

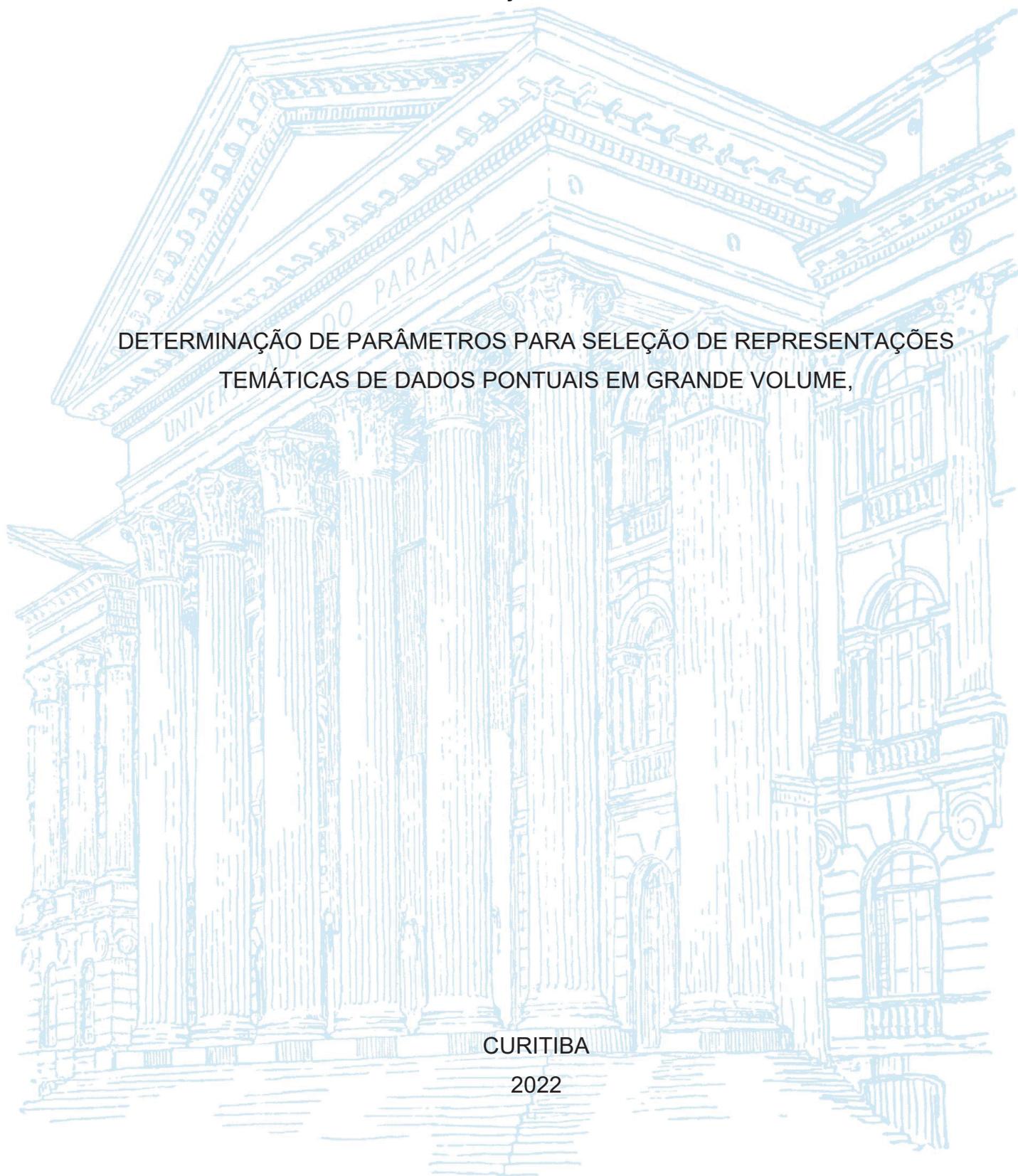
UNIVERSIDADE FEDERAL DO PARANÁ

RAPHAEL GONÇALVES DE CAMPOS

DETERMINAÇÃO DE PARÂMETROS PARA SELEÇÃO DE REPRESENTAÇÕES  
TEMÁTICAS DE DADOS PONTUAIS EM GRANDE VOLUME,

CURITIBA

2022



RAPHAEL GONÇALVES DE CAMPOS

DETERMINAÇÃO DE PARÂMETROS PARA SELEÇÃO DE REPRESENTAÇÕES  
TEMÁTICAS DE DADOS PONTUAIS EM GRANDE VOLUME

Tese apresentada como requisito parcial à obtenção do título de Doutor em Ciências Geodésicas no curso de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra da Universidade Federal do Paraná – UFPR.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Silvana Philippi Camboim

Co-Orientador: Prof. Dr. João Vitor Meza Bravo

CURITIBA

2022

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
UNIVERSIDADE FEDERAL DO PARANÁ  
SISTEMA DE BIBLIOTECAS – BIBLIOTECA CIÊNCIA E TECNOLOGIA

Campos, Raphael Gonçalves de.

Determinação de parâmetros para seleção de representações temáticas de dados pontuais em grande volume. / Raphael Gonçalves de Campos. – Curitiba, 2022.

1 recurso on-line : PDF.

Tese (Doutorado) – Universidade Federal do Paraná, Setor de Ciências da Terra, Programa de Pós-Graduação em Ciências Geodésicas.

Orientadora: Profa. Dra. Silvana Philippi Camboim.

Coorientador: Prof. Dr. João Vitor Meza Bravo.

1. Geodésia. 2. Cartografia. 3. Mapas. 4. Dados de pesquisa. I. Camboim, Silvana Philippi. II. Bravo, João Vitor Meza. III. Universidade Federal do Paraná. Programa de Pós-Graduação em Ciências Geodésicas. IV. Título.



MINISTÉRIO DA EDUCAÇÃO  
SETOR DE CIÊNCIAS DA TERRA  
UNIVERSIDADE FEDERAL DO PARANÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS  
GEODÉSICAS - 40001016002P6

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação CIÊNCIAS GEODÉSICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **RAPHAEL GONÇALVES DE CAMPOS** intitulada: **DETERMINAÇÃO DE PARÂMETROS PARA SELEÇÃO DE REPRESENTAÇÕES TEMÁTICAS DE DADOS PONTUAIS EM GRANDE VOLUME**, sob orientação da Profa. Dra. SILVANA PHILIPPI CAMBOIM, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 25 de Julho de 2022.

Assinatura Eletrônica

26/07/2022 17:14:57.0

SILVANA PHILIPPI CAMBOIM  
Presidente da Banca Examinadora

Assinatura Eletrônica

28/07/2022 10:48:15.0

CLAUDIA ROBBI SLUTER  
Avaliador Externo (UNIVERSIDADE DO RIO GRANDE DO SUL)

Assinatura Eletrônica

28/07/2022 10:47:25.0

ANDREA FARIA ANDRADE  
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

28/07/2022 12:57:56.0

ANDRE LUIZ ALENCAR DE MENDONÇA  
Avaliador Externo (UNIVERSIDADE FEDERAL DO AMAZONAS)

Centro Politécnico - Caixa Postal 19001 - CURITIBA - Paraná - Brasil

CEP 81531-980 - Tel: (41) 3361-3153 - E-mail: cpgcg@ufpr.br

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.

Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 209720

Para autenticar este documento/assinatura, acesse <https://www.prppg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp>  
e insira o código 209720

## **AGRADECIMENTOS**

Agradeço à minha família pelo suporte e incentivo.

Agradeço aos meus colegas de laboratório pelo companheirismo e apoio.

Agradeço a minha orientadora Silvana Philippi Camboim e ao meu coorientador João Vitor Meza Bravo pelos conhecimentos transmitidos e paciência.

Agradeço a todos os professores que de alguma forma me transmitiram conhecimentos.

Agradeço à Universidade Federal do Paraná pela oportunidade da formação acadêmica.

Agradeço à CAPES pela bolsa que me possibilitou ter dedicação exclusiva

Agradeço a todos aqueles que de alguma forma contribuíram para que esta etapa fosse concluída.

## RESUMO

A era dos dados traz possibilidades e desafios às diversas áreas do conhecimento. No contexto da Cartografia, possibilita a representação de diversos fenômenos com alta taxa de resolução temporal, no entanto, exige a adequação de técnicas de representação cartográficas e ferramentas para a automatização da construção de mapas com maior celeridade e eficiência. A teoria que dá suporte para a construção de mapas temáticos a partir de dados pontuais (MacEachren e Dibiase, 1991) foi desenvolvida em um contexto tecnológico diferente do atual e não foram encontradas evidências experimentais prévias específicas da sua validação. Majoritariamente, estudos que avaliaram a eficiência dos diferentes tipos de mapas na solução de tarefas negligenciaram as características apontadas na teoria original: a suavidade e continuidade. Neste sentido, resta verificar como tais conceitos influenciam na eficiência da visualização de agrupamentos diferentes tipos de mapas, para que possam ser utilizados na escolha da técnica adequada para cada conjunto de dados. Para realizar esta verificação, o problema é decomposto em três experimentos. O primeiro consistiu em definir medidas quantitativas para suavidade e continuidade e comparar com as percepções de cartógrafos experientes. No segundo, baseado nas medidas quantitativas, foi desenvolvido um algoritmo para gerar dados pontuais sintéticos com as características de suavidade e continuidades predefinidas. Com a possibilidade de gerar dados pontuais com as características controladas, foi construído um experimento para verificar a eficiência de diferentes tipos de mapas, aplicados a dados com diferentes características de suavidade e continuidade, na comunicação de agrupamentos. Com este último experimento, apoiado nas percepções dos participantes, foi possível produzir um modelo de técnicas de representação cartográfica temática por suavidade e continuidade semelhante ao proposto por MacEachren e Dibiase (1991). Pode-se dizer, baseado nos resultados obtidos, que a hipótese deste trabalho foi confirmada, ou seja, os conceitos de suavidade e continuidade tem uma influência positiva na escolha das técnicas de representação a serem utilizadas.

**Palavras-chave:** Cartografia. Cartografia Temática. Suavidade. Continuidade.

## ABSTRACT

The age of data brings possibilities and challenges to the various areas of knowledge. In Cartography, it enables the representation of various phenomena with high temporal resolution rate. However, it requires adapting cartographic representation techniques and tools to automate maps' construction with greater speed and efficiency. The theory that supports the construction of thematic maps from point data (MacEachren and Dibiase, 1991) was developed in a technological context different from the current one, and no specific previous experimental evidence of its validation was identified. Primarily, studies that evaluated the efficiency of different types of maps in solving tasks neglected the characteristics pointed out in the original theory: smoothness and continuity. In this regard, it is still necessary to verify how such concepts influence the efficiency of visualization of clusters of different types of maps so that they can be used in choosing the appropriate technique for each data set. The problem was decomposed into three experiments to perform this verification. The first one consisted in defining quantitative measures for smoothness and continuity and comparing them with the perceptions of experienced cartographers. In the second one, based on the quantitative measures, an algorithm was developed to generate synthetic point data with predefined smoothness and continuity characteristics. Finally, with the possibility of generating point data with the controlled characteristics, an experiment was designed to verify the efficiency of different maps, applied to data with different smoothness and continuity characteristics in the communication of clusters. With this last experiment, supported by the participants' perceptions, it was possible to produce a model of thematic cartographic representation techniques with smoothness and continuity similar to the one proposed by MacEachren and Dibiase (1991). Based on the results obtained, the hypothesis of this work was confirmed: the concepts of smoothness and continuity positively influence the choice of representation techniques to be used.

**Keywords:** Cartography. Thematic Cartography. Smoothness. Continuity.

## LISTA DE FIGURAS

FIGURA 1 – MODELOS DE DISTRIBUIÇÃO DE DADOS E AS TÉCNICAS ADEQUADAS PARA SUA REPRESENTAÇÃO TEMÁTICA.....	16
FIGURA 2 – FLUXOGRAMA.....	20
FIGURE 3 – (A) MODELS OF GEOGRAPHIC PHENOMENA ARRANGED ALONG DISCRETE-CONTINUOUS AND ABRUPT- SMOOTH CONTINUA. (B) A SET OF SYMBOLIZATION METHODS APPROPRIATE FOR THESE MODELS. ....	25
FIGURE 4 – ALGORITHM DESIGN.....	27
FIGURE 5 – WEB VIEWER FOR COMPARING TWO DIFFERENT SPATIAL DISTRIBUTIONS. THE PERSPECTIVE REPRESENTATIONS ARE FOR DATASET A (LEFT) AND B (RIGHT) .....	36
FIGURE 6 – THE PARAMETERS DISPERSION CALCULATED BY THE ALGORITHM.....	38
FIGURE 7 – WEB MAPS PRESENTED DURING THE INTERVIEW.....	38
FIGURE 8 – FREQUENCY OF EACH TYPE OF RESPONSE, BASED ON THE RANK MADE BY THE INDIVIDUALS FOR THE LEVEL OF SMOOTHNESS.....	39
FIGURE 9 – FREQUENCY OF EACH TYPE OF RESPONSE, BASED ON THE RANK MADE BY THE INDIVIDUALS FOR THE LEVEL OF CONTINUITY. ....	40
FIGURE 10 – EDUCATION IN CARTOGRAPHY, MAP PRODUCTION FREQUENCY, AND DIVERGENCE WITH THE ALGORITHM'S ORDERING.....	44
FIGURE 11 – RELATIONSHIP BETWEEN THE POSITION DIVERGENCE AND THE CALCULATED VALUES FOR THE CONTINUITY VARIABLE.....	46
FIGURE 12 – RELATIONSHIP BETWEEN THE POSITION DIVERGENCE AND THE CALCULATED VALUES FOR THE SMOOTHNESS VARIABLE. ....	46
FIGURE 13 – (A) MODELS OF GEOGRAPHIC PHENOMENA ARRANGED ALONG DISCRETE-CONTINUOUS AND ABRUPT-SMOOTH CONTINUA. (B) A SET OF SYMBOLISATION METHODS ARE APPROPRIATE FOR THESE MODELS.....	53
FIGURE 14 – FLOWCHART CONTAINING THE STEPS FOR GENERATING SYNTHETIC POINT DATA.....	55
FIGURE 15 – POSSIBLE GRID TYPES FORMED BY CELLS IN THE FORMAT OF REGULAR POLYGONS WITHOUT OVERLAPS AND SPACES BETWEEN CELLS. ....	57

FIGURE 16 – EXAMPLE OF A GRID WITH 67% CONTINUITY, THREE OF THE NINE CELLS BEING EMPTY.....	58
FIGURE 17 – ILLUSTRATION OF HOW THE SMOOTHNESS OF A HEXAGONAL GRID IS CALCULATED. ....	59
FIGURE 18 – REPRESENTATION OF CASES WHERE THE CELL VALUE IS SMALLER THAN ITS NEIGHBOURS' VALUES (A), AND THE CELL VALUE IS LARGER THAN ITS NEIGHBOURS (B). ....	62
FIGURE 19 – (A) GRID WITH DEFINED VALUES FOR MINIMUM SMOOTHNESS (B) GRID WITH VALUES TO OBTAIN SMOOTHNESS EQUAL TO 0%. ....	63
FIGURE 20 – CELLS WITH MORE THAN 50% OF THEIR NEIGHBOURS WITH A VALUE OF 0.....	64
FIGURA 21 – POINT DISTRIBUTIONS AND THE GRIDS GENERATED BY THE ALGORITHM. WHERE, S - SMOOTHNESS, C - CONTINUITY, CS - CELL SIZE AND NP - NUMBER OF POINTS PER POLYGON.....	67
FIGURA 22 – RECORTES DE SETORES UTILIZADOS PARA A GERAÇÃO DOS MAPAS.....	79
FIGURA 23 – DISTRIBUIÇÃO COM 100% DE CONTINUIDADE E 55% DE SUAVIDADE.....	81
FIGURA 24 – DISTRIBUIÇÃO COM 100% DE CONTINUIDADE E 85% DE SUAVIDADE.....	81
FIGURA 25 – DISTRIBUIÇÃO COM 100% DE CONTINUIDADE E 70% DE SUAVIDADE.....	81
FIGURA 26 – DISTRIBUIÇÃO DAS CARACTERÍSTICAS DOS DADOS UTILIZADOS.....	83
FIGURA 27 – ESPECIFICAÇÕES DE CONSTRUÇÃO DE CADA TIPO DE MAPA.....	84
FIGURA 28 – ARQUITETURA CLIENTE-SERVIDOR UTILIZADA PARA O DESENVOLVIMENTO DA PLATAFORMA DE APLICAÇÃO DOS TESTES .....	86
FIGURA 29 – DIAGRAMA DE CASOS DE USO DO SISTEMA.....	87
FIGURA 30 – DIAGRAMA DE CLASSES .....	88
FIGURA 31 – REPRESENTAÇÃO DAS SITUAÇÕES CONSIDERADAS PARA O CÁLCULO DO FATOR DE EFICIÊNCIA.....	91
FIGURA 32 – DIFERENTES TIPOS DE MAPAS CONSTRUÍDOS COM OS DADOS COM CARACTERÍSTICAS DE SUAVIDADE 85% E CONTINUIDADE 100%.....	93

FIGURA 33 – CLUSTERS MATEMÁTICOS REFERENTES AOS DADOS DA FIGURA 32.....	95
FIGURA 34 – PÁGINA INICIAL DO TESTE.....	95
FIGURA 35 – CARACTERIZAÇÃO DO PARTICIPANTE .....	96
FIGURA 36 – INSTRUÇÕES PARA REALIZAR O TESTE.....	97
FIGURA 37 – VÍDEO PARA APRENDIZADO DO USO DE FERRAMENTAS WEB.....	97
FIGURA 38 – MAPA PARA APRENDIZADO DAS FERRAMENTAS DA PLATAFORMA.....	98
FIGURA 39 – GRÁFICO DE BARRAS DAS MÉDIAS DE EFICIÊNCIA DE CADA TIPO DE MAPA PARA UM CONJUNTO DE DADOS ESPECÍFICO (S – SUAVIDADE, C – CONTINUIDADE).....	99
FIGURA 40 – GRÁFICO DE BOLHAS REPRESENTANDO A PERFORMANCE DOS TIPOS DE MAPAS CONFORME A SUAVIDADE E CONTINUIDADE DE CADA CONJUNTO DE DADOS.....	100
FIGURA 41 – MODELO DE PREVISÃO DAS MELHORES TÉCNICAS DE REPRESENTAÇÃO PARA INTERVALOS DE SUAVIDADE E CONTINUIDADE...	103

## LISTA DE TABELAS

TABLE 1 – EXPERIMENTAL HYPOTHESES .....	34
TABLE 2 – LEVEL OF DIVERGENCE BETWEEN THE CARTOGRAPHER'S AND THE ALGORITHM'S POSITIONS, FOR SMOOTHNESS .....	42
TABLE 3 – LEVEL OF DIVERGENCE BETWEEN THE CARTOGRAPHER'S AND THE ALGORITHM'S POSITIONS, FOR CONTINUITY .....	43
TABLE 4 – COMPARISON OF THE MINIMUM SMOOTHNESS OBTAINED BY THE BRUTE-FORCE ALGORITHM AND THE HEURISTIC ALGORITHM .....	65
TABLE 5 – COMPARISON OF THE MINIMUM SMOOTHNESS OBTAINED BY THE BRUTE FORCE ALGORITHM AND THE HEURISTIC ALGORITHM WITH THE NEW STEP.....	66
TABELA 6 – TÉCNICAS DE REPRESENTAÇÃO CARTOGRÁFICAS TEMÁTICAS ADEQUADAS PARA A SUAVIDADE E CONTINUIDADE .....	78
TABELA 7 – CARACTERÍSTICAS DOS CONJUNTOS DE DADOS SÍNTEÉTICOS UTILIZADOS NO TESTE. ....	82
TABELA 8 – TÍTULO DA TABELA .....	101
TABELA 9 – QUALIDADE DAS REGRESSÕES PARA CADA TIPO DE MAPA.....	102

## LISTA DE SIGLAS

CSS – *Cascading Style Sheets*

2D – Bidimensional

GI Science – *Geospatial Information Science*

GPU – *Graphics Processing Units*

HTTP – *Hypertext Transfer Protocol*

IBGE – Instituto Brasileiro de Geografia e Estatística

LCD – *Liquid Crystal Display*

LISA – *Local Indicator Of Spatial Association*

SQL – *Structured Query Language*

UCD – User-Centered Design

WMS – *Web Map Service*

TIFF – *Tagged Image File Format*

## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO .....</b>	<b>16</b>
1.1	OBJETIVOS .....	19
1.1.1	Objetivo Geral.....	19
1.1.2	Objetivos Específicos.....	19
1.2	ESTRUTURA DA TESE .....	20
<b>2.</b>	<b>COMPUTING SMOOTHNESS AND CONTINUITY TO GENERATE THEMATIC MAPS FROM HIGH-VOLUME POINT DATA.....</b>	<b>23</b>
2.1	ABSTRACT .....	24
2.2	INTRODUCTION .....	24
2.3	METHOD .....	27
2.3.1	System Design .....	28
2.3.1.1	Input data .....	28
2.3.1.2	Scale definition.....	29
2.3.1.3	Creategrid and count point in polygon .....	29
2.3.1.4	Filter data .....	30
2.3.1.5	Compute proportion of empty cells.....	30
2.3.1.6	Compute average of differences between cells.....	31
2.4	EXPERIMENT .....	33
2.4.1	Participants .....	35
2.5	TASKS, APPARATUS, AND ANALYSIS .....	35
2.6	RESULTS AND DISCUSSION .....	37
2.6.1	Algorithm output.....	37
2.6.2	Experiment results .....	39
2.6.2.1	Hypothesis 1: Cartographers agree on the order of smoothness.....	39
2.6.2.2	Hypothesis 2: Cartographers agree on the order of continuity .....	40
2.6.2.3	Hypothesis 3: Cartographers agree with the algorithm on the order of smoothness.....	41

2.6.2.4 Hypothesis 4: Cartographers agree with the algorithm on the order of continuity .....	43
2.6.2.5 Hypothesis 5: The time of formal education in cartography or the frequency of map production brings the reading of the concepts closer .....	44
2.6.2.6 Hypothesis 6: The further apart the smoothness and continuity measure values, the greater the agreement between cartographers and the algorithm. ....	45
2.7 CONCLUSIONS .....	46
2.8 DATA AND CODES AVAILABILITY STATEMENT .....	47
2.9 DISCLOSURE STATEMENT .....	48
2.10 FUNDING .....	48
2.11 REFERENCES .....	48
<b>3. SYNTHETIC HIGH-VOLUME POINT DATA GENERATION BASED ON SMOOTHNESS AND CONTINUITY .....</b>	<b>49</b>
3.1 ABSTRACT .....	50
3.2 INTRODUCTION .....	50
3.2.1 Part one of the solution: From the evolution of cartographic representations to Geo Big Data .....	52
3.2.2 Continuity and Smoothness .....	53
3.2.3 Synthetic Geospatial Data .....	54
3.2.4 Research Problem .....	54
3.3 METHODS AND RESULTS .....	55
3.3.1 Starting settings .....	57
3.3.2 Quantification of smoothness and continuity parameters .....	57
3.3.3 Generation of synthetic spatial data .....	59
3.3.4 Continuity definition .....	60
3.3.5 Definition of smoothness .....	60
3.4 CONCLUSIONS .....	69
3.5 REFERENCES .....	70

<b>4.</b>	<b>RECONHECIMENTO VISUAL DE AGRUPAMENTOS PELOS USUÁRIOS PARA VALIDAÇÃO DE RANQUEAMENTO DE REPRESENTAÇÕES TEMÁTICAS</b>	<b>72</b>
4.1	RESUMO .....	72
4.2	ABSTRACT .....	73
4.3	INTRODUÇÃO.....	74
4.4	MATERIAIS E MÉTODOS.....	77
4.3.1	Materiais.....	77
4.3.2	Experimento.....	77
4.3.2.1	Definição dos Mapas .....	78
4.3.2.2	Definição da Base Cartográfica .....	78
4.3.2.3	Gerando os Dados Sintéticos.....	80
4.3.2.4	Definição das Técnicas de Representação Temática.....	83
4.3.2.5	Definição da Paleta de Cores.....	84
4.3.3	Desenvolvimento da Plataforma de Teste.....	85
4.3.3.1	Diagrama de Casos de Uso.....	87
4.3.3.2	Diagrama de Classes.....	88
4.3.3.3	Interface .....	89
4.3.4	Definição da Medida de Eficiência.....	90
4.3.5	Análise das Regiões Percebidas .....	91
4.5	RESULTADOS .....	92
4.4.1	Mapas Construídos e <i>Clusters</i> Matemáticos .....	93
4.4.2	Aplicação Web Construída.....	95
4.4.3	Resultados do experimento .....	98
4.6	CONCLUSÃO .....	106
4.7	REFERÊNCIAS .....	107
<b>5.</b>	<b>GENERAL CONCLUSIONS.....</b>	<b>109</b>
<b>6.</b>	<b>REFERÊNCIAS.....</b>	<b>113</b>
<b>7.</b>	<b>APÊNDICES.....</b>	<b>115</b>

APÊNDICE 1 – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)  
115

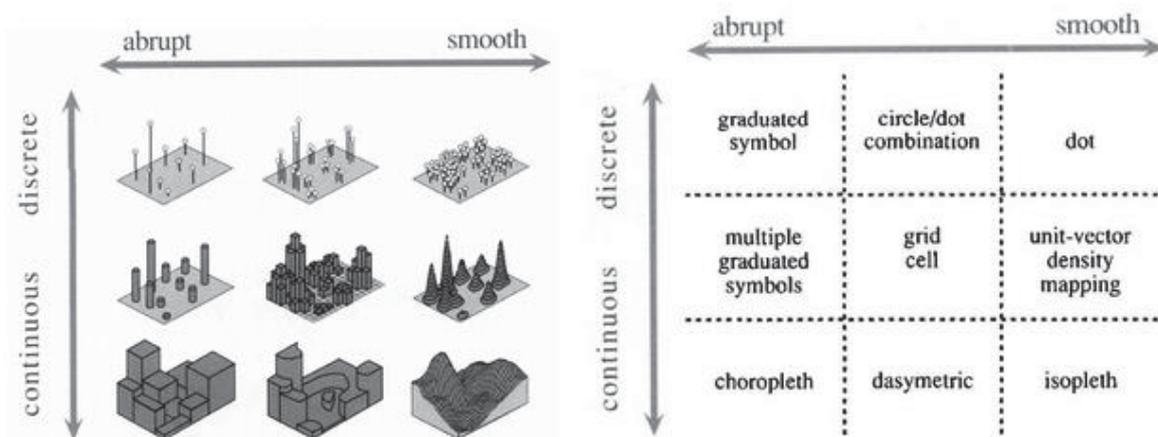
APÊNDICE 2 – MAPAS UTILIZADAS NO TESTE DE PERCEPÇÃO..... 118

## 1. INTRODUÇÃO

A natureza dos fenômenos geográficos é um fator fundamental a ser considerado no processo de criação de representações cartográficas temáticas (ROBINSON *et al.*, 1995; SLOCUM *et al.*, 2009; SLUTER, 2008). Sabe-se que a dimensão espacial dos fenômenos geográficos é a principal maneira de se descrever sua natureza (FOREST, 1999). Entretanto, podemos fazê-lo por uma outra perspectiva, considerando-se a valoração da suavidade e da continuidade espacial (MACEACHREN E DIBIASE, 1991; SLOCUM *et al.*, 2009).

O conceito de suavidade e continuidade de um fenômeno geográfico foi apresentado por Maceachren e Dibiase (1991) para auxiliar a escolha de técnicas de representação cartográfica temática, no que tange a formulação de mapas para dados agregados. Especificamente, a suavidade descreve a dependência ou independência entre os valores do fenômeno através do espaço, com variação ao longo dos eixos suave-abrupto. Paralelamente, a continuidade reflete o preenchimento do espaço geográfico pelo fenômeno, o que difere do sentido matemático estrito, ainda que isto seja fator essencial na escolha do método de representação. MacEachren e Dibiase (1991) classificam os fenômenos em três classes para cada característica: Suavidade - suave, meio suave-abrupto, abrupto; Continuidade - contínuo, meio contínuo-discontínuo, discontínuo. Estes autores construíram um modelo que conecta as características do fenômeno às técnicas mais adequadas à sua representação (figura 1).

FIGURA 1 – MODELOS DE DISTRIBUIÇÃO DE DADOS E AS TÉCNICAS ADEQUADAS PARA SUA REPRESENTAÇÃO TEMÁTICA



FONTE: Adaptado de MacEachren e DiBiase (1991).

As técnicas de representações temáticas são combinadas com as características do fenômeno com base nas metáforas visuais que os fenômenos evocam (ROTH *et al.*, 2019). O propósito de se utilizar metáforas visuais é construir representações que sejam acessíveis à cognição humana (SKUPIN, 1998; FABRIKANT E BUTTENFIELD, 1991). Para auxiliar na identificação das características dos fenômenos MacEachren e DiBiase (1991) trazem exemplos, que combinam com as classes de cada característica. Entretanto, a definição destas características é subjetiva e depende exclusivamente da análise visual do cartógrafo, além de não existirem evidências empíricas que demonstrem a efetividade dos mapas gerados baseado na teoria de suavidade e continuidade.

Esta teoria foi postulada em um contexto tecnológico diverso da atualidade, no qual os desafios eram representar dados pouco volumosos em mapas 2D. Por essa razão, não se sabe se sua fundamentação ainda serve à tomada de decisão para a escolha da técnica de representação temática ideal, considerando este novo contexto. Especificamente, nos tempos atuais, há uma geração massiva de dados, que são denominados *Big Data* (SAGIROGLU E SINANC, 2013). O *Big Data* é caracterizado por Laney (2001) segundo 3 dimensões: volume, variedade e velocidade. O volume refere-se à quantidade dos dados que são gerados todos os dias; a variedade remete aos diferentes tipos de dados, estruturados e não estruturados; e por fim, a velocidade é o quão rápido os dados são gerados e devem ser processados para suprir as necessidades dos usuários. Dentre esses dados, grande parte tem uma localização espacial associada e pode ser denominado de *Geo Big Data* (GOODCHILD, 2013).

A grande disponibilidade de dados geoespaciais oportuniza que os cartógrafos representem os fenômenos espaciais com alto grau de qualidade (*et al.*, 2016). É notório que as pessoas usam cada vez mais os mapas e esses produtos têm - cada vez mais - um papel importante para a tomada de decisão em problemas do dia a dia como, por exemplo, o suporte a resgates em desastres (SCHOLZ, 2018). Contudo, quando representamos quantidades massivas de dados pontuais em escala reduzida, há o desafio, uma vez que a formação de imagens mentais não é favorecida pelo volume exagerado de dados, que gera problemas nas condições geométricas de

generalização (MCMASTER E SHEA, 1992). No caso de dados pontuais, podemos citar dois principais problemas: coalescência e sobreposição.

A sobreposição ocorre quando os dados têm as mesmas coordenadas. Em consequência das suas precisões ou no caso de dados geocodificados, estes podem ser associados a um mesmo endereço. Já a coalescência se dá quando a distância entre os pontos, representados em uma mesma escala, é menor que a resolução da tela dos dispositivos. Neste caso, os símbolos dos pontos se tocam no processo de representação, ocasionando a supracitada condição geométrica (MCMASTER E SHEA, 1992).

Admitindo que os problemas da representação dos dados pontuais do *Geo Big Data* excedem as limitações da nossa capacidade visual para exploração e análise dos dados, evidencia-se a necessidade de métodos de representação que levem em consideração o conhecimento sobre limites do nosso processamento de informação, cognição e percepção (LI *et al.*, 2016). Assim, os diferentes métodos de representação temática na Cartografia devem nos permitir visualizar padrões e diferenças espaciais, bem como derivar os significados de conjuntos de dados maciços e complexos (ROBINSON *et al.*, 2017). O conceito de suavidade e continuidade foi desenvolvido estimulado por um problema semelhante, porém para dados em menor quantidade. Assim, entende-se que esses conceitos podem iluminar o caminho à solução do problema de se representar grandes volumes de dados pontuais.

Dessa forma, entende-se que resta ainda responder “qual a competência da suavidade e continuidade em descrever a natureza dos fenômenos espaciais volumosos, para a escolha de técnicas de representação temática adequadas ao contexto do *Geo Big Data*?”

A hipótese que se levanta nesta tese baseia-se no entendimento de que a suavidade e a continuidade são propriedades que descrevem a natureza dos fenômenos espaciais, quer seja para dados espaciais em pequena quantidade ou no contexto de grandes volumes do Big Data. Assim, espera-se que a suavidade e a continuidade ainda sejam conceitos válidos para a determinação de técnicas mais adequadas à construção de mapas temáticos, com o objetivo de visualizar agrupamento de dados, mas com um certo nível de competência. Compreende-se que este nível de competência se associa ao volume, à velocidade e à variedade dos dados espaciais no contexto do Big Data.

Nesta tese, testou-se esta hipótese mediante à construção de três artigos, cujos objetivos culminam na resposta ao problema de pesquisa levantado anteriormente. No primeiro artigo gera-se medidas quantitativas para caracterizar a suavidade e a continuidade de distribuições de dados pontuais, complementarmente comparamos este método com percepções de cartógrafos experientes sobre a suavidade e continuidade. Com os resultados do primeiro artigo, construímos um segundo trabalho cujo objetivo foi construir um algoritmo para geração de dados pontuais sintéticos com suavidade e continuidade pré definidas. Este último trabalho nos permitiu criar dados sintéticos com as características de suavidade e continuidade controladas. Com base no conhecimento gerado nos dois primeiros artigos, elaboramos o terceiro estudo com o intuito de avaliar a eficiência de diferentes técnicas de representação cartográfica temática aplicadas a dados com diferentes características de suavidade e continuidade. Estes trabalhos estão apresentados nos capítulos que seguem.

## 1.1 OBJETIVOS

### 1.1.1 Objetivo Geral

Determinar a influência da suavidade e continuidade na escolha de Representações Temáticas, ranqueadas pela eficiência da visualização de agrupamentos em dados pontuais volumosos.

### 1.1.2 Objetivos Específicos

- 1) Estabelecer medidas quantitativas que descrevam as características de suavidade e continuidade de distribuições espaciais dos fenômenos;
- 2) Verificar se as medidas retratam a percepção de usuários experientes, quanto às características de suavidade e continuidade;
- 3) Desenvolver um algoritmo para geração de dados espaciais sintéticos baseado nas características de suavidade e continuidade;
- 4) Desenvolver uma interface para capturar as percepções de regiões dos usuários;

5) Gerar dados sintéticos com diferentes suavidades e continuidades e construir mapas com diferentes técnicas para cada dados sintéticos gerado;

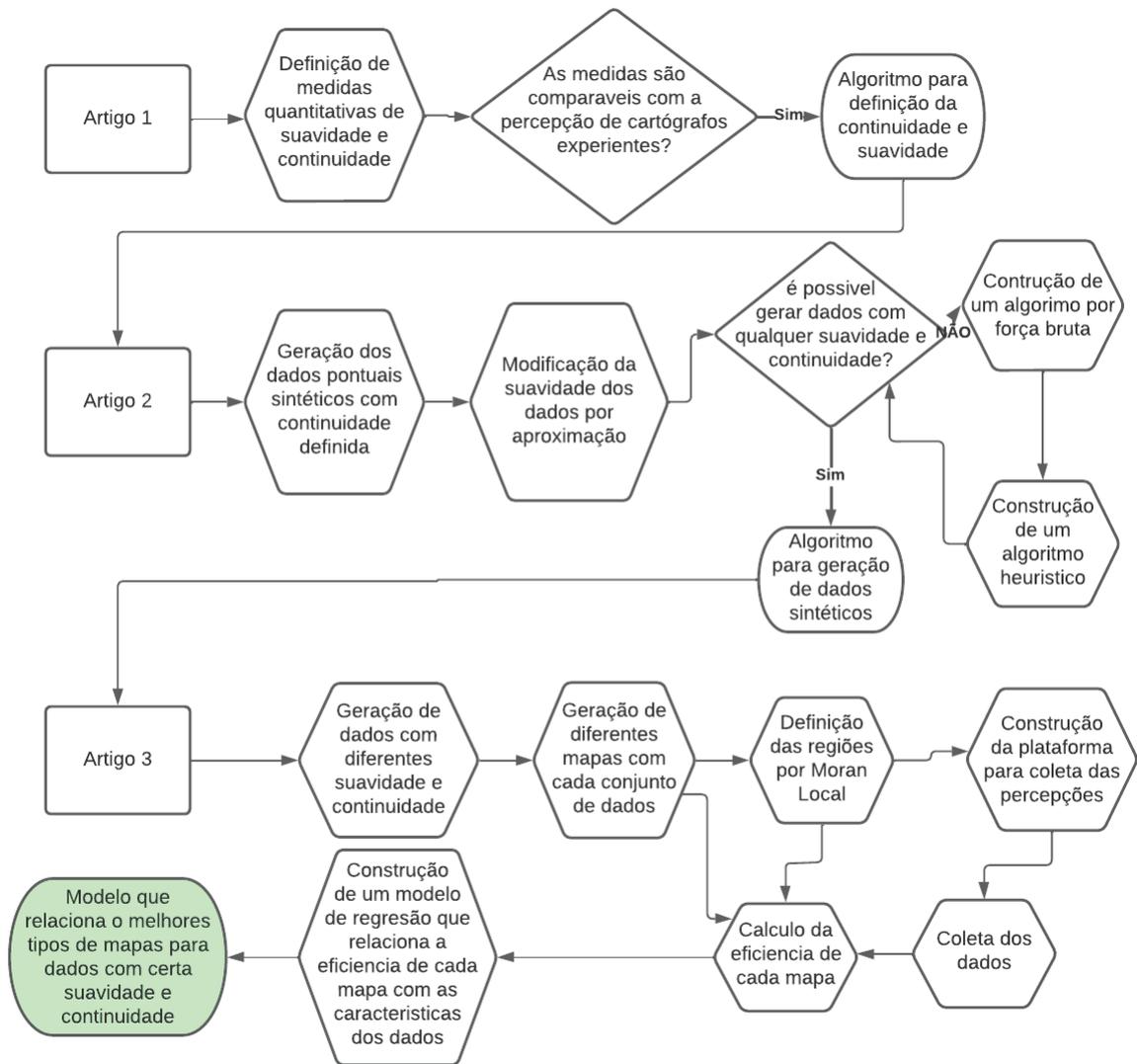
6) Avaliar a eficiência da percepção de regiões com maior ocorrência de fenômenos;

7) Propor um ranqueamento das representações cartográficas indicadas segundo o modelo, para um conjunto de dados volumoso real.

## 1.2 ESTRUTURA DA TESE

Esta tese foi desenvolvida por meio da execução de três experimentos, o fluxograma abaixo (Figura 2) descreve as etapas de cada experimento e seus relacionamentos.

FIGURA 2 – FLUXOGRAMA



FONTE: O autor (2022).

No artigo “*Computing smoothness and continuity to generate thematic maps from high-volume point data*”, apresentado no capítulo 2, desenvolvemos um método automático para quantificar a suavidade e continuidade dos dados pontuais agregados em unidades geográficas. Para quantificar essas características utilizamos uma grade hexagonal para realizar a contagem de pontos por célula, para então quantificar a continuidade pela proporção de células vazias que a grade apresenta e a suavidade pela proporção média de diferenças entre as células vizinhas da grade. Estas medidas foram implementadas no *software QGIS*, utilizando a linguagem de programação *Python 3*. Comparamos nosso método com a decisão de cartógrafos experientes, relacionando a ordem de um conjunto de distribuições geográficas, referente a suavidade e continuidade, decididas pelos cartógrafos com aquela definida

automaticamente pelo algoritmo. Os resultados indicaram concordância entre as decisões dos cartógrafos e a decisão do algoritmo.

Com a definição de medidas para quantificar estes parâmetros, surgiu a necessidade de se obter dados com as características de suavidade e continuidade com variadas características. Para experimentos que dependem destas características, além de definir a quantidade e continuidade, é necessário obter dados com estas características variadas, de forma a ter o controle dessas variáveis.

Impulsionado por essa necessidade, desenvolvemos uma aplicação para gerar dados sintéticos com suavidade e continuidade pré-determinadas, o qual é descrito no artigo intitulado de “*Synthetic high-volume point data generation based on Smoothness and Continuity*”, apresentado no capítulo 3. A aplicação foi desenvolvida utilizando a linguagem programação *Python* para ser executada no *software QGIS 3*.

A estratégia para a geração dos dados sintéticos foi utilizar uma grade hexagonal para dar suporte a geração dos dados, de forma a facilitar a definição dos relacionamentos entre regiões do espaço geográfico. Foi utilizada uma função do *Python 3*, que geram números aleatórios baseado em informações do sistema, para definir a localização das células da grade que apresentam valor igual a zero, assim definindo a continuidade. Da mesma forma, as relações entre células vizinhas são definidas aleatoriamente, porém não foi possível determinar de forma direta os valores das células para que a suavidade da grade fosse a desejada. Foi utilizada uma estratégia por aproximação para definição dos valores das células, que em cada modificação do valor de uma célula, é recalculada a suavidade e avaliado se a suavidade calculada é semelhante à suavidade desejada. Para que o algoritmo gere-se dados com qualquer valor de suavidade, no caso de suavidade muito baixa (< 35%), o algoritmo por aproximação na convergência para esse valor de continuidade, então foi necessário desenvolver dois algoritmos auxiliares: Algoritmo de força bruta e um algoritmo baseado em Heurísticas para encontrar a mínima suavidade de uma grade.

O primeiro possibilitou observar padrões na grade quando a suavidade era mínima, e por motivo de performance, este algoritmo é inviável para grades com muitas células. Com isto, foram definidas heurísticas para gerar grades com valores que apresentam mínima suavidade em tempo computacional aceitável e obtemos resultados muito próximos dos resultados obtidos pelo algoritmo por força bruta. Os conhecimentos obtidos com o desenvolvimento dos artigos supracitados

possibilitaram o desenvolvimento do último experimento, que contém a ideia central desta tese.

No capítulo 4 é apresentado o artigo intitulado “Reconhecimento Visual de Agrupamentos pelos Usuários para Validação de Ranqueamento de Representações Temáticas”. Neste experimento, busca-se entender a eficiência de diferentes técnicas de representação cartográfica temática na comunicação de regiões que têm maior densidade de incidência do fenômeno, quando aplicadas a fenômenos com diferentes características de suavidade e continuidade. Para isto, foram construídos vários conjuntos de dados pontuais com diferentes características de suavidade e continuidade, utilizando o *script* desenvolvido no experimento 2. Para cada conjunto de dados, foram construídos seis tipos de mapas, que segundo MacEachren e Dibiase (1991), são indicados para fenômenos com diferentes características de suavidade e continuidade. Para avaliar a eficiência destas representações, foi desenvolvida uma aplicação *web* para coletar as percepções de usuários. Estas percepções foram comparadas com regiões definidas pelo método geoestatístico Índice de Moran Local, desta forma foi possível avaliar a eficiência das representações em comunicar as regiões quando aplicadas a fenômenos com determinada suavidade e continuidade.

Todos os *softwares* utilizados nesta pesquisa são livres e os *softwares* aqui desenvolvidos são disponibilizados no repositório do *GitHub*: <https://github.com/LabgeolivreUFPR/Smoothness-and-Continuity-in-Geospatial-Data>, de forma a propiciar a reprodução destes experimentos.

## 2. COMPUTING SMOOTHNESS AND CONTINUITY TO GENERATE THEMATIC MAPS FROM HIGH-VOLUME POINT DATA

Raphael Gonçalves de Campos<sup>a \*</sup>,  
João Vitor M. Bravo<sup>b</sup> and  
Silvana Philippi Camboim<sup>c</sup>

<sup>a,c</sup>Department of Geomatics , Federal University of Parana, Curitiba, Brazil;

<sup>b</sup> Institute of Geography, Federal University of Uberlandia, Uberlandia, Brazil.

<sup>a\*</sup> Email - raphageoc@gmail.com, ORCID - <https://orcid.org/0000-0001-5409-2877>

<sup>b</sup> ORCID - <https://orcid.org/0000-0002-5457-3192>

<sup>c</sup> ORCID - <https://orcid.org/0000-0003-3557-5341>

## 2.1 ABSTRACT

Geo Big Data enables the cartographic representation of phenomena with high spatial detail and high temporal updating. This is a consequence of developing various technologies related to geodetic positioning, telecommunications, and smartphones. This brings the need to build tools to assist the generation of cartographic representations at almost the same speed as the data are generated. Therefore, we developed an automatic method to quantify point data's smoothness and continuity aggregated into geographic units. It is a stage of cartographic design that involves a subjective decision dependent on the cartographer's visual analysis. We compared our method with the decision of experienced cartographers, relating the order of a set of geographic distributions decided by the cartographers with that one automatically defined by the algorithm. The results indicate high agreement between cartographers' decisions and the algorithm's decision. However, we detected disagreements among cartographers, which has caused disturbances in the results.

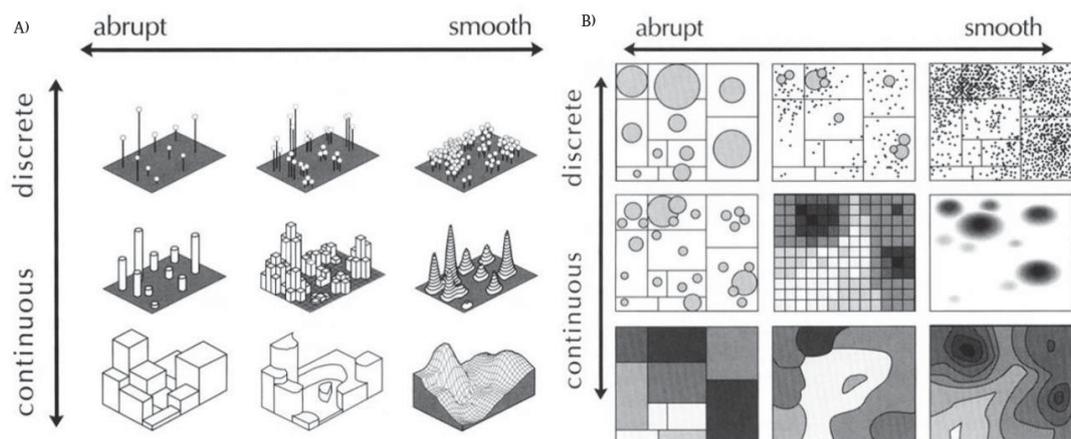
Keywords: Geo Big Data, Smoothness, Continuity, Point data, Visualisation

## 2.2 INTRODUCTION

Cartographers are now facing the emerging challenge of representing voluminous sets of geographic data within the technological context of Big Data (Griffin, Robinson, et al. 2017, Robinson et al. 2017). Therefore, here we developed a method to optimise choosing the most appropriate cartographic representation for visualising regions in large-volume point datasets, based on the automatic quantification of smoothness and continuity concepts (MacEachren and Dibiase, 1991). We understand that these concepts were first proposed for 2D-static cartographic representations made with a limited set of point data: a context far different from what we see in the current technological frame.

According to MacEachren and DiBiase (1991), Smoothness and Continuity must be considered when choosing the technique for designing a thematic map (figure 3). The Smoothness describes the dependence or independence between the phenomenon values in the spatial units, with variation along the axes smooth-abrupt. Continuity reflects the filling of geographic space by the phenomenon, which differs from the strict mathematical sense, even though the nature of the phenomenon is an essential factor in choosing the representation method. Both characteristics can vary for the same distribution depending on the scale at which they are represented, so a suitable representation method for distribution at a specific scale may not be ideal for a different scale.

FIGURE 3 – (A) MODELS OF GEOGRAPHIC PHENOMENA ARRANGED ALONG DISCRETE-CONTINUOUS AND ABRUPT- SMOOTH CONTINUA. (B) A SET OF SYMBOLIZATION METHODS APPROPRIATE FOR THESE MODELS.



SOURCE: MacEachren, 1992, Figures 8 and 9 modified.

Notably, the visualisation of geospatial data – including the Smoothness and Continuity - is a complex task once it is dependent on the cognitive skills of individuals reading maps and cartographers making the maps (MacEachren 1991, 1995). Therefore, Cartographers usually decide about the appropriate map-design decision based on the users' needs, cognitive background, the map use context, and the map-reading tasks (Ooms et al., 2015; Roth et al., 2017; Griffin et al., 2017). However, the map use research has long addressed the difficulties inherent in communicating such complex information, even because there are no 'best solutions' that could be applied to the map-design process.

Therefore, we consider that the subjectivity of the map design process is one of those variables interfering with the communication between the map-reader and the cartographer. Nowadays, such a problem is more significant than ever, once map-users also could be mapmakers, or the so-called 'producers' (Budhathoki et al., 2008; Elwood et al., 2012; Olteanu-Raimond et al., 2017). Besides, designing a map is a time-consuming process, which has subjective decisions in the course that is dependent on the cartographer's skills of interpreting the user's requirements for designing adequate solutions (Sluter et al., 2017).

Nowadays, exploring maps is a relatively simple task because technological tools allow people to navigate and explore the world (Griffin and Fabrikant, 2012). Additionally, geospatial data is widely available, which allows for detailed representations at large scales. However, when observing the Smoothness and Continuity of voluminous point data, it is hard to visually assess where on the continuum between variables a given dataset is to determine the technique to be employed, especially when a human makes the decision. That means human eyes have limits and different cognitive backgrounds, making even those experienced cartographers select different approaches for defining the boundaries associated with voluminous point data.

Again, cartographers still lack in developing graphical strategies for representing voluminous point datasets (Coetzee and Rautenbach, 2017). Moreover, the conventional techniques once created for designing thematic maps - conceived in a different technological context - are now struggling with the emergence of new, challenging, and voluminous questions posed by the Big Data era (Lee and Kang, 2015; Robinson et al., 2017; Coetzee and Rautenbach, 2017; Shi et al., 2018). Furthermore, given the nature of Big Data, such as large volume, speed in creation, and variety of types, it requires developing new methods for processing data to produce maps. Robinson et al. (2017) listed several research opportunities related to the challenges posed by Big Data in the context of Cartography. Among them is the need to combine computational methods and best practices in cartographic representation, such as computational solutions to automatically assess the phenomenon's characteristics to define the best representation method.

On that account, we argue that in the super-dimensioned context of big geospatial data, cartographers should be assisted by computational systems while

deciding how to make a map that meets user needs, especially when handling large geographic datasets.

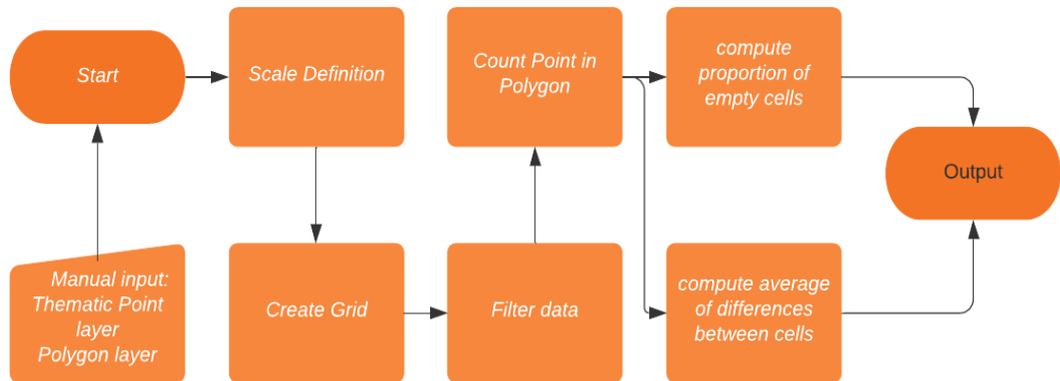
Developing solutions based on objective indicators is a convenient way to prevent subjective or wrong decisions while making a map. That means we would have reproducible methods and equal decisions based on patterns, quantities, and characteristics unreachable by the human eyes or mind. Specifically, developing automated and reproducible decisions is needed to make useful maps within the geospatial big data context. Furthermore, the algorithms could learn from experienced cartographers, observing a similar but limited set of data, and then apply to a much bigger context. However, we still do not understand if there is agreement among experienced cartographers while deciding whose best design choices are for producing a map. Additionally, we do not know if the concepts thought in the past apply to this new technological era.

For that reason, we designed a computational tool to calculate two measures to characterise the Smoothness and Continuity of the spatial distributions of point phenomena aggregated into geographic units. We also test the level of agreement of experienced cartographers and the algorithm. Finally, we aim to collect information for designing tools to facilitate – and prevent subjective choices on - the decision-making process of designing maps with those voluminous datasets.

## 2.3 METHOD

We built this paper by two main phases (Figure 4). In the first phase, we developed a tool to evaluate the Continuity and Smoothness of geographic data distributions based on the factors recommended by Machearen and DiBiase (1991). In the second phase, we created an experiment with cartographers specialising in thematic cartography to test the central hypothesis of this research.

FIGURE 4 – ALGORITHM DESIGN



SOURCE: Authors (2021).

### 2.3.1 System Design

The tool for calculating the Continuity and Smoothness was developed in Python 3.x language within QGIS 3.x. This way, by using the internal libraries of the software, the external dependencies are reduced, besides allowing the reproducibility of this research, since it is open-source software and available for several operating systems.

Figure 2 shows the main steps performed by the algorithm, which will be described in detail in the following items. The complete code can be accessed in <https://drive.google.com/file/d/1EaUfrj16PbCTtIRRqQSOLX2ommPqYEm1/view?usp=sharing>.

#### 2.3.1.1 Input data

The dataset used for this experiment was the database of requests to Curitiba's city government, called 'Base 156'. This database contains the requests generated through phone 156, the main communication channel between citizens and the City Hall of Curitiba. It includes all demands directed to the Secretariats and Bodies of the Municipal Administration and is available at the Open Data Portal of the city (<https://www.curitiba.pr.gov.br/dadosabertos/>)

Before the script execution, the user is expected to insert all the layers via the QGIS interface. These layers include the points grouped in a geographic division

and an area of interest to clip the analysis. In this work, we used a polygon representing the boundaries of the city of Curitiba to define the area of interest.

### 2.3.1.2 Scale definition

After entering the data, the first step is to define the scale of the representation. This determination is made through a function used to adjust the representation to the largest scale that allows the visualisation of the entire dataset. The result depends on the device screen and the defined size of the canvas.

Specifically, we use method `setExtent(extent)`, which sets the viewing area of the layers (`mapCanvas`) based on coordinates of the lower-left corner and upper right corner of the target area. The coordinates are obtained from the bounding box of the input database. After this definition, the scale at which the viewer has been configured is accessed by the `scale()` method. The following is a code snippet for defining the scale:

```
extent = polygon.extent()

iface.mapCanvas().setExtent(extent)

scale =iface.mapCanvas().scale()
```

After this step, the algorithm executes the spatial analysis by creating the grid and counting points on the polygons.

### 2.3.1.3 Creategrid and count point in polygon

The initial premise of this work is the calculation of the point events counts according to a grid. First, a hexagonal grid is generated based on the bounding box of the point data layer and the scale defined before. Next, the circle radius in which each hexagonal cell is circumscribed has been defined as 0.5 mm on the scale. Finally, the generated grid is processed to exclude cells not within or intersecting the area of interest.

The grid on which the data is aggregated was generated using QGIS Processing framework, using the following code:

```

parameter_dictionary={'EXTENT':extent,'HSPACING':step,'CRS':crs,'VSPACING':step, 'TYPE':type, 'OUTPUT':'memory:grade_contage'}
grid = processing.run('qgis:creategrid', parameter_dictionary)

```

Then, the grid was clipped by a polygon representing the area of interest, thus eliminating unnecessary cells that are not important in the analysis of the distribution of the phenomenon:

```

parameter_dictionary={'INPUT':grid,'OVERLAY':layer_polygon,'OUTPUT':'memory:grade_cut'}

grid=processing.run('qgis:clip',parameter_dictionary)

```

#### 2.3.1.4 Filter data

The data in 'database 156' is structured in tabular format, where each row has a field that identifies which subject that request refers to and the requester's location. The number of different topics is finite and pre-determined by the request registration system. Then, the script we created is designed to read the data and filter by subject, thus generating subsets of data with different geographical distributions. For each subset, the continuity and smoothness parameters were calculated.

Specifically, a field filtered the dataset to generate different spatial distributions. First, the unique values belonging to this field were found with the `uniqueValues()` method. Then the list with unique values is traversed, and each element generates a filter, which is applied to the database with the `setSubsetString(filter)` method. Finally, the result is saved in a file for each filtering, containing the number of points per cell and the calculated Continuity and Smoothness parameters. The following code illustrates the filtering process:

```

unique_values = layer.uniqueValues(field)

filter ='field name='+unique_values[i]+'
layer.setSubsetString(filter)

```

#### 2.3.1.5 Compute proportion of empty cells

The analysis to determine the Continuity of a spatial distribution comprises visualising an abstract surface, where gaps, tears, and overlaps indicate discontinuities (Machearn and DiBiase, 1991). To quantify the discontinuities of distribution, we create a measure based on the number of geographical units that have no occurrence of the phenomenon to be represented. In order to enable comparison between two distributions of different data, the measure divides the sum of empty geographical units by the total number of units into which the geographical space is divided.

The QGIS Processing toolbox was also used to count points in polygons. The following is the code fragment for performing the counting of points within a polygon:

```
parameter_dictionary={'POLYGONS':grid,'POINTS':layer_filtered,'FIELD':
'count','OUTPUT':archive_path}

count_points=
processing.run('qgis:countpointsinpolygon',parameter_dictionary )
```

### 2.3.1.6 Compute average of differences between cells

The Smoothness of geographical data distribution is relative to the difference of each geographical unit regarding its neighbours, i.e. the more significant the discrepancy between values, the less smooth this relationship will be. In this way, we propose a measure consisting of the average value differences between neighbouring geographical units. For this measure to be used to compare different distributions, the differences of values are divided by the maximum difference before calculating the average. First, the calculation is done excluding the geographical units with values equal to zero and which have no difference to the values of their neighbours. This step decreases the correlation between Smoothness and Continuity since values equal to zero are accounted for in the latter parameter.

Therefore, we calculated the Continuity by extracting all cells with the count attribute equal to zero, using the algorithm with the id 'qgis:extractbyattribute'. A code fragment for performing the grid cell extraction by attribute is shown below:

```
parameter_dictionary='INPUT':grid,'FIELD':'count','OPERATOR':0,'VALUE':
:0,'OUTPUT':'memory:filter0'}
```

```
grid_filtered=
processing.run('qgis:extractbyattribute',{parameter_dictionary})
```

After obtaining the layer with cells without the occurrence of the phenomenon, the `featureCount()` method was applied to retrieve the number of remaining cells. Then, it is also calculated the number of cells with zero value divided by the total number of cells.

The smoothness parameter is calculated by computing the average of the differences between neighbouring cells. For this, the algorithm computes the grid cell's centroid and generates a Delaunay triangulation. Next, it creates lines between the centroids of neighbouring cells. The following demonstrates the pieces of code to perform the described tasks.

```
parameter_dictionary={'INPUT':grid,'OUTPUT':'memory:centroid'}

centroid = processing.run('native:centroids',parameter_dictionary)

parameter_dictionary = {'INPUT':centroid,'OUTPUT':'memory:delau_cent'}

delau_cent = processing.run('qgis:delatriangulation',
parameter_dictionary)
```

After creating the rows connecting neighbouring cells, the count attributes of each cell at the start and end of the lines are determined. These values are required to calculate the difference between the values of the neighboring cells. To retrieve the points at the ends of the rows, an SQL query was performed, whose code is presented below:

```
sql= 'select id, st_startpoint(geometry) from 'lines''
parameter_dictionary={'INPUT_DATASOURCES':[lines],'INPUT_QUERY':sql,'
OUTPUT':'memory:startpoints'}
proc_start=processing.run ('qgis:executesql',parameter_dictionary)
```

The cell attributes are joined to the lines' start and end points using the algorithm id 'joinattributesbylocation', which joins layer attributes based on spatial relationships. The following example contains code for performing attribute joins by spatial location.

```

parameter_dictionary=
['INPUT':points,'JOIN':grid,'PREDICATE':[0,1,2,3,4],'JOIN_FIELDS':['co
nt'],'METHOD':1,'DISCARD_NONMATCHING':
'false','OUTPUT':'memory:startpoints_join']]

join_attributes=
processing.run('qgis:joinattributesbylocation',parameter_dictionary)

```

The start and end points have the line's attribute, so the counting attributes are carried over to the row and make it possible to make the difference between the counting at the start and end point. The following code is an example of the field calculator function.

```

parameter_dictionary=
{'INPUT':layer,'FIELD_NAME':'dif_count','FIELD_TYPE':0,'FIELD_LENGTH':
10,'FIELD_PRECISION':3,'FORMULA':'abs(\'startcount\'-
\'endcount\')','OUTPUT':'memory:difcount'}

count_norm=
processing.run('qgis:fieldcalculator',parameter_dictionary)

```

After obtaining the count difference, the values were normalized by the maximum of all differences. These values were finally averaged. Below is an example of the calculation of statistics for a layer.

```

parameter_dictionary={'INPUT_LAYER':count_norm,'FIELD_NAME':'norm_cou
nt','MEAN':'mean'}

statistics=
processing.run('qgis:basicstatisticsforfields',parameter_dictionary)

```

After finalizing the tool development stage, we designed an experiment with cartographers who are experts in Thematic Cartography. This phase is described in the following items.

## 2.4 EXPERIMENT

The experiment allowed us to verify the level of agreement between experienced cartographers and the algorithm for automatic classification of datasets

concepts of Smoothness and Continuity. In this way, both computer systems and map designers would have at their disposal an objective measure to assist in deciding which representation form is most suitable for a given distribution of large point datasets.

While assessing the level of agreement between cartographers and the algorithm, we also looked at the experts' concordance. In doing so, we aimed to understand the level of subjectivity associated with the map-reading for evaluating the Smoothness and Continuity of large point datasets. Again, with higher levels of disagreement, we could endorse the use of the algorithm we developed, as it provides a standardisation that removes the subjectivity inherent in the cognitive background of humans.

Therefore, we have established small experimental hypotheses to achieve the broader one. The hypotheses tested here are presented in Table 1.

TABLE 1 – EXPERIMENTAL HYPOTHESES

h1	Cartographers agree with the algorithm on ordering datasets based on Smoothness (from least smooth to smoothest).
h2	Cartographers agree with the algorithm ordering datasets based on Continuity (from most abrupt to most continuous).
h3	Cartographers agree with the algorithm ordering considering the concept of Smoothness.
h4	Cartographers agree with the algorithm ordering considering the concept of Continuity
h5	The time or level of education in Cartography produce natural groups of Cartographers. They agree on smoothness and continuity evaluation.
h6	The further apart the smoothness and continuity measure values, the greater the agreement between Cartographers and the algorithm.

SOURCE: Authors (2021).

Meanwhile, by testing these experimental hypotheses, we understand we could establish the benefits of using the algorithm we created for classifying datasets on the Smoothness and Continuity concepts.

### 2.4.1 Participants

Due to the nature of this research, we have only interviewed cartographers with experience in teaching or researching Cartography disciplines, especially Thematic Cartography. In doing so, we have focused on a restricted set of subjects with experience in dealing with GI Science. Because we designed the experiment for Portuguese native speakers, the set of individuals was even more restricted. Therefore, we invited individuals via e-mail lists and personal messages gathered from graduate programs focused on GI Science in Brazil. We only accepted responses from those individuals with PhD or master's degrees. We asked participants to tell us how long they teach or research this topic and how often they produce maps. We understand that, in doing so, we could stratify the individuals by years of theoretical and practical experience. We better assessed the balance between theoretical and practical knowledge by collecting this data, allowing us to group and classify cartographers by similarities in such characteristics.

Thus, twenty-four cartographers participated in this experiment. The sample of interviewees reported a minimum of five years and a maximum of thirty-three years of professional experience with Cartography, with an average of 14 years. Furthermore, the interviewees reported living in different parts of Brazil, with a particular incidence of the sample of residents in Curitiba. Also, all the interviewees are involved with or have completed post-graduate degrees in Cartography.

## 2.5 TASKS, APPARATUS, AND ANALYSIS

Before starting the main experiment, we conducted a pre-interview with a small group of cartographers. This preliminary interview has enabled us to detect variables we should observe carefully and test the test's flow. Here we have asked the participants to classify datasets based on the concepts of Smoothness and Continuity, looking at a large dataset of points aggregated into hexagonal grids (as shown in Figure 5). As they freely spoke and acted, we noticed we should present the concepts of Smoothness and Continuity as they appear in the literature, once there were disagreements among the specialists' understanding. We also observed that it should be essential to investigate the theoretical and practical backgrounds of

the specialists. These preliminary results also showed that we should investigate the perception of Smoothness and Continuity separately, as mappers had considerable difficulty assessing the two concepts simultaneously. Therefore, we designed the main experiment flow.

FIGURE 5 – WEB VIEWER FOR COMPARING TWO DIFFERENT SPATIAL DISTRIBUTIONS. THE PERSPECTIVE REPRESENTATIONS ARE FOR DATASET A (LEFT) AND B (RIGHT)



SOURCE: Authors (2021).

The participants were first invited to assign the consent form to answer the questionnaire within the main experiment. Then, sequentially, the participants accessed the identification section, where we collected information about how long they used to teach/research GI Science disciplines and how often they produced maps. After that, individuals started to answer the main section. At this point, we presented a small text and a figure exposing the concepts of Smoothness and Continuity, gathered from MacEachren and DiBiase (1991). We also explained how to access the maps we produced for the interview via the address [http://www.labgeolivres.ufpr.br/teste\\_raphael/](http://www.labgeolivres.ufpr.br/teste_raphael/). To guide them in manipulating the interactive maps, we made a video tutorial available at [https://youtu.be/FK8\\_sfV9J0c](https://youtu.be/FK8_sfV9J0c). Then, the participants accessed the interface (Figure 5) and started to execute the ordering tasks we demanded.

Participants accessed an interface containing two interactive maps. They could change which six datasets were represented in the left and right maps in the

upper bar. Thus, they could compare the six datasets two by two until they concluded the order of Smoothness and the order of Continuity. The available map (Figure 3) was created based on the point count on the grid for each phenomenon, plotted through an unclassified colour ramp, and the point count depicted on the Z-axis, forming a three-dimensional rendering. Users were free to rotate the map on the three-dimensional axes, to zoom in and out, but did not access the map legend.

The first task was to sort the distributions presented within the maps from the most DISCRETE to the most CONTINUOUS. Again, individuals should mark the column referring to the position that the dataset belongs. Then, they compared and sorted all the six datasets. The second task was equal to that first one, but sorting the distributions presented on the maps from the most ABRUPT to the most SMOOTH.

It is essential to point out that we performed the experiment online without controlling external variables. In the final analysis stage, we evaluated each distribution and established the difference between the order defined by the participant and the order defined by the algorithm. Next, we calculated the average of the residuals for each participant and for each distribution, where the average equal to zero expresses the total agreement between users and the algorithm. We then investigated the distribution of these results' distribution using frequency histograms and scatter plots.

## 2.6 RESULTS AND DISCUSSION

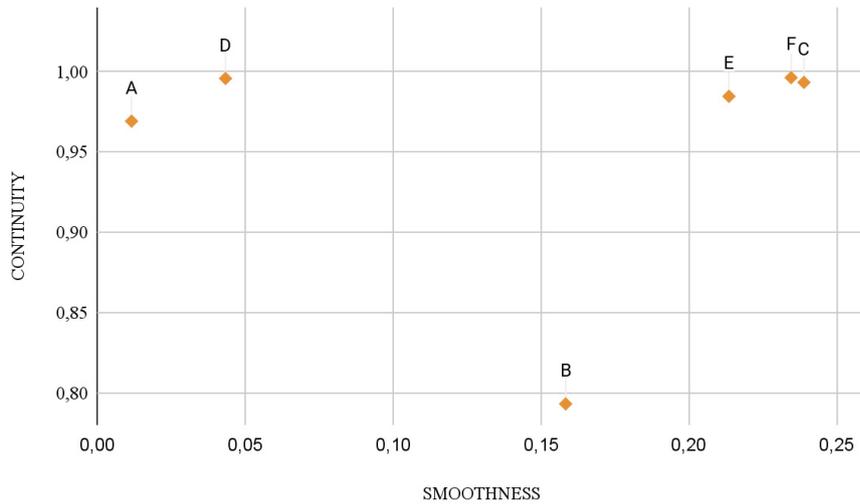
In this section, we initially report the results of the algorithm implementation and the calculated Smoothness and Continuity parameters of the generated spatial distributions. Subsequently, we detail the experiment results done with the cartographers, according to the construction of the experimental hypotheses described in the method.

### 2.6.1 Algorithm output

Figure 6 presents the dispersion graph of the datasets concerning the computed parameters. The vertical axis represents the variation in Continuity, and

the horizontal axis the variation in Smoothness, and the higher the values on the axes, the less continuous and less smooth is the spatial distribution.

FIGURE 6 – THE PARAMETERS DISPERSION CALCULATED BY THE ALGORITHM.

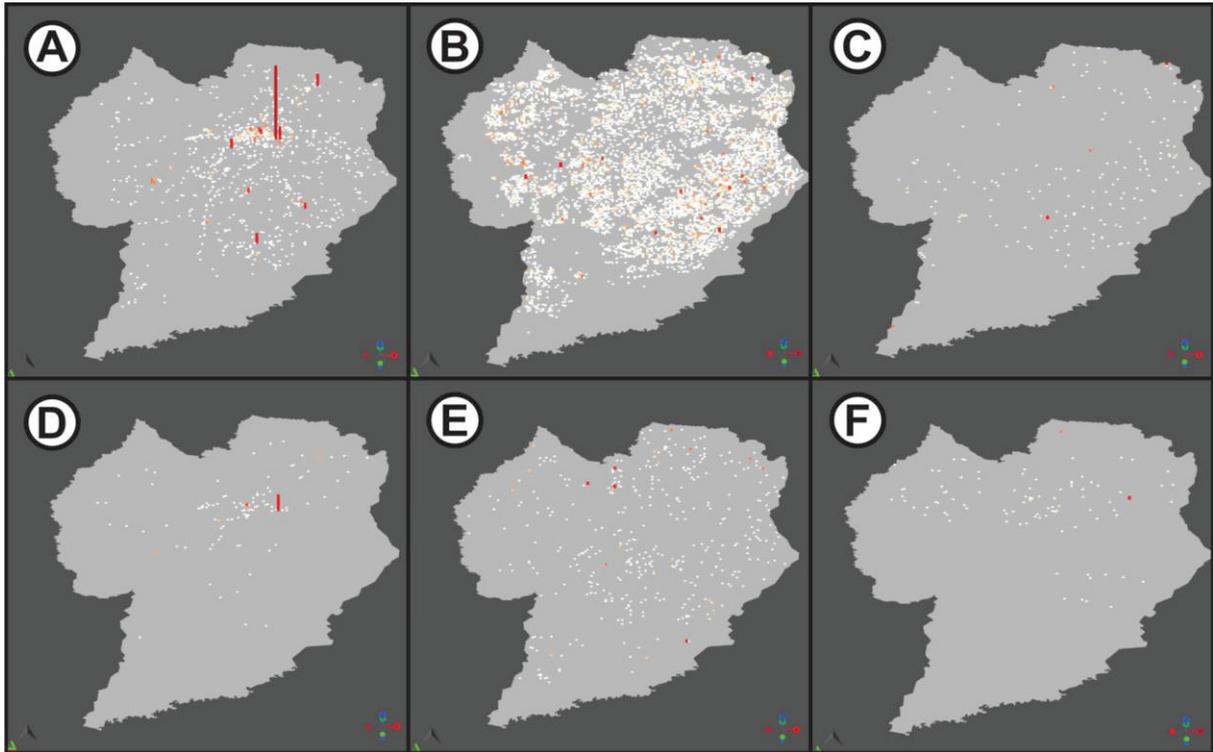


SOURCE: Authors (2021).

The parameters calculated for the datasets chosen from base 156 range in Continuity from 0.793 to 0.996 and from 0.012 to 0.239 in Smoothness. In Figure 7, it is noticeable the formation of 3 distinct groups. These clusters suggest that the spatial distributions belonging to each group have similar measures of Smoothness and Continuity. This grouping can be observed in Figure 5, in which the data are represented by a combination of visual variables: perspective height and lightness. Dataset A and D have more prominent peaks and clusters in similar locations. Dataset C, E, and F do not exhibit peaks, and the occurrences of the phenomena are well dispersed. The spatial distribution of dataset B differs because it contains no peaks and is more continuous than the others.

Given these data sets, we designed the experiment with the cartographers, and its outcomes are presented in the following topics.

FIGURE 7 – WEB MAPS PRESENTED DURING THE INTERVIEW.



SOURCE: Authors (2021).

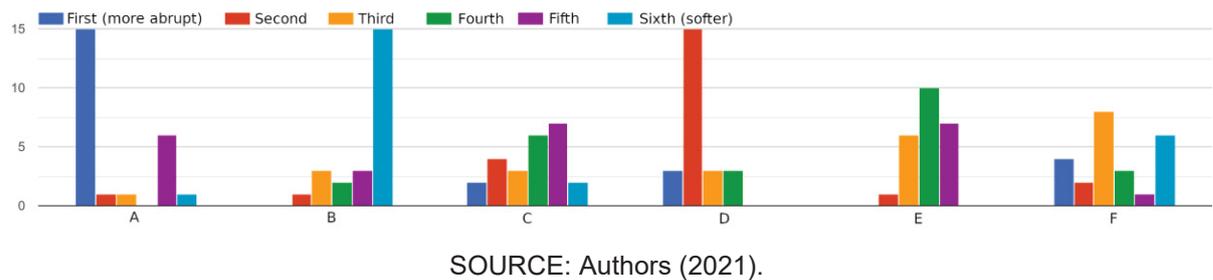
## 2.6.2 Experiment results

Here the results are described according to the order of the experimental hypotheses described in the method. A frequency histogram is used to evaluate the variation between the answers of the cartographers.

### 2.6.2.1 Hypothesis 1: Cartographers agree on the order of smoothness

Figure 8 shows the frequency histograms of the participants' responses regarding the position in the ordered set, according to the smoothness characteristic, of all spatial distributions. Again, the more concentrated the answers are in a bar of the histogram, the greater the agreement among participants. Moreover, the disagreement is characterised by the dispersion of the answers in different histogram bars.

FIGURE 8 – FREQUENCY OF EACH TYPE OF RESPONSE, BASED ON THE RANK MADE BY THE INDIVIDUALS FOR THE LEVEL OF SMOOTHNESS

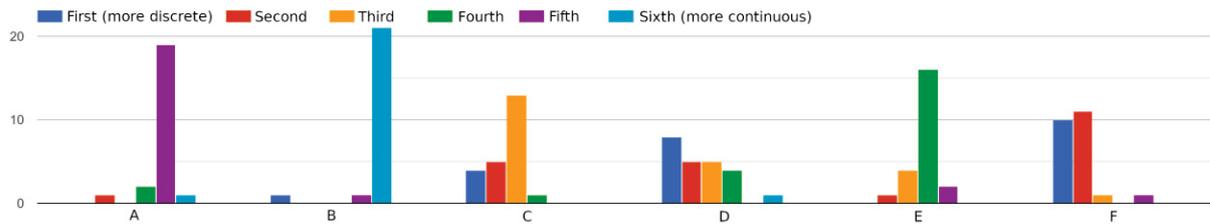


As shown in Figure 8, there is no agreement on the order of distribution C and F, with the participants' answers being distributed over all the position possibilities. For datasets A, B, and D, the responses are mainly concentrated in one position, denoting agreement among the participants. The agreement between the responses for data E is not as explicit, but the responses are concentrated in the fourth position with similar variations for the third and fifth positions. These results mean that Smoothness's perception is distinct among the cartographers we interviewed, considering the spatial variation of the phenomena represented in the situation elaborated in this research. This condition shows no explicit agreement even when there is high-level formal education in the same field of knowledge.

#### 2.6.2.2 Hypothesis 2: Cartographers agree on the order of continuity

Figure 9 shows the frequency histograms of the participants' responses regarding the position in the ordered set, according to the Continuity characteristic. The answers on Continuity's order are less dispersed than the answers on Smoothness. The worst result was in defining the position of the D dataset, as they presented answers in five of the six possible order options. The highest agreement is observed for datasets A, B, and E. There is a predominance of answers for dataset C with the third position, but it is not unanimous. There is a split of answers for data F between the first and second positions: here, we consider this condition inexpressive since the confusion occurred between neighbouring positions.

FIGURE 9 – FREQUENCY OF EACH TYPE OF RESPONSE, BASED ON THE RANK MADE BY THE INDIVIDUALS FOR THE LEVEL OF CONTINUITY.



SOURCE: Authors (2021).

In general, these results show that the concept of Continuity - considering the spatial variation of the phenomena represented in the situation elaborated in this research - is perceived in a similar, but not identical, way by the cartographers we interviewed. This characteristic denotes that, possibly, this concept may have less variation in the visual interpretation of the specialists.

### 2.6.2.3 Hypothesis 3: Cartographers agree with the algorithm on the order of smoothness

To analyse the agreement between the order of Smoothness and Continuity defined by comparing the participants' responses to the algorithm (hypotheses 3 and 4), we calculated the position divergence by a subtraction operation between the position defined by the algorithm and the position indicated by the participant. Then, the percentage of responses of each possible position divergence was determined across all spatial distributions. Tables 2 (Smoothness) and 3 (Continuity) show the percentages of answers referring to the number of positions diverging from the order defined by the algorithm. The higher the percentages in the position divergence zero, the more agreement between participants and the algorithm is. In both cases, the different intensities of blue per column indicate the variation of the percentages; the darkest indicates the highest value and the lightest the lowest value.

For the case of Table 2, the distributions A, D, and E presented the highest percentage of answers agreeing with the order defined by the algorithm, of which A and D obtained more than half of the agreeing answers. On the other hand, spatial distribution B presented the worst result, as 62.5 % of the answers presented the position divergence -3, demonstrating an agreement among the cartographers we interviewed but a disagreement between the choice of the cartographers and the algorithm. Meanwhile, distributions C and F presented divergence between the

orders defined by the cartographers and the orders defined by the algorithm compared to the orders defined by the cartographers.

Given these results, we understand that, in general, the concept of Smoothness was interpreted differently by the cartographers we interviewed and the algorithm we developed, especially when there is disagreement among the cartographers themselves.

TABLE 2 – LEVEL OF DIVERGENCE BETWEEN THE CARTOGRAPHER'S AND THE ALGORITHM'S POSITIONS, FOR SMOOTHNESS

Smoothness						
Position divergence	A (%)	B (%)	C (%)	D (%)	E (%)	F (%)
-5	4,2					
-4	25			0		
-3	0	62,5		0		
-2	4,2	12,5		8,3	0	
-1	4,2	8,3		12,5	29,2	25
0	62,5	12,5	8,3	66,7	41,7	4,2
1		4,2	29,2	12,5	25	12,5
2		0	29,2		4,2	33,3
3			12,5		0	4,2
4			16,7			20,8
5			4,2			

SOURCE: Authors (2021).

TABLE 3 – LEVEL OF DIVERGENCE BETWEEN THE CARTOGRAPHER'S AND THE ALGORITHM'S POSITIONS, FOR CONTINUITY

Continuity						
Position divergence	A (%)	B (%)	C (%)	D (%)	E (%)	F (%)
-5						0
-4				4,2		4,2
-3			0	0		0
-2			0	16,7	0	4,2
-1	4,2		4,2	20,8	8,3	45,8
0	83,3	91,7	58,3	25	70,8	45,8
1	8,3	4,2	20,8	33,3	16,7	
2	0	0	16,7		4,2	
3	4,2	0			0	
4	0	0				
5		4,2				

SOURCE: Authors (2021).

#### 2.6.2.4 Hypothesis 4: Cartographers agree with the algorithm on the order of continuity

In Table 4, we can see that the highest percentages are concentrated on zero divergences or close, so there is less discrepancy between the participants' answers and the algorithm on Continuity compared to the smoothness variable (Table 3).

The only data that the most responses did not agree with the algorithm on Continuity were datasets D and F, since the latter obtained the two highest percentages in the 0 and -1 position divergence, thus indicating less divergence. On the other hand, Dataset A, B, and E had over 70% agreement, while 58.3% in dataset C.

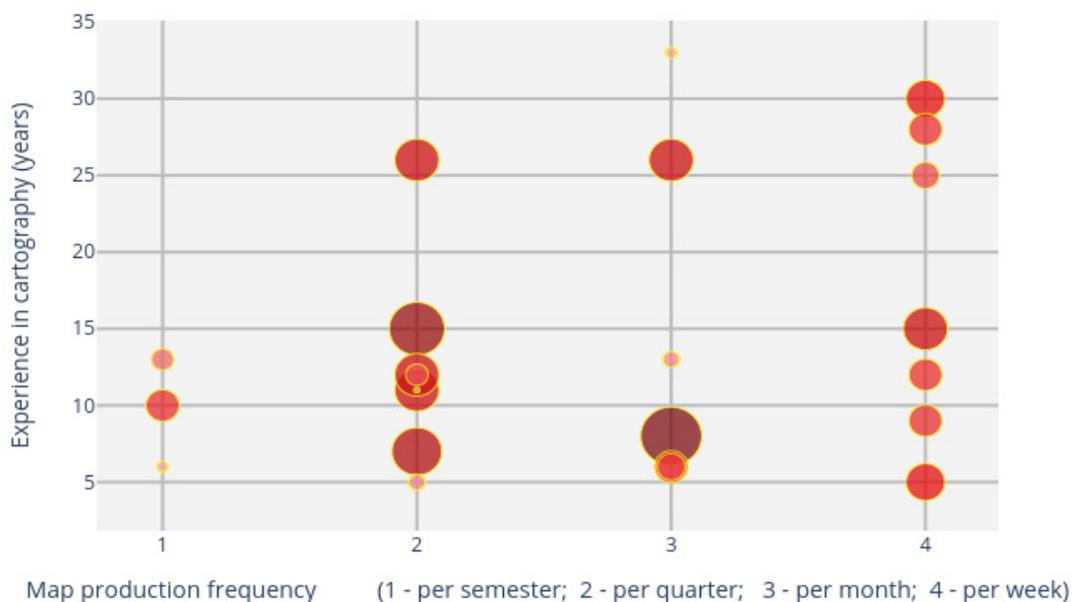
The outcomes show that the algorithm and the cartographers arrive at similar results regarding the concept of Continuity. This finding may indicate that Continuity

is a variable possibly less dependent on each respondent's varied cognitive conditions, a fact that needs to be studied further.

#### 2.6.2.5 Hypothesis 5: The time of formal education in cartography or the frequency of map production brings the reading of the concepts closer

To investigate the relationships of the variables time of formal education and frequency of map construction with the divergence between users' answers and the algorithm, a bubble chart (from Figure 10) was constructed. In Figure 10, the horizontal axis of the chart defines the reference for map building frequency, where each position refers to an alternative in the form answered by the participants. The higher the position on the horizontal axis, the more frequent is the map production. The vertical axis represents the participants' length of experience in cartography. Each circle represents one participant, and its size is the average modulus of the position divergences of Smoothness and Continuity across all datasets.

FIGURE 10 – EDUCATION IN CARTOGRAPHY, MAP PRODUCTION FREQUENCY, AND DIVERGENCE WITH THE ALGORITHM'S ORDERING



SOURCE: Authors (2021).

Thus, when observing Figure 10, it is possible to conclude that there is no relationship between cartographic experience and average position divergence because there is no tendency for the radius of the circles to decrease as the cartographers' experience increases. Regarding the frequency of map production, the participants with the lowest frequency showed the lowest maximum value. The group with the highest frequency in map production appeared to have a smaller variation of position divergence and presented the second-lowest maximum position divergence among the groups. These results show that the hypothesis raised is not valid since there is no clear relationship between formal education in cartography and agreement with the algorithm. In contrast, the people who scored highest on map production frequency also had the lowest residue discrepancies compared to each other.

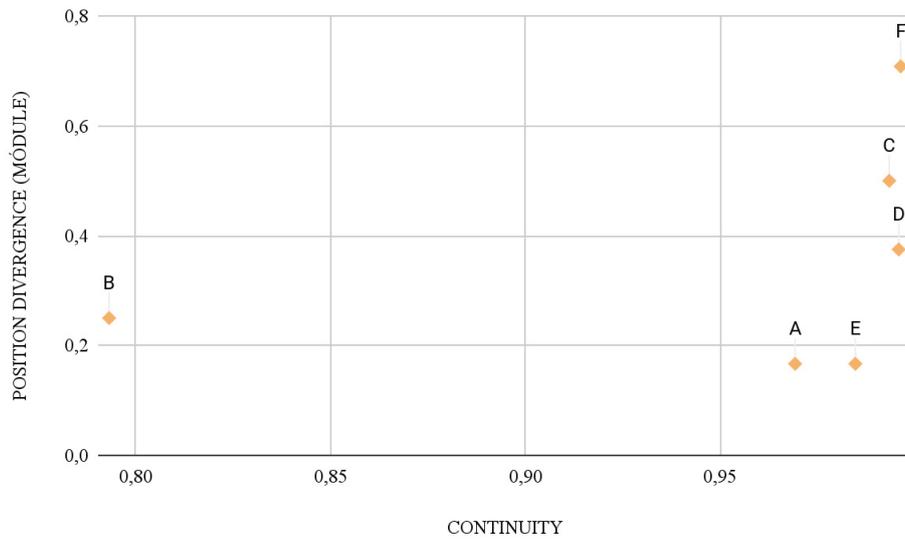
2.6.2.6 Hypothesis 6: The further apart the smoothness and continuity measure values, the greater the agreement between cartographers and the algorithm.

The scatter plots in Figures 11, and 12 show the relationship between the similarity of computed features and the modulus of positional divergence for all the datasets used.

In defining the order of Continuity by the participants, the divergence of position with the algorithm increases as the data F, C and D are close on the horizontal axis (Figure 9) that is, when the data are similar in Continuity, the mappers have trouble defining the order.

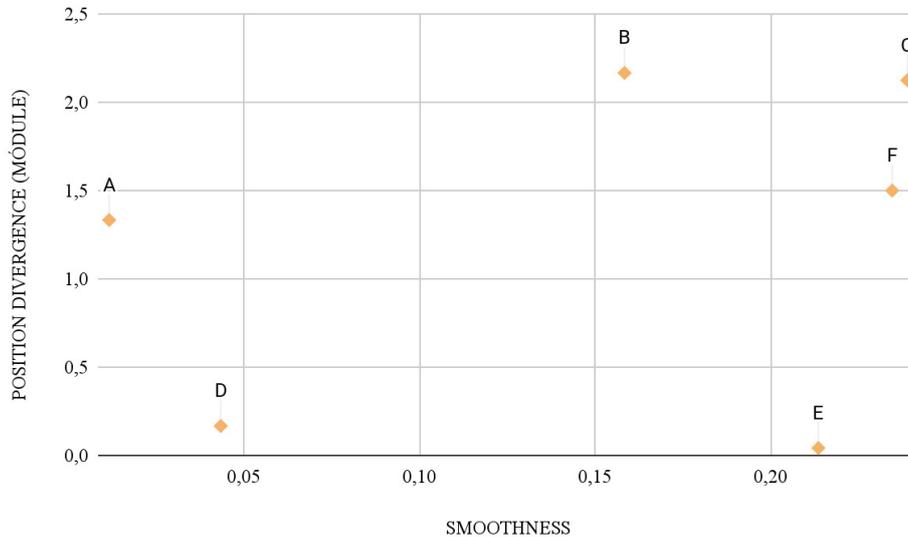
In Figures 11 and 12, we can observe a similar behaviour, with distribution C and F - the most similar in Smoothness - presenting the highest divergence. However, there is a situation that we consider a noise: distribution B has a high position divergence without having any data with similar Smoothness to it. This difference means that all the datasets that showed the most significant divergences between participants and algorithm, both in Smoothness and Continuity, have similar characteristics, except distribution B in Smoothness. However, it is possible to state that distribution B presents the most considerable number of clusters, making the visual analysis of the spatial dependence between the various cells of the grid difficult.

FIGURE 11 – RELATIONSHIP BETWEEN THE POSITION DIVERGENCE AND THE CALCULATED VALUES FOR THE CONTINUITY VARIABLE



SOURCE: Authors (2021).

FIGURE 12 – RELATIONSHIP BETWEEN THE POSITION DIVERGENCE AND THE CALCULATED VALUES FOR THE SMOOTHNESS VARIABLE.



SOURCE: Authors (2021).

## 2.7 CONCLUSIONS

Overall, the algorithm proved to be fit for defining the smoothness and continuity characteristics of spatial data distributions. The differences found seem to

be correlated to the cognitive-visual thresholds of the participants, and their experience in cartography does not appear to influence the results. That said, the algorithm can help standardise some design decisions similarly to humans and, therefore, can lead to the most effective communication.

There is a partial agreement among the experts on the order of Smoothness and Continuity of the spatial data distributions. The bias observed is the same between the experts and the algorithm. The exception was the definition of Smoothness of one distribution, which obtained agreement among the experts and disagreement between experts and algorithm. The characteristic that distinguishes this dataset from the others is the high quantity of clusters, making the expert's evaluation difficult since it is necessary to analyse the neighbourhoods of these clusters to determine the spatial dependence of the spatial distribution. Therefore, future research can be conducted in this field of knowledge.

The cause of the disagreements seems to be associated with the similarity between the characteristics of Smoothness and Continuity. So, it is detected that the intrinsic characteristics of the data impact the results obtained in this experiment. However, the time of formal education in cartography does not seem to be a variable correlated with the agreement in reading the concepts of Continuity and Smoothness. In the case of the frequency of map construction, there is stability in the divergence between participants producing maps more frequently with the algorithm. However, there is no evidence that they are lower than the disagreement in the groups with a lower map construction frequency.

Thus, we conclude that there is a need to develop solutions that reconcile or help the visual interpretation of the concepts of Smoothness and Continuity. However, as this condition is complex, considering the variation in the cognitive conditions of human beings, the solution is given by the automation proposed here in this article. It is viable and can be an attractive alternative since it is detached from human decision-making.

## 2.8 DATA AND CODES AVAILABILITY STATEMENT

The data and codes that support the findings of this study are available at <https://drive.google.com/file/d/1EaUfrj16PbCTtIRRqQSOLX2ommPqYEm1/view?usp=sharing>. This storage location is temporary to maintain anonymity.

## 2.9 DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

## 2.10 FUNDING

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

## 2.11 REFERENCES

Budhathoki, N.R., Bruce, B., and Nedovic-Budic, Z., 2008. Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal*.

Coetzee, S. and Rautenbach, V., 2017. A Design Pattern Approach to Cartography with Big Geospatial Data. *Cartographic Journal*, 54 (4), 301–312.

Dados Abertos - Consulta - Prefeitura de Curitiba [online], 2022. Available from: <https://www.curitiba.pr.gov.br/dadosabertos/> [Accessed 2 Mar 2022].

Elwood, S., Goodchild, M.F., and Sui, D.Z., 2012. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, 102 (3), 571–590.

Griffin, A.L. and Fabrikant, S.I., 2012. More maps, more users, more devices means more cartographic challenges. *Cartographic Journal*.

Griffin, A.L., Robinson, A.C., and Roth, R.E., 2017. Envisioning the future of cartographic research. *International Journal of Cartography*.

Griffin, A.L., White, T., Fish, C., Tomio, B., Huang, H., Sluter, C.R., Bravo, J.V.M., Fabrikant, S.I., Bleisch, S., Yamada, M., and Picanço, P., 2017. Designing across map use contexts: a research agenda. *International Journal of Cartography*, 3 (sup1), 90–114.

Lee, J.G. and Kang, M., 2015. Geospatial Big Data: Challenges and Opportunities. *Big Data Research*, 2 (2), 74–81.

MacEachren, A.M., 1991. The role of maps in spatial knowledge acquisition. *The Cartographic Journal*, 28 (2), 152–162.

MacEachren, A.M., 1992. Visualizing Uncertain Information. *Cartographic Perspectives*, (13), 10–19.

MacEachren, A.M., 1995. How maps work : representation, visualization, and design, 513.

MacEachren, A.M. and DiBiase, D., 1991. Animated Maps of Aggregate Data: Conceptual and Practical Problems. *Cartography and Geographic Information Systems*, 18 (4), 221–229.

Olteanu-Raimond, A.M., Hart, G., Foody, G.M., Touya, G., Kellenberger, T., and Demetriou, D., 2017. The Scale of VGI in Map Production: A Perspective on European National Mapping Agencies. *Transactions in GIS*, 21 (1), 74–90.

Ooms, K., De Maeyer, P., and Fack, V., 2015. Listen to the map user: Cognition, memory, and expertise. *Cartographic Journal*, 52 (1), 3–19.

Robinson, A.C., Demšar, U., Moore, A.B., Buckley, A., Jiang, B., Field, K., Kraak, M.J., Camboim, S.P., and 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.

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

Shi, W., Zhang, A., Zhou, X., and Zhang, M., 2018. Challenges and Prospects of Uncertainties in Spatial Big Data Analytics. *Annals of the American Association of Geographers*, 108 (6), 1513–1520.

Sluter, C.R., van Elzakker, C.P.J.M., and Ivánová, I., 2017. Requirements Elicitation for Geo-information Solutions. *Cartographic Journal*, 54 (1), 77–90.

### **3. SYNTHETIC HIGH-VOLUME POINT DATA GENERATION BASED ON SMOOTHNESS AND CONTINUITY**

Raphael Gonçalves de Campos<sup>a \*</sup>,

João Vitor M. Bravo<sup>b</sup> and

Silvana Philippi Camboim<sup>c</sup>

<sup>a,c</sup>Department of Geomatics , Federal University of Parana, Curitiba, Brazil;

<sup>b</sup> Institute of Geography, Federal University of Uberlandia, Uberlandia, Brazil.

<sup>a\*</sup> Email - raphageoc@gmail.com, ORCID - <https://orcid.org/0000-0001-5409-2877>

<sup>b</sup> ORCID - <https://orcid.org/0000-0002-5457-3192>

<sup>c</sup> ORCID - <https://orcid.org/0000-0003-3557-5341>

### 3.1 ABSTRACT

The construction of appropriate maps depends on the insertion of the user's needs in all its stages. And to understand the processes involved in map reading and exploration, it is essential to propose tests. Cartographers throughout the 20th century sought to understand the relationships between the symbology applied to maps and the human capacity to perceive them. This has culminated in guidelines incorporated into the cartographic design to guide cartographers in the map making process. However, there are other aspects related to the nature of the phenomena to be considered in cartographic design regarding the representation of voluminous point data. According to MacEachren and Dibiase (1991), choosing the cartographic techniques to represent aggregated point data must consider the characteristics of smoothness and continuity of the data. However, in the studies conducted to determine the efficiency of maps in communicating patterns, smoothness and continuity are neglected. It is understood that even though we are in the era of Big Data, it is not trivial to find data with the desired characteristics and that the characteristics of smoothness and continuity depend on the visual analysis of the cartographer. Motivated by the need to obtain data with defined smoothness and continuity, we built an algorithm to generate synthetic point data based on the pre-defined smoothness and continuity characteristics.

### 3.2 INTRODUCTION

Cartographic research developed from Arthur Robinson's functional approach (Montello, 2002). In this view, the focus of scientific development was

based on the generation of knowledge that culminated in creating maps intended for a specific purpose and audience. These fundamentals guided the way researchers reasoned their investigations, especially in Thematic Cartography and UCD (User-Centered Design) (Roth, Ross and MacEachren, 2005).

Cartographic research that aims to develop new representations within the context of UCD depends on the evaluation of users to proceed in the task of building methods that make the reading of maps efficient, considering the concepts of transferability and relevance (Griffin et al., 2017). The relationship between map representation techniques and users occurs with the proposition of tests. Tests with map users use methods applied in controlled use environments and methods that observe users in the natural environment of map use (Suchan and Brewer, 2000). The insertion of user needs as a fundamental element in all stages of developing geoinformation solutions, such as thematic maps, promotes the construction of products suitable for use (Sluter, Elzakker and Ivanova, 2016).

The set of procedures for building a thematic map, which encompasses mental and physical processes, constitutes the Map/Cartographic Design process (Slocum, 2009). One of the main steps in Cartographic Design is to recognise the users' needs and the contexts in which the thematic maps will be used. To do this, cartographers interview users so that solutions are suitable for use. In addition, Cartographers often use techniques from other fields of knowledge to understand whether the developed product is suitable for use. For example, one can apply usability engineering methods to collect information to understand users and their tasks in order to improve the cartographic design (Nivala Sarjakoski, Tapani Sarjakoski 2007; WICKENS 1992); one can verify the workload required to read a map or a set of symbols on a map (Cognitive Load); among other possibilities.

Cartographers also depend on the data parameterisation to allow the map's construction to make cartographic products suitable for use. This condition must establish a common baseline for the evaluation of variables. In the case of Geo Big Data visual analysis systems, we need these data with specific characteristics to develop geoinformation solutions that meet user needs. Although there is a high volume and speed generation of spatial data, this data is not always accessible for various reasons, such as privacy laws, copyrights, and others (Rubinfeld and Gal, 2016). There is also the problem of source variability, which makes the data parameterisation process an unsolved challenge. By integrating Thematic

Cartography's solutions to Geo Big Data representation, we find a possible way to solve part of this problem.

### 3.2.1 Part one of the solution: From the evolution of cartographic representations to Geo Big Data

Throughout the second half of the twentieth century, research with map users was related to experimental psychology, seeking to relate the size of graphic symbols and colours, among other graphic aspects, with the human ability to perceive and understand them. For example, Flannery (1971) investigated the sizing of proportional point symbols in the representation of quantitative data. Slocum (1983) developed a cluster perception prediction model for proportional point maps. Provin (1977) analysed the effect of different options for constructing count maps, such as point size, on the intended stimulus on the perception experienced.

Although the representation of punctual data has been discussed since the early days of the development of Cartography as a science, the dynamics of technological evolution and its link with the representation of spatial data in digital media made this theme remain a challenge. In this sense, the Geo Big Data has brought additional challenges to the development of cartographic visual representation since it is necessary for new methods which encompass the inherent characteristics of Geo Big Data and allow overviews of the data and the extraction of insights (Robinson et al., 2017)

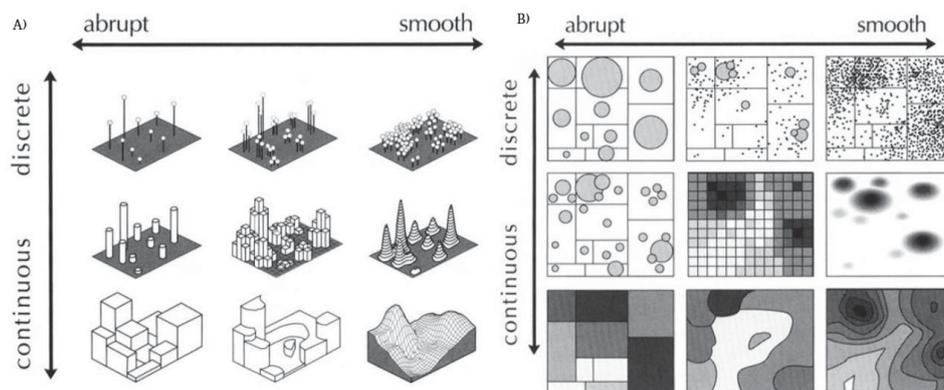
Big Data is a phenomenon of data generation caused by a set of technological developments, predominantly driven by the development of technologies for data collection, the deployment of communication infrastructure, and the development of communication applications. Such developments occasioned data generation with three characteristics: velocity - data generated all the time; Volume - large amounts of data; Variety - data with different formats, structured, semi-structured and unstructured (Laney, 2001; Robinson et al., 2017). Considering these characteristics, one of the challenges of developing research on the visual representation of geoinformation in the Big Data era relates to the complicated task of gathering voluminous databases compatible with the cartographic concepts and variables to be tested. From understanding how data characteristics influence the

choice of representation for the best understanding of users, producing such representations can be automated to fit existing data, including artificial intelligence tools. For example, in the case of point data, the concepts of continuity and smoothness (Maceachren and DiBiase, 1991) are the focus of the discussion of this article.

### 3.2.2 Continuity and Smoothness

Continuity and smoothness describe how a geographical phenomenon is distributed in space (Slocum et al., 2009). Continuity describes whether the phenomenon occurs in isolated locations or tends to happen everywhere. Maceachren and DiBiase (1991) proposed these concepts to influence a cartographic representation technique choice (Figure 13) that evokes the same visual metaphor as spatial distributions of data (Roth et al., 2019).

FIGURE 13 – (A) MODELS OF GEOGRAPHIC PHENOMENA ARRANGED ALONG DISCRETE-CONTINUOUS AND ABRUPT-SMOOTH CONTINUA. (B) A SET OF SYMBOLISATION METHODS ARE APPROPRIATE FOR THESE MODELS.



SOURCE: MACEACHREN, 1992, FIGURES 8 AND 9 MODIFIED.

Facing this paradigm of geospatial data generation, Campos, Bravo and Camboim (2022) investigated metrics for the smoothness and continuity characteristics of voluminous point data and observed the complexity in determining these characteristics, by experienced cartographers, through visual perception.

Because of the difficulty of knowing variations of voluminous data and representing them, it is natural that it is difficult to evaluate whether the cartographic products or geoinformation solutions created in the light of this technological paradigm are suitable for use. This condition depends on the human capacity to process the voluminous information. However, some works neglect the characteristics of the data in their research. For example, Słomska-Przech and Gołębiowska (2021) conducted a comparative study in which they evaluated the execution of tasks using different types of thematic cartographic representations, but the data used to generate the representations were not discriminated regarding their characteristics. Here we understand that a possible solution to this problem is the generation of synthetic data because they would allow the standardisation and control of the data set used to create the representation.

### 3.2.3 Synthetic Geospatial Data

The term synthetic data can mean both fully simulated data for application in testing, and data generated from one data or a combination of data, retaining some characteristics (Census,2022). The definition of synthetic data is variable depending on where it is applied. Synthetic data have wide applications as a method of statistical disclosure control, in which some - or all - observed data have been replaced by synthetic values generated from models based on the original data, thus attempting to maintain the anonymity of the data (General and specific utility measures for synthetic data). In artificial intelligence, synthetic data are being used to fill the need for millions of labelled data to train neural networks (Melo et. al., 2022). Within cartography, synthetic data is widely used. The construction of terrain models commonly occurs through data interpolation (Lam, 1983). We understand that this is a possible solution to the problems of Thematic Cartography because it facilitates the control of research variables related to the distribution of geospatial data.

### 3.2.4 Research Problem

In this paper, we answer "How do we generate synthetic data with the controlled-varying smoothness and continuity characteristics?" and "What aspects of

a computational tool should be developed in order to optimise the process of generating voluminous synthetic data-whose varying smoothness and continuity properties-serve the purpose of testing map users?"

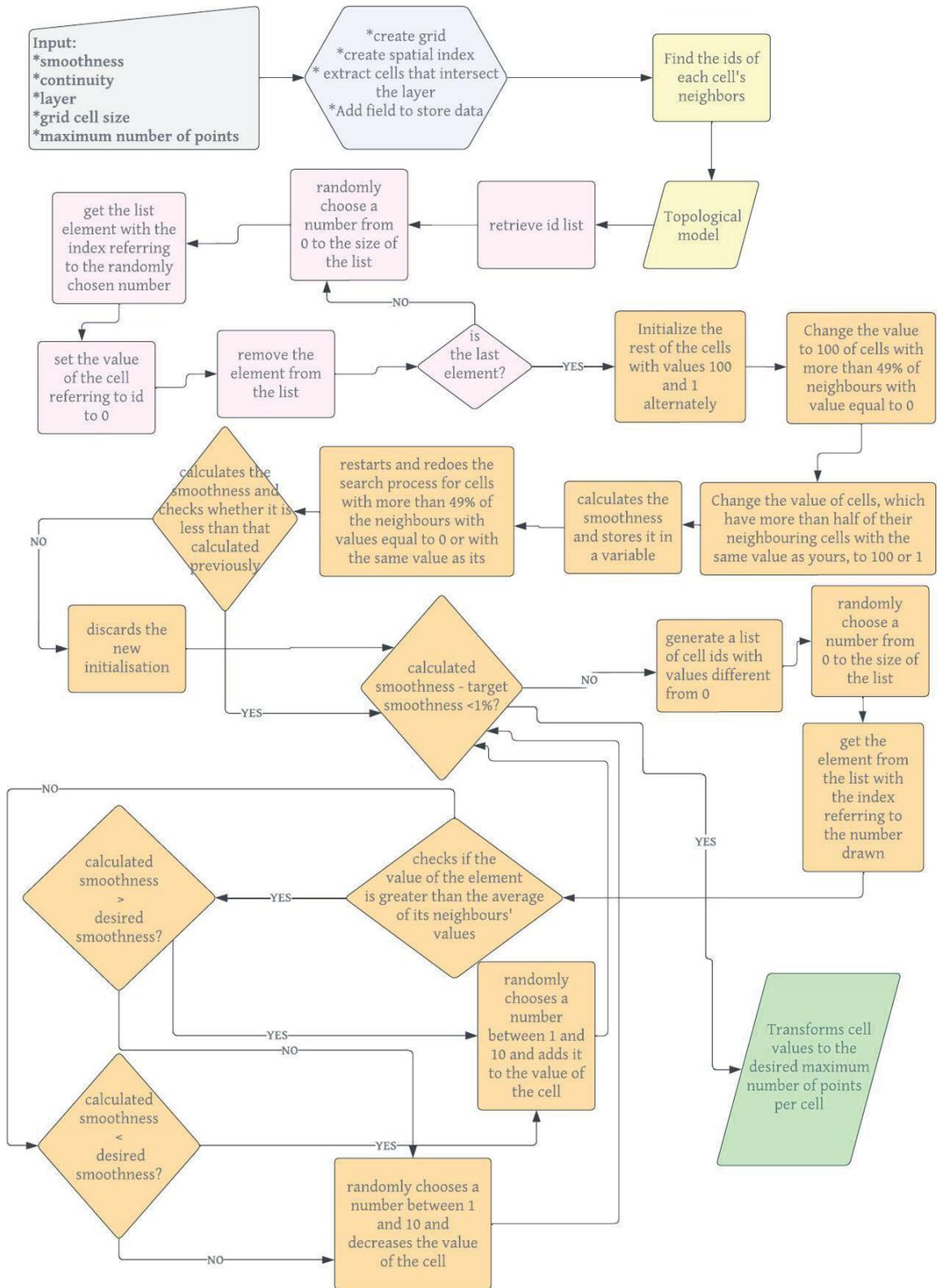
Thus, we present a computational solution that generates synthetic point-voluminous geographic data from a geographic dataset's smoothness and continuity characteristics in the following topics.

### 3.3 METHODS AND RESULTS

Synthetic data simulates real data by maintaining the similarity of some features between these data (Jaewook et. al., 2016). So, we developed an algorithm that generates geographic point data with the smoothness and continuity characteristics defined by a user. The algorithm was implemented in python to run as a QGIS script, so we used the QGIS internal libraries and its interface for data input.

In the flowchart in (Figure 14), we present the algorithm's development steps that generate synthetic point data with smoothness and continuity variation. In the following items, we describe the steps in detail the solution we created can be reproduced.

FIGURE 14 – FLOWCHART CONTAINING THE STEPS FOR GENERATING SYNTHETIC POINT DATA



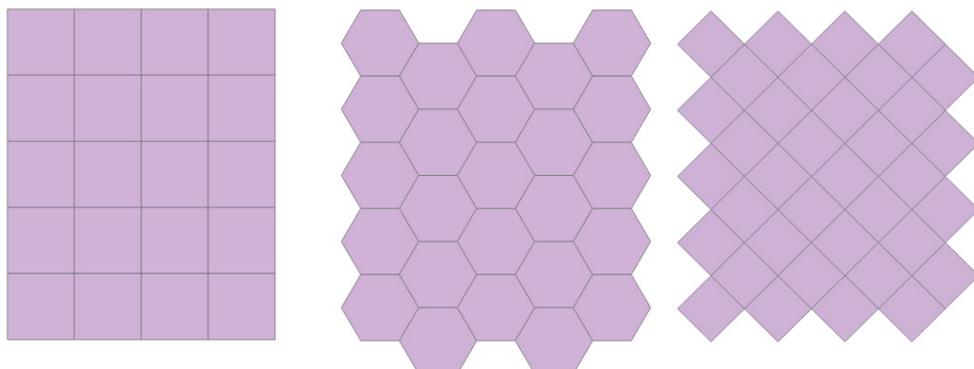
SOURCE: Authors (2022).

### 3.3.1 Starting settings

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 different 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 15). Considering that within a cell, the quantity of points to be generated is uniform and random, in the grid, linear patterns will be formed in the direction of the sides of the polygons. Then the hexagons will increase the possibilities of generating patterns to be formed in the grid and, consequently, a greater variety of spatial distributions.

FIGURE 15 – POSSIBLE GRID TYPES FORMED BY CELLS IN THE FORMAT OF REGULAR POLYGONS WITHOUT OVERLAPS AND SPACES BETWEEN CELLS.

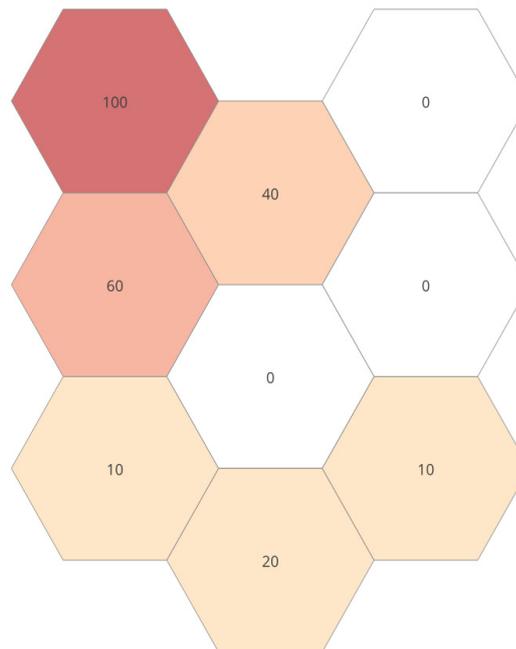


SOURCE: Authors (2022).

### 3.3.2 Quantification of smoothness and continuity parameters

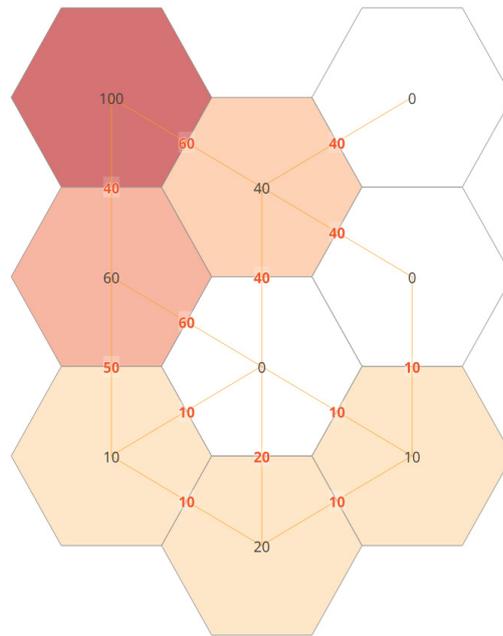
Knowing how to quantify smoothness and continuity is necessary to build an algorithm for generating synthetic spatial data with predefined values. In this work, we used the method for quantification developed by Campos, Bravo and Camboim (2022), in which the continuity is quantified by computing the ratio of cells of a grid that do not present occurrences of the phenomenon by the total number of grid cells (Figure 16). First, the difference in the phenomenon occurrence from each grid cell to each neighbouring cell is calculated for smoothness. After that, the sum of the absolute differences is calculated and divided by the number of neighbours. In the case of neighbouring cells in the grid that have no occurrence of the phenomenon, the differences are not included in the smoothness calculation (Figure 17). Thus, the correlation between smoothness and continuity measures is reduced.

FIGURE 16 – EXAMPLE OF A GRID WITH 67% CONTINUITY, THREE OF THE NINE CELLS BEING EMPTY.



SOURCE: Authors (2022).

FIGURE 17 – ILLUSTRATION OF HOW THE SMOOTHNESS OF A HEXAGONAL GRID IS CALCULATED.



SOURCE: Authors (2022).

### 3.3.3 Generation of synthetic spatial data

The algorithm was developed considering the following input information: smoothness and continuity parameters, vector layer, grid cell size to be generated and the maximum number of points per cell.

The vector layer will define the extent of the region where the grid and data will be generated. The grid cell size limits the minimum scale, at which point data can be represented with the characteristics of smoothness and continuity. This definition comes as a consequence of the modifications of these characteristics according to the representation scale, avoiding that points are created too close, generating coalescence.

After inputting the initial information, the algorithm generates a hexagonal grid within the area of the inserted layer. It then generates a topological model to facilitate operations on the grid cells, thus reducing the need for spatial queries, and our algorithm widely uses the information of the cell's neighbourhood. The steps for the generation of the topological model are described below.

- 1) Add to the grid layer a field to insert the synthetic data.

- 2) Build a spatial R-tree index to speed up spatial queries.
- 3) Obtain the bounding box of each feature
- 4) Select the features intersecting the bounding box of each cell. Then test whether they touch the geometry of the feature.
- 5) Create the topological model, composed of the id of each feature and a list of all neighbour features id.

Once the topological model has been built, synthetic data generation begins. Because continuity is independent of smoothness, the first step is to fill the grid with values that match the desired continuity.

### 3.3.4 Continuity definition

Given the required continuity percentage as input, the number of empty cells the grid will contain is calculated. The calculation is performed as follows:

$$c_v = \frac{ctnd * nt_c}{100}$$

$c_v$  – empty cells

$ctnd$  – continuity

$nt_c$  – total number of cells

Due to the possibility of the calculated value not being an integer, the value is rounded to the nearest integer. The next step is to define which grid cells will have no value. Finally, the chosen cells' indexes are randomly drawn using a Python function, which uses information from the operating system to generate random numbers. The randomisation process is defined according to the following steps:

### 3.3.5 Definition of smoothness

Empirically, it is noticeable that the process of characterising the smoothness of a grid is not as straightforward as the characterisation of continuity. Smoothness characterises the relationship between the values of neighbouring cells, so each

hexagonal cell has six neighbours except grid edge cells, and every modification of a cell's value modifies its relationship with all its neighbours. Then, an iterative algorithm was developed to characterise the smoothness of the data by an approximation process. The approximation algorithm works according to the procedures listed:

1) The process initialises with a random value between 1 and 100, except for cells that will have values equal to 0.

2) It defines randomly which cell will have its value changed, excluding cells that have a value equal to 0.

3) Calculates the smoothness.

4) Calculates the influence of each cell on the overall smoothness. This means if the cell value is smaller or larger than the average of the neighbouring cell values.

5) Defines a random factor to be applied to the cell value, ranging from 1 to the difference between the current and desired smoothness.

a) If the smoothness of the grid needs to be increased and the neighbours' average is less than the cell value (Figure 18a), then decrease the cell value by applying a negative factor.

b) If it is necessary to increase the smoothness of the grid and the average of the neighbours is greater than the cell value (Figure 18b), then increase the cell value by applying a positive factor.

c) If it is necessary to decrease the smoothness of the grid and the average of the neighbours is smaller than the cell value (Figure 18a), then one increases the cell value by applying a positive factor.

d) If it is necessary to decrease the smoothness of the grid and the average of the neighbours is greater than the cell value (Figure 18b), then decrease the cell value by applying a negative factor.

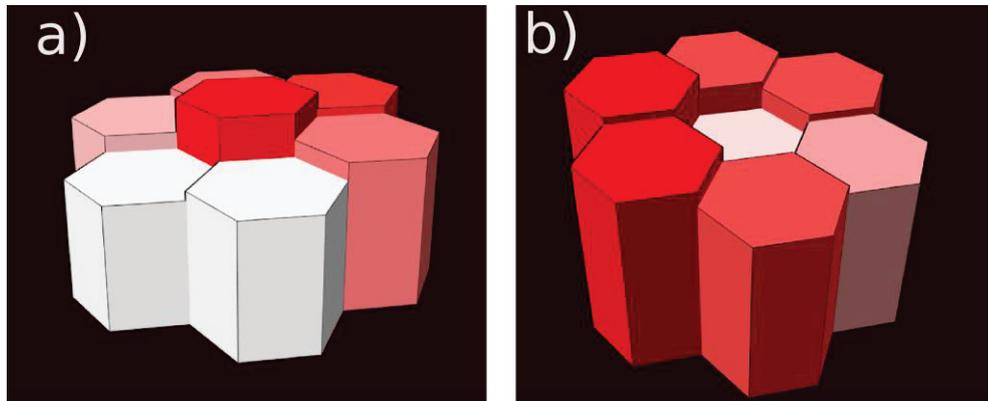
6) Repeat from step 2 until:

$$s_a - s_p \leq 1 \%$$

$s_a$  – *current grid smoothness*

$s_p$  – *desired grid smoothness*

FIGURE 18 – REPRESENTATION OF CASES WHERE THE CELL VALUE IS SMALLER THAN ITS NEIGHBOURS' VALUES (A), AND THE CELL VALUE IS LARGER THAN ITS NEIGHBOURS (B).



SOURCE: Authors (2022).

This initial algorithm proved inefficient when the required smoothness values were low. In general, when the desired smoothness values were less than 40%, the algorithm took a long time to converge or did not converge at all. So, a new algorithm was developed to fill in the grid values to achieve the smallest possible smoothness.

Hence, we created an algorithm that uses "brute force" to find values to satisfy the least smoothness condition. The script tests all combinations of values in the grid cells until it finds the lowest smoothness. Because the smoothness between two neighbouring cells is a percentage - calculated by taking the absolute difference between the values and dividing by the maximum variation in the grid - the brute force algorithm will only use values 1 and 100 to speed up the calculations.

The brute force algorithm works in the following sequence:

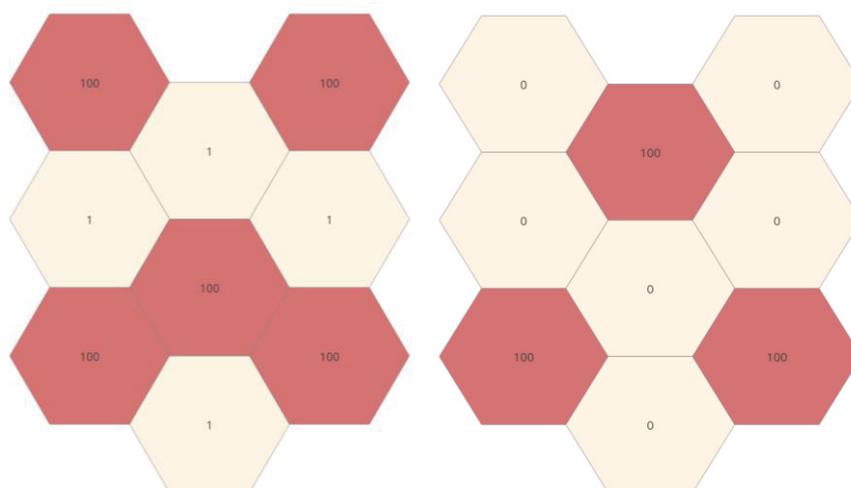
1. Perform a combinatorial analysis with all possible combinations of values 1 and 100 that the grid can assume.
2. Initialize a variable with an infinite value to store the minimum smoothness value.
3. Assignment of a combination of values to the grid.
4. Calculate the smoothness.
5. If the calculated smoothness is less than the one stored in the variable, assign the calculated value to the variable.
6. Repeat from step 3 until test all possible combinations.

This second algorithm can be rather cumbersome, which will be filled with values other than 0. This behaviour occurs because the number of combinations grows exponentially as the number of cells increases. In this case, the number of combinations ( $nc$ ) is given by:  $nc = 2^n$ , where "n" is the number of cells in the grid. In a grid of 64 cells, there will be approximately  $4 \cdot 10^9$  combinations. However, this algorithm helped us evaluate the minimum smoothness value behaviour for different small grids.

From our experiments, we can highlight that, in the first instance, the case in which continuity is 100% were tested, so all cells are filled with values other than 0. As a result, two patterns appear in the distribution of the cell values. The first is that if the value of the cell in the upper left corner is 100, then the value of the cell just below it is 1 and below it is 100, continuing in this manner. In the second pattern, if the first cell's value is 100, then the cell to the right and at the top will be 1, and the values below will follow the pattern 100, 1, 100 in succession. Figure 19 (A) illustrates the patterns described above.

The minimum smoothness depends on the number of shared connections that the grid cells have. So different grids will have different minimum smoothness. For example, grids with a smoothness of 0% have only cells with values equal to 100, which are neighbours with only cells with values equal to 0, as shown in Figure 19 (B).

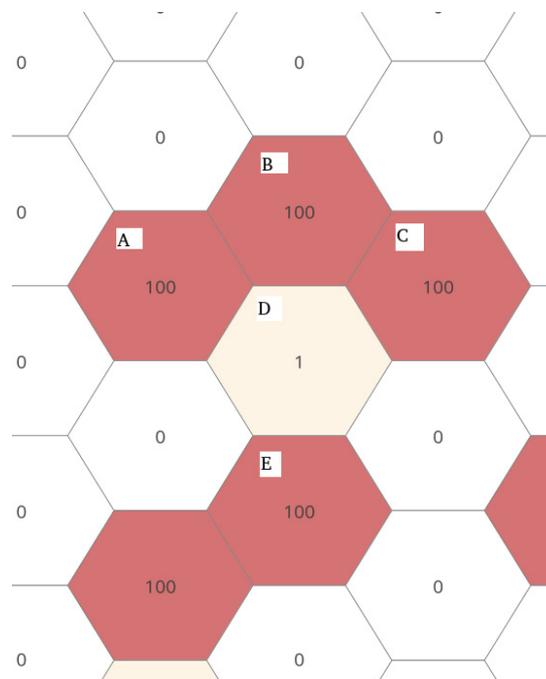
FIGURE 19 – (A) GRID WITH DEFINED VALUES FOR MINIMUM SMOOTHNESS (B) GRID WITH VALUES TO OBTAIN SMOOTHNESS EQUAL TO 0%.



SOURCE: Authors (2022).

The brute-force algorithm was also applied in the case of continuity other than 100%. However, there is no well-defined fixed pattern for finding the minimum smoothness of the grid. This is because all cells with a value bigger than zero have more than 50% of neighbours with values different from their own. For example, the cell with a value equal to 1, which has six neighbours, of them, at least three will have a value equal to 100. This fact implies that cells with more than half of their neighbours with a value equal to 0 will necessarily have a value equal to 100. Figure 20 demonstrates these patterns, where cells A, C, B and E have more than 50% of neighbours with a value equal to 0. Consequently, cell D will have a value equal to 1, thus minimising smoothness.

FIGURE 20 – CELLS WITH MORE THAN 50% OF THEIR NEIGHBOURS WITH A VALUE OF 0.



SOURCE: Authors (2022).

Based on these two patterns and on the grid filling pattern with 100% continuity - which will provide a sub-minimal solution - we created a heuristic

algorithm to find values that make the grid smoothness minimal, according to the steps listed below:

1. Initialisation with the values 1 and 100, following the pattern observed in grids with continuity equal to 100%
2. Change the value to 100 for the cells that have 50% or more of their neighbours with a value equal to 0.
3. Search for cells that less than half of their neighbours have values different from their own. The values of the neighbours and the cell are different from 0.

Datasets with different continuity values were generated to verify the results of this algorithm. Also, the brute force algorithm was applied, so the results of the two algorithms can be compared. 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 4.

TABLE 4 – COMPARISON OF THE MINIMUM SMOOTHNESS OBTAINED BY THE BRUTE-FORCE ALGORITHM AND THE HEURISTIC ALGORITHM

Continuity	Minimum smoothness (%) by brute-force algorithm	Minimal smoothness (%) heuristic algorithm	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: Authors (2022).

In Table 4, it is observed that the most discrepant value between the two algorithms was generated by the data with 80% continuity, which obtained approximately a 6% difference in smoothness between the two algorithms. Applying

a different initialisation to this grid, in which the first cell's value is changed from 1 to 100 and maintaining the fill pattern (100,1, 100...), 20.1 smoothness was achieved, which is identical to the result achieved by the brute force algorithm. Then, another procedure was added to the heuristic algorithm, which tests different initialisations to achieve the minimum smoothness. Table 2 shows the results of applying this new step to the data in Table 4. Table 5 highlights the divergence in the smoothness obtained by the two algorithms. Only the grid with 70% continuity was not achieved the same result as the brute force algorithm.

TABLE 5 – COMPARISON OF THE MINIMUM SMOOTHNESS OBTAINED BY THE BRUTE FORCE ALGORITHM AND THE HEURISTIC ALGORITHM WITH THE NEW STEP.

Continuity	Minimum smoothness (%) by brute-force algorithm	Minimal smoothness (%) heuristic algorithm	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: Authors (2022).

After populating the grid with values to achieve minimum smoothness, the algorithm is applied by approximation to achieve the desired smoothness. With this, it is possible to generate random points within the polygon. This way, a set of points with the spatial distribution with the desired smoothness and continuity characteristics is obtained. In addition, a parameter that defines the maximum value of points a polygon will contain was added to have more control over the output point data. Because, for the sake of simplicity, the algorithm's numerical variation of a phenomenon is defined from 1 to 100. Thus, to transform the default phenomenon numerical variation to the desired one and maintain the smoothness and continuity characteristics, the following equation is used:

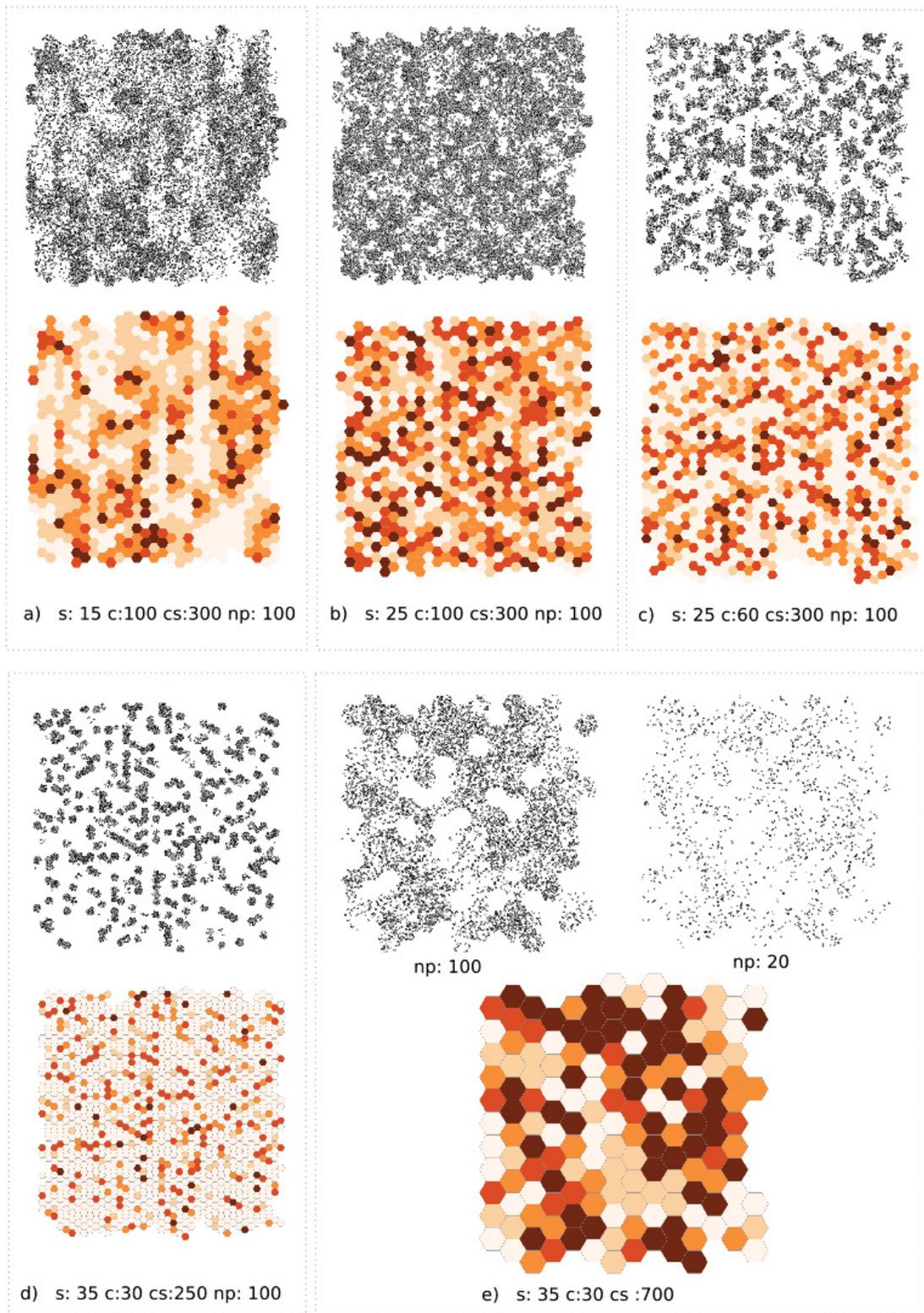
$$v = \frac{v_d * v_p}{100}$$

$v_d$  – value set by the algorithm

$v_p$  – Maximum user – defined point value

However, setting the maximum number of points changes the smoothness of the distribution. As our algorithm uses the value 100 as the maximum and 1 as a minimum to simplify calculations, there are rounded values in the process. In some tests, when changing the maximum to 20 points per cell, the distribution showed differences of 2% in smoothness. To find the new smoothness value, one can count the points generated on the grid and find the new smoothness value. Figure 21 shows the results of distributions generated by our algorithm.

FIGURA 21 – POINT DISTRIBUTIONS AND THE GRIDS GENERATED BY THE ALGORITHM.  
WHERE, S - SMOOTHNESS, C - CONTINUITY, CS - CELL SIZE AND NP - NUMBER OF POINTS  
PER POLYGON.



SOURCE: Authors (2022).

Although the algorithm is built to generate point data with predefined smoothness and continuity, it aims to provide data with a "natural" aspect. Because

of this, it is challenging to generate clusters when the continuity is low and the cell size is small (approximately 300m at scale 1:50000, for the case study), considering the total grid size, which can be seen in figures 10b and 10d. Increasing the grid cell size, as in figure 10e  $np = 100$ , can be mitigated this issue.

However, when the grid cell is large (approximately 700m at scale 1:50000, for the case study), there is a tendency for the hexagonal shape of the cells to appear in the data distribution (10e ( $np = 100$ )). In this case, one can mitigate this problem by controlling the number of points in the cell to obtain a more natural distribution. For example, figure 10e ( $np = 20$ ) shows the same spatial distribution as 10e ( $np = 100$ ) but with fewer points, which reduces the perception of hexagons in the points.

On the other hand, when aiming for very low smoothness, close to 40%, almost linear patterns will likely appear in the horizontal direction of the grid. It is a consequence of solving the problem of minimum smoothness of a hexagonal grid previously presented. Finally, as required by the appearance of the point distributions, it is necessary to test the values to achieve the desired result empirically.

### 3.4 CONCLUSIONS

In this paper, we develop a method to generate voluminous synthetic point data from varying the continuity and smoothness of spatial data distribution. Additionally, we introduce a parameter to disguise the synthetic origin of the data, thus creating a natural-looking distribution. This last condition favours the development of tests with users in a simulated realistic situation.

Our algorithm is designed to find distributions with all possible variations of smoothness and continuity. However, extreme cases of smoothness and continuity are rare, considering natural geographical phenomena. Therefore, the heuristic nature of the algorithm, relative to the search for achieving the minimum smoothness, will not significantly impact its operation.

Due to the algorithm's characteristic of generating data by regions, there is the possibility of dividing the area of interest, generating data with different characteristics for each region. This condition can favour the creation of simulated models of geographic phenomena in a way similar to the occurrence of a given real

one. Thus, research that aims to run tests with map users considering these phenomena can take advantage of this condition that makes the simulated model closer to the actual situation. For example, to obtain data similar to human settlement-related data, one can use population density data to assist in the generation of the bulky point synthetic data.

In future work, we intend to optimise the algorithm, especially in initialising the grid values, to speed up the convergence to the desired smoothness. Another improvement we intend to make is to use a function for positioning points inside the polygons to obtain even more natural-looking data.

In summary, the concept of creating synthetic geospatial data leads the direction for several applications, including a more prominent role for Artificial Intelligence in cartographic representations.

### 3.5 REFERENCES

FLANNERY, J. J. THE RELATIVE EFFECTIVENESS OF SOME COMMON GRADUATED POINT SYMBOLS IN THE PRESENTATION OF QUANTITATIVE DATA. <https://doi.org/10.3138/J647-1776-745H-3667>, v. 8, n. 2, p. 96–109, 2006. University of Toronto Press . Disponível em: <<https://www.utpjournals.press/doi/10.3138/J647-1776-745H-3667>>. Acesso em: 22/6/2022.

GRIFFIN, A. L.; WHITE, T.; FISH, C.; et al. Designing across map use contexts: a research agenda. **International Journal of Cartography**, v. 3, n. sup1, p. 90–114, 2017. Taylor & Francis. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/23729333.2017.1315988>>. Acesso em: 2/3/2022.

LAM, N. S. N. Spatial interpolation methods: A review. **American Cartographer**, v. 10, n. 2, p. 129–150, 1983.

LANEY, D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. **Application Delivery Strategies**, 2001.

MACEACHREN, A. M.; DIBIASE, D. Animated Maps of Aggregate Data: Conceptual and Practical Problems. **Cartography and Geographic Information Systems**, v. 18, n. 4, p. 221–229, 1991. Taylor & Francis Group. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1559/152304091783786790>>. Acesso em: 2/3/2022.

- MONTELLO, D. R. Cognitive map-design research in the twentieth century: Theoretical and empirical approaches. **Cartography and Geographic Information Science**, v. 29, n. 3, p. 283–304, 2002. American Congress on Surveying and Mapping.
- NIVALA, V. I.; SARJAKOSKI, A.-M. T.; SARJAKOSKI, T. T. User-Centred Design and Development of a Mobile Map Service. , p. 109–123, 2005.
- ROBINSON, A. C.; DEMŠAR, U.; MOORE, A. B.; et al. Geospatial big data and cartography: research challenges and opportunities for making maps that matter. **International Journal of Cartography**, v. 3, n. sup1, p. 32–60, 2017. Taylor and Francis Ltd.
- ROTH, R. E.; KELLY, M.; UNDERWOOD, N.; et al. Interactive & Multiscale Thematic Maps: A Preliminary Study. **AbICA**, v. 1, p. 315, 2019. Copernicus GmbH. Disponível em: <<https://ui.adsabs.harvard.edu/abs/2019AbICA...1..315R/abstract>>. Acesso em: 23/6/2022.
- ROTH, R. E.; ROSS, K. S.; MACEACHREN, A. M. User-Centered Design for Interactive Maps: A Case Study in Crime Analysis. **ISPRS International Journal of Geo-Information 2015, Vol. 4, Pages 262-301**, v. 4, n. 1, p. 262–301, 2015. Multidisciplinary Digital Publishing Institute. Disponível em: <<https://www.mdpi.com/2220-9964/4/1/262/htm>>. Acesso em: 27/6/2022.
- RUBINFELD, D., 2000. Reference guide on multiple regression. Reference manual on scientific evidence, 179, pp.425-469.
- RUBINFELD, D.; GAL, M. Access Barriers to Big Data. . Disponível em: <<https://www.researchgate.net/publication/317994296>>. Acesso em: 16/5/2022.
- SLOCUM, T. A. Predicting visual clusters on graduated circle maps. **American Cartographer**, v. 10, n. 1, p. 59–72, 1983.
- SLOCUM, T. A.; MCMASTER, R. B.; KESSLER, F. C.; HOWARD, H. H. **Thematic Cartography and Geovisualization**. Pearson Prentice Hall, 2008.
- SŁOMSKA-PRZECH, K.; PANECKI, T.; POKOJSKI, W. Heat maps: Perfect maps for quick reading? comparing usability of heat maps with different levels of generalization. **ISPRS International Journal of Geo-Information**, v. 10, n. 8, p. 562, 2021. Multidisciplinary Digital Publishing Institute. Disponível em: <<https://www.mdpi.com/2220-9964/10/8/562/htm>>. Acesso em: 10/5/2022.
- SLUTER, C. R.; VAN ELZAKKER, C. P. J. M.; IVÁNOVÁ, I. Requirements Elicitation for Geo-information Solutions. **Cartographic Journal**, v. 54, n. 1, p. 77–90, 2017. Taylor and Francis Ltd. Disponível em: <<https://repositorio.unesp.br/handle/11449/173231>>. Acesso em: 2/3/2022.

#### 4. RECONHECIMENTO VISUAL DE AGRUPAMENTOS PELOS USUÁRIOS PARA VALIDAÇÃO DE RANQUEAMENTO DE REPRESENTAÇÕES TEMÁTICAS

Raphael Gonçalves de Campos<sup>a\*</sup>,  
João Vitor M. Bravo<sup>b</sup> and  
Silvana Philippi Camboim<sup>c</sup>

<sup>a,c</sup>Department of Geomatics , Federal University of Parana, Curitiba, Brazil;

<sup>b</sup> Institute of Geography, Federal University of Uberlandia, Uberlandia, Brazil.

<sup>a\*</sup> Email - raphageoc@gmail.com, ORCID - <https://orcid.org/0000-0001-5409-2877>

<sup>b</sup> ORCID - <https://orcid.org/0000-0002-5457-3192>

<sup>c</sup> ORCID - <https://orcid.org/0000-0003-3557-5341>

##### 4.1 RESUMO

Os mapas foram empregados como ferramentas para representar fenômenos e extrair informações ao longo da história. Então, é esperado que sejam ferramentas fundamentais neste novo contexto tecnológico, em que dados são gerados em grande quantidade, variedade e velocidade. Uma tarefa a ser realizada utilizando os mapas, que foi amplamente pesquisada, é a visualização de agrupamentos. Muitos estudos existentes buscam compreender se as tarefas a serem realizadas nos mapas, incluindo o reconhecimento de agrupamento, eram executadas de forma diversa com diferentes escolhas de representação. Entretanto, há pouca compreensão numérica sobre o vínculo entre a natureza dos fenômenos e as formas mais indicadas para representação dos dados. Ao se estabelecer uma relação quantitativa, é possível buscar aplicação de ferramentas computacionais neste processo. MacEachren and Dibiase, 1991 recomendam a escolha da técnica conforme a suavidade e continuidade, em uma teoria adota amplamente. O primeiro passo desse processo é a proposição de medidas para definir a suavidade e continuidade (Campos, Bravo e Camboim, 2022 em avaliação) e algoritmos que geram conjuntos de dados com essas características definidas (Campos, Bravo e

Camboim,2022 a submeter). A terceira etapa visa suprir de evidências experimentais de que mapas construídos sob a teoria da suavidade e continuidade geram mapas mais eficientes, do ponto de vista da execução de tarefas. Posto isto, neste trabalho procuramos evidências do relacionamento entre as técnicas de representação cartográfica temáticas e os dados com diferentes características de suavidade e continuidade, no contexto da tarefa de percepção de regiões onde há maior ocorrência de fenômenos. Para isto, coletamos percepções de participantes em mapas de diferentes tipos, construídos com dados com diferentes suavidades e continuidade. O estudo foi realizado em junho de 2022 com 101 participantes. Os resultados basearam então a exploração relacionamentos entre técnicas e dados utilizando a técnica de regressão linear múltipla, o que possibilitou gerar o ranqueamento das representações, dada a suavidade e continuidade de determinado conjunto de dados.

#### 4.2 ABSTRACT

Maps have been used to represent phenomena and extract information throughout history. So, they are expected to be fundamental tools in this new technological context, in which data are generated in great quantity, variety and speed. One task to be performed using maps, which has been widely researched, is the visualisation of clusters. Many existing studies seek to understand whether the tasks performed on maps, including cluster recognition, were performed differently with different representation choices. However, there is little numerical understanding of the link between the nature of phenomena and the most suitable forms of data representation. Establishing a quantitative relationship makes it possible to seek the application of computational tools in this process. MacEachren and Dibiase, 1991 recommend the map technique choice according to smoothness and continuity in a widely adopted theory. The first step of this process is the proposition of measures to define smoothness and continuity (Campos, Bravo and Camboim,2022 in evaluation) and algorithms that generate datasets with these defined characteristics (Campos, Bravo and Camboim,2022 to be submitted). The third stage aims to supply experimental evidence that maps constructed under the smoothness and continuity theory generate more efficient maps from the point of view of task execution. In this work, we seek evidence of the relationship between thematic cartographic

representation techniques and data with different characteristics of smoothness and continuity in the context of the task of perceiving regions where there is a more significant occurrence of phenomena. For this, we collected perceptions of participants on maps of different types, constructed with data with different smoothness and continuity. The study was conducted in June 2022 with 101 participants. The results were then based on the exploration of relationships between techniques and data using the multiple linear regression techniques, which made it possible to generate the ranking of the representations, given the smoothness and continuity of a given data set.

### 4.3 INTRODUÇÃO

No atual panorama tecnológico, dados são gerados em todo lugar e a qualquer momento, em grande quantidade, alta velocidade e com grande variedade. Essa é a conjuntura do que se denominada *Big Data* (LANEY, 2001). Especificamente no campo da Cartografia, o *Geo Big Data* trata das informações espaciais as quais se enquadram no contexto conceitual do *Big Data*. Pode-se dizer que essa grande variedade, volume e velocidade tornam difícil a tarefa de se criar soluções visuais que representem a variação espacial dos fenômenos geográficos.

É notório que o mapa é uma ferramenta indispensável para extração de informações neste novo contexto, a geovisualização poderia permitir a análise do Geo Big Data, quando os computadores não são capazes de apresentar soluções diretas (LI *et al.*, 2016). O mapa é uma interface para a realidade (KOLACNY, 1969), e a interação cartográfica é o diálogo entre um humano e um mapa mediado por meio de um dispositivo computacional, que permite a transformação de ambas as partes (ROTH, 2012). Os mapas nos ajudam a compreender melhor as relações geoespaciais dos dados (KRAAK E ORMELING, 2010). Dentre as principais relações a se identificar, no contexto de análise visual, é a percepção de regiões (AMAR *et al.*, 2005).

Considerando os fenômenos pontuais, gerados em abundância no contexto do *Big Data*, as regiões podem ser definidas como áreas onde a densidade de ocorrência dos fenômenos são semelhantes. A percepção de regiões em diferentes tipos de mapas protagonizou como objeto de vários estudos na segunda metade do

século 20. Por exemplo, Slocum (1983) estudou a visualização de *clusters* em mapas de símbolos pontuais proporcionais na tentativa de entender como a percepção dos agrupamentos ocorria, considerando um grupo de pessoas entrevistadas. Assim, este autor levantou a hipótese de que se a percepção das regiões fosse diferente entre os usuários e o produtor dos mapas, poderia se modificar a representação para melhorar a comunicação. Posteriormente, Sadahiro (1997) desenvolveu um modelo de predição de percepção de regiões para representações de objetos pontuais, o qual considera três fatores: concentração, proximidade e mudança de densidade. Porém, no caso de grande quantidade de pontos a se representar, quase todos os pontos aparecerão clusterizados; o que é um problema para aqueles pesquisadores contemporâneos ao contexto do *Geo Big Data*. Mak e Coulson (1991) avaliaram o desempenho na tarefa de percepção de regiões utilizando dados classificados e não classificados, usando mapas coropléticos. Segundo o estudo destes autores, não foram encontradas diferenças significativas no desempenho dos entrevistados nesta tarefa. A pesquisa de Mak e Coulson (1991) difere dos estudos citados anteriormente, pois investiga a variabilidade de percepção entre os usuários, enquanto outros autores tentavam prever como os usuários visualizavam certa representação.

Diferentemente das pesquisas anteriores, Lewandowski *et al.* (1993) avaliaram a variabilidade de percepção de regiões entre técnicas. Neste caso foram avaliados 4 tipos de representações: mapas coropléticos com escala de cores sequencial e divergente, mapas de pontos de contagem, mapa corocromáticos e mapas de pizza. Os mapas coropléticos com esquemas de cores sequenciais foram os que geraram menos variação na percepção de clusters entre os usuários. Também foi verificado que ao mudar o tamanho da unidade geográfica na representação de pontos de contagem havia uma variação significativa na quantidade de *clusters* percebidos. Quanto maior a área das unidades de enumeração ocorria mais divergências da percepção de clusters entre os usuários. Słomska-Przech e Gołębiowska (2021) reiteram o resultado obtidos por Lewandowsky *et al.* (1993), ao descobrir que os mapas coropléticos apresentaram melhores resultados quando foram comparados a outras técnicas de representação temática, quanto a eficiência na realização de tarefas de leituras de mapas.

Entretanto, os trabalhos relatados nesta problematização não consideram as características da distribuição do fenômeno na avaliação das diferentes técnicas de

representação cartográfica temática, conforme é prescrito por MacEachren e DiBiase (1991). Neste caso, estes autores sugerem que a escolha do tipo de mapa a ser construído deve se basear nas características de suavidade e continuidade da distribuição espacial de um fenômeno. Essas características descrevem como o fenômeno varia através do espaço (suavidade) e o preenchimento do espaço geográfico pelo fenômeno (continuidade). E é neste raciocínio e nestes conceitos que acreditamos residir a solução para o caso de dados pontuais volumosos, conforme ocorre nos fenômenos mapeados no contexto do *Big Data*.

Nesse sentido, a suavidade e a continuidade podem variar conforme a escala de representação do fenômeno. Campos, Camboim e Bravo (2022) definiram medidas para quantificar essas características, além de avaliarem a concordância entre as medidas quantitativas e a percepção das características por especialistas em cartografia. Estes autores observaram que, de modo geral, há concordância entre o resultado obtido com o algoritmo e as respostas dos cartógrafos(as) quando inquiridos sobre a ordem de suavidade e continuidade de um conjunto de fenômenos. Entretanto, existem situações nas quais até experientes cartógrafos(as) não conseguem perceber as características de continuidade e suavidade. Como a pesquisa de Campos, Bravo e Camboim (2022) foi conduzida no contexto de dados pontuais volumosos, é necessário avançar no conhecimento das características de suavidade e continuidade quanto às suas influências no processo de construção de mapas temáticos.

Assim sendo, devido ao novo contexto tecnológico, em que a velocidade de geração dos dados requer a produção de mapas quase na mesma velocidade, a geração de medidas para quantificar a suavidade e continuidade podem ser valiosas ferramentas na construção do projeto cartográfico. Em contrapartida, não há indícios empíricos de que os mapas construídos se submetendo aos critérios da teoria formulada por MacEachren e DiBiase (1991), sejam mais eficientes, uma vez que foram criados à luz de um contexto tecnológico muito diferente daquele do *Geo Big Data*.

Então, neste artigo investiga-se qual a capacidade dos usuários de mapas em reconhecer visualmente os padrões agrupamentos de fenômenos espaciais volumosos, considerando-se representações temáticas com diferentes configurações de valores para suavidade e continuidade. Em posse deste conhecimento, conseguimos ranquear a eficiência dessas representações variadas,

no que tange o estímulo do processo de leitura do fenômeno espacial, adequadamente.

#### 4.4 MATERIAIS E MÉTODOS

O objetivo principal deste trabalho é baseado em testar se usuários conseguem perceber padrões em mapas temáticos. Nesse sentido, este capítulo será decomposto em três principais etapas. A primeira etapa consiste na descrição dos materiais a serem utilizados para a construção dos mapas e da plataforma. Posteriormente, é descrito o experimento, e, por fim, são detalhadas as decisões tomadas na construção dos mapas a serem testados, a implementação da plataforma para aplicação dos testes e as análises dos dados.

##### 4.3.1 Materiais

Os seguintes dados, *softwares* e linguagens de desenvolvimento de aplicações web são utilizados neste trabalho:

- Dados: Base de setores censitários do Paraná e São Paulo (2022);
- *Software* e bibliotecas: *QGIS* 3.22, *Proj4*, *Plugin QGIS* Gerador de dados espaciais sintéticos (Campos, Camboim e Bravo, 2022), *Geoda* 1.16, *Python* 3.8, *PostGIS* 3.0, *PostgreSQL* 13, *Geoserver* 2.20, *Nginx* 1.21, *Docker* 20.10, *OpenLayers* 6.14, *Bootstrap* 5, *Django* 3.2;
- Linguagens de desenvolvimento de aplicações *web*: *Python*; *Javascript*; *HTML*; *CSS*; *SQL*.

##### 4.3.2 Experimento

Para verificar a dependência entre a natureza do fenômeno e as técnicas de representações temáticas, no contexto da percepção de regiões, entende-se a necessidade de avaliação de fenômenos com diferentes suavidades e continuidades. O modelo proposto por Maceachren e Dibiase (1991) orienta utilizar técnicas de representações conforme as características de suavidade e continuidade apresentadas na Tabela 6.

TABELA 6 – TÉCNICAS DE REPRESENTAÇÃO CARTOGRÁFICAS TEMÁTICAS ADEQUADAS PARA A SUAVIDADE E CONTINUIDADE

	ABRUPTO		SUAVE
DISCRETO	símbolos pontuais proporcionais	combinação entre símbolos pontuais proporcionais e pontos de contagem	pontos de contagem
	múltiplos símbolos pontuais proporcionais por unidade	“Chorodot” - Técnica coroplética aplicada a grades	mapa de calor
CONTINUO	coroplético	dasimétrico	isoplético

FONTE: O autor (2022).

#### 4.3.2.1 Definição dos Mapas

Posto que o artigo original propõe que a escolha da técnica de representação de acordo com modelos de distribuições e exemplos de fenômenos equivalentes, a definição das características de um fenômeno depende da habilidade visual e do conhecimento do cartógrafo. Então, é um processo experimental e sujeito às interpretações do observador. Porém, Campos, Bravo e Camboim (2022) desenvolveram uma aplicação baseada em QGIS para gerar dados pontuais sintéticos com suavidade e continuidade definidas. Embora não se conheça os limiares desses parâmetros para definir a qual técnica cada distribuição está associada, vamos utilizar essa aplicação para geração dos dados para construir os mapas, para então ordenar os resultados de acordo com a percepção dos usuários.

#### 4.3.2.2 Definição da Base Cartográfica

O algoritmo gerador de dados sintéticos requer os seguintes dados obrigatórios de entrada: uma camada contendo as delimitações da área de estudo, tamanho da célula da grade em que os dados serão gerados e os parâmetros de

suavidade e continuidade. Então, a primeira decisão a se tomar é a camada de divisão do território. Para esta escolha utilizou-se as seguintes premissas:

- Ser uma unidade geográfica real, trazendo realismo para o experimento;
- As unidades geográficas devem ter tamanhos semelhantes, de forma a representarr as características dos dados pontuais;
- Utilizar a divisão do território em diversas regiões, de forma a poder testar vários mapas com as mesmas características de suavidade e continuidade sem que ocorra aprendizado entre os mapas pelo participante.

Foi explorada a base de setores censitários fornecida pelo IBGE para encontrar recortes com as características supracitadas. O experimento foi elaborado para ser realizado em computadores *desktop* ou *notebook*, que possuem telas com no mínimo de 11,6”, que são as menores telas disponíveis em lojas especializadas. Então o recorte de setores deve ser representado em uma escala em que possa ser visualizado todo o recorte e os setores sejam visíveis. O recorte modelo foi extraído da base de setores censitários da cidade de Curitiba por análise visual, posteriormente foi calculado a área dos setores e calculado o desvio padrão das áreas, para verificar a conformidade. Em seguida foram extraídas as caixas delimitadoras dos setores contidos no recorte e calculado as médias dos comprimentos dos lados das caixas. A escala foi definida de forma a representar o lado com menor média - que tinha o comprimento de 160 m - em 4 mm. Assim, definindo a escala em 1:40.000 e o recorte de setores foi novamente recortado para ser inteiramente representado em uma tela de 11”.

Deste novo recorte de setores foi extraída a caixa delimitadora, com a qual foi utilizada para navegar sobre a camada de setores e retirar nove recortes. Estes recortes foram analisados conforme o recorte modelo quanto a semelhança das áreas dos seus componentes. Então, foram escolhidos mais dois recortes, além do recorte modelo, os quais estão representados na Figura 22.

FIGURA 22 – RECORTES DE SETORES UTILIZADOS PARA A GERAÇÃO DOS MAPAS



FONTE: O autor (2022).

A escolha da projeção cartográfica depende do propósito do mapa, na maioria dos casos da cartografia temática é considerável aplicar uma projeção equivalente (SLOCUM *et al.*, 2009). Robinson *et al.* (1995) indicam as projeções equivalentes para casos de comparações geográficas em um mapa, o que é a base da tarefa do nosso experimento. Então, optamos por utilizar a projeção equivalente de *Albers* para o Brasil, que tem os seguintes parâmetros no formato *proj4*:

```
+proj=aea +lat_0=-12 +lon_0=-54 +lat_1=-2 +lat_2=-22 +x_0=0 +y_0=0 +ellps=GRS80
+units=m +no_defs.
```

A próxima etapa consiste na definição do tamanho da célula da grade, isto deve-se pela característica da aplicação que gera os dados pontuais dentro de uma grade hexagonal. Nesse sentido, procuramos calcular o tamanho da célula baseado no tamanho dos setores. Para isto, foi gerado a caixa delimitadora dos setores, extraímos o comprimento dos lados das caixas e calculamos as estatísticas para a largura e altura das caixas. Para evitar que setores maiores influenciassem no resultado, utilizamos a menor mediana das larguras e alturas, que resultou em um tamanho de célula de 160 metros.

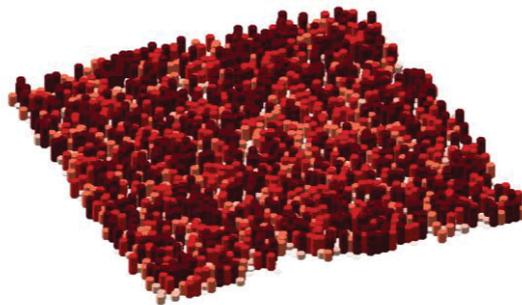
#### 4.3.2.3 Gerando os Dados Sintéticos

As últimas informações necessárias para gerar os dados espaciais sintéticos são as características de suavidade e continuidade. Embora seja desconhecido o limiar dos parâmetros para definir se um fenômeno é abrupto ou suave, ou ainda, se é contínuo ou descontínuo, podemos estipular valores progressivos de suavidade e

continuidade, com o intuito de avaliar os diferentes espectros das escalas de suavidade e continuidade.

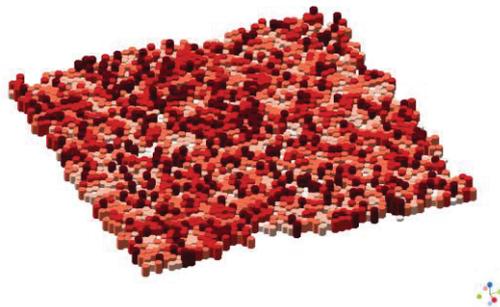
Ao construir mapas com diferentes suavidade e continuidades, foi possível observar que os limiares se aproximam dos valores de suavidade de 85%, 70% e 55%, e isto poderá ser explorado mais detalhadamente a partir dos resultados. Para a continuidade foi percebido limiares entre 30%, 60% e 100%. A determinação dos limiares não entra no escopo deste artigo, mas apenas a ordenação ou ranqueamento das representações mais eficientes versus os parâmetros de cada distribuição testada. As Figuras 23, 24 e 25 representam distribuições com 100% de continuidade e 55%, 85% e 70% de suavidade, respectivamente.

FIGURA 23 – DISTRIBUIÇÃO COM 100% DE CONTINUIDADE E 55% DE SUAVIDADE



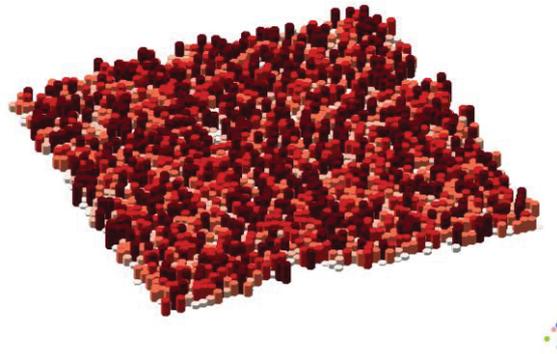
FONTE: O autor (2022).

FIGURA 24 – DISTRIBUIÇÃO COM 100% DE CONTINUIDADE E 85% DE SUAVIDADE



FONTE: O autor (2022).

FIGURA 25 – DISTRIBUIÇÃO COM 100% DE CONTINUIDADE E 70% DE SUAVIDADE



FONTE: O autor (2022).

Definidos os recortes de setores censitários e as suavidades e continuidades, determinamos um conjunto de nove dados sintéticos para serem empregados no teste (Tabela 7):

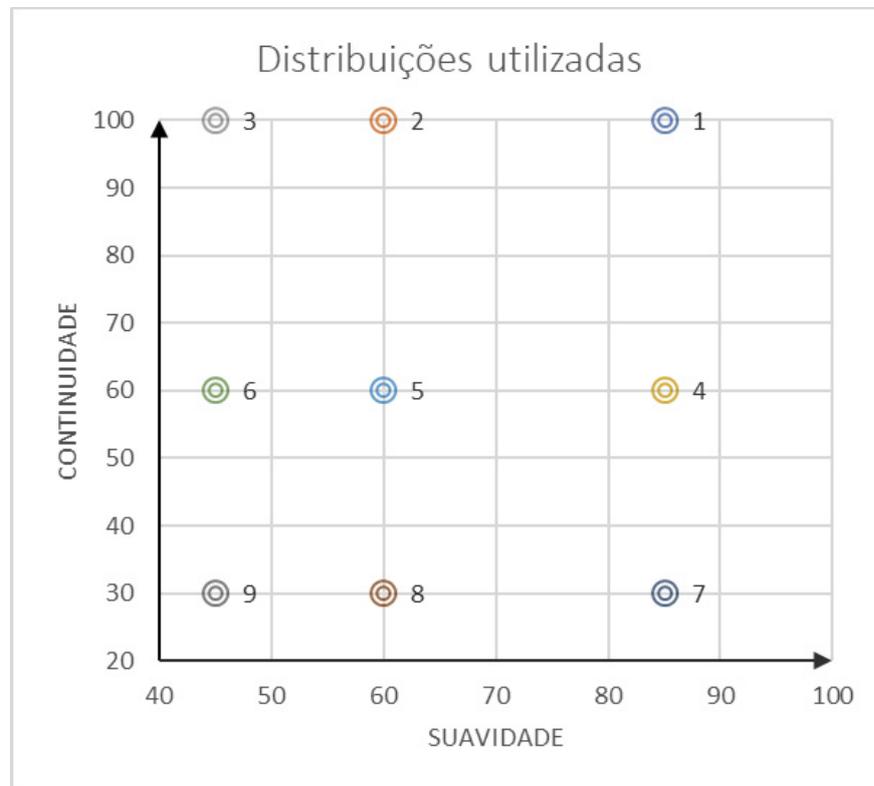
TABELA 7 – CARACTERÍSTICAS DOS CONJUNTOS DE DADOS SÍNTEÉTICOS UTILIZADOS NO TESTE.

Conjunto de Dados	Suavidade	Continuidade
1	85	100
2	60	100
3	45	100
4	85	60
5	60	60
6	45	60
7	85	30
8	60	30
9	45	30

FONTE: O autor (2022).

A distribuição dos dados no espaço bidimension da suavidade-continuidade são representadas na Figura 26.

FIGURA 26 – DISTRIBUIÇÃO DAS CARACTERÍSTICAS DOS DADOS UTILIZADOS



FONTE: O autor (2022).

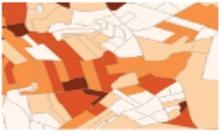
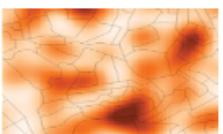
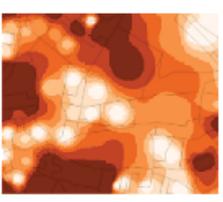
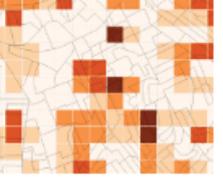
#### 4.3.2.4 Definição das Técnicas de Representação Temática

A combinação de variáveis visuais, primitivas gráficas, métodos de interpolação e formas de agregar dados nos permite a geração de numerosos tipos de mapas. Pillay *et al.* (2019) apresentam uma versão de um modelo estendido de técnicas adequadas ao tipo de fenômeno, com a introdução de outras técnicas de representação temática. No entanto, para este experimento considera-se que podemos identificar a correlação entre técnicas e características dos fenômenos utilizando apenas seis técnicas: mapas coroplético, mapas de símbolo pontuais proporcionais, mapa isoplético, mapa de pontos de contagem, mapas de calor e mapas “chorodot”.

Para cada técnica de representação foi definido os parâmetros para a geração dos mapas, porém todos os mapas de uma técnica terão a mesma configuração. Na Figura 27 são descritas as decisões inerentes a cada técnica

quanto às recomendações encontradas na literatura e decisões para uma melhor representação, a partir de análises visuais.

FIGURA 27 – ESPECIFICAÇÕES DE CONSTRUÇÃO DE CADA TIPO DE MAPA

Tipo de Representação	Parâmetro	Valor	Ilustração de Exemplo (fora de escala)
<b>Mapa coroplético:</b>	Número de classes:	5	
	Unidades de enumeração:	Setores Censitários	
	Técnica de classificação:	Natural Breaks (Jenks) - QGIS3	
	Rampa de cores:	Oranges	
	Normalização:	Área	
	Espessura do traço das unidades de enumeração:	0.1 mm	
	Cor do traço das unidades de enumeração:	#4f4f4f	
<b>Mapa de calor</b>	Raio:	200 metros	
	Forma do kernel:	Quártico	
	Tamanho do pixel:	10 metros	
	Base Cartográfica:	Setores Censitários	
	Cor do traço dos polígonos da base:	#232323	
	Espessura do traço dos polígonos da base:	0.1 mm	
	Opacidade dos polígonos da base	30%	
Modo de fusão de cores da base cartográfica e tema	Multiplicar layer (QGIS)		
<b>Mapa isoplético</b>	Geometria interpolada:	Centróide dos setores censitários	
	Atributo interpolado:	valor normalizado pela área	
	Método de interpolação:	IDW com potência 3	
	Classificação dos dados:	Quartil	
	Número de classes:	7	
	Tamanho do pixel:	10 metros	
	Base Cartográfica:	Setores Censitários	
	Cor do traço dos polígonos da base:	#232323	
Espessura do traço dos polígonos da base:	0.1 mm		
Opacidade dos polígonos da base	30%		
Modo de fusão de cores da base cartográfica e tema	Multiplicar layer (QGIS)		
<b>Mapa de símbolos pontuais proporcionais</b>	Dados:	valor normalizado pela área	
	Método de classificação:	Não classificado	
	Dimensionamento do símbolos:	Dimensionamento aparente (*)	
	Maior raio:	5 mm	
	Cor do preenchimento dos símbolos pontuais:	#e77d2b	
	Opacidade dos símbolos pontuais:	75%	
	Cor do traço dos símbolos pontuais:	#fff7c8	
	Base Cartográfica:	Setores Censitários	
	Espessura do traço dos polígonos da base:	0.1 mm	
Cor do traço dos polígonos da base:	#4f4f4f		
<b>Mapa de pontos de contagem</b>	Valor do ponto (N):	5	
	Tamanho do ponto:	3pt (**)	
	Cor do ponto:	#df4f05	
	Base Cartográfica:	Setores Censitários	
	Cor do traço dos polígonos da base:	#232323	
	Espessura do traço dos polígonos da base:	0.1 mm	
	Opacidade dos polígonos da base:	100%	
Modo de fusão de cores da base cartográfica e tema	Queima (QGIS)		
<b>Mapa "chorodot" (grade quadrada)</b>	Tamanho da célula:	160 m	
	Número de classes: 5;	5	
	Técnica de classificação:	Natural Breaks (Jenks) - QGIS3	
	Rampa de cores:	Oranges	
	Espessura do traço da linha da grade:	0.0 mm	
	Base Cartográfica:	Setores Censitários	
	Cor do traço dos polígonos da base:	#232323	
	Espessura do traço dos polígonos da base:	0.1 mm	
Opacidade dos polígonos da base	30%		
Modo de fusão de cores da base cartográfica e tema	Multiplicar layer (QGIS)		

(\*) Expoente de Flanery (1971);

(\*\*) definido com o auxílio do guia de seleção de densidade de pontos pseudo-aleatórios, desenvolvido por Kimerling (2009);

FONTE: O autor (2022).

#### 4.3.2.5 Definição da Paleta de Cores

Neste trabalho, o foco é a comparação sobre a escolha das representações, portanto as demais escolhas foram realizadas de forma a otimizar a comparação, não entrando as variáveis das escolhas seguintes nas variações ou nas análises. Serão definidas as especificações de cores, formato e características de imagem, além da projeção cartográfica a ser utilizada.

A escolha das cores é um fator decisivo para construir bons mapas. No caso de mapas que têm a variável visual valor de cor, esta decisão é crítica. Por este motivo, utilizamos o *software Collor Brewer 2* (<https://colorbrewer2.org>), que é uma ferramenta online que auxilia na escolha de esquemas de cores que são adequados conforme o nível de medida dos dados, quantidade de classes a ser representada, a mídia em que o mapa será apresentado e a acessibilidade a daltônicos (BREWER, HATCHARD E HARROWER, 2003).

Este experimento foi delineado para ser executado em navegadores *web*, então a variável que mais se assemelha ao nosso caso é a mídia LCD. O teste será aplicado a todos sem nenhuma distinção, o nível de medida dos nossos dados é numérico e utilizaremos cinco classes nos mapas coropléticos. Neste caso, um dos esquemas de cores adequados ao nosso caso de uso é “*oranges*”, o qual foi escolhido para os mapas que utilizam a variável visual valor de cor e utilizaremos suas cores para simbolizar os mapas com outras variáveis visuais.

#### 4.3.3 Desenvolvimento da Plataforma de Teste

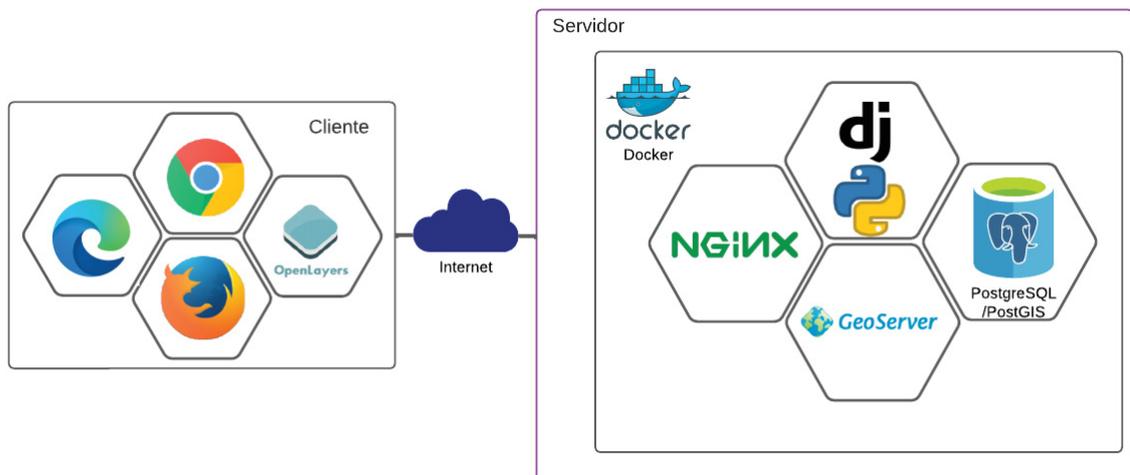
Para ser desenvolvida a plataforma e a aplicação do teste, este trabalho foi submetido e aprovado no Comitê de ética em Pesquisa do setor de ciências da saúde da UFPR - CEP/SD/UFPR. Como condicionante para a participação do teste é necessário a concordância do participante ao Termo de Consentimento Livre e Esclarecido (TCLE) (Apêndice 1), que foi construído baseado nas características desta pesquisa e aprovado pelo Comitê de ética.

Para a apresentação dos mapas aos participantes e a coleta de suas percepções foi desenvolvida uma aplicação *web*. O experimento foi realizado em ambiente online, similar à utilização contemporânea de mapas digitais. A escolha deste tipo de teste se deu pela abrangência e número de participantes, em detrimento de realizar em condições controladas como um teste presencial

Para tanto, foi proposta uma arquitetura Cliente-servidor (Figura 28). Isto permitirá um maior número de usuários para participar do teste, pois o usuário poderá realizar o teste em seu computador no próprio navegador, sem a necessidade de instalação de *software* adicional. No entanto, será imposta a utilização de monitores maiores ou iguais a 13”.

No servidor, cada componente do sistema é construído dentro de *containers docker*. Os *containers* permitem “empacotar” uma aplicação inteira, isolando de outras aplicações, de forma a eliminar problemas relacionados com dependências de bibliotecas, assim é facilitada a implantação da aplicação em qualquer ambiente computacional. No contexto de reprodutibilidade de pesquisa, os *containers* permitem o compartilhamento dos códigos e *softwares* utilizados nesta pesquisa, sem a necessidade de conhecimentos aprofundados para fazer sua implantação.

FIGURA 28 – ARQUITETURA CLIENTE-SERVIDOR UTILIZADA PARA O DESENVOLVIMENTO DA PLATAFORMA DE APLICAÇÃO DOS TESTES



FONTE: O autor (2022).

A interação do usuário com o mapa acontece no navegador *Web*, que exibe a página desenvolvida em *HTML/CSS/Javascript* com inclusão da interface do mapa desenvolvida com biblioteca *OpenLayers*, a qual permitire a realização de desenhos feitos pelos usuários, que são produtos da sua percepção. A página web será estruturada com a linguagem *HTML*, estilizada com a linguagem *CSS* e as demais funcionalidades programadas com a linguagem *Javascript*.

A comunicação entre o cliente e aplicação é executada pelo servidor *NGINX*, o qual recebe solicitações do cliente e designa para o subcomponente responsável.

Neste sistema, quando o cliente solicitar um dado espacial, será redirecionado para o container em que está o servidor de mapas *Geoserver*.

O *Geoserver* disponibiliza os geoserviços, como *WMS (Web Map Service)*. O geoserviço *WMS* permite requisições de dados geográficos renderizados na forma de imagem utilizando diversos formatos. Foi definido o uso de imagens no formato *TIFF*, que são renderizadas e servidas com o *Geoserver*. Por uma análise visual foi definido a qualidade da imagem para 148 *ppi*, que geram visualizações semelhantes às visualizadas no momento da sua criação.

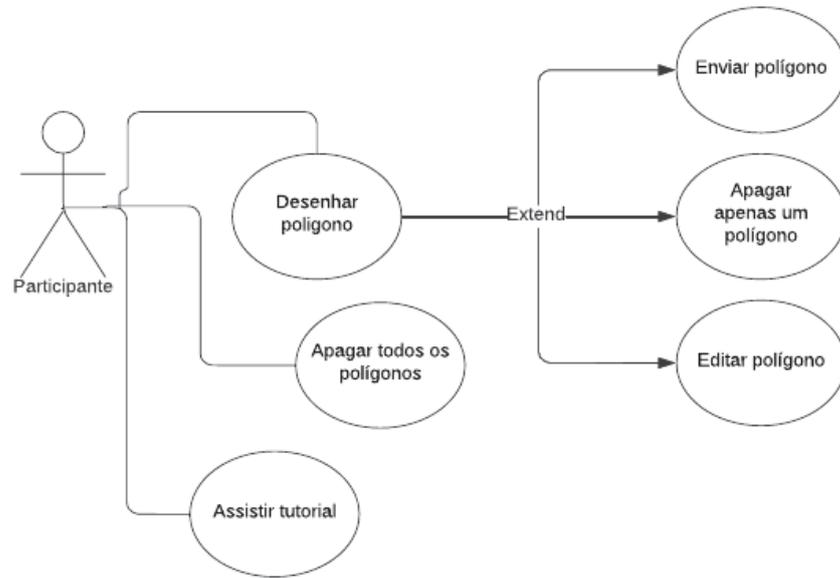
Nesta aplicação, os dados geográficos são armazenados em dois sistemas: Sistema de Gerenciamento de banco de dados *PostgreSQL* com sua extensão espacial *PostGIS* e sistemas de arquivos. Por motivos de performance, as imagens dos mapas serão acessadas pelo geoserver no sistema de arquivos, no caso dos polígonos referentes às regiões identificadas pelos participantes, estão armazenados no *PostgreSQL/PostGIS*. Também são armazenadas neste banco de dados, as configurações da aplicação web desenvolvidas com o *framework Django*.

Neste trabalho, utilizamos para desenvolvimento backend com a assistência do *framework Django*, em conjunto com uma série de bibliotecas já desenvolvidas em *Python*, inclusive para aplicações geoespaciais.

#### 4.3.3.1 Diagrama de Casos de Uso

Para o desenvolvimento da interface foram pressupostos alguns casos de uso (Figura 29) que possibilitaram a realização do teste. A princípio, o participante precisa aprender o funcionamento das ferramentas da plataforma, portanto foram elaborados vídeos curtos para cada funcionalidade da plataforma. A funcionalidade principal da plataforma é realizar o desenho das regiões percebidas, então além de ter este caso de uso, são originados casos de usos que podem ou não serem executados após o desenho das regiões. Se o participante não estiver satisfeito com o resultado dos seus desenhos ele pode editar o polígono, assim como apagar um ou todos os polígonos. E por fim pode enviar seu desenho para o servidor.

FIGURA 29 – DIAGRAMA DE CASOS DE USO DO SISTEMA

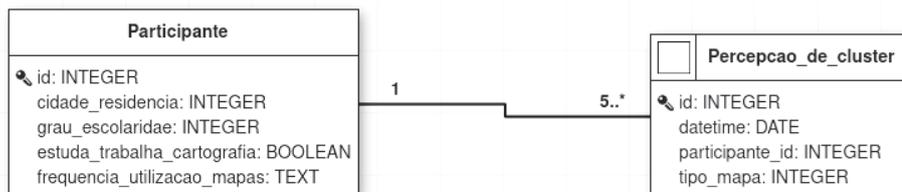


FONTE: O autor (2022).

#### 4.3.3.2 Diagrama de Classes

O modelo do Banco de Dados (figura 30) apresenta apenas duas classes,. A primeira refere-se aos participantes e seus atributos são compostos pelas informações da sua caracterização. A outra classe refere-se às percepções dos usuários de cada mapa e tem como chave estrangeira o identificador de qual participante pertence, além de data e hora da sua inserção no sistema e qual tipo de mapa foi visualizado.

FIGURA 30 – DIAGRAMA DE CLASSES



FONTE: Os autores (2022).

#### 4.3.3.3 Interface

A próxima etapa é o projeto da interface a ser utilizada. Para a realização deste estudo é necessário no mínimo três templates diferentes: Página para apresentação do termo de consentimento livre e esclarecido; Formulário para caracterização dos participantes; *Web* mapa.

A primeira página a ser apresentada aos participantes contém as informações legais necessárias para a execução do teste. Na página há um botão para o *download* do TCLE. Esta página é acessada pelo participante se ele estiver acessando de um computador *desktop* ou *notebook*, pois no código é introduzido uma funcionalidade que bloqueia dispositivos *mobiles*. Se o participante concordar com o TCLE, ele será direcionado para a página que contém o formulário de caracterização do participante.

Se o participante responder todas as questões e enviar as respostas ele será redirecionado para o *web* mapa. Sobreposto ao *web* mapa aparecerá um popup com instruções sobre os procedimentos para realização do teste. Após a leitura das instruções serão apresentados os vídeos com os tutoriais sobre as funcionalidades. A visualização dos vídeos é opcional, porém no *web* mapa há um botão para reabrir às instruções e aos vídeos. Quando o participante fecha o *pop up* contendo os vídeos, o *web mapa* é transformado para apresentação em tela cheia. O primeiro mapa que é apresentado tem a função de treinamento para o usuário aprender a usar as funcionalidades dos mapas. Sobreposto ao *web* mapa são posicionados botões para cada caso de uso.

A cada envio dos desenhos são verificados se o usuário desenhou sobre o mapa e se o servidor conseguiu armazenar a resposta, para então mudar o tipo de mapa para o participante. Para cada participante é apresentado sete mapas diferentes, considerando o mapa de treinamento. Foi desenvolvido um *script* para extrair amostras aleatórias com seis mapas do universo geral dos mapas a serem testados, com a restrição de não ocorrer mapas sequenciais construídos com o mesmo recorte de setores, para minimizar a influência do arranjo dos setores na percepção dos padrões.

A interatividade com os mapas é controlada, desabilitando as funcionalidades no *OpenLayers*. Os controles de “*zoom*” e “*pan*” são desabilitados para que todos os participantes tenham a mesma visão global dos mapas.

#### 4.3.4 Definição da Medida de Eficiência

Para a mensuração da eficiência da representação cartográfica na visualização de agrupamentos pelos usuários, precisamos primeiramente definir os *clusters* "reais" formados pelos dados sintéticos criados para cada um dos mapas. Estes foram definidos através de estatística espacial, usando o Índice de Moran Local. O índice de Moran Local é um tipo de estatística LISA (*Local Indicator Of Spatial Association*), que tem duas principais características: fornece uma estatística para cada local associado a uma medida de significância; e soma das estatísticas locais é igual a estatística global de autocorrelação espacial (ANSELIN, 1995). Neste caso, a soma dos indicadores de cada local é igual ao índice de Moran Global.

Para cada observação, o índice de Moran local fornece indícios de similaridade entre valores em torno dessa observação. O cálculo do índice de moran é realizado com a seguinte equação:

$$I_i = z_i * \sum_j \omega_{ij} * z_j$$

$\omega_{ij}$  – valor da matriz de vizinhança para a região *i* com a região *j* em função da distância

$z_i$  – desvio do valor do polígono de referência em relação a média

$z_j$  – desvio do valor do polígono vizinho em relação a média

Para nossa aplicação utilizaremos o *software Geoda* para a definição das regiões. As regiões foram definidas com significância igual a 0.05 e 999 permutações. O resultado é armazenado em cada setor da base, com a identificação do tipo de região a que aquele polígono pertence. Os tipos de regiões possíveis são *High-High*, *High-Low*, *Low-Low* e *Low-High*. Porém, para este trabalho vamos selecionar as regiões *High-High*.

Após a definição das regiões nos dados brutos, é calculada a eficiência que cada participante obteve em suas análises visuais. Foi requisitado o desenho de cinco regiões em cada mapa analisado, então são analisadas as regiões desenhadas baseado na existência ou não de *clusters* naquele local. É calculado um fator de eficiência baseado em 3 situações na comparação entre as regiões calculadas e desenhadas (Figura 31):

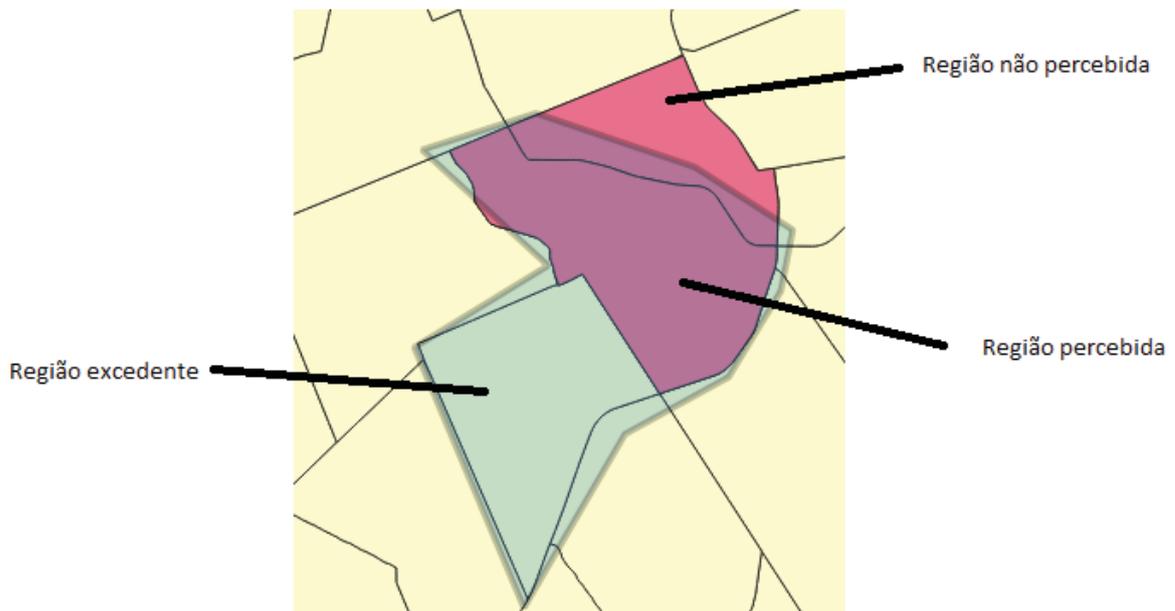
- Não existe *cluster* que intercepte a região desenhada. Então calcula-se a razão entre as regiões desenhadas que não existem clusters matemáticos pelo número total de regiões desenhadas. É um fator positivo (+) que pode ser entendido como taxa de regiões percebidas erroneamente;

- É calculada a razão entre as áreas das regiões que foi percebida e a área das *clusters* matemáticos que intersectam com a região percebida. Esse é um fator negativo (-) e pode ser entendido como a taxa de acerto na percepção da *cluster* matemática;

- É calculada a razão entre a área da região que foi desenhada pela área da região que percebida. Esse é um fator positivo (+) e pode ser entendido como um excesso ao determinar a região.

Cada mapa tem um fator de eficiência, que é determinado a partir do cálculo da média de todas as análises que este mapa foi submetido. Quanto maior for este fator pior será sua eficiência.

FIGURA 31 – REPRESENTAÇÃO DAS SITUAÇÕES CONSIDERADAS PARA O CÁLCULO DO FATOR DE EFICIÊNCIA



FONTE: Os autores (2022).

#### 4.3.5 Análise das Regiões Percebidas

Para explorar os resultados obtidos, as eficiências foram agrupadas por conjunto de dados testados e pela representação cartográfica utilizada. Então, foram construídos gráficos de barras com as médias das eficiências agrupadas para visualizar a distribuição das respostas por conjunto de dados. Em seguida, para verificar em qual local do espaço suavidade-continuidade cada mapa apresenta maiores médias, foram construídos gráficos de bolhas com dados agrupados por tipo de mapa e valores de suavidade-continuidade, em que o tamanho das bolhas representa a eficiência do mapa em cada conjunto de dados.

E por fim, foi realizada uma regressão linear múltipla, de forma a descrever as relações entre múltiplas variáveis explicativas - a suavidade e continuidade - na determinação de um processo ou comportamento, ou seja, na eficiência de cada mapa (EFROYMSON, 1960). O modelo estatístico de uma regressão linear múltipla com  $k$  variáveis explanatórias é (HOFFMANN, 2016):

$$Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + u_j \quad j = 1, \dots, n$$

$Y_j$  é a variável dependente em função das variáveis explanatórias

$(X_{ij} \quad i = 1, \dots, k)$

$\beta_k$  são os coeficientes de regressão

$\alpha$  é o coeficiente de regressão constante

$u_j$  é o erro ou resíduo referente a diferença entre o valor  $Y$  real e o  $Y$  previsto

Uma solução de regressão múltipla fornece o "melhor" valor dos mínimos quadrados desses coeficientes para uma amostra específica de observações. A solução também fornece uma medida da confiabilidade de cada um dos coeficientes para que possam ser feitas inferências sobre os parâmetros da população da qual a amostra de observações foi retirada (EFROYMSON, 1960). A partir do modelo matemático construído com a regressão linear múltipla, é possível construir um gráfico de predição, que baseado em um valor de suavidade e continuidade prevê qual técnica pode ser mais adequada.

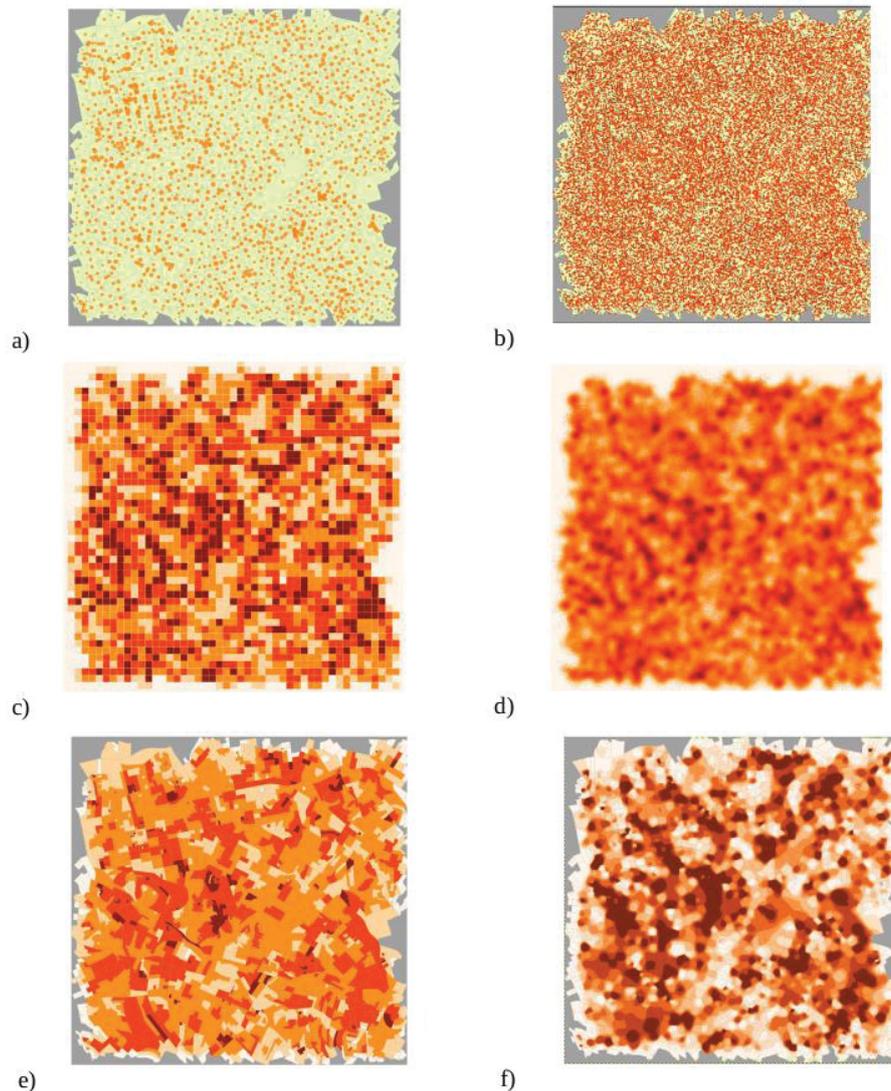
#### 4.5 RESULTADOS

Nesta seção, primeiramente são descritos os mapas construídos para o teste e as regiões que foram definidas posteriormente. Na sequência é apresentada a aplicação de coleta de dados. E por último são apresentados os resultados do experimento de percepção de regiões.

#### 4.4.1 Mapas Construídos e *Clusters* Matemáticos

Para cada distribuição gerada com o algoritmo que gera dados sintéticos, foram construídos seis mapas. Então, foram utilizadas 6 distribuições, totalizando 56 mapas diferentes. Na Figura 32 são apresentados os diferentes tipos de mapas gerados com a distribuição com 100% de continuidade e 85% de suavidade, considerada como uma distribuição contínua e suave. Ao fazer uma comparação visual entre os mapas gerados, nota-se que há uma semelhança de regiões que podem ser percebidas nos mapas c), d), e) e f). E as piores representações são os b) e a). Os demais mapas que foram construídos são apresentados no Apêndice 2.

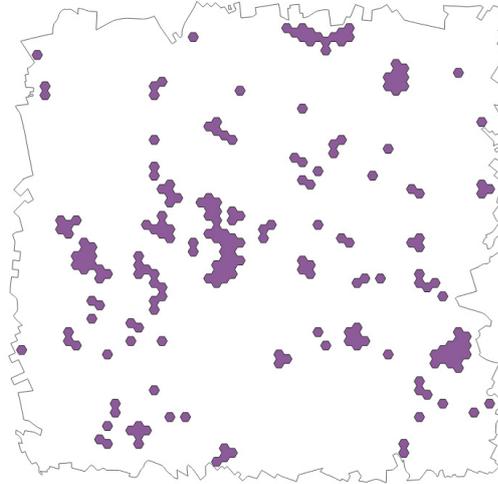
FIGURA 32 – DIFERENTES TIPOS DE MAPAS CONSTRUÍDOS COM OS DADOS COM CARACTERÍSTICAS DE SUAVIDADE 85% E CONTINUIDADE 100%



FONTE: Os autores (2022).

Para cada distribuição foram geradas as *clusters* matemáticas no *software Geoda*. Porém, este *software* retorna apenas a indicação de qual grupo um polígono pertence, sendo necessário processar a camada e dissolver as geometrias baseado no atributo que indica o grupo. Também é necessário excluir as regiões que não são interessantes para este experimento, ou seja, manter apenas regiões de tipo “*High-High*”. Na Figura 32 são apresentados os clusters matemáticos já processados, referentes à distribuição com suavidade 85% e continuidade 100%, idêntica a distribuição usada nos mapas na Figura 33.

FIGURA 33 – CLUSTERS MATEMÁTICOS REFERENTES AOS DADOS DA FIGURA 32



FONTE: Os autores (2022).

#### 4.4.2 Aplicação Web Construída

A seguir serão apresentadas as páginas *web* na sequência que os participantes desta pesquisa serão expostos durante a aplicação do teste. A primeira página (Figura 34) esclarece o participante sobre informações da pesquisa e possibilita o *download* do Termo de consentimento livre e esclarecido (TCLE).

FIGURA 34 – PÁGINA INICIAL DO TESTE






### Pesquisa sobre percepção de regiões em mapas temáticos.

Você é convidado(a) a colaborar voluntariamente com minha pesquisa por meio da realização deste teste. Os dados coletados serão utilizados como insumos para a pesquisa do doutorando Raphael Gonçalves de Campos, aluno do curso de Pós-Graduação em Ciências Geodésicas, na Universidade Federal do Paraná, sob orientação da Profª Dra. Silvana Philippi Camboim e do Prof. Dr. João Vitor Meza Bravo.

**É necessária a utilização de notebook ou desktop para a realização do teste!**

Acesse o termo de consentimento livre e esclarecido: [Download!](#)

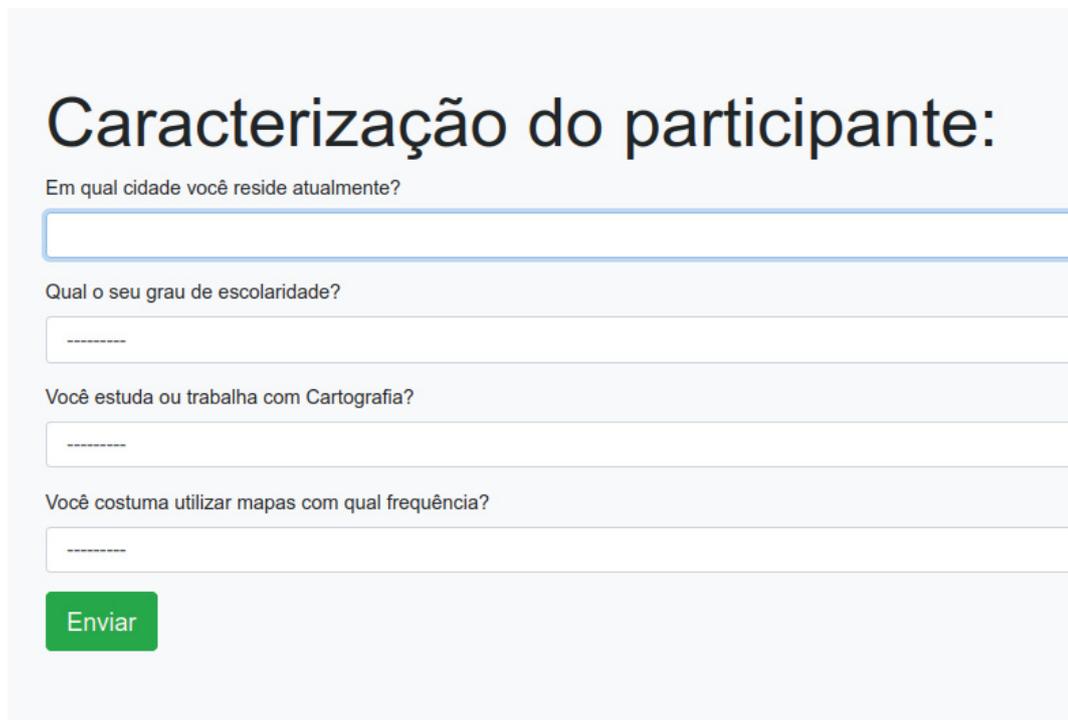
Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa (CEP) em Ciências Humanas e Sociais da Universidade Federal do Paraná (UFPR).  
 Número do Parecer: 5.426.02  
 CAAE: 57999622.2.0000.0102

Ao clicar no botão "Aceito" você afirma que leu o termo de consentimento livre e esclarecido e concorda em participar do teste para a pesquisa de percepção cartográfica, visto que a preservação do anonimato é garantida, e que as informações obtidas serão utilizadas exclusivamente para a realização desta pesquisa. Sua decisão de participação é voluntária, podendo parar a qualquer momento.

FONTE: Os autores (2022).

A próxima imagem retrata o formulário para inserção das informações de caracterização dos participantes. O sistema é desenvolvido para avaliar se há dados em cada campo antes de submeter.

FIGURA 35 – CARACTERIZAÇÃO DO PARTICIPANTE



Caracterização do participante:

Em qual cidade você reside atualmente?

Qual o seu grau de escolaridade?

Você estuda ou trabalha com Cartografia?

Você costuma utilizar mapas com qual frequência?

Enviar

FONTE: Os autores (2022).

As Figuras 36 e 37 apresentam as instruções para realizar o teste e os vídeos para aprendizado do uso das ferramentas do *web* mapa, respectivamente. As informações mais relevantes para a execução do teste são destacadas com o intuito que o participante entenda as tarefas a serem realizadas. Os vídeos tutoriais são apresentados em “tela cheia” para propiciar a identificação dos procedimentos retratados no tutorial. O participante não é obrigado a assistir todos os vídeos. Porém, o primeiro vídeo apresentado para o usuário é automaticamente iniciado, com a intenção de estimular o usuário a assisti-lo.

FIGURA 36 – INSTRUÇÕES PARA REALIZAR O TESTE

**Instruções (Leia com atenção!)**

A seguir será apresentado um conjunto de vídeos para aprendizado das ferramentas da interface, você poderá acessar os vídeos a qualquer momento do teste pelo botão de ajuda;

Posteriormente, será apresentada uma sequência de mapas para realizar as tarefas, observe as instruções abaixo para executá-las:

- Cada mapa representa um fenômeno distinto;
- Serão apresentados 7 mapas, dos quais o primeiro será utilizado para treinamento;
- Observe cada mapa e desenhe no mapa regiões onde há maiores ocorrências do fenômeno, dê preferência para as maiores regiões;
- Desenhe 5 regiões por mapa;

obs:

Para desenhar no mapa aperte o botão "desenhar", após clique no mapa e arraste.

Você poderá excluir cada polígono clicando no botão "selecionar e apagar", após clique sobre o polígono a ser excluído;

Pode excluir todos os polígonos desenhados ao clicar no botão "apagar tudo"

Entendi

FONTE: Os autores (2022).

FIGURA 37 – VÍDEO PARA APRENDIZADO DO USO DE FERRAMENTAS WEB

Exemplos de como utilizar as funções do teste

Como desenhar sobre o mapa e enviar o desenho

0:00 / 0:37

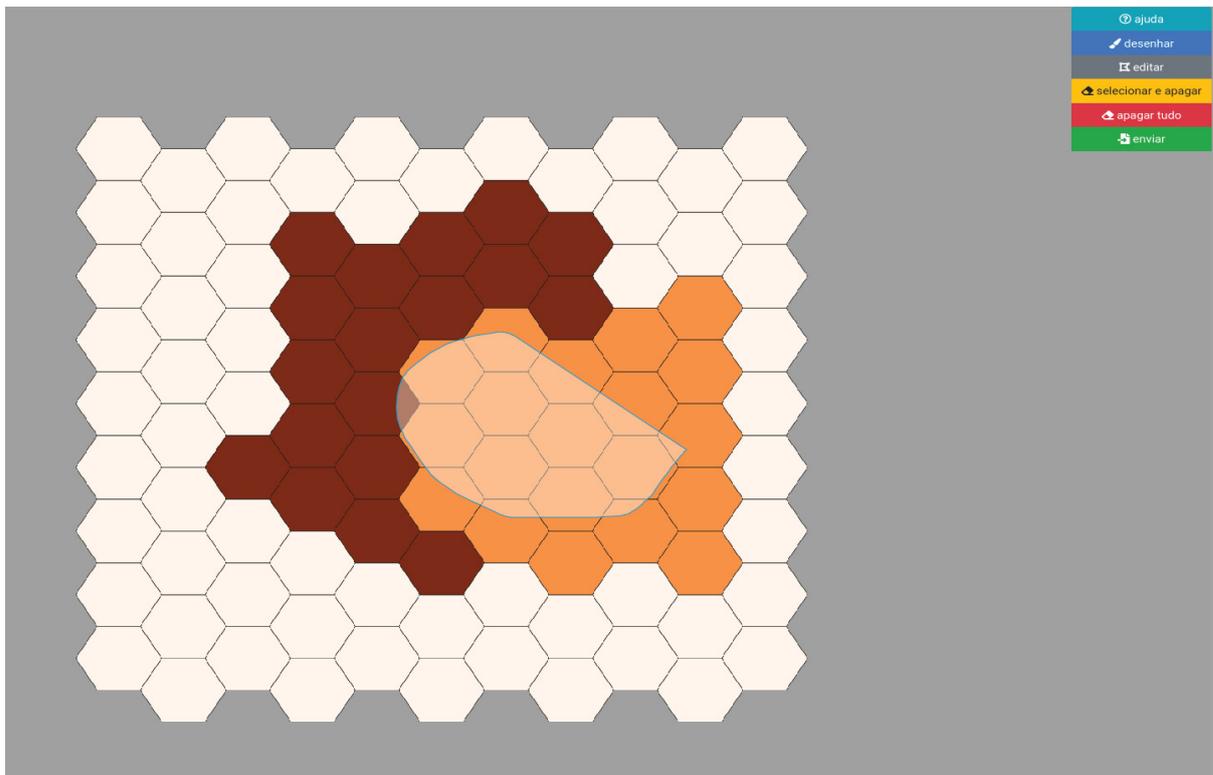
Fechar

FONTE: Os autores (2022).

A Figura 37 demonstra o mapa *web*, na opção de mapa para aprendizado do participante. Os botões estão dispostos no canto superior direito, de forma a não sobrepor os mapas apresentados. O *web* mapa é apresentado ao usuário em “tela cheia” e com a função de desenho ativada, de forma a não esforçar o usuário a procurar a função. O apontador do *mouse* é modificado na função desenho para

identificar que a função está ativa. A interação com o botão que apaga todos os polígonos desenhados exige a confirmação da ação por parte do usuário, para que o usuário não o execute acidentalmente (Figura 38).

FIGURA 38 – MAPA PARA APRENDIZADO DAS FERRAMENTAS DA PLATAFORMA



FONTE: Os autores (2022).

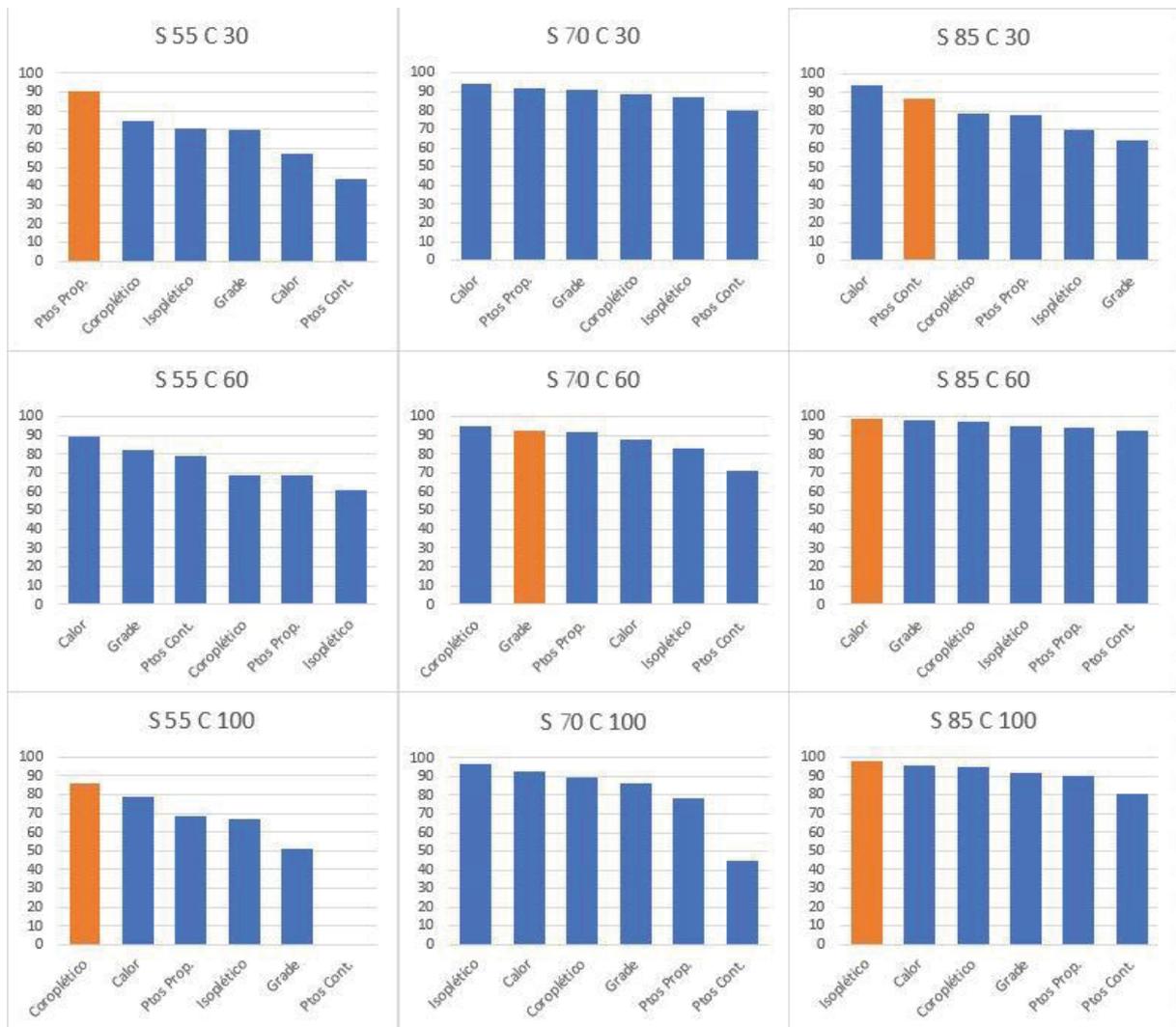
#### 4.4.3 Resultados do experimento

O experimento teve a participação de 101 participantes, que realizaram o desenho de 3.982 regiões distribuídas nos 56 mapas. A distribuição de respostas por mapa foi desigual, variando de 9 a 16 por mapa. Isto é consequência das desistências do teste antes da sua conclusão. Para simplificar a análise dos dados, o fator de efetividade foi invertido e escalado, de modo que os valores variam entre 0 a 100 e os maiores valores passassem a indicar mais eficiência.

Para explorar os dados, as respostas foram agrupadas por distribuição e por representação, para então calcular a média do fator por mapa. A (Figura 39) apresenta o gráfico de colunas referente aos dados agrupados e cada coluna

representa a média do fator modificado para cada tipo de mapa. No gráfico as colunas que estão simbolizadas com a cor laranja são as técnicas de representação cartográfica temática pressupostas a serem indicadas para representar as distribuições de dados com aquelas características. Vale ressaltar que não foi construído um tipo de mapa adequado para cada distribuição.

FIGURA 39 – GRÁFICO DE BARRAS DAS MÉDIAS DE EFICIÊNCIA DE CADA TIPO DE MAPA PARA UM CONJUNTO DE DADOS ESPECÍFICO (S – SUAVIDADE, C – CONTINUIDADE)

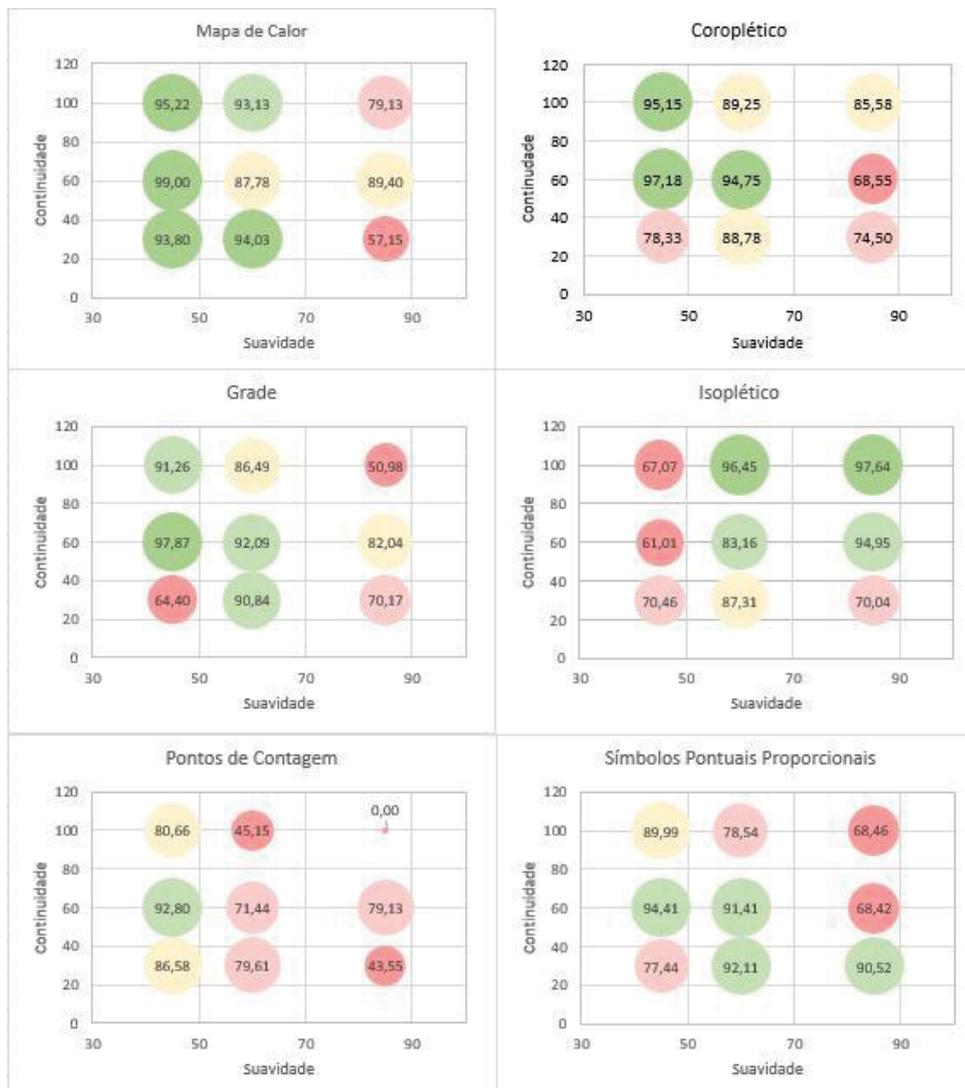


FONTE: Os autores (2022).

Obtemos o primeiro indício positivo que o tipo de mapa influencia na realização da tarefa, pois 66,6% dos casos a técnica pressuposta teve a maior influência. E nos demais casos a técnica pressuposta ficou ranqueada em segundo lugar. O mapa de calor foi mais eficiente em aproximadamente 44% dos casos.

Uma outra análise foi feita para verificar em quais regiões do espaço bidimensional suavidade-continuidade cada técnica obteve maior eficiência. Para isto, os fatores de eficiência modificados foram agrupados por técnicas de representação e por suavidade e continuidade. Então, foi construído um gráfico de bolhas (Figura 40), em que o eixo vertical se refere a continuidade, o eixo horizontal é referente a suavidade e o tamanho dos círculos é proporcional aos fatores de eficiência modificados.

FIGURA 40 – GRÁFICO DE BOLHAS REPRESENTANDO A PERFORMANCE DOS TIPOS DE MAPAS CONFORME A SUAVIDADE E CONTINUIDADE DE CADA CONJUNTO DE DADOS



FONTE: Os autores (2022).

É possível verificar que os mapas de calor são mais eficientes quando a suavidade não é tão alta. Os mapas coropléticos são mais eficientes quando a

continuidade é alta e a suavidade é baixa. Os mapas “chorodot” (grades) têm mais eficiência em suavidade baixa e continuidade alta, também quando a suavidade é média e a continuidade é média e baixa. Os mapas isopléticos têm maior eficiência quando a continuidade é alta e a suavidade é média e alta. Os mapas de pontos de contagem têm maior eficiência quando a suavidade é baixa. Os símbolos pontuais proporcionais têm sua eficiência afetada quando a continuidade é alta.

Por fim, para avaliar a relação entre a eficiência dos mapas e as características de suavidade e continuidade foi construído um modelo de regressão linear múltipla para cada tipo de mapa, na Tabela 8, são apresentados os parâmetros de regressão para cada tipo de mapa.

TABELA 8 – PARÂMETROS DA REGRESSÃO PARA CADA TIPO DE MAPA

	$\alpha$	S ( $\beta_1$ )	C ( $\beta_2$ )
Isoplético	20,919	0,712	0,160
Pontos Proporcionais	63,759	0,383	-0,112
Coroplético	44,729	0,467	0,1320
Calor	32,967	0,693	0,097
Grade	41,819	0,559	-0,005
Pontos Contagem	-14,634	1,515	-0,426

FONTE: Os autores (2022).

Na Tabela 8 podemos observar que os parâmetros de regressão para a suavidade (S) influenciam mais os modelos de todos os tipos de mapas do que os parâmetros de regressão para a continuidade (C). Para a regressão são gerados índices de qualidade, apresentados na Tabela 9.

TABELA 9 – QUALIDADE DAS REGRESSÕES PARA CADA TIPO DE MAPA

	R-Múltiplo	Erro Padrão			valor-P			Resíduos padrão (médio)	F de significação
		$\alpha$	S	C	$\alpha$	S	C		
Isoplético	74,4%	32,846	0,432	0,185	0,387	0,052	0,251	-6,106E-15	0,089
Pontos Proporcionais	60,9%	20,205	0,266	0,113	0,020	0,200	0,363	1,875E-15	0,292
Coroplético	72,8%	16,372	0,215	0,092	0,034	0,073	0,201	2,090E-15	0,104
Calor	74,3%	20,396	0,268	0,115	0,157	0,042	0,428	-1,566E-15	0,090
Grade	46,7%	32,846	0,432	0,185	0,250	0,243	0,981	0	0,477
Pontos de Contagem	80,3%	41,763	0,549	0,235	0,738	0,033	0,119	0	0,04

FONTE: Os autores (2022).

A qualidade caracterizada pelo índice R-Múltiplo demonstra a correlação entre os fatores reais de eficiência dos mapas e os fatores preditos pelo modelo de regressão, então quanto mais perto de 100% for o R-Múltiplo, os fatores reais estão mais correlacionados aos fatores previstos. Neste caso, apenas os mapas de Pontos Proporcionais e os Mapas de Grade tiveram valores menores que 70%, sendo o último com pior resultado. A causalidade não pode ser inferida apenas pela análise de dados – em vez disso, deve-se inferir que existe uma relação causal com base em uma teoria causal que explica os resultados (RUBINFELD, 2000). Porém, os resultados reiteram o modelo desenvolvido por Maceachren e Dibiase (1991).

O valor-P fornece a significância estatística de cada parâmetro a ser considerado no modelo de regressão. Assim, busca saber quais variáveis estão relacionadas com a variável dependente, ou seja, traz indicativos se a suavidade ou continuidade está relacionada ao fator de eficiência de cada mapa. Para que exista significância, o valor-P deve ser  $< 0,05$  (à 95%). Os parâmetros dos modelos que

exibiram indicativos significativos foram:  $\alpha$  para os mapas coropléticos e pontos pontuais proporcionais; Suavidade para os mapas de calor, isopléticos e pontos de contagem. Os parâmetros de regressão da continuidade não apresentaram significância para nenhum tipo de mapa. Entretanto, a significância pode aumentar conforme o tamanho da amostra aumenta (RUBINFELD, 2000), deste modo, é necessário um cenário com mais participantes para analisar a significância destes parâmetros.

O F de significação avalia se o modelo proposto é útil para explicar a variável dependente, ou seja, busca identificar se pelo menos uma das variáveis independentes está relacionada à variável dependente. Assim, o F de significação deve ser  $< 0,05$  para que o modelo seja considerado útil, visto que avalia a significância estatística geral do modelo estimado. O modelo que apresentou resultado satisfatório neste teste foi o mapa de pontos de contagem.

Os modelos de regressão desenvolvidos permitem fazer a predição dos fatores de eficiência para cada mapa, então podemos prever os mapas mais eficientes para diferentes características de suavidade e continuidade. Então, foram previstos os melhores mapas para a suavidade e continuidades de 0% a 100%, com intervalos de 10%, que são exibidos na (Figura 41).

FIGURA 41 – MODELO DE PREVISÃO DAS MELHORES TÉCNICAS DE REPRESENTAÇÃO PARA INTERVALOS DE SUAVIDADE E CONTINUIDADE

		Suavidade										
		0	10	20	30	40	50	60	70	80	90	100
Continuidade	0	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Prop/Cont	Cont.	Cont.	Cont.
	10	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Prop/Cont	Cont.	Cont.	Cont.
	20	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Cont.	Cont.	Cont.
	30	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Calor/Prop	Calor/Cont	Cont.	Cont.
	40	Prop.	Prop.	Prop.	Prop.	Prop.	Prop.	Calor/Prop	Calor/Prop	Calor/Prop	Cont.	Cont.
	50	Prop.	Prop.	Prop.	Prop.	Coro/Prop	Coro/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Cont	Cont.
	60	Prop.	Prop.	Coro/Prop	Coro/Prop	Coro/Prop	Coro/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Cont	Calor/Cont
	70	Coro/Prop	Coro/Prop	Coro/Prop	Coro/Prop	Coro/Prop	Coro/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Cont
	80	Coro/Prop	Coro/Prop	Coro/Prop	Coro/Prop	Coro	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop
	90	Coro	Coro	Coro	Coro	Coro	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop
	100	Coro	Coro	Coro	Coro	Coro	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop	Calor/Prop

FONTE: Os autores (2022).

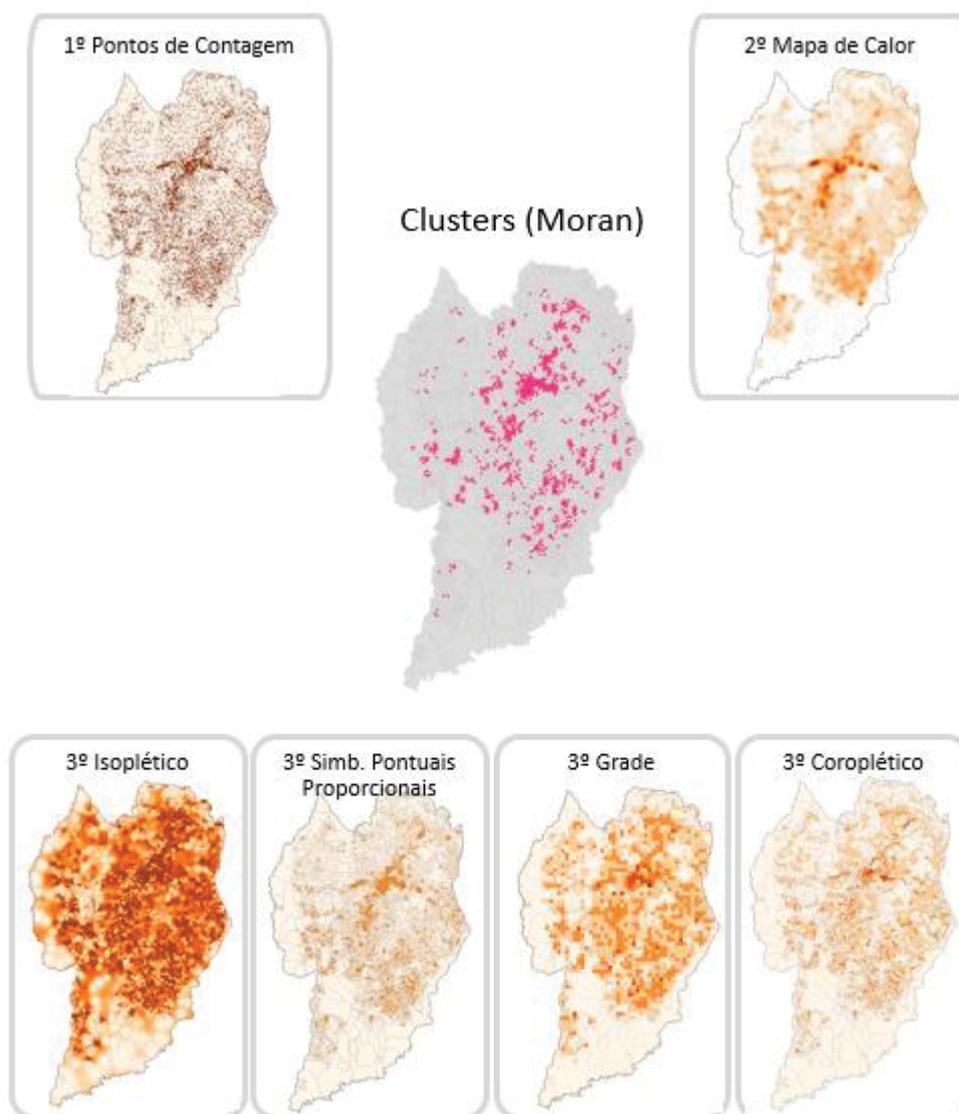
Os códigos utilizados em cada intervalo correspondem a técnicas de representação temática da seguinte forma:

- Prop – Mapas de pontos proporcionais
- Cont – Mapas de pontos de contagem
- Calor – Mapas de Calor
- Coro – Mapas Coropléticos
- Isso – Mapas Isopléticos

Nos locais do modelo acima em que há dois mapas, a diferença dos fatores de eficiência entre as melhores técnicas foi de menos de 5%. A única técnica que não é contemplada no modelo é o “chorodot” (grade). No entanto, há uma similaridade entre o modelo previsto pelos modelos de regressão e o modelo definido por Maceachren e Dibiase (1991). Para examinar o modelo construído, utilizamos uma base real de dados pontuais, denominada de Base 156 (<https://www.curitiba.pr.gov.br/dadosabertos/busca/>). Os dados foram agregados em uma grade hexagonal, posteriormente foram calculadas as suas características de

suavidade e continuidade. Os dados apresentavam a suavidade de 98% e a continuidade de 39%, que, segundo o modelo, o mapa mais indicado para visualizar regiões com alta densidade de eventos é o mapa de pontos de contagem. Então, foram aplicadas aos dados todas as técnicas de representação temática estudadas neste trabalho, além de calcular os clusters matemáticos. Os mapas contruídos e os clusters calculados são apresentados na figura 42.

FIGURA 42 – MAPAS CONSTRUÍDOS COM OS DADOS DA BASE 156 E OS CLUSTERS MATEMÁTICOS CALCULADOS PELO ÍNDICE DE MORAN LOCAL



FONTE: Os autores (2022).

Embora os mapas da Figura 42 não foram testados com usuários, pode-se perceber que no mapa de pontos de contagem e no mapa de calor as regiões

formadas tem maior similariedade com os clusters matemáticos. Ressalta-se que os dados base 156 foram escolhidos ao acaso, não foi realizado nenhum tipo de escolha ou filtragem dos dados.

#### 4.6 CONCLUSÃO

Neste trabalho, conseguimos desenvolver uma plataforma para testes em ambiente Web. Nesta plataforma apresentamos um conjunto de mapas cuja variação da suavidade e da continuidade foram representadas por intermédio de diferentes técnicas de representação temática. Assim conseguimos avaliar o conjunto de representações, quanto à sua adequabilidade às características de suavidade e continuidades, conforme os modelos de MacEachren e DiBiase (1991).

Com base nos resultados encontrados, conseguimos comprovar, com base na amostra de indivíduos entrevistados, que o processo de seleção de técnicas adequadas à representação temática de um conjunto de dados funciona também para o contexto do *Big Data*. Isso ocorre, pois, nos *rankings* os métodos de representação que supostamente seriam adequados àquelas características de dados obtiveram o primeiro ou a segundo melhor posição. O mapa de calor obteve destaque no ranqueamento em 44,44% das configurações de suavidade e continuidade; concluímos ocorreu porque o algoritmo para geração do mapa de calor foi desenvolvido justamente para visualizar regiões com maior densidade de ocorrência de pontos, suportando, a visualização dos padrões de regiões com maior ocorrência massiva de pontos.

Os usuários foram parcialmente capazes de reconhecer os padrões de agrupamento de pontos em representações como os mapas feitos com a técnica de pontos de contagem. Isso pode ter ocorrido porque nesta técnica há menor generalização da representação pontual, uma vez que próprio intuito da técnica é representar fenômenos espaciais por intermédio de pontos, em seus respectivos locais de ocorrência. Assim, entende-se que esta técnica não foi capaz de estimular a leitura adequada de um fenômeno pontual volumoso. Em contrapartida, as demais técnicas - exceto o mapa de calor - tiveram melhor eficiência quando a suavidade e a continuidade tinham valores próximos ao que a literatura indicava para a escolha destas técnicas.

Com os resultados obtidos nas entrevistas, considerando-se as configurações de suavidade e continuidade que foram testadas, foi possível criar um modelo para determinação da técnica de representação temática mais adequada à detecção de regiões com maior densidade de ocorrências de pontos. Apesar do modelo de regressão ter sido treinado com um número reduzido de amostras, i.e., respostas de indivíduos entrevistados, aqui conseguimos inferir que é viável utilizar a suavidade e a continuidade enquanto medidas de variação que permitem a escolha do método de representação temática.

#### 4.7 REFERÊNCIAS

AMAR, R.; EAGAN, J.; STASKO, J. Low-level components of analytic activity in information visualization. **Proceedings - IEEE Symposium on Information Visualization, INFO VIS**, p. 111–117, 2005.

ANSELIN, L. Local Indicators of Spatial Association—LISA. **Geographical Analysis**, v. 27, n. 2, p. 93–115, 1995. Disponível em: <<http://doi.wiley.com/10.1111/j.1538-4632.1995.tb00338.x>>. Acesso em: 11/12/2019.

BREWER, C. A.; HATCHARD, G. W.; HARROWER, M. A. ColorBrewer in print: A catalog of color schemes for maps. **Cartography and Geographic Information Science**, v. 30, n. 1, p. 5–32, 2003.

CARR, D. B.; OLSEN, A. R.; WHITE, D. Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data. <http://dx.doi.org/10.1559/152304092783721231>, v. 19, n. 4, p. 228–236, 2013. Taylor & Francis Group . Disponível em: <<https://www.tandfonline.com/doi/abs/10.1559/152304092783721231>>. Acesso em: 6/7/2022.

HOFFMANN, R. Análise de regressão: uma introdução à econometria. **Análise de regressão: uma introdução à econometria**, 2016. Universidade de São Paulo. Escola Superior de Agricultura Luiz de Queiroz.

KOLÁČNÝ, A. Cartographic information—a fundamental concept and term in modern cartography. **Cartographic Journal**, v. 6, n. 1, p. 47–49, 1969.

KIM, J.; FOX, M. F. J.; FIELD, A. R.; NAM, Y. U.; GHIM, Y. C. Conditions for generating synthetic data to investigate characteristics of fluctuating quantities. **Computer Physics Communications**, v. 204, p. 152–158, 2016. Elsevier B.V. Disponível em: <<http://arxiv.org/abs/1510.07457>>. Acesso em: 6/7/2022.

KRAAK, MENNO-JAN; ORMELING, F. **Visualization of Spatial Data Third Edition**. 2010.

LEWANDOWSKY, S.; HERRMANN, D. J.; BEHRENS, J. T.; et al. Perception of clusters in statistical maps. **Applied Cognitive Psychology**, 1993.

LI, S.; DRAGICEVIC, S.; CASTRO, F. A.; et al. Geospatial big data handling theory and methods: A review and research challenges. **ISPRS Journal of Photogrammetry and Remote Sensing**, 2016a.

M. A. Efronymson, "Multiple Regression Analysis," In A. Ralston and H. S. Wilf, Eds., *Mathematical Methods for Digital Computers*, John Wiley, New York, 1960. - References - Scientific Research Publishing. Disponível em: <[https://www.scirp.org/\(S\(351jmbntvnsjt1aadkposzje\)\)/reference/ReferencesPapers.aspx?ReferenceID=430544](https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=430544)>. Acesso em: 4/7/2022.

MACEACHREN, A. M.; DIBIASE, D. Animated Maps of Aggregate Data: Conceptual and Practical Problems. **Cartography and Geographic Information Systems**, v. 18, n. 4, p. 221–229, 1991. Taylor & Francis Group. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1559/152304091783786790>>. Acesso em: 2/3/2022.

MAK, K.; COULSON, M. R. C. Map-user response to computer-generated choropleth maps: Comparative experiments in classification and symbolization. **Cartography and Geographic Information Systems**, v. 18, n. 2, p. 109–124, 1991.

MELO, C. M.; TORRALBA, A.; GUIBAS, L.; et al. Next-generation deep learning based on simulators and synthetic data. **Trends in Cognitive Sciences**, v. 26, n. 2, p. 174–187, 2022. Elsevier Ltd.

PILLAY, L.; SCHAAB, G.; COETZEE, S.; RAUTENBACH, V. A comprehensive workflow for automating thematic map geovisualization from univariate big geospatial point data. **Proceedings of the ICA**, 2019.

PROVIN, R. W. The perception of numerosness on dot maps. **American Cartographer**, v. 4, n. 2, p. 111–125, 1977.

ROTH, R. E. Cartographic interaction primitives: Framework and synthesis. **Cartographic Journal**, v. 49, n. 4, p. 376–395, 2012.

SADAHIRO, Y. Cluster perception in the distribution of point objects. **Cartographica**, v. 34, n. 1, p. 49–61, 1997.

SLOCUM, T. A. Predicting visual clusters on graduated circle maps. **American Cartographer**, v. 10, n. 1, p. 59–72, 1983.

SLOCUM, T. A.; MCMASTER, R. B.; KESSLER, F. C.; HOWARD, H. H. **Thematic Cartography and Geovisualization**. Pearson Prentice Hall, 2008.

SŁOMSKA-PRZECH, K.; PANECKI, T.; POKOJSKI, W. Heat maps: Perfect maps for quick reading? comparing usability of heat maps with different levels of generalization. **ISPRS International Journal of Geo-Information**, v. 10, n. 8, p. 562, 2021. Multidisciplinary Digital Publishing Institute. Disponível em: <<https://www.mdpi.com/2220-9964/10/8/562/htm>>. Acesso em: 10/5/2022.

SUCHAN, T. A.; BREWER, C. A. Qualitative Methods for Research on Mapmaking and Map Use. **Professional Geographer**, v. 52, n. 1, p. 145–154, 2000.

What Are Synthetic Data? .Disponível em: <<https://www.census.gov/about/what/synthetic-data.html>>. Acesso em: 6/7/2022.

WICKENS, C. D.; HELTON, W. S.; HOLLANDS, J. G.; BANBURY, S. Engineering psychology and human performance. **Engineering Psychology and Human Performance**, p. 1–566, 2021. Taylor and Francis. Disponível em: <<https://www.taylorfrancis.com/books/mono/10.4324/9781003177616/engineering-psychology-human-performance-christopher-wickens-william-helton-justin-hollands-simon-banbury>>. Acesso em: 6/7/2022.

## 5. GENERAL CONCLUSIONS

No primeiro estudo foi possível quantificar as características de suavidade e continuidade de fenômenos pontuais volumosos e, ainda, verificar aderência dos valores encontrados com a capacidade de reconhecimento visual de cartógrafos experientes. Encontramos limitações no julgamento para definir o ordenamento das distribuições de dados geográficos conforme a variação das características de suavidade e continuidade. Vimos que o tempo de experiência do cartógrafo não influencia o julgamento sobre o nível de continuidade e de suavidade de um fenômeno geográfico. Assim, identificamos que havia a necessidade de se gerar uma solução de geoinformação que padronizasse a escolha da técnica de representação temática mais adequada, segundo a variação das condições de suavidade e de continuidade, conforme propuseram MacEachren e Dibiase (1991).

Os resultados do primeiro estudo nos instigam a continuar em um segundo estudo, no qual gerou-se uma solução computacional que cria dados sintéticos com diferentes valores de suavidade e continuidade. Isso feito, haveria subsídio à proposição de novas entrevistas - além daquelas do primeiro estudo - que

consolidassem nosso conhecimento. Assim, no segundo estudo desenvolvemos tal solução computacional que gerava dados pontuais sintéticos volumosos. O desafio residia em transformar a distribuição dos dados para que eles apresentassem as características de suavidade conforme as desejadas, i.e., com total controle do operador. Concluimos que a geração dos dados sintéticos - conforme feito na solução computacional - favorece nosso objetivo final, uma vez que era possível controlar os valores de suavidade e continuidade do conjunto de dados. Assim, partimos para o terceiro estudo, em posse das ferramentas necessárias para se determinar os limiares de influência da suavidade e a continuidade na decisão do método de representação temática mais adequado.

Dessa forma, nesta tese ratificamos que o modelo de MacEachren e Dibiase (1991), baseado nas metáforas visuais, funciona para o contexto do Geo Big Data, especificamente quando há grande velocidade na geração de dados que são volumosos. Apesar de ter sido gerado em um contexto tecnológico completamente diferente daquele no qual esta tese foi concebida, o modelo proposto por estes autores é reproduzível e fornece subsídio consistente à tomada de decisão dos cartógrafos. Especificamente, encontramos indícios empíricos de que a teoria das metáforas visuais de MacEachren e Dibiase (1991) é um método confiável à seleção de técnicas de representação temática de agrupamentos de ocorrência de fenômenos pontuais em mapas, considerando-se o novo contexto tecnológico e as injunções que utilizamos na criação dos testes.

Nesse sentido, o ranqueamento da adequabilidade das técnicas de mapeamento temático ao contexto de representação de clusters, demonstrou que a suavidade e a continuidade são indicadores fidedignos da descrição da natureza dos fenômenos. Essa condição vincula a suavidade e a continuidade ao status de marcadores representativos que, de certa forma, cooperam com o cartógrafo na decisão do projeto do mapa temático. Nosso modelo de regressão comprovou essa hipótese, considerando-se os dados que foram testados, as configurações de suavidade e continuidade que foi gerado e, também, do treinamento dado pelas respostas dos usuários que entrevistamos.

Entretanto, aqui encontramos que a continuidade é menos competente na determinação do método de representação temática mais adequado. Por conseguinte, é, também, menos competente ao descrever a natureza dos fenômenos geográficos, quando comparada à suavidade. Os resultados que

ratificam essa conclusão estão vinculados ao terceiro estudo que fizemos, especificamente ressaltados pela regressão múltipla. Dessa forma, ao definir a contribuição da suavidade e da continuidade na escolha da técnica de representação temática mais adequada à visualização de agrupamentos em dados pontuais volumosos, cumprimos com o objetivo geral desta tese.

Durante o desenvolvimento desta pesquisa, entende-se que estudos futuros deverão testar maior quantidade de usuários, bem como, maior quantidade de configurações de suavidade e continuidade para a geração dos mapas temáticos dos testes. Tais estudos deverão buscar avaliar diferentes espectros de usuários, i.e., pessoas com e sem conhecimento em cartografia, para que haja ampliação do entendimento de como cada mapa temático pode interferir na leitura da ocorrência de dados pontuais volumosos. Conforme recomenda a literatura (GRIFFIN *et al.*, 2017) é importante que as entrevistas ocorram em ambiente realístico.

Ainda, estudos futuros deverão focalizar atenção na otimização do algoritmo para geração de dados sintéticos. Notamos que a convergência para o valor de suavidade pretendido do algoritmo que foi gerado, tem um custo computacional elevado do ponto de vista do tempo de processamento. Essa condição pode ter sido favorecida pela forma de implementação em ambiente *desktop*, ou, também, pela utilização de máquinas com poucos recursos avançados de *hardware*. Para melhorar esse cenário, as pesquisas futuras podem desenvolver soluções baseadas em processamento paralelo, ou processamento GPU (*Graphics Processing Units*), como possíveis soluções à incompetência dos hardwares, sem incrementar grandes custos às demais configurações dos computadores pessoais.

Durante as entrevistas, observamos que há necessidade de otimização da interface para teste com usuários desenvolvido em ambiente *Web*. A solução desenvolvida para que o teste ocorresse em ambiente *Web* se mostrou adequada do ponto de vista da plataforma escolhida, uma vez que, ao disparar o convite na internet, rapidamente se obteve um número satisfatório de respostas. Entretanto, os resultados poderiam ter sido melhores caso a limitação da execução não tivesse sido vinculada à computadores com telas 15 polegadas. Neste caso, não foi permitido que usuários de celulares fizessem o teste. Entende-se que, em um mundo no qual a utilização de dispositivos móveis é predominante, especialmente no uso dos mapas (GRIFFIN e FABRIKANT 2012), caso a interface do teste permitisse a participação de usuários destes dispositivos, atingiríamos uma maior

quantidade de pessoas, com características também diversas. Entende-se que esta condição pode ser facilmente resolvida por pesquisas futuras, que, tendo realizado essa transformação, possibilitarão que os testes se ampliem e sejam realizados em ambientes realísticos. Do ponto de vista da ampliação da quantidade de pessoas que se sentem motivadas a colaborar com entrevistas dessa natureza, entende-se que as pesquisas futuras também podem favorecer a criação de cenários e tarefas embutidas em contextos de uso gamificados.

Considerando cenários mais realísticos e diversos àqueles que foram testados com os dados descritos nos experimentos, os pesquisadores que desenvolverão suas investigações nesta linha devem pensar em testar conjuntos de dados diferentes para ver se há repetição dos padrões que observamos e preencher a lacuna que deixamos de se testar a variedade dos dados no contexto do Big Data. Devem, também, testar maior quantidade de tarefas de leitura de mapas (BOARD, 1978) desde as menos complexas até as mais complexas, observando o impacto da variação da suavidade e da continuidade na seleção de técnicas de mapeamento temático destinadas à execução dessas tarefas em específico. Esta seria uma abordagem funcional, baseada no raciocínio de Robinson (1952).

Por fim, as pesquisas futuras podem testar novas variáveis além da suavidade e da continuidade. Poderiam ser foco destes estudos o tamanho dos clusters, o número de clusters ou variáveis de complexidade, que são medidas em função da dificuldade que representa uma determinada variável à leitura do mapa (OLSON, 1975). Poderiam ser incorporadas técnicas de mensuração do esforço cognitivo demandado na leitura dos agrupamentos, segundo as diferentes técnicas de representação temática. Com esta última recomendação, aliando-se ao modelo de determinação da técnica mais adequada que elaboramos nesta tese - baseado na modelo de regressão, poderia se ter mais certeza da viabilidade da técnica selecionada, uma vez que vincularia-se ao ponto de vista estatístico da distribuição dos dados, o ponto de vista do usuário e seu esforço em ler os mapas.

## 6. REFERÊNCIAS

BOARD, C. MAP READING TASKS APPROPRIATE IN EXPERIMENTAL STUDIES IN CARTOGRAPHIC COMMUNICATION. <https://doi.org/10.3138/AG15-V252-3726-W346>, v. 15, n. 1, p. 1–12, 2006. University of Toronto Press . Disponível em: <<https://www.utpjournals.press/doi/10.3138/AG15-V252-3726-W346>>. Acesso em: 6/7/2022.

FORREST, D. Geographic Information: Its Nature, Classification, and Cartographic Representation. <https://doi.org/10.3138/D2U8-7218-6G57-4127>, v. 36, n. 2, p. 31–53, 2006. University of Toronto Press . Disponível em: <<https://www.utpjournals.press/doi/10.3138/D2U8-7218-6G57-4127>>. Acesso em: 6/7/2022.

GOODCHILD, M. F. The quality of big (geo)data: <https://doi.org/10.1177/2043820613513392>, v. 3, n. 3, p. 280–284, 2013. SAGE Publications Sage UK: London, England. Disponível em: <[https://journals.sagepub.com/doi/full/10.1177/2043820613513392?casa\\_token=9ia4u2PWpz4AAAAA%3A4x45D8-0s5\\_ZqdRLZrBuJheovtFXiqBgvgn5Xeg5AcqZoizgFkNtREnihMg1p\\_pQidx0CzHyplvr bA](https://journals.sagepub.com/doi/full/10.1177/2043820613513392?casa_token=9ia4u2PWpz4AAAAA%3A4x45D8-0s5_ZqdRLZrBuJheovtFXiqBgvgn5Xeg5AcqZoizgFkNtREnihMg1p_pQidx0CzHyplvr bA)>. Acesso em: 5/7/2022.

GRIFFIN, A. L.; FABRIKANT, S. I. More maps, more users, more devices means more cartographic challenges. **Cartographic Journal**, Nov. 2012.

GRIFFIN, A. L.; ROBINSON, A. C.; ROTH, R. E. Envisioning the future of cartographic research. **International Journal of Cartography**, 30. Oct. 2017a. Taylor and Francis Ltd. Disponível em: <<https://pennstate.pure.elsevier.com/en/publications/envisioning-the-future-of-cartographic-research>>. Acesso em: 2/3/2022.

LANEY, D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. **Application Delivery Strategies**, 2001.

LI, S.; DRAGICEVIC, S.; CASTRO, F. A.; et al. Geospatial big data handling theory and methods: A review and research challenges. **ISPRS Journal of Photogrammetry and Remote Sensing**, 2016a.

MACEACHREN, A. M.; DIBIASE, D. Animated maps of aggregate data: Conceptual and practical problems. **Cartography and Geographic Information Systems**, v. 18, n. 4, p. 221–229, 1991.

ROBINSON, A. C.; DEMŠAR, U.; MOORE, A. B.; BUCKLEY, A.; JIANG, B.; FIELD, K.; KRAAK, M. J.; et al. Geospatial big data and cartography: research challenges and opportunities for making maps that matter. **International Journal of Cartography**, v. 3, n. sup1, p. 32–60, 2017. Taylor and Francis Ltd.

ROBINSON, A. H. The Look of Maps: An Examination of Cartographic Design. **The American Cartographer**, v. 13, n. 3, p. 280–280, 1986. Disponível em: <<https://www.tandfonline.com/doi/full/10.1559/152304086783899881>>. Acesso em: 6/7/2022.

ROBINSON, A. H.; MORRISON, J. L.; MUEHRCKE, P. C.; KIMERLING, A. J.; GUPTILL, S. C. Elements of Cartography: Sixth Edition. , v. 6, n. 2, p. 674, 1995. Wiley. Disponível em: <[http://journals.lww.com/soilsci/Abstract/1960/08000/Elements\\_of\\_Cartography\\_\\_Second\\_Edition.22.aspx](http://journals.lww.com/soilsci/Abstract/1960/08000/Elements_of_Cartography__Second_Edition.22.aspx)>. Acesso em: 6/7/2022.

ROTH, R. E.; KELLY, M.; UNDERWOOD, N.; et al. Interactive & Multiscale Thematic Maps: A Preliminary Study. **AbICA**, v. 1, p. 315, 2019. Copernicus GmbH. Disponível em: <<https://ui.adsabs.harvard.edu/abs/2019AbICA...1..315R/abstract>>. Acesso em: 23/6/2022.

SAGIROGLU, S.; SINANC, D. Big data: A review. **Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013**, p. 42–47, 2013.

SCHOLZ, S.; KNIGHT, P.; ECKLE, M.; MARX, S.; ZIPF, A. Volunteered Geographic Information for Disaster Risk Reduction—The Missing Maps Approach and Its Potential within the Red Cross and Red Crescent Movement. **Remote Sensing 2018, Vol. 10, Page 1239**, v. 10, n. 8, p. 1239, 2018. Multidisciplinary Digital Publishing Institute. Disponível em: <<https://www.mdpi.com/2072-4292/10/8/1239/htm>>. Acesso em: 5/7/2022.

SHEA, K. S.; MCMASTER, R. B. Cartographic Generalization in a Digital Environment: When and How to Generalize. **Proceedings Auto-Carto 9**.

SLOCUM, T. A.; MCMASTER, R. B.; KESSLER, F. C.; HOWARD, H. H. **Thematic Cartography and Geovisualization**. Pearson Prentice Hall, 2008.

SLUTER, C. R. Uma abordagem sistêmica para o desenvolvimento de projeto cartográfico como parte do processo de comunicação cartográfica. **Portal de Cartografia das Geociências**, v. 1, n. 1, p. 01–20, 2008. Disponível em: <<http://www.uel.br/revistas/uel/index.php/portalcartografia/article/view/1365/1084>>.

## 7. APÊNDICES

### APÊNDICE 1 – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Pesquisa: Efeitos da suavidade e continuidade na visualização de agrupamentos em mapas temáticos.

O(A) Sr(a) está sendo convidado a participar do projeto de pesquisa " Efeitos da suavidade e continuidade na visualização de agrupamentos em mapas temáticos", cujo pesquisador responsável é Prof<sup>a</sup>. Dra. Silvana Philippi Camboim.

1. O(A) Sr(a) está sendo convidado porque é importante para nossa pesquisa entender como os diferentes mapas temáticos aplicados a diferentes dados facilitam a sua percepção de agrupamentos em mapas temáticos.
2. O principal objetivo do projeto é avaliar as técnicas de representação cartográfica temática quando aplicadas a diferentes dados geográficos.
3. Caso aceite participar da pesquisa sua participação consiste em: Preencher um formulário com algumas perguntas relacionadas a região que reside e sua experiência em utilizar mapas; Após, será apresentado alguns mapas e será requisitado para desenhar na tela, utilizando *mouse* ou "*touchpad*", os agrupamentos (regiões no mapas que parecem iguais) nos mapas.
4. O tempo previsto para responder os formulários e executar as tarefas é de 20 minutos.
5. O(A) Sr(a). tem plena liberdade de recusar-se a participar ou retirar seu consentimento, em qualquer fase da pesquisa.
6. Caso após preencher e enviar o questionário o(a) Sr(a) desejar retirar seu consentimento para uso dos dados, deve entrar em contato com o

pesquisador responsável, que lhe enviará resposta confirmando ciência de sua decisão. Para possibilitar a identificação e a exclusão das informações, imprima esse TCLE e apresente ao pesquisador responsável.

7. Toda pesquisa com seres humanos envolve riscos aos participantes. Nesta pesquisa os riscos para o(a) Sr.(a) são desconforto, cansaço ou tédio, principalmente relacionado à ansiedade de utilizar a plataforma para desenhar os agrupamentos e devido ao tempo total para realizar as tarefas, para minimizar estes riscos a plataforma de desenho de agrupamentos será construída de forma a facilitar a realização das tarefas, assim minimizando os riscos inerentes à execução.
8. Também são esperados os seguintes benefícios com esta pesquisa: A construção de mapas temáticos mais eficazes e eficientes.
9. Se julgar necessário, o(a) Sr(a) dispõe de tempo para que possa refletir sobre sua participação, consultando, se necessário, seus familiares ou outras pessoas que possam ajudá-los na tomada de decisão livre e esclarecida.
10. O(A) Sr(a). pode entrar em contato com a pesquisadora responsável Silvana Philippi Camboim a qualquer tempo para informação adicional no endereço [silvanacamboim@gmail.com](mailto:silvanacamboim@gmail.com).
11. O(A) Sr(a). também pode entrar em contato com o Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal do Paraná (CEP/UFPR) e com a Comissão Nacional de Ética em Pesquisa (CONEP), quando pertinente. O CEP/UFPR fica na Rua Padre Camargo nº 285 – 1º andar – 1ª sala à esquerda – CEP 80060-240 – Bairro Alto da Glória – Curitiba – PR, Fone: (41) 3360-7259, E-mail: [cometicasd2@gmail.com](mailto:cometicasd2@gmail.com) e [cometica.saude@ufpr.br](mailto:cometica.saude@ufpr.br).

12. Recomendamos o(a) Sr.(a). imprimir este TCLE e guardá-lo como comprovante de seu consentimento e dos termos aqui descritos, ou fazer download em pdf.

Ao clicar no botão abaixo [Próxima], o(a) Senhor(a) declara que é maior de idade, legalmente capaz e concorda em participar da pesquisa nos termos apresentados neste TCLE, e iniciará a resposta ao questionário. Caso não concorde em participar, apenas feche essa página no seu navegador. Caso desista da participação antes de finalizar o formulário basta não enviar ao final.

CONSENTIMENTO PÓS-INFORMAÇÃO: Li e concordo em participar da pesquisa.

CONTATOS:

Pesquisador responsável: Prof<sup>ª</sup>. Dra. Silvana Philippi Camboim - email: silvanacamboim@gmail.com

Aluno de doutorado: Raphael Gonçalves de Campos - email: raphageoc@gmail.com

Coorientador: João Vitor Meza Bravo - email: jvmbravo@gmail.com

Comitê de Ética em Pesquisa com Seres Humanos do Setor de Ciências da Saúde da UFPR | CEP/SD Rua Padre Camargo, 285 | 1º andar | Alto da Glória | Curitiba/PR | CEP 80060-240 | cometica.saude@ufpr.br – telefone (041) 3360-7259

## APÊNDICE 2 – MAPAS UTILIZADAS NO TESTE DE PERCEPÇÃO

Os nomes dos arquivos são compostos pelos seguintes elementos:

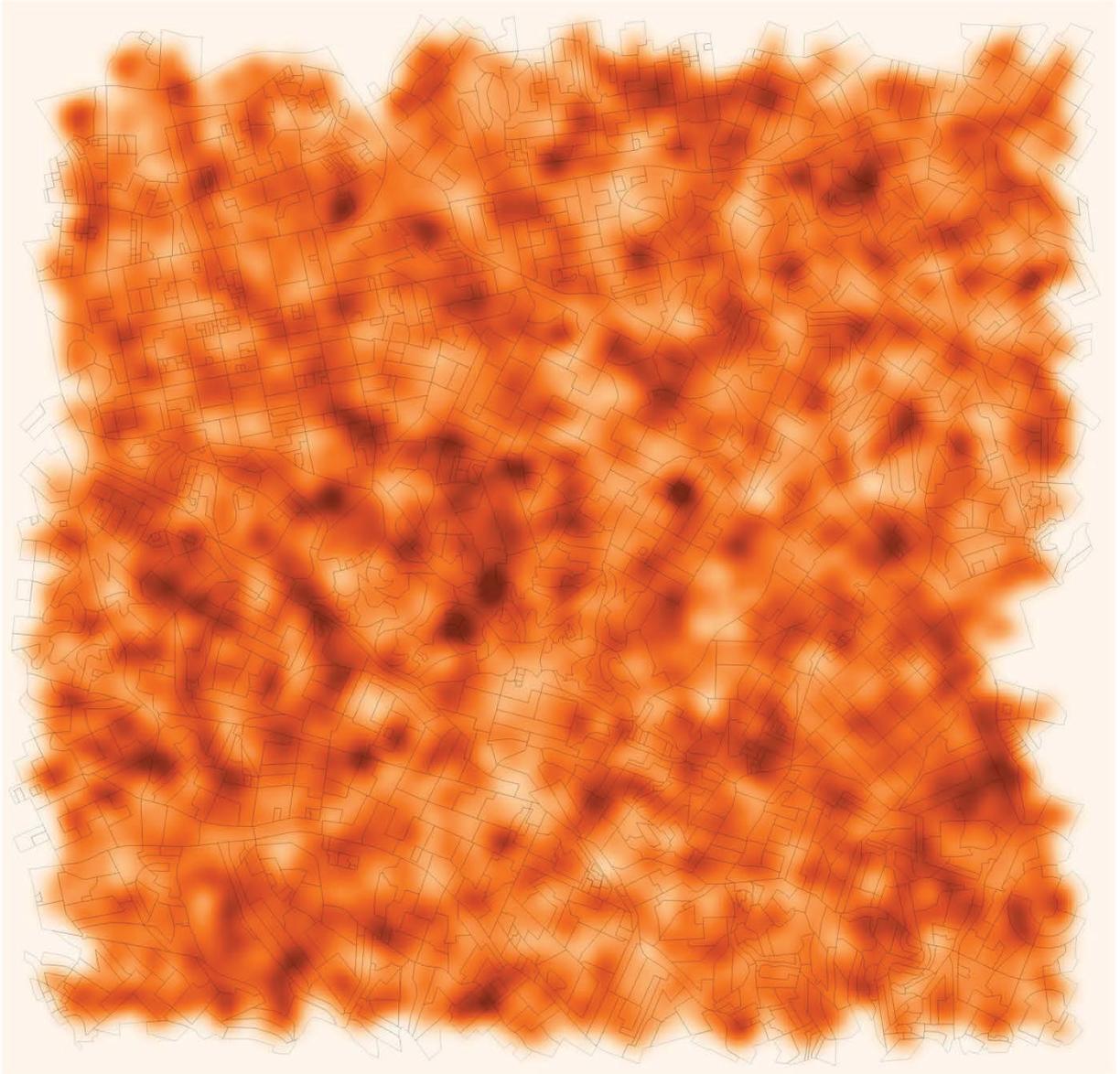
1. Pelo recorte que o mapa foi construído, por exemplo set7.
2. Pela suavidade dos dados, por exemplo s85
3. Pela continuidade dos dados, por exemplo c100.
4. Identificação da técnica:
  - a. calor - Mapas de calor;
  - b. coro - Mapas coropléticos;
  - c. grade - Mapas “chorodot”;
  - d. iso - Mapas isopléticos;
  - e. ptsconta - Mapas de pontos de contagem;
  - f. ptsprop - Mapas de pontos proporcionais;

A lista de imagens utilizadas neste projeto é a seguinte e está vinculada ao número das páginas das imagens abaixo.

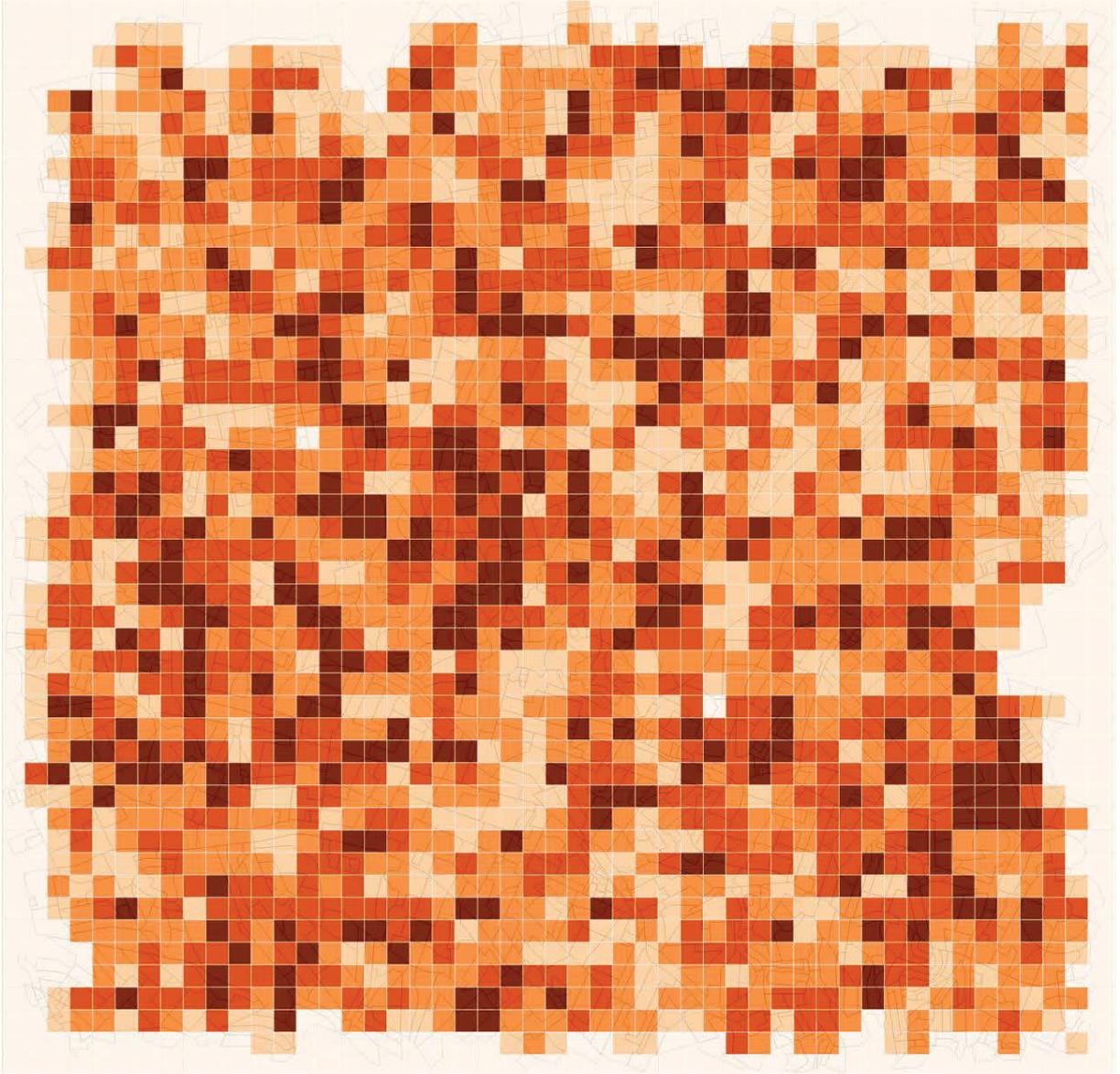
1. set7\_s85c100\_calor (pg. 98)
2. set7\_s85c100\_coro (pg. 99)
3. set7\_s85c100\_grade (pg. 100)
4. set7\_s85c100\_iso (pg. 101)
5. set7\_s85c100\_ptsconta (pg. 102)
6. set7\_s85c100\_ptsprop (pg. 103)
7. set7\_s70c100\_calor (pg. 104)
8. set7\_s70c100\_coro (pg. 105)
9. set7\_s70c100\_grade (pg. 106)
10. set7\_s70c100\_iso (pg. 107)
11. set7\_s70c100\_ptsconta (pg. 108)
12. set7\_s70c100\_ptsprop (pg. 109)
13. set7\_s55c100\_calor (pg. 110)
14. set7\_s55c100\_coro (pg. 111)
15. set7\_s55c100\_grade (pg. 112)
16. set7\_s55c100\_iso (pg. 113)

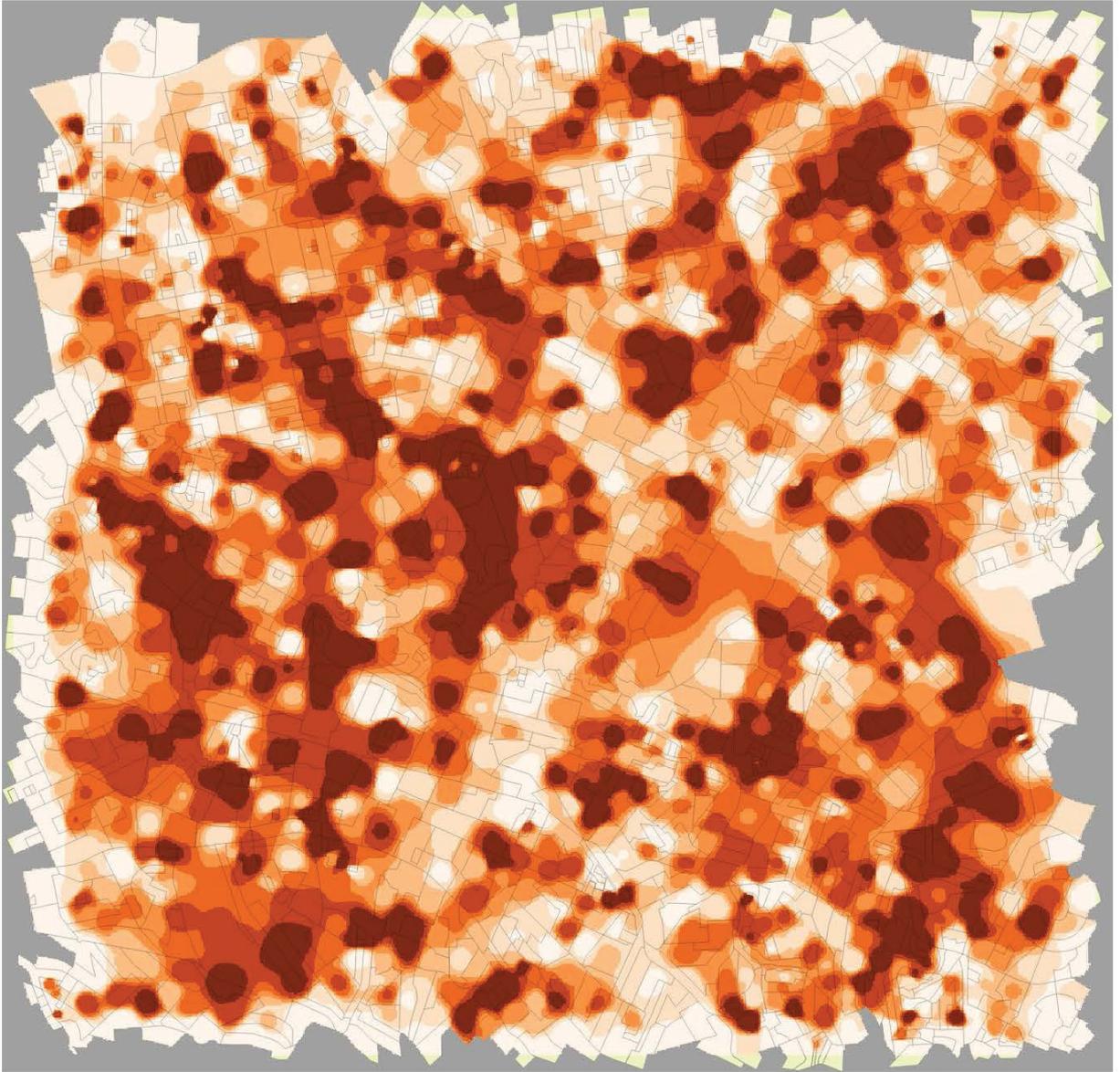
17. set7\_s55c100\_ptsconta (pg. 114)
18. set7\_s55c100\_ptsprop (pg. 115)
19. set8\_s85c60\_calor (pg. 116)
20. set8\_s85c60\_coro (pg. 117)
21. set8\_s85c60\_grade (pg. 118)
22. set8\_s85c60\_iso (pg. 119)
23. set8\_s85c60\_ptsconta (pg. 120)
24. set8\_s85c60\_ptsprop (pg. 121)
25. set8\_s70c60\_calor (pg. 122)
26. set8\_s70c60\_coro (pg. 123)
27. set8\_s70c60\_grade (pg. 124)
28. set8\_s70c60\_iso (pg. 125)
29. set8\_s70c60\_ptsconta (pg. 126)
30. set8\_s70c60\_ptsprop (pg. 127)
31. set8\_s55c60\_calor (pg. 128)
32. set8\_s55c60\_coro (pg. 129)
33. set8\_s55c60\_grade (pg. 130)
34. set8\_s55c60\_iso (pg. 131)
35. set8\_s55c60\_ptsconta (pg. 132)
36. set8\_s55c60\_ptsprop (pg. 133)
37. set9\_s85c30\_calor (pg. 134)
38. set9\_s85c30\_coro (pg. 135)
39. set9\_s85c30\_grade (pg. 136)
40. set9\_s85c30\_iso (pg. 137)
41. set9\_s85c30\_ptsconta (pg. 138)
42. set9\_s85c30\_ptsprop (pg. 139)
43. set9\_s70c30\_calor (pg. 140)
44. set9\_s70c30\_coro (pg. 141)
45. set9\_s70c30\_grade (pg. 142)
46. set9\_s70c30\_iso (pg. 143)
47. set9\_s70c30\_ptsconta (pg. 144)
48. set9\_s70c30\_ptsprop (pg. 145)
49. set9\_s55c30\_calor (pg. 146)
50. set9\_s55c30\_coro (pg. 147)

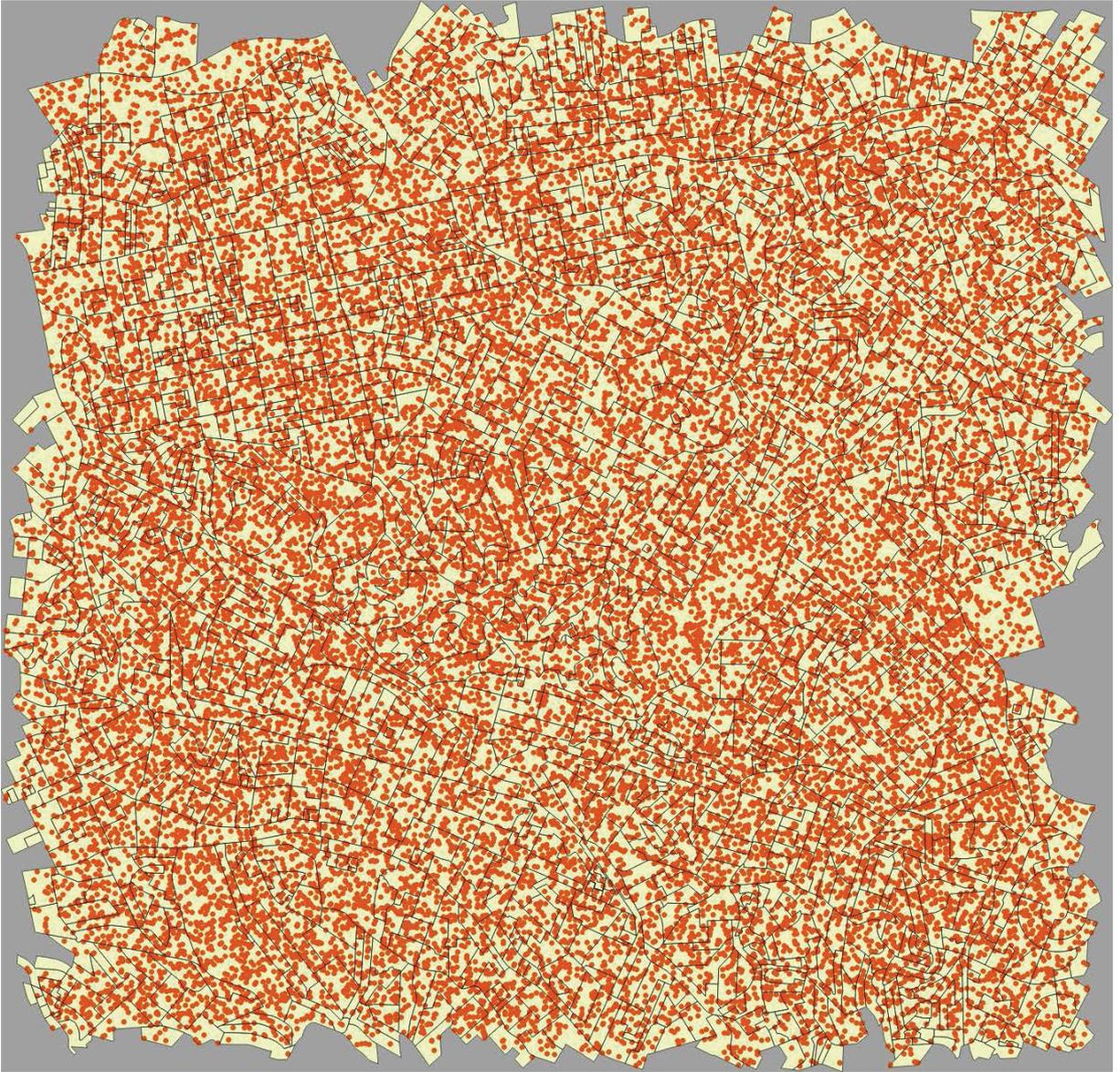
51. set9\_s55c30\_grade (pg. 148)
52. set9\_s55c30\_iso (pg. 149)
53. set9\_s55c30\_ptsconta (pg. 150)
54. set9\_s55c30\_ptsprop (pg. 151)

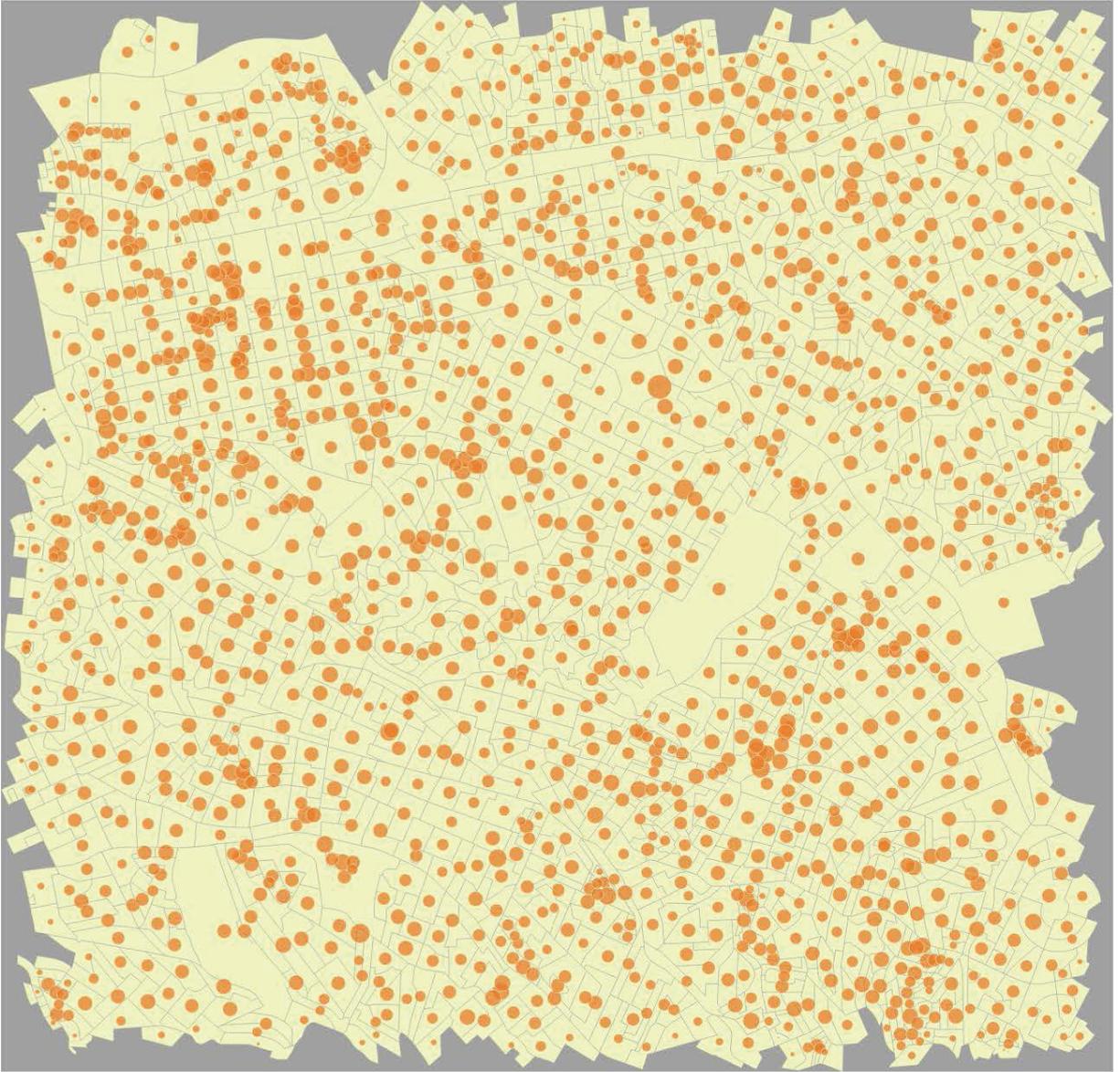


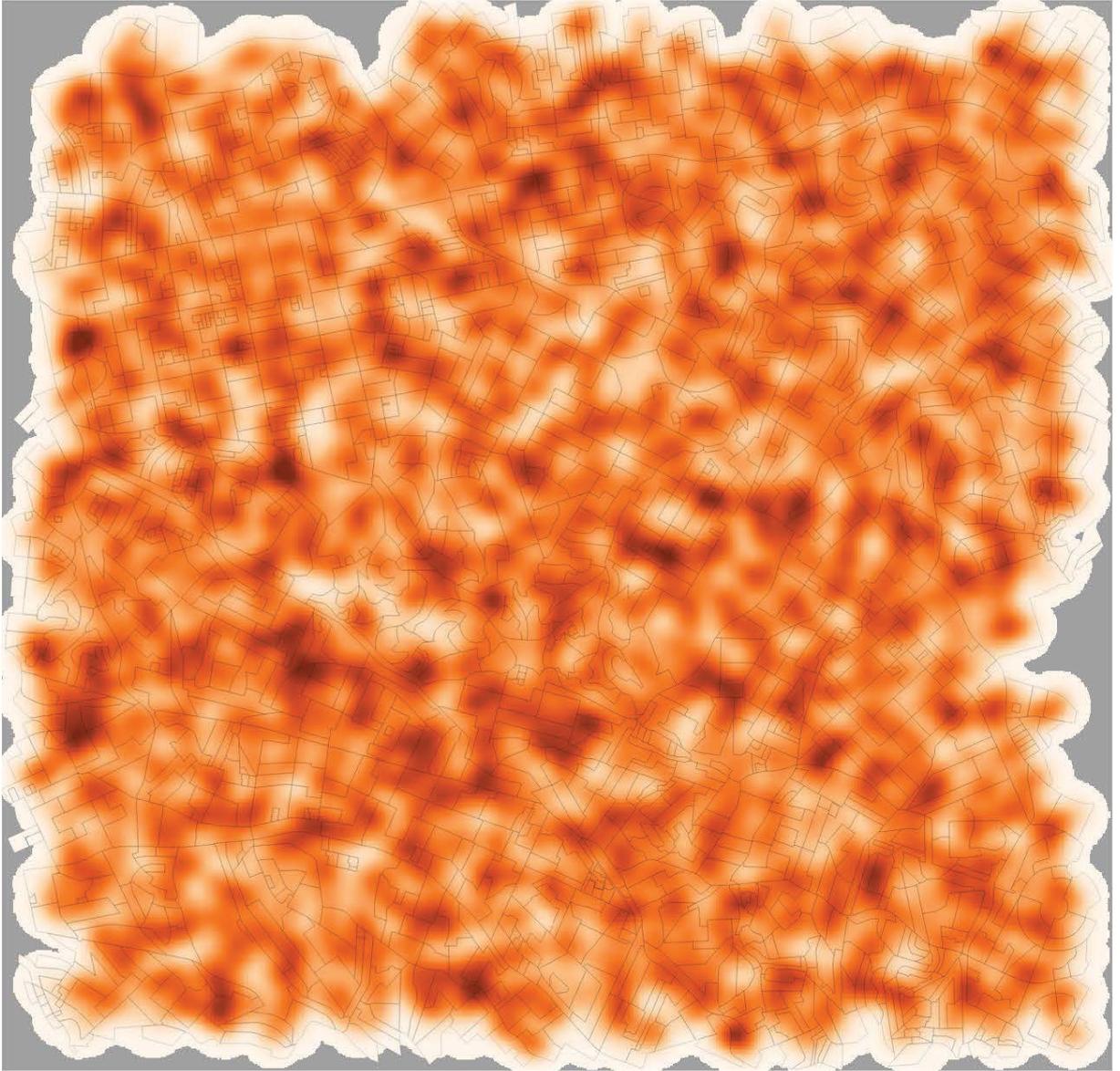




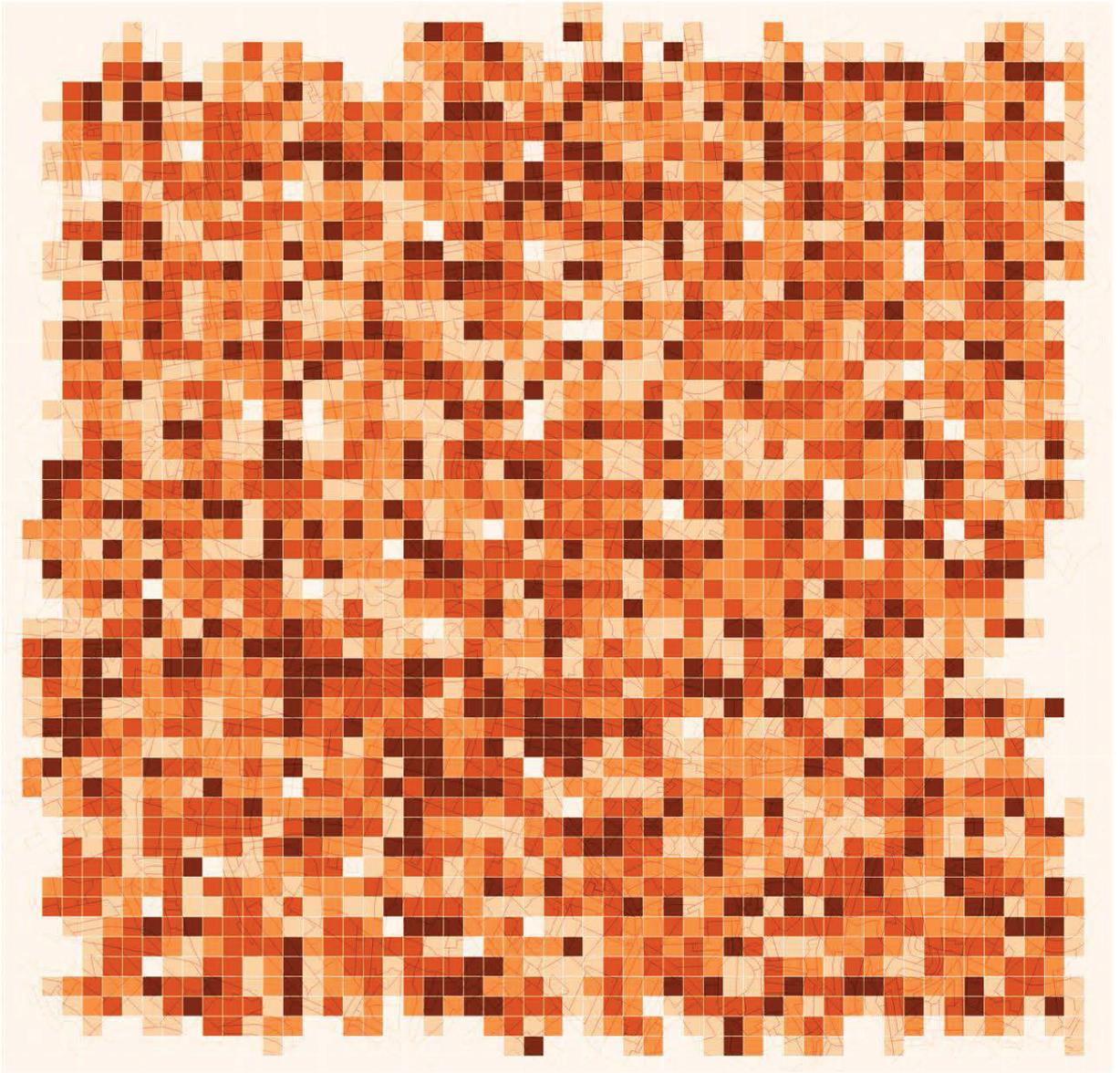


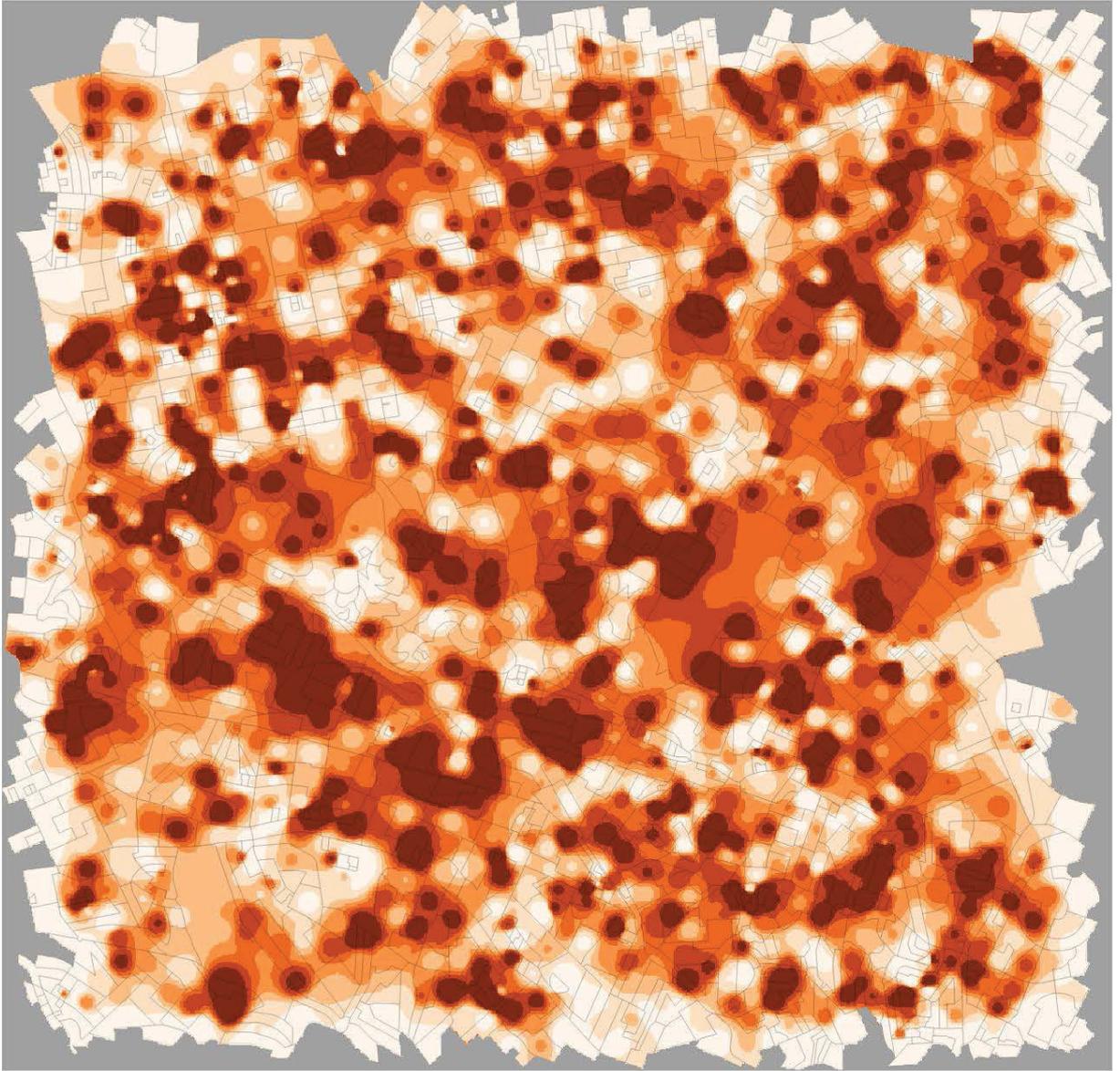


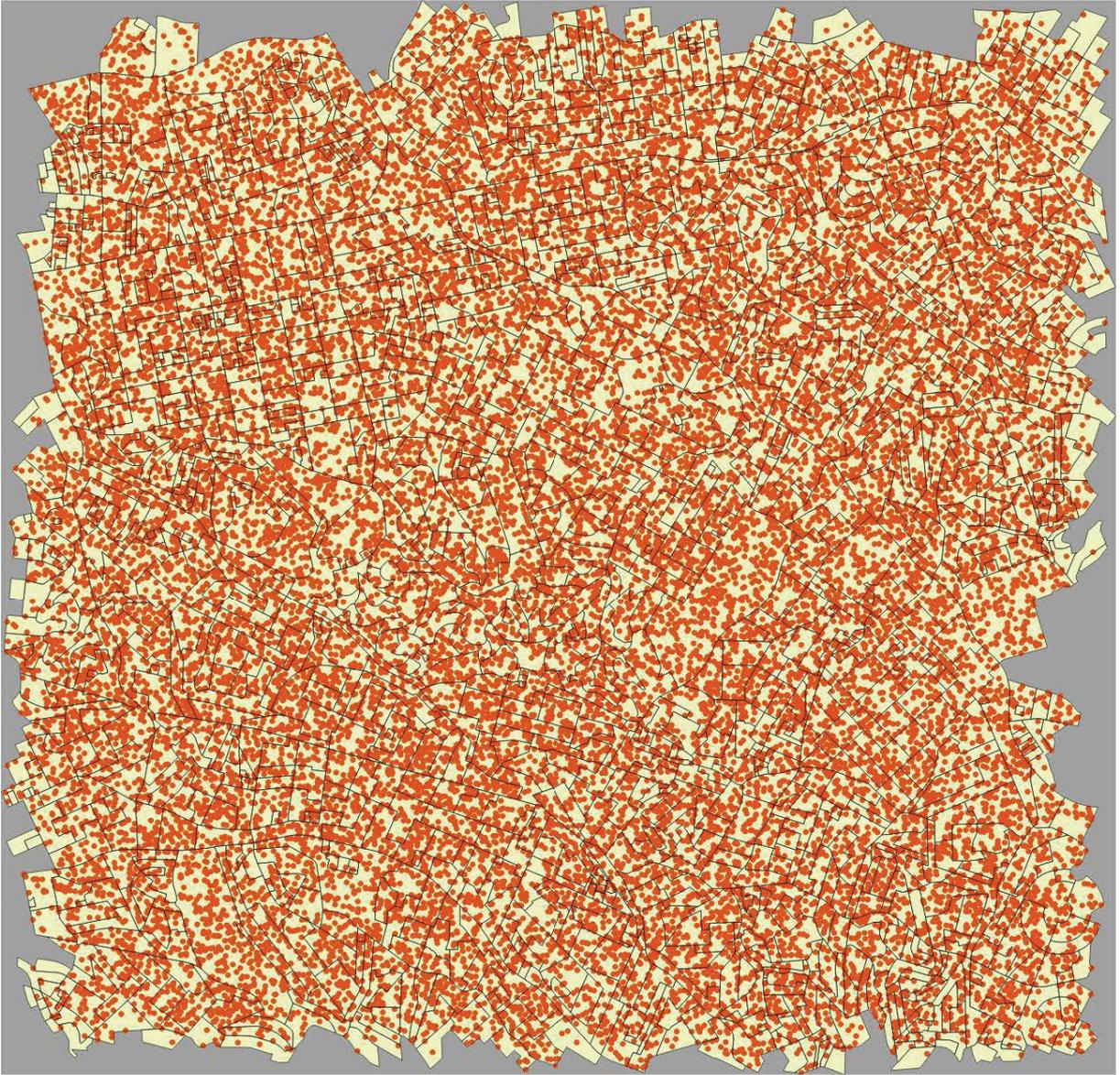


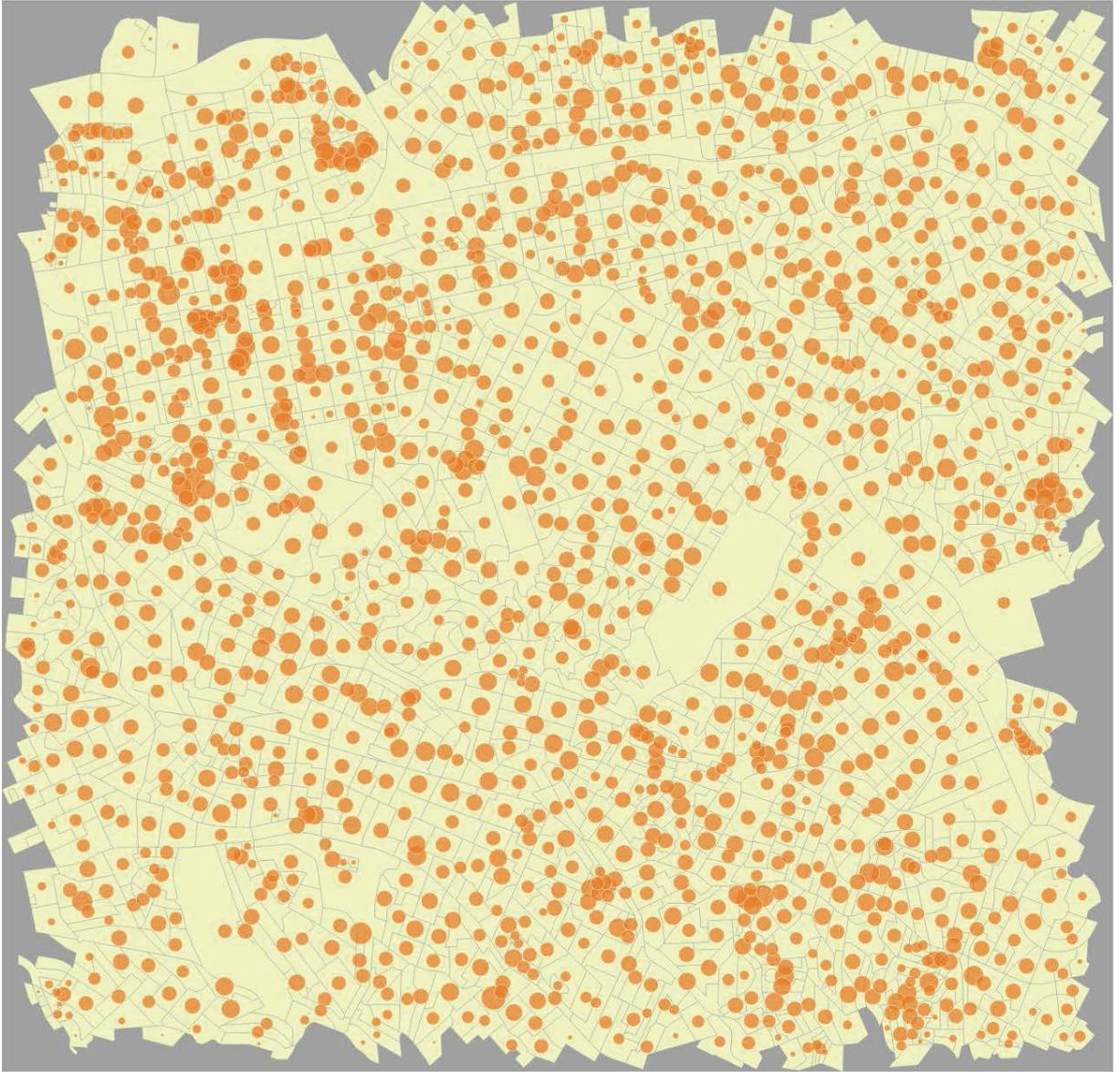


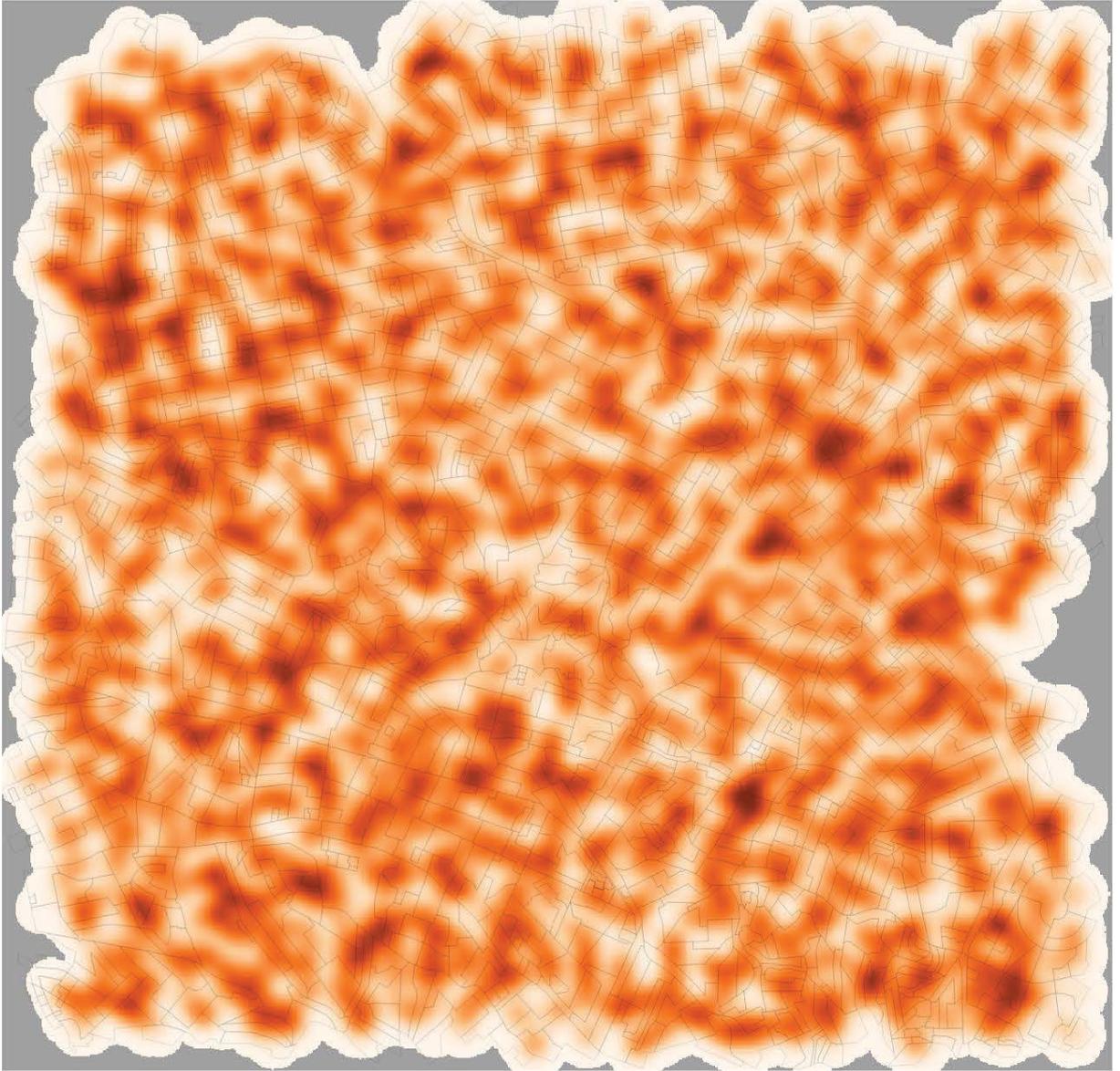




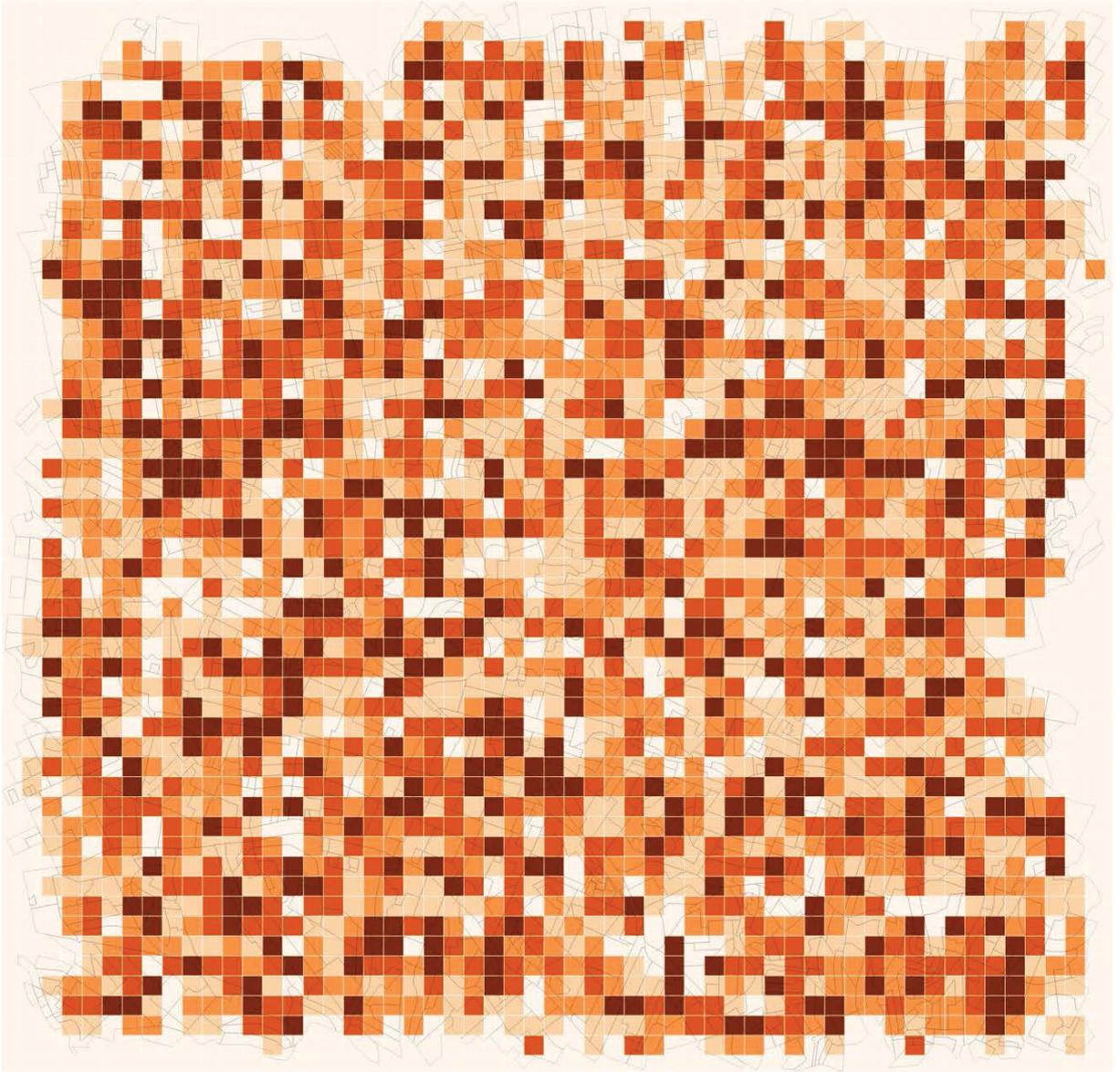


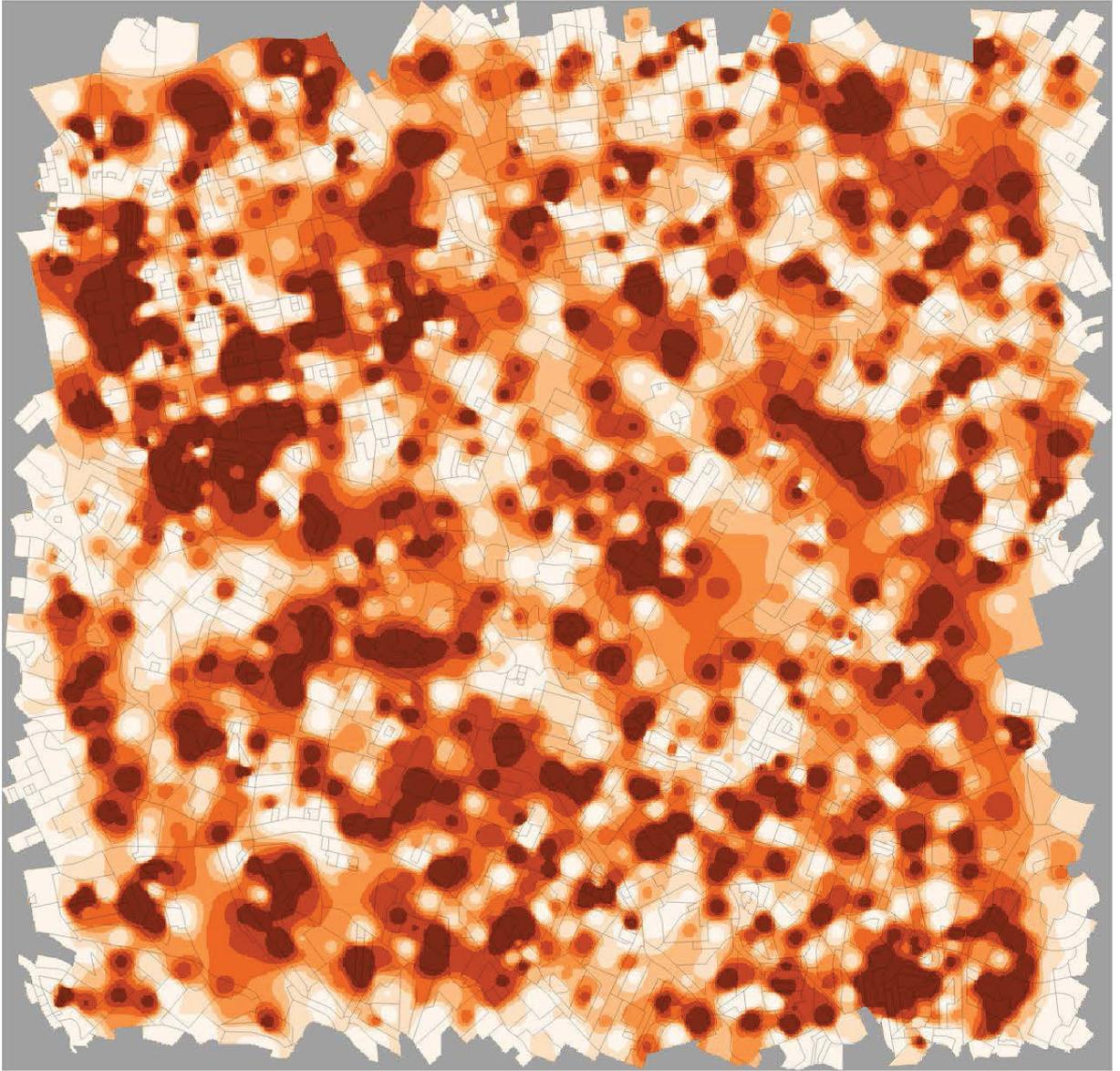


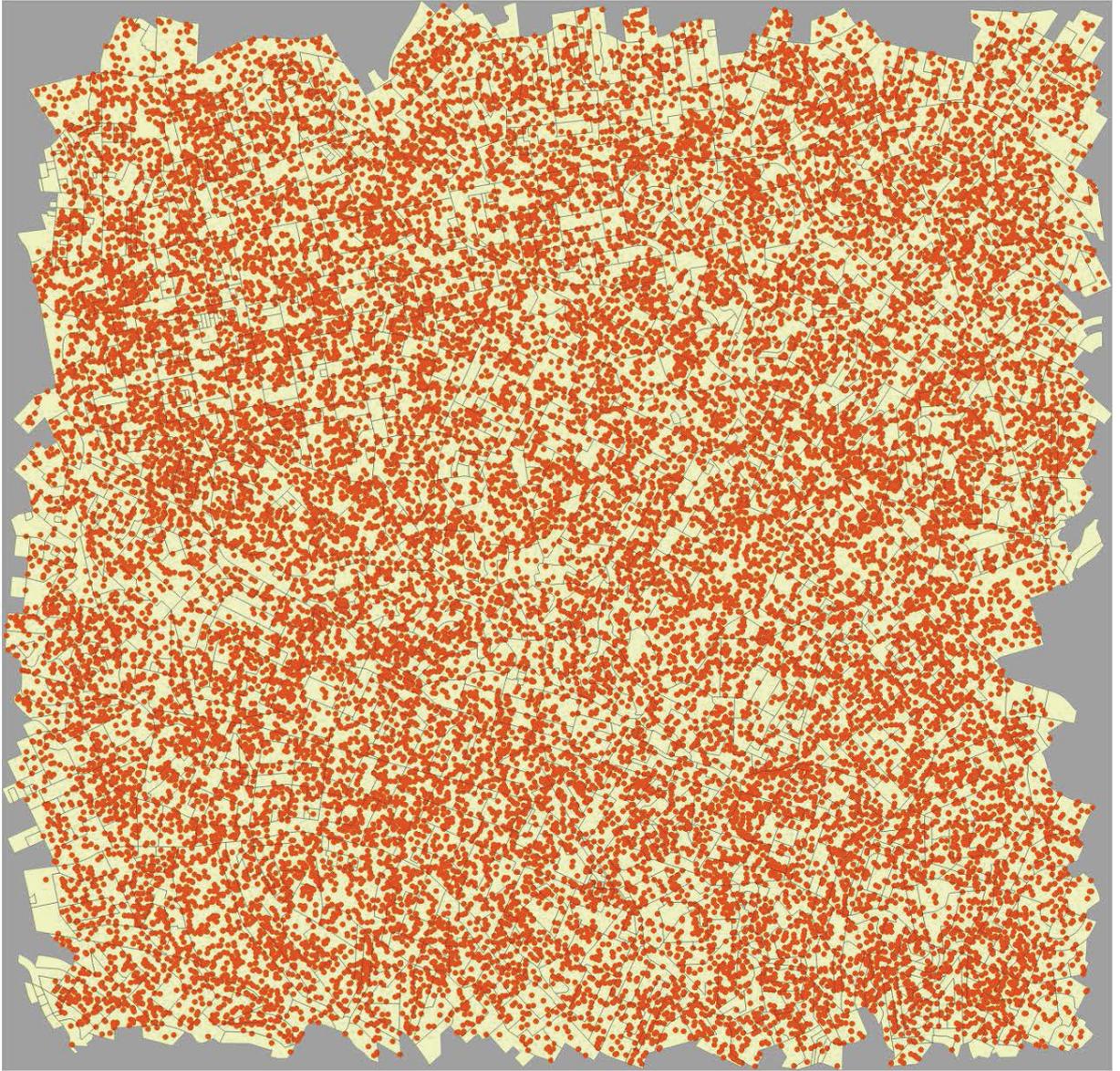


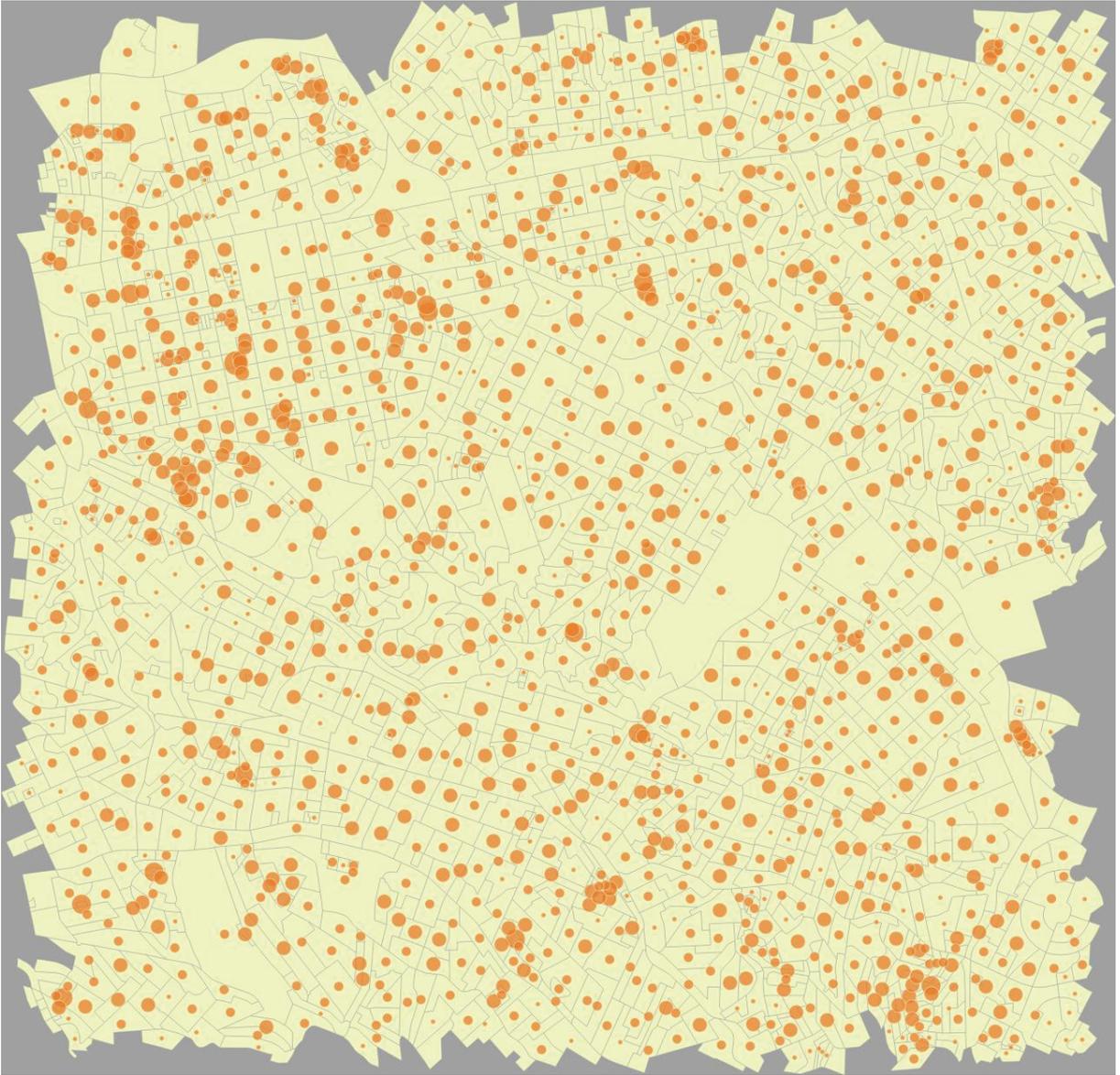


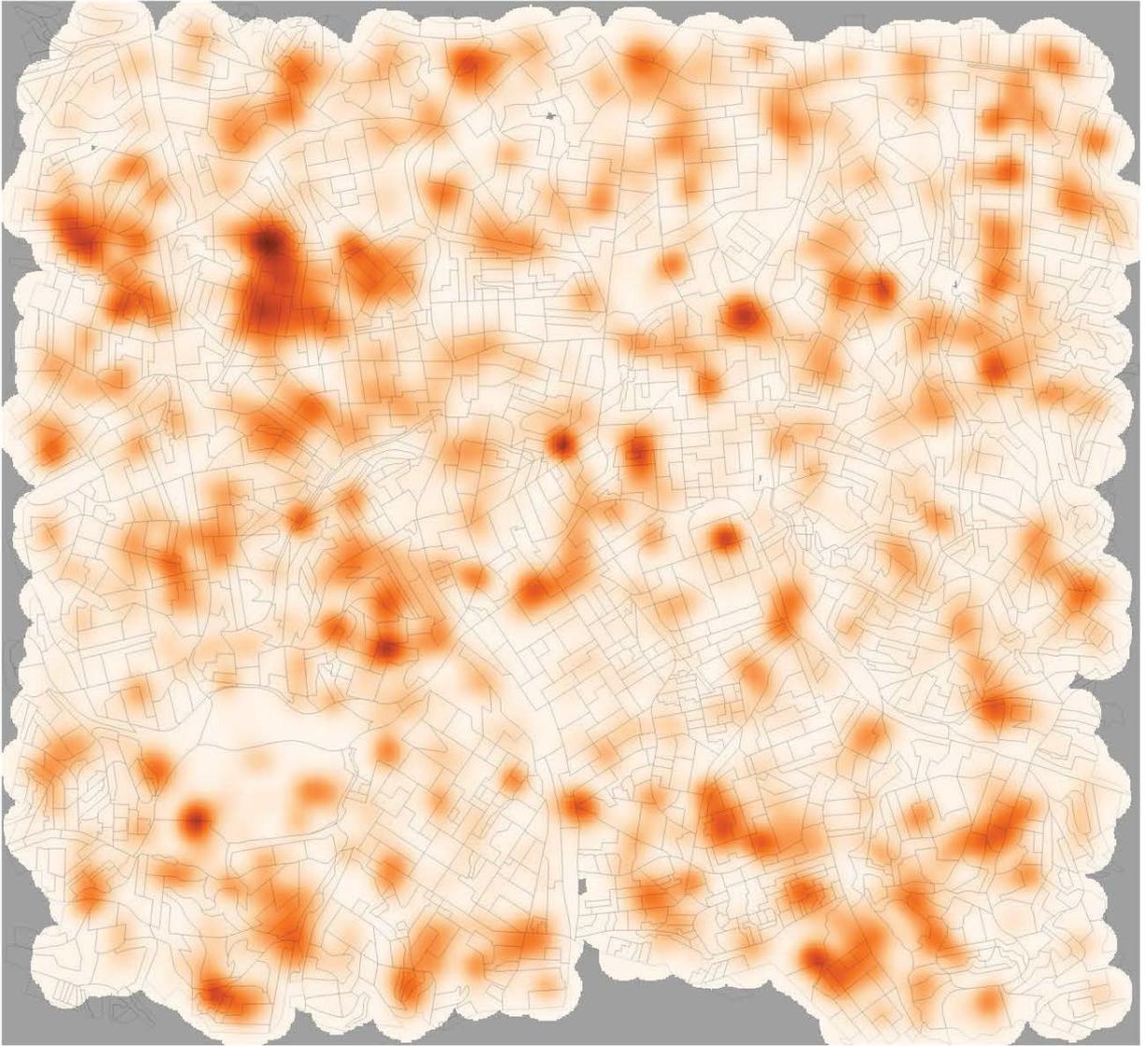




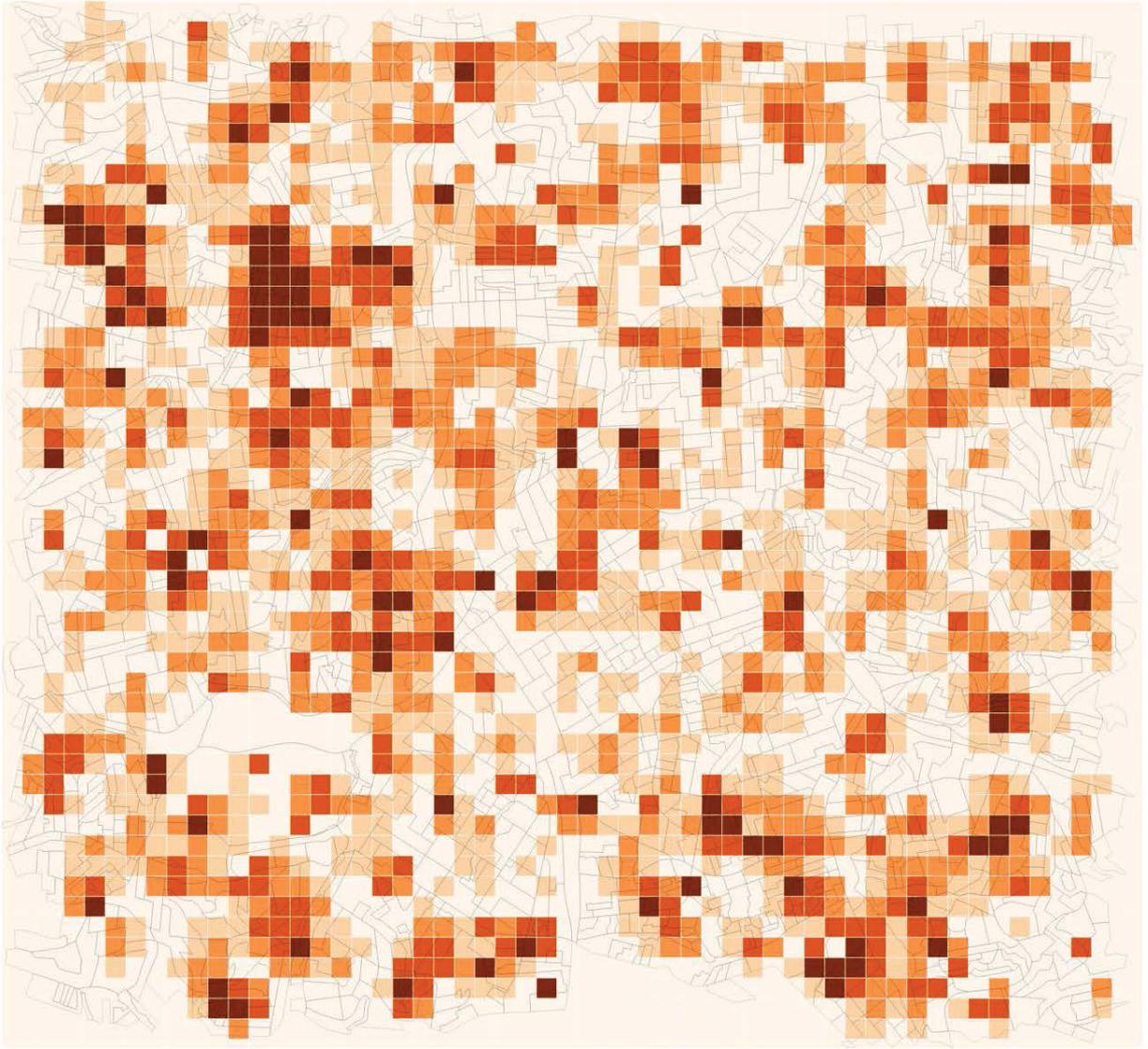


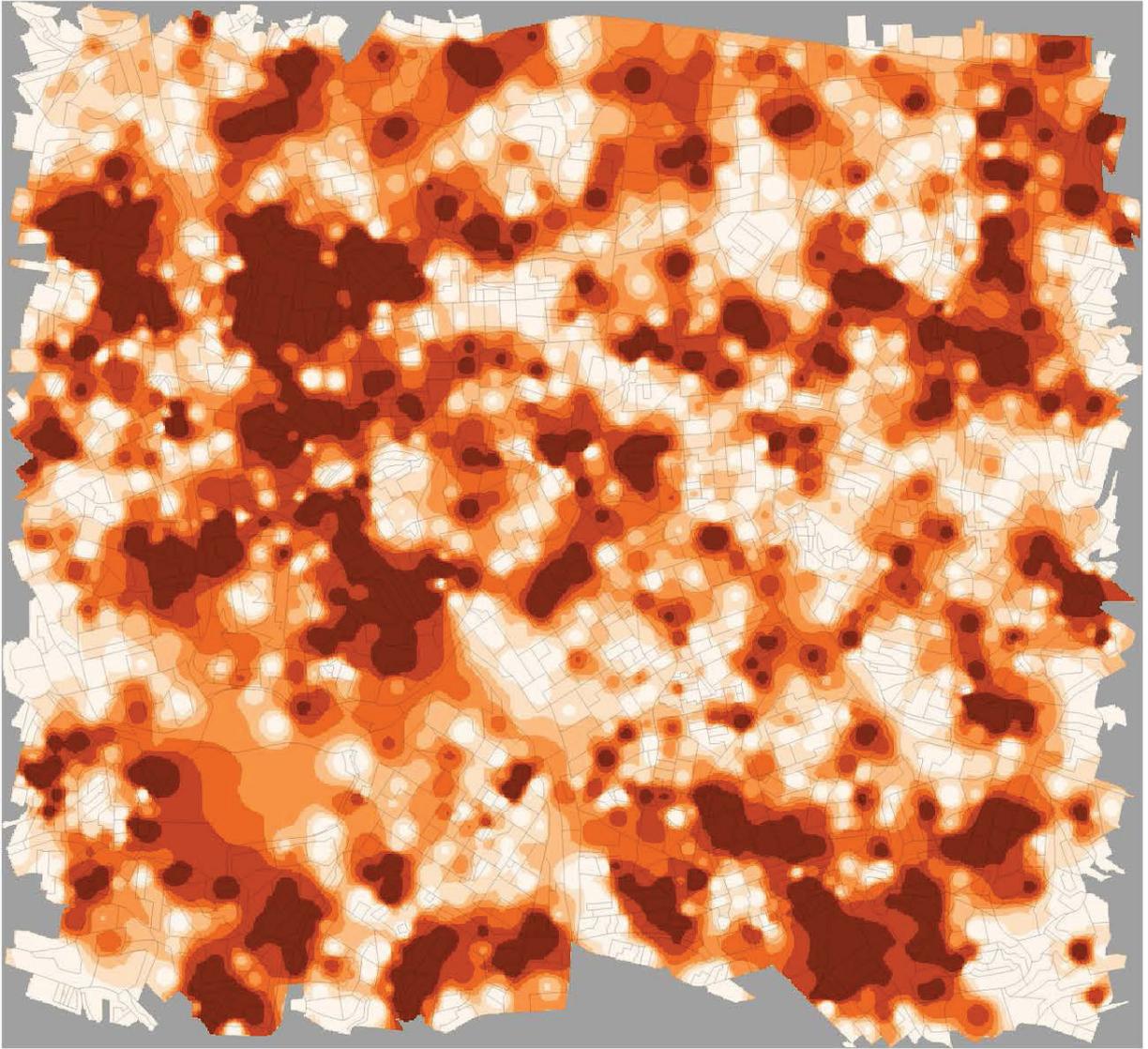


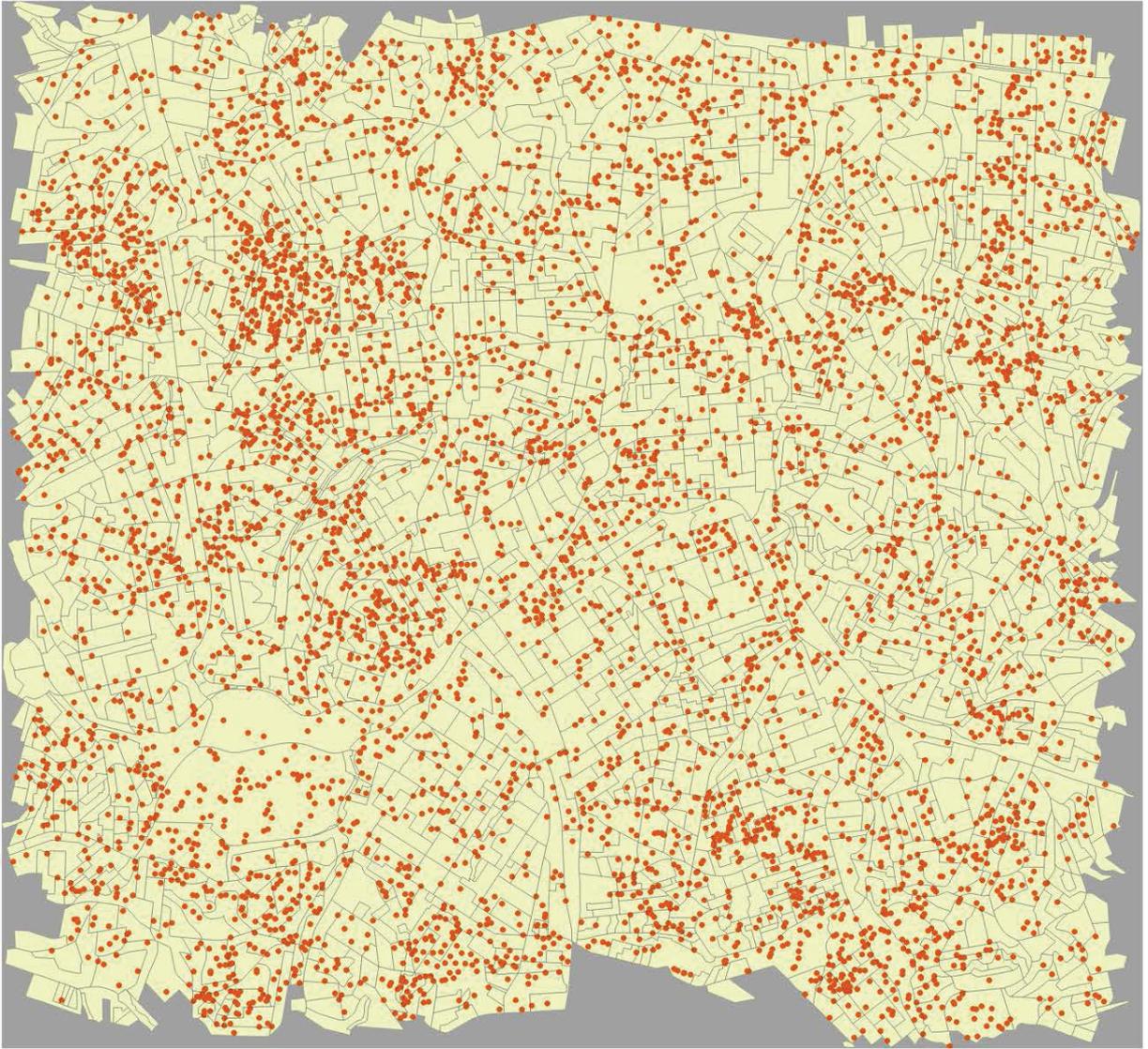


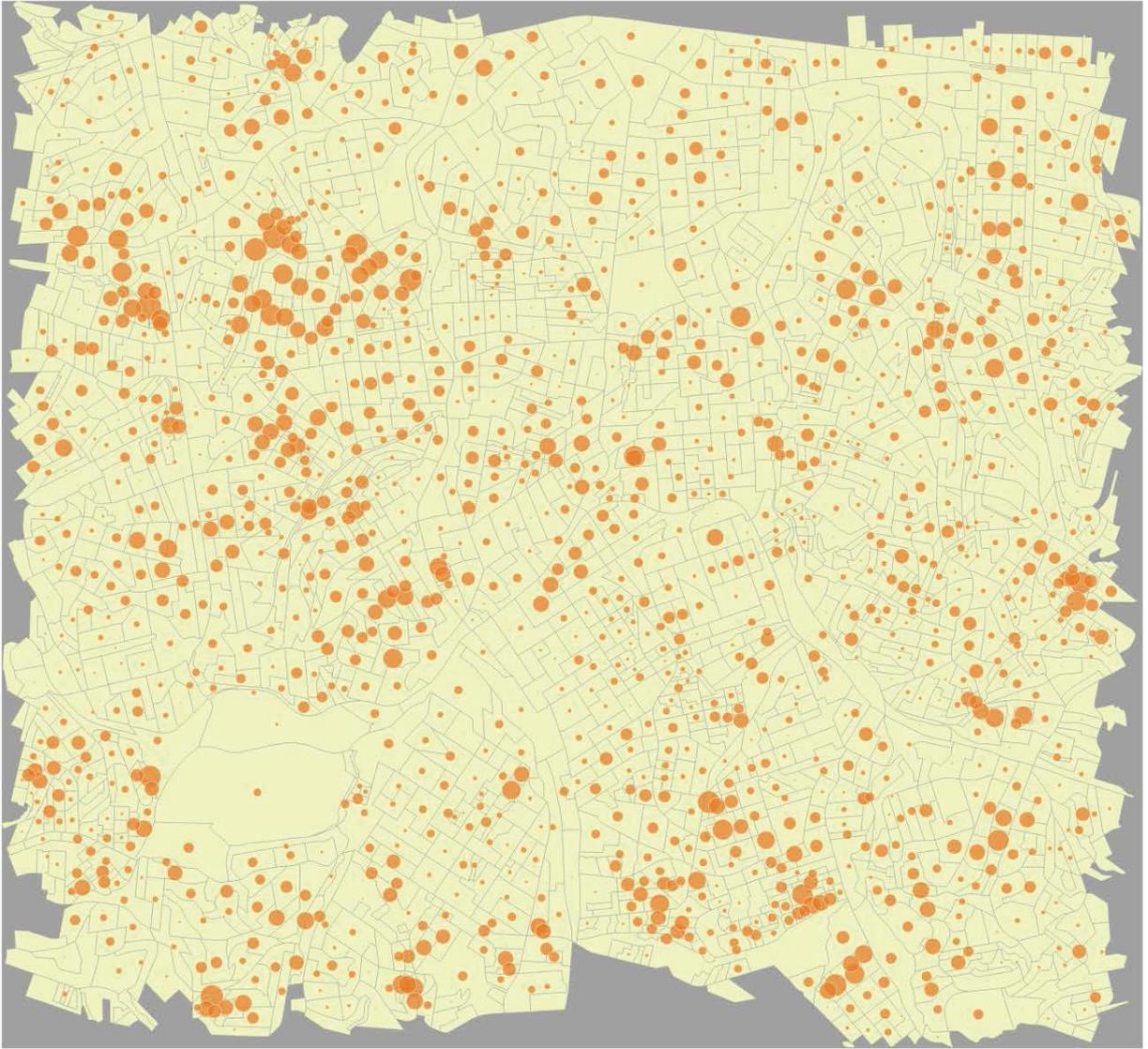


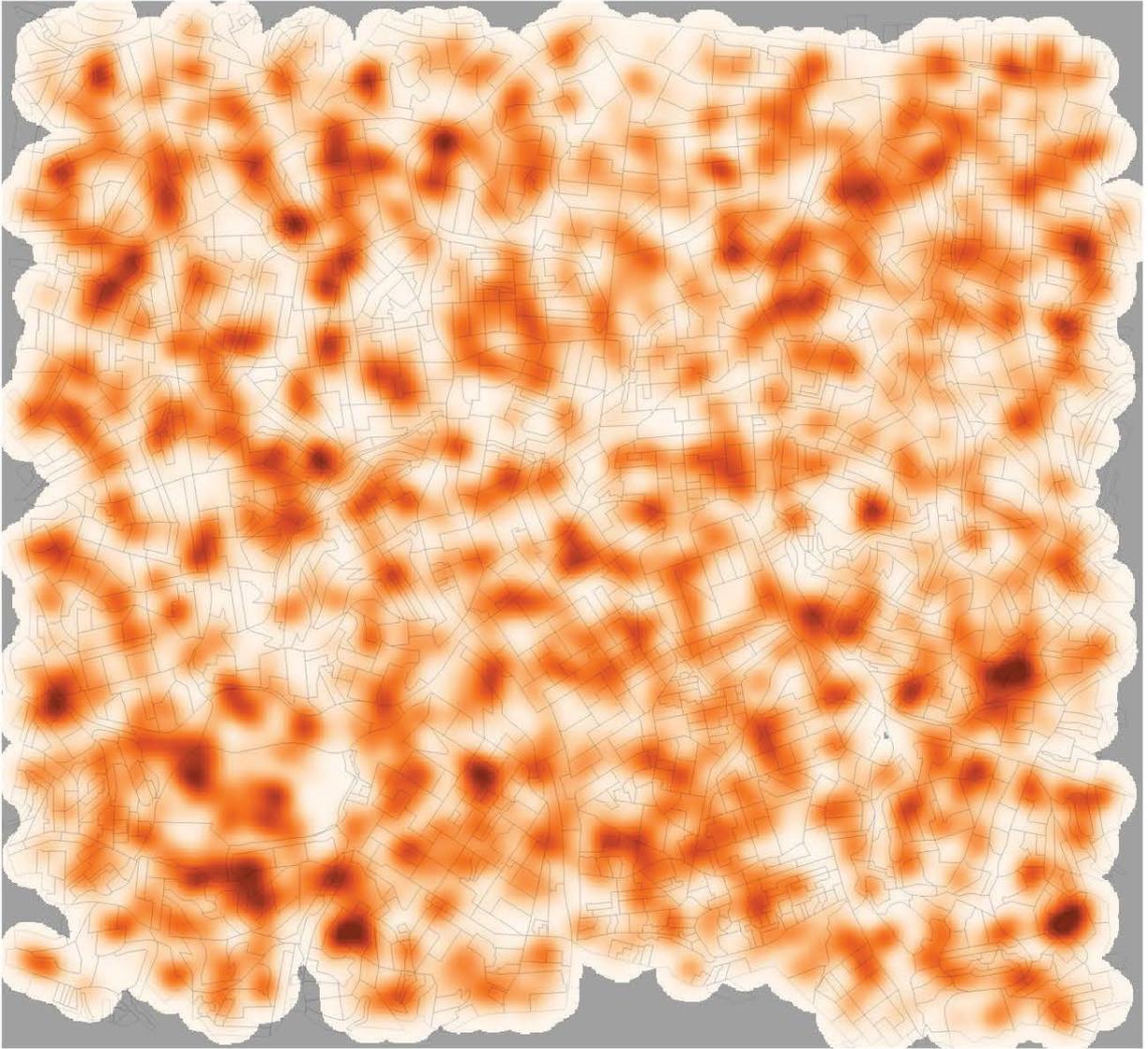




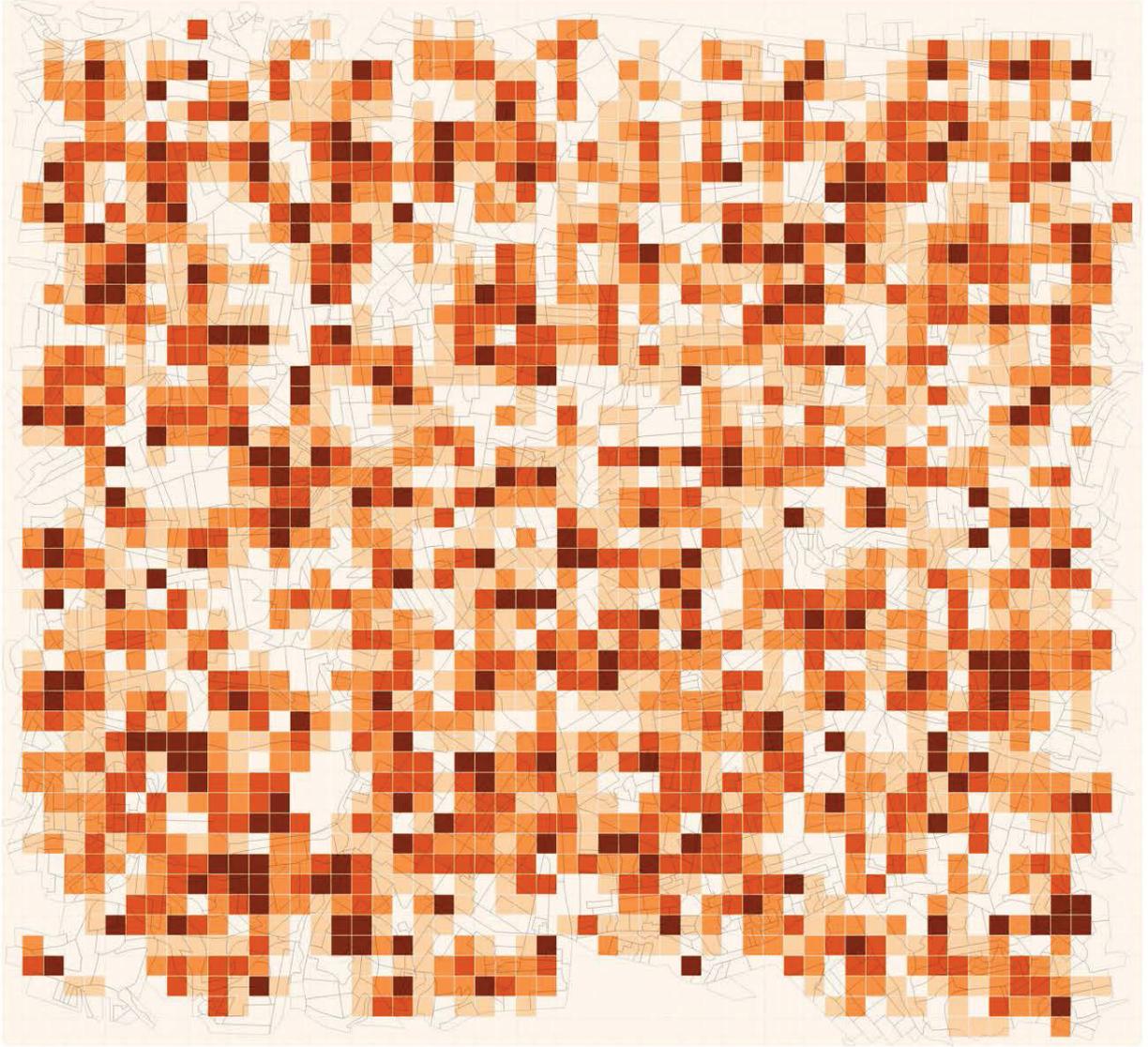


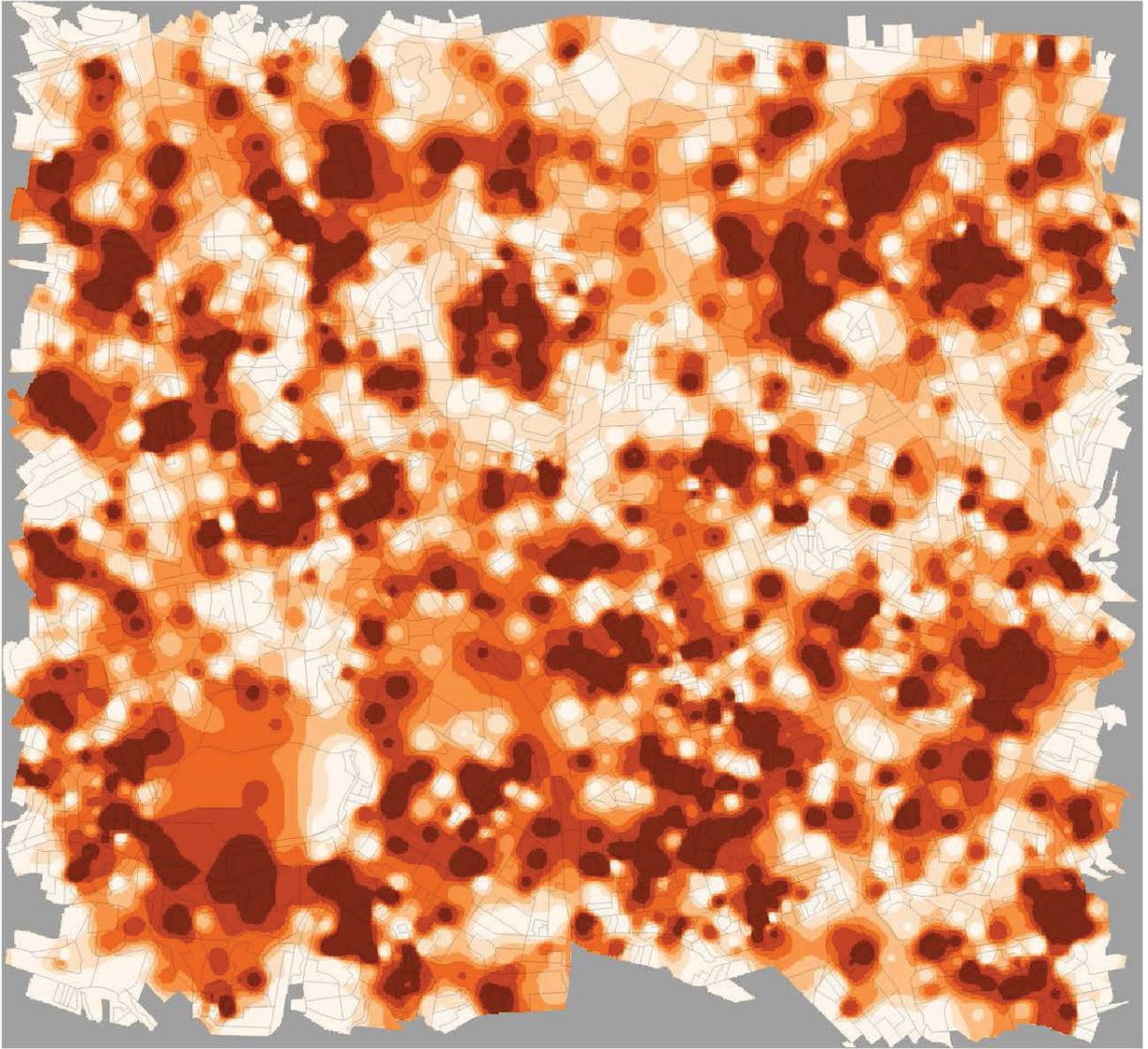


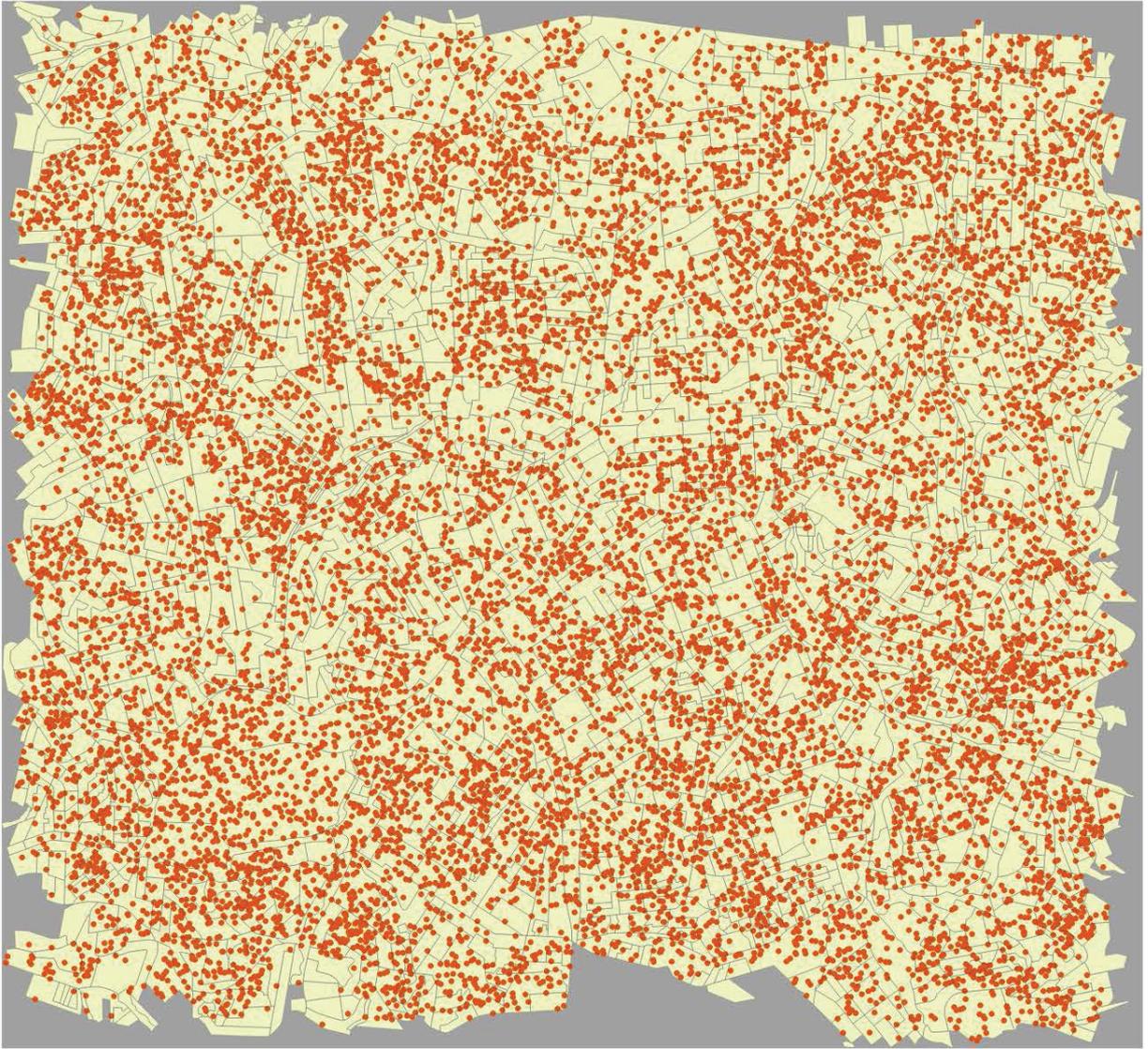




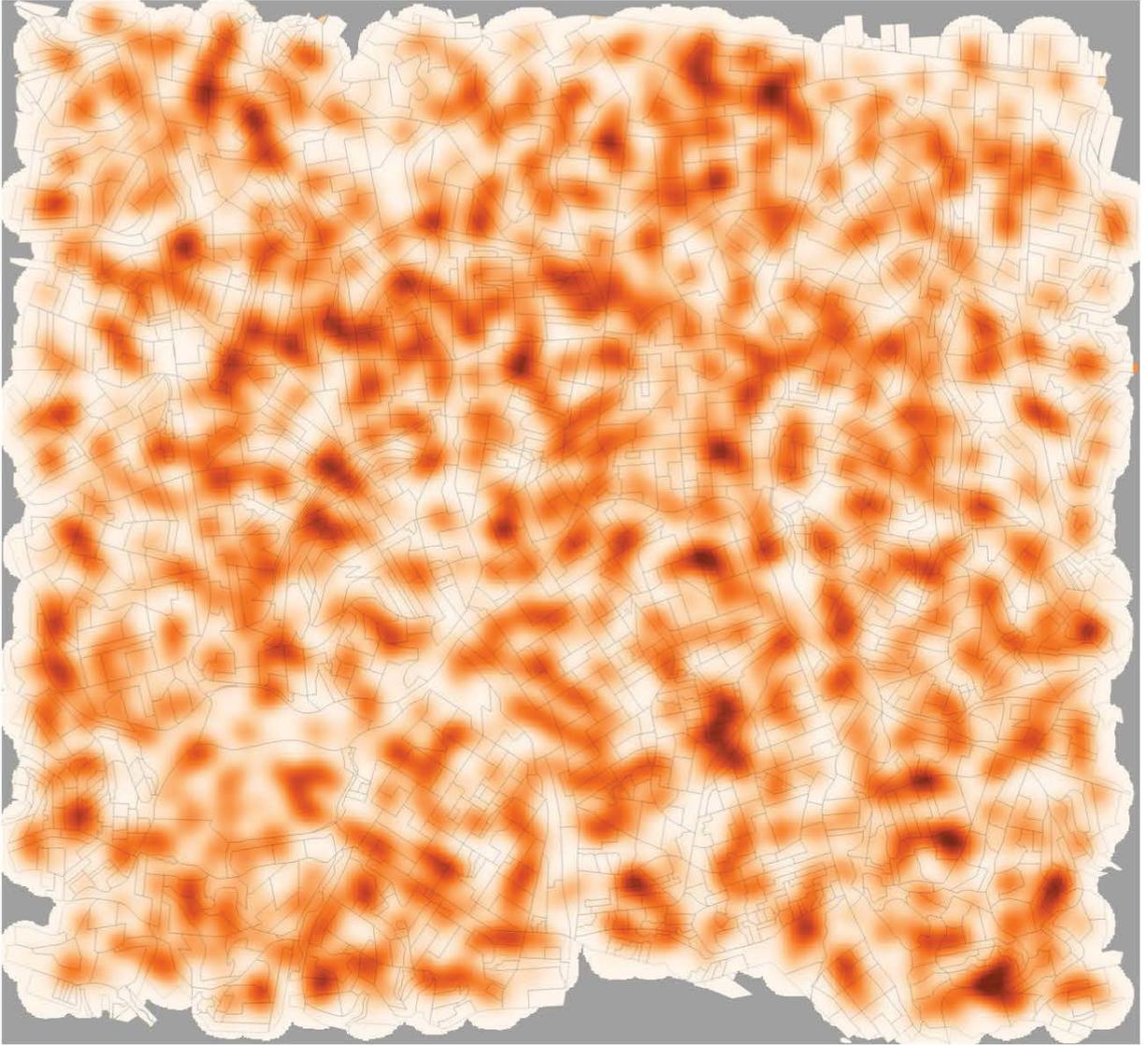




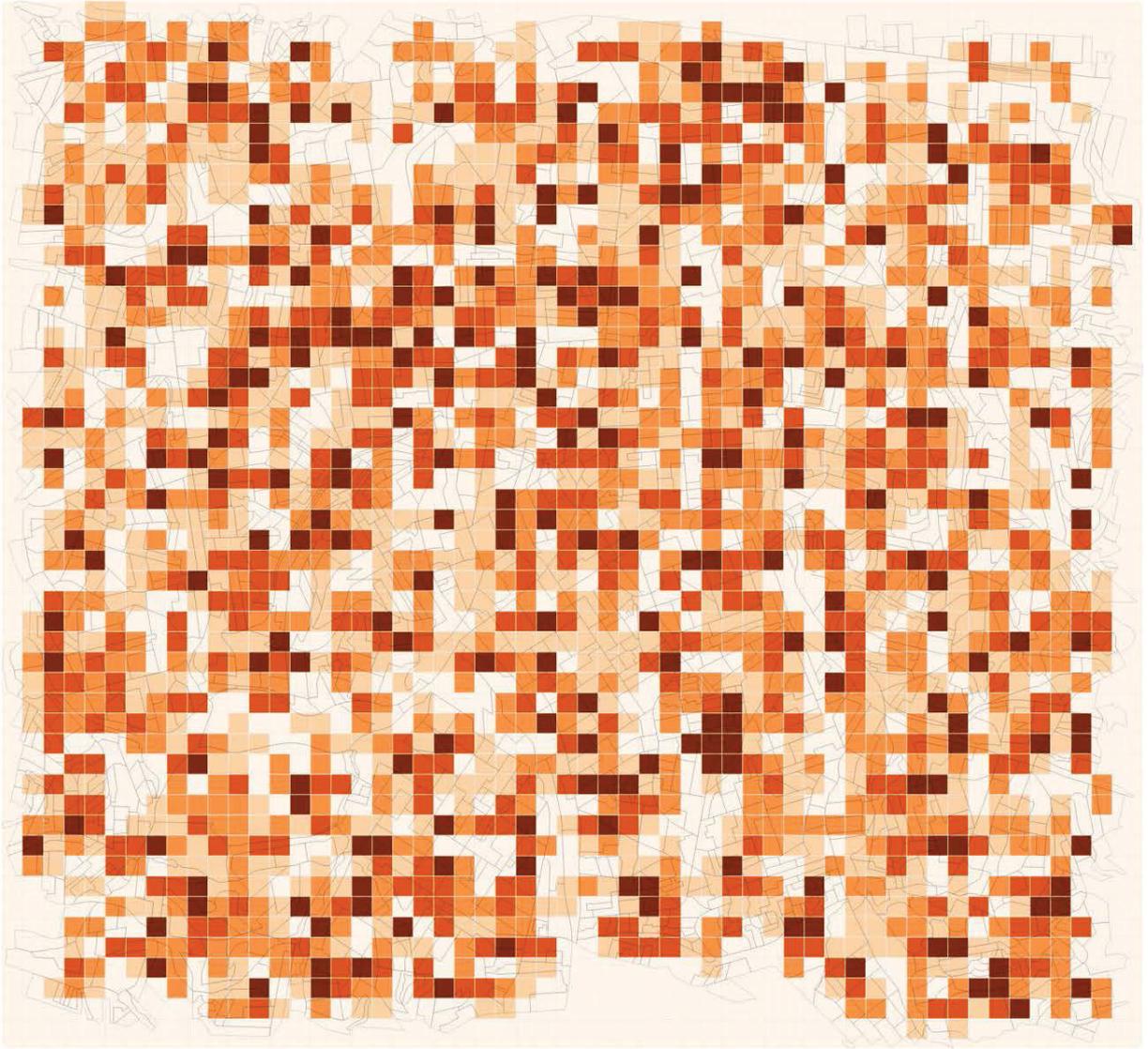


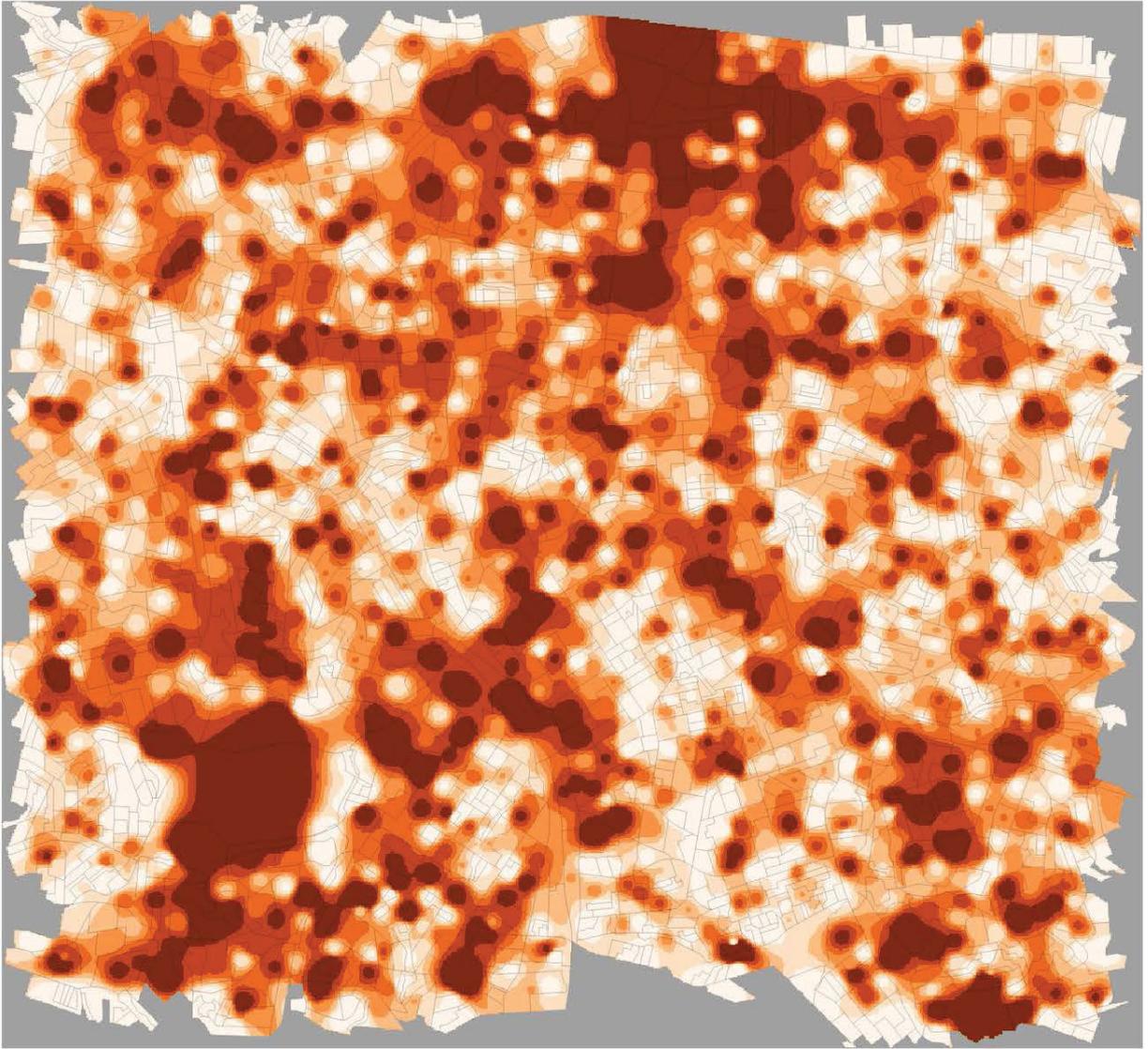


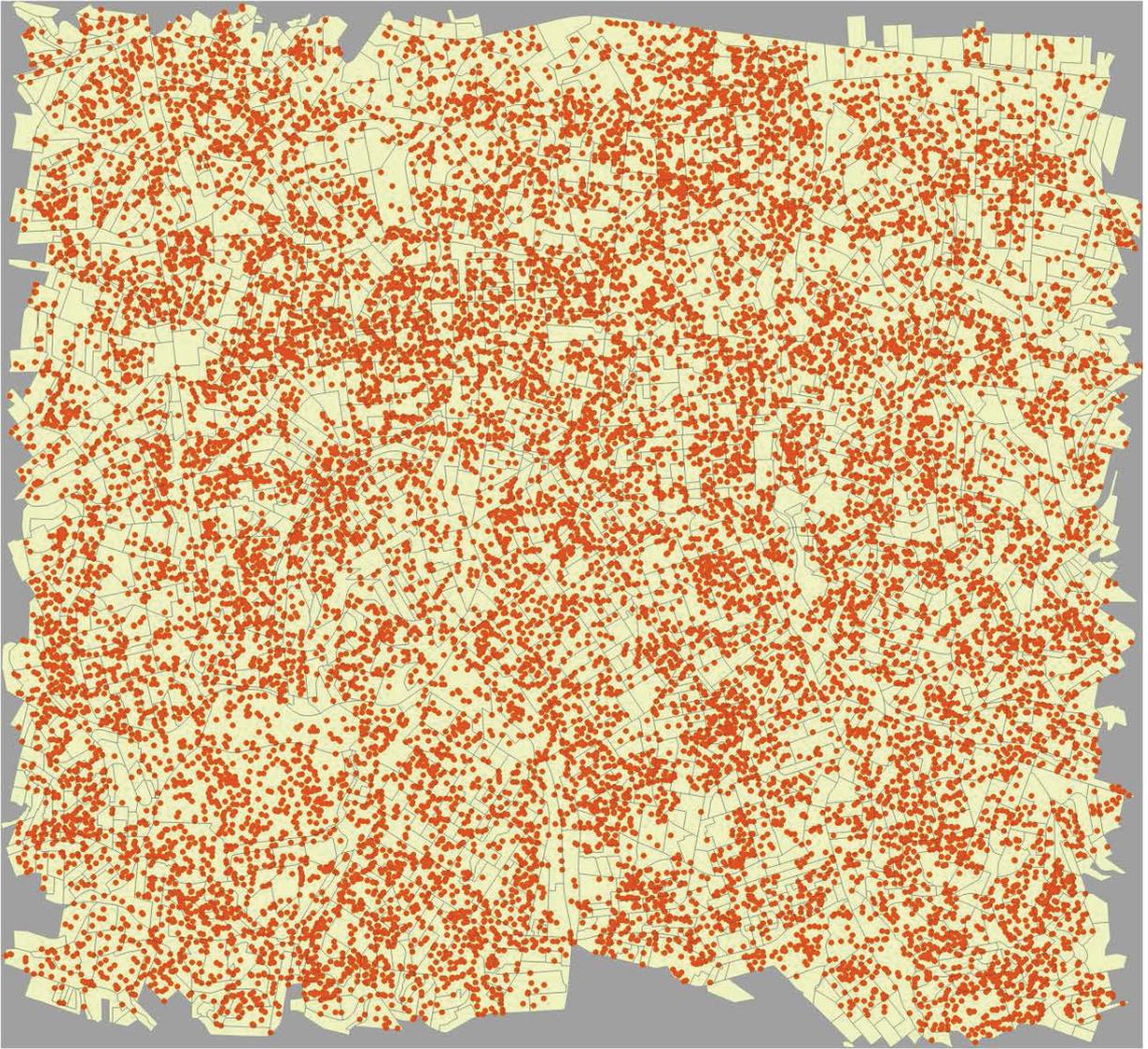




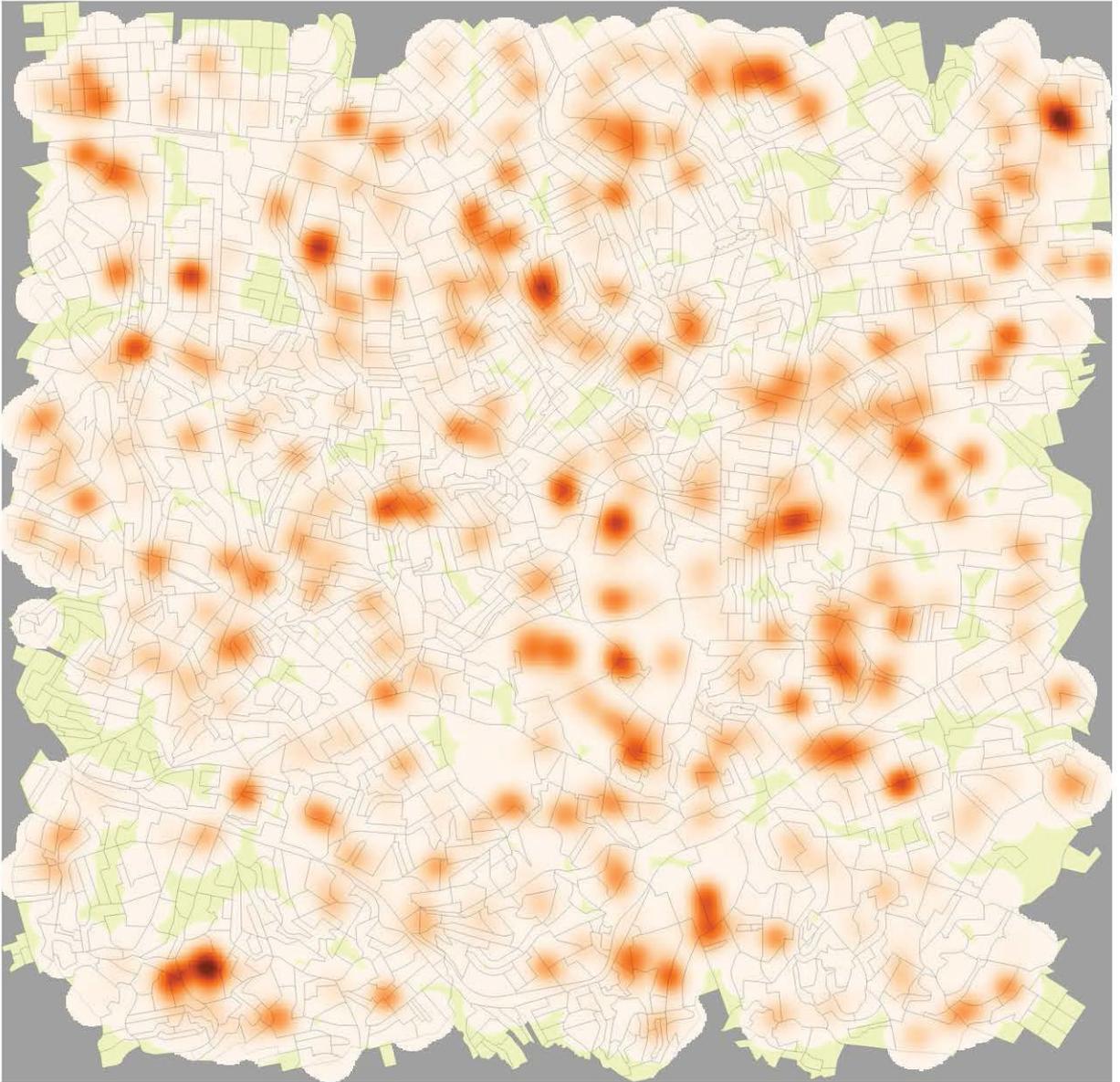




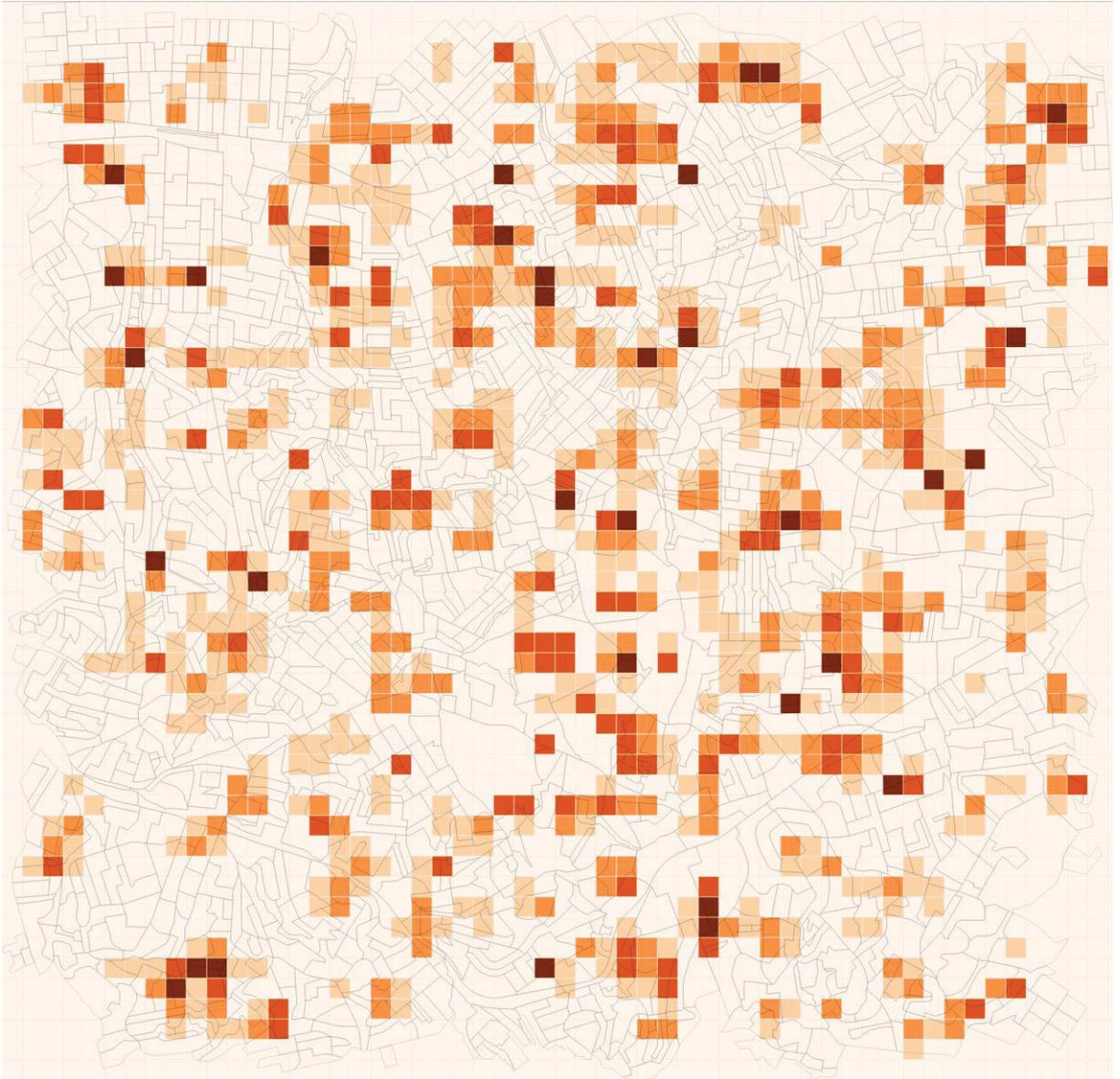


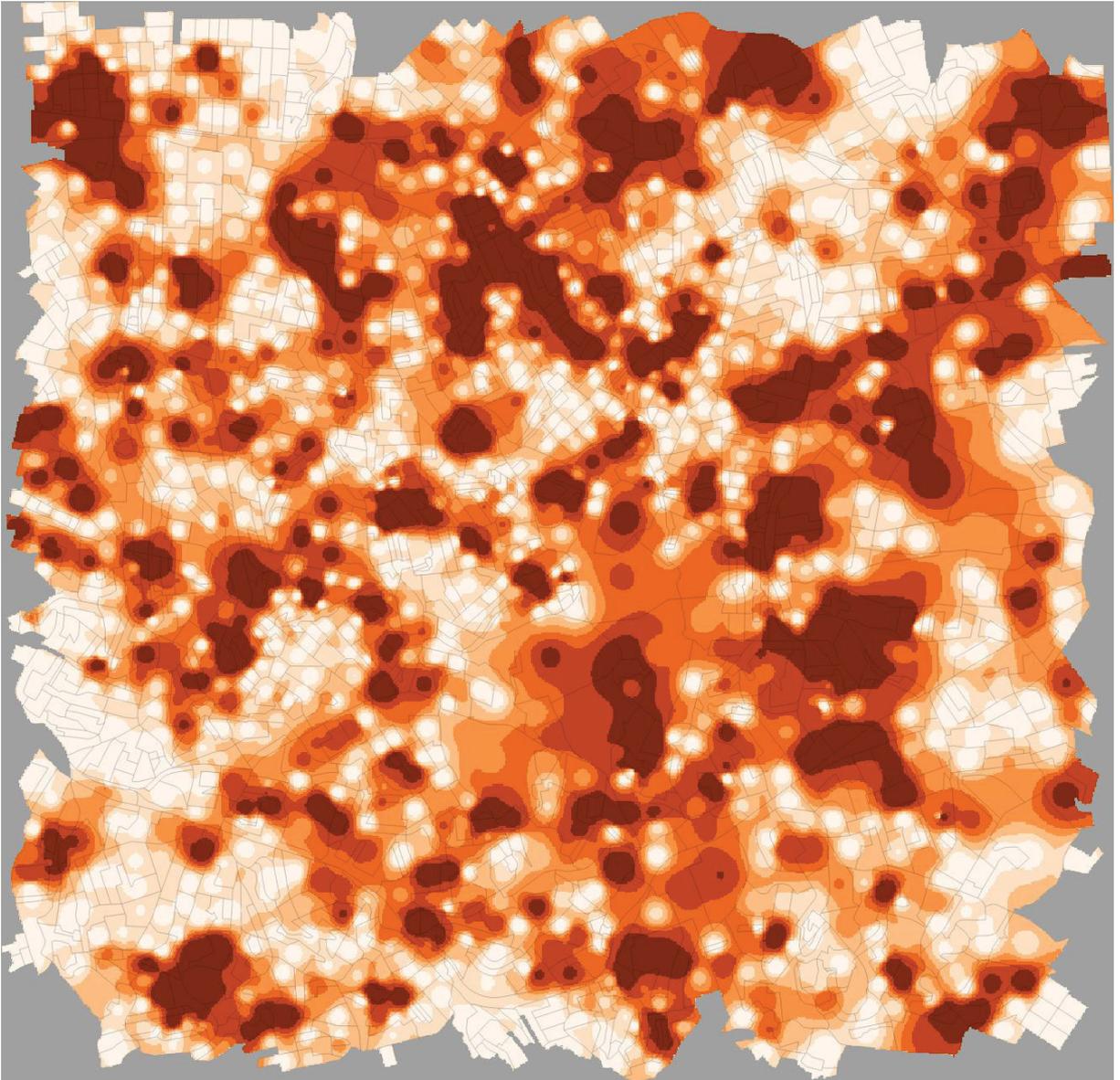


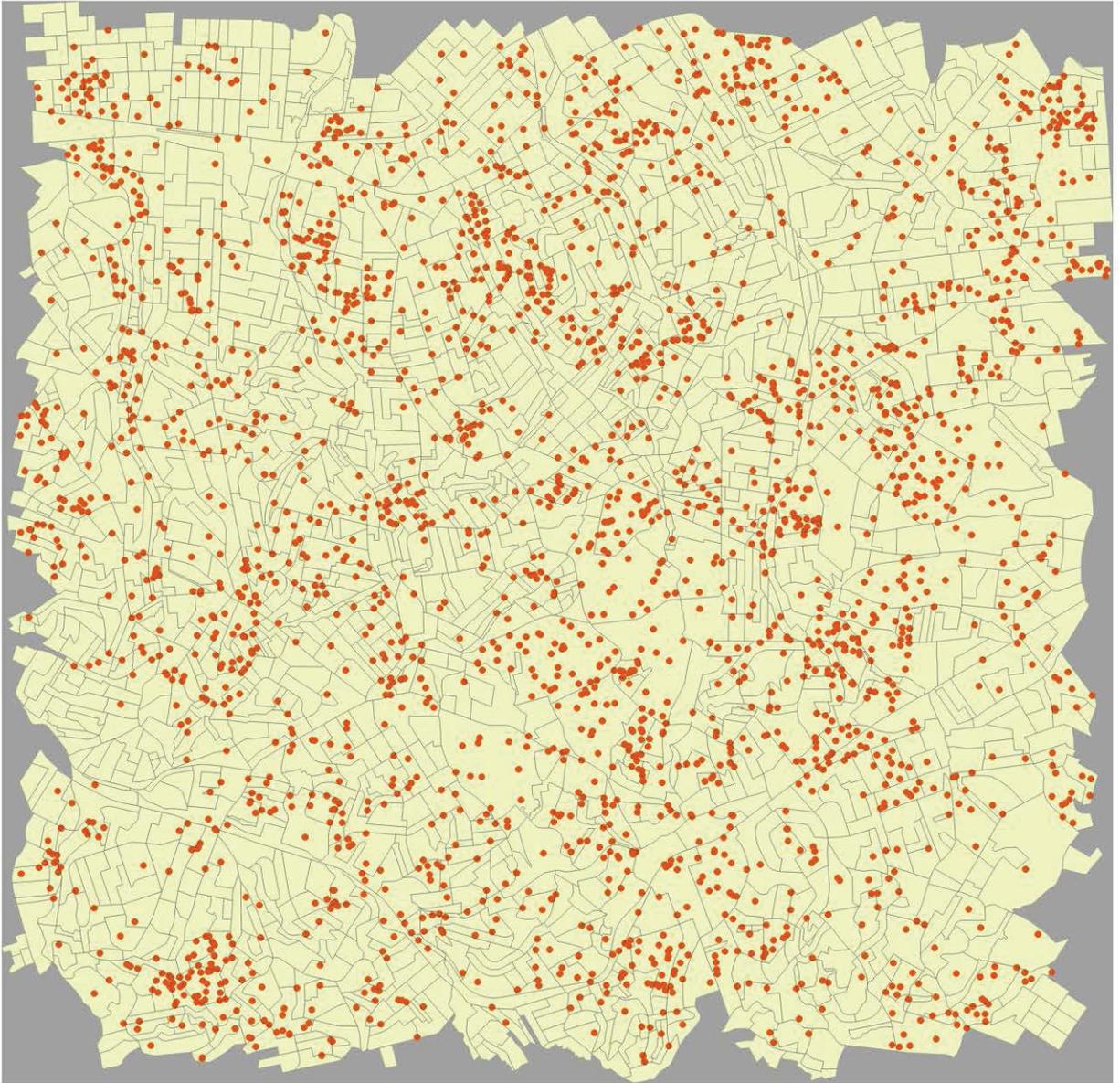


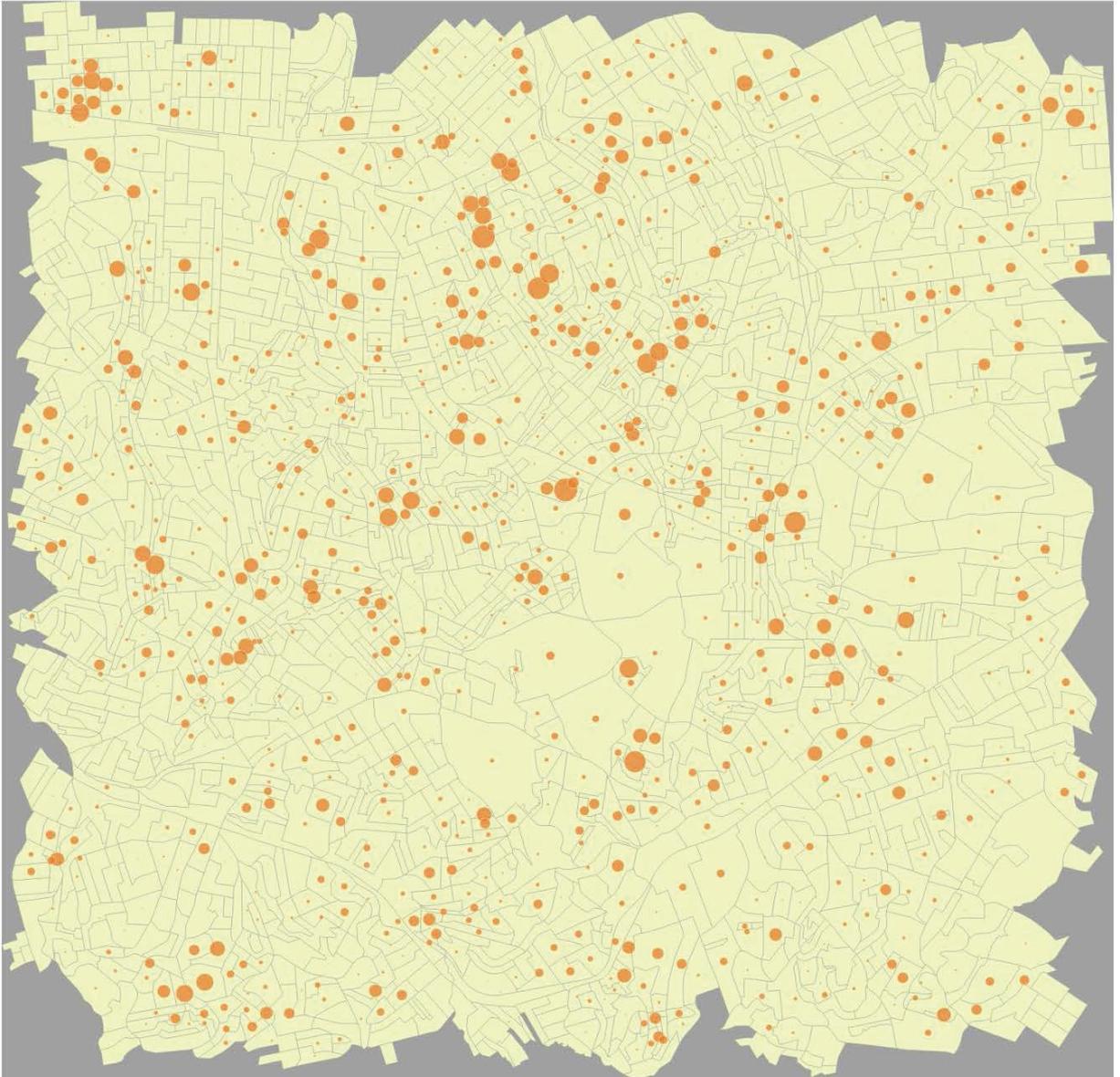


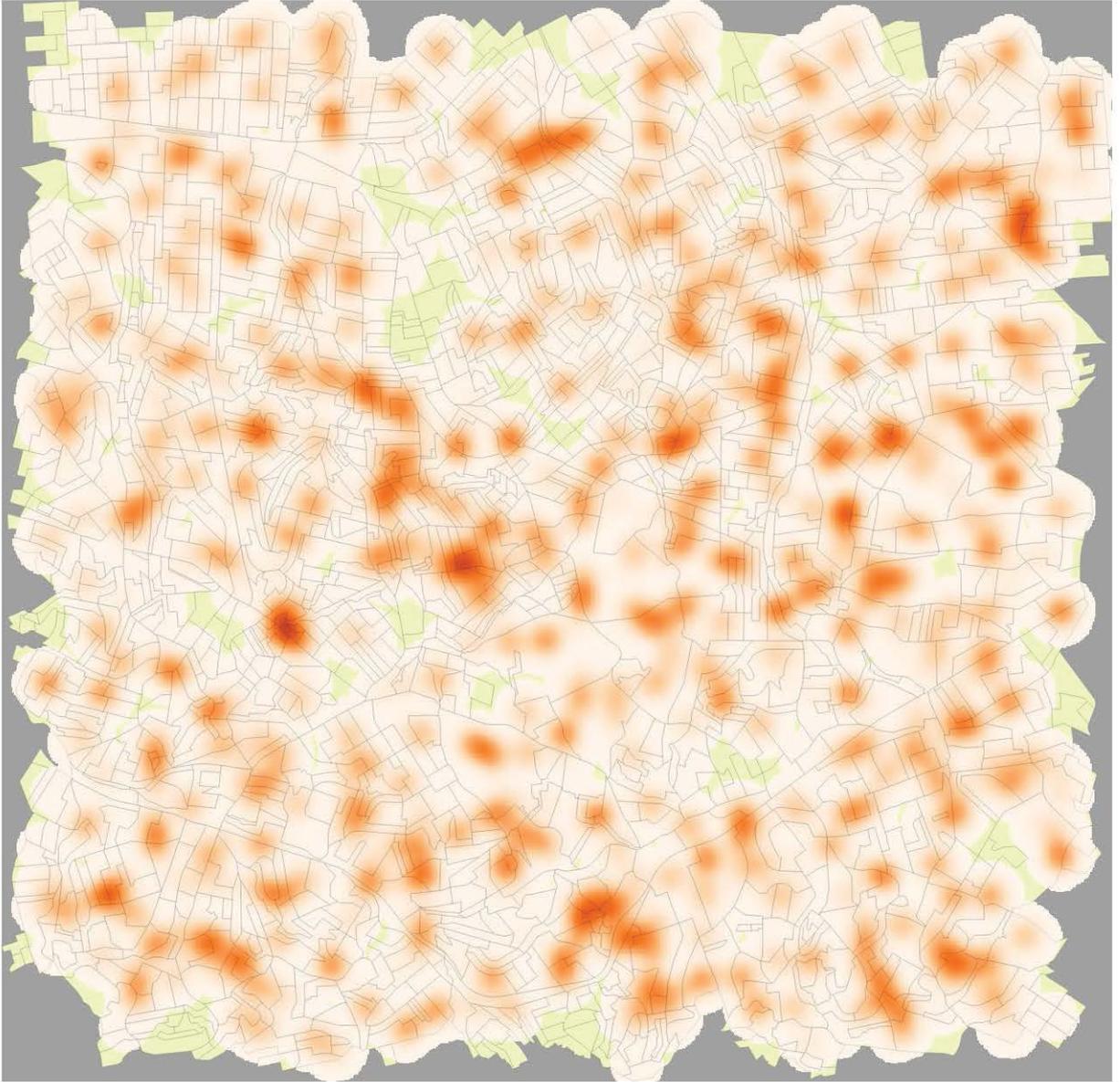


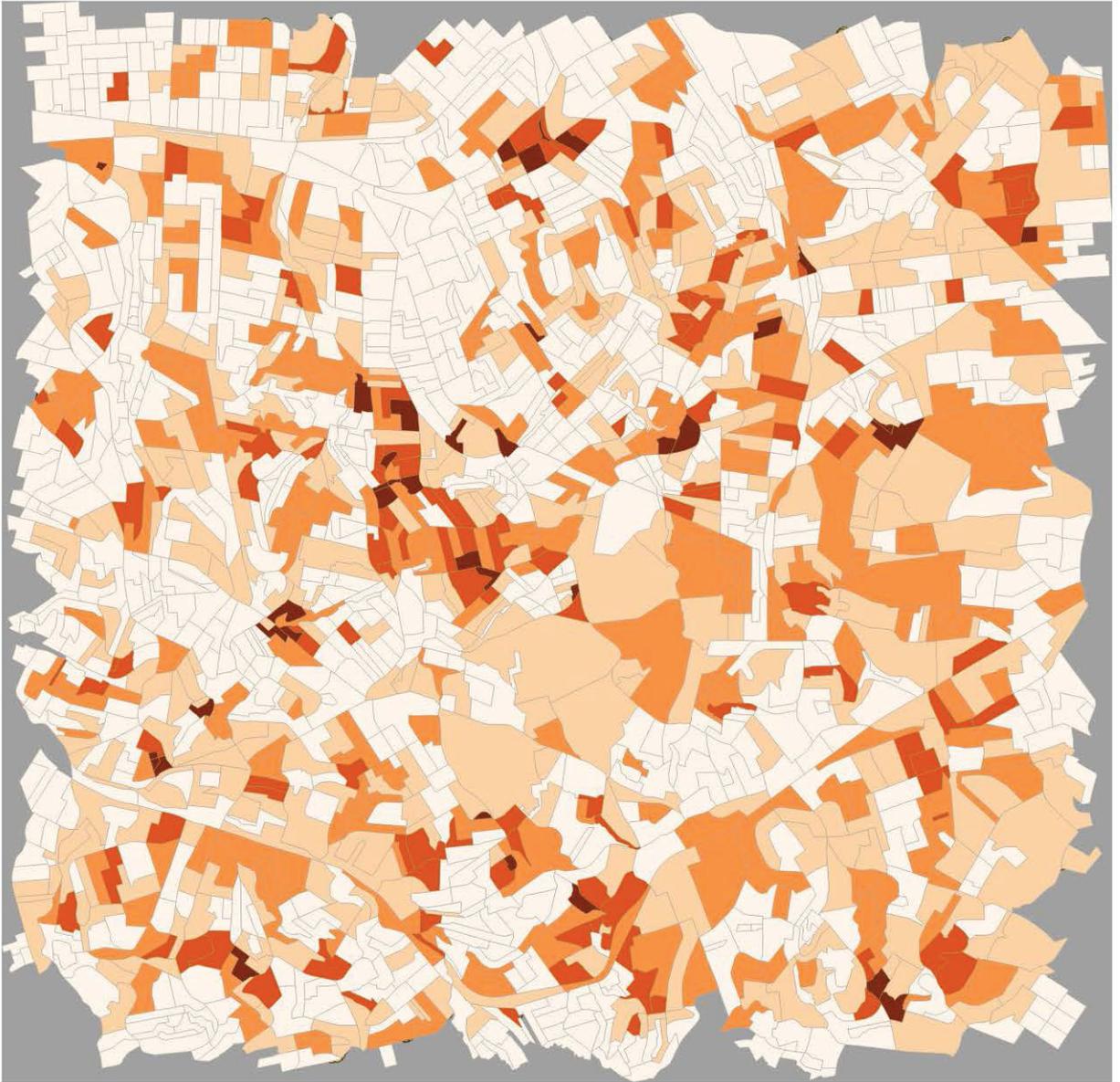


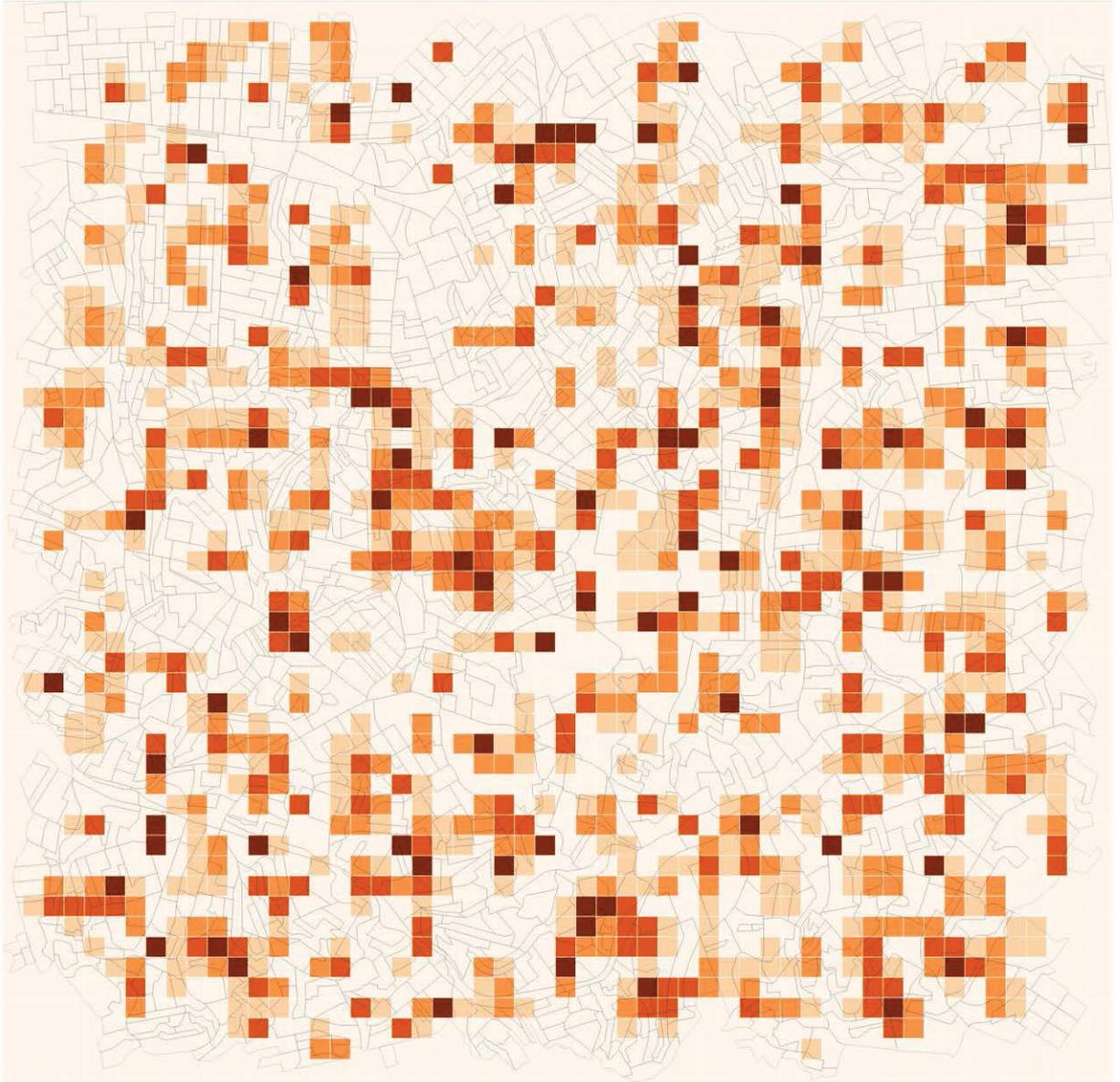


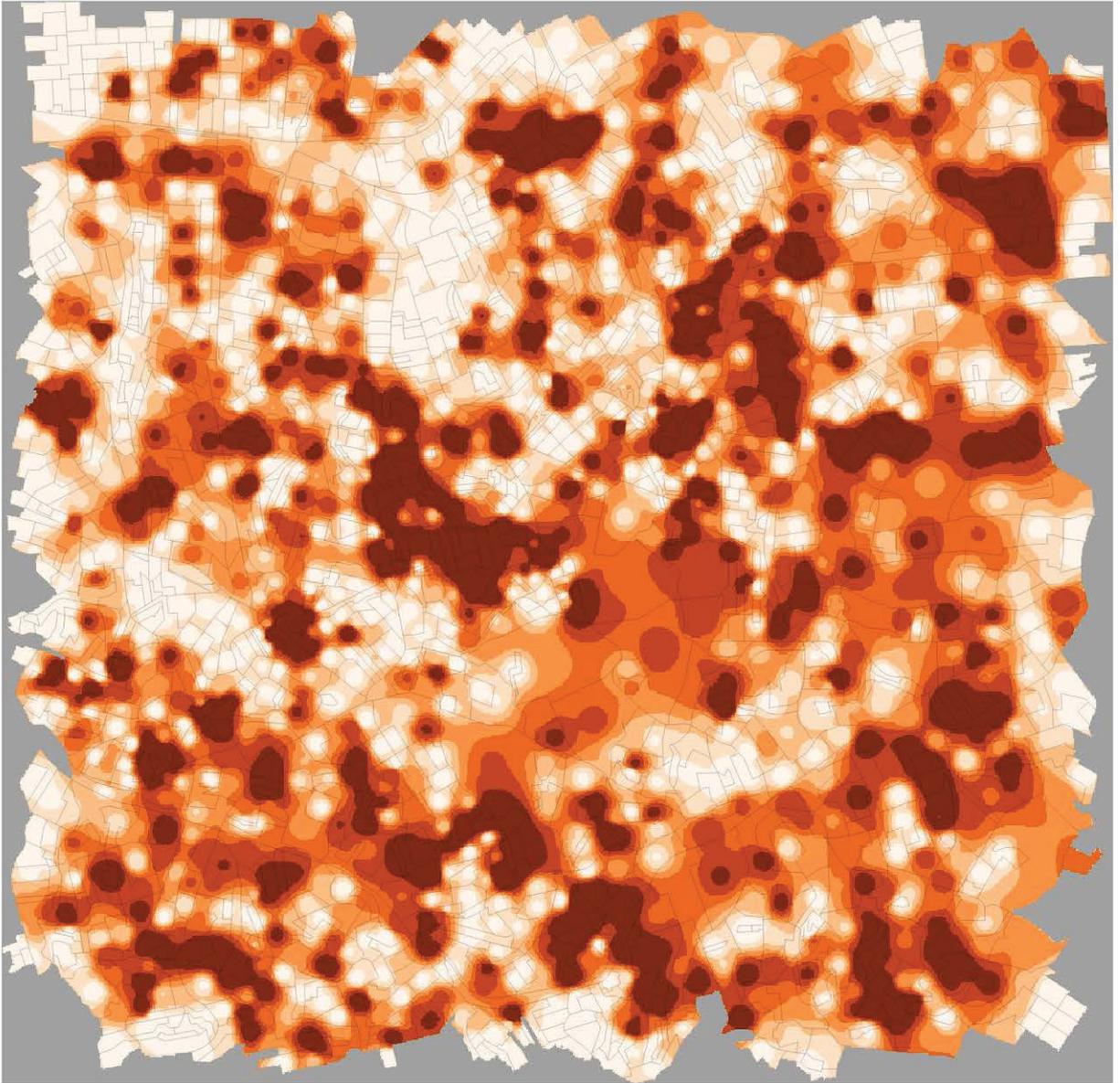


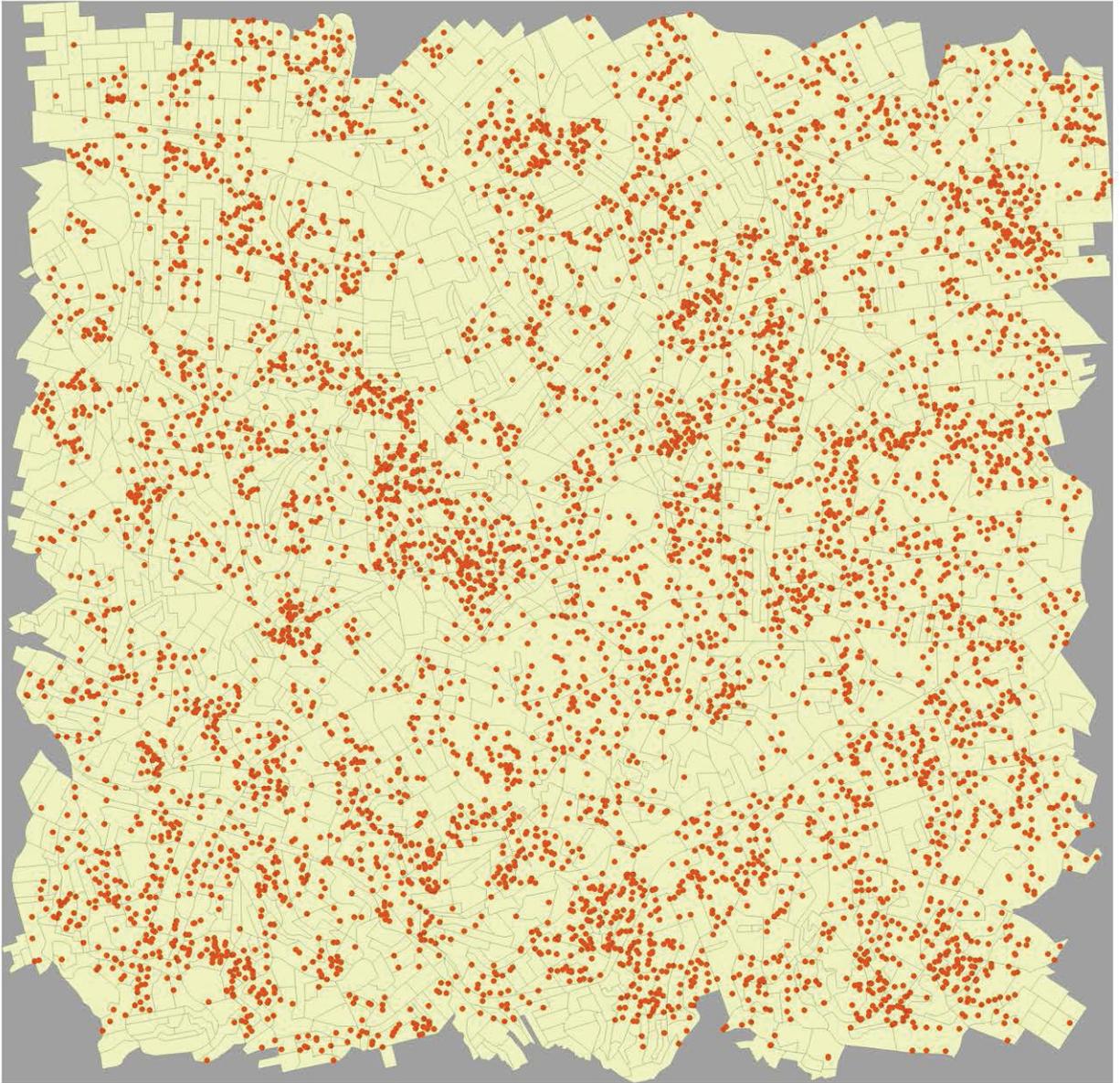




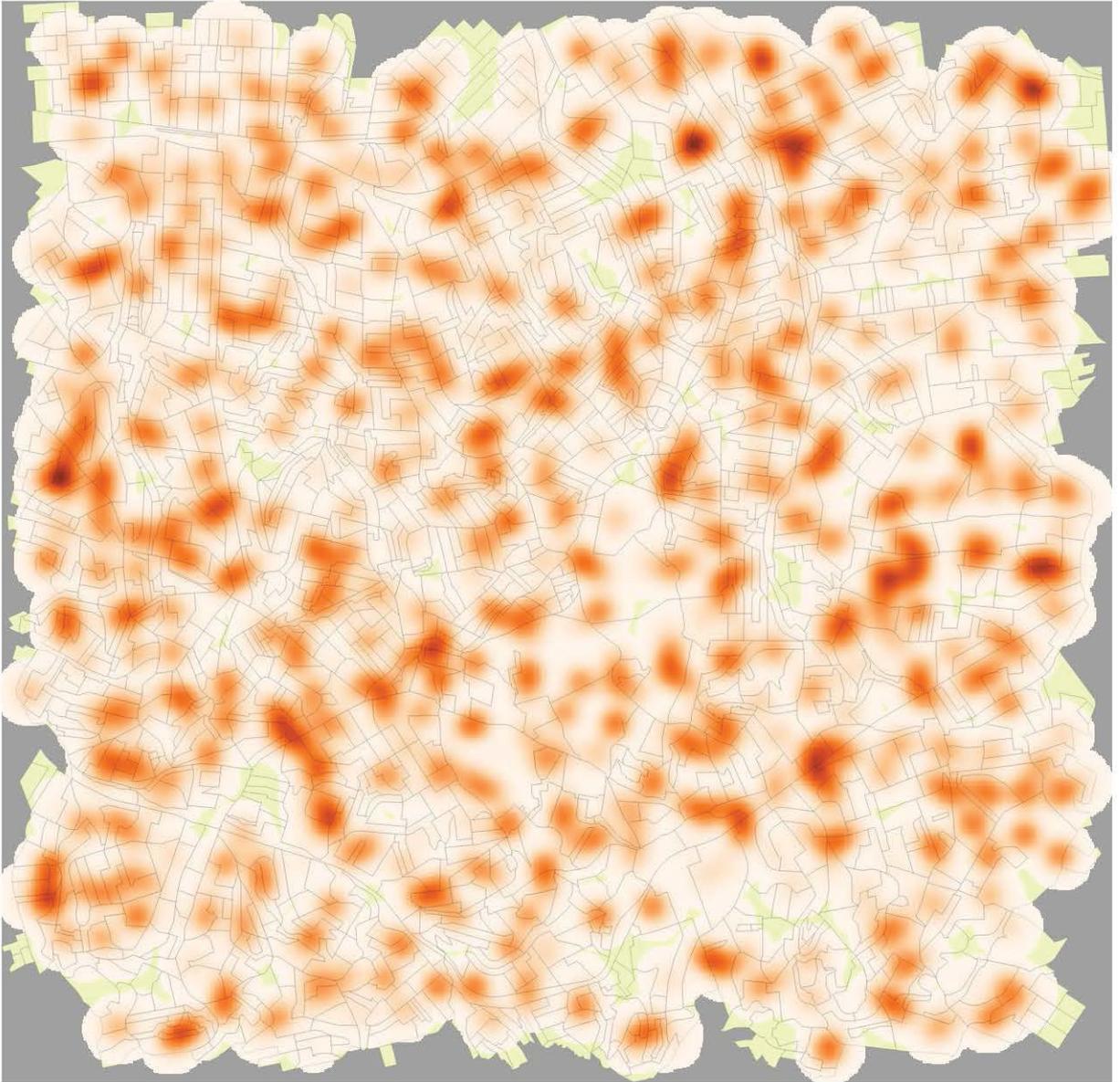


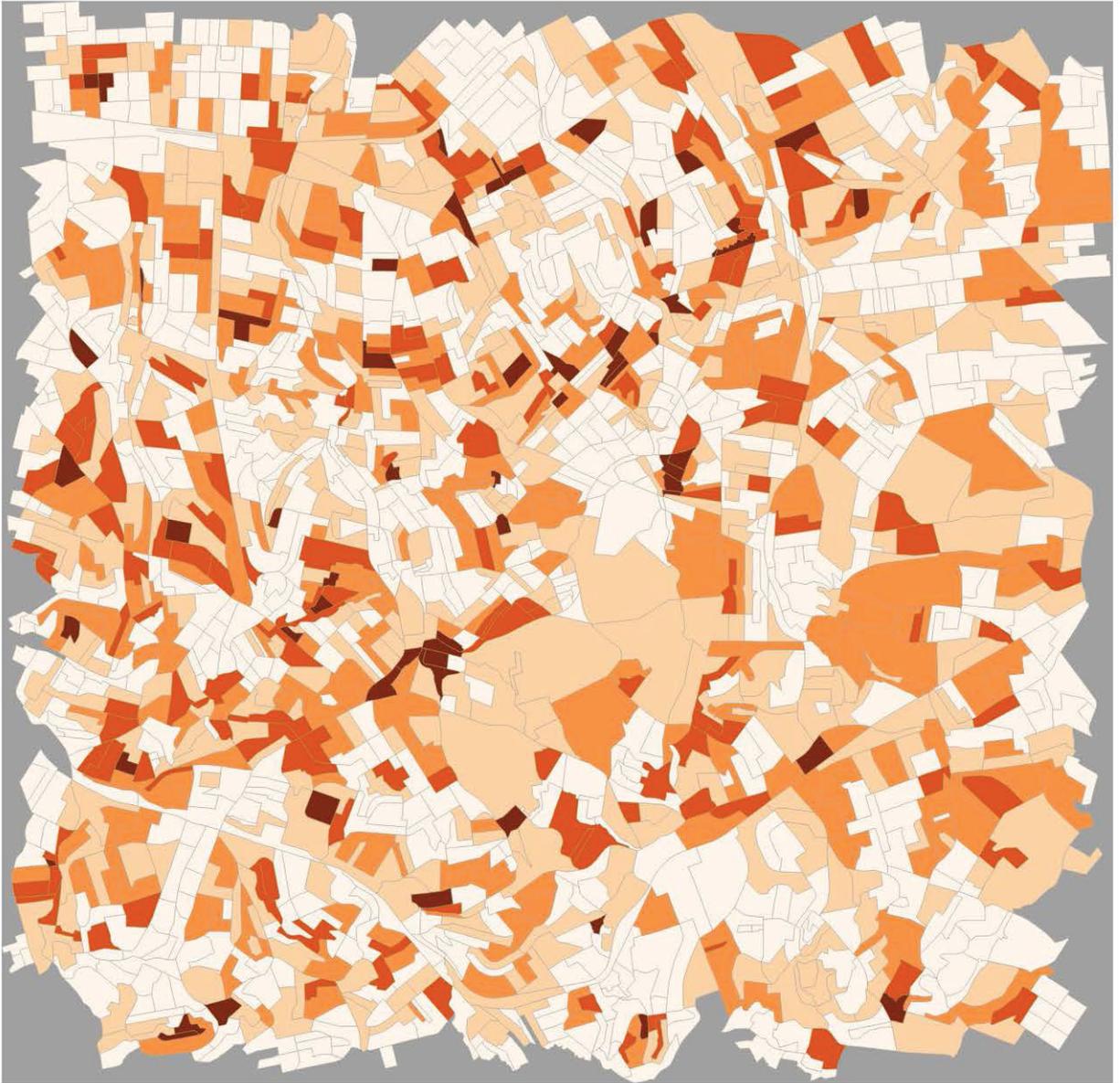


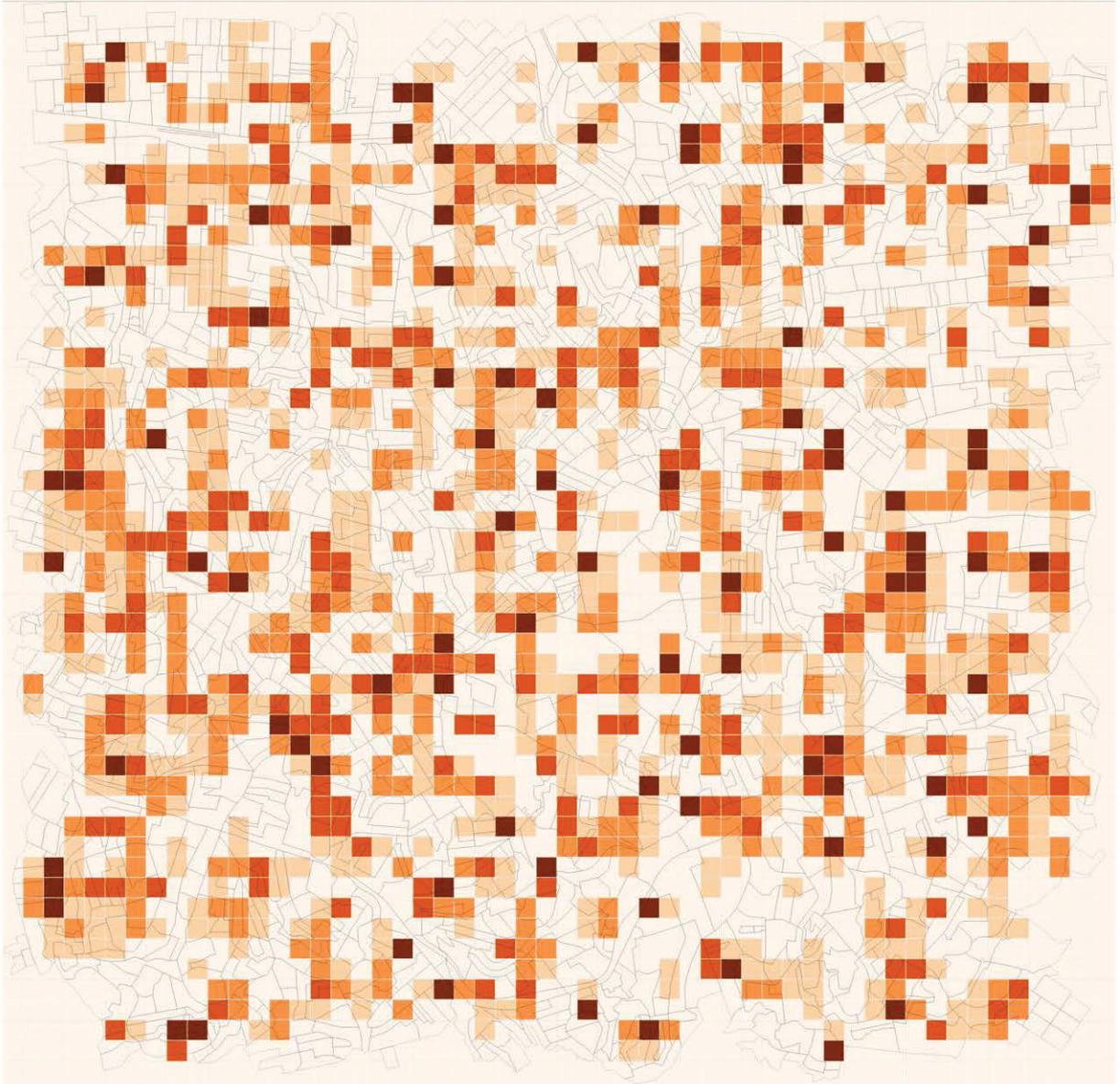


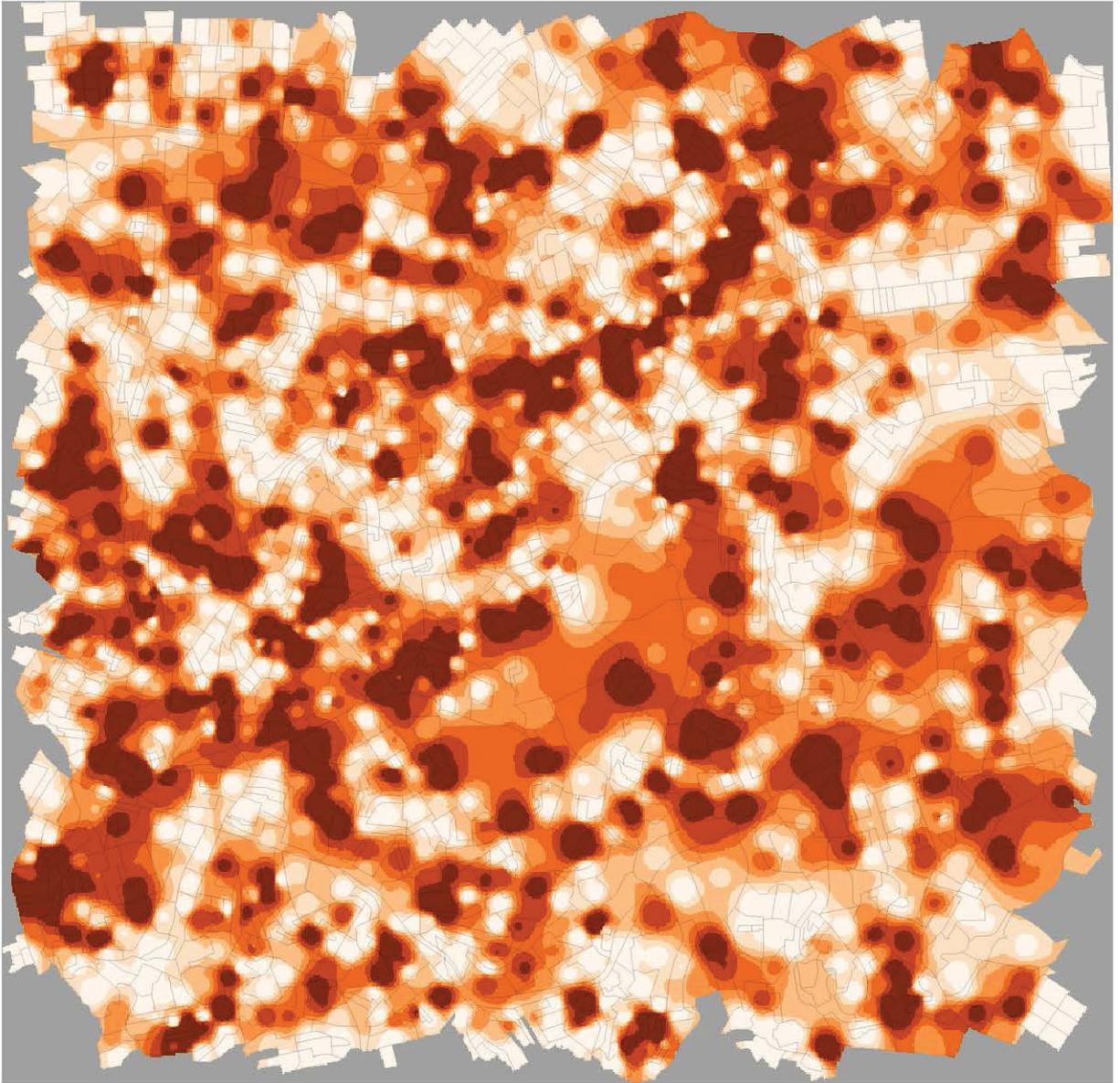


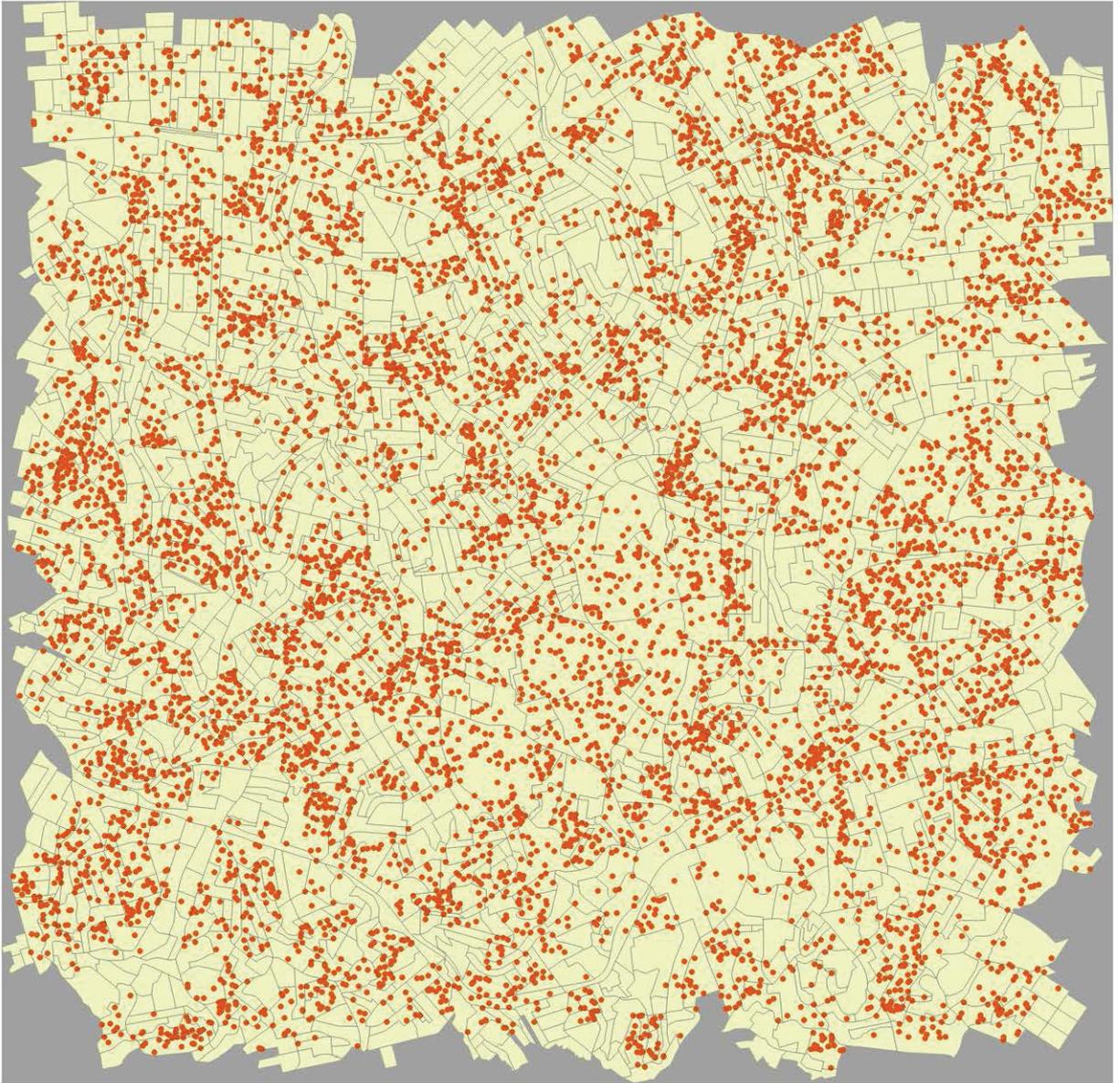


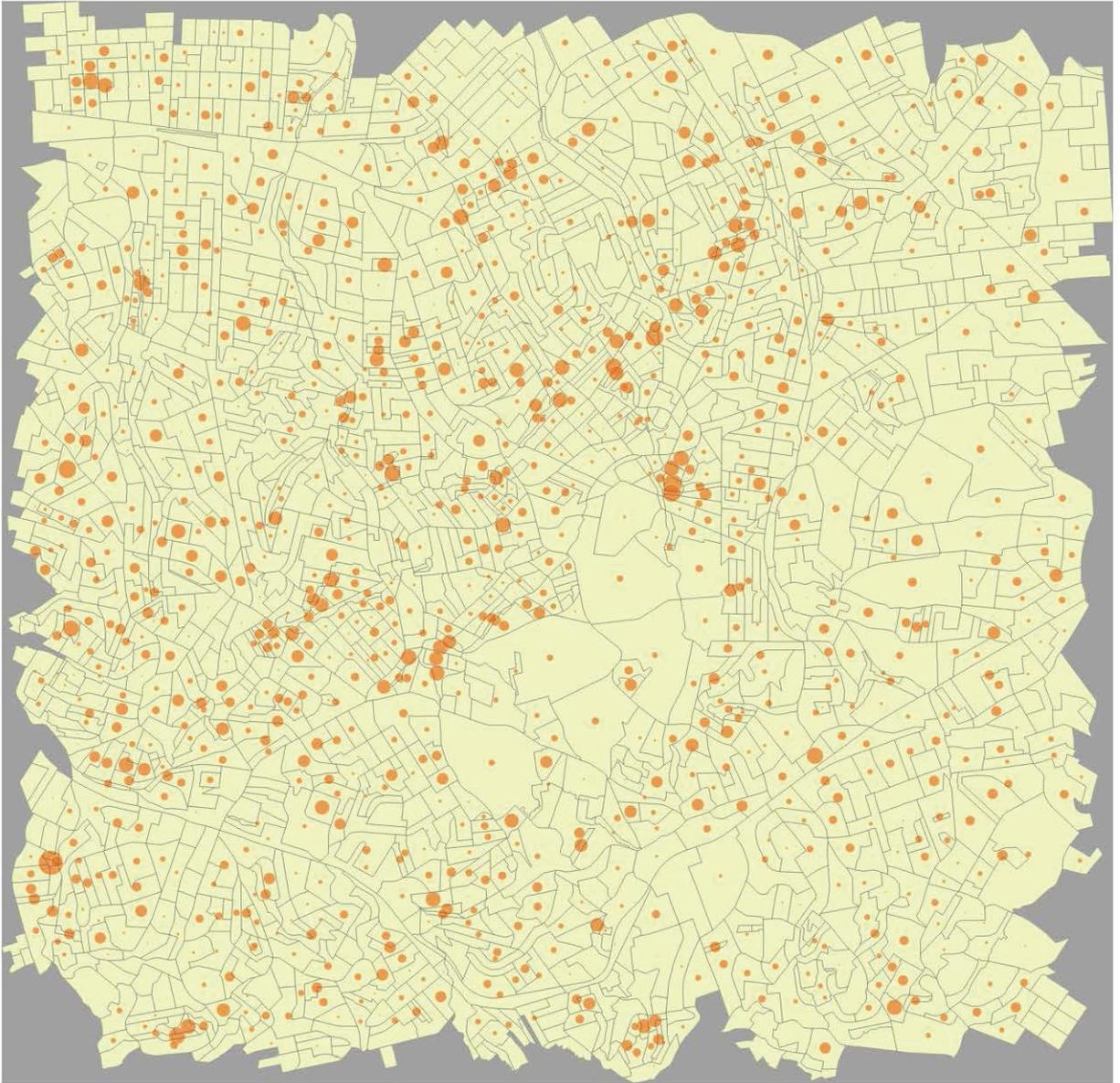




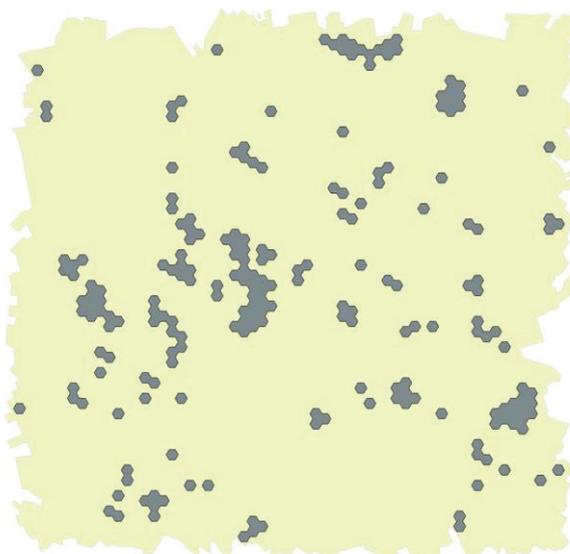




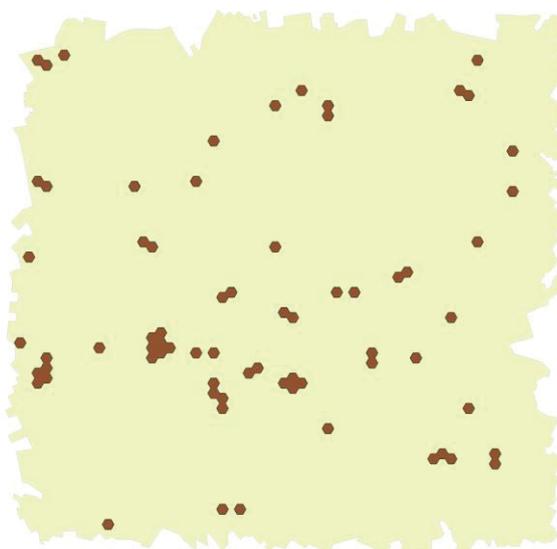




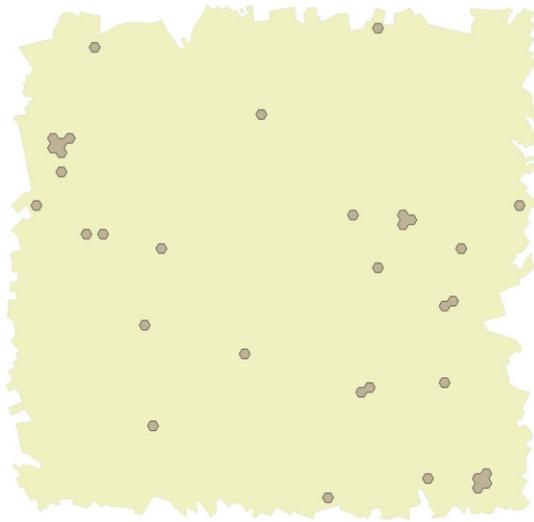
APÊNDICE 3 – REGIÕES DEFINIDAS POR MORAN LOCAL POR CADA CONJUNTO DE DADOS



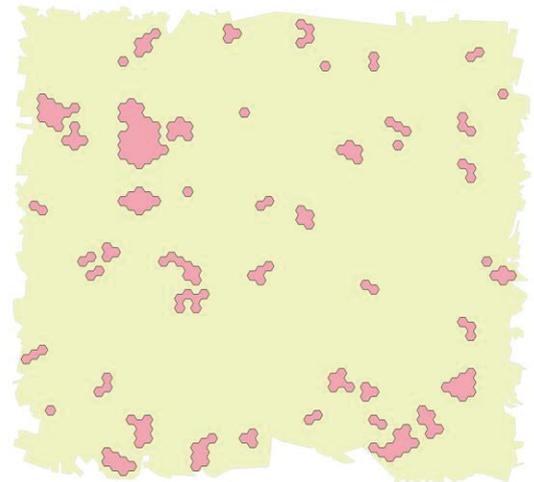
Conjunto de dados - Set7\_s85c100



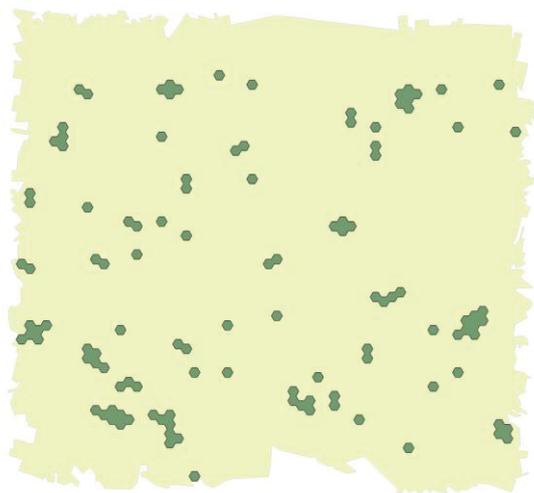
Conjunto de dados - Set7\_s70c100



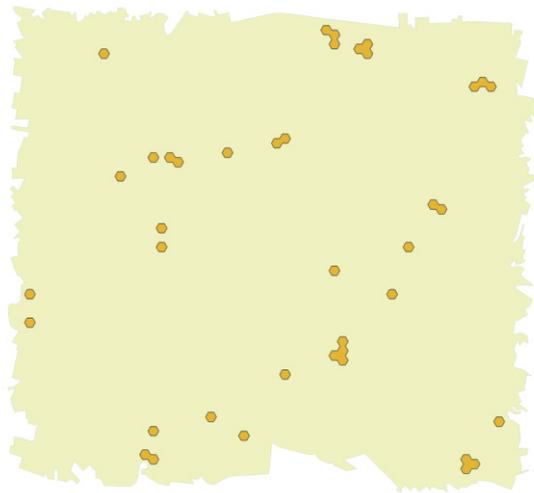
Conjunto de dados - Set7\_s55c100



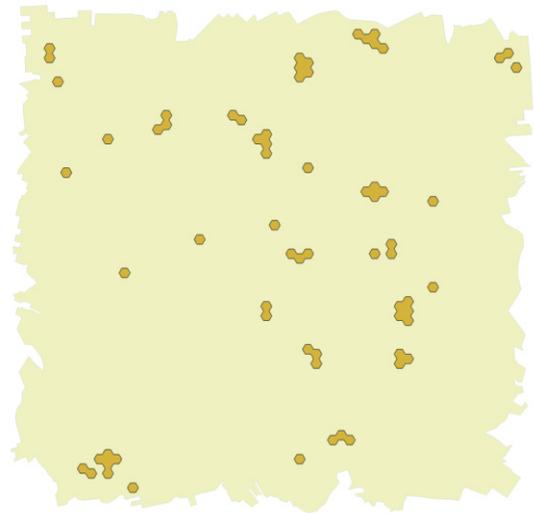
Conjunto de dados - Set8\_s85c60



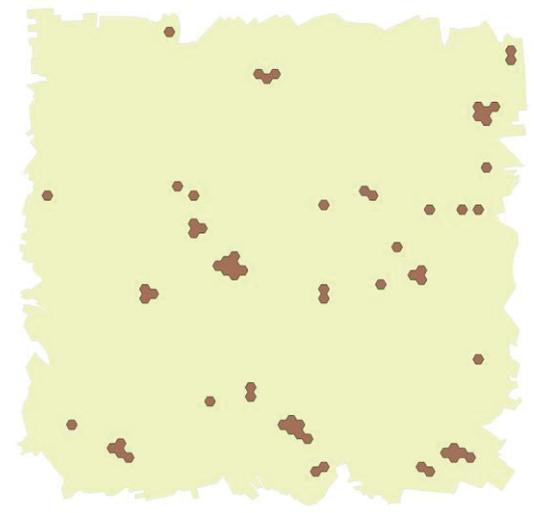
Conjunto de dados - Set8\_s70c60



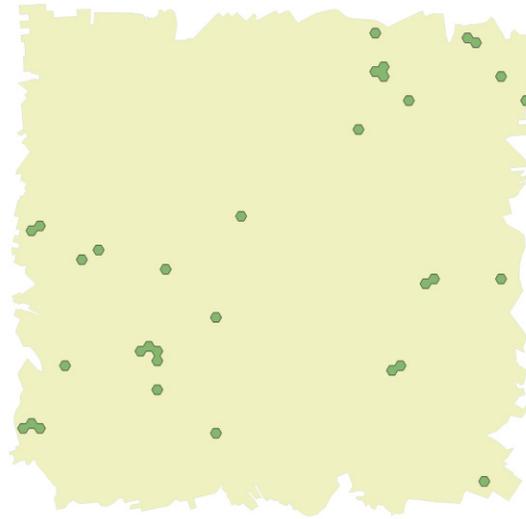
Conjunto de dados - Set8\_s55c60



Conjunto de dados - Set9\_s85c30



Conjunto de dados - Set9\_s70c30



Conjunto de dados - Set9\_s55c30