Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Syn-SC: Generating High-Volume Synthetic Point Data with Target Continuity and Smoothness

Syn-SC: Gerando Dados Pontuais Sintéticos de Alto Volume com Continuidade e Suavidade Especificadas

Raphael Gonçalves de Campos [1], João Vitor Meza Bravo [2] e Silvana Philippi Camboim [3]

1 Federal University of Paraná, Curitiba, Brazil. raphageoc@gmail.com.
ORCID: https://orcid.org/0000-0001-5409-2877
2 Federal University of Uberlândia, Uberlândia, Brazil. jvmbravo@gmail.com.
ORCID: https://orcid.org/0000-0002-5457-3192
3 Federal University of Paraná, Curitiba, Brazil. silvanacamboim@gmail.com.
ORCID: https://orcid.org/0000-0003-3557-5341

Received: 07.2025 | Accepted: 10.2025

Abstract: When point data is aggregated and depicted on maps, the choice of thematic mapping technique depends on the cartographer's skill and the spatial structure of the data. MacEachren and DiBiase (1991) argued that two structural variables deserve explicit attention: continuity (the proportion of space occupied by events) and smoothness (the degree of variation between neighbouring locations). However, empirical studies seldom explicitly isolate these variables because real-world datasets rarely span a convenient range of values. Syn-SC closes this gap; it is a self-contained QGIS 3 Processing plug-in which synthesizes high-volume point datasets, and the continuity and smoothness of these can be defined independently. The Scale Assistant tool tessellates any area of interest with size-adaptive hexagons and reports a two-value smoothness window — floor and ceiling — defined by the binary extreme levels 1 and 100. A rule-based dispatcher then selects one of three generative solvers: an exhaustive brute-force solver for small grids of up to 16 cells, a checkerboard heuristic for large grids, and an iterative optimizer that rapidly converges whenever the requested smoothness lies within the reported window or below. Benchmarks demonstrate that Syn-SC matches the requested continuity precisely, achieves smoothness targets within ±1 percentage point and generates sets approaching one million points in seconds. Syn-SC, therefore, provides cartographers, usability researchers, and AI developers with shareable, perceptually parameterised point datasets that were previously unavailable.
Keywords: Smoothness. Continuity. Map Perception. Geo-big Data. Synthetic Data.

Resumo: Quando os dados pontuais são agregados e representados em mapas, a escolha da técnica de mapeamento temático depende da habilidade do cartógrafo e da estrutura espacial dos dados. MacEachren & DiBiase (1991) argumentaram que duas variáveis estruturais merecem atenção explícita: continuidade (a proporção do espaço ocupado pelos eventos) e suavidade (o grau de variação entre locais vizinhos). No entanto, estudos empíricos raramente isolam explicitamente essas variáveis, pois os conjuntos de dados do mundo real raramente abrangem uma faixa conveniente de valores. O Syn-SC preenche essa lacuna; é um plug-in de processamento QGIS 3 independente que sintetiza conjuntos de dados pontuais de alto volume, e a continuidade e suavidade desses dados podem ser definidas independentemente. A ferramenta Scale Assistant (Assistente de escala) divide qualquer área de interesse em hexágonos adaptáveis ao tamanho e relata uma janela de suavidade de dois valores — piso e teto — definida pelos níveis extremos binários 1 e 100. Um distribuidor baseado em regras seleciona então uma das três soluções generativos: uma solução exaustiva de força bruta para pequenas grades de até 16 células, uma heurística de tabuleiro de xadrez para grades grandes e um otimizador iterativo que converge rapidamente sempre que a suavidade solicitada está dentro da janela relatada ou abaixo dela. Benchmarks demonstram que o Syn-SC corresponde precisamente à continuidade solicitada, atinge as metas de suavidade dentro de ±1 ponto percentual e gera conjuntos que se aproximam de um milhão de pontos em segundos. O Syn-SC, portanto, fornece aos cartógrafos, pesquisadores de usabilidade e desenvolvedores de IA conjuntos de dados de pontos compartilháveis e parametrizados perceptualmente que antes não estavam disponíveis.
Palavras-chave: Suavidade. Continuidade. Percepção Cartográfica. Geo-big Data. Dados Sintéticos.

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-78917

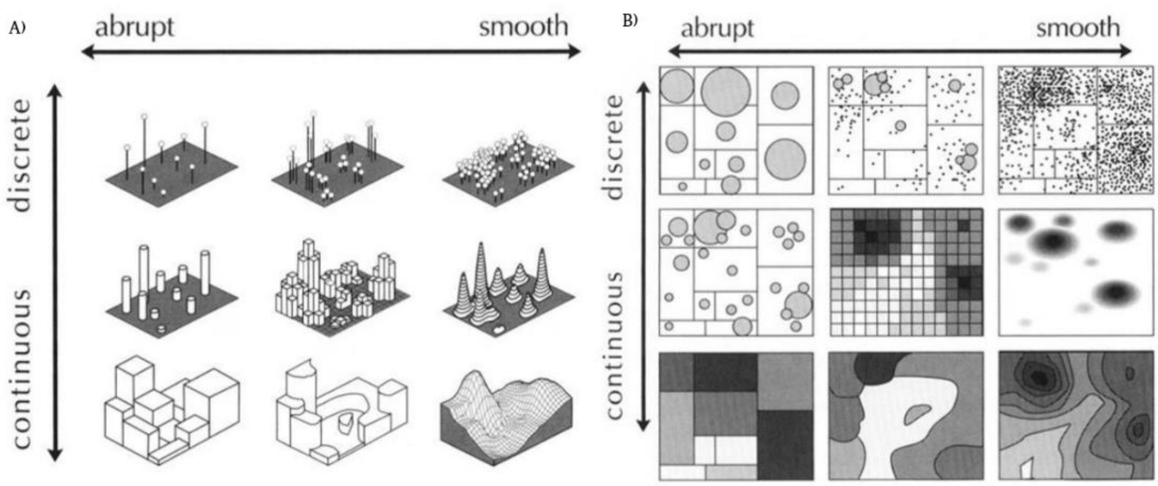
1 INTRODUCTION

Since the 1950s, cartographic research has been developed from Arthur Robinson's functional approach (Montello, 2002), focusing on generating knowledge that culminated in creating maps intended for a specific purpose and audience (Elzakker & Griffin, 2013; Robinson et al., 2023). These fundamentals guided researchers' investigations, especially in Thematic Cartography under the UCD (User-Centred Design) approach (Roth et al., 2015, 2017). Although those studies produced robust guidelines for symbolisation, they did so with relatively small, well-behaved datasets.

The present era of Geo Big Data, marked by unprecedented velocity, volume, and variety, reopens fundamental questions about the most effective ways to represent point phenomena, which may now number in the millions. Interestingly, this challenge also invites us to take a fresh look at experimental work from the 1970s, 1980s and 1990s, when cartographic researchers drew on methods from experimental psychology to study the perceptual effects of graphic variables. Studies by Flannery (1971), Provin (1977) and Slocum, (1983) examined the influence of symbol size, density and clustering on readers' accuracy, while subsequent research investigated cognitive workload and colour perception. Their findings provided robust, transferable principles, yet these principles have generally not been applied in a big data landscape. Revisiting these principles with contemporary processing power and interactive visual analytics environments offers an opportunity to provide well-grounded, technologically updated solutions to today's high-volume problems.

MacEachren & DiBiase, 1991 emphasised the importance of continuity — the extent to which a phenomenon is spatially pervasive, ranging from isolated occurrences to near-ubiquity — and smoothness — the abruptness of change between neighbouring locations — in guiding the choice of visual technique, particularly for aggregated point distributions. Figure 1 reproduces their continua and the associated symbolisation methods (after MacEachren (1992)). Continuity and smoothness describe how a geographical phenomenon is distributed in space (Slocum et al., 2022); these properties remain important even when datasets are streamed in real-time from heterogeneous sources.

Figure 1 - (A) Models of geographic phenomena arranged along discrete-continuous and abrupt-smooth continua. (B) A set of symbolisation methods is appropriate for these models



Source: MacEachren (1992).

However, two issues have limited new empirical work on these properties. Firstly, there's a scarcity of point datasets that exhibit the diversity of continuity and smoothness needed for systematic experimentation (Roth et al., 2019). Real-world examples, when available, rarely cover the full spectrum of these properties. Secondly, although seminal perceptual studies from the 1970s to the 1990s continue to inform cartographic practice, there are opportunities to re-examine their insights in the context of today's data volumes and interactive visual analytics environments. By introducing synthetic geospatial datasets whose continuity and smoothness can be prescribed on a large scale, this study aims to encourage further engagement with these studies while also providing the practical data needed for rigorous experimentation.

Synthetic geospatial data provides a practical route to experimental control, provided it can be generated at scale while faithfully encoding the required spatial properties. This study, therefore, tackles two questions: (i) how can point datasets be generated so that continuity and smoothness can be prescribed independently across a wide range, and (ii) which computational strategies deliver those targets fast enough for large-volume user studies? The guiding hypothesis is that a tiered, spatial-stochastic workflow implemented as a QGIS Processing plugin can achieve the requested characteristics within a margin of error of ±1 percentage point. This paper introduces Syn-SC, an open-source plugin that (a) tessellates any area of interest with a size-adaptive hexagonal grid, (b) reports a two-value smoothness window that bounds feasible targets, (c) routes generation to one of three solvers—brute force, checker-board heuristic, or iterative optimiser—according to grid size and the target's position within that window, (d) verifies that the finished grid meets the requested continuity and smoothness, and (e) exports a ready-to-use point layer for cartographic experiments.

2 RELATED WORK

Geospatial big-data research agendas emphasise that cartographers now confront unprecedented velocity, volume and variety and must devise empirical methods that scale accordingly (Robinson et al., 2017). However, systematic experimentation is hampered because real point datasets are often unavailable, owing to privacy regulations, proprietary constraints or heterogeneous provenance, and seldom cover the full combination of spatial properties one wishes to test.

One approach to overcoming these limitations is the use of synthetic data. In statistics, Rubin's pioneering proposal to release multiple-imputed synthetic microdata for disclosure limitation laid the conceptual foundations (Raghunathan et al., 2003). Since then, synthetic datasets have become mainstream in data science, supporting the preservation of privacy, benchmarking, and bias diagnosis. Geospatial researchers are following suit: the Python-based RADIAN tool, for instance, generates configurable point and polygon datasets for teaching and rapid prototyping (Gorry & Mooney, 2025).

Synthetic data has long been proposed as a remedy when real datasets are sparse, confidential or too cumbersome to share. Early studies focused on population-scale micro-simulation to replicate census characteristics (e.g., Hermes & Poulsen, 2012_review of spatial microdata methods) and on privacy-oriented public-use files that replace sensitive counts with model-based draws (Quick & Waller (2018) Bayesian spatio-temporal synthesiser for US mortality data). These streams establish that synthetic records can preserve analytical utility while bypassing disclosure and licensing barriers, but they target demographic inference rather than cartographic experimentation. Closer to the needs of this study are task-specific generators that allow researchers to configure spatial patterns for benchmarking, teaching, or rapid prototyping. Vu et al. (2022) Spider framework offers six canonical point distributions with affine transformations and a web interface for reproducible downloads. Mannino & Abouzied (2019) Synner adopts a mixed-initiative design interface, allowing users to sketch attribute distributions with immediate visual feedback guiding the synthesis process.

While such tools provide valuable control over density, skewness, and geometry, they do not expose higher-level perceptual parameters; in particular, they cannot set continuity and smoothness independently, nor do they embed cartographic metrics to verify these properties.

Within cartography, Campos et al. (2021) proposed quantitative metrics for continuity and smoothness in voluminous point clouds. They observed that expert cartographers struggle to visually assess these properties once the dataset size exceeds a few thousand observations. However, there is still no open framework that produces synthetic point datasets whose continuity and smoothness can be independently prescribed and empirically verified. Existing stochastic generators—typically based on Poisson or Matérn processes (Dirrler et al., 2020) — seldom incorporate explicit continuity constraints. Recent usability experiments comparing choropleth, graduated-symbol, isoline or hex-bin maps (Arnold et al., 2017; Kawakami et al., 2024; Słomska-Przech & Gołębiowska, 2021; Wallner & Kriglstein, 2020) likewise overlook continuity and smoothness, treating the spatial distribution of the test data as an incidental property rather than an experimental factor..

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-78917

Addressing these shortfalls, the Syn-SC QGIS Plugin introduced in the next section combines spatial-stochastic processes with heuristic optimisation to generate point datasets at a Big-Data scale while honouring user-specified continuity and smoothness targets, thereby enabling experiments that reconnect classic perceptual insights with contemporary Geo Big Data challenges.

3 METHODS

The following sections describe Syn-SC, an open-source QGIS processing plugin that creates large synthetic point datasets with continuity and smoothness matching user-specified targets. The source code is openly archived on Github (https://anonymous.4open.science/r/syn_sc-210C).

3.1 Plug-in architecture and control flow

The pipeline, summarised in Figure 2, proceeds through five macro-steps. The generator starts when the user supplies five parameters: Area of Interest (AOI) layer L_{AOI} , scale, target continuity c_{tgt} , target smoothness s_{tgt} and maximum points per cell $p_{\max}$. The Scale Assistant constructs a hexagon grid, derives the two-value smoothness window and dispatches generation to one of three solvers: Brute, Heuristic or Iterative. All solvers share the same point-realisation routine. An optional Metric Evaluator verifies that the resulting grid meets the requested continuity (c_{tgt}) and smoothness (s_{tgt}).

Table 1 – Syn-SC Plugin Processing Algorithms

Label	Processing algorithm	Purpose
SA	Scale Assistant	Constructs a size-adaptive hexagon grid, reports the binary smoothness window, and recommends a solver.
BRU	Syn-SC Brute	Exhaustive search for grids ≤ 16 cells.
ITE	Syn-SC Iterative	Gradient-descent optimiser for intermediate targets on medium-sized grids.
HEU	Syn-SC Heuristic	Checker-board generator that reaches the binary floor on large grids or when very low SM targets are requested.
ME	Syn-SC Metric Evaluator	Verifies that the continuity and smoothness of any output layer match the specification.

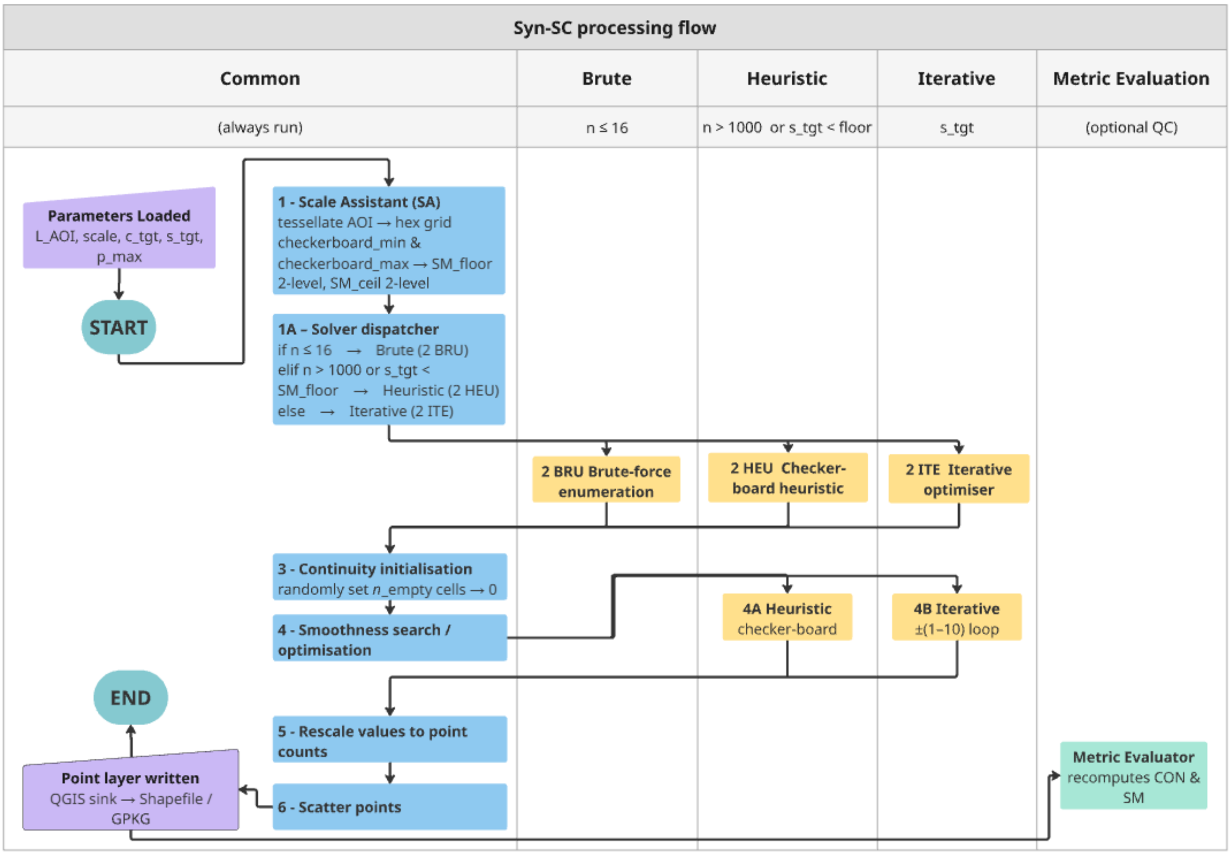
Source: The authors (2025).

3.1.1 GRID CHOICE

The parameters smoothness and continuity define gaps between the occurrences of the phenomena (continuity) and the variation of the phenomenon between different regions of space (smoothness). Thus, we understand it is convenient to use a grid formed by hexagons to generate data with these characteristics. Furthermore, the grid of hexagons allows the discretisation of space in portions with equal area values, which enables the control of the number of points generated between distinct parts of the region of interest.

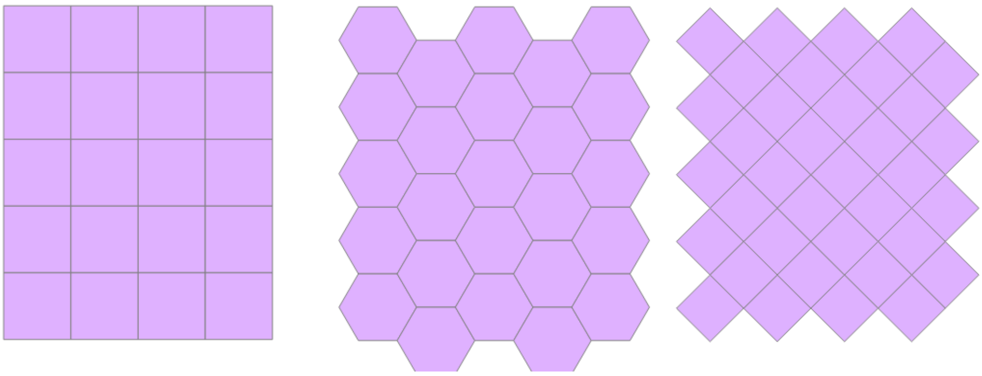
Among the regular polygons that cover a plane without overlaps and spaces between them (Carr et al., 1992), the hexagon is the one with the most sides (6) when compared to diamonds and rectangles (Figure 3). Given that the number of points generated within a cell is both uniform and random, linear patterns are expected to develop along the edges of the polygons in the grid. Then, the hexagons will increase the possibilities of generating patterns in the grid and, consequently, create a greater variety of spatial distributions.

Figure 2 - Control flow of the Syn-SC plug-in



Source: The authors (2025).

Figure 3 - Possible grid types formed by cells in the format of regular polygons without overlaps and spaces between cells



Source: The authors (2025).

3.1.2 QUANTIFICATION AND DE-CORRELATION OF SMOOTHNESS AND CONTINUITY

To generate synthetic spatial data with prescribed levels of continuity c_{tgt} and smoothness s_{tgt}, the algorithm must first measure—and then independently control—each characteristic. We adopt the procedure proposed by Campos et al. (2021).

Continuity (CON) is the proportion of populated cells in the grid, as formalised in Eq. (1):

$$\mathrm{CON} = \frac{n_{pop}}{n_{\text{total}}} = 1 - \frac{n_{\text{empty}}}{n_{\text{total}}} \tag{1}$$

where n_{empty} is the number of cells whose value equals zero, n_{pop} is the number of cells whose value is non-zero and n_{total} is the total number of cells (Figure 4A). Larger CON therefore indicates a more continuous (less fragmented) spatial distribution, whereas smaller values signify greater fragmentation.

For every populated cell i, the smoothness (SM) is the mean absolute difference between its value v_i and the values v_j of its first-order neighbours that are also populated, as in Eq. (2) and Figure 4B:

$$\mathrm{SM_i} = \frac{1}{k_i} \sum_{j \in \mathcal{N}(i)} |v_i - v_j| \tag{2}$$

where $\mathcal{N}(i)$ is the set of $k_i$valid neighbours.

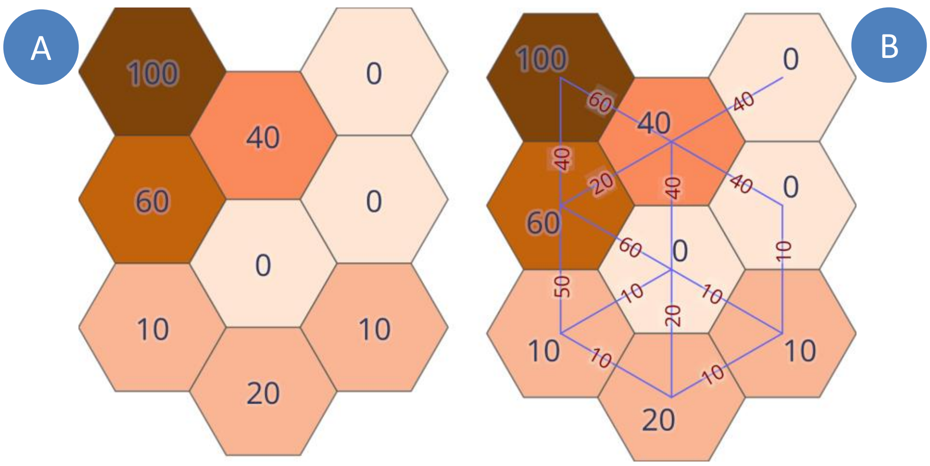
The global smoothness score in Eq. (3) is the arithmetic mean of $\mathrm{SM_i}$ over all populated cells:

$$\mathrm{SM} = \frac{1}{n_{\mathrm{pop}}} \sum_{i=1}^{n_{\mathrm{pop}}} SM_i \tag{3}$$

with n_{pop} the number of non-empty cells.

By excluding empty cells from the neighbourhood calculations, the method prevents the spatial pattern that defines continuity from influencing the local difference estimates that define smoothness. This separation decorrelates the two metrics, allowing the code to nudge CON and SM toward their respective targets c_{tgt} and s_{tgt} in two independent loops.

Figure 4 - (A) Example of a grid with continuity CON = 67%, three of the nine empty cells. (B) Depiction of the methodology used to calculate the smoothness (SM) of a hexagonal grid. The number on each edge is the absolute difference $|v_i - v_j|$, SM is calculated without counting empty cells



Source: The authors (2025).

3.2 Continuity initialisation (common to all solvers)

Given the target continuity c_{tgt}, the script first converts it to the required count of empty cells:

$$n_{empty}^{req} = n_{total} \frac{100 - c_{tgt}}{100} \tag{4}$$

and rounds the result to the nearest integer.

A Python Random Number Generator (RNG) is used to draw the corresponding number of cell IDs; those cells have their *value* set to 0, while all others are assigned a provisional random value in [1,100]. Because continuity is independent of smoothness, this step is performed once before any smoothness optimisation begins.

3.3 Smoothness optimisation

3.3.1 ITERATIVE OPTIMISER

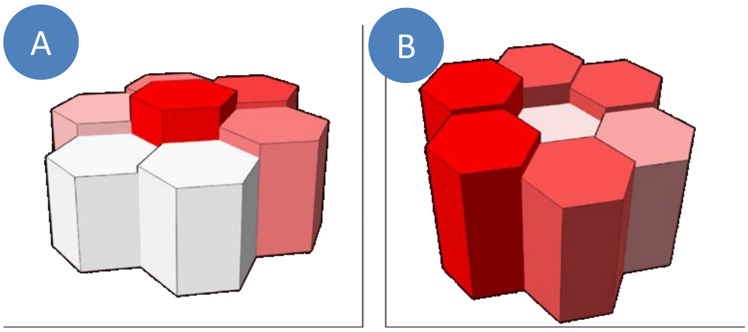
Smoothness is a neighbourhood metric, so changing one cell affects up to six neighbours. The script,

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-78917

therefore, applies an iterative approximation loop:

a) Draw a random populated cell i.
b) Compute the current global smoothness using Equation (3).
c) Determine whether SM must increase or decrease to reach s_{tgt} .
d) Perturb v_i by a random integer $\Delta \in [1,10]$:
e) increase v_i (add Δ) when SM < s_{tgt} and v_i is below the mean of its neighbours,
f) decrease v_i(subtract Δ) in the symmetric cases, as illustrated in Figure 5a–b.
g) Repeat until $\mid SM - s_{tgt} \mid \leq 0.01 * s_{tgt}$, i.e., global smoothness is within 1% of the target.

Figure 5 - Representation of cases where the cell value is smaller than its neighbours' values (a) and the cell value is larger than its neighbours ' values (b)



Source: The authors (2025).

3.4 Minimum attainable smoothness

When $s_{tgt} \leq SM_{floor}$ 2-level, the optimisation loop in item 3.3.1 often stalls or diverges because every perturbation of a cell value v_i affects up to six neighbour pairs simultaneously. To establish the true lower bound of global smoothness, an exhaustive brute-force search was implemented for small grids, and a lightweight checkerboard heuristic was derived for production-sized grids.

3.4.1 BRUTE-FORCE ENUMERATION (SMALL GRIDS)

The brute-force script assigns only the extreme values 1 and 100 to populated cells because the pair-wise smoothness term in Eq. (5)

$$\mid vi - vj \mid = 99 \tag{5}$$

is minimized when the absolute value of the difference is 99.

For a fully populated grid (c_{tgt}=100%) containing n_{total}, the number of candidate grids is

$$N_{comb} = 2^{n_{total}} \tag{6}$$

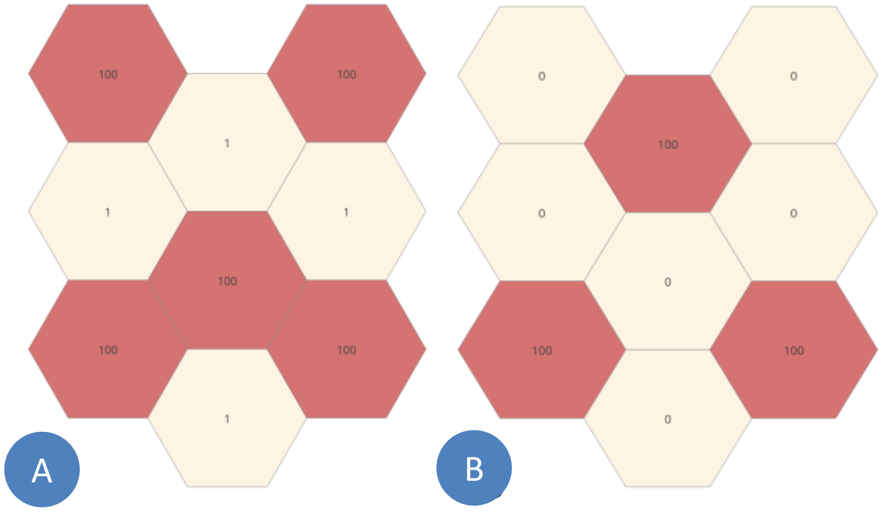
which grows exponentially (e.g., $n_{total} = 64 \implies N_{comb} \approx 4 \times 10^{19}$).

The algorithm iterates over every combination, computes SM (Eq. 2), and stores the minimum encountered value.

a) Pattern 1. If the upper-left cell is 100, the column beneath alternates 1 - 100 - 1 …
b) Pattern 2. If the upper-left cell is 100, the row to its right alternates 1 - 100 - 1 …

Both patterns (Figure 6A) yield the same global minimum for the fully populated case. A smoothness of exactly 0% arises only when every populated cell is surrounded exclusively by empty cells (Figure 6B), an arrangement that is impossible once two populated cells are adjacent to each other.

Figure 6 – (A) Grid with defined values for minimum smoothness (B) Grid with values to obtain smoothness equal to 0%



Source: The authors (2025).

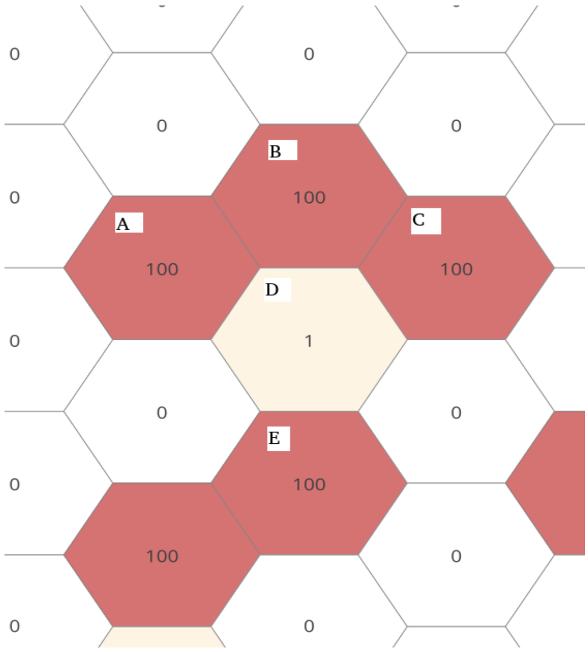
3.4.2 BRUTE-FORCE WITH c_{tgt} < 100 %

When c_{tgt}<100%, an additional rule emerges: any populated cell that has ≥ 50 % empty neighbours must take the value 100. Conversely, populated cells whose neighbour set is ≥ 50 % populated alternate between 1 and 100 to keep SM low (Figure 7). Although brute force still guarantees the absolute minimum, the number of possible grids grows so quickly that the method becomes infeasible once the total cell count exceeds about 20.

When continuity is lower than 100 %, some cells are fixed at 0, so the set of free cells has a size $n_{\text{pop}} = \frac{c_{\text{tgt}}\, n_{\text{total}}}{100}$. The search space now contains $2^{n_{pop}}$ combinations, but the same exponential explosion applies: beyond n_{pop} ≈ 20, exhaustive search is no longer practical.

The brute-force results revealed one additional rule: any populated cell with≥ 50% empty neighbours must take the value 100. All other populated cells alternate between 1 and 100 to maintain a low SM (Figure 8). These observations underpin the heuristic below.

Figure 7 - Cells with more than 50% of their neighbours with a value of 0



Source: The authors (2025).

3.4.3 HEURISTIC MINIMUM-SMOOTHNESS ALGORITHM

Based on the two empirical patterns, we devised a three-step heuristic:

a) Initial checkerboard: assign 1 and 100 in a row-and-column chessboard pattern to every populated cell (works for any continuity).
b) High-contrast reinforcement: for each populated cell with ≥ 50 % empty neighbours, set its value to 100.
c) Low-contrast correction: scan remaining populated cells; if *all* populated neighbours share the central value, flip half of them to the opposite extreme until at least one neighbour differs.

The procedure runs in $O(n_{pop})$ time and, in practice, achieves a global SM no more than 0.5 percentage points above the brute-force optimum on test grids of up to 16 × 16.

4 RESULTS AND DISCUSSION

4.1 Minimum-smoothness validation

Datasets with different continuity values were generated to verify the results of this algorithm. Additionally, the brute force algorithm was applied, enabling a comparison of the results from the two algorithms. We generated datasets with the following continuities: 10, 20, 30, 40, 50, 60, 70, 80, and 90. The brute force and heuristic algorithms were applied for each dataset, and the results are presented in Table 2.

Table 2 - Comparison of the minimum smoothness obtained by the brute-force algorithm and the the first Heuristic version

Continuity	Minimum smoothness (%) by brute-force algorithm	Minimal smoothness (%) heuristic v1	Difference
10	0	0	0
20	17.7	17.7	0
30	12.5	12.5	0
40	20.1	22.9	-2.8
50	14.0	14.0	0
60	22.3	24.5	-2.2
70	29.5	31.6	-2.1
80	27.7	33.6	-5.8
90	30.7	30.7	0

Source: The authors (2025).

The largest discrepancy (−5.9 pp) occurs at 80% continuity. Re-initialising that grid by flipping the first cell from 1 to 100 (maintaining the 100–1–100 pattern) reduces smoothness to 20.1%, exactly matching the brute-force result. This re-initialisation has, therefore, been added as a pre-step (Heuristic v2). Table 2 shows the results of applying this new step to the data in Table 1.

With the new pre-step, the heuristic matches brute-force in every case except for 70% continuity (Table 3); the remaining error is −2.06 pp. The 0.5 percentage-point convergence threshold set earlier is therefore met.

Table 3 - Comparison of the minimum smoothness obtained by the brute force and the heuristic algorithms with the new step

Continuity	Minimum smoothness (%) by brute-force algorithm	Minimum smoothness (%) heuristic v2	Difference
40	20.09	20.09	0.00
60	22.36	22.36	0.00
70	29.50	31.56	-2.06
80	27.75	27.75	0.00

Source: The authors (2025).

4.2 Effect of the point-cap parameter

Before export, the optimiser's cell values $v_i \in [1,100]$ are rescaled to the user-defined p_{max} by:

$$v_i' = v_i * p_{max}/100 \tag{8}$$

Points are then placed uniformly at random inside each populated hexagon, with v_i determining the exact count. Because the optimiser never "sees" p_{max}, this post-hoc scaling introduces a small rounding error: some low values are truncated to zero, and high values are compressed. In our tests, setting p_{max}=20 altered the global smoothness by no more than two percentage points—a drift that disappeared once SM was recomputed on the output point layer.

4.3 Iterative convergence inside and outside the two-value window

In a grid consisting of 8370 cells, for c_{tgt} = 60% the Scale Assistant reports SM_{floor} = 68.4% and SM_{ceil} = 68.9% . Running the Iterative solver with an ambitious s_{tgt} = 30% demonstrates that intermediate values (1–100) allow the optimiser to beat the binary floor:

```
Initial SM = 68.69 % (target 30.00)
Iter 500: SM = 61.33 %
Iter 1000: SM = 55.25 %
Iter 2000: SM = 48.01 %
Iter 3000: SM = 43.01 %
Iter 4000: SM = 39.40 %
Iter 5000: SM = 36.65 %
Iter 6000: SM = 34.71 %
Iter 7000: SM = 33.22 %
Iter 8000: SM = 32.27 %
Iter 9000: SM = 31.55 %
Iter 10000: SM = 31.10 %
Iter 11000: SM = 30.78 %
Iter 12000: SM = 30.57 %
Iter 13000: SM = 30.44 %
Iter 14000: SM = 30.32 %
Finished after 14172 iterations: SM = 30.30 %→ tolerance met
```

The solver converged in 14172 iterations and 135.36 s on a commodity laptop, confirming that targets **below** the two-value window are achievable when all integer levels are available. The number of points generated was 986444 with p_{max} = 100.

4.4 Visual assessment at map scale 1 : 50 000 and implications for perceptual testing

Figure 8 shows four representative outputs rendered at the working map scale of 1:50,000. Each panel contains (top) the realised point cloud and (bottom) the corresponding hexagons coloured by cell value; captions list the target smoothness s_{tgts}, continuity c_{tgt}, hexagon diameter d_{hex} and point cap p_{max} .

With medium hexagons (d_{hex}≈300 m), the prescribed parameters are met, and the point pattern appears visually natural (Fig. 8A,8B and 8C). Two scale-dependent effects nevertheless require attention:

- Low continuity with small cells (c_{tgt}≤30, d_{hex}≈300 m) limits intra-cell clustering, yielding an over-uniform cloud (Fig. 8E). Increasing d_{hex} or relaxing the continuity target restores a more believable spatial texture.

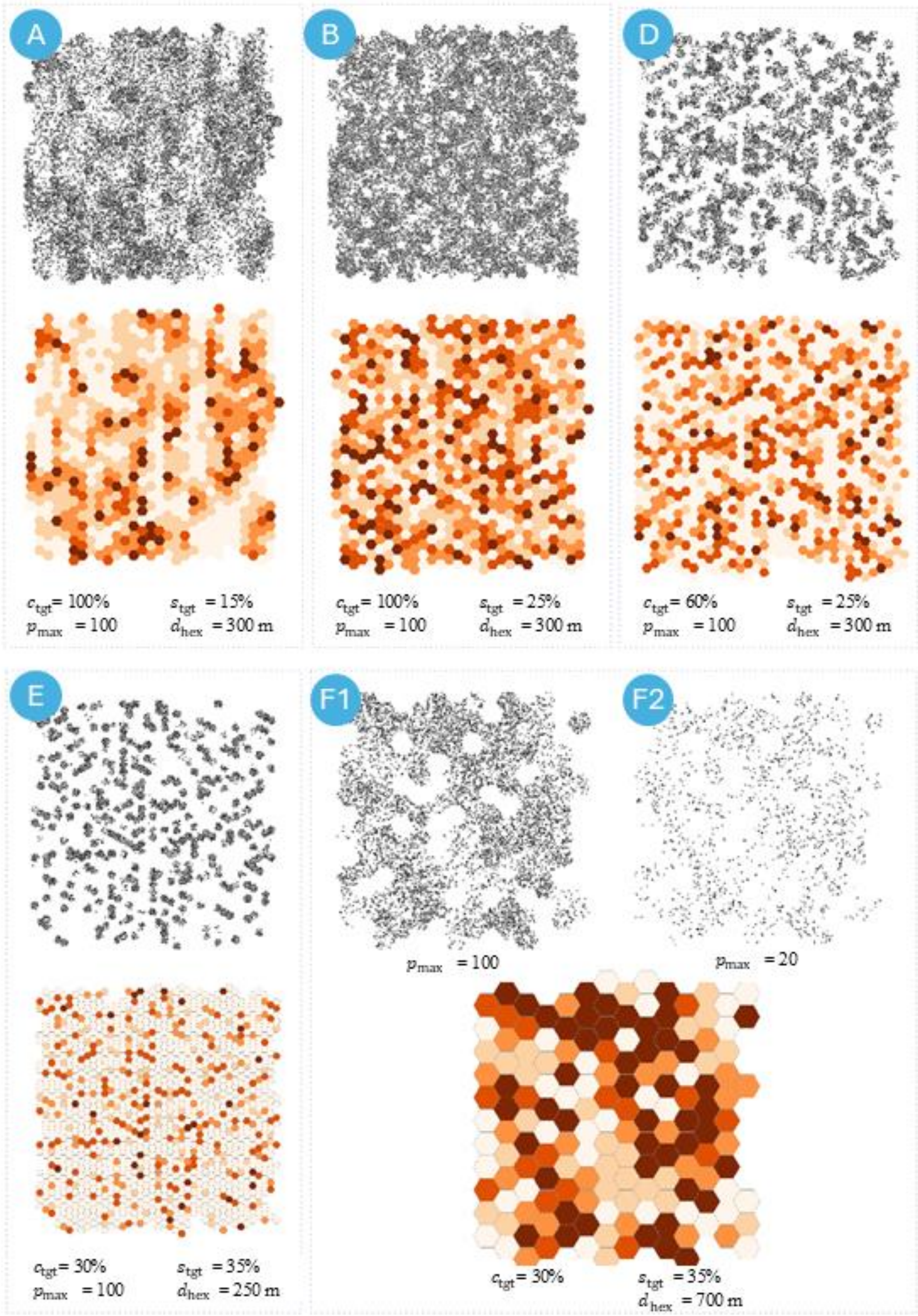
 DOI: http://dx.doi.org/10.14393/rbcv77n0a-78917

- Large cells combined with a high point cap (d_{hex}≈700 m, p_{max}=) reveal the hexagonal lattice (Fig. 8F1). Lowering the cap to p_{max}=20 breaks up the honeycomb pattern without altering continuity or smoothness (Fig. 8F2).

When the target smoothness approaches the theoretical minimum (s_{tgt}≈40%), faint horizontal banding may emerge, a direct consequence of the optimal checkerboard arrangement on a hexagonal grid.

These aesthetic effects are crucial for perceptual user studies, the principal use case for Syn-SC: any visible artefact risks biasing participants' judgements of symbol size, density, or texture. Researchers are therefore advised to (i) select d_{hex} and pmax so that the underlying mesh is imperceptible at the display scale and (ii) maintain continuity above roughly thirty per cent when realistic clustering is desired. With these adjustments, an experimenter can deliver stimuli that look convincingly natural while still preserving the strict, reproducible control over continuity and smoothness that the algorithm provides.

Figure 8 - Synthetic point clouds generated with Syn-SC at map scale 1:50000



Source: The authors (2025).

5 CONCLUSIONS

Syn-SC has been presented as an open-source QGIS plug-in that generates voluminous point clouds whose continuity (proportion of populated cells) and smoothness (mean absolute contrast between neighbouring cells) can be adjusted independently. A third parameter—the point-per-cell cap—limits local density and disguises the synthetic origin of the data, producing clouds that appear natural to the casual observer and are, therefore, suitable for perceptual experiments.

The validation exercises demonstrate that a tiered search strategy is effective: brute-force enumeration achieves the exact minimum smoothness for very small grids, a lightweight checkerboard heuristic remains within half a percentage point of that optimum on production-sized grids, and the standard iterative optimiser converges rapidly whenever lies inside the two-value window reported by the Scale Assistant or lower values. Even when the cap is reduced to 20 points per cell, global smoothness drifts by no more than two percentage points and recalculating the metric on the realised points corrects the discrepancy without affecting continuity.

Because the generator operates on any polygon layer, an area of interest can be subdivided, and each region can be assigned its own parameter set. This capability enables the simulation of heterogeneous phenomena—such as settlement density gradients—while maintaining complete control over the underlying statistics, thereby providing realistic yet shareable test beds for cartographic usability studies, data-hungry AI models, and classroom demonstrations.

Future work will focus on accelerating the initial value assignment for low-smoothness targets, introducing alternative point-placement kernels that permit sub-cell clustering, and extending the framework to polygonal and time-varying phenomena. Taken together, these enhancements will further integrate classic cartographic principles with contemporary Geospatial Big Data analytics, expanding the range of applications for synthetic geospatial data.

Acknowledgments

To the Graduate Program in Geodetic Sciences of the Federal University of Paraná (UFPR) and the Coordination for the Improvement of Higher Education Personnel (CAPES) for their encouragement of research and financial support (Process 1541928).

Authors Contribution

Author Raphael Gonçalves de Campos participated in the conceptualization, application development, article writing, and investigation. Author Silvana Philippi Camboim participated in the application development, writing, investigation, review, and editing of the article. Author João Vitor Meza Bravo participated in the writing, review, and editing of the article.

Conflicts of Interest

No conflict of interest.

References

Arnold, N. D., Jenny, B., & White, D. (2017). Automation and evaluation of graduated dot maps. International Journal of Geographical Information Science, 31(12), 2524–2542. https://doi.org/10.1080/13658816.2017.1359747

Campos, R. G. de, Paiva, C., Bravo, J. V. M., & Camboim, S. P. (2021). A proposition to define boundaries based on the smoothness and the continuity of voluminous point data phenomena. Abstracts of the ICA, 3, 1–2. https://doi.org/10.5194/ica-abs-3-42-2021

Carr, D. B., Olsen, A. R., & White, D. (1992). Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data. Cartography and Geographic Information Systems, 19(4), 228–236.

https://doi.org/10.1559/152304092783721231

Dirrler, M., Dörr, C., & Schlather, M. (2020). A generalization of Matérn hard-core processes with applications to max-stable processes. Journal of Applied Probability, 57(4), 1298–1312. https://doi.org/10.1017/jpr.2020.66

Elzakker, C. P. J. M. van, & Griffin, A. L. (2013). Focus on geoinformation users: Cognitive and use/user issues in contemporary cartography. GIM International, 27(8), 20–23.

Flannery, J. J. (1971). The relative effectiveness of some common graduated point symbols in the presentation of quantitative data. Cartographica, 8(2), 96–109. https://doi.org/10.3138/J647-1776-745H-3667

Gorry, P., & Mooney, P. (2025). RADIAN – A tool for generating synthetic spatial data for use in teaching and learning. Cartography and Geographic Information Science, 52(3), 314–330. https://doi.org/10.1080/15230406.2024.2377981

Hermes, T. B., & Poulsen, C. N. (2012). Spatial microsimulation for the generation of synthetic populations: A review. Journal of Geographical Systems, 14(4), 437–467. https://doi.org/10.1007/s10109-012-0164-8

Kawakami, Y., Yuniar, S., & Ma, K.-L. (2024). HexTiles and Semantic Icons for MAUP-Aware Multivariate Geospatial Visualizations (arXiv:2407.16897). arXiv. https://doi.org/10.48550/arXiv.2407.16897

MacEachren, A. M. (1992). Visualizing Uncertain Information. Cartographic Perspectives, 13, Artigo 13. https://doi.org/10.14714/CP13.1000

MacEachren, A. M., & DiBiase, D. (1991). Animated Maps of Aggregate Data: Conceptual and Practical Problems. Cartography and Geographic Information Systems, 18(4), 221–229. https://doi.org/10.1559/152304091783786790

Mannino, M., & Abouzied, A. (2019). Is this Real? Generating Synthetic Data that Looks Real. Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, 549–561. https://doi.org/10.1145/3332165.3347866

Montello, D. R. (2002). Cognitive research in GIScience: The role of user-centered design and user studies. Transactions in GIS, 6(1), 1–11. https://doi.org/10.1111/1467-9671.00091

Provin, R. W. (1977). The Perception of Numerousness on Dot Maps. The American Cartographer, 4(2), 111–125. https://doi.org/10.1559/152304077784080374

Quick, H., & Waller, L. A. (2018). Using spatiotemporal models to generate synthetic data for public use. Spatial and Spatio-temporal Epidemiology, 27, 37–45. https://doi.org/10.1016/j.sste.2018.08.004

Raghunathan, T. E., Reiter, J. P., & Rubin, D. B. (2003). Multiple imputation for statistical disclosure limitation. Journal of official statistics, 19(1), 1.

Robinson, A. C., Demšar, U., Moore, A. B., Buckley, A., Jiang, B., Field, K., Kraak, M. J., Camboim, S. P., & Sluter, C. R. (2017). Geospatial big data and cartography: Research challenges and opportunities for making maps that matter. International Journal of Cartography, 3(sup1), 32–60. https://doi.org/10.1080/23729333.2016.1278151

Robinson, A. C., Kettunen, P., Delazari, L., & Çöltekin, A. (2023). New directions for the state of the art and science in Cartography. International Journal of Cartography, 9(2), 143–149. https://doi.org/10.1080/23729333.2023.2216334

Roth, R. E., Çöltekin, A., Delazari, L., Filho, H. F., Griffin, A., Hall, A., Korpi, J., Lokka, I., Mendonça, A., Ooms, K., & Elzakker, C. P. J. M. van. (2017). User studies in cartography: Opportunities for empirical research on interactive maps and visualizations. International Journal of Cartography, 3(sup1), 61–89. https://doi.org/10.1080/23729333.2017.1288534

Roth, R. E., Kelly, M., Underwood, N., Lally, N., Vincent, K., & Sack, C. (2019). Interactive & Multiscale Thematic Maps: A Preliminary Study. Abstracts of the ICA, 1, 1–2. https://doi.org/10.5194/ica-abs-1-315-2019

Roth, R. E., Ross, K. S., & MacEachren, A. M. (2015). User-Centered Design for Interactive Maps: A Case Study in Crime Analysis. ISPRS International Journal of Geo-Information, 4(1), Artigo 1. https://doi.org/10.3390/ijgi4010262

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-78917

Slocum, T. A. (1983). Predicting Visual Clusters on Graduated Circle Maps. The American Cartographer, 10(1), 59–72. https://doi.org/10.1559/152304083783948168

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2022). Thematic Cartography and Geovisualization (4ª ed.). CRC Press.

Słomska-Przech, K., & Gołębiowska, I. M. (2021). Do Different Map Types Support Map Reading Equally? Comparing Choropleth, Graduated Symbols, and Isoline Maps for Map Use Tasks. ISPRS International Journal of Geo-Information, 10(2), Artigo 2. https://doi.org/10.3390/ijgi10020069

Vu, T., Migliorini, S., Eldawy, A., & Belussi, A. (2022). Spatial Data Generators. In: Spatial Gems, Volume 1 (1ª ed., Vol. 46, p. 13–24). Association for Computing Machinery. https://doi.org/10.1145/3548732.3548736

Wallner, G., & Kriglstein, S. (2020). Multivariate Visualization of Game Metrics: An Evaluation of Hexbin Maps. Proceedings of the Annual Symposium on Computer-Human Interaction in Play, 572–584. https://doi.org/10.1145/3410404.3414233

Biography of the main author

Raphael Gonçalves de Campos was born in Curitiba, Paraná, Brazil. He holds a degree in Cartographic and Surveying Engineering from the Federal University of Paraná (UFPR). He received his Master's and Doctoral degrees in Geodetic Sciences from the Graduate Program in Geodetic Sciences of the Federal University of Paraná (PPGCG/UFPR).

Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional – CC BY. Esta licença permite que outros distribuam, remixem, adaptem e criem a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito pela criação original.