.

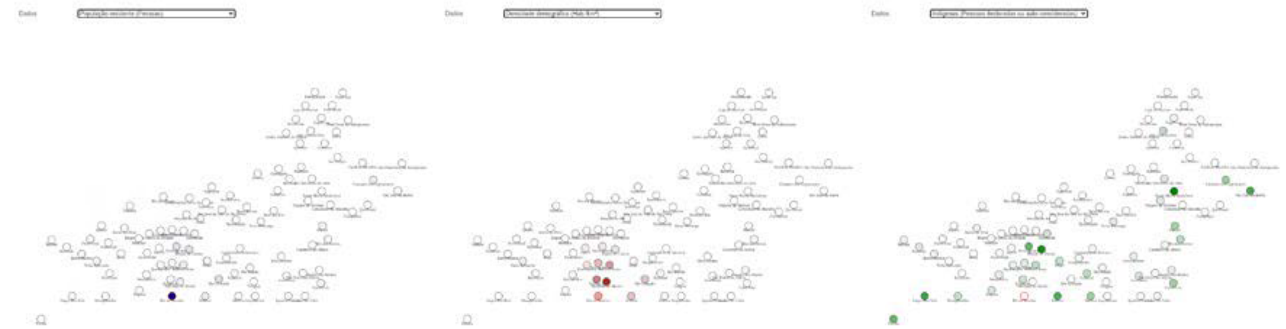

Figure 17 Carto-centric presenting data from the 2023 Census for Rio de Janeiro state municipalities, capital-centered.

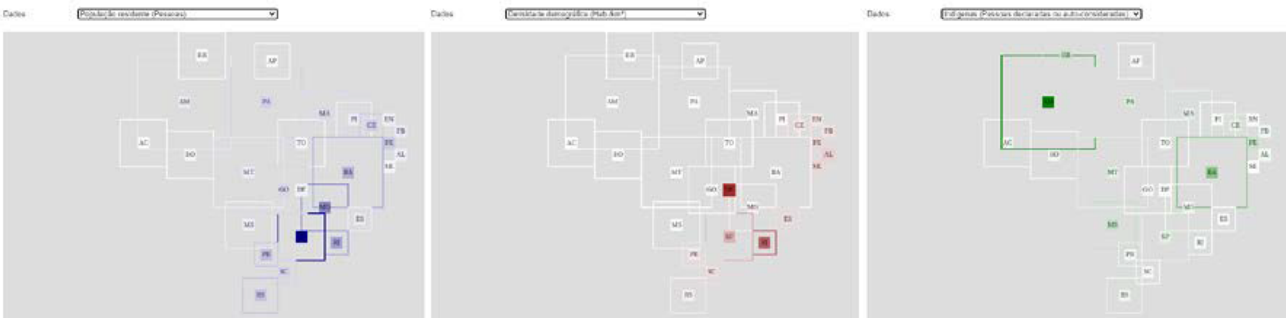

Figure 18 Carto-step presenting data from the 2023 Census for Brazilian Federation Units.

Figure 19 Three-dimensional carto-step presenting data from the 2023 Census for Brazilian Federation Units.

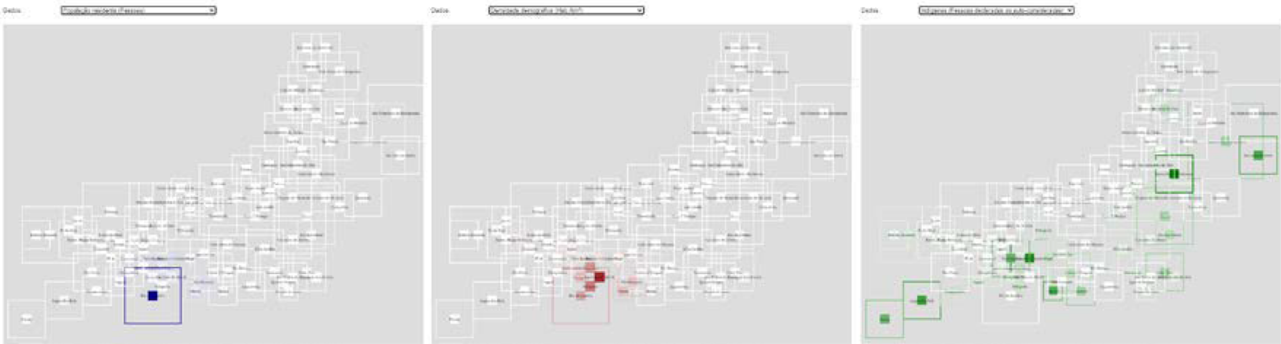

Figure 20 Carto-step presenting data from the 2023 Census for Rio de Janeiro state municipalities.

4 Conclusions, applications and future work

Through the generated presentations, it is possible to confirm the visualizations' validity in helping the comprehension of data sets relating to the represented geographical units: obviously, neither every proposed visualization will suit any and all territories, with the most appropriate option remaining on the representation producer's choice; however, increasing the available options will always be desirable, in order to favor data interpretation through cartograms.

As pointed out, the choices of values used in the tuples for the sizing and positioning of the units were individually arbitrated, according to essentially visual adequacy criteria, such as avoiding overlapping of units and maintain the closest possible proximity to their relative geographical positions: for territories with a small number of units, this definition in "manual" mode is still viable; however, when applying this task to larger federation units (for example, the most municipalities' State is Minas Gerais, having 853 units to represent), workload for value-defining becomes inappropriate.

From this perspective, Observable's usage is justified by the possibility of subsequently developing an online graphical interface that allows the loading of new sets of data for the cartograms, especially those originating from the Population Census carried out in Brazil at 2023. For such an online interface, it is also envisaged the possibility of loading additional maps from files in widely used formats such as SHP, KML or KMZ, to be converted into the special cartograms presented here.

Moving forward with this proposal, we sought to develop this interface in applications that offer simplified options in relation to raw coding, which is beyond the scope of this work: Observable has modules that allow the manipulation of files in shapefile format, but this resource proved to be limited in the initial stages of development.

Getting around such limitations, an alternative that has proven to be interesting in favor of developers not used to raw code is the KNIME Analytics Platform (2024), based on the Eclipse platform, which allows the creation of visual workflows composed of modules connected by inputs and outputs with simplified parameter configuration, and which is free and open development. Also, for the appropriate manipulation of source files, it is necessary to have a broad understanding of the vector map formats used today, and for this reason, Geographic Information Systems (GIS) applications such as QGIS and ArcGIS are important tools for this comprehension, in order to achieve the final goal of offering a simplified and user-friendly application.

The first versions produced using KNIME required a small readaptation of the data structures, since the paradigm explored for storing and manipulating values in Observable was based on multidimensional arrays, while in KNIME two-dimensional tables are used for such functions. Below are the first flowcharts for the carto-centric and carto-step proposals already presented, which, once converted from the resources present in the KNIME platform, will give rise to interactive applications with interfaces similar to the corresponding screenshots.

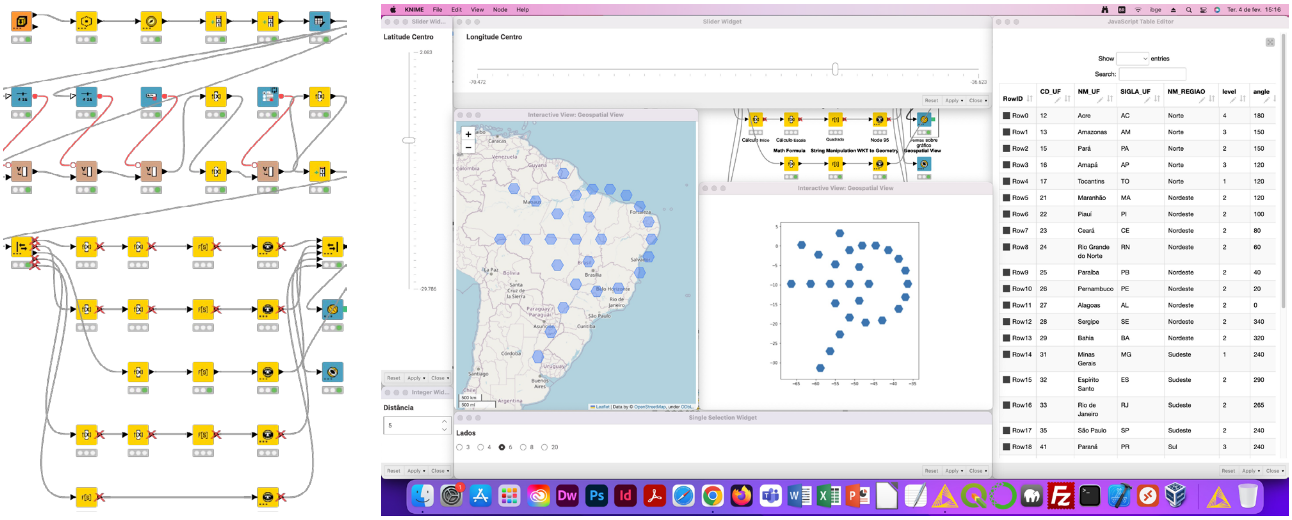

Figure 21 KNIME flowchart and screenshot for the carto-centric proposal.

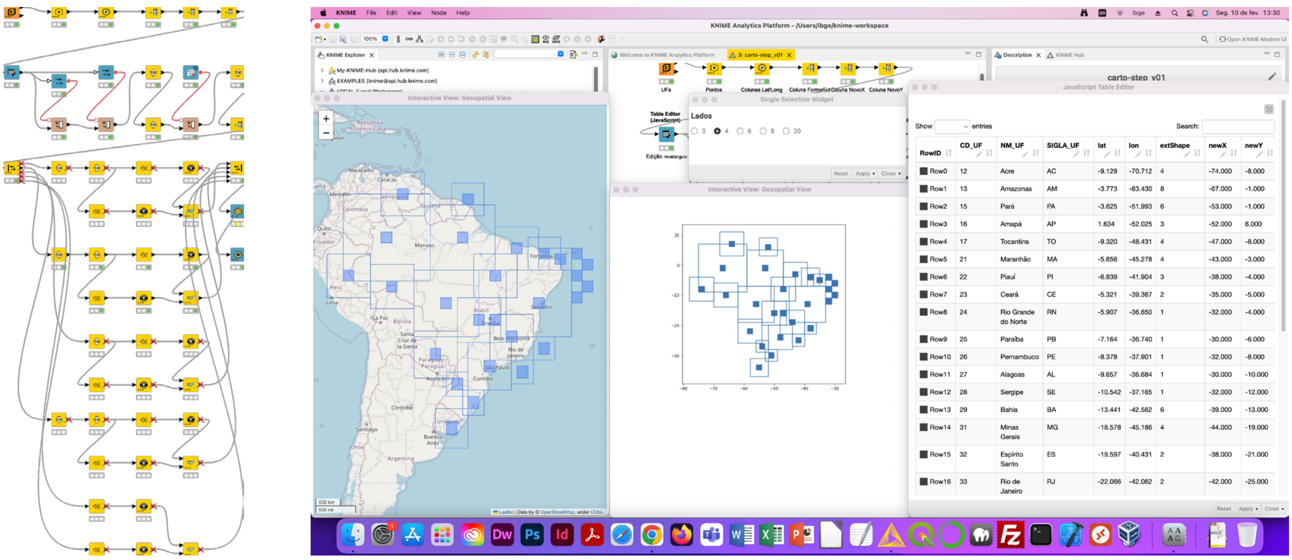

Figure 22 KNIME flowchart and screenshot for the carto-step proposal.

References

Andrienko, G., Andrienko, N., & Savinov, A. (2001). Choropleth maps: classification revisited. In *17th International Congress on Acoustics – ICA* (Conference proceedings). Rome, Italy (pp. 1209–1219).

Barkowsky, T., Latecki, L. J., & Richter, K. F. (2000). Schematizing maps: Simplification of geographic shape by discrete curve evolution. In C. Freksa, C. Habel, W. Brauer, & K. F. Wender (Eds.), *Spatial cognition II: Integrating abstract theories, empirical studies, formal methods, and practical applications* (pp. 41–53). Springer. https://doi.org/10.1007/3-540-45460-8_4

Barreto, M. C., Kosminsky, D., & Esperança, C. (2018). Hexagonal hierarchical cartogram: Towards a thematic map of Brazil. *InfoDesign – Revista Brasileira de Design da Informação*, *15*(1), 45–62. https://doi.org/10.51358/id.v15i1.563

Bertin, J. (1983). *Semiology of graphics*. University of Wisconsin Press.

Knime. (2024). *Open for Innovation | KNIME*. https://www.knime.com/

Munzner, T. (2014). *Visualization analysis and design*. CRC Press.

Observable. (2024). *Observable: Build expressive charts and dashboards with code*. https://observablehq.com/

Pinheiro, T. M., & Esperança, C. (2018). Interactive visualization of trivariate georeferenced data. *Journal of Information and Data Management*, *9*(3), 245. https://doi.org/10.5753/jidm.2018.2043

Speckmann, B., & Verbeek, K. (2010). Necklace maps. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 881–889. https://doi.org/10.1109/TVCG.2010.180

Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2). Cheshire, CT: Graphics Press.

Tversky, B. (2000). Some ways that maps and diagrams communicate. In C. Freksa, C. Habel, W. Brauer, & K. F. Wender (Eds.), *Spatial cognition II: Integrating abstract theories, empirical studies, formal methods, and practical applications* (pp. 72–79). Springer. https://doi.org/10.1007/3-540-45460-8_6

Yau, N. (2013). *Data points: Visualization that means something*. John Wiley & Sons.

About the authors

Luiz Nascimento
luizpaulo@cos.ufrj.br
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brasil

Regina Reznik
regina.reznik@ibge.gov.br
Instituto Brasileiro de Geografia e Estatística
Rio de Janeiro, Brasil

Tarsus Pinheiro
tarsus.pinheiro@ibge.gov.br
Instituto Brasileiro de Geografia e Estatística
Rio de Janeiro, Brasil

Cláudio Esperança
esperanc@cos.ufrj.br
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brasil

Ricardo Farias
rfarias@cos.ufrj.br
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brasil

Submission date/*Artigo recebido em*: 10/5/2024
Approvement date/*Artigo aprovado em*: 24/2/2025