

**ESTRUTURA FAMILIAR E PADRÕES ESPACIAIS DA ESCOLHA RESIDENCIAL\***

**Tatiana Kolodin Ferrari**

Centro de Desenvolvimento e Planejamento Regional (CEDEPLAR) da Universidade Federal de Minas Gerais (UFMG)

E-mail: [tatianak.ferrari@gmail.com](mailto:tatianak.ferrari@gmail.com)

**Antônio Miguel Vieira Monteiro**

Laboratório de investigação em Sistemas Socioambientais (LiSS). Coordenação Geral de Observação da Terra (CGOBT). Instituto Nacional de Pesquisas Espaciais (INPE)

E-mail: [miguel.monteiro@inpe.br](mailto:miguel.monteiro@inpe.br)

**Pedro Vasconcelos Maia do Amaral**

Centro de Desenvolvimento e Planejamento Regional (CEDEPLAR) da Universidade Federal de Minas Gerais (UFMG)

E-mail: [pedroamaral@cedeplar.ufmg.br](mailto:pedroamaral@cedeplar.ufmg.br)

**RESUMO:** A ocupação residencial das áreas urbanas ocorre de maneira heterogênea, em que se observam aglomerações de grupos similares em determinadas localizações. A estrutura familiar é apontada na literatura como um potencial fator para determinar a localização residencial. A hipótese é que as distintas estruturas familiares têm necessidades específicas de consumo de espaço e de acessibilidade, além do que a renda deve ser considerada de forma *per capita* na função de restrição. Para verificar essa hipótese, o presente trabalho busca explorar a localização espacial de diferentes estruturas familiares no município de São Paulo e observar uma possível relação entre a constituição familiar e os atributos construtivos. A metodologia deste estudo é baseada em análise exploratória de dados espaciais e método de análise de *clusters*. Os principais resultados encontrados evidenciam a existência de padrões de localização com base na estrutura familiar e a renda domiciliar *per capita* como um importante indutor da localização.

**Palavras-chave:** Localização residencial; Padrões espaciais; Estrutura familiar.

**Classificação JEL:** R14; R20; J12.

**FAMILY STRUCTURE AND SPATIAL PATTERNS OF RESIDENTIAL CHOICE**

**ABSTRACT:** The residential land use of urban areas presents spatial heterogeneity, in which agglomerations of similar groups are observed in certain locations. Household structure is pointed out in the literature as a potential factor in determining residential location. The hypothesis is that different household structures have specific needs of consumption of space and accessibility. In addition to that the income restriction should consider the number of members in the household. To verify this hypothesis, the present work investigates the distribution of household structures in the city of São Paulo and observes a possible relationship between household composition and constructive attributes. The methodology of this study is based on exploratory spatial data analysis and cluster analysis method. The main results show the existence of location patterns based on family structure and per capita household income as an important inducer of location.

**Keywords:** Residential location; Spatial patterns; Household structure.

**JEL Codes:** R14; R20; J12.

## 1. Introdução

De acordo com o Censo Demográfico de 2010, 84,4% da população brasileira se encontra nas áreas urbanas (IBGE, 2010), indicando um alto grau de concentração. Essa aglomeração espacial faz com que a atividade residencial tenha grande importância na alocação de usos do solo urbano e se torne responsável por boa parte da configuração espacial das cidades. Em vista disso, o entendimento do comportamento de escolha do local de moradia tem se tornado primordial para o planejamento urbano e para a formulação de políticas públicas.

Cada parcela do solo de uso residencial possui características próprias, assim como também sofre influência dos atributos locais e de vizinhança, fazendo com que cada indivíduo tenha que realizar uma escolha, ponderando as diferentes características. De acordo com os modelos tradicionais da Economia Urbana, os indivíduos realizam um *trade-off* entre acessibilidade, espaço e amenidades locais no seu processo de escolha residencial (FUJITA, 1989).

O trabalho de Alonso (1964) foi pioneiro na modelagem do uso do solo urbano. Tendo como base o sistema de lances, o estudo mostra que os indivíduos escolhem seu local de residência maximizando sua função de utilidade, função essa que depende do tamanho do lote, da distância ao centro da cidade e da restrição orçamentária.

Ao se analisar a configuração dos espaços urbanos, nota-se que, em geral, esses são heterogêneos e são organizados por aglomerações de determinados grupos similares em determinadas localizações. O trabalho de Cecchini et al. (2019) mostra que a atividade imobiliária nas áreas metropolitanas tem se tornado complexa, de forma que está sendo menos influenciada pelas forças econômicas tradicionais (forças de aglomeração) e se tornando mais associada à dinâmica demográfica, de migração, de gentrificação, de acessibilidade e de atrativos culturais/naturais.

Nos grandes centros urbanos, é possível observar um fenômeno de diferenciação espacial com base nas características sociodemográficas. O presente trabalho tem especial interesse em explorar os diferenciais de localização residencial que ocorrem com base na estrutura familiar constituída. Nesse aspecto, o trabalho de Beckmann (1973) foi pioneiro ao estender o modelo de Alonso (1964) considerando o problema da estrutura familiar na escolha residencial. Mas, apesar de vários estudos terem apontado uma relação entre a estrutura familiar e o espaço para habitação, poucos trabalhos têm procurado entender como essa estrutura familiar está associada à localização residencial (JUNG; YANG, 2016).

Dessa forma, é central para ampliar a compreensão da estruturação do espaço intraurbano entender por que determinadas famílias e grupos populacionais escolhem certas localizações de residência e não outras. São essas escolhas que produzem a paisagem urbana. Neste estudo, buscam-se uma exploração e uma caracterização desses padrões de localizações e uma discussão das consequências desses na estruturação do espaço intraurbano de São Paulo. Assim, o objetivo deste trabalho está em responder: Em que medida a estrutura familiar se diferencia espacialmente na cidade? Se isso ocorre, quais características podem desempenhar um papel para essa diferenciação na escolha residencial?

Para a análise, busca-se primeiramente identificar a existência de um padrão de localização, dado o tipo de estrutura domiciliar, através da utilização de análises estatísticas e espaciais. A metodologia deste estudo é baseada em análise multivariada e exploratória de dados. Os resultados encontrados evidenciam os diferentes padrões de localização. Procura-se ainda verificar a existência de características comuns das construções entre as áreas através da análise de *clusters*, a fim de inferir características que podem contribuir para o padrão de localização das famílias.

O trabalho contribui para a literatura ao mostrar que a estrutura familiar é um fator determinante do local de moradia das famílias. Por exemplo, a presença de filhos, ou seja, de dependentes na estrutura familiar, apresentou-se como um fator importante para a localização dessas famílias na borda da cidade, como consequência temos uma demanda diferenciada no local por serviços e tipo de habitação. Essa constatação é relevante, pois, do ponto de vista das políticas públicas, esses padrões geram diferentes demandas locais que devem ser analisadas nas formulações de políticas habitacionais e de planejamento urbano.

Além desta introdução, o trabalho traz uma revisão da literatura na seção 2; apresenta a metodologia utilizada, descrevendo a área de estudo, a classificação da estrutura domiciliar utilizada, a base de dados e o método de análise aplicado, na seção 3; mostra os resultados na seção 4; e, por fim, a seção 5 aborda as principais conclusões.

## 2. Teoria urbana e localização residencial

O principal objetivo de estudo da economia urbana está no entendimento da estrutura interna das cidades, que analisa o mercado de terras, a fim de entender como estão alocados os diferentes agentes e as atividades econômicas (THISE, 2011).

Dentre os diferentes usos do solo urbano, o uso residencial engloba uma parcela importante de consumo do espaço urbano, sendo que, em geral, mais da metade do uso do solo urbano é alocado para essa atividade. Assim, o entendimento da estrutura espacial e das características da alocação residencial é importante para o planejamento urbano das cidades.

O bem terra possui características distintas dos demais bens econômicos. A terra é um bem indestrutível, imóvel e sua oferta é fixa, sendo a função de oferta inelástica no curto prazo, para determinado espaço considerado. Além disso, cada parcela do solo possui características únicas, que refletem suas características físicas e naturais, assim como propriedades institucionais. A ação do homem e o uso destinado a cada parcela do solo afeta o uso e o valor das terras vizinhas, ou seja, cada unidade do bem terra possui externalidades sobre as demais unidades (FUJITA, 1989). Dadas essas características, a pergunta que surge é: como os indivíduos e as famílias decidem onde irão residir?

Desde a década de 1960, um número crescente de estudos tem se concentrado em investigar os determinantes das escolhas de localização residencial urbana e sua influência no surgimento de padrões espaciais capazes de afetar o cotidiano dos habitantes urbanos. A base teórica da abordagem dominante atual para a localização residencial urbana tem suas raízes no modelo de Alonso (1964). Seguindo os princípios defendidos por esses modelos, uma ordem única e eficiente é alcançada por meio de escolhas residenciais que equilibram um *trade-off* entre consumo de habitação e custos de deslocamento para o trabalho.

A ideia é que cada família, ao decidir o local de sua residência, busca maximizar sua utilidade, pondera a acessibilidade ao CBD em relação à sua necessidade de espaço da habitação, estando sujeita a restrições orçamentárias e de tempo. Partindo-se de uma cidade monocêntrica, a ideia é que, quanto maior for a acessibilidade ao CBD, maior será o preço pelo espaço de habitação. As famílias possuem uma renda fixa ( $Y$ ) e a cada período gastam no consumo de bens ( $z$ ) e no consumo da sua unidade residencial ( $s$ ). Assim, dada uma certa distância  $r$  ao CBD, a escolha residencial pode ser representada como:

$$\begin{aligned} & \max_{r,z,s} U(z, s), \\ & \text{sujeito a:} \\ & z + R(r)s = Y - T(r) \end{aligned} \tag{1}$$

em que  $R(r)$  se refere ao preço por unidade de terra na distância  $r$  e  $T(r)$  ao custo de transporte em  $r$ , considerando que  $r > 0$ ,  $z > 0$ ,  $s > 0$ .

Como resultado, os modelos de cidade monocêntrica apresentam uma estrutura em círculos, em que cada classe econômica ocupa determinado anel ao redor da cidade, sendo a dispersão da população uma função unicamente da distância ao CBD.

Esse é o modelo básico que trata a escolha residencial de forma simplificada, deixando elementos importantes de fora da análise. Os espaços urbanos são heterogêneos e possuem uma diferenciação espacial com base em atributos e características dos indivíduos, apontando que existem outras características que condicionam a escolha locacional além desse *trade-off* apontado.

Beckmann (1973) foi o primeiro a trazer o problema da estrutura familiar para a escolha residencial. O grupo familiar possui membros que trabalham ( $n$ ) e membros que são dependentes

desses ( $d$ ). Quanto maior o tamanho da família, maior será a necessidade de consumo de espaço de habitação. Em particular, quanto maior o número de dependentes, os que não trabalham, em relação aos membros que trabalham, menor será o lance de aluguel que uma família conseguirá pagar por um determinado espaço a uma distância  $r$ . Assim, o que Beckmann (1973) procurou responder é como as famílias de diferentes tamanhos e renda irão se localizar e o quanto de espaço vão demandar.

Fujita (1989) apresenta a formalização do modelo de Beckmann (1973) sobre o problema de maximização da utilidade da família. Considerando o tempo dos membros que trabalham na família ( $t$ ), que precisa ser dividido entre o tempo de trabalho ( $t_w$ ), o tempo de lazer ( $t_l$ ) e o deslocamento ( $b_r$ ), temos o problema de maximização da escolha residencial da família como:

$$\begin{aligned} & \max_{r,z,s,t_l,t_w} U(z, s, t_l; d, n), \\ & \text{sujeito a:} \\ & z + R(r)s + nar = Y_N + nW_{tw} \\ & e \\ & t_l + t_w + b_r = t \end{aligned} \quad (2)$$

em que a primeira restrição se refere à renda, em que  $z$  é o consumo de bens,  $R(r)s$  o aluguel por unidade de terra na distância  $r$ ,  $n$  o número de pessoas que trabalham e  $ar$  o custo de deslocamento, que deve ser igual à renda ( $Y_N$ ) mais o salário por hora de trabalho ( $nW_{tw}$ ). A segunda restrição se refere ao tempo dos membros que trabalham na família.

Como resultado, o modelo mostra que: i) quanto maior o número de dependentes na família, maior será sua distância ao CBD, devido à necessidade de maior número de unidades de terra; ii) para uma família que não possui dependentes, a localização é independente do tamanho da família.

Estudos recentes têm enfatizado o padrão espacial que emerge de características culturais, demográficas e de consumo em nível das famílias. Jung e Yang (2016) discutem o processo de suburbanização e o papel da estrutura familiar nesse processo. Os autores partem do pressuposto de que a estrutura familiar está associada com o local de residência, uma vez que a composição e o tamanho da família geram diferentes necessidades. Realizando uma análise para 29 áreas metropolitanas nos Estados Unidos, Jung e Yang (2016) encontraram evidências que suportam a hipótese de que há diferença na estrutura familiar entre a área central e o subúrbio. Particularmente, observaram que indivíduos sozinhos são mais comuns nas áreas centrais e que famílias nucleares estão mais presentes nos subúrbios do que nas áreas centrais.

Dessa forma, tem-se que:

*“one-person households tend to live in central cities. Young adults are known to reside in central city locations due to the neighborhood liveliness and temporary housing consumption patterns. Neighborhood environments such as “proximity to restaurants, mixed-used land uses, and night life” attract young householders. Also, “temporary housing needs” make young one-person householders remain in central city [...] In contrast, married-couple households tend to live in suburbia. Historically, suburbia has been portrayed as an ideal place for married-couple households that have children. The favorable living conditions of suburbia, the provision of safety, a peaceful and ample space, and physical comfort, meet parents’ expectation of raising children”.* (JUNG; YANG, 2016, p. 4).

Buzar et al. (2005) apresentam uma discussão sobre o tema, enfatizando o potencial de estudo abordando diferentes aspectos da estrutura familiar na geografia humana. Dentro dessa discussão, mostram que a ocorrência de concentração de alguns tipos de estruturas familiares em determinadas partes da cidade gera diferenças territoriais em termos de consumo de bens e serviços e na produção do espaço urbano.

Existe uma relação entre o local construído e as características sociodemográficas de sua população. Assim, se existe um padrão de localização com base na estrutura familiar, há demandas

construtivas específicas com base nas suas características, de forma que tal padrão de localização das famílias pode ser impulsionado ou impulsionar determinados padrões construtivos modelando o espaço urbano.

A discussão sobre a suburbanização apresentada por Jung e Yang (2016) pode representar a preferência das famílias nucleares por maior espaço de habitação, de forma que o mercado imobiliário para atender essas famílias cria áreas que atendem aos seus desejos nos subúrbios da cidade. Consequentemente, surge também toda uma infraestrutura, como escolas e creches, para atender a esse tipo de família em específico. Por outro lado, o valor do metro quadrado mais elevado no centro das cidades induz o mercado imobiliário a uma verticalização no padrão de construção, sendo um motivo para afastar certas famílias dessas áreas.

Apesar dessa discussão, pouca importância e estudos empíricos levam em conta a estrutura familiar para explicar o padrão residencial nas cidades. Mais do que isso, Cecchini et al. (2019) chamam atenção que a pesquisa ligando mecanismos locais de expansão urbana (e econômica) e transformações sociodemográficas, como um caminho de desenvolvimento imobiliário e segmentação em mercados de construção, é cada vez mais necessária para fornecer estruturas interpretativas sobre as mudanças espaciais nas metrópoles.

### 3. Metodologia

#### 3.1 Área de estudo

A área de estudo deste trabalho é a cidade de São Paulo, capital do estado de São Paulo. Cidade mais populosa do Brasil, com pouco mais de 11 milhões de habitantes, distribuídos sobre um território de 1.523,3 Km<sup>2</sup>.<sup>1</sup>

A cidade é dividida em nove zonas administrativas (Figura 1), sendo a zona do centro, centro-sul e o começo da zona oeste<sup>2</sup> consideradas as principais áreas de negócio e comércio da cidade, caracterizando-se como o CBD da região.

#### 3.2 Base de dados e algumas especificações

O trabalho parte da hipótese de que existem padrões de localização com base na estrutura familiar. Assim, primeiramente, faz-se necessário classificar esses diferentes grupos.

Considerando a ideia de arranjos familiares<sup>3</sup>, criou-se uma primeira tipologia a partir do chefe do domicílio<sup>4</sup>. Da identificação do chefe do domicílio, são definidas as demais posições dos moradores, sendo: cônjuge ou companheiro, filho ou enteado e demais parentes. Do exposto, foram criadas quatro categorias:

- Arranjo Matrimonial: casal, independente do sexo, unidos matrimonialmente ou por união estável, sem filhos.
- Arranjo Nuclear: casal, independente do sexo, unidos matrimonialmente ou por união estável, com a presença de filhos.
- Arranjo Monoparental: domicílios que tem a presença de apenas um dos pais na criação dos filhos.

<sup>1</sup> IBGE. Dados do Censo Demográfico, 2010. Disponível em: <https://censo2010.ibge.gov.br/sinopse>.

<sup>2</sup> Cabe ressaltar que essa área é similar aos CBDs identificados por Campos (2017).

<sup>3</sup> Segundo Medeiros e Osorio (2001), por arranjo entende-se a combinação de pessoas classificadas segundo diferentes categorias de parentesco (inclusive não parentes) em grupos e/ou subgrupos de residentes em uma determinada unidade domiciliar”.

<sup>4</sup> Segundo o IBGE (2010), o chefe do domicílio se define como a pessoa (homem ou mulher), de 10 anos ou mais de idade, que é o responsável pelo domicílio, ou é assim considerada pelos demais moradores.

- Arranjo Composto: o domicílio composto traz a noção de que o domicílio não abrange apenas o casal e filhos. Pessoas agregadas e com outros vínculos familiares podem estar integradas ao domicílio.

Cabe ressaltar que foi também considerado o arranjo unipessoal. Apesar de os indivíduos que moram sozinhos não serem considerados como arranjos familiares, mas sim como um arranjo “não família”, esses mesmos foram incluídos neste estudo. Sua inclusão é importante devido ao elevado crescimento dessa forma de domicílio nas últimas décadas (ALVES; CAVENAGHI, 2012).

Além de analisar a distribuição espacial, procuramos identificar áreas homogêneas com relação ao padrão construtivo de forma que poderia ser um efeito e/ou afetar a ocorrência de determinado padrão de localização desses grupos. Para esse fim, foram selecionadas as seguintes dimensões do espaço construído: tipo de construção, número de cômodos, renda domiciliar *per capita*.

O tipo de construção indica a preferência dos indivíduos por determinado perfil de densidade local. Locais centrais possuem maior densidade por área, em que predominam as construções verticais, ou seja, prédios em detrimento de casas. Sendo o oposto nos subúrbios.

Já o número de cômodos nos remete à questão do tamanho da família e da preferência por espaço. Como o valor do metro quadrado é mais caro nas regiões centrais, em geral, residências com um número maior de cômodos tendem a ser construídas em áreas periféricas. Exceção ocorre em locais de assentamento informal, os quais são alguns dos principais mecanismos de espraiamento urbano, que se caracterizam por casas pequenas de pouco planejamento.

Para captar essa peculiaridade, considera-se também na análise a renda domiciliar *per capita*. Apesar de se tratar de uma característica do domicílio, a variável é uma *proxy* para o valor do imóvel, mostrando a capacidade de as famílias de determinada renda residirem no local e indicando também valor do aparato construtivo do local.

Acredita-se que essas variáveis afetem a decisão locacional das famílias, assim, a identificação de áreas homogêneas com determinadas características desses parâmetros pode mostrar o porquê da preferência de certos arranjos por determinadas áreas e apontar tendências de localização.

Os dados foram obtidos pelo Censo Demográfico (IBGE, 2010). Neste artigo, foram utilizadas as áreas de ponderação como recorte populacional. No Censo Demográfico de 2010, foram usados métodos e sistemas automáticos de formação de áreas de ponderação que conjugam critérios como tamanho, contiguidade e homogeneidade em relação a um conjunto de características populacionais e de infraestrutura conhecidas. A fração amostral dos domicílios variou conforme o tamanho da população residente em cada município. O município de São Paulo contou com 310 áreas de ponderação, com fração amostral efetiva de 8,81%.

Foram contabilizadas pouco mais de 3,5 milhões de famílias no município de São Paulo. Essas foram então classificadas de acordo com os arranjos descritos anteriormente e, para a análise, transformadas em taxa por mil famílias, de forma que as diferentes áreas de ponderação pudessem ser comparadas. Assim, o cálculo da taxa de arranjo para cada área de ponderação é dado por:

$$\frac{n^{\circ} \text{ de famílias de determinado arranjo na área de ponderação } i}{\text{total de famílias residentes na área de ponderação } i} \cdot 1000 \quad (3)$$

com  $i = 1, 2, \dots, 310$ .

Figura 1 - Zoneamento administrativo da cidade de São Paulo



Fonte: Prefeitura Municipal de São Paulo (PMSP)<sup>5</sup>.

<sup>5</sup> Disponível em: <https://www.prefeitura.sp.gov.br/cidade/secretarias/subprefeituras/subprefeituras/mapa/index.php?p=250449>. Acesso em: 28/01/2019.

### 3.3 Método de análise

A configuração de localização dos arranjos, dada sua estrutura, será explorada com a utilização de técnicas de análise espacial de áreas. Segundo Bailey e Gatrell (1995), o principal interesse nesse tipo de análise está em inferir se há alguma tendência ou algum padrão espacial nos valores dos atributos de um dado conjunto de áreas.

O índice de autocorrelação espacial  $I$  de Moran e os valores de estatística descritiva são utilizados para medir a posição e dispersão dos arranjos sobre as áreas de ponderação da cidade de São Paulo.

O índice de Moran ( $I$ ) mede a autocorrelação espacial através do produto dos desvios em relação à média ( $y^* - y$ ). A sua estimativa é dada por:

$$I = n \frac{\sum \sum W_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum (y_i - \bar{y})^2} \quad (4)$$

em que  $n$  é o número de observações.  $W_j$  é uma matriz de vizinhança para o par  $i$  e  $j$ , e  $y$  se refere à média. Como o interesse está em observar se há comportamento similar entre as áreas de fronteira direta, não importando a direção, adotou-se uma matriz de vizinhança do tipo *Queen* com um grau de contiguidade.

De uma forma geral, o índice de Moran presta-se a um teste cuja hipótese nula é de independência espacial; nesse caso, seu valor seria zero. Valores positivos (entre 0 e +1) indicam correlação direta e negativos, (entre 0 e -1) correlação inversa.

O índice  $I$  de Moran é um indicador global que retorna apenas um único valor de associação espacial para o conjunto todo de dados. Entretanto, é de interesse obter valores específicos para cada área do conjunto de dados.

Para analisar de forma individual, os dados de cada arranjo foram espacializados, utilizando-se a divisão em três classes geradas por quebra natural de Jenks, a qual determina a melhor divisão de valores entre as classes. Em seguida, utilizou-se o Indicador de Associação Espacial Local (LISA), que produz um valor específico para cada área, permitindo a identificação de agrupamentos de áreas com atributos semelhantes (*clusters*) e objetos anômalos (*outliers*). Seu cálculo é derivado do índice de  $I$  de Moran, só que utilizando os desvios em relação à média como uma medida de covariância. De maneira formal, temos:

$$I_i = \frac{(y_i - \bar{y}) \cdot W_{z_i}}{\sigma^2} \quad (5)$$

sendo,  $W_{z_i}$  o valor médio dos desvios dos objetos vizinhos de  $i$  e  $\sigma^2$  a variância da distribuição dos valores dos desvios.

Além disso, tem-se o interesse em verificar se há características espaciais similares entre as áreas de ponderação que condicionam ou são condicionadas pela presença de determinadas estruturas de arranjos domiciliares. Nesse aspecto, foi utilizada a análise de *clusters*, cujo objetivo está em agrupar um determinado conjunto de dados de acordo com algum critério de similaridade. A ideia básica é criar conjuntos que tenham alta similaridade entre si, mas que também apresentem alta dissimilaridade com os objetos que não pertencem ao *cluster*.

A metodologia empregada foi de *cluster* hierárquico pelo método aglomerativo. De acordo com Kaufman e Rousseeuw (2009), esse método parte de todos os elementos separados entre si, designada de etapa 0, em que o número de *clusters* é igual ao número de observações. No caso deste trabalho, em que as observações são as áreas de ponderação do município de São Paulo, partimos de 310 *clusters*. As áreas de ponderação são então agrupadas sequencialmente de acordo com a sua similaridade, dado critério de distância euclidiana no campo das variáveis e método de agrupamento baseado em dissimilaridade máxima.

Os resultados dos *clusters* hierárquicos são apresentados graficamente em um dendograma, cujos nós representam cada etapa do processo de agrupamento. Entretanto, esse método não retorna um

número ótimo de *clusters* a ser considerado. A sua determinação se deu pela análise de demais indicadores através do algoritmo *Nb-Clust*. Esse algoritmo foi desenvolvido por Charrad et al. (2012) para o programa R, justamente para determinar a quantidade ótima de *clusters*, com base na avaliação simultânea de 30 índices.

## 4. Resultados

### 4.1 Análise da distribuição espacial

Apresentamos nesta seção os resultados da análise dos dados e sua distribuição espacial. Como mostra a Tabela 1, das 3,5 milhões de famílias do município de São Paulo, a maioria, 1,4 milhões, são do tipo nuclear. Assim, é de se esperar um maior percentual médio de sua distribuição sobre as diferentes áreas de ponderação. Também é o que possui a maior dispersão dos dados, com desvio padrão de 7,66. O arranjo do tipo nuclear é o mais frequentemente encontrado em todas as áreas de ponderação, possuindo um percentual médio de 40,15%, bastante superior aos demais grupos. O índice global de Moran mostra que existe autocorrelação espacial significativa na localização dos arranjos sobre o território de São Paulo, sendo bastante elevado principalmente entre os unipessoais (0,851) e o arranjo nuclear (0,825) (Tabela 1).

**Tabela 1 - Estatística descritiva da distribuição dos arranjos domiciliares segundo área de ponderação, São Paulo, 2010**

Estatísticas	Unipessoal	Matrimonial	Nuclear	Monoparental	Composto
Total de famílias	512 mil	541 mil	1.422 mil	472 mil	596 mil
Mínimo	5,87	8,93	13,96	5,76	8,01
Máximo	45,27	22,96	54,29	19,53	23,45
Média	14,45	15,23	40,15	13,33	16,81
Desvio-Padrão	7,52	3,27	7,66	2,45	2,83
Mediana	11,89	14,62	41,50	13,46	17,01
Variância	56,64	10,75	58,80	6,02	8,03
<i>I</i> Moran	0,851	0,675	0,825	0,499	0,397

Fonte: IBGE, 2010. Elaboração própria.

O padrão de localização dos arranjos unipessoais fica bastante claro no Figura 2. Grande parte das áreas de ponderação de São Paulo apresentam baixos percentuais de indivíduos morando sozinhos, com exceção de alguns poucos com taxas superiores a 30%, todos localizados na zona central da cidade. O índice local LISA apresenta o padrão alto-alto na região central e o padrão baixo-baixo nas bordas norte, leste e sul da cidade de São Paulo. Apesar dos altos percentuais no centro da cidade, a média da distribuição dos arranjos unipessoais é de apenas 14,45%, decorrência dos baixos percentuais nos arredores da cidade. O desvio-padrão desse grupo foi de 7,52, valor elevado que indica a existência de uma maior dispersão dos dados (Tabela 1).

Analogamente, os domicílios matrimoniais também apresentam concentração no centro da cidade (Figura 3), no entanto, com menor intensidade e maior similaridade entre as áreas de ponderação, tendo um desvio padrão de apenas 3,27 (Tabela 1). Vale ressaltar que o arranjo do tipo matrimonial tende a ser um tipo de arranjo instável, no sentido de que muitos casais sem filhos classificados nessa categoria, no momento, podem estar em transição para um arranjo do tipo nuclear e com isso já procurem locais mais semelhantes a esse outro<sup>6</sup>.

<sup>6</sup> Como apresentado por Jung e Yang (2016), casais recém-casados tendem a morar próximos a centros de emprego, mas sua localização tende a mudar a partir do nascimento de filhos, sendo, portanto, um arranjo localmente instável.

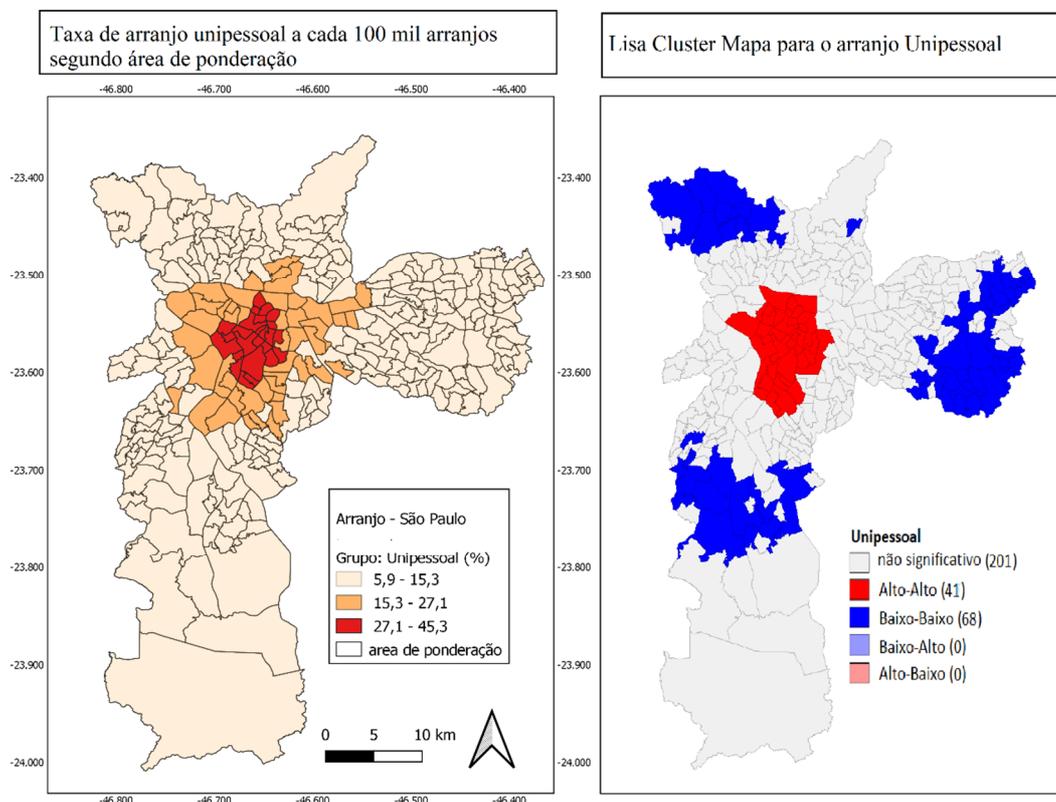
As famílias do tipo nucleares (Figura 4) têm padrão de localização nos subúrbios, com padrão alto-alto nas zonas sul, noroeste e leste.

Com relação aos arranjos monoparentais, esses possuem uma menor representatividade. A sua participação nas áreas de ponderação não passa de 19,53%, com média de 13,33% e mediana no mesmo patamar, 13,46%. O índice global de Moran foi de 0,499. Conforme observado no Figura 5, esses possuem uma tendência de se localizar nas bordas da cidade, afastados do centro, mas não têm um padrão tão homogêneo. Quatro *clusters* de padrão alto-alto foram identificados pelo índice local LISA: 1) no extremo da zona norte, abrangendo as subprefeituras de Perus, Pirituba, Brasilândia e Cachoeirinha; 2) no extremo da zona leste, principalmente nas subprefeituras de Cidade Tiradentes, Guaianazes e Itaim Paulista; 3) na zona sul, nas subprefeituras de Campo Limpo e M'Boi Mirim; e 4) e também zona sul, mais a leste, na subprefeitura de Cidade Ademar.

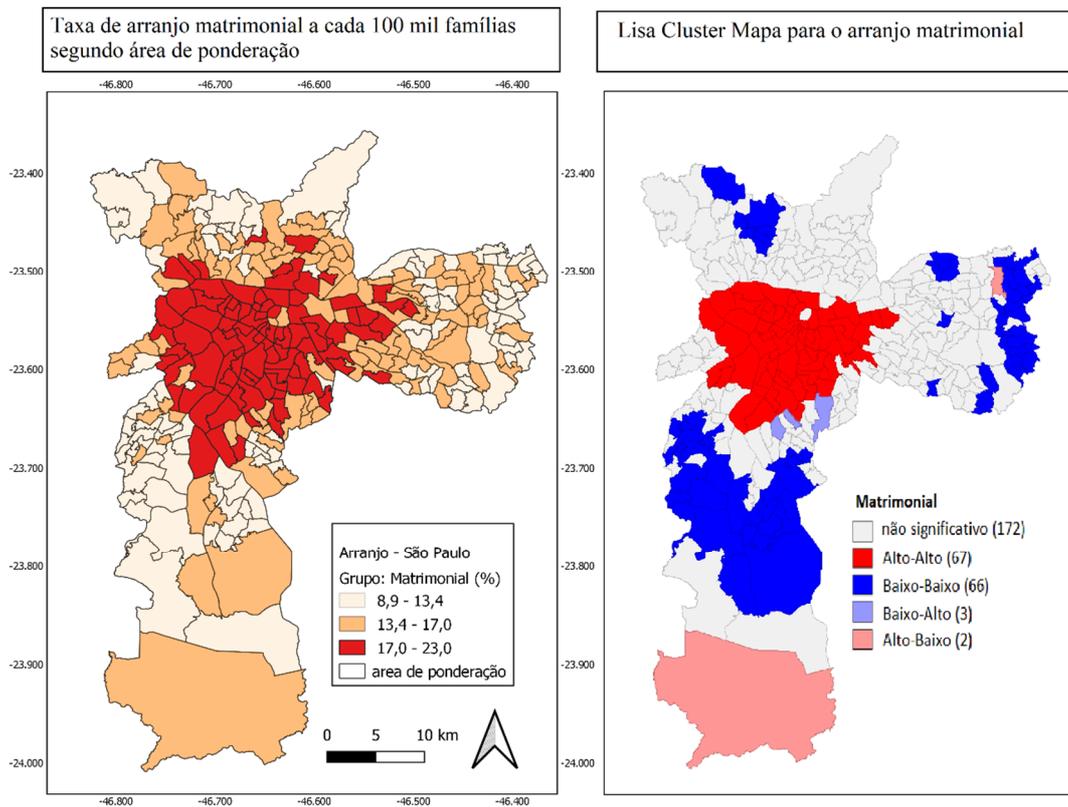
Por fim, os arranjos compostos apresentaram o menor índice de autocorrelação global *I* de Moran, sendo de 0,397. A distribuição do arranjos compostos (Figura 6) também mostra uma tendência a se localizar fora do centro da cidade. Nesse caso, as áreas de ponderação que apresentam as maiores participações, de 18,9% a 23,4%, não são contíguas. O índice local LISA identificou um *cluster* de padrão alto-alto principalmente na região das subprefeituras de Vila Maria e da Mooca.

O fato aqui verificado de que os unipessoais e matrimônios têm uma maior concentração no centro, enquanto os demais arranjos tendem a se localizar afastados desse pode estar ligado a características do local e das construções. O centro da cidade, em geral, é caracterizado por um elevado preço do metro quadrado, construções mais antigas, com menores amenidades ambientais e menor segurança. A preferência de famílias maiores e com filhos em se localizar nos extremos da cidade pode estar ligada a um menor preço do metro quadrado, o que permite adquirir residências maiores, mais condizentes com esse tipo de arranjo familiar. A próxima seção investiga essas hipóteses.

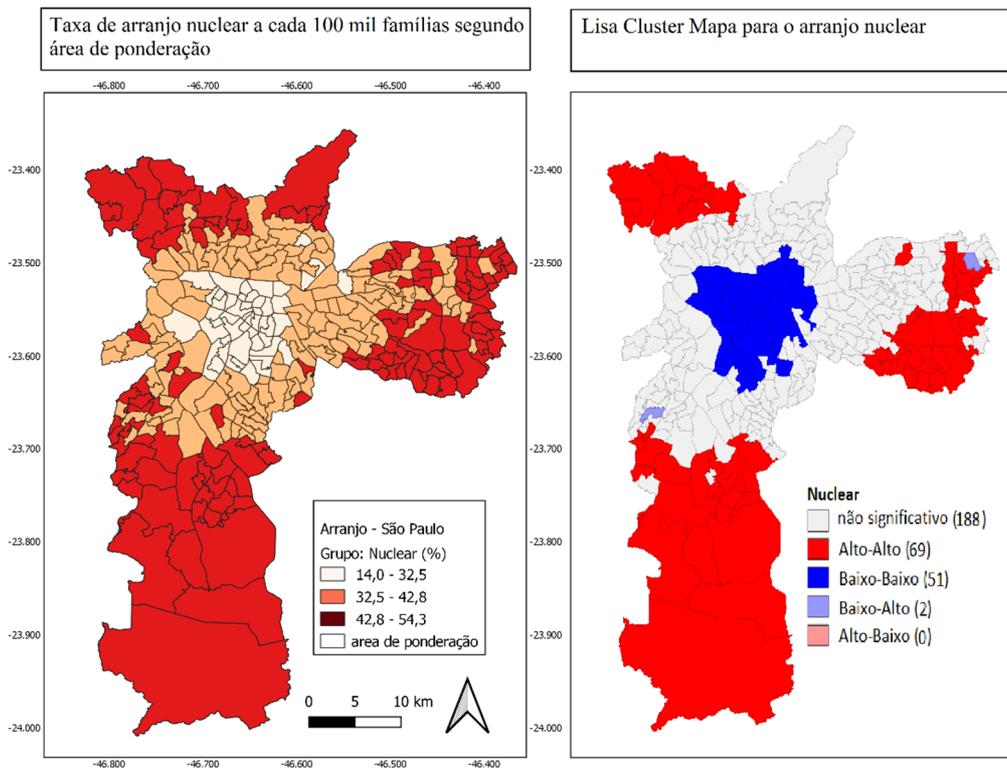
**Figura 2 - Distribuição do arranjo unipessoal e *cluster* espacial por área de ponderação, São Paulo, 2010**



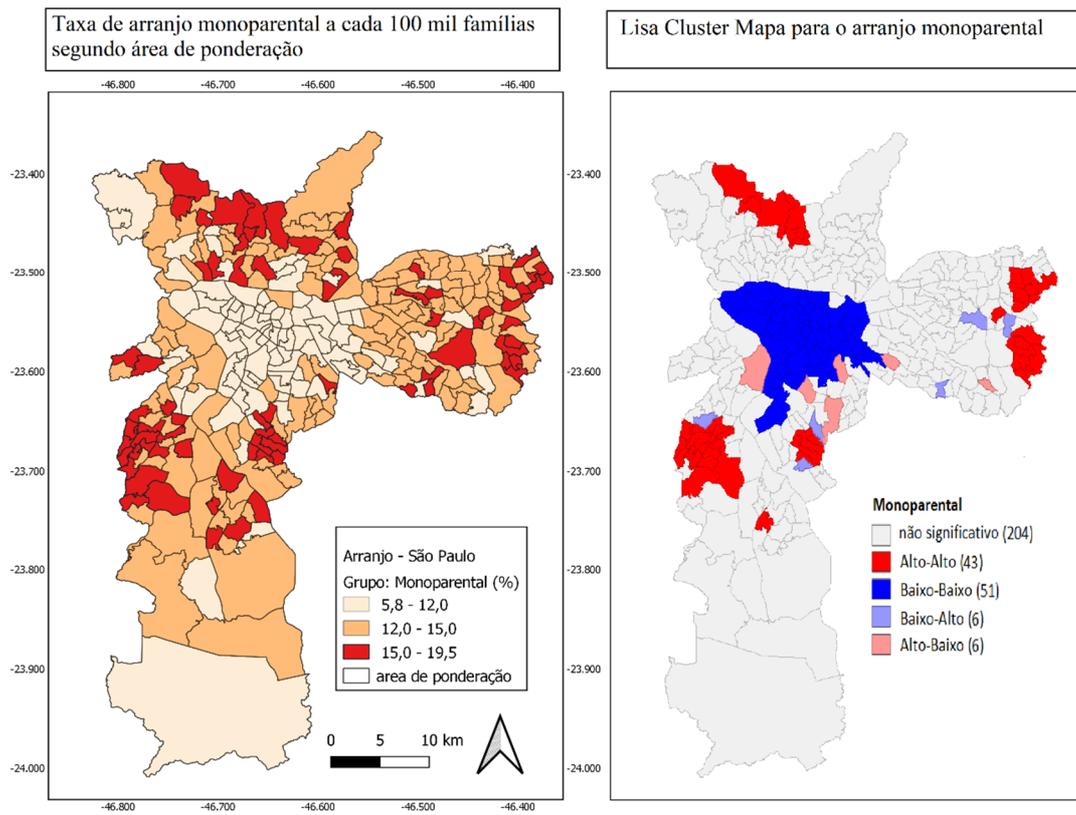
**Figura 3 - Distribuição do arranjo matrimonial e *cluster* espacial por área de ponderação, São Paulo, 2010**



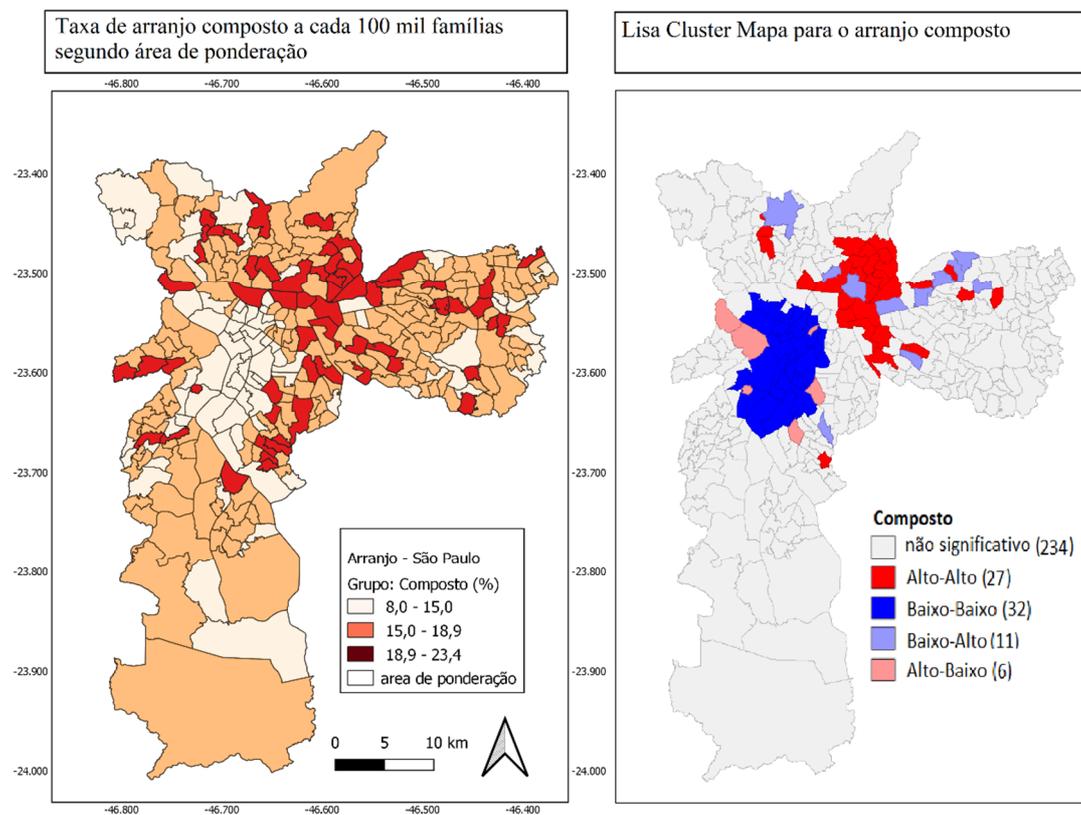
**Figura 4 - Distribuição do arranjo nuclear e *cluster* espacial por área de ponderação, São Paulo, 2010**



**Figura 5 - Distribuição do arranjo monoparental e *cluster* espacial por área de ponderação, São Paulo, 2010**



**Figura 6 - Distribuição do arranjo composto e *cluster* espacial por área de ponderação, São Paulo, 2010**



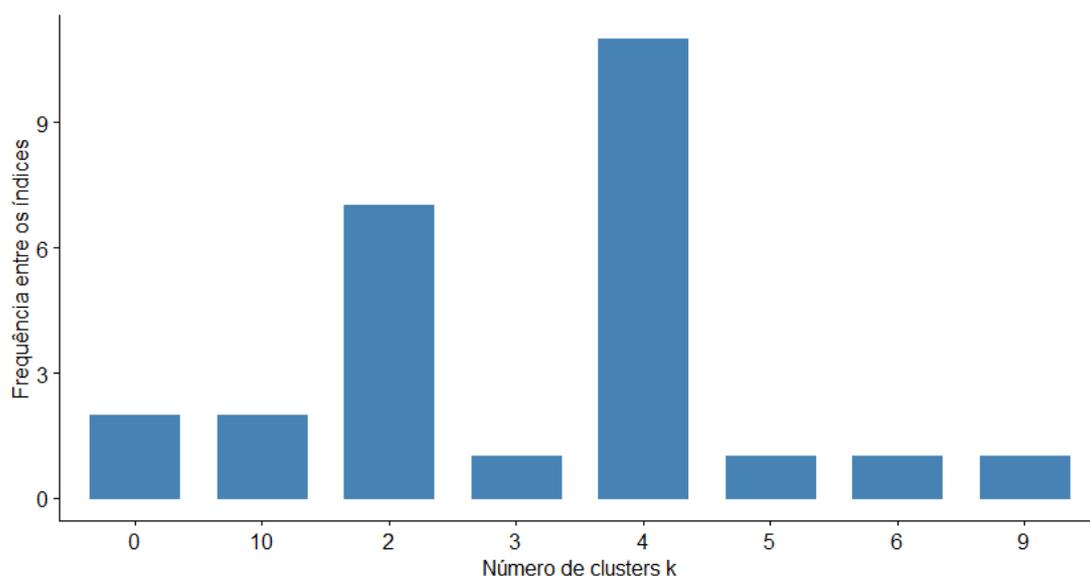
#### 4.1 Análise de clusters

A análise de *cluster* hierárquico visa mostrar as similaridades e divergências no padrão espacial construtivo, observando a média de número de cômodos, o perfil vertical ou horizontal das construções e a renda dos domicílios nas áreas de ponderação da cidade de São Paulo.

Os dados foram padronizados e a similaridade calculada utilizando a distância Euclidiana. Existem diferentes métodos para vincular as variáveis a uma árvore de *cluster* hierárquica. O trabalho realizou essa ligação utilizando cinco algoritmos: máximo, mínimo, pela média, centroide e variância de Ward. Em seguida, foi calculada a correlação entre as distâncias geradas por cada método e a distância dos dados originais, com o objetivo de determinar qual desses métodos reflete com maior precisão os dados originais. O método de máximos, ou completo, retornou o maior coeficiente de correlação, sendo de 0,876. Nesse algoritmo, a distância entre dois *clusters* é definida como o valor máximo de todas as distâncias emparelhadas entre os elementos no *cluster* 1 e os elementos no *cluster* 2 (KASSAMBARA, 2017). Os resultados do dendograma gerado são apresentados na Figura 8.

Outra questão na análise de *clusters* é determinar o número ótimo de agrupamentos. Os resultados da aplicação do algoritmo desenvolvido por Charrad et al. (2012) estão resumidos na Figura 7, o qual mostra que 11 índices determinaram que quatro *clusters* seria o ótimo, em seguida, tivemos 7 índices determinando uma quantidade de dois *clusters*.

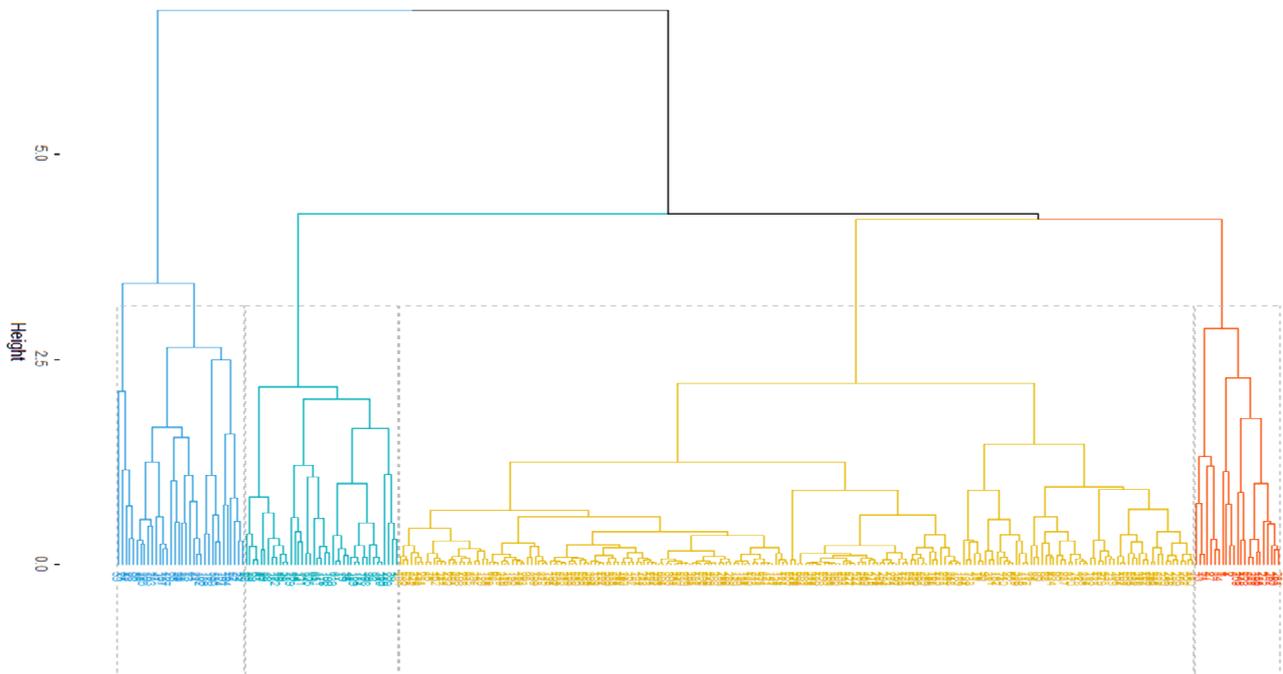
**Figura 7 - NbClust análise para o melhor número de clusters**



Dessa forma, quatro *clusters* foram estabelecidos com base na média do número de cômodos, padrão vertical das construções e a renda domiciliar *per capita*. As cores apresentadas no dendograma da Figura 8 estabelecem a divisão desses grupos.

A maior parte das áreas de ponderação (212) fazem parte do *cluster* em amarelo, grupo caracterizado por baixa renda, padrão horizontal das construções e baixo número de cômodos na residência. Por outro lado, o agrupamento em vermelho possui a menor quantidade de elementos no grupo, apenas 23 áreas de ponderação. Esse agrupamento, assim como o amarelo, tem baixa média de renda domiciliar *per capita* e residências com poucos cômodos, mas se distingue do *agrupamento amarelo* pelo padrão de construções do local, sendo que possuem média de verticalização de 73%. A Tabela 2 apresenta os valores médios de número de cômodos, renda e padrão de construção, medido pelo percentual de construções verticais.

**Figura 8 - Cluster Hierárquico dos padrões de construção por áreas de ponderação, São Paulo, 2010**



**Tabela 2 - Características médias dos clusters**

Cluster	Cor de representação	Áreas de Ponderação	Cômodos	Renda (R\$)	Construção Vertical (%)
1	Vermelho	23	5,0	1758,82	73,25
2	Verde	41	6,6	2822,70	47,30
3	Azul	34	7,9	5694,26	74,62
4	Amarelo	212	4,8	888,96	11,71

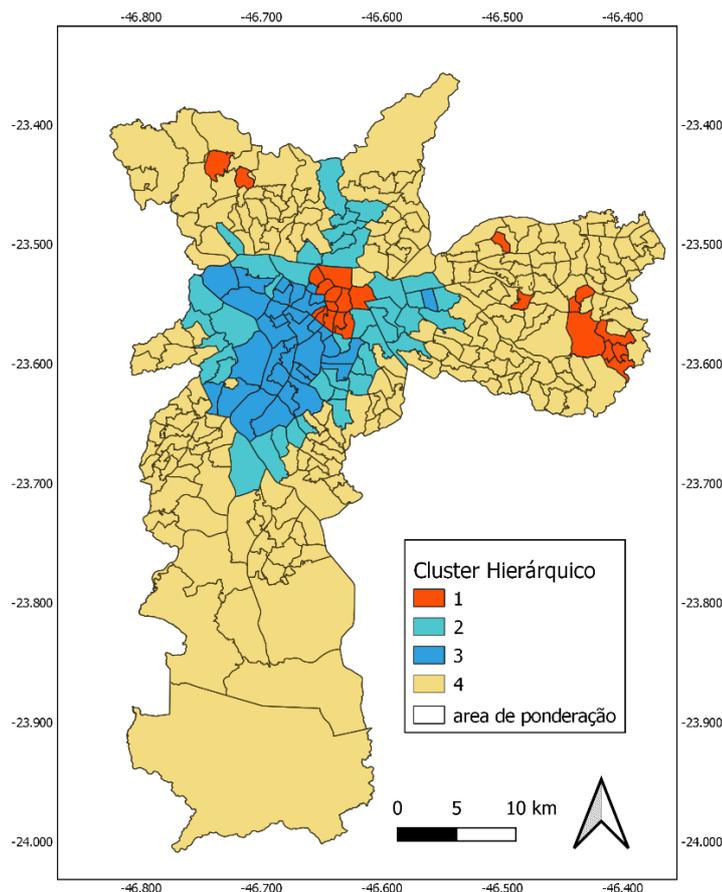
Já os outros dois *clusters*, representados pelas cores verde e azul no dendograma, possuem elevado número de cômodos na residência. O que os diferencia é que o *cluster* em azul, com um total de 34 áreas de ponderação, é caracterizado por áreas de alta renda e padrão bastante verticalizado das construções, enquanto o *cluster* em verde, com 41 áreas de ponderação, remete a locais de renda média e um padrão de construção tendendo a casas, visto que o grau de verticalização é em média de 47%. Para entender como esses padrões construtivos se relacionam com os arranjos domiciliares, os resultados foram especializados, conforme apresenta a Figura 9.

Como se esperava, a região central é mais valorizada, de forma que concentra os domicílios com maior poder aquisitivo. Nota-se que as aglomerações geradas estão fortemente ligadas à distribuição de renda no município, assimilando-se com o padrão de anéis derivado do modelo básico de Alonso (1964).

O padrão espacial das famílias nucleares, monoparentais e compostas nas bordas pode estar ligado justamente a uma menor renda domiciliar *per capita*, visto que esses arranjos possuem maior número de dependentes. Uma extensão para a análise é estratificar os domicílios segundo o número de membros e a proporção dos dependentes no meio familiar, visto que seu padrão locacional indica estar fortemente ligado à renda *per capita*, conforme aponta o modelo de Beckmann (1973). Tal fato aponta a tendência dos arranjos unipessoais de residirem no centro, os quais, possuindo uma maior renda, conseguem se estabelecer nos locais mais valorizados. Importante notar uma peculiaridade na cidade de São Paulo de que residências com maior número de cômodos estão localizadas em áreas de

maior densidade, isto é, com padrão construtivo vertical. Esperava-se que a busca por mais espaço levassem as famílias a residirem em locais com maior presença de casas afastadas do centro, mas a análise apontou que locais de baixa verticalidade tendem a possuir uma média menor de número de cômodos nas residências. O número de cômodos parece estar correlacionado à renda dos domicílios. As áreas de ponderação do *cluster* em azul possuem uma renda domiciliar *per capita* média acima de R\$ 5 mil e uma média de 8 cômodos na residência. Já o grupo do *cluster* em amarelo de menor renda (média de R\$ 888,96) possui em média cinco cômodos em suas residências.

**Figura 9 - Espacialização dos *clusters* hierárquicos, São Paulo**



Uma hipótese é de que as famílias de melhor renda, que buscam a suburbanização como modo de vida, podem estar indo para outros municípios da RMSP, uma vez que essa é altamente interligada, de forma que esse movimento somente no município não seja percebido. No caso de São Paulo, o deslocamento das famílias para as bordas do município parece estar mais ligado a um padrão de renda e de verticalização das construções.

## 5. Considerações finais

A literatura de economia urbana mostra que existe uma relação entre a estrutura familiar constituída e a escolha residencial sobre o espaço urbano. Assim, o primeiro objetivo foi identificar o padrão de localização desses arranjos sobre o território do município de São Paulo.

Em uma análise preliminar, ficou clara a existência de padrões de localização. Em geral, as famílias unipessoais e matrimoniais tendem a se localizar no centro da cidade, sendo essa última mais dispersa. Já as famílias nucleares e monoparentais localizam-se mais afastadas no centro. Famílias

com mais membros, denominadas aqui de compostas, também estão afastadas do centro, localizando-se na região norte e leste da cidade.

A análise de *cluster* do padrão construtivo local estratificou o conjunto de dados em quatro grupos: 1) uma área de alta renda, padrão vertical e elevado número de cômodos na área central do município de São Paulo; 2) circundando esse grupo, temos um *cluster* de renda média, elevado número de cômodos e uma área proporcional em relação a casas e apartamentos; 3) circundando essa, ocupando a maior parte do território, temos um *cluster* de áreas de renda domiciliar *per capita* baixa, padrão horizontal e residências menores; por fim, 4) sem um padrão espacial claro, observa-se um *cluster* de áreas de renda domiciliar *per capita* baixa, residências menores, mas altamente verticalizado.

As análises sugerem a renda domiciliar *per capita* como principal indutor para a localização das estruturas familiares. Existe uma alta correlação entre a renda domiciliar *per capita* e o número de dependentes no domicílio, de forma que o tamanho da família e o número de dependentes passam a ser importantes além do seu arranjo familiar típico. A hipótese é de que os arranjos unipessoais, por não possuírem dependentes, conseguem arcar com um custo maior de moradia na região central do município, com relação aos demais arranjos. Dessa forma, o modelo de Beckmann (1973) parece ser um caminho correto para se tratar a questão.

O padrão construtivo também ficou claro, sendo altamente verticalizado para os arranjos unipessoais e de padrão mais horizontal para os demais arranjos. Mas isso não indica uma preferência por espaço, visto que o tamanho da residência não é um atributo das áreas mais afastadas. Assim, o *trade-off* entre espaço e acessibilidade ao CDB não foi claro na configuração espacial da cidade de São Paulo.

Assim, apesar do mesmo padrão de configuração espacial das estruturas familiares ser encontrado entre este trabalho e de Jung e Yang (2016) nas cidades americanas, as explicações que levam a essa configuração podem ser bem diferentes. Vale ressaltar a diferença nos padrões de periferização entre os dois países. Nos EUA, temos um padrão de suburbanização voltado para o alto consumo do espaço de habitação, caracterizado por amenidades naturais e elevados custos de residência e transporte. O que explica que Jung e Yang (2016) também encontrem forte relação com a renda, mas no padrão americano o aumento de renda é associado com a preferência de residência no subúrbio. No Brasil, temos um duplo movimento de periferização, um tal como o padrão americano, mas na sua maioria esse processo é marcado pela exclusão, sem controle urbano ou social nas periferias, o que resulta em baixo consumo de espaço residencial nas periferias e maior valorização das áreas perto ao centro da cidade.

Os resultados encontrados neste trabalho mostram que a estrutura residencial é um fator a determinar o local de moradia das famílias. Aqui, analisamos apenas a estratificação de acordo com a estrutura familiar, mas a análise de outras estratificações seria interessante em trabalhos futuros a fim de trazer mais evidências do padrão de localização. Do ponto de vista das políticas públicas, esses padrões geram diferentes demandas locais que devem ser analisadas nas formulações de políticas habitacionais e de planejamento urbano. A presença de filhos, ou seja, de dependentes na estrutura familiar, apresentou-se como um fator importante para a localização dessas famílias na borda da cidade, como consequência temos uma demanda diferenciada no local por serviços e tipo de habitação. Este trabalho possui uma série de limitações tanto com relação aos dados como metodológicos. O objetivo foi trazer um primeiro olhar para esse fator que ainda é pouco explorado na literatura. Trabalhos futuros seriam interessantes no sentido de verificar estatisticamente fatores como habitacionais, de vizinhança, cultural, entre outros, a influenciar nesse padrão de localização. Além disso, diferentes estratificações podem ser consideradas, assim como uma análise temporal para observar mudanças e se há uma intensificação de aglomeração espacial das estruturas familiares ao longo do tempo.

## Referências

- ALONSO, W. *Location and land use*. Harvard University Press Cambridge, 1964.
- ALVES, J. E. D.; CAVENAGHI, S. *Tendências demográficas, dos domicílios e das famílias no Brasil*. Aparte: Inclusão Social em Debate, Instituto de Economia, Universidade Federal do Rio de Janeiro, 2012.
- BAILEY, T. C.; GATRELL, A. C. *Interactive spatial data analysis*. Longman Scientific & Technical Essex, 1995.
- BECKMANN, M. J. Equilibrium models of residential land use. *Regional and Urban Economics*, v. 3, n. 4, p. 361-368, 1973.
- BUZAR, S., OGDEN, P. E.; HALL, R. Households matter: the quiet demography of urban transformation. *Progress in Human Geography*, v. 29, n. 4, p. 413-436, 2005.
- CAMPOS, R. The housing market in the municipality of São Paulo: a spatial hedonic prices approach. *Nova Economia*, v. 27, n. 1, p. 303-337, 2017.
- CECCHINI, M.; ZAMBON, I.; SALVATI, L. Housing and the city: A spatial analysis of residential building activity and the socio-demographic background in a mediterranean city, 1990-2017. *Sustainability*, v.11, n.2, p.375-402, 2019.
- CHARRAD, M.; GHAZZALI, N.; BOITEAU, V.; NIKNAFS, A. Nbclust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, v.61, n.6, p.1-36, 2014.
- FUJITA, M. *Urban economic theory: land use and city size*. Cambridge University Press, 1989.
- IBGE. Instituto Brasileiro de Geografia e Estatística. *Censo Demográfico*. 2010. Acesso em: julho de 2018. Disponível em: <<http://www.ibge.gov.br/home/>>.
- JUNG, G.; YANG, T. C. Household structure and suburbia residence in us metropolitan areas: Evidence from the american housing survey. *Social Sciences*, v.5, n.4, p. 74-92, 2016.
- KASSAMBARA, A. *Practical guide to cluster analysis in R: unsupervised machine learning*. STHDA, Alboukadel Kassambara, 2017.
- KAUFMAN, L.; ROUSSEEUW, P. J. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009.
- MEDEIROS, M.; OSORIO, R. *Arranjos domiciliares e arranjos nucleares no Brasil: classificação e evolução de 1977 a 1998*. Instituto de Pesquisa Econômica Aplicada, 2001. (Texto para Discussão, n. 788).
- THISE, J. F. Geografia econômica. In: CRUZ, B. O.; FURTADO, B. A.; MONASTERIO, L. R. JÚNIOR (eds.). *Economia regional e urbana: teorias e métodos com ênfase no Brasil*. Instituto de Pesquisa Econômica Aplicada, 2011.

## ORCID

Tatiana Kolodin Ferrari  <https://orcid.org/0000-0003-1234-1235>

Antônio Miguel Vieira Monteiro  <https://orcid.org/0000-0003-1477-1749>

Pedro Vasconcelos Maia do Amaral  <https://orcid.org/0000-0002-2505-035X>