

SER-301: ANÁLISE ESPACIAL DE DADOS GEOGRÁFICOS

Bárbara Maria Giacom Ribeiro

RELATÓRIO DE ATIVIDADES LABORATÓRIO Nº 3: ANÁLISE DE PADRÕES DE ÁREAS

1 INTRODUÇÃO

O *Laboratório 3* teve como objetivo a identificação de padrões espaciais de valores observados na distribuição de eventos cuja localização está vinculada a áreas.

Por meio de técnicas apropriadas, buscou-se determinar a existência de padrões espaciais nos valores observados. Estes padrões dizem respeito à distribuição de eventos cuja localização está associada a áreas (delimitadas por polígonos). Este caso ocorre com muita frequência quando se trabalha com fenômenos agregados por municípios, bairros ou setores censitários (como população, mortalidade e renda, por exemplo). Neste caso, não se dispõe da localização exata dos eventos, mas de um valor agregado por área.

A forma usual de apresentação dos padrões de áreas é o uso de mapas coloridos com o padrão espacial do fenômeno – mapas temáticos ou cloropléticos. O mapa da Figura 1 ilustra um exemplo no qual aplicou-se a técnica de *Agrupamento* (ou *Agregação*), onde os dados (Percentagem de Idosos por Bairro) foram agregados por *sextis*, ou seja, dividido em seis classes, de modo que cada classe possuísse aproximadamente o mesmo número de distritos; e por intervalos de valores iguais (*passos iguais*), ou seja, cada uma das seis classes possui a mesma variação entre os menores e maiores valores.

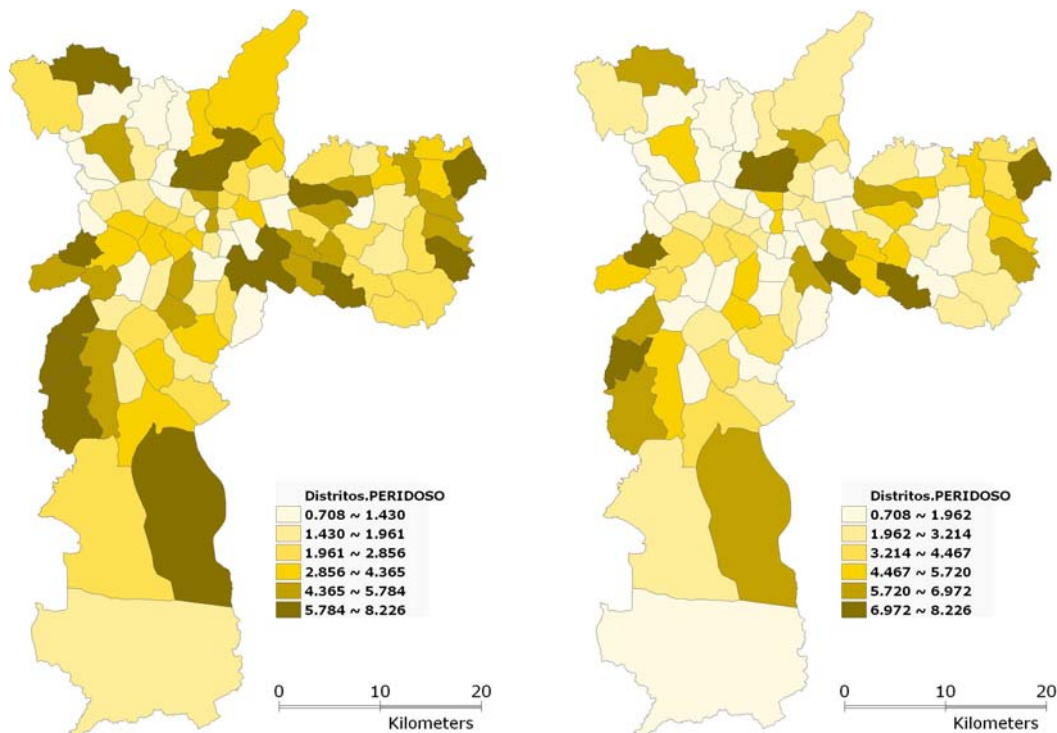


Figura 1 – visualização do atributo Percentagem de Idosos em sextis (A) e em classes de mesmo intervalo de valores (B) – software TerraView 3.2.1.

2 DADOS

Os dados a serem utilizados referem-se ao índice de exclusão social para os distritos de São Paulo, extraídos do trabalho “*Mapa de Exclusão / Inclusão Social na Cidade de São Paulo*”, resultado de uma pesquisa multi-institucional liderada pela prof^ª. Aldaiza Sposati - PUC-SP (SPOSATI, 1996).

3 DESENVOLVIMENTO

Os passos desenvolvidos neste *Laboratório 3* resumem-se em: (1) Técnicas de Agrupamento, e (2) Técnicas de Autocorrelação Espacial.

3.1. Técnicas de Agrupamento

3.1.A - Agrupamento por Passos Iguais

3.1.B - Agrupamento por Quantil

3.1.C - Agrupamento Estatístico

3.2. Técnicas de Autocorrelação Espacial

3.2.A - Índice Global de Moran – I-Moran

3.2.B - Diagrama de Espalhamento de Moran

3.2.C - Box Map

3.2.D - Gráfico de Barras $Z \times Wz$

3.2.E - Índice Local de Associação Espacial (LISA)

3.2.F - LISA Map

3.2.G - Moran Map

3.2.H - Média Espacial Móvel

3.2.I - Estatísticas G_i e G_i^* (GETIS e ORD, 1992)

O *banco de dados* utilizado foi o Sampa, e o *projeto* SaoPaulo, com projeção UTM/SAD69, determinado pelas seguintes coordenadas: 45° 39' 35,58" O e 23° 9' 11,44" S; 47° 11' 20,16" O e 24° 30' 50,68" S.

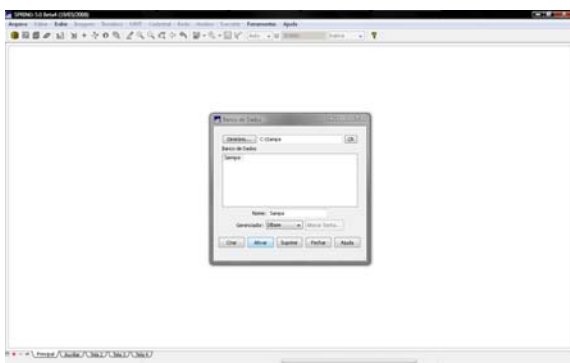


Figura 2 – ativação do *banco de dados* Sampa.

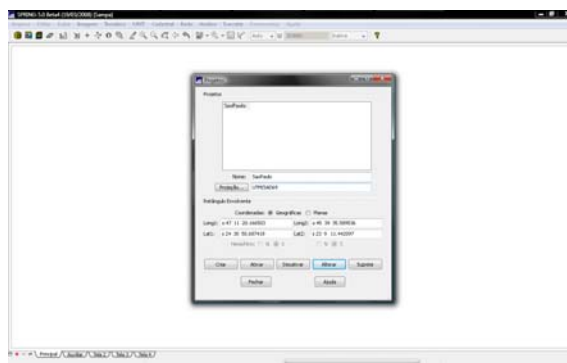


Figura 3 – ativação do *projeto* SaoPaulo dentro do banco de dados Sampa.

Inicialmente ativou-se o *banco de dados* e o *projeto* com os respectivos dados. Os dados agregados por áreas (polígonos dos Distritos de São Paulo) puderam ser visualizados através da *Tabela de Atributos* (Figura 4).

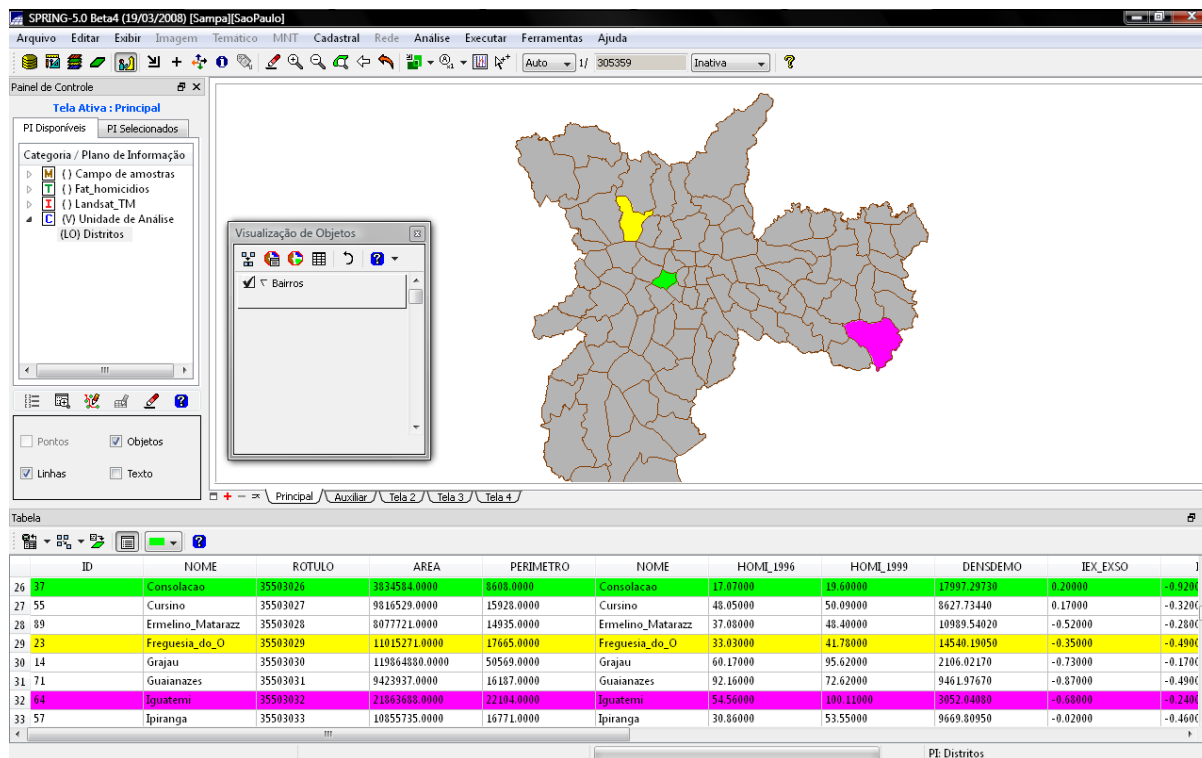


Figura 4 – visualização dos atributos relativos aos Distritos do município. Os Distritos em verde, amarelo e magenta (Consolação, Freguesia do Ó e Iguatemi) encontram-se selecionados; as colunas subsequentes ao nome do Distritos contém os dados de homicídios, densidade demográfica, etc.

3.1 Técnicas de Agrupamento

As *Técnicas de Agrupamento* baseiam-se em técnicas convencionais de visualização cartográfica, utilizando a estatística não-espacial. A apresentação dos dados requer muito cuidado, pois mapas coloridos podem levar a resultados distintos e, conseqüentemente, a várias interpretações sobre os mesmos dados.

No SPRING, as *Técnicas de Agrupamento* são aplicadas por meio da interface de *Visualização de Objetos*, selecionando a opção *Agrupamento*, conforme ilustra a Figura 5. O módulo de *Agrupamento* controla a forma de simbolizar objetos geográficos baseado nos valores de seus atributos característicos. Encontram-se implementadas as principais técnicas de agrupamento descritas por Dent (1985): passo igual, quantil, estatístico e valor único.

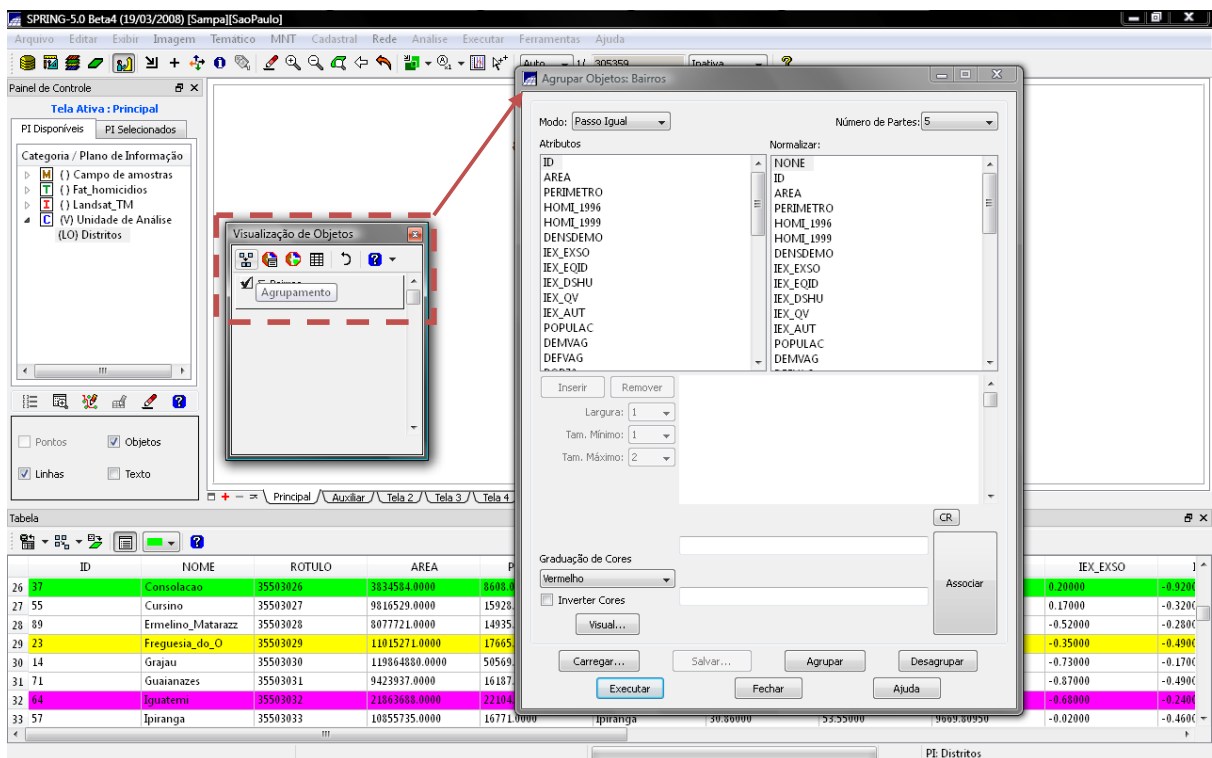


Figura 5 – Visualização de Objetos >> Agrupamento.

3.1.A -Agrupamento por Passos Iguais

Esta forma de agrupamento gera n grupos contendo o mesmo *range* de valores em cada grupo. Seu cálculo é feito da seguinte maneira:

$$Intervalo = \frac{V_{max} - V_{min}}{n} \quad (1)$$

onde:

$Intervalo$ = intervalo de valores de cada grupo

V_{max} = valor máximo do atributo

V_{min} = valor mínimo do atributo

n = número de grupos

Neste caso, o valor mínimo do intervalo de dados (Porcentagens de Idosos) é 0,7094 e o valor máximo, 8,225. O intervalo foi dividido em cinco classes, conforme mostrado na Figura 6.

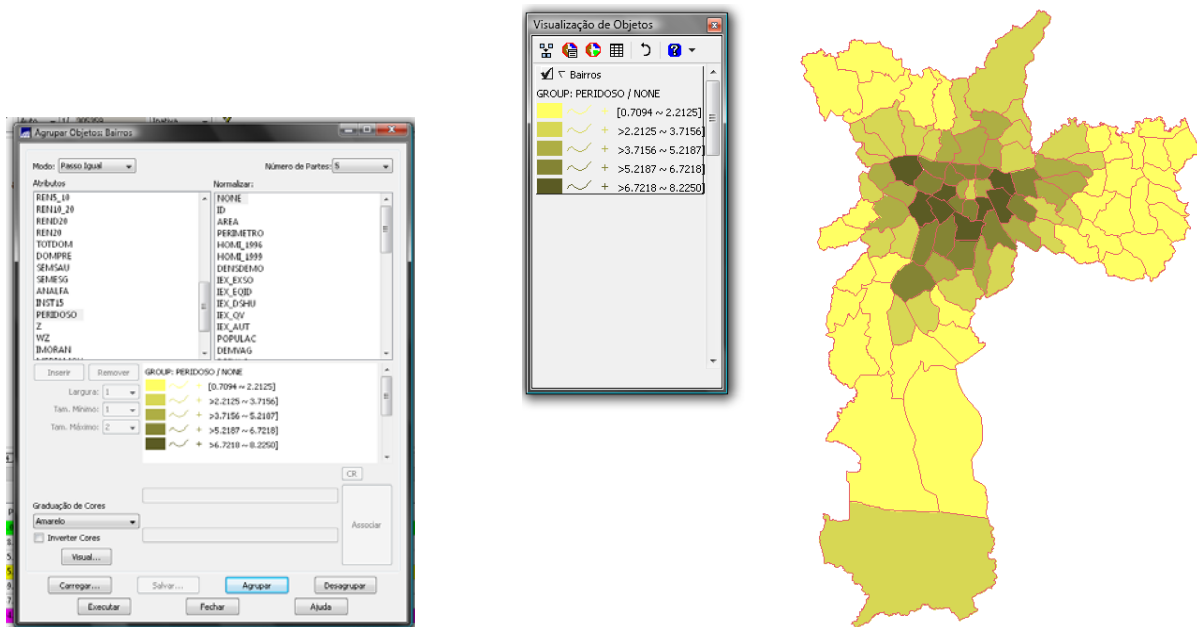


Figura 6 – O mapa cloroplético mostra o atributo Porcentagem de Idosos dividido em cinco classes, pelo método de agrupamento por *Passos Iguais*.

3.1.B -Agrupamento por Quantis

A técnica de *Agrupamento por Quantis* gera n grupos contendo o mesmo número de objetos em cada grupo. É uma forma de separar grupos de forma percentual. O intervalo de valores de cada grupo é calculado através do ordenamento de objetos pelo atributo selecionado. Após o ordenamento, calcula-se o número de objetos em cada grupo, que é igual ao número total de objetos dividido pelo número de grupos que se deseja.

$$N_{objs} = \frac{N_{Total}}{n} \quad (2)$$

onde:

N_{objs} = número de objetos de cada grupo

N_{Total} = número total de objetos

n = número de grupos.

O *range* de cada grupo é o intervalo de valores encontrados entre os objetos localizados na disposição de ordenamento.

$$sort(n-1) \geq Range(n) < sort(n) \quad (3)$$

onde:

$Range(n)$ = intervalo do n ésimo grupo

$sort(n-1)$ = valor do atributo relativo à ordem $n-1$ do ordenamento

$sort(n)$ = valor do atributo relativo à n ésima ordem do ordenamento.

Para o atributo Porcentagens de Idosos utilizado neste *Laboratório 3*, o intervalo de valores de cada grupo será diferente um do outro, entretanto, cada grupo conterà o

mesmo número de Distritos (áreas às quais estão associados os atributos). Então, por meio do *Agrupamento por Quantil*, utilizando $n = 5$, obtêm-se cinco classes onde 20% dos Distritos com as menores porcentagens de idosos estão na classe 1, 20% dos Distritos com as menores porcentagens seguintes estão nas classes 2, 3 e 4, e os últimos 20% dos Distritos, com as maiores porcentagens de idosos, constituem a classe 5 (Figura 7).

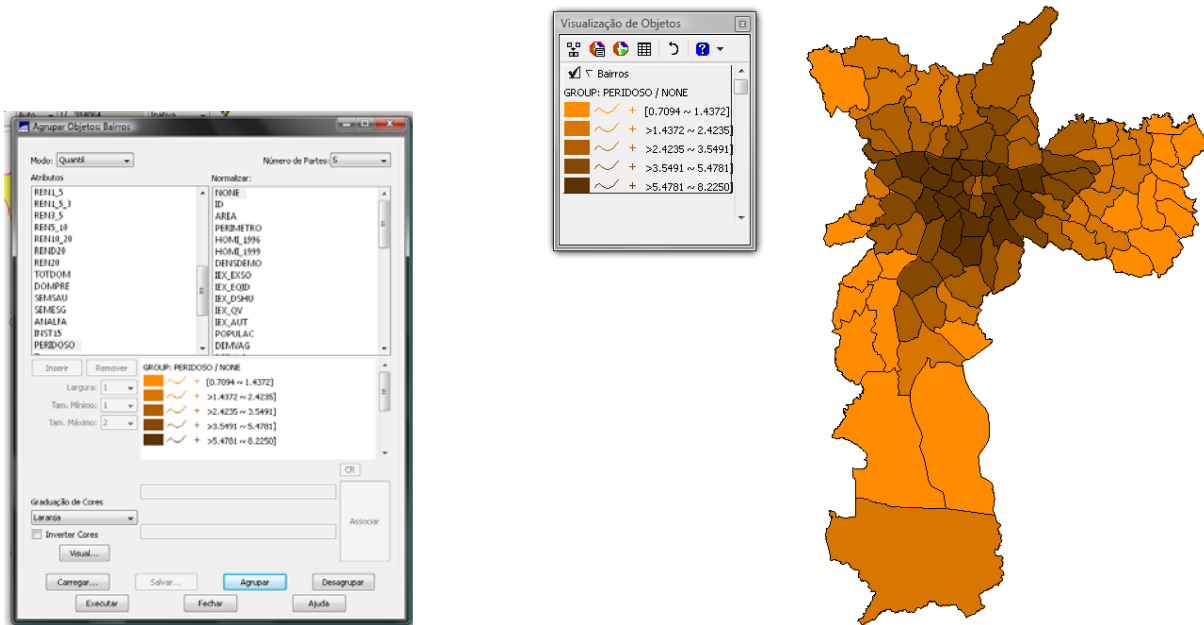


Figura 7 – O mapa cloroplético mostra o atributo Porcentagem de Idosos dividido em cinco classes, pelo método de agrupamento *por Quantis*.

3.1.C - Agrupamento Estatístico

Por meio do *Agrupamento Estatístico*, são geradas classes separadas pelo valor do *desvio padrão*. Nesse agrupamento o principal divisor é a *média* μ . A partir dela, são definidos grupos acima e abaixo da *média*, em intervalos de valor que podem ser de um *desvio padrão* ($1d$), meio ($\frac{1}{2}d$) ou um quarto ($\frac{1}{4}d$) de desvio padrão. O número de classes resultantes acima e abaixo dependerá do valor da quantidade de desvio padrão escolhido. Portanto, o número de classes será dependente deste valor e também dos valores existentes nos atributos dos objetos em questão.

A distribuição do atributo selecionado (Porcentagem de Idosos) foi dividida, acima e abaixo da *média*, em seis faixas com dimensões de 1 *Desvio Padrão* (Figura 8). O procedimento foi repetido com a definição de nove faixas com dimensões de $\frac{1}{2}$ *Desvio Padrão* (Figura 9).

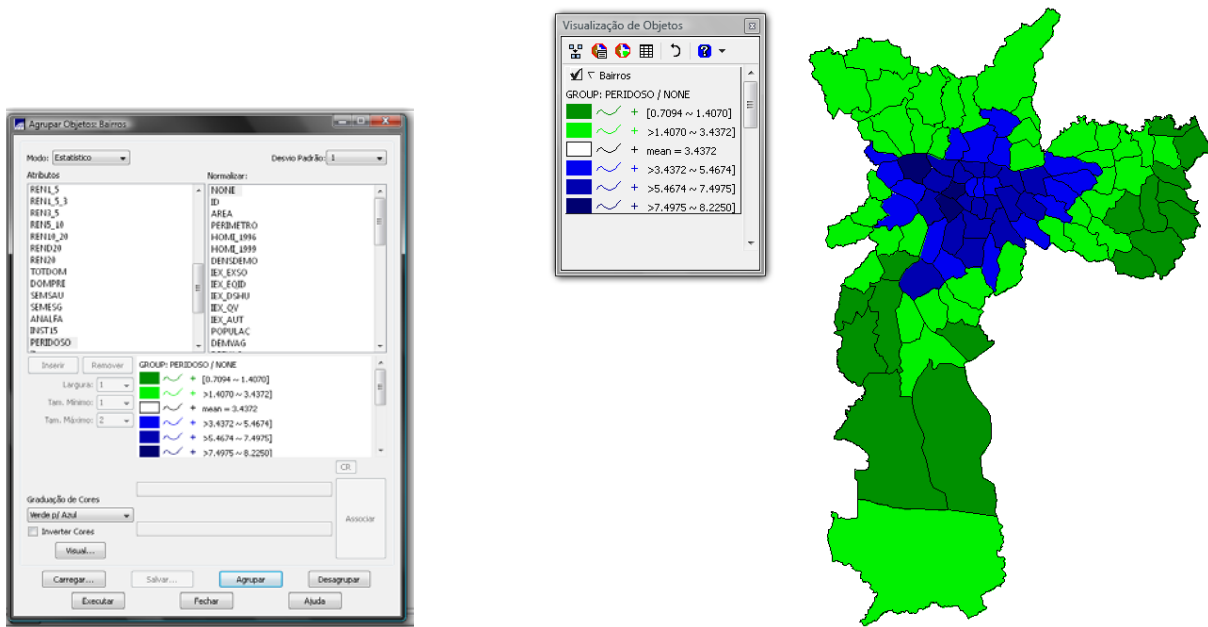


Figura 8 – O mapa cloroplético mostra o atributo Porcentagem de Idosos dividido em seis classes de 1 desvio padrão, pelo método de agrupamento *Estatístico*. A média do atributo corresponde ao valor 3,4372, que não corresponde a porcentagem de idosos de nenhum distrito (classe vazia).

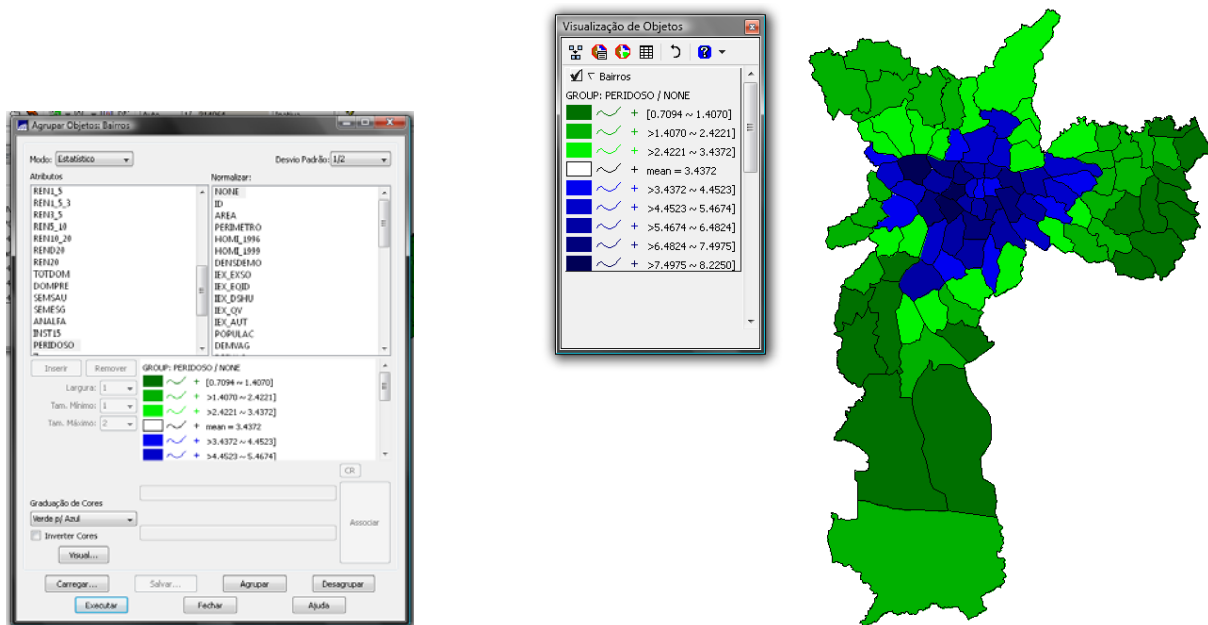


Figura 9 – O mapa cloroplético mostra o atributo Porcentagem de Idosos dividido em nove classes de $\frac{1}{2}$ desvio padrão, pelo método de agrupamento *Estatístico*.

3.1.D - Agrupamento por Valor Único

Esse agrupamento gera grupos que possuem valores distintos. Normalmente é utilizado para separar objetos que possuem características bem definidas como, por exemplo, seco ou úmido, quente ou frio, isento ou não isento, devedor ou não devedor, etc. É uma forma de agrupamento onde o intervalo não é calculado. Ele depende somente

do valor de seu atributo. O número de grupos gerados depende da quantidade de diferentes valores existentes.

Para agrupamentos feitos por *Passo Igual* e *Quantil*, a opção *Número de Partes*, permite agrupá-los em até 15 partes. Para o modo *Estatístico*, a opção *Desvio Padrão*, permite escolher intervalos, entre grupos, de 1, 1/2 ou 1/4 de desvio padrão. Já para agrupamento por *Valor Único* não há opção de divisão em número de partes ou intervalos. O sistema se encarrega de gerar grupos automaticamente em função dos valores a serem analisados.

Por se tratar de um atributo de valores no conjunto dos Reais, foram criadas tantas classes quanto os valores de porcentagens de idosos existentes, neste caso, 96 classes (que correspondem aos 96 Distritos do município, dada a precisão de oito casas decimais) (Figura 10).

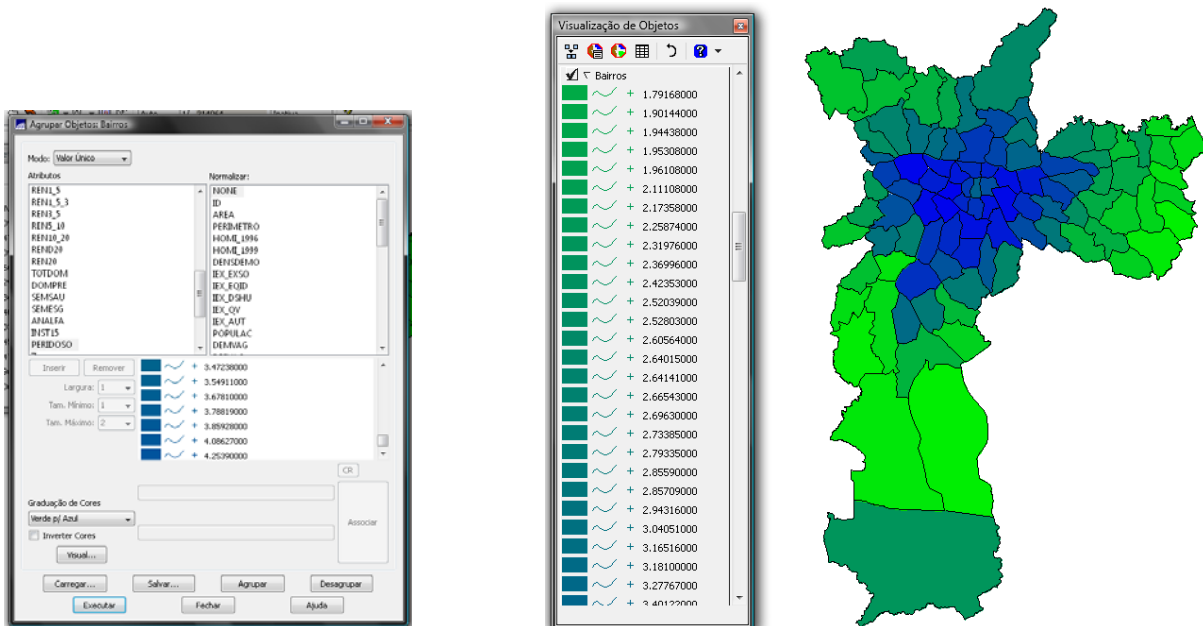


Figura 10 – O mapa cloroplético mostra o atributo Porcentagem de Idosos dividido em 96 classes, pelo método de agrupamento *por Valor Único*.

3.2 Técnicas de Autocorrelação Espacial

As *Técnicas de Autocorrelação Espacial* são técnicas de análise exploratória em dados espaciais (cadastral, contendo objetos-área), que fornece índices de associação espacial e algumas possibilidades de sua visualização, com a intenção de permitir ao usuário identificar agrupamentos espaciais (*clusters*), casos atípicos (*outliers*) e diferentes

regimes espaciais existentes no plano de informação. O conceito central deste módulo é a autocorrelação espacial.

As técnicas de estatística espacial de áreas foram desenvolvidas para tentar identificar regiões onde a distribuição dos valores possa apresentar um padrão específico associado a sua localização geográfica. A informação que se busca é quanto se é parecido com o vizinho próximo e se é diferente do vizinho distante. Neste *Laboratório 3*, foram utilizadas técnicas de estatística espacial que fornecem indicadores globais e locais de autocorrelação. Cada polígono, ou unidade de área, terá um determinado valor de dependência espacial. Por exemplo, o indicador de associação espacial local Moran (I_i) (ANSELIN, 1995) explora o grau de dependência espacial a partir de uma estimativa de segunda ordem, ou seja, um tipo de covariância espacial entre os polígonos; os indicadores G_i e G_i^* (GETIS e ORD, 1992) baseiam-se em estimativas de primeira ordem, como um tipo de estimador de médias móveis.

3.2.A- Índice Global de Moran

O *Índice Global de Moran* fornece uma medida geral da associação espacial existente no conjunto dos dados, variando de $[-1, 1]$. Dados com baixa associação espacial resultam em um índice baixo, próximos a zero. Valores positivos (0 a +1) e negativos (-1 a 0) indicam autocorrelação espacial positiva e negativa, respectivamente.

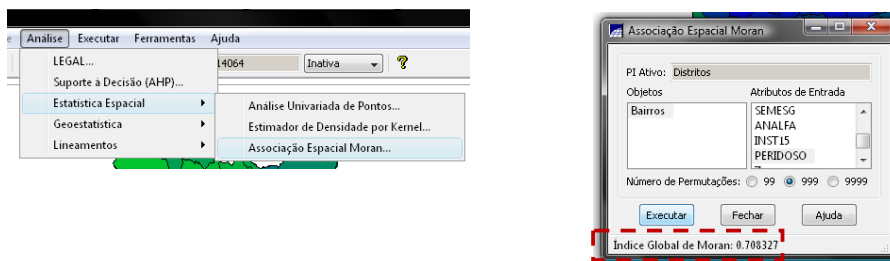


Figura 11 – Análise >> Estatística Espacial >> Associação Espacial Moran.

O valor do *Índice Global de Moran* é reportado no rodapé da interface: 0,7083. Este valor aponta para a existência de autocorrelação espacial positiva.

Ao executar se executar a aplicação do *Índice Global de Moran*, são acrescentadas novas colunas na *Tabela de Atributos* do objeto Bairros, conforme mostra a Figura 12.

Os indicadores globais do *Índice de Moran* estabelecem, para todo conjunto de dados, um único valor como medida da associação espacial. Por outro lado, os

indicadores locais produzem um valor específico para cada objeto, conhecidos por agrupamentos de atributos semelhantes, ou *clusters*; quando anômalos, são denominados *outliers*.

PERIMETRO	NOME	VI...	VE...	ISDI...	EX...	EC...	D...	C...	A...	PU...	VN...	FV...	DP...	IT...	MI...	REI...	N...	IL...	NE...	N5...	LI...	NE...	IN...	TD...	MI...	MS...	VE...	PERIDOSO	Z	WZ	IMORAN	MEDIAMOV	BOXMAP	LISAMAP	MORANMP
32346.0000	Anhanguera	3...	3...	3...	1.43456	-2.00265	-1.71620	0.83390	-1.81168	2.00000	0.00000	0.00000	
11100.0000	Aricanduva	2...	4...	1...	2.42353	-1.01368	-0.77357	0.19026	-0.81359	2.00000	0.00000	0.00000	
11896.0000	Artur_Alvim	4...	6...	1...	1.95308	-1.48413	-0.55459	0.19970	-0.70951	2.00000	0.00000	0.00000	
10860.0000	Barra_Funda	3...	4...	2...	0...	6.65331	3.21610	2.09879	1.63772	2.28501	1.00000	1.00000	1.00000	
7216.0000	Bela_Vista	3...	2...	2...	0...	5.78350	2.34629	3.36480	1.91550	3.19505	1.00000	3.00000	1.00000	
11048.0000	Belem	3...	3...	8...	7.15939	3.72218	1.94727	1.75859	2.16913	1.00000	2.00000	1.00000	
9989.0000	Bom_Retiro	3...	4...	9...	5.43427	2.19796	1.46102	0.88543	1.73759	1.00000	1.00000	1.00000	
8611.0000	Bras	3...	9...	9...	4.36546	0.92825	2.74591	0.61823	2.44222	1.00000	2.00000	1.00000	

Figura 12 – Novas colunas adicionadas à *Tabela de Atributos Bairros*:
 Z = desvio do atributo *PERIDOSO* em relação a *média* ($\mu = 3,4372$);
 WZ = médias locais ou dos vizinhos; $IMORAN$ = Índice Local de Moran;
 $MEDIAMOV$ = média móvel espacial; $BOXMAP$ = Box Map;
 $LISAMAP$ = LISA Map; $MORANMAP$ = Moran Map.

3.2.B- Diagrama de Espalhamento de Moran

Este dispositivo permite visualizar o comportamento dos dados utilizando um gráfico de espalhamento, onde os valores de desvio dos atributos em relação à média (Z), são associados ao eixo X , e o valor da média dos seus vizinhos (WZ), ao eixo Y .

Para execução do *Diagrama de Espalhamento de Moran*, seleciona-se o atributo Z e o WZ (com o botão esquerdo do *mouse*), e seleciona-se a opção Gráfico, conforme ilustrado na Figura 13 a seguir:

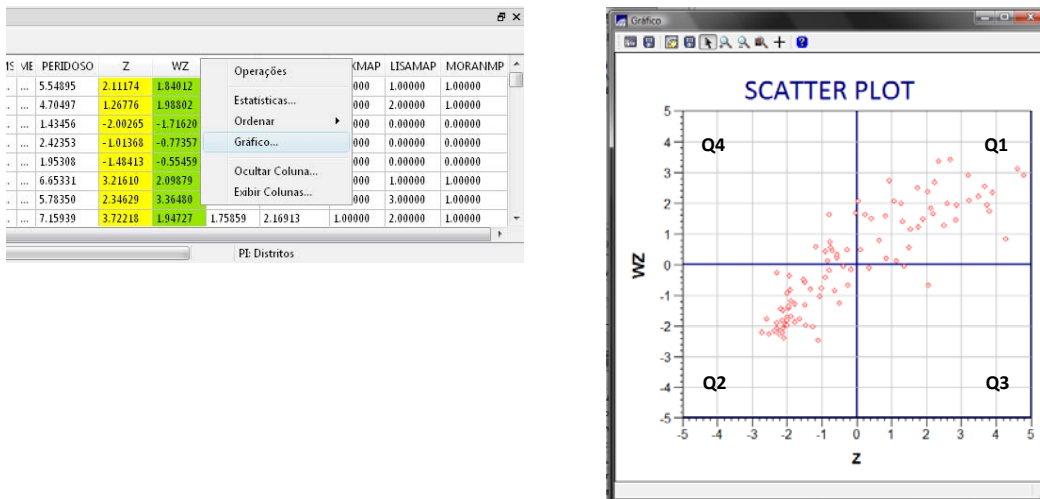


Figura 13 – *Diagrama de Espalhamento de Moran*.

Os quadrantes podem ser interpretados como:

- Q1 (valores positivos e médias positivas) e Q2 (valores negativos e médias negativas):

indicam pontos de *associação espacial positiva*, no sentido que uma localização possui vizinhos com valores semelhantes.

• Q3 (valores positivos e médias negativas) e Q4 (valores negativos e médias positivas): indicam pontos de *associação espacial negativa*, no sentido que uma localização possui vizinhos com valores distintos, indicando pontos de transições entre diferentes padrões espaciais ou pontos de não estacionariedade do atributo.

3.2.C- Box Map

Uma forma alternativa ao *Diagrama de Espalhamento de Moran* é sua apresentação em forma de mapa, no qual cada polígono (Distrito) é apresentado indicando-se seu quadrante no diagrama de espalhamento. No *Box Map*, cada objeto recebe uma cor correspondente ao quadrante a que pertence.

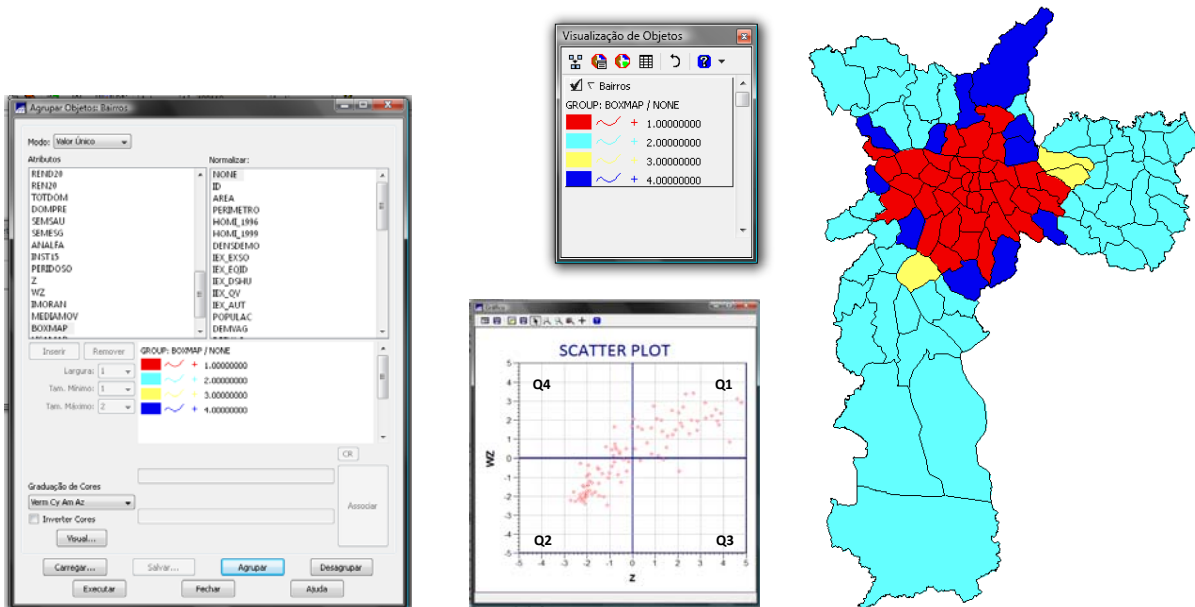


Figura 14 – *Box Map*: valores do *Diagrama de Espalhamento de Moran* de cada Distrito – mapa cloroplético executado por *Agrupamento por Valor Único*.

Os pontos localizados em Q3 e Q4 podem ser vistos como extremos, tanto por estar afastados da reta de regressão linear, como por indicar regiões que não seguem o mesmo processo de dependência espacial das demais observações. Estes pontos marcam regiões de transição entre regimes espaciais distintos.

3.2.D - Gráfico de Barras Z x Wz

Este dispositivo permite a visualização simultânea do valor relacionado ao atributo do objeto e do valor correspondente à sua respectiva vizinhança, com o uso de duas barras gráficas sobre a área correspondente ao objeto no mapa. A altura das barras são proporcionais aos valores do atributo do objeto e à média dos vizinhos. Ambas informações, podem ser obtidas das colunas na *Tabela de Objetos: Z e WZ*.

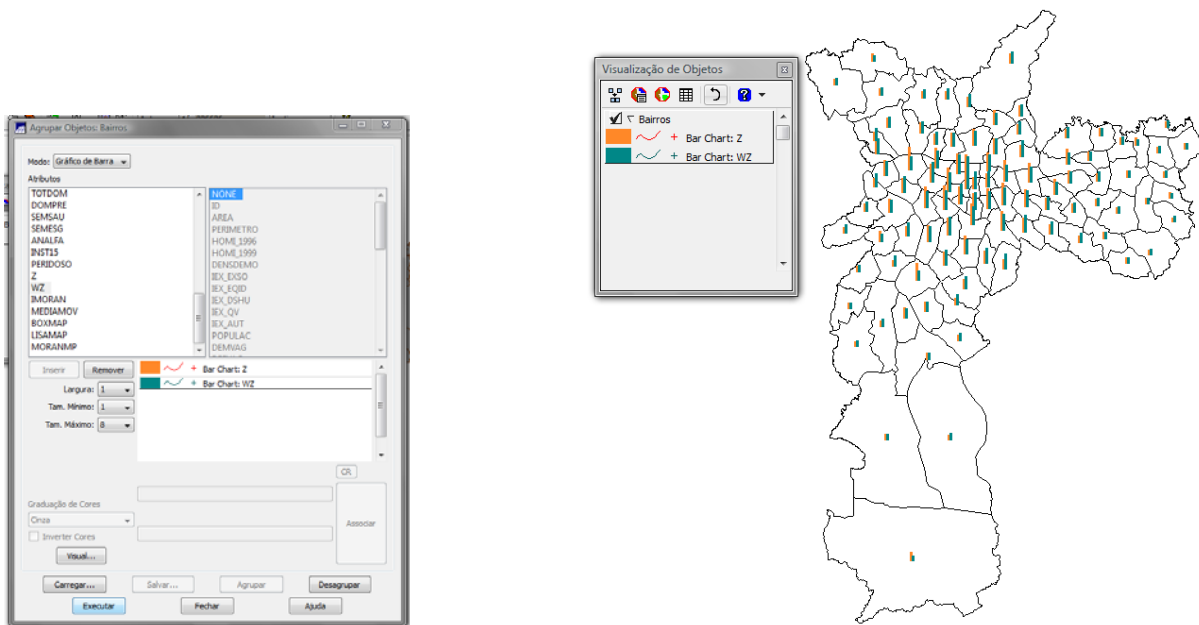


Figura 14 – Gráfico de Barras Z x Wz: os valores dos atributos Z e WZ de cada Distrito são plotados no mapa com os limites de cada unidade de área. O mapa foi executado por meio de *Agrupamento – Gráfico de Barras*.

3.2.E - Índice Local de Associação Espacial (LISA)

Os *indicadores globais de autocorrelação espacial*, como o *Índice de Moran*, fornecem um único valor como medida da associação espacial para todo o conjunto de dados, que é útil como caracterização de toda a região de estudo. Por contraste, muitas vezes é desejável examinar padrões numa escala de maior detalhe, para verificar se a hipótese de estacionariedade do processo verifica-se localmente. Para tanto, é preciso utilizar indicadores de associação espacial que possam ser associados às diferentes localizações de uma variável distribuída espacialmente.

Quando usados em conjunto com indicadores globais, eles refinam o conhecimento sobre os processos que dão origem à dependência espacial, pois permitem encontrar “bolsões” de dependência espacial que não são evidenciados pelos índices globais.

Os *indicadores locais* produzem um valor específico para cada objeto, permitindo assim, a identificação de agrupamentos de objetos com valores de atributos semelhantes (*clusters*), objetos anômalos (*outliers*) e de mais de um regime espacial.

Índice de Moran Local

O *Índice de Moran Local* foi proposto por Luc Anselin (1995) como uma ferramenta estatística para testar a autocorrelação local e para detectar objetos espaciais com influência no indicador *Moran Global*. Esta família estatística trabalha a partir da estimativa de segunda ordem do comportamento dos seus dados, em outras palavras, a partir da análise das covariâncias entre as diferentes unidades de área. Enquanto o *Índice Global de Moran* informa o nível de interdependência espacial entre todos os polígonos em estudo, o *Índice Local de Moran* avalia a covariância entre um determinado polígono e uma dada vizinhança definida em função de uma distância d .

Em análises de associação espacial, assumir uma estabilidade estrutural ou estacionariedade pode ser complicado, especialmente quando um grande número de observações são utilizadas, (ANSELIN, 1995), é possível então a utilização do *Índice Local de Moran* como uma ferramenta estatística que possibilita uma indicação sobre a extensão da significância de um *cluster* de iguais valores.

Anselin define o *Índice Local de Moran* como o produto do resíduo no polígono de referência com a média local dos resíduos dos seus vizinhos adjacentes. Desta forma, valores significativamente altos indicam altas probabilidades de que haja locais de associação espacial tanto de regiões com altos valores como com baixos valores associados.

O *Índice Local de Moran* pode ser escrito como:

$$I_i(d) = \frac{(n-1)}{s^2} \sum_j w_{ij}(d) (x_j - \bar{x}) \quad \text{para } j \neq i \quad (4)$$

onde:

I_i = *Índice Local de Moran*

d = medida de distância estabelecida pelo modelo de vizinhança

$w_{ij}(d)$ = ponderador da matriz de vizinhança W para o par i e j

x_i e x_j = valores encontrados na posição i e suas vizinhas j s

\bar{x} = média amostral global

A matriz W_{ij} é a que define os vizinhos de um certo polígono. Esta matriz pode ser gerada de diferentes maneiras, ou seja, existem várias possibilidades de determinação de quem são ou não vizinhos. Neste *Laboratório*, adotou-se como critério de vizinhança a *adjacência* (ou *contigüidade*), de modo que os valores correspondentes aos polígonos

com borda em comum assumem valor 1 na matriz de vizinhança, e aqueles que não apresentam este tipo de relação recebem valor nulo.

A interpretação mais direta do *Índice Local de Moran* é aquela onde valores significativamente altos e positivos apontam a presença de um *cluster* tanto de valores iguais tanto altos como baixos, já valores significativamente baixos indicam um regime espacial de desigualdade na região, uma espécie de padrão *anti-cluster*, ou seja, zonas de transição entre um determinado regime espacial e outro.

A Figura 15 mostra os valores do *Índice Local de Moran* sob forma de mapa cloroplético.

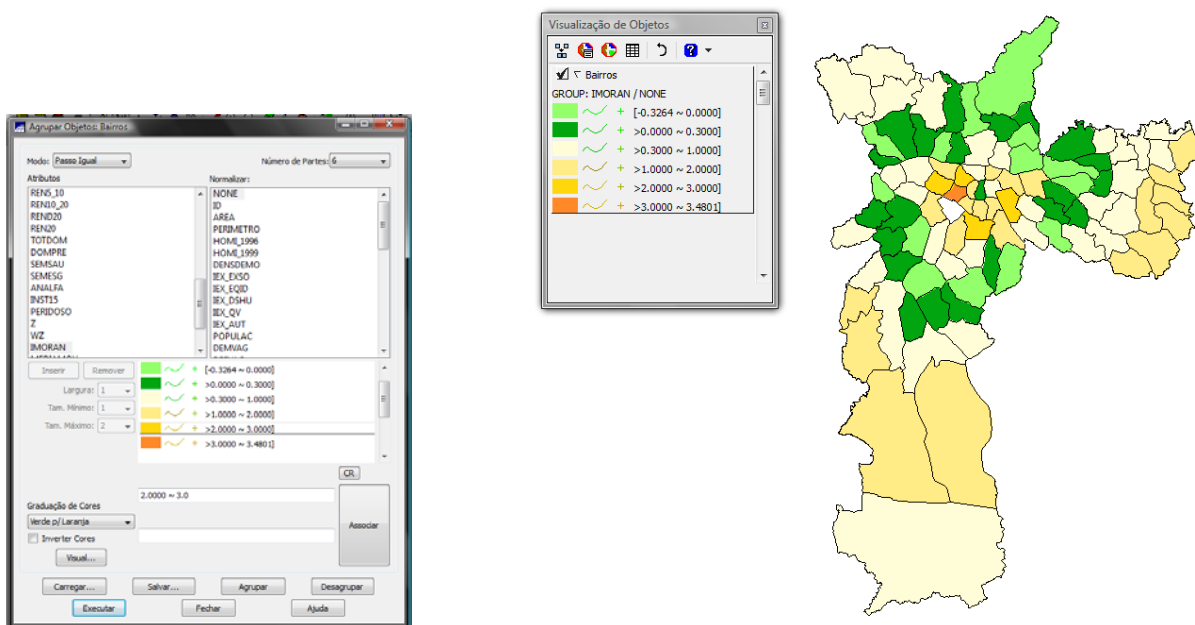


Figura 15 – mapa cloroplético dos valores do *Índice Local de Moran* de cada Distrito.

O padrão apontado pelo indicador *Índice Local de Moran* revela dois regimes espaciais distintos, um na área mais central e nos extremos sul, leste e nordeste (*cluster*); e outro antagônico, formando um anel periférico no município (*anti-cluster*).

O *Índice Local de Moran* aponta a presença de uma espécie de anel de polígonos com baixos valores justamente nas áreas de transição entre o regime espacial da região central com o regime espacial das periferias. Este tipo de informação é só é possível no *Índice Local de Moran*, sendo este o motivo de se utilizar tal índice em conjunto com as estatísticas G_i , pois possibilita observar o comportamento espacial das variáveis de maneira mais global.

3.2.F - LISA Map

Na geração do *LISA Map*, a avaliação da significância é feita comparando-se os valores de LISA encontrado com uma série de valores obtidos por meio de permutações dos valores dos atributos dos vizinhos (número de permutações definido: 999) sob a hipótese nula (não existência de autocorrelação espacial). Uma vez determinada a significância estatística do *Índice Local de Moran*, faz-se útil gerar um mapa indicando as regiões que apresentam correlação local significativamente diferente do resto do dados. Estas regiões podem ser vistas como “bolsões” de não-estacionariedade, pois tratam-se de áreas com dinâmica espacial própria e que merecem análise detalhada. Na geração do *LISA Map*, os valores do *Índice Local de Moran* são classificados em três grupos (ANSELIN, 1995):

- não significantes
- com significância de 5% (1,96 σ)
- com significância de 1% (2,54 σ)
- com significância de 0,1% (3,20 σ)

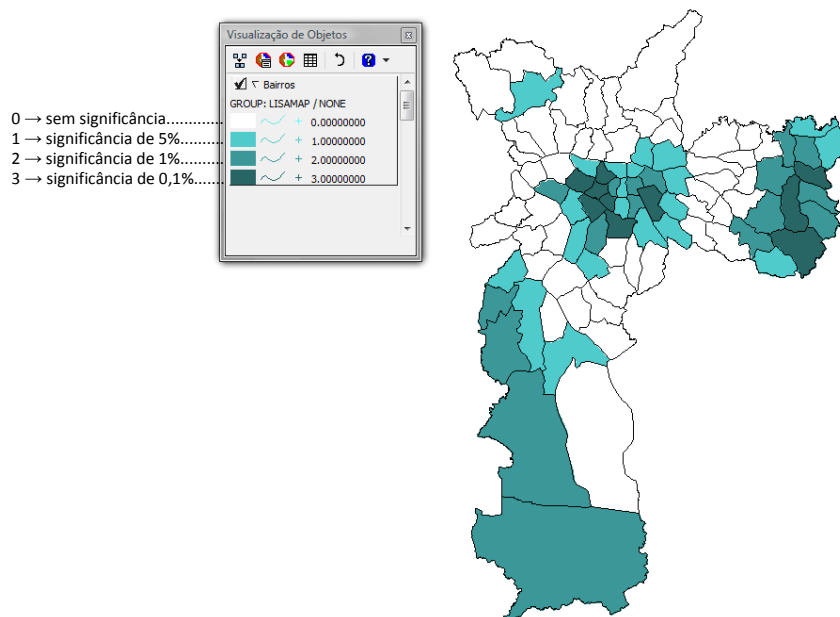


Figura 16 – *LISA Map*: indicação clara de uma forte polarização centro-periferia indicando a presença de “bolsões”.

3.2.G - Moran Map

O *Índice Local de Moran* é equivalente ao coeficiente de regressão linear que indica a inclinação da reta de regressão de W_z em Z . O *Diagrama de Espalhamento de Moran* pode ser apresentado na forma de um mapa temático (cloroplético) bidimensional,

em que cada polígono é apresentado com uma cor que corresponde à sua localização em um dos quadrantes do *Diagrama de Espalhamento*.

O *Índice de Moran* é uma medida global da autocorrelação espacial, pois indica o grau de associação espacial presente no conjunto de dados. Entretanto, uma vez calculado, é importante estabelecer sua validade estatística.

No *Moran Map*, de modo semelhante ao *LISA Map*, somente os objetos para os quais os valores de LISA foram considerados significantes ($\alpha \leq 0,05$) são apresentados. Os objetos são então classificados em quatro grupos, conforme o quadrante aos quais pertencem no *Diagrama de Espalhamento de Moran*. Os demais objetos foram classificados como "sem significância".

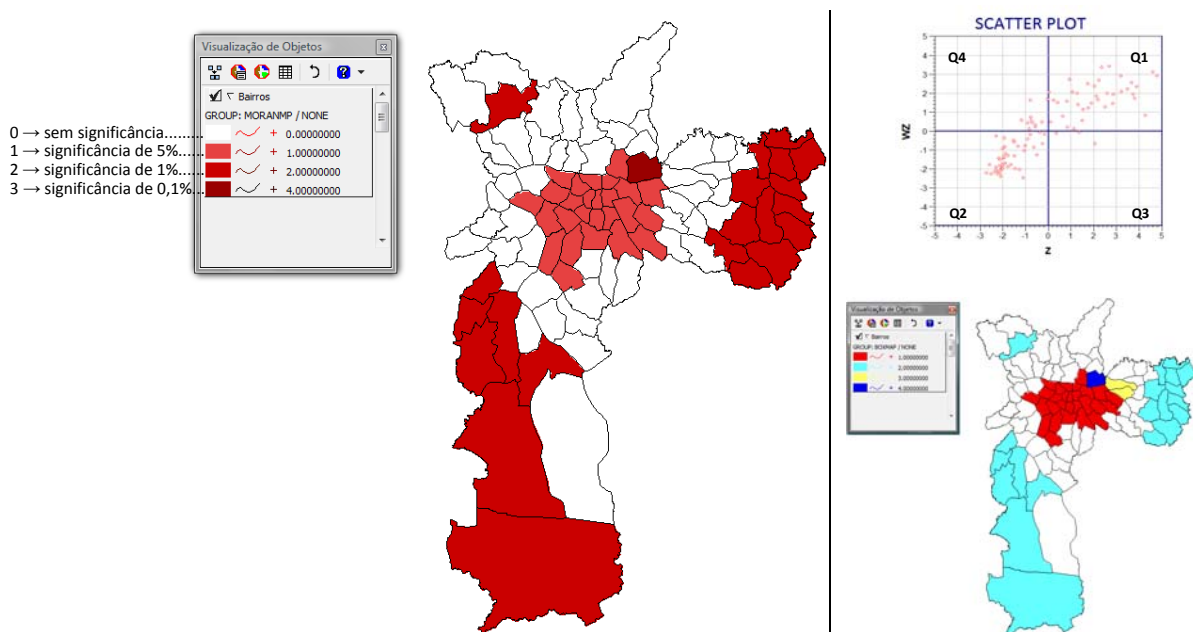


Figura 17 – *Moran Map*: indicação da significância dos valores de Moran para cada Distrito. Ao lado, *Diagrama de Espalhamento de Moran* com todos os pontos plotados (com e sem significância), e abaixo, *Box Map* somente com os valores cuja significância está abaixo de 5% ($\alpha \leq 0,05$).

3.2.H - Média Espacial Móvel

O método de *Média Espacial Móvel* é uma técnica que explora o valor médio μ_i do atributo na região de estudo (primeira ordem). Trata-se de uma técnica simples de exploração da variação da tendência espacial dos dados, por meio do cálculo da média dos valores vizinhos, o que reduz a variabilidade espacial, uma vez que a operação tende a produzir uma superfície com menor flutuação que os dados originais. A média móvel μ_i associada ao atributo z_i , relativo a i 'ésima área, pode ser calculada a partir de elementos

w_{ij} da matriz normalizada de proximidade espacial, W , tomando-se simplesmente a média dos vizinhos.

$$\bar{x}_i = \sum_{j \in N_i} w_{ij} x_j \quad (6)$$

A matriz de vizinhança comumente utilizada para geodados zonais define vizinhança a partir da propriedade topológica de contigüidade, assim, W é uma matriz binária (0, 1), onde 1 está associado as zonas com fronteira em comum, e 0, àquelas sem esta propriedade.

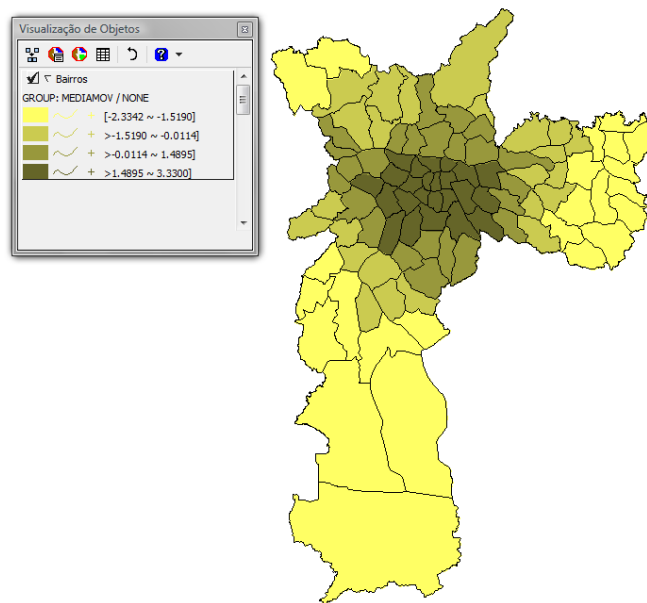


Figura 18 – Mapa cloroplético produzido por *Agrupamento por Quantil* do atributo *Média Espacial Móvel*.

3.2.1 - Estatísticas G_i e G_i^*

A estatística G_i e G_i^* é uma ferramenta estatística apresentada por Getis e Ord (1992) como um indicador de associação espacial local. Apesar da estatística G se aproximar da estatística *Moran* quanto ao objetivo, as informações que ambas apontam são conceitualmente diferentes: enquanto o *Índice Local de Moran* é baseado na análise das covariâncias entre áreas, o *Índice G* é, na realidade, uma somatória de valores vizinhos definidos a partir de uma matriz de vizinhança W_{ij} (GETIS e ORD, 1992).

Desta forma, G_i e G_i^* podem ser escritos como:

$$G_i = \frac{\sum_{j \neq i}^n w_{ij}(d)x_j}{\sum_{j \neq i}^n x_j} \quad \text{para } i \neq j \quad (7)$$

$$G_i^* = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j} \quad \text{para } i \neq j \quad (8)$$

onde:

W_{ij} corresponde à matriz binária e simétrica que define a vizinhança entre as áreas: é um valor na matriz de proximidade para região i com a região j em função da distância (d);

x_i e x_j são os valores dos atributos considerados nas áreas i e j ;

d é distância entre pontos;

n o número de áreas (polígonos).

A única diferença existente entre o indicador G_i e G_i^* é o fato do valor do polígono referência ser incluído no cálculo do G_i^* e não no cálculo do G_i . Neste *Laboratório*, a matriz de vizinhança utilizada para o cálculo dos índices G_i e G_i^* foi a mesma utilizada para a definição do *Índice Local de Moran*: apenas áreas com bordas comuns foram consideradas vizinhas.

Diferentemente do *Índice Local de Moran*, a estatística G_i e G_i^* apresenta uma interpretação mais direta sobre como os dados estão distribuídos no espaço. Uma vez que estes indicadores são compostos por uma somatória de valores de atributos e não desvios em relação à média como no *Moran*, a observação de valores significativamente altos de G_i e G_i^* apontam a existência de altos índices de ocorrência deste atributo, sendo o oposto, um indício de agrupamento de valores baixos.

Uma das comparações mais evidentes entre a estatística G e a *Moran* é a de que, apesar de medirem coisas fundamentalmente diferentes, ambas supõem normalidade em suas distribuições para construir intervalos de confiança para a hipótese nula de total ausência de autocorrelação local.

Em circunstâncias típicas, a hipótese nula é a de que um certo quadro de valores x dentro de uma distância d é uma amostra aleatória independente de sua posição geográfica, então assumindo-se uma distribuição normal dos resultados dos G_i podemos construir intervalos de confiança para um determinado valor Z_i , onde:

$$Z_i = \{ G_i(d) - E [G_i(d)] \} / (\text{var } G_i(d))^{1/2} \quad (9)$$

A estatística $G(d)$ mede a concentração total ou falta de concentração de todos os pares de (x_i, x_j) desde que j seja vizinho de i . O *Índice Moran*, por outro lado mede a correlação entre cada x_i com seus vizinhos x_j . Desta forma, e como poderá ser observado nos resultados obtidos com estes indicadores no caso de São Paulo, é interessante

adotar ambos indicadores na identificação de autocorrelação espacial uma vez que estes trazem informações complementares.

A partir deste ponto, as estatísticas ou medidas de associação espacial G_i e G_i^* foram calculadas utilizando o software Space Stat, e a visualização e análise dos resultados em forma de mapas foi realizada no sistema SPRING.

Para utilização do software, os dados relativos à Porcentagem de Idosos foi exportada nos formatos Space Stat: a *Tabela de Atributos* (tabela.txt) e o outro a *Matriz de Proximidade* (matriz.gal). A Figura 19 mostra este procedimento:

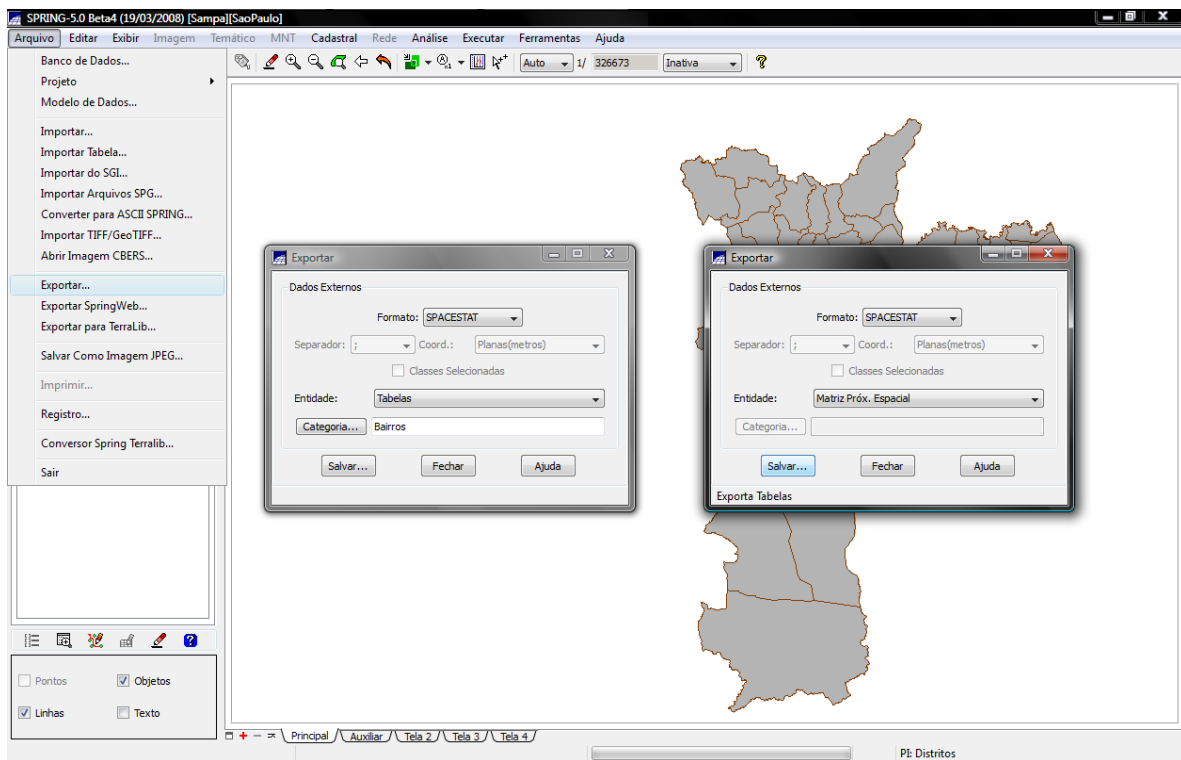


Figura 19 – Exportação da *Tabela de Atributos* e da *Matriz de Proximidade* para o formato SpaceStat.

Iniciando os procedimentos no Space Stat, a tela inicial é apresentada (Figura 20). É definido o diretório de trabalho (e:\SaoPaulo\Temp).



Figura 20 – Tela inicial do SpaceStat.

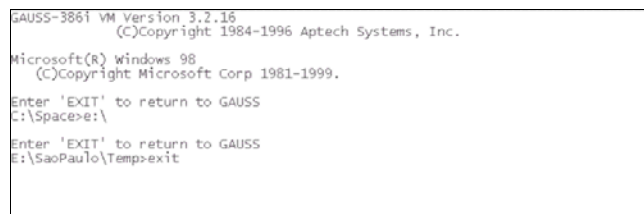


Figura 21 – Definição do diretório de trabalho.

```

SPACESTAT  OPTIONS
-----
ESC Return to menu          Press option key or move cursor and press Return
-----
1 Interactive Operation:    YES
2 Output to a File:        NO
3 Long Output:             NO
4 Indicator Variable:      YES          Variable: GEOID
5 Report Format:            COMMA DELIMITED
6 Convergence Criterion:   0.0000010000
7 Number of Permutations:  99
8 Maximum Iterations:      100
9 Random Number Seed:      397204094

```

Figura 22 – Configuração dos dados.

A *Tabela de Atributos* (tabela.txt) foi convertida em dois novos arquivos: bairros.dat e bairros.dht. Estes são formatos intrínsecos e necessários para o Space Stat trabalhar. O “Report” mostrado na Figura 24 aparece sobre a *Tabela de Atributos* após esta operação.

```

creating a data set from an Ascii file
Enter the name of the Ascii file
File name is: tabela.txt
Enter the data set filename (do not include .DAT or .DHT)
File name is: bairros_

```

Figura 23 – Criação dos novos arquivos: bairros.dat e bairros.dht.

```

The GAUSS dataset bairros has been created
It contains 96 observations on 44 variables
The variables contained in the data set are:
  GEOID DENSDEMO IEX_EXSO IEX_EQID IEX_DSHU IEX_QV IEX_AUT POPULAC
  DEMVAG DEJVAG POP70 TOTCH CHMULH S_REND REN1_5 REN1_5_3
  REN3_5 REN5_10 REN10_20 REND20 REN20 TOTDOM DMPRE SEMSAU
  SEMESG ANALFA INST15 PERIDOSO PERSREND PERREN20 PERSESG PERDPRE
  PERINS15 Z WZ IMORAN MEDIAMOV BOXMAP LISAMAP MORANMP
  G_PERIDO Z_PERIDO P_PERIDO S_PERIDO

```

Figura 24 – Report sobre a *Tabela de Atributos*.

Para calcular a estatística G_i digitou-se E 5 3 2 e ENTER¹; então, é visualizada a tela mostrada na Figura 25. A seguir, digitou-se bairros e ENTER (Figura 25). Definiu-se o arquivo da *Matriz de Proximidade* (matriz.gal) (Figura 26).

Em seguida, definiu-se a variável para qual a estatística G_i seria calculada: peridoso e ENTER (Figura 27).

¹ Para calcular a estatística G_i^* , digitou-se E 5 4 2 e ENTER, e então, todos os passos subsequentes foram repetidos para esta estatística. O mapa de significância G_i^* (S_PERIDO) é mostrado na Figura 32.

```

INTERACTIVE PROBLEM FILE CREATION

Problem no.  1

Data set for problem

Enter the data set filename (do not include .DAT or .DHT),
or press Return for directory listing: bairros

```

Figura 25 – Calculando a estatística G_i .

```

Spatial weights for problem

Enter the filenames for the spatial weights from the following list
or press Return to stop

.FMT files in default directory:

.GAL files in default directory:
  BAIRROS  CÓPIAD~1

.GWT files in default directory:

weights file: matriz
weights file:

```

Figura 26 – Definição do arquivo de ponderação matriz.gal.

```

Variables for problem

Choose the variable(s) from the following list
(each variable name should correspond exactly to one in the list):
  GEOID  DENSDEMO  IEX_EXSO  IEX_EQID  IEX_DSHU  IEX_QV  IEX_AUT
  POPULAC  DEMVAG  DEFVAG  POP70  TOTCH  CHMULH  S_REND
  REN1_5  REN1_5_3  REN3_5  REN5_10  REN10_20  REND20  REN20
  TOTDOM  DOMPRES  SEMSAU  SEMESG  ANALFA  INST15  PERIDOSO
  PERSREND  PERREN20  PERSESG  PERDPRE  PERINS15  Z  WZ
  IMORAN  MEDIAMOV  BOXMAP  LISAMAP  MORANMP  G_PERIDO  Z_PERIDO
  P_PERIDO  S_PERIDO

Enter the variable name, or press Return to stop
Variable name: peridoso
Variable name: _

```

Figura 27 – Definição da variável peridoso.

G-I STATISTICS FOR SPATIAL ASSOCIATION							
DATA SET: BAIRROS		WEIGHTS: MATRIZ		VARIABLE: PERIDOSO			
LARGEST (MOST POSITIVE Z) VALUES				SMALLEST (MOST NEGATIVE Z) VALUES			
GEOID	G-I	Z	PROB	GEOID	G-I	Z	PROB
40	0.127017	4.2911	0.0000	69	0.0226153	-2.7554	0.0059
38	0.122179	4.0163	0.0001	86	0.0410726	-2.5582	0.0105
96	0.104908	3.8224	0.0001	64	0.0194348	-2.4630	0.0138
42	0.148607	3.8099	0.0001	84	0.028214	-2.3784	0.0174
37	0.0986282	3.4158	0.0006	4	0.0208781	-2.3545	0.0185
36	0.0983584	3.3550	0.0008	62	0.0286379	-2.3395	0.0193
75	0.094934	3.0881	0.0020	90	0.021771	-2.2701	0.0232
95	0.0945506	3.0640	0.0022	66	0.02277	-2.2092	0.0272
58	0.107834	3.0225	0.0025	9	0.0487229	-2.1137	0.0345
59	0.105327	2.8452	0.0044	12	0.0320244	-2.1119	0.0347

Figura 28 – Relatório final apresentado.

A tela de relatório da Figura 28 exibe, do lado esquerdo, apenas os dez valores Z

mais *positivos*, com seus respectivos valores de *GEOID*, G_i e níveis de significância (*PROB*). Do lado direito, os dez valores Z mais *negativos*, também com seus respectivos valores de *GEOID*, G_i e níveis de significância (*PROB*).

O valor Z é definido como:

$$Z_i = (G_i - E[G_i]) / \sigma(G_i) \quad \text{para } i=1, \dots, n \quad (9)$$

onde:

n = número de polígonos

Um valor Z positivo e significativo para a estatística G_i indica agrupamento espacial de alto valores; por outro lado, um valor Z negativo e significativo indica agrupamento espacial de baixo valores.

Neste caso, do lado esquerdo, os três maiores valores Z são 4.2911 para vizinhança 40 (*GEOID*), 4.0163 para vizinhança 38 (*GEOID*) e 3.8224 para vizinhança 96 (*GEOID*). Todos são altamente significantes, o que indica um forte agrupamento de alto valores de Peridoso (a variável em estudo) concentrado nestas localizações.

Além da tela de relatório, o Space Stat gera também um arquivo ASCII no diretório determinado inicialmente (e:\SaoPaulo\Temp). O nome deste arquivo está relacionado com o tipo de operação executada, (a estatística G_i): G_i _matri.txt. Este arquivo contém o valor da variável indicadora (*GEOID*), e para cada variável, seu valor original (*PERIDOSO*), o valor da estatística G_i ou G_i^* (G_PERIDO), os valores Z (Z_PERIDO), suas probabilidades (P_PERIDO), e um indicador de significância (S_PERIDO) – de modo que é atribuído o valor 3 se $p < 0.001$; 2 se $p < 0.01$ e 1 se $p < 0.05$ (5%).

Após a criação do arquivo .txt com os dados da estatística G_i , importou-se este arquivo para o SPRING, como uma *Tabela de Atributos* do objeto Bairros (Figura 29).

O resultado desta operação pode ser visto na tabela de atributos, conforme ilustra a Figura 30.

Um resultado interessante de ser apresentado e analisado refere-se ao mapa de significância do indicador de associação espacial G_i . Ele é ilustrado pela Figura 31.

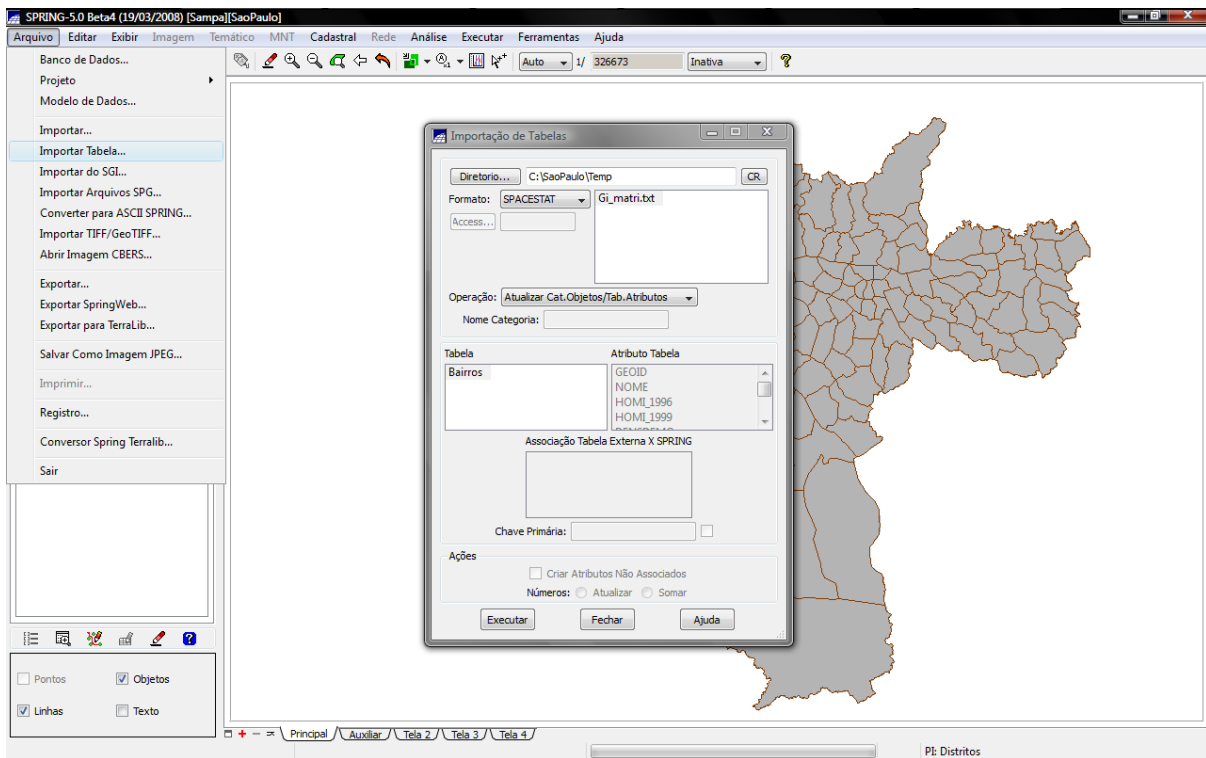


Figura 29 – Importação de Gi_matri.txt como uma Tabela de Atributos dos objeto Bairros.

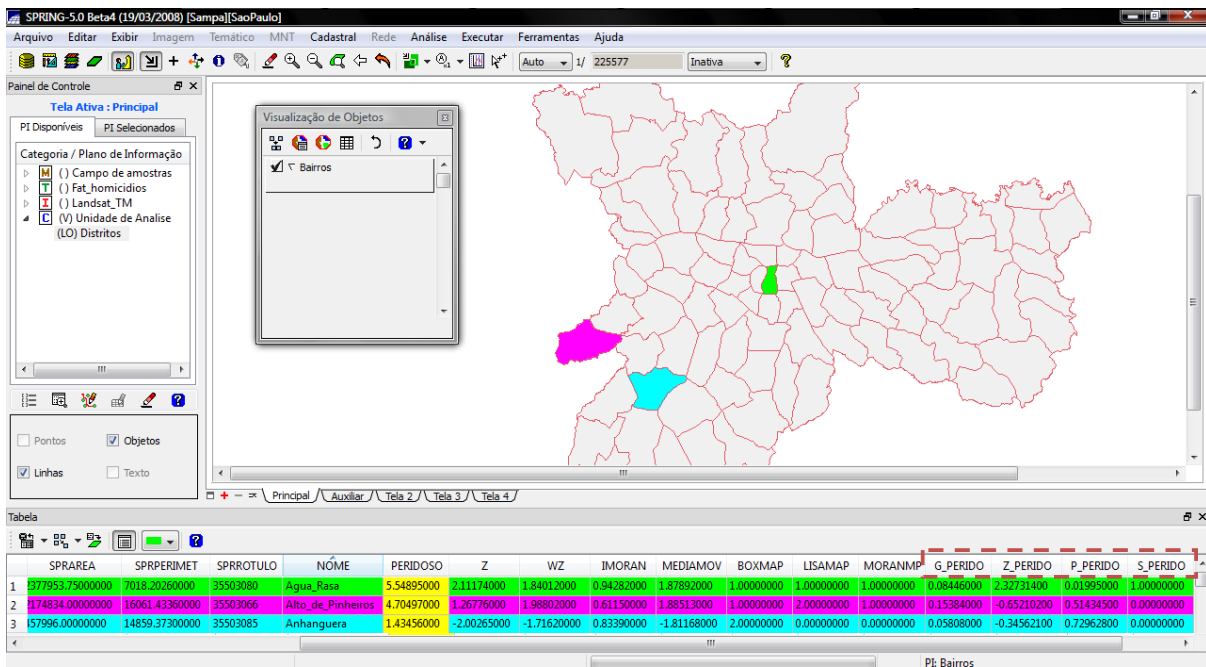


Figura 30 – Tabela de Atributos dos objeto Bairros com as novas colunas referentes a estatística G_i .
 G_PERIDO : refere-se a estatística G_i ; Z_PERIDO : é o valor Z , igual a $(G_i - E[G_i]) / \sigma G_i$;
 P_PERIDO : a probabilidade da estatística G_i ; S_PERIDO : é um indicador de significância, sendo: 3 se $p < 0.001$; 2 se $p < 0.01$; 1 se $p < 0.05$ e 0 sem significância.

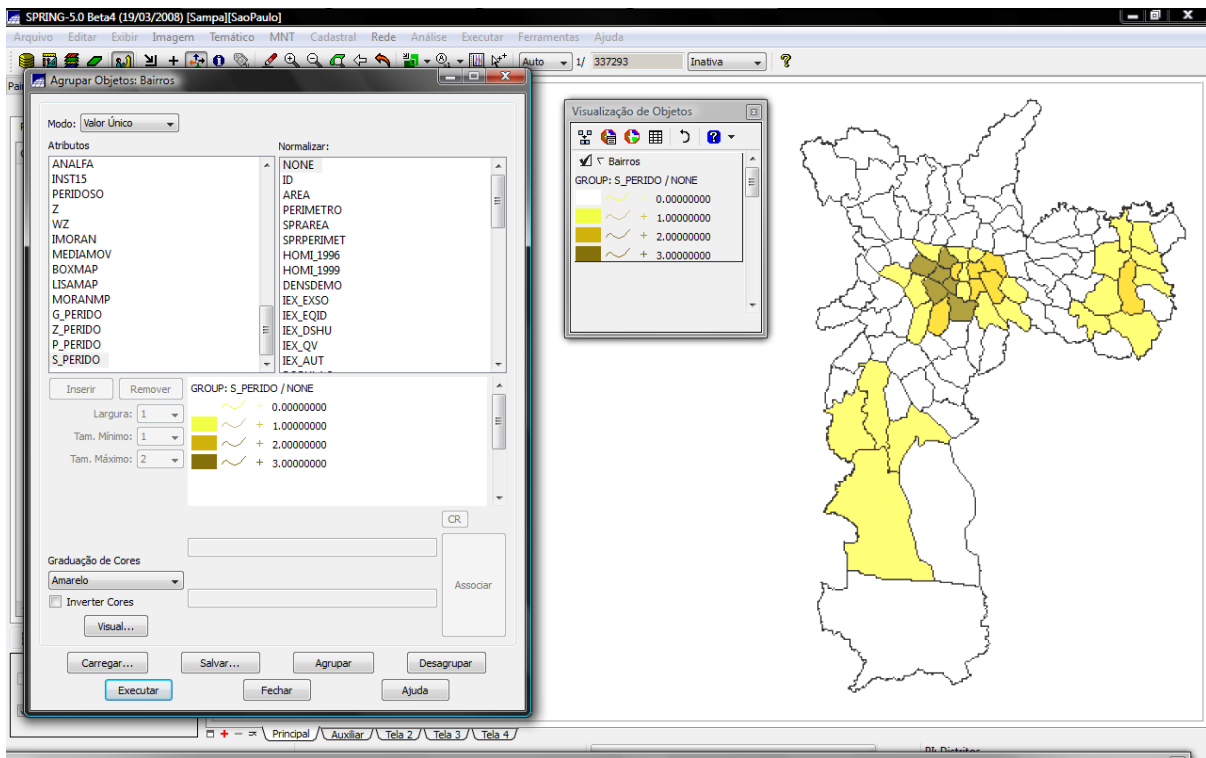


Figura 31 – Mapa cloroplético produzido por *Agrupamento por Valor Único* do atributo S_PERIDO, que indica a significância do valor Z calculado pela estatística G_i .

O mesmo procedimento foi realizado para a estatística G_i^* (ver página 20). O mapa de significância do indicador de associação espacial G_i^* mostrado na Figura 32.

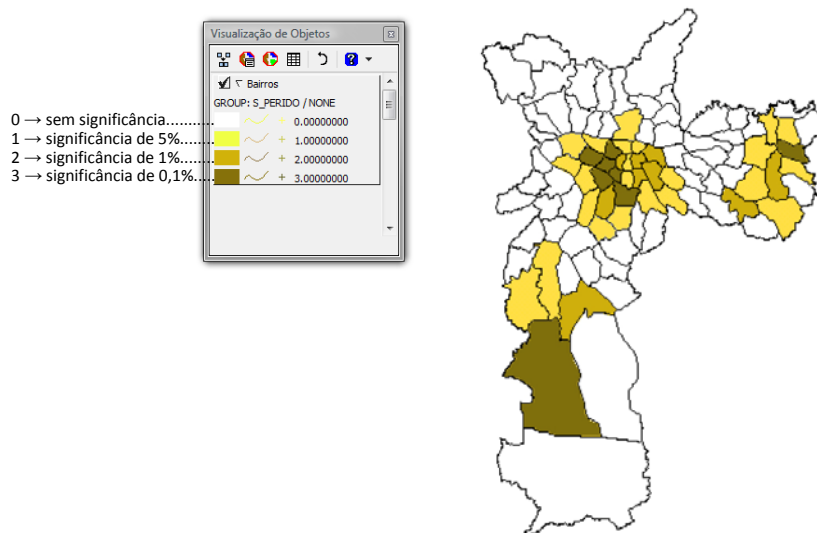


Figura 31 – Mapa cloroplético produzido por *Agrupamento por Valor Único* do atributo S_PERIDO, que indica a significância do valor Z calculado pela estatística G_i^* .

4 CONCLUSÃO

Por meio deste *Laboratório 3*, foi possível ter contato e compreender o ferramental disponível para análise espacial em unidades de áreas nos Sistemas de Informação Geográfica. Entende-se que estes tipos de análises espaciais são processos complexos, uma vez que toda uma manipulação de indicadores de pesquisas, verificação do suporte gráfico (mapa coroplético) e pesquisa de ferramentas estatísticas constituem peças fundamentais para a coerência das análises.

Referências Bibliográficas

- ANSELIN, L., *Local Indicators of Spatial Association – LISA*. **Geographical Analysis**, v. 27, n. 2, April 1995.
- GETIS, A.; ORD, J.K., *The Analysis of Spatial Association by Use of Distance Statistics*. **Geographical Analysis**, v. 24, n. 3, July 1992.
- SPOSATI, A., **Mapa da Exclusão / Inclusão da Cidade de São Paulo**. São Paulo, Editora da PUC-SP, 1996.