

Indicadores

Neste trabalho, os indicadores foram definidos basicamente de quatro formas: pela proporção de um evento específico em relação ao total de eventos (número de chefes de domicílio mulheres pelo número de chefes de domicílio total, por exemplo), por possuir uma característica específica (estar em área de proteção aos mananciais ou não), pelo padrão espaço-temporal (manter baixas taxas de homicídio em uma década) ou pelo número, médio ou total, de um evento (tempo médio de deslocamento casa-trabalho), divididos em seis dimensões - conforme tabela 1.

Tabela 1
Indicadores

Dimensão/Indicadores	Descrição	Ano/Período	Fonte
CONDIÇÕES AMBIENTAIS			
Áreas de risco	Conter área de risco geológico, por categorias	2009-2010	SMSP / IPT
Proteção aos mananciais	Estar em Áreas de Proteção aos Mananciais ou não	2006	EMPLASA
CONDIÇÕES HABITACIONAIS			
Aglomerado subnormal	Ser um aglomerado subnormal ou não	1991-2010	IBGE
Densidade demográfica	Razão do número de residentes pela área do setor	1980-2010	IBGE
Domicílios permanentes	Proporção de Domicílios Particulares Permanentes em relação ao total de domicílios	1980-2010	IBGE
Domicílios improvisados	Proporção de Domicílios Particulares Improvisados em relação ao total de domicílios	1980-2010	IBGE
Expansão urbana	Momento no qual a área urbanizada recobre mais de dois terços do setor censitário, por faixas de tempo, ou setor rural	1881-2010	EMPLASA / IBGE
Verticalização	Proporção de apartamentos em relação ao total de domicílios	1980-2010	IBGE
CONDIÇÕES SANITÁRIAS E DE HIGIENE			
Atendimento de água	Porção de domicílios com abastecimento de água na rede geral	1991-2010	IBGE
Atendimento de esgoto	Porção de domicílios com instalação sanitária	1991-2010	IBGE
Coleta de lixo	Porção de domicílios com coleta de lixo	1991-2010	IBGE
MOBILIDADE URBANA			
Viagens	Número de viagens	2007	Metrô/SP
Deslocamento	Tempo médio de deslocamento da casa ao trabalho	2007	Metrô/SP
PADROES CRIMINAIS			
Homicídio	Não haver registro de homicídio doloso, as taxas de homicídio são sempre altas, as taxas de homicídio são sempre baixas ou outros	2000-2008	SSP/SP

Continua...

Dimensão/Indicadores	Descrição	Ano/Período	Fonte
PERFIL POPULACIONAL			
População residente	Número de residentes em Domicílios Particulares Permanentes	1980-2010	IBGE
População masculina jovem	Proporção de residentes na faixa de 10 a 19 anos em relação ao total de residentes	1980-2010	IBGE
Chefe de família mulheres	Proporção de chefes de domicílio mulheres em relação ao total de chefes	1980-2010	IBGE
Chefe de família alfabetizado	Proporção de chefes de domicílio alfabetizados em relação ao total de chefes	1980-2010	IBGE
Chefe de família de alta renda	Proporção de chefes de domicílio com rendimento nominal mensal de mais de 20 salários mínimos em relação ao total de chefes	1991-2010	IBGE

Outras variáveis poderiam ser escolhidas, visando caracterizar as áreas chaves segundo diferentes enfoques. A opção apresentada foi orientada pelos trabalhos desenvolvidos pelo NEV-USP, pela bibliografia especializada levantada sobre o assunto pesquisado e pela perspectiva de analisar simultaneamente um conjunto de indicadores sensíveis às condições coletivas de vida na cidade de São Paulo, e, por isso, selecionaram-se indicadores com menor correlação¹.

CONDIÇÕES AMBIENTAIS

“Áreas de risco” foi criada com informações da Secretaria Municipal de Coordenação das Subprefeituras (SMSP) e do Instituto de Pesquisas Tecnológicas (IPT) e “proteção aos mananciais”² com informações da EMPLASA. As áreas de risco são subdivididas em cinco categorias, que variam de acordo com o risco de ocorrência de eventos em situações de chuva intensa e/ou prolongada³: zero (R0: não existem áreas de risco), baixo (R1: não se espera a ocorrência de acidentes), médio (R2: é reduzida a possibilidade de ocorrência de acidentes), alto (R3: é possível a ocorrência de acidentes) e muito alto (R4: é muito provável a ocorrência de acidentes). Por sua vez, as áreas de proteção e recuperação dos mananciais são territórios definidos e delimitados pela a lei estadual 9866, promulgada em 1997 - uma legislação cogente à questão dos recursos hídricos, entretanto sua importância não se deve unicamente ao que tange o meio ambiente. É interessante destacar que a utilização de limites naturais para definir fronteiras que permitam ações de

¹ Quando existe forte correlação entre as variáveis estudadas (variáveis independentes) pode ocorrer multicolinearidade. A multicolinearidade é um problema no ajuste do modelo que pode causar impactos na estimativa dos parâmetros.

² A “proteção aos mananciais” foi determinada pela proporção da área de um setor contida nas unidades de conservação (EMPLASA). Assim, são considerados em áreas de proteção os setores que apresentavam ao menos um terço dos seus territórios nessas unidades.

³ No total foram georreferenciadas 1.179 áreas na cidade de São Paulo. Destarte, havendo no setor mais de uma área de risco, com diferentes categorias, foi considerado o maior risco identificado.

fiscalização é uma atitude estratégica para áreas de preservação, sobretudo quando essas estão conectadas a áreas com propriedades particulares, situação inerente a um contexto que envolve espaços urbanos.

Essas informações tornam-se significativas quando se verifica que áreas de risco estão relacionadas à ocupação desordenada do solo e que as condições socioeconômicas estão diretamente relacionadas com as características propriamente ambientais destas localidades. Breilh et al. (1987)⁴, por exemplo, verifica que “a contradição entre os caracteres benéficos ou destrutivos que podem ter os elementos geoecológicos expressa também as contradições da ordem social estabelecida”. Além disso, ao examinar as áreas de proteção aos mananciais, constatamos que no município de São Paulo cerca de 36% de suas terras estão nessas áreas: precárias condições de moradia e saneamento básico e deficiência dos serviços públicos são alguns dos elementos que compõem a paisagem, sendo contínuo o crescimento de loteamentos irregulares nessas áreas. Os problemas das áreas de proteção tornam-se intrincados e se agravam com o passar dos anos. E cada vez é mais notório o enredamento deste cenário mediante a insuficiência de ações concretas.

CONDIÇÕES HABITACIONAIS

O desenvolvimento da condição habitacional foi mensurado com os dados censitários. As variações dos “aglomerados subnormais”⁵, “densidade demográfica”⁶, “expansão urbana”⁷, “domicílios permanentes”, “domicílios improvisados”⁸ e “verticalização”⁹,

⁴ Breilh J, Granda E, Campaña A, Betancourt O. **Ciudad y muerte infantil**. Investigación sobre el deterioro de La salud en el capitalismo atrasado: un método. Quito: CEAS; 1987.

⁵ Conforme tipologia empregada pelo IBGE (Censo Demográfico, 2010), “Aglomerado subnormal” é um conjunto constituído de, no mínimo, 51 unidades habitacionais (barracos, casas...) carentes, em sua maioria de serviços públicos essenciais, ocupando ou tendo ocupado, até período recente, terreno de propriedade alheia (pública ou particular) e estando dispostas, em geral, de forma desordenada e densa. Por se tratarem de áreas problemáticas sob o aspecto da legalidade da terra e caracterizadas por uma ocupação desordenada, os aglomerados subnormais são normalmente locais com deficiência dos serviços de infraestrutura urbana e com oferta insuficiente de equipamentos públicos.

⁶ As densidades demográficas (pessoas residentes pela área do setor censitário, em Km²) foram subdivididas em quatro categorias: baixa, média, alta e muito alta (tomando-se como referência os valores dos *quartis*).

⁷ A “expansão urbana” é determinada de duas maneiras. Entre 1881 e 2002, corresponde à proporção da área de um setor coberta pela uma mancha urbana de um período (Veja Anexo 1). Assim, são considerados “urbanos” os setores que apresentavam áreas urbanizadas cobrindo ao menos um terço dos seus territórios. Após 2002, o setor é classificado como “urbano” ou “rural” segundo classificação do Censo de 2010 (IBGE) - ou seja, se é “urbano” em 2010 (e não havia sido considerado dessa maneira antes, mediante informação da EMPLASA) considera-se que foi urbanizado entre 2003 e 2010, se não é “rural”. Por fim, o resultado foi subdividido nas seguintes faixas: Urbanizado até 1929, de 1930 a 1949, de 1950 a 1962, de 1963 a 1974, de 1975 a 1985, de 1986 a 1997, de 1998 a 2010 e rural.

⁸ Conforme tipologia empregada pelo IBGE (Censo Demográfico, 2010), “Domicílio Particular Permanente” é o domicílio que foi construído a fim de servir exclusivamente para habitação e tinha a finalidade de servir de moradia a uma ou mais pessoas. “Domicílio Particular Improvisado” é aquele localizado em uma edificação que não tenha dependências destinadas exclusivamente à moradia (por exemplo, dentro de um bar), como também os locais inadequados para habitação e que estavam ocupados por moradores.

⁹ Verticalização é a relação entre apartamentos (definidos por estarem localizados em edifício: de um ou mais andares, com mais de um domicílio, servidos por espaços comuns (*hall* de entrada, escadas, corredores,

todas, possuem o atributo de serem sensíveis a transformações na forma de ocupação habitacional (consolidada ou não, adequada ou desordenada), na tendência de alteração da estrutura urbana, na demanda por recursos ambientais e sociais, na qualidade dos serviços públicos essenciais, e no espaço físico (que pode possibilitar maior ou menor número de inter-relações pessoais e, conseqüentemente maior ou menor número de demandas sociais, disputas, conflitos e crimes).

CONDIÇÕES SANITÁRIAS E DE HIGIENE

Os indicadores “atendimento de água”, “atendimento de esgoto” e “coleta de lixo” foram elaborados com os dados censitários de 1991, 2000 e 2010. Assim, gerados pela proporção de domicílios particulares permanentes com abastecimento de água na rede geral, instalação sanitária ou coleta de lixo, com relação ao total de DPP. A importância desses está no fato de expressarem a qualidade da infraestrutura urbana e a oferta de serviços públicos essenciais para o bem-estar da população.

MOBILIDADE URBANA

Esses indicadores buscam ser uma *proxy*¹⁰ das condições de deslocamento. A importância desses indicadores decorre da grande disparidade no afluxo de pessoas em diferentes localidades da cidade, especialmente em pontos turísticos ou lugares que atraem ou produzem grande número de viagens por trabalho, estudo, lazer, saúde etc. A comparação desses fluxos permite examinar as carências no atendimento da demanda de transporte e do tempo gasto neles e observar um importante aspecto da qualidade de vida dos indivíduos, pois o tempo de viagem está relacionado às atividades remuneradas e à satisfação de necessidades básicas, como de lazer, por exemplo. Ademais, se o movimento pendular da população for significativo, ele pode prejudicar a acessibilidade, agravar as condições de segurança ou, até mesmo, interferir nas relações interpessoais¹¹.

Para retratar a mobilidade na perspectiva intraurbana foram utilizados dados da pesquisa Origem - Destino (Metrô/SP, 2007) e dados censitários da população residente (IBGE, 2010 – estimado para 2007) para gerar dois indicadores: “Viagens” para medir o número de viagens produzidas no setor censitário e “deslocamento” para aferir o tempo médio

portaria ou outras dependências); de dois ou mais andares em que as demais unidades eram não residenciais; e de dois ou mais pavimentos com entradas independentes para os andares) e domicílios totais. Em 1991, apartamento inclui apartamento isolado, apartamento em conjunto popular e apartamento em aglomerado.

¹⁰ Visto que *proxy* é uma mensuração aproximada que se presume guardar relação de pertinência com a variável real. Nesse caso, as condições de deslocamento.

¹¹ Do ponto de vista estatístico, uma grande população flutuante pode inflar as taxas e distorcer as estimativas que tenham apenas a população residente como denominador.

de duração dessas viagens casa-trabalho em minutos¹². No presente estudo, as “viagens” foram calculadas por meio da média ponderada entre o índice de mobilidade parcial (número de viagens da Zona O/D, dividido pelo número de setores que a compõe) e o índice de mobilidade esperada (número de habitantes do setor, multiplicado por 1,95 – valor do índice viagens por habitantes da Região Metropolitana de São Paulo), com peso ao índice de mobilidade parcial igual ao índice viagens por habitantes da Zona O/D, a qual o setor censitário pertence. Os tempos de “deslocamento” foram estimados com o emprego da abordagem metodológica Bayesiana, baseada na aplicação do estimador espacial bayesiano¹³. Inicialmente, com os dados pontuais coletados das viagens¹⁴, mensurou-se o tempo médio por setor censitário (sendo possível gerar essa informação para 8.098 setores, 42,7% do total). Para os setores restantes foi atribuído o tempo médio das viagens produzidas, da Zona O/D a qual pertenciam. Esses tempos foram recalculados por meio do estimador bayesiano empírico local, o qual utiliza as informações dos vizinhos geográficos do setor para medir, neste caso, o seu tempo de deslocamento¹⁵.

PADRÕES CRIMINAIS

O indicador “homicídio” foi baseado nos padrões espaço-temporais da ocorrência de homicídios dolosos no Município de São Paulo (Cf. Nery et al., 2012)¹⁶. Foram identificados sete padrões de distribuição espacial, ou seja, sete regimes espaciais, para a ocorrência de homicídios dolosos, considerando as taxas de homicídios dentro de cada setor censitário e nos setores adjacentes. Três desses regimes são empregados na constituição dos indicadores: setores com altas taxas de homicídios em uma vizinhança com altas taxas; setores com baixas taxas de homicídios em uma vizinhança com baixas taxas; e setores com nenhum homicídio registrado no período 2000-10. Isso porque esses regimes demonstram estar mais fortemente associados com a presença de organizações criminosas e com aspectos sociodemográficos específicos – setores caracterizados pela

¹² Foram consideradas todas as viagens. Pelos motivos trabalho e escola, qualquer viagem independentemente da distância percorrida; para os demais motivos, as viagens a pé foram aferidas quando a distância superava 500 metros.

¹³ Cf. Assunção RM, Barreto S, Guerra H, Sakurai E. Mapas de taxas epidemiológicas: Métodos estatísticos. **Cadernos de Saúde Pública**, 1998;14:713-723.

¹⁴ A Pesquisa O/D de 2007 abrangeu os 39 municípios da Região Metropolitana de São Paulo, teve 54.571 domicílios visitados, totalizando 92 mil entrevistados, 30 mil entrevistas válidas e 431.658 endereços referentes a locais de residência, trabalho, escola; além de origens e destinos de viagens e pontos de transferência entre modos. Essas informações redundam em 196.523 dados pontuais coletados, sendo 151.169, 76,9%, na capital.

¹⁵ Esse tempo, corrigido, é menos instável, uma vez que leva em conta no seu cálculo tanto a informação do setor como a informação dos setores que o circunda.

¹⁶ Nery MB, Peres MFT, Cardia N, Vicentin D, Adorno S. Regimes espaciais: dinâmica dos homicídios dolosos na cidade de São Paulo entre 2000 e 2008. **Rev Panam Salud Publica**. 2012;32:405–12.

sobreposição de carências habitacionais, educacionais e de subsistência, por exemplo, tendem a ter homicídios com maior frequência.

A utilização de apenas um único indicador na categoria “padrões criminais” decorre do homicídio ser um dos mais confiáveis registros criminais, com menor subnotificação, e de sua capacidade de refletir a variabilidade da mortalidade violenta intraurbana. É importante ressaltar que a validade desse indicador não está relacionada apenas ao fato identificar padrões de mortalidade violenta, pois ele também é sensível às alterações socioeconômicas e demográficas.

PERFIL POPULACIONAL

Também baseado nos dados censitários, o perfil habitacional é constituído pela evolução da “população residente”, “população masculina jovem”, “chefe de família mulheres”, “chefe de família alfabetizado” e “chefe de família de alta renda”¹⁷. Cada um desses indicadores mostra-se fundamental para avaliar mudanças nas migrações interurbanas, na vulnerabilidade social¹⁸ e na satisfação das necessidades de subsistência¹⁹.

Análise fatorial

Para estabelecer as características dos setores empregamos os dezenove indicadores propostos, representando seis dimensões: mobilidade urbana, padrões criminais, condições sanitárias e de higiene, condições ambientais, condição habitacional e perfil populacional. Vale lembrar que cada um dos indicadores possui diferente número de medidas temporais (alguns com dados de 1980 a 2010 e outras com apenas uma medida). Foi utilizado um modelo de análise temporal que considera este efeito temporal (*Dinamic Factor Analysis*²⁰) e um escore fatorial único foi determinado para cada indicador. A partir daí, foram calculados os escores fatoriais de cada setor censitário, em cada período, gerando um único fator. Esse método permite a integração e análise de um conjunto de variáveis, tendo por objetivo a construção de índices, capazes de sintetizar os diferentes aspectos descritos univariadamente, que serviram para a análise de *cluster* subsequente.

¹⁷ Assim como a estimativa da população, os indicadores do “perfil habitacional” também foram ponderados considerando os dados de domicílios.

¹⁸ Existem indícios de que lugares com grandes concentrações de homens jovens, bem como de famílias que tem como chefe mulheres ou pessoas não alfabetizadas, tendem a apresentar maiores taxas de vitimização à violência e mais difícil acesso à cidadania e aos direitos sociais.

¹⁹ Verifica-se que a maior proporção de chefes de família alfabetizados ou com alto rendimento pode indicar maior capacidade de uma comunidade local satisfazer demandas por cultura e inserção política e social.

²⁰ Cf. Geweke, J. F., Singleton, K. J., Maximum likelihood “confirmatory” factor analysis of economic time series, *International Economic Review* 22 (1981) 37-54 In: Aigner, D. J., Goldberger, A. S. (eds.), **Latent Variables in Socio-economic Models**, North-Holland, Amsterdam (1977), p.365-382.

Todos os dados utilizados foram apresentados em função da sua forma de mensuração. As variáveis categóricas foram descritas por porcentagens e as variáveis contínuas por meio de média e desvio padrão em cada um dos anos coletados. Para as variáveis que possuem dados coletados por mais de 1 ano (indicadas na Tabela 1) foi utilizada uma técnica de análise fatorial para séries temporais (*time series factor analysis*), utilizando o pacote “tsfa” escrito em linguagem R. Com esta análise foi possível agrupar a série temporal composta por uma mesma variável em uma sequência de anos em uma única coluna no banco de dados que representa a variação desta variável ao longo do tempo. Logo, no momento de descrever o efeito dos *clusters* gerados em função dessas variáveis agrupadas, temos de considerar que essas variáveis tiveram uma variação ao longo do tempo maior ou menor em função de cada *cluster*.

Posteriormente a realização das análises fatoriais para as variáveis que possuem séries temporais, elas foram utilizadas em um modelo de análise de *cluster* (método k-means) juntamente com variáveis que foram coletadas apenas uma vez. As variáveis utilizadas nesse modelo foram descritas na Tabela 1. Lembrando que dos 18.953 setores censitários da capital, 1.421 foram retirados da análise por ausência de dados. Assim, totalizando 17.532 setores (92,5%), dos quais podemos descrever as características a seguir.

Análise de *cluster*

Para análise de *cluster* foram utilizados dois métodos: *Cluster* Hierárquico (para definir o número de *clusters* ideal) e *Cluster* K-means (para identificar quais setores correspondem a cada *cluster*)²¹. Inicialmente, esta análise classificou todos os setores censitários em *clusters*. A partir daí, identificando o centroide de cada um deles, podemos verificar quais são os setores mais próximos do centroide, ou seja, quais setores são os mais representativos de cada *cluster*. Esse método considera os critérios de homogeneidade dos setores em relação ao conjunto de indicadores apresentados anteriormente.

A análise de *cluster* k-means segmenta os setores censitários em função de grupos gerados pelo agrupamento de todas as variáveis incluídas no modelo em um número de agrupamentos preestabelecidos²². O número de *cluster* ideal é definido com base em dois critérios principais: os *clusters* devem ter número semelhante de setores e o modelo deve maximizar a importância de cada variável para a discriminação dos agrupamentos. Ou seja, a maior parte das variáveis utilizadas para gerar cada um dos *clusters*, deve ter diferença estatisticamente significativa entre os *clusters*.

²¹ O índice de significância adotado para todas as análises deste estudo foi de 1% e foram utilizados os softwares R versão 15.1.2, SPSS 20.0 e Stata 11.0.

²² Hair, J. F., Anderson, R. E., Tatham, R. L., Black, W. C. **Análise Multivariada de Dados**. 5. ed. Porto Alegre: Artmed, 2005.

Como todas as variáveis são consideradas conjuntamente no modelo de *cluster*, algumas apresentam impacto maior do que outras para a criação dos agrupamentos. Na tabela a seguir é apresentada a ordem de importância destas variáveis para os *clusters*.

Tabela 2
Ordem de importância das variáveis para o *clusters*

Indicadores	Dimensão	Importância relativa
Expansão urbana	Condições habitacionais	100
Chefe de família alfabetizado	Perfil populacional	50
Chefe de família de alta renda	Perfil populacional	46
Verticalização	Condições habitacionais	38
Domicílios permanentes	Condições habitacionais	28
Densidade demográfica	Condições habitacionais	27
População masculina jovem	Perfil populacional	20
Chefe de família mulheres	Perfil populacional	15
Coleta de lixo	Condições sanitárias e de higiene	14
Atendimento de água	Condições sanitárias e de higiene	14
Atendimento de esgoto	Condições sanitárias e de higiene	14
Aglomerado subnormal	Condições habitacionais	14
Áreas de risco	Condições ambientais	11
Domicílios improvisados	Condições habitacionais	10
Proteção aos mananciais	Condições ambientais	10
População residente	Perfil populacional	8
Viagens	Mobilidade urbana	6
Deslocamento	Mobilidade urbana	6
Homicídio	Padrões criminais	3

Por fim, o último passo foi realizar a descrição dos *clusters* em função das variáveis utilizadas para gerá-los. Para verificar se existe diferença entre as variáveis selecionadas em função dos *clusters*, foram utilizados Análises de Variância (ANOVA) com posthoc Bonferroni e testes qui-quadrado de contingência e os dados foram apresentados em tabelas de contingência e também com uso de gráficos de barra de erro com intervalo de confiança de 99%. Com estas análises foi possível fazer a descrição dos *clusters* e propor nomes para cada um deles.

CLUSTERIZAÇÃO

No processo de determinação dos *clusters* foram utilizados dezenove indicadores (Anexo 2), gerando oito agrupamentos (Anexo 3). Esses agrupamentos exibem a característica de serem similares intragrupos e distintos intergrupos, portanto diferentes em função das médias de cada indicador em um grupo determinado.

Agrupamentos gerados no processo de clusterização

