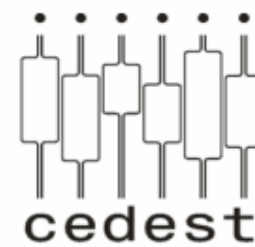




PUC/SP



# Território e Desigualdade: Desafios Metodológicos na Análise de Dados Socioterritoriais

Gilberto Câmara

Coordenação de Observação da Terra - INPE

CEDEST – Centro de Estudos de Desigualdades Socioterritoriais

Mesa Redonda “A Estatística e as Grande Questões Sociais Brasileiras”, SINAPE, Caxambu, julho de 2004



Licença de Uso: Creative Commons Atribuição-Use Não-Comercial-Compartilhamento  
<http://creativecommons.org/licenses/by-nc-sa/2.5/br/>

# CEDEST

- Centro de Estudos de Desigualdades Socio-Territoriais
  - Projeto Financiado pela FAPESP (2001-2004)
  
- Missão CEDEST
  - Investigar as desigualdades socioterritoriais através de análise espacial para *construir indicadores de diagnóstico e predição* em políticas sociais.
  - Estabelecer um diálogo permanente entre análise sociológica e técnicas matemático-computacionais, para uma crítica substantiva e subjetiva dos procedimentos analíticos.

# Colaboradores em Estudos Estatísticos

- Suzana Druck (EMBRAPA/Cerrados)
- Paulo Justiniano Ribeiro (UFPr)
- Corina Freitas (INPE)
- Trevor Bailey (Univ. Exeter – Reino Unido)
- Renato Assunção (UFMG)
- Marília Sá Carvalho (FIOCRUZ)

# Motivação

- A exclusão social existe?
- Como podemos representar quantitativamente conceitos sociológicos como ‘exclusão social’, ‘vulnerabilidade’, ‘desigualdade’?
- Qual a contribuição da Estatística nas questões sociais brasileiras?
- Como a estatística pode operar em ambiente multidisciplinar?

# Motivação

- O que exatamente acontece quando buscamos representações computacionais para conceitos como ‘exclusão social’, ‘vulnerabilidade’, ‘desigualdade’?
- Será que o ambiente computacional engendra processos de descoberta e invenção distintos das metodologias tradicionais das ciências sociais?
- *Que critérios deve satisfazer um conceito sociológico para que seja utilizável em estudos quantitativos para subsidiar políticas públicas universais e territoriais?*

# Como representar a realidade?



## Como representar conceitos socio-territoriais?

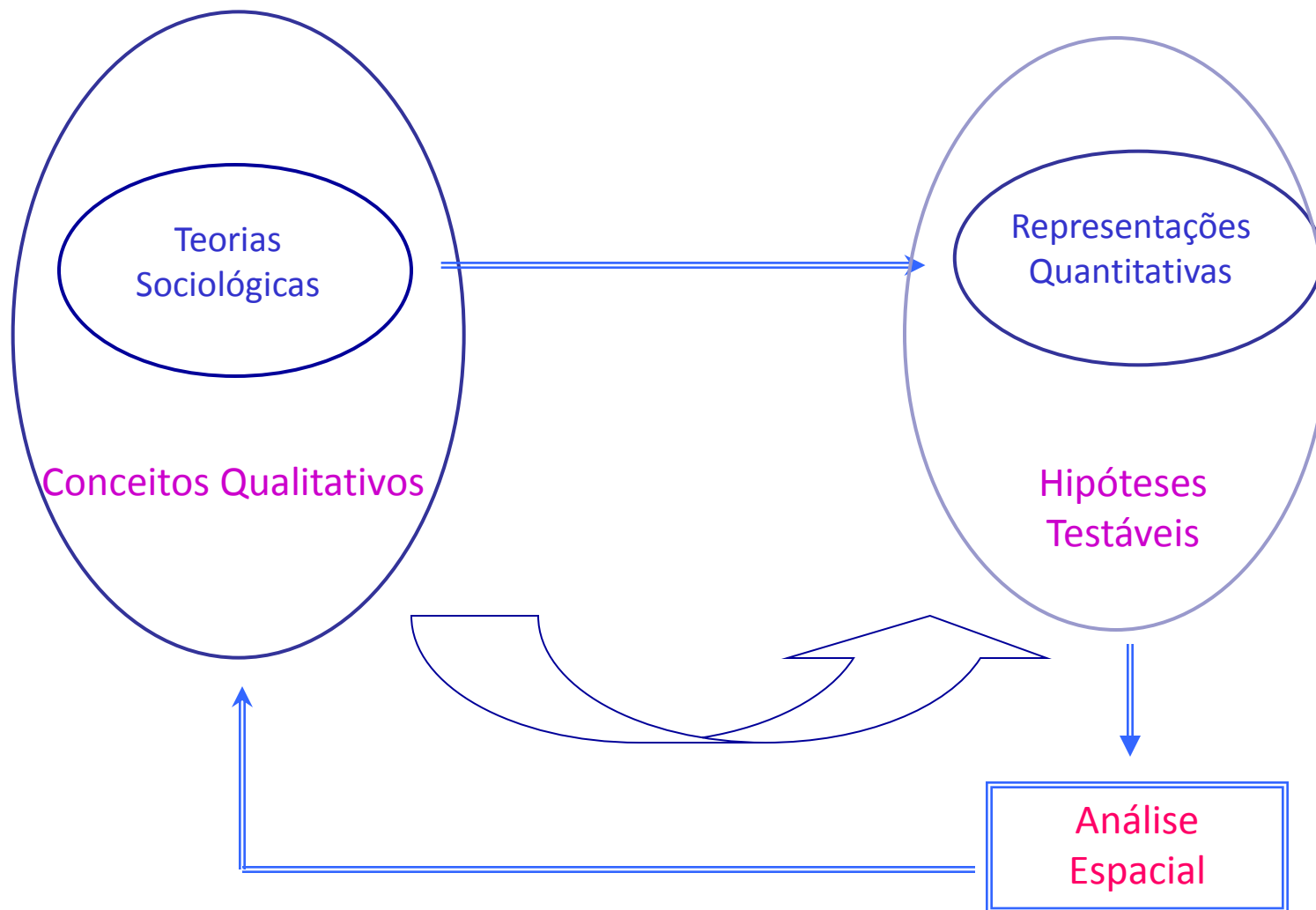
- O conceito deve ser passível de ser associado a propriedades mensuráveis.
- Estas propriedades devem ser medidas no território e devem permitir diferenciação entre a população estudada.
- Os resultados quantitativos e os modelos matemáticos utilizados devem ser validados em estudos de campo, que devem incluir dimensões objetivas e subjetivas do problema.

# Conhecimento como Construção

- Especialistas de áreas de Políticas Públicas
  - Teorias gerais sobre os fenômenos
  - Conceitos organizadores de sua pesquisa (como 'exclusão' ou 'vulnerabilidade').
- Dos conceitos para a representação computacional
  - Modelos inferenciais quantitativos.
  - Testes de validação e de corroboração com Análise Espacial.
- Uso dos resultados numéricos
  - Corroborar ou ajudar a rejeitar conceitos sociológicos qualitativos.



# Conhecimento como Construção

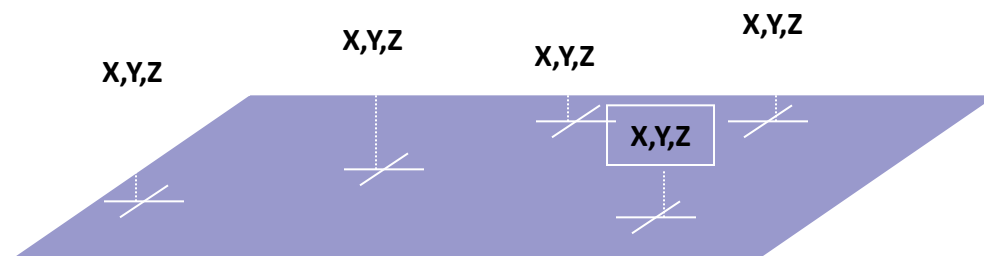


# Espaço como Elemento de Análise

- “Primeira Lei da Geografia”
  - “Todas as coisas se parecem, coisas mais próximas são mais parecidas que aquelas mais distantes” (Tobler)
  - Como medir a semelhança entre objetos no espaço ?
- Variabilidade Espacial
  - Aumento da variância de uma grandeza com a distância (variograma e correlograma)
  - Grau de similaridade entre vizinhos (índices de autocorrelação espacial)

# Dos Dados às Representações

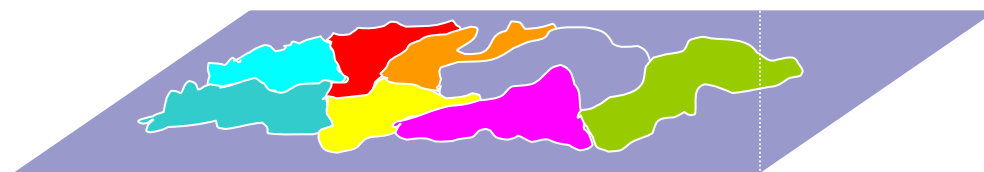
Eventos / Amostras



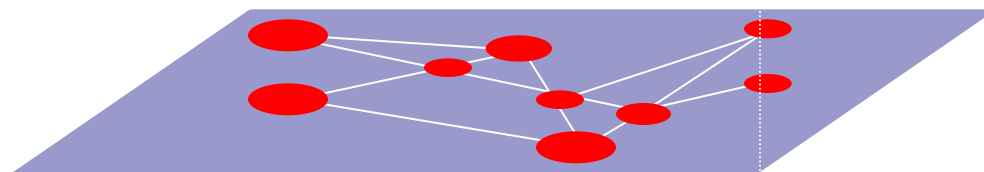
Superfícies / Grades Regulares



Dados de Área- Polígonos



Redes e Dados de Fluxo

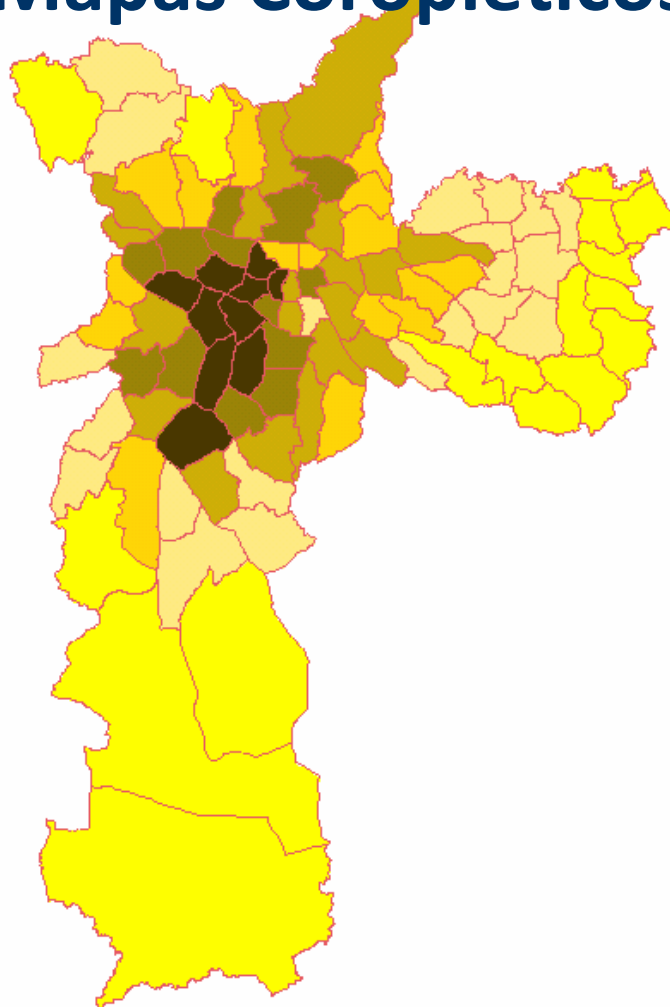


# Fotos Aéreas

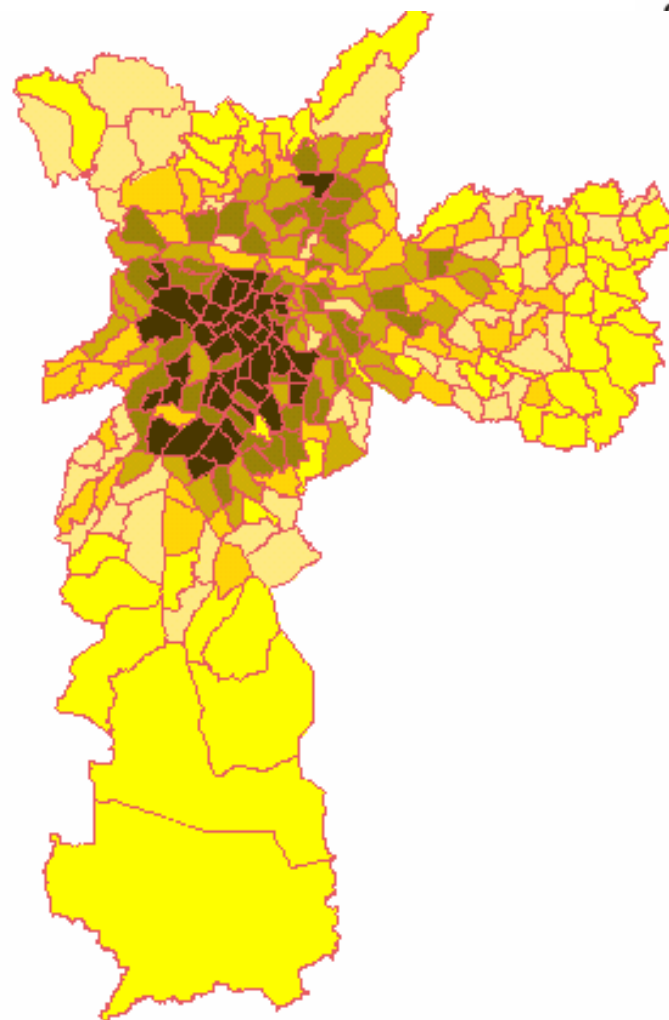


Favela da maré, Rio de Janeiro - 2001

# Mapas Coropléticos



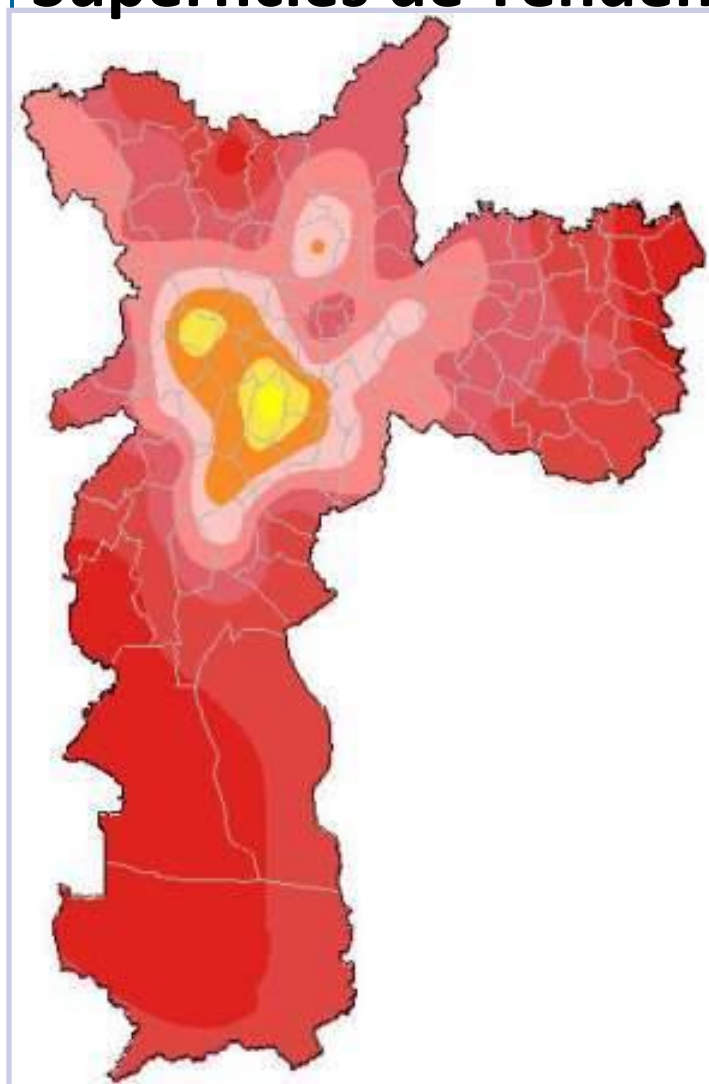
São Paulo - 96 districts per capita income



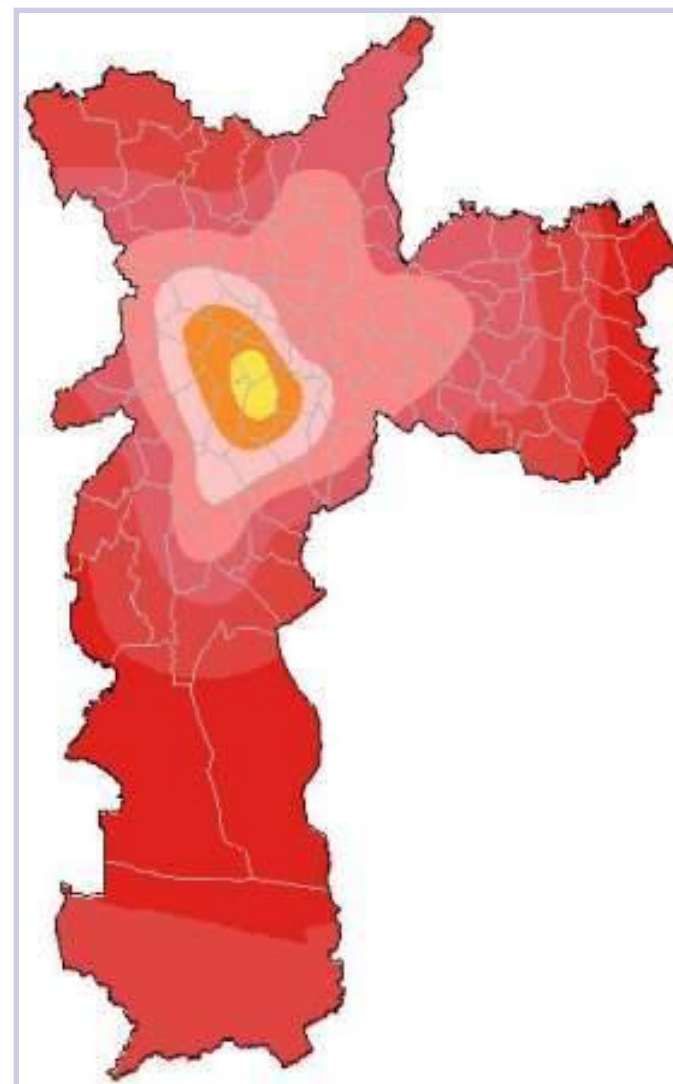
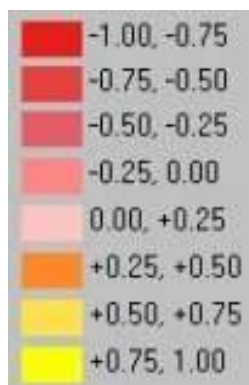
São Paulo – 270 survey areas per capita income



# Superfícies de Tendência



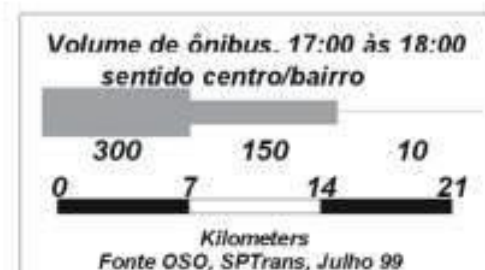
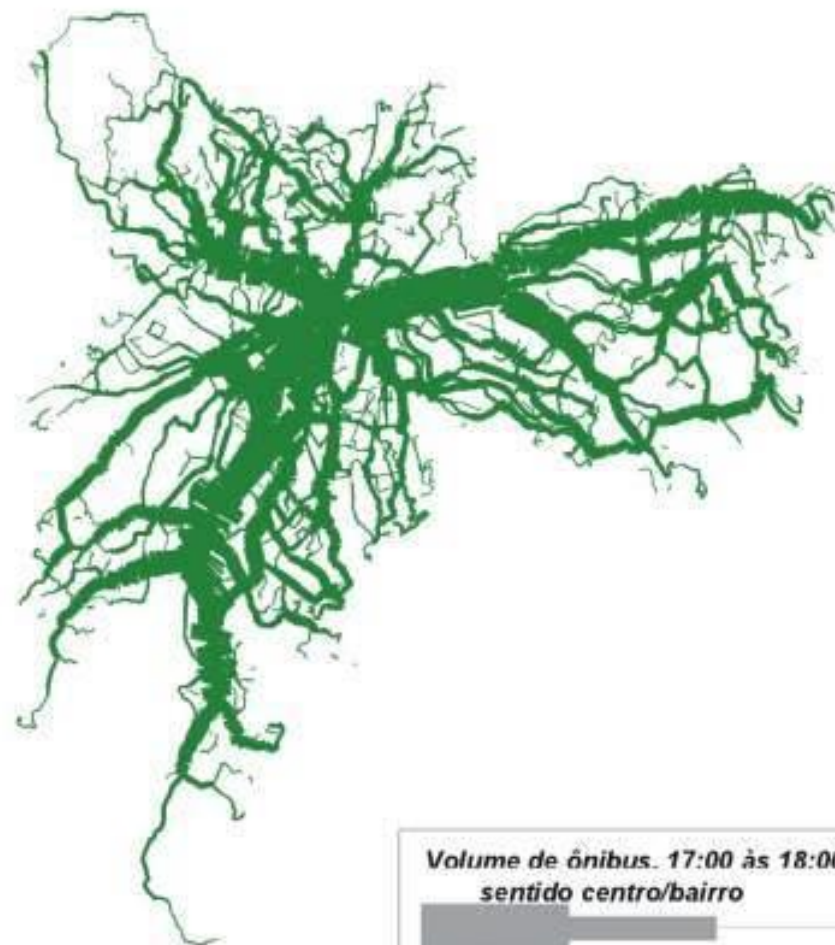
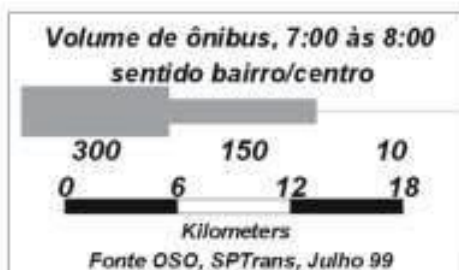
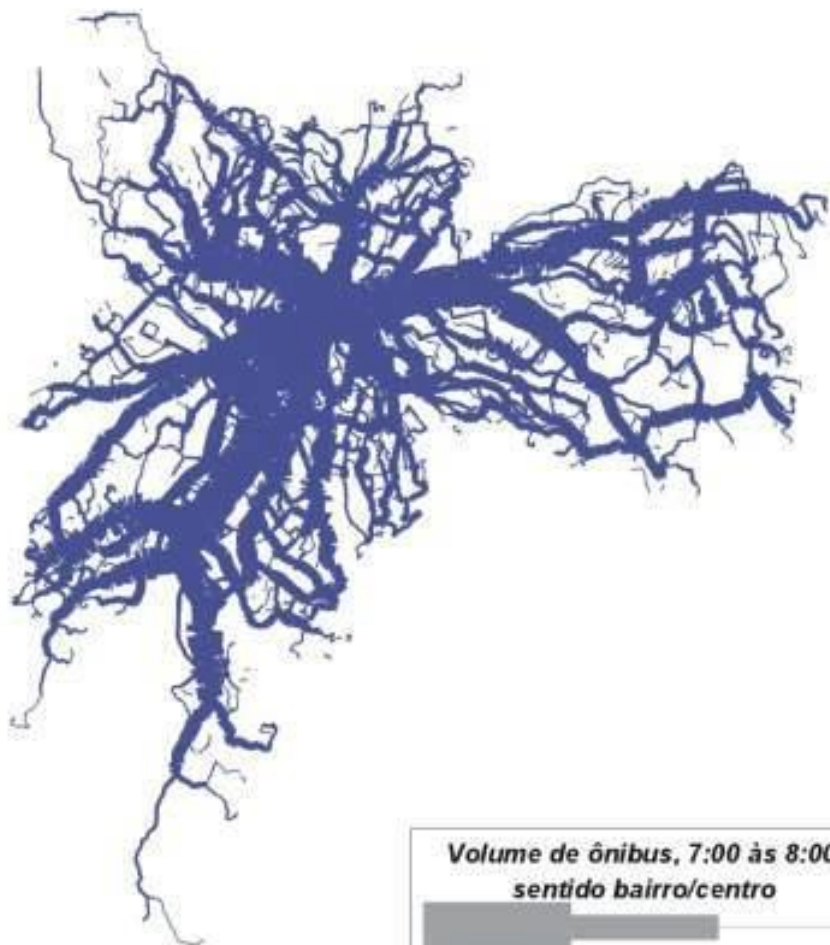
iex



Social Exclusion 1995

Social Exclusion 2002

# Fluxos



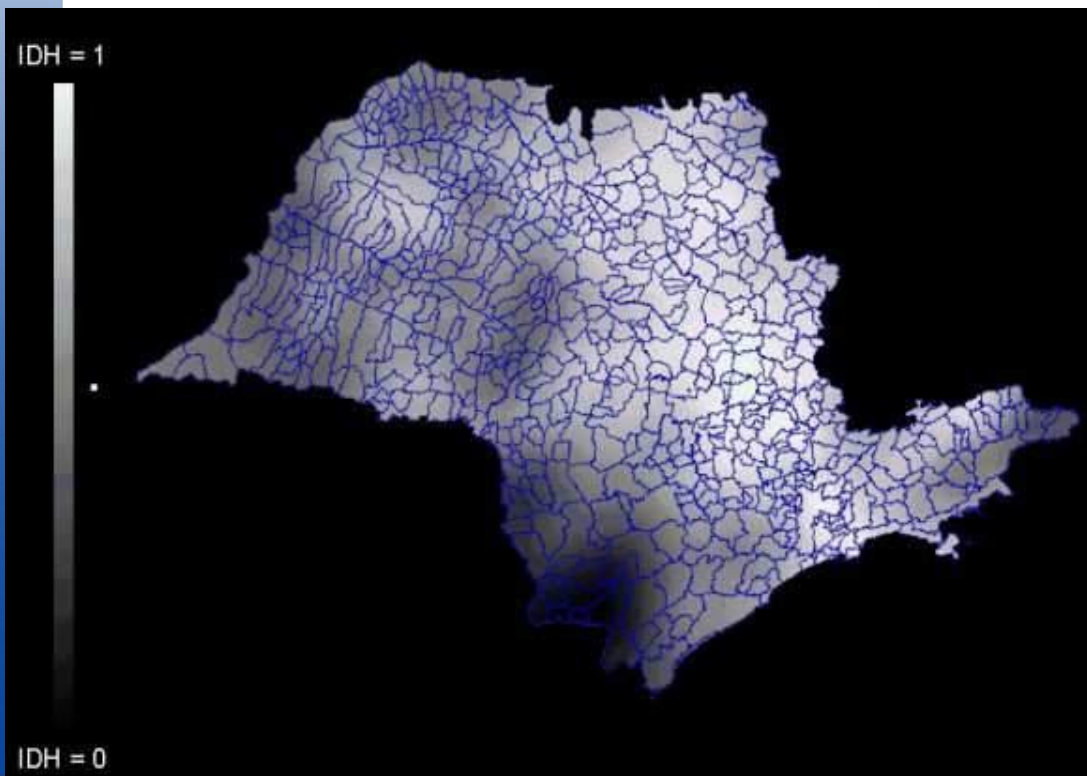
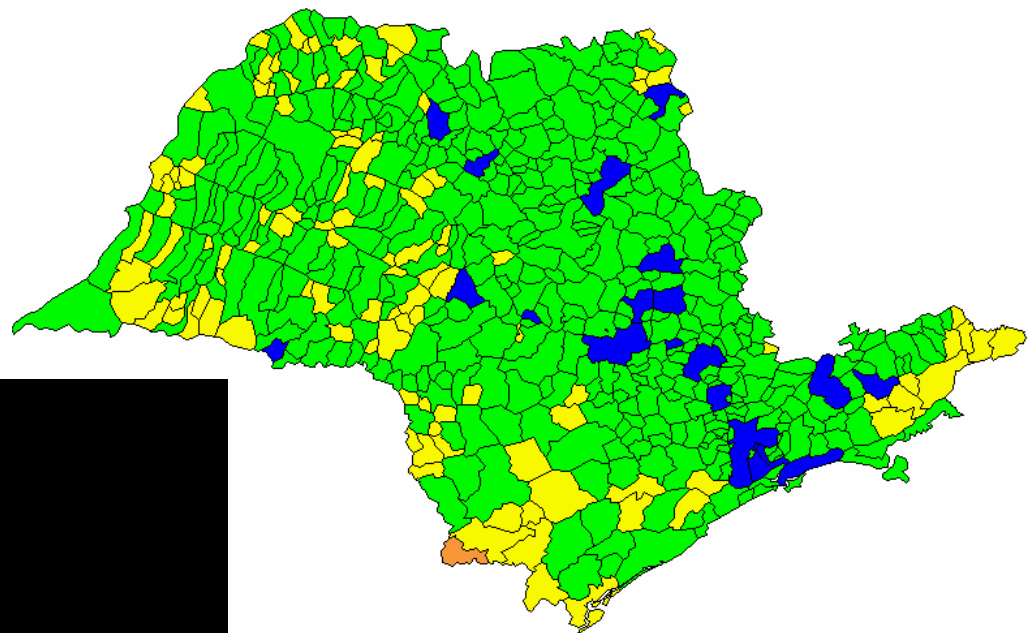
# Topografia Social: Das Áreas às Superfícies

- Boa parte dos estudos de políticas públicas está associada a fenômenos contínuos no espaço
  - Risco de homicídio
  - Densidade demográfica
  - Distribuição de renda
- Dados sócio-econômicos são coletados pontualmente, mas agregados por partição espacial
  - Setor censitário, distritos, municípios
  - Partições espaciais são unidades de coleta e não unidades de análise
- Agregamento de dados em regiões esconde variações internas, e pode criar limites artificiais



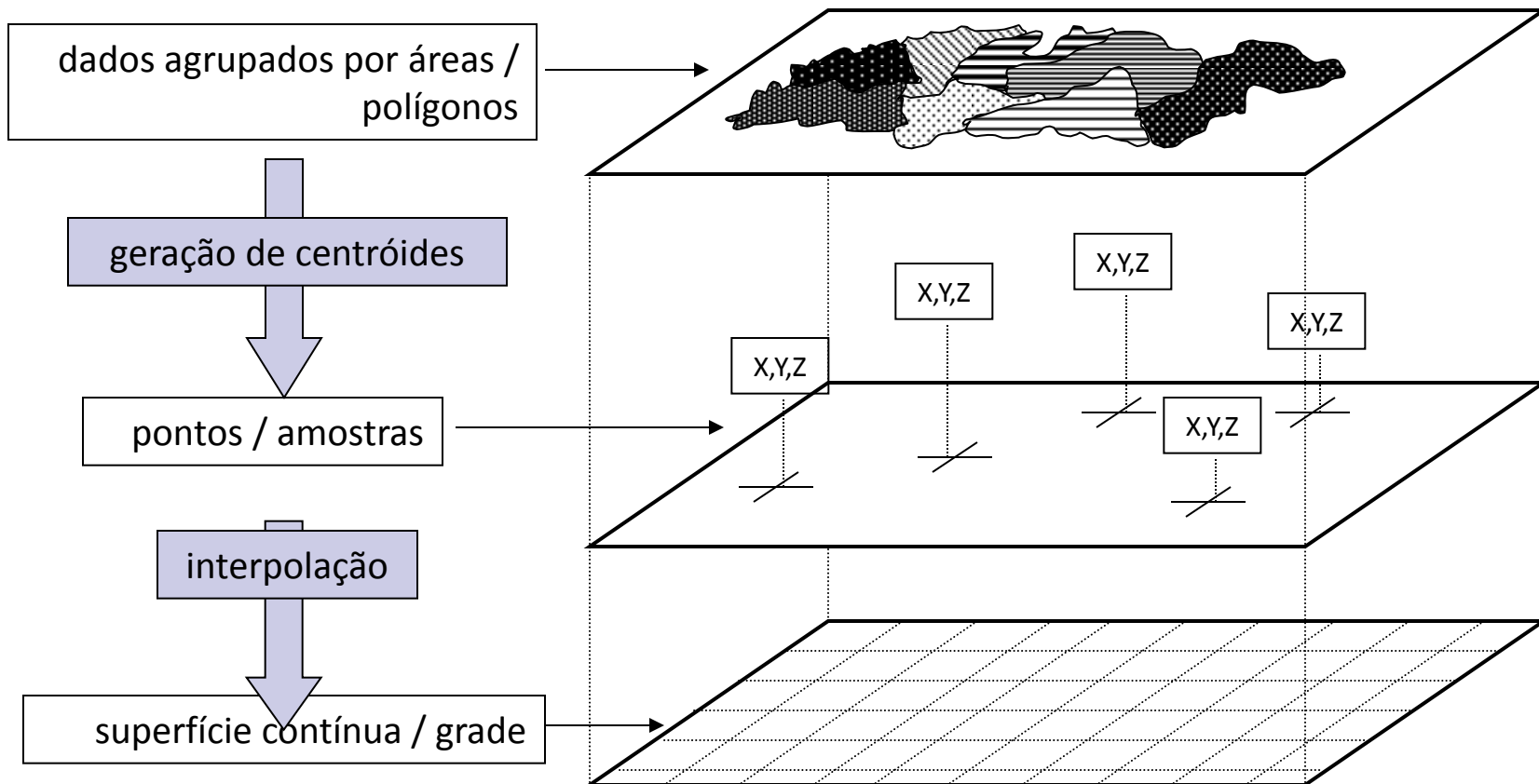
# Topografia Social: Percepções do Espaço

Espaço como uma subdivisão planar

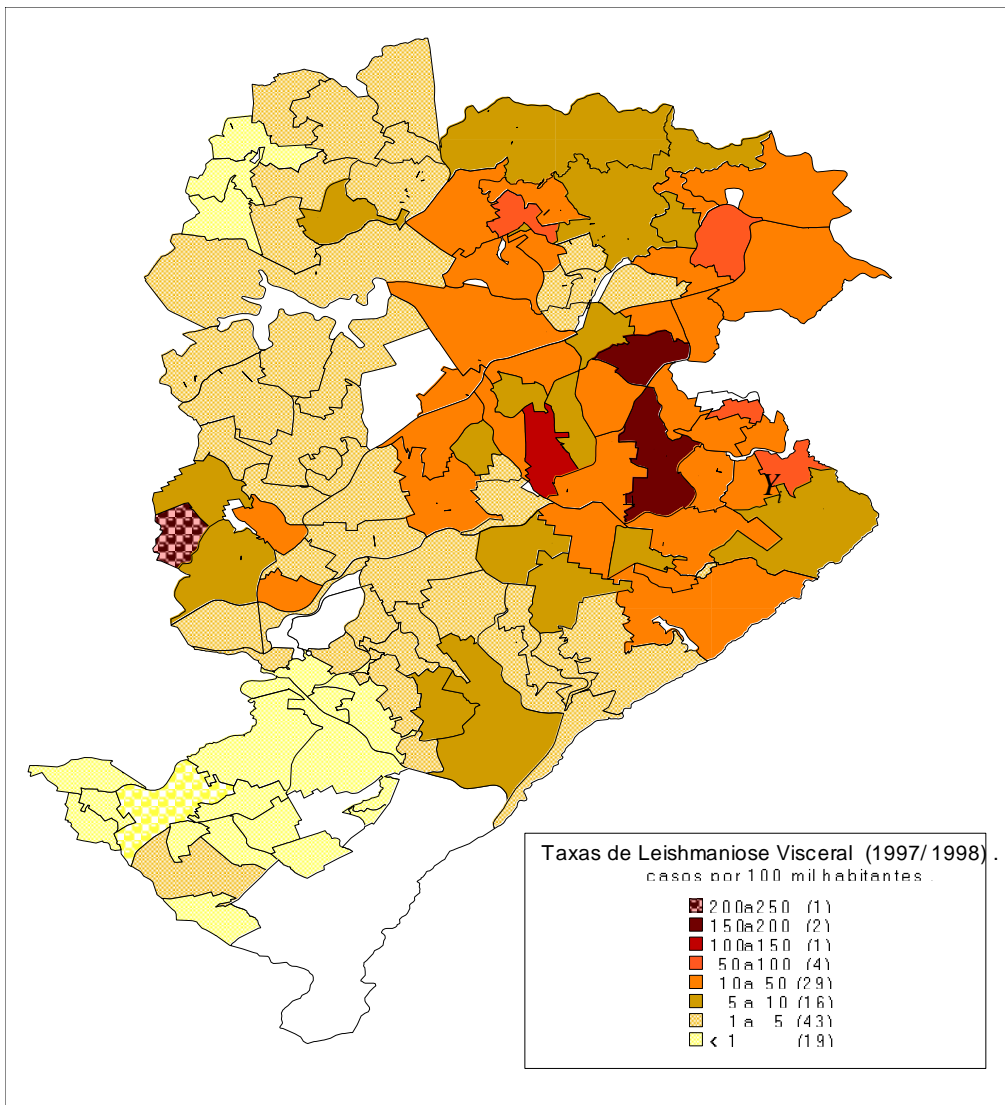


Espaço como uma superfície contínua

# Das Áreas às Superfícies



# Models of Discrete Spatial Variation

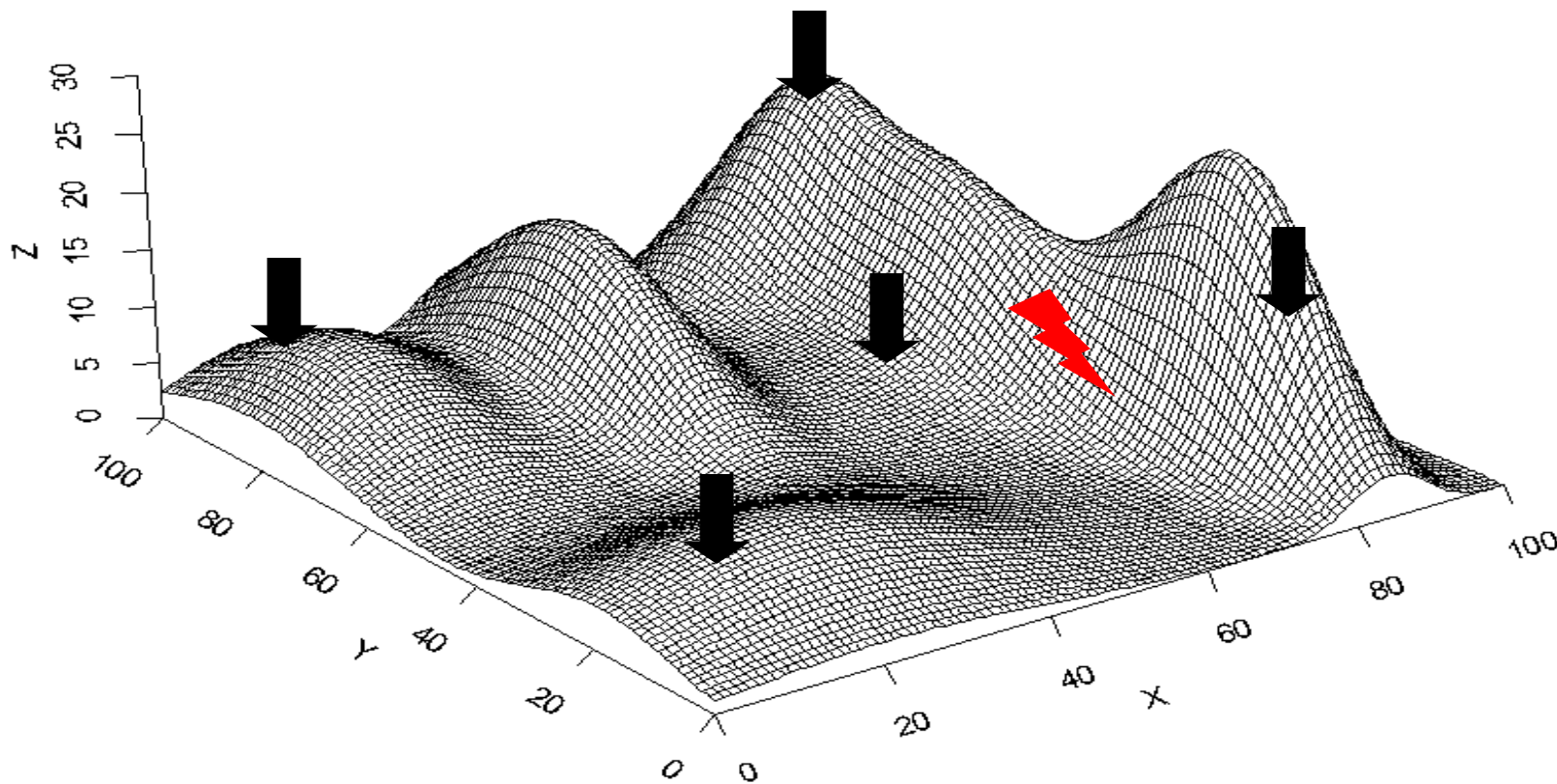


$Z_i =$  Random variable in area  $i$

- n° of ill people
- n° of newborn babies
- per capita income

# Models of Continuous Spatial Variation

Temperature, Water ph, soil acidity...

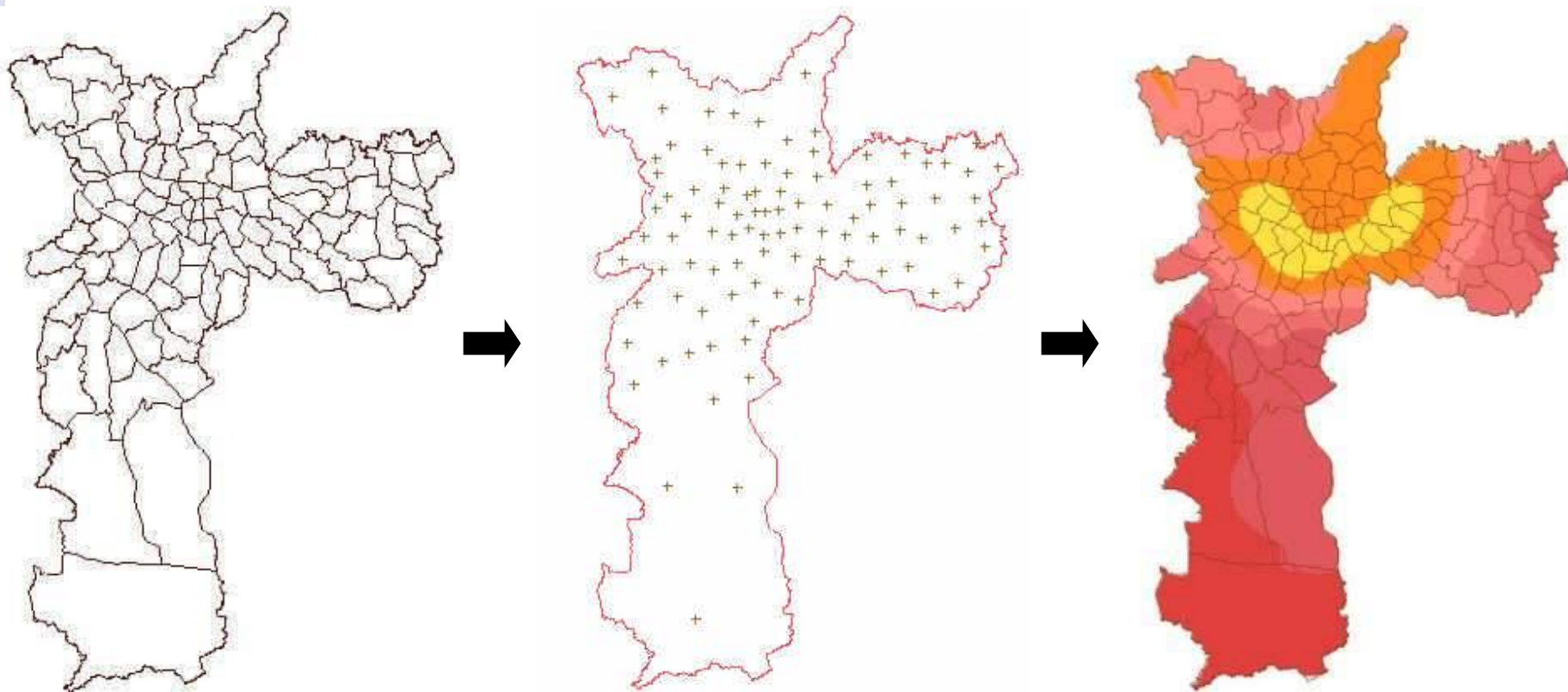


Sampling stations in locations marked by

Location to predict value: shown as



# From Areas to Surfaces

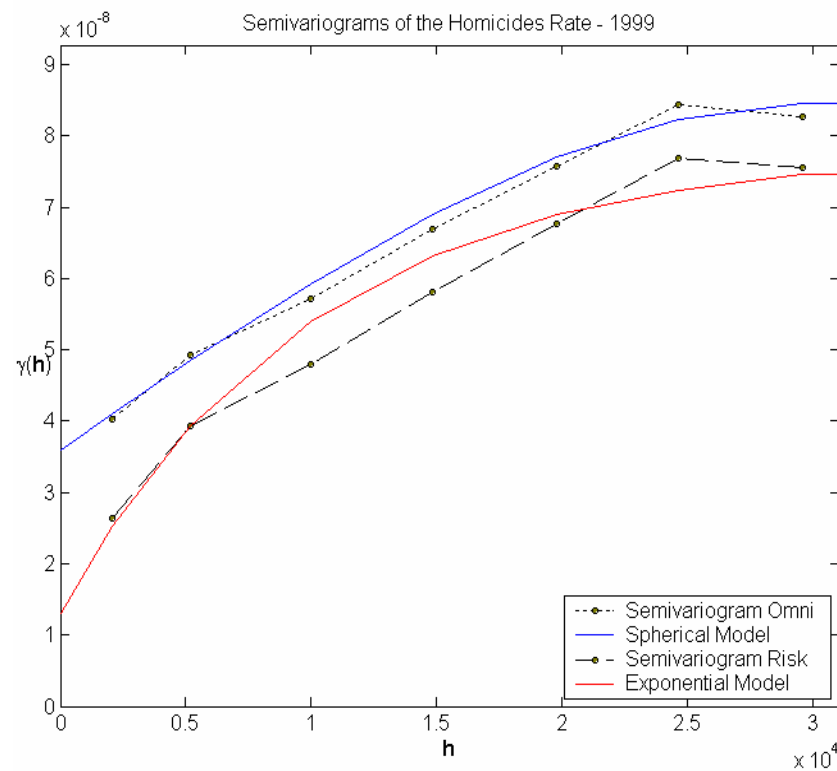
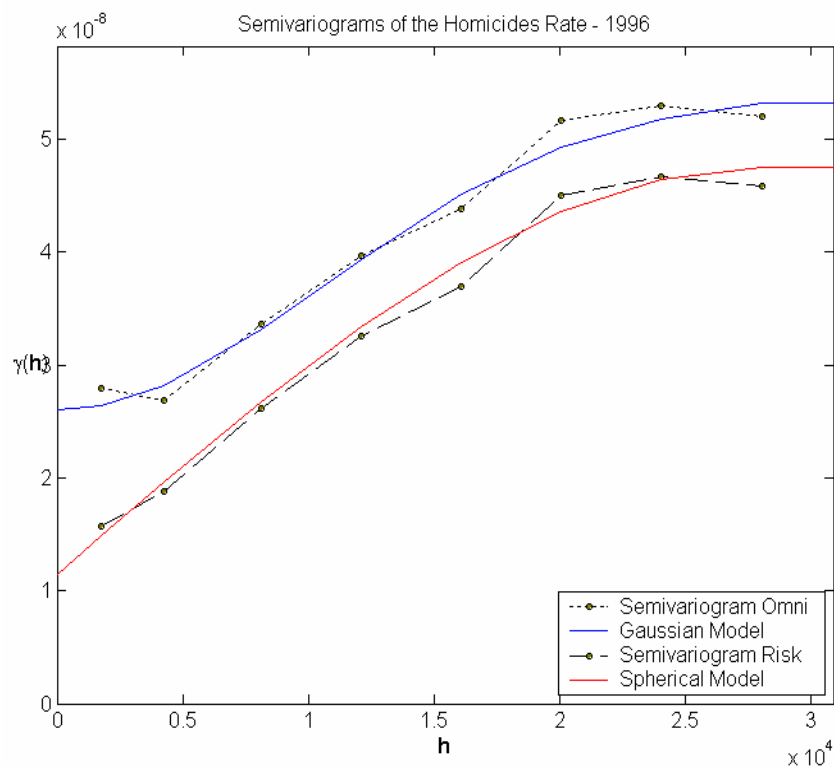


# Como Interpolar as Superfícies?

- Métodos não-paramétricos (“data-driven”)
  - Não fazem hipóteses sobre modelos estatísticos
  - “Deixe os dados falarem”
  
- Métodos paramétricos simples
  - Geoestatística “ordinária”
  - Comportamento simplificado (“estacionário”)
  - Distribuições simples (gaussianas)
  
- Métodos paramétricos gerais
  - Distribuições arbitrárias (e.g., binomial, Poisson)
  - “model-based geostatistics”

# Aplicação / Dados Estatísticos

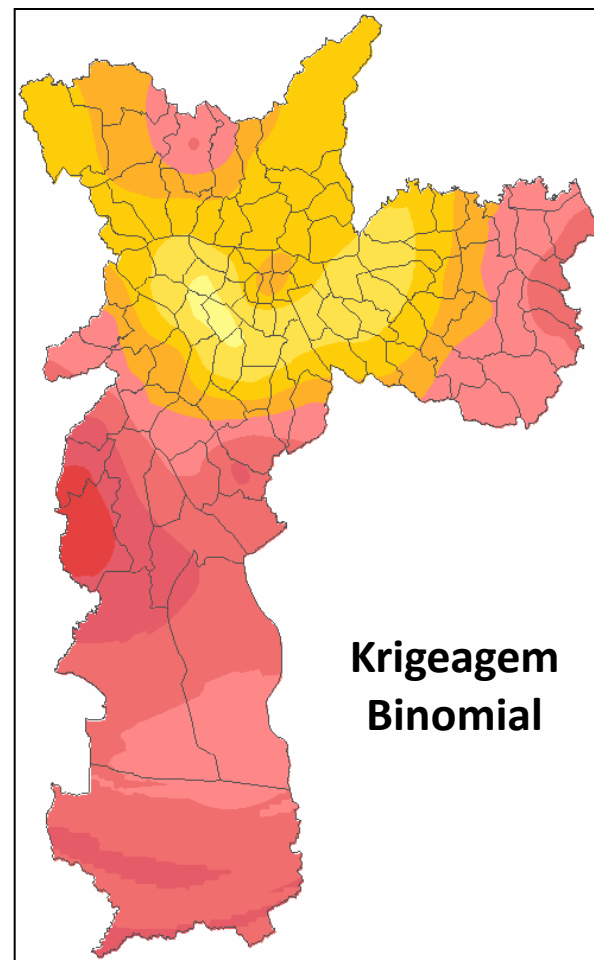
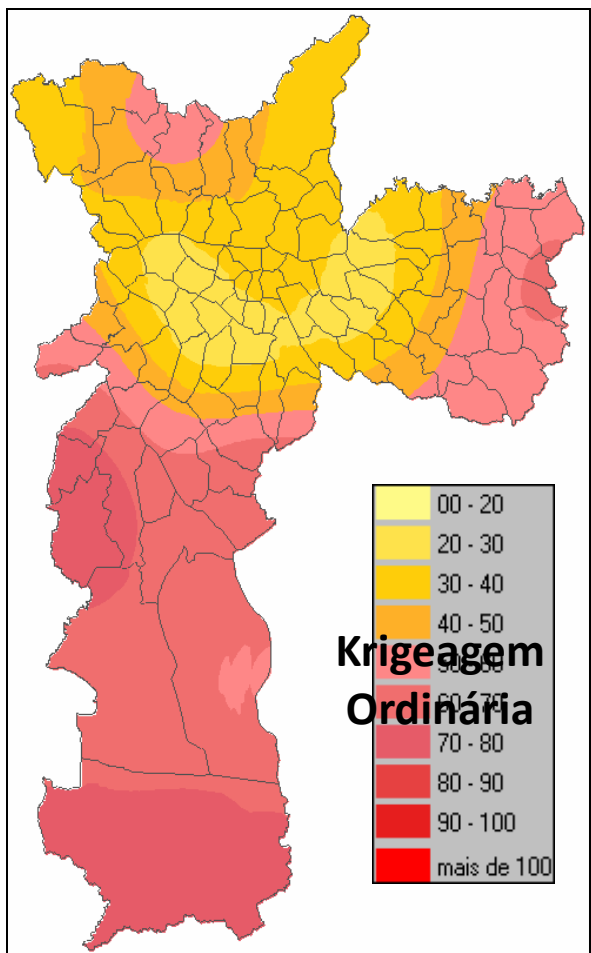
## Variografia de Risco



Correção do variograma para levar em conta a exposição relativa de cada área

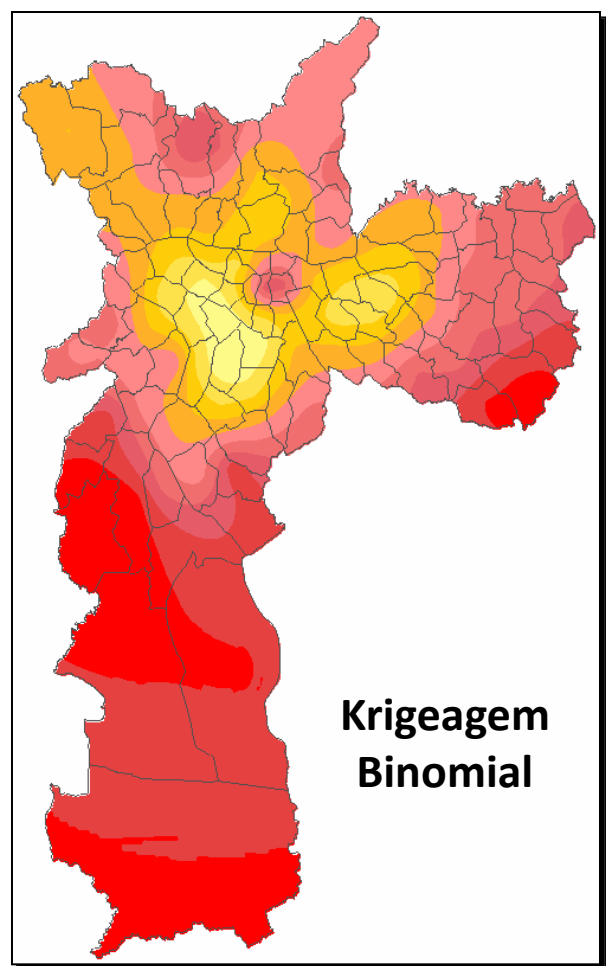
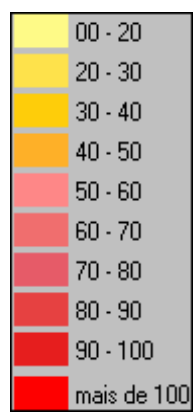
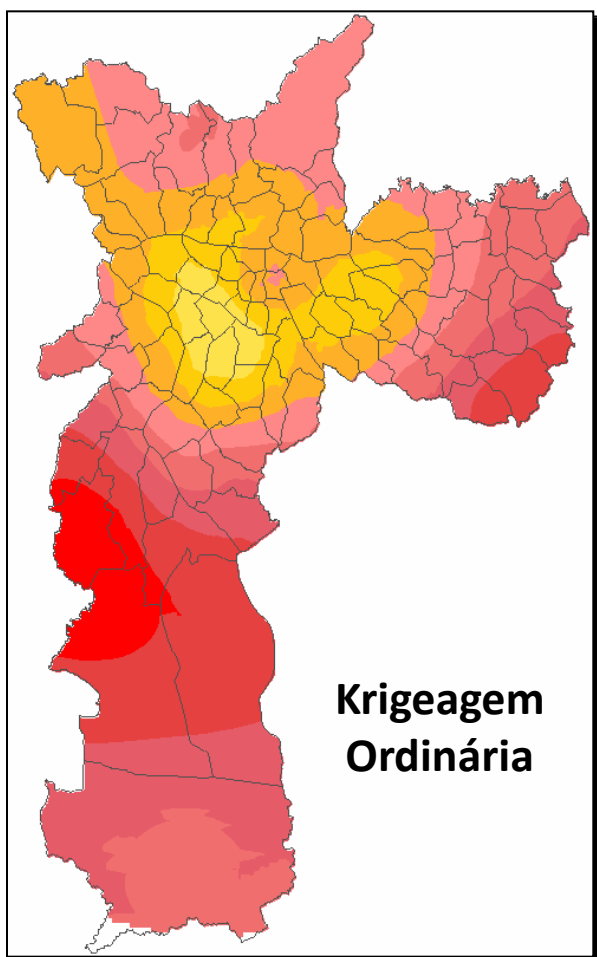


# Aplicação / Comparação entre Superfícies Obtidas por Krigeagem Ordinária e Binomial no ano de 1996





# Aplicação / Comparação entre Superfícies Obtidas por Krigeagem Ordinária e Binomial no ano de 1999



# Regressão Espacialmente Ponderada

Técnica de análise exploratória

Extensão do modelo de regressão tradicional para o modelo com parâmetros estimados localmente.

Betas variam espacialmente

Exemplo: Relação entre homicídios e analfabetismo em São Paulo

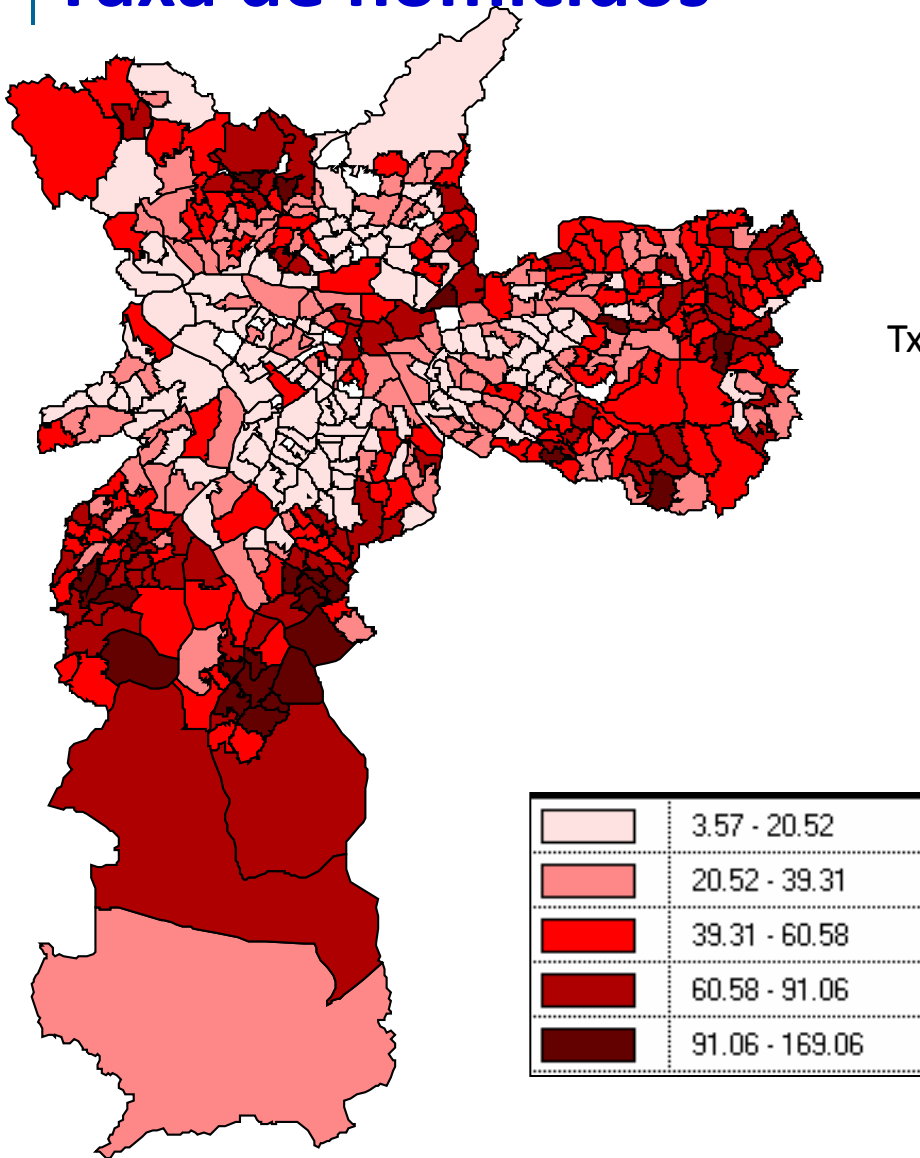
$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \xi_i$$

$$\beta = \begin{bmatrix} \beta_{0(u_i, v_i)} & \beta_{0(u_i, v_i)} & \beta_{0(u_i, v_i)} & \dots & \beta_{0(u_i, v_i)} \\ \beta_{0(u_i, v_i)} & \beta_{0(u_i, v_i)} & \beta_{0(u_i, v_i)} & \dots & \beta_{0(u_i, v_i)} \\ \dots & \dots & \dots & \dots & \dots \\ \beta_{0(u_i, v_i)} & \beta_{0(u_i, v_i)} & \beta_{0(u_i, v_i)} & \dots & \beta_{0(u_i, v_i)} \end{bmatrix}$$

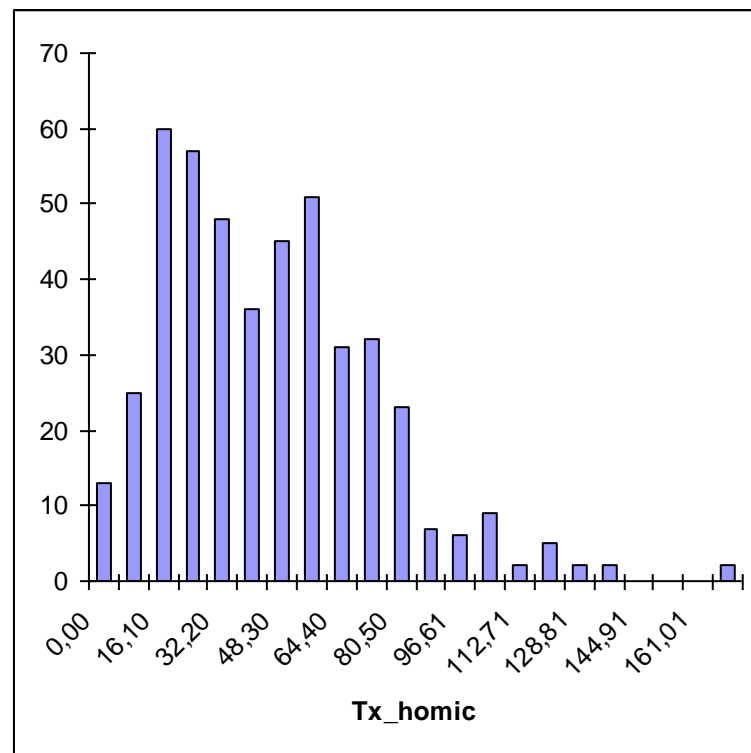
$$\beta(i) = (X^T W(i) X)^{-1} X^T W(i) Y$$

$$W(i) = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & w_{in} \end{bmatrix}$$

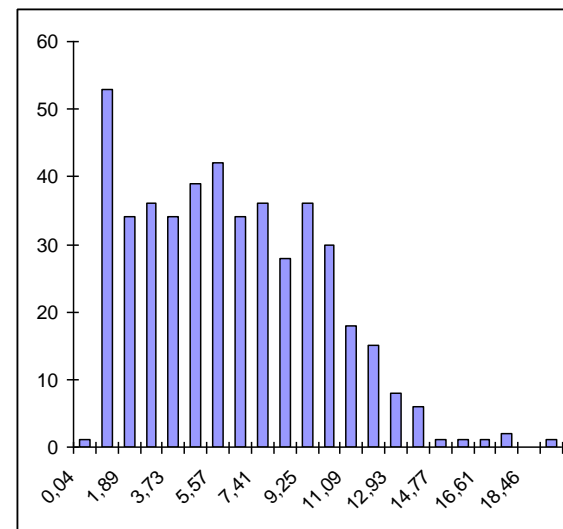
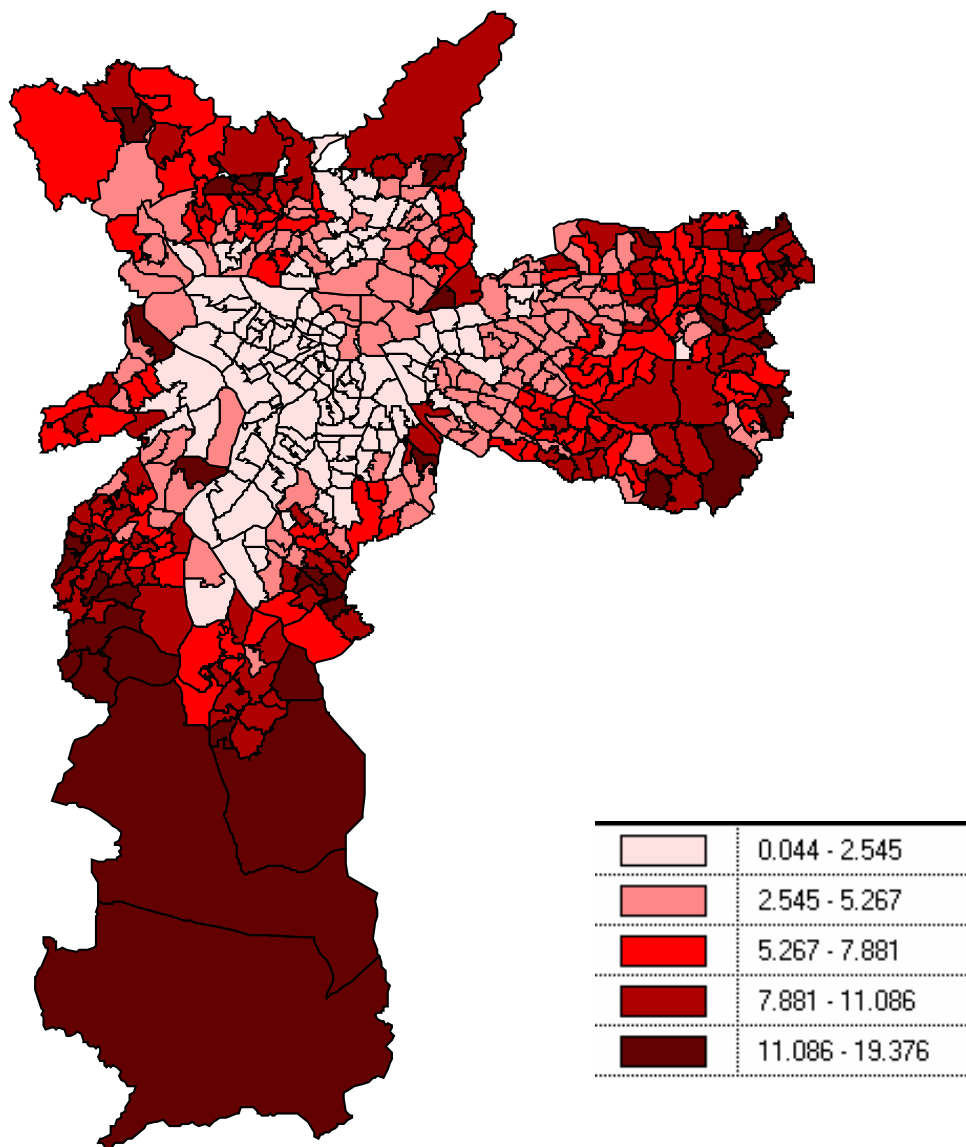
# Taxa de homicídios



$Tx\_homic = \frac{\text{count homicide events (2001)} * 100.000}{\text{population (census, 2000)}}$



# Porcentagem de chefes de família analfabetos



# Regressão espacialmente ponderada

## Modelo geral

$$TX\_HOM = \beta_0 + \beta_1(\% X\_Analf) + \varepsilon$$

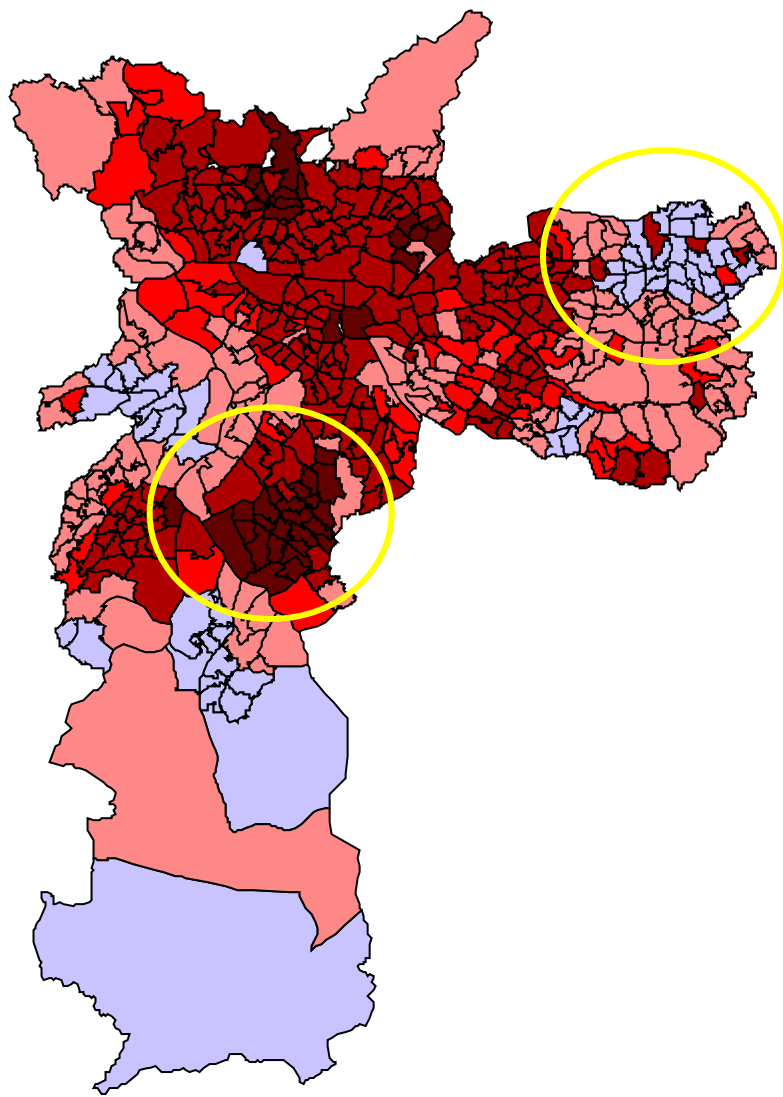
$$R^2 = 0.356$$

## Modelo espacial

$$TX\_HOM = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)(\% X\_Analf) + \varepsilon$$

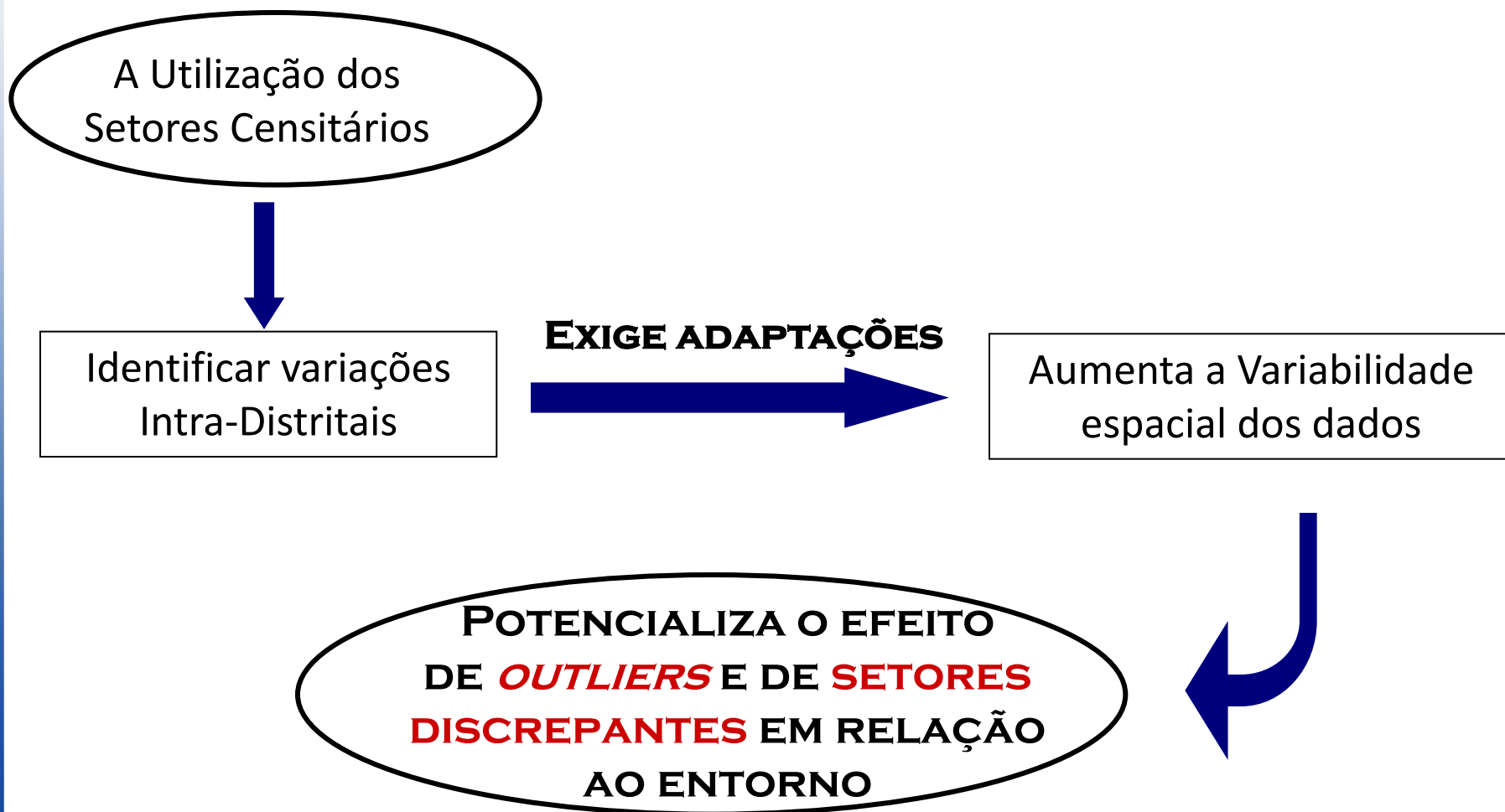
$$R^2 = 0.699$$

## Variação Espacial dos Fatores (Betas)



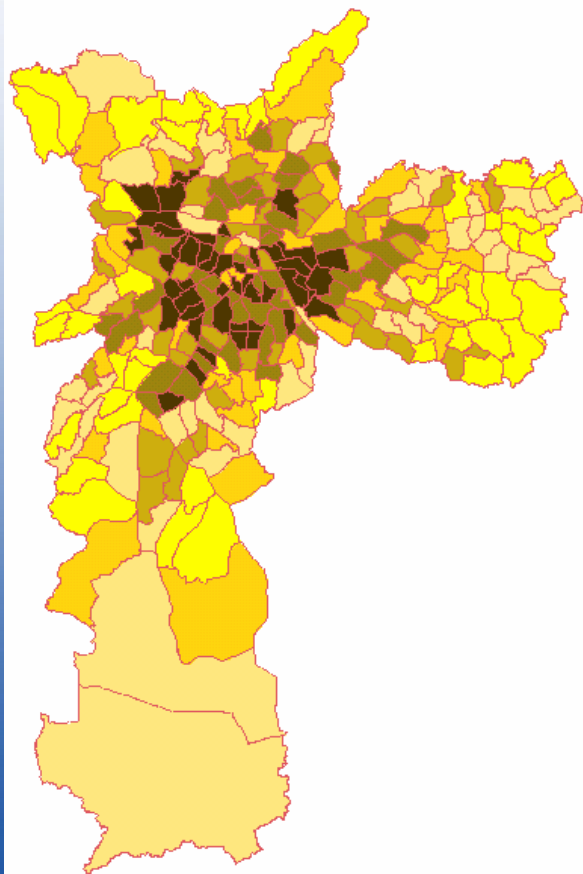
- Regressão Espacial: indica diferenças significativas na relação entre taxas de homicídio e chefes de família analfabetos em São Paulo

# A Questão da Escala

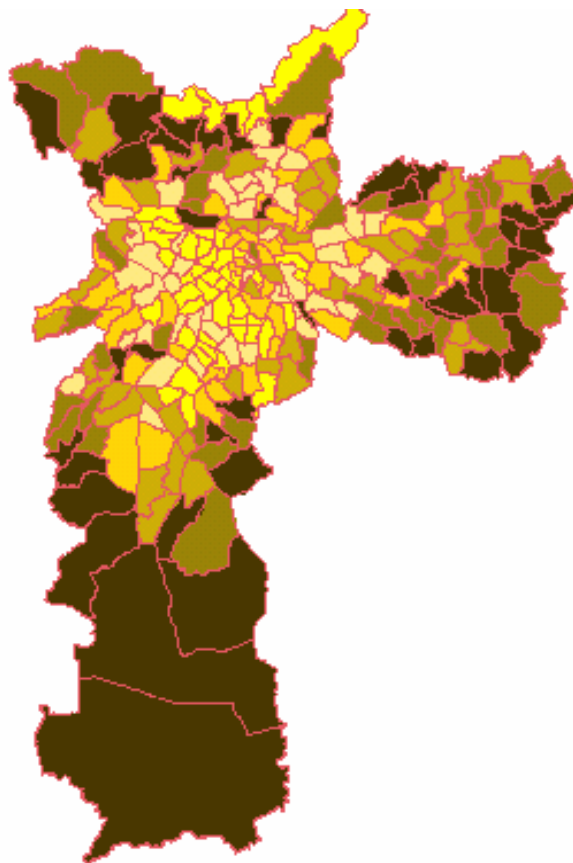


# EFEITOS DE ESCALA

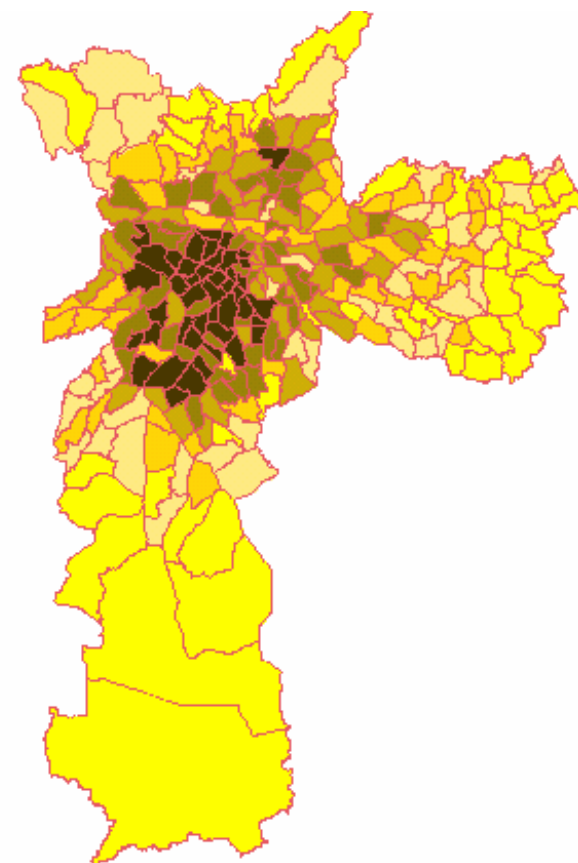
## 270 ZONAS OD97 DO MSP



**População  
>60 anos**



**População não  
alfabetizada**

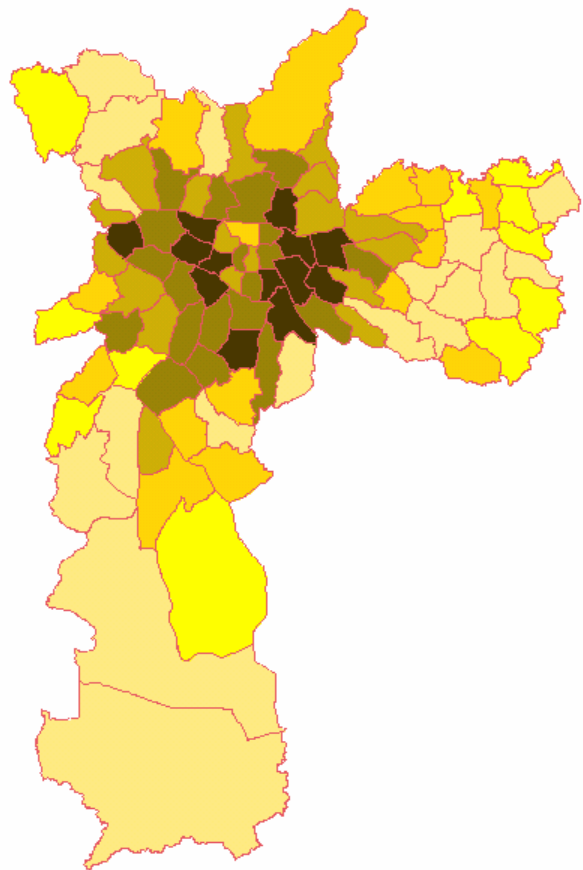


**Renda  
*per capita***

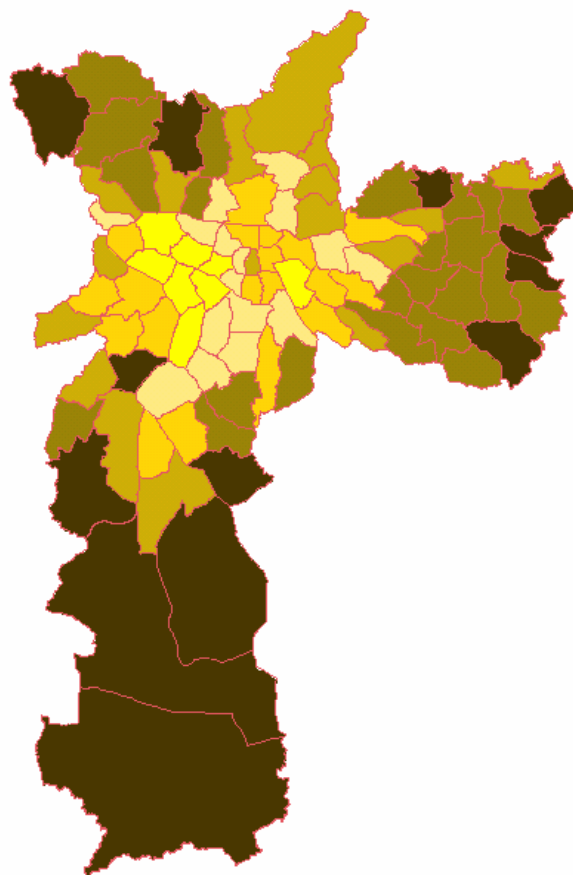


# EFEITOS DE ESCALA

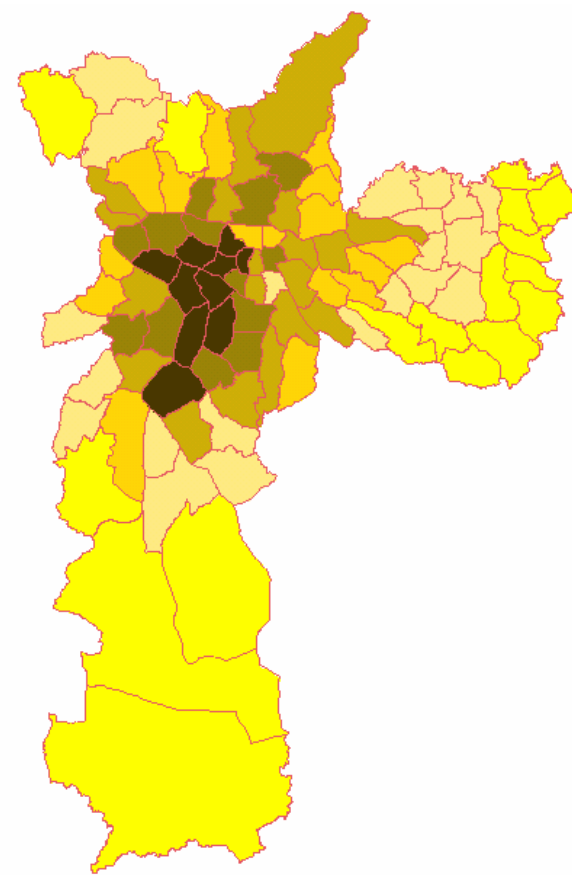
## 96 DISTRITOS DO MSP



**População  
>60 anos**



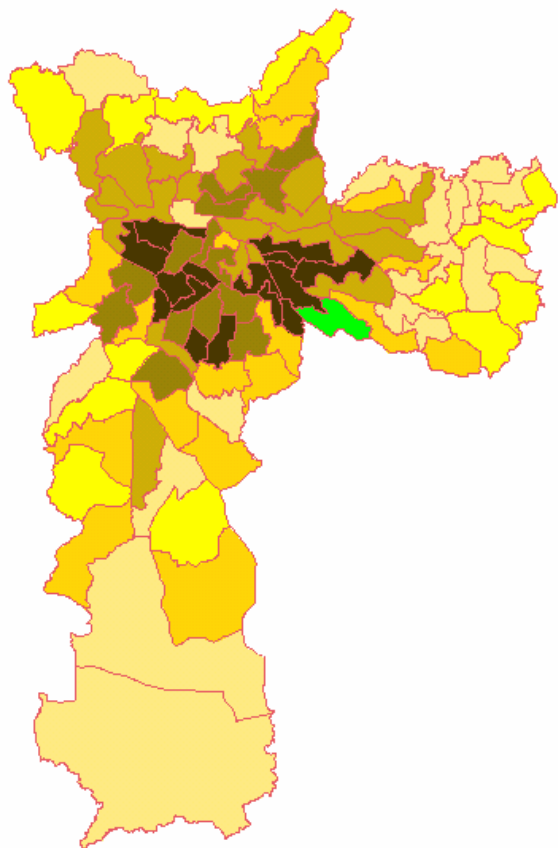
**População não  
alfabetizada**



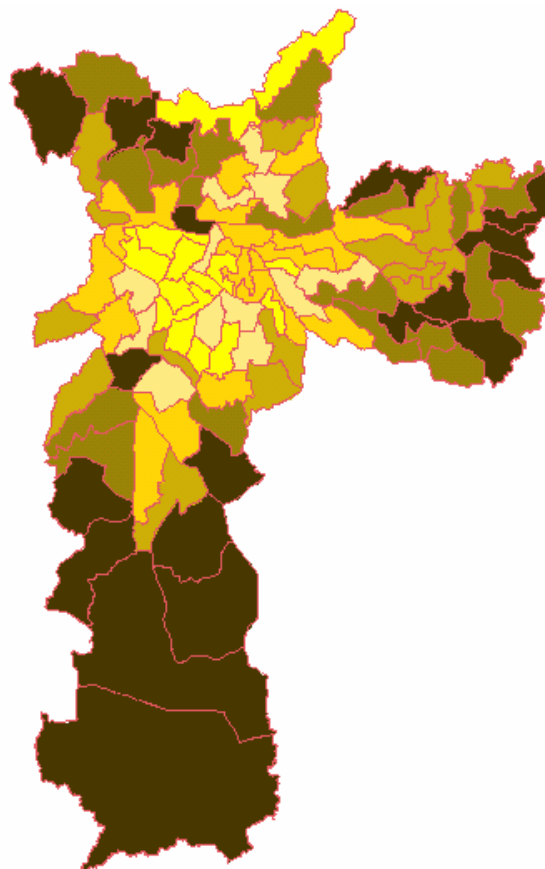
**Renda  
*per capita***

# EFEITOS DE ESCALA

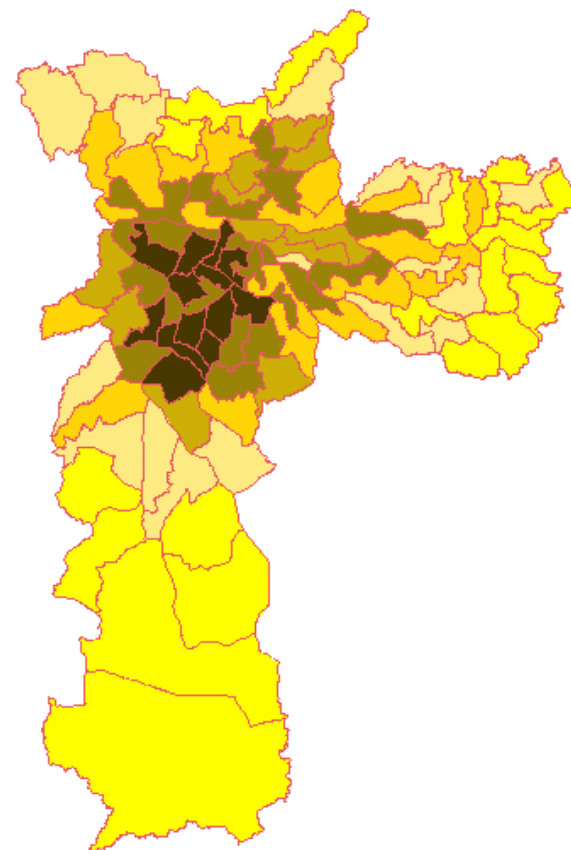
## 96 ZONAS RENDA HOMOGENEAS DO MSP



**População  
>60 anos**



**População não  
alfabetizada**



**Renda  
*per capita***

# EFEITOS DE ESCALA

**Comparação entre as matrizes de correlação para as variáveis selecionadas**

VARIÁVEIS

**A) Porcentagem da população com 60 anos ou mais**

**B) Porcentagem da população não alfabetizada**

**C) Renda individual per capita em reais**

\*

|                   |          |          |          |          |
|-------------------|----------|----------|----------|----------|
| <b>ZONAS OD97</b> |          | <b>A</b> | <b>B</b> | <b>C</b> |
|                   | <b>A</b> | 1        | -0,53    | 0,53     |
|                   | <b>B</b> | -0,53    | 1        | -0,59    |
|                   | <b>C</b> | 0,53     | -0,59    | 1        |

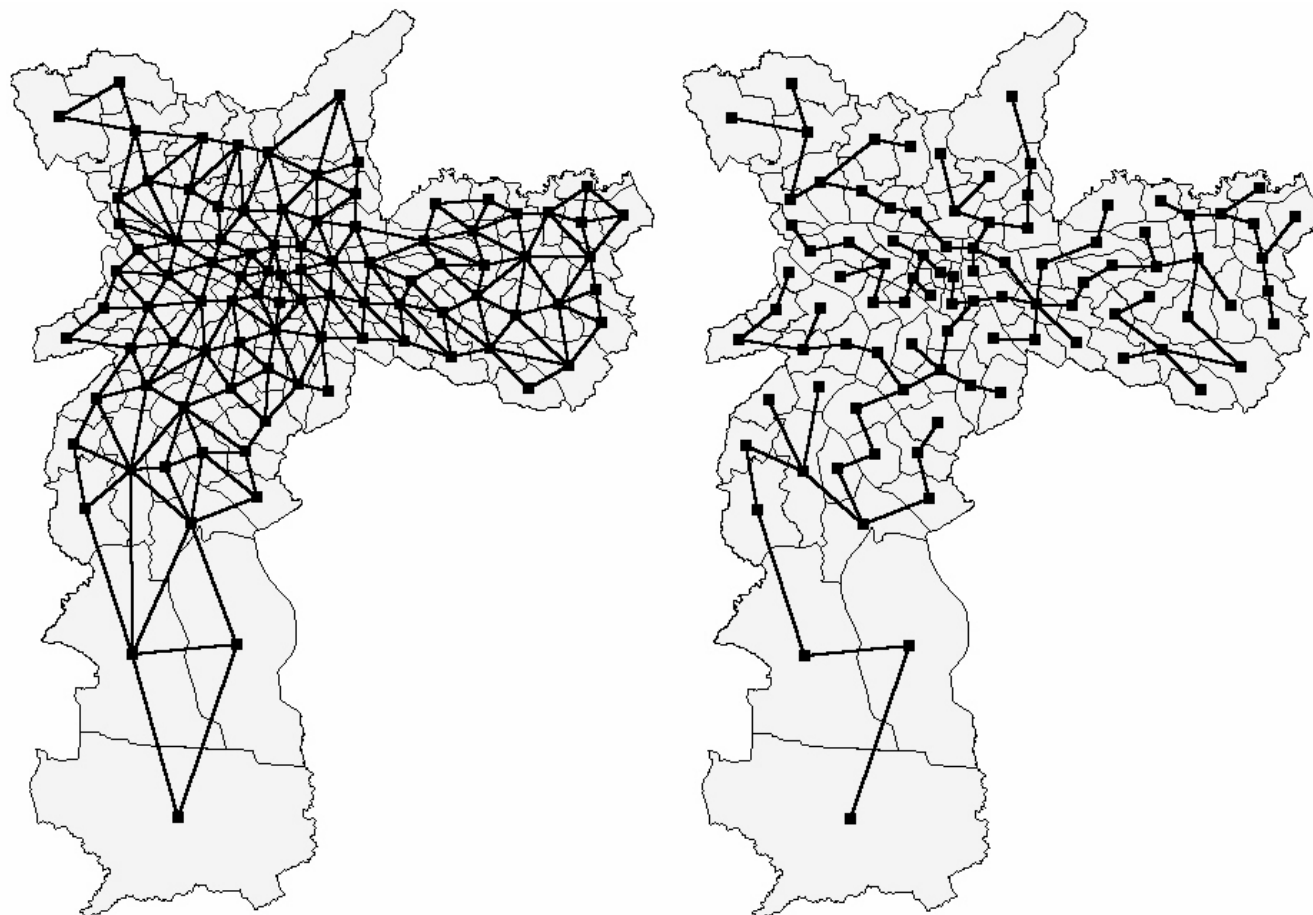
|                  |          |          |          |          |
|------------------|----------|----------|----------|----------|
| <b>DISTRITOS</b> |          | <b>A</b> | <b>B</b> | <b>C</b> |
|                  | <b>A</b> | 1        | -0,81    | 0,65     |
|                  | <b>B</b> | -0,81    | 1        | -0,78    |
|                  | <b>C</b> | 0,65     | -0,78    | 1        |

|                    |          |          |          |          |
|--------------------|----------|----------|----------|----------|
| <b>ZONAS RENDA</b> |          | <b>A</b> | <b>B</b> | <b>C</b> |
|                    | <b>A</b> | 1        | -0,52    | 0,57     |
|                    | <b>B</b> | -0,52    | 1        | -0,72    |
|                    | <b>C</b> | 0,57     | -0,72    | 1        |

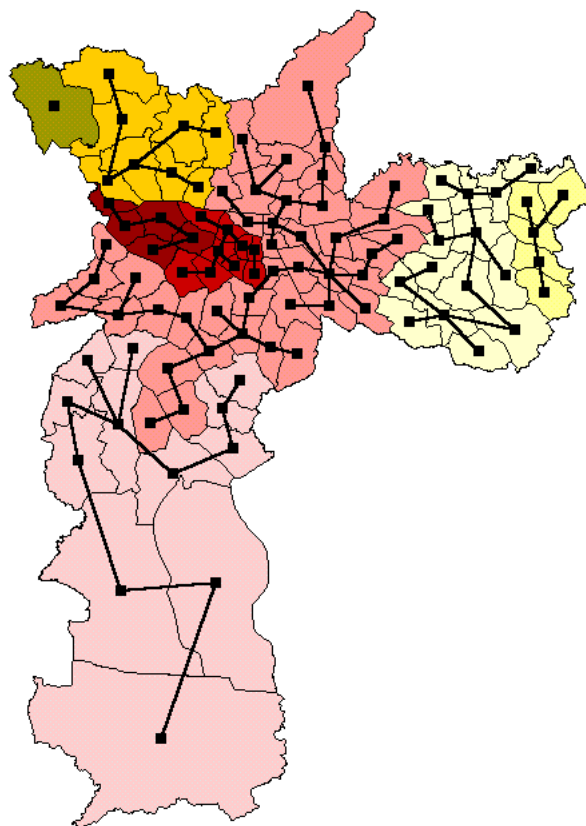
## Diferenciação do Espaço: Regionalização

- Reagregar N pequenas áreas (a mais fina resolução disponível) em M regiões maiores de maneira a reduzir os efeitos de escala.
  
- Uma solução possível
  - Algoritmos de agregação com restrições espaciais
  - SKATER (Renato Assunção e equipe LESTE/UFMG)
  - Otimização do SKATER em ambiente de SIG (Marcos Neves e equipe INPE)

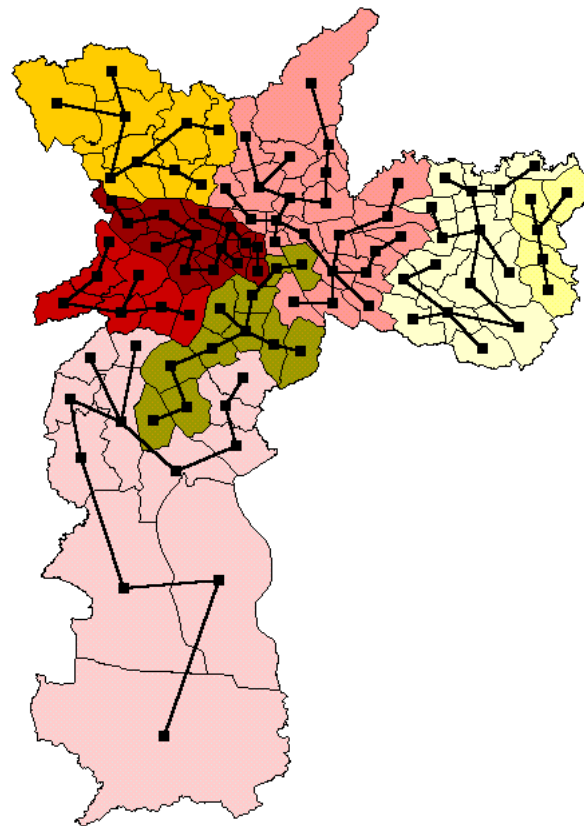
# Regionalização: Mapas como Grafos



# Regionalização: Mapas como Grafos



Agregamento em áreas



Agregamento em áreas com restrição de população mínima

# Relógios, Nuvens ou Formigas?

## ■ Relógios

- Paradigmas: lei de Newton (leis físicas que governam o mundo)

## ■ Nuvens

- Modelos estocásticos
- Suporte: Teoria de sistemas caóticos

## ■ Formigas

- Modelos emergentes
- Suporte: teoria de sistemas complexos
- Exemplos: automata celulares

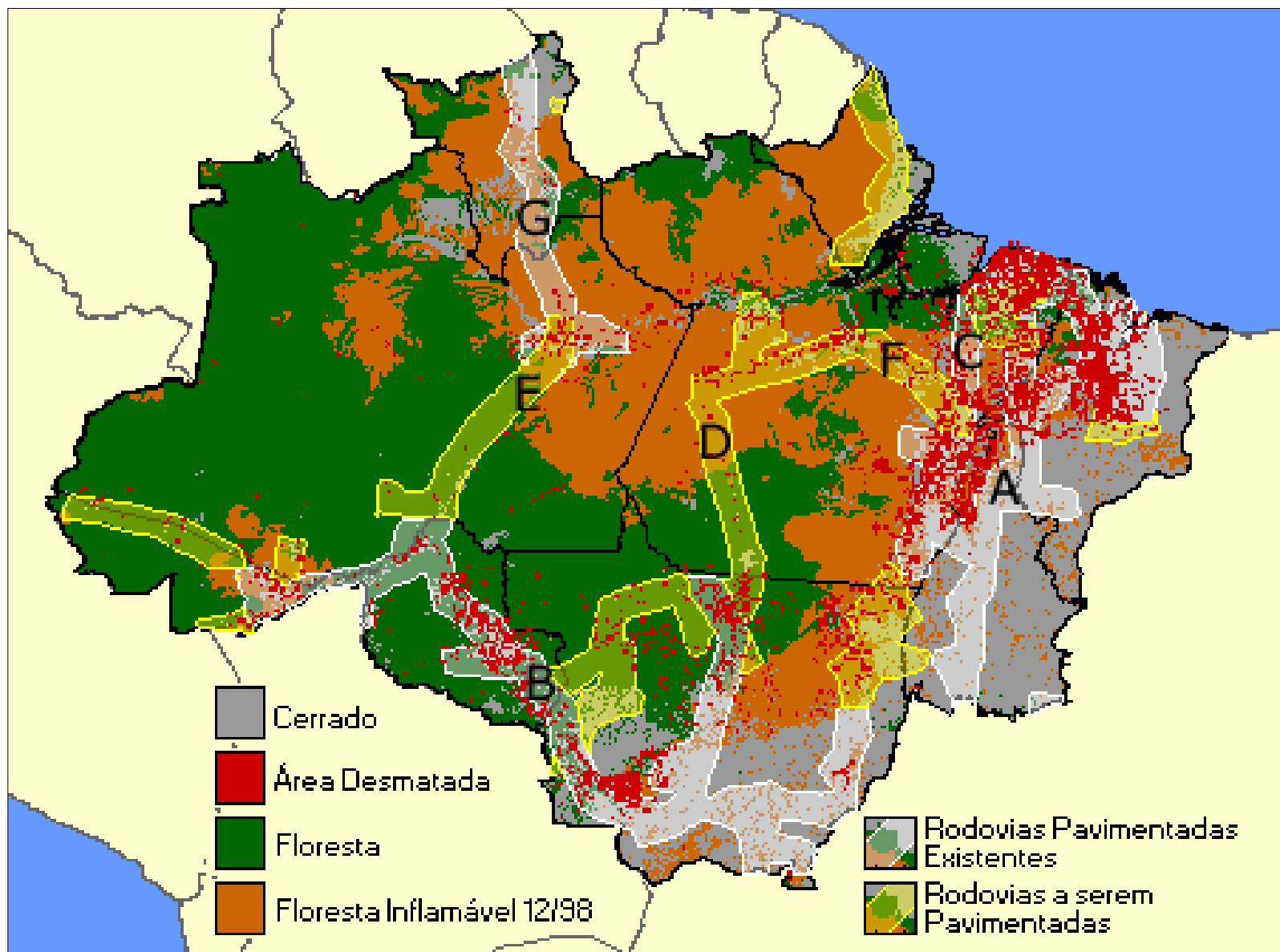
# Podemos conhecer o passado....

Estimativa do Desmatamento da Amazônia (INPE)

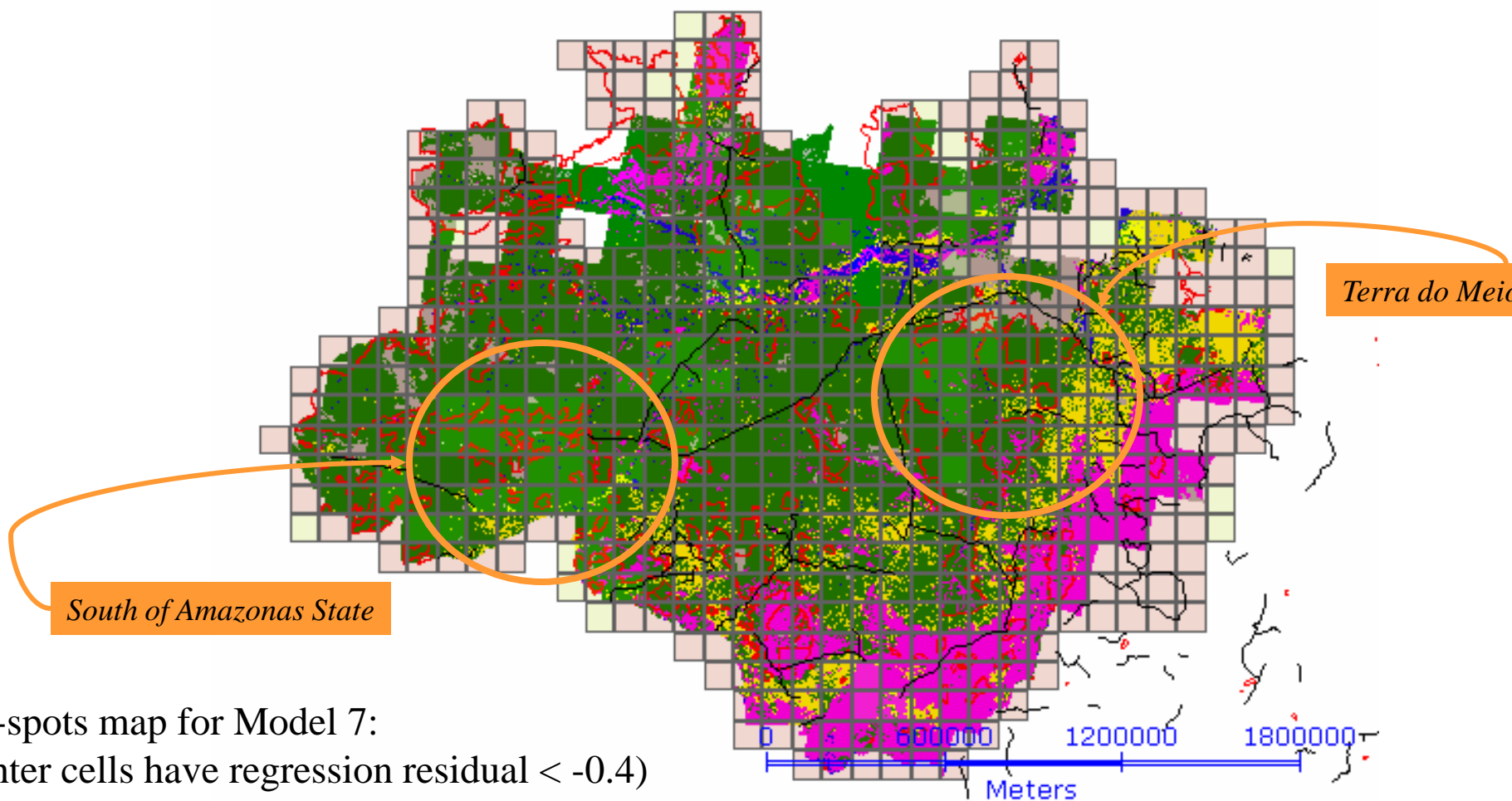




# O que nos reserva o futuro?



# Modelos de Uso da Terra: Áreas Previstas de Expansão do Desmatamento



Hot-spots map for Model 7:  
(lighter cells have regression residual  $< -0.4$ )

# Categorias de fatores candidatos

(~ 90 variáveis)

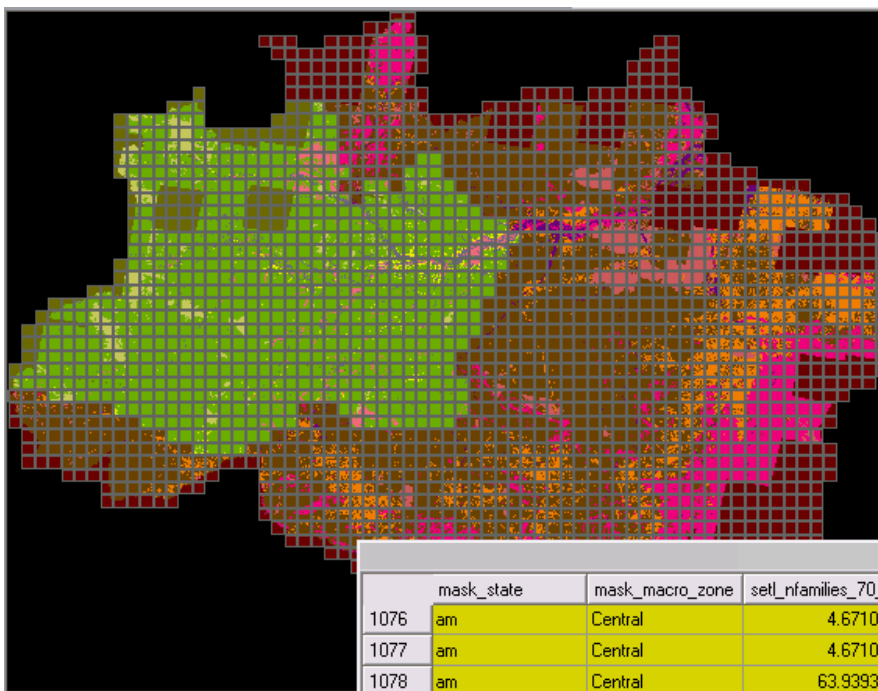
|  |  |
|--|--|
| <b>Demográficos</b>                        | Densidade populacional e migração.   |
| <b>Estrutura Agrária</b>                   | Indicadores de distribuição de terras entre pequenos, médios e grandes estabelecimentos rurais (em número e área). |
| <b>Acessibilidade e conexão a mercados</b> | Medidas de conexão a mercados nacionais e portos via rede viária.<br>Distância a estradas, rios e centros urbanos. |
| <b>Políticos</b>                           | Assentamentos, áreas de proteção e crédito.  |
| <b>Atividades econômicas</b>               | Distância a pólos madeireiros e depósitos minerais.  |
| <b>Ambientais</b>                          | Qualidade do solo (fertilidade, textura), relevo e clima (precipitação, umidade, temperatura).                     |

# Fatores Correlacionados ao Desmatamento

- Sete fatores estão relacionados à variação de 83% das taxas de desmatamento na Amazônia nos últimos anos:
  - (a) Estrutura Agrária (2 fatores): percental de área ocupada por grandes fazendas e número de pequenas propriedades.
  - (b) Ocupação Populacional (1 fator): densidade de população.
  - (c) Condições do Meio Físico (2 fatores): Precipitação média e percentual de solos férteis.
  - (d) Infraestrutura (1 fator): distância a estradas.
  - (e) Presença do Estado (1 fator): percentagem de áreas indígenas

# Ambientes Computacionais para Modelagem

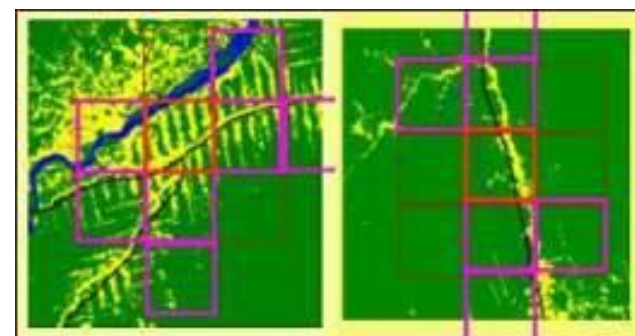
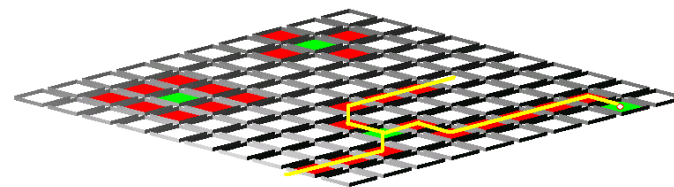
## Espaços celulares



|             | mask_state | mask_macro_zone | setl_nfamilies_70_99 | setl_area_70_99 | agr |
|-------------|------------|-----------------|----------------------|-----------------|-----|
| 1076        | am         | Central         | 4.671096             | 146.23648       |     |
| 1077        | am         | Central         | 4.671096             | 146.23648       |     |
| 1078        | am         | Central         | 63.939396            | 23501.954167    |     |
| 1079        | am         | Central         | 81.582006            | 29565.766222    |     |
| 1080        | pa         | Central         | 12.805476            | 1287.076729     |     |
| 1081        | pa         | Central         | 13.10852             | 1329.578364     |     |
| 1082        | pa         | Central         | 13.10852             | 1329.578364     |     |
| <b>1083</b> | pa         | Central         | 11.466334            | 1163.013824     |     |

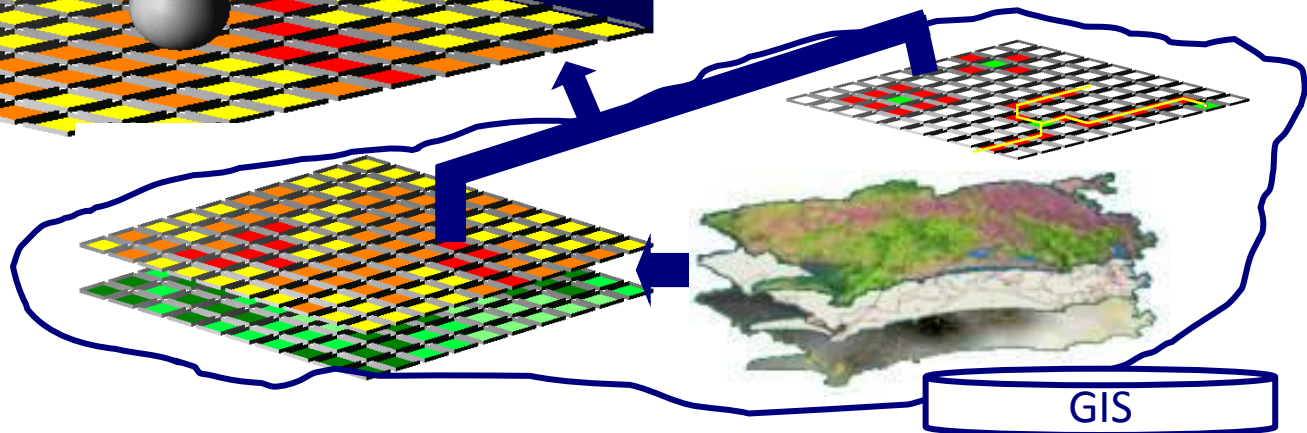
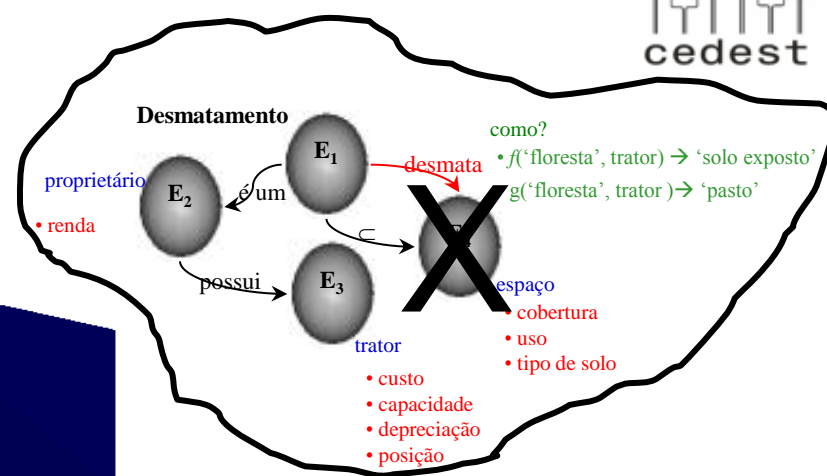
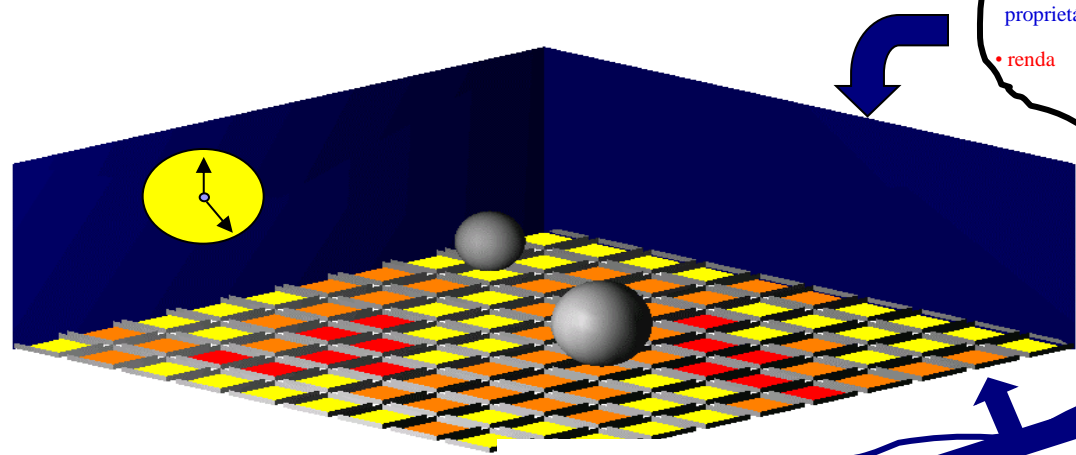
### Componentes

- conjunto de células georeferenciadas
- identificador único
- vários atributos por células
- matriz genérica de proximidade - GPM



superfície discreta de células retangulares multivaloradas possivelmente não contíguas

# O modelo ambiental

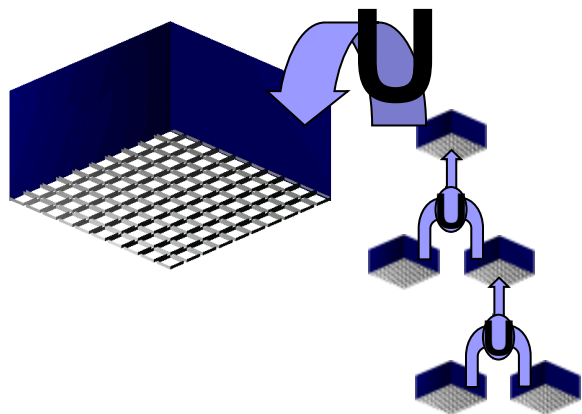


■ Um ambiente possui 3 submodelos:

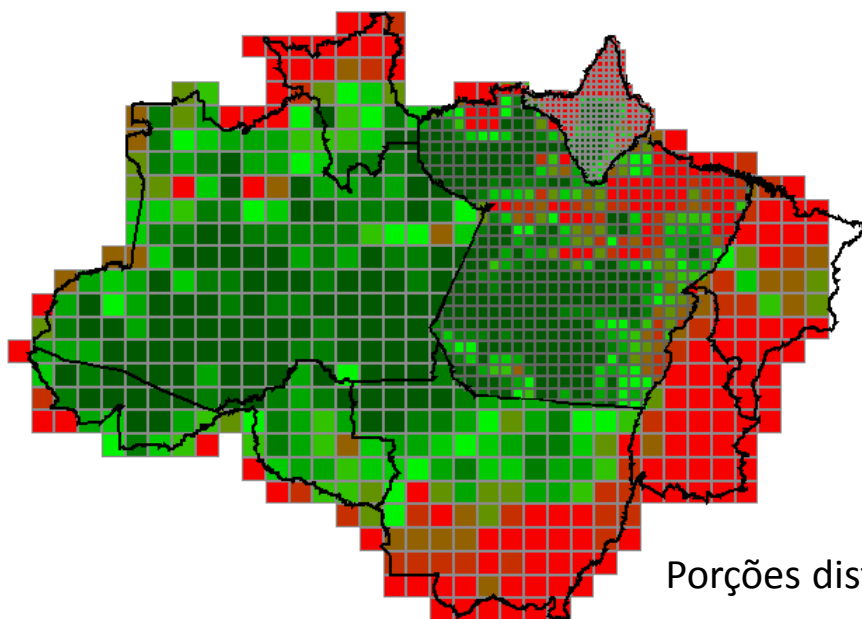
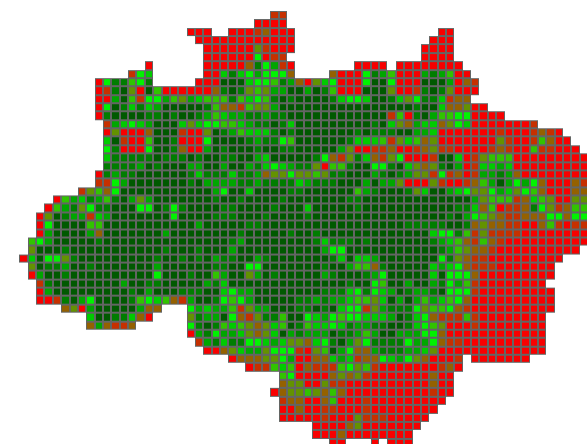
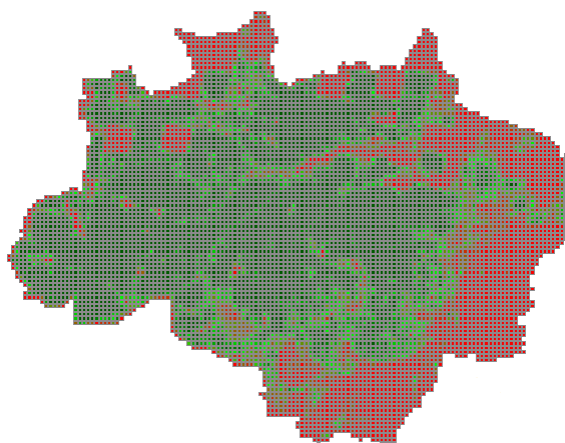
- **Modelo Espacial:** espaços celulares + regiões + GPM
- **Modelo Comportamental:** teoria de sistemas + autômatos celulares híbridos + agentes situados
- **Modelo Temporal:** simulador de eventos discretos definidos de forma recorrente

■ A estrutura espacial e temporal é compartilhada por vários agentes.

# A estrutura do espaço é heterogênea



Ambientes definidos de forma recorrente



É possível construir modelos **multiescalas**

Porções distintas do espaço podem ter escalas diferentes



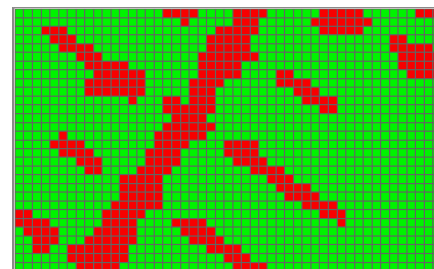
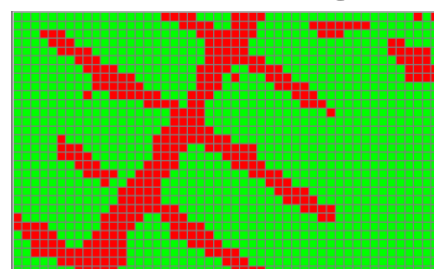
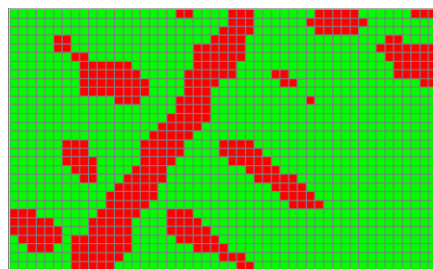
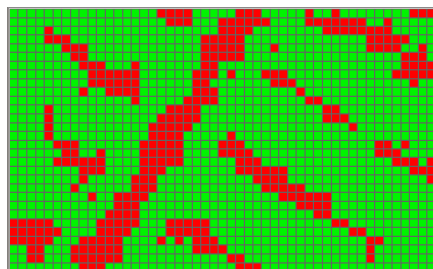
# Ambiente Computacional de Modelagem: TerraLib

Realidade

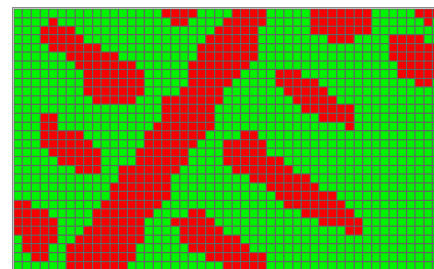
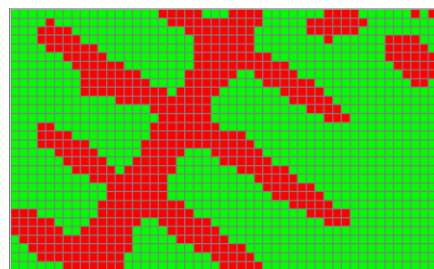
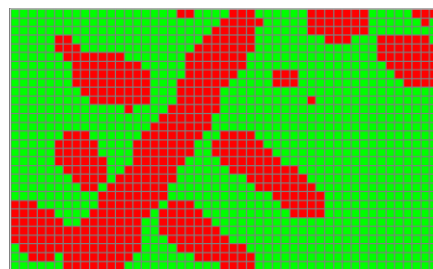
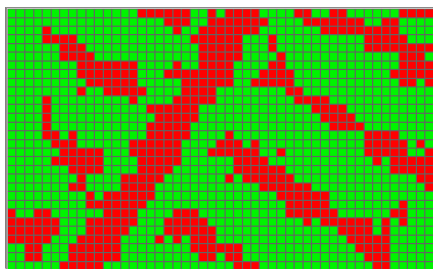
Moore

GPM

GPM+Lote



1988



1991

Geoinfo (Aguiar, 2003),  
Submetido GIScience (Câmara et al, 2004)



# Limites para Modelar a Realidade

fonte: John Barrow



# Sistema de Avaliação e Controle da Criminalidade em Ambiente Urbano (TerraCrime)

Baseado no Sistema



TerraView

**SENASP**



### Tela de visualização do Mapa de Kernel das distribuição de ocorrências em Porto Alegre

TerraCrime 1.0 - [Tela de Visualização]

Arquivo Exibir Plano Vista Tema Análise Operação Janela

Desenhar

Bancos de Dados

- PortoAlegre.mdb
  - Arruamento
  - Bairros
  - Bairros\_1
  - Bairros\_2
  - Bairros\_3
  - Bairros\_4

Vistas

- UTM\_SAD9\_052
  - Ocorrencias\_Ago20
  - Distritos
  - min
  - Arruamento
  - TM7\_2002Abr02
  - botanico
  - bairros
  - TODAS
  - bairrosMinimo

| ID | SPRPERIMET | SPRROTULO | SPRNome       | LENGTH | CHAVE   | CODLOGR | CATEG | LOGR                            | PREP | TIPO | SMF_P_ID | SMF_P_F |
|----|------------|-----------|---------------|--------|---------|---------|-------|---------------------------------|------|------|----------|---------|
| 1  | 0          |           |               | 0      | 0       | 0       |       |                                 |      | 0    | 0        |         |
| 2  | 89,3388    | 1         | 1 89,339      | 20     | 8280034 | AV      |       | ASSIS BRASIL                    | 0    |      | 10766    |         |
| 3  | 504,7866   | 10        | 10 504,787    | 45     | 8885014 | AC      |       | ASSIS BRASIL-AUTO ESTRADA OESTE | 0    |      | 0        |         |
| 4  | 231,9991   | 100       | 100 231,999   | 14204  | 8484016 | R       |       | DIRETRIZ SETECENTOS SETE        | 4    |      | 2        |         |
| 5  | 34,0052    | 1000      | 1000 34,005   | 831    | 8682064 | R       |       | JOSE MENDES                     | 0    |      | 348      |         |
| 6  | 191,7412   | 10000     | 10000 191,741 | 9758   | 8177198 | R       |       | ENG ADOLFO STERN                | 0    |      | 2        |         |
| 7  | 94,044     | 10001     | 10001 94,044  | 9759   | 7977168 | R       |       | CABRAL                          | 0    |      | 1390     |         |
| 8  | 197,0091   | 10002     | 10002 197,009 | 9760   | 8077026 | R       |       | GIORDANO BRUNO                  | 0    |      | 26       |         |

Desenhar sobre a tela de visualização

## Tecnologia em Saúde Pública: Epidemiologia Espacial

As Tecnologias da Informação Espacial no apoio ao Controle de Endemias: Dos “Lugares das Doenças” às “Doenças do Lugar”

<http://saudavel.dpi.inpe.br>



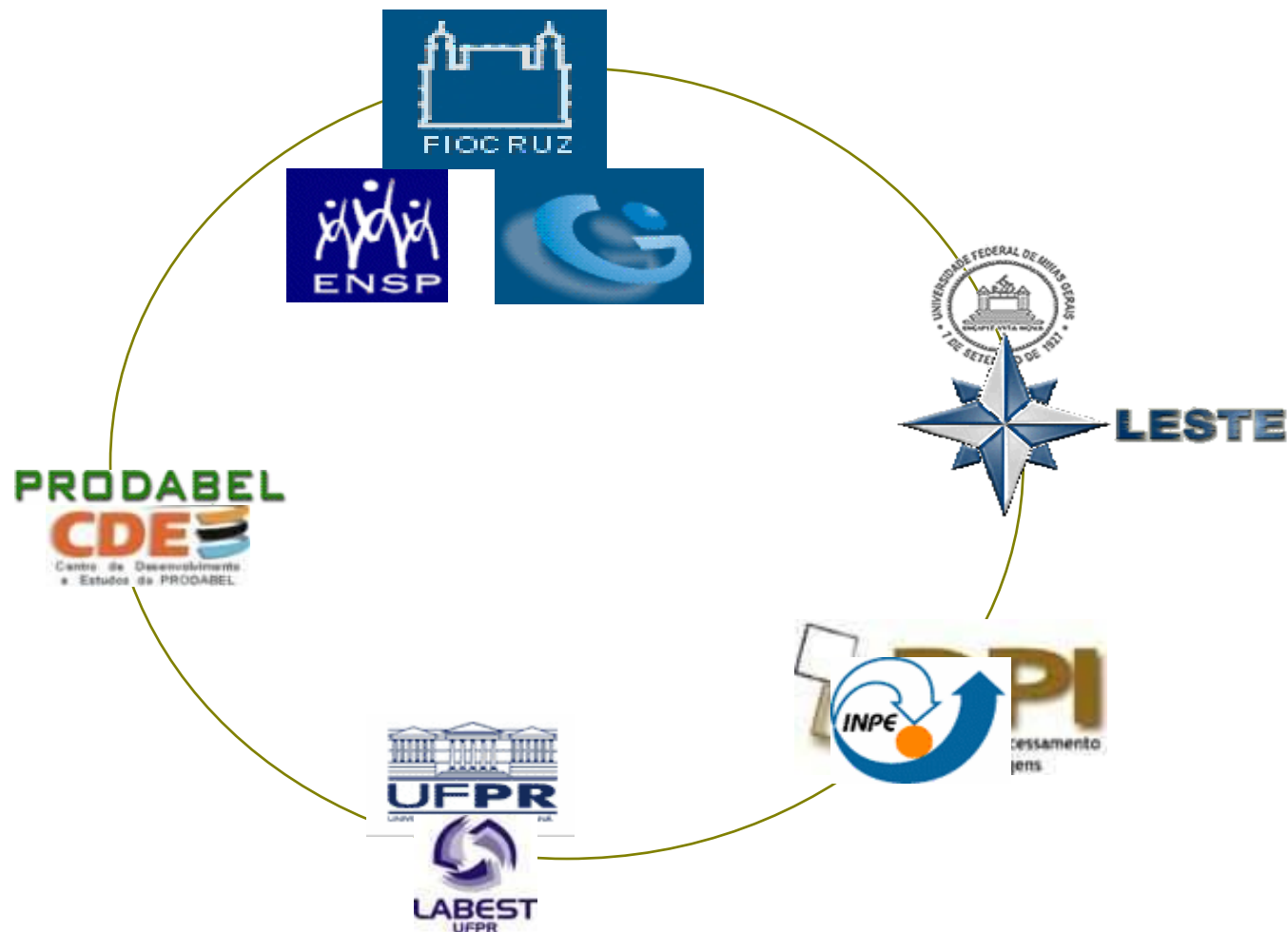
**CTINFO**  
Fundo Setorial

**FINEP**

**CNPq**  
Conselho Nacional de Desenvolvimento Científico e Tecnológico

Processo CNPq **55.2044/2002-4**  
Projeto aprovado pela Chamada Conjunta MCT/SEPIN -  
FINEP - CNPq 01/2002,  
Programa de Apoio à Pesquisa e Desenvolvimento e  
Inovação em Tecnologia de Informação - PDI-TI.  
**4 Anos – Ano 1**

# Rede SAUDAVEL: Opera uma Rede Inter(Intra)-Institucional





# Integração Produtos SAUDAVEL com R GECEA/UFPR

Grupo de Estatística Computacional e Espacial Aplicadas/ UFPR  
(Paulo Justiniano Ribeiro Jr)



aRT : (**API R-TerraLib**) é um aplicativo utilizado para fazer a integração entre o software **R** e a biblioteca **TerraLib**, combinando informações estatísticas e geoespaciais.

myR: é uma biblioteca escrita em C++ para o acesso às funcionalidades do **R**, sem precisar entender o seu funcionamento interno. Com o myR é possível chamar as funções e algoritmos da **TerraLib** de dentro de um aplicativo C++.

Integração Terralib - R

# TerraStat: Biblioteca de Procedimentos Estatísticos Espaciais Avançados em C++

Parceria DPI/INPE e LESTE/UFMG



TerraStat : (Algoritmos TerraLib) é um conjunto de algoritmos para clusters espaço-temporais através de métodos estatísticos e outras soluções, implementadas em C++ e a incorporados como *algoritmos* em TerraLib

# Os desafios para a Ciência Estatística na análise das Questões Sociais Brasileiras

## ■ Avanços Metodológicos

### □ Modelos espaciais multivariados

- (Assunção e Reis, “Multiple cancer sites incidence rates estimation using a multivariate Bayesian model”, *International Journal of Epidemiology*, 2004)

### □ Modelagem de fenômenos contínuos

- (Diggle e Ribeiro, “Model based geostatistics”)

### □ Modelagem espaço-temporal

- Como compatibilizar as nuvens com as formigas?



# Os desafios para a Ciência Estatística na análise das Questões Sociais Brasileiras

## ■ Avanços Tecnológicos

- Integração de técnicas estatísticas em sistemas de informação geográfica
- Uso de software livre
  - TerraCrime, Integração R-TerraLib
  - Palestra de Paulo Ribeiro no SINAPE 2004

## ■ Benefícios Potenciais

- Suporte a ambiente multidisciplinar
- Redução substancial de custo
- Integração maior da Estatística às questões sociais