

# Ciência de Dados Geoespaciais

## Parte III – Análise Espacial

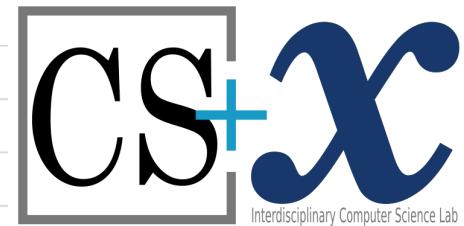
Análise: correlação espacial, padrões de pontos, redes

Clodoveu Davis

Departamento de Ciência da Computação

Instituto de Ciências Exatas

Universidade Federal de Minas Gerais



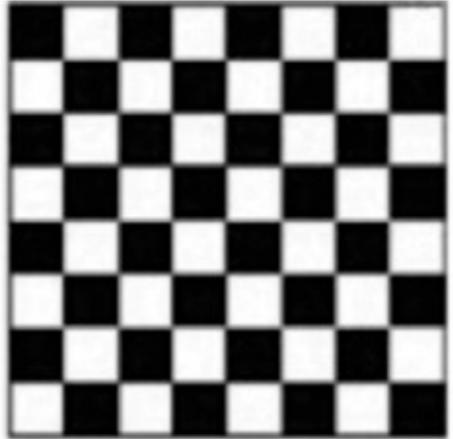
# Ciência de Dados Geoespaciais

- **Intro: Geoinformática**
- **Parte I – Aquisição**
  - Posicionamento geográfico: coordenadas e projeções; medições;
  - Fontes de dados geográficos
  - Fontes de dados não-geo com referências a localização
  - Codificação de dados geográficos (esquema físico)
- **Parte II – Engenharia de Dados**
  - Modelagem de dados
  - Gerenciamento de dados
  - Tratamento e transformações
  - Integração
  - Geocodificação, georreferenciamento, geolocalização, geotagging
- **Parte III – Investigação e Análise**
  - Problemas específicos de dados geoespaciais
  - Análise exploratória
  - **Análise espacial: padrões de pontos, redes, autocorrelação espacial**
  - Mineração de dados geográficos
- **Parte IV – Visualização e Decisão**
  - How to Lie with Maps: cartografia e detalhes cartográficos
  - Mapas de densidade kernel e mapas de calor
  - Cartogramas
  - Esquemáticos
  - Animações e mapas interativos

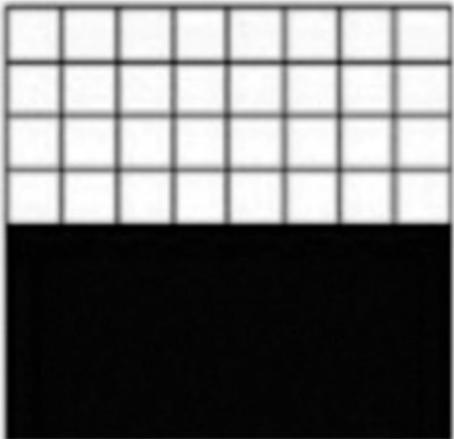
# Autocorrelação espacial

- Quantifica a semelhança entre a localização de objetos espaciais e seus atributos
- Se objetos que são semelhantes em localização (i.e., próximos) são também semelhantes nos valores de seus atributos, então existe **autocorrelação espacial positiva**
  - No caso inverso, existe **autocorrelação espacial negativa**
- **Autocorrelação espacial nula** ocorre quando os valores dos atributos independem da localização

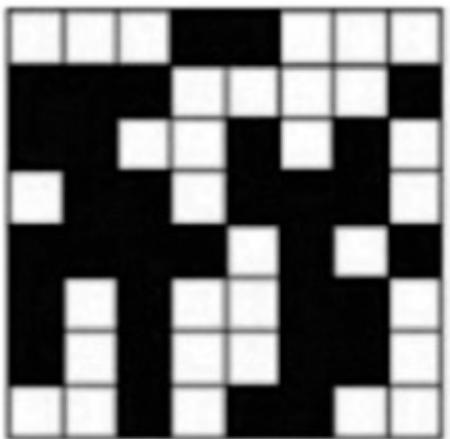
# Autocorrelação espacial



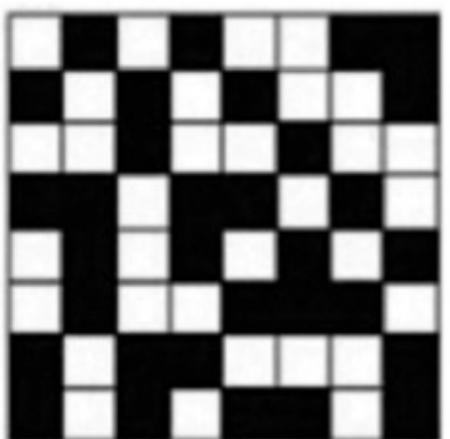
A  $i = -1.000$



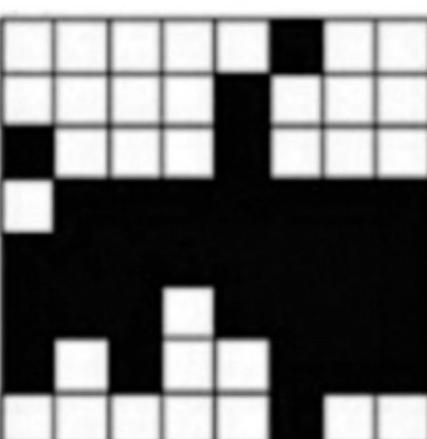
B  $i = +0.857$



C  $i = 0.000$



D  $i = -0.393$



E  $i = +0.393$

Extrema  
autocorrelação  
negativa

Extrema  
autocorrelação  
positiva

Independência  
espacial

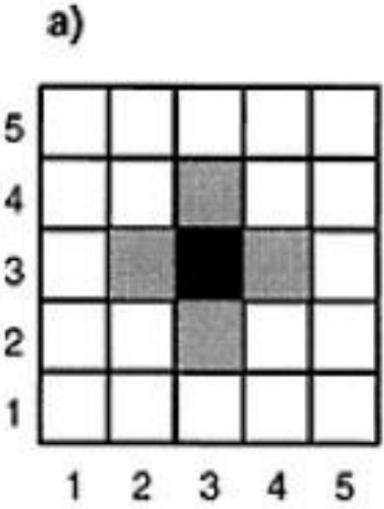
Arranjo disperso

Agrupamento  
espacial

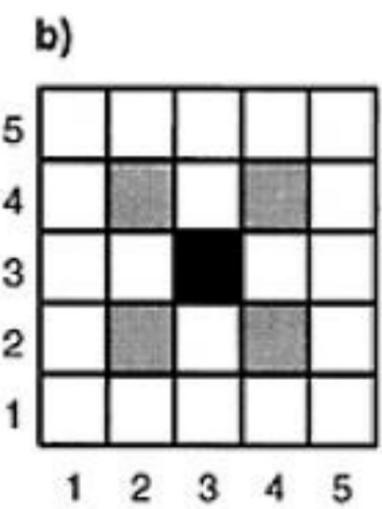
# Índice Globais de Autocorrelação Espacial

- I de Moran
  - Varia entre -1 (extrema autocorrelação negativa) e 1 (extrema autocorrelação positiva); 0 = ausência de correlação; independente da posição espacial
  - Semelhante à correlação de Pearson, tendo uma variável espacial em um dos eixos
  - É necessário definir uma medida de vizinhança espacial
- Geary
  - Varia entre 0 (máxima autocorrelação positiva), 1 (ausência de autocorrelação) e um valor positivo  $> 1$  indica autocorrelação negativa
  - Moran é considerado mais “poderoso”

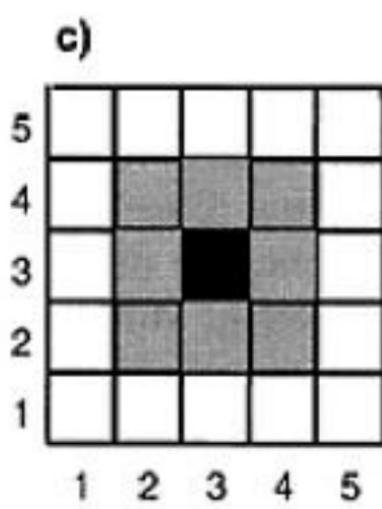
# Vizinhança espacial - Moran



Rook



Bishop

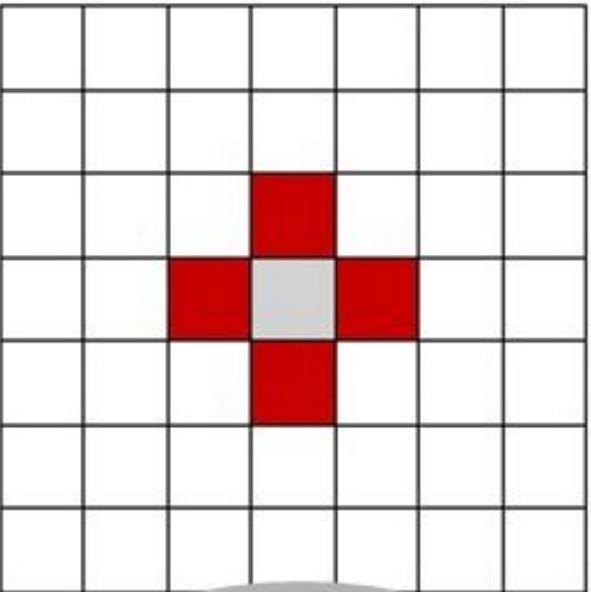


Queen

- É construída uma matriz de pesos, para ponderar a influência dos elementos vizinhos sobre o central
- Pode-se ainda considerar os k vizinhos mais próximos (KNN)

1<sup>st</sup>  
order

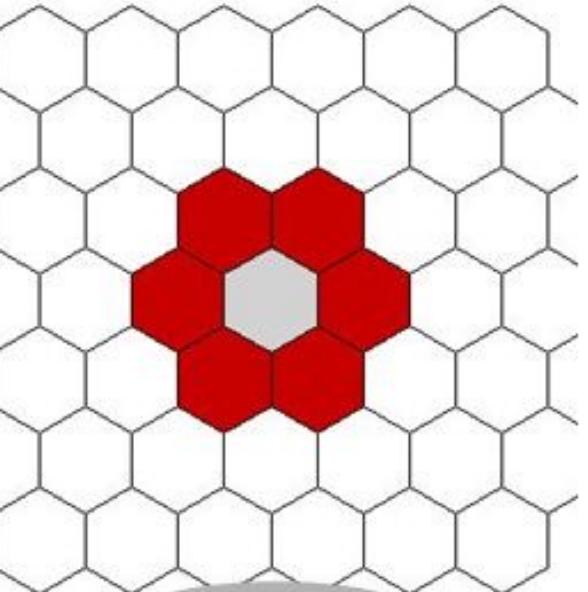
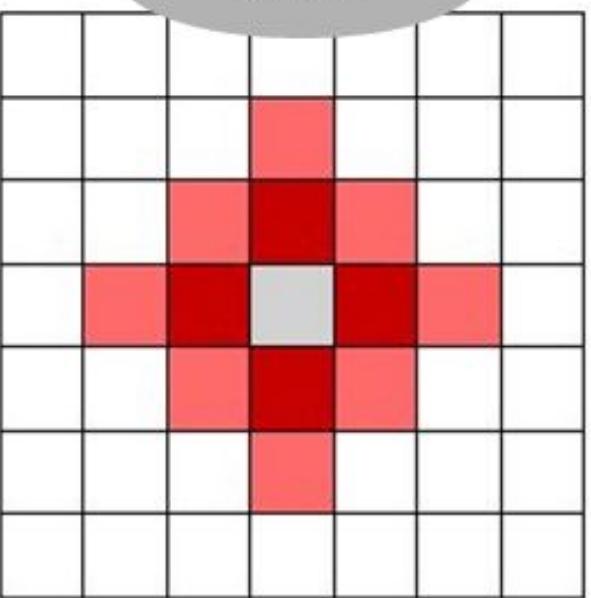
Nearest  
neighbor



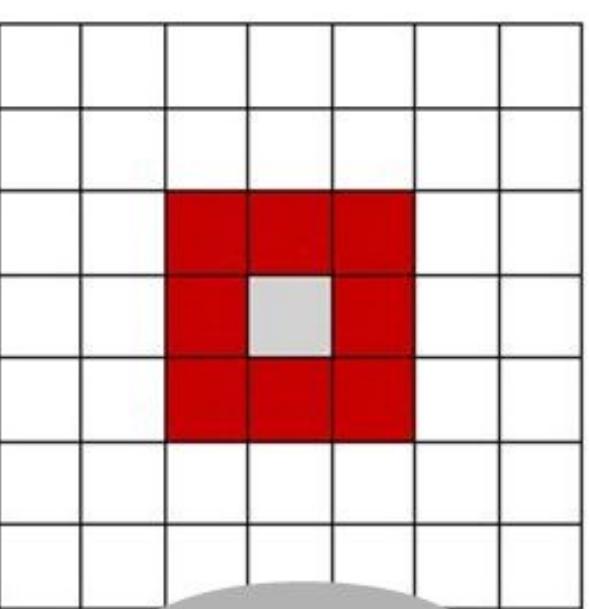
rook

2<sup>nd</sup>  
order

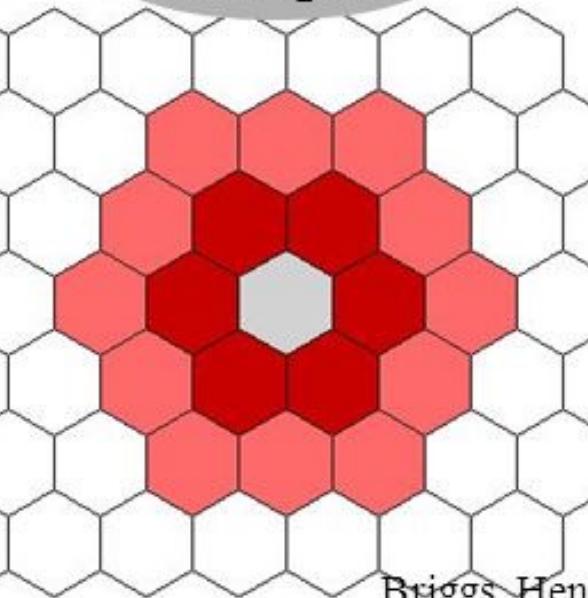
Next  
nearest  
neighbor



hexagon

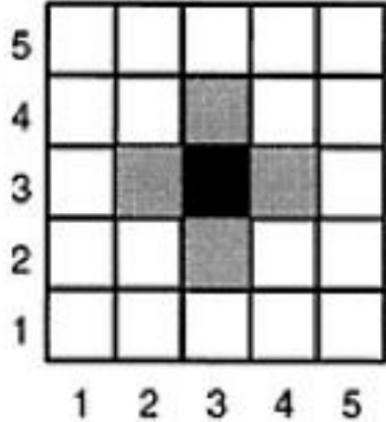


queen

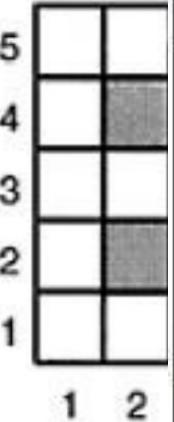


# Vizinhança espacial

a)

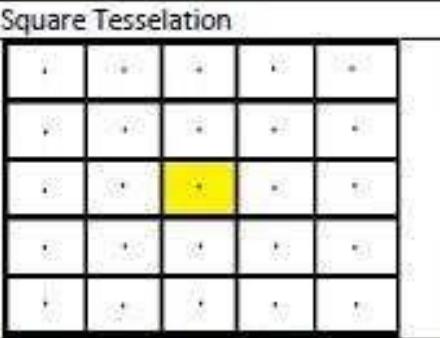


b)



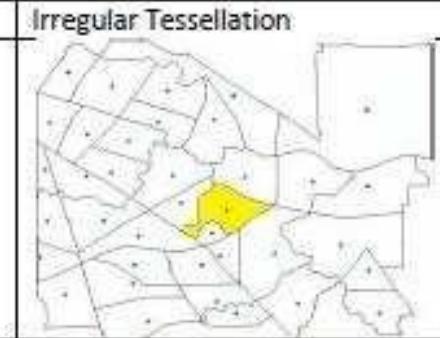
Rook

Bis



Neighbors for the given yellow region are shown.

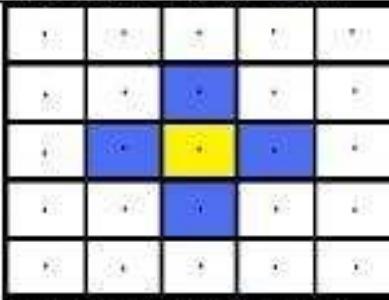
\*Note only one neighbor difference in rook and queen's cases for the irregular tessellation



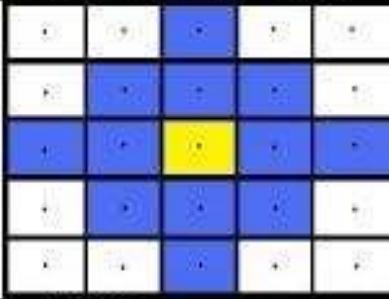
Rook's case: neighbors are connected along an edge

Queen's case: neighbors are connect on edges and at points

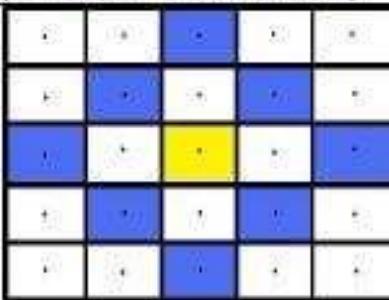
Rook's Adjacency (1<sup>st</sup> order)



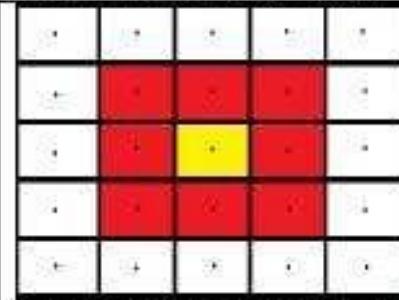
Rook's Adjacency (2<sup>nd</sup> order)



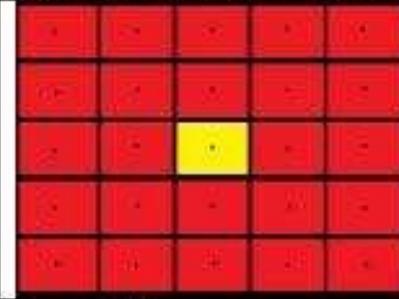
Rook's Adjacency (2<sup>nd</sup> order 1<sup>st</sup> order removed)



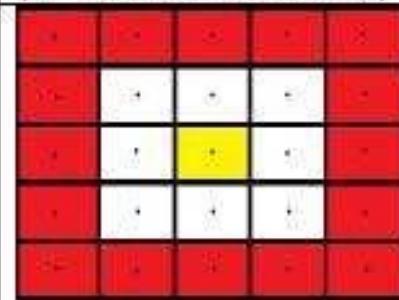
Queen's Adjacency (1<sup>st</sup> Order)



Queen's Adjacency (2<sup>nd</sup> Order)



Queen's Adjacency (2<sup>nd</sup> Order 1<sup>st</sup> order removed)



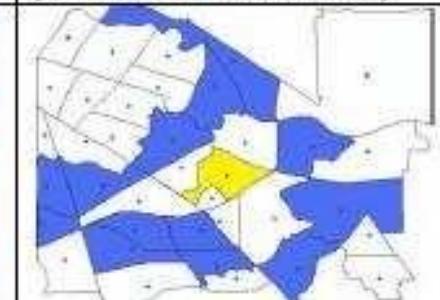
Rook's Adjacency (1<sup>st</sup> order)



Rook's Adjacency (2<sup>nd</sup> order)



Rook's Adjacency (2<sup>nd</sup> order 1<sup>st</sup> order removed)



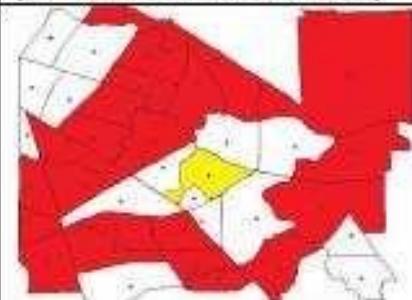
Queen's Adjacency (1<sup>st</sup> Order)

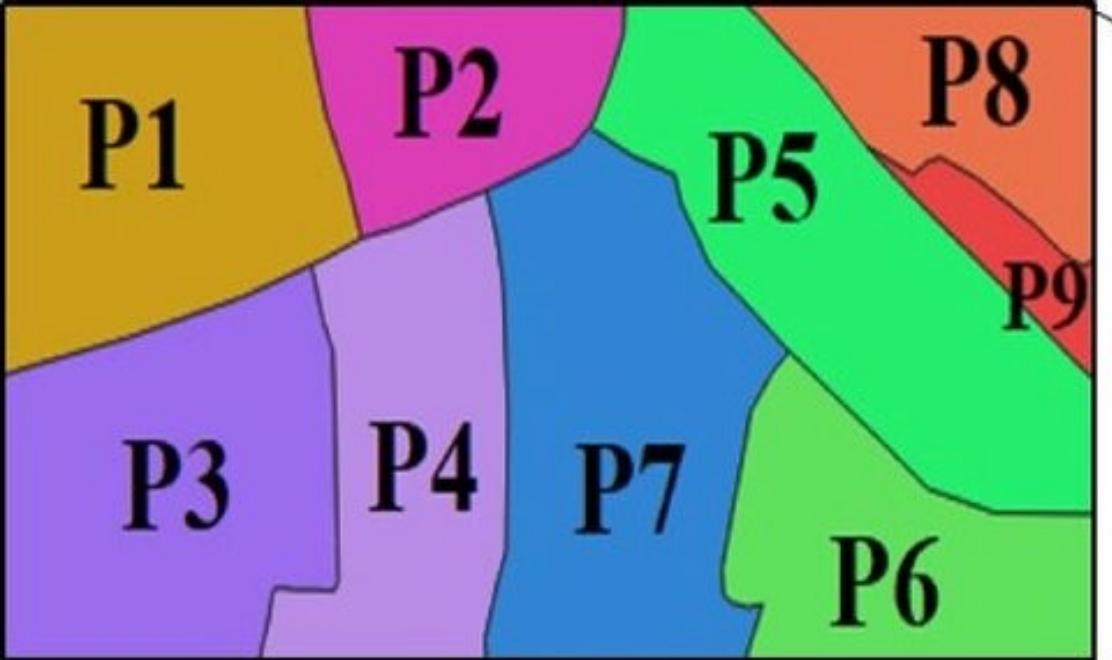


Queen's Adjacency (2<sup>nd</sup> Order)

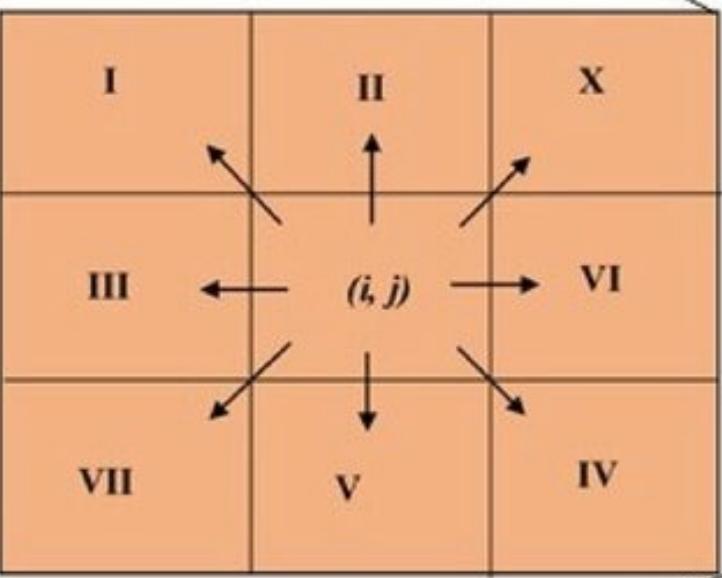


Queen's Adjacency (2<sup>nd</sup> Order 1<sup>st</sup> order removed)





	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	0	1	1	1	0	0	0	0	0
P2	1	0	0	1	1		1	0	0
P3	1	1		1	0	0	0	0	0
P4	1	1	1	0	0	0	1	0	0
P5	0	1	0	0	0	1	1	1	1
P6	0	0	0	0	1	0	1	0	0
P7	0	1	0	1	1	1	0	0	0
P8	0	0	0	0	1	0	0	0	1
P9	0	0	0	0	1	0	0	1	0



$$I = \frac{N \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij}) \sum_{i=1}^n (x_i - \bar{x})^2}$$

- Where:

$\frac{N}{X}$  is the number of observations (points or polygons)  
 $X$  is the mean of the variable  
 $X_i$  is the variable value at a particular location  
 $X_j$  is the variable value at another location  
 $w_{ij}$  is a weight indexing location of  $i$  relative to  $j$

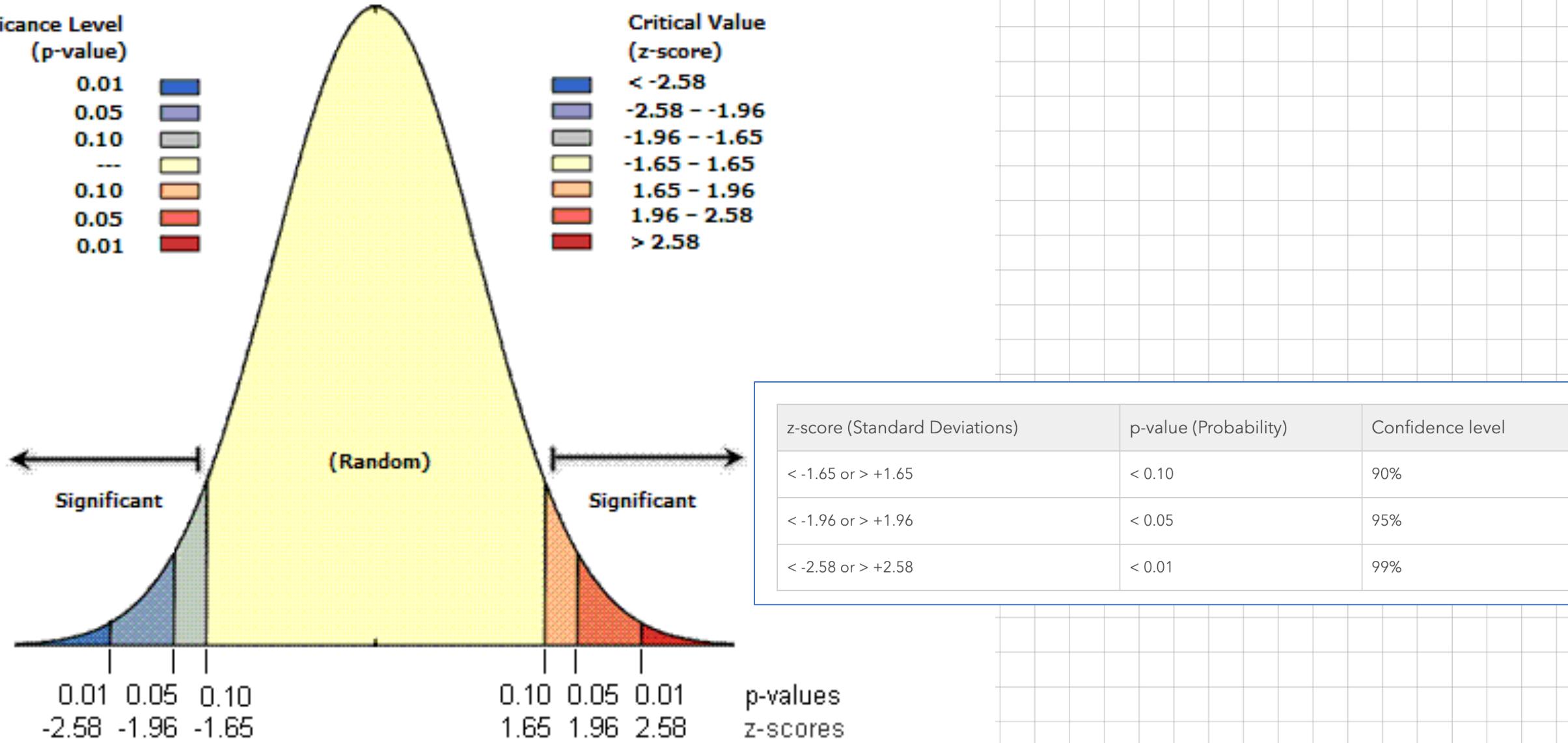
Material from Prof. Briggs UT Dallas

# Teste de pseudossignificância

- Usado para estimar a significância do índice
- São geradas muitas permutações dos valores dos atributos associados às regiões espaciais
  - Cada permutação produz um novo arranjo espacial
  - Como só um dos arranjos é o real, gera-se uma distribuição empírica do índice
- Um p-value é calculado: indica a probabilidade de que a **hipótese nula seja válida**
  - **Hipótese nula na autocorrelação espacial: distribuição espacial aleatória**
  - Corresponde aproximadamente à área sob a curva dessa distribuição
  - p-values pequenos indicam alta significância do índice
- Pode-se calcular um z-score também: ~ número de “desvios-padrões” em que se encontra o índice na distribuição produzida

# Teste de pseudossignificância – análise

- p-value não é estatisticamente significante (i.e., é “alto”)
  - Hipótese nula não pode ser rejeitada; os processos subjacentes à variável podem ser espacialmente aleatórios
- p-value é estatisticamente significante e z-score é positivo
  - A distribuição espacial da variável indica maior agrupamento do que se os processos subjacentes fossem espacialmente aleatórios
- p-value é estatisticamente significante e z-score é negativo
  - A distribuição espacial da variável indica maior espalhamento do que se os processos subjacentes fossem espacialmente aleatórios



# Exemplo: eleições 2022

- Dados obtidos no site do TSE:

<https://dadosabertos.tse.jus.br/dataset/resultados-2022>

- Selecionados dados de MG
- Transformação: código de município TSE -> código de município IBGE
- Filtro: candidatos a deputado federal
- Filtro: candidatos eleitos
- Agregação: soma dos votos por seção até o nível de município
- Integração: junção com uma tabela espacial de municípios, contendo dados de população em 2010
- Resultado: uma view por candidato eleito, contendo sua votação em cada município de MG

```
create table eleicoes_2022_mg (
DT_GERACAO date ,
HH_GERACAO varchar(20) ,
ANO_ELEICAO int ,
CD_TIPO_ELEICAO int ,
NM_TIPO_ELEICAO varchar(50) ,
NR_TURNO int ,
CD_ELEICAO int ,
DS_ELEICAO varchar(50) ,
DT_ELEICAO date ,
TP_ABRANGENCIA_ELEICAO char(3) ,
SG_UF char(2) ,
SG_UE char(2) ,
NM_UE varchar(30) ,
CD_MUNICIPIO int ,
NM_MUNICIPIO varchar(100) ,
NR_ZONA int ,
.
.
.
(...)
```

```
copy eleicoes_2022_mg  
from '/Users/clodoveu/Downloads/votacao_candidato_munzona_2022/vot'  
delimiter ';'  
csv header null '#NULO#' encoding 'ISO-8859-1'
```

```
create table munic_tse_ibge (
cod_tse int,
uf char(3),
nome_munic varchar(200),
capital int,
cod_ibge int)
```

```
copy munic_tse_ibge  
from '/Users/clodoveu/Downloads/municípios_brasileiros_tse.csv'  
delimiter ','  
csv header
```

```
alter table eleicoes_2022_mg add column cod_ibge int
```

```
create table mg_2022_depfed as
(select cod_ibge, nr_candidato, nm urna_candidato, sum(qt_votos_nominais_validos) as votos_total
from eleicoes_2022_mg em
where nr_candidato >= 1000 and nr_candidato <= 9999 and ds_sit_tot_turno LIKE 'ELEITO%'
group by cod_ibge, nr_candidato, nm urna_candidato
)
```

```
create view dep_1313 as
(select *, (1.0 * votos_total / pop2010) as perc_votos
from mg_2022_depfed d, munbrasilpop m
where d.cod_ibge = m.mun
and nr_candidato = 1313)

create view dep_2222 as
(select *, (1.0 * votos_total / pop2010) as perc_votos
from mg_2022_depfed d, munbrasilpop m
where d.cod_ibge = m.mun
and nr_candidato = 2222)

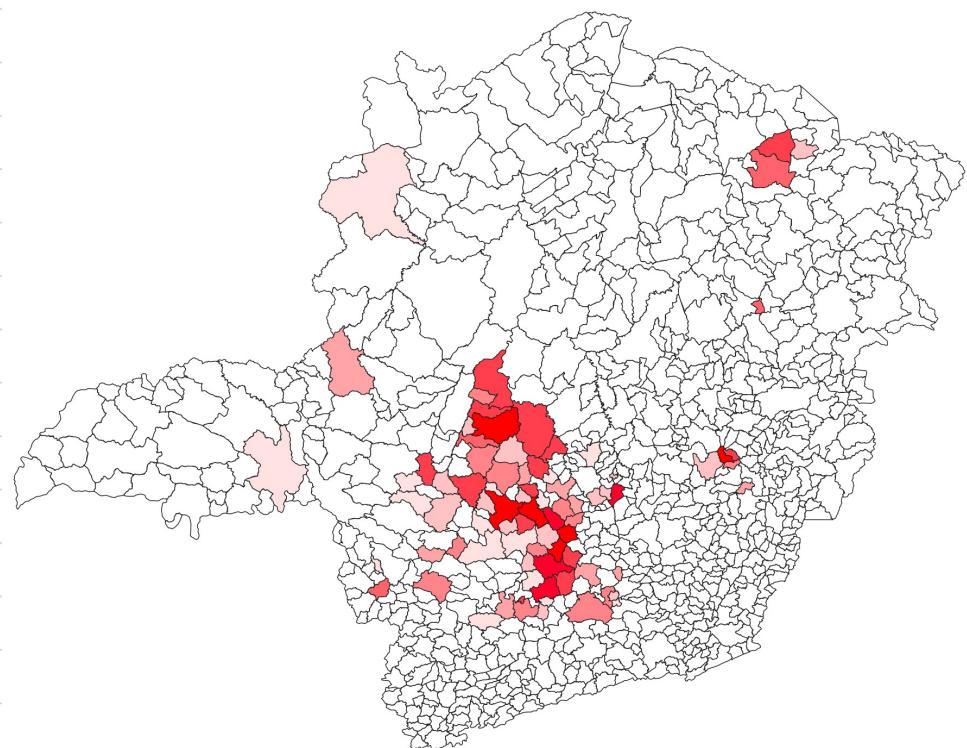
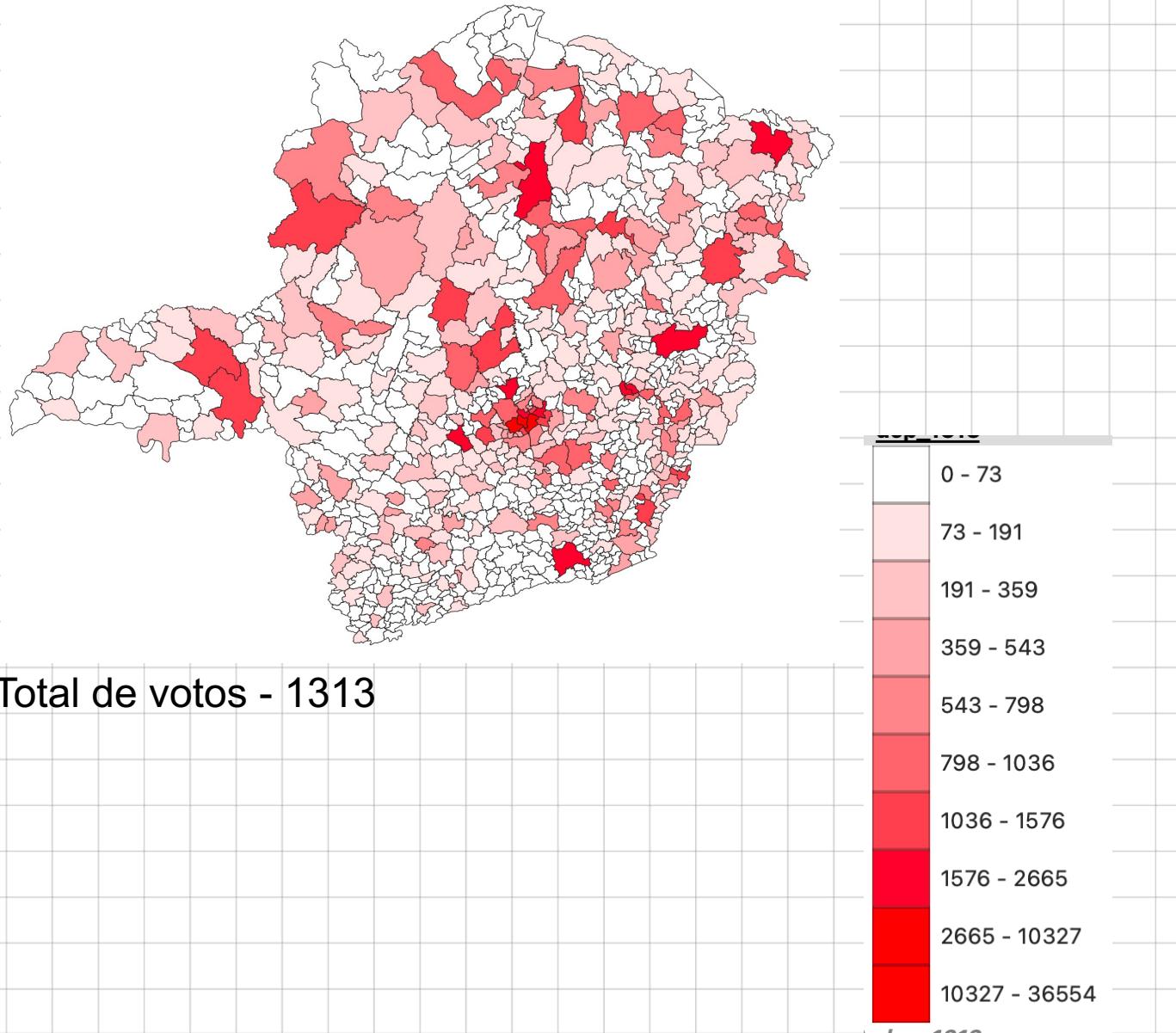
(...)
```

Variável incluída: percentual (aprox.) de votos no município = votos/pop2010

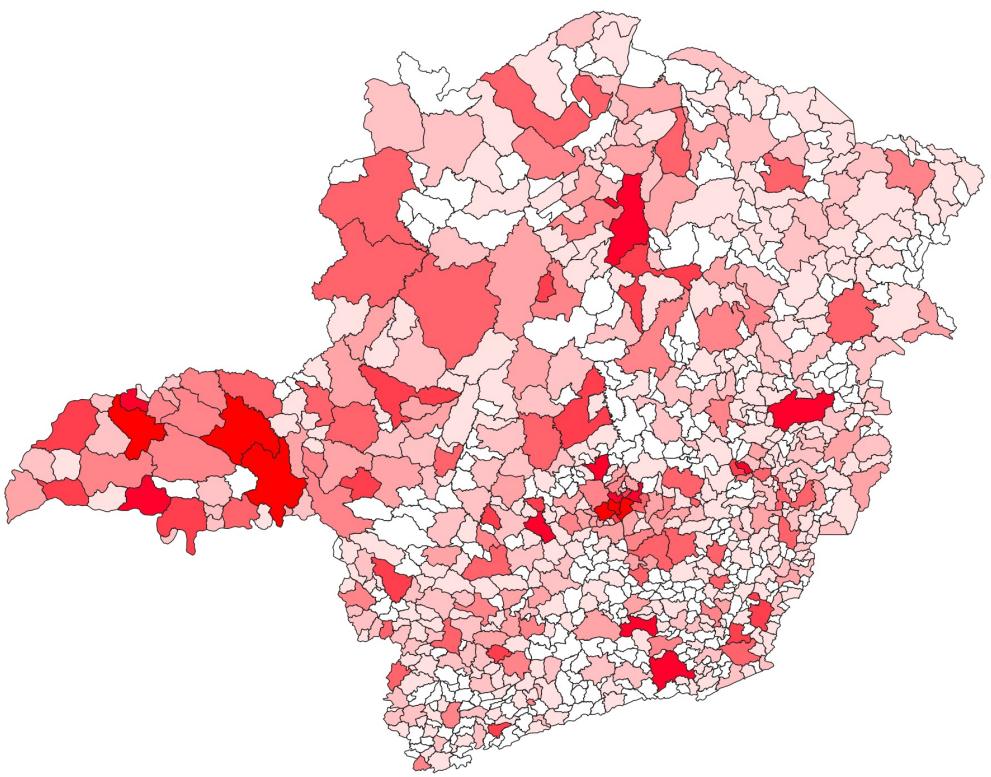
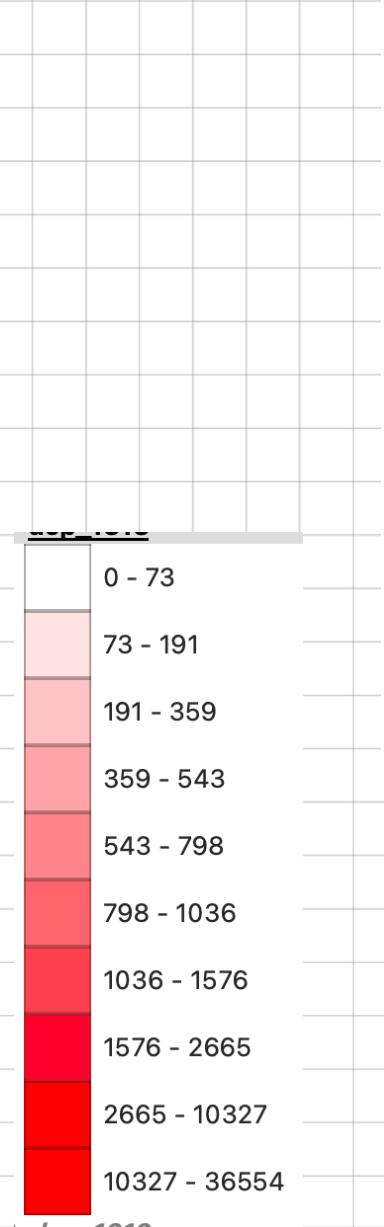
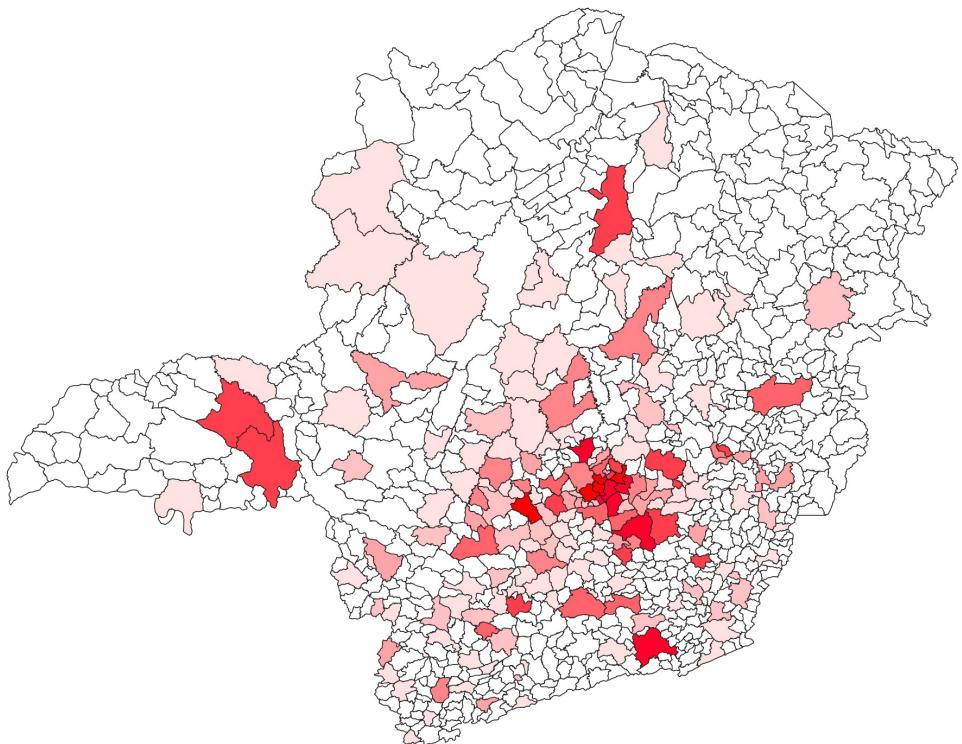
## Selecionados para análise:

- 1212 – Duda Salabert (PDT) (2º lugar)
- 1313 – Rogério Correia (PT) (5º lugar)
- 2222 – Nikolas Ferreira (PL) (1º lugar)
- 2260 – Domingos Sávio (PL) (28º lugar - ex-prefeito)
- 7040 – André Janones (Avante) (4º lugar)

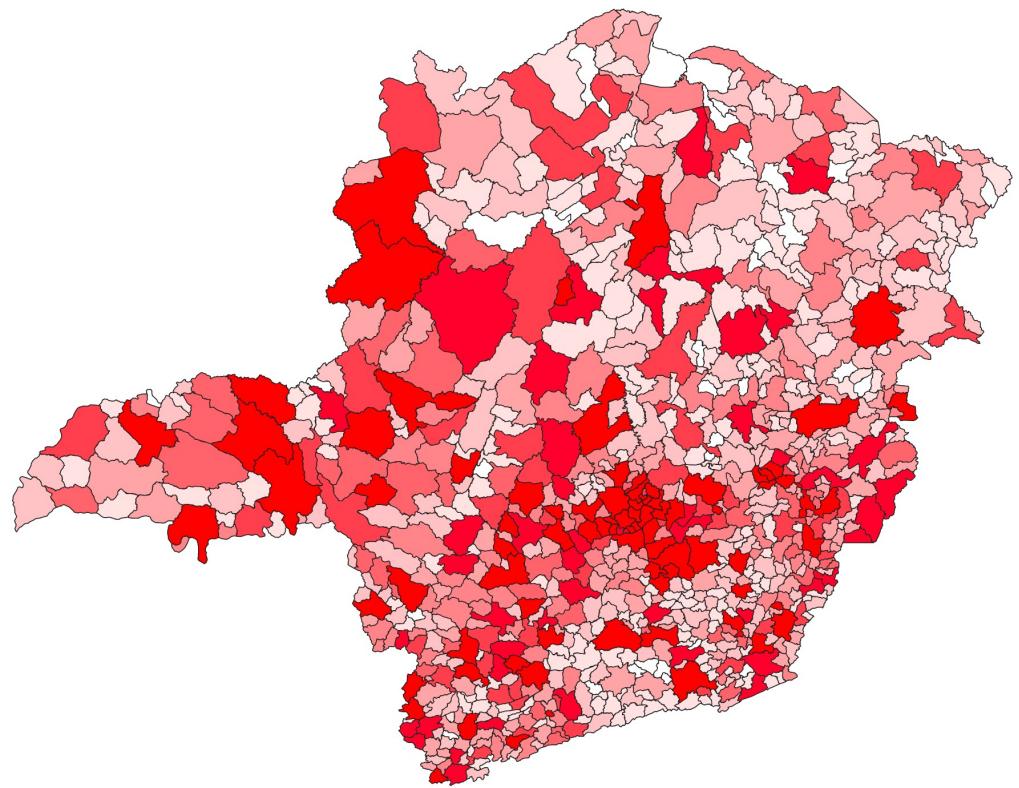
# Análise visual exploratória



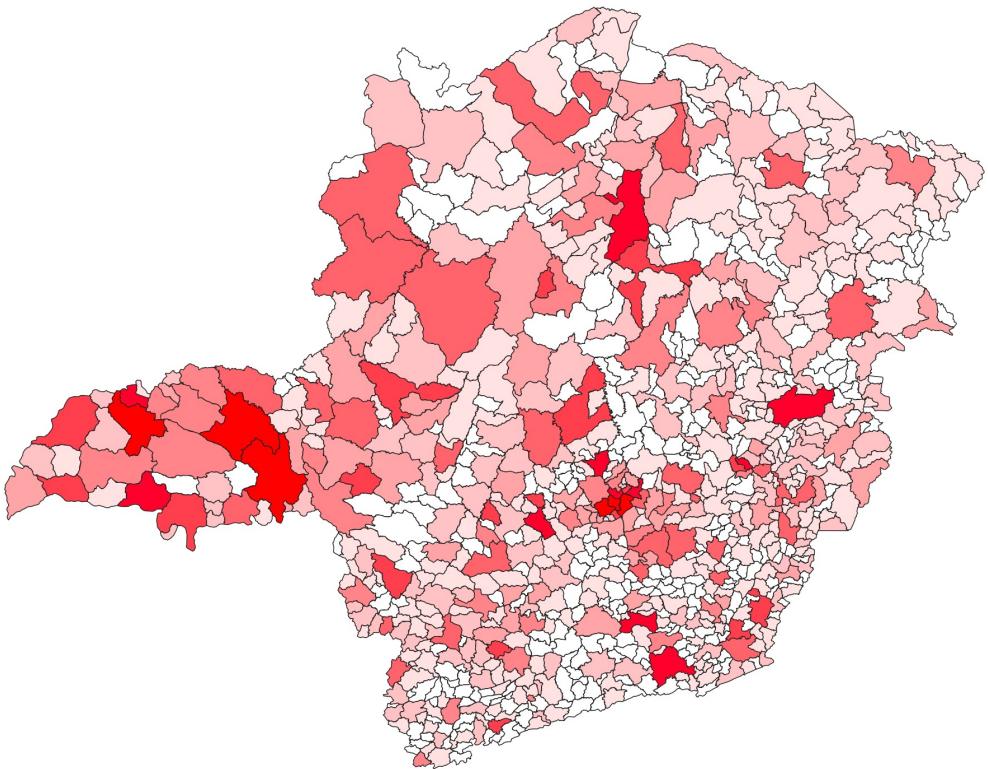
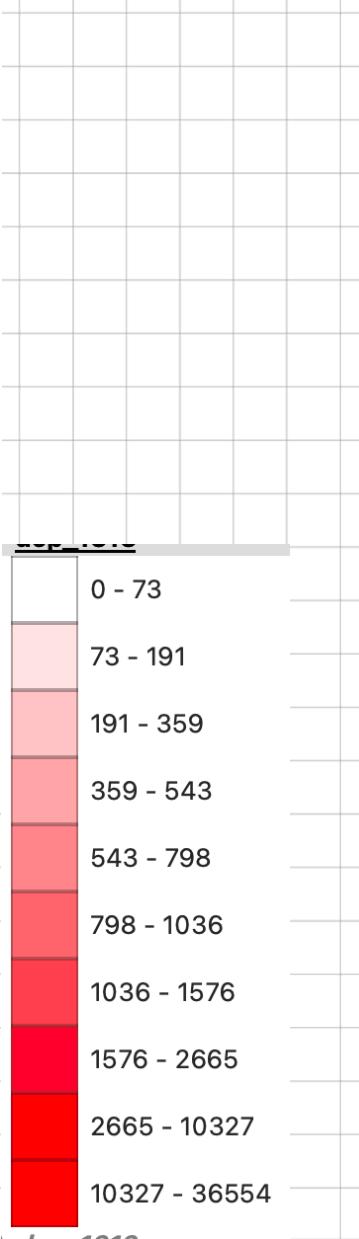
# Análise visual exploratória



# Análise visual exploratória



Total de votos - 2222

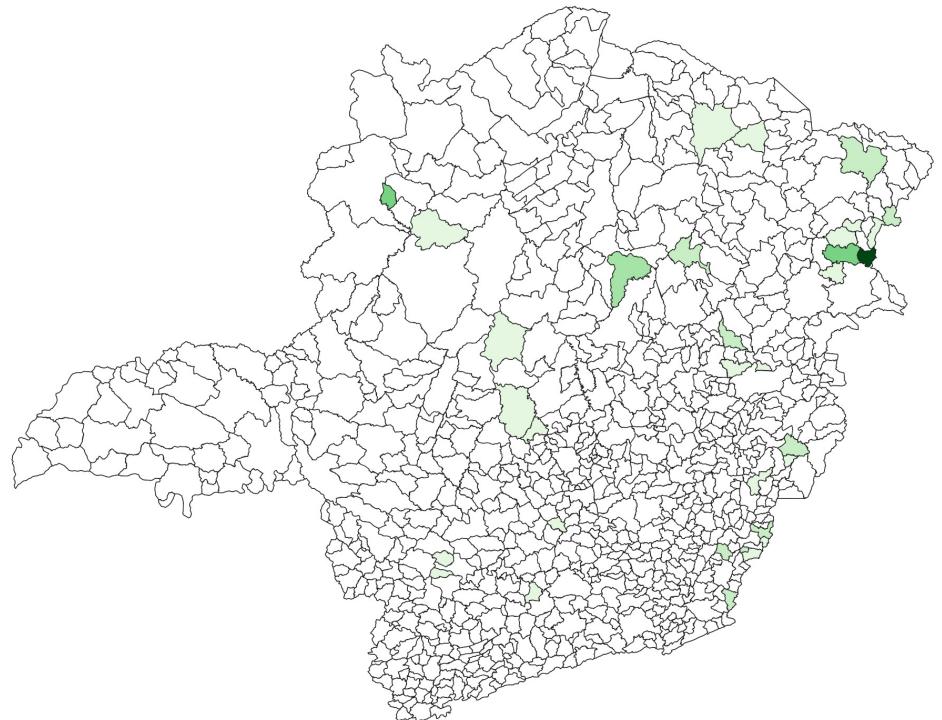


Total de votos - 7040

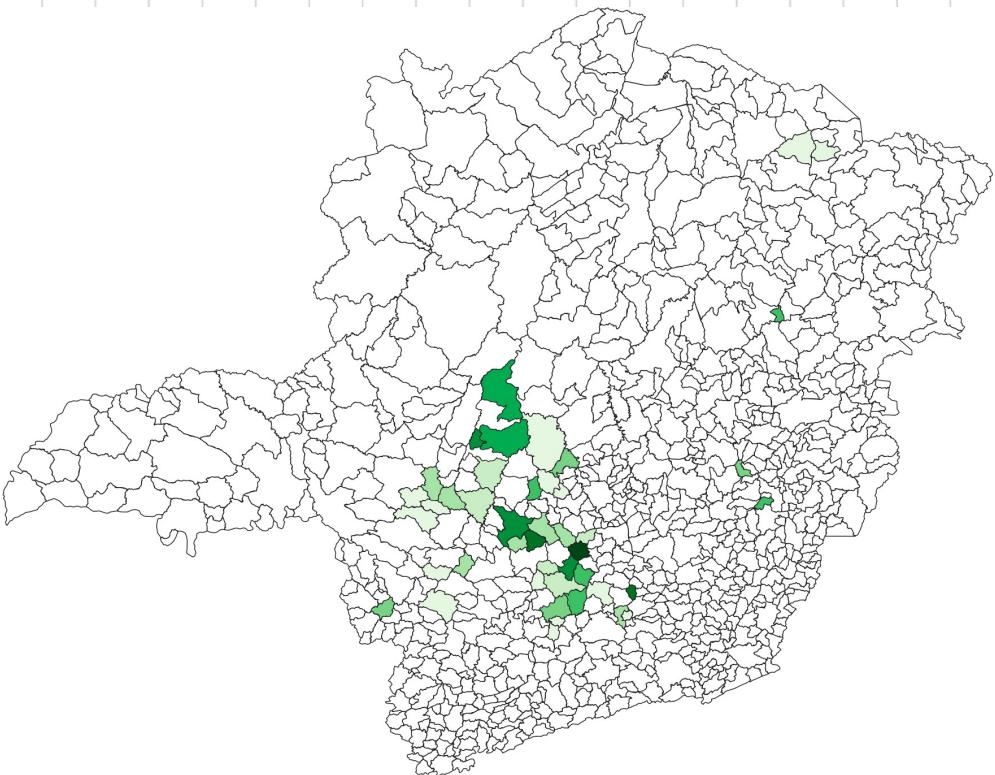
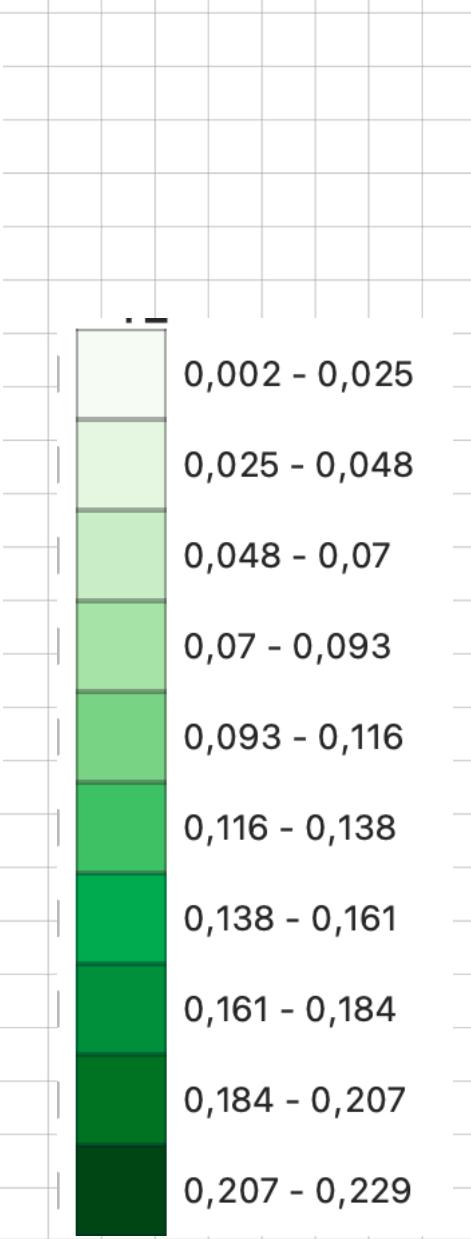
## Problemas/dúvidas

- Se a escala de cores do mapa não é constante, não se consegue comparar visualmente os mapas
- Mesmo sendo a mesma, há muitos polígonos (853)
- Além disso, pode ocorrer saturação na última faixa da legenda, ocultando variações acima do valor máximo
- Municípios de grande território chamam a atenção visualmente, mas podem ter população/eleitorado pequeno
- Alguns municípios de grande eleitorado são muito pequenos
- > Normalizar pelo eleitorado. Aproximação: população

# Análise visual exploratória

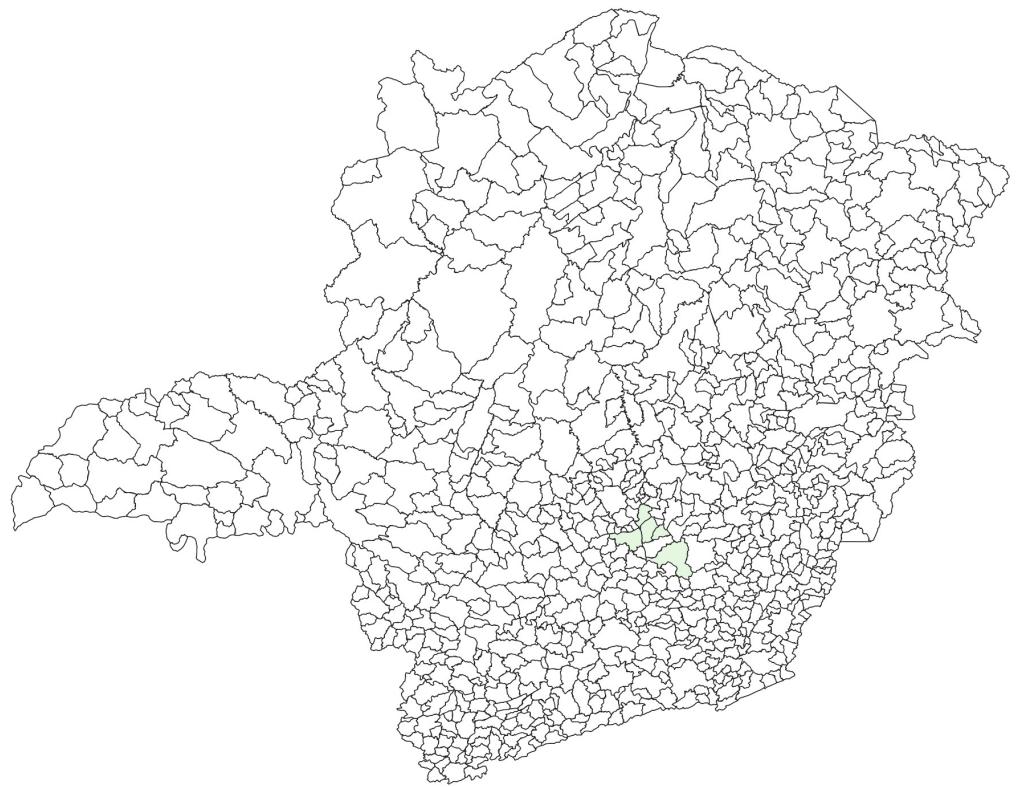


Percentual de votos - 1313

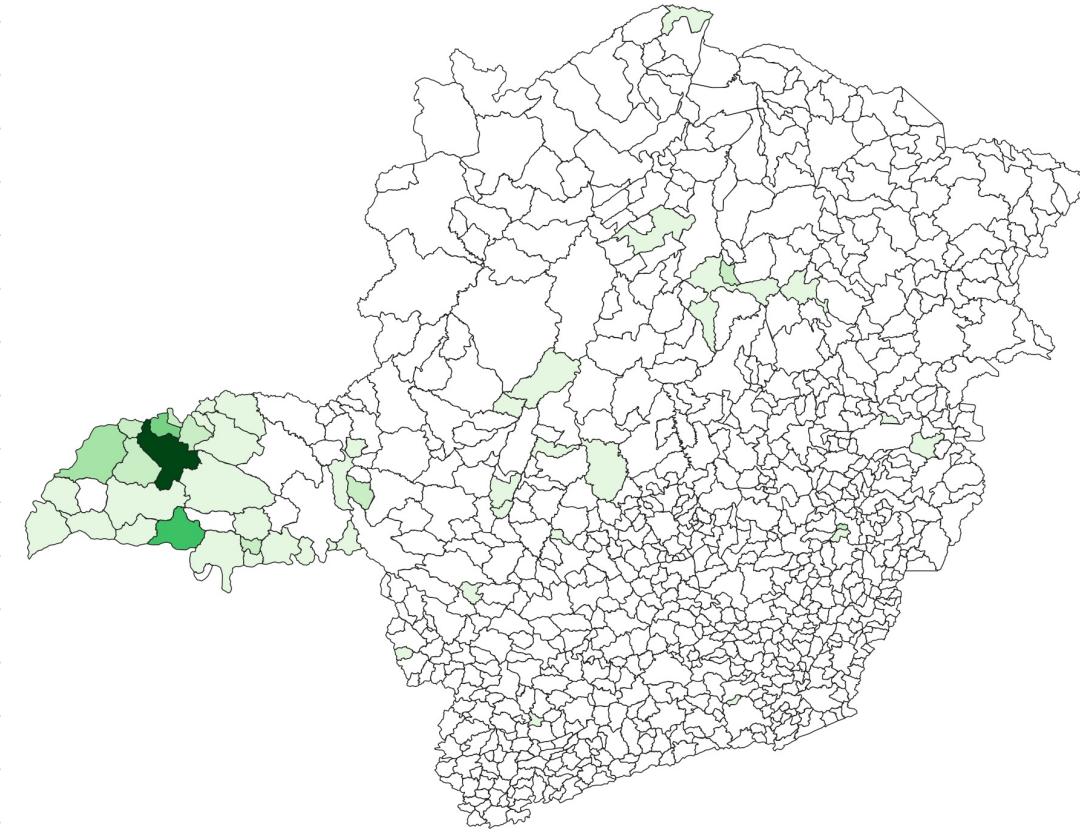
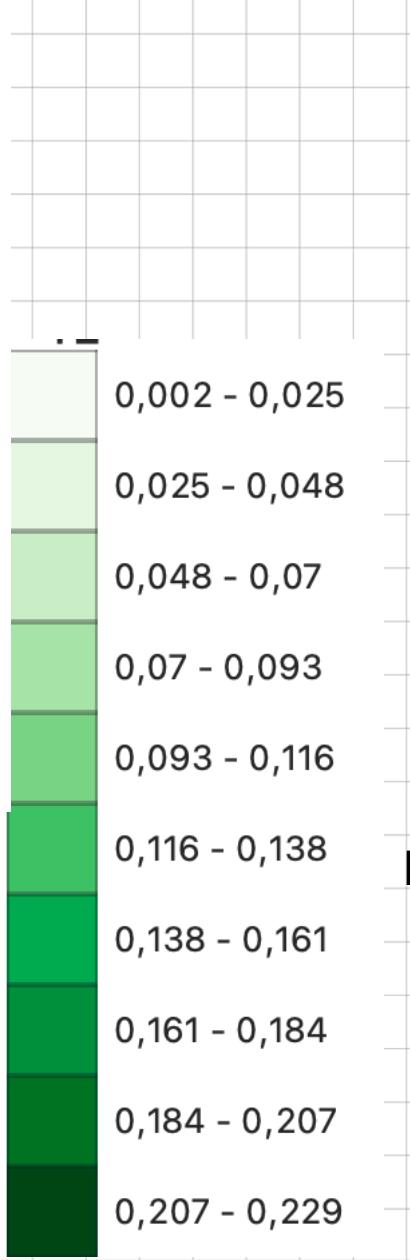


Percentual de votos - 2260

# Análise visual exploratória

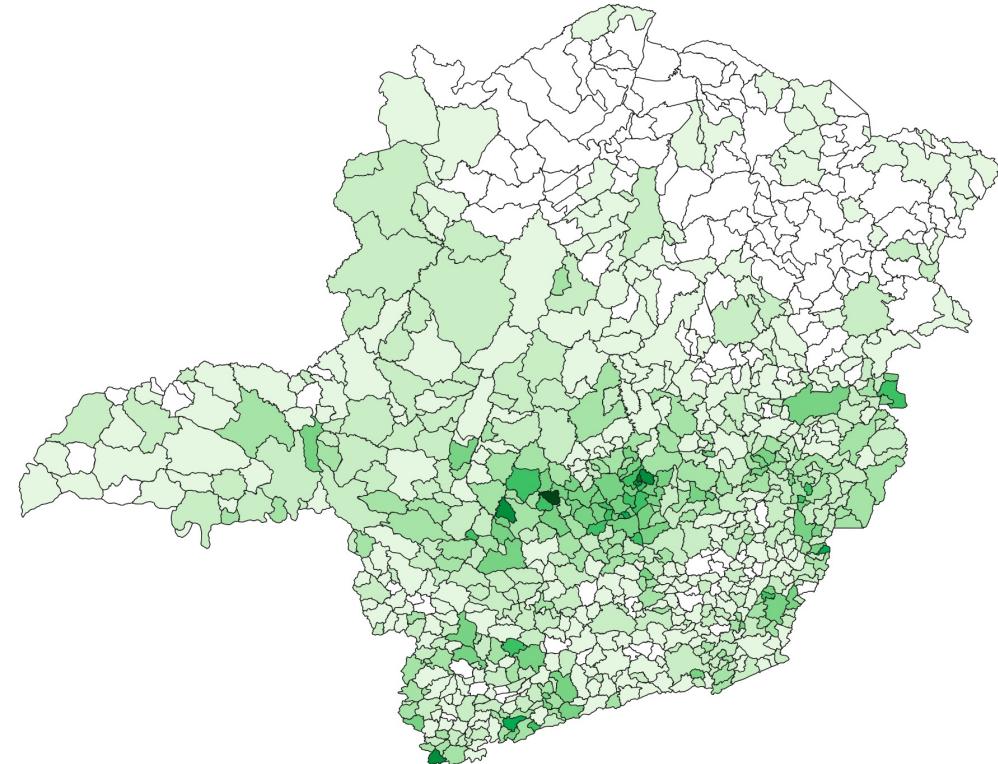


Percentual de votos - 1212

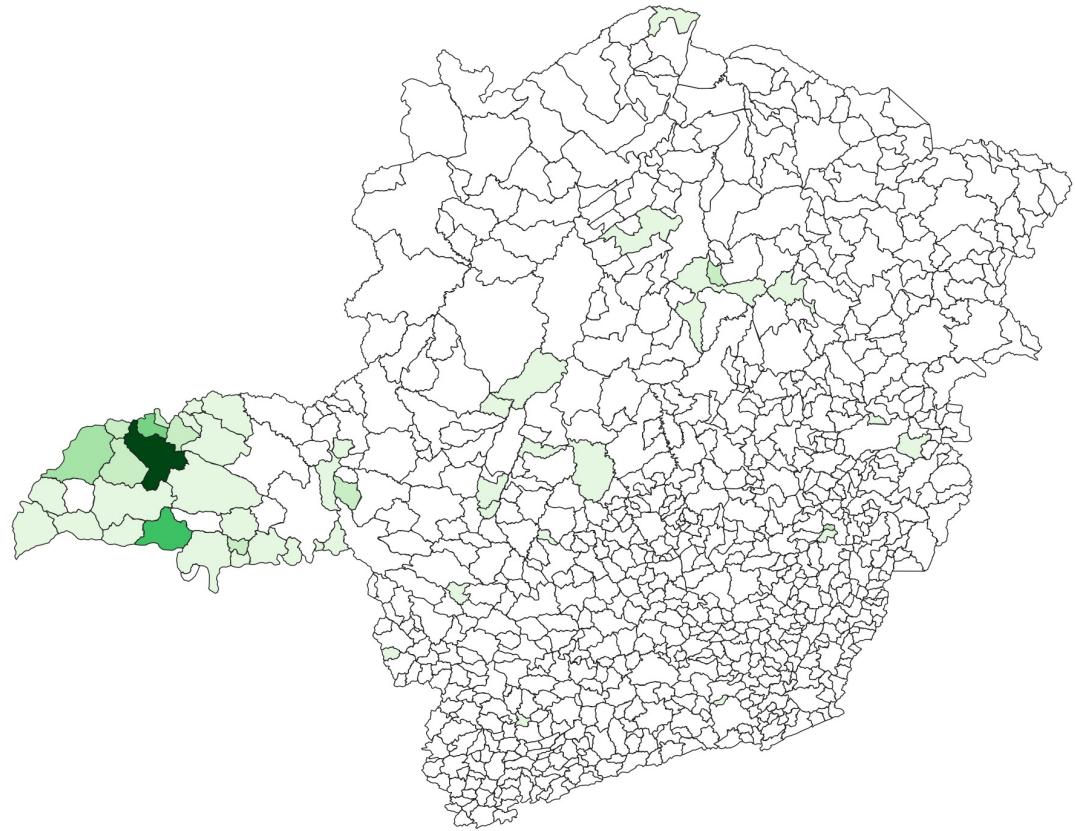
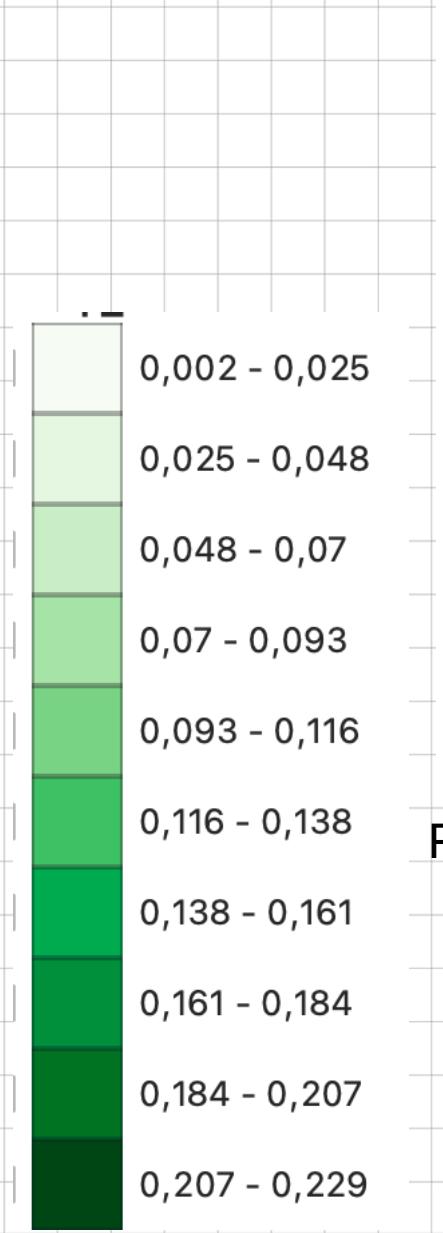


Percentual de votos - 7040

# Análise visual exploratória



Percentual de votos - 2222



Percentual de votos - 7040

## Problemas/dúvidas

- Ainda há dificuldade com a escala de cores, pela grande variação entre os candidatos
- Máximos:
  - 1212 → 4,1% (Ouro Preto)
  - 1313 → 34,9% (Umburatiba)
  - 2222 → 22,9% (Nova Serrana)
  - 2260 → 30,3% (Ribeirão Vermelho)
  - 7040 → 23,9% (Ituiutaba)

# Análise de autocorrelação espacial

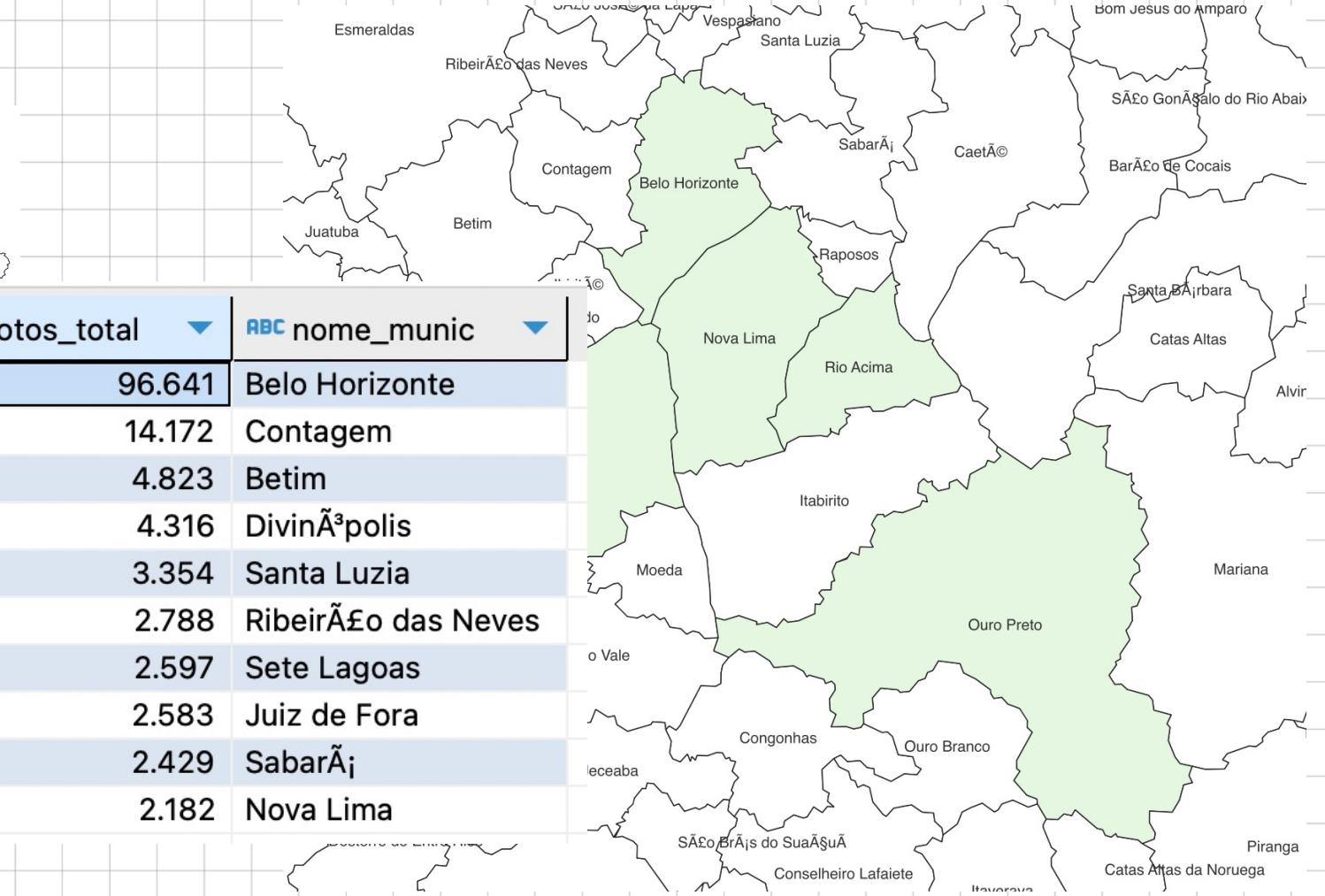
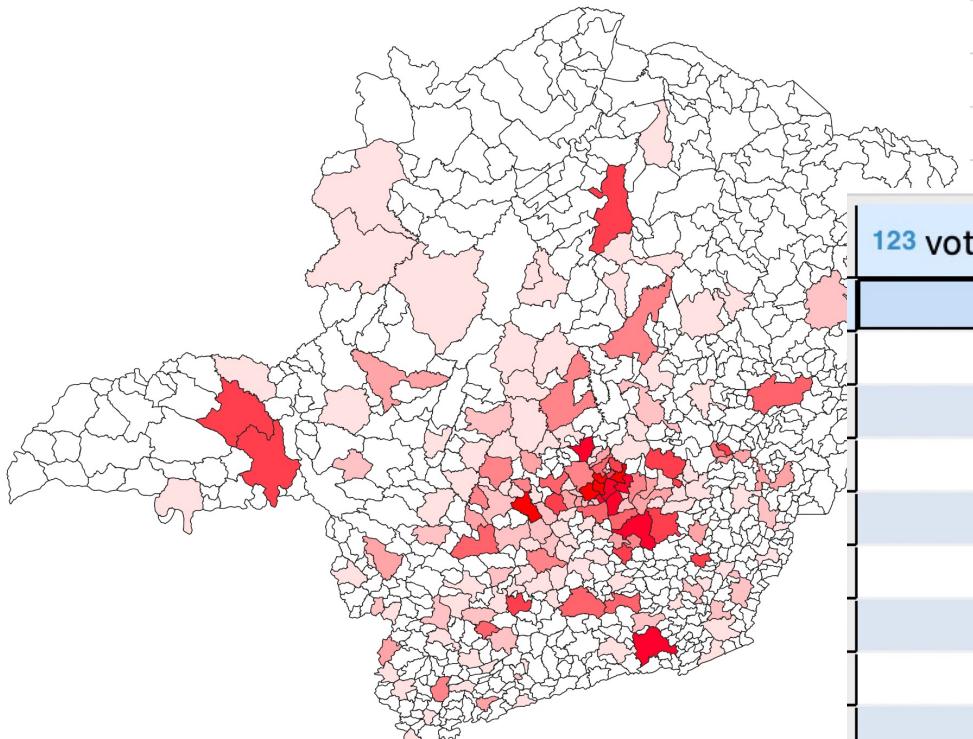
- A origem dos votos recebidos pelos candidatos é regionalizada?
- Existem grupos regionais de municípios que os elegem?

# Análise de autocorrelação espacial

Candidato	I Moran	P-value	Z-score	Obs
1212	0,71937	< 0.001	34,265	Concentrado
1313	0,16318	< 0.001	7,816	Distribuído
2222	0,60399	< 0.001	28,779	Concentrado
2260	0,34493	< 0.001	16,459	Regionalizado
7040	0,45814	< 0.001	21,543	Regionalizado

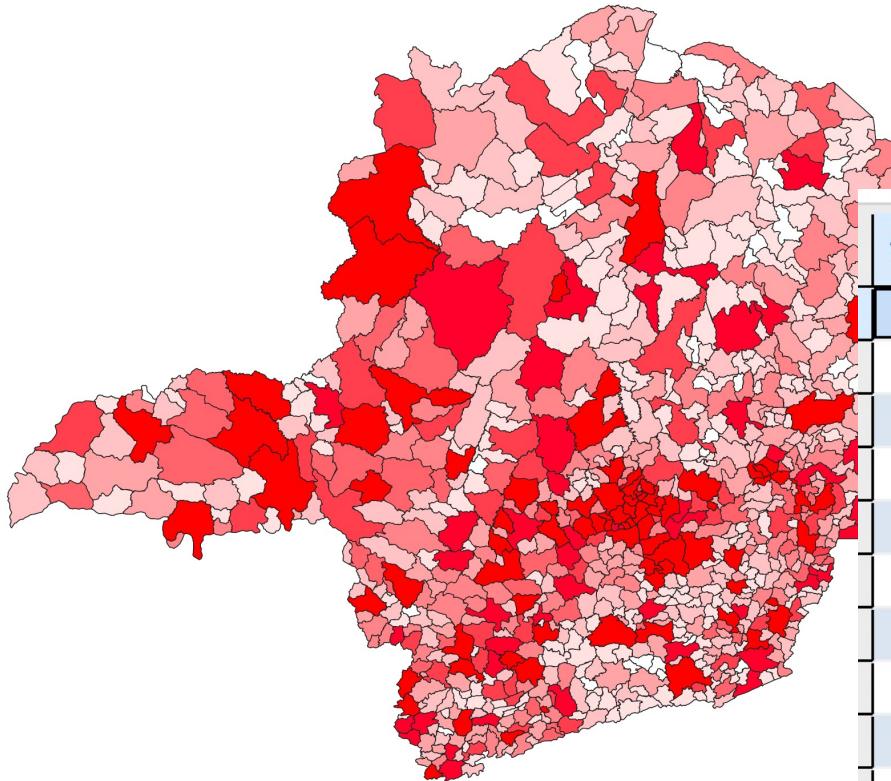
- Variável: percentual de votos no município
- Contiguidade: QUEEN

1212

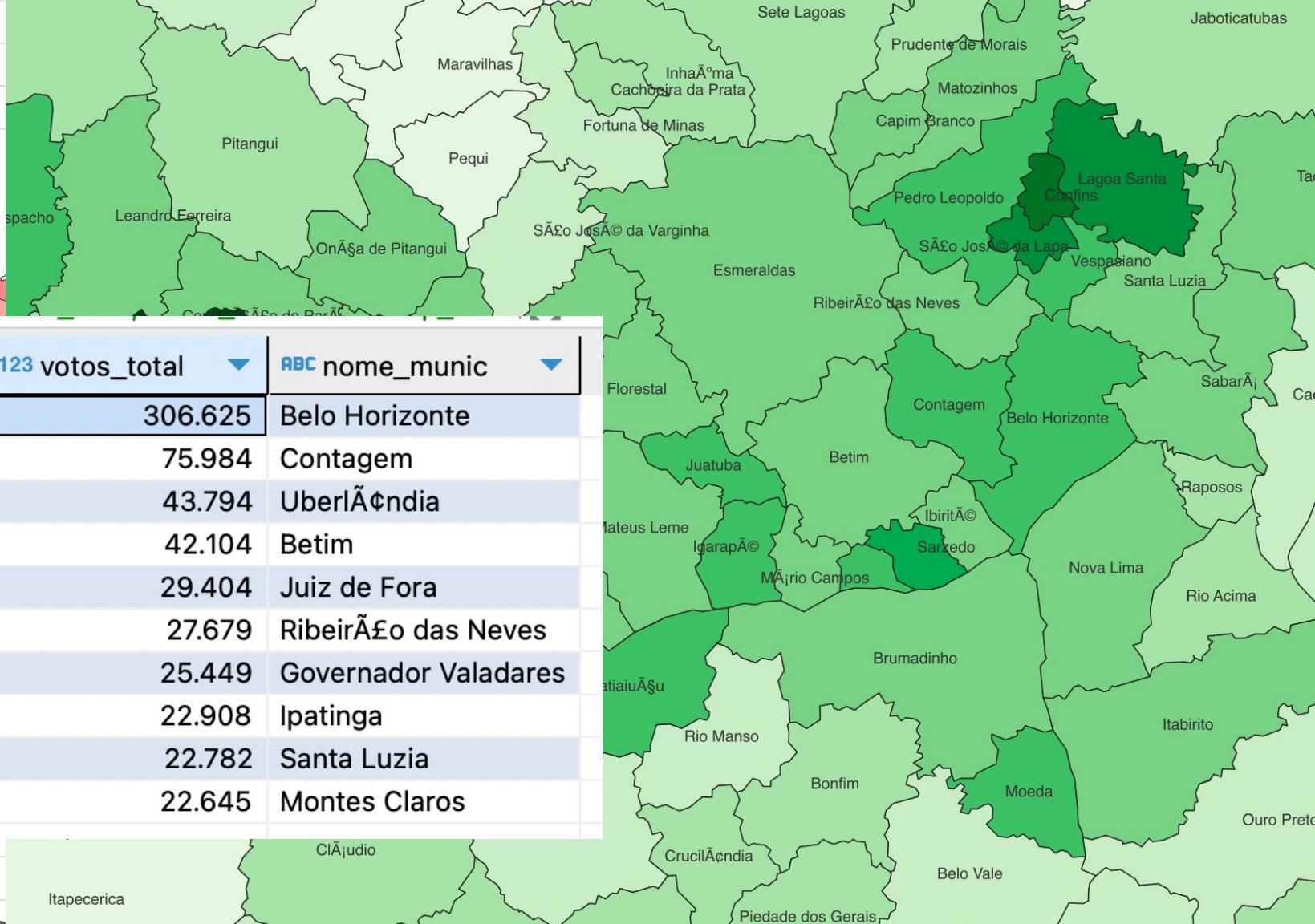


- Votos vindos de várias regiões do estado, mas apenas RMBH é significativa

**2222**

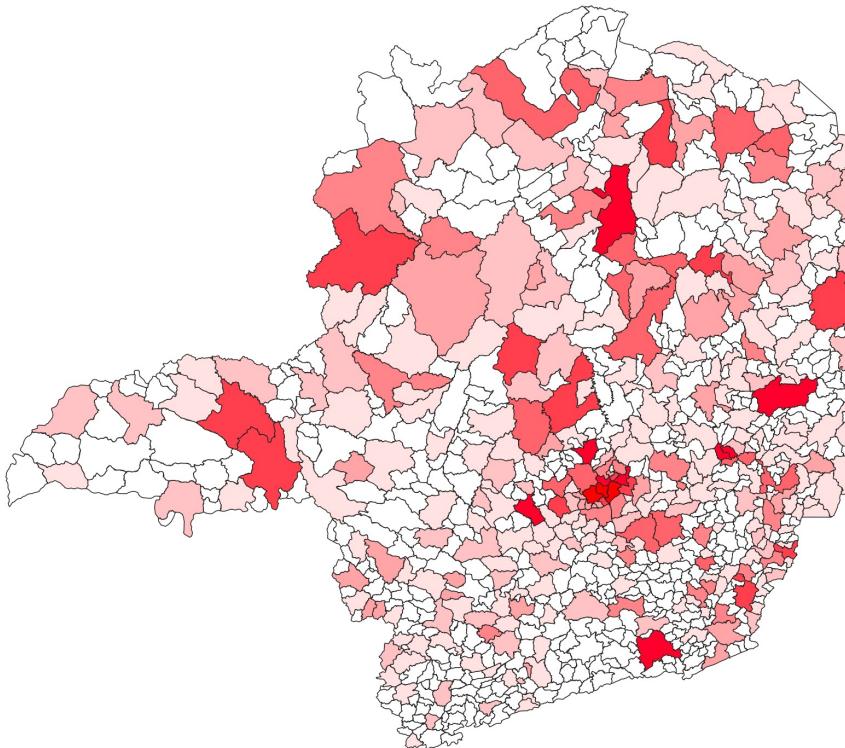


123 votos_total	ABC nome_munic
306.625	Belo Horizonte
75.984	Contagem
43.794	Uberlândia
42.104	Betim
29.404	Juiz de Fora
27.679	Ribeirão das Neves
25.449	Governador Valadares
22.908	Ipatinga
22.782	Santa Luzia
22.645	Montes Claros

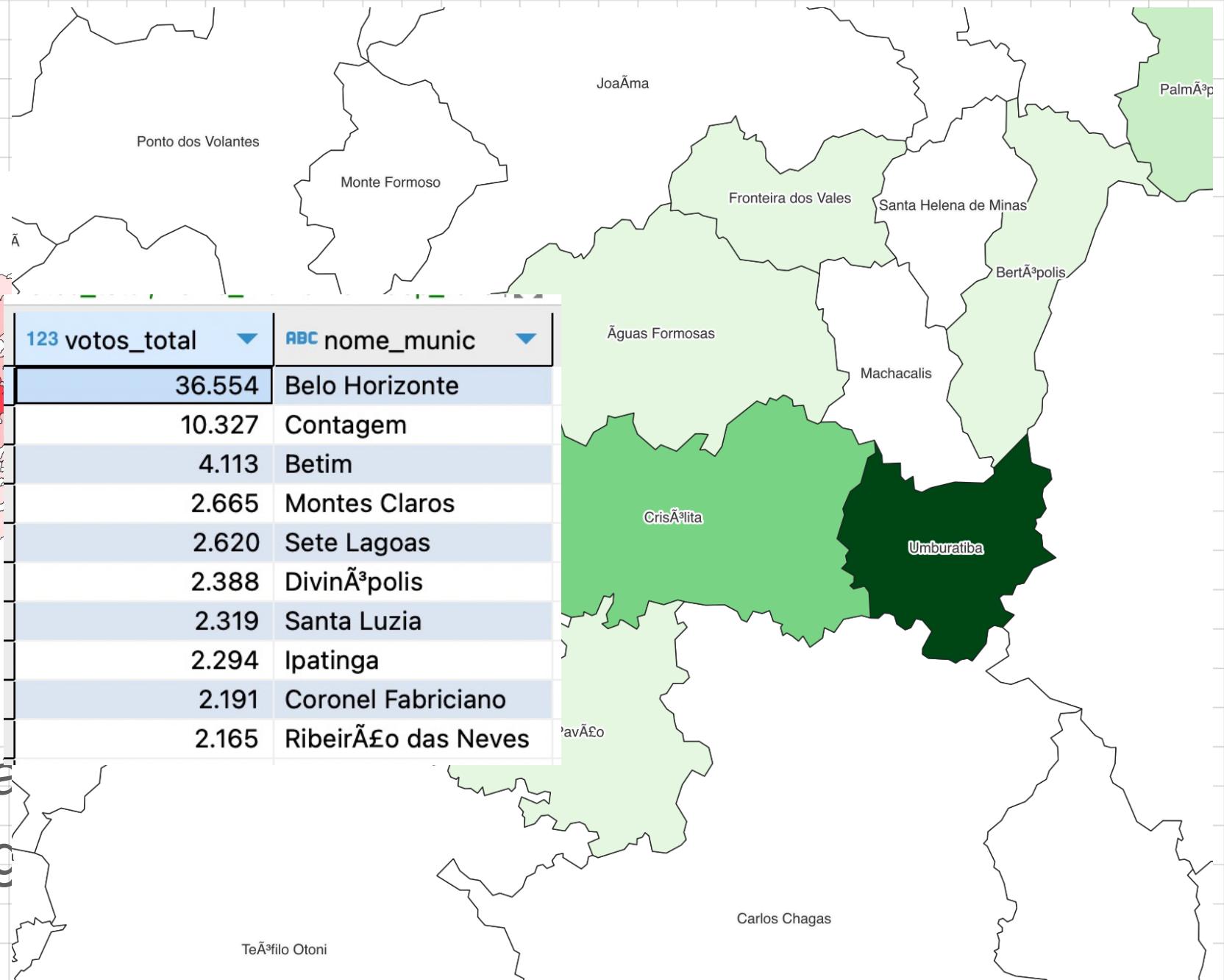


- Votos em todo o estado
  - Maior volume de votos em municípios mais populosos

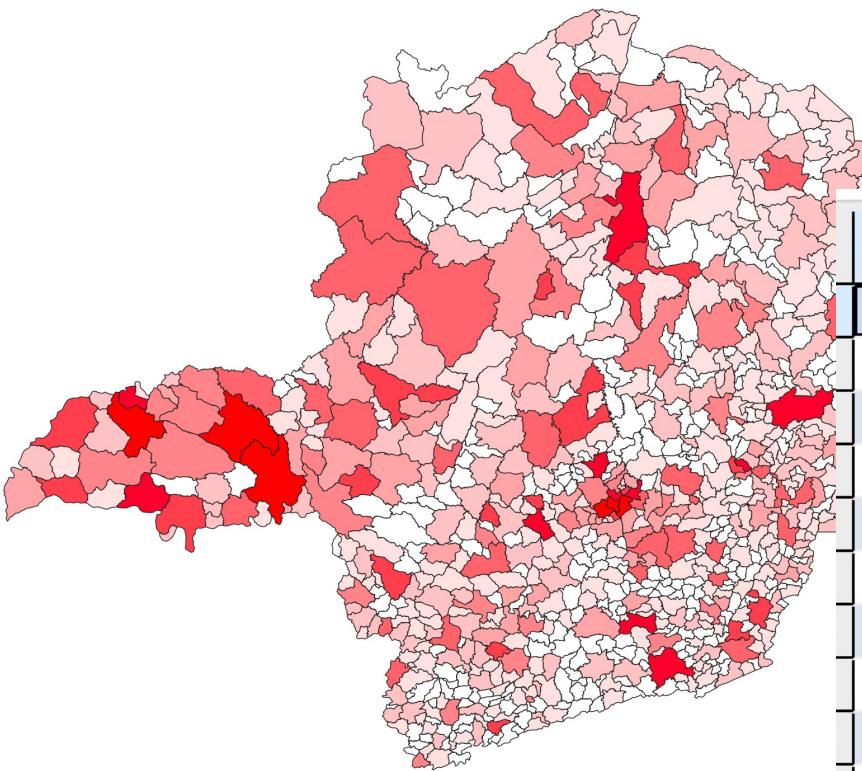
1313



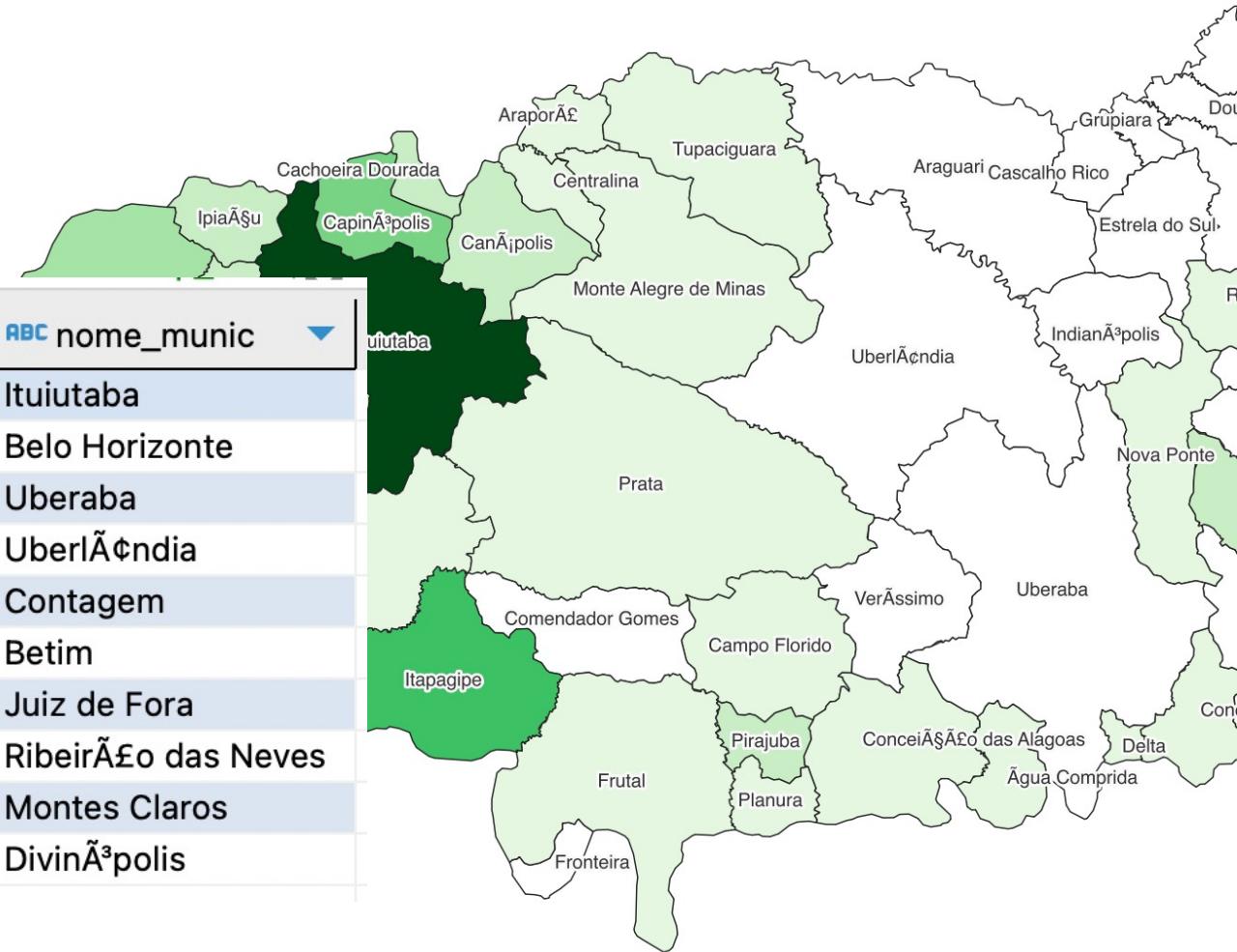
Votos vindos de va  
região central é sig  
em outra região



7040

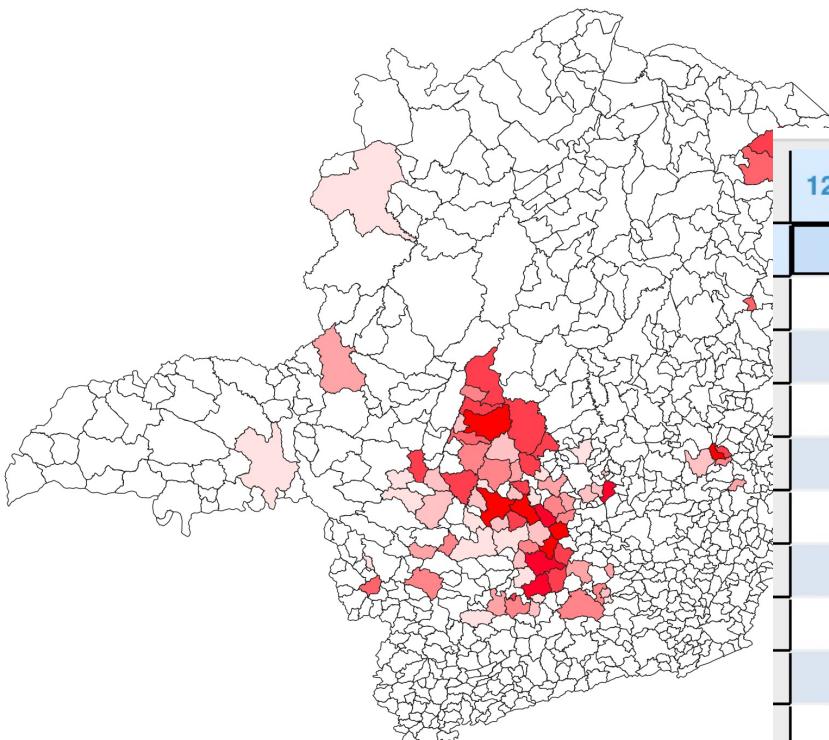


123 votos_total	ABC nome_munic
23.200	Ituiutaba
13.793	Belo Horizonte
6.081	Uberaba
4.957	Uberlândia
4.357	Contagem
3.491	Betim
2.582	Juiz de Fora
2.419	Ribeirão das Neves
2.192	Montes Claros
2.088	Divinópolis

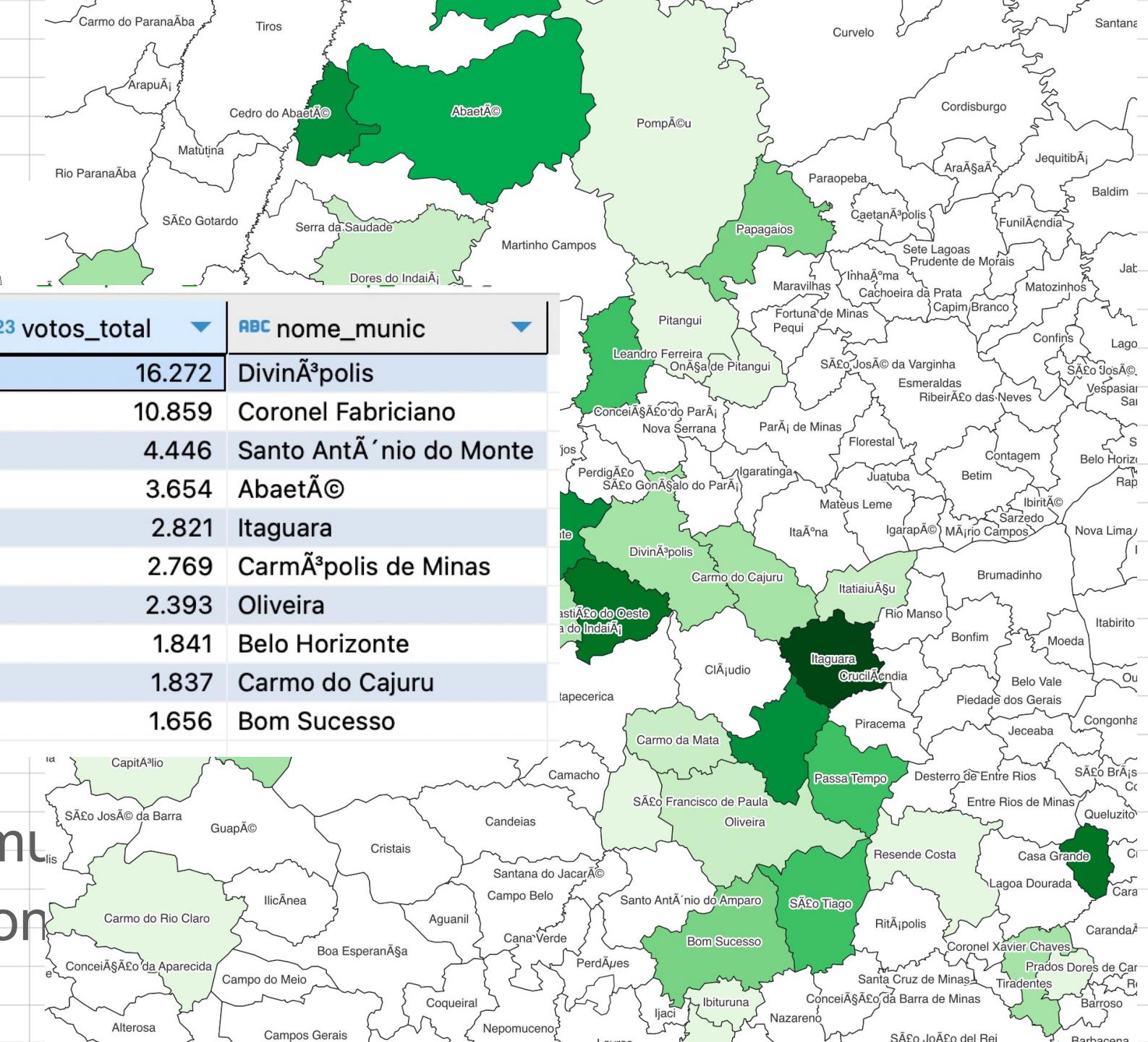


- Votos em muitos m...
- Concentração de relevância no interior (interior) e votos em municípios maiores

2260



123 votos_total	ABC nome_munic
16.272	Divinópolis
10.859	Coronel Fabriciano
4.446	Santo Antônio do Monte
3.654	Abaeté
2.821	Itaguara
2.769	Carmópolis de Minas
2.393	Oliveira
1.841	Belo Horizonte
1.837	Carmo do Cajuru
1.656	Bom Sucesso



- Votos em poucos municípios
- Concentração regionalizada

## Moran Local

## Local Indicators of Spatial Association (LISA)

- O I de Moran oferece um único valor para quantificar a associação espacial: indicador global
- Indicadores locais buscam refinar essa medida
  - Permitem a identificação de padrões de associação espacial significativos
  - São uma decomposição do índice global

## Moran Local

The Local Moran's  $I$  statistic of spatial association is given as:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{i,j} (x_j - \bar{X}) \quad (1)$$

where  $x_i$  is an attribute for feature  $i$ ,  $\bar{X}$  is the mean of the corresponding attribute,  $w_{i,j}$  is the spatial weight between feature  $i$  and  $j$ , and:

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n (x_j - \bar{X})^2}{n - 1} \quad (2)$$

- $I_i > 0 \rightarrow$  clusters de objetos de valores similares  
(vizinhos semelhantes)
- $I_i < 0 \rightarrow$  clusters de objetos de valores distintos  
(valores altos cercados de valores baixos, ou vice-versa)

# Scatterplot de Moran

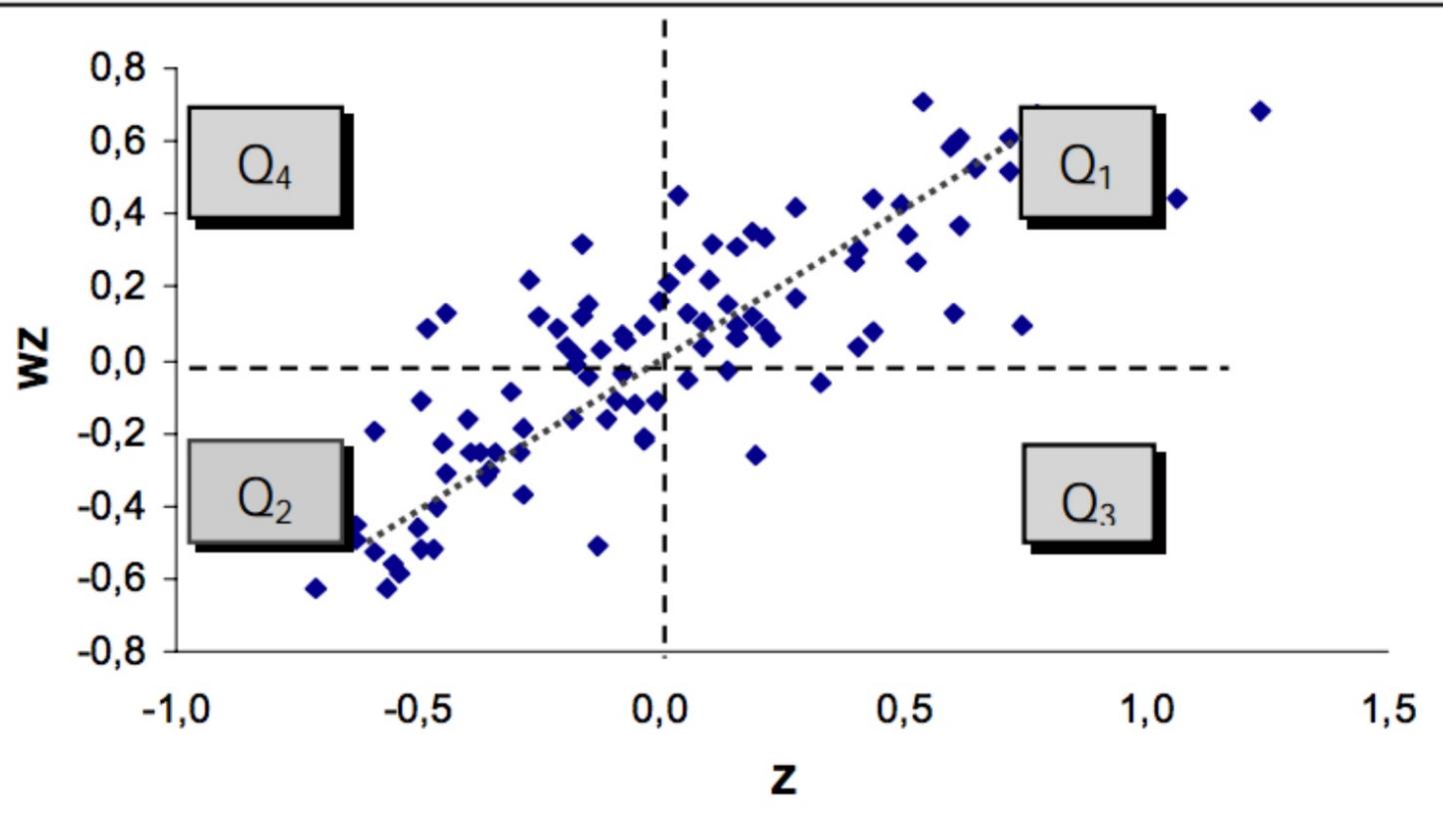


Figura 5-15 Diagrama de Espalhamento de Moran para o índice de exclusão/inclusão social de São Paulo, censo de 1991.

Plotar z-score versus wz  
(média dos z-scores dos vizinhos)

Q1 (High/High) e Q2  
(Low/Low) – associação  
espacial positiva,  
vizinhos semelhantes

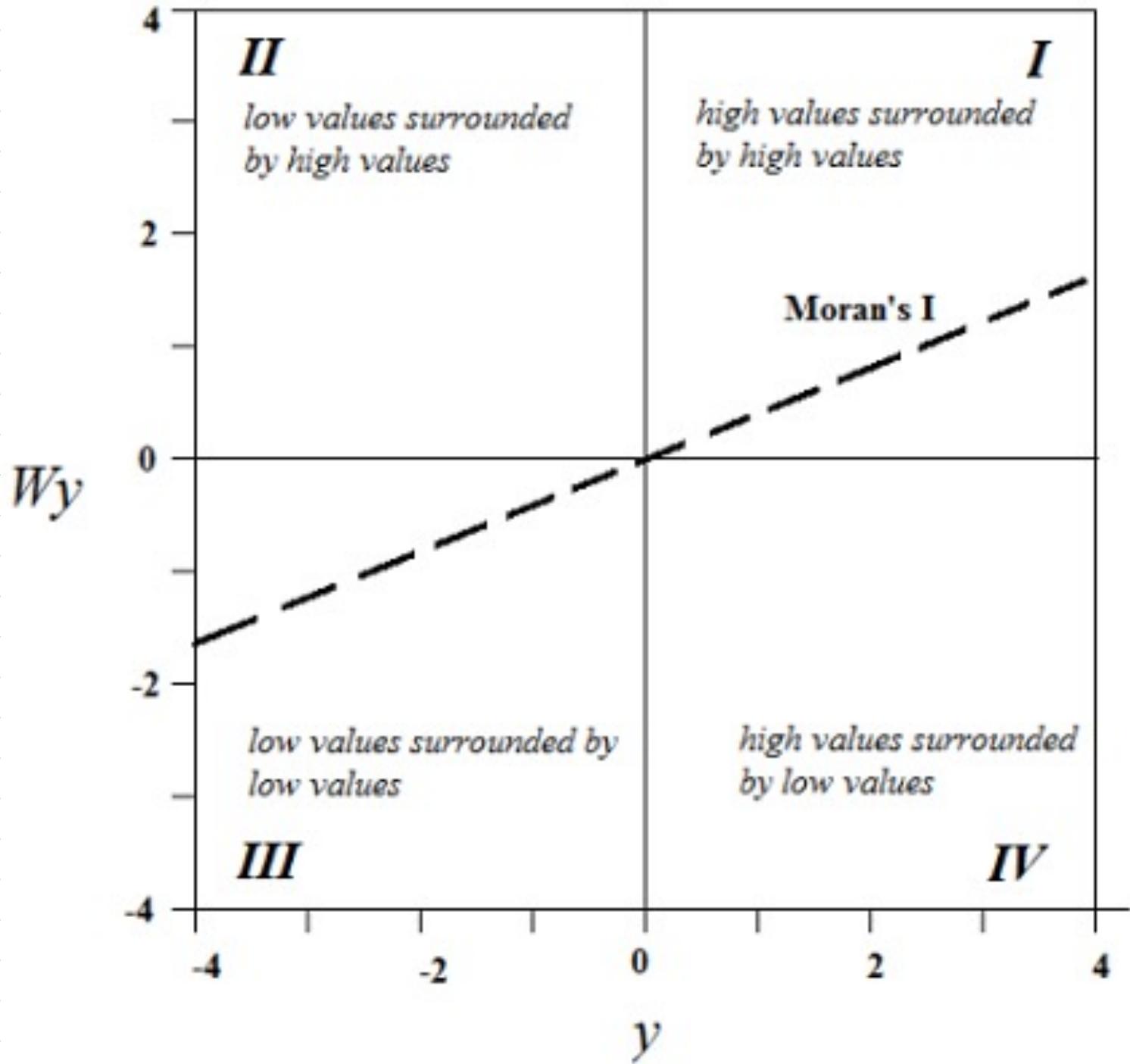
Q3 (High/Low) e Q4  
(Low/High) – associação  
espacial negativa,  
vizinhos possuem valores  
distintos

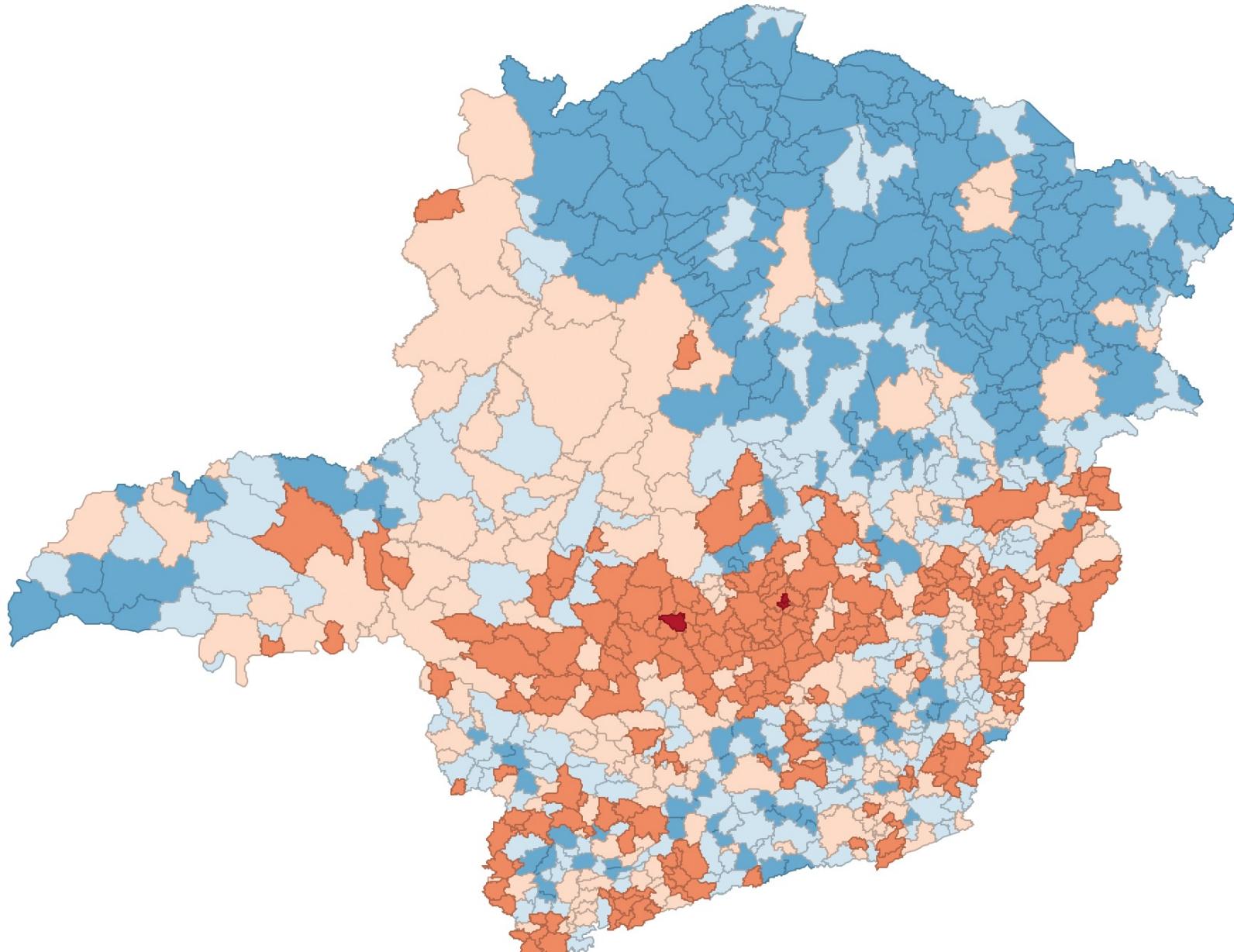
A inclinação da reta de  
tendência é exatamente  
o I de Moran

Valores padronizados  
em  $y$  e  $Wy$ : observação  
menos a média

$y$ : cada observação

$Wy$ : média ponderada  
das observações nos  
vizinhos; pesos  
definidos segundo a  
matriz de vizinhança





Hinge=3.0: perc_votos	
Lower outlier (0)	[-inf : -0.081]
< 25% (213)	[-0.081 : 0.030]
25% - 50% (213)	[0.030 : 0.045]
50% - 75% (214)	[0.045 : 0.067]
> 75% (210)	[0.067 : 0.178]
Upper outlier (3)	[0.178 : inf]

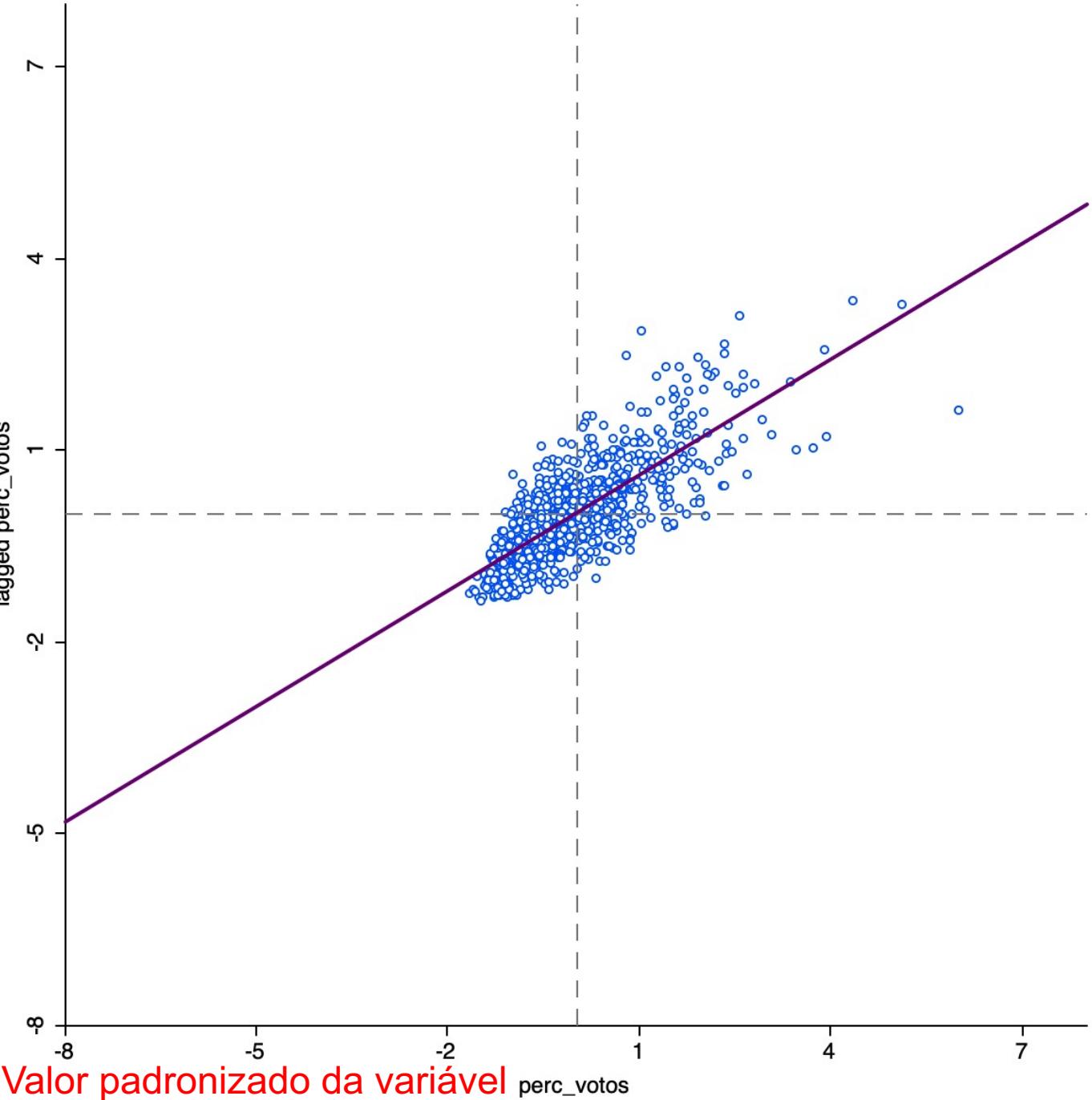
Hinges  $\approx$  linhas do boxplot  
Neste caso, considerando  
outliers os dados que  
estiverem a mais de  $3 * \text{IQR}$

Lower:  $Q1 - 3 * \text{IQR}$   
Higher:  $Q3 + 3 * \text{IQR}$

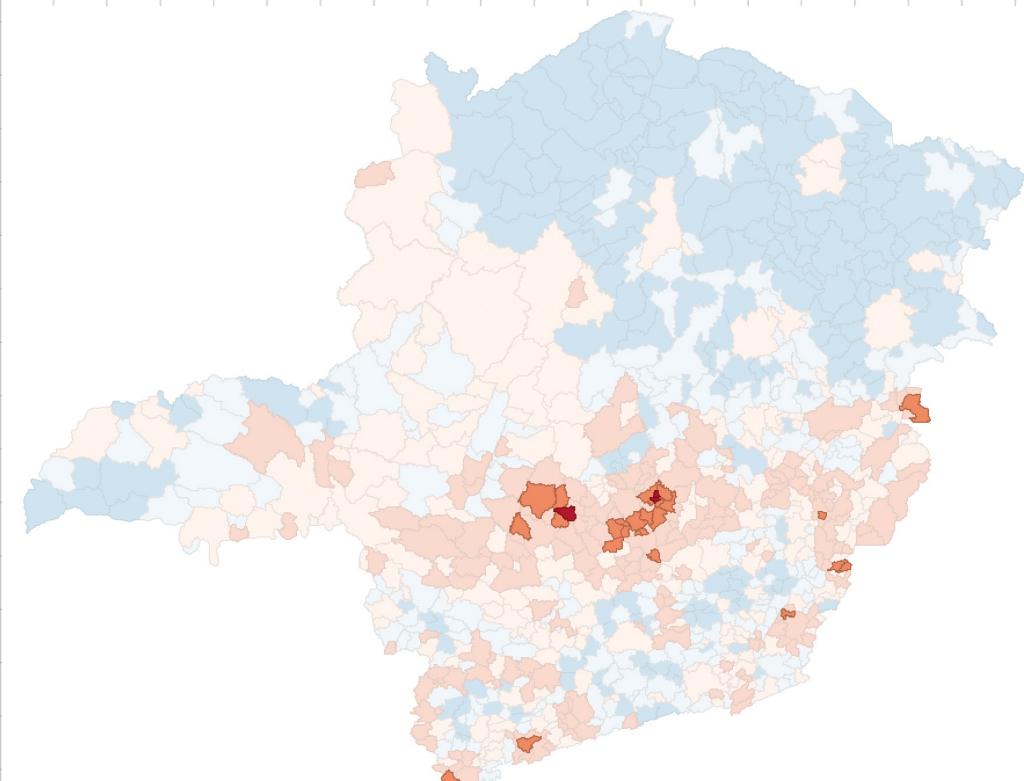
# Scatterplot de Moran

2222

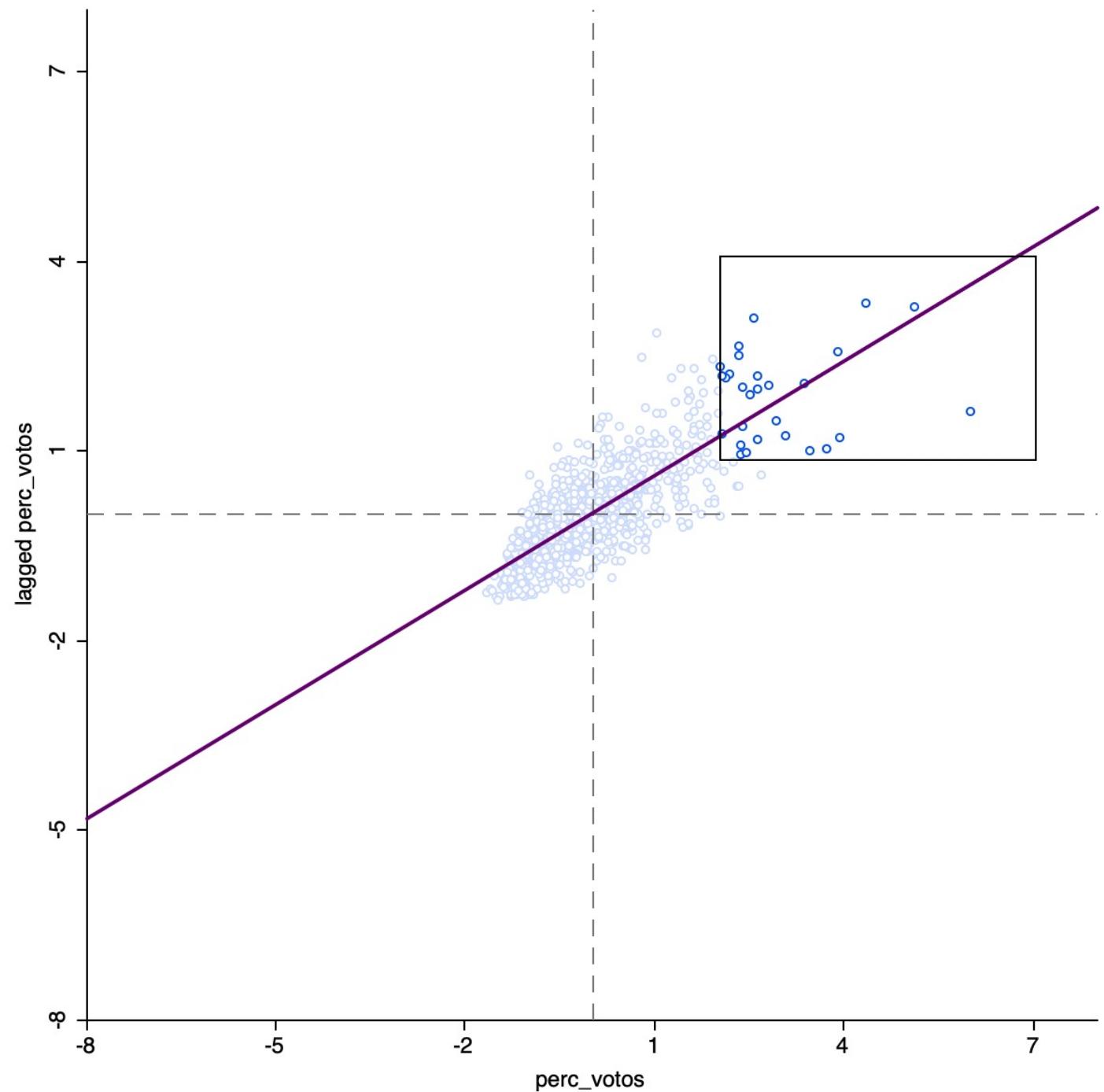
Média ponderada padronizada dos vizinhos  
(matriz de vizinhança)



Moran's I: 0.604

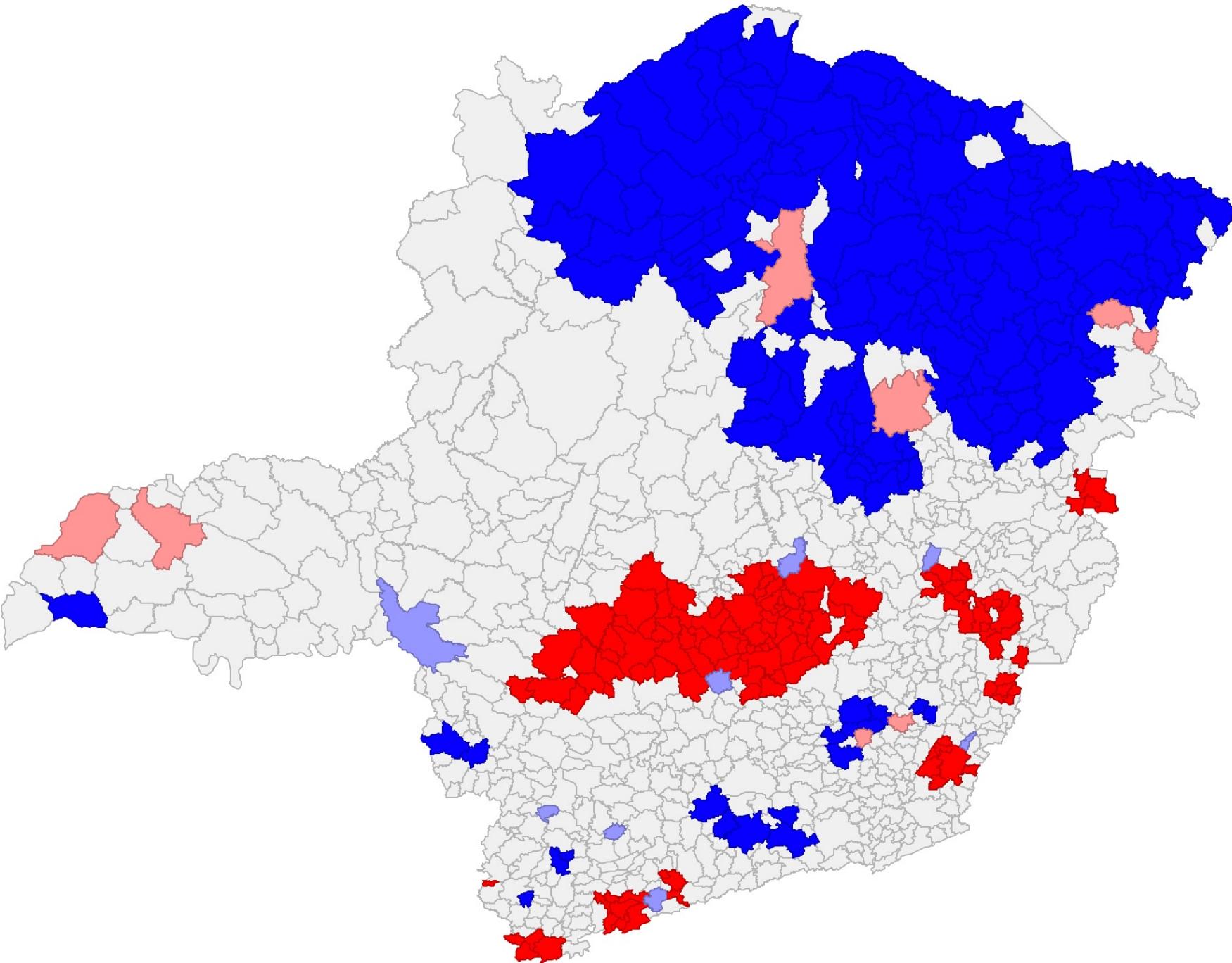


2222

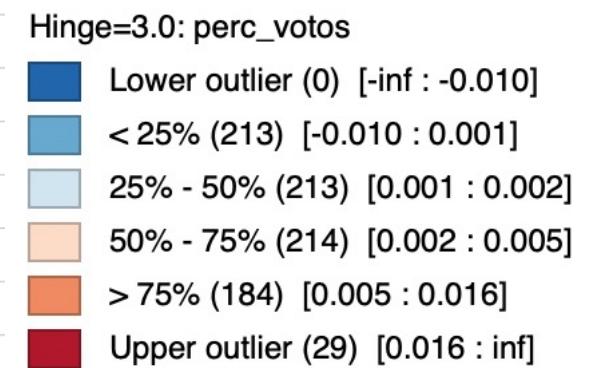
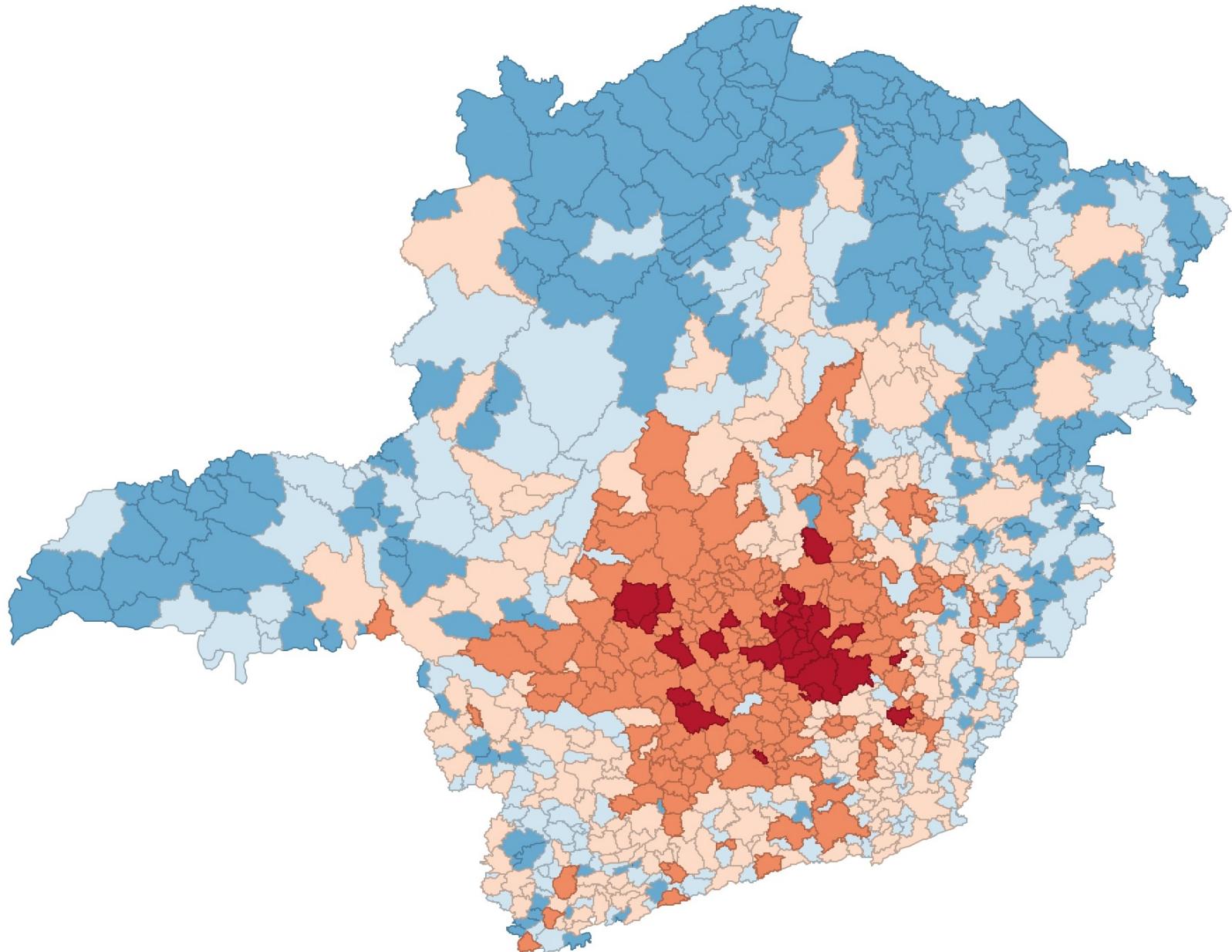


LISA Cluster Map: geoda

- Not Significant (557)
- High-High (118)
- Low-Low (162)
- Low-High (8)
- High-Low (8)

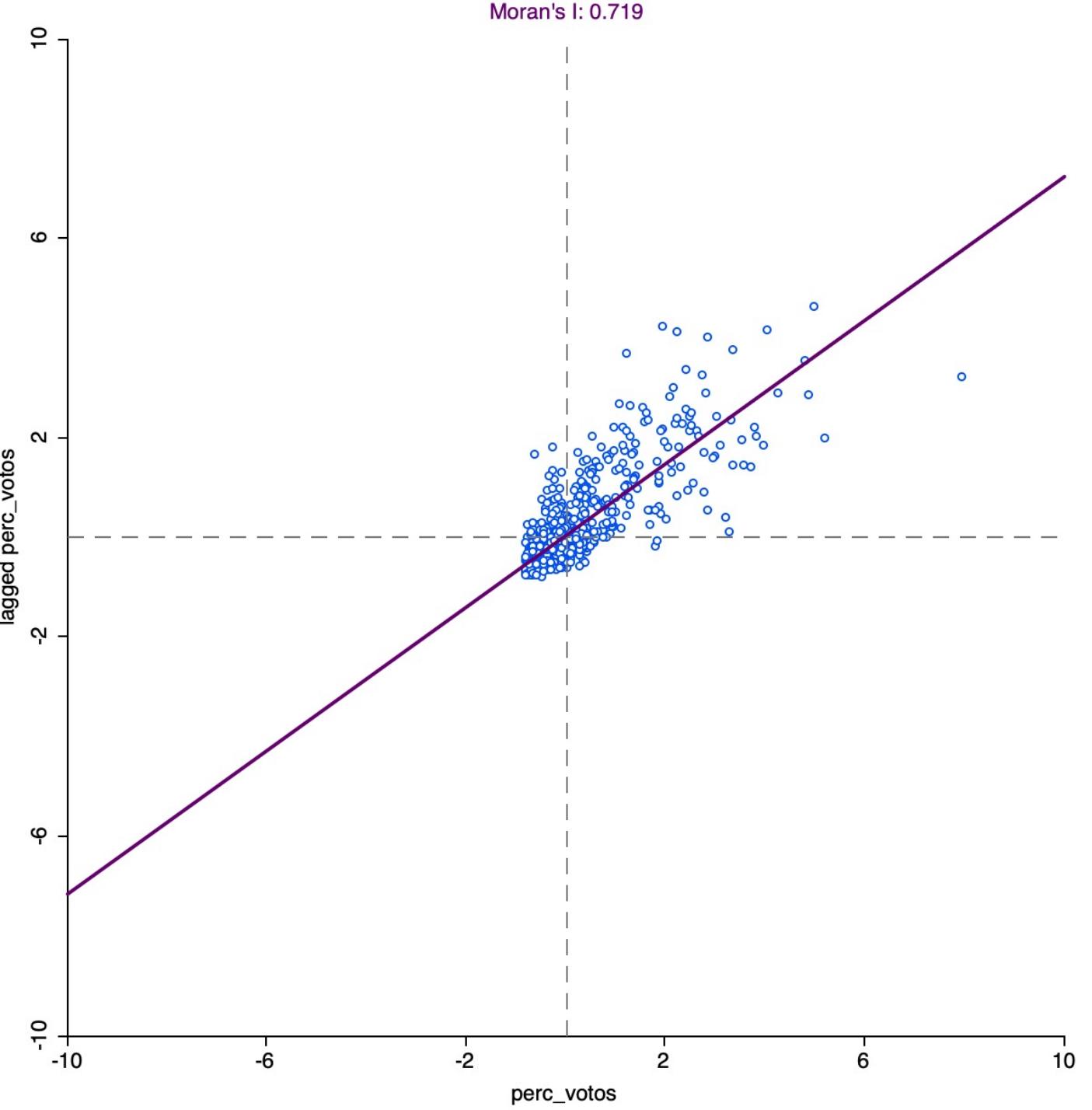


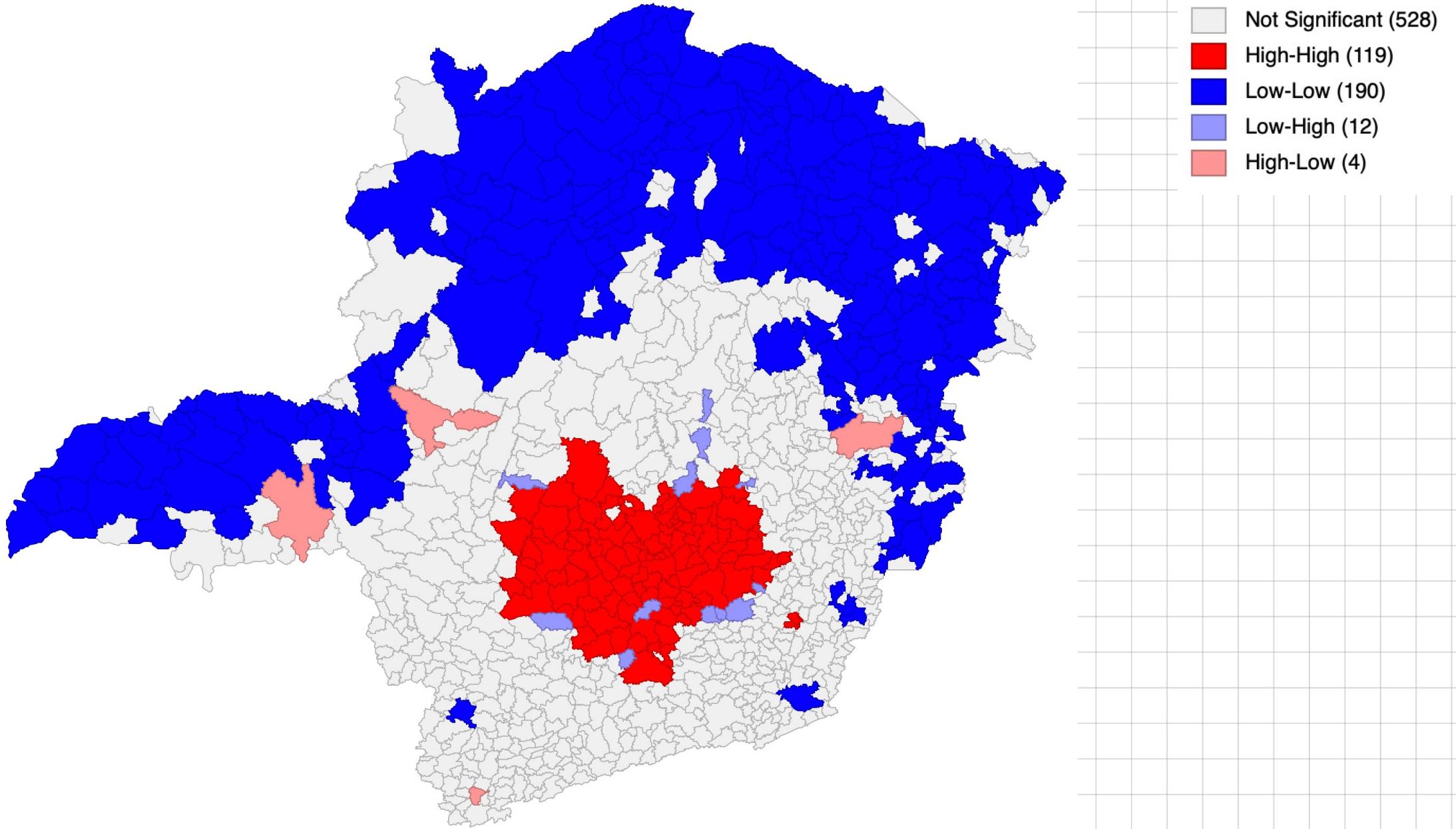
2222

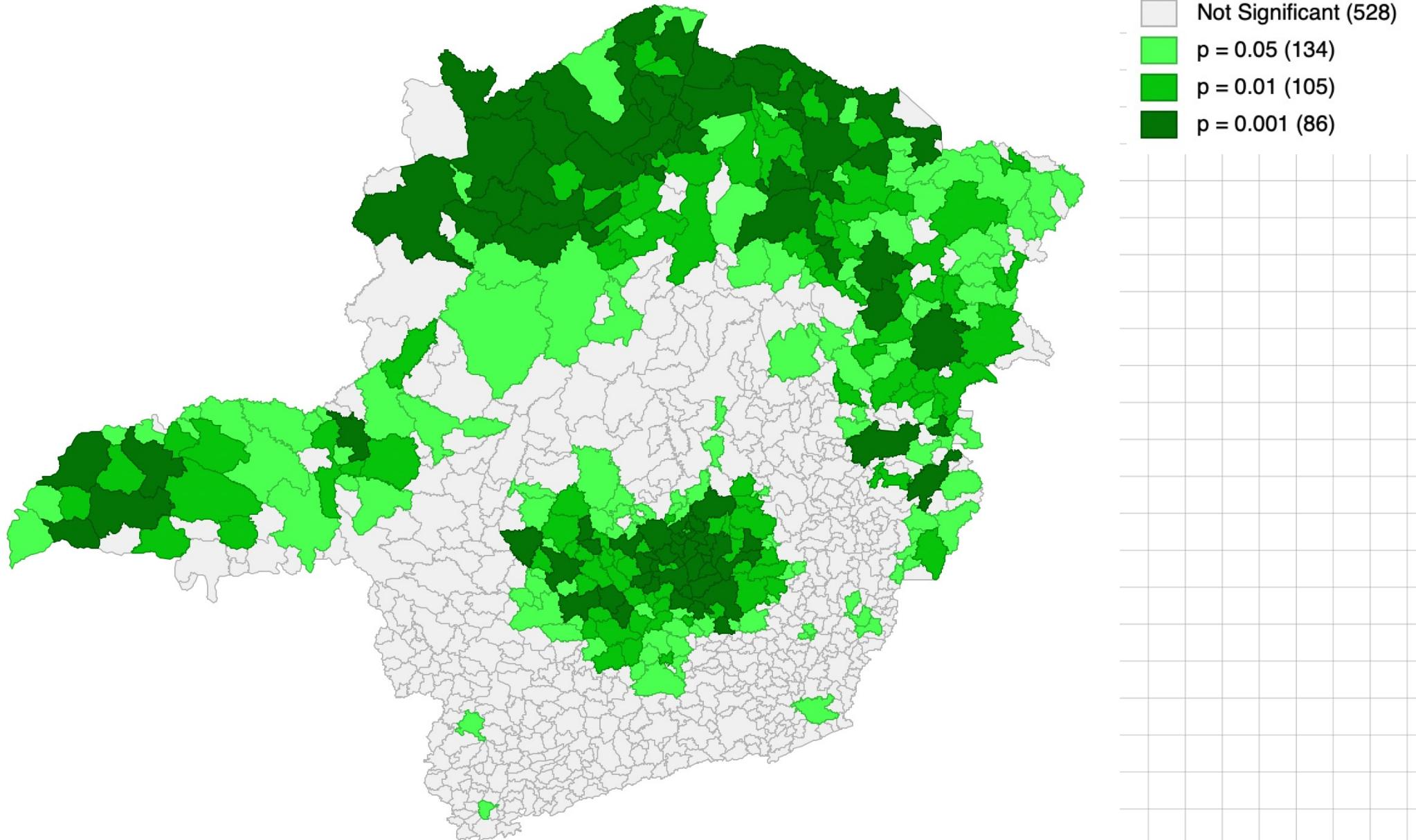


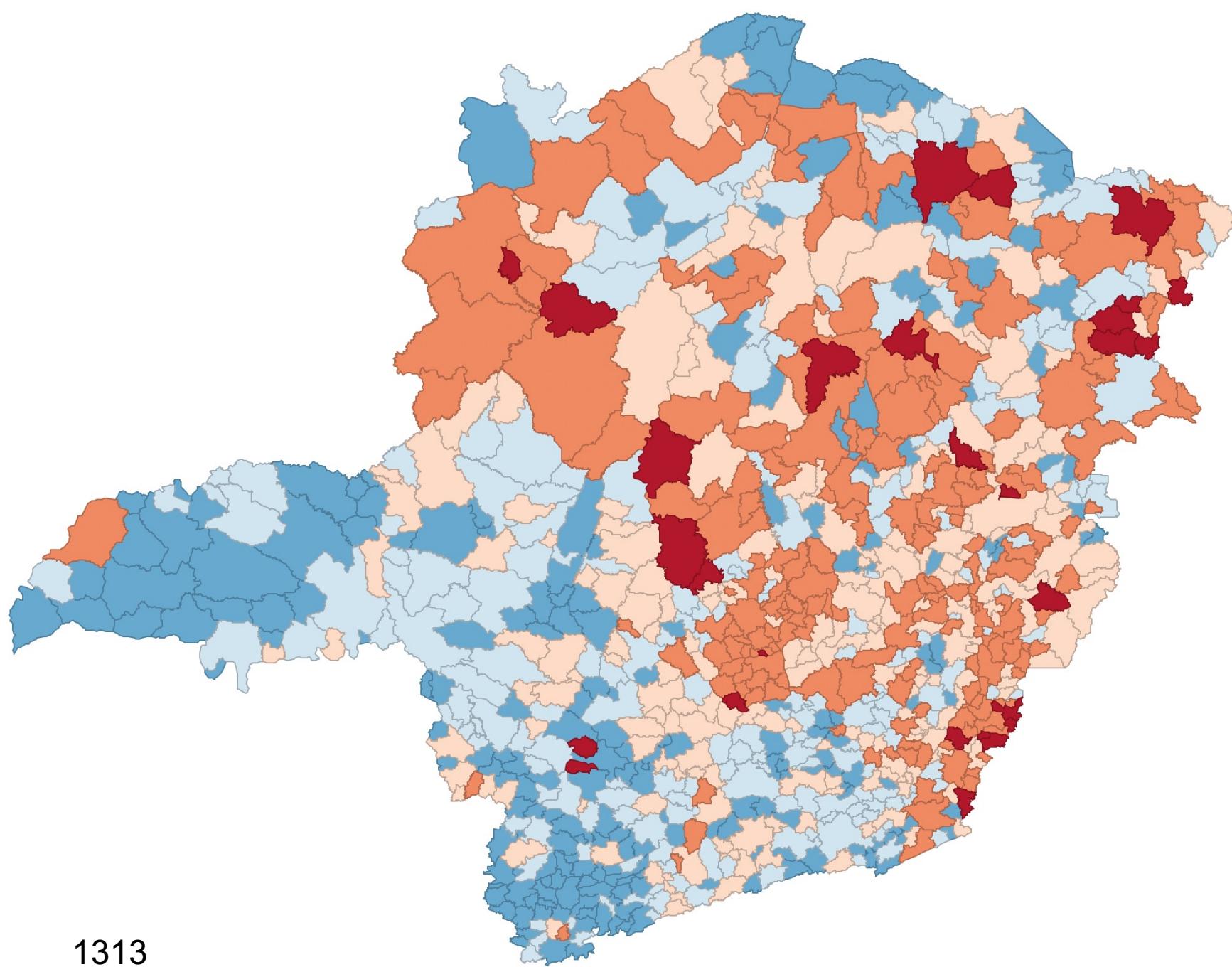
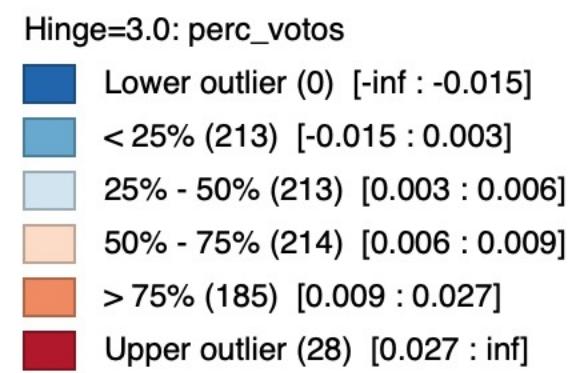
Dep. 1212

1212



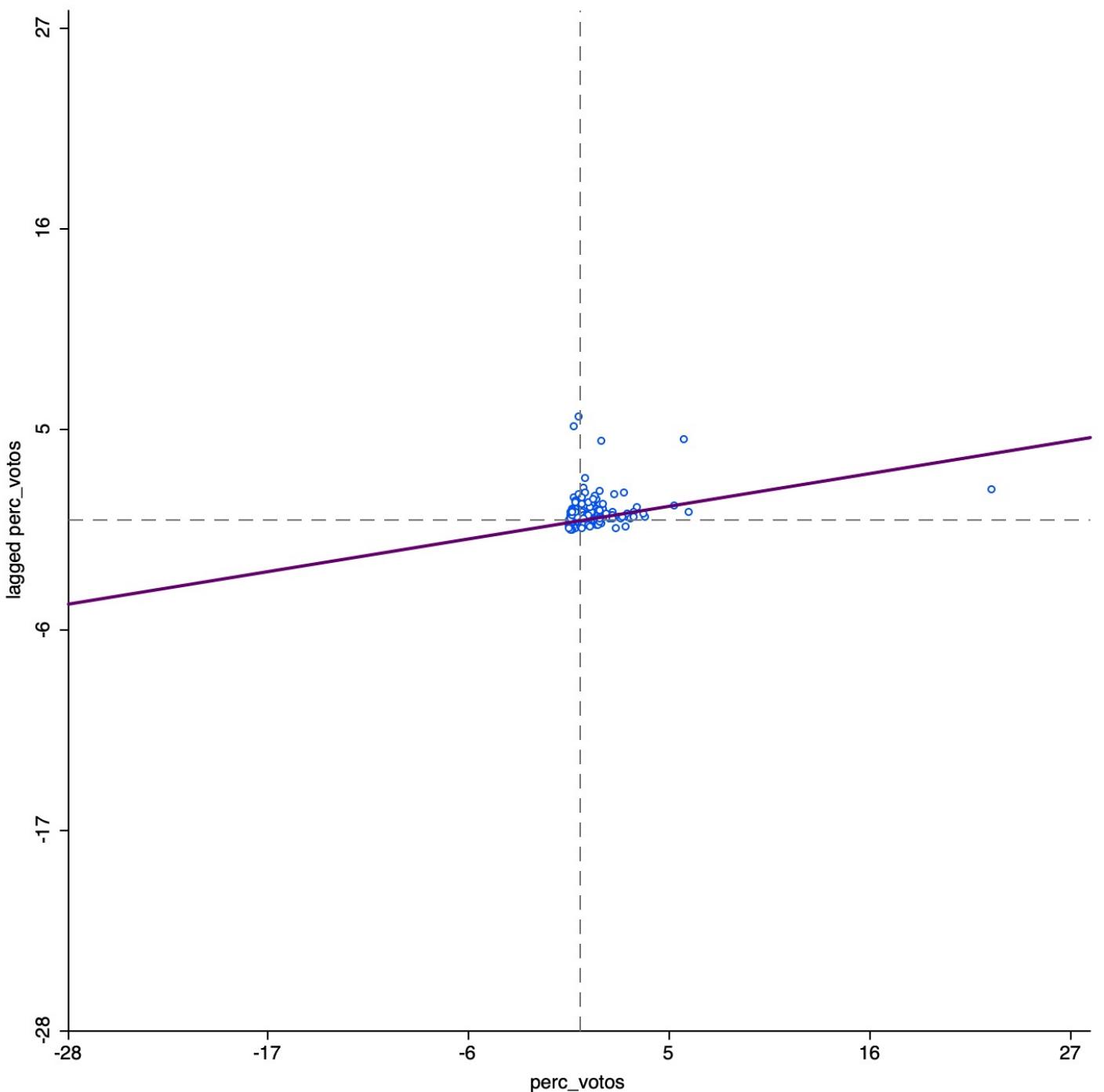


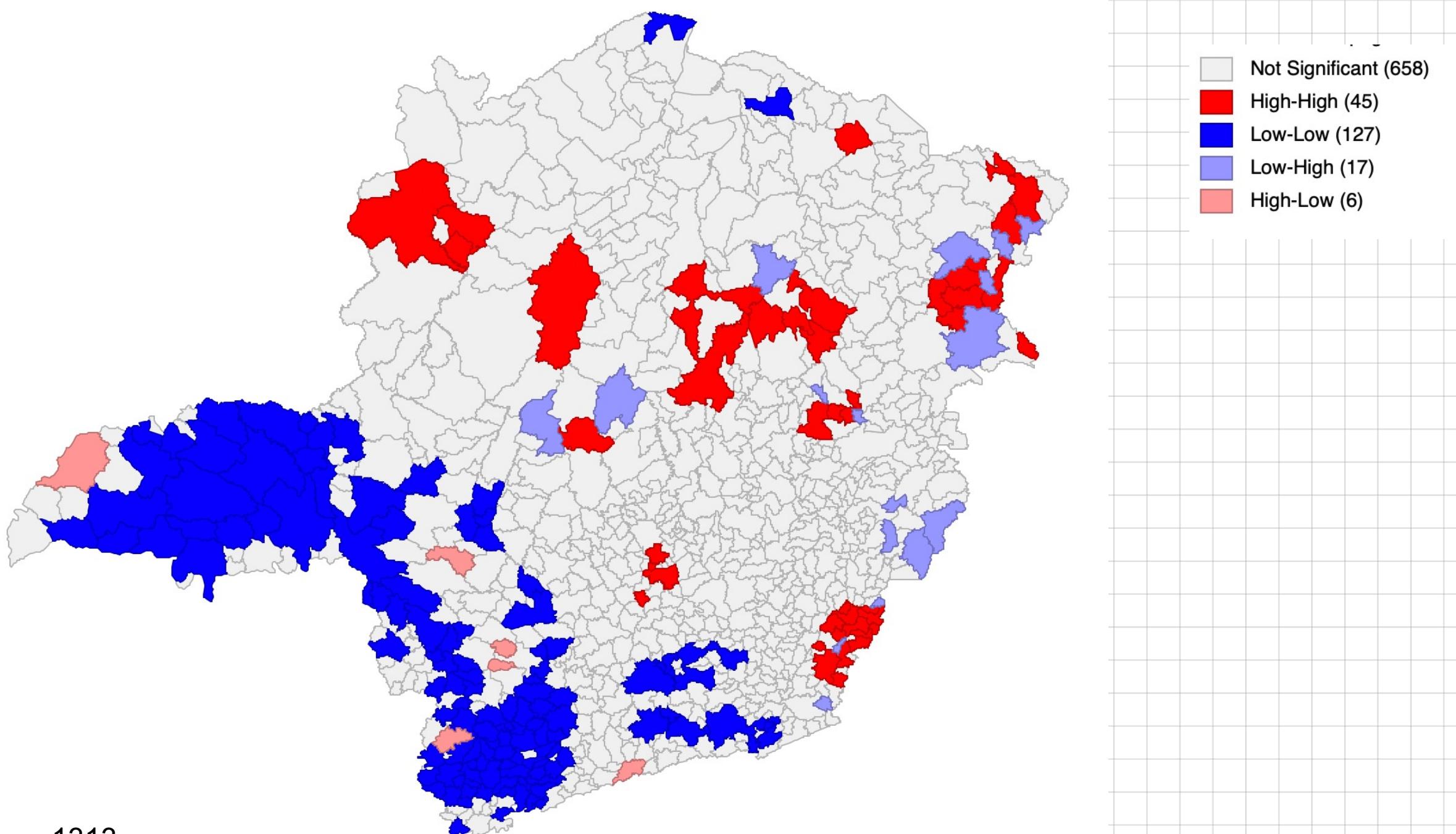




1313

Moran's I: 0.163





# Análise de padrões de pontos

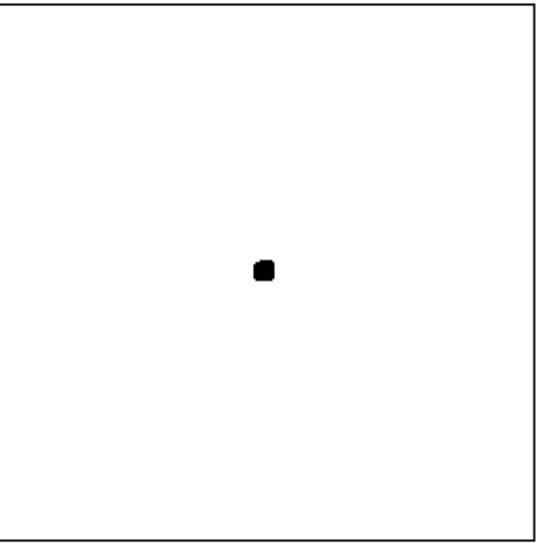
- Dados pontuais podem corresponder a
  - amostras de algum processo/fenômeno (OMT-G: amostras)
  - objetos distintos (OMT-G: pontos)
- No caso de amostras, o interesse está em estimar o valor de alguma variável em pontos onde não há amostra
  - Interpolação
  - Transformação para outras representações de geocampo
- No caso de objetos distintos, o interesse está em avaliar os padrões espaciais de distribuição dos pontos
  - Concentrações, densidade, clustering
  - Hot spot analysis

# Análise de padrões de pontos

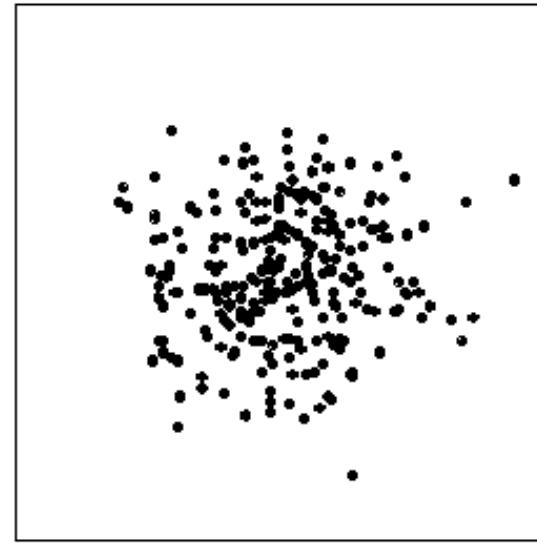
- Análise exploratória visual
  - Mapa de pontos
  - Mapa de calor / densidade / kernel
    - Cuidado: parâmetros!
- Efeitos de primeira ordem
  - Variação de intensidade em unidade de área
  - Densidade baseada na localização absoluta
- Efeitos de segunda ordem
  - Interação baseada na distância entre objetos
  - Posicionamento relativo

**256 pontos**

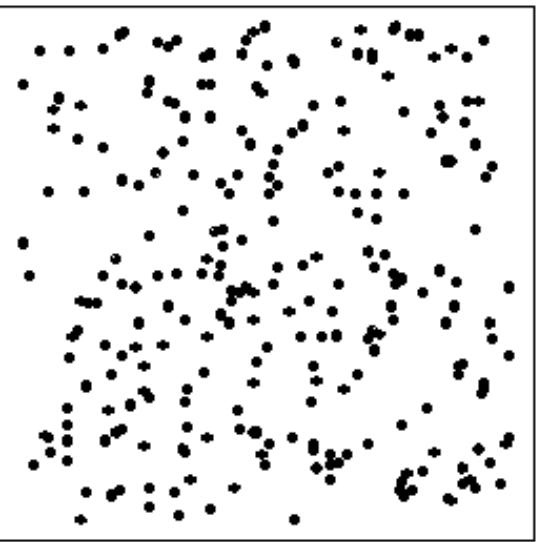
**clustered**



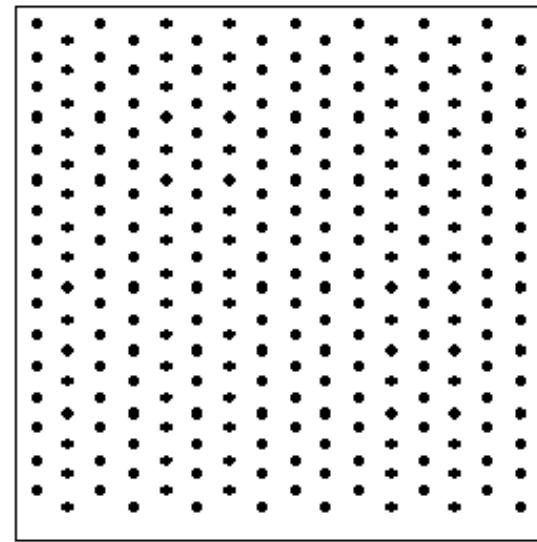
**normal**



**random**



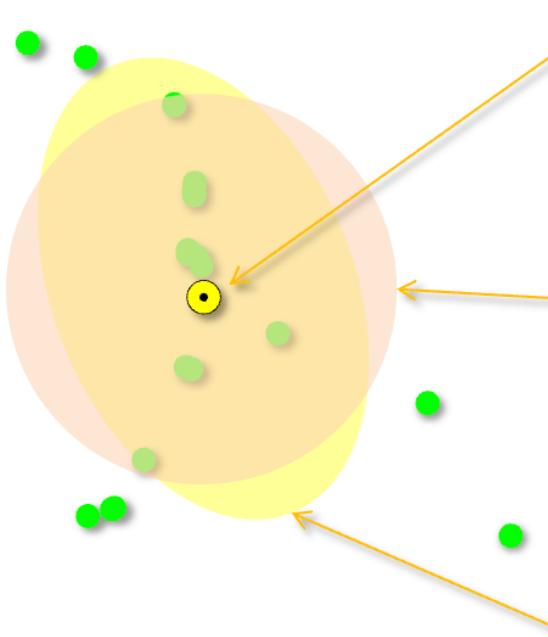
**regular**



# Análise de padrões de pontos

- Centrografia
  - Geração de estatísticas summarizadoras como o centro médio, distância padrão e elipse desviacional padrão

# Centrografía



**Mean center**  
computed average X  
and Y coordinate  
values.

$$\bar{s} = \left( \frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right)$$

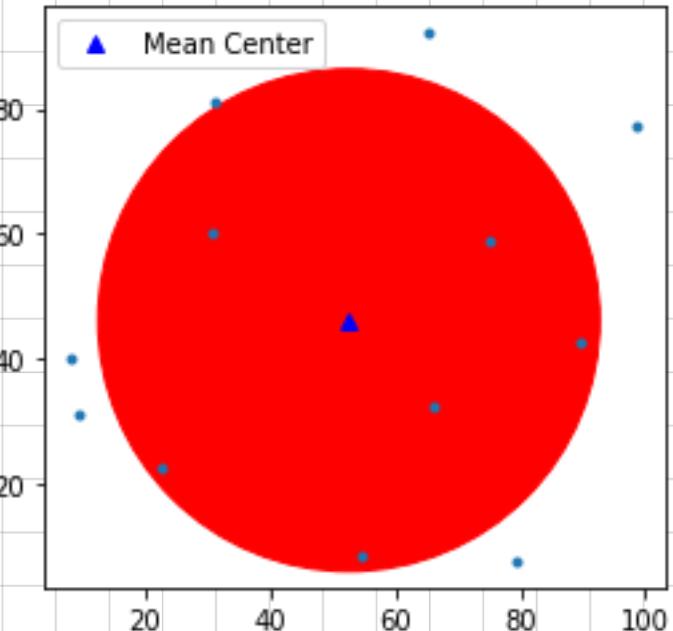
**Standard distance**  
measure of the variance  
between the average  
distance of the features  
to the mean center.

$$d = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu_x)^2 + (y_i - \mu_y)^2}{n}}$$

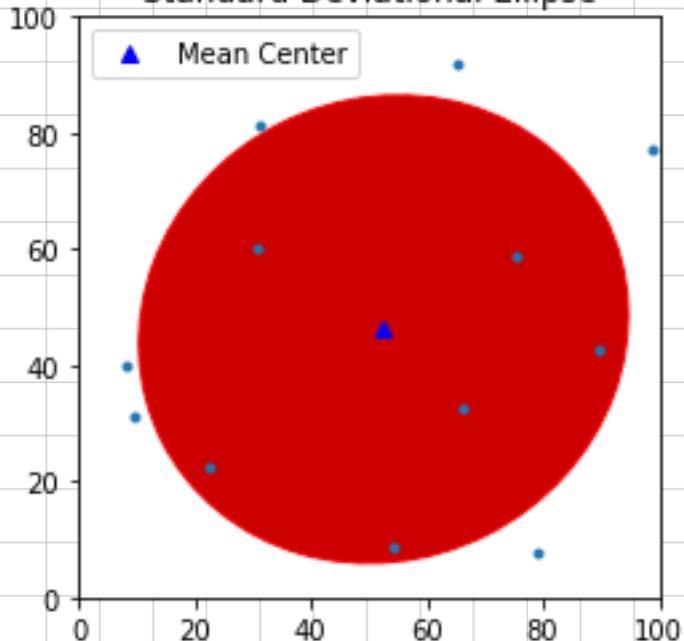
**Standard deviational  
ellipse**  
separate standard  
distances for each  
axis.

$$\left. \begin{array}{l} d_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu_x)^2}{n}} \\ d_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \mu_y)^2}{n}} \end{array} \right\}$$

Standard Distance Circle

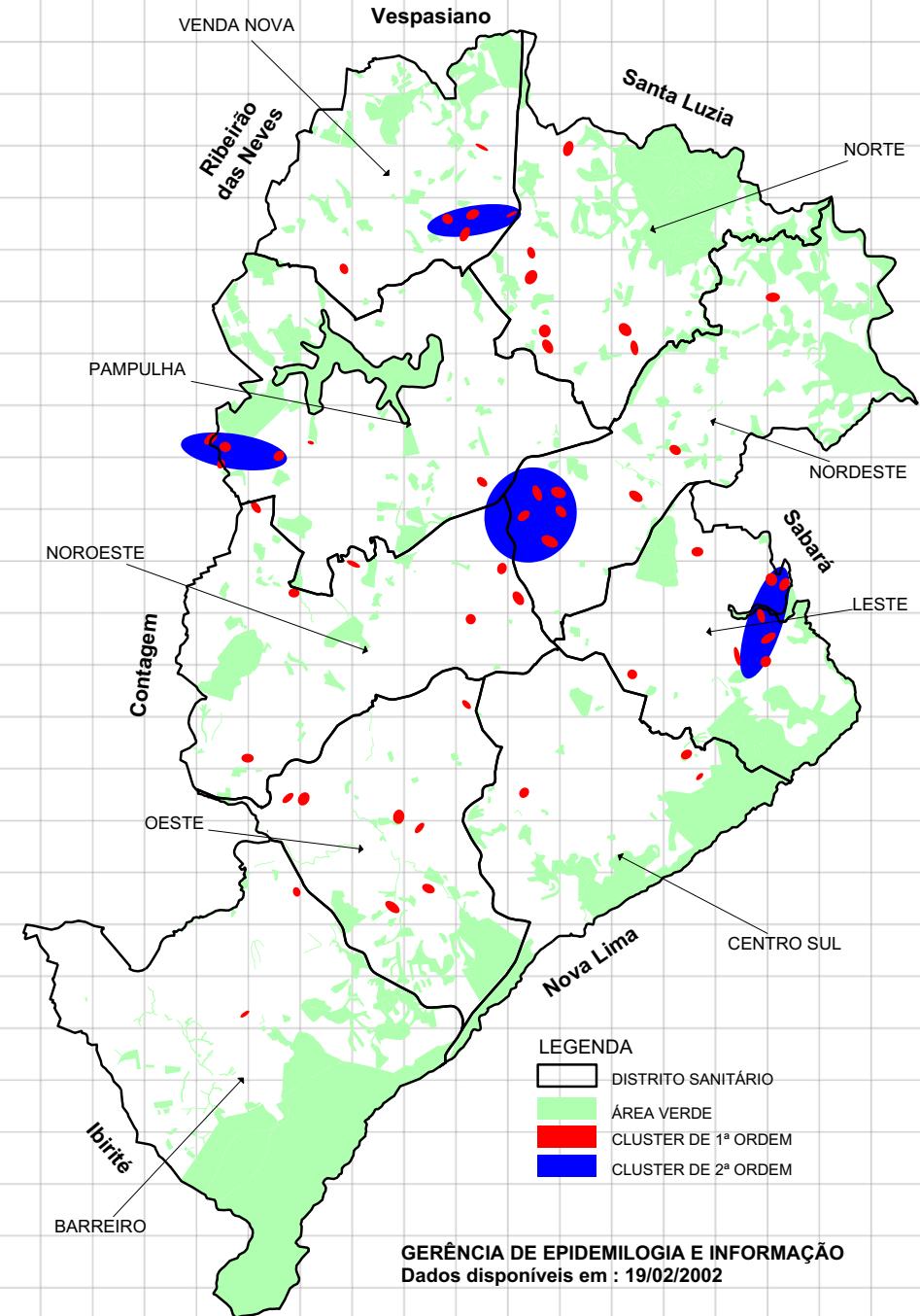


Standard Deviational Ellipse



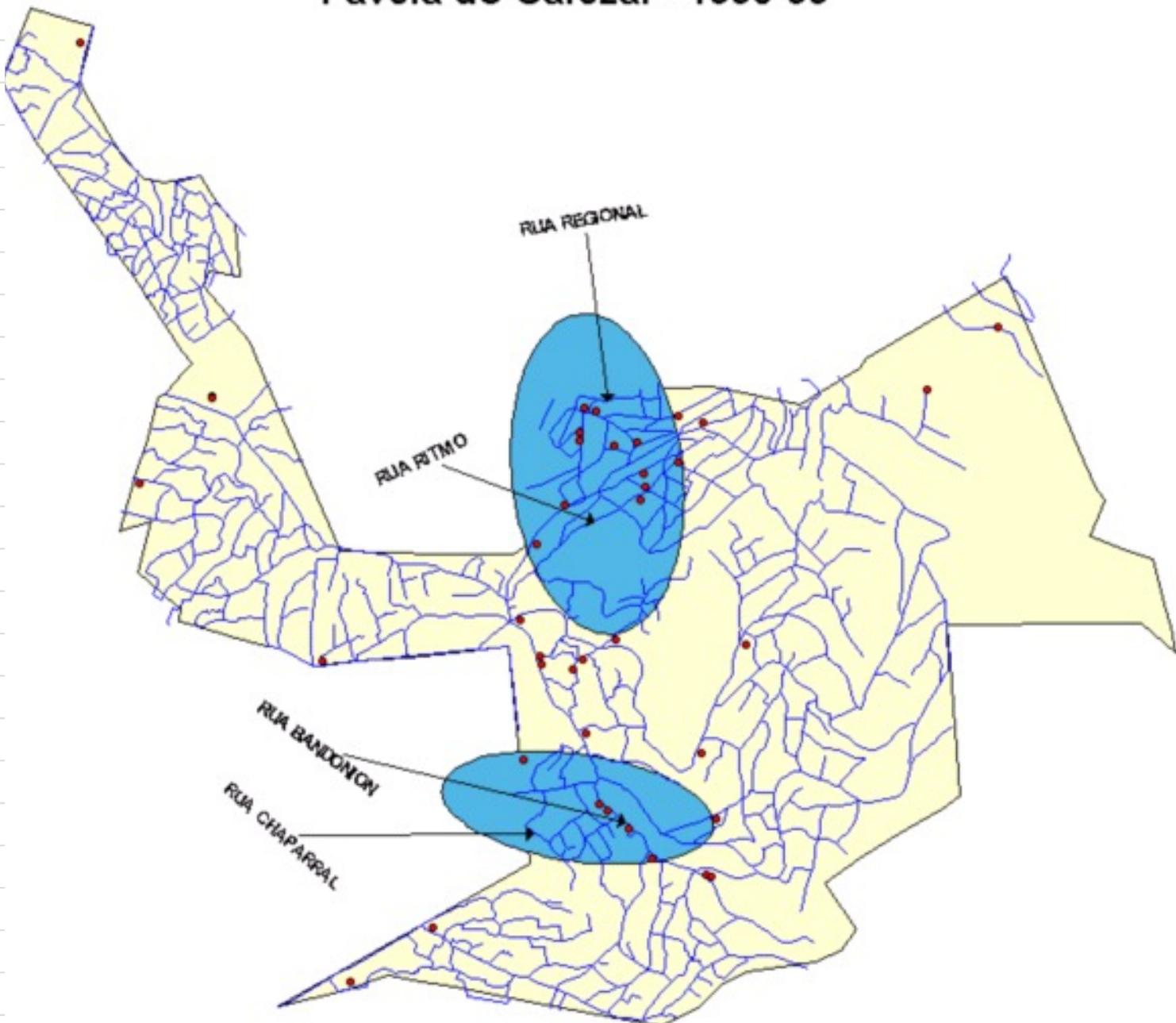
# Centrografia

- Casos de dengue
- Clusters de 1<sup>a</sup> e 2<sup>a</sup> ordens



# Centrografia

Conglomerados de maior risco de Homicídios  
Favela do Cafezal - 1995-99

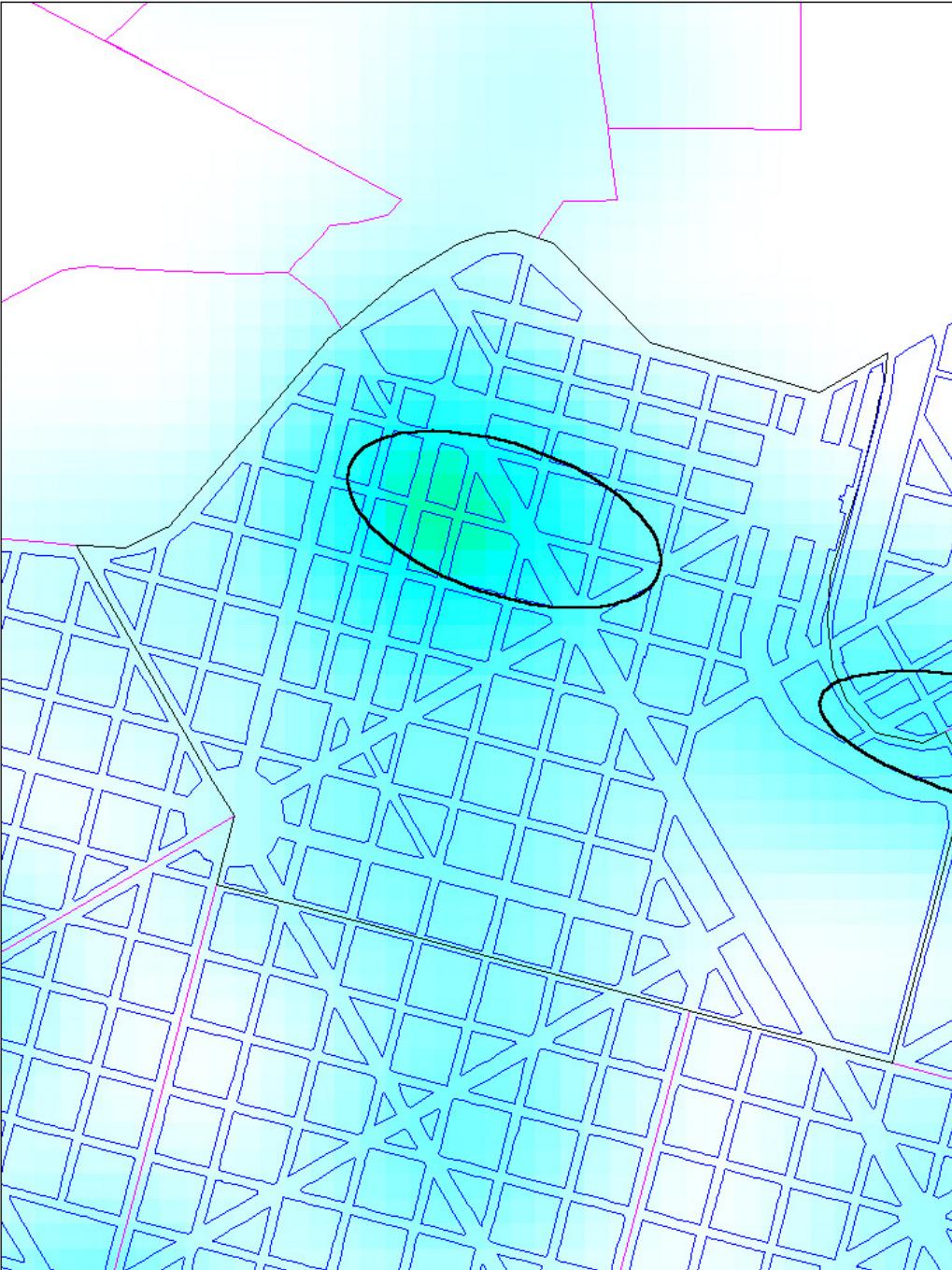


7h-8h

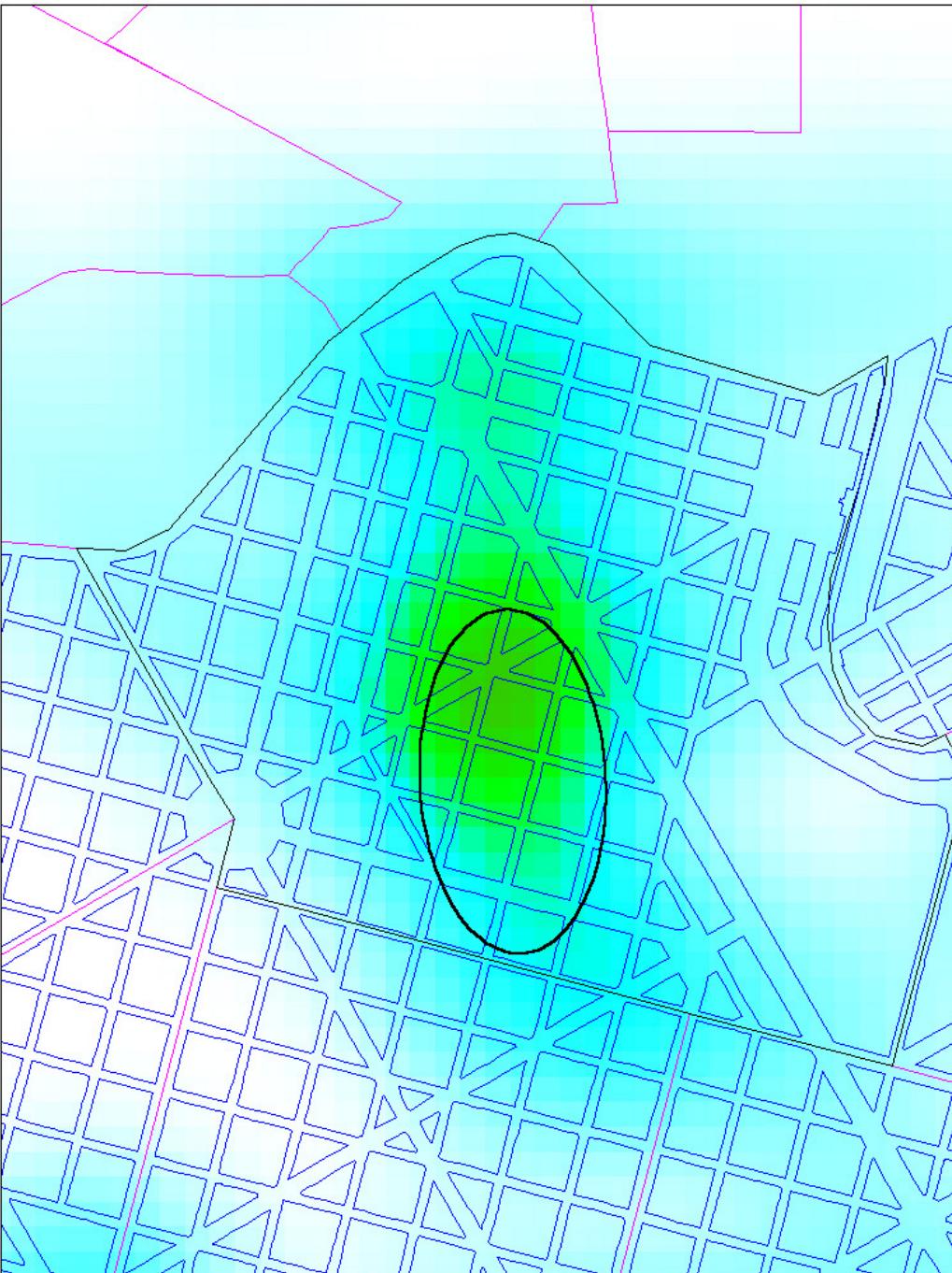


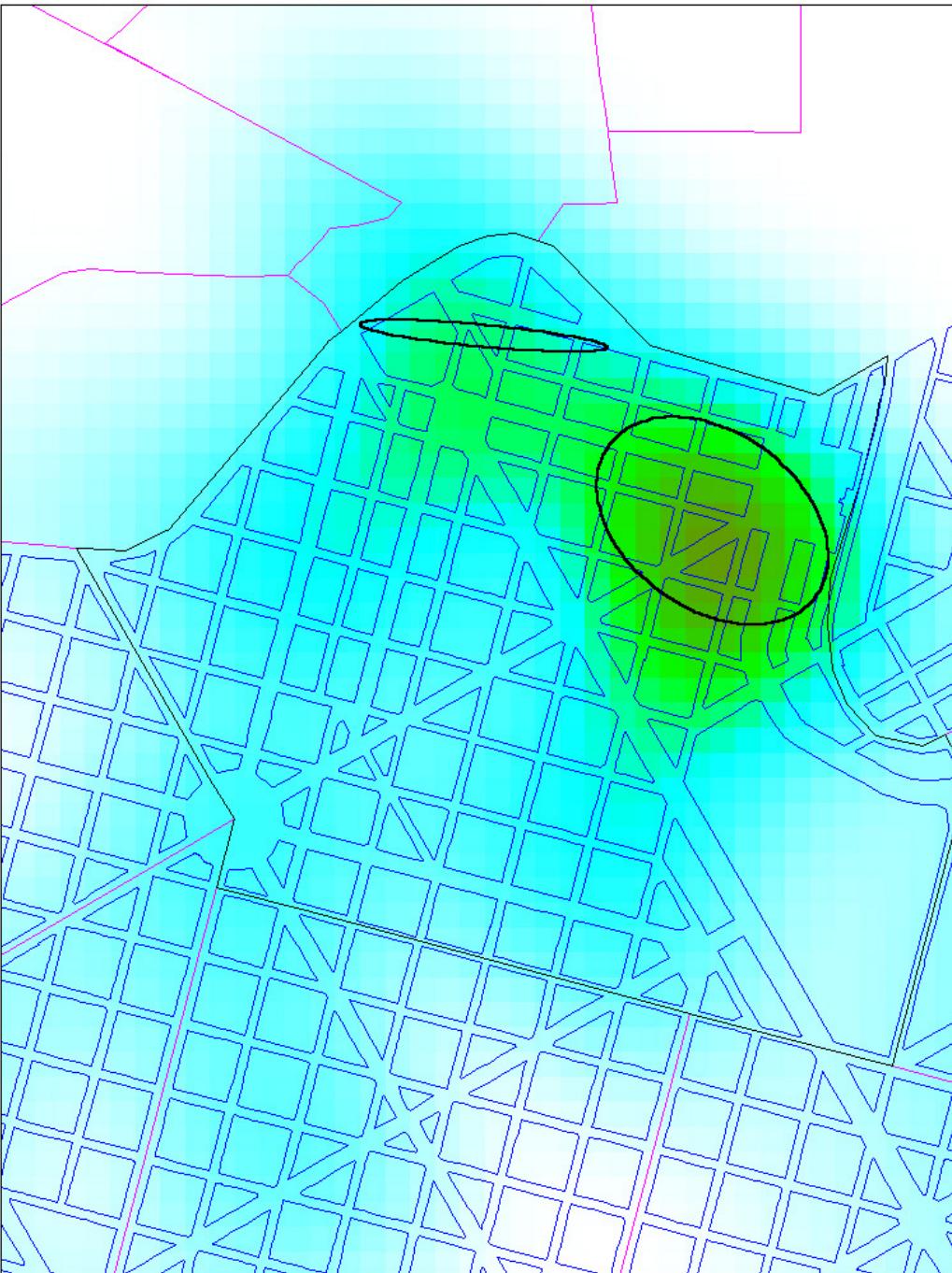


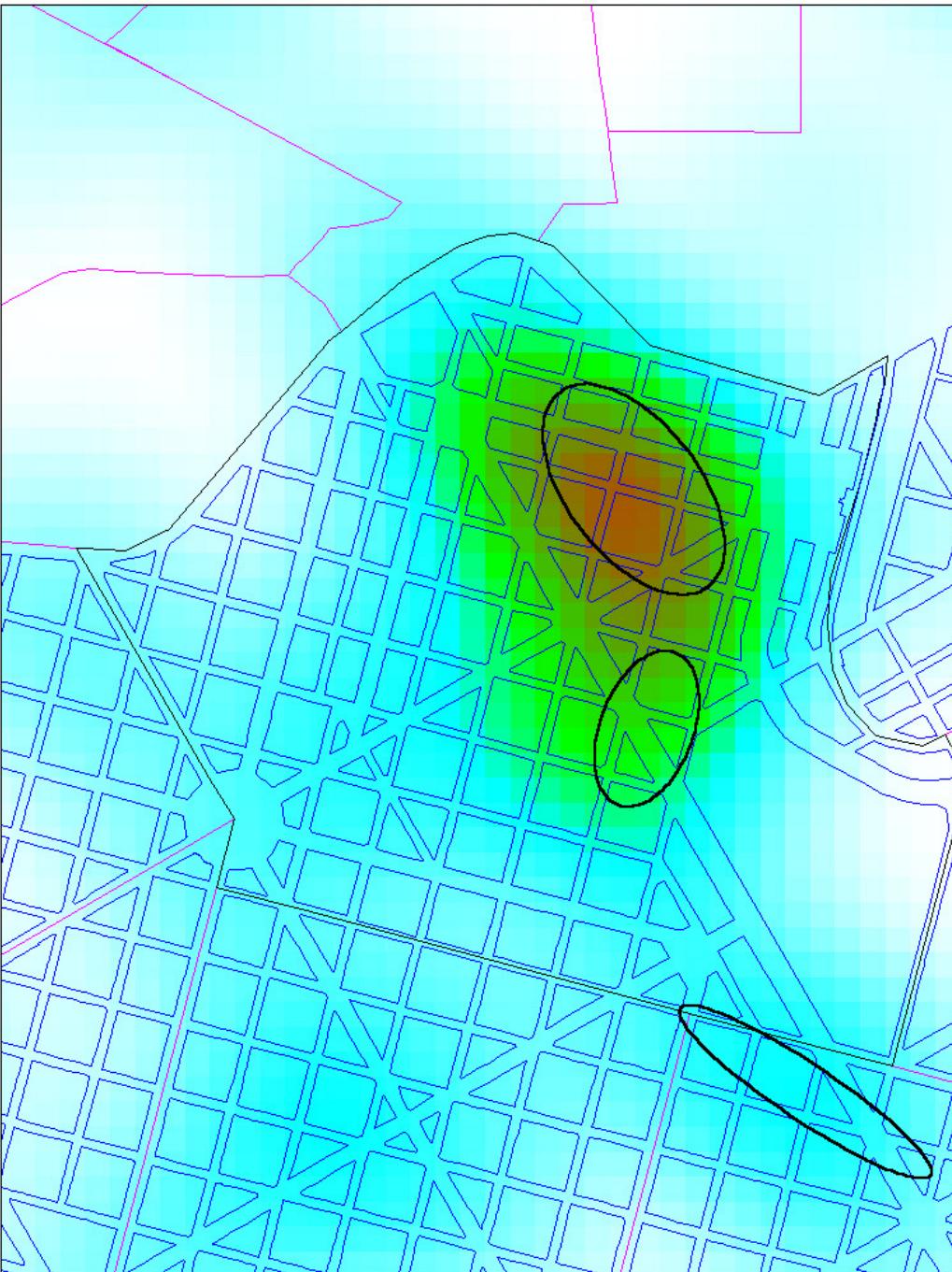


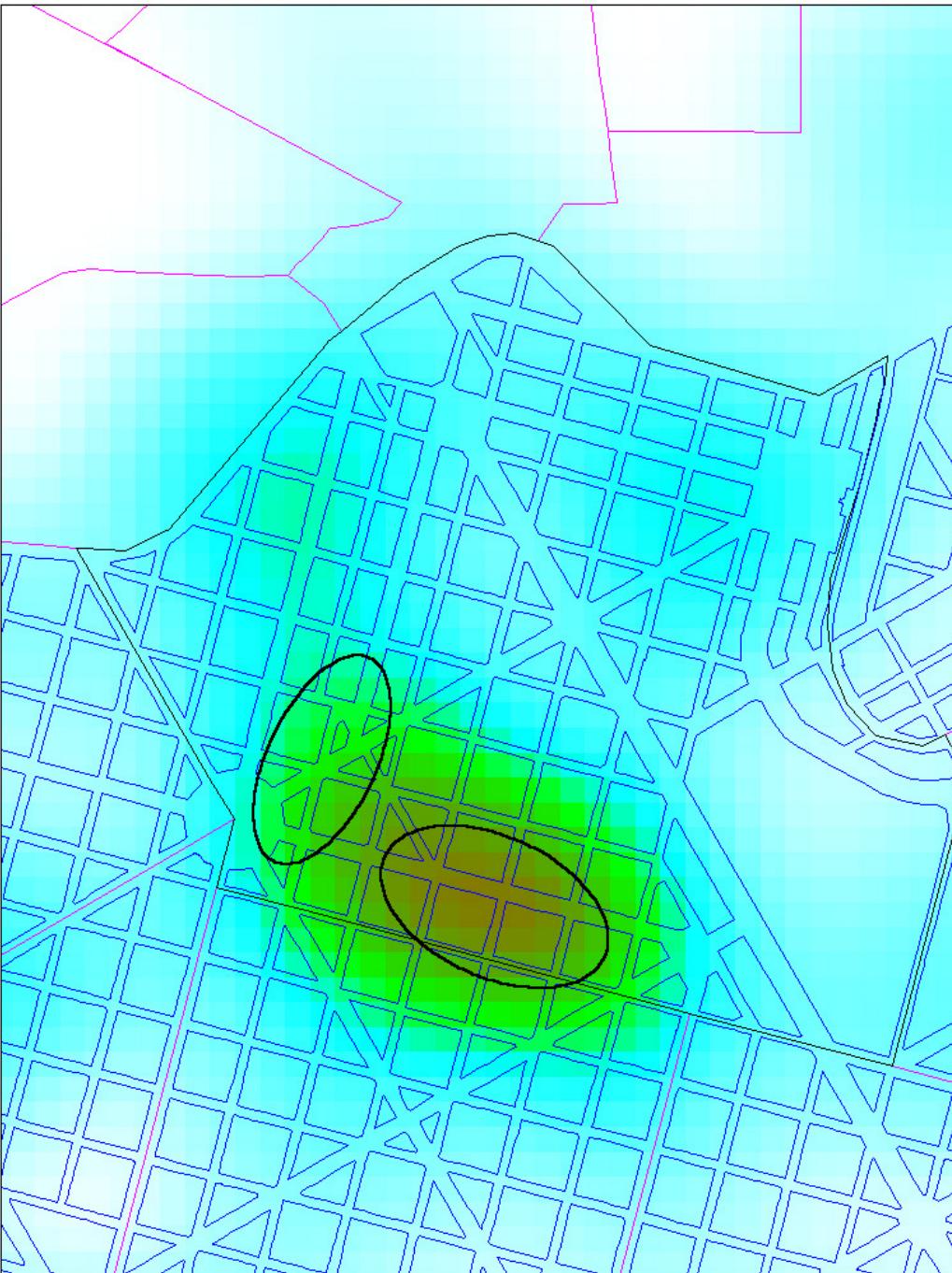


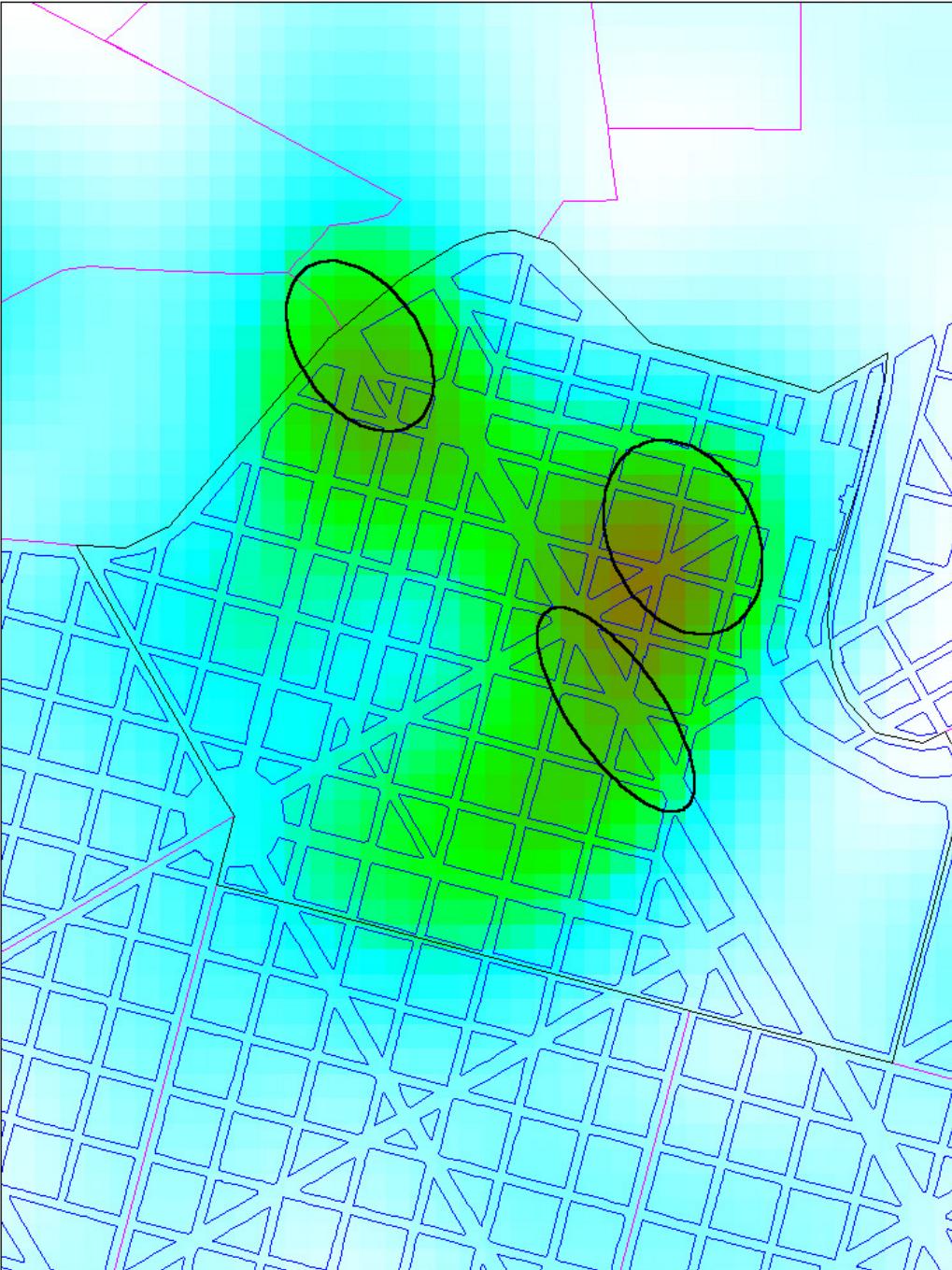


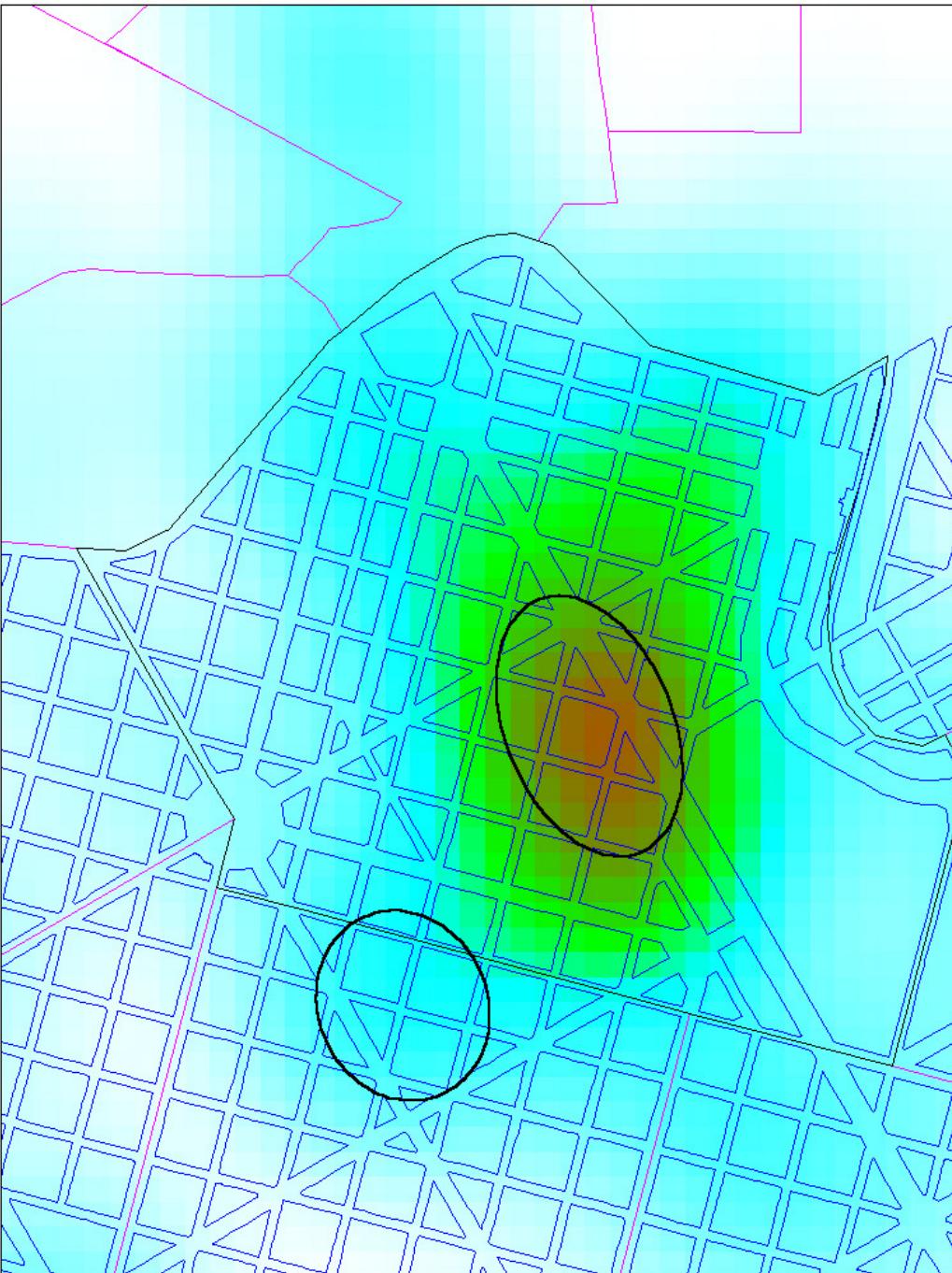




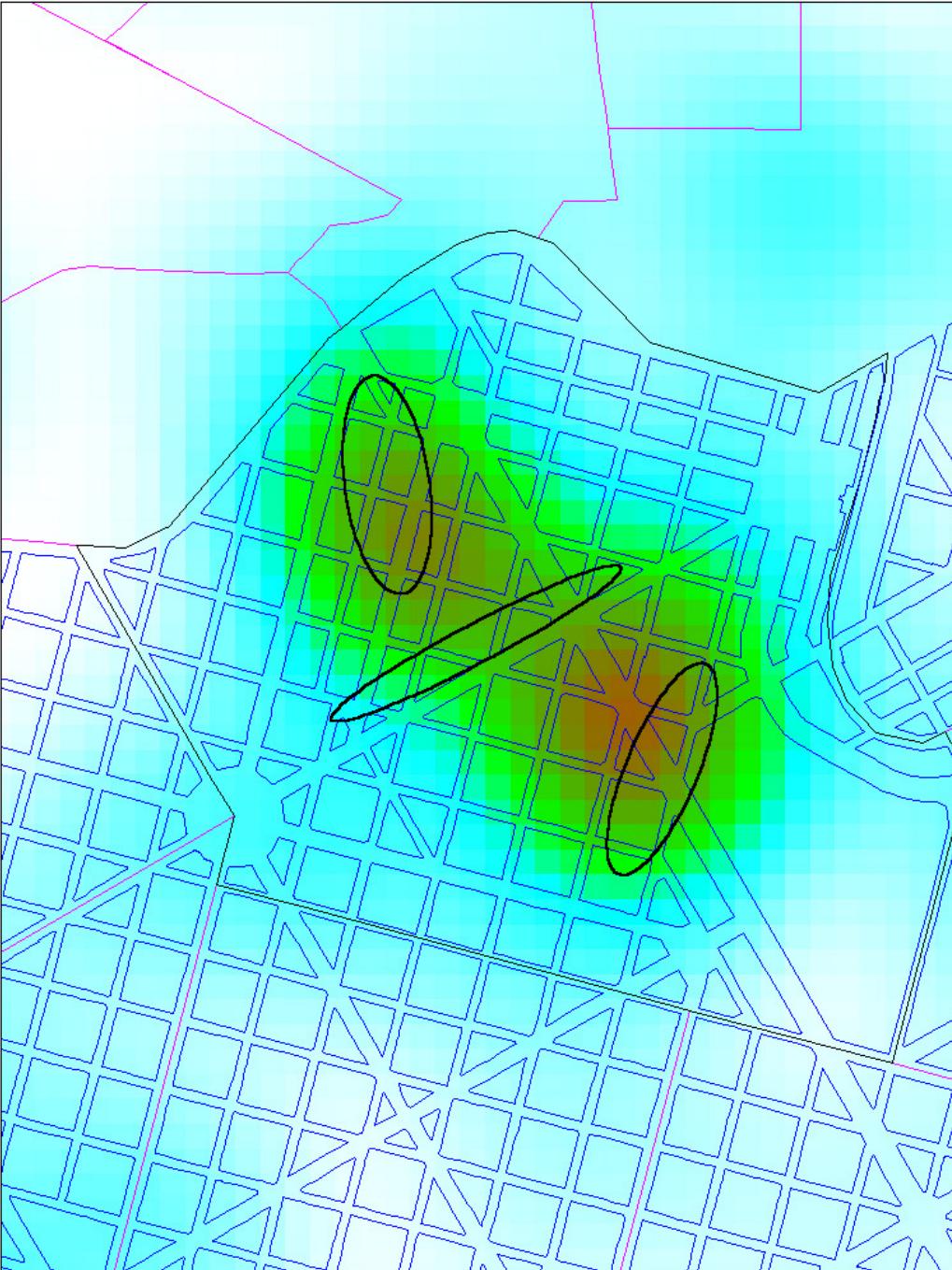


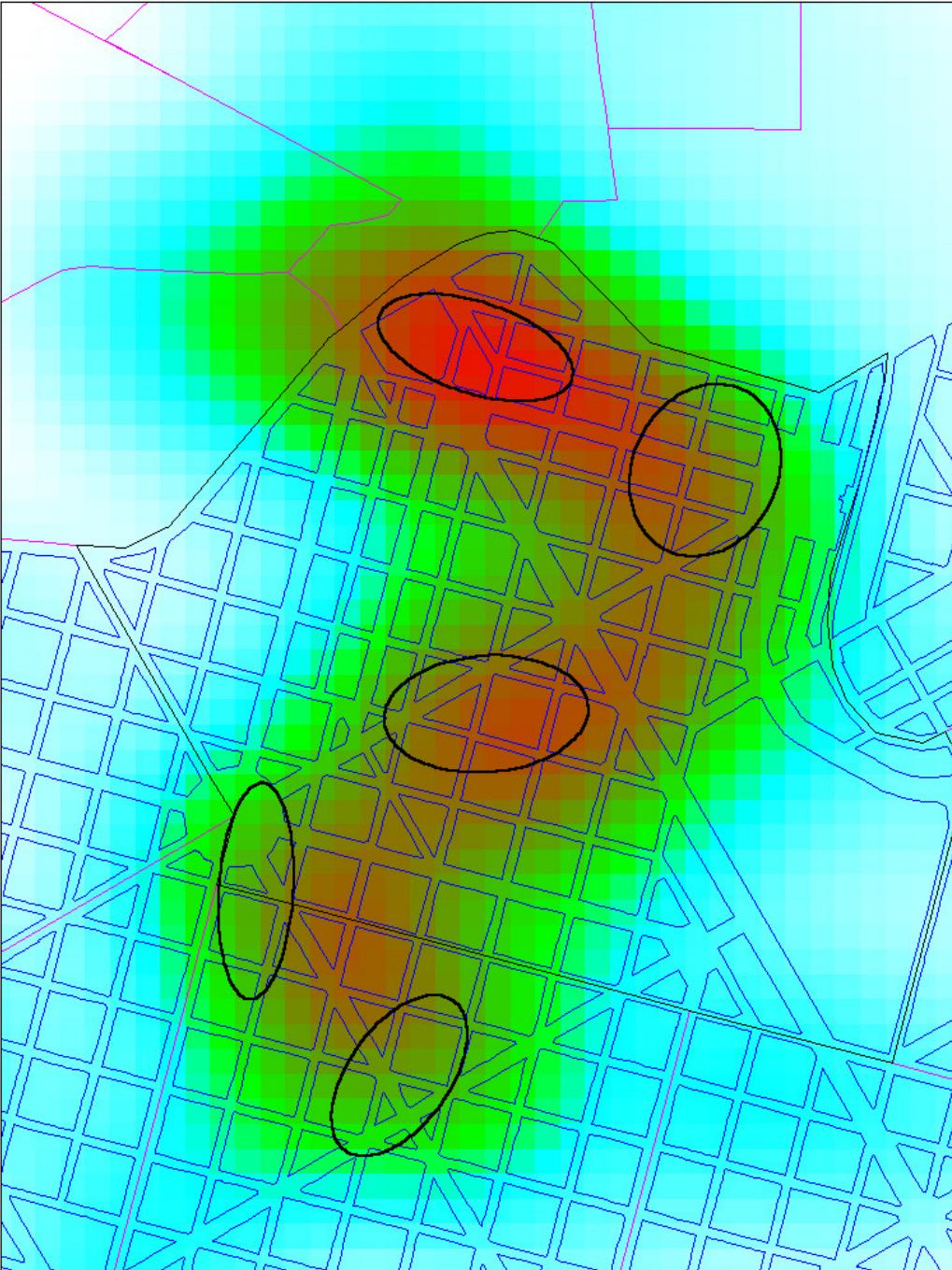


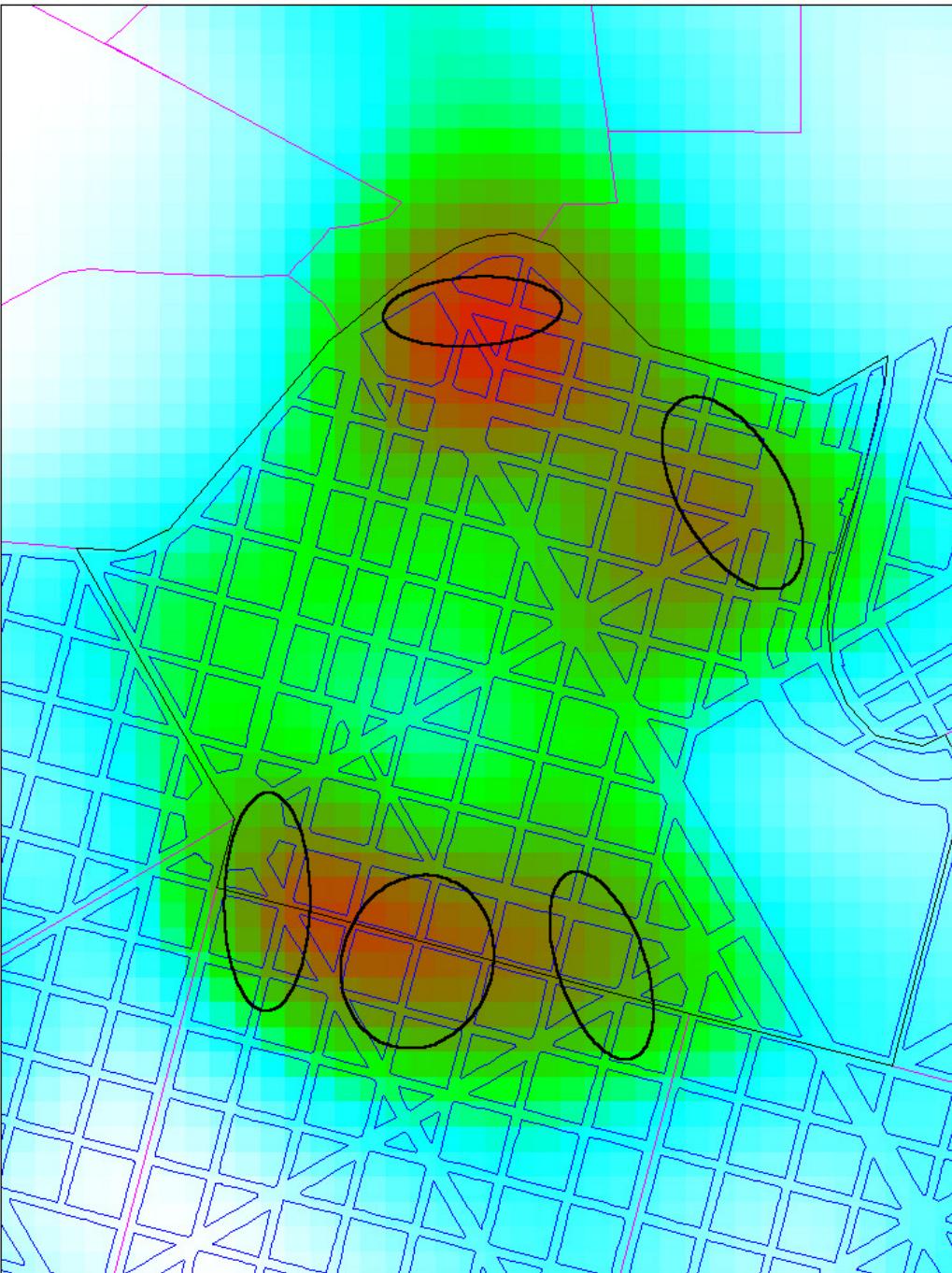


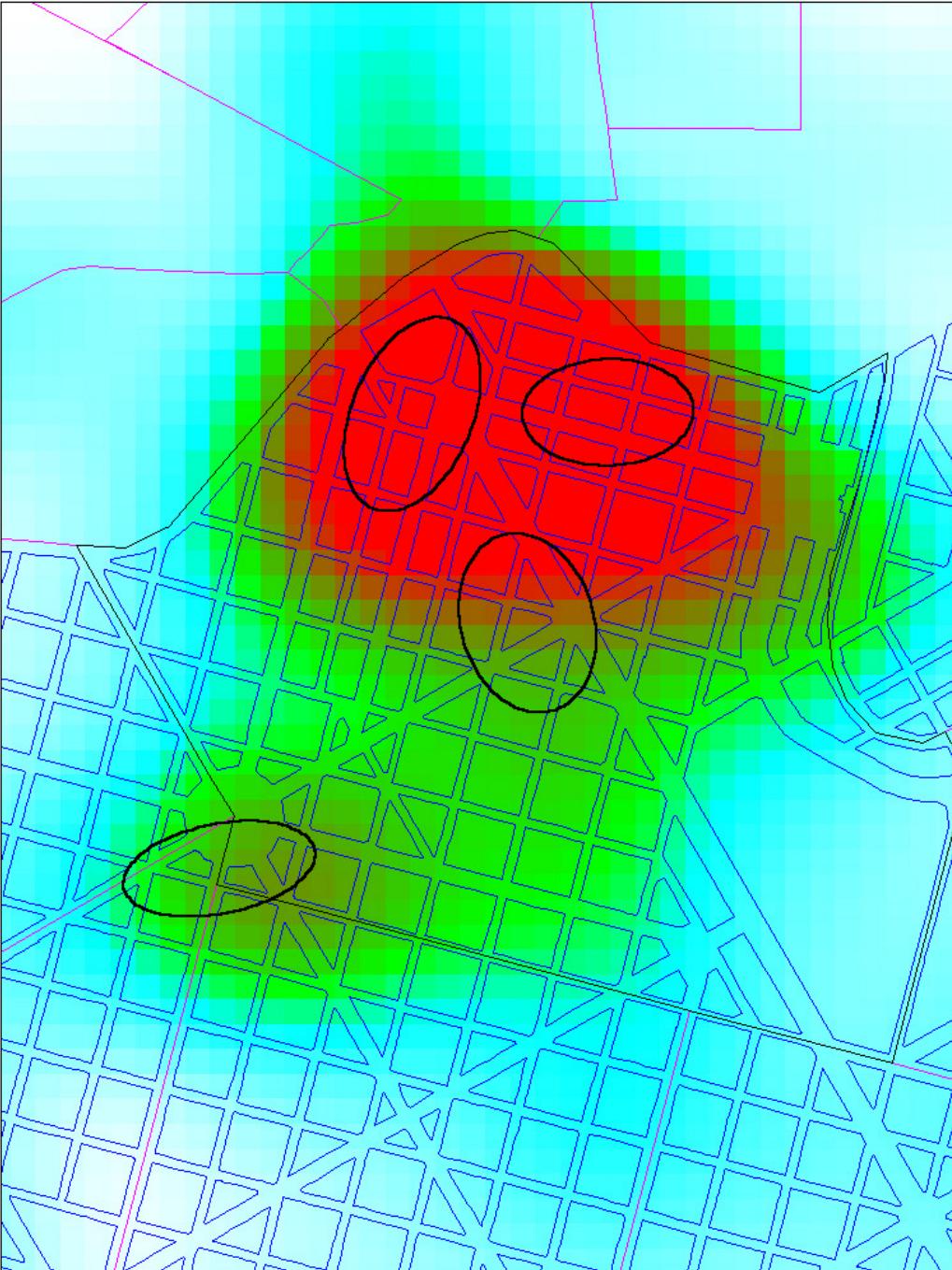


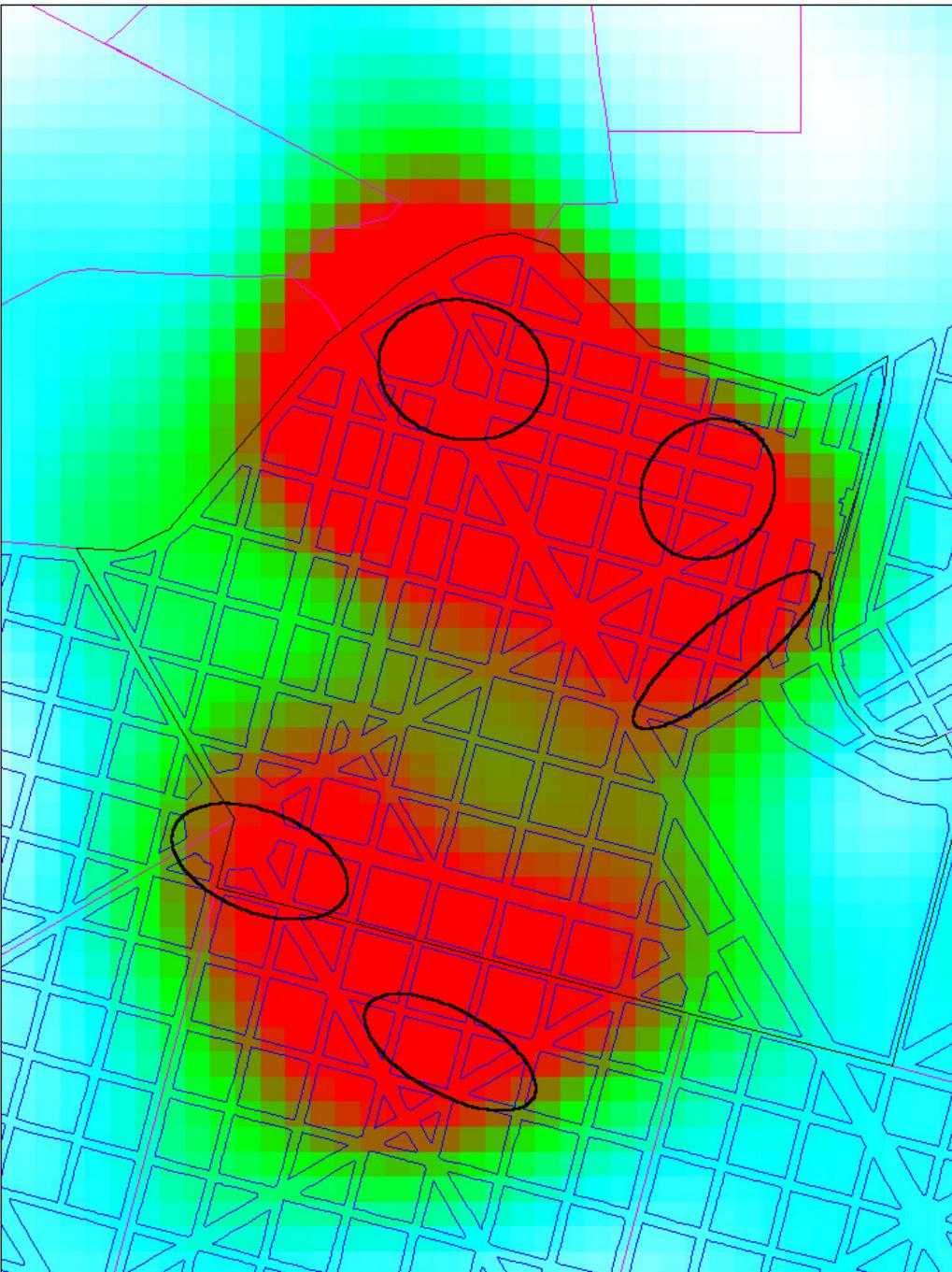


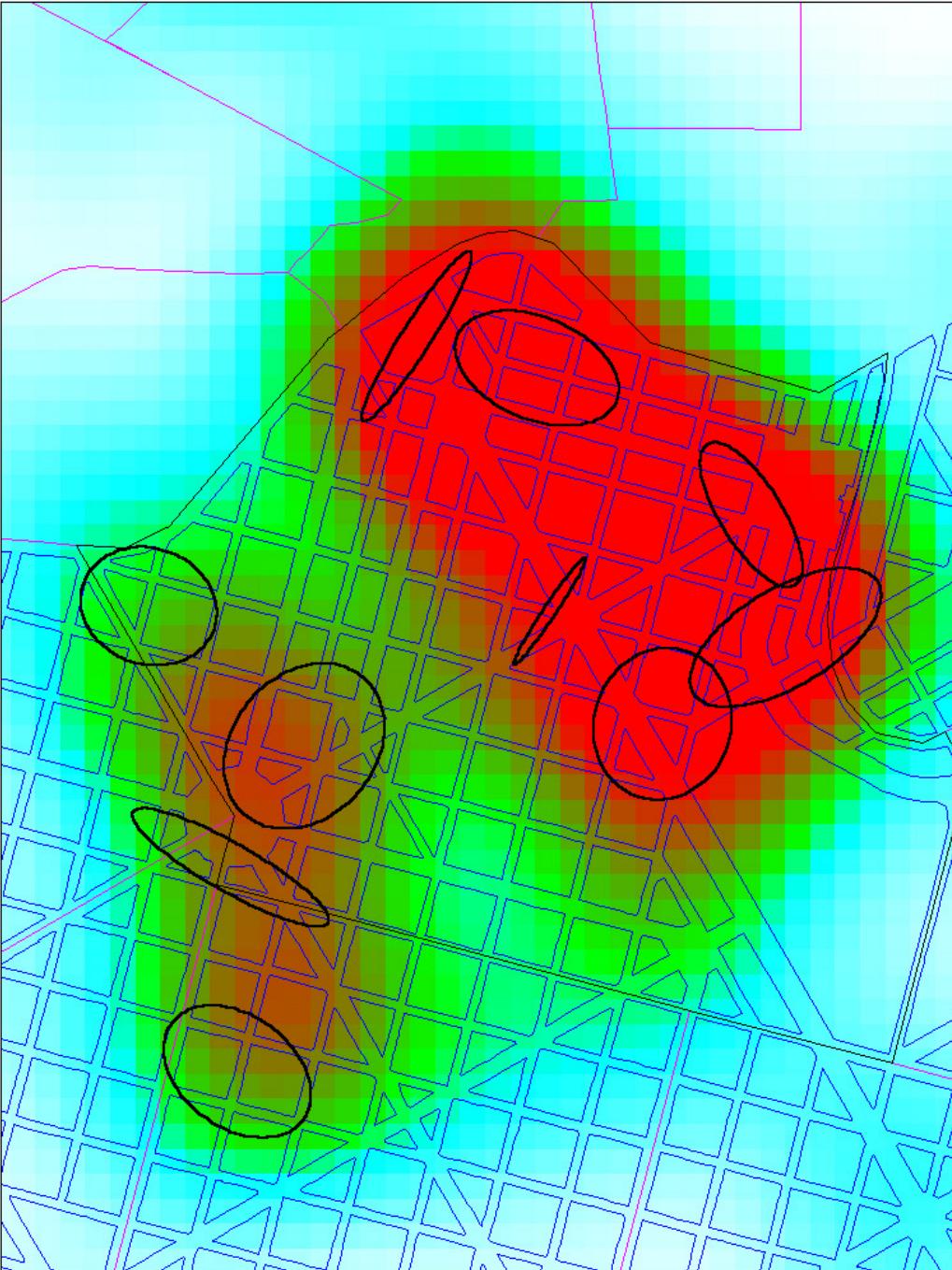


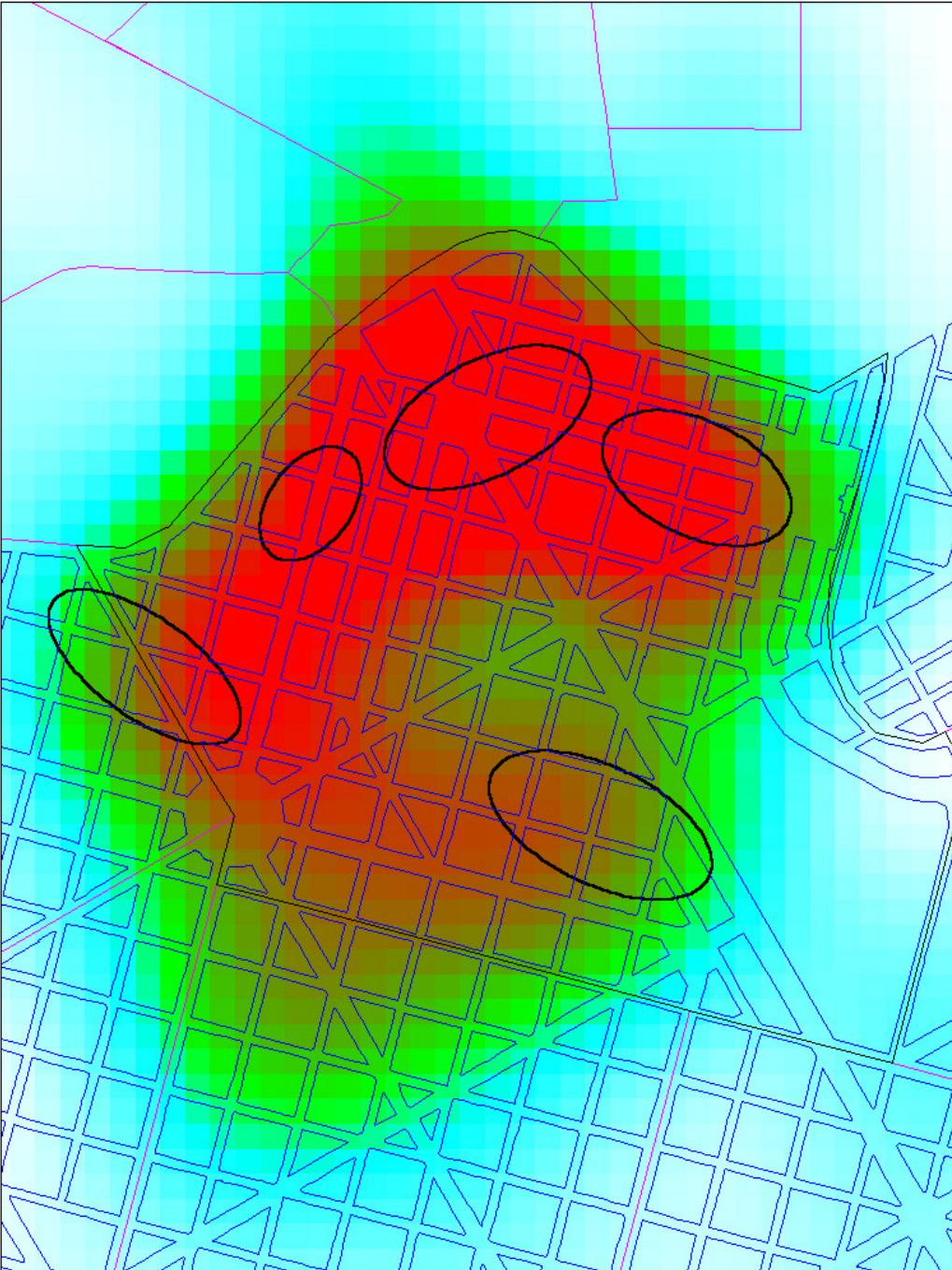


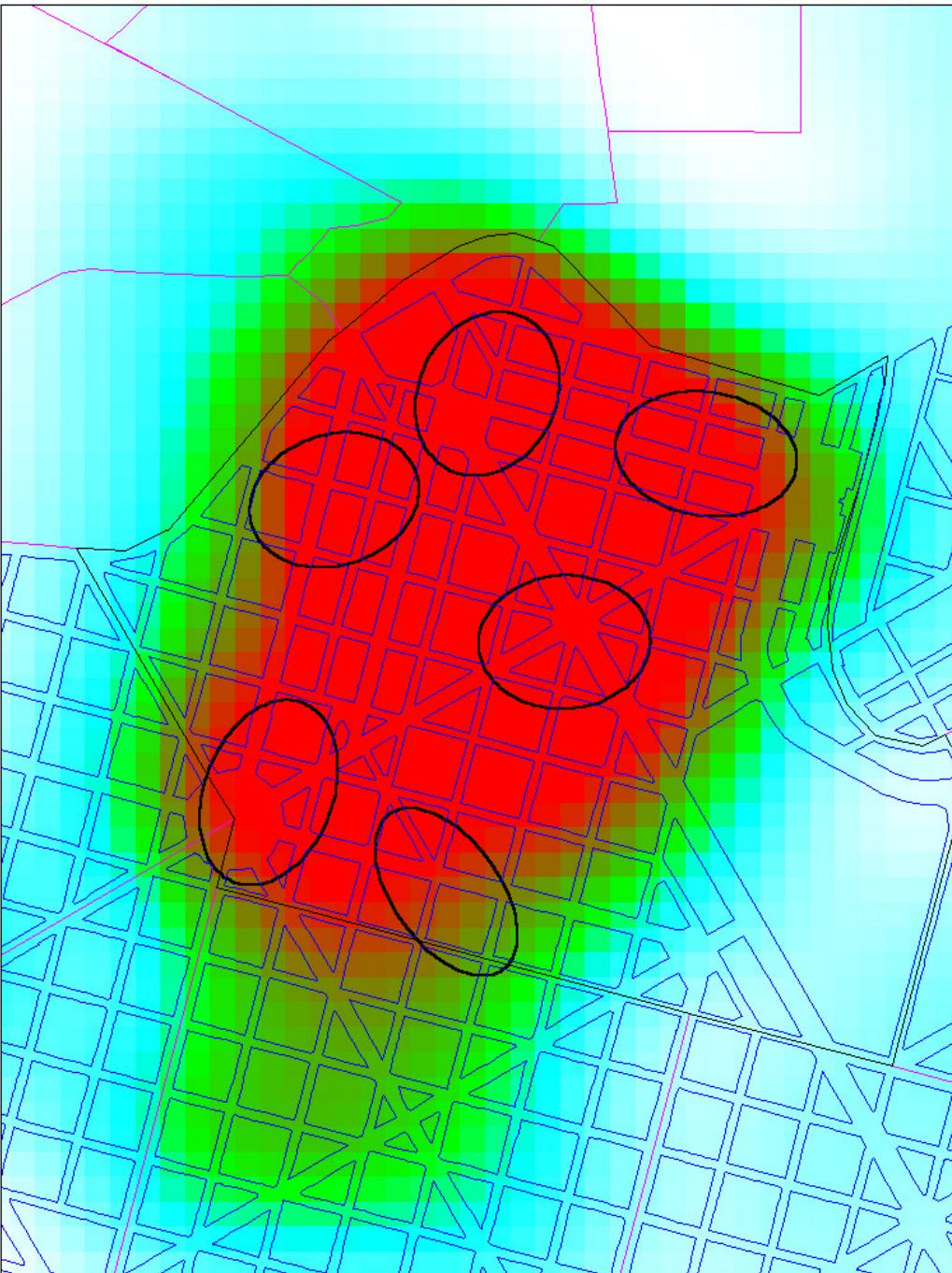


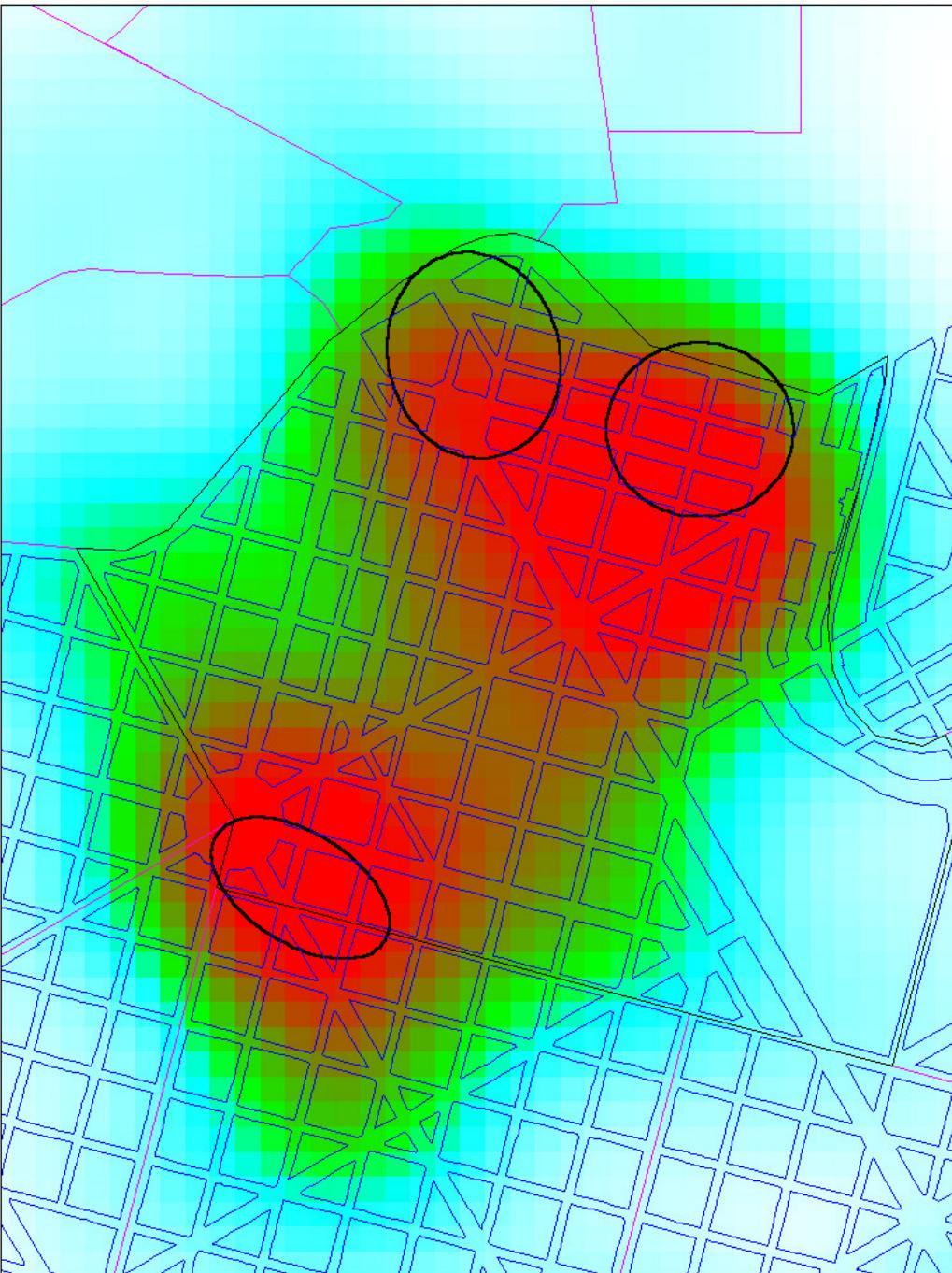


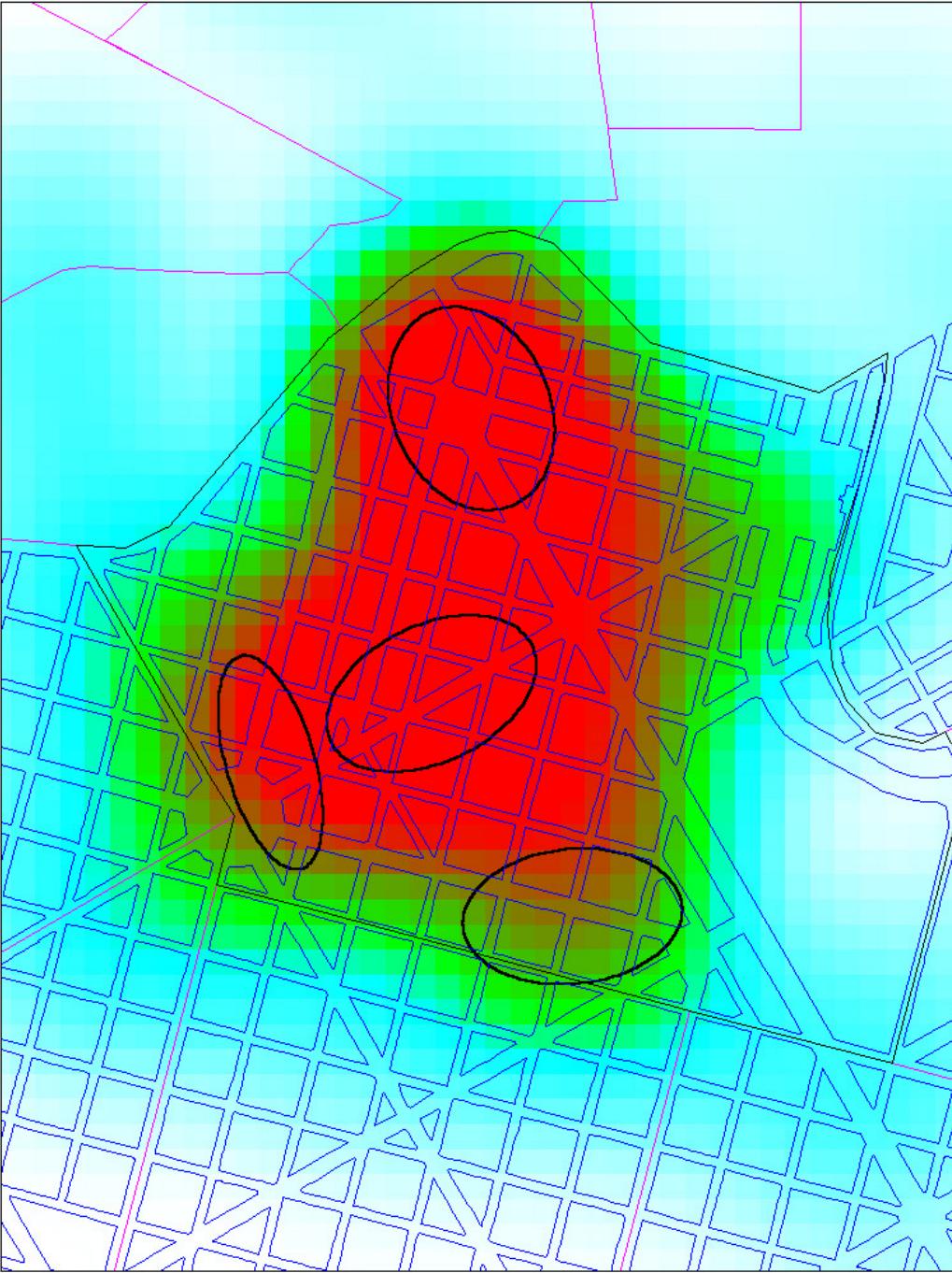


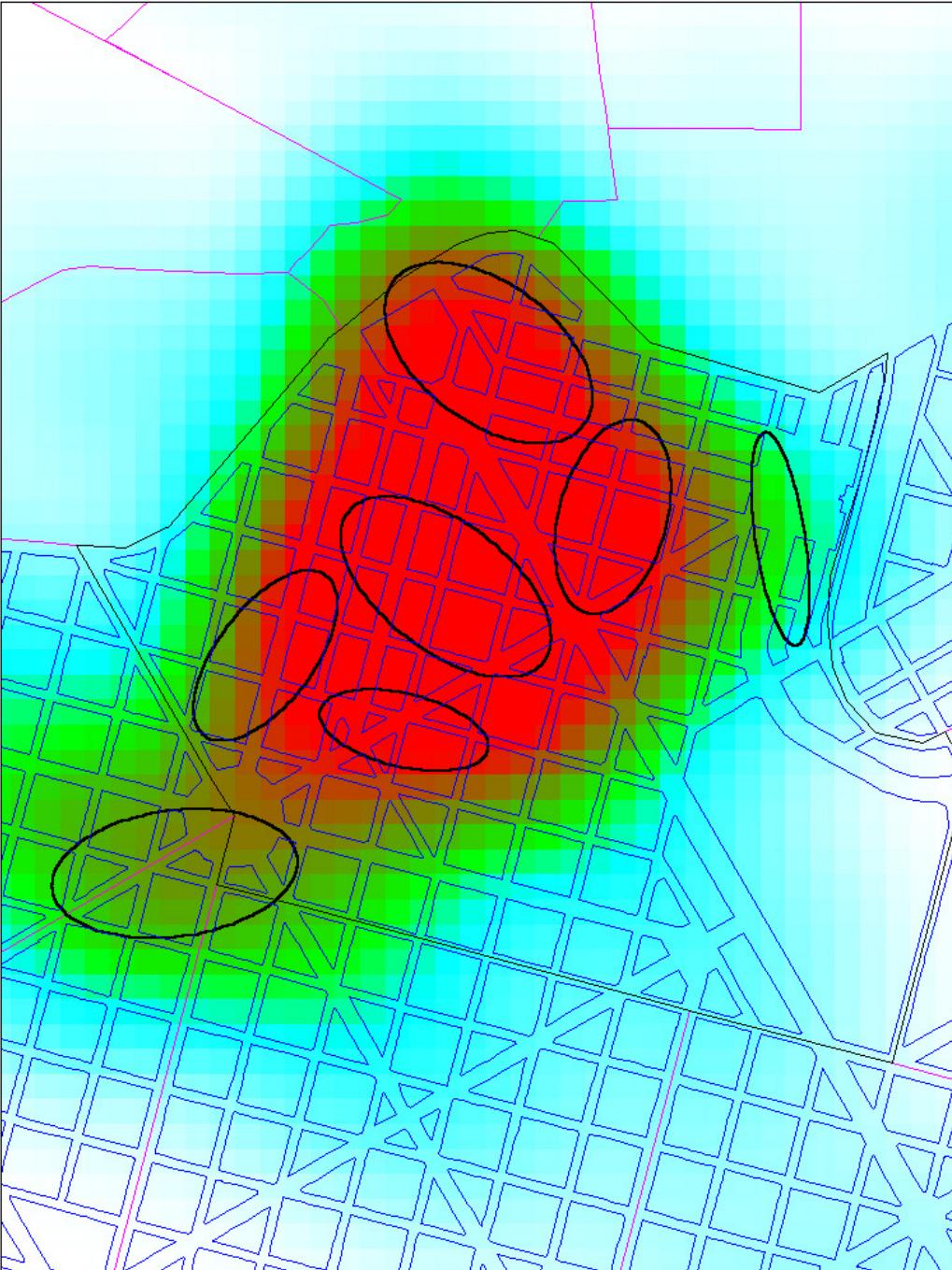


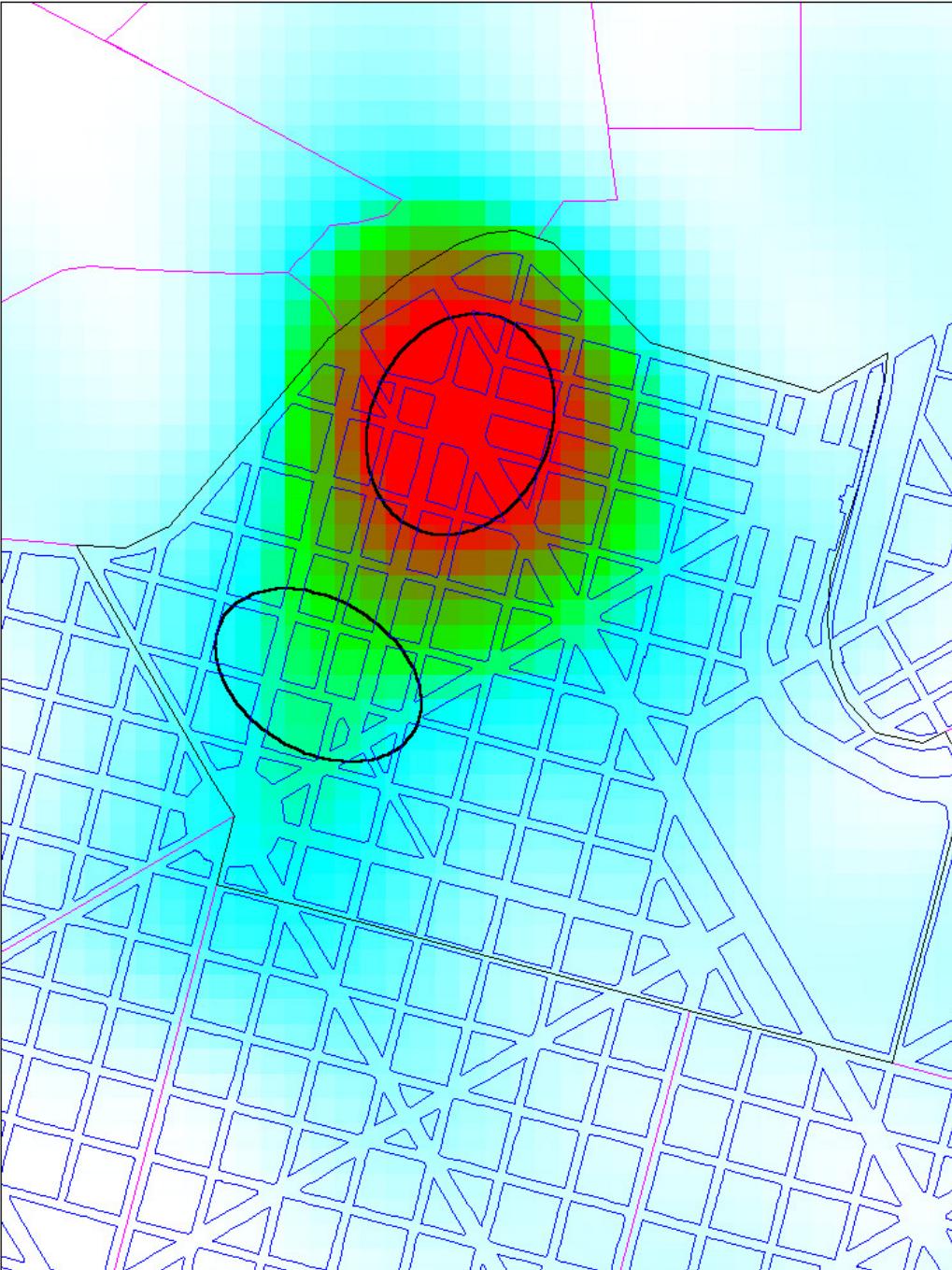












7h-8h

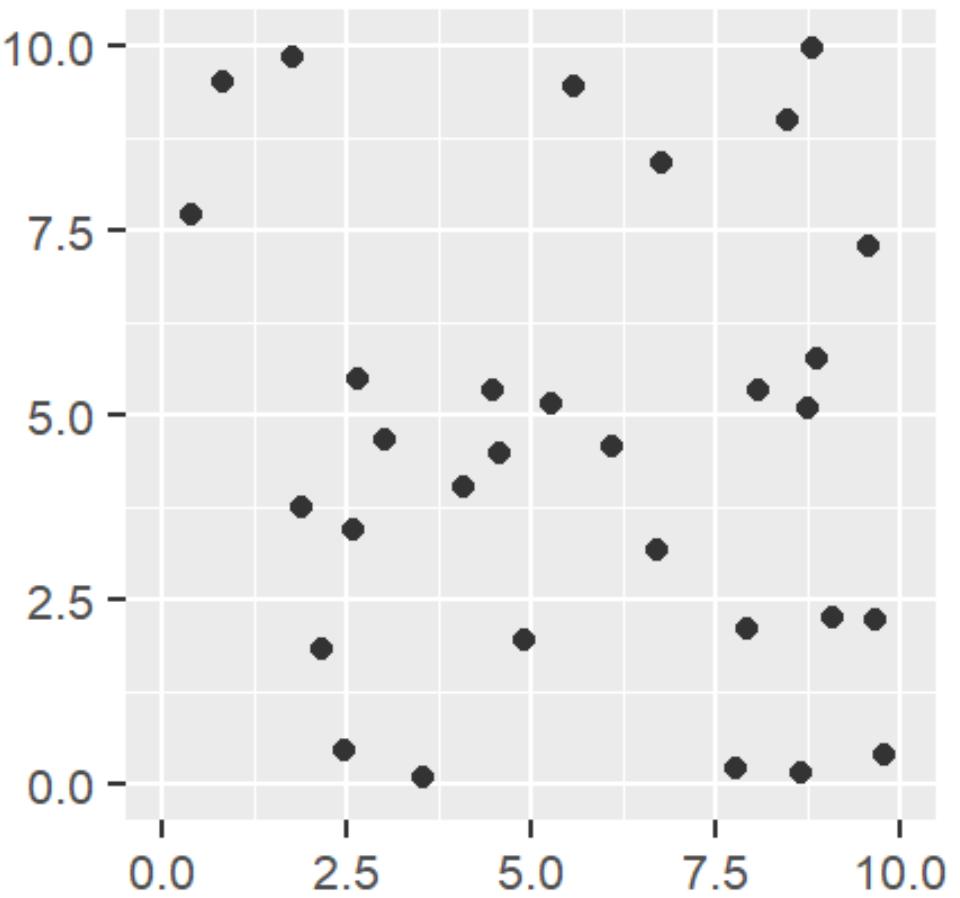


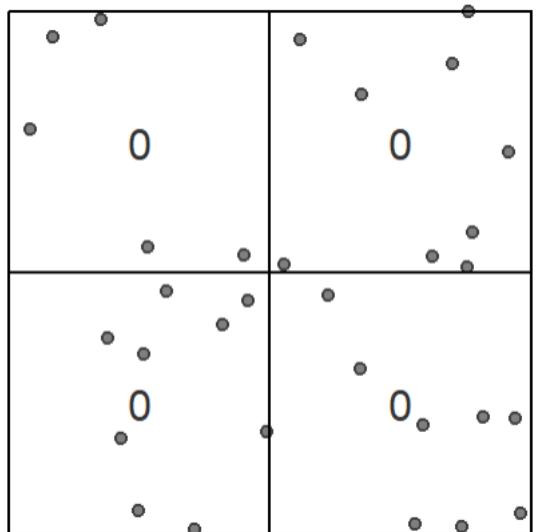
# Análise de padrões de pontos

- Análise baseada em densidade (primeira ordem)
  - Densidade global (número de pontos / área de interesse)
  - Densidade local
    - Dividir o espaço em partes (*quadrats*) e calcular a densidade de pontos em cada uma
  - Densidade de kernel (heat maps)
    - Bandwidth + acumulação

Densidade global

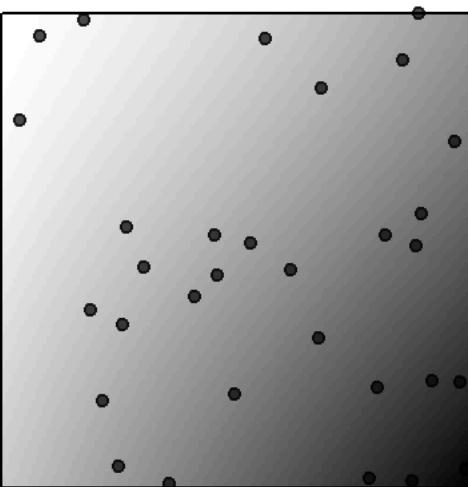
$$\hat{\lambda} = \frac{n}{a}$$



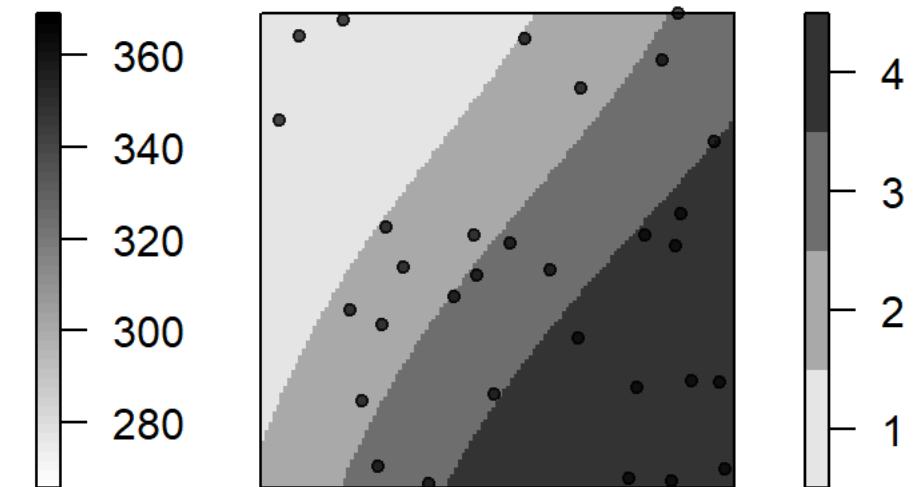


Inicial

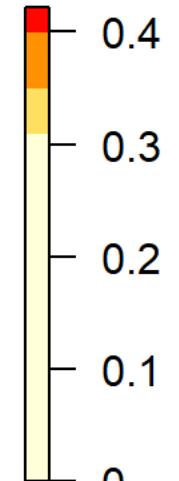
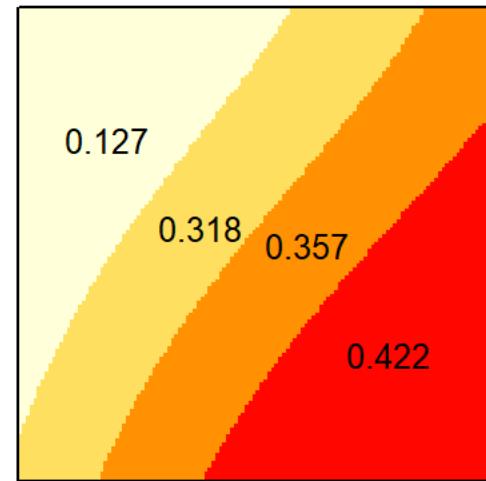
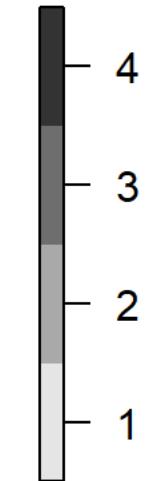
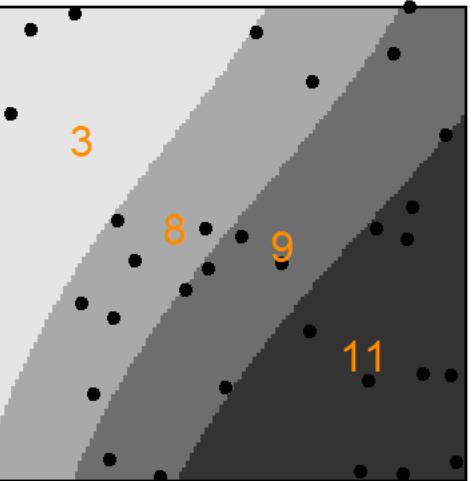
Densidade em quadrats



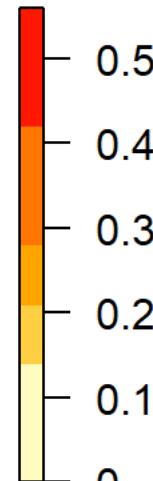
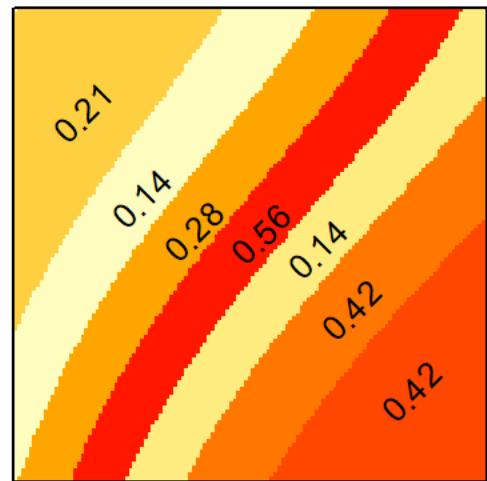
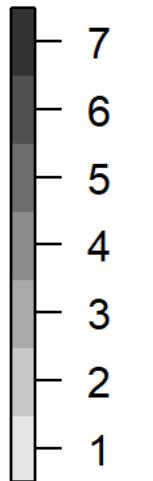
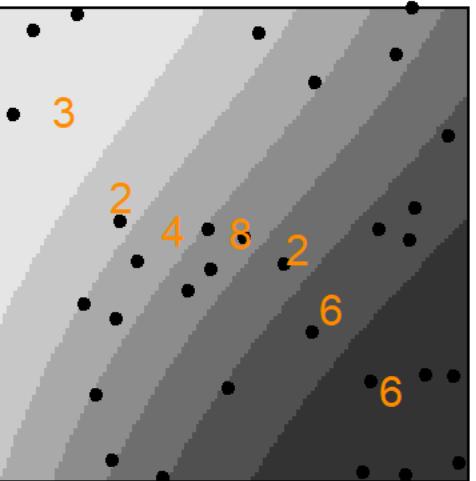
Transformado para tesselação  
após classificação por um  
atributo covariado (elevação)



Transformado para isolinhas/isoáreas



Número de pontos por sub-região – densidade por subregião

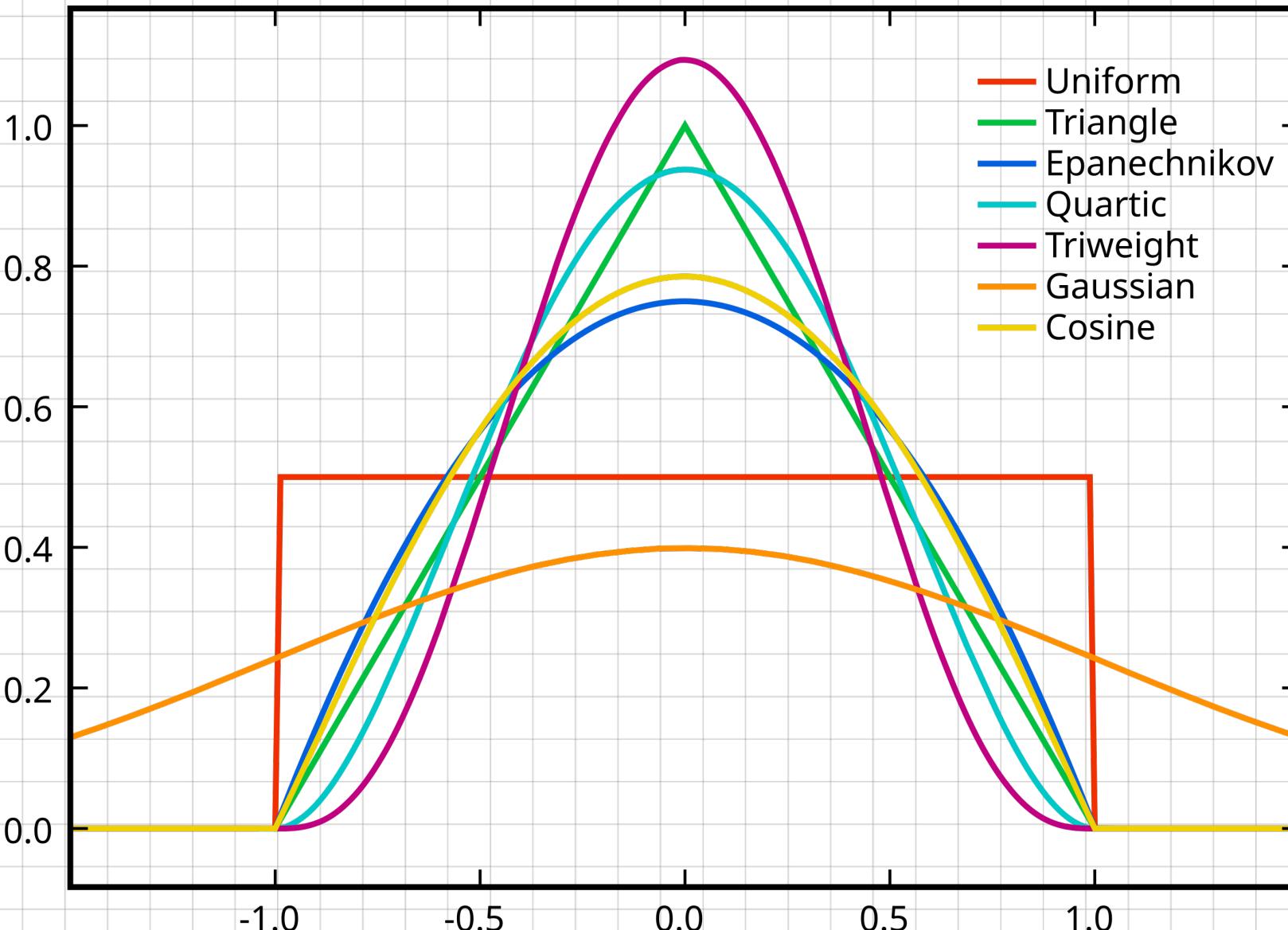


Idem – mais subregiões

# Mapas de densidade de kernel / heatmaps

- Entrada: conjunto de pontos
  - + atributo que indique peso/importância
- Parâmetros:
  - Tamanho do bandwidth
  - Resolução do raster final
  - Função de decaimento a partir de cada ponto
  - Escala de cores: se térmica → heatmap

# Kernel / funções de decaimento



## Kernel / heatmap exemplo

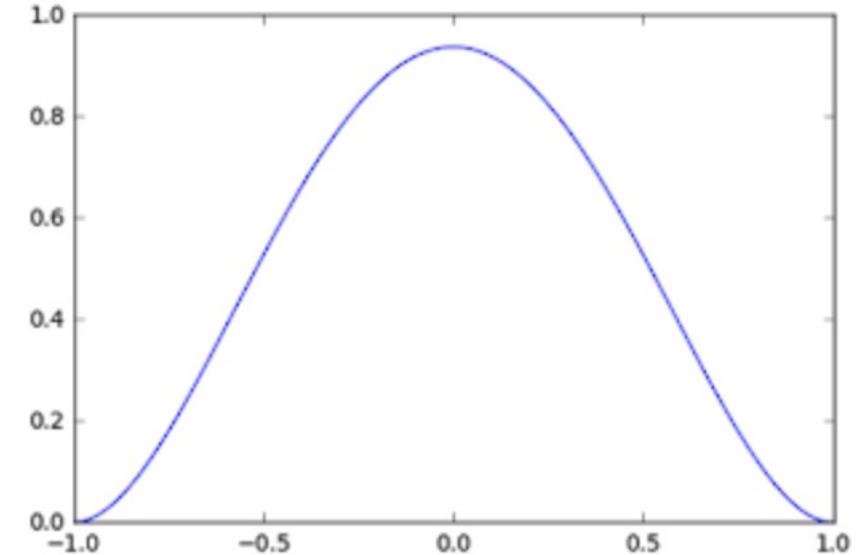
- Função de densidade de kernel quártica:

$$P(x) = \frac{15}{16}(1 - d^2)^2$$

- Mesma função, considerando peso/intensidade do ponto:

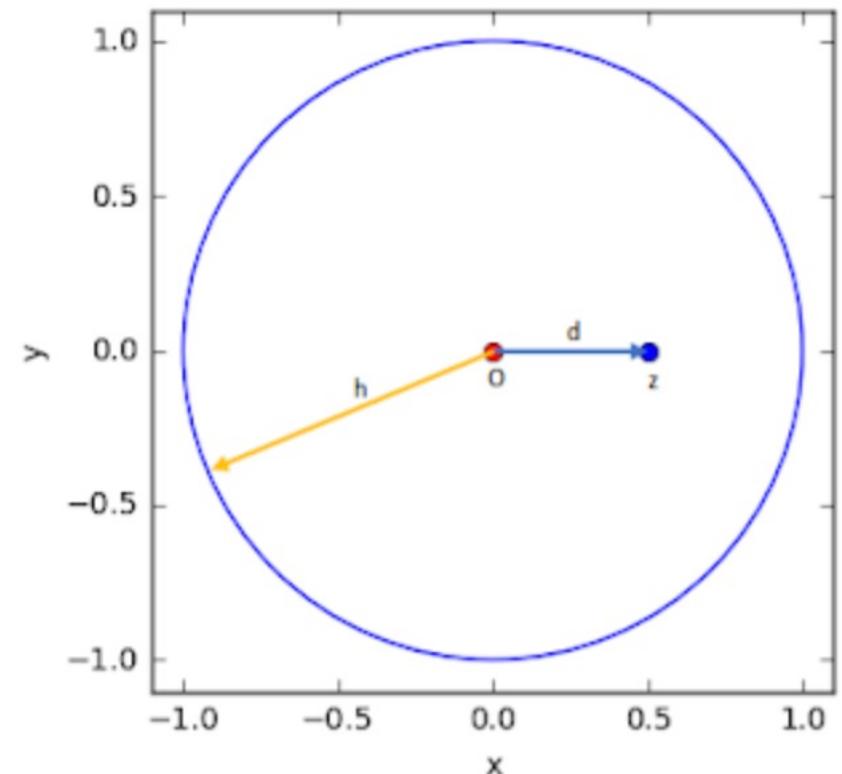
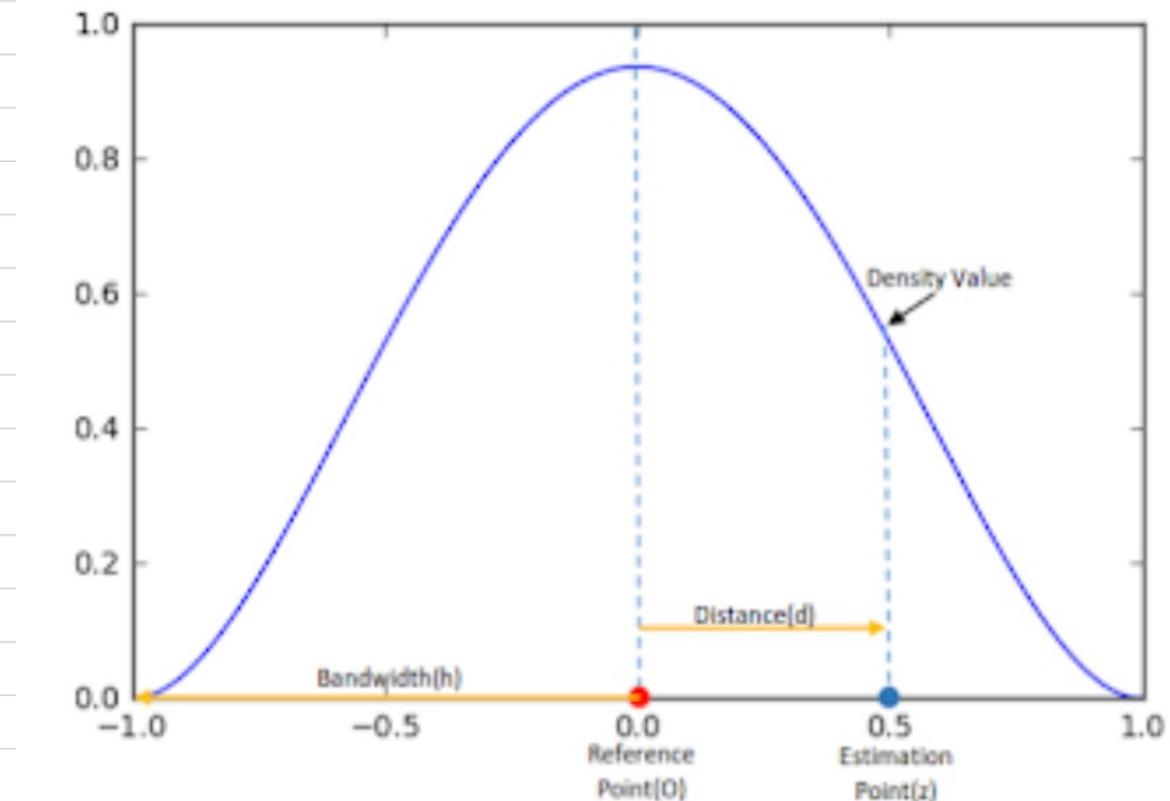
$$P(x) = KWI \frac{15}{16}(1 - d^2)^2$$

*K = constante, W = peso, I = intensidade*



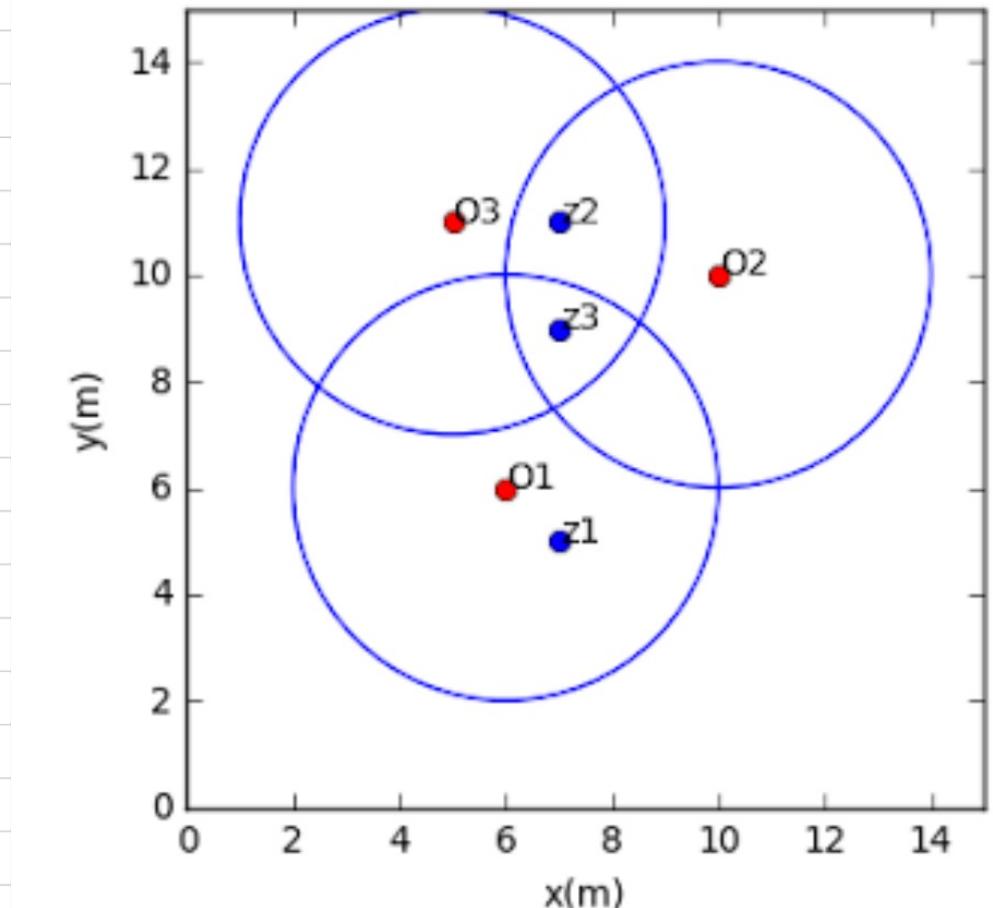
# Kernel / heatmap exemplo

- Estimativa de densidade de kernel



## Kernel / heatmap exemplo

- Acumulação de efeitos
- Normalização pela distância (bandwidth)



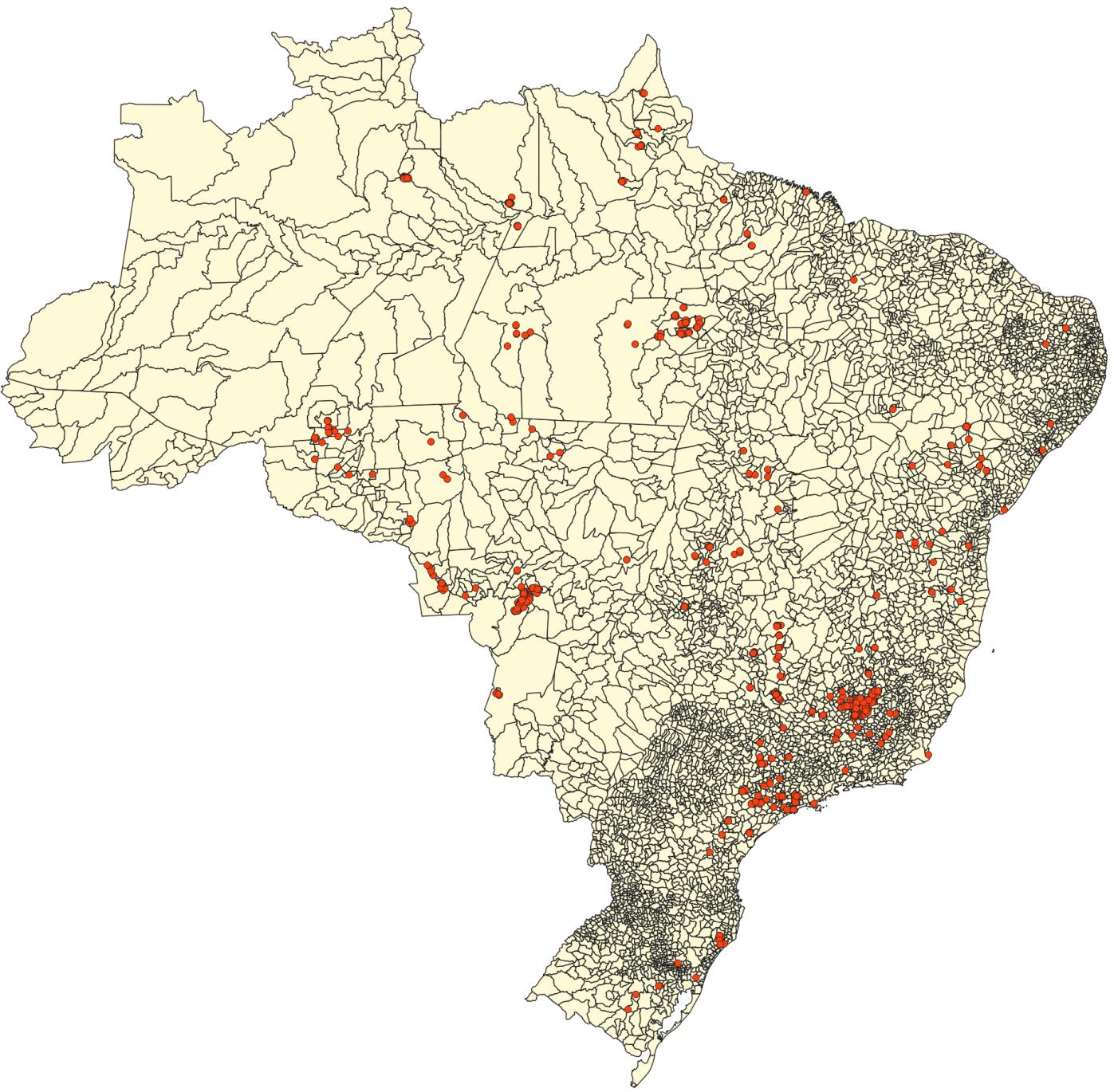
$$P(x) = \sum_{i=1}^n \frac{15}{16} (1 - d_i^2)^2$$

$d_i$  é a distância entre o ponto interpolado e a amostra  $i$

Normalizar dividindo pelo bandwidth (4m na figura)

# Exemplo QGIS

- Dados SIGBM/ANM



# QGIS exemplo

The screenshot shows the QGIS interface with a color ramp and symbology settings for a heatmap layer.

**Color Ramp:** A horizontal color bar at the top of the symbology panel, transitioning from blue to red, representing the heatmap's intensity gradient.

**Symbology Settings:** The "Layer Properties - sigbm — Símbologia" dialog shows the following parameters:

- Gradiente de cores:** A color bar matching the ramp above.
- Raio:** 10,000000
- Valor máximo:** Automático
- Ponderar pontos por:** (empty field)
- Qualidade da renderização:** Melhor

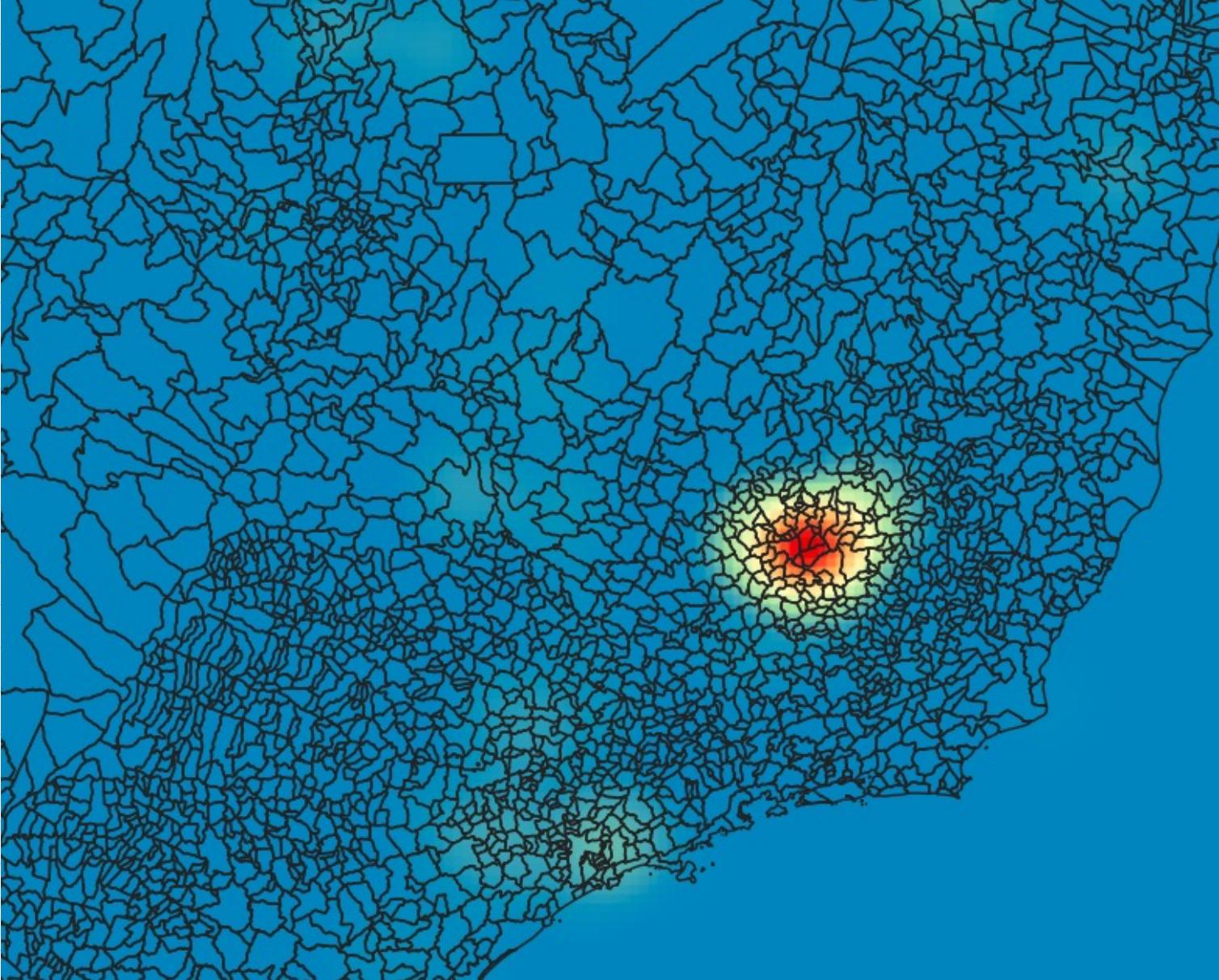
**Scale Units:** A floating menu on the right lists scale units: Milímetros, Pontos, Pixels, Metros na Escala (selected), Unidades do Mapa, and Polegadas. An arrow points from the "Metros na Escala" option to the "Milímetros" unit in the Raio field.

**Annotations:**

- Escala térmica, tipicamente** (Heatmap scale, typically) - points to the color ramp.
- Default: bandwidth em mm no mapa!** (Default: bandwidth in mm on the map!) - points to the Raio field.
- Indicação de atributo para intensidade, ou equação usando atributos** (Indication of attribute for intensity, or equation using attributes) - points to the Qualidade da renderização field.
- Resolução da tesselação regular para o resultado** (Regular tessellation resolution for the result) - points to the bottom right corner of the symbology panel.

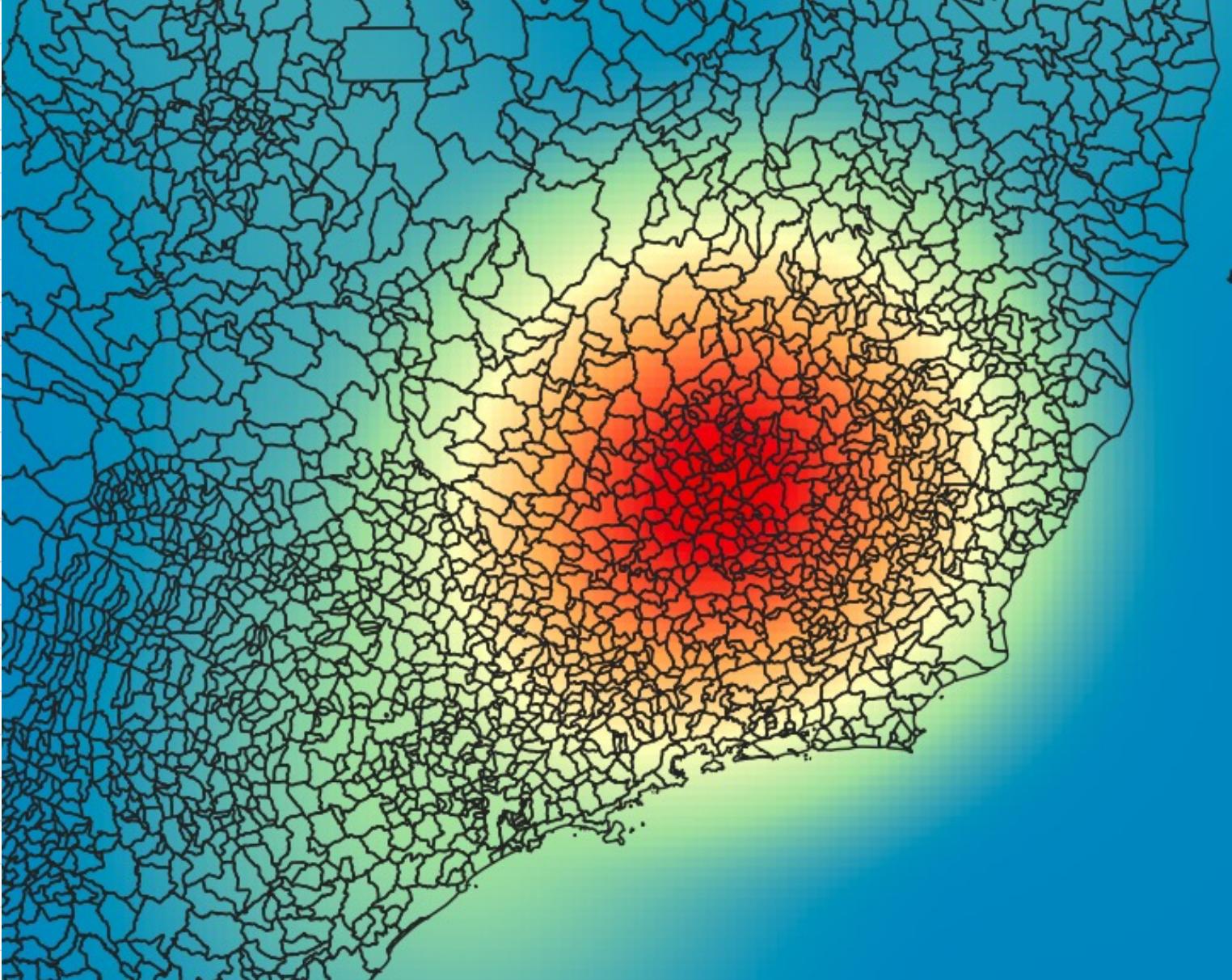
# QGIS exemplo

- Bandwidth = 100km
- Risco alto -> peso 10; risco médio -> peso 5; outros -> peso 1

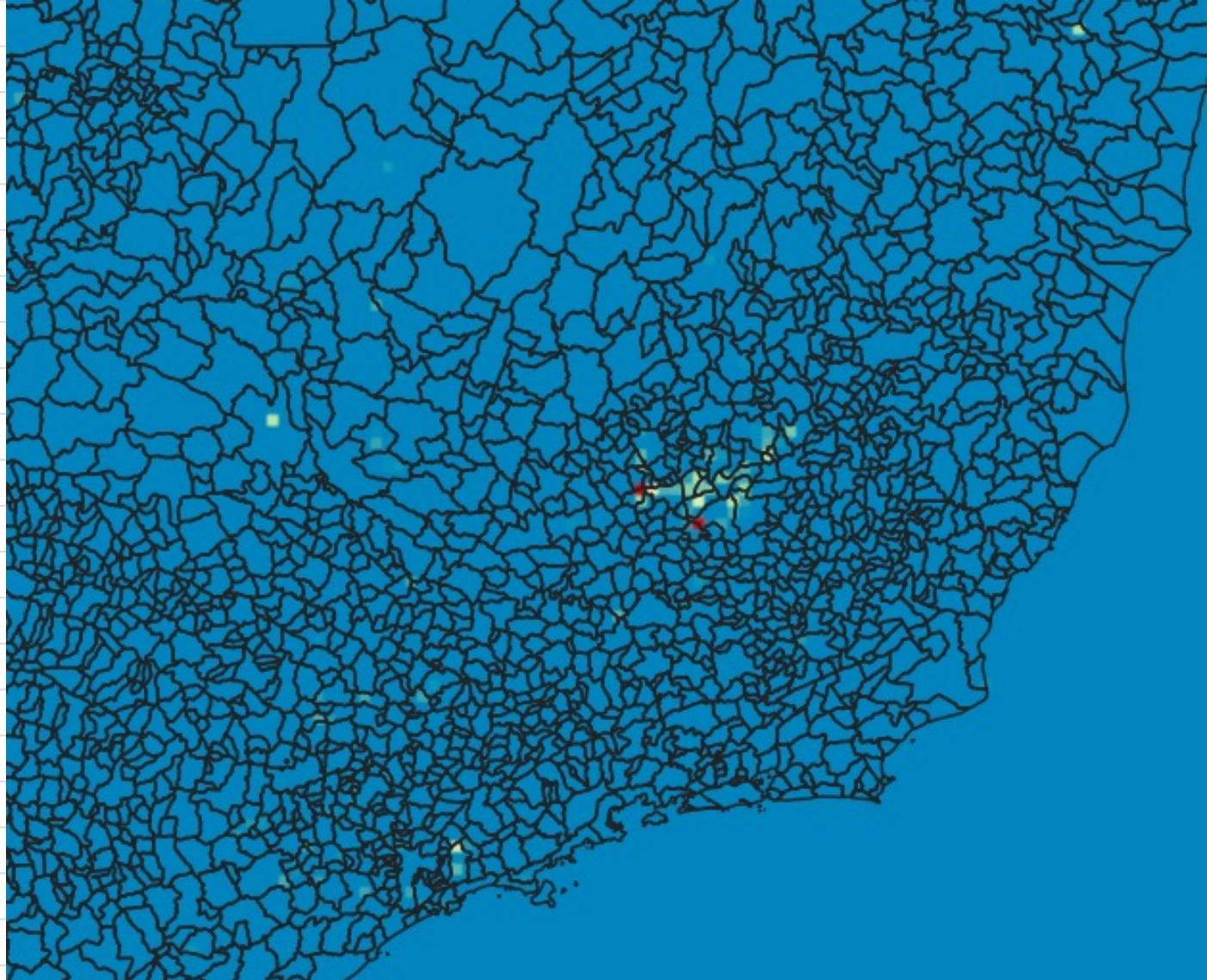
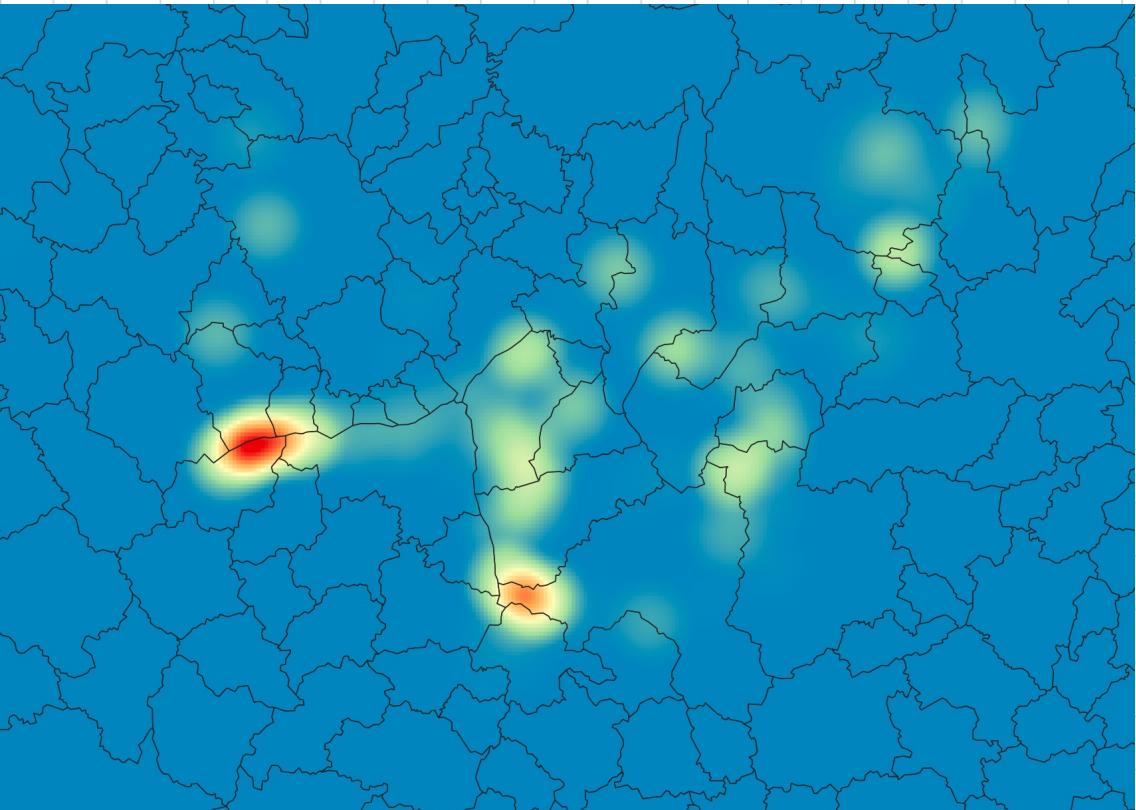


# QGIS exemplo

- Bandwidth = 500km
- Risco alto -> peso 10; risco médio -> peso 5; outros -> peso 1



## QGIS exemplo



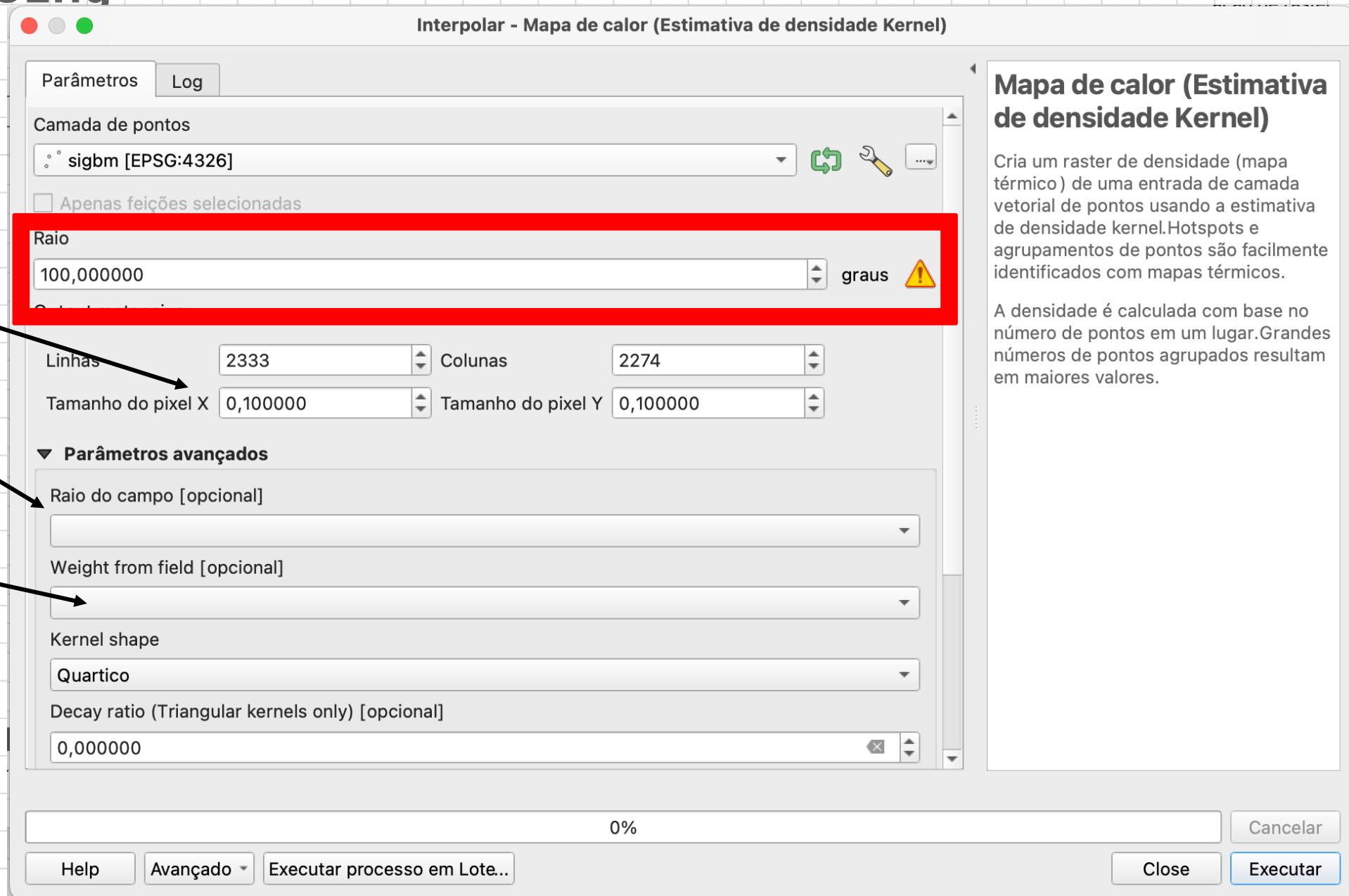
- Bandwidth = 10km
- Risco alto -> peso 10; risco médio -> peso 5; outros -> peso 1

# QGIS: Processing

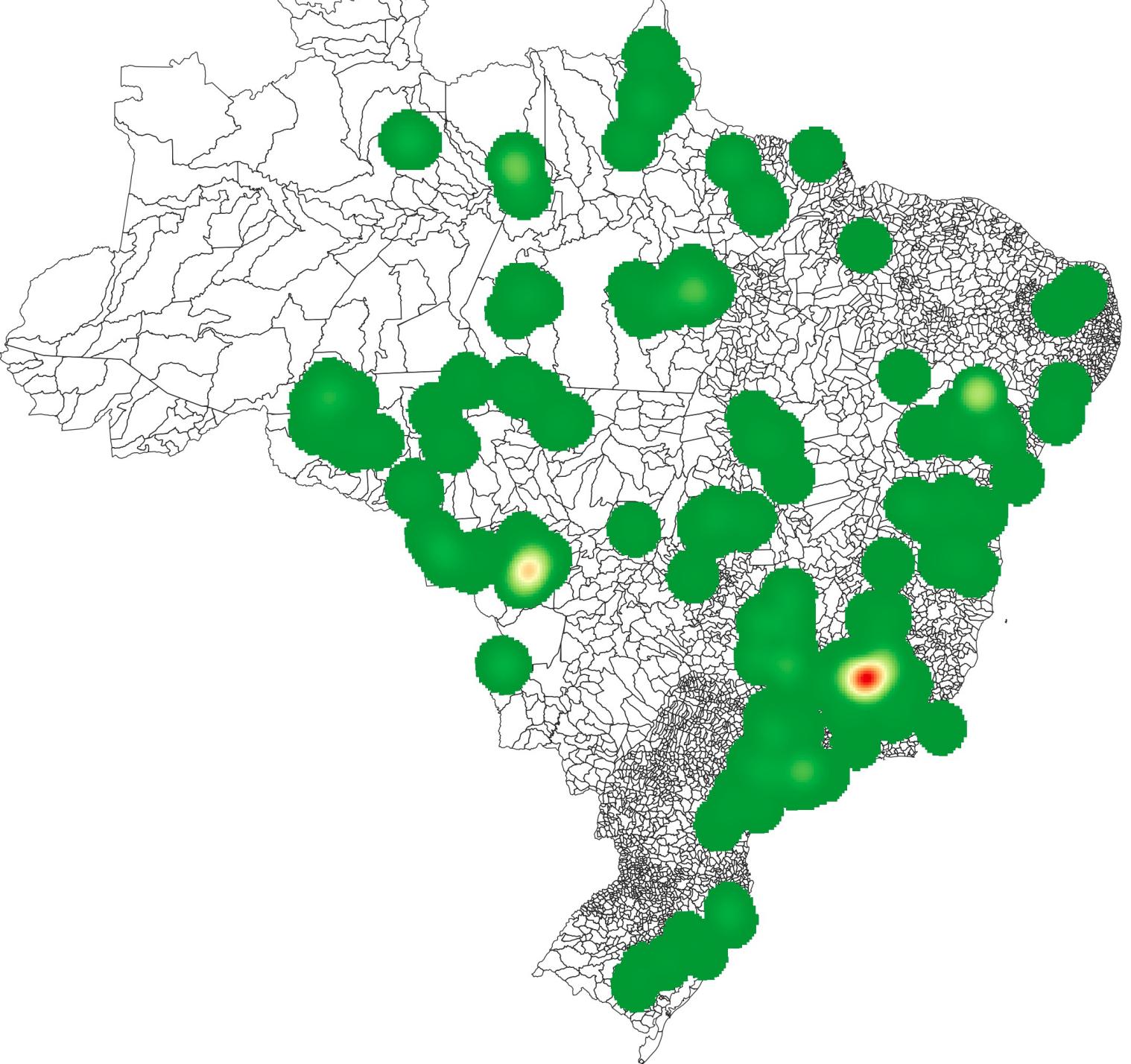
Resolução da tesselação

Bandwidth derivado de atributo

Peso derivado de atributo



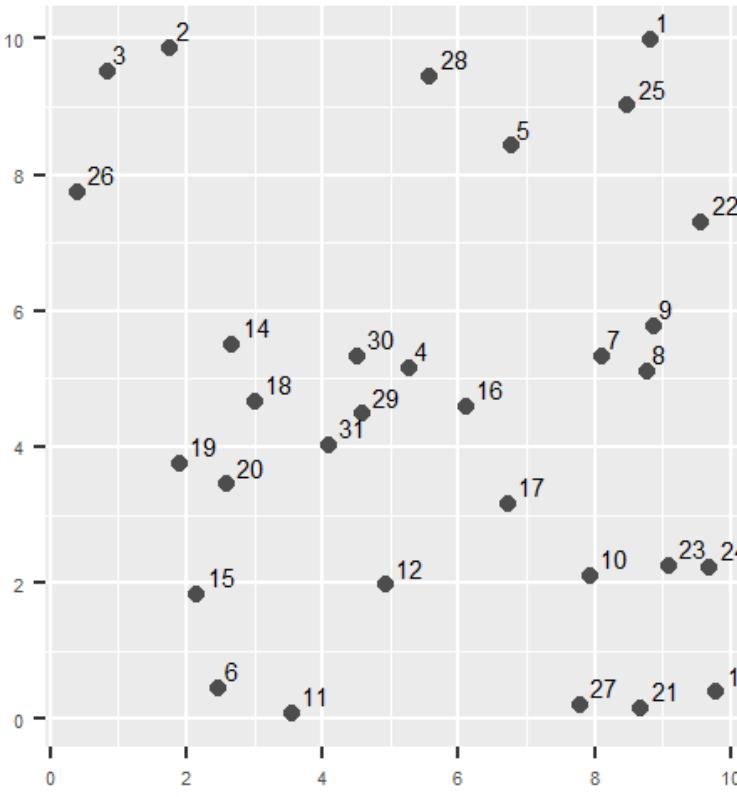
• Bandwidth = 1 grau



# Kernel / heatmap

- Problemas
  - Raio do bandwidth: se o conjunto de pontos está representado com coordenadas geográficas, o valor do raio seria em graus
  - Algumas implementações permitem definir o bandwidth em unidades de display, como percentual da tela ou pixels, o que não faz sentido espacial, e ainda faz com que o mapa mude de aspecto quando se muda a escala de visualização
  - Balanceamento entre bandwidth e valor da variável usada como peso
  - Escolha da função de decaimento
- Na escolha dos parâmetros, considerar a natureza do que está sendo representado pelos pontos (amostras)

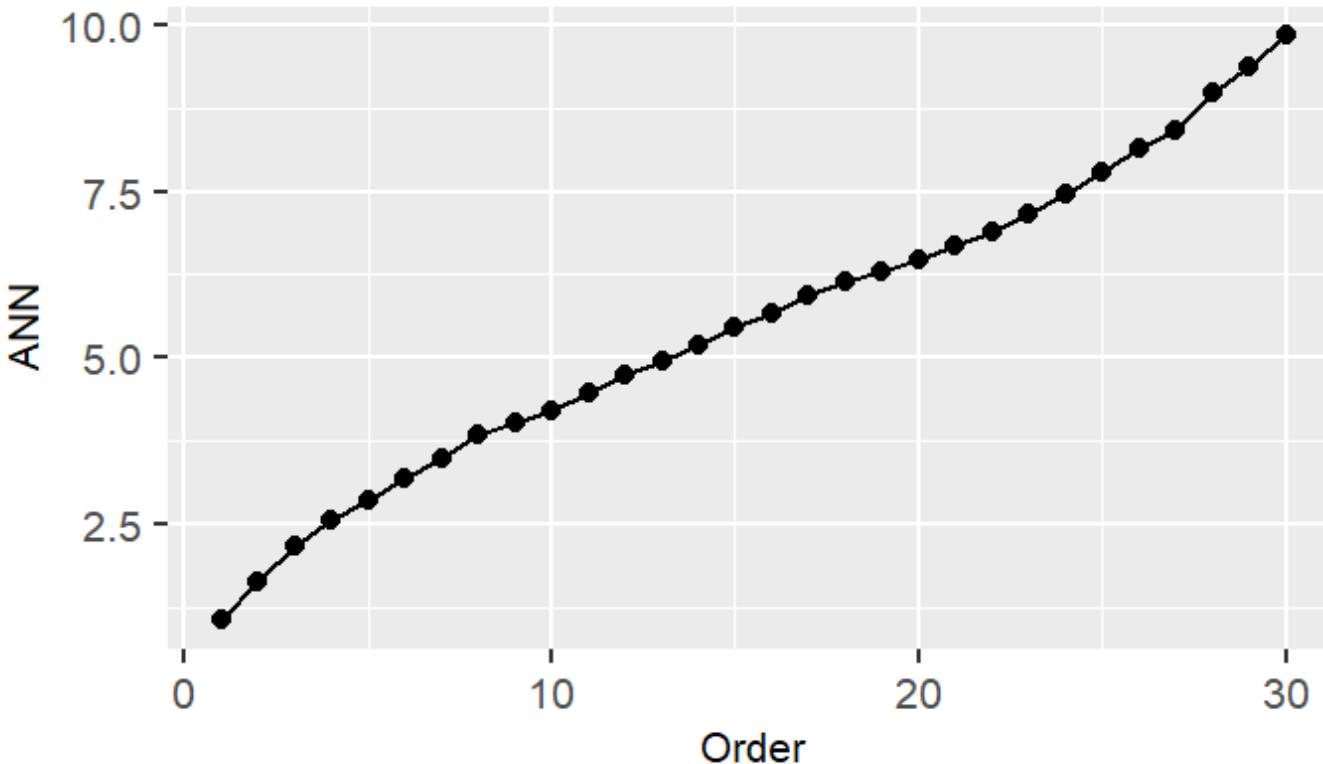
# Vizinho mais próximo médio – Average Nearest Neighbor (ANN)



From	To	Distance	From	To	Distance
1	25	1.03	11	6	1.13
2	3	0.99	12	17	2.15
3	2	0.99	13	21	1.14
4	30	0.78	14	18	0.91
5	28	1.57	15	6	1.42
6	11	1.13	16	4	1.02
7	8	0.7	17	16	1.55
8	9	0.67	18	14	0.91
9	8	0.67	19	20	0.75
10	23	1.16	20	19	0.75

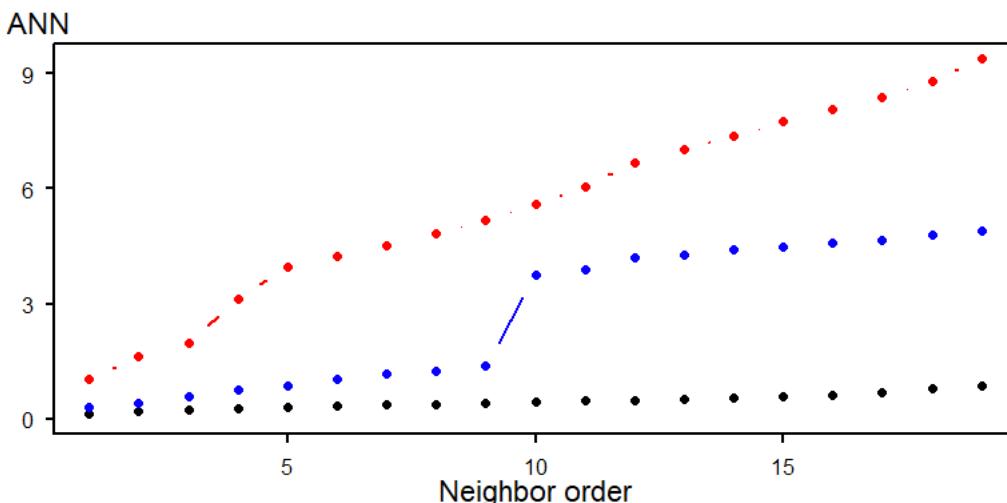
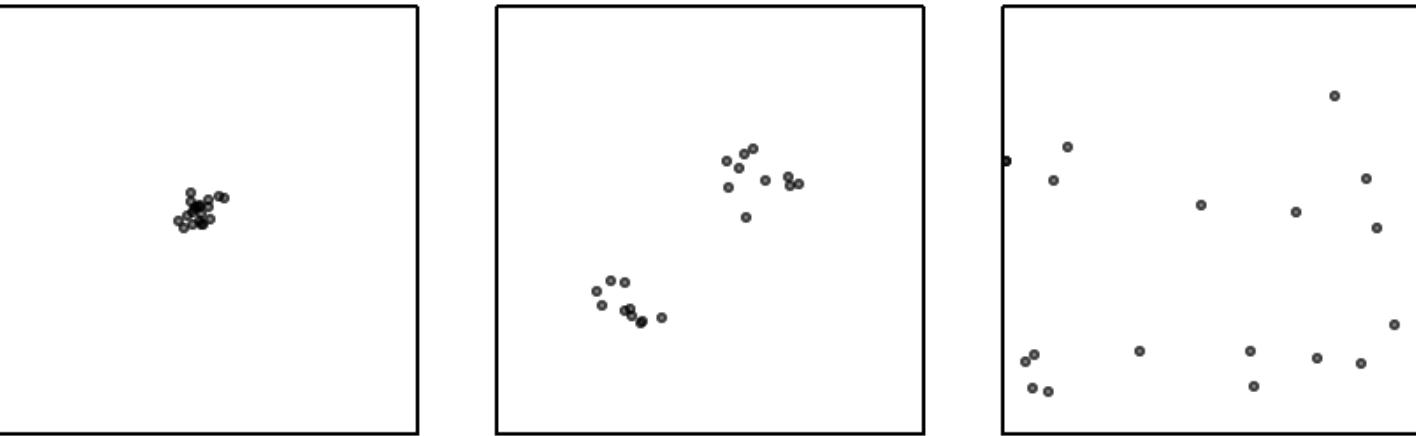
- Identificar o vizinho mais próximo, e registrar a distância até ele

# Average Nearest Neighbor



- Gerar gráfico mostrando a evolução das distâncias até o vizinho mais próximo, 2º mais próximo, 3º...

# Average Nearest Neighbor

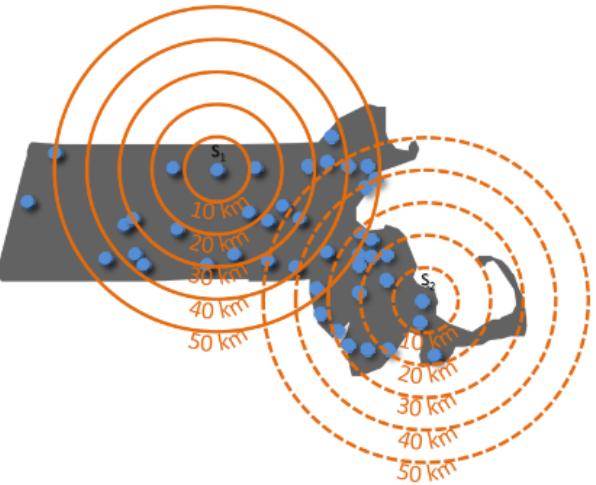


Disperso

Dois clusters

Concentrado

# Função K



Distance band (km)	# events from $S_1$	# events from $S_2$	# events from $S_i$	K
10	0	1	...	0.012
20	3	5	...	0.067
30	9	14	...	0.153
40	17	17	...	0.269
50	25	23	...	0.419

- Sumário das distâncias entre pontos para todas as distâncias, em faixas

clodoveu@dcc.ufmg.br



Links



vCard