

ANÁLISE ESPACIAL E GEOPROCESSAMENTO

*Gilberto Câmara
Antônio Miguel Monteiro
Marília Sá Carvalho*

11.1 INTRODUÇÃO

Uma das operações mais utilizadas em sistemas de informação geográfica é a apresentação espacial de variáveis como população, índices de qualidade de vida e mapas de vendas de empresa. Basta dispor de uma planilha em Excel ou dBase e de uma base geográfica (como um mapa de municípios); o SIG "automaticamente" nos apresenta um mapa colorido (que os cartógrafos chamam pomposamente de "cloroplético") com o padrão espacial do fenômeno.

Grande parte dos usuários limita seu uso de GIS a essas operações de visualização, tirando conclusões intuitivas. Mas é possível ir muito além. Quando visualizamos um padrão espacial, é muito útil traduzi-lo em considerações objetivas: o padrão que observamos é aleatório ou apresenta uma agregação definida? Esta distribuição pode ser associada a causas mensuráveis?

Epidemiologistas coletam dados sobre ocorrência de doenças. A distribuição dos casos de uma doença formam um padrão no espaço? Existe associação com alguma fonte de poluição? Evidência de contágio?

A polícia deseja investigar se existe algum padrão espacial na distribuição de roubos. Roubos que ocorrem em determinadas áreas estão correlacionados com características sócio econômicas dessas áreas?

Geologistas desejam estimar a extensão de um depósito mineral em uma região a partir de amostras. Pode-se usar essas amostras para construir um mapa de contaminação ?

Queremos analisar uma região geográfica para fins de zoneamento agrícola. Como escolher as variáveis explicativas (p.ex., o solo, a vegetação e a geomorfologia) e determinar qual a contribuição de cada uma delas para a obtenção de um mapa resultante?

Todos esses problemas podem ser caracterizados genericamente como questões de Análise Espacial de Dados Geográficos. Para resolvê-los, este livro apresenta um conjunto de ferramentas computacionais, que deverão auxiliar os analistas a determinar as evoluções espacial e temporal de um fenômeno geográfico e as interrelações entre diferentes fenômenos.

Tome-se um exemplo concreto para explicitar os conceitos acima. Em 1854, Londres estava sofrendo uma grave epidemia de cólera, doença sobre a qual na época não se conhecia a forma de contaminação. Numa situação onde já haviam ocorrido mais de 500 mortes, o doutor John Snow teve um “estalo”: colocar no mapa da cidade a localização dos doentes de cólera e dos poços de água, que eram a fonte principal de água dos habitantes da cidade (vide Figura 11.1).



Figura 11.1 - Mapa de Londres com casos de cólera (pontos) e poços de água (cruzes) (adaptado de Tufte, 1983).

Com a espacialização dos dados, o doutor Snow percebeu que a maioria dos casos estava concentrada em torno do poço da “Broad Street” e ordenou a sua lacração, o que contribuiu em muito para debelar a epidemia. Este caso forneceu evidência empírica para a hipótese (comprovada posteriormente) de que o cólera é transmitido por ingestão de água contaminada. Este é uma situação típica onde a relação espacial entre os dados muito dificilmente seria inferida pela simples listagem dos casos de cólera e dos poços. O mapa do doutor Snow passou para a História como um dos primeiros exemplos que ilustra bem o poder explicativo da análise espacial.

11.2 COMPONENTES DA ANÁLISE ESPACIAL

O processo de análise de dados espaciais incluem métodos de visualização, métodos exploratórios para investigar algum padrão nos dados e métodos que auxiliem a escolha de um modelo estatístico e a estimação dos parâmetros desse modelo (Carvalho, 1997). Podemos dividir as ferramentas da Análise Espacial em *seleção, manipulação, análise exploratória e confirmação (modelagem)* (Anselin, 1999).

Por *seleção*, entendemos os processos de navegação num banco de dados geográficos, realizando consultas e apresentando mapas cloropléticos simples. O processo de *manipulação* envolve todas as funções que criam dados espaciais. Aqui encontramos o poder da Álgebra de Mapas em Geoprocessamento, com uma capacidade extremamente ampla de produzir novos mapas.

As técnicas de *análise exploratória* permitem descrever e visualizar distribuições espaciais, descobrir padrões de associação espacial (aglomerados espaciais), sugerir a existência de instabilidades espaciais (não-estacionariedade) e identificar observações atípicas (*outliers*). As técnicas de análise exploratória aplicadas a dados espaciais são essenciais ao desenvolvimento das etapas da modelagem estatística espacial, em geral muito sensível ao tipo de distribuição, à presença de valores extremos e à ausência de estacionariedade (Carvalho, 1997).

As técnicas de *análise confirmatória* envolvem o conjunto de modelos de estimação e procedimentos de validação, necessários para implementar análise multi-variadas com componentes espaciais.

Um conceito chave na Análise Espacial é a dependência espacial, e sua formulação matemática, a autocorrelação espacial. Estas noções partem do que Waldo Tobler chama de "a primeira lei da geografia": "*todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes*". O que quer nos dizer este princípio ? que nada na natureza (como na vida real) acontece por acaso. Se encontramos poluição num trecho de um lago, é provável que locais próximos a esta amostra também estejam poluídos.

O fenômeno de autocorrelação espacial pode ser entendido como uma situação em que observações próximas no espaço possuem valores similares (correlação de atributos). O desafio da Análise Espacial é medir objetivamente este relacionamento.

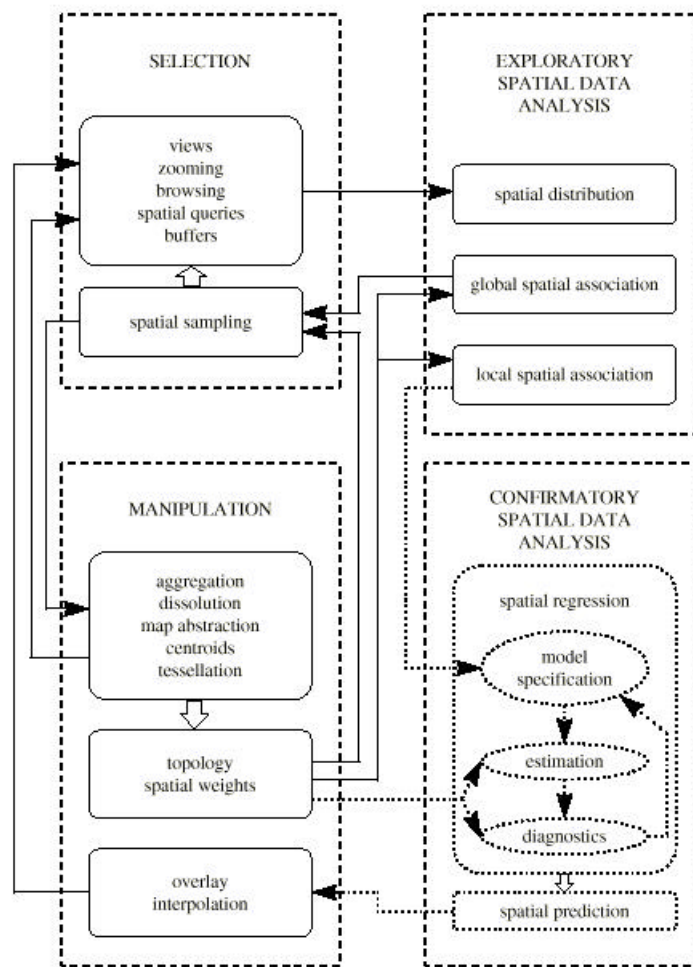


Figura 11.2 – Componentes da Análise Espacial (FONTE: Anselin (1999))

11.3 TIPOS DE DADOS EM ANÁLISE ESPACIAL

De uma forma geral, os problemas de análise espacial lidam com duas classes de dados, num contexto semelhante à divisão geral dos dados de Geoprocessamento em *campos* e *objetos*¹: dados *ambientais* e dados *sócio-econômicos*. Por dados ambientais (ou naturais) referimo-nos a todos os tipos de dados resultantes de levantamento de recursos naturais, e que incluem mapas geológicos, topográficos, ecológicos, fitogeográficos e pedológicos. Chamamos de dados sócio-econômicos todos aqueles decorrentes de levantamento associados recursos humanos, como dados de saúde, demografia, imóveis e rotas de transporte.

No caso de dados ambientais, estes são disponíveis usualmente como um conjunto de valores, que podem estar regularmente ou irregularmente distribuídas, e que podem ser modelados como amostras de uma superfície contínua. Neste contexto, a abordagem geoestatística (descrita no capítulo 12) representa um conjunto de técnicas para reconstruir uma superfície, com base numa caracterização da similaridade espacial entre as amostras (o variograma). Central à abordagem da geo-estatística é o conceito de *estacionariedade*, que supõe um comportamento homogêneo da autocorrelação espacial na região de estudo. Como dados ambientais são resultantes de fenômenos naturais (como os processos geológicos) a hipótese de estacionariedade é sustentada com frequência até certo ponto surpreendente.

Para os dados sócio-econômicos, a situação é mais complexa. Em grande medida, tratam-se de dados associados a levantamentos populacionais, como censos e estatísticas de saúde, e que originalmente se referem a indivíduos localizados em pontos específicos do espaço. Por razões de confidencialidade e de tratamento estatístico, estes dados são agregados em unidades de análise, usualmente delimitadas por polígonos fechados (setores censitários, zonas de endereçamento postal, municípios). A suposição implícita neste caso é que as regiões apresentadas são intrinsecamente homogêneas e que as mudanças significativas só ocorrem nos limites (Martin, 1995). Evidentemente, esta é uma falsa premissa, pois as unidades de levantamento são definidas por critérios operacionais (setores censitários) ou políticos (municípios) e não há qualquer garantia que a distribuição das variáveis socio-econômicas seja homogênea dentro destas unidades.

¹ A divisão entre campos e objetos (apresentada no capítulo 2) reflete uma visão abstrata dos dados geográficos. Neste capítulo, procuramos dar forma concreta a estes conceitos.

Deste modo, podemos postular que a simples apresentação de dados sócio-econômicos como mapas temáticos é insuficiente, de forma geral, para caracterizar o fenômeno em estudo. Acrescente-se a isto alguns efeitos colaterais:

- Em países com grande disparidade de renda como o Brasil, é freqüente que estejam agregados numa mesma região de coleta grupos sociais distintos (favelas e áreas nobres). Assim, o indicador calculado pode representar uma média entre populações diferentes.
- Em estudos de detalhe (como incidência de doenças), muitas vezes as unidades amostrais apresentam número de casos insuficientes para tratamento estatístico. Esta situação requer o uso de ajustes como estimadores bayesianos (Bailey e Gatrell, 1995) para gerar uma distribuição ajustada ou técnicas de agrupamento para obter regiões maiores com número representativo de casos (Carvalho et al., 1998).
- Muitas regiões de estudo (como as grandes regiões metropolitanas brasileiras) apresentam unidades amostrais com grandes diferenças em população e área. Neste caso, tanto a apresentação em mapas cloropléticos como o cálculo simples de indicadores populacionais podem levar a distorções nos indicadores obtidos e será preciso utilizar técnicas de ajuste de distribuições.

Deste modo, a escolha das unidades de coleta e análise é parte crucial do uso de dados socio-econômicos em Geoprocessamento. Estas unidades devem apresentar resolução - definida a partir da menor área para a qual estão disponíveis informações - adequada ao fenômeno que se deseja estudar. Quanto menor a escala, maior a população e a área da unidade de estudo, menor a resolução e, portanto, menor a homogeneidade interna e a capacidade de distinguir diferenças. Aumentar a escala e a resolução traz outros problemas: à medida em que diminui a área e a população, diminui também a ocorrência do evento estudado. Assim, a contrapartida do aumento na homogeneidade é a instabilidade dos indicadores nos grupos (Carvalho, 1997).

O problema da interrelação entre a área de estudo e os resultados mensurados é chamado de MAUP ("modifiable area unit problem" - problema da unidade de área modificável) pois, mudando a escala e as dimensões da unidade de análise, podemos chegar a conclusões bastante diferentes. Dados socio-econômicos disponíveis na forma de polígonos (entidades) sempre apresentarão este problema, independentemente da forma como os limites tiverem sido definidos (Martin, 1995).

Para tratamento num GIS, os dados sócio-econômicos podem ser abordados em duas perspectivas distintas:

- Como um conjunto de polígonos homogêneos, com atributos descritivos para cada sub-região.
- Como um conjunto de amostras (onde cada amostra está associada ao centróide da unidade de levantamento) que representam uma superfície contínua.

Estas perspectivas estão ilustradas na Figura 11.3, em que se comparam as duas formas de conceitualização de dados sócio-econômicos, ilustradas para o caso do IDH (índice de desenvolvimento humano) para o estado de São Paulo, com base no censo de 1991. Na figura 11.3(a), apresentamos um mapa cloroplético construído dividindo-se o valor do IDH em cinco classes, onde se nota um efeito típico de mapas deste tipo: os municípios com IDH entre 0.65 e 0.80 (apresentados em verde) dominam visualmente o mapa, e dão uma impressão de grande homogeneidade no interior do estado. Esta impressão é desfeita ao analisarmos a figura 11.3(b), que apresenta uma superfície contínua que estima o IDH, interpolada a partir dos centróides dos municípios com uso de geoestatística. A "superfície do IDH" mostra uma forte heterogeneidade no estado, com um eixo polarizador que liga São Paulo a Ribeirão Preto, e um acentuado gradiente de queda do IDH nas direções perpendiculares a este eixo, permitindo uma interpretação bastante diferente do mapa anterior.

Outra aplicação importante do uso de superfícies para representação de dados sócio-econômicos é o problema de reterritorialização, em que desejamos transferir dados entre dois recobrimentos espaciais distintos. Neste sentido, cabe lamentar que a maioria dos GIS utilizados para tratamento de dados sócio-econômicos esteja limitado à apresentação de mapas cloropléticos, sem oferecer ao usuário ferramentas adicionais de tratamento dos dados.

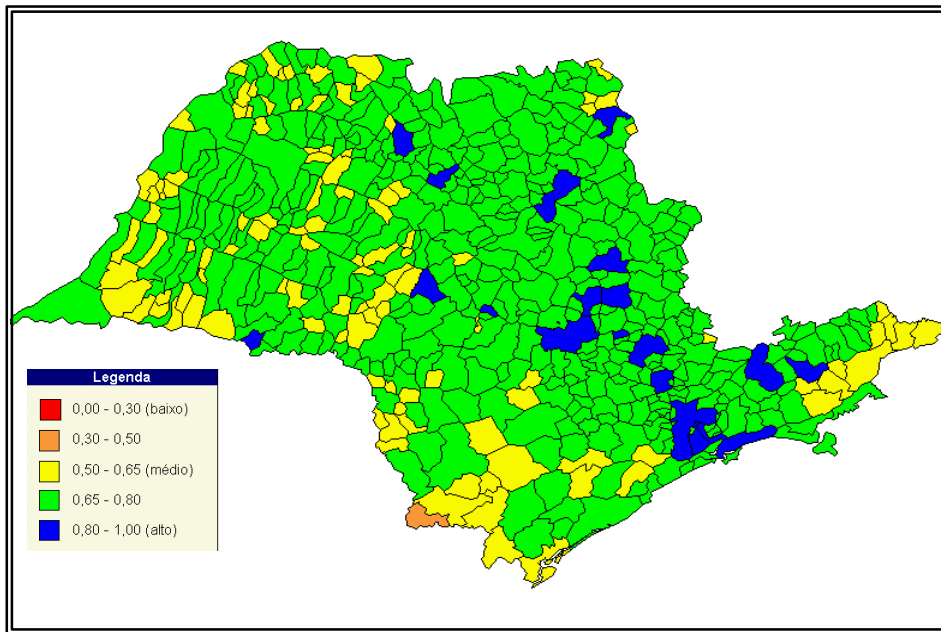


Figura 11.3 (a) Mapa Poligonal do Índice de Desenvolvimento Humano para o Estado de São Paulo (censo de 1991). FONTE: PNUD (1998).

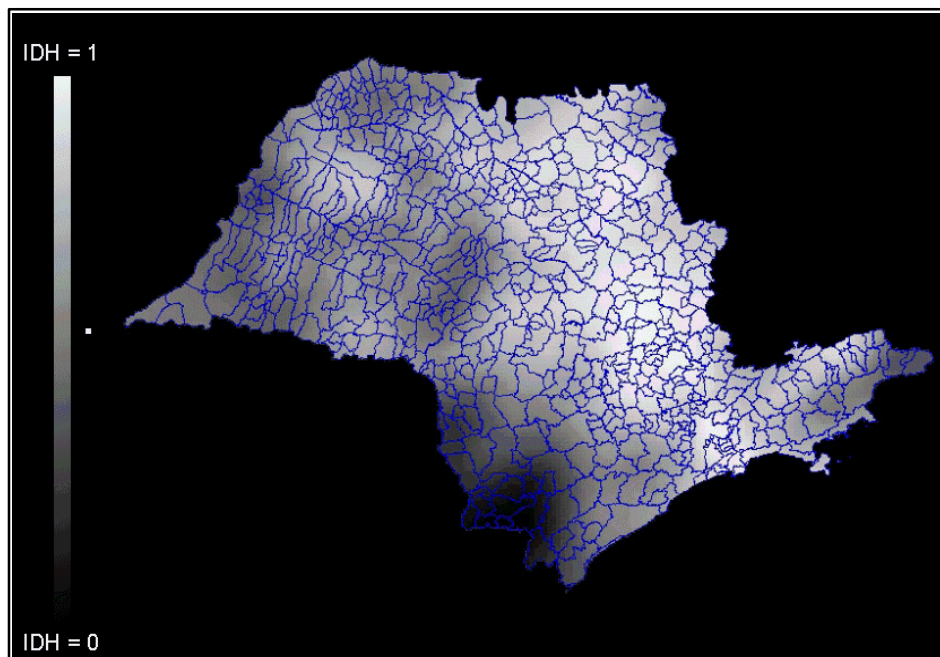


Figura 11.3(b) Distribuição do IDH no estado de São Paulo, estimada por interpolação geoestatística.

BIBLIOGRAFIA

- ANSELIN, L. (1998). "Exploratory spatial data analysis in a geocomputational environment". In: Longley, Brooks, McDonnell, *Geocomputation: A Primer*. London, Macmillan, pp. 77–94.
- ANSELIN, LUC (1999). "Interactive techniques and Exploratory Spatial Data Analysis". In: P. Longley, M. Goodchild, D. Maguire and D. Rhind (eds.), *Geographical Information Systems: principles, techniques, management and applications*. Cambridge: Geoinformation International, 1999.
- BAILEY, T.C., GATRELL, A.C. (1995)., *Interactive spatial data analysis*. Essex, Longman Scientific & Technical.
- CARVALHO, M.S. (1997) *Aplicação de Métodos de Análise Espacial na Caracterização de Áreas de Risco à Saúde*. Tese de Doutorado em Engenharia Biomédica, COPPE/UFRJ. (Internet: <www.procc.fiocruz.br/~marilia>).
- CARVALHO, M. S., CRUZ, O. G., NOBRE, F. F. (1998) "Fuzzy classification applied to modeling socioeconomic profiles for spatial health analysis." In: *Proceedings of the EIS'98 - International ICSC Symposium on ENGINEERING OF INTELLIGENT SYSTEMS*, Tenerife, CD-ROM, paper number 812-170.
- MARTIN, D. (1995) *Geographic Information Systems: Socioeconomic Applications*. 2 ed., London, Routledge.
- MARTIN, D. (1996) An assessment of surface and zonal models of population *International Journal of Geographical Information Systems* 10, 973-989
- MARTIN, D. (1998) "Optimizing census geography: the separation of collection and output geographies." *International Journal of Geographical Information Science* 12, 673-685.