

**UNIVERSIDADE FEDERAL DA PARAÍBA - UFPB**

**JEFFERSON DA ROCHA AUGUSTO**

**ANÁLISE ESTATÍSTICA E ESPACIAL DE HOMICÍDIOS OCORRIDOS  
ENTRE 2007 E 2011 NO MUNICÍPIO DE JOÃO PESSOA – PB**

JOÃO PESSOA-PB  
2011.2

UNIVERSIDADE FEDERAL DA PARAÍBA – UFPB  
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA – CCEN  
DEPARTAMENTO DE ESTATÍSTICA - DE

**JEFFERSON DA ROCHA AUGUSTO**

**ANÁLISE ESTATÍSTICA E ESPACIAL DE HOMICÍDIOS OCORRIDOS  
ENTRE 2007 E 2011 NO MUNICÍPIO DE JOÃO PESSOA – PB**

Trabalho de conclusão de curso apresentado a banca examinadora de avaliação do curso de Bacharelado em Estatística da Universidade Federal da Paraíba, em cumprimento às exigências para obtenção do grau de Bacharel.

Orientador:  
Prof. Dr. Ronei Marcos de Moraes  
Depto. Estatística – UFPB

**JEFFERSON DA ROCHA AUGUSTO**  
Bacharel em Estatística

**ANÁLISE ESTATÍSTICA E ESPACIAL DE HOMICÍDIOS OCORRIDOS  
ENTRE 2007 E 2011 NO MUNICÍPIO DE JOÃO PESSOA – PB**

Trabalho de conclusão de curso apresentado a banca examinadora de avaliação do curso de Bacharelado em Estatística da Universidade Federal da Paraíba, em cumprimento às exigências para obtenção do grau de Bacharel.

Aprovado em 19/12/2011, pela banca examinadora:

---

Orientador: Prof. Dr. Ronei Marcos de Moraes

---

Prof. Dr. Joab de Oliveira Lima

---

Prof. Dr. Hemílio Fernandes Campos Coêlho

---

Profª. Ma. Izabel Cristina Alcantara de Souza

## AGRADECIMENTOS

A todos os professores do departamento de estatística da Universidade Federal da Paraíba que lutam pelo crescimento e qualidade do curso de bacharelado em especial aos professores:

Ronei Marcos de Moraes, pelo comprometimento com as disciplinas lecionadas e pela orientação prestada para a realização desse trabalho;

Joab de Oliveira, Izabel Cristina e Hemílio Fernandes por ter aceitado participar da banca examinadora.

Ao Cel Wolgrand Pinto (Chefe do estado maior estratégico), por ter me concedido a oportunidade de estágio na coordenadoria de estatística e avaliação da Polícia Militar da Paraíba.

A Ten.-cel Valtânia Ferreira (coordenadora do EM-7 PMPB), por ter me orientado no setor, pelos conselhos dados e por ser uma das pessoas mais atenciosa e determinada que já conheci.

Aos colegas de curso Natália, Thelmo, Abner e Thadeu que estiveram comigo em boa parte das disciplinas.

Aos colegas do LEAPIG, pela ajuda em alguns trabalhos, pela atenção e principalmente pela amizade.

A minha família, por sempre acreditar que eu chegaria até aqui.

A Deus, por me fazer acreditar no sentido da vida.

## RESUMO

O homicídio é a ocorrência de maior impacto na sociedade por ser o maior dano que uma vítima pode ter. Visto que em João Pessoa, capital do estado da Paraíba, o número de ocorrências desse caso vem se elevando consideravelmente nos últimos anos, pois a cidade apresenta o maior índice de homicídios do estado (50% dos homicídios nos últimos 10 anos), segundo dados do CIOP (Centro Integrado de Operações Policiais) e a segunda maior taxa para as capitais brasileiras em 2010 (Anuário Brasileiro de Segurança Pública), foi feito um estudo com o objetivo de acompanhar o número de ocorrências durante 2007 e 2011 e entender melhor algumas características dessa ocorrência. Além de serem usados para análise dos dados de ocorrências distribuições de frequência, testes não paramétricos e análise de séries temporais, foram utilizadas algumas técnicas de geoprocessamento para a identificação de lugares críticos e possibilitar uma melhor visualização nos dados. Para a identificação de conglomerados, o método Scan foi o que melhor identificou conglomerados de valores altos na região noroeste e sudeste do município. Também foi observado que 93,9% das vítimas de homicídios são do sexo masculino, 94,1% são cometidos por arma de fogo e 70,8% ocorrem no turno da madrugada. Foram identificadas regiões de alta incidência de ocorrências para dados pontuais do ano 2011, principalmente na zona norte e foi estimado um aumento preocupante no número de ocorrências em 2012, 15,5% em relação a 2011.

Palavras-Chave:

Aglomeración espacial, Análise estatística, homicídios.

## Sumário

LISTA DE FIGURAS .....	7
LISTA DE TABELAS .....	8
LISTA DE GRÁFICOS .....	9
1 INTRODUÇÃO .....	10
2 OBJETIVO.....	11
2.1 Geral.....	11
2.2 Específicos .....	11
3 METODOLOGIA .....	12
3.1 Projeção Populacional.....	12
3.2 Risco Relativo .....	12
3.3 Análise de Dados de Área .....	13
3.4 Matriz de Proximidade Espacial .....	14
3.5 Índice Global de Autocorrelação de Moran ( <i>I</i> ) .....	15
3.6 Índices Locais de Associação Espacial .....	17
3.7 Índice Global de Autocorrelação de Getis e Ord .....	17
3.8 Método de Varredura Scan.....	19
3.9 Método Besag e Newell .....	20
3.10 Processos Pontuais .....	20
3.11 Função intensidade de ocorrências.....	23
3.12 Estimador de Intensidade de Kernel.....	24
3.13 Método dos vizinhos próximos (propriedade de segunda ordem).....	26
3.13.1 Função <i>G</i> .....	26
3.13.2 Função <i>F</i> .....	27
3.13.3 Função <i>J</i> .....	28
3.13.4 Função <i>K</i> .....	28
3.13.5 Função <i>L</i> .....	29
3.14 Teste Qui-Quadrado de Aderência.....	29
3.15 SÉRIES TEMPORAIS.....	30
3.15.1 Suavização Exponencial de Holt (SEH).....	30
3.16 Área de Estudo .....	31

3.16.1	Mapa de Divisão Política .....	31
3.17	Softwares Utilizados .....	32
3.17.1	R.....	32
3.17.2	Spring .....	33
3.18	Banco de Dados.....	33
4	RESULTADOS.....	34
4.1	Homicídios ocorridos entre 2007 e 2010 .....	34
4.1.1	Análise de Autocorrelação Espacial para Dados de Área .....	34
4.2	Homicídios ocorridos entre agosto e novembro de 2011 .....	39
4.2.1	Análise Descritiva .....	39
4.2.2	Análise Pontual .....	45
4.3	Previsão para 2012 por Alisamento Exponencial de Holt.....	49
5	CONCLUSÃO .....	51
6	REFERÊNCIAS.....	52
	APÊNDICES.....	56
	APÊNDICE I – NÚMERO DE OCORRENCIA DE HOMICÍDIOS POR BAIRRO .....	i
	APÊNDICE II – INCIDÊNCIA DE HOMICÍDIOS POR BAIRRO .....	ii
	APÊNDICE III - PARTE 1 – DIAGRAMA DE EMPARALHAMENTO DE MORAN.....	iii
	APÊNDICE III - PARTE 2 – DIAGRAMA DE EMPARALHAMENTO DE MORAN.....	iv
	APÊNDICE IV – LISA MAPA .....	v
	APÊNDICE V – Gi .....	vi
	APÊNDICE VI – Gi* .....	vii
	APÊNDICE VII – STATÍSTICASCAN .....	viii
	APÊNDICE VIII – BESAG&NEWELL MODIFICADO .....	ix
	APÊNDICE IX - SCRIPT UTILIZADO NO R .....	x
	APÊNDICE X – LOCALIZAÇÃO PONTUAL DOS HOMICÍDIOS OCORRIDOS ENTRE AGOSTO E NOVEMBRO DE 2011.....	xxi
	ANEXO .....	77
	MAPA DE JOÃO PESSOA .....	77

## LISTA DE FIGURAS

Figura 1	Matriz de Proximidade Espacial .....	15
Figura 2	Processo Regular .....	21
Figura 3	Processo Aleatório.....	22
Figura 4	Processo Heterogêneo .....	22
Figura 5	Estimador de Kernel .....	25
Figura 6	Localização da Área de Estudo .....	31
Figura 7	Divisão Política de João Pessoa .....	32
Figura 8	Cálculo da matriz de proximidade espacial .....	36
Figura 9	Ocorrência de homicídios durante os meses de agosto, setembro, outubro e novembro de 2011 .....	44
Figura 10	Risco relativo de homicídios durante os meses de agosto, setembro, outubro e novembro de 2011 .....	45
Figura 11	Estimador de intensidade de Kernel .....	48

## LISTA DE TABELAS

Tabela 1	Medidas descritiva dos homicídios .....	35
Tabela 2	Índice de auto correlação de Moran .....	37
Tabela 3	Índice de auto correlação de Getis&Ord .....	37
Tabela 4	Frequência por dia da semana durante os meses agosto, setembro, outubro e novembro de 2011.....	41
Tabela 5	Medidas Descritivas da Variável Idade Durante os Meses Agosto, Setembro, Outubro e Novembro de 2011 .....	43
Tabela 6	Frequência de Homicídios por Bairro Durante os Meses Agosto, Setembro, Outubro e Novembro de 2011 .....	44
Tabela 7	Valor Estimado para 2012 e Intervalo de Confiança .....	50

## LISTA DE GRÁFICOS

Gráfico 1	Box-plot dos homicídios .....	34
Gráfico 2	Frequência de homicídios por mês .....	39
Gráfico 3	Histograma de homicídios durante o mês .....	40
Gráfico 4	Frequência de homicídios por turno .....	41
Gráfico 5	Frequência de homicídios por dia da semana .....	42
Gráfico 6	Frequência de homicídios por sexo .....	42
Gráfico 7	Histograma da variável idade durante os meses agosto, setembro, outubro e novembro de 2011 .....	43
Gráfico 8	Frequência de homicídios por tipo de arma .....	43
Gráfico 9	Função G .....	46
Gráfico 10	Função F .....	46
Gráfico 11	Função J .....	47
Gráfico 12	Função L .....	47
Gráfico 13	Alisamento Exponência de Holt .....	49
Gráfico 14	Ajuste de Normalidade para o Alisamento Exponencial de Holt .....	50
Gráfico 15	Resíduos versus Valores estimados para o Alisamento Exponencial de Holt .....	50

## 1 INTRODUÇÃO

Um dos grandes motivos que confronta o bem estar da população é a insegurança causada pela violência urbana. Entende-se por violência urbana os atos que confrontam a lei e à ordem pública e acaba tirando a liberdade de ir e vir de várias pessoas, impondo-lhes medo insegurança e apreensividade. Mas, a violência não fica só nas ruas, ela também pode estar nos bares, clubes, escolas, no trânsito ou até mesmo nas residências. A violência pode ser classificada de diversas formas, desde o confronto verbal até a eliminação da vida alheia.

A violência, de um modo geral, é difícil de ser modelada por falta de registros, pois muitas vítimas não registram o ato sofrido, seja agressão, crimes contra o patrimônio, tentativa de homicídio, etc. Segundo WAISELFISZ (1989), em uma pesquisa realizada no Distrito Federal, no caso de violência física, só 6,4% dos jovens denunciaram à polícia, nos casos de assalto ou furto, apenas 4%; no caso de violência de trânsito, só 15%.

Para um combate eficaz à violência, é necessário saber onde ela se encontra. A violência em certos lugares pode ser representada pelo número de homicídios, apesar de os lugares em que ocorrem homicídios com mais frequência, serem diferentes, por exemplo, dos lugares onde ocorrem furtos a residências. Em geral, os homicídios ocorrem em áreas mais carentes e os furtos a residências ocorrem em áreas mais nobres, mas apenas os homicídios aparecem em muitos estudos como indicador principal de violência por esse ser um caso de extrema violência, é o maior dano que uma pessoa pode ter, além dos danos psicológicos que ficam com os familiares e amigos da vítima e também por termos dados mais confiáveis desses casos.

Estudos recentes feitos pela ONG (Organização Não Governamental) Fórum Brasileiro de Segurança Pública e divulgados na 5ª Edição do Anuário Brasileiro de Segurança Pública mostraram que foram registrados em 2010 na Paraíba 1.438 homicídios dolosos, o que a colocou como sendo o segundo estado brasileiro com a maior taxa para esse tipo de ocorrência (38,2 por 100 mil habitantes), tendo um dos mais crescentes índices (22,4% em relação a 2009). Nesse mesmo estudo, a Paraíba apresentou a 3ª maior taxa de Crimes Violentos Letais Intencionais (38,76 por 100 mil habitantes) que agrega as ocorrências de Homicídio Doloso, Latrocínio e Lesão Corporal seguida de Morte. Boa parte desses casos são ocorridos no município de João Pessoa, capital do estado da Paraíba.

Visto o crescente aumento no número de ocorrências de homicídios em João Pessoa, esse trabalho tenta verificar o comportamento da ocorrência de homicídios no município e modelar espacialmente essas ocorrências em busca de lugares críticos.

## **2 OBJETIVO**

### **2.1 Geral**

Analisar espacialmente e descritivamente as ocorrências de homicídios no município de João Pessoa, capital do Estado da Paraíba – Brasil.

### **2.2 Específicos**

- Acompanhar o número de ocorrências anuais de homicídios no município entre 2007 e 2011;
- Verificar bairros de maiores riscos;
- Identificar bairros com padrões similares de ocorrência;
- Identificar perfil das vítimas;
- Verificar possíveis conglomerações espaciais nas ocorrências de homicídios;
- Localizar áreas com maior incidência de homicídios;

### 3 METODOLOGIA

#### 3.1 Projeção Populacional

Para a aplicação de algumas técnicas será necessário estimar a população total para o meio do ano em estudo. No Brasil, a contagem da população atualmente é feita em um intervalo periódico de aproximadamente cinco anos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). As últimas contagens populacionais no Brasil foram feitas em 2007 e em 2011, foi necessário coletar essas informações para poder ser aplicado as projeções populacionais.

Será usado o método de projeção geométrica, que é indicado para estimar populações em um período curto de tempo, nesse método a população do município é alterada em uma taxa percentual calculada com base no total da população em dois instantes distintos de tempo.

A população projetada para o ano  $t$  é

$$P_t = (P_1)[(1+r)^z] \quad (1)$$

onde  $P_1$  é a população no ano inicial,  $z$  é o número de anos decorridos entre  $P_1$  e  $P_t$  e  $r$  é a taxa de crescimento que é dada pela equação (2), onde  $r$  é a taxa de crescimento,  $P_b$  é a população no ano base e  $y$  é o número de anos no período.

$$r = \left( \frac{P_1}{P_b} \right)^{(1/y)} - 1 \quad (2)$$

Pela falta de algumas informações, foi calculada apenas a taxa de crescimento de todo o município e foi considerado que todos os bairros do município têm essa mesma taxa de crescimento. Foi usado às informações da população de João Pessoa nos anos de 2007, ano em que houve a contagem da população, e em 2010, ano em que houve o censo demográfico, no cálculo da taxa de crescimento. A partir da taxa de crescimento obtida foi possível estimar a população para cada bairro do município de 2007 a 2011.

#### 3.2 Risco Relativo

Em certos estudos, o número de ocorrências não é uma informação muito relevante se não for considerada a população que esteve exposta ao risco, pois é sempre esperado que em áreas com maior aglomeração de pessoas tenham um maior número de ocorrências de um evento.

Em estudos espaciais geralmente é mais informativo uma medida chamada taxa de incidência, que é a razão entre o número de ocorrências de um evento em determinada área pela população total da região para o meio do período das ocorrências.

Para efeito comparativo, uma taxa de incidência pode ser considerada alta ou baixa se termos como base outra taxa, então foram calculadas as razões entre a taxa de incidência de cada bairro de João Pessoa pela taxa de incidência de todo o município, essa razão é chamada de risco relativo ( $RR$ ):

$$RR = \frac{\text{Taxa de incidência da área}}{\text{Taxa de incidência de toda região em estudo}} \quad (3)$$

Se o valor de  $RR = 1$ , então a área tem taxa semelhante a da região que a sobrepõe, ou seja, o número esperado de ocorrências é satisfatório em relação à região. Se  $RR > 1$  o número de ocorrências da área é superior ao padrão esperado de toda a região e se  $RR < 1$ , então o número de ocorrências da área é inferior ao valor esperado para aquele local.

### 3.3 Análise de Dados de Área

Seja  $x_1, x_2, \dots, x_n$  atributos associados as áreas  $s_1, s_2, \dots, s_n$ , independente das localizações espaciais das áreas é importante verificar o comportamento dos valores dos atributos  $x_i, i=1, \dots, n$ . A análise exploratória, resumidamente, é um conjunto de técnicas que resumem ou descrevem a distribuição de um conjunto de dados observados. Em estudos de área, serve tanto para escolha do agrupamento que será usado nos mapas coropléticos, que serão apresentados mais adiante, como também para simplificar o cálculo de alguns indicadores caso a variável tenha uma distribuição normal.

Para o caso de dados de áreas, onde temos um vetor de atributos associados a localizações delimitadas espacialmente, pode-se observar: medida de tendência central, como média, moda, mediana; medidas de dispersão, como amplitude, variância ou desvio padrão, covariância, correlação e quantis; valores extremos em busca de valores distantes dos padrões dos dados; estacionariedade no processo, verificação de parâmetros estatísticos constantes.

Os dados de áreas usualmente são apresentados através de mapas coropléticos. Mapas coropléticos são mapas onde os valores dos atributos associados as áreas são representados através de cores de modo a facilitar sua interpretação.

Muitos SIG's (Sistemas de Informação Geográfica) vem com algumas opções de criação automática de classes na geração de consultas (visualização de dados no mapa). Por exemplo, no Spring temos os seguintes métodos de geração de classes:

- Passo Igual – os valores do atributo são distribuídos em intervalos de amplitudes iguais e cada intervalo é atribuído uma cor.
- Quantil – as amplitudes dos intervalos são definidas de maneira que tenha o mesmo número de objetos em cada grupo.
- Estatístico – é calculada a média do vetor de atributos e os grupos são divididos em quantidades de desvio padrão em relação a média.
- Valor Único – muitas vezes temos atributos dicotômicos ou com poucos valores diferenciados (menos de 6), nesse caso cada valor do atributo pode ser representado no mapa por cores diferenciadas.

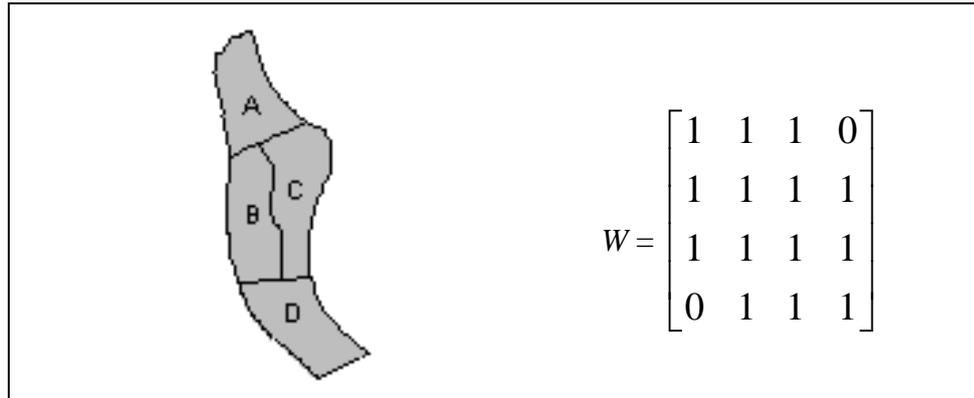
Entretanto, ao realizar um consulta em um mapa coroplético é necessário ter uma boa atenção no intervalo que está sendo usado, pois, diferentes intervalos podem gerar diferentes interpretações.

### 3.4 Matriz de Proximidade Espacial

Em análise de dados de área, geo-objetos são áreas distintas e delimitadas espacialmente. Seja  $A_1, A_2, \dots, A_n$  geo-objetos no espaço  $S$ , a matriz de proximidade espacial mostra a relação de proximidade existente entre esses geo-objetos. A matriz de proximidade espacial é uma matriz de dimensão  $n \times n$ , onde cada elemento  $w_{ij}$  da matriz guarda uma informação numérica de proximidade existente entre o geo-objeto da linha  $i$  e o geo-objeto da coluna  $j$ , para  $i$  e  $j$  variando de 1 até  $n$ .

Um critério utilizado em alguns SIG's é a criação de uma matriz binária  $W$  construída da seguinte maneira: se um geo-objeto  $i$  tem fronteira com o geo-objeto  $j$  então  $w_{ij} = 1$ , se essa condição não é satisfeita então  $w_{ij} = 0$ . Os elementos da diagonal principal de  $W$  mostra a relação de proximidade dos próprios geo-objetos, normalmente é utilizado o valor 1 para essa relação,

Exemplo:



**Figura 1:** Matriz de Proximidade Espacial

Como A faz fronteira com B e C então  $W_{12}=1$ , A e  $W_{13}=1$  e como A não faz fronteira com D então  $W_{14} = 0$  formando a primeira linha da matriz, como B e C tem fronteira com todos geo-objetos então todos os elementos da segunda e terceira linha de  $W$  é igual a 1 e como D não tem fronteira com A então  $w_{41}= 0$ .

### 3.5 Índice Global de Autocorrelação de Moran ( $I$ )

O coeficiente de correlação espacial  $I$  de Moran foi proposto em 1948. A hipótese subjacente é a aleatoriedade espacial, isto é, não existe dependência espacial nos dados, ou seja, o valor de um atributo em uma área não depende do valor do mesmo atributo nas áreas próximas. Este índice fornece um único valor como medida da associação espacial para todo o conjunto de dados, o que é útil na caracterização da região de estudo como um todo. Ele foi criado inicialmente para verificar padrões espaciais de doenças (Moran, 1948; Moran, 1950; Cliff & Ord, 1981), mas a técnica  $I$  de Moran também pode ser utilizada para investigar o componente espaço-temporal (Cliff & Ord, 1981) e verificar a dependência espacial de outras variáveis no espaço. A estatística de Moran calculada para verificar a independência de uma variável no espaço é semelhante ao coeficiente de correlação de Pearson, mas tem o acréscimo das informações da matriz de proximidade espacial e é dado por

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

onde  $w_{ij}$  é a matriz de proximidade espacial do geo-objeto  $i$ ,  $y_i$  é o valor do atributo  $i$  e  $\bar{y}$  é a média do vetor de atributos.

O índice  $I$  é indicado quando o vetor de atributo assume as propriedades:

- Independência espacial;
- Normalidade (existência do primeiro e do segundo momentos da distribuição);
- Processo estacionário de 1ª Ordem: Média  $\bar{z}$  constante.

Se  $W$  for gerada de forma que a soma dos elementos de cada linha seja igual a um, o índice  $I$  pode ser simplificado em

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_{i=1}^n (Z_i - \bar{Z})^2}, \quad (5)$$

onde  $Z_i$  é o valor normalizado [ $N(\mu = 0$  e  $\sigma^2 = 1)$ ] do atributo do geo-objeto  $i$  e  $\bar{Z}$  é a média do vetor de atributos,

Valores de  $I$  próximos de zero, indicam a inexistência de autocorrelação espacial entre os valores dos geo-objetos e seus vizinhos. Valores positivos desse índice indicam correlação direta, ou seja, o valor do atributo de um objeto tende a ser semelhante aos valores dos seus vizinhos. Valores negativos para o índice, por sua vez, indicam autocorrelação inversa.

Um dos aspectos mais relevantes com relação ao índice  $I$  de Moran é estabelecer sua validação estatística. Em outras palavras: será que os valores medidos representam correlação espacial significativa?

A significância do teste foi medida através de simulação. Primeiramente foi calculado e armazenado o índice  $I$  para os dados observados, depois foi feita varias permutações no vetor de atributos e para cada permutação foi calculado um novo valor de  $I$  e computado quantas vezes o valor de  $I$  para cada permutação gerada é maior do que o valor de  $I$  medido sem permutação.

O  $p$ -valor será:

$$p = \frac{NVM}{N + 1}, \quad (6)$$

onde  $NVM$  é o número de vezes que o valor de  $I$  com a variável permutada supera o valor de  $I$  da variável original e  $N$  é o número de permutações realizadas.

A Hipótese  $H_0: I = 0$  (não existe autocorrelação espacial entre os bairros) será rejeitada se  $p$  for menor que o nível de significância.

### 3.6 Índices Locais de Associação Espacial

Este índice fornece uma medida de associação para cada geo-objeto no espaço. Os indicadores locais  $I_i$  de Moran são dado por (Anselin, 1996):

$$I_i = \frac{(y_i - \bar{y}) \sum_{j=1}^n w_{ij} (y_j - \bar{y})}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} \quad (7)$$

Se  $I_i > 0$ , o geo-objeto  $i$  tem valor semelhante a sua vizinhança, se  $I_i < 0$  o geo-objeto  $i$  tem vizinhança com valores distintos e se  $I_i = 0$  os valores das regiões próximas ao geo-objeto  $i$  são aleatórios.

Normalizando as variáveis o indicador reduz-se a

$$I_i = \frac{z_i \sum_{j=1}^n w_{ij} z_j}{\sum_{j=1}^n z_j^2} \quad (8)$$

### 3.7 Índice Global de Autocorrelação de Getis e Ord

Outro método para verificação da presença de autocorrelação espacial foi recentemente proposto por Getis e Ord (1992). A estatística  $G$  é calculada através da matriz de proximidade espacial em função de uma distância crítica  $d$  a partir do local, ou seja, serão computados apenas os valores dos atributos que estão dentro de uma distância  $d$ . A estatística  $G$  para uma distância crítica  $d$  escolhida é:

$$G(d) = \frac{\sum_i \sum_j w_{ij}(d) y_i y_j}{\sum_i \sum_j y_i y_j} \quad (9)$$

Em que  $w_{ij}(d)$  são os elementos da matriz de proximidade espacial e  $y_i$  é o valor do atributo associado ao geo-objeto  $i$ .

Um valor positivo e significativo de  $G(d)$  indica aglomeração espacial de valores elevados, enquanto um valor negativo e significativo indica agrupamento espacial de valores pequenos (Anselin, 1992).

Getis e Ord sugerem ainda dois tipos de estatística semelhantes às medidas tradicionais de autocorrelação espacial em que avaliam um padrão local de agrupamentos,  $G_i(d)$  e  $G_i^*(d)$  os valores dessas estatísticas são:

$$G_i(d) = \frac{\sum_{j=1}^n w_{ij}(d) y_j}{\sum_{j=1}^n y_j}, j \neq i \quad (10)$$

e

$$G_i^*(d) = \frac{\sum_{j=1}^n w_{ij}(d) y_j}{\sum_{j=1}^n y_j} \quad (11)$$

A diferença entre  $G_i(d)$  e  $G_i^*(d)$  é que  $G_i(d)$  considera como nulo o valor de proximidade do atributo  $j$  com ele mesmo, então  $G_i(d)$  não utiliza os elementos da diagonal principal da matriz de proximidade. Em outras palavras, o valor da própria observação não está incluído na estatística.

Essa estatística indica se certo local está cercado por uma vizinhança de valores baixos ou valores altos para a variável em questão. Este índice é simplesmente uma relação entre a soma dos valores nas localizações vizinhas e a soma de todos os valores do atributo.

A significância de  $G$  também pode ser calculada através de permutação e é representada no mapa da seguinte maneira:

- Negativo\*\*\* - Índice local negativo com p-valor menor que 0,005;
- Negativo\*\* - Índice local negativo com p-valor entre 0,005 e 0,025;
- Negativo \* - Índice local negativo com p-valor entre 0,025 e 0,05;
- Negativo - Índice local negativo com p-valor acima de 0,05;
- Positivo - Índice local positivo com p-valor acima de 0,05;
- Positivo\* - Índice local positivo com p-valor entre 0,025 e 0,05;

- Positivo\*\* - Índice local positivo com p-valor entre 0,005 e 0,025;
- Positivo\*\*\* - Índice local com p-valor menor que 0,005.

### 3.8 Método de Varredura Scan

Como em muitos casos não se conhece a distribuição do vetor de atributos, KULLDORFF e NARGAWALLA (1995) propuseram o método não paramétrico para a detecção de conglomerados. Diferente dos métodos apresentados anteriormente, este utiliza a informação da população total em cada geo-objeto na busca de regiões com valores similares de certa variável, essas regiões podem ser chamadas de agrupamentos, conglomerados ou clusteres.

Se tivermos uma região dividida em pequenas áreas, como, por exemplo, bairros em um município, suas áreas serão representadas pelos seus centroides (centro do geo-objeto). Para o cálculo da estatística de varredura scan são computados esses centróides através de suas coordenadas geográficas (Latitude e Longitude), o número de casos de ocorrência de dado evento na região considerada ( $C$ ) e a população exposta ao risco no período sob toda região em estudo ( $N$ ). É gerado conglomerados circulares a partir dos centróides e calculado a probabilidade de ocorrência do evento dentro do conglomerado ( $p$ ) e fora dele ( $q$ ). Assumindo que as contagens são variáveis aleatórias independentes e identicamente distribuídas e seguem distribuição de Poisson, o valor da estatística scan será:

$$KN = \max_{z \in Z} \frac{L(z, \hat{p}(z), \hat{q}(z))}{L_0} \quad (12)$$

onde  $z$  é o conjunto de todos os possíveis candidatos a conglomerados,  $\hat{p}(z)$  e  $\hat{q}(z)$  são estimativas de  $p(z)$  e  $q(z)$  respectivamente,  $L_0$  é dado por (13)

$$L_0 = \frac{C^C (M - C)^{M-C}}{M^M} \quad (13)$$

e  $L(z)$  é definido como:

$$L(z) = \frac{e^{-p \cdot n_z - q(M - n_z)}}{C!} p^{c_z} q^{C - c_z} \prod_i c_i \quad (14)$$

em que  $c_z$  é o número de casos no círculo  $z$ ,  $c_i$  o número de casos no círculo  $i$  e  $n_z$  é o número de indivíduos em risco no círculo  $z$ .

### 3.9 Método Besag e Newell

A ideia do método de Besag e Newell é similar ao método Scan, é definido um conglomerado circular em cada centroide dos geo-objetos e ampliado o raio dessa região, incluindo novos centróides, sendo que o critério de parada de crescimento do raio é que a região circular contenha pelo menos  $k$  casos em seu interior. Seja  $C$  o número total de casos em toda região em estudo e  $Y$  a população total exposta ao risco na região. A estatística  $L$  do teste é:

$$L = \min\{j : C_{j(i)} \geq k\} \quad (15)$$

Considerando como  $M_j$  a população observada em uma área  $j$ , o nível de significância do teste é mostrado na seguinte expressão (COSTA e ASSUNÇÃO, 2005):

$$p_k(i) = P(L_i \leq l_i) = 1 - \sum_{j=1}^{k-1} \frac{(M_{j(i)} C / M)^j}{j!} \exp(-M_{j(i)} C / M) \quad (16)$$

Se o p-valor calculado é menor que a significância adotada a aglomeração encontrada será considerada significativa. Uma modificação do método Besag e Newell para poder ser comparado com o método Scan é dada pela estatística  $T$ , que elimina a dependência do número de casos e da significância do método (COSTA e ASSUNÇÃO, 2005):

$$T = \min_k \{ \min_{Z_k \in Z_k} [p_k(i)] \} \quad (17)$$

Sendo  $Z_k$  o conjunto de todas as regiões circulares  $z_k$  centradas nos centróides e contendo  $k$  casos próximos.

### 3.10 Processos Pontuais

Uma das grandes desvantagens em se utilizar análise por áreas é a generalização dos lugares de ocorrências do evento estudado, principalmente se as áreas forem de grande extensão. A existência de áreas com maior ocorrência é bem melhor percebida se for considerado a localização exata da ocorrência do evento (Latitude e Longitude), denominados de processos pontuais. Além do mais, esse tipo de análise segue melhor a 1ª Lei da Geografia, que diz que coisas próximas são mais parecidas que coisas mais distantes.

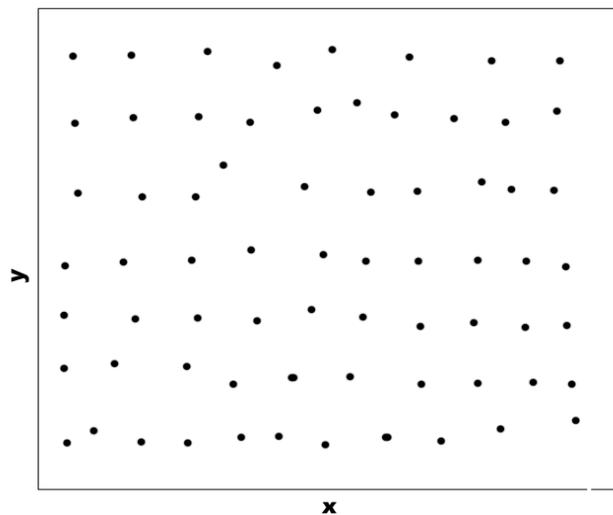
Processos pontuais são processos em que a variável aleatória é a localização espacial da ocorrência de um evento de interesse dada pelas coordenadas geográficas (latitude e longitude).

Existem dois tipos de processos pontuais. Os simples, quando observamos apenas uma característica da ocorrência do evento, e os marcados, quando temos mais de uma característica na ocorrência de um evento. Para esse estudo foi utilizado apenas a análise pontual simples.

O objetivo da análise pontual é determinar se certos eventos tem sua ocorrência espacial aleatorizada ou se existe uma tendência na localização de sua ocorrência.

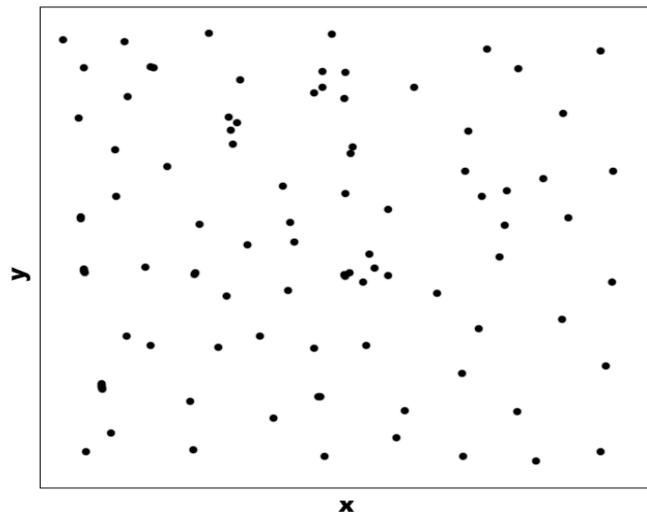
Existem três padrões gerais de pontos dispersos no espaço geográfico:

- Regular - Quando os pontos estão dispersos de forma uniforme, com distâncias semelhantes entre si, indicando que existe um fator que interfere na localização das ocorrências.



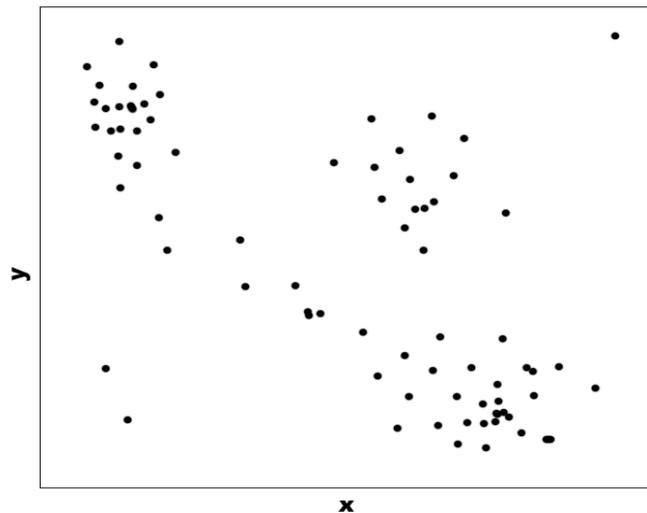
**Figura 2:** Processo Regular

- Homogêneo - Quando os pontos estão dispersos de forma aleatória, em forma de um único processo de Poisson, o que indica que as ocorrências podem acontecer em qualquer lugar do espaço geográfico, ou seja, não existe nenhum fator significativo que influêncie na localização da ocorrência do evento. Além disso, temos que as propriedades estatísticas do processo, média e variância são constantes no espaço  $S$ , para várias localizações  $s$  em  $S$  os valores da média e da variância serão idênticos.



**Figura 3:** Processo Aleatório

- Heterogêneo - Existe mais de um processo de Poisson atuando no espaço geográfico. Muitos postos estão concentrados, formando conglomerados enquanto que existem grandes áreas sem ocorrência do evento ou com pouca ocorrência.



**Figura 4:** Processo Heterogêneo

Embora ao observar as figuras apresentadas, fica evidente a diferença do tipo dos processos, na prática nem sempre isso é possível de ser confirmado apenas com a observação da localização dos pontos. Para diferentes observadores um mesmo processo pode ser classificado como aleatório ou possuir agrupamentos. Então, para evitar discordância quanto à interpretação dos dados será apresentado mais adiante funções que auxiliam a identificação de presença de conglomerados. Essas funções são baseadas em estatísticas descritivas que podem ser analisada individualmente ou em conjunto.

Seja  $A$  uma superfície contínua, a ocorrência de um evento  $z$  em  $A$  é representada graficamente por um ponto. Representamos por  $z(x) = 1$  se existe um ponto no lugar  $x$ , ou  $z(x) = 0$  caso contrário. Pela grande extensão e continuidade do espaço  $A$ , usualmente a  $P(z(x)=1)$  é zero (RIPLEY, 2004). Mas, se dividirmos o espaço  $A$  em pequenas áreas limitadas  $S_{i's}$ , todas formando uma partição de  $A$  e se consideramos que os números de ocorrências dos eventos nas áreas  $S_{i's}$ , são independentes e identicamente distribuídos e definimos como  $Z(S)$  o número esperados de pontos na área  $S$ , podemos comparar o número esperado de ocorrências de um evento por unidade de área de  $S$  com o número observado na região  $S$ . Sob a hipótese que temos um processo homogêneo de Poisson, ou seja, um modelo com aleatoriedade espacial completa (CSR –Complete Spatial Randomness), basta então testar essa hipótese para tomar conclusões sobre possíveis conglomerações ou podemos simplesmente melhor analisar a distribuição de pontos no espaço através da intensidade de pontos.

Um Processo Homogêneo de Poisson (PHP) é um processo completamente aleatório a qual tem duas características importantes, são estacionários e isotrópicos (CRESSIE, 1993). Um processo pontual  $Y(S)$  em  $R^2$  é estacionário se sua média e variância não se alteram sobre toda região  $A$ , então  $Y(S_i)$  não depende da localização de  $Y(S_j)$ , pois será o mesmo, para qualquer  $k$  inteiro, na região  $S_i: i=1, \dots, k$ . Dizemos, então, que o processo é invariante quanto a sua translação. Já se o processo assume comportamento semelhante para qualquer direção em  $A$ , ou seja, a distribuição conjunta de  $Y(S_1), \dots, Y(S_k)$  é invariante a rotação por um ângulo arbitrário  $\theta$ , dizemos que o processo é isotrópico.

### 3.11 Função intensidade de ocorrências

A função intensidade de ocorrência  $\lambda(S)$  ela mostra a probabilidade de existir pelo menos uma ocorrência na área  $S$ , equação (18), onde  $N_s$  é o número de ocorrência de um evento no espaço  $S$ .

$$\lambda(S) = \frac{P(N_s > 0)}{\text{área de } S} \quad (18)$$

A modelagem de um processo pontual é feita através de sua esperança  $E[N(S)]$  e a covariância  $E[N(S_i), N(S_j)]$ . Segundo Cressie (1993), quando uma pequena região  $ds$  em torno

de  $S$  tende a zero, a função de intensidade em (18) em um ponto, para propriedades de primeira ordem, é definida por:

$$\lambda(s) = \lim_{|ds| \rightarrow 0} \left\{ \frac{E[N(ds)]}{|ds|} \right\} \quad (19)$$

Existem vários métodos para a estimação da função intensidade, alguns utilizam o efeito de primeira ou de segunda ordem do processo. O efeito de primeira ordem (efeito global) se refere à estacionariedade do processo, ou seja, a variação no valor médio do processo se o processo é estacionário  $\lambda(u)$  é constante. O efeito de segunda ordem (Efeitos Locais ou de pequena escala) corresponde a estrutura de correlação espacial do processo, representam a dependência espacial do processo e assume que a matriz de covariância do processo não depende da direção.

Pela propriedade de segunda ordem, a intensidade conjunta em duas regiões infinitesimais  $|ds_i|$  e  $|ds_j|$  que contêm os pontos  $i$  e  $j$ ,  $\lambda(s_i, s_j)$  é definida por:

$$\lambda(s_i, s_j) = \lim_{\substack{|ds_i| \rightarrow 0 \\ |ds_j| \rightarrow 0}} \left\{ \frac{E[N(ds_i)N(ds_j)]}{|ds_i| |ds_j|} \right\} \quad (20)$$

se o processo é estacionário e isotrópico  $\lambda(s_i, s_j)$  depende apenas da distância entre  $s_i$  e  $s_j$ .

### 3.12 Estimador de Intensidade de Kernel

A estimativa da função intensidade dada pelo estimador de Kernel é calculada através de uma função bi-dimensional em torno dos pontos gerando uma superfície com valores proporcional à intensidade de amostras por unidade de área. A função Kernel contabiliza todos os pontos dentro de uma região e faz uma ponderação através das distâncias entre o centro dessa região e as localizações dos eventos dentro da região. Seja  $(s_1, \dots, s_n)$  as localizações dos  $n$  eventos em uma região  $A$ , a intensidade  $\lambda(s)$  de uma localização genérica  $s$  é estimada por [CRESSIE, 1993]:

$$\hat{\lambda}_\tau(s) = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{d_i}{\tau}\right) \quad (21)$$

onde  $\tau > 0$  definido como o raio de influência ou largura da banda que é responsável pelo alisamento da superfície gerada,  $k(\cdot)$  é uma função de ponderação (função Kernel).

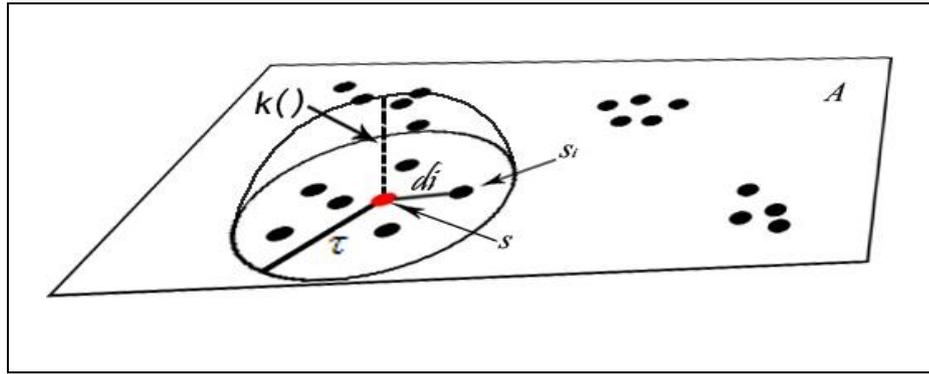


Figura 5: Estimador de Kernel

A função Kernel pode ser do seguinte tipo [CRESSIE,1993; MARTINEZ, 2001]:

Quadro 1: Algumas Funções Kernel

	Função	Exemplo
Uniforme	$k(t) = \begin{cases} \frac{1}{2}, &  t  \leq 1 \\ 0, &  t  > 1 \end{cases}$	
Biponderado / quartic	$k(t) = \begin{cases} \frac{15}{16}(1-t^2)^2, &  t  \leq 1 \\ 0, &  t  > 1 \end{cases}$	
Gaussiano	$k(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right)$	

### 3.13 Método dos vizinhos próximos (propriedade de segunda ordem)

A intensidade do processo estimada anteriormente resulta em uma superfície onde é possível identificar zonas de aglomerações, mas pequenas áreas de aglomeração sempre existiram em um processo aleatório, logo o resultado gerado pelo estimador de intensidade de Kernel pode ser não conclusivo se não verificado em conjunto com algumas outras técnicas, que levam em consideração as propriedades de segunda ordem do processo (dependência espacial).

#### 3.13.1 Função $G$

Seja  $r$  uma distância maior que zero (raio de influência) definida pelo observador, a Função  $G(r)$  é uma função baseada na medida de distância de um evento escolhido aleatoriamente a os seus vizinhos mais próximos (evento-para-evento). Esta estatística estima à probabilidade de que a distância de um evento escolhido aleatoriamente ao seu vizinho mais próximo seja menor ou igual a  $r$ .

A função  $G$  é dada por (22),

$$G(r) = \frac{\sum_{i=1}^n I(r_i \leq r)}{n} \quad (22)$$

onde  $r_i$  é a distância de um ponto  $s$  ao seu vizinho,  $s_i$ , mais próximo a qual a distância seja menor que  $r$ ,  $n$  é o número de pontos na região e  $I(\cdot)$  é a função indicadora do evento  $A_i$ . A função  $G(r)$  é uma distribuição acumulada de  $r$ .

Para concluir sobre a presença de conglomerados faz-se a plotagem dessa função de distribuição acumulada. Se o gráfico dessa função apresentar um crescimento rápido para os pequenos valores de distância indicará um forte indício da presença de interações entre as áreas, indicando a existência de conglomerados. Se em pequenos valores de distância o crescimento for lento e conseqüentemente rápido nos valores altos de distância indicará regularidade no processo.

Podemos também comparar a distribuição dos pontos observados com uma distribuição simulando um processo com aleatoriedade espacial completa (CSR), podemos ainda gerar várias simulações e calcular os valores máximos e mínimos para cada distância

gerando limites superiores e inferiores (envelopes) da função. Sob a hipótese de CSR a função de distribuição de  $G$  é dada por:

$$G(r) = 1 - e^{-\lambda\pi r^2} \quad (23)$$

onde  $\lambda$  é o número esperado de pontos por unidade de área.

A distribuição simulada de  $G(r)$  é estimada por:

$$G_s(r) = \frac{\sum_i^k \hat{G}_i(r)}{k} \quad (24)$$

onde,  $i=1, \dots, k$  são simulações independentes feitas da função através da hipótese CSR e  $k$  é o número total de simulações. Espera-se que esse gráfico tenha um comportamento de uma reta com inclinação de  $45^\circ$ .

Os valores mínimos,  $L(r)$ , e máximos,  $U(r)$ , estimados da função  $G(r)$  são:

$$U(r) = \max \{ \hat{G}_i(r) \}, i = 1, \dots, k \quad (25)$$

$$L(r) = \min \{ \hat{G}_i(r) \}, i = 1, \dots, k \quad (26)$$

Verificando o gráfico da distribuição  $G(r)$  com a distribuição de  $G_s(r)$ , se  $G(r)$  crescer rapidamente no início e supera a curva  $U(r)$  então indicará a presença de aglomerações na região. Se a curva  $G(r)$  estiver entre as curvas  $L(r)$  e  $U(r)$  indicará aleatoriedade e se  $G(r)$  for inferior a  $L(r)$  indicará regularidade entre os pontos.

### 3.13.2 Função F

Assim como a função  $G$ , a função  $F$  leva em sua estatística medidas de distância entre pontos, só que a distância calculada para essa função é a distância chamada de ponto amostral-para-evento que é a distância de um ponto escolhido aleatoriamente em uma região  $A$  para os eventos mais próximos  $S_i, i, \dots, k$ . Sua função de distribuição é dada por:

$$F(r) = 1 - e^{-\lambda\pi r^2} \quad (27)$$

onde  $\lambda$  é a intensidade e  $r$  é a distância limite de proximidade.

A função  $F(r)$  é estimada por [CRESSIE, 1993]:

$$\hat{F}(r) = \frac{\sum_{i=1}^n I(r_i^* \leq r)}{n} \quad (28)$$

Sendo  $r_i^*$  a distância do evento  $i$  para o evento mais próximo em  $A$ .  $\hat{F}(r)$  mostra a probabilidade de a distância de um ponto selecionado aleatoriamente ao seu vizinho mais próximos seja menor ou igual  $r$ .

O gráfico dessa função mostrará a presença de aglomeração se mostra um crescimento rápido no final da curva ou ainda podemos plotar o gráfico de  $\hat{F}(r)$  pela distribuição simulada  $\hat{F}_s(r)$  e os limites mínimos,  $L(r)$  e  $U(r)$ :

$$L(r) = \min \{ \hat{F}_i(r) \}, i = 1, \dots, k \quad (29)$$

$$U(r) = \max \{ \hat{F}_i(r) \}, i = 1, \dots, k \quad (30)$$

### 3.13.3 Função J

A função  $J$  é uma relação entre a função  $G$  e a função  $F$  e é dada por [KERSCHER, 1999]:

$$J(r) = \frac{1 - G(r)}{1 - F(r)} \quad (31)$$

Sob a hipótese CSR o valor esperado da função  $J$  é 1.  $J(r) < 1$  mostra que os efeitos  $G(r)$  unido com  $J(r)$  foram significativos para na indicação de presença de conglomerados e se  $J(r) > 1$  indica regularidade.

### 3.13.4 Função K

A função  $K$  tem uma medida mais eficaz da dependência espacial em processos com intervalo amplos e é definida pela equação (32) [CRESSIE, 1993]:

$$K(h) = \lambda^{-1} E(\#) \quad (32)$$

Onde  $h$  é um evento arbitrário,  $\lambda^{-1}$  é o número de eventos esperados por unidade de área e  $\#$  é o número de eventos extras contidos a uma distância  $h$  de um evento arbitrário. Considerando estacionariedade e isotropia no processo, a função  $K$  é definida por [CRESSIE, 1993]:

$$\hat{K}(h) = \hat{\lambda}^{-1} \frac{\sum_{i=1}^n \sum_{j=1, i \neq 0}^n W(S_i, S_j)^{-1} I(\|S_i - S_j\| \leq h)}{N} \quad (33)$$

Sendo  $h > 0$ ,  $S_i$  para  $i=1, \dots, n$  e  $S_j, j=1, \dots, n$  eventos ocorridos na região  $A$ ,  $\hat{\lambda} = \frac{N}{|A|}$  é a estimativa da função intensidade de  $\lambda$ ,  $N$  é o número total de eventos ocorridos na região  $A$  e  $W(S_i, S_j)$  a proporção da circunferência centrada em  $s_i$  e que é interna a  $A$ . Existirá aglomeração espacial se  $K(h) > \pi h^2$ .

### 3.13.5 Função L

Com o intuito de facilitar a interpretação do gráfico  $K$  foi proposto a função  $\hat{L}(h)$  definida por [CRESSIE, 1993]:

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h \quad (34)$$

Sob CSR podemos construir os envelopes dessa função. Se o gráfico de  $\hat{L}(h)$  em relação as distâncias apresentar valores positivos e acima do limite superior indicará a presença de aglomeração espaciais e se for obtido valores negativos e abaixo do limite inferior indicará regularidade espacial.

### 3.14 Teste Qui-Quadrado de Aderência

O teste foi introduzido pelo britânico Karl Pearson em 1900. O Teste de Pearson avalia se os valores de um conjunto de dados com distribuição multinomial são iguais a certos valores especificados.

Considerando  $H_0: \pi_j = \pi_{j0}, j = 1, \dots, c$ , onde  $\sum_j \pi_{j0} = 1$ . Sob a hipótese  $H_0$  os valores esperados são  $\mu_j = n\pi_{j0}, j=1, \dots, c$ . O estatística do teste proposto por Person é

$$\chi^2 = \sum_{j=1}^c \frac{(n_j - \mu_j)^2}{\mu_j} \quad (35)$$

Grandes diferenças entre  $n_j$  e  $\mu_j$  irá resultar em valores  $\chi^2$  grande. A Região crítica do teste é

$$RC = \{(x_1, x_2, \dots, x_n), \chi^2 \geq \chi^2_{(c-1), \alpha}\}. \quad (36)$$

Então, para testar a hipótese nula basta comparar o valor calculado da estatística com o valor tabelado para  $c-1$  graus de liberdade. Sob a hipótese nula o p-valor é  $P(\chi^2_{(c-1)} \geq \chi^2_{cal})$ .

### 3.15 SÉRIES TEMPORAIS

Série temporal é qualquer conjunto de observação que esta ordenada em um período de tempo (Moretin e Tolo, 2006). O principal objetivo da série temporal é descrever o comportamento de dados para poder gerar valores futuros (fazer previsões).

A maioria dos métodos de previsão baseia-se na ideia de que observações passadas contêm informações sobre o padrão de comportamento da série temporal. O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros da série.

Será apresentado o método de suavização exponencial de Holt (SEH) para análise de séries temporais. Essa metodologia permite que valores futuros de uma série sejam previstos tomando por base apenas seus valores presente e passados. Isto é feito explorando a correlação temporal que existe geralmente entre os valores exibidos pela série.

#### 3.15.1 Suavização Exponencial de Holt (SEH)

O método de Suavização Exponencial de Holt é utilizado para analisar séries que apresentam tendência. Ou seja, uma série do tipo

$$Z_t = \mu + T_t + a_t, \quad t = 1, \dots, N \quad (37)$$

onde  $\mu$  representa o nível,  $T$  representa a tendência e  $a$  é o resíduo aleatório com média zero e variância constante ( $\sigma_a^2$ ).

O método utiliza duas constantes de suavização, uma para modelar o nível da série e outra para modelar a tendência.

Para um instante  $t$  o valor do nível da série é estimado por (38) e a tendência da série é estimado por (39).

$$\bar{Z}_t = AZ_t + (1-A)(\bar{Z}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1 \text{ e } t = 2, \dots, N, \quad (38)$$

$$\hat{T}_t = C(\hat{Z}_t - \bar{Z}_{t-1}) + (1-C)\hat{T}_{t-1}, \quad 0 < C < 1 \text{ e } t = 2, \dots, N, \quad (39)$$

onde  $A$  e  $C$  são denominadas constantes de suavização.

A previsão é feita adicionando-se ao valor  $\bar{Z}_t$  a tendência  $\hat{T}_t$  multiplicada pelo número de passos que se deseja prever ( $h$ ), ou seja:

$$\hat{Z}_{t+h}(h) = \bar{Z}_t + h\hat{T}_t, \quad \forall h > 0, \quad (40)$$

Para determinação das constantes de suavização escolhe-se os valores do vetor  $(A, C)$  tal que a soma dos quadrados dos erros de previsão sejam mínimos.

### 3.16 Área de Estudo

A área que foi utilizada nesse estudo é a cidade de João Pessoa, capital do estado da Paraíba, situada no nordeste brasileiro. A cidade é destacada por ter uma vasta área de vegetação e possuir o ponto mais oriental das Américas situado na Ponta do Seixas. De acordo com o IBGE o município possui área de aproximadamente 211 Km<sup>2</sup> e 723.155 habitantes. O município está localizado entre as coordenadas geográficas 34°46'52 a 35°01'08 de longitude oeste e 6°59'53 a 7°16'06 de latitude sul e é limitado ao norte pelo Rio Sanhauá e ao sul pelo Rio Gramame, onde se situam áreas de mangue pertencentes ao município e banhado a leste pelo oceano Atlântico.

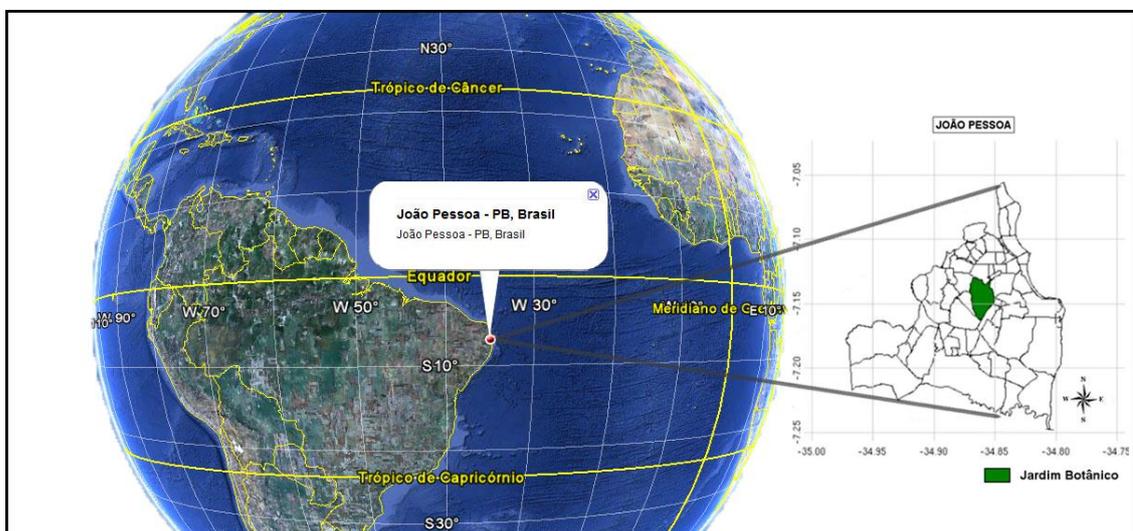
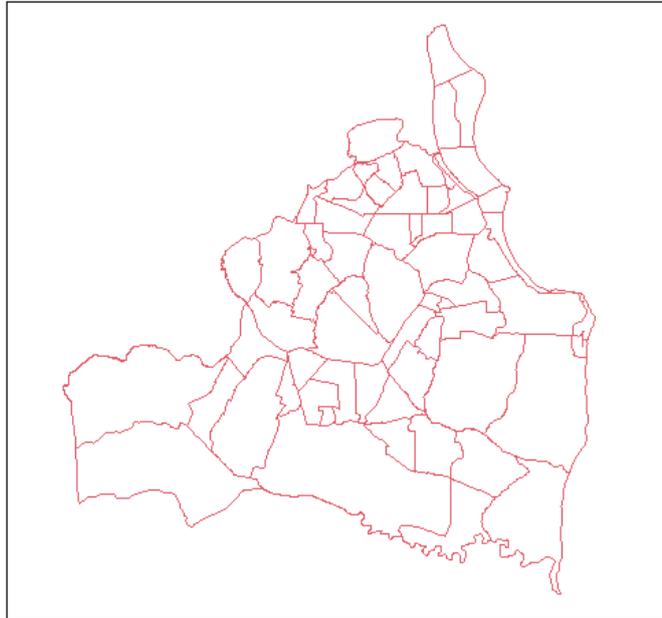


Figura 6 – Localização da Área de Estudo - Adaptação do Google Earth

#### 3.16.1 Mapa de Divisão Política

O mapa que contém a divisão política e delimitações dos bairros do município de João Pessoa foi cedido pela Secretaria de Planejamento da Prefeitura Municipal de João Pessoa,

referente ao ano 2007. O mapa está no formato “.shp” e pode ser visualizado na Figura 2. Com ele foi possível atribuir esse plano de informação com as ocorrências dos homicídios por bairro, população total em cada bairro, coordenadas das ocorrências de homicídios, entre outras informações.



**Figura 7**– Divisão Política de João Pessoa

**Fonte:** *Prefeitura Municipal de João Pessoa (2007)*

### 3.17 Softwares Utilizados

#### 3.17.1 R

O R é um ambiente para computação estatística e gráfica que pode ser obtido gratuitamente, sob os termos de licença da GNU (General Public License) no endereço eletrônico "<http://www.r-project.org/>". A versão utilizada foi a 2.9.0.

Os pacotes utilizados foram:

Spatstat – Pacote com várias funções para plotagem de dados espaciais, análise exploratória de dados, simulação, amostragem espacial, diagnóstico, modelo e inferência formal (<http://www.spatstat.org/>).

DCluster – Pacote com conjunto de funções para detecção de conglomerados espaciais (<http://cran.r-project.org/web/packages/DCluster/index.html>).

### **3.17.2 Spring**

O Spring é um sistema de informação geográfica com as funções de processamento de imagem, análise espacial, modelagem numérica de terreno e consulta a banco de dados espaciais. O SPRING é um projeto do INPE / DPI (Divisão de Processamento de Imagens) podendo ser adquirido gratuitamente pela Internet (“<http://www.dpi.inpe.br/spring>”), bastando se cadastrar na própria página. A versão utilizada foi a 5.1.7.

### **3.18 Banco de Dados**

O banco de dados disponível foi cedido pelo CIOP com informações de ocorrências de homicídios por bairro de 2007 a 2010 no município de João Pessoa, mas foi proposto um estudo mais detalhado com informações mais recentes possíveis, pois o homicídio é um processo que se altera constantemente durante o tempo, tanto em frequência como no espaço, então além das informações de 2007 a 2010 foi decidido acrescentado os meses de agosto, setembro, outubro e novembro de 2011 com novas variáveis. Com isso, o banco de dados foi complementado com atualizações diárias através das informações do Instituto de Polícia Científica (IPC) e do Centro Integrado de Operações Policiais (CIOP). Além dos dados coletados por essas duas instituições, através das referências registradas pelo CIOP, buscou-se a localização exata (coordenadas geográficas) do local do crime violento, uma vez que tais dados possibilita a elaboração de análise mais precisa e confiável, pois refletirá melhor o fenômeno em estudo.

Os dados do IPC são cedidos através de planilhas eletrônicas com informações como idade, data e hora de entrada do corpo da vítima no GEMOL, cidade e local da ocorrência dos homicídios sendo que, devido à natureza do IPC, os dados mencionados necessitam de algumas correções ou complementações. Sendo assim, junto aos dados do IPC, foram acrescentadas informações das fichas de ocorrências geradas diariamente pelo CIOP, com informações mais próximas da hora em que o fato foi ocorrido e confirmações de idade e sexo da vítima além de melhores descrições do endereço em que o fato ocorreu. Através dessas descrições e de mapas georreferenciados, foi possível coletar as coordenadas geográficas de boa parte das ocorrências.

## 4 RESULTADOS

Como nos dados coletados de ocorrências de homicídios no município de João Pessoa para o ano 2011 foi acrescentado uma nova fonte de informação, foi decidido separar os dados desse ano dos demais, dividindo o trabalho em duas seções: Uma para os dados do período de 2007 a 2010 e outra para os dados de 2011. Além dessas duas seções, foi criada uma nova seção com uma estimativa das ocorrências para o ano 2012.

### 4.1 Homicídios ocorridos entre 2007 e 2010

#### 4.1.1 Análise de Autocorrelação Espacial para Dados de Área

Como o padrão de homicídios em um lugar pode variar muito em relação ao tempo, pois o que era considerado um número alto de ocorrências, por exemplo, 10 anos atrás hoje pode ser considerado como um número baixo, para a construção das classes nos mapas coropléticos apresentados no Anexo I foi feito um gráfico Box Plot (Gráfico 1) e observado que não houveram grandes alterações na distribuição dos homicídios nos bairros no município de João Pessoa. Com isso, foram construídos mapas coropléticos com intervalos iguais para cada ano, possibilitando uma melhor comparação entre eles. É também possível verificar a assimetria positiva das observações, mas ainda assim foi feito o teste de normalidade de Kolmogorov-Smirnov, obtendo como p-valor menor que 0,01 para todas as variáveis, a hipótese de normalidade é rejeitada com  $\alpha=0,01$ .

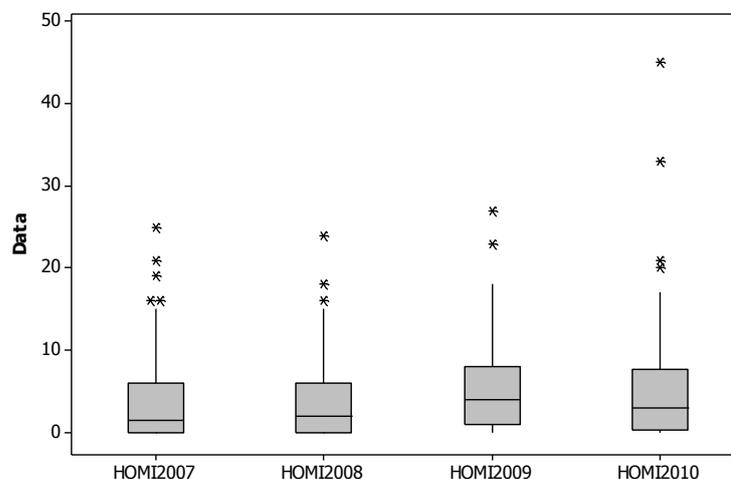


Gráfico 1 - Box-plot dos homicídios

**Tabela 1 - Medidas descritiva dos homicídios**

VARIÁVEL	DESVIO								
	MÉDIA	PADRÃO	SOMA	MINIMO	Q1	MEDIANA	Q3	MÁXIMO	MODA
HOM_2007	4,078	5,810	261	0	0,00	1,50	6,00	25	0
HOM_2008	4,172	5,382	267	0	0,00	2,00	6,00	24	0
HOM_2009	5,547	5,944	355	0	1,00	4,00	8,00	27	0
HOM_2010	5,800	8,100	371	0	0,25	3,00	7,75	45	0

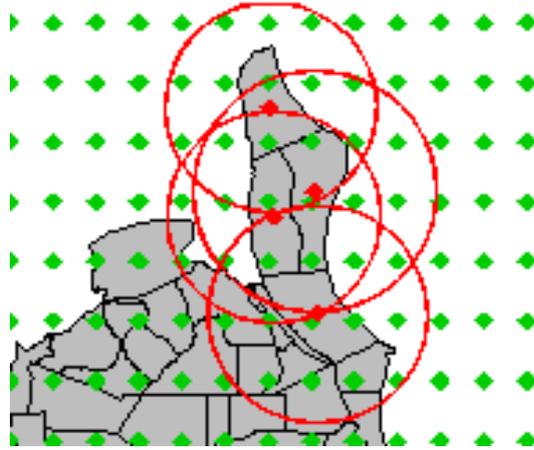
Fonte: Calculado com base nos dados brutos do CIOP e IPC

A Tabela 1 apresenta que, ao passar dos anos, os dados ficaram mais dispersos em relação à média, houve um aumento no número de ocorrências, conseqüentemente um aumento da média e apesar de o valor mais frequente de ocorrências por bairro ser zero para todos os anos, a frequência desse valor vem diminuindo.

No Apêndice I é possível encontrar os mapas coropléticos com número de ocorrência de homicídios por bairros. Percebe-se que os bairros: Mangabeira, Mandacarú e Cristo Redentor obtiveram um alto número de ocorrência (15 ou mais) e permaneceram durante os quatro anos consecutivos. O bairro do Bessa, que em 2007 não houve nenhuma ocorrência, chegou em 2010 com um dos mais altos valores de ocorrência.

Mapas mais informativos são encontrados no Apêndice II, que mostra a taxa de incidência de homicídios por bairro do município. Nele, destaca-se o bairro de Mangabeira que, apesar de ser um dos bairros com maior número de ocorrências de homicídios, apresentou uma taxa de incidência abaixo da média do município. Bairros que apresentaram um número baixo em relação a alguns bairros do município, Ilha do Bispo, Altiplano Cabo branco e Distrito Industrial, apresentaram taxas de incidências preocupantes durante esses quatro anos. Bairros como Centro, São José, Varadouro e Valentina, como no número de ocorrências, continuaram com o padrão alto de homicídios mostrado pela taxa de incidência.

A matriz de proximidade espacial foi calculada baseada em distâncias variáveis de acordo com o perímetro do bairro. Para o município de João Pessoa, se fosse gerado uma malha de poucos pontos, a matriz de proximidades seria calculado como mostra a Figura 8:



**Figura 8** - Cálculo da matriz de proximidade espacial

Foi gerada uma grade de pontos, mostrada pelos pontos verdes e armazenado os centróides dos bairros (pontos vermelhos). O raio da circunferência foi calculado com base no perímetro de cada bairro. Para definir se os pontos estão próximos ou não do bairro basta verificar se ele está dentro da circunferência referente ao seu centroide. De uma forma intuitiva, basta contar quantos pontos estão entre as interseções das circunferências para atribuir os pesos.

No exemplo da Figura 8, dos pontos que estão próximos ao bairro do Bessa existe 5 pontos que também estão próximos ao bairro do Altiplano, 8 pontos próximos ao Jardim Oceania e nenhum ponto próximo a Manáira. Logo, esses serão os valores de proximidades para esses três bairros em relação ao bairro do Bessa. Assim, sucessivamente pode ser dado os valores de proximidades para todos os outros bairros do município.

Foi gerada uma malha com 10000 pontos, parte da matriz de proximidade espacial gerada é mostrada abaixo:

171	0	0	0	0	0	0	0	0	18
0	160	60	88	0	0	0	0	0	0
0	60	162	133	64	0	13	0	0	0
0	88	133	211	66	0	18	0	0	0
0	0	64	66	170	46	108	0	0	0
0	0	0	0	46	56	56	0	0	0
0	0	13	18	108	56	569	48	12	180
0	0	0	0	0	0	48	97	41	97
0	0	0	0	0	0	12	41	47	47
18	0	0	0	0	0	180	97	47	938

Aplicando a estatística de Moran ( $I$ ) para os dados de homicídios no município de João Pessoa, foram obtidos os índices de autocorrelação mostrados na Tabela 2.

**Tabela 2** - Índice de autocorrelação de Moran

Ano	<i>I</i>	<i>p-valor</i>
2007	0,1289	0,0594
2008	0,1231	0,0792
2009	0,1120	0,0990
2010	0,0667	0,5446

Fonte: Calculado com base nos dados brutos do CIOP e IPC

Para todos os anos é mostrado um valor positivo da estatística *I*, o que diz que existe autocorrelação de homicídios para os bairros no município, o número de ocorrências de homicídio em um bairro tende a ser semelhante aos bairros vizinhos, mas pelo *p-valor*, considerando  $\alpha = 0,05$ , essas autocorrelações não são significativas. Logo, em geral, as ocorrências de homicídios são distintas para cada bairro. Considerando esse mesmo nível de significância o valor da estatística *G* é significativo apenas nos anos 2007 e 2008 como mostra a Tabela 3, indicando a presença de bairros correlacionados para esses dois anos.

**Tabela 3** - Índice de autocorrelação de Getis&Ord

Ano	<i>G</i>	<i>p-valor</i>
2007	0,0252	0,0166
2008	0,0237	0,0199
2009	0,0206	0,0897
2010	0,0220	0,1827

Fonte: Calculado com base nos dados brutos do CIOP e IPC

No Apêndice III, temos o diagrama de emparelhamento de Moran, que é um gráfico onde no eixo x mostra os valores dos desvios do atributo em relação a média e o eixo y mostra os valores das médias locais do atributo para os 4 anos em questão. As cores vermelha e laranja representadas no mapa, mostra os bairros que se situam entre o primeiro e o segundo quadrante indicando bairros com alta correlação espacial positiva, indicando que uma localização possui vizinhos com valores semelhantes. Nele, se encontra uma maior concentração de bairros a noroeste da cidade com a cor vermelha e na região nordeste da cidade encontra-se uma maior concentração de bairros representados pela cor laranja. Os bairros representados pelas cores mais claras são os bairros que se situam no 3º e 4º quadrante do diagrama, mostrando bairros com associação espacial negativa, ou seja, vizinhos com valores distintos. Também é possível observar que ao passar dos anos o número de bairros com associação negativa vem aumentando, isso pode ser explicado pelo fato que o número de bairros com ocorrências nulas, que existia em boa parte da cidade, está diminuindo, logo a cidade está passando a ter um padrão de homicídios cada vez mais heterogêneo.

O Apêndice IV mostra o índice local de autocorrelação espacial de Moran. Os bairros que permaneceram com auto correlação significativa durante os 4 anos foram Varjão e Cristo Redentor, indicando que esses tem valores semelhantes entre si e diferenciados possível observar claramente que o número de bairros com significância espacial vem diminuído durante dos seus vizinhos. É esses quatro anos.

O resultado índice  $G_i$  e  $G_i^*$  é mostrado no Apêndice V e VI, respectivamente, onde observa-se que os bairros: Mangabeira, Cristo Redentor e Mandacarú obtiveram valores positivos e significantes durante os quatro anos, sendo que Mangabeira para 2007 e 2008, Cristo Redentor em 2009 e Mandacarú obtiveram  $G_i$  e  $G_i^*$  com p-valor abaixo de 0,005, ou seja, esses bairros estão cercado por uma vizinhança de valores baixos para a região. Além desses três bairros apareceram com índices positivos e significativos o bairro do Oitizeiro em 2007 e Valentina em 2008 e 2009.

Os mapas com os valores das estatísticas Scan são apresentados no Apêndice VII. O mapa foi construído admitido que os conglomerados não poderiam superar 10% da população e que o nível de significância, com base nos mapas de riscos, deveria ser de  $\alpha=0,05$ . Os mapas mostram a presença de conglomerados na parte noroeste do município sendo que esse vem diminuindo sua extensão ao decorrer dos anos e em 2010 aparecem Gramame e Valentina como conglomerados. Observa-se também conglomerados que persistiram durante todo o período com os bairros: Varadouro, Roger, Treze de Maio, Mandacarú, Ipês, João Agripino, São José e Bairro das Indústrias.

No Apêndice VIII encontra-se os mapas mostrando os bairros com valores da estatística Besag&Newell significativos com o centróide na cor vermelha. E ele se mostrou similar ao método de varredura Scan, acusando a presença de conglomerados na região noroeste do município, sendo que este inclui um número menor de municípios.

Do ponto de vista epidemiológico, já que o alvo maior é localizar conglomerados de bairros com maior risco de homicídios, o método de estatística scan, para esse caso, se mostrou superior a todos os outros, identificando mais precisamente conglomerados de valores elevados.

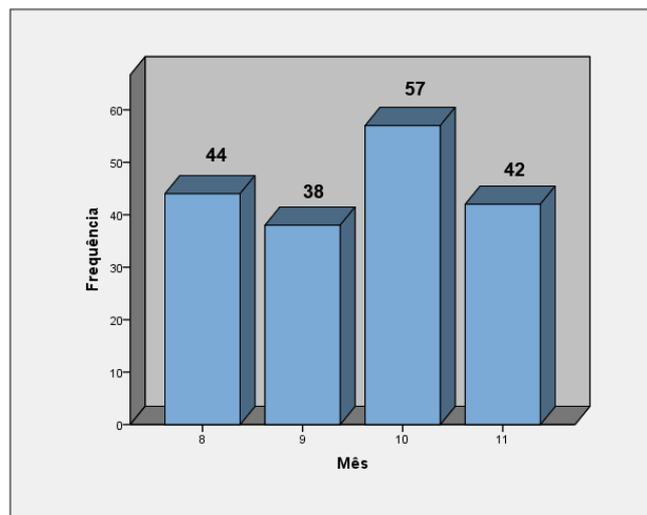
## 4.2 Homicídios ocorridos entre agosto e novembro de 2011

Os dados utilizados nessa análise são referentes a informações registradas pelo CIOP, por ter horário e data mais próximo do fato ocorrido. Por falta de algumas informações do CIOP, o banco de dados foi complementado com informações registradas pelo IPC.

Foram consideradas apenas as informações de data e hora registradas pelo CIOP, apesar de ter esse registro apenas em 75% das informações no banco de dados montado. As informações de data e hora do IPC não foram consideradas por haver uma diferença mediana de 3 horas e 54 minutos a mais que os dados do CIOP com desvio padrão de 5 horas e 29 minutos, ou seja, os horários registrados pelo CIOP são mais próximos do fato ocorrido.

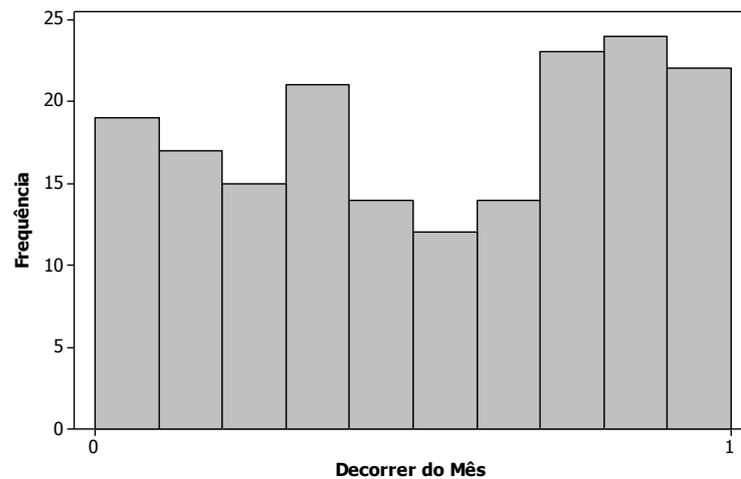
### 4.2.1 Análise Descritiva

Durante os meses de agosto, setembro, outubro e novembro do ano de 2011 foram registrados 181 casos de homicídios, sendo que o maior número de ocorrências foi no mês de Outubro (57 homicídios). Os valores mensais de ocorrência pode ser observado no Gráfico 2.



**Gráfico 2:** Frequência de Homicídios por Mês

Durante esse período ocorreram em média 1,48 homicídios por dia, ou seja, um homicídio a cada 16 horas e 11 minutos. Para se ter uma melhor noção da distribuição de homicídios durante o mês foi montado o histograma mostrado no Gráfico 3, como os meses em questão não tem o mesmo número de dias foi necessário proporcionar os dias das ocorrências.



**Gráfico 3** – Histograma de homicídios para o período dos meses

**Quadro 2** - Teste Qui-Quadrado para o período do mês

Qui-Quadrado	8,659
Grau de Liberdade	9
p-valor	0,469

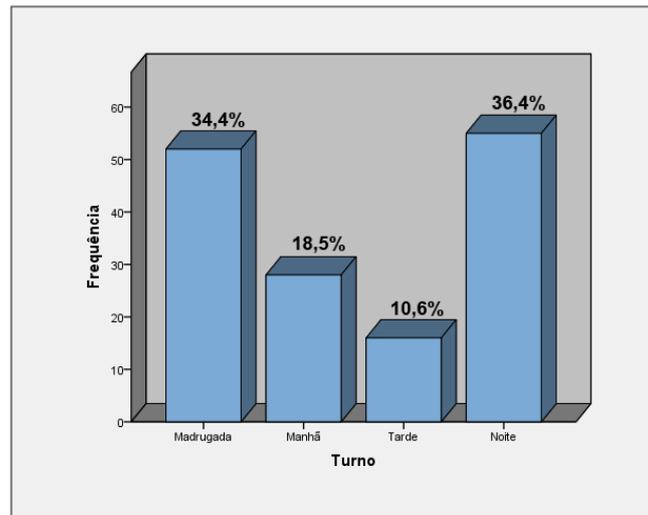
**Fonte:** Calculado com base nos dados brutos do CIOP e IPC

Observa-se no histograma que no final dos meses a frequência de ocorrência é maior e no meio essa frequência mais baixa, principalmente no início da segunda metade, dando a impressão que é mais fácil haver homicídios no final de mês, mas o teste qui-quadrado, na Quadro 2, mostra, com 95% de confiança, que essas diferenças não são significativas. Esse teste é adequado quando temos uma variável dividida em várias categorias e queremos testar se todas as categorias contêm a mesma proporção de ocorrência do evento, ou se cada categoria tem uma proporção especificada, através da comparação entre os valores esperados e os valores observados de cada categoria.

Com as informações dos horários das ocorrências de homicídios que foram registrados pelo CIOP, foram criados os seguintes turnos:

- 1- Madrugada - Das 00h00min às 06h00min;
- 2- Manhã - Das 06h00min às 12h00min;
- 3- Tarde - Das 12h00min às 18h00min;
- 4- Noite- Das 18h00min às 24h00min.

No Gráfico 4 mostra a frequência de ocorrências durante esses turnos.



**Gráfico 4** – Frequência de homicídios por turno

Observa-se que o turno da tarde é menos propenso a ocorrência de homicídios, pois teve a menor frequência. Já os turnos noite e madrugada foi onde mais ocorreram homicídios (70,8%). O valor da estatística  $\chi^2$  (qui-quadrado) no Quadro 3 mostra que essas frequências são distintas (usando  $\alpha=0,05$ ).

**Quadro 3** - Teste Qui-Quadrado para os turnos

Qui-Quadrado	28,311
Grau de liberdade	3
p-valor	0,00

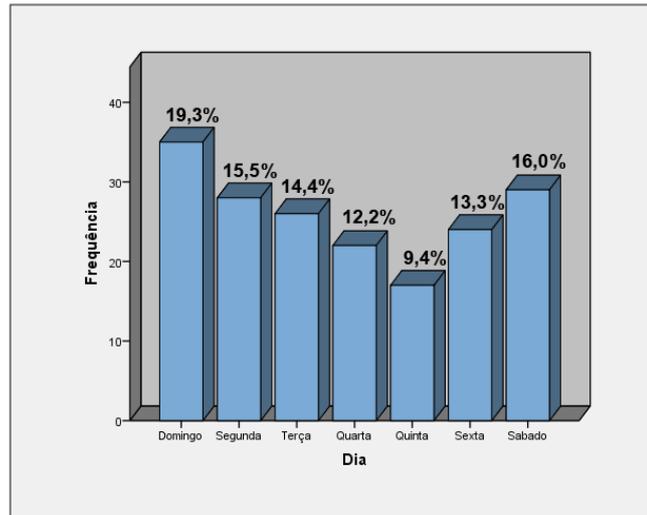
**Fonte:** Calculado com base nos dados brutos do CIOP e IPC

A frequência de homicídios ocorridos durante esses quatro meses estudados, onde observa-se que os dias em que mais ocorrem homicídios é no domingo (19,3%), no sábado (16%) e na segunda (28%) e o dia que foi obtido uma menor frequência foi a quinta-feira (9,4%). O teste  $\chi^2$  mostra que essas diferenças não são significantes, ou seja, a ocorrência de homicídios durante o dia da semana se distribui uniformemente.

**Tabela 4** – Frequência de homicídios por dia da semana

Dia	Dom	Seg	Ter	Qua	Qui	Sex	Sáb	Total
Frequência	35	28	26	22	17	24	29	181

**Fonte:** Calculado com base nos dados brutos do CIOP e IPC



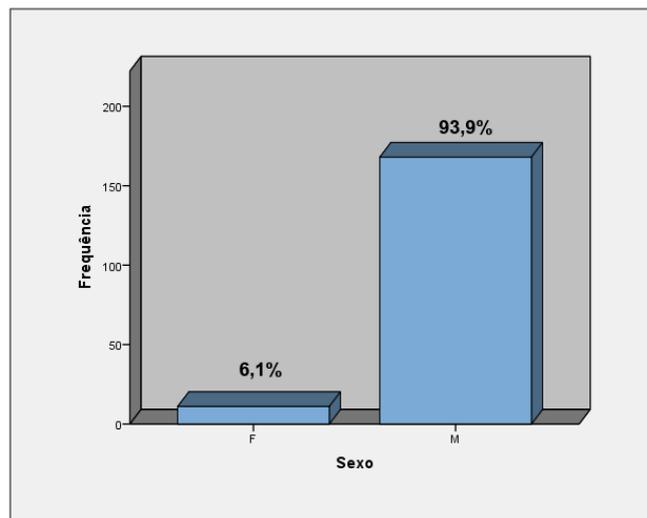
**Gráfico 5** – Frequência de homicídios por dia da semana

**Quadro 4** - Teste qui-quadrado para os dias da semana

Qui-Quadrado	7,536
Grau de Liberdade	6
p-valor.	0,274

**Fonte:** Calculado com base nos dados brutos do CIOP e IPC

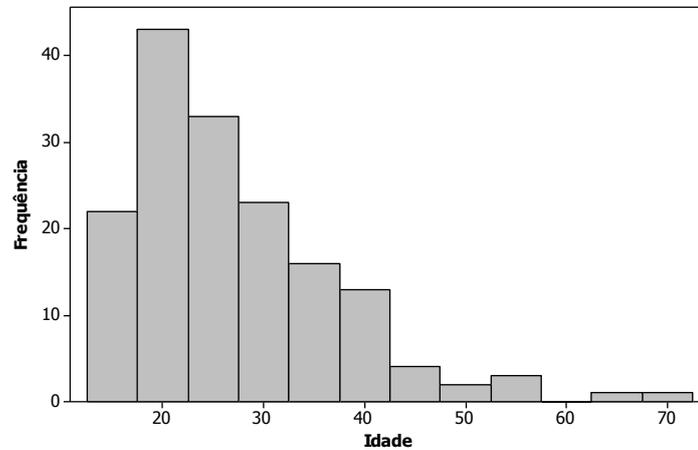
Durante todos os quatro meses foram registrados apenas 6,1% homicídios ocorridos com vítimas do sexo feminino, deixando evidente que o sexo masculino são mais propenso a ser vítima desse caso.



**Gráfico 6** – Frequência de homicídios por sexo

A Tabela 5 mostra que pelo menos metade das vítimas tem idade inferior a 24 anos e o desvio padrão mostra que existe uma grande variação em torno da média. As vítimas tem

idade mais frequente em torno de 20 anos e vítimas com idade mais elevadas, como já era de se esperar, são menos frequente.



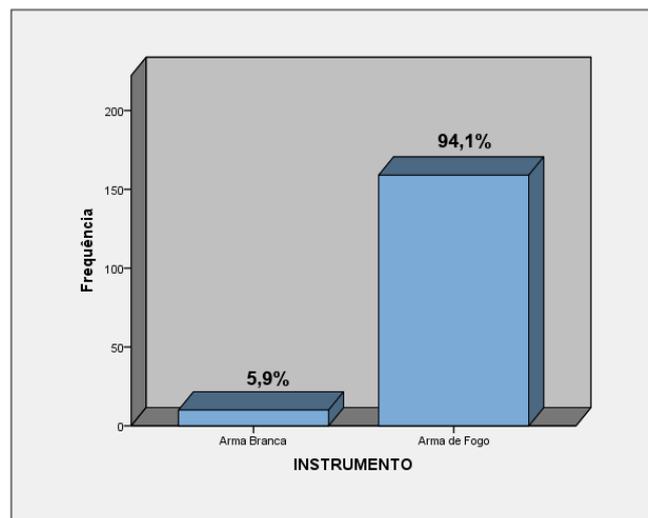
**Gráfico 7 - Histograma da Variável Idade**

**Tabela 5 - Medidas Descritivas da Variável Idade**

Variável	Média	StDev	Mínimo	Q1	Mediana	Q3	Máximo	Amplitude
<b>Idade</b>	27,161	10,328	13	19	24	32,5	70	57

Fonte: Calculado com base nos dados brutos do CIOP e IPC

O Gráfico 8 mostra a frequência de homicídios por tipo de arma, onde observa-se a grande predominância de homicídios cometidos por armas de fogo (94,1%).



**Gráfico 8 – Frequência de homicídios por tipo de arma**

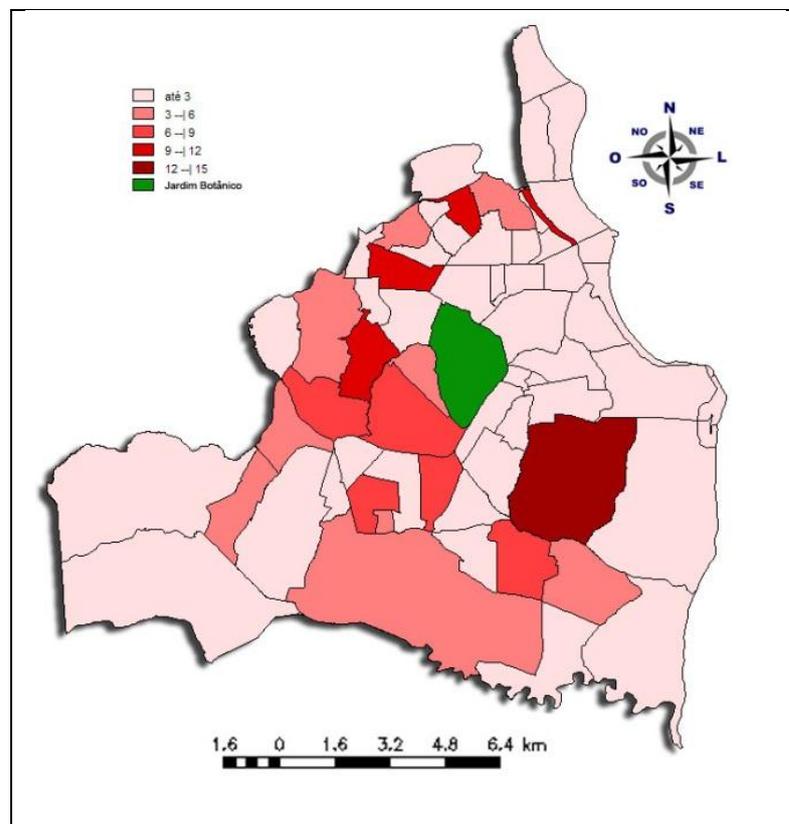
Os Bairros com o maior número de homicídios registrados estão sendo mostrados na Tabela 6.

**Tabela 6** – Frequência de Homicídios por Bairro

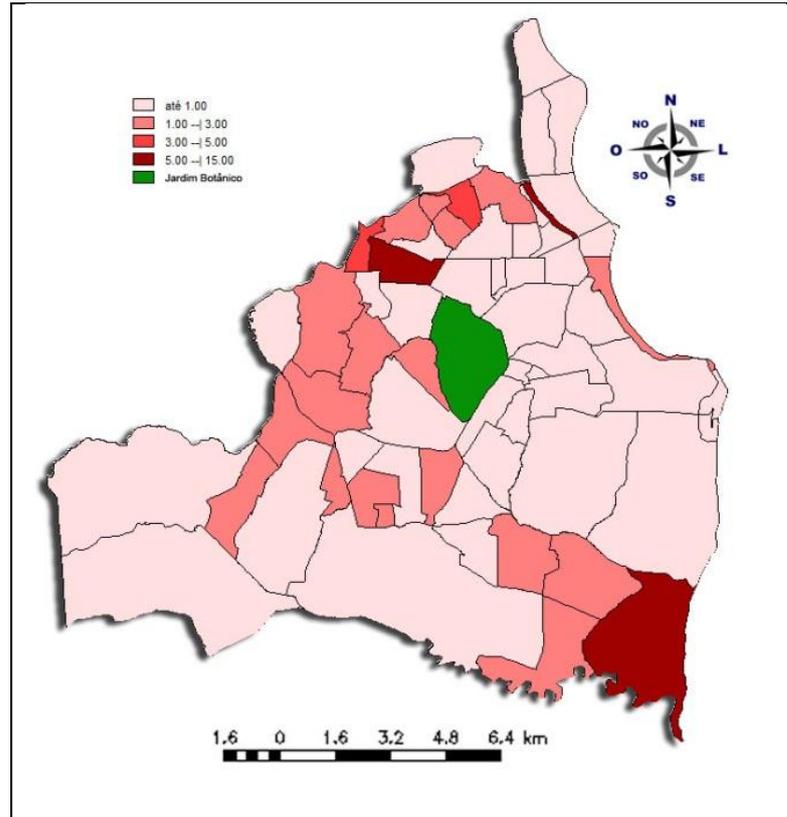
Posição	Bairro	Frequência	Percentual
1º	MANGABEIRA	15	8,1
2º	MANDACARÚ	11	5,9
3º	SÃO JOSÉ	11	5,9
4º	CENTRO	10	5,4
5º	CRUZ DAS ARMAS	10	5,4

Fonte: Calculado com base nos dados brutos do CIOP e IPC

A frequência de homicídios para os outros bairros do município pode ser visualizada na Figura 9.

**Figura 9** - Ocorrência de homicídios

Além dos cinco bairros mencionados na Tabela 3, observam-se ainda com números de ocorrências elevados os bairros Funcionários, Cristo Redentor, Ernesto Geisel, Oitizeiro, Valentina, Bessa e Roger.



**Figura 10** –Riscos relativos de homicídios

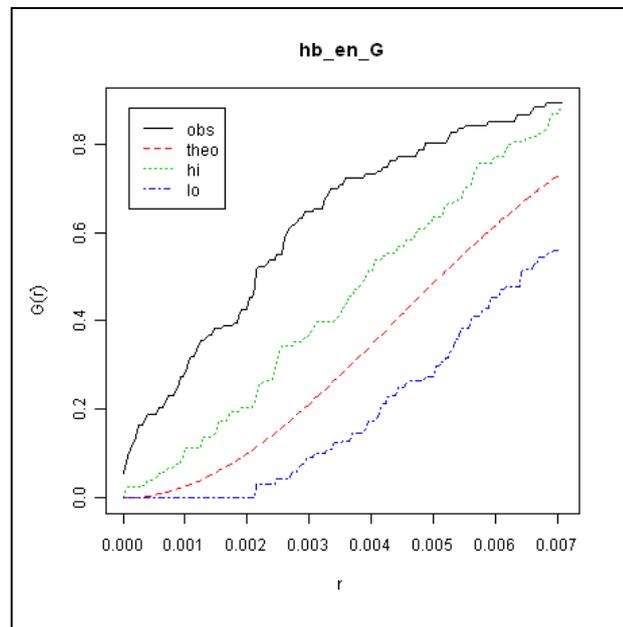
A Figura 10 mostra os riscos relativos. Durante esses quatro meses os bairros que apresentaram um maior risco, representados no mapa pela cor mais forte, são Barra de Gramame, Centro e São José. Alguns bairros, como Mangabeira e Cristo Redentor, que apresentaram um valor auto de ocorrência, devido existir um número superior de moradores no local, foi obtido um risco relativo abaixo do risco do município. Já outros, como Barra de Gramame e Muçumagro, apesar de haver poucas ocorrências registradas, foi obtido um auto risco relativo por esses bairros serem pouco habitados.

#### 4.2.2 Análise Pontual

Foram usados dados pontuais de ocorrências de homicídios obtidos através de análise de dados fornecidos pelo CIOP e IPC, para os meses de agosto, setembro, outubro e novembro. Devido a não clareza das informações na descrição das localizações das ocorrências, foi possível obter a localizações pontuais de apenas 80% das ocorrências registradas.

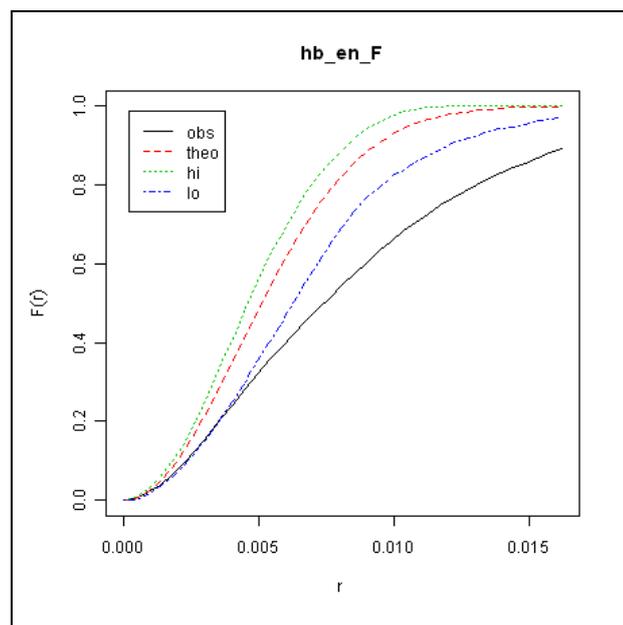
Primeiramente foram aplicadas as técnicas de 2ª ordem para verificar a presença de aglomerados espaciais no município de João Pessoa e em seguida, confirmado a presença de

clusters, foi calculado o estimador de densidade de kernel para verificar os lugares de maior intensidade de homicídios. O gráfico da função  $G$  é mostrado no Gráfico 9.



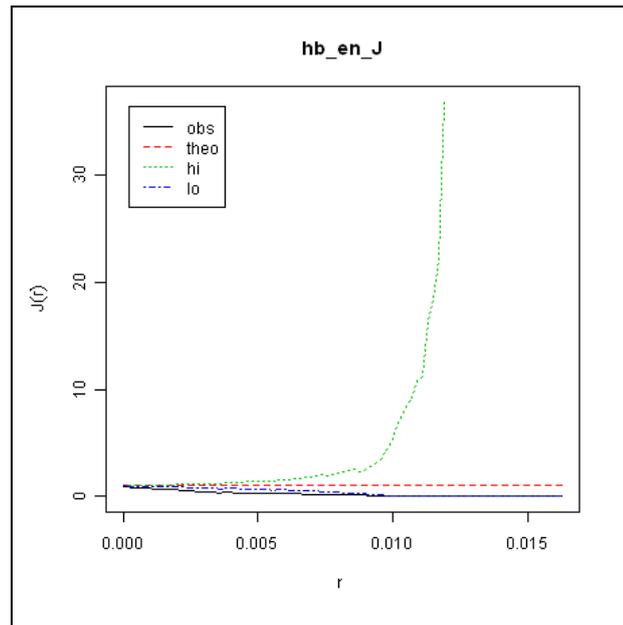
**Gráfico 9:** Função  $G$

A função  $G$  mostra que os valores da função para os pontos observados (linha preta) estão todos acima do envelope superior representado pela curva pontilhada da cor verde. Desde as menores distâncias já foi obtido uma frequência bem superior a frequência esperada. Isso indica a presença de conglomerados na região e conseqüentemente os homicídios não são distribuídos de forma aleatória na região, existem áreas mais propensas a essa ocorrência. Isso também é confirmado pelo gráfico  $F$  na Gráfico 10, onde observa-se que os valores de  $F$  observados estão abaixo do limite do inferior (linha azul) dado pelo envelope.



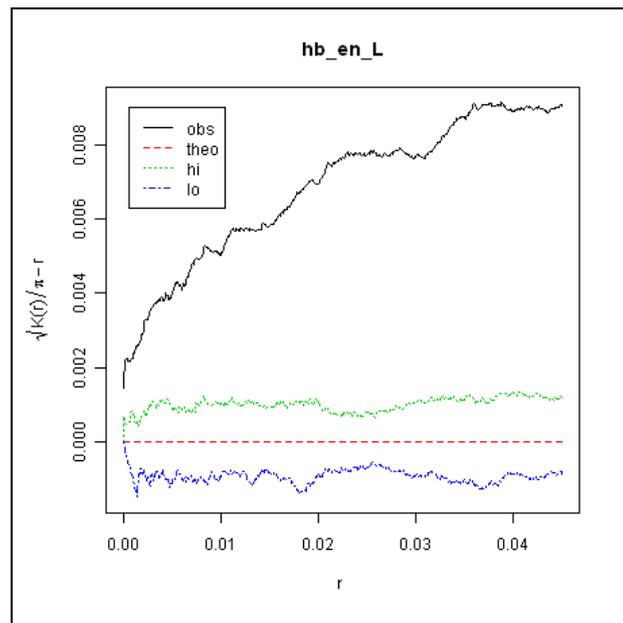
**Gráfico 10:** Função  $F$

Apesar de estar um pouco mais difícil de enxergar, pelo limite superior do envelope ser muito alto para as distâncias maiores, o Gráfico da função  $J$ , representado no Gráfico 11, mostra que os valores da função  $J$  observada, para as baixas distâncias, são inferiores significativamente a os valores estimados se o processo estivesse sob CSR, indicando a existência de conglomerados no município.



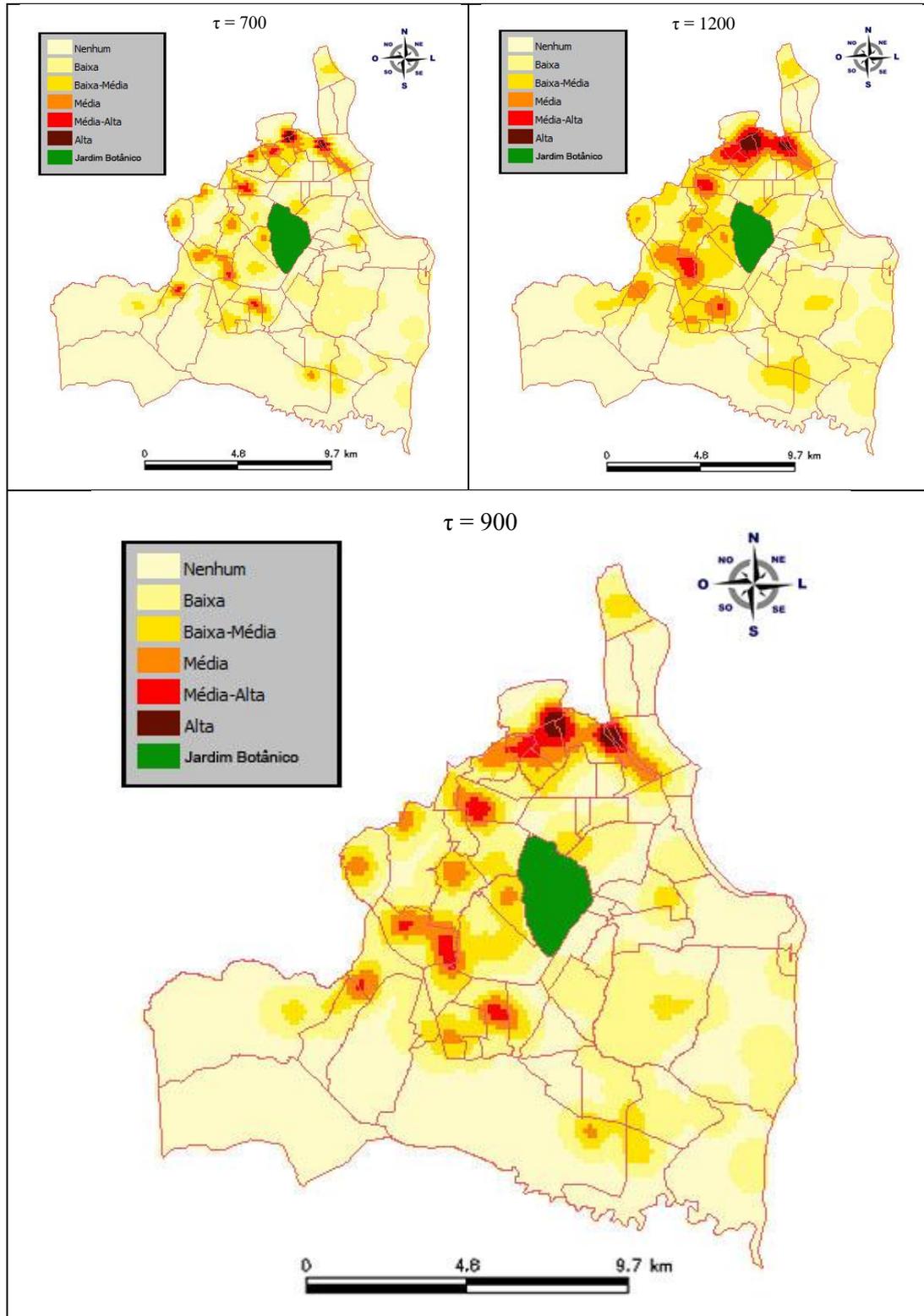
**Gráfico 11:** Função  $J$

A função  $L$  não deixa duvida na presença de conglomerados, para todas as distâncias o valor da função  $L$  para os dados observados foi superior ao limite do envelope estimado.



**Gráfico 12:** Função  $L$

Confirmado a presença de conglomerados espaciais no município resta saber onde eles estão localizados, isso é possível pelo estimador de intensidade de kernel, como os métodos acusaram indício de conglomeração para todas as distâncias, foi variada o valor da largura da banda de modo a obter um mapa que facilite interpretação e isso é mostrado na Figura 11.



**Figura 11** - Estimador de intensidade de Kernel

Foi feito o estimador de Kernel com diferentes larguras de banda e foi observado que ao aumentar o valor de  $\tau$  os valores de incidência vai aumentando pela região. Foram obtidas regiões de alta incidência na região norte do bairro São José e bairros que ligam essa região: Bairro dos Ipês, João Agripino, e parte de oeste de Manaíra.

Apresentou ainda um alto índice na parte sudeste do Alto do Céu, Nordeste de Mandacarú e Noroeste do Bairro dos Ipês. Essas foram as regiões com a maior incidência de homicídios.

### 4.3 Previsão para 2012 por Alisamento Exponencial de Holt

Foram utilizados o número anual de ocorrências registrados pelo Ciop partindo de 2001 e verificado que essa série apresenta uma tendência crescente durante o tempo e não possui sazonalidade. Com isso, foi escolhida a utilização do modelo de alisamento exponencial de Holt para esse caso. Devido a uma mudança no sistema de cadastro do CIOP durante o final de 2010 e início de 2011 foi perdido algumas informações mensais desse período e foi utilizado esse mesmo método para recupera-las.

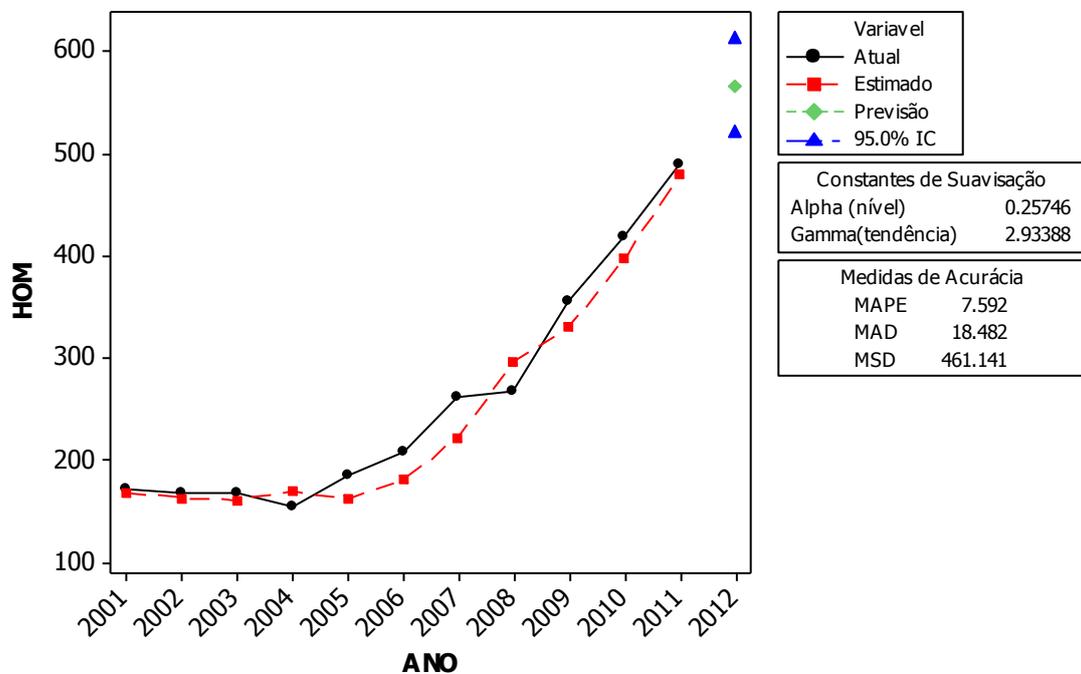


Gráfico 13 – Alisamento Exponencial de Holt

**Tabela 7 – Valor Estimado para 2012 e Intervalo de Confiança**

ANO	Estimado	Limite Inferior	Limite Superior
2012	566,446	521,166	611,725

Fonte: Calculado com base nos dados brutos do CIOP e IPC

No Gráfico 12, é observado pelos pontos pretos a série de número de registros de homicídios anuais pelo CIOP, a linha tracejada vermelha são os valores estimados pelo método, o ponto verde mostra o valor estimado e os pontos azuis mostra um intervalo de 95% de confiança para essa estimação. Também são mostrados os valores das constantes de suavizações ótimos obtidos e medidas de acurácia para esse modelo. O valor estimado para 2012 também é mostrado na Tabela 7, onde mostra que é previsto pelo modelo que 566 casos de homicídios sejam registrados no ano de 2012. O que equivale um aumento de 15,5% de ocorrências.

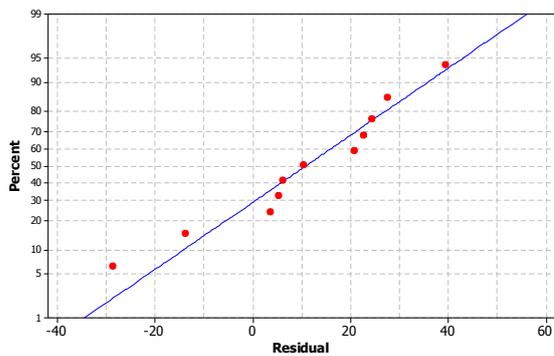


Gráfico 14 – Ajuste de Normalidade para o Alisamento Exponencial de Holt

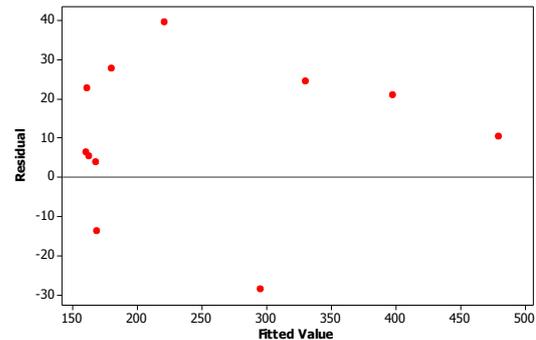


Gráfico 15 - Resíduos versus Valores estimados para o Alisamento Exponencial de Holt

Como se observa no gráfico no gráfico 14 e 15 a normalidade e homocedasticidade dos resíduos para a série estimada pelo modelo de Holt é satisfeita.

## 5 CONCLUSÃO

Na análise por Área feita durante o período de 2007 a 2010 foi possível destacar as áreas de maior risco relativo de homicídios. Apesar de existirem áreas com um elevado número de homicídios, o risco relativo mostrou que esse é um padrão já esperado e destacou como tendo os maiores riscos os bairros: Centro, São José, Varadouro e Valentina.

Tanto o Índice *G* como o índice *I* rejeitaram a hipótese de aglomeração espacial no município para os dados de Área. Já o método de varredura scan e Besag&Newell obtiveram padrões similares e identificaram a presença de conglomerados espaciais na região noroeste e sudeste.

Apesar de muitos estudos enfatizarem o número de homicídios ocorridos na idade mais jovem, durante os meses de agosto, setembro, outubro e novembro observou-se que as vítimas não eram tão jovens assim, mesmo o teste qui-quadrado rejeitando a hipótese de igualdade de frequência entre as idades, já era de se esperar esse padrão, pela quantidade de população nessas faixas de idade. A maioria dos homicídios tem como vítima pessoas do sexo masculino (93,9%), são cometidos por arma de fogo (94,1%) e ocorrem durante o turno da noite e madrugada (70,8%).

A análise pontual dos últimos meses decorridos, para todas as técnicas utilizadas, verificou-se que no município existem regiões mais propensas a ocorrência de homicídios. Pelos estimadores de 2ª ordem e o estimador de intensidade de Kernel mostrou a intensidade de ocorrências distribuída em todo o município, onde observou-se que na maior parte do seu território a intensidade é baixa e em apenas alguns pontos, principalmente na zona norte, chega a ter uma taxa muito elevada.

Durante os dez últimos anos foi verificado um crescimento do número de registros de homicídio pelo CIOP de 185% e foi estimado pelo modelo de alisamento exponencial de Holt um crescimento de 15,5% de ocorrências em relação a 2011, um resultado no mínimo preocupante para o município, o que leva a necessidade de tomada de atitudes para que isso não chegue a ocorrer.

## 6 REFERÊNCIAS

Almeida, E. S., Haddad, E. A. e Hewings, G. J. D. (2005). The Spatial Pattern of Crime in Minas Gerais: An Exploratory Analysis. *Revista de Economia Aplicada*, 9(1).

Anselin, L. (1992). *SPATIAL DATA ANALYSIS WITH GIS: AN INTRODUCTION TO APPLICATION IN THE SOCIAL SCIENCES* - National center for Geographic Information and Analysis, University of California, Technical Report 92-10

Anuário Brasileiro de Segurança Pública, 5ª Ed, 2011, Copyright, Fórum Brasileiro de Segurança Pública, ISSN 1983-7634, disponível em <http://www2.forumseguranca.org.br/node/26939>, acessado em dezembro de 2011.

COSTA, M.A.; ASSUNÇÃO, R.M. A fair comparison between the spatial scan and the Besag-Newell disease clustering tests. *Environmental and Ecological Statistics*, Springer, v.12, p. 301-319, 2005.

CRESSIE, N. A. C. *Statistics for spatial data revised edition* – Wiley series in probability and mathematical statistics. New York: John Wiley & Sons, Inc., 1993.

Druck, S.; Carvalho, M.S.; Câmara, G.; Monteiro, A.V.M. (eds) "Análise Espacial de Dados Geográficos". Brasília, EMBRAPA, 2004 (ISBN: 85-7383-260-6). Disponível em <http://www.dpi.inpe.br/gilberto/livro/analise/>, Acessado em dezembro de 2011.

FURTADO, P. *Análise Espacial e Classificação de Imagens Aplicadas ao Estudo do Dengue em Relação às Áreas de Vegetação na Cidade de João Pessoa* – PB. Dissertação de Mestrado. 2008.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Sistema IBGE de Recuperação Automática – SIDRA, home Page: [www.ibge.gov.br/](http://www.ibge.gov.br/) Acessado em outubro de 2011.

Johnson, R. A.; Wichern, D. W.; *Applied Multivariate Statistical Analysis*, Prentice-Hall, 4th edition, 1998.

KERSCHER, M.; PONS-BORDERIA, M. J.; SCHMALZING, J.; TRASARTI -BATTISTONI, R.; BUCHERT, T.; MARTINEZ, V. J. and VALDARNINI, R. A global descriptor of spatial pattern interaction in the galaxy distribution. *The Astrophysical Journal*, v. 513, n. 2, p. 543-548, 1999.

KULLDORFF, M.; NAGARWALLA, N. Spatial disease clusters: detection and inference. *Statistics in Medicine*, v. 14, p. 799-810, 1995.

MORETTIN, P. A; TOLOI, C. M. C., *Análise de Séries Temporais*, 2ª ed, São Paulo, Editora Blucher, 2006, ISBN 978-85-212-0389-6.

Neves, M. C., Ramos, F. R., Camargo, E. C. G., Câmara, G. e Monteiro, A. M. (2000). *Análise Exploratória Espacial de Dados Sócio-Econômicos de São Paulo*. DPI – INPE – Disponível em

[http://www.dpi.inpe.br/gilberto/papers/marcos\\_gisbrasil2000.pdf](http://www.dpi.inpe.br/gilberto/papers/marcos_gisbrasil2000.pdf). Acessado em outubro de 2011.

PREFEITURA MUNICIPAL DE JOÃO PESSOA. Secretaria de Planejamento. Diretoria do Geoprocessamento e Cadastro Urbano (SEPLAN). Mapa de João Pessoa. João Pessoa, 2007.

R DEVELOPMENT CORE TEAM, R: A language and environment for statistical computing, Vienna, Austria: R Foundation for Statistical Computing, [online] 2005. Disponível em: <[http://geosere.ccr.ufsm.br/sites/default/files/sites/default/files/arquivos\\_documentos/refman.pdf](http://geosere.ccr.ufsm.br/sites/default/files/sites/default/files/arquivos_documentos/refman.pdf)>. Acesso em: 06 de dezembro 2011.

Ripley, B.D. (2004). SPATIAL STATISTICS. University of London, Copyright © 1981, John Wiley & Sons, Inc., Hoboken, New Jearsey.

Siegel, Jacob e Swanson (2004). Population Projections (eds.), The Methods and Materials of Demography. San Diego: ElsevierAcademic Press.

The Comprehensive R Archive Network. URL <http://www.cran.r-project.org>

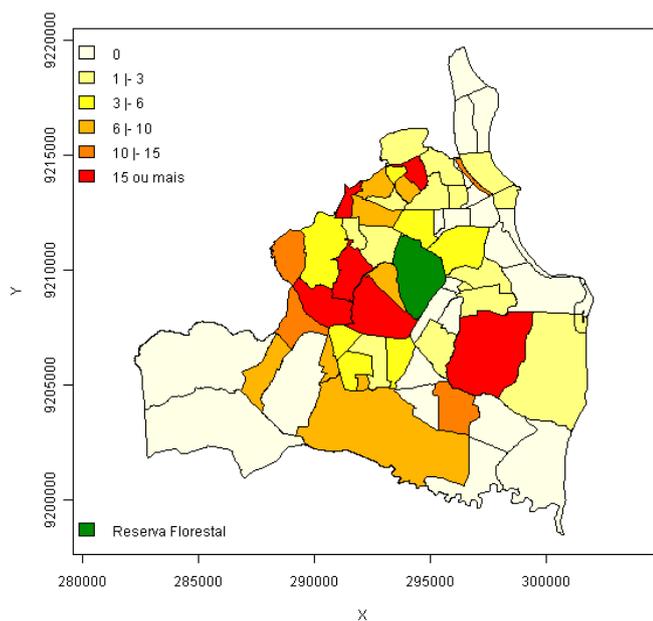
WAISELFISZ, J. J. Juventude, violência e cidadania. Os jovens de Brasília. São Paulo: Cortez/UNESCO, 1998.

WAISELFISZ, J. J. Mapa da violência 2011: Os jovens do Brasil. 1º ed, São Paulo: Instituto Sangari; Brasília, DF: Ministério da Justiça, 2011.

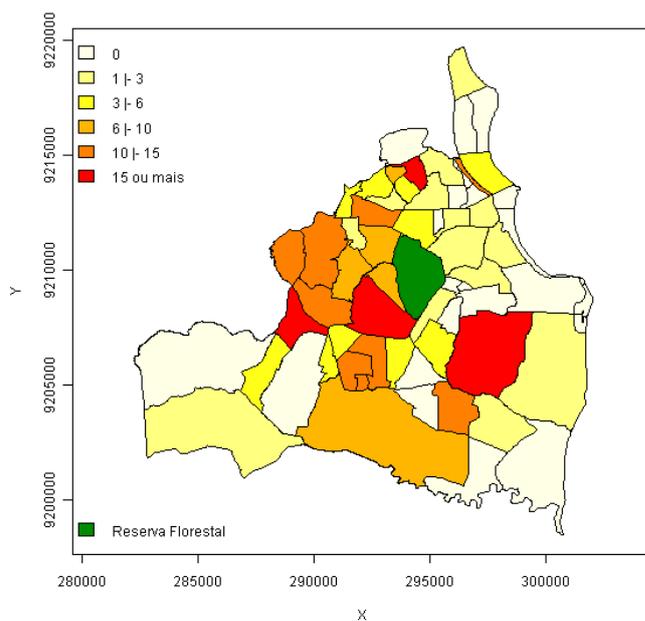
## **APÊNDICES**

## APÊNDICE I – NÚMERO DE OCORRÊNCIA DE HOMICÍDIOS POR BAIRO

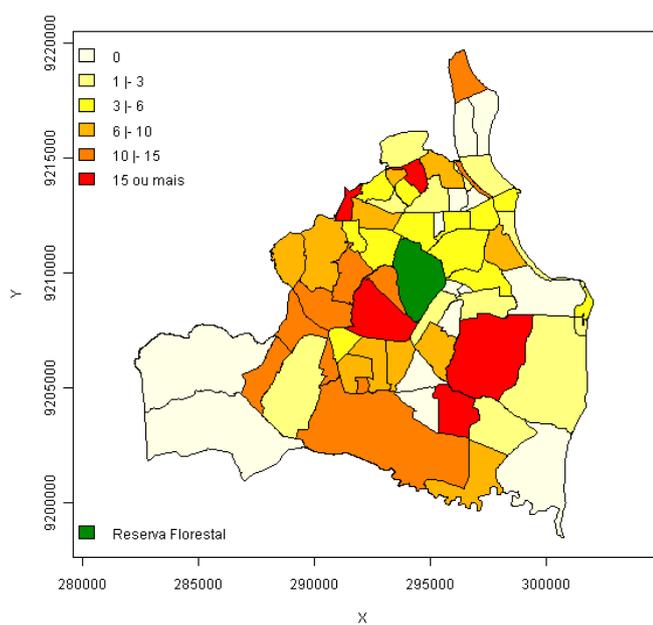
2007



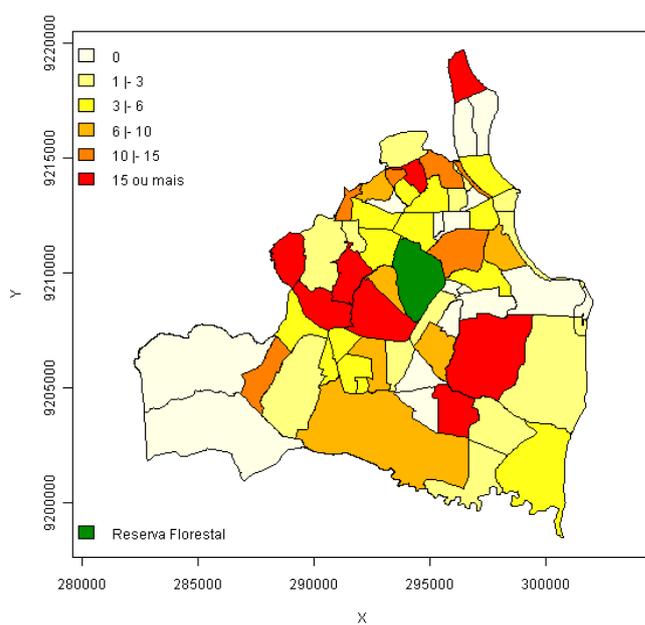
2008



2009

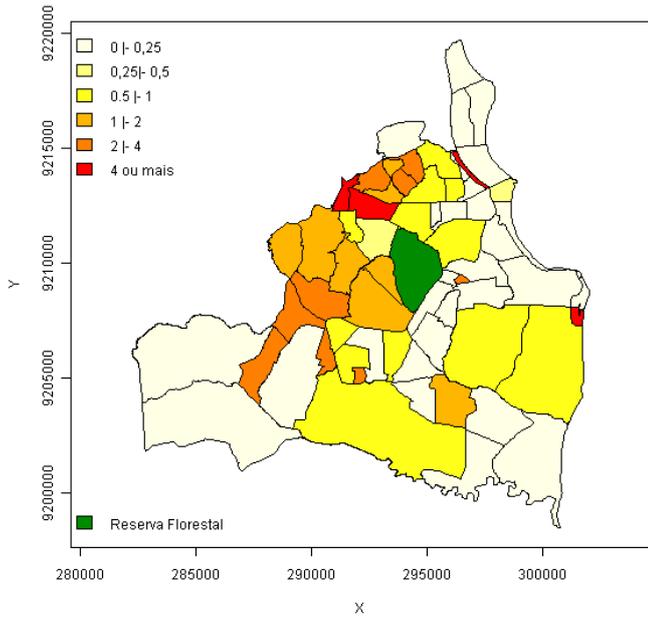


2010

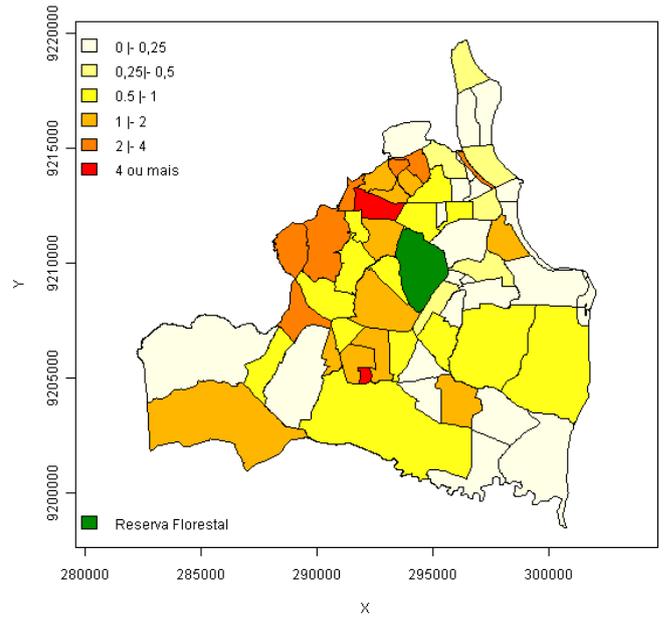


## APÊNDICE II – INCIDÊNCIA DE HOMICÍDIOS POR BAIRRO

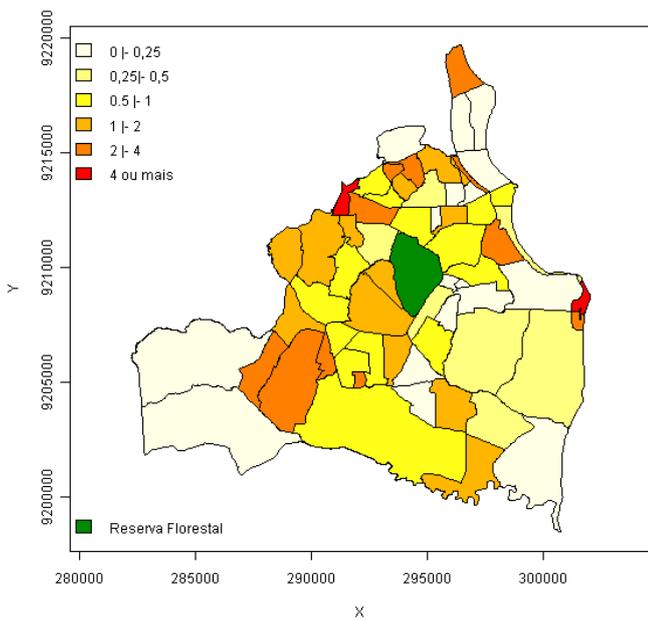
2007



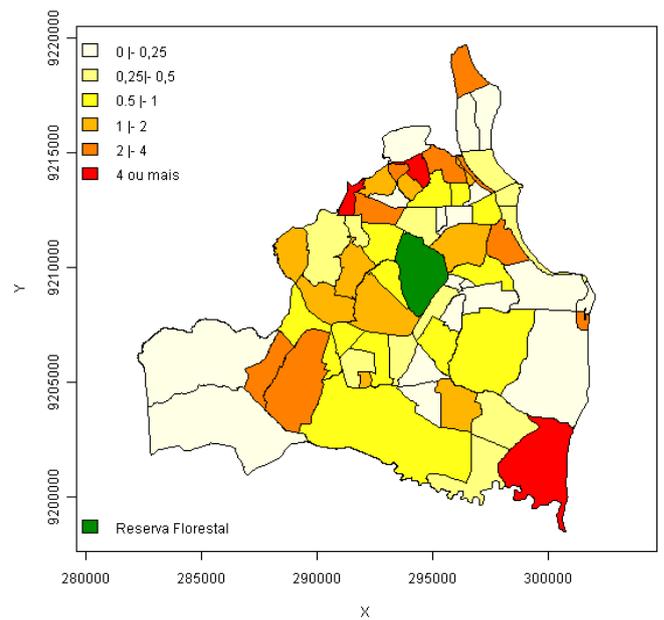
2008



2009

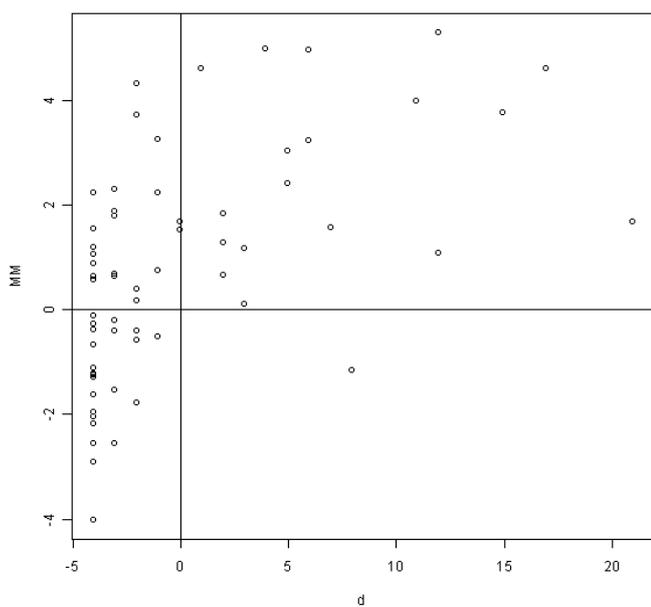


2010

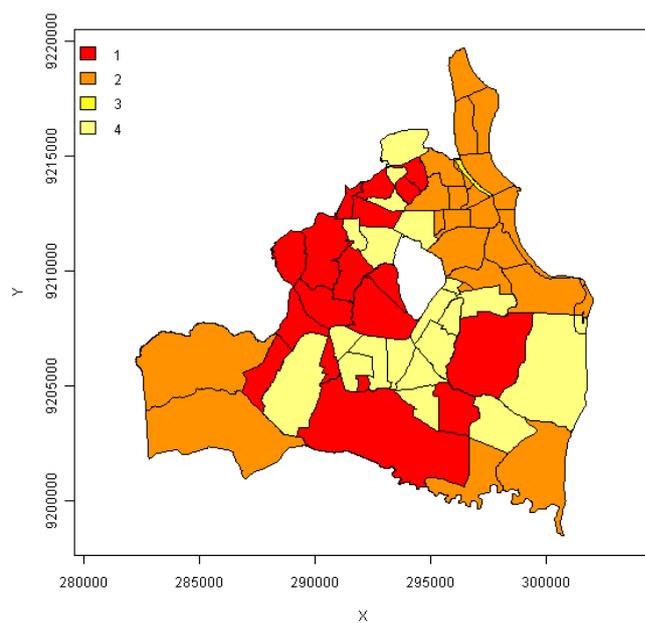


## APÊNDICE III - PARTE 1 – DIAGRAMA DE EMPARALHAMENTO DE MORAN

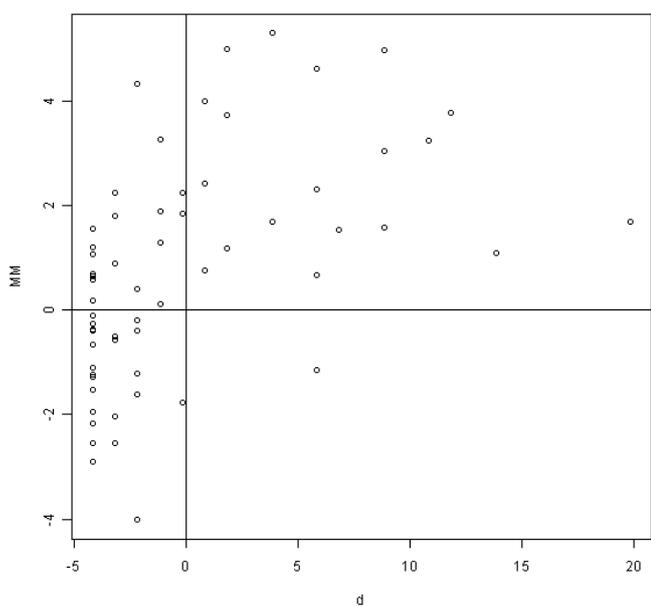
2007



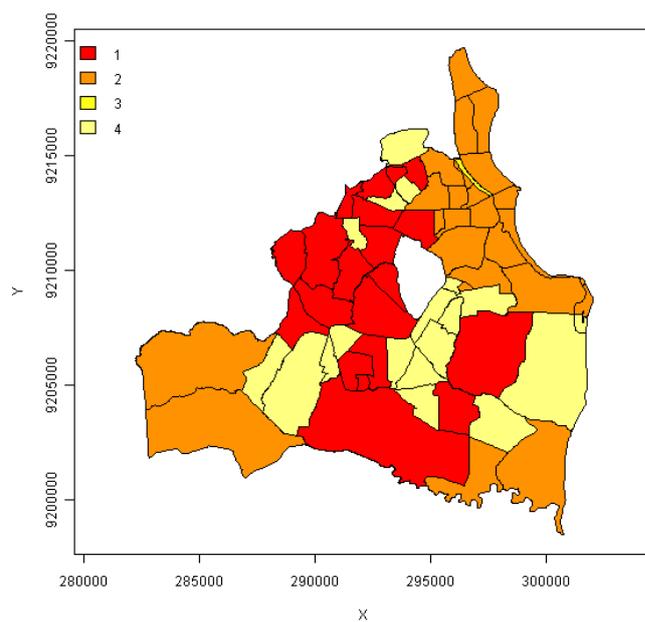
2007



2008

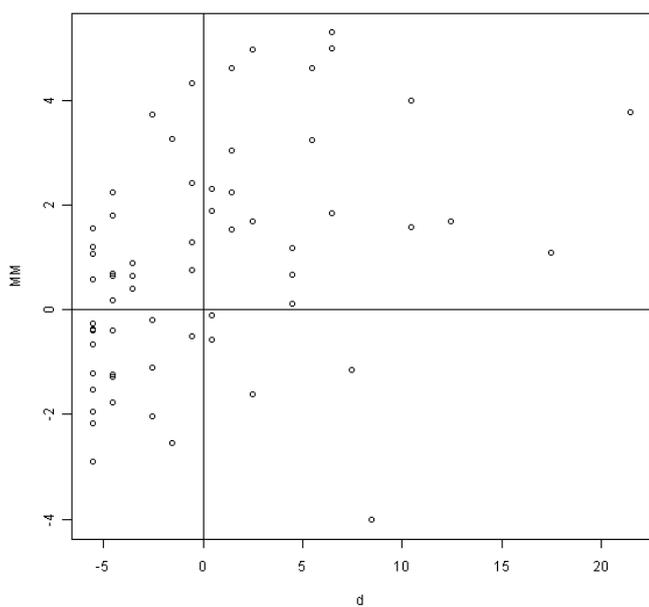


2008

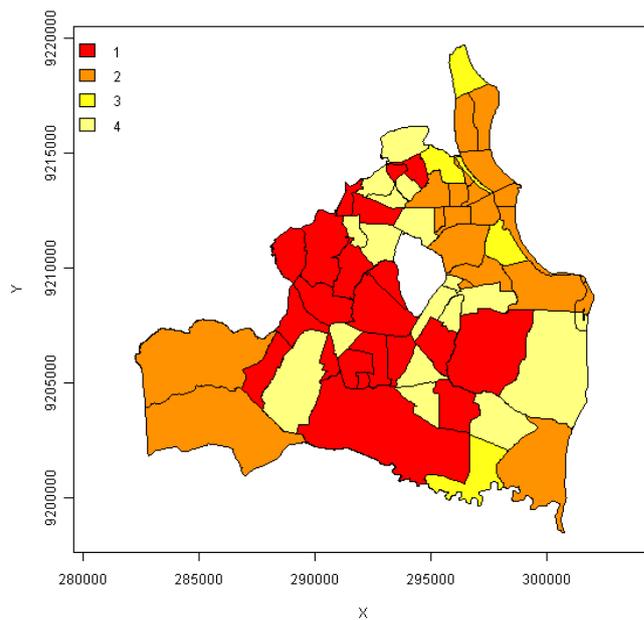


## APÊNDICE III - PARTE 2 – DIAGRAMA DE EMPARALHAMENTO DE MORAN

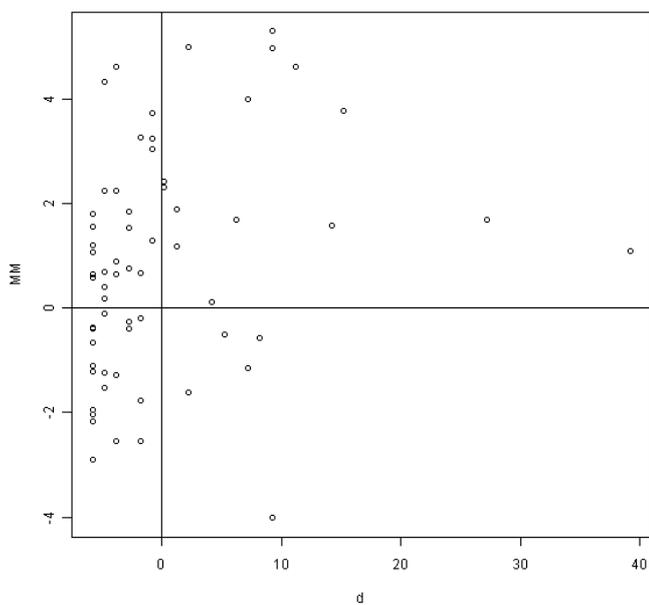
2009



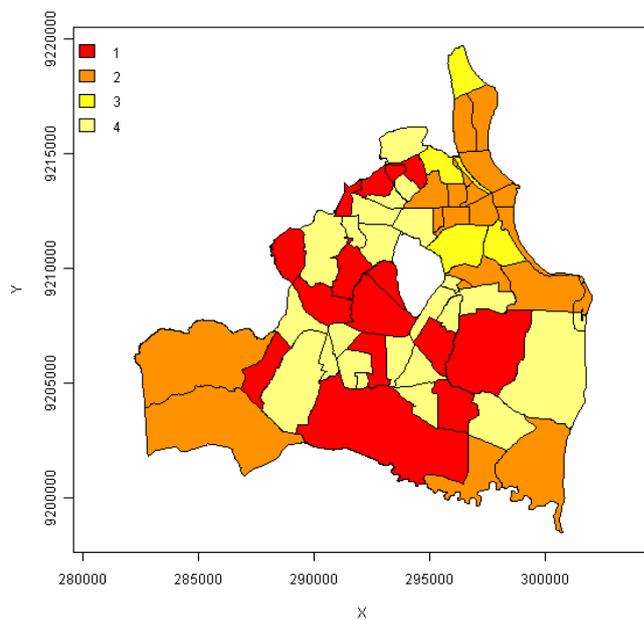
2009



2010

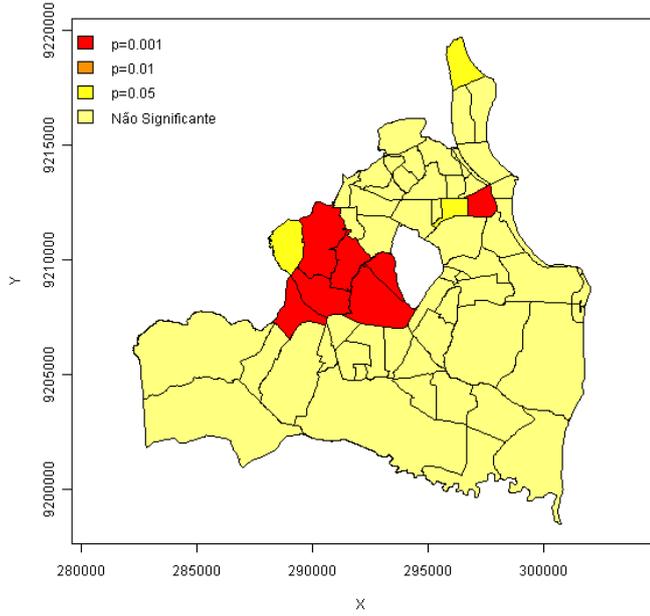


2010

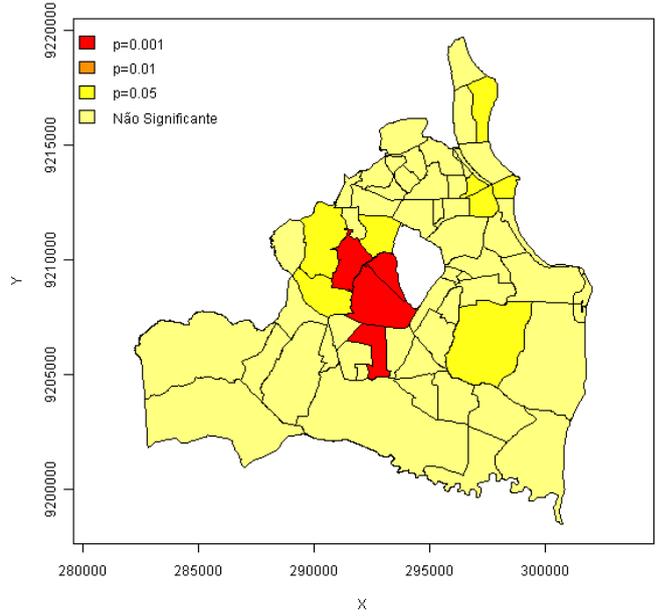


### APÊNDICE IV – LISA MAPA

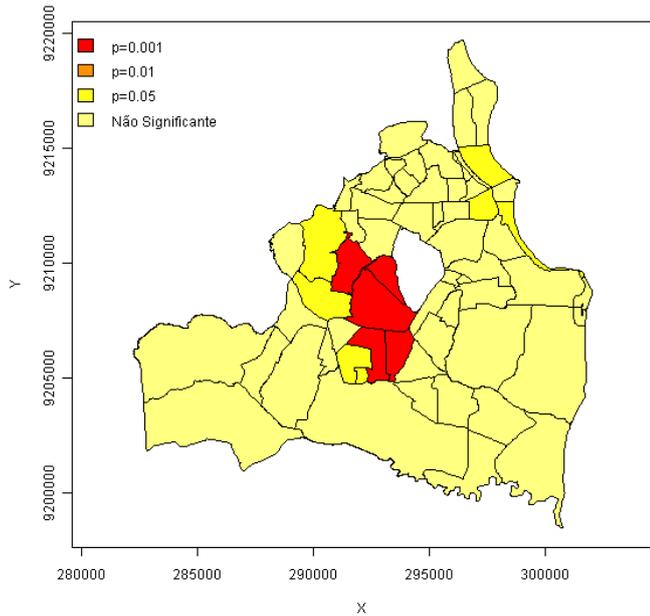
2007



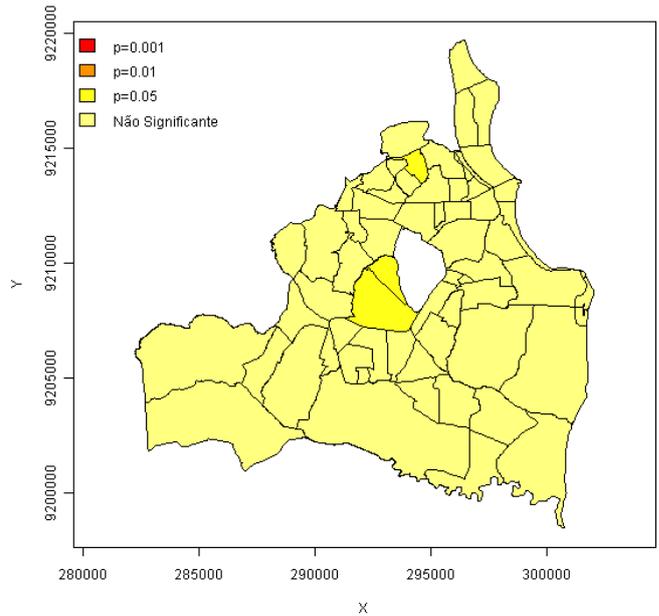
2008



2009

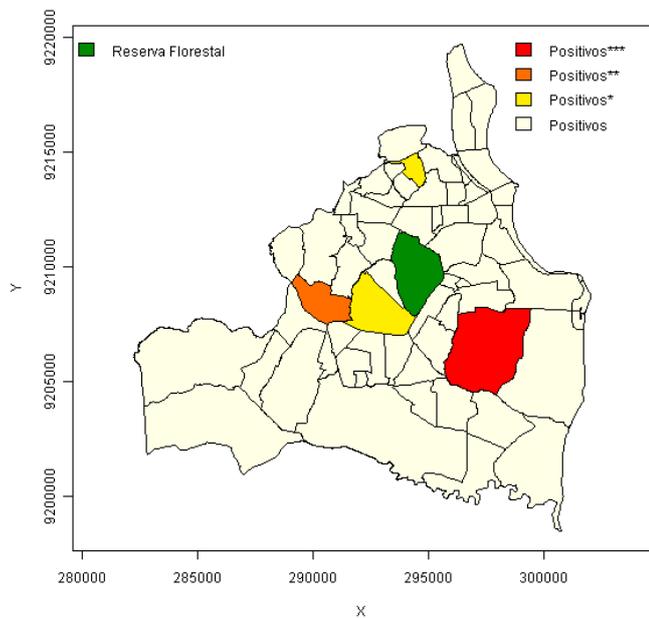


2010

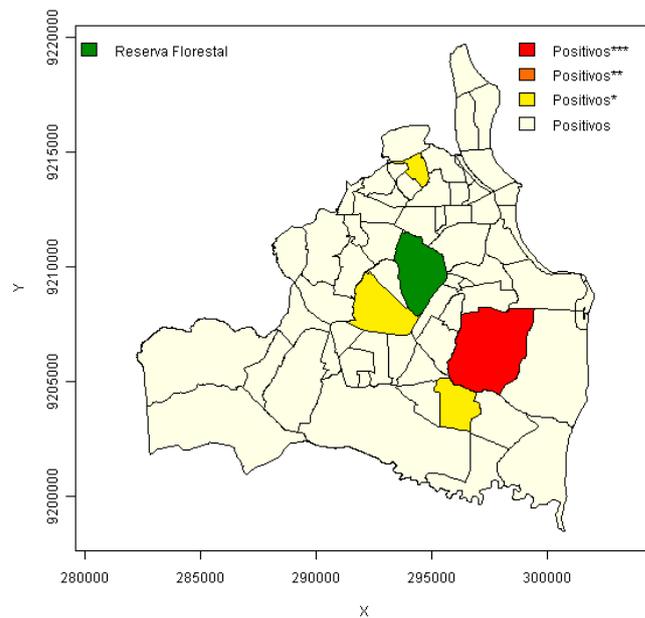


## APÊNDICE V – Gi

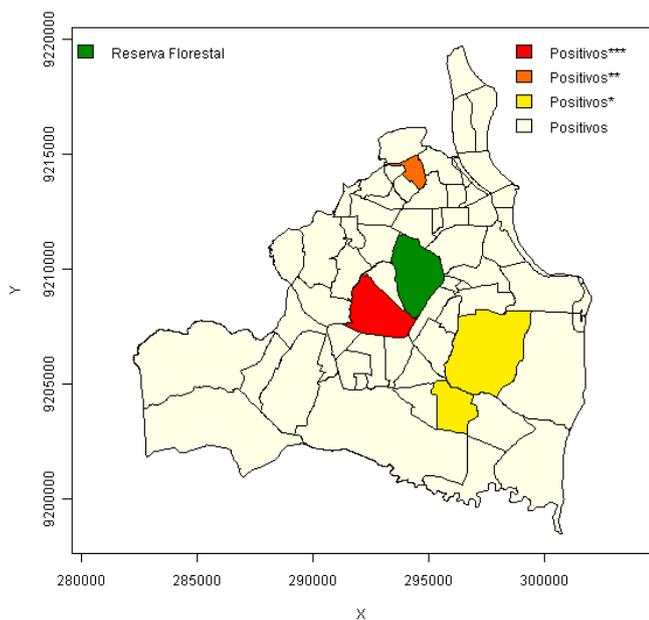
2007



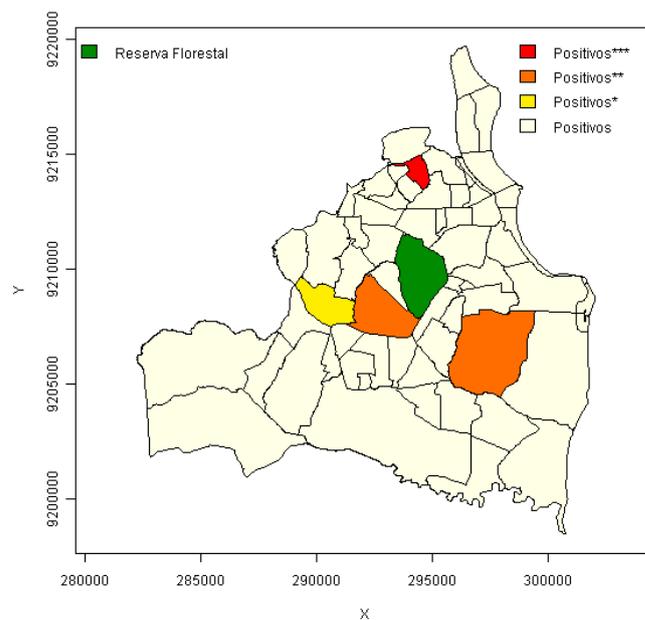
2008



2009

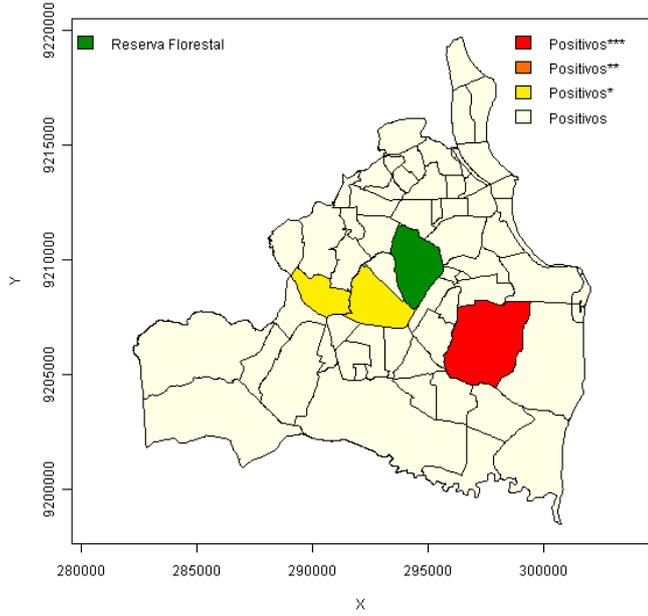


2010

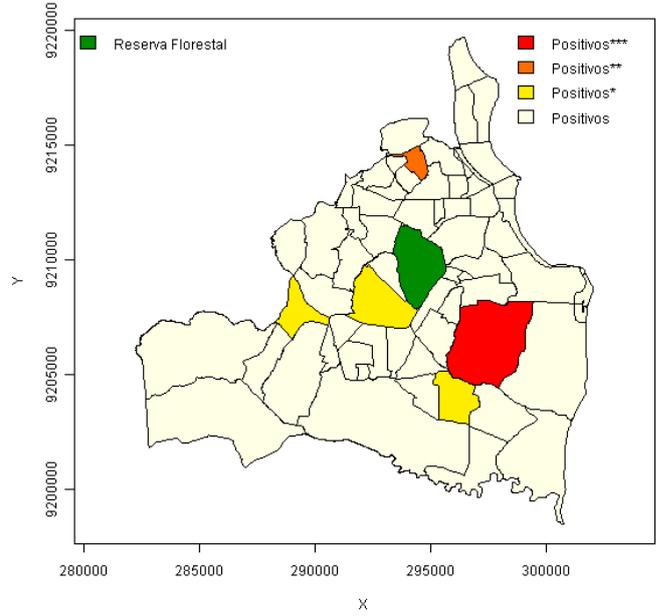


### APÊNDICE VI – GI\*

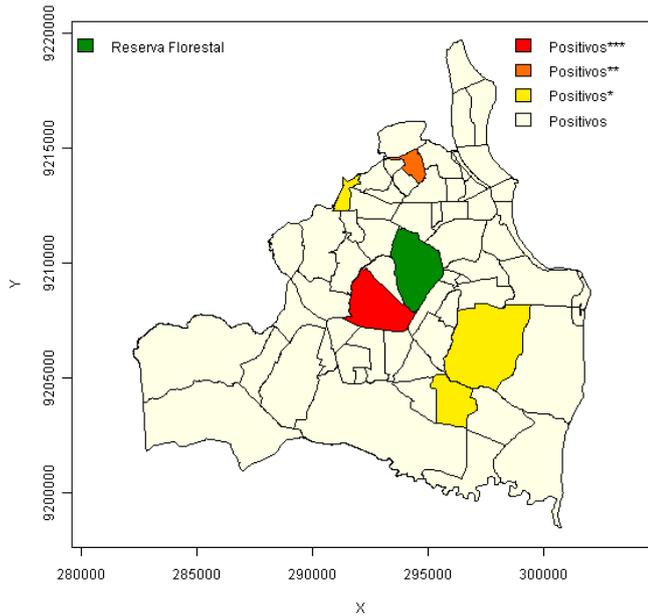
2007



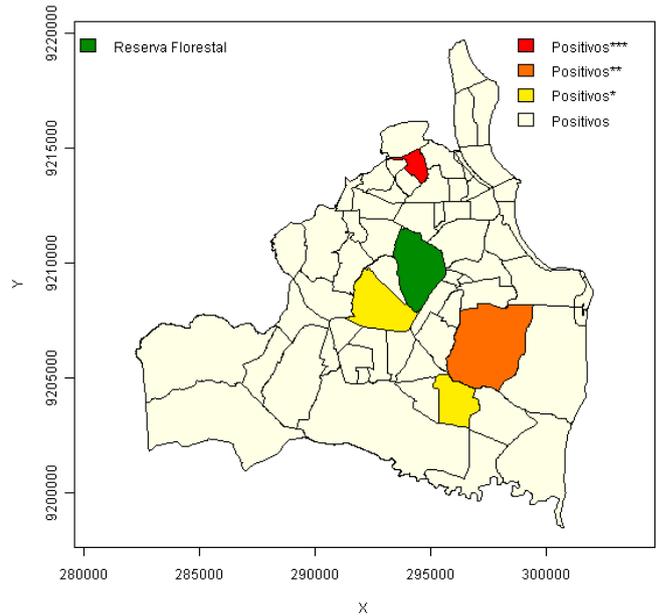
2008



2009

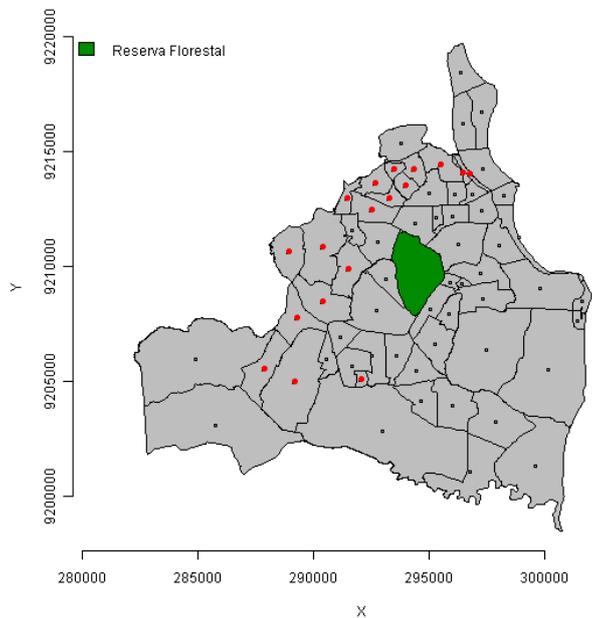


2010

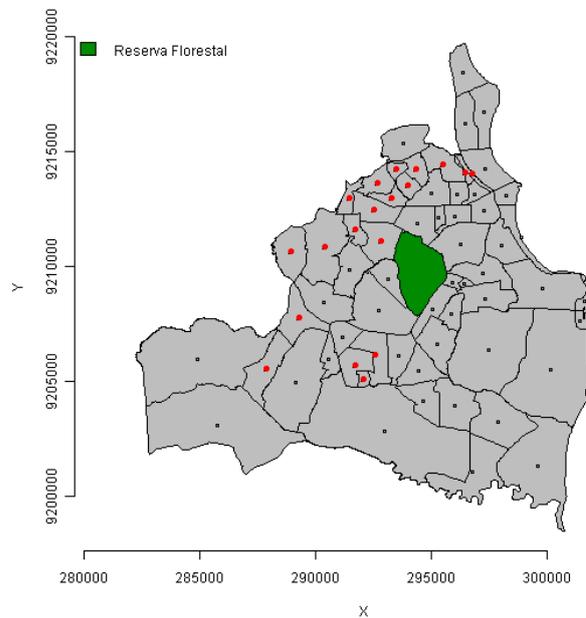


## APÊNDICE VII – STATÍSTICAS CAN

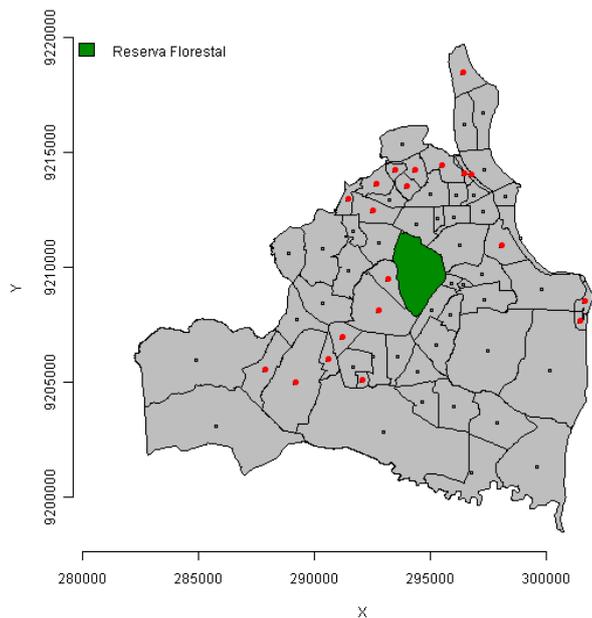
2007



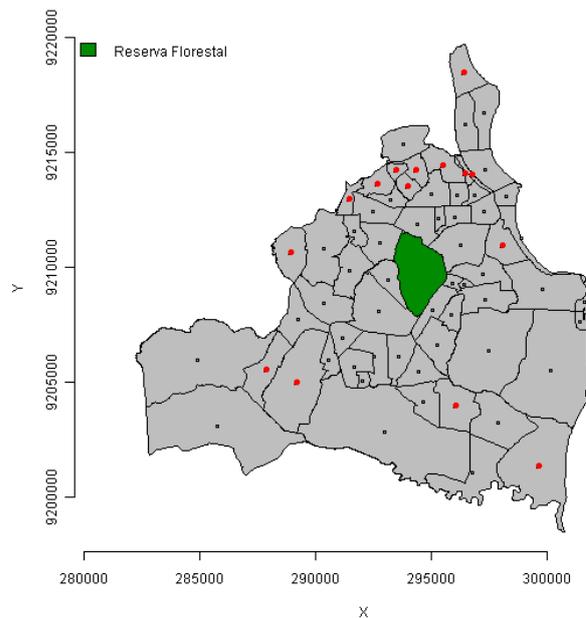
2008



2009

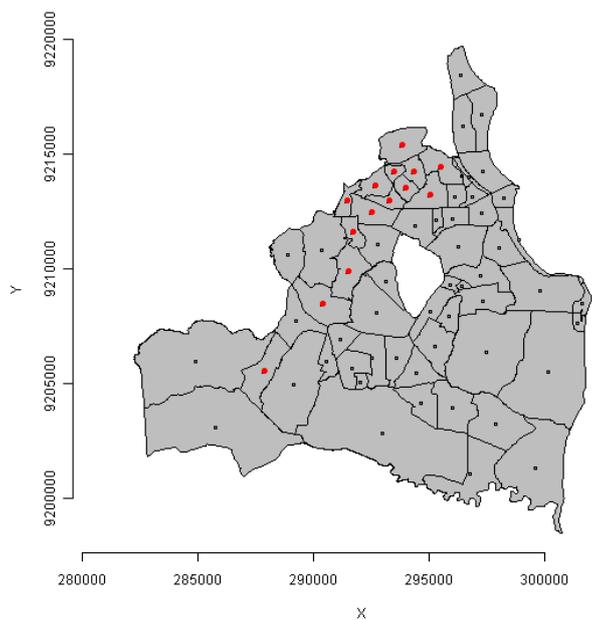


2010

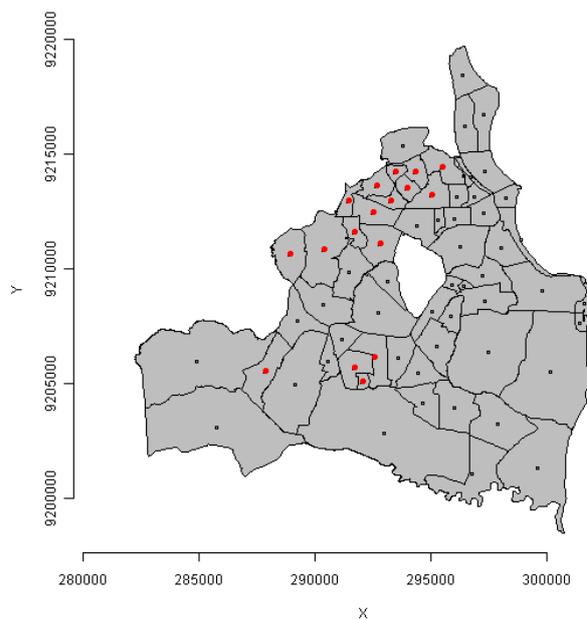


**APÊNDICE VIII – BESAG&NEWELL MODIFICADO**

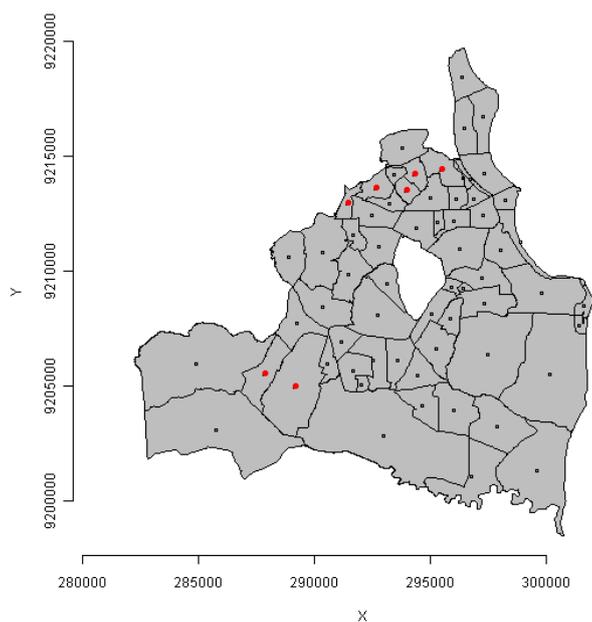
2007



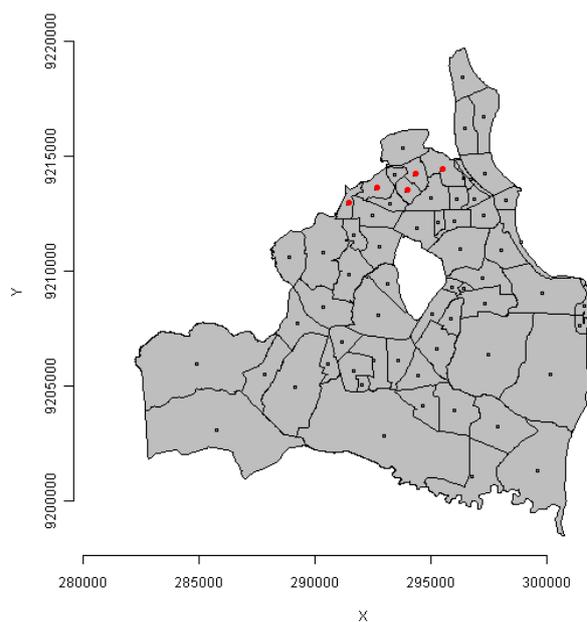
2008



2009



2010



APÊNDICE IX - SCRIPT UTILIZADO NO R

#Análise Espacial com o R

```
require(DCluster)
#CARREGAR MAPA
mapa=read.shape('E:/Monografia/Shape/homi_jpa.shp',dbf.data = TRUE, verbose=TRUE, repair=FALSE)
dados=mapa$att.data
```

```
#juntar dados ao mapa
mapa = Map2poly(mapa, as.character(dados[,1]))
```

#MAPA DE HOMICÍDIOS PARA 2010

```
classes.hom = c(-Inf,1,3,6,10,15,Inf)
legenda.hom = c("0", "1 |- 3", "3 |- 6", "6 |- 10", "10 |- 15", "15 ou mais")
palette(heat.colors(20)) #voltar a paleta normal >palette("default")
cores.hom = c(20,18,16,11,8,1)
mapa.hom = c(cores.hom[findInterval(dados$HOM_2010, classes.hom, all.inside=TRUE)], "green4")
plot(mapa,xlab="X", ylab="Y", col=mapa.hom, forcefill=FALSE)
legend(x="topleft", legend=legenda.hom, fill=cores.hom, bty="n")
legend(x='bottomleft', legend=c('ReservaFlorestal'), fill='green4', bty='n')
dev.print(png,filename="E:/Monografia/Hom_2010.png", width=587, height=586)
```

#MAPA DE RISCO RELATIVO PARA 2010

```
classes.hom = c(-Inf,0.25,0.5,1,2,4,Inf)
legenda.hom = c("0 |- 0,25", "0,25|- 0,5", "0.5 |- 1", "1 |- 2", "2 |- 4", "4 ou mais")
palette(heat.colors(20)) #voltar a paleta normal >palette("default")
cores.hom = c(20,18,16,11,8,1)
mapa.hom = c(cores.hom[findInterval(dados$RR_2010, classes.hom, all.inside=TRUE)], "green4")
plot(mapa,xlab="X", ylab="Y", col=mapa.hom, forcefill=FALSE)
legend(x="topleft", legend=legenda.hom, fill=cores.hom, bty="n")
legend(x='bottomleft', legend=c('ReservaFlorestal'), fill='green4', bty='n')
dev.print(png,filename="E:/Monografia/RR_2010.png", width=587, height=586)
```

#Calculo da Matriz de Proximidades

```
#Função para criar malha
malha<- function(inix,fimx,iniy,fimy,n){
hx<- (fimx-inix)/(n-1)
hy<- (fimy-iniy)/(n-1)
```

```
y <- seq(iniy,fimy,hy)
x <- rep(inix,n)
```

```
for(i in 1:n-1){
y <- c(y,seq(iniy,fimy,hy))
x <- c(x,rep(x[1]+i*hx),n)
}
M <- matrix(c(x,y),ncol=2)
return(M)
}
```

#EXEMPLO

```
#m <- malha(1,10,20,25,10)
#plot(m[,1],m[,2])
#####
```

# GERAÇÃO DA MATRIZ

```
xy<- matrix(c(dados$X, dados$Y), ncol=2) #matriz com coodenadas dos centróides
P <- malha(280000,305000,9190000,9225000,30)
#points(P[,1],P[,2])
n <- dim(P)[1] # coordenadas dos pontos gerados da malha
```

```

perim<- dados$PERIMETER
d <- perim/3 #distância
c <- dim(dados)[1]
I <- matrix(c(rep(0,n*c)),ncol=c)
MC <- xy # matriz com centroides
for(i in 1:n){
  for(j in 1:c){
    if(d[j] <sqrt((P[i,1]-MC[j,1])^2+(P[i,2]-MC[j,2])^2)) I[i,j] <- 0 else I[i,j] <- 1
  }
}

M <- matrix(c(rep(0,c*c)),ncol=c)
for(k in 1:c){
  for(j in 1:c){
    cont<- 0
    for(i in 1:n){
      if (I[i,k]==1 && I[i,j]==1) cont<- cont + 1
    }
    M[k,j] <- cont
  }
}
#gravar matriz em um arquivo txt
write.table(M, 'E:/Monografia/matriz_prox.txt', sep = '\t', col.names = F, row.names = F)

#Definições de Algumas funções que serão utilizadas
#multiplicação de matriz
mult<- function(u,v){
  #se u ou v for vetor então transformar em matriz
  u<- as.matrix(u)
  v <- as.matrix(v)
  #teste condição de multiplicação
  du<- dim(u)
  dv<- dim(v)
  if(du[2]!= dv[1]) return("erro") else

x <- matrix(c(rep(0,du[1]*dv[2])),ncol=dv[2])
for(i in 1:du[1]){
  for(k in 1:dv[2]){
    soma<- 0
    for(j in 1:du[2]){
      soma<- u[i,j]*v[j,k] + soma
    }
    x[i,k] <- soma } }

return(x)
}

##Proporcionar matriz
normm<- function(W){
n <- dim(W)[2]
w <- matrix(c(integer(n*n)),ncol=n)
for (j in 1:n){
  s <- sum(W[j,])
  for (i in 1:n){
    w[j,i] <- (W[j,i]/s)
  }
}
return(w)
}

```

```
###CALCULO DO INDICE GLOBAL DE MORAN
```

```

W <- read.table('E:/Monografia/matriz_prox.txt')
Y <- dados$HOM_2010

yb<- mean(Y)
n <- length(Y)
s1 <- 0
for(i in 1:n){
  for(j in 1:n){
    s1<- s1 + (W[i,j]*(Y[i]-yb)*(Y[j]-yb))
  }
}

s2 <- 0
for(i in 1:n){
  s2 <- s2 + (Y[i]-yb)^2
}

s3 <- 0
for(i in 1:n){
  for(j in 1:n){
    s3<- s3 + W[i,j]
  }
}

I <- (n/(2*s3))*(s1/s2)

#####
# testar significância

x <- c(1:n)
nis<- 0
r <- 100
Ip<- c(rep(0,r))

s3 <- 0
for(i in 1:n){
  for(j in 1:n){
    s3 <- s3 + W[i,j]
  }
}

for(k in 1:r){
  Y2 <- Y
  x <- sample(x,n)
  for(i in 1:n){
    Y2[i] <-Y[x[i]]
  }

  y2b<- mean(Y2)

  n <- dim(dados)[1]

  s1<- 0
  for(i in 1:n){
    for(j in 1:n){
      s1 <- s1 + (W[i,j]*(Y2[i]-y2b)*(Y2[j]-y2b))
    }
  }
}

```

```

s2 <- 0
for(i in 1:n){
  s2 <- s2 + (Y2[i]-yb)^2
}
Ip[k] <- (n/(2*s3))*(s1/s2)
if(Ip[k] > I) nis<- nis+1
}
p <- nis/(r+1)
hist(c(I,Ip))
dev.print(png,filename="E:/Monografia/pvalor_moran.png", width=587, height=586)

Y <- dados$HOM_2010

# NORMALIZAR MATRIZ DE PROXIMIDADES
w <- normm(W)

####
#indice local de moran
yb<- mean(Y)
n <- length(Y)
I <- c(integer(n))
for(i in 1:n){
s1<- 0
s2<- 0
for(j in 1:n){
s1 <- s1 + W[i,j]*(Y[j]-yb)
s2<- s2 + (Y[j]-yb)^2
}
I[i] <- ((Y[i]-yb)*s1)/(s2/n)
}

# testar significância
r <- 50
P2 <- c(integer(n))
Y2 <- Y
x <- c(1:n)
nis<- c(integer(n))

for(k in 1:r){
Ip<- c(integer(n))
x <- sample(x,n)
for(i in 1:n){
Y2[i] <- Y[x[i]]
}
for(i in 1:n){
s1 <- 0
s2 <- 0
for(j in 1:n){
s1 <- s1 + W[i,j]*(Y2[j]-yb)
s2 <- s2 + (Y2[j]-yb)^2
}
Ip[i] <- ((Y[i]-yb)*s1)/(s2/n)
if(Ip[i] > I[i]) nis[i] <- nis[i]+1
}
}
P2<- nis/(r+1) #

classes = c(-Inf,0.001,0.01,0.05,Inf)
legenda = c("p=0.001", "p=0.01", "p=0.05", "Não Significante")
palette(heat.colors(20)) #voltar a paleta normal >palette("default")

```

```

cores = c(1,9,16,18)
mapa.hom = c(cores[findInterval(P2, classes, all.inside=TRUE)], "green4")
plot(mapa,xlab="X", ylab="Y", col=mapa.hom, forcefill=FALSE)
legend(x="topleft", legend=legenda, fill=cores, bty="n")
dev.print(png,filename="E:/Monografia/mapa_LISA2010.png", width=587, height=586)

#DIAGRAMA DE EMPARALHAMENTO DE MORAN
# Função Média Local
MM <- function(W,Z){
  Z <- as.matrix(Z)
  n <- length(Z)
  Wb<- c(integer(n))
  for(i in 1:n){
    Wb[i] <- mult(W[i,],Z)/sum(W[i,])
  }
  return(Wb)
}

##CALCULAR DESVIOS
desv<- function(d){
  m <- mean(d)
  for(i in 1:length(d)){
    d[i] <- d[i]-m
  }
  return(d)
}

Y <- dados$HOM_2010 #####entrada do vetor de atributos

d <- desv(Y)
MM <- MM(W,d)
plot(d,MM)
abline(v=0)
abline(h=0)
dev.print(png,filename="E:/Monografia/Diagrama_moran_hom2010.png", width=587, height=586)

Q <- c(integer(n))
for(i in 1:n){
  if(MM[i]>=0 && d[i]>=0) Q[i] <- 1
  if(MM[i]<0 && d[i]<0) Q[i] <- 2
  if(MM[i]<0 && d[i]>=0) Q[i] <- 3
  if(MM[i]>=0 && d[i]<0) Q[i] <- 4
}

classes = c(1,2,3,4,5)
legenda = c("1", "2", "3", "4")
palette(heat.colors(20)) #voltar a paleta normal >palette("default")
cores = c(1,9,16,18)
mapa.hom = c(cores[findInterval(Q, classes, all.inside=TRUE)], "green4")
plot(mapa,xlab="X", ylab="Y", col=mapa.hom, forcefill=FALSE)
legend(x="topleft", legend=legenda, fill=cores, bty="n")
dev.print(png,filename="E:/Monografia/mapa_Diagrama_moran_hom2010.png", width=587, height=586)

```

```
#####
###INDICE LOCAL DE GETIS E ORD
##### Gi*

gi <- function(Y,W){
  n <- length(Y)
  Gi <- c(integer(n))
  s <- sum(Y)
  for(i in 1:n){
    soma <- 0;
    for(j in 1:n){
      soma <- soma + W[i,j]*Y[j]
    }
    Gi[i] <- soma/s
  }
  return(Gi)
}

##### Nivel de Significância de Gi*
sigGi <- function(Gi,r,W){
  n <- length(Gi)

  maior <- matrix(integer(n*r),ncol=r)
  Gip <- matrix(integer(n*r),ncol=r)
  Gip2 <- c(integer(n))
  for(k in 1:r){
    Y2 <- c(integer(n))
    x <- c(1:n)
    x <- sample(x,n)
    for(i in 1:n){
      Y2[i] <- Y[x[i]]
    }
    Gip2 <- gi(Y2,W)
    for(j in 1:n){
      Gip[j,k] <- Gip2[j]
      if(Gip[j,k] > Gi[j]) maior[j,k] <- 1
    }
  }

  P2 <- c(integer(n))
  for(i in 1:n){
    if((r/2)>sum(maior[i,])) P2[i] <- (1 + sum(maior[i,]))/(1 + r) else P2[i] <- (1 + r - sum(maior[i,]))/(1 + r)
  }
  return(P2)
}

#Chamada das funções para Gi
W <- read.table('E:/Trabalho Geoestatística/matriz_prox.txt')
w <- normm(W)
Y <- dados$HOM_2007
Gi_local <- gi(Y,w)
sig_Gi <- sigGi(Gi_local,500,w)

#mapa
classes.localG = c(0, 0.005, 0.025, 0.05, Inf)
legenda.localG = c("Positivos***", "Positivos**", "Positivos*", "Positivos")
palette(heat.colors(20))
cores.localG = c(1, 7, 14, 20)
mapa.localG = c(cores.localG[findInterval(sig_Gi, classes.localG, all.inside=TRUE)], "green4")
```

```

plot(mapa, xlab="X", ylab="Y", col=mapa.localG, forcefill=FALSE)
legend(x="topright", legend=legenda.localG, fill=cores.localG, bty="n")
legend(x="topleft", legend=c("Reserva Florestal"), fill="green4", bty="n")
dev.print(png,filename="E:/Trabalho Geoestatística/mapa_gi2_hom2007.png", width=587, height=586)

```

```

##### Gi
# excluir o próprio elemento da coluna
W2 <- W
n <- dim(W)[1]
for(i in 1:n){
  W2[i,i] <- 0
}
#Chamada da função Gi*
Y <- dados$HOM_2010
Gi_local2 <- gi(Y,W2)
sig_Gi2 <- sigGi(Gi_local2,200,W2)

#mapa
classes.localG = c(0, 0.005, 0.025, 0.05, Inf)
legenda.localG = c("Positivos****", "Positivos**", "Positivos*", "Positivos")
palette(heat.colors(20))
cores.localG = c(1, 7, 14, 20)
mapa.localG = c(cores.localG[findInterval(sig_Gi2, classes.localG, all.inside=TRUE)], "green4")
plot(mapa, xlab="X", ylab="Y", col=mapa.localG, forcefill=FALSE)
legend(x="topright", legend=legenda.localG, fill=cores.localG, bty="n")
legend(x="topleft", legend=c("Reserva Florestal"), fill="green4", bty="n")
dev.print(png,filename="E:/Trabalho Geoestatística/mapa_gi_hom2010.png", width=587, height=586)

```

```

## índice global de Getis e Ord
GGlobal <- function(Y,W){
  n <- length(Y);
  s1 <- 0;
  for(i in 1:n){
    for(j in 1:n){
      s1 <- s1 + W[i,j]*Y[i]*Y[j]
    }
  }

  s2 <- 0;
  for(i in 1:n){
    for(j in 1:n){
      s2 <- s2 + Y[i]*Y[j]
    }
  }
  return(s1/s2)
}

```

```

#Resultado
# 2007 - 0.0252 ;2008 - 0.0237 ; 2009 - 0.0206 ; 2010 - 0.0220

```

```

#TESTAR SIG DE sigG
sigG <- function(G,r,W){

```

```

n <- dim(W)[1]
maior <- c(integer(r))

Gp <- c(integer(r))
Y2 <- c(integer(n))
for(k in 1:r){
  x <- c(1:n)
  x <- sample(x,n)
  for(i in 1:n){
    Y2[i] <- Y[x[i]]
  }
  Gp[k] <- GGlobal(Y2,W)
}
for(i in 1:r){
  if(Gp[i] > G) maior[i] <- 1
}

if((r/2)>sum(maior)) p <- (1 + sum(maior))/(1 + r) else p <- (1 + r - sum(maior))/(1 + r)

return(p)
}

```

```

#Chamada da função GGlobal
W <- read.table('E:/Trabalho Geoestatística/matriz_prox.txt')
w <- normm(W)

Y <- dados$HOM_2010
Global_G <- GGlobal(Y,w)
sigG(Global_G,300,w)

#resultado(2007-0.0166/2008-0.0199/2009-0.0897/2010-0.1827)

```

## # ESTATÍSTICA SCAN

```

#Para o ano 2010
# Gerar banco de dados no formato DCluster
DADOS = data.frame(Observed = dados$HOM_2010)
DADOS = cbind(DADOS, Expected = dados$POP2010*sum(dados$HOM_2010)/sum(dados$POP2010))
DADOS = cbind(DADOS, Population = dados$POP2010,x = dados$X, y = dados$Y)

##### MAPA PARA 3% #####
mle = calculate.mle(DADOS, model = 'poisson')
knresults = opgam(data = DADOS, thegrid = DADOS[,c('x','y')], alpha = .05, set.idxorder=TRUE, iscluster =
kn.iscluster, fractpop = .03, R = 1000, model = 'poisson', mle = mle)

plot(mapa, xlab = "X", ylab = "Y", bty = "n")
points(DADOS$x, DADOS$y, pch = 176)
points(knresults$x, knresults$y, col = 'red', pch = 19)
legend(x='topleft', legend=c('ReservaFlorestal'), fill='green4', bty='n')

dev.print(png,filename="E:/Monografia/est_scan.2010_3p.png", width=587, height=586)

```

```
write.table(knresults, 'E:/Monografia/est_scan.2010_3p.txt', sep = '\t', col.names = T, row.names = F)
```

```
# BESAG E NEWELL
# B & N deve usar o centróides como grid.
# O tamanho do cluster é 20.
# 100 simulações bootstrap são realizadas
# Poisson é o modelo utilizado nas simulações bootstrap para gerar
# observações.
# Nível significiance é 0.05

DADOS = data.frame(Observed = dados$HOM_2010)
DADOS = cbind(DADOS, Expected = dados$POP2010*sum(dados$HOM_2010)/sum(dados$POP2010))
DADOS = cbind(DADOS, Population = dados$POP2010,x = dados$X, y = dados$Y)

bnresults<-opgam(DADOS, thegrid=DADOS[,c("x","y")], alpha=.05,
iscluster=bn.iscluster, set.idxorder=TRUE, k=20, model="poisson",
R=100, mle=calculate.mle(DADOS) )

plot(mapa, xlab = "Longitude", ylab = "Latitude",bty = "n")
#Plot all the centroids
points(DADOS$x, DADOS$y, pch = 176)
#Plot signifiant centroids in red
points(bnresults$x, bnresults$y, col="red", pch=19)
dev.print(png,filename="E:/Monografia/Besag_Newell_2007_5p.png", width=587, height=586)
```

## #ANÁLISE DE PONTOS

```
require(DCluster)
#CARREGAR MAPA
mapa=read.shape('E:/Users/Jefferson/Documents/Arquivos
Spring/shape/shape_jp_2007correto/jp_bairros_2_region.shp',dbf.data = TRUE, verbose=TRUE, repair=FALSE)
#dados=mapa$att.data

mapa2=read.shape('E:/Users/Jefferson/Documents/Arquivos Spring/Municipios.shp',dbf.data = TRUE,
verbose=TRUE, repair=FALSE)

dados=read.table('J:/coord.txt', header = TRUE)
#TRANSFORMAR COODENADASGEOGRAFICAS EM PLANAS
#dados do tipo (s 7 10 00.49, o 34 53 05.27)
transcoord<- function(dados){
n <- dim(dados)[1]
g <- c(integer(n))
m <- c(integer(n))
s <- c(integer(n))
c <- matrix(c(integer(2*n)), ncol = 2)

for(i in 1:n){
g[i]<-as.numeric(substr(dados[i,1],2,2))
m[i]<-as.numeric(substr(dados[i,1],3,4))/60
```

```

s[i]<-as.numeric(substr(dados[i,1],5,9))/3600
c[i,1] <- -(g[i]+m[i]+s[i])
}
for(i in 1:n){

g[i]<-as.numeric(substr(dados[i,2],2,3))
m[i]<-as.numeric(substr(dados[i,2],4,5))/60
s[i]<-as.numeric(substr(dados[i,2],6,10))/3600
c[i,2] <- -(g[i]+m[i]+s[i])
}
return(c)
}

coord<- transcoord(dados)
write.table(coord, 'J:/coord2.txt', sep = '\t', col.names = T, row.names = F)

#
#plot(mapa)
#points(coord[,2],coord[,1], pch=19, bg="red")
#poly1 = mapa$Shapes[[95]]$verts
#points(poly1[,1],poly1[,2], pch=19, bg="red")
#plot(coord[,2],coord[,1])

#identify(coord[,2],coord[,1],n=2) #identificar pontos
#x11() #nova janela

library(spatstat)
poly1 = mapa2$Shapes[[96]]$verts
### Criar um poligono para o spatstat
n = dim(poly1)[1]
poly2 = poly1[-n,]
n1 = dim(poly2)[1]
poly2 = as.matrix(poly2)
poly3 = poly2
for(i in 1:n1) {
poly2[i, 1] = poly3[n1+1-i, 1]
poly2[i, 2] = poly3[n1+1-i, 2]
}
poly2 = list(x = poly2[,1], y = poly2[,2])

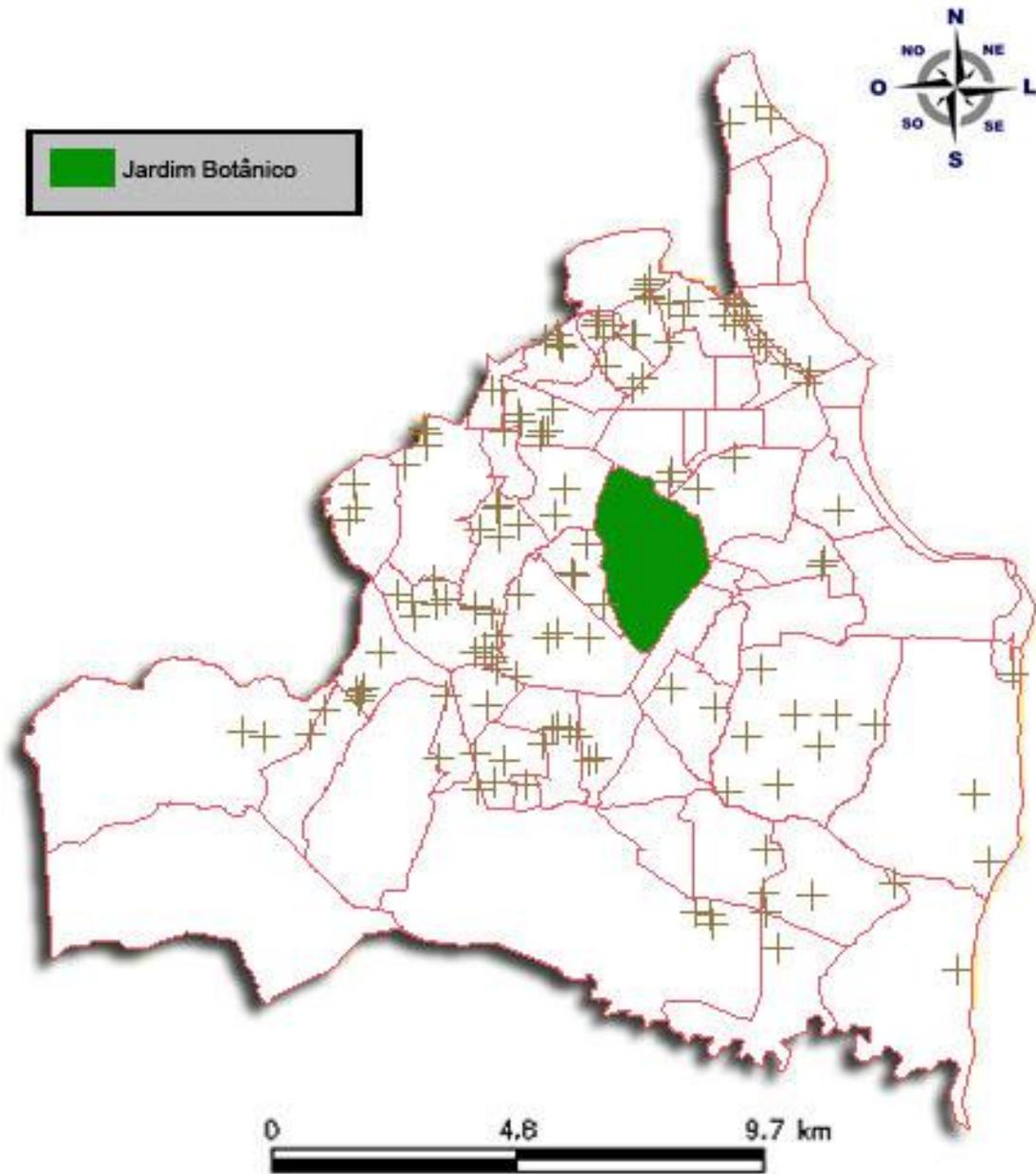
xb = coord[,2]
yb = coord[,1]
HB = ppp(xb,yb, poly = poly2)
#X <- ppp(xb, yb, c(-35.00,-34.76), c(-7.26,-7.05))
X <- ppp(xb, yb, c(0,0), c(1,1));plot(X)
### FUNÇÕES G, F, J, L
## Função G
hb_en_G = envelope(HB, fun=Gest, nsim=500)
plot(hb_en_G)

# Salvar o gráfico da Função G
dev.print(png, filename = "J:/Funcoes kernel/Funcao_G.png", width=440, height=450)
## Função F
hb_en_F = envelope(HB, fun=Fest, nsim=500)
plot(hb_en_F)
# Salvar o gráfico da Função F
dev.print(png, filename = "J:/Funcoes kernel/Funcao_F.png", width=440, height=450)
## Função J
hb_en_J = envelope(HB, fun=Jest, nsim=500)

```

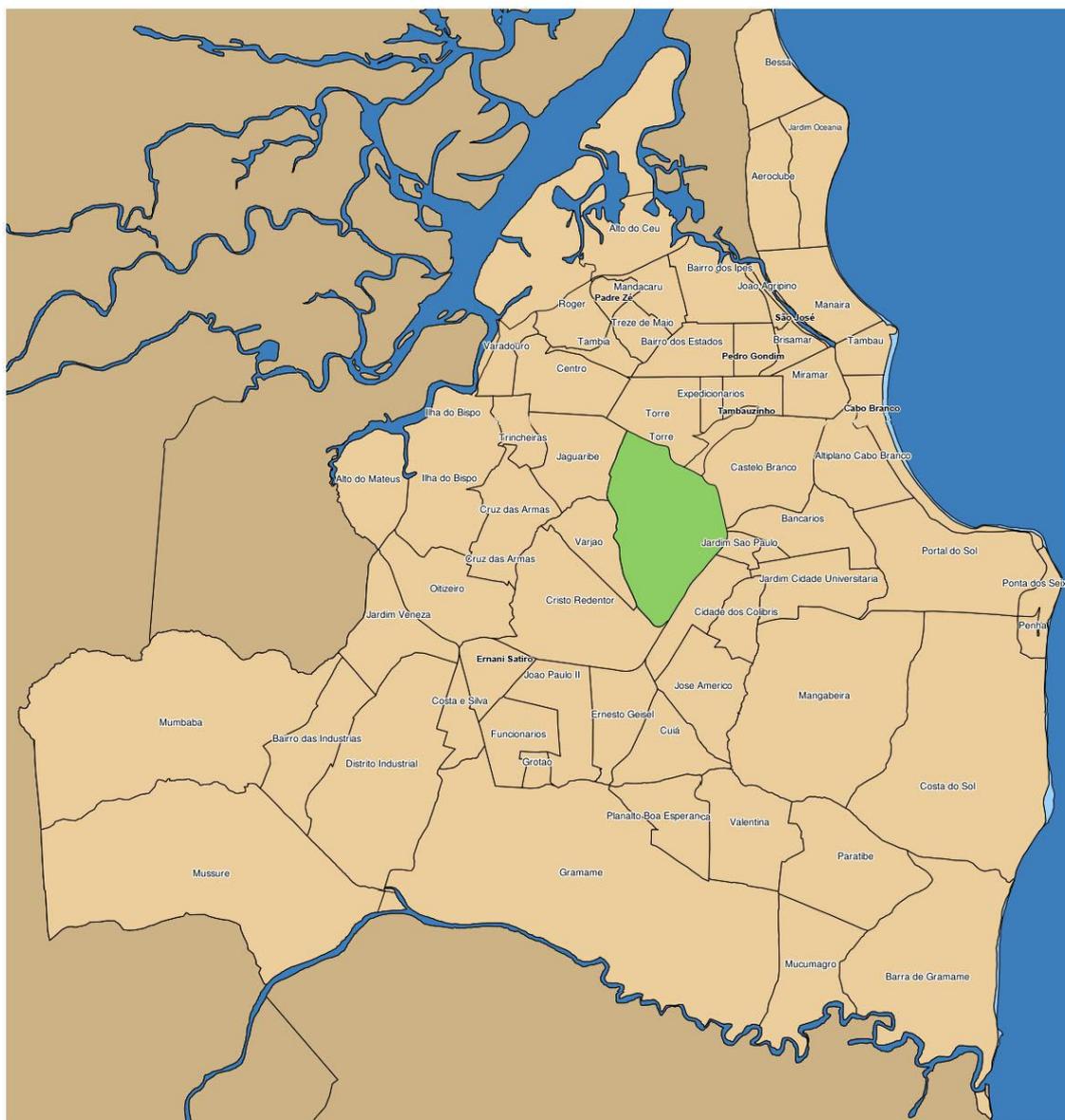
```
plot(hb_en_J)
# Salvar o gráfico da Função J
dev.print(png, filename = "J:/Funções kernel/Funcao_J.png", width=440, height=450)
## Função L
hb_en_L = envelope(HB, fun=Kest, nsim=50)
plot(hb_en_L, sqrt(.pi) - r ~ r)
# Salvar o gráfico da Função L
dev.print(png, filename = "J:/Funções kernel/Funcao_L.png", width=440, height=450)
```

**APÊNDICE X - LOCALIZAÇÃO PONTUAL DOS HOMICÍDIOS OCORRIDOS ENTRE AGOSTO  
E NOVEMBRO DE 2011**



**ANEXO**  
**PREFEITURA MUNICIPAL DE JOÃO PESSOA**  
**SECRETARIA DE PLANEJAMENTO**  
**DIRETORIA DE GEOPROCESSAMENTO E CADASTRO URBANO**

**MAPA DE JOÃO PESSOA**



0 0,5 1 1,5 2 2,5 km