
Geoprocessamento e Pesquisas de Mercado

Por **MundoGEO** | 0100, 31 de Julho de 2008

Por *Eduardo de Rezende Francisco*

Quem nunca foi abordado por um pesquisador interessado em saber sua opinião ou satisfação sobre um produto ou serviço? Ou sobre sua intenção de voto em um candidato a uma eleição política? Acho que a maioria não foi, mas, acreditem, muitos já foram.

Sem dúvida, as pesquisas realizadas em campo (entrevistas na rua, em pontos de grande fluxo de pessoas, em domicílios, em locais de trabalho, em lojas, por telefone, etc.) são um grande instrumento para que a perspectiva do consumidor, cliente, contribuinte ou cidadão seja coletada e utilizada como insumo para gestão, em seu sentido mais amplo. Seu uso, combinado com Sistemas de Informação Geográfica (GIS), pode ser um grande diferencial para quem conhece geotecnologia.

A característica mais importante desse instrumento é que a coleta de informações de pesquisas de campo está sempre associada a uma localização no espaço. Essa localização pode ser o próprio local da entrevista ou uma característica do entrevistado (endereço residencial ou comercial, por exemplo).

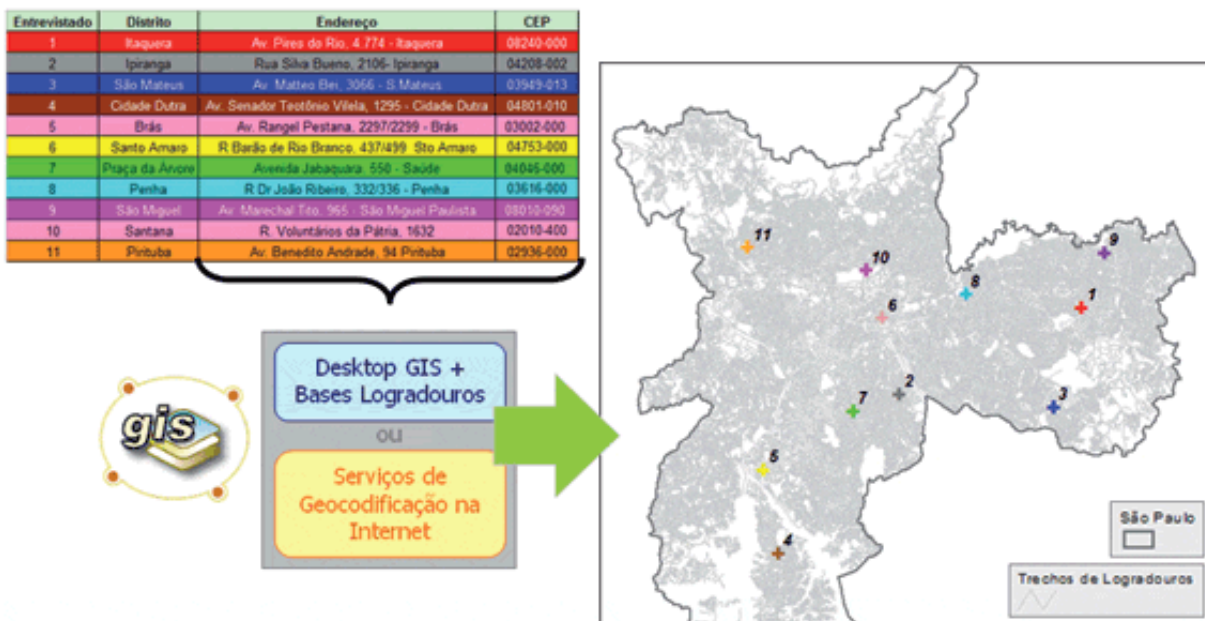
Pesquisas realizadas sob encomenda, ou que sejam totalmente planejadas pelos solicitantes, podem obter essas informações diretamente, ou buscá-las em um cadastro mais amplo com as características dos entrevistados (se forem clientes da empresa, por exemplo). Com o amplo barateamento de dispositivos computacionais e de localização, as pesquisas recentes já podem ser realizadas por pesquisadores dotados de GPS, que no momento da entrevista já coletam as informações desejadas e associam-nas à posição geográfica de sua realização. O recente Censo Agropecuário do IBGE foi conduzido dessa maneira.

A coleta da localização espacial normalmente se dá a partir de um mapa em campo (assinalado), com o uso de GPS, com a anotação do endereço em que a entrevista foi realizada, ou a complementação do endereço do entrevistado a partir de sua identificação em um banco de dados específico.

A associação de um endereço a uma localização no espaço é a principal porta de entrada da contribuição que os GIS podem dar a aplicações ligadas a geomarketing e gestão pública ou privada. Já tratamos disso em inúmeras oportunidades (ver, por exemplo, o artigo “Eixo de Logradouros – Conceitos e Benefícios” nas edições 41 e 42 da revista InfoGEO). Basta que tenhamos uma base de trechos de logradouros e um serviço de busca por endereço ou geocodificação, que conseguimos colocar um endereço no mapa. Endereços (ruas e números) e/ou CEPs permitem essa associação, que nem sempre é trivial se temos endereços incompletos, incorretos ou bases de dados antigas, sem as informações de nomes de ruas e limites de numeração.

Qualquer ferramenta Desktop GIS de mercado permite que se faça esse trabalho. É necessário, no entanto, que se tenham mapas atualizados de trechos de logradouro, o que é a parte mais demorada e custosa do processo. Uma alternativa altamente recomendada é usar e abusar dos serviços de busca por endereço nos sites de mapas da internet. Desde o mais famoso Google Earth até os similares nacionais, todos mantêm bases atualizadas das principais cidades brasileiras. Muitos outros sites permitem que se faça o processo batch de geocodificação – você pode passar uma lista ou tabela de endereços e o sistema irá retornar-lhe as coordenadas geográficas dos endereços fornecidos. Além disso, o Google Maps API disponibiliza esse recurso para quem quiser incorporá-lo em uma aplicação.

A figura a seguir descreve como esse processo pode ser realizado e quais são os elementos importantes que devem estar associados à base de trechos de logradouros.



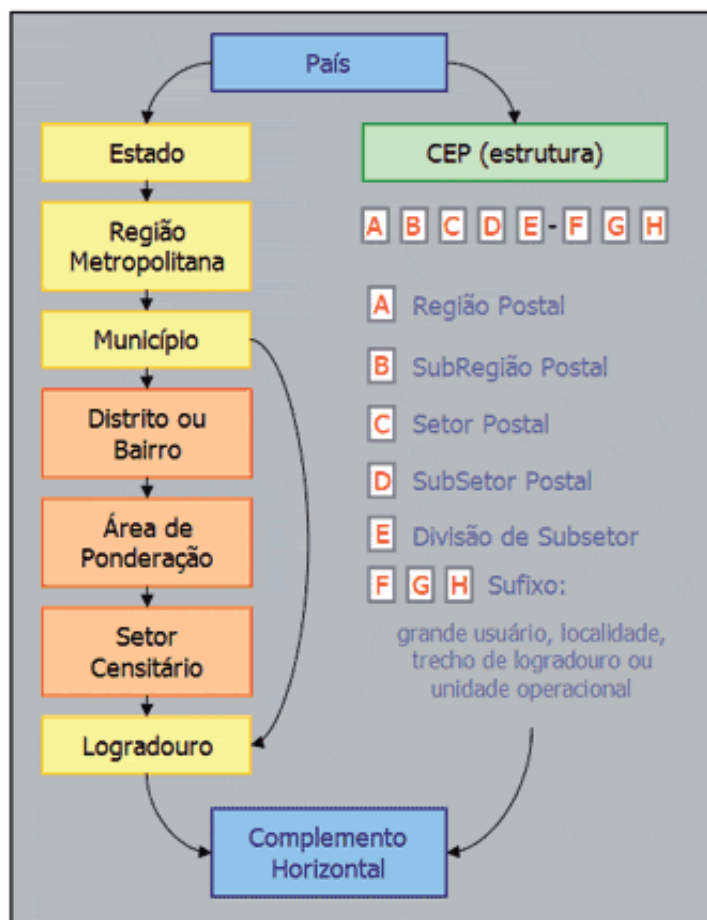
Geocodificação (por endereço ou CEP) em São Paulo

O interessante em usar dados de pesquisas é que nem sempre as informações espaciais estão na granularidade desejada. Muitas pesquisas de campo não são “espacialmente orientadas”, apesar de serem, essencialmente, “espacialmente coletadas”. Nesses casos, as informações da localização espacial do elemento coletado acabam sendo abrangentes, vagas, imprecisas ou mesmo faltantes. Em outras palavras, nem sempre conseguimos chegar a pontos no mapa. É muito comum termos como referência espacial simplesmente o município ou distrito em que a pesquisa foi realizada. Algumas pesquisas destacam apenas o CEP do entrevistado, ou procuram cobrir regiões de CEPs de três ou cinco dígitos. Dessa forma, a representação da observação em um espaço geográfico não pode ser dada deterministicamente como um ponto. A confidencialidade da informação do entrevistado, necessária em algumas pesquisas, muitas vezes colabora para que isso aconteça.

Normalmente isso não representa um problema quando precisamos simplesmente espacializar a informação para a produção de mapas temáticos. É comum realizarmos uma série de entrevistas em uma determinada região, a partir de um planejamento amostral que garanta que aquele número de entrevistas realmente a represente, e depois utilizarmos os resultados sumarizados (por exemplo, média de satisfação) para termos uma informação agregada para a região. Assim, o que estamos mapeando é a região, e não o entrevistado diretamente.

Muitos dados secundários, oriundos de pesquisas sistemáticas do IBGE (PNAD e POF, por exemplo) ou de outros institutos, são representativos de municípios ou regiões metropolitanas do país. No entanto, em seus bancos de dados é fornecida a informação da área de ponderação (conjunto de setores censitários), ou do distrito em que reside o entrevistado, mesmo que esse conjunto de entrevistas não seja representativo da região. Já o Censo Demográfico 2000 do IBGE disponibiliza em sua pesquisa da amostra um banco de dados representativo dessas áreas, conforme discutimos nas edições 43 e 47 da InfoGEO.

O processo de geocodificação, descrito anteriormente, pode se dar em níveis diferentes, e normalmente apenas o endereço completo ou CEP de 8 dígitos está vinculado a um ponto. Os demais níveis são áreas ou polígonos (municípios, distritos, setores postais). A figura abaixo mostra essa hierarquia geográfica. Dependendo da necessidade de mapeamento, saber que determinada entrevista foi realizada em um distrito ou município ou setor postal pode ser mais do que suficiente.



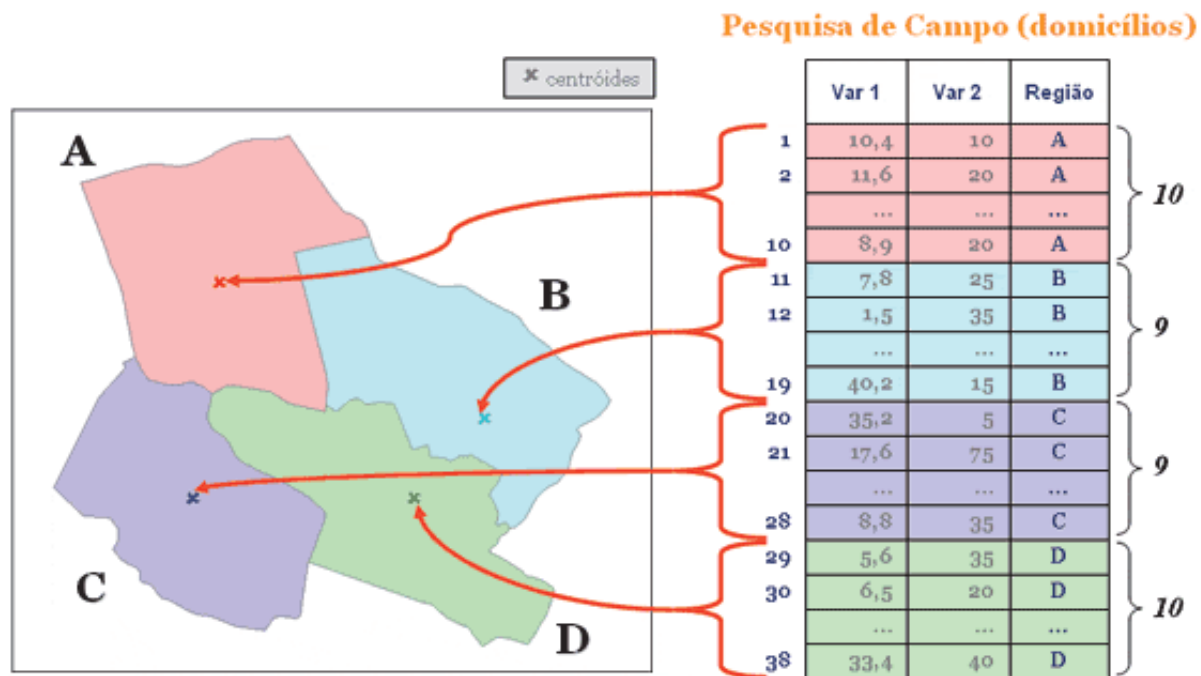
Níveis de Geocodificação

Um pouco mais além: pesquisas e geoestatística

Mesmo que estejamos espacializando dados de pesquisas utilizando regiões de localização, é comum associarmos essas observações a pontos – normalmente aos centróides dos polígonos que descrevem a observação. Assim, se tivermos mais de uma observação coletada por região (o que é muitíssimo comum), acabamos por ter várias observações situadas em um mesmo ponto – o centróide da região pesquisada.

Isso pode representar um problema se estamos pensando em utilizar técnicas de estatística

espacial. Muitos modelos tradicionais, como o Spatial Auto-Regressive model (SAR), não permitem que tenhamos duas observações em um mesmo ponto. Outros modelos, como o Geographically Weighted Regression (GWR), associam pesos máximos para as observações de mesma localização, e pesos zerados para observações situadas em distritos relativamente distantes para suas amostras locais – o que, na verdade, não parece ser uma ponderação razoável e realística. Em outras palavras, quando simplificamos nossa representação espacial, associando várias observações a um mesmo ponto (centróide), passamos a tratá-los, todos, como situados no mesmo endereço, o que não representa, obviamente, a realidade.



Associação de pesquisas aos centróides das regiões

Por conta do isso, algumas técnicas de alocação de pontos podem ser alternativas para o uso simples dos centróides. Este estudo foi denominado Point Allocation Inside Polygons, com alguns resultados apresentados no Colóquio de Geoinformação SIRC 2007, na Nova Zelândia.

Uma primeira alternativa de alocação de pontos é a mais intuitiva: aleatoriamente dentro do polígono. Uma segunda, mais sofisticada, leva em consideração informações que sabemos a priori sobre a região: a distribuição de densidade populacional, ou de domicílios, por exemplo. O interessante desta última abordagem é que essas informações normalmente estão disponíveis a todos, a custos baixos ou inexistentes. Se estamos buscando distribuir entrevistas realizadas dentro de um município, conhecemos, pelo Censo Demográfico do IBGE, a distribuição de domicílios dos setores censitários do município. Setores com mais domicílios têm mais chances de serem sorteados para uma pesquisa.



Alocação de pontos dentro de polígonos

Se sabemos como foi o processo de distribuição amostral da pesquisa, melhor ainda. Podemos excluir áreas rurais ou até tentar reproduzir o processo de estratificação por distritos e sorteio dos setores censitários que terão entrevistas. O conhecimento da localização de lagos, reservas florestais e outros elementos naturais ajuda nesse processo, pois certamente não teremos entrevistas nesses locais.

As alternativas de alocação podem se sofisticar ainda mais, a partir das próprias informações da pesquisa. Outros aspectos sócio-econômico-demográficos, coletados durante as entrevistas, podem servir de base para outros modelos de distribuição. Isso sem dúvida torna o processo de alocação aleatório de pontos bem mais complexo.

Tudo isso pode estar simplesmente adicionando “ruído” aos dados, uma vez que estamos “sugerindo” localizações para os pontos dentro das áreas, ao invés de simplesmente colocá-los em uma posição central. O ideal é que essa técnica venha acompanhada de um processo exaustivo de iteração, repetição da alocação, e que os resultados finais sejam sumários dessa aplicação repetida.

Vejam a seguir uma análise realizada com dados da Pesquisa Abradee 2004 de Satisfação do Cliente Residencial, tradicional do setor elétrico, aplicada no município de São Paulo com 662

entrevistas realizadas em 75 dos 96 distritos da cidade. O estudo buscava medir o quanto o consumo de energia elétrica explica a renda domiciliar, através de modelos estatísticos de regressão. Utilizando técnicas tradicionais (regressão linear), foram obtidos 19,8% de explicação, o que é um valor relativamente baixo. Através do uso da técnica GWR, alocando as entrevistas nos centróides de distritos, esse valor sobe para 45,4%, um número bem mais alto. Porém, quando utilizamos GWR alocando os pontos através das alternativas 1 (totalmente aleatória) e 2 (baseada na concentração de domicílios), os modelos foram processados mil vezes para cada distribuição de pontos, e foram obtidos os seguintes valores médios para o coeficiente de explicação: ele cai para 39,4% e 40,3%, respectivos às alternativas, com um desvio padrão de 3,0% a 3,2%.

Os valores são mais baixos, mas são mais realísticos! A conclusão que podemos chegar é que o GWR, aplicado a pesquisas de campos da forma mais usual, (centróides) fornece resultados que devem ser vistos com certa desconfiança, uma vez que muitas das entrevistas realizadas não foram feitas no ponto central das regiões pesquisadas. Os resultados com a técnica de alocação de pontos parecem ser bem mais factíveis.

Independentemente da sofisticação que busquemos no uso especializado de pesquisas de campo, é importante sempre termos em mente que o mapeamento dessas informações trará um benefício maior do que tabelas e gráficos estatísticos, apoiando tomadas de decisão e auxiliando programas de gestão. Afinal, um mapa vale mais do que mil palavras.



Eduardo de Rezende Francisco

Mestre e doutorando em administração de empresas pela FGV-Eaes

Bacharel em Ciência da Computação pelo IME-USP

Atua em GIS, business intelligence, pesquisas de mercado e estratégias de marketing na AES Eletropaulo

Consultor em geomarketing, geoestatística e microcrédito

Presidente da Gita Brasil (www.gita.org.br)

eduardo.francisco@aes.com