

# Programa de Pós-graduação em Geografia Física

## FLG5127 – Métodos e Técnicas de Pesquisa e Redação Científica

### Aula 6

- **Análise espacial de dados**

# Premissa

“In a real sense we do not map data. Instead, we map theories using data to present them in place.(...)”

**Koch (2011, p.5)**

# Natureza do Dado Espacial

## Georeferenciado

- dados de atributos com localização geográfica
- latitude-longitude, UTM (GPS)
- referência espacial: municípios, setores censitários, CEP, etc.

# O que há de **ESPECIAL** em dados **ESPACIAIS**?

- Localização, localização, localização
- **onde** faz a diferença
- Dependência é regra
- Escala espacial é importante

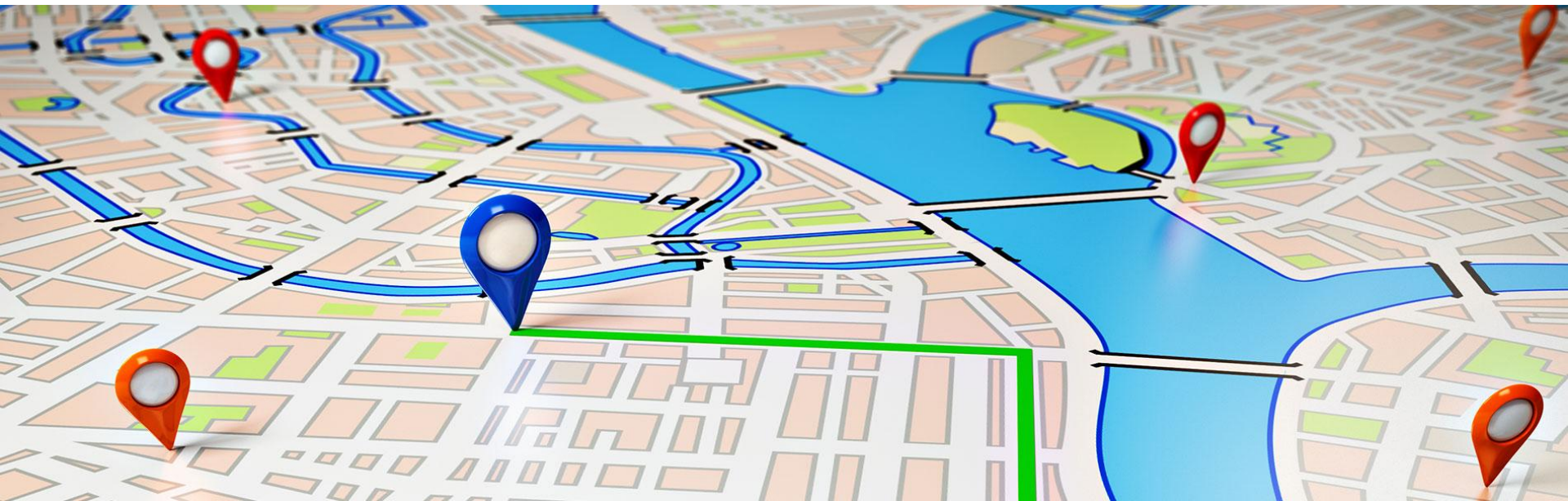
**Tudo que está localizado na Terra pode ser mapeado e analisado em um SIG.**



Image: Free use. Taken from: <http://balaarka.com/wp-content/uploads/2015/09/53mx3P5I.jpeg>



Dados ambientais (*e.g.*, fontes de poluição) podem ser coletadas por meio de receptor GPS (*Global Positioning System*) e plotado em um SIG como uma camada.



Endereços em um banco de dados podem ser geocodificados. Isto significa que o endereço é convertido para uma coordenada geográfica de referência (*e.g.*, latitude e longitude) e pode ser plotado em um SIG como uma camada de informação. A geocodificação pode ser feita usando um banco de dados de rua com endereços ou usando a base do Google Maps.

# O que é análise espacial?

- estudo quantitativo de fenômenos que são localizados no espaço.
- “... quando os dados são espacialmente localizados e se considera explicitamente a possível importância de seu arranjo espacial na análise ou interpretação dos resultados” (Bailey & Gatrell, 1995).
- Diversas operações realizadas em um SIG são também chamadas análise espacial.



# O que é e o que não é estatística espacial

➤ **TODO** dado possui referência temporal e espacial.

➤ Não basta que o dado seja “espacial”:

A regressão linear usual de taxas de mortalidade municipais de Minas contra a renda per capita municipal **NÃO** é parte de estatística espacial.

➤ Embora os municípios possuam localização espacial, isto não foi usado na análise acima.

➤ Estatística espacial é propriedade do método de análise, e não do dado utilizado na análise.

# Quando usar?

- se existirem perguntas ou hipóteses sobre o mecanismo gerador dos dados que envolvam alguma característica espacial ou geográfica;
- quando o evento em estudo é gerado por fatores ambientais de difícil detecção a nível do indivíduo;
- quando o evento em estudo e os fatores relacionados têm distribuição espacialmente condicionada;
- no estudo de trajetórias entre localidades.

# Primeira Lei da Geografia

*“everything is related to everything else, but near things are more related than distant things”* (Tobler, 1970, p.236)

Proximidade = similaridade

# Dependência espacial

➤ *“Independência é um pressuposto muito conveniente que faz grande parte da teoria estatística matemática tratável. Entretanto, modelos que envolvem dependência estatística são frequentemente mais realísticos.” (Cressie, 1991)*

**Dependência espacial:  
um problema?**

A dependência espacial  
viola a premissa de  
**independência** em  
muitos testes  
estatísticos!

# CLASSES DE PROBLEMAS

	<b>Tipos de Dados</b>	<b>Exemplo</b>	<b>Problemas Típicos</b>
<b>Análise de Padrões Pontuais</b>	<b>Eventos Localizados</b>	<b>Ocorrência de Doenças</b>	<b>Determinação de Padrões</b>
<b>Análise de Superfícies</b>	<b>Amostras de Campo</b>	<b>Distribuição de Minerais</b>	<b>Interpolação de Superfície</b>
<b>Análise de Áreas</b>	<b>Entidades e Atributos</b>	<b>Dados Censitários</b>	<b>Relacionamento das entidades Regressão</b>

# Técnicas de Análise Espacial para eventos - Pontos

## Spatial Analytic Techniques

Centro médio/distância padrão

Elipse por desvio-padrão

Análise de gradiente

Análise de vizinho mais próximo

Teste da taxa variância/média

Análise quadrática

Agrupamento espacial

Agrupamento espaço-temporal



# Técnicas de Análise Espacial para eventos - Linhas

## **Spatial Analytic Techniques**

**Caminho aleatório**

**Vetores**

**Teoria gráfica (nodalidade, conectividade, dispersão, hierarquia nodal, análise de fluxo)**

# Técnicas de Análise Espacial para eventos - Áreas

## **Spatial Analytic Techniques**

**Quocientes locacionais**

**Razões padronizadas de mortalidade**

**Probabilidade de Poisson**

**Agrupamento espacial**

**Agrupamento espaço-temporal**

**Medidas de autocorrelação**

**Agrupamento hierárquico**

# Técnicas de Análise Espacial para eventos - Superfícies

## Spatial Analytic Techniques

### Isolinhas

**Análise de superfície de tendência  
(polinômios de séries de potência,  
séries de Fourier)**

# Técnicas de Análise Espacial para eventos - Comparação entre mapas

## Spatial Analytic Techniques

**Curvas de Lorenz**

**Coeficiente de correspondência de área**

**Coeficiente de correlação**

**Diferença entre mapas**

# Pontos

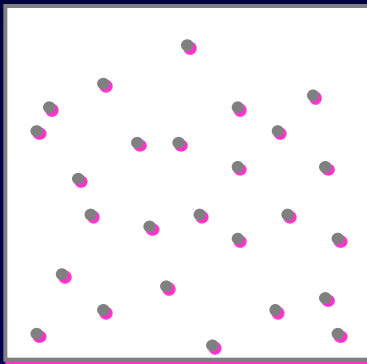
➤ como **eventos**

crimes (endereços), acidentes (localização)

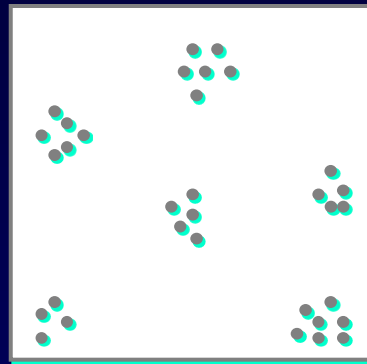
- análise de padrão de pontos: Estimador de Intensidade (“Kernel Estimation”); Vizinho mais Próximo; Função K

# Distribuição de Pontos - *Arranjos Espaciais*

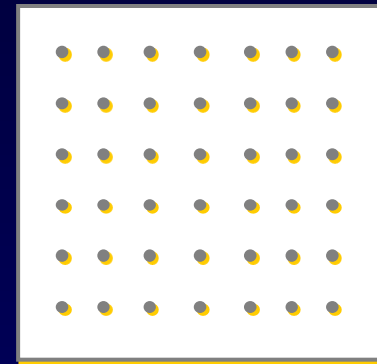
- Identificar padrão sistemático



*Aleatório*



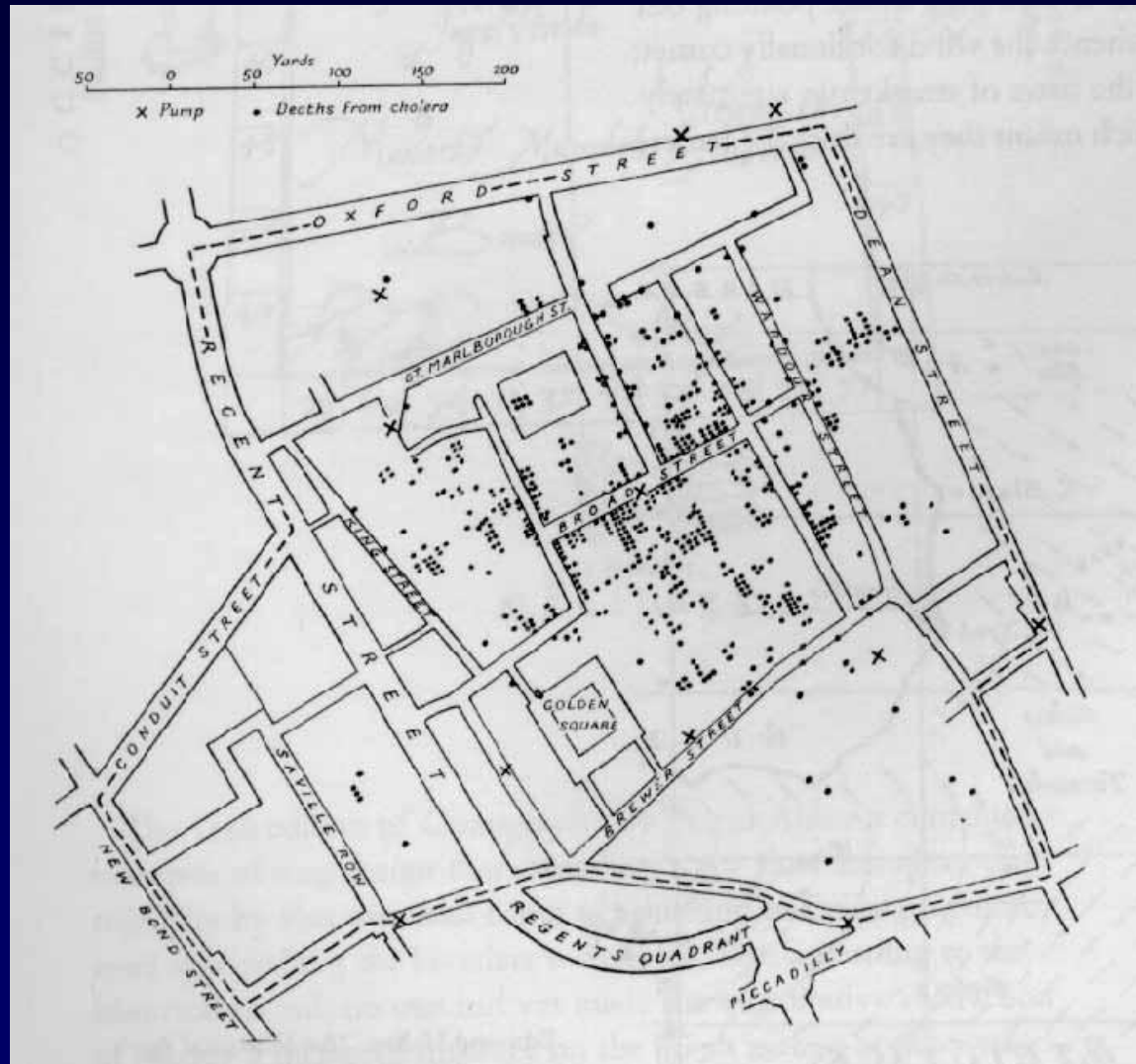
*Agrupado*



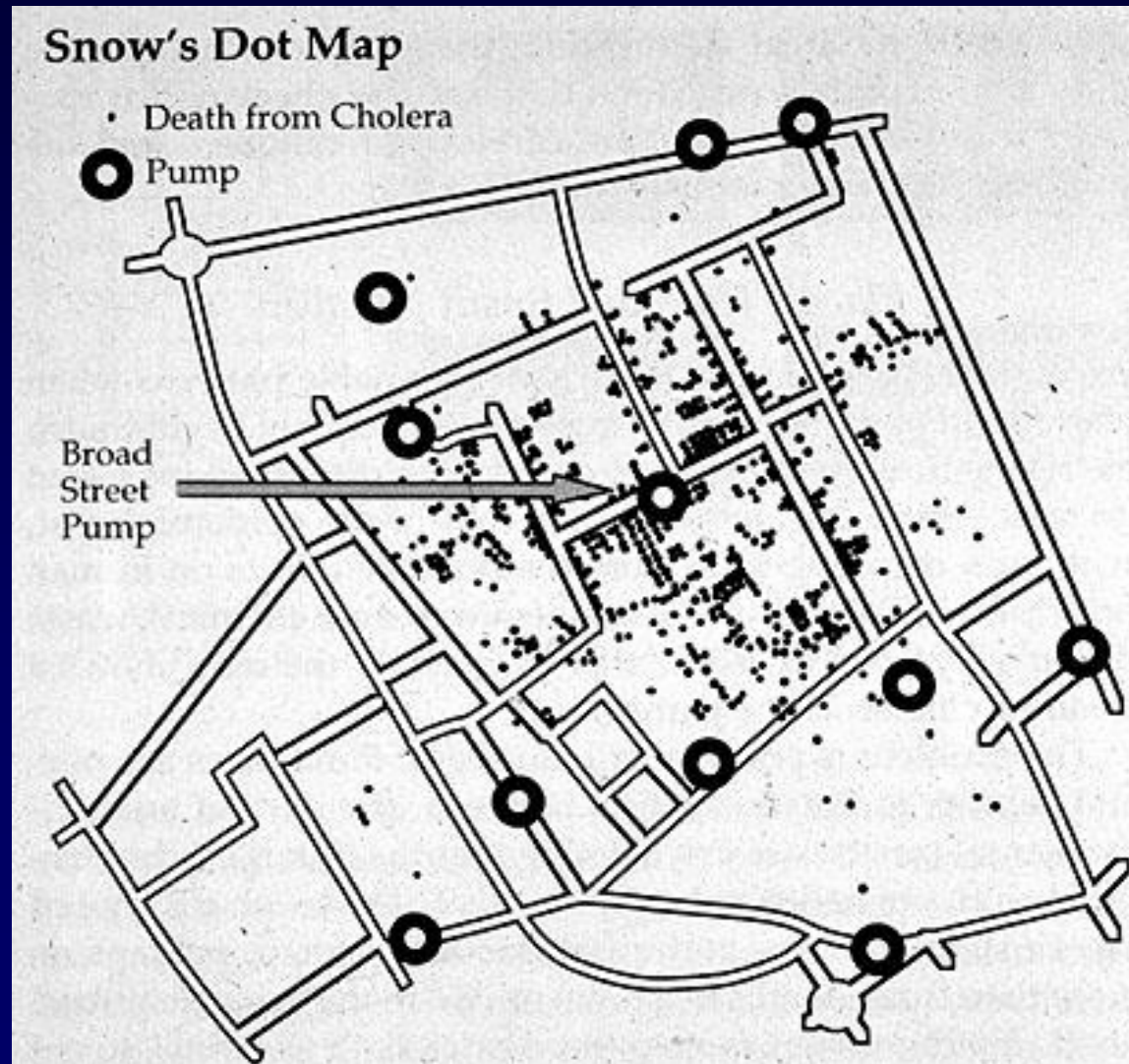
*Regular*

- Ocorrência de clusters ou clusterização

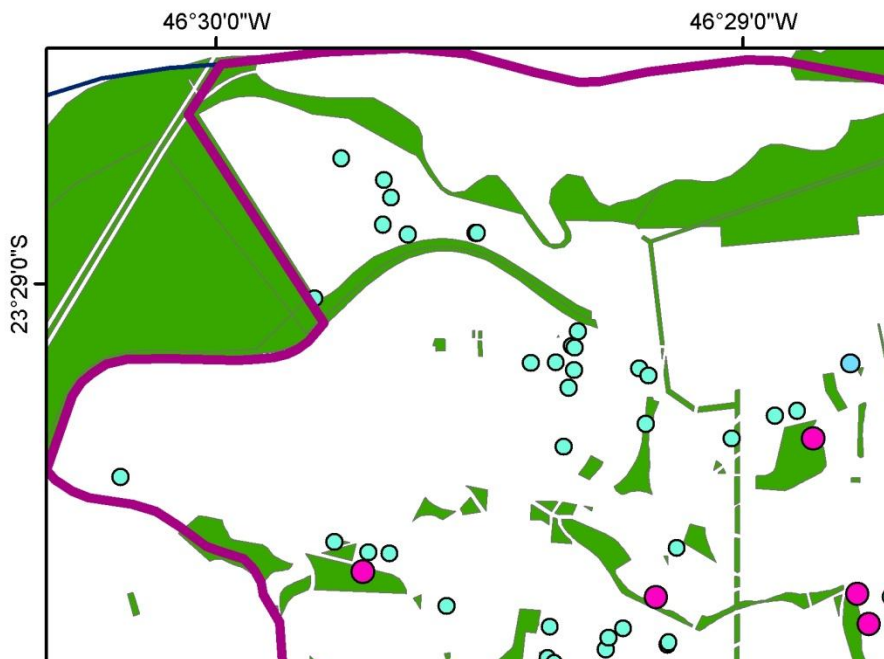
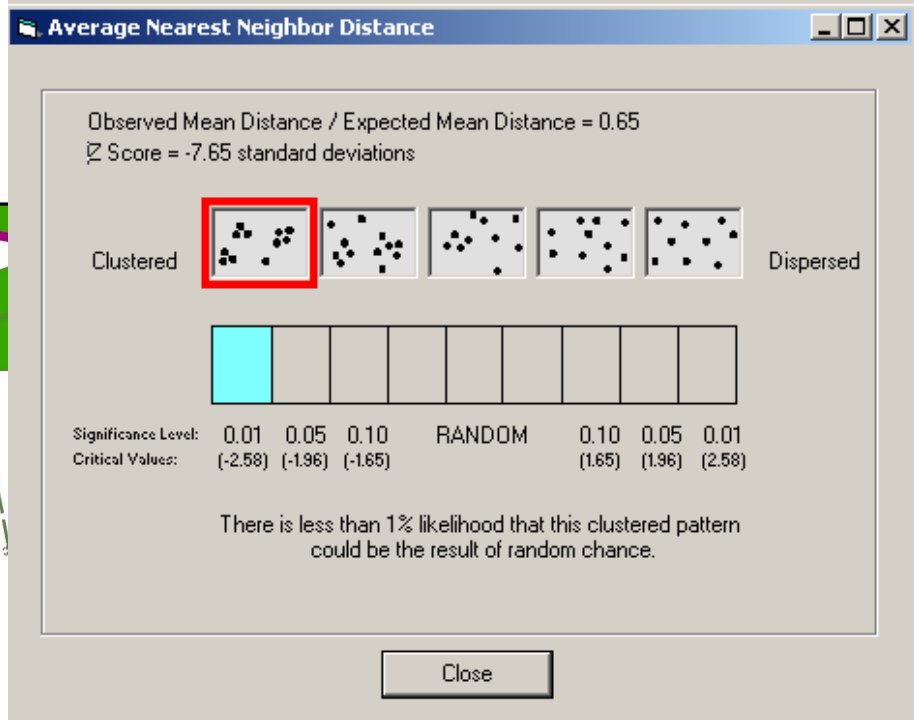
# Endereços plotados no mapa como pontos







# Endereços plotados no mapa como pontos

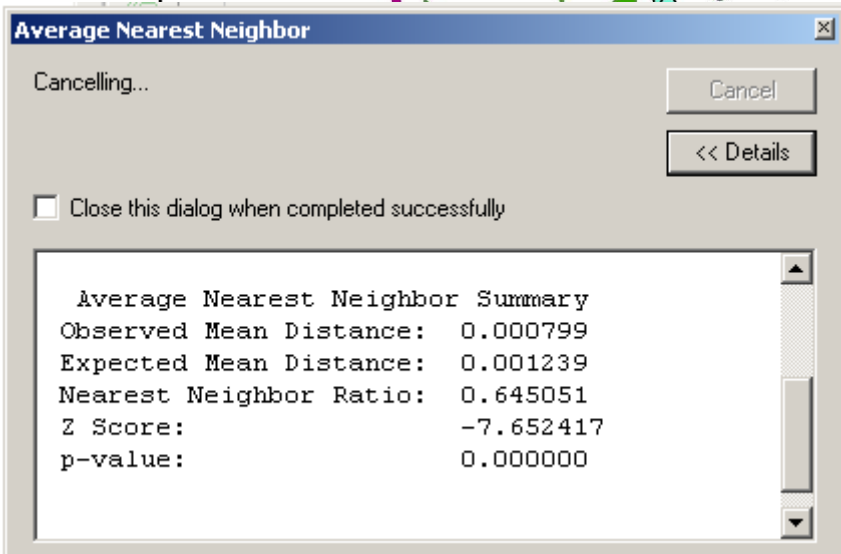
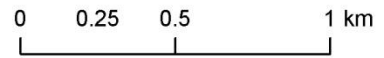






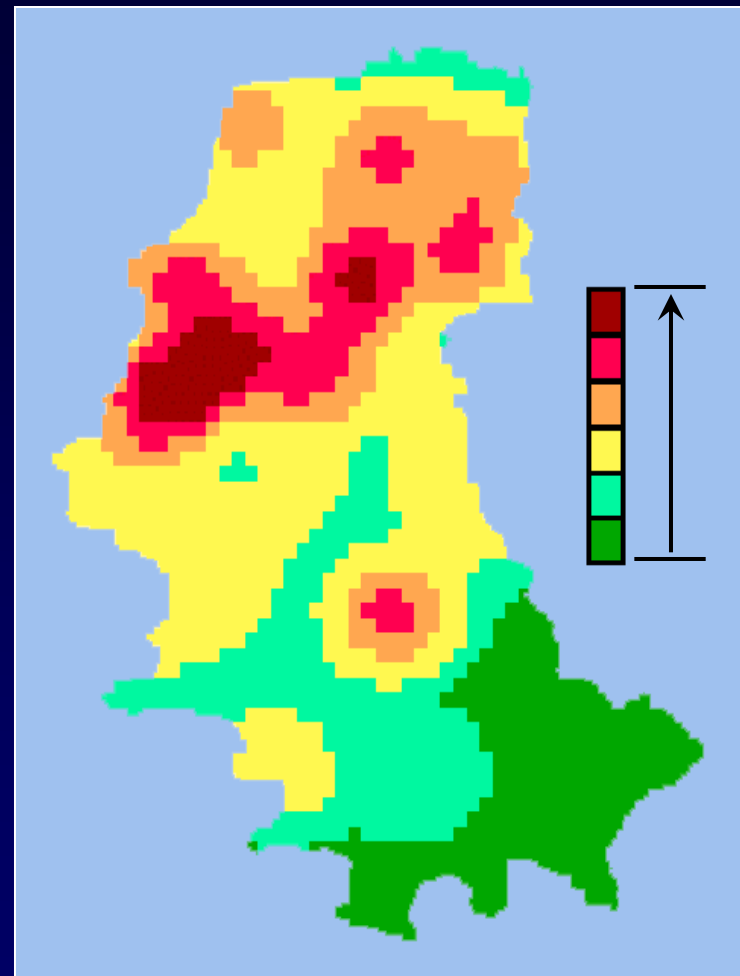
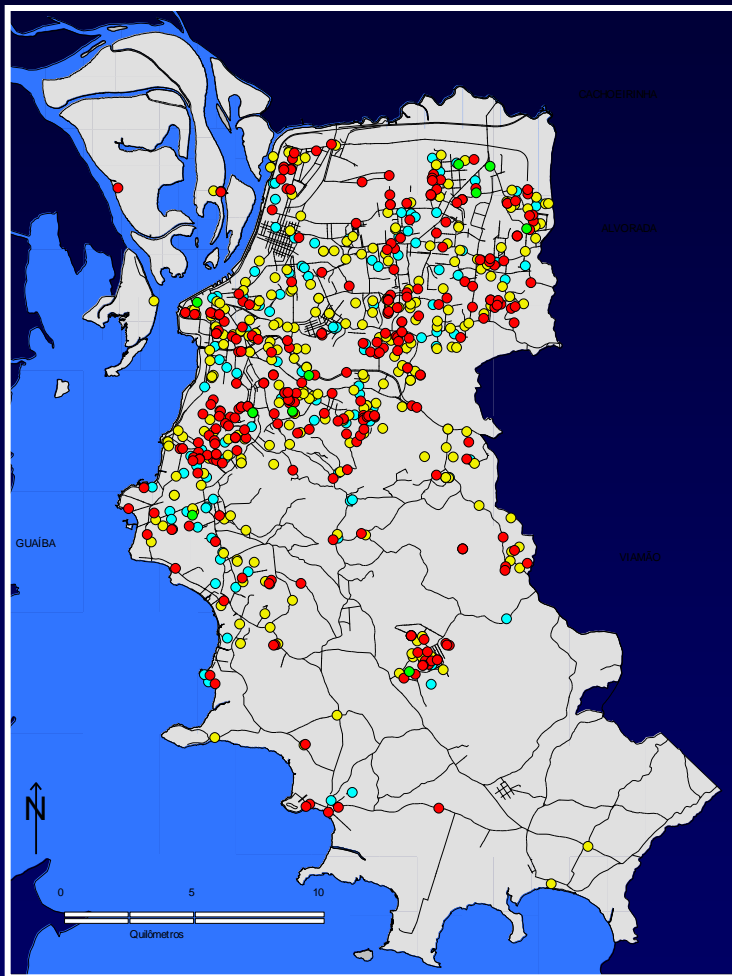
### Prática de caminhada como lazer

-  equipamentos\_esportivos
-  atinge\_lazer
-  Pracas\_do\_MSP\_Proj
-  Distr\_EM



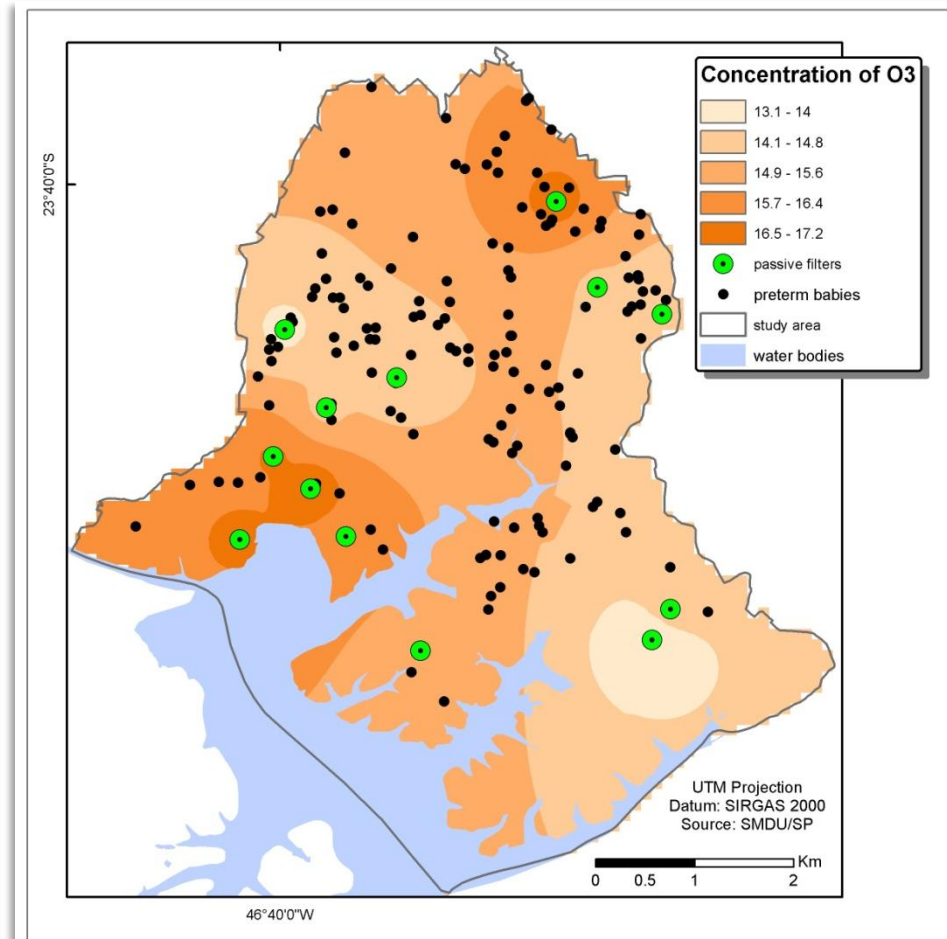
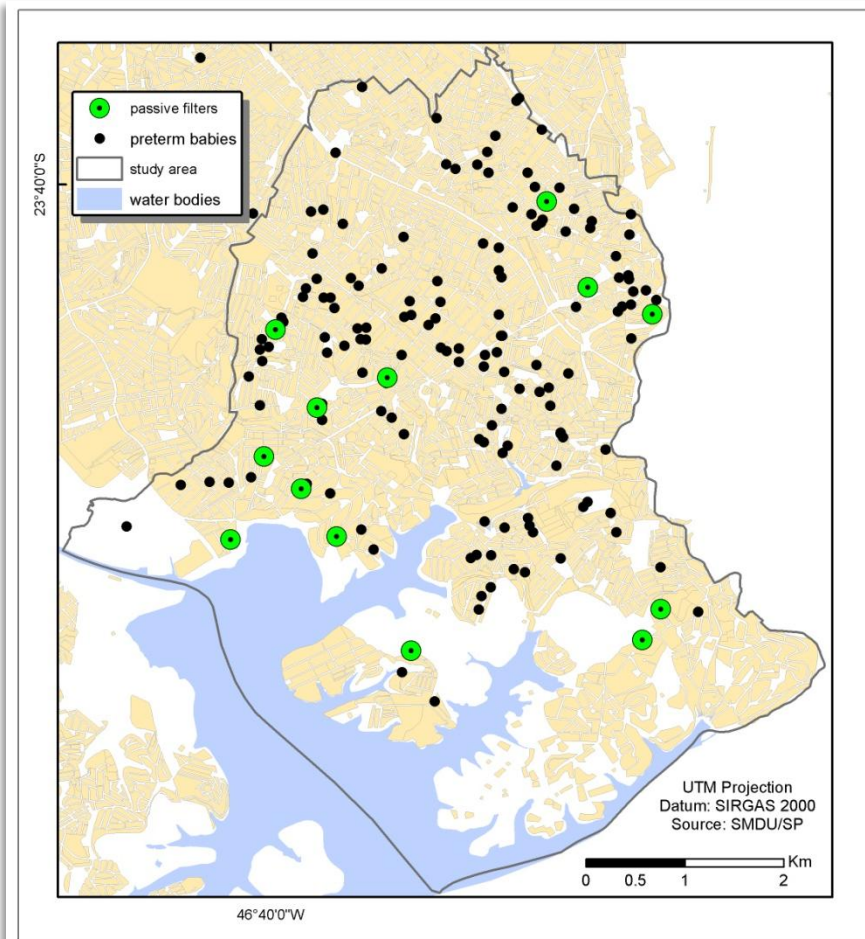
# Kernel Estimation

## Mapeando a violência - localização pontual



Legenda: ● Homicídios / ● Acidentes de transporte / ● Suicídios

# Preterm deliveries and Ozone in São Paulo (interpolation techniques)



# Agrupamentos (“clusters”)

- podem ser causados por diferentes fatores
  - agentes infecciosos, contaminação ambiental localizada, efeitos colaterais de tratamentos, etc.
- pontos ou áreas
- controle de fatores (distribuição populacional e outras co-variáveis que podem criar agregados)

# Exemplo de aplicação da estatística de varredura na área ambiental

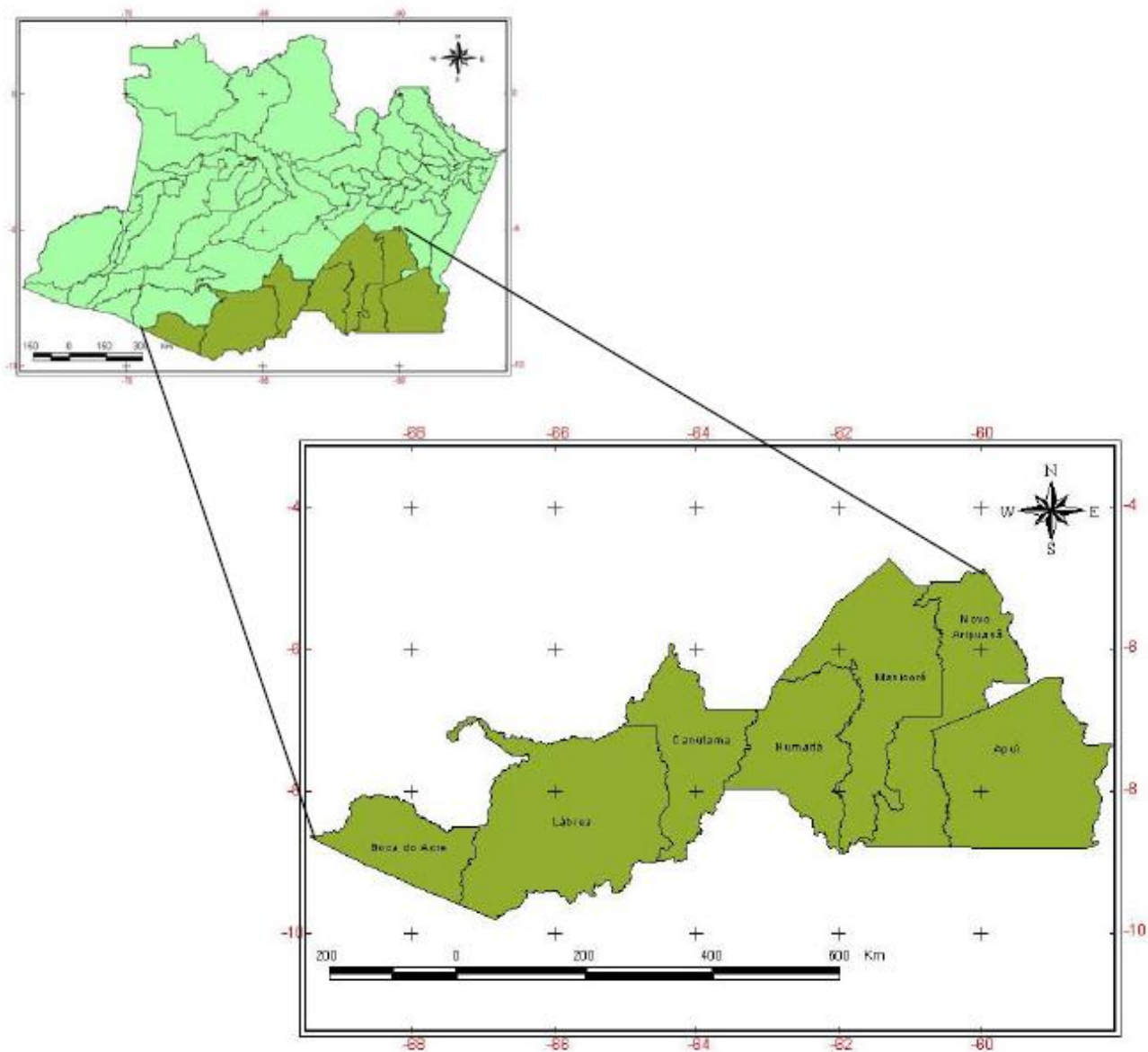
BALIEIRO, A.A.S. *Detecção de conglomerados dos alertas de desmatamentos no estado do Amazonas usando estatística de varredura espaço-temporal*. Dissertação de Mestrado, Universidade Federal de Viçosa, 2008. 87p.

## **Objetivo geral do trabalho:**

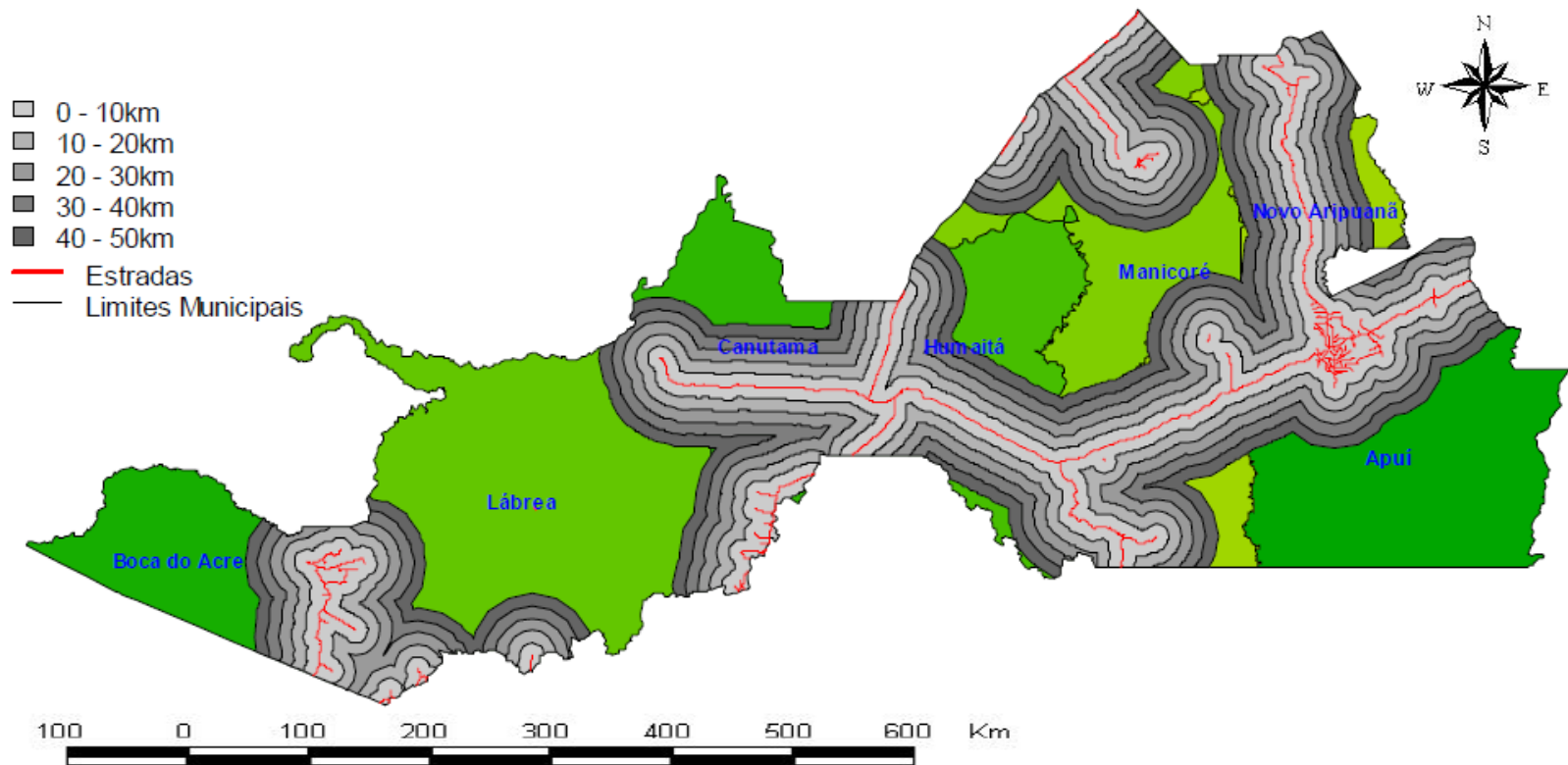
**- Testar a existência de conglomerados espaço-temporais de alerta de desmatamento em alguns municípios da região sul do Amazonas que deveriam ser priorizados para controle**

## **Resultados**

- identificar conglomerados ativos e históricos**
- Foi possível constatar que a proximidade das estradas foi a covariável que mais explicou os conglomerados**

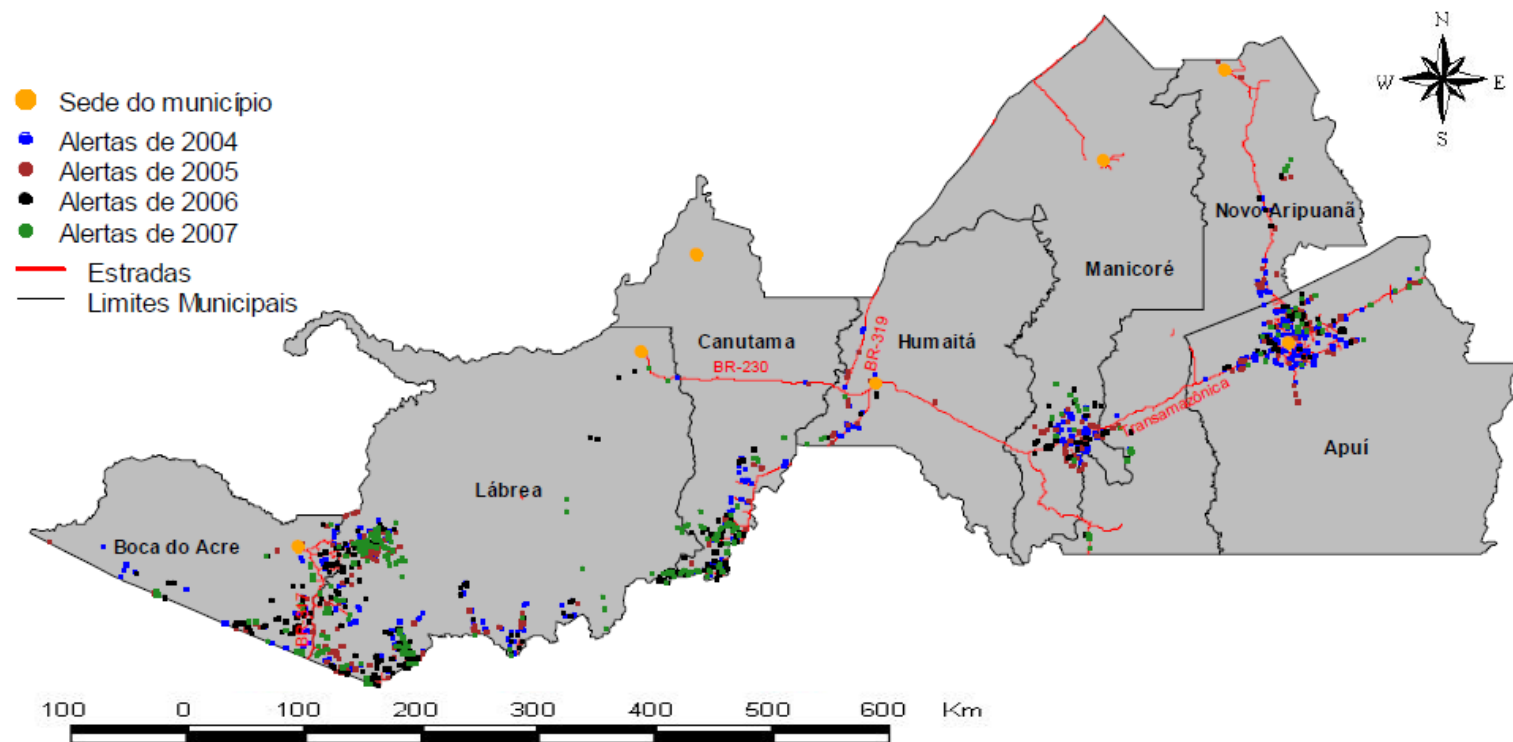


**Figura 3:** Localização no estado do Amazonas dos municípios de Boca do, Lábrea, Canutama, Humaitá, Manicoré, Novo Aripuanã e Apuí.



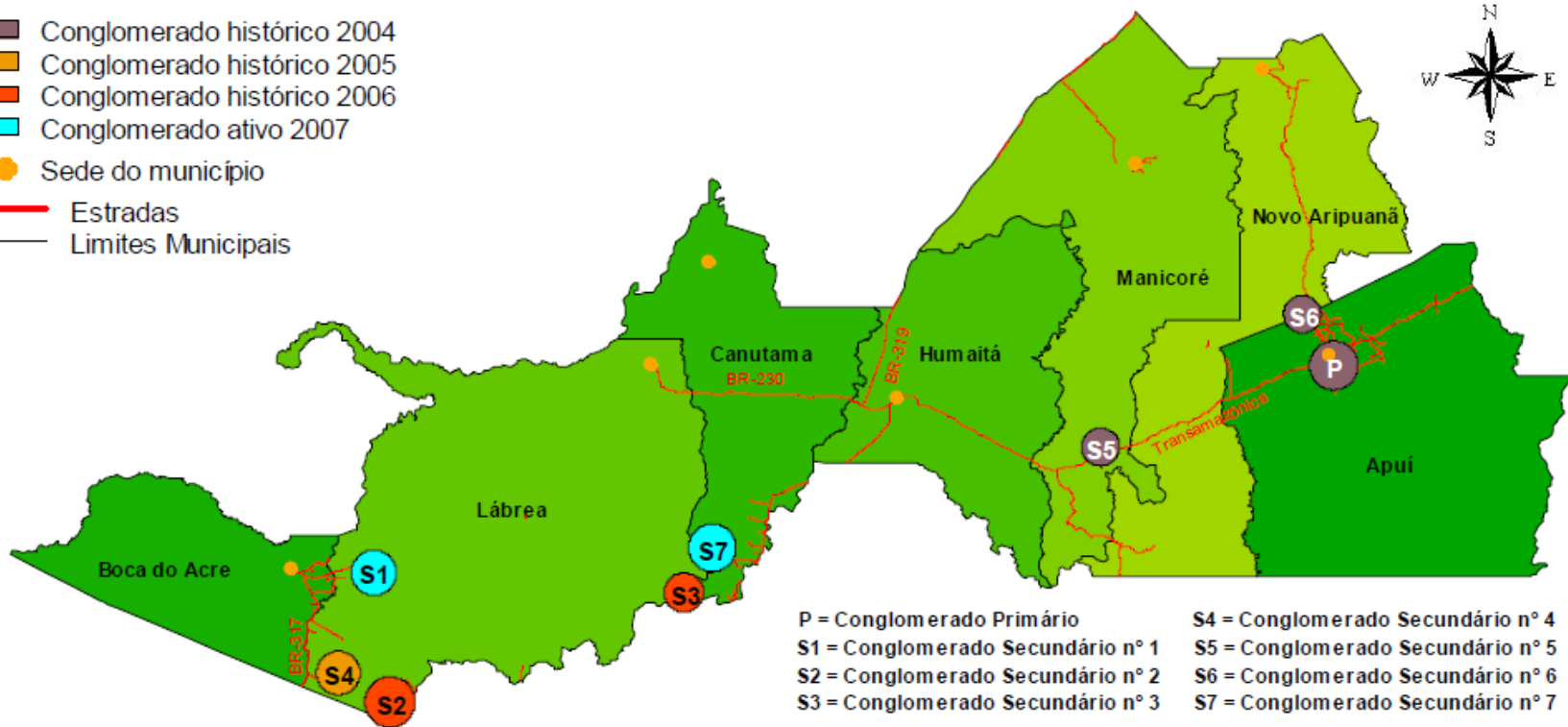
**Figura 4:** Mapa de alguns municípios da região sul do Amazonas, contendo as máscaras (*buffer*s) para as distâncias das estradas.





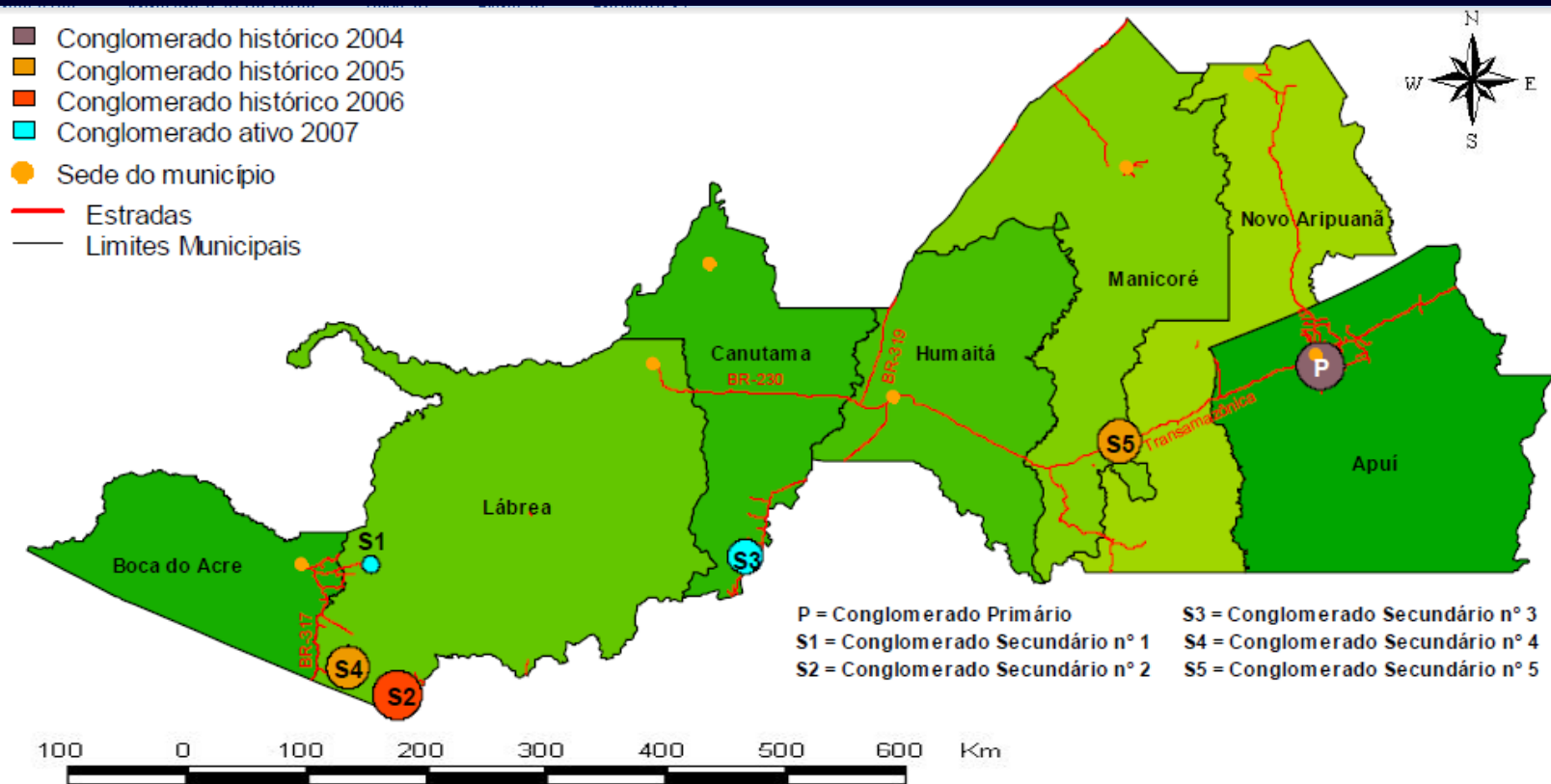
**Figura 9:** Mapa de alguns municípios da região sul do Amazonas contendo os alertas de desmatamento no período de 2004 a 2007.

- Conglomerado histórico 2004
- Conglomerado histórico 2005
- Conglomerado histórico 2006
- Conglomerado ativo 2007
- Sede do município
- Estradas
- Limites Municipais



- P = Conglomerado Primário
- S4 = Conglomerado Secundário nº 4
- S1 = Conglomerado Secundário nº 1
- S5 = Conglomerado Secundário nº 5
- S2 = Conglomerado Secundário nº 2
- S6 = Conglomerado Secundário nº 6
- S3 = Conglomerado Secundário nº 3
- S7 = Conglomerado Secundário nº 7





**Figura 15:** Mapa de alguns municípios da região sul do Amazonas, contendo os conglomerados detectados, a partir da inserção das covariáveis área desmatada, área protegida e distância de estradas, por ano, tamanho e hierarquia.

# Pontos

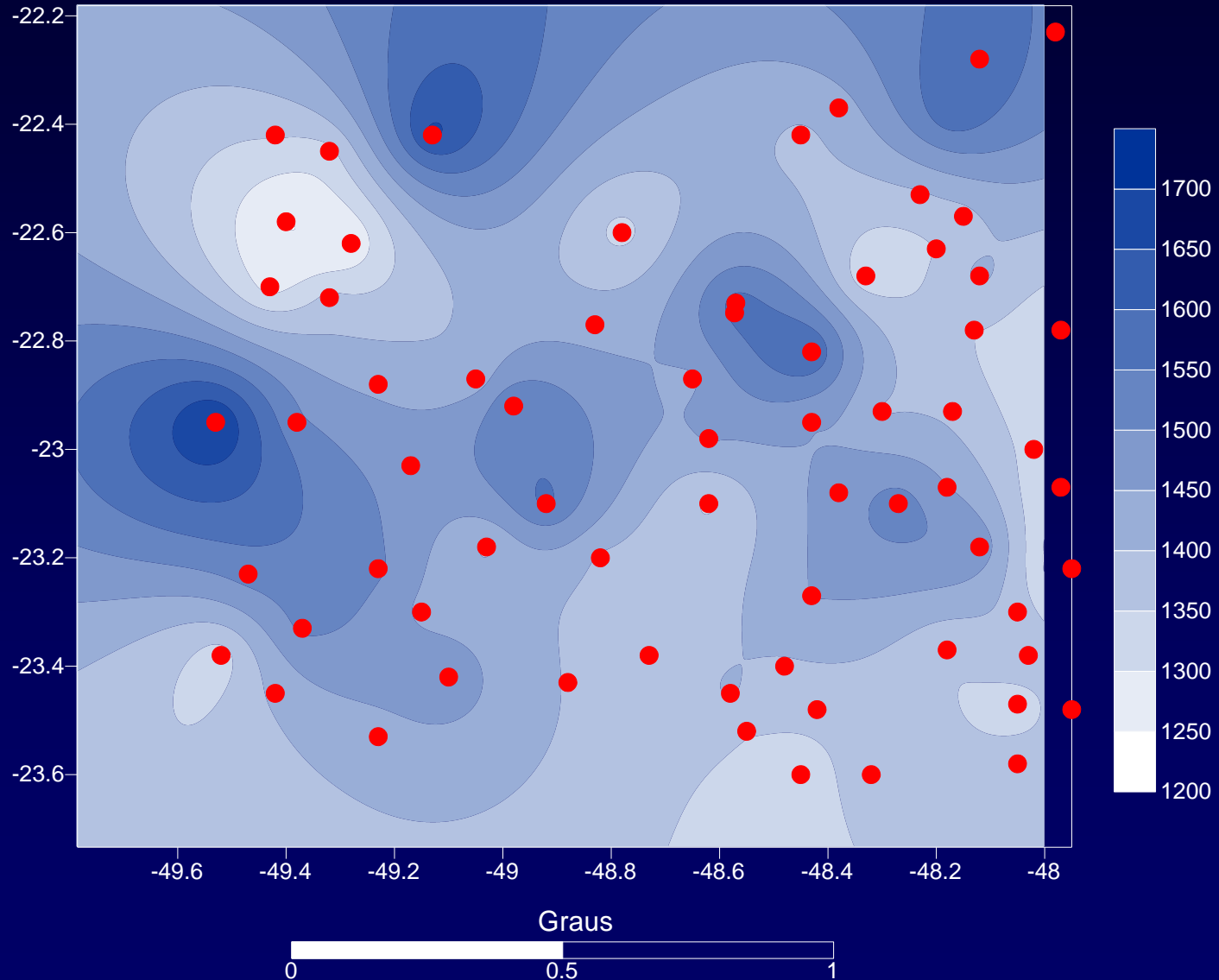
➤ como **amostras da superfície**

estações climatológicas, de monitoramento da qualidade do ar, de nível de ruído

- geoestatística

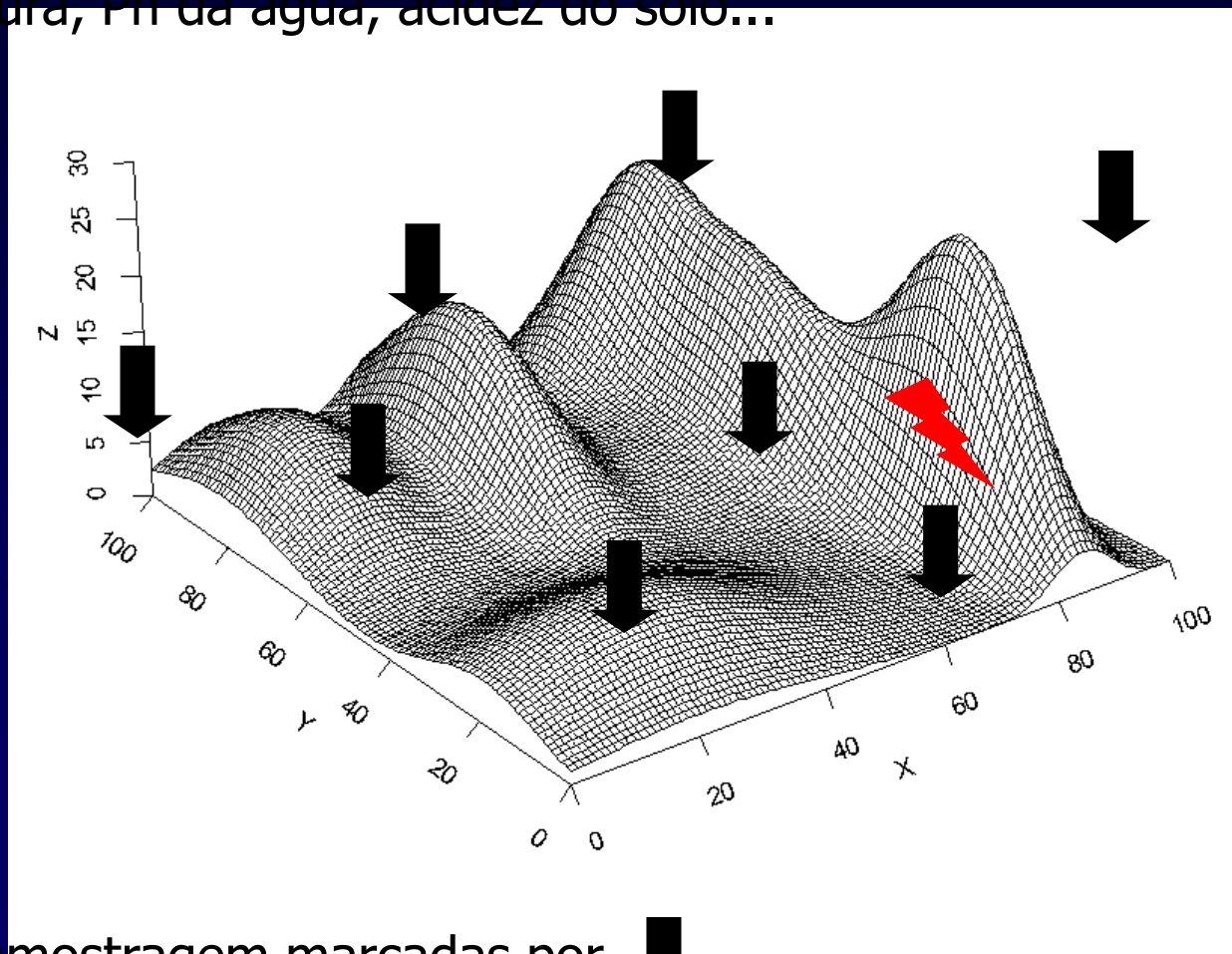
# Análise Geostatística

## Precipitação média (1960-1999)



# MNT – modelo numérico de terreno

Temperatura, Ph da água, acidez do solo...



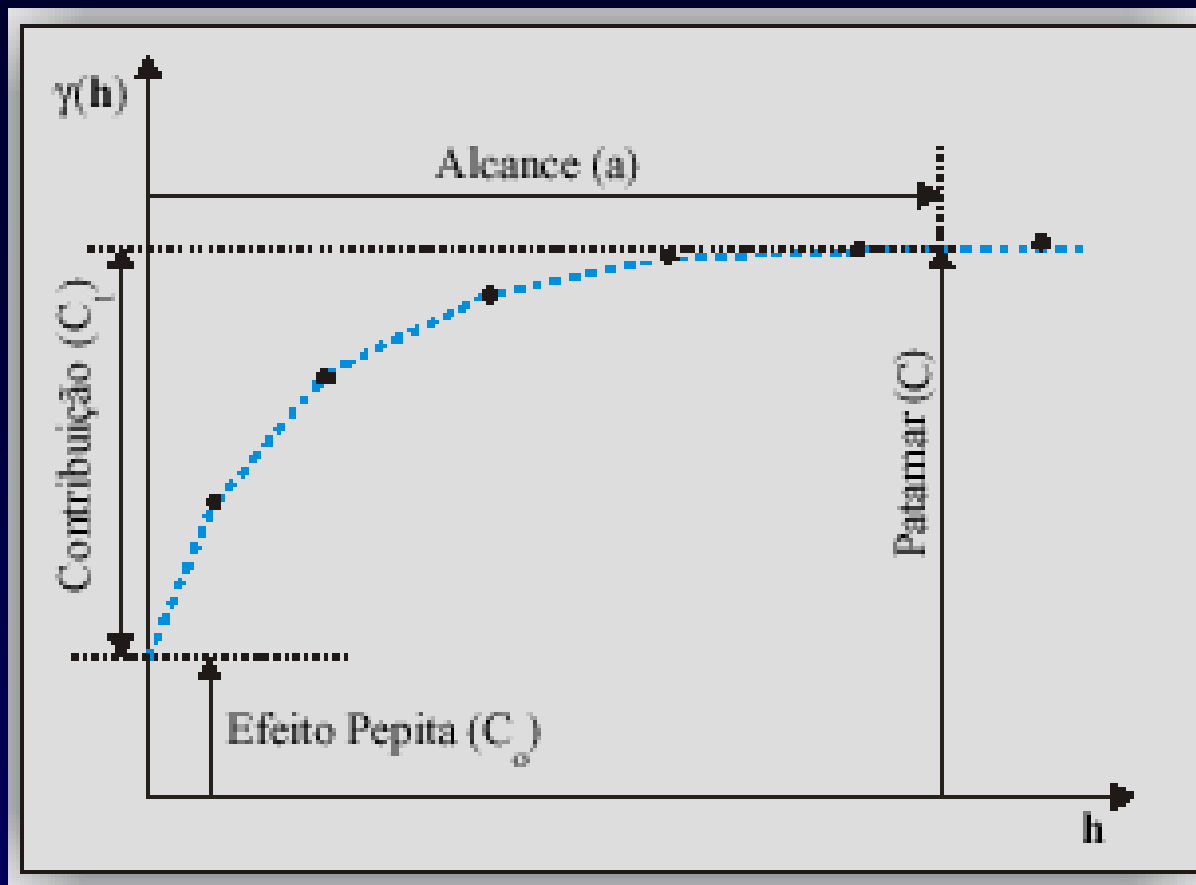
Estações de amostragem marcadas por  
Exemplo de valor predito mostrado como

Fonte: Renato Assunção (UFMG/Brasil)

A semi-variância é a medida de quão similar são os pontos próximos uns dos outros e a taxa na qual esta relação diminui com a distância.

A semi-variância de um conjunto de dados em uma distância  $h$  da localização  $x$ , dado  $n$  pontos caindo nesta categoria de distância e abreviado por  $\gamma^*(h)$  é um valor derivado dos dados e definido como:

$$\gamma^*(h) = \frac{\sum_{i=1}^n [y(x_i) - y(x_i + h)]^2}{2n}$$

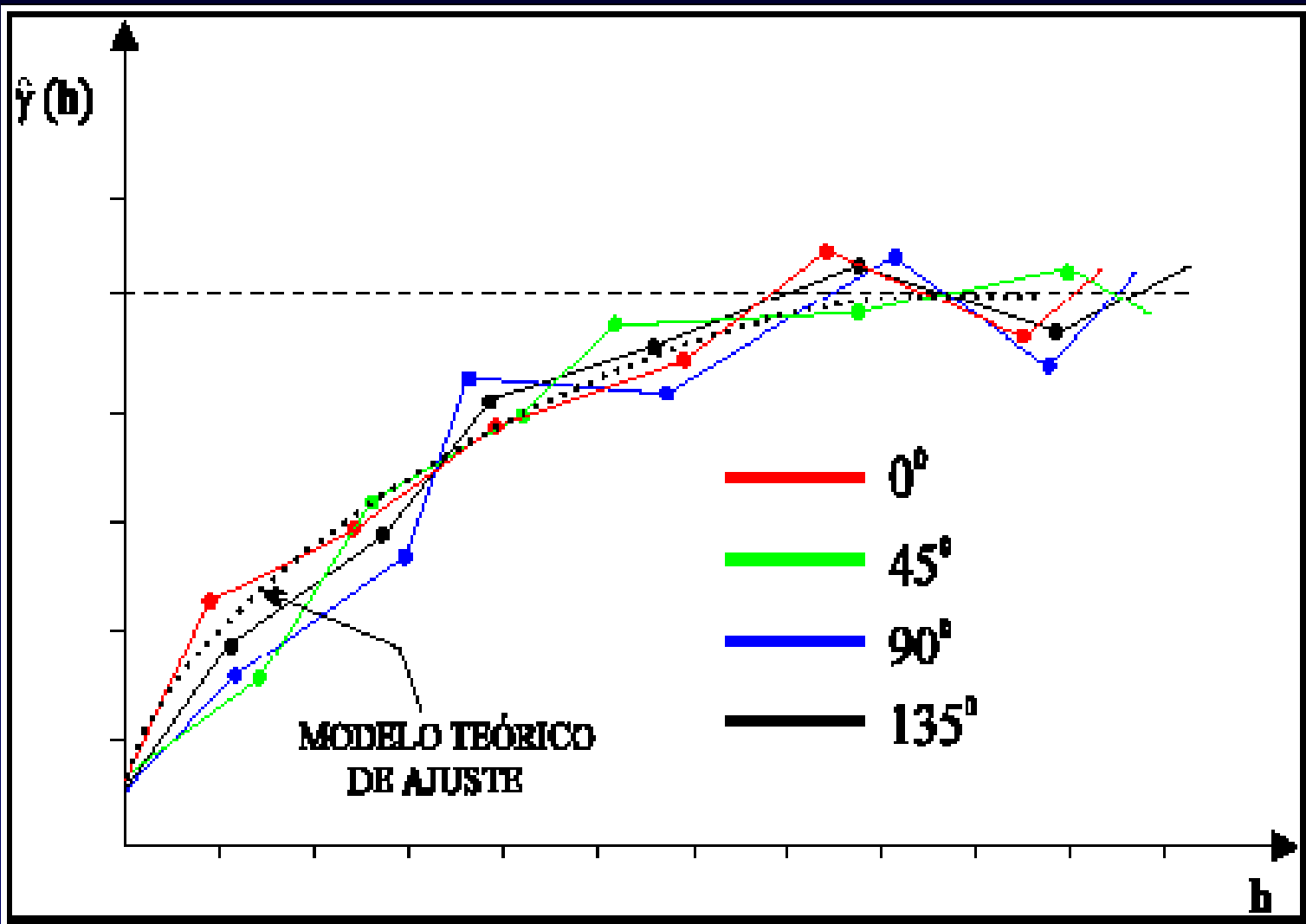


Forma característica de um semivariograma (LANDIM, 2005)

O efeito pepita ( $C_0$ ) é o ponto inicial da curva, onde a curva toca o eixo  $\gamma$ , quando  $h=0$ . O patamar ( $C$ ) é o valor de  $\gamma$  máximo da curva, o ponto em que não existe mais nenhuma correlação entre as variáveis, sendo assim a variância do conjunto de dados. O alcance ( $a$ ) é o ponto máximo onde existe autocorrelação espacial das variáveis.

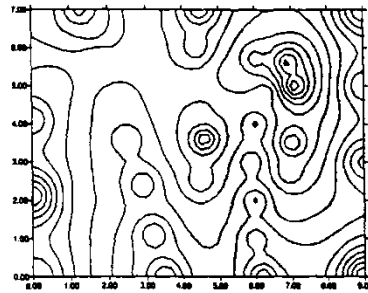


Figura 2 - Representação gráfica de semivariogramas isotrópicos.

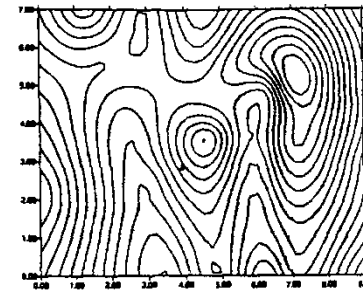


Representação gráfica de semivariogramas isotrópicos.

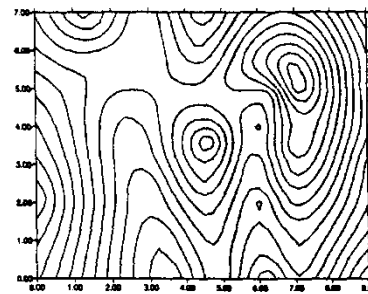
## Alguns métodos de interpolação por geoestatística



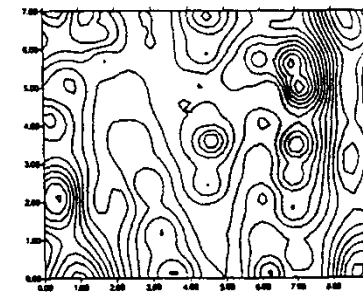
Inverse Distance



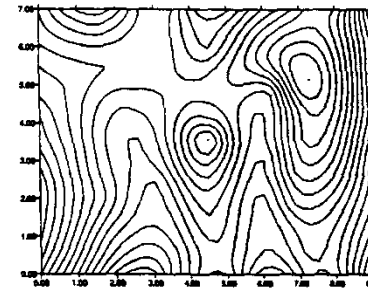
Radial Basis Functions



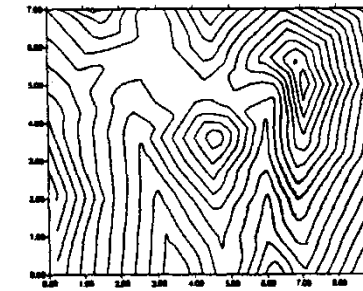
Kriging



Shepard's Method



Minimum Curvature



Triangulation

*A comparison of the different gridding methods. For these examples, the same file "demogrid.dat" was used, and all the defaults for the various methods were accepted. This data set contains 47 data points, irregularly spaced over the extent of the map. The data point locations are indicated with dots on the maps.*

# ÁREAS

## Padrões de autocorrelação espacial

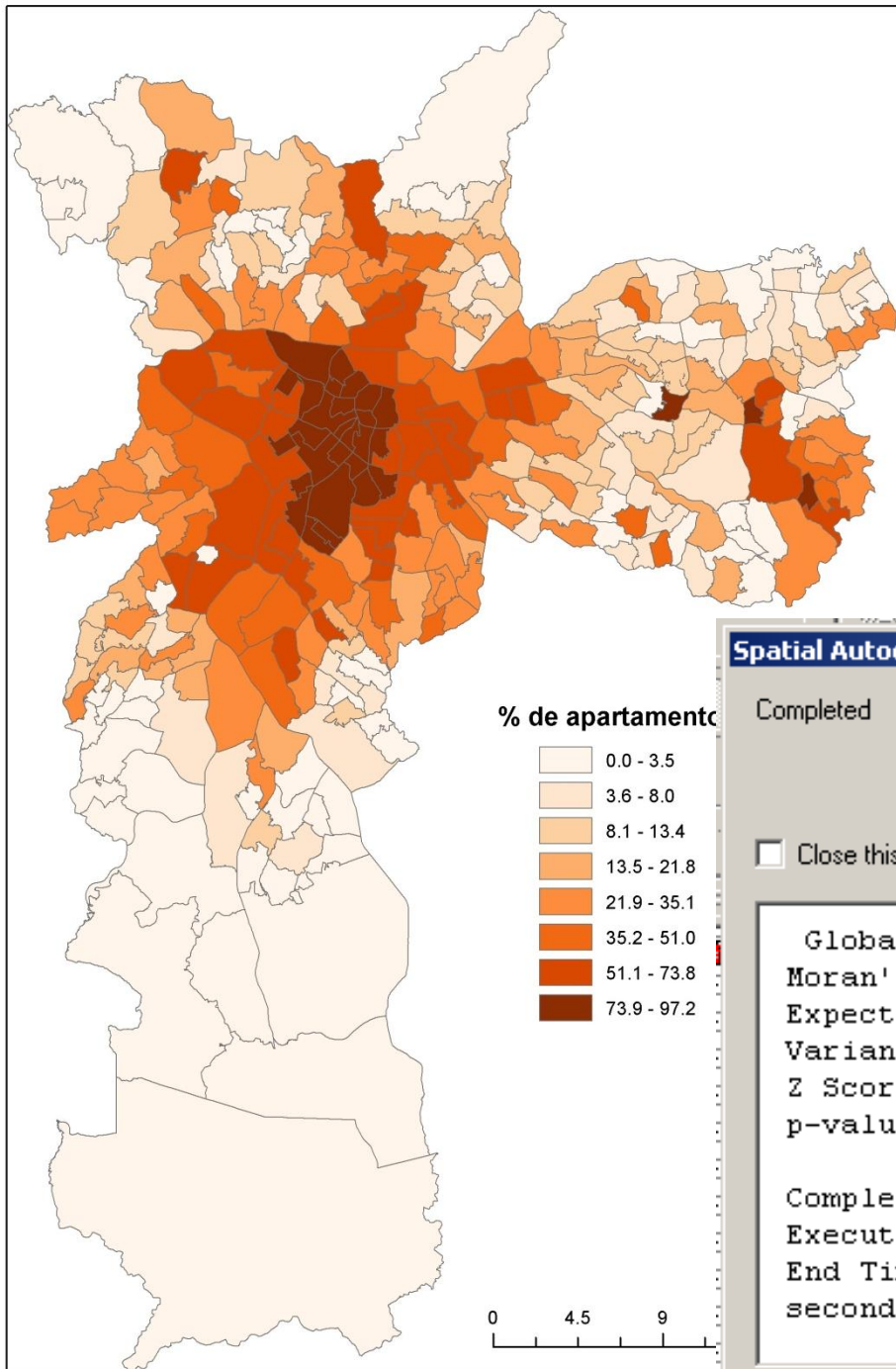


Negativa

Nula

Positiva

# Porcentagem de domicílios do tipo apartamento no Município de São Paulo



## Spatial Autocorrelation (Morans I)

Completed

Close

<< Details

Close this dialog when completed successfully

### Global Moran's I Summary

Moran's Index: 0.743968  
Expected Index: -0.003236  
Variance: 0.001198 |  
Z Score: 21.586833  
p-value: 0.000000

Completed script SpatialAutocorrelation...  
Executed (SpatialAutocorrelation) successfully.  
End Time: Fri Oct 02 14:06:22 2015 (Elapsed Time: 4.00 seconds)

# Moran's I

- Medida clássica de autocorrelação espacial
- Medidas de correlação entre  $X$  e  $X_{(\text{vizinhos})}$
- Usa a magnitude de desvio da média e não apenas se os valores são maiores ou menores do que a média
- Depende da definição de unidade de vizinhança

# Moran's I

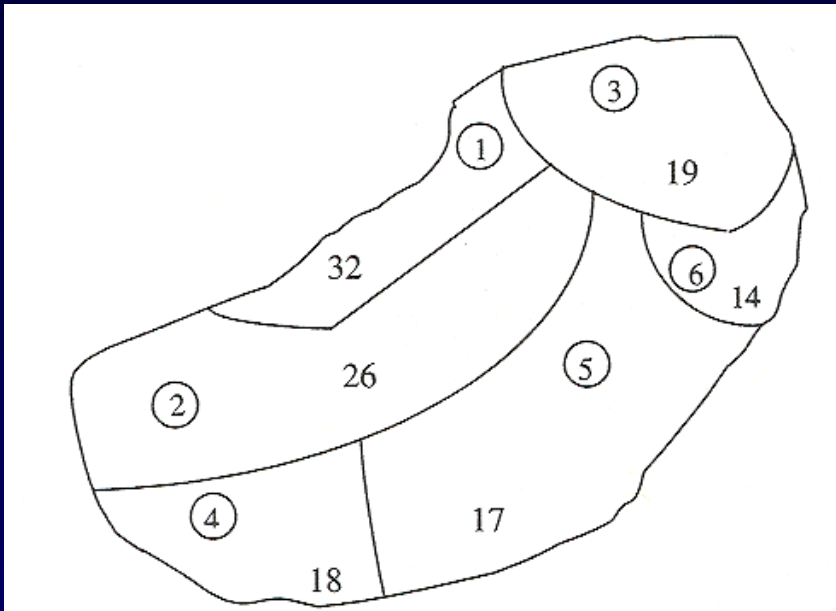
$$I = \frac{N \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{J \sum_i \sum_j w_{ij} \sum_i (x_i - \bar{x})^2}$$

onde,  $N$  é o número de pontos no estudo,  $J$  o número de pares de pontos examinados,  $x_i$  o valor (e.g., incidência da doença) do ponto  $i$ ,  $x_j$  é o valor do ponto  $j$  ( $x_i$  e  $x_j$  são dois pontos de vizinhos próximos) e  $w_{ij}$  é a matriz de ponderação

# Coeficiente I de Moran

- $I$  é aproximadamente  $\text{corr}(X, Y)$  onde  $Y$  é a média dos vizinhos
- Varia de -1 a +1 e 0=ausência de correlação
- $Y$  é também chamado de spatial lag or  $w_X$

# Moran's $I$ com vizinhos definidos como unidades contíguas



unit	X	Y or w_X
1	32	$(26+19)/2$
2	26	$(32+19+18+17)/4$
3	19	$(32+26+17+14)/4$
4	18	$(26+17)/2$
5	17	$(26+19+18+14)/4$
6	14	$(19+17)/2$



# Análise de áreas

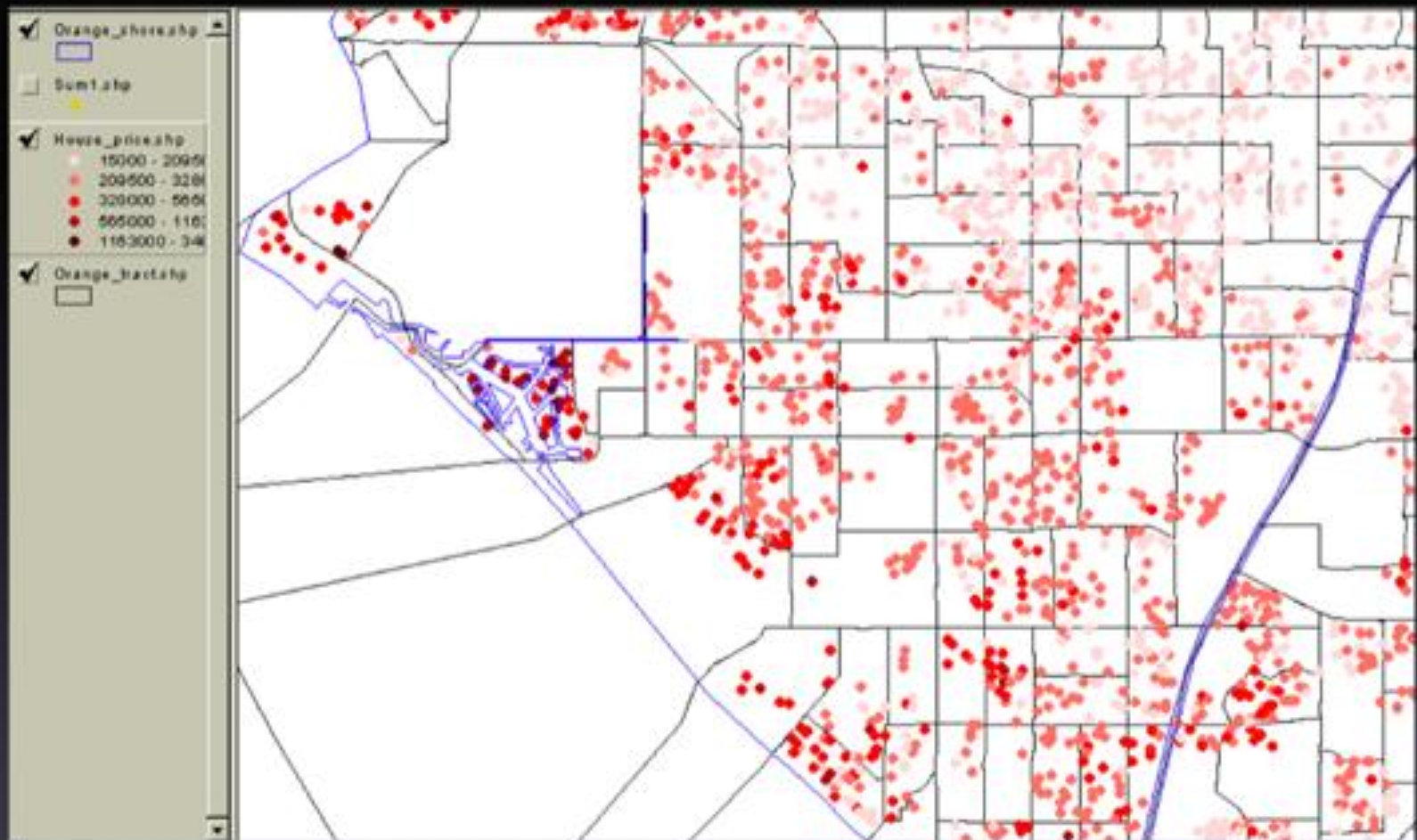
- dados oriundos de levantamentos populacionais
- supõe-se homogeneidade interna (agrupamentos aleatórios de indivíduos/moradas que tendem a ser semelhantes em relação a outras áreas)
- as unidades são definidas por critérios operacionais
- em diversas regiões, as unidades amostrais apresentam diferenças importantes em população e área

# Efeitos de escala na análise de dados de área

- Tendência de agregação geográfica para preservar a confidência de registros individuais
- No Censo, os dados são agregados por setor censitário
- cada setor corresponde à capacidade de levantamento do recenseador, variando em torno de 200 a 400 domicílios

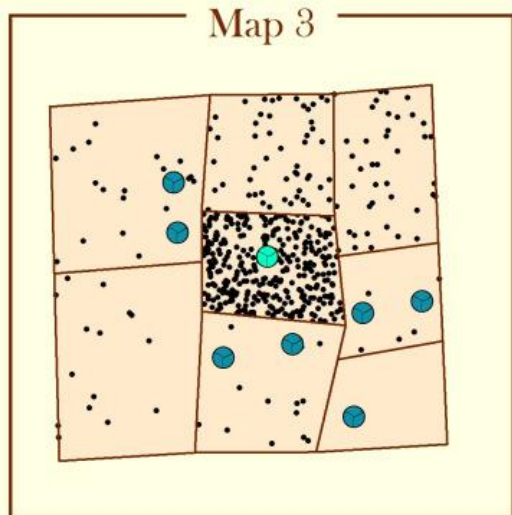
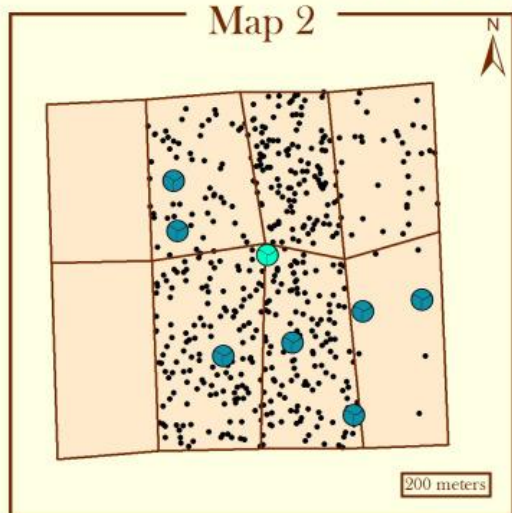
# Problema com dados agregados por área

- Para uma mesma população estudada, a definição espacial das fronteiras afeta os resultados obtidos
- Pode-se obter resultados diferentes simplesmente alterando as fronteiras destas zonas – (**MAUP** – *modifiable areal unit problem*) “problema da unidade de área modificável”



## Spatial Heterogeneity of House Prices Within Tracts Orange County, CA

# Revisiting John Snow's Cholera Map



- Cholera Deaths**
- 1 - 2
  - 3 - 5
  - 6 - 10
  - 11 - 15
- Broad St. pump** ●
- Other pumps** ●
- Aggregated areas:**
- 1 dot = 1 case
- Dots do not show actual location

The map above (map 1) shows Cholera cases in London during an outbreak of the disease in 1854. The data was collected by Dr. John Snow (1813-1858) and in fact, the background map is his own map, that he prepared for the Board of Health. By mapping individual Cholera cases he made the connection between a contaminated water pump (on Broad Street) and the spread of the disease.

One of the challenges of mapping diseases is choosing the appropriate areal unit. These are usually census tracts or zip codes, where all cases that fall within that particular area are collected. Maps 2 & 3 show two types of areas where cases from map 1 have been counted within each area. Comparing maps 2 & 3 there is a striking difference in the number of cases and spatial distribution. This problem is referred to as the Modifiable Areal Unit Problem (MAUP).



John Snow

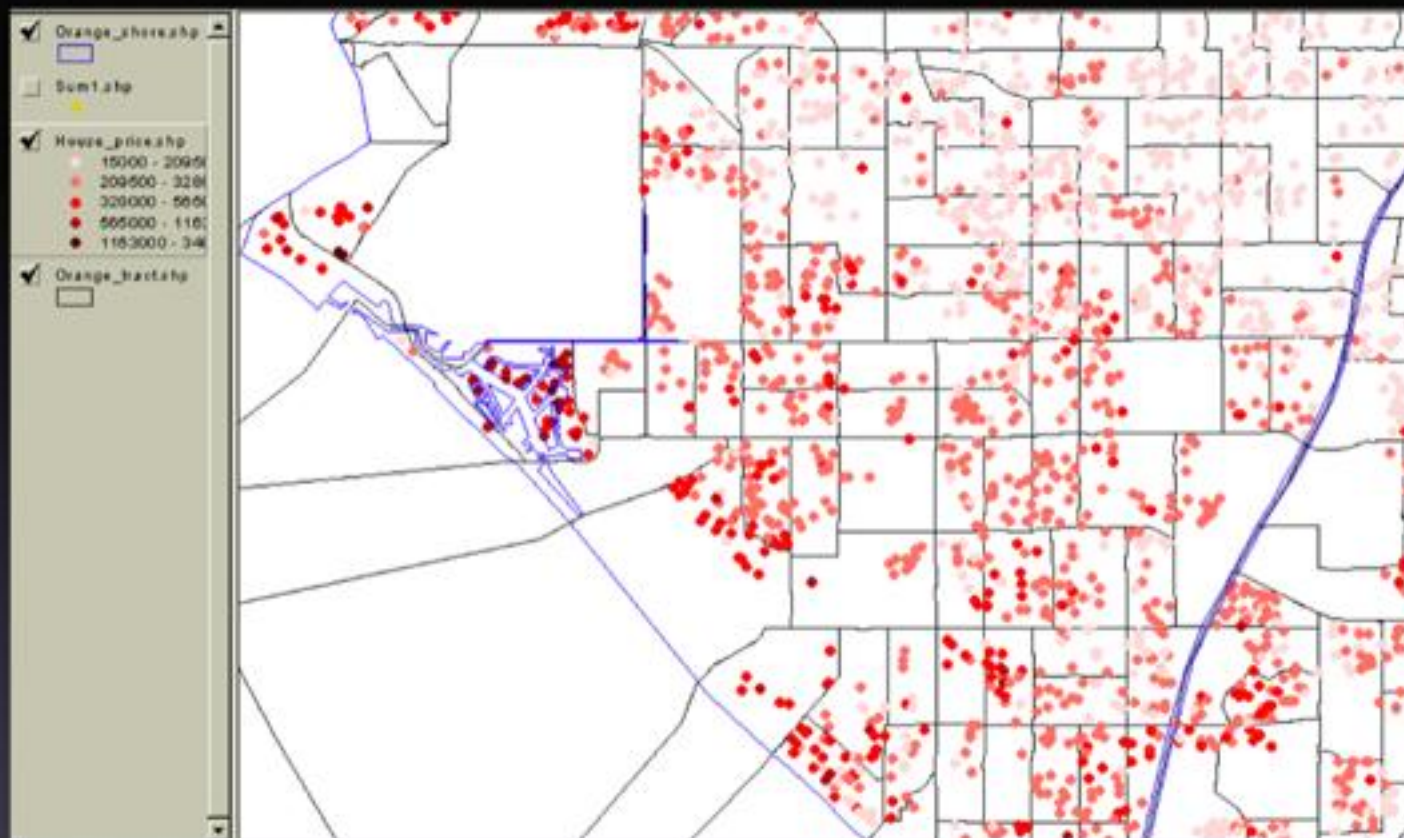
# Problema com dados agregados por área

- Falácia ecológica: devido aos efeitos de escala e de agregação de áreas, os coeficientes de correlação podem ser inteiramente diferentes no indivíduo e nas áreas.

Envolve a conclusão IMPRÓPRIA de relacionamentos a nível individual a partir de resultados agregados ao nível de unidade de área

# FALÁCIA ECOLÓGICA

- Os resultados estatísticos têm validade dependente da unidade de área e do reconhecimento dos problemas existentes nas conclusões decorrentes de dados agregados;
- Propriedade inerente aos dados agregados por áreas;
- A agregação tende a AUMENTAR a correlação entre as variáveis e REDUZIR flutuações estatísticas.



## Spatial Heterogeneity of House Prices Within Tracts Orange County, CA



# Interpretação e Limitações

Métodos exploratórios **não explicam**

- sugerem padrões potencialmente interessantes
- sugerem hipóteses
- *Clusters e Outliers*
- são univariados
- podem mascarar associação multivariada, problemas de escala, ignorar heterogeneidade etc.