FUNDAÇÃO GETULIO VARGAS

ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

ELIANA LINS MORANDI

CLOSE BUT FAR: THE REAL DISTANCE BETWEEN BLACKS AND WHITES IN THE LARGEST BRAZILIAN CITY

SÃO PAULO

2020

ELIANA LINS MORANDI

CLOSE BUT FAR: THE REAL DISTANCE BETWEEN BLACKS AND WHITES IN THE LARGEST BRAZILIAN CITY

Dissertação apresentada à Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas para obtenção do título de Mestre em Administração Pública e Governo.

Linha de Pesquisa: Governo e Sociedade Civil em Contexto Subnacional

Orientador: Marco Antonio Teixeira

Coorientador: Gustavo A. A. L. Fernandes

SÃO PAULO

2020

Morandi, Eliana Lins.

Close but far : the real distance between blacks and whites in the largest Brazilian city / Eliana Lins Morandi. - 2020.

104 f.

Orientador: Marco Antônio Carvalho Teixeira.
Co-orientador: Gustavo Andrey de A. L. Fernandes.

Dissertação (mestrado CMAPG) – Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo.

1. Negros - Segregação - São Paulo (SP). 2. Discriminação racial - São Paulo (SP). 3. Negros - Condições sociais - São Paulo (SP). 4. Identidade racial - São Paulo (SP). 5. São Paulo (SP) - Questão racial. I. Teixeira, Marco Antônio Carvalho. II. Fernandes, Gustavo Andrey de A. L.. III. Dissertação (mestrado CMAPG) – Escola de Administração de Empresas de São Paulo. IV. Fundação Getulio Vargas. V. Título.

CDU 572.96(816.11)

Ficha Catalográfica elaborada por: Raphael Figueiredo Xavier CRB SP-009987/O
Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

ELIANA LINS MORANDI

CLOSE BUT FAR: THE REAL DISTANCE BETWEEN BLACKS AND WHITES IN THE LARGEST BRAZILIAN CITY

Dissertação apresentada à Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas para obtenção do título de Mestre em Administração Pública e Governo.

Data de aprovação: 17 de fevereiro de 2020.

Marco Antonio Teixeira
Orientador - FGV EAESP

Gustavo Andrey de Almeida Lopes Fernandes
Coorientador - FGV EAESP

Eduardo Cesar Leão Marques
DCP-USP

Eduardo de Rezende Francisco
FGV EAESP

Flávia da Fonseca Feitosa
UFABC

FINANCIAL SUPPORT

This study was financed by the Brazilian National Council for Scientific and Technological Development (CNPq) and by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES) - Finance Code 001.

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq) e com apoio parcial da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

AGRADECIMENTOS

Agradeço à Milena, que me acompanhou de perto nesses dois anos em um processo psicoterapêutico fundamental à manutenção da minha saúde mental e à conclusão desta pesquisa com todo fôlego que consegui nela depositar; ao grupo de estudos do CEPESP, em especial, à Tainá Pacheco, pelas sugestões ao projeto de pesquisa, pela paciência e ajuda fundamental na implementação das matrizes de tempo de deslocamento, e ao Guilherme Russo pelo entusiasmo e dedicação na leitura do trabalho e nos comentários; à família que comigo convive todos os dias, especialmente à Michele, por ter me aproximado do lugar que ela ocupa e da perspectiva segundo à qual ela vê e é vista, e por ter feito com que eu pudesse enxergar mais de perto, ganhando forma, carne e osso, o que tanto lemos sobre racismo nos papers e nos livros; ao Rodolfo, pela escuta sincera nos momentos de angústia que permeiam qualquer projeto de pesquisa, e pelos pitacos na implementação de alguns algoritmos; às(aos) amigas(os) Gabriela, Livia, Tainá, Camila, Leo, Juliana, Filipe, Lara, Guilherme e Caio, e às(aos) demais colegas que tive o prazer de conhecer nesse processo curto e intenso de pós-graduação e que fizeram a caminhada uma pouco mais divertida; aos membros das bancas de qualificação e defesa, pelo tempo e atenção, e pelas sugestões. Especialmente, aos professores Eduardo Francisco e Gustavo Fernandes, que acompanharam esta pesquisa um pouco mais de perto.

RESUMO

A democracia racial é, há muito, um mito questionado no Brasil. Todavia, níveis médios e baixos de segregação têm sido encontrados para metrópoles brasileiras, em contraste com uma vasta literatura qualitativa indicando separação social entre negros e brancos no Brasil. Isso pode resultar de um problema de mensuração, visto que a maior parte dos estudos utiliza o Índice de Dissimilaridade em sua versão aespacial (D), ainda que sua consistência seja amplamente questionada: D não leva em conta relações de vizinhança entre as unidades geográficas, é sensível à forma como a área de estudo é subdividida e pode falhar em capturar certos padrões de segregação, a depender da escala na qual este fenômeno acontece. Como áreas onde predomina um padrão de micro-segregação estão pouco documentadas pela literatura empírica quantitativa sobre cidades brasileiras, essa pesquisa busca enriquecer a discussão teórica sobre segregação no Brasil. D é comparado com uma medida espacial de Entropia (REARDON; O'SULLIVAN, 2004) na parte Centro-Oeste da Região Metropolitana de São Paulo – região de urbanização mais antiga e onde se encontra o centro financeiro e comercial da metrópole. Usando dados do último censo demográfico (2010), a Entropia foi calculada considerando tanto a distância euclidiana, quanto o tempo de deslocamento (Open Street Maps) para estabelecer relação de vizinhança entre os setores censitários, sendo o tempo de deslocamento uma melhor *proxy* para a distância entre dois pontos em um espaço urbano. Ambas as distribuições (de D e da Entropia) foram trianguladas com a descrição estatística da condição socioeconômica dos grupos raciais, com uma análise qualitativa da paisagem por meio de imagens de satélite e da vista da rua (Google Maps), com a composição racial de escolas e com a segregação escolar na mesma área, calculada com dados do censo escolar (2010). Os resultados sugerem que brancos têm escolaridade mais elevada e salários mais altos que a população negra: 27% dos brancos chegam à universidade, contra 8.7% de pardos e 11.5% de pretos; 15% dos brancos estão entre os salários 10% mais altos, em contraste com 2% e 3% de pardos e pretos, respectivamente. Salários mais elevados permitem certas escolhas em termos de serviços. No que se refere à educação, escolas privadas e racialmente homogêneas são claramente uma escolha das famílias de classe média e alta, já que a educação pública é gratuita e universal. Essa escolha se estende ao local de residência. Áreas com baixa Entropia (alta segregação) estão correlacionadas com o precentual de residentes brancos (correlação parcial de cerca de 0.60) e de domicílios de alta renda (correlação parcial de cerca de 0.23). D não foi capaz de capturar este aspecto da distribuição racial na mesma área, porque este indicador não leva em conta a composição racial das unidades geográficas vizinhas. Além disso, bairros ricos e brancos são marcados pela existência de condomínios fechados e com infraestrutura de segurança. Portanto, brancos de classes média e alta parecem ter desenvolvido suas próprias formas de auto-segregação que parecem caracterizar um *white flight* à brasileira, ainda a ser testado em estudos futuros.

Palavras-chaves: Raça; Segregação; Segregação Racial; São Paulo.

ABSTRACT

Brazil being a racial democracy has been for long a questionable myth. However, researchers have been finding low and medium levels of racial segregation in Brazilian cities, in contrast with a wide qualitative literature documenting strong social separation between blacks and whites in Brazil. This might be a measurement problem, since most of the research on the topic measures segregation through the aspatial Dissimilarity Index (D), even though D has been shown to be questionable. For instance, it doesn't take neighborhood relationship between geographic units into account and heavily rely on their boundaries. It might also fail to capture certain segregation patterns, depending on the scale in which segregation takes place. In addition, areas where micro-segregation is predominant in Brazilian cities lack documentation in a more quantitative manner. Hence, this research is an attempt to enrich the theoretical discussion on how the phenomenon of segregation takes place in Brazilian metropolises by comparing D with a measure of Spatial Entropy (REARDON; O'SULLIVAN, 2004) in the Center-Western region of São Paulo Metropolitan Area – the region of oldest urbanization and where the Central Business District is located. Using the last Brazilian demographic census data (2010), Entropy was calculated with both straight line distance and travel time (Open Street Maps) in order to establish neighborhood relationship between census tracts – travel time being a better proxy for the social distance between two individuals in an urban setting. Both D and Entropy distributions were triangulated with a statistical description of racial groups' socioeconomic conditions, qualitative landscape analysis through satellite and street images (Google Maps), as well as the schools' racial composition and school segregation index in the same area, calculated using 2010 school census data. Results suggest that whites are more educated and earn higher wages than blacks: 27% of whites attend college *versus* 8.7% of browns and 11.5% of blacks; 15% of whites earn the 10% highest wages while this is the case for 2% and 3% of brown and black population, respectively. Higher income levels allow choices in terms of service provision. Regarding education, the racially homogeneous private schools are clearly a choice of white middle- and upper-class families, since public schools are free of charge and the Brazilian state guarantees universal access to basic education (K-12). Those choices are extended to the place of residence as well. Areas with low Entropy levels (high segregation) are partially correlated to the percentage of high-income households (≈ 0.60) and percentage of white residents (≈ 0.23). D was not able to capture this aspect of racial distribution in the same area, because this index only takes into account the racial composition within the geographic unit, but not its neighboring units. Those rich white neighborhoods are also marked by the existence of private condominiums well equipped with security infrastructure. Hence, the Brazilian white middle- and upper-classes seem to have developed its own forms of self-segregation that might characterize a white flight *à la brésilienne* to be tested in further research.

Key-words: Race; Segregation; Segregation measurement; São Paulo.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Brazil had been known for being a racial democracy, a myth built from Gilberto Freyre ideas and widely questioned nowadays. In spite of being a miscegenated nation, it is still a developing and unequal country, and race is an important dimension of those inequalities. In this context, urban space constitutes an important arena in which those inequalities take place. Segregation, understood in a simple way as the separation of social groups in space, is one of the possible lens through which urban inequalities can be analyzed.

More segregated cities have reduced chances of social contact between members of different groups and usually engenders an unequal access to urban resources (BARROS; FEITOSA, 2018; VILLAÇA, 2011; TORRES, 2004; MARICATO, 1996). São Paulo has been largely described as a segregated city marked by different segregation mechanisms (CALDEIRA, 2000) leading to multiple segregation patterns. Thus, a comprehensive description of the phenomenon should take different measures into account. Qualitative literature describes São Paulo as a segregated city. However, research measuring racial segregation in São Paulo has mostly found low and medium segregation levels. This might be a measurement problem since the most popular index, the Dissimilarity Index (D) has many limitations and doesn't always succeed in capturing social distance, specially when there is physical proximity with no social proximity.

Barros and Feitosa (2018) is, to the best of my knowledge, the single work measuring racial segregation with spatial indices and cross-analyzing more than one index. They develop an in-depth methodological discussion on segregation measurement, although the authors do not detail the broader theoretical discussion on its causes and consequences in the specific contexts – London and São Paulo. Even though, the measures they apply and their insights should be taken into account in further developments in this theoretical debate.

Hence, this research applies Reardon and O'Sullivan's (2004) Spatial Entropy measure using both euclidean distance and travel time between census tracts when establishing neighborhood relationship between tracts. This measure is compared with the aspatial Dissimilarity Index. Those are triangulated with school segregation, demographic and school census data and a qualitative analysis of landscape through satellite and street view images. Thus, this research aims to be a theoretical contribution to the characterization of segregation in the Brazilian metropolitan context, as well as a methodological contribution to the residential segregation measurement applying Brazilian public data, as well as open travel time data.

This research is structured in 7 chapters. After this Introduction, Chapter 2 discusses racial classification as a basis to the debate on segregation measurement. In Chapter 3, the concept of segregation is discussed, as well as the main streams that have been debating this phenomenon, with a focus on the case of São Paulo Metropolitan Area. In Chapter 4, demographic and school administrative data is analyzed in order to understand the contrasts between racial groups; urban contrasts are highlighted through a LISA Map and qualitative analysis of Google Maps images. In Chapter 5, the empirical strategy here adopted is detailed. In Chapter 6, racial residential micro-segregation indices are described, compared and triangulated with school segregation and other demographic and administrative data. Finally, conclusions are highlighted in Chapter 7.

2 RACE & CLASSIFICATION

The concept of race has for long been an object of political dispute and source of power exercise of one group of people over others. Defining ourselves in contrast to the *others* – or conversely, defining *others* in contrast to ourselves – has been present in human history since ancient times, although under other terminologies. More recently, racial classification was the basis for the emergence of nation states and eugenics movements and policies starting in Europe in the XIXth century and spreading through much of the Western world, not long after the term race was used for the first time in 1684 by a French physician (HACKING, 2007).

In the context of the development of the eugenics movements, mixed racial backgrounds were devalued as degenerated, which jeopardized the Brazilian racial identity. Brazil was considered the ideal of degeneration in the tropical world as a consequence of miscegenation resulting from depraved behaviour. This was upheld by many foreign intellectuals from the rich world and became a social issue when the idea was received by local elites (STEPAN, 2004). For instance, Compte de Gobineau stated the inferiority of the Brazilian "best families" – referring to the elite – impregnated by the ugliness, laziness and infertility because of the miscegenation process in place in his *Essay sur l'inégalité des races humaines* (TELLES, 2004).

As a response, the first eugenic society in Brazil was founded in 1918, one decade after its British counterpart. The movement started in the state of São Paulo and mobilized almost the entire small intellectual elite of the state capital and the countryside. The scientific character aimed by the movement gave it the air of modern times, aligned to the ideal of a recently proclaimed Republic, even though Brazilian universities hadn't any Science department yet. Medicine schools such as the Oswaldo Cruz Institute leaded the eugenic studies in Brazil and were cherished by the Brazilian society for the success of public health campaigns, such as vaccination against smallpox, bubonic plague and yellow fever (STEPAN, 2004).

However, according to Stepan (2004), it was not the Kehl's[1] theory that became a consensus, but rather the anthropological and anti-racist racial ideology attributed to Gilberto Freyre did. Freyre's key ideas defining Brazil as a racial democracy have strongly predominated in the national narrative from the 1930s until the 1970s, when social scientists started to systematically question the validity of this theory until today (STEPAN, 2004; TELLES; 2004).

Despite the uncountable attempts to define race from a scientific or biological approach, race cannot so far be biologically defined, with clear-cut genetic or geographic boundaries (BARBUJANI, 2005). In addition, racial classification systems exist much before human DNA could be mapped by scientists and apply to many quotidian situations where any direct appeal to scientific arguments isn't necessary. Race is a social construct and classification happens in all

[1] Kehl was the the Brazilian eugenicist movement leader.

interactions in daily life. It could be the family name in the United States, where the one drop of blood rule has been the criterion for assigning people into the "black" group; or the phenotype, the most important criterion of racial classification in Brazil. Not randomly, race variable in Brazilian census is called "cor ou raça" (color or race), highlighting the importance of the skin tone, a phenotypic attribute, for racial classification. About the Brazilian case, Osório (2003) states:

> "Society doesn't need to know how black is a person or his/her ancestors. It suffices to know if his/her physical appearance in his/her relational context turns him/her a possible subject to direct or structural discrimination. One has never heard about a private condominium doorman who has asked for a technical report or an electronic microscope to decide whether to require that a visitor of dark skin tone enter through the service/back door." (OSORIO, 2003, p. 08. Freely translated by the author).[2]

In this context, where phenotype is more important than ancestry and where miscegenation prevails, racial boundaries can become quite fluid and racial classification, uncertain. According to Rangel (2008, p. 18): "[i]n general, the probability that two parents of same skin color having kids that are either darker or lighter than themselves ranges from 50% to 67.5% (except for the ones that are either 100% black or 100% white, of course)."

Barth (1998) argues that most of the stratified classification systems allow mobility to a certain degree, based on the scale that define its hierarchy. Ethnic groups may not be open to this penetration because criteria are more restrictive. For instance, based on the individual's phenotype. However, the author admits and exemplifies cases in which ethnic boundaries are not clear as a consequence of within group variations. Ethnic classification – understood as self-ascription and ascription of others into certain ethnic categories – happens as a consequence of interaction rather than contemplation of "objective" classification criteria. It takes into account not only origin but also current identity (p. 29).

In addition, Telles (2004) argues that Brazil, as a continent-sized country, has blurry racial boundaries that vary according to the region. The North of the country is inhabited by a population of darker skin tone, while the Southeast and mostly the South have a larger share of population with European ancestry and lighter skin tone. Thus, the same person between black and white stereotypes could fall in different classes depending on the region (s)he finds (her)himself (*e.g.* being classified as black or *pardo*[3] in the South, but as white in the North

[2] Black people in private and mostly white-inhabited condominiums are usually mistaken for maids or other domestic employees in Brazil. Domestic employees – no matter white or black – have always used the back doors and "service elevators", because they were obligated to, or because it became a habit or a desirable behaviour – even after discrimination in the use of elevators became illegal in Brazil. – Thus, black people are usually associated with maids or other domestic employees by Brazilian doormen, who frequently require them to use the "service doors". In addition, this is not well or naturally received by many black people nowadays, who gradually become more and more empowered, and more present in middle and upper Brazilian classes.

[3] A person of mixed black and white race background.

or Northeast): race is not absolute, it is relational. Blurry boundaries entail an identification problem since there will be a set of individuals that might be racially classified in different ways according to context. This is a limitation of this research.

No matter the criteria, classification – racial classification included – is a cognitive tool to help one to interpret the world, make sense (a certain sense) out of social reality. Stereotypes helps us judge, make decisions and act more quickly. For instance, Loury (2002) distinguishes two types of discrimination: statistical discrimination is the association of a group's characteristics – or beliefs or narratives about that group – to an individual belonging to that group in the absence of specific information about the individual; the author contrasts this type with a second: the taste for discrimination, consisting on an aversion or distaste about a certain group, prejudice.

Those narratives about a certain group are powerful as they are self-reinforcing, capable of producing *truths*. Beliefs or narratives about a certain group are not built by an isolated individual, but rather by society at a larger scale. If there are not objective natural categories or *kinds of people*, there are socially built classification systems. Hacking (2007) proposes two main mechanisms through which he explains classification and creation of kinds of people. The first, "making people up", is about how sciences are able to create and spread categories of people; the second is the "looping effect" and concerns the way in which classification interacts with classified people. For instance, a certain people subject to the Persian Empire was defined as a people of craftsmen by the administration. Each year, this people would pay tribute to the king with their best crafts, trying to outperform their last year's presents to the court. Gradually, this people would really become the craftsman of the empire, even though they were not necessarily the most skillful craftsmen at the beginning. Similarly, stereotypes of American slaves were not only a belief for their masters, but also experienced by the enslaved themselves. Hence, the black power movement efforts targeted both to foster a new self-conception of blacks and to change the conception of the others about the blacks, specially the powerful others.

Hacking (2007) highlights five elements that play an important role in this framework: (i) the classification system; (ii) the individuals, that are (self-)assigned into the categories; (iii) institutions: well established organisations that frame or hold up the classification system in place; (iv) knowledge about each category, for instance, main characteristics of a certain people or racial group – which may easily fall into stereotypes –, and (v) experts, such as officers of the administration and scientists.

Under this framework, one can easily notice the central role of the state as a powerful institution, and that is not a recent conquer of the modern States. Hacking (2007) states about the Persian Empire:

> "(..) We can well imagine that Darius's captains chose to categorise his subjects for the convenience of administration. The subjects were not classified in exactly this way before they were conquered. Geography, language, allegiances, previous social

> cohesion, bodily structure, and skin colour would all have been grounds for forming classifications, and in some cases, those kinds of people would not have existed, *as a kind of people*, until they had been so classified, organised and taxed." (HACKING, 2007, p.288).

In modern states, categories can be a tool for managing aid to vulnerable groups, such as aid to the poor, or policies focused on racial or gender minorities. Departments concerned with demography are usually part of this process. About the United States, Nobles (2000) asserts:

> "The Census bureau has escaped inquiry both as a state institution that determines the benefits and penalties of racial membership through the data it collects and as a place where racial categories themselves are constructed. That perception that census agencies and census categories are at some remove from politics ensures that a deeper theoretical appreciation of how the census supports racial discourse and how census racial data serve public policy is blunted." (NOBLES, 2000, p. 17).

The author argues that public policy, an important driver of citizenship, is formulated based on public data and justify data collection on race (NOBLES, 2000). In Brazil, the *Instituto Brasileiro de Geografia e Estatística* – IBGE has been responsible for the demographic census since 1967[4], but demographic censuses were held longer before. In the occasion of the first census in 1872 – even before the slavery was abolished in 1888 – there were four categories of race: *preto* (black), *pardo* ("brown" or mixed-race), *branco* (white) and *caboclo* (native Brazilian or native Brazilian descendent). According to Osório (2003), census applied categories that were commonly used by people in daily life, although more complex subcategories or composed descriptions were just as frequent, such as *preta mina*, the noun designating the color or race (*preta*) associated with an adjective indicating the person's origin in Africa (*mina*). In the second census, *pardo* was replaced by *mestiço*. After that, census ceased the collection of racial data until 1940. The category *amarelo* ("yellow") was established in 1960[5] in order to comprise the incoming Asian migrants, but there was no specific category for native Brazilians or indigenous people (*indígena*) until 1991, the same year the variable was renamed, shifting from *cor* (color) – the variable name since 1940 – to *cor e raça* (color and race). Thus, IBGE's five current categories are in place since 1991: *preto*, *pardo*, *amarelo*, *branco* and *indígena*.

For long, miscegenation and the absence of state-sponsored segregation in Brazil – at least not legally such as in the South of the United States – were the basis of Freyre's ideology of racial democracy. This has disallowed positive policies deemed unnecessary for most of the Brazilian history since slavery was abolished. Scholars and activists at the time considered

---

[4] CPDOC-FGV. Retrieved from http://www.fgv.br/cpdoc/acervo/dicionarios/verbete-tematico/fundacao-instituto-brasileiro-de-geografia-e-estatistica-ibge in Dec. 16, 2019.

[5] IBGE. *Séries históricas e estatísticas*. Retrieved from https://seriesestatisticas.ibge.gov.br/series.aspx?vcodigo=POP106 in March 08, 2020.

the census a destroyer of racial democracy for collecting data on race, rather than a tool for better policies targeting the guarantee of basic rights. IBGE cross-tabulating of color/race and socioeconomic variables played an important role in the shift of racial narrative in Brazil from a racial democracy towards a racially stratified society, which has been the mainstream in academic debate for the last few decades (NOBLES, 2000).

3 SEGREGATION

3.1 Definitions and streams

Segregation has been defined in multiple ways by the literature. However, one can identify a common ground between all those definitions: the word has been used to identify separation between social groups and, sometimes, the definition also evokes social inequality. Torres (2004) defines segregation in a generic way as the degree to which a certain social or ethnic group concentrates itself in a certain area. Reardon (2006) defines segregation more precisely as the extent to which individuals pertaining to different groups live in different sub-units within a region (REARDON, 2006, p. 134). However, Torres (2004) highlights the usually imprecise use of the concept in Brazilian literature, as a synonym for inequality, social disparities, discrimination or other related phenomena. It is not a surprise that many scholars have been using the term as a synonym for inequality, since segregation and inequality are two phenomena that manifest themselves in a conterminous and endogenous way. However, this research adopts a more precise concept, meaning separation of social groups in space, noticeably in urban space.

Three streams or frameworks approaching segregation under these terms can be identified. The first is concerned about identifying the structural roots of segregation in Western societies. The second targets the description of causes and consequences, and sometimes the measurement of social impacts of segregation. Finally, the third stream is composed by scholars who have been trying to measure the phenomenon of interest.

The first stream adopts a Marxist approach to analyze segregation and, more generally, the urban phenomenon. Authors aligned to this stream defend that one must take into account the historical and material conditions in order to fully understand the urban phenomenon (CASTELLS, 1972). Thus, the city is seen as the arena where tension between capital – that profit from urban land increasing prices – and the working class and the poor, expelled to areas with less urban services and infrastructure (KOWARICK, 1979; HARVEY, 2012).

Definition of segregation under this perspective emphasizes hierarchy among territories inside the city. According to Castells (1972), segregation consists on the homogeneity within territorial units in contrast to heterogeneity between them, related to the idea of an existing hierarchy between those units (CASTELLS, 1977[1972], p. 16 *apud* FRANÇA, 2017). For instance, richer/poorer, more/less educated, more/less "sketchy" neighborhoods.

The second stream is concerned about more immediate causes and consequences of segregation. It is also the most affluent stream in Brazil. Main causes of segregation pinpointed by the literature are developers' business strategy, such as speculation and microeconomics in real state markets (TORRES et al, 2002; MARICATO, 1996); technology, such as the urban transportation infrastructure (RAMOS, 2014)[1]; state action (or inaction) and public policies (or lack of public policies) (ROLNIK, 1997; CALDEIRA, 2000); and neighborhood tipping – the willingness to pay more in order to live closer to your alike, or not (ABRAMO, 2007; SCHELLING, 1969). For instance, Zhang (2011) formalizes this game theory approach with Schelling's multiple-neighborhood tipping model showing the tendency towards complete segregation, even when every individual's preference is to live in a mixed-race environment: segregation at the macro-level does not reflect individual preferences in multiple-neighborhood settings.

Caldeira (2000) and Rolnik (1997) argue that the state has been (and still is) a protagonist of the urban phenomenon in São Paulo, noticeably of segregation. Double-standards have been adopted for zoning rules in the inner city – strongly regulated – and in the fringes of town – left unregulated (*e.g.* to be occupied by dwellers without any land ownership paper, or freely used by speculators). Caldeira also highlights the importance of housing policy in the second half of the XX$^{\text{th}}$ century, mostly financing housing for middle-class and lower-middle-class families. Hence, São Paulo inner city, with better provision of services, became too expensive for the poor and the working class. More recently, the increasing fear of violence and willingness for social separation among the middle- and upper-class led to the multiplication of private condominiums or *fortified enclaves*. Thus, segregation structure in São Paulo has been marked by three main mechanisms that can be summarized as follows (CALDEIRA, 2000):

- Concentration (1890-1940): high concentration of residents in the central urban area. Rich and poor lived geographically close to each other, although in different conditions and social circles.
- Peripheralization (1940-1980): as the state increased regulation with a double-standard enforcement – central urban area strongly controlled while suburban and rural areas were left unregulated or unenforced – poor and working class were pushed towards periphery because of the high prices of land and rent.
- Fortified Enclaves (1980-...): the fear of violence leads middle and upper-classes to enclosure themselves in private condominiums well equipped with security infrastructure. Through this strategy, they are able to live near *favelas* or *cortiços* without any social contact with those neighbors.

Main consequences highlighted by Brazilian literature can be associated with peripherilization: lower income, poorer infrastructure and services, and longer commutes (MARICATO, 1996; TORRES & MARQUES, 2001). However, peripheries are heterogeneous and some of

[1] Ramos (2014) explains that building or enlarging road systems creates the possibility for segregation through urban spread. When income-elasticity of demand for housing is higher than income-elasticity of marginal cost of transportation, white suburbs may occur (*e.g.* white suburbs in the United States). This would be what Torres (2004) called self-segregation. Ramos (2014) argues this is not the case of most of Latin-American cities at least not at the same moment (mid-XX$^{\text{th}}$ century).

its parts, more vulnerable than others. Torres & Marques (2001) identify the hyper-peripheries: peripheral areas of the metropolis that accumulate vulnerabilities or risk factors other than the previously mentioned: environmental risks and ill-provided housing conditions. Peripheral areas are also more vulnerable to violence (MARICATO, 1996; CALDEIRA, 2000). Maricato summarizes:

> "Segregation is one of the most important facets of social exclusion (...) Less access to urban services and infrastructure (poor public transportation and sewage collection, no draining service, low access to basic supplies and health services, education and childcare, higher exposition to floods and slides) sums up with less jobs (...) and professionalization opportunities, more exposition to violence (marginal or police violence), racial discrimination, discrimination against women and children, less access to justice and leisure. (...)." (MARICATO, 1996).[2]

As segregation is at the root of unequal access to material and symbolic resources of the city, it is a mechanism of inequalities reinforcement (GRAFMEYER, 1994; WILSON, 1987 *apud* FRANÇA, 2017).

In contrast with the above mentioned authors, who adopt a normative perspective that approaches segregation as a problem or a cause of problems – and *solely* problems, Sabatini (2003) proposes an analytical perspective that doesn't consider segregation a problem, but a phenomenon. He argues that segregation is not necessarily bad, since it can also entail beneficial consequences, such as (i) facilitating preservation of urban subcultures, or (ii) social mobilisation, because people with similar life conditions are closer to each other. Hence, social mobilization is facilitated in more segregated urban settings. These might be true for social-economic or class segregation, but what about ethnic or racial segregation in the Brazilian metropolitan context?

One can argue that in the context of the last demographic census (2010), the most marked separation in Brazil is between blacks and whites, both groups representing large portions of the overall population, unlike most of the American cities. In 2010, blacks represented approximately 40% of São Paulo Metropolitan Area (SPMA) population while whites, 60% (Table 2). In addition, both racial groups have been living in Brazil for centuries and share a fair amount of cultural common ground, specially among the middle- and lower-classes, that are also more racially mixed than the upper-class (Table 4 in Chapter 4).

In this context, segregation has not necessarily preserved but has firstly catalyzed the emergence of peripheral subcultures and forms of artistic expression that have been ways of resilience and identity building for peripheral youth and of political mobilization – aligned to Sabatini's argument (ii) but not fully contemplated by argument (i) –, such as the worldwide known hip-hop since the 1970s and Brazilian funk since the 1990s in Southeastern Metropolises

[2] Freely translated by the author.

of Rio and São Paulo (WELLER, 2004; SCANDIUCCI, 2006, FACINA, 2009; LOPES; FACINA, 2012; CALDEIRA, 2014). Those are related to segregation and have surely enriched the Brazilian urban cultural scene, which does not justify the conditions to which the marginalized population has been subject in the last half century. These conditions largely described by the above mentioned literature are not solely related to segregation, but all the other conterminous factors that are hard to separate from segregation itself – violence, state neglect, lack of services and infrastructure, lack of education and job opportunities and stigmatization.

Following the tone set by Sabatini (2003), the main hypothesis to be tested by this research is an attempt to characterize the specificities of racial segregation in Brazil in a more systematic way, taking into account a description of how race interacts with income. Unlike developed countries where racial minorities and low-income population segregated in poorly served ghettos represent a small proportion of the overall population, low-income households are the largest share of Brazilian homes. In this context, the upper-middle class and the rich are the exception and the ones with power to choose to segregate themselves in private fortresses with the money to replace public with privately supplied services. They also happen to be mostly white, while low-income population is racially mixed. This can be perceived in urban space, through landscape, through the distribution and use of public services and through segregation indices, the main tool to systematically test this hypothesis (discussed by the following third stream).

Finally, the third stream – the measurement stream – is what França (2017) called a "more operational definition", consisting on the degree to which different groups live apart from each other (MASSEY & DENTON, 1988; GRAFMEYER, 1994; TORRES, 2004; MARQUES & TORRES, 2005).

Marcuse (2005) highlights the importance of the processes that generate certain patterns of uneven distribution of subgroups in the space and define twenty precise concepts related to those processes. Indeed, the most used concepts and, thus, segregation indices are concerned about the picture but are unable to take into account the processes that engendered these spatial distributions. The quantification or measurement approach is unable to capture by itself the segregation structural forms, its causes or consequences:

> "Although this kind of work, which has been taken up enthusiastically by Duncan (Duncan and Duncan, 1955) and, more recently, by the University of Wisconsin group, (Schnore, 1965) expresses the articulation between social differentiation and the configurations of space, it cannot explain the production of these forms. To do so, it would have to be related to the rest of the elements structuring the form and rhythms of an urban area." (CASTELLS, 1972, pp. 122-123) .

Even though, measurement must be understood as a small and foundational stone of the path to diagnose and fully understand segregation. Under the mainstream rationale of modern

Western States, a phenomenon that cannot be measured cannot be adequately addressed. Hence, improving measurement will be an important component of this research. A more detailed description of this third stream is presented in the next section.

3.2 Segregation Measurement

Feitosa (2005) presents a comprehensive literature review on segregation measurement methods. She identifies three main periods of development of segregation measures: dichotomous measures (1940s-1950s); dichotomous measures applied to multiple groups (since the 1970s); and spatial measures (since the 1980s).

The first period is characterized by dichotomous measures mostly used to compute black-white segregation in the context of pre-Civil Rights United States, where there were segregationist policies against blacks in place. The Dissimilarity Index (DUNCAN; DUNCAN, 1955), the Exposition and Isolation indices (BELL, 1954; LIEBERSON, 1969; MASSEY & DENTON, 1988) are the most notorious examples.[3]

The aspatial Dissimilarity Index can be interpreted as the proportion of the population that would have to move to other territorial units in order to make group proportion in each territorial unit (*e.g.* a district) equals to the overall group proportion in the region of study (*e.g.* the entire city). It varies from 0 to 1, 0 being complete evenness in the geographic distribution of both groups and 1 complete segregation (this interpretation is not valid for the spatial form of this index). The Dissimilarity Index originally allow only two groups in the model. Let them be groups m and n. Then, following Reardon and O'Sullivan's (2004) notation, the Dissimilarity Index (D) can be defined as:

$$D = \frac{1}{2}\sum_{q=1}^{Q} \mid \frac{m_q}{T_m} - \frac{n_q}{T_n} \mid$$

Where q is each territorial unit of the area of study (R), m_q stands for the number of individuals from group m living in q and T_m is the total population of group m in R (the same for group n).

The Exposition and Isolation indices (P^*) are variants of the same measure. Exposition is a measure of the extent to which an individual of group m is likely to encounter an individual of another group (n) (*exposure*) or a member of her/his own group m (*isolation*) in her/his local environment p. In the aspatial version of these indices, local environment p is usually the census tract or the more disaggregated level in which data is available. The more isolated or exposed

[3] Those are here defined in detail because they were the most used indices and will be the base for further discussion on their classification and limitations in this section. Indices to be mentioned later in this section will not be mathematically defined before the methodological section, where they may be defined as needed.

m is, the higher the values of P^*. The aspatial version of these indices is defined in the chosen notation (REARDON; O'SULLIVAN, 2004) as follows:[4]

- ***exposure of*** m ***to*** n***:***

$$mP_n^* = \int_{q \in R} \frac{m_q}{T_m} \pi_{qn}$$

- ***isolation of*** m***:***

$$mP_m^* = \int_{q \in R} \frac{m_q}{T_m} \pi_{qm}$$

Where m_q is the number of individuals of group m living in q; q in the aspatial approach is usually the census tract; π_{qm} is the proportion of group m in q, given by $\pi_{qm} = \frac{\tau_{qm}}{\tau_q}$, where τ_q is the population density in point q, given by a fraction of inhabitants per area unit and τ_{qm} is the group m population density in point q.

A second wave started during the 1970s, when scholars were interested in more complex social taxonomies – with more than two groups –, but kept using the already known dichotomous measures. Literature started to criticize this approach since the multiplication of the groups complexified analysis and interpretation under the old measures. At this moment, new indices that admitted multiple groups started to be proposed, such as the Entropy Index (THEIL; FINIZZA, 1971), the Residential Segregation Index (JARGOWSKY, 1996) and the generalized versions of the Dissimilarity (MORGAN, 1975) and Exposition indices (JAMES, 1986).

At that moment, one limitation held: spatial disposition of geographic units matters and, hence, should be taken into account in the segregation measurement. For instance, the **checkerboard** and the **grid problem**s affect aspatial Dissimilarity Index (WHITE, 1983). Massey & Denton (1988) explain the checkerboard problem as follows:

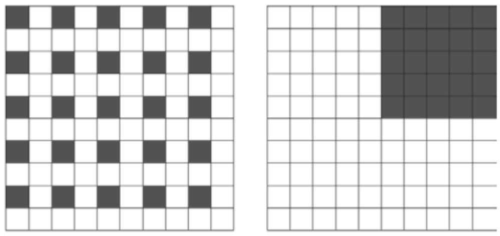
> "(...) suppose we have two urban areas with the same number of minority members, who comprise the same proportion of the total population. In each place, no minority member shares a common residential area with a majority member, all minority areas are located the same average distance from the central business district, and all areas are of the same geographic size. In this case, both urban areas would display identical measures of evenness, exposure, concentration, and centralization. However, if all minority areas in one of the urban areas were contiguous to one

[4] Authors use integral symbol to denote summation.

> another, but in the other area they were separated from one another, then we would probably consider the former urban area to be more segregated, since all minority members live within one single homogeneous ghetto, compared to the latter area, where they reside in minority neighborhoods that are scattered throughout the urban area." (MASSEY & DENTON, 1988, pp. 293-294).

Figure 1 exemplifies two situations in which the above mentioned segregation index would have the same value, although one would expect a higher level of segregation to be computed for the situation on the right.

Figure 1 – The checkerboard problem



Source: Feitosa (2005, p. 46), adapted from Wong (2003, p. 67).

The grid problem, on the other hand, concerns the size of the territorial units to which the Dissimilarity Index is sensible. Larger units lead to lower values – because there is more space for diversity to be captured if the geographical unit is larger – while smaller units, or data in higher resolution causes higher levels of Dissimilarity Index. A way to avoid this issue is to privilege relational interpretation in analysis (*e.g.* comparing the same area, with the same territorial divisions across time), in detriment of absolute interpretations.

These weaknesses have led to the third and last period identified by Feitosa (2005). This period has begun in the 1980s and is characterized by the development of spatial measures – not randomly at the same time improvements in both computers' hardware and software spread access to Geographic Information Systems (GIS) in academic community. Spatial Dissimilarity (WONG, 1993), Isolation and Exposure indices (MORGAN, 1983), as well as the spatial Proximity Index (WHITE, 1983; REARDON; O'SULLIVAN, 2004) were proposed. Auto-correlation measures, such as the Moran's I (FRANK, 2003) are also spatial measures that can be applied to segregation, although they are a more general tool of spatial analysis. The LISA Map, for instance, is the spatial tool most frequently applied to the analysis of segregation in Brazil, a discussion to be deepened in the next section.

Unlike aspatial indices, the spatial measurement methods take into account how the data is distributed throughout the space. The neighborhood matrix is the most used tool in order to operationalize this sort of analysis. The matrix allows the researcher to take into account the relationship between geographic units. For instance, it can be a simple dummy matrix with 1 representing contiguity between units and zero, non-contiguity; or a slightly more complex matrix with a spatially defined kernel density function of the euclidean distance between geographic units. How this relationship between units is established can be called the *proximity function*. In practical terms, by taking into account data in neighboring units, spatial indices do not rely as heavily on geographic units' boundaries as their aspatial counterparts.

Another strength of spatial indices is the flexibility given by the proximity function, to be defined by the researcher. This function might be thought according to context and data structure and availability. This function also defines the scale in which segregation is to be measured, which allows for adjustments in order to measure macro- or micro-segregation, but also leads to an issue analogous to the grid problem: functions that allow longer distance relationships between sub-units result in lower segregation levels, while short-range functions result in higher values.

Most of the segregation literature encompasses distance-decay functions of the euclidean distance between geographic units. More recent works, however, have been exploring other metrics that might be more realistic in terms of the opportunity cost of being present in a certain area of the city, such as the travel time. For instance, Farber et al. (2015) develop a measure of exposure that decomposes the social interaction potential (SIP) based on the concept of *joint accessibility*.

3.3 How has segregation been measured in Brazil

Despite the various ways through which segregation has been tested worldwide, there are only a few studies concerned with segregation measurement in Brazil, and even fewer that measure racial segregation. Efforts to measure segregation in Brazil started in the early 1990s with Telles (1992). He applies the aspatial dissimilarity index to the 1980 Brazilian census data. He analyses most of the Brazilian metropolitan areas. His results regarding São Paulo Metropolitan Area show moderate levels of residential segregation (indices below 0.6) – white-black segregation being the highest and brown-black segregation being the lowest (compared to brown-white and white-black). This applies to five out of the seven metropolitan areas in São Paulo state.

Further work starting at the beginning of the 2000s shifted segregation measurement in Brazil towards socio-economic groups, mostly based on income and education level (RAMOS, 2004; TORRES, 2004; FEITOSA, 2005; 2010; GARCÍA-LÓPEZ; MORENO-MONROY, 2016), or based on the Erikson-Goldthorpe-Portocarero (EGP) classes, an occupational-based social classification (MARQUES; SCALON; OLIVEIRA, 2008; MARQUES, 2014; FRANÇA, 2017). Torres (2004) applies aspatial dissimilarity index only. Moran's I (global and local) is the most used spatial technique among those works. Feitosa (2005; 2010) and Ramos (2014) are the only applying variants of the spatial dissimilarity and isolation indices. Although Reardon and O'Sullivan's review highlights that information theory index is the most reliable according to their eight criteria (REARDON; O'SULLIVAN, 2004), Barros & Feitosa (2018) and Gacia-Lopez & Moreno-Monroy (2016) are the only works using this index in order to calculate residential segregation in Brazil. The overview of the empirical work measuring segregation in Brazil is presented in Table 1.

Table 1 – Segregation measurement in Brazil

Ref.	Segregation criterium	Groups	Location	Main Data Sources	Indices
Telles (1992)	Racial and Class	Brown, Black and White.	35 (out of a total of 40) Brazilian Metropolitan Areas (those with population size above 200 000 inhabitants).	1980 Census (IBGE)	Aspatial Dissimilarity Index
Telles (2004) – Chap. 08	Racial	Whites and non-whites, White-Black and White-Brown.	10 largest metropolitan areas of Brazil	1980 Census (IBGE)	Aspatial Dissimilarity Index
Ramos (2004)	Socio-economic	Continuous variables: low education and employment density.	SPMA (SP)	Origin-Destionation survey (SP state Metro); Census (IBGE); Satellite images (Landsat5-TM, Landsat7-TM e SPOT-HRV made available by INPE); Exclusion/Inclusion Map (Sposati, 1996, 2000).	LISA Maps and LISA Scatterplot (local Moran's I and local standardized G*). Doesn't require categorical variables.
Torres (2004)	Socio-economic	Education (0-3 years of study vs. 11 years of study or more; 0-3 vs. 15 years of study or more) and income groups (0-3 m.w. vs. 10 m.w. or more ; 0-3 m.w. vs. 15 m.w. or more; 0-3 m.w. vs. 20 m.w. or more), based on the characteristics of the head of the household.	SPMA (SP)	1991 and 2000 Census (IBGE)	Aspatial Dissimilarity Index
Feitosa (2005)	Socio-economic	Head of household income and school levels	São José dos Campos (SP)	1991 and 2000 Census (IBGE)	Spatial dissimilarity, isolation and exposure indices – adpated from Reardon and O'Sullivan, (2004) –, and residential segregation index (ISR). Gaussian kernel distance function (bandwidth 400 and 2000 meters).

Carvalho & Barreto (2007)	Racial	Brown, Black and White.	Salvador	2000 Census (IBGE)	Aspatial Dissimilarity Index
Marques, Scalon & Oliveira (2008)	Class	Occupational categories, the EGP classes (Erikson, Goldthorpe and Portocarrero, 1979). A socioeconomic index – the ISEI index, a continuous variable composed by occupational, income and education criterea – completes the analysis.	São Paulo & Rio de Janeiro Metropolitan Areas	2000 Census (IBGE)	Local and Global Moran's I. EGP classes applied to the Global Moran's I, and ISEI applied to the local Moran's I.
Feitosa (2010)	Socio-economic	Income categories (given by 2000 Brazilian Census)	São José dos Campos (SP)	2000 Census (IBGE), 'Survey for Urban Planning Instrumentation and Evaluation of the Housing Deficit in São José dos Campos', and Property advertisements collected from local newspapers printed from March 2002 to February 2003.	Spatial dissimilarity, isolation and exposure indices.
Marques (2014)	Class	Occupational categories, the EGP classes	SPMA (SP)	1991, 2000 and 2010 Census	Global Moran's I and Aspatial Dissimilarity Index
Ramos (2014) – Chap. 02	Socio-economic	School (4 groups) and income levels (7 groups)	SPMA (SP) – focus on the urbanized area	2000 Census (IBGE)	Global and local Spatial Dissimilarity Index and Spatial Isolation Index, with gaussian kernel distance function (radius: 1000 meters)
Garcia-Lopez & Moreno-Monroy (2016)	Socio-economic	9 groups based on head of household income split in bins (measured in m.w.)	35 Brazilian urban areas	2000 and 2010 Census (IBGE)	Rank-order Information Theory Index
França (2017)	Racial and Class	6 racial-class groups, combining EGP classes (3 groups: lower-, middle- and upper-class) with 2 racial groups – whites and blacks (blacks and browns).	SPMA (SP)	2000 and 2010 Census (IBGE)	Aspatial Dissimilarity Index, Location Quotient, LISA Maps

Fernandes (2017)	Racial	Black (brown and black) and White	São Paulo (SP) – municipality only	2010 Census and School Census	Aspatial Dissimilarity Index
Barros & Feitosa (2018)	Racial	4- and 2-group racial systems	London and São Paulo Metropolitan Areas	2011 UK and 2010 Brazilian Censuses	Dissimilarity and Information Theory Index

Torres (2004) pinpoints that most of Brazilian scholars who try to measure this phenomenon focus on the socioeconomic segregation, rather than racial. This might be a consequence of the belief that racial segregation in Brazil exists as long as it is conterminous to the socioeconomic condition (*e.g.* poverty is positively correlated with darker skin tone, while wealth is correlated with whiteness), such as witnessed by Telles in a conversation with a Brazilian sociologist (TELLES, 2004). However, it has been changing in recent years: 6 out of the 14 works listed in Table 1 are concerned with racial segregation, three of them written in the last three years.

Furthermore, Telles as well as França challenge the above mentioned assumption cross-analyzing race and socio-economic condition. Both test racial segregation among members of the same income group (TELLES, 1992) or same class (FRANÇA, 2017). They find lower residential segregation levels among the very poor or the lower-class. Telles explains that economic constraints leave no room for neighborhood choice based on skin color of neighbors. On the other hand, França (2017) applies a mixed-method approach. Using D, location quotient and some other spatial analysis (LISA maps) he shows that middle-class blacks and whites are more distant than the same racial groups in other classes – even more distant than in the upper-class –, meaning there are mechanisms of social separation that operate other than economic constraints. His qualitative analysis applying a social-network approach corroborates with these results.

Other efforts to measure racial segregation in Brazilian cities were made by Carvalho & Barreto (2007) , Fernandes (2017) and Barros and Feitosa (2018). Barros and Feitosa (2018) – the only work applying spatial racial residential segregation indices – find low global indices to SPMA. However, local analysis reveal the nuanced distribution throughout the metropolises. SPMA central area in which CBD is located is marked by high dissimilarity as well as low diversity (local Information Theory Index) due to the absence of non-whites. Carvalho & Barreto (2007) measures residential segregation in Salvador and Fernandes (2017) analyzes school segregation in the municipality São Paulo. Carvalho & Barreto (2007), França (2017) and Telles (1992) find moderate levels of residential segregation when analyzing aspatial dissimilarity index in Brazil using the US standards – indices below 0.6 are considered moderate. This is a comparison problem, since urban formation and thus segregation structures are very different in both countries (or at least when comparing Brazilian and Northern US cities). However, this might also be a measurement issue, related to the previously discussed limitations of D. Then, why does Fernandes (2017) find high levels of school segregation with high impact on wages? School segregation is a better proxy for social proximity when compared with residential segregation, since school is necessarily a place of social contact, while living in the same block or neighborhood is not a guarantee of any form of sociability in a city of walls, such as São Paulo. An illustrative example might be helpful to understand the phenomenon: a famous landscape in São Paulo presented in Figure 2 – Paraisópolis[5], a favela surrounded by a high-income neighborhood (Morumbi).

Figure 2 – Paraisópolis

Source: Tuca Vieira (2002).

This is not the single case of favelas neighboring luxurious condominiums in São Paulo. There are many locations where people living a few hundred meters from each other live very different and separated lives, as shown in detail in Chapter 4.

In addition, 45.86% of households in São Paulo Metropolitan Area have declared to earn a *per capita* income lower than 1 minimum wage in 2010, while only 3.24% had a monthly *per capita* income of 10 minimum wages or more.[6] Despite the high percentage of households below 1 minimum wage of *per capita* income, it is important to highlight that the minimum wage at the moment 2010 census was run worth R$ 510, according to federal legislation approved in June, 2010,[7] the equivalent of US$ 283.24.[8] This amount is far below the needed by those families to afford frequenting the same private spaces as those in households classified here as high-income (10 *per capita* minimum wages or more). Even the middle-class has changed the way of living and moving around the urban space. For instance, Frúgoli Jr. (1995, p. 76) highlights the shift of the middle-class preferences towards private spaces and private means of transportation in São Paulo, in order to avoid public spaces and its social diversity.

Thus, this geographic discontinuity of *per capita* household income might be a good indicator of where social distance can be spotted, despite the short physical distance. In these

5 Image retrieved from https://www.tucavieira.com.br/A-foto-da-favela-de-Paraisopolis. Access in Dec. 11, 2019.
6 Data from IBGE 2010 demographic census. Calculations by the author.
7 Retrieved from http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Lei/L12255.htm. Access in Dec. 12, 2019.
8 Source of the conversion rate: the Brazilian Central Bank. Conversion date: July 1st, 2010.

cases, aspatial Dissimilarity Index might not be able to capture social distance through residential data, depending on the scale in which data for race is aggregated and how the geographic units were defined.

Even though, as middle- and upper-class children usually attend private schools in Brazil while children living in favelas, usually coming from low-income families, attend free public schools, social separation between groups is well marked at the school level. Private schools are mostly white, while public ones are racially mixed – as shown in section 4.3 –, the reason why high levels of school racial segregation can be identified even when measured with aspatial Dissimilarity Index. For instance, Fernandes (2017) found low levels of school racial segregation when considering only public schools in the calculations, but high levels when taking private schools into account in the municipality of São Paulo.

Dissimilarity Index being flawed in translating social proximity or distance when applied to residential data is a problem highlighted by the literature (FRANÇA, 2017; GRIGORYEVA & RUEF, 2015). For instance, Telles (1992) highlights the hypothesis that domestic workers living in their white middle-class employers' homes may attenuate the results in the Southern Brazilian regions. Analogously, in the South of the United States, middle- and upper-class whites lived in houses facing the main streets, while black dwellers were concentrated in small inner alleys in the same blocks as whites. However, this physical proximity didn't mean social proximity. Through Dissimilarity Index, social scientists were unable to capture this backyard segregation pattern, the reason why Southern cities had been documented as having lower levels of residential segregation compared to the Northern American cities. Although, the South is historically known for its higher levels of discrimination and for its segregationist policies. With a new index they propose, specifically fitted to the format in which historical data has been structured, Grigoryeva and Ruef (2015) were able to measure segregation at a smaller scale and show the higher levels of racial segregation in Southern cities.

In spite of the intention of reflecting social separation between groups through physical distance between them (a proxy variable), this correlation is not always accurate. The reason why São Paulo is still considered a moderately segregated city in terms of race when measurement is made through Dissimilarity Index might be similar to the one Grigoryeva and Ruef (2015) highlight. Different patterns of segregation require different indices. This and being the largest metropolitan area in Brazil are the reasons why SPMA is a good choice as laboratory for testing the adequacy of racial residential segregation measurement methods.

As Fernandes (2017), França (2017) and Barros & Feitosa (2018) are, to the best of my knowledge, the only three recent researches measuring racial segregation in São Paulo, the later being the only to apply spatial indices in order to measure racial segregation in a Brazilian city, one should explore the nuances of those spatial indices through the urban area.

4 CLOSE BUT FAR: A PICTURE OF SPMA URBAN CONTRASTS

An initial analysis of racial and social stratification in SPMA is necessary in order to understand the relationship between race and class in the chosen metropolitan context. Thus, a statistical description of income and education level by racial group was adopted. Two 2010 Brazilian census data bases of the entire SPMA were analyzed: Results of the Universe and Sample Results. The former contains basic data, such as race, gender, age and household income counts aggregated by census tract, while the later contains more detailed results, such as education level and monthly wages that are made available at the individual level. 2010 school census was also adopted in order to describe racial composition of schools in basic education (K-12). These analysis are presented in sections 4.2 and 4.3, respectively.

A second analysis consists on the investigation of which areas are potentially those in which physical proximity does not necessarily entail social proximity. A LISA Map of household income is here implemented using using 2010 demographic census with that purpose – and not racial composition of tracts, as pinpointed in section 3.3. Local bivariate Moran's I (ANSELIN, 1995) was computed to each census tract in GeoDa. Following its documentation, the index can be defined as:

$$I_{B,i} = cx_i \sum_j w_{ij} y_j$$

> x being the percentage of high-income households (10 *per capita* minimum wages or more) and y being the percentage of low-income households (less than 1 *per capita* minimum wage); w_{ij} is the element in the neighborhood matrix relating i and j census tracts.

Hence, areas with high-high clusters – tracts with a high percentage of low-income households close to tracts with a high percentage of high-income households – were qualitatively analyzed in more detail through Google Maps satellite and Google Street View images. Those results are presented in section 4.4.

4.1 Racial composition

SPMA is the largest Metropolitan Area in South America, with 19,683,975 inhabitants according to the last demographic census (2010). It is composed by 39 municipalities, comprising 29,943 census tracts (724 rural and 29,219 urban census tracts). Only 297 census tracts had no race or income data collected. Table 2 shows SPMA total population by racial group in 2010.

Table 2 – SPMA racial composition

Racial group	Population	Percentage
White (*Branco*)	11,574,507	58.86%
Brown (*Pardo*)	6,445,797	32.78%
Black (*Preto*)	1,266,344	6.44%
Yellow (*Amarelo*)	355,251	1.81%
Native Brazilian (*Indígena*)	21,049	0.11%
Not declared	1,367	0.01%

Source: Data from IBGE (2010). Elaboration by the author.

Total population in census tracts without racial data available sums 19,660 inhabitants. Dropping those and those who refrain from declaring their race, 2010 Brazilian census still has racial data for 99.89% of SPMA inhabitants.

4.2 Income & education

Since results of the survey applied to all households in Brazil does not disclose information at the individual level, an exploration of the sample survey results will be needed in order to show how race and socioeconomic conditions interact in SPMA. Table 3 shows the relative frequency of literacy, of individuals whose higher school level ever attended is high school, and relative frequency of individuals who have been at least to college within each group.

The percentage of blacks, browns and whites who goes or went to high school are similar (around 30%), although the percentage of whites who goes or went to college or attended higher studies is more than double than this percentage in both other above mentioned groups: 26.98% for the whites, against 8.69% and 11.53% for browns and blacks, respectively. Yellows are the ones with the lowest percentage of illiteracy (2.3%) and high school as being the higher school level attended (22.9%). This group also presents the higher relative frequency of members who go or who have been at least to college (42.5%). Native Brazilians are the group with the lowest levels of formal education, which might be correlated to their culture and lack of higher education policies focusing in their specificities.[1]

Regarding the economic status, the picture doesn't change much. The cross-tabulation of race and income of those who have income and are in the range of 25 to 65 years of age[2] enables the description of the racial composition of the income extremes – the 10% higher and the 10% lower incomes –, as well as the income composition of each racial group.

In terms of the racial groups' relative frequencies inside each income extreme, 84.7% of

1 Statistical summary of all these variables is presented in Table 17.
2 This age range was chosen because after 25, most of people have already entered the job market and have not yet retired.

Table 3 – Schooling by racial groups – SPMA residents with 25 years of age or older

Group	Literacy	High school	College
Black (*Preto*)	93.1%	30.48%	11.53%
Brown (*Pardo*)	93.23%	29.88%	8.69%
White (*Branco*)	96.73%	28.77%	26.98%
Yellow (*Amarelo*)	97.71%	22.87%	42.47%
Native Brazilian (*Indígena*)	93.57%	32.07%	13.71%
Not declared	57.01%	0%	0.23%
Total	95.36%	29.1%	20.4%

Source: Data from IBGE (2010). Elaboration by the author.

the individuals in the 10% richer are *brancos*; 7.94% are *pardos*; 5.20% are *amarelos*; 2.11% are *pretos* and 0.04% is *indígena*. Meanwhile, the percentage of those who earn the 10% lower incomes that are *brancos* and *pardos* is nearly the same: 45.90% and 43.71%, respectively. 9.13% of the 10% poorer are *pretos*; 1.04% is *amarelo* and 0.22% is *indígena*. Thus, while the high-income population is predominantly white – 89.9% of the 10% richer are *brancos* or *amarelos* –, poverty is racially mixed, with 46.94% of *brancos* or *amarelos* and 52.84% of *pretos* or *pardos*.

As *brancos* and *pardos* are the largest groups in SPMA, one must also take into account the relative frequencies of those income extremes in each racial group in order to better assess each group's socioeconomic status. *Brancos* and *amarelos* are the group with the higher relative frequency of individuals falling into the higher income stratum: 27% of *amarelos* have a monthly income higher than R$ 4,000.00 – the inferior limit to the 10% higher income stratum –, while 15% of the *brancos* are in the same situation. Simultaneously, they are the groups with the smaller portion of members falling into the 10% lower income stratum, having a maximum income of R$ 510, as shown in Table 4.

Table 4 – Frequency of income extremes in racial groups – SPMA residents between 25 and 65 years of age with income

	10% lower	10% higher	Pop. with income
Black (*Preto*)	12%	3%	33149
Brown (*Pardo*)	14%	2%	145701
White (*Branco*)	8%	15%	262555
Yellow (*Amarelo*)	6%	27%	8672
Native Brazilian (*Indígena*)	15%	3%	608
Not declared	0%	0%	1

Source: Data from IBGE (2010). Elaboration by the author.

Thus, in a broad way, the yellows or *amarelos*, Asian descendants, benefit from the highest socioeconomic level. Lesser (2000) states that success and prestige is usually associated with being white in Brazil, in a way that successful are usually classified as white, regardless of whether they have a clear European ancestry or not.

> "This is just the most recent formulation of the great Brazilian paradox – that policies constructed to re-make Brazil as "white" in fact created a multi-cultural society. Today the *nikkei*[3] and Syrian-Lebanese communities are broadly successful in the economic, political, military and artistic arenas. Both groups seem more part of the Brazilian nation than do, for example, Afro-Brazilians or Polish-Brazilians." (LESSER, 2000, p. 11).

Throughout the last century, people with Asian origin – Japanese, Jews, Syrian-Lebanese – were able to claim their whiteness within Brazilian society. This, in addition to the low percentage of native Brazilians in SPMA, makes this dichotomous white–black racial system seems to be adequate to describe social reality in the case of São Paulo, as highlighted by Fernandes (2017). Thus, for simplification purposes, most of the analysis will be made based on this two groups: black refering to *pretos* and *pardos*, and white, to *brancos* and *amarelos*.

4.3 Schools

As stated in section 3.3, schools are an important space of sociability. There are two main types of schools in Brazil: public and private. Public basic education (K-12) is universal and free of charge in Brazil. Thus, parents that choose to enroll their children in private schools, even though they need to pay for it, are an example of the Weak Axiom of Revealed Preference (SAMUELSON, 1938). Some hypothesis might explain this choice, such as the location, absence of cases of violence within the school environment, pedagogical approach, prestige of the school or the higher quality claimed by private schools. For instance, private schools all over the country compete for the best ranking in ENEM, the Brazilian national exam used by federal universities to select their future students, or the number of students that succeeded to enter a medical school (a six-year and the most competitive undergraduate course in Brazil). Those factors might also vary according to class (BALLION, 1986; NOGUEIRA, 1998 *apud* RESENDE et al., 2011). The important aspect of this context is that public schools largely receive students coming from low- and medium-income families creating a first noticeable separation of class and race between citizens since their early childhood.

In 2010, there were 438 schools without a single black (preto or pardo) student in SPMA, in contrast with 49 schools without a single white (branco or amarelo) student. While the later group of schools had only 53 enrollments of black students (pretos or pardos), 348 native Brazilian students and 1361 for which race was not declared, the first group of schools counted

[3] Descendants of Japanese.

with a sum of 15421 enrollments of whites, 313 native Brazilians and 13781 students with no race declared. Thus, there is a clear asymmetry of school racial composition and this might be explained by private schools.

Table 5 shows the relative frequencies of racial group of students in each school type: all, private and public schools. In the entire SPMA, 78% of enrollments with no race declared came from public schools (Table 6). However, private schools reported relatively more enrollments without race declaration: 35.8% of enrollments in public schools had no race declared *versus* 39.9% in private schools (Table 5).

Almost half of the students in private schools are white. Brown students are the second group more frequent in private schools. However, this Table has to be confronted with Table 6, since blacks, yellows and native Brazilians are smaller groups and are inevitably going to figure among the less frequent ones when relative frequency is calculated over the total students in each school type. Table 6, on the other hand, presents the relative frequency of enrollment by school type within each racial group. As a consequence, one can notice that the group with relatively more students enrolled in private schools are the yellows, Asian descendants, corroborating with what has been previously mentioned about their socioeconomic status, which seems to be, one more time, the most privileged of all groups in SPMA.

In summary, student body in private schools can be considered whiter and more homogeneous than public schools'. Since parents in Brazil get to choose the private school in which their children are going to enroll – which is not possible for parents with children in public schools, in which students are enrolled according to their address of residence or spots available for new enrollments –, this can be considered a phenomenon of *self*-segregation. This is further explored in more depth later in this chapter.

Table 5 – Relative frequencies of racial groups within each school type in SPMA

	Private	Public	All schools
Black (*Preto*)	1.53%	3.52%	3.13%
Brown (*Pardo*)	8.37%	24.32%	21.16%
White (*Branco*)	48.73%	35.72%	38.31%
Yellow (*Amarelo*)	1.33%	0.37%	0.56%
Native Brazilian (*Indígena*)	0.15%	0.26%	0.24%
Not declared	39.86%	35.81%	36.61%
Total	100%	100%	100%

Source: School Census – IPEA (2010). Elaborated by the author.

Table 6 – Frequencies of students enrolled in each school type

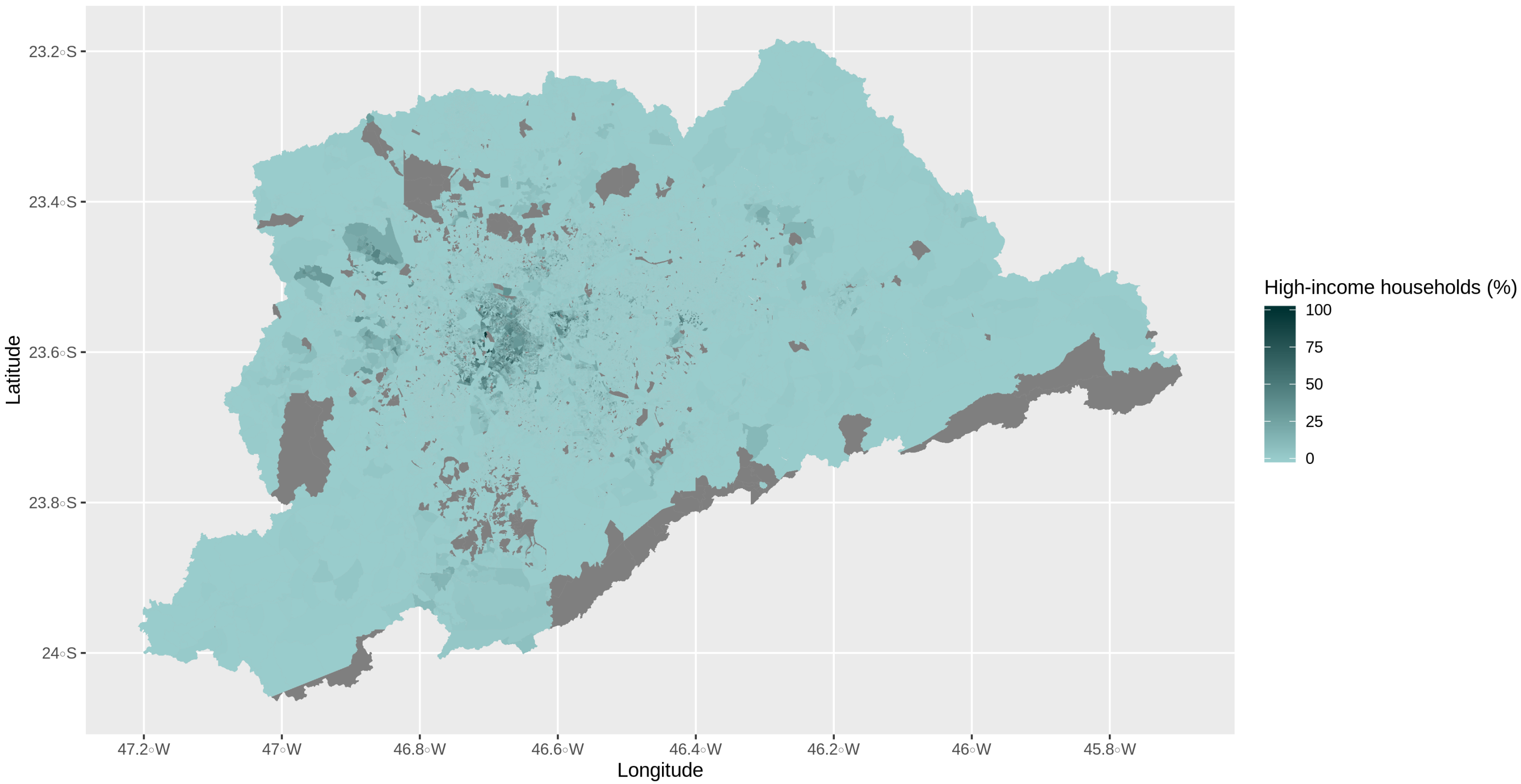
Group	School type	Relative frequency	Absolute frequency
Black (*Preto*)	public	90.28%	149130
Black (*Preto*)	private	9.72%	16054
Brown (*Pardo*)	public	92.16%	1029103
Brown (*Pardo*)	private	7.84%	87494
White (*Branco*)	public	74.78%	1511852
White (*Branco*)	private	25.22%	509984
Yellow (*Amarelo*)	public	53.05%	15719
Yellow (*Amarelo*)	private	46.95%	13910
Native Brazilian (*Indígena*)	public	87.33%	10838
Native Brazilian (*Indígena*)	private	12.67%	1572
Not declared	public	78.42%	1515463
Not declared	private	21.58%	416932

Relative frequencies calculated within racial groups. Source: School Census – IPEA (2010). Elaborated by the author.

4.4 Urban landscape and segregation scales

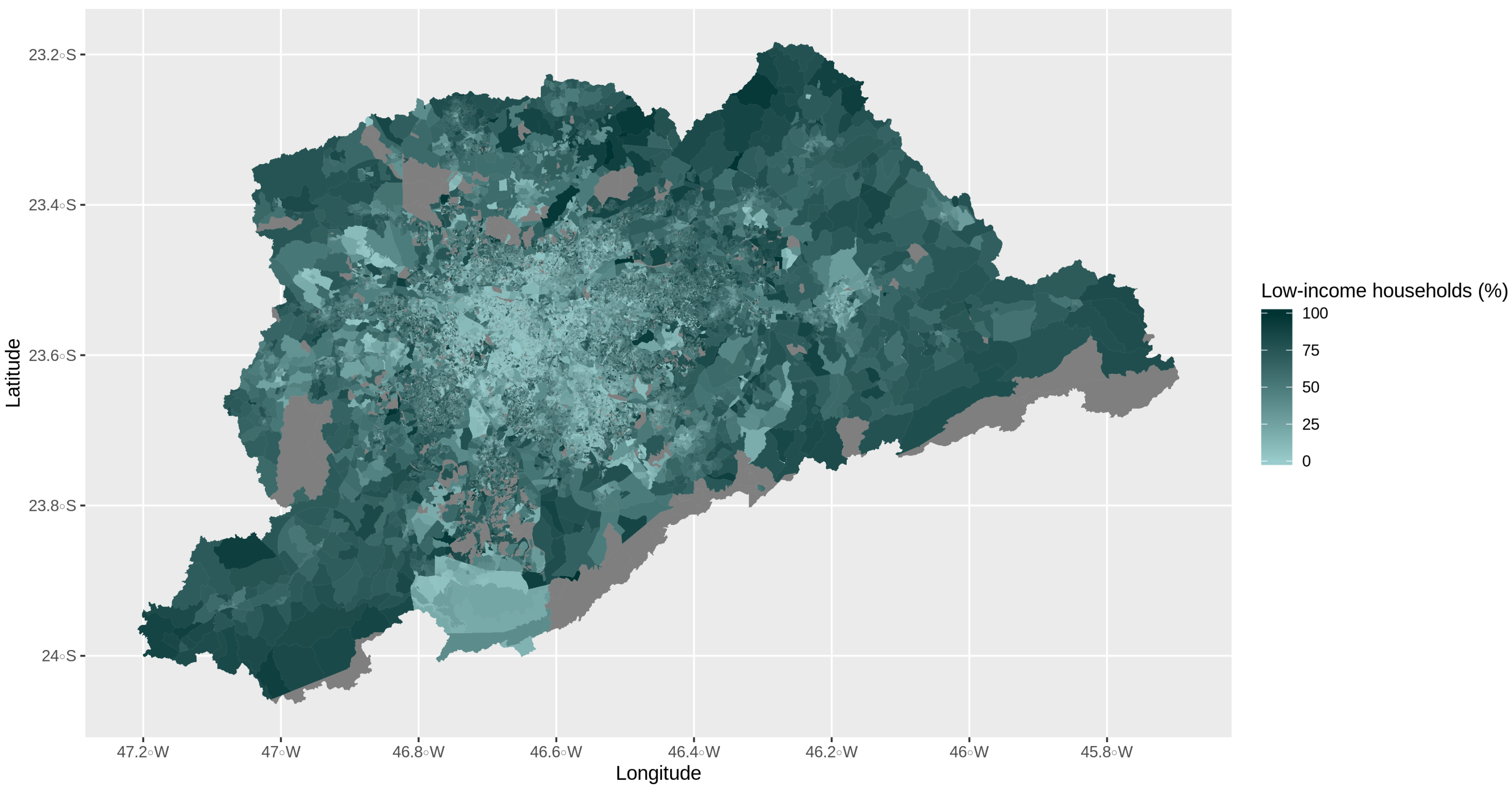
All these inequalities are not apart from space. Geography is indispensable for a comprehensive description of the phenomenon. A macro-segregation pattern in São Paulo has been well documented by the literature: segregation through peripheralization, expelling the poor and the working class towards the fringes of urban space (KOWARICK, 1979; VILLAÇA, 1998; MARICATO, 1996; CALDEIRA, 2000). This pattern is easier to identify through contrasts in urban landscape and maps. Figures 3 and 4 show the percentage of high- and low-income households in census tracts in SPMA, respectively. Grey areas are polygons without income data.

Figure 3 – Percentage of households with 10 *per capita* minimum wages of income or more in SPMA census tracts



Source: Data from IBGE (2010). Elaboration by the author.

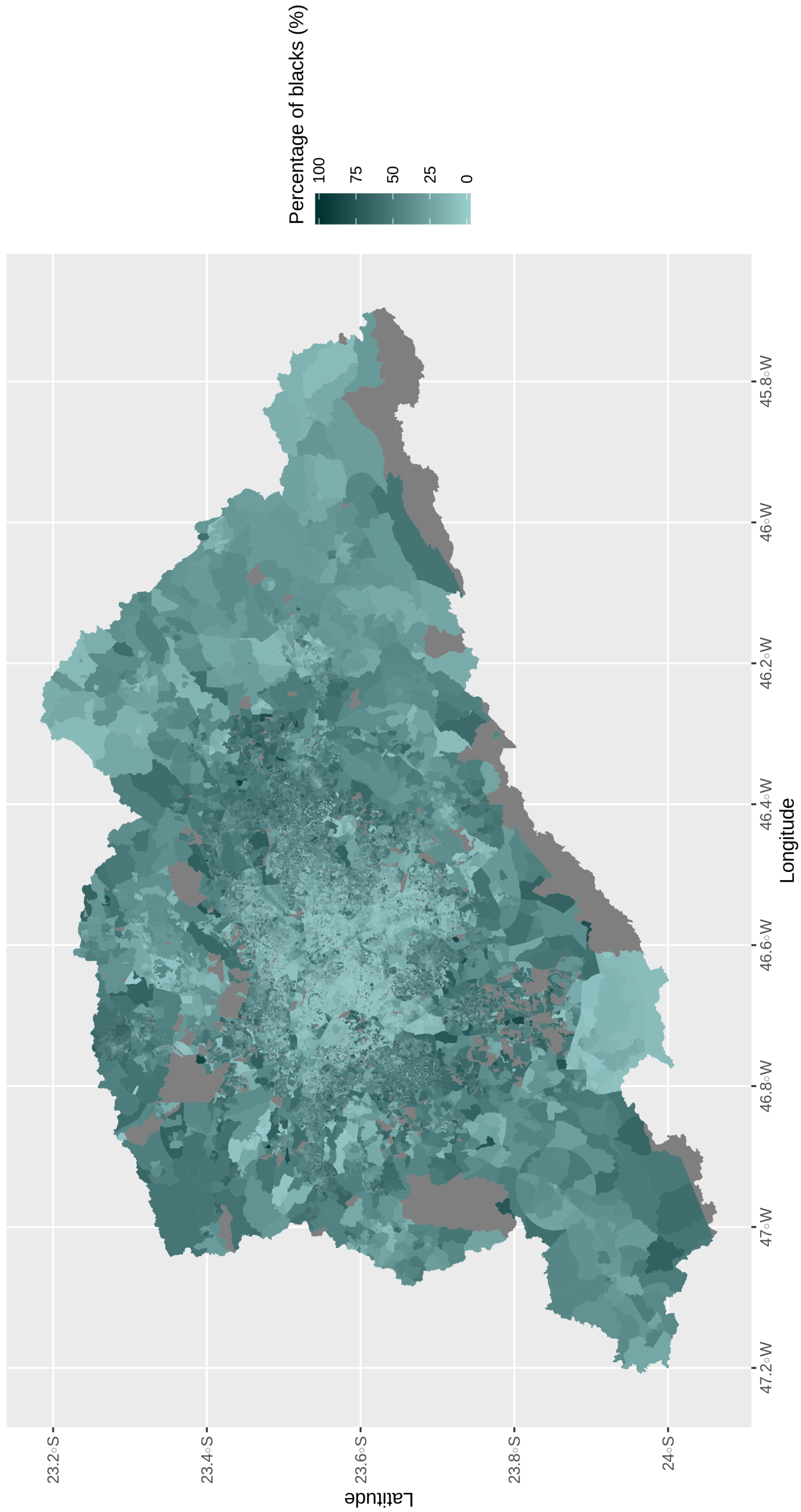
Figure 4 – Percentage of households with less than 1 *per capita* minimum wage of income in SPMA census tracts



Source: Data from IBGE (2010). Elaboration by the author.

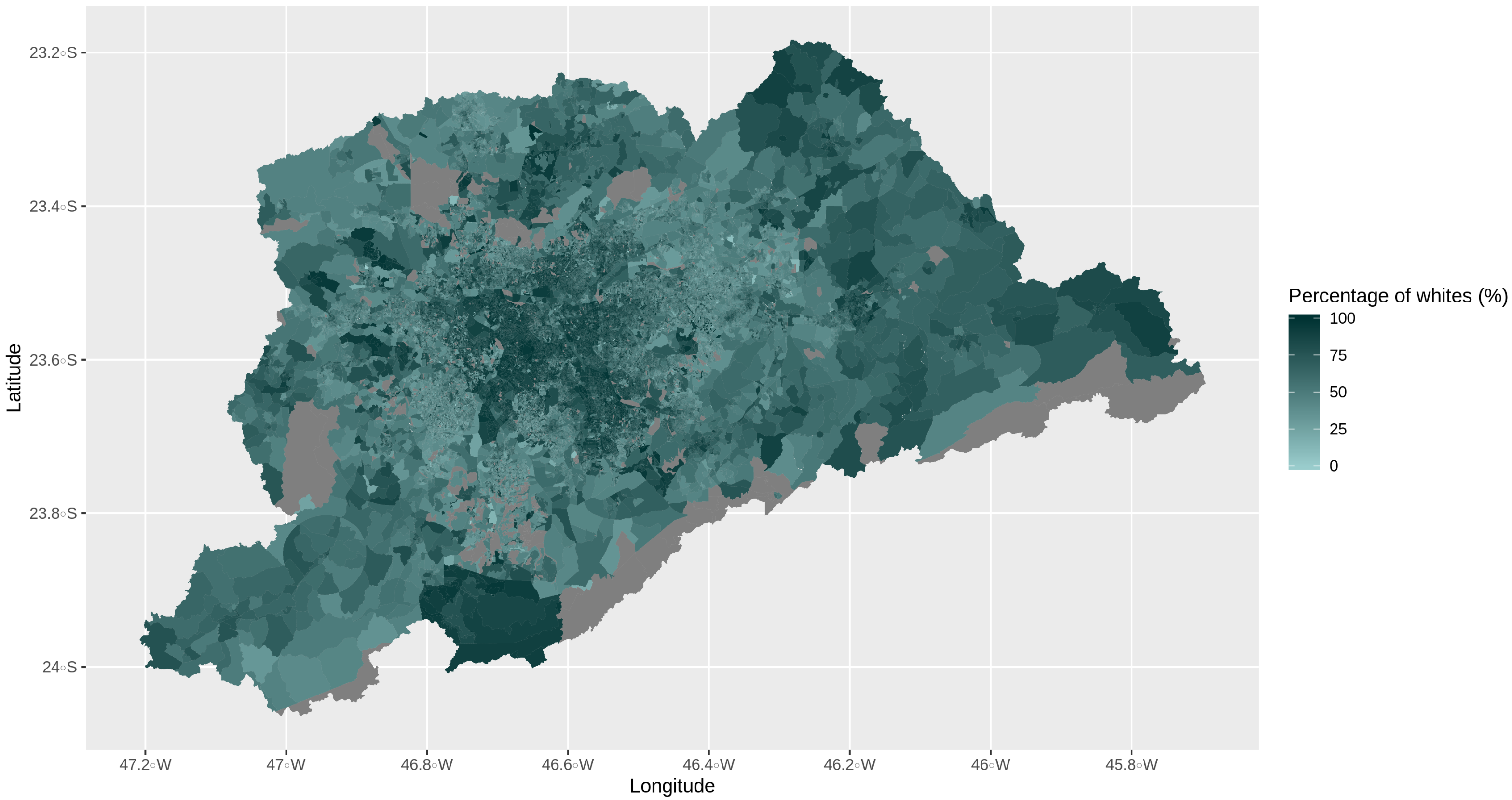
It is also easy to see how these income spatial distributions are correlated with race, specially how blacks do not inhabit the center of the metropolis, the same area where high-income households are concentrated (Figure 5). Meanwhile, whites (*brancos* and *amarelos*) inhabit the central area, although they are also present in peripheral areas, as shown by Figure 6 (grey areas are polygons without data on race). In addition, whites are very numerous and distributed throughout SPMA territory – whites and yellows correspond to 60.67% and blacks and browns to 39.22% of the population in SPMA (Table 2). Thus, whites tend to live in areas where they are more than 50%. Figure 7 presents the boxplot of the percentage of white and yellow, and brown and black residents in census tracts.

Figure 5 – Percentage of black (*pretos* and *pardos*) inhabitants in SPMA census tracts



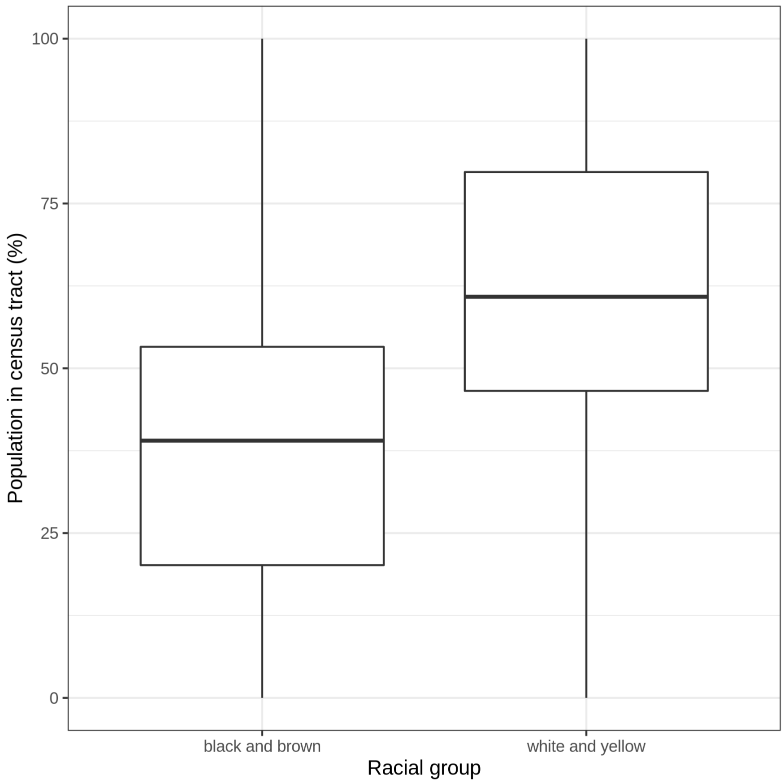
Source: Data from IBGE (2010). Elaboration by the author.

Figure 6 – Percentage of white (*brancos* and *amarelos*) inhabitants in SPMA census tracts



Source: Data from IBGE (2010). Elaboration by the author.

Figure 7 – Percentage of residents in census tract by racial supergroup

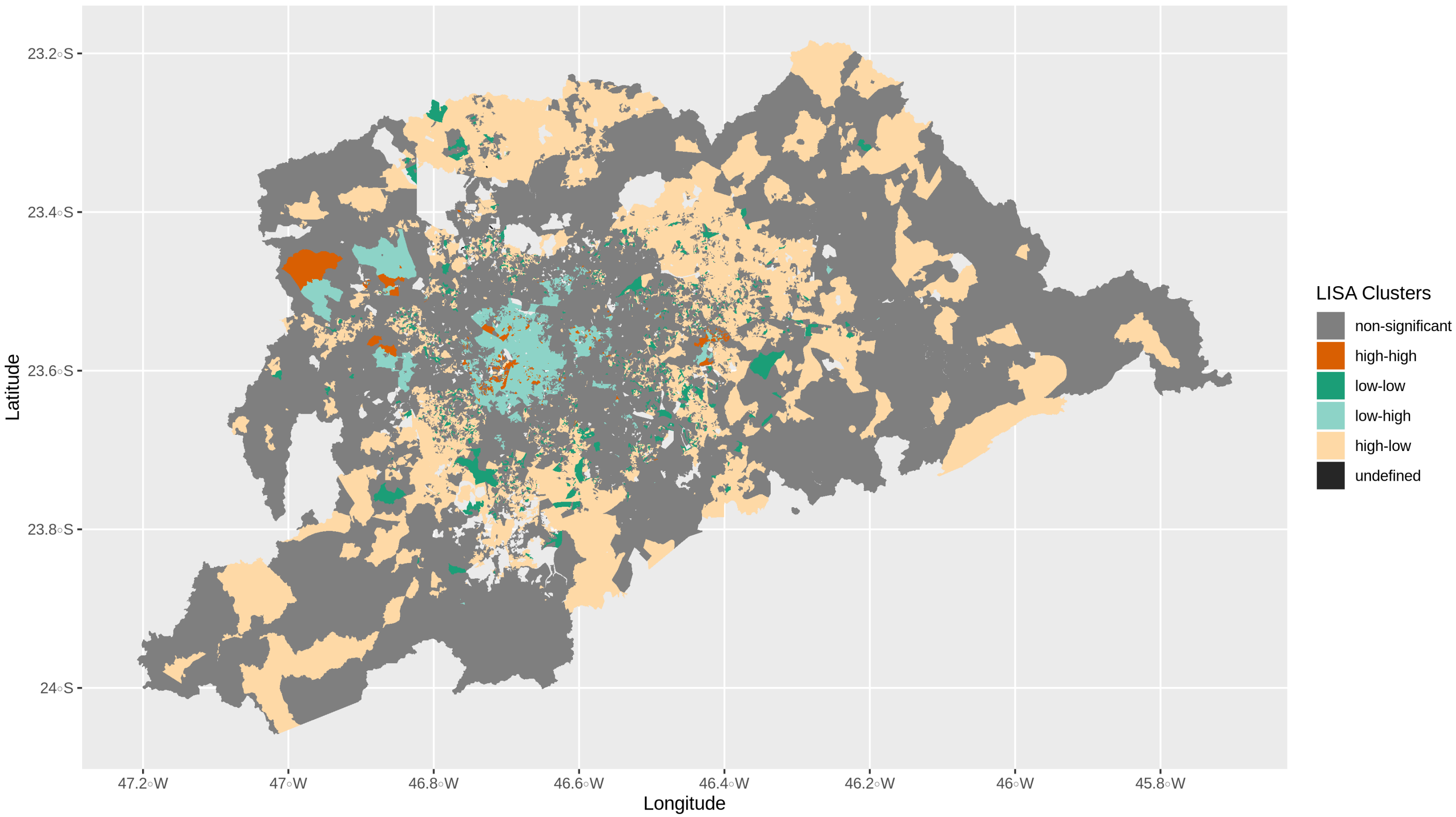


Source: Data from IBGE (2010). Elaboration by the author.

While macro-segregation seems evident, micro-segregation might be harder to spot, since it can be more easily hidden by high buildings, walls and other obstacles in urban space. In order to test this pattern, a preliminary exploration using LISA Maps (Bivariate Local Moran's I) shows spatial correlation between percentage of low-income households (less than 1 *per capita* minimum wage) and percentage of high-income households in census tracts (10 *per capita* minimum wages or more). Income was used in this case, instead of race, because the intention of this analysis is to clearly show social separation, which is caused by the lack of conviviality in physical space, for instance, sharing the same leisure and education establishments.

Figure 8 shows the LISA Map for the entire SPMA. In order to establish neighborhood relationship between census tracts, a maximum distance of 100 meters was considered. Statistical summary of data used to generate the LISA Map is available in Table 16 of Appendix A.

Figure 8 – LISA Map: Percentage of high- and low-income households in census tracts

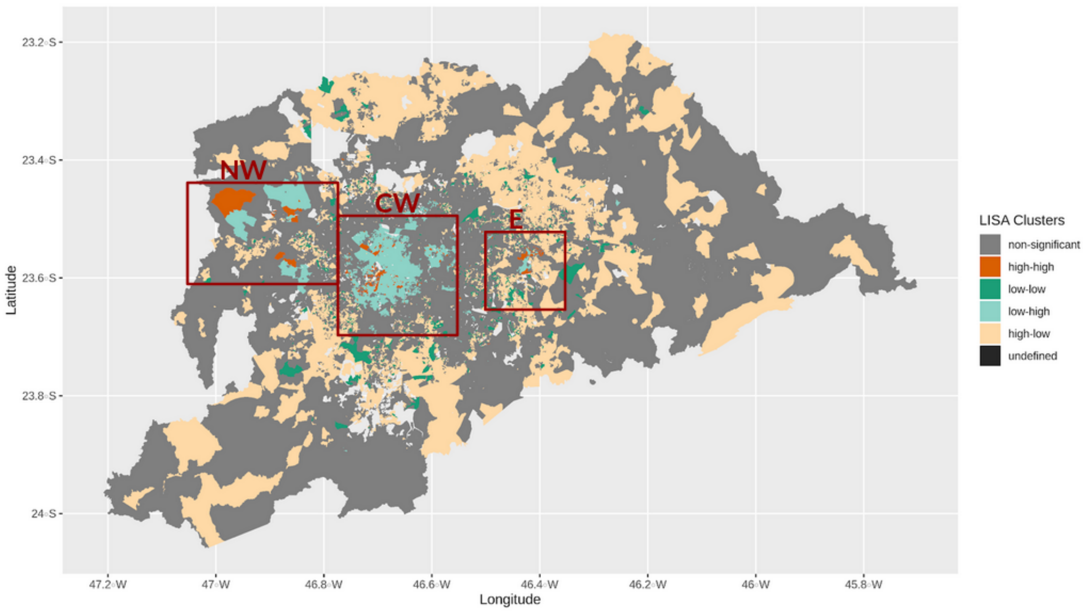


Source: Data from IBGE (2010). Calculations by the author.

There were 299 census tracts in high-high clusters (tracts with high percentage of low-income households close to tracts with high percentage of high-income households), which is the sort of situation previously described. In addition, there were 826 tracts in low-low clusters; 2,617 in low-high clusters (low percentage of low-income and high percentage of high-income); 6,847 in high-low clusters (high percentage of low-income and low percentage of high-income); 18,969 census tracts with non-significant results, and 40 undefined census tracts, meaning those could not have neighborhood relationship established with any other tract because of problems in the polygons' boundaries in the shapefile.[4] Transparent tracts are tracts without income data, thus dropped from the analysis.

A zoomed vision of these clusters can help characterize each of them in detail. In order to do so, three main groups of clusters were highlighted: High-high clusters in the Northwest, in the Center-West and in the East of SPMA, as shown by Figure 9. Those groups of clusters were separated by the author taking into account administrative divisions of the municipality of São Paulo (CW and E). NW concentrates the clusters that fall outside this municipality.

Figure 9 – LISA Map: Percentage of high- and low-income households in census tracts – Three groups of clusters



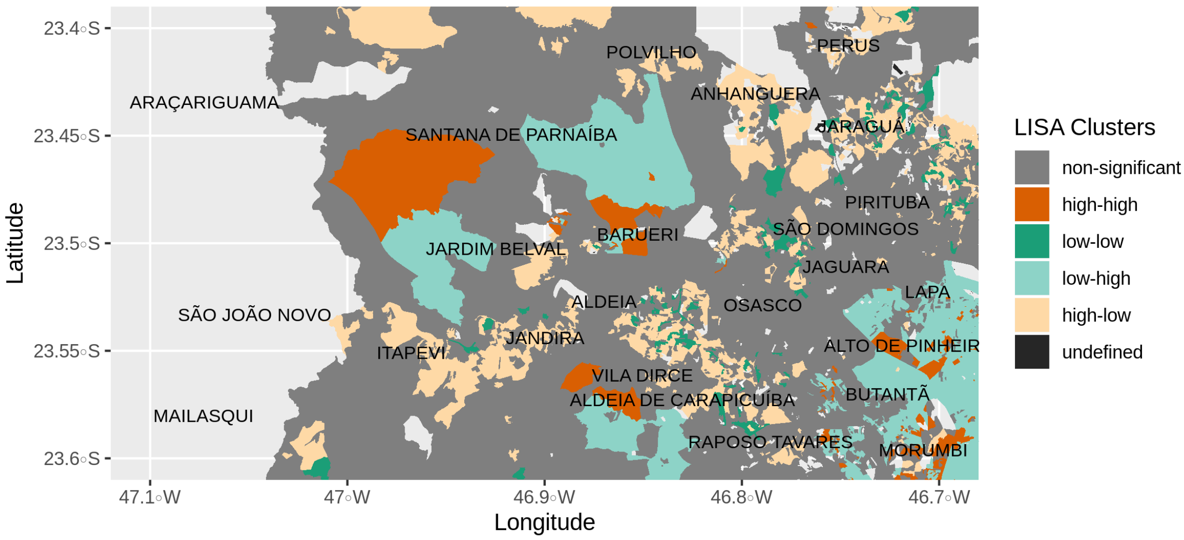
Source: Data from IBGE (2010). Calculations by the author.

4 Higher tolerance distances for neighborhood were tested in order to check if those 40 undefined tracts would vanish or be reduced, but they didn't even for unreasonably large distances.

4.4.1 High-high clusters in the Northwest

The Northwest area has been stage of a quite recent phenomenon of urban sprawl compared to the other two. It is characterized by the multiplication of gated communities, condominiums inhabited by the upper- and upper-middle class. Those are well equipped with private security and leisure infrastructure, reducing the need of its inhabitants to circulate in public areas, which are considered to be unsafe. Alphaville and Aldeia da Serra are two examples of this sets of private neighborhoods. Figure 10 shows the zoomed LISA Map of this region, displaying districts names.

Figure 10 – Northwestern SPMA LISA Map – Labels with districts names

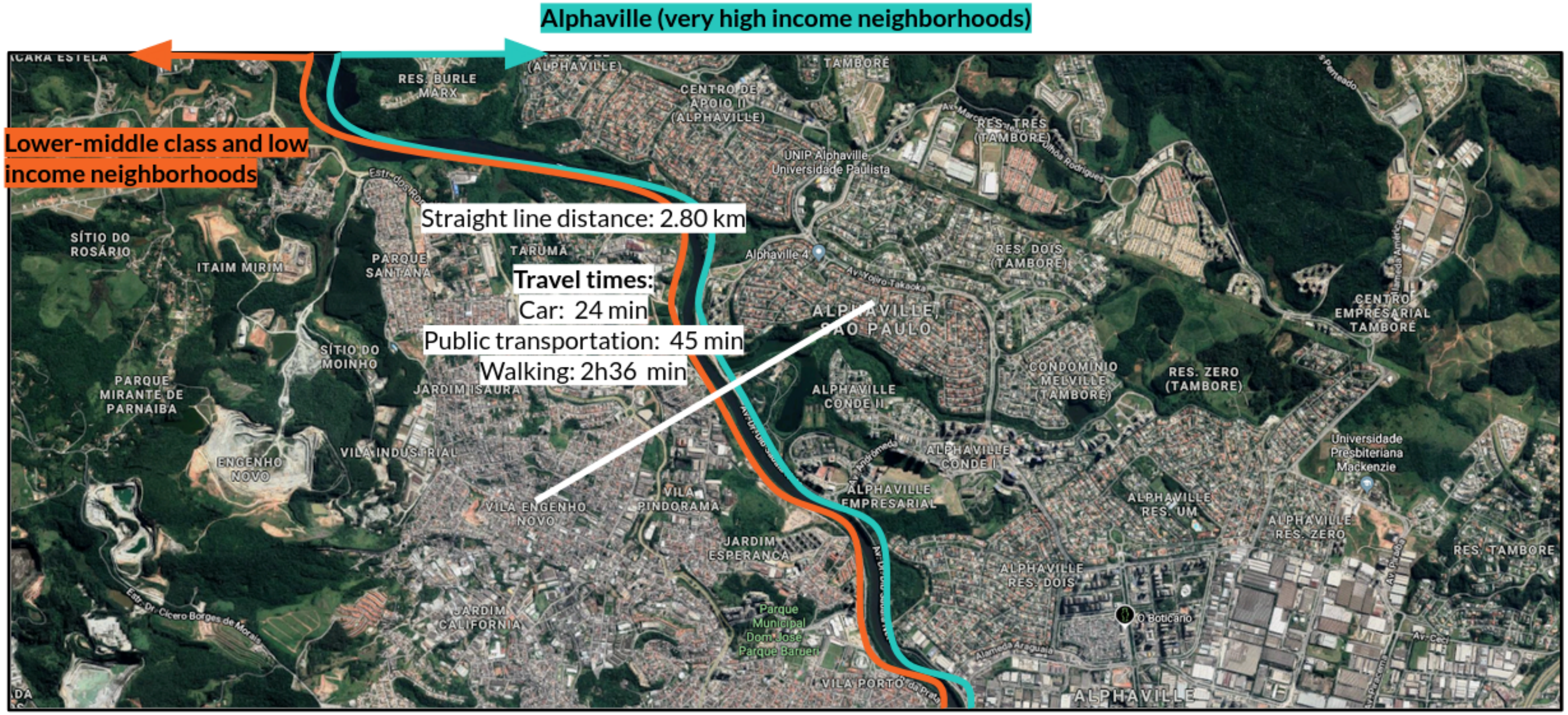


Source: Data from IBGE (2010). Analysis by the author.

Alphaville was the first and is the most emblematic example of fortified enclaves in Brazil. It started in São Paulo and is now present in 20 Brazilian states, as well as Libon and Sintra in Portugal. It was first built during the 1970s by private initiative. Initially, Albuquerque & Takaoka – the enterprise responsible for the project – invested in industrial plots. The municipality of Barueri seemed to be a good choice because it is not far from the capital and production could easily outflow through the Castello Branco highway. However, entrepreneurs soon realized the potential of their business model to commercial and residential purposes. The first residential plot was launched in 1979 and followed a rhythm of one plot per year, with a resounding success. At first, publicity focused in social status of the Alphaville high-income inhabitants as the main value of the brand, which has shifted towards better life quality since the 2000s – proximity to nature and leisure infrastructure, in contrast with the pollution, chaos and violence of the neighboring big city. However, the social homogeneity prevailing within its walls and the safety seem to be the constant values that continue to be preached throughout the four decades since the first plot was launched. Security was a special concern of the chief engineer Takaoka and a subject highlighted by Alphaville inhabitants in interviews (GUERRA, 2013).

These private neighborhoods were built close to smaller towns that already existed, in which income is much lower than within those gated communities (Figure 12). This is the reason why they figure among the high-high clusters. Some examples may be helpful to paint this picture. Figure 12 refers to the two sides indicated in Figure 11. In Figure 11, one can observe a clear geographical barrier between those areas, the Tietê river. This is one case in which straight line distance might be mistaken as a measure of the social distance between individuals in the opposite sides of the white segment. Also, travel times took as examples here show that this area seems to privilege private means of transportation, with a travel time by public transportation that is almost double of the travel time by car. In Figure 13, the difference of densities between the two neighborhoods is evident. While Vila Dirce is very dense, Aldeia de Carapicuíba is sparse and greener, with curve streets in order to reduce the traffic flow through the neighborhood.

Figure 11 – Alphaville and Barueri – Satellite image



Source: satellite images, distance and travel time: Google Maps. Nov. 23, 2019. Elaboration by the author.

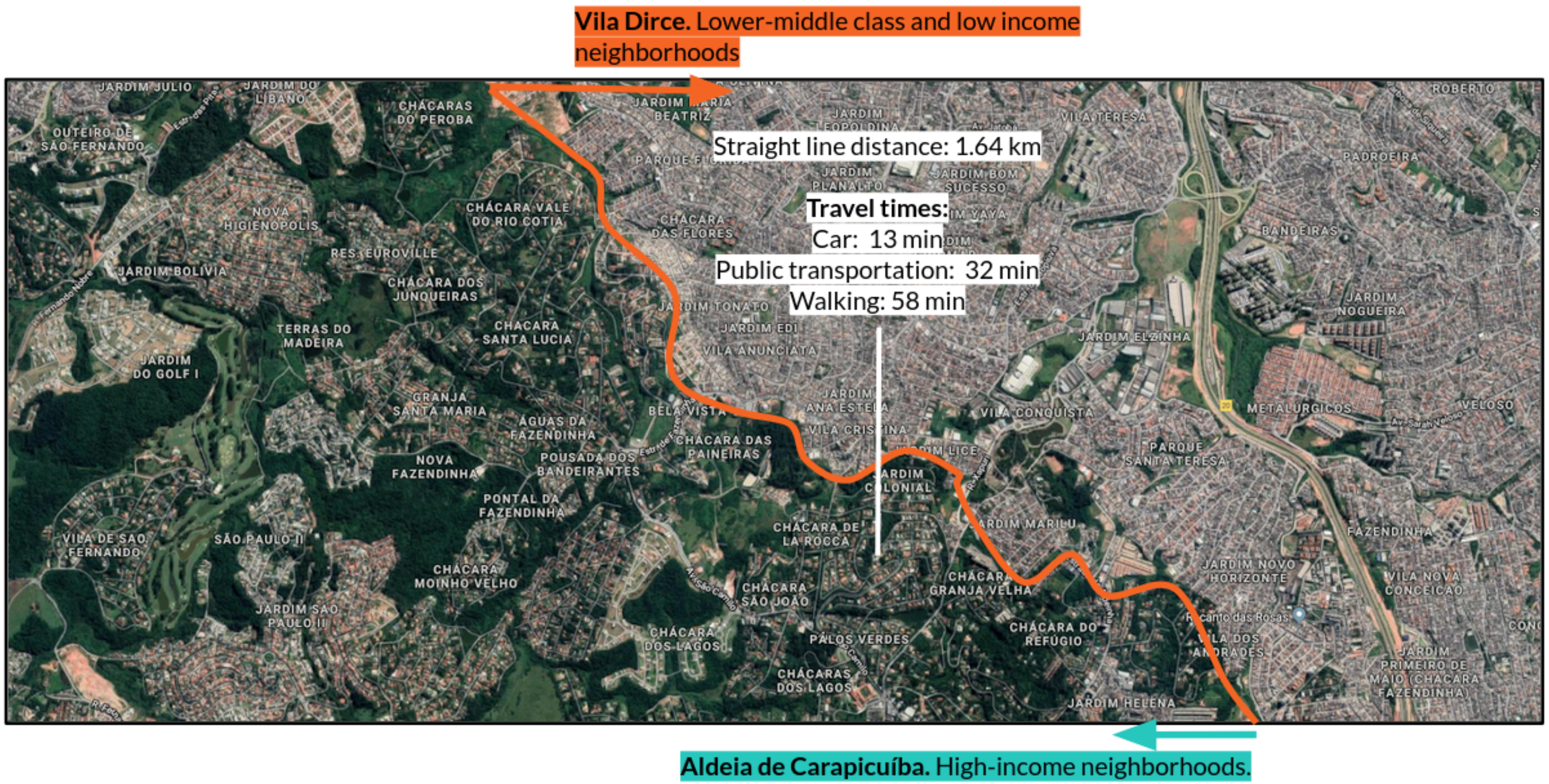
Figure 12 – Alphaville and Barueri – Street View

From up to bottom: Rua José Maria de Abreu and Rua Paraná (Barueri).
Source: Google Street View (23 nov. 2019).

Entrances of two Alphaville's gated communities.
Source: Google Street View (23 nov. 2019).

Source: Google Street View (23 nov. 2019).

Figure 13 – Aldeia de Carapicuíba and Vila Dirce – Satellite image



Source: satellite images, distance and travel time: Google Maps. Nov. 23, 2019. Elaboration by the author.

Figure 14 – Aldeia de Carapicuíba and Vila Dirce – Street View

Rua Cavalcante (Jardim Donato / Vila Dirce, Carapicuíba).
Source: Google Street View (23 nov. 2019).

Rua Ouro Preto (Aldeia de Carapicuíba).
Source: Google Street View (23 nov. 2019).

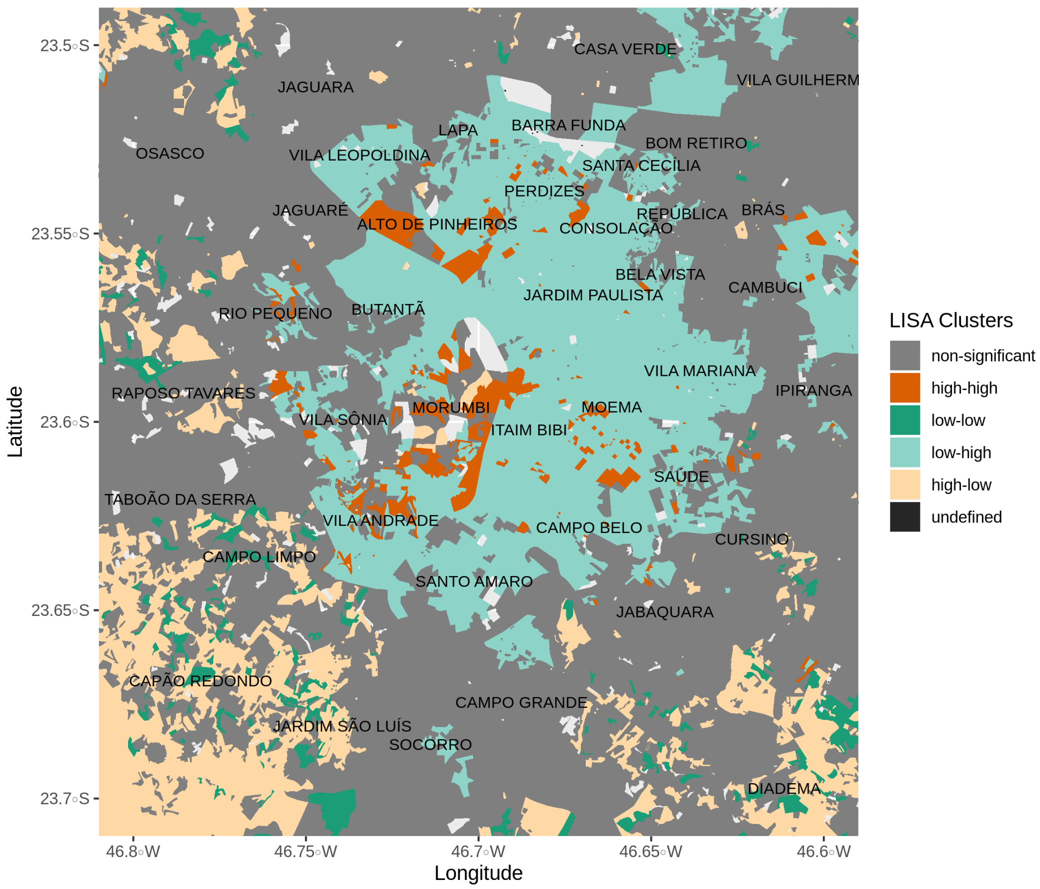
Source: Google Street View (23 nov. 2019).

4.4.2 High-high clusters in the Center-West

The area shown in Figure 15 is the one with the oldest occupation comprising many historical neighborhoods, as well as the largest economic center of the metropolis. According to Hermann and Haddad (2005), São Paulo can be defined as a duocentric city: the first center being Sé-*Avenida* Paulista and *Avenida* Berrini–*Avenida* Faria Lima being the second. Most of the job opportunities concentrate in those two axes, and both of them are comprised in Figure 15. Considering the 15 years past since Hermann and Haddad stated the above, one could add that new centers have been developed in the meantime. For instance, there are new industrial and commercial parks in Alphaville, that is also becoming an important area in the tertiary sector of the metropolitan economy. Anyway, this Center-Western SPMA continues to be the most important in economic terms.

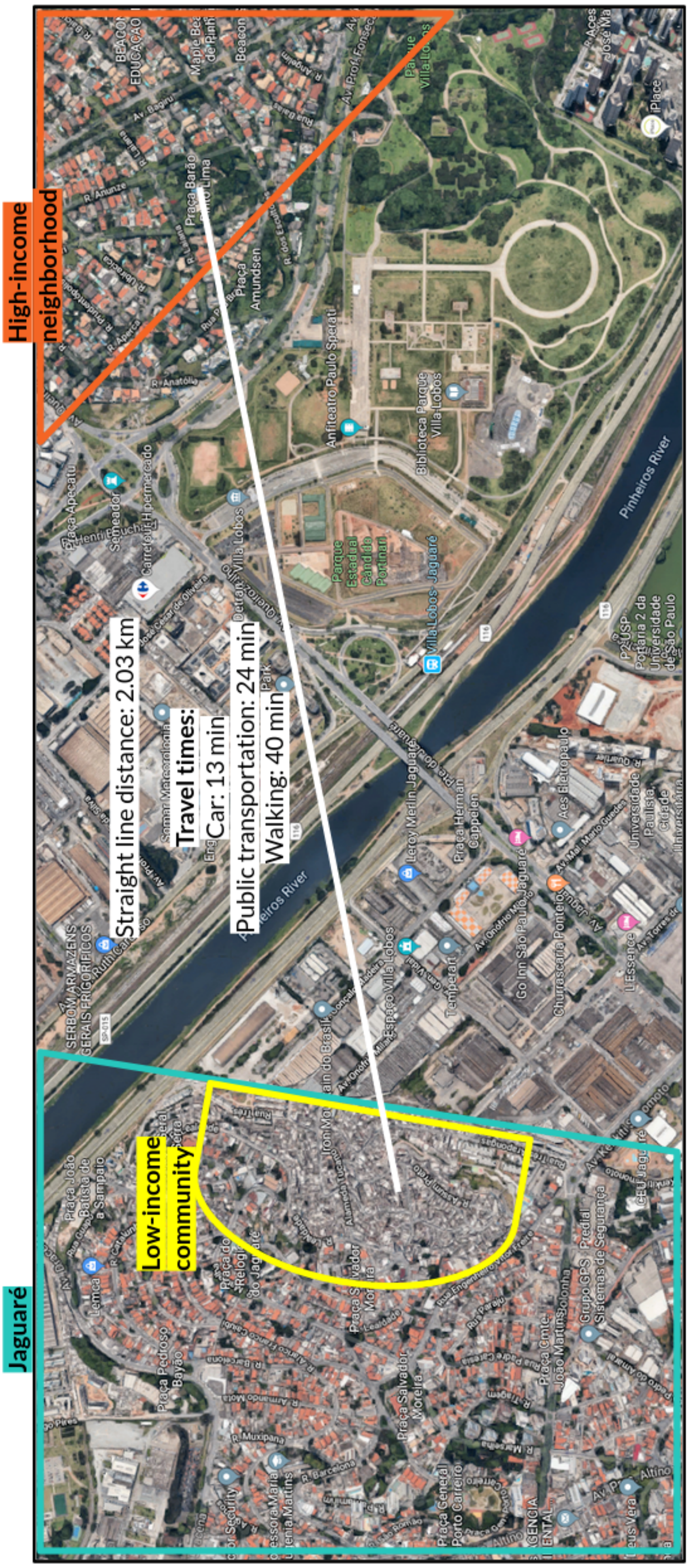
Figures 16 and 17 constrast Pinheiros and Jaguaré, another case in which there is a river between the high- and the low-income neighborhoods, and Figure 18 presents Morumbi and Paraisópolis satellite image (this is the area spotted by Figure 2).

Figure 15 – Center-Western SPMA LISA Map – Labels with districts names



Source: Data from IBGE (2010). Analysis by the author.

Figure 16 – Pinheiros and Jaguaré – Satellite image



Source: satellite images, distance and travel time: Google Maps. Nov. 23, 2019. Elaboration by the author.

Figure 17 – Pinheiros and Jaguaré – Street View

Rua Carcara (Jaguaré). Source: Google Street View (23 nov. 2019).

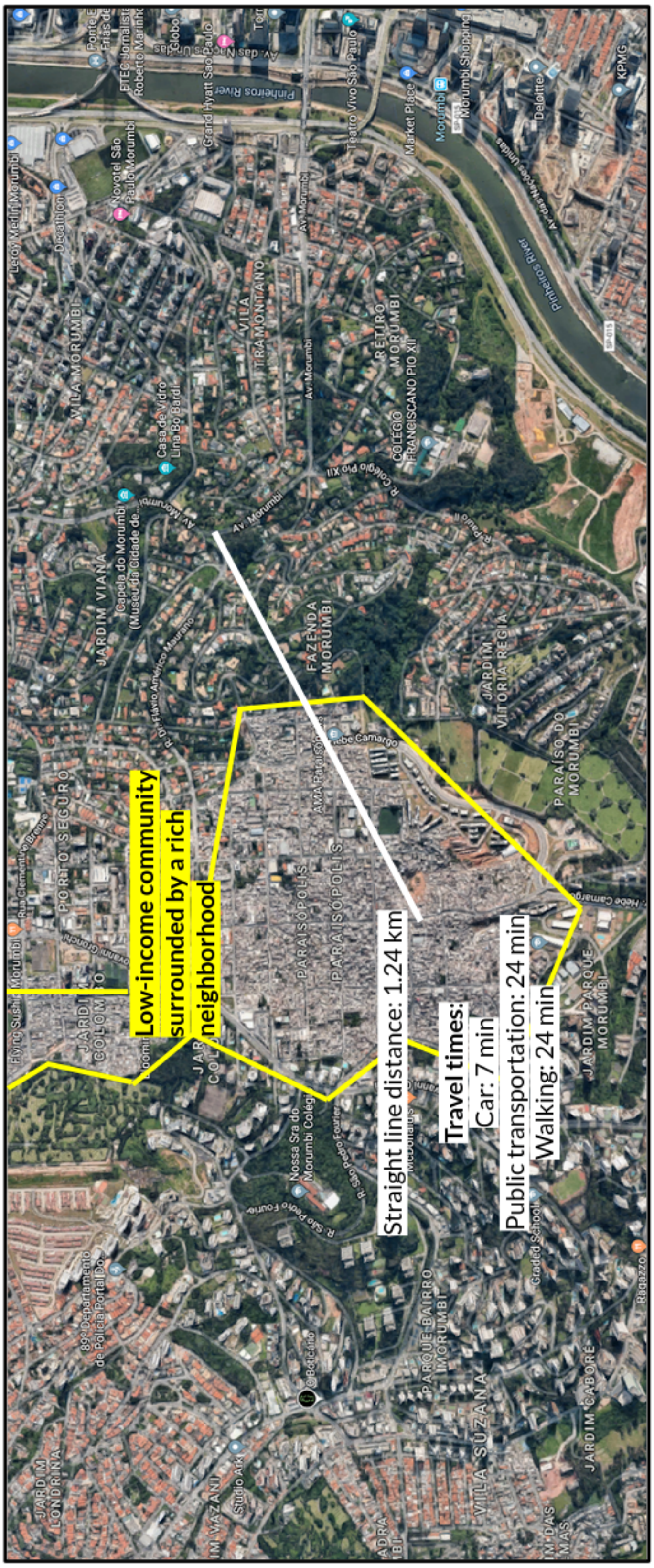
Alameda Tucano (Jaguaré). Source: Google Street View (23 nov. 2019).

Rua Antônio Bicudo (Pinheiros). Source: Google Street View (23 nov. 2019).

Praça Coronel Caetano del Cioppo (Pinheiros). Source: Google Street View (23 nov. 2019).

Source: Google Street View (23 nov. 2019).

Figure 18 – Morumbi and Paraisópolis – Satellite image



Source: satellite images, distance and travel time: Google Maps. Nov. 23, 2019. Elaboration by the author.

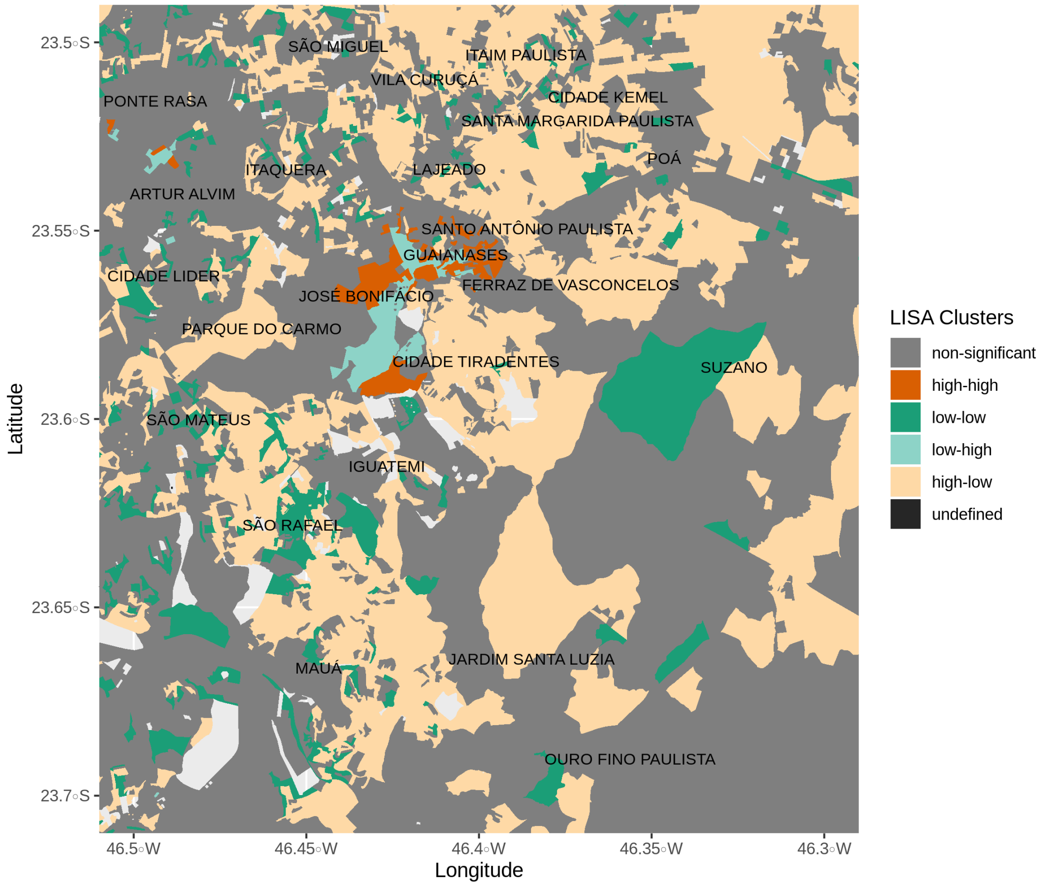
4.4.3 High-high clusters in the East

High-high clusters in the East of SPMA – displayed in Figure 19 – are located in *Zona Leste* (East zone) of the municipality of São Paulo. The urbanization of this area has been marked by different housing policies for low-income families noticeably since the 1970s. It was characterized by the lack of urban services, such as public transportation which has made and can still make commuting to workplaces a long travel. It is what Brazilian scholars have called periphery. However, the improvements in the supply of services combined to the continuing urban sprawl towards the East gives this area nuances. Peripheries are also marked by contrast and heterogeneity.

The area with the clearest contrast regarding urban landscape is Nossa Senhora do Carmo. There are middle and upper-middle class families living a few meters apart from low-income households and even favelas in this neighborhood. It is clear in the satellite image how density and rooftop colors change from one area to another: roofs in the right side of Figure 20 are predominantly orange and occupation is less dense compared to the area circled in orange. The later has mostly grey rooftops, indicating another type of construction process, marked by informality and self-construction. This is visible in the examples shown in the two images on the left of Figure 21.

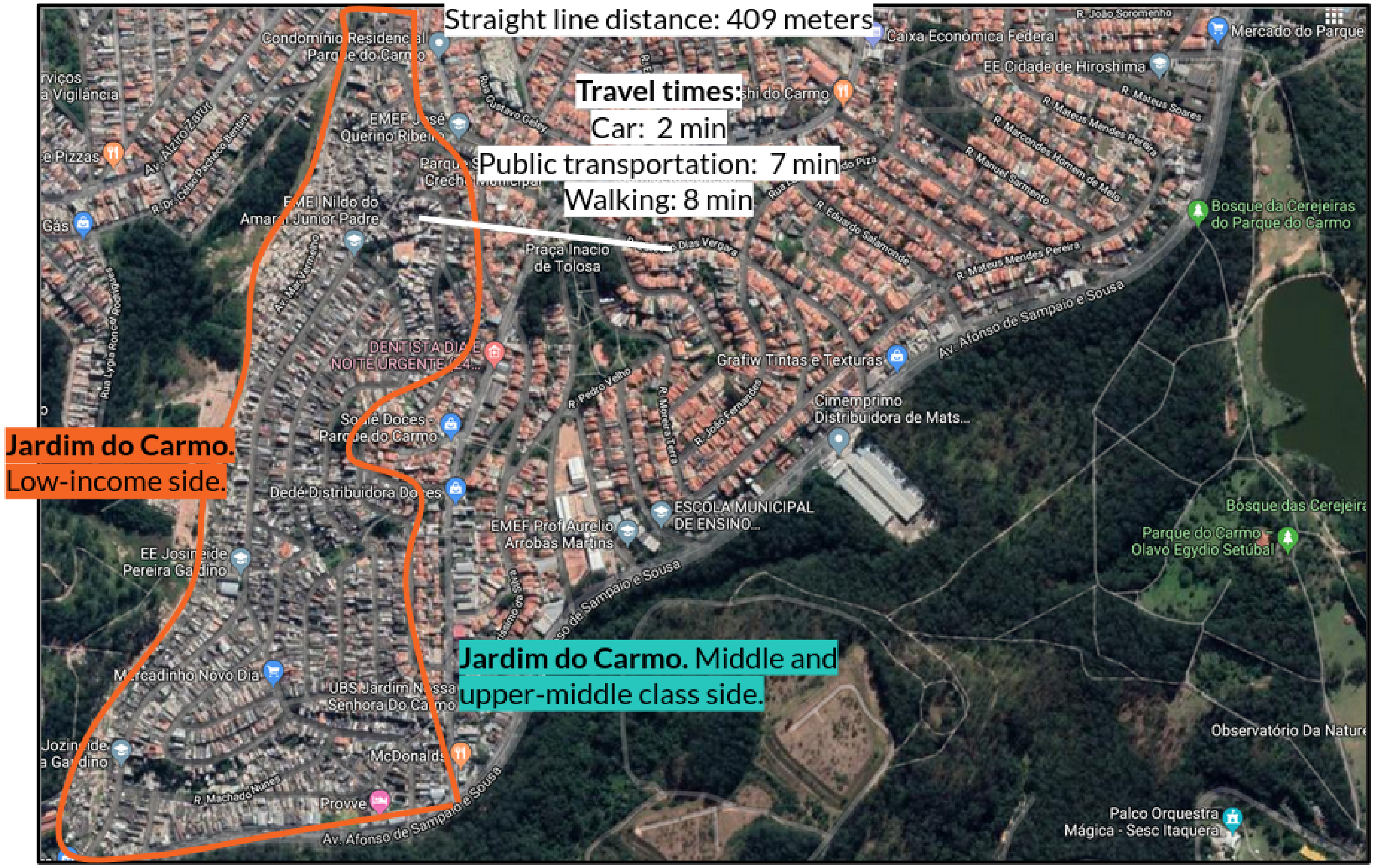
Although located in the rectangle in Figure 19, Jardim Nossa Senhora do Carmo doesn't fall exactly on the orange polygons indicating high-high clusters. High-high clusters area is represented by the satellite image in Figure 22. However, no visibly rich buildings could be found in its surroundings. What calls one's attention, though, are the multiple buildings provided by housing policies. Since housing units are distributed to very vulnerable families, the income contrast between usual middle and lower-middle class houses and families inhabiting those state-provided apartment buildings is less clear in the urban landscape. It would be probably clearer if housing policies hadn't been implemented. Figure 23 displays some of those apartment buildings on the left and usual houses on the right. In any case, the clusters in the East can still be considered the ones with the less clear contrasts if compared to the other two.

Figure 19 – Eastern SPMA LISA Map – Labels with districts names



Source: Data from IBGE (2010). Analysis by the author.

Figure 20 – Jardim Nossa Senhora do Carmo – Satellite image



Source: satellite images, distance and travel time: Google Maps. Dec. 18, 2019. Elaboration by the author.

Figure 21 – Jardim Nossa Senhora do Carmo – Street View

Rua Joaquim Meira de Siqueira (Jardim Nossa Senhora do Carmo).
Source: Google Street View (18 dec. 2019).

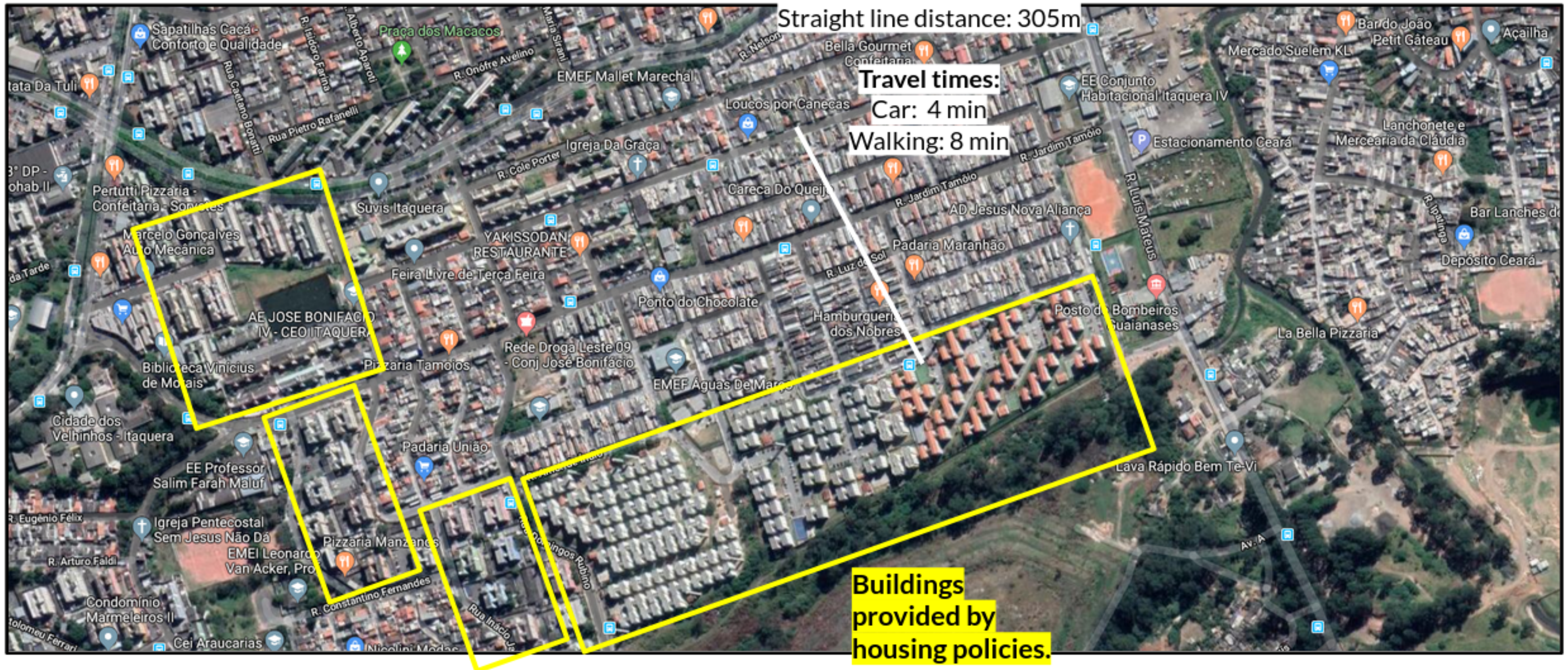
Rua Manuel Sarmento (Jardim Nossa Senhora do Carmo).
Source: Google Street View (18 dec. 2019).

Rua Jerônimo de Abreu do Vale (Jardim Nossa Senhora do Carmo).
Source: Google Street View (18 dec. 2019).

Rua Estevão Dias Vergara (Jardim Nossa Senhora do Carmo).
Source: Google Street View (18 dec. 2019).

Source: Google Street View (18 dec. 2019).

Figure 22 – Guaianases – Satellite image



Source: satellite images, distance and travel time: Google Maps. Dec. 18, 2019. Elaboration by the author.

Figure 23 – Guaianases – Street View

Rua Fascinação (Guaianases). Building provided by housing policy.
Source: Google Street View (18 dec. 2019).

Rua Frei Antônio Fraggiano (Guaianases).
Source: Google Street View (18 dec. 2019).

Source: Google Street View (18 dec. 2019).

5 METHOD

The empirical strategy presented in this chapter was built based on the characterization of racial inequalities and urban contrasts in SPMA presented in the last chapter. The segregation measures are discussed in the next section in order to justify the choice of an index. In the following sections, data structure and robustness test are detailed.

5.1 Choosing the measures

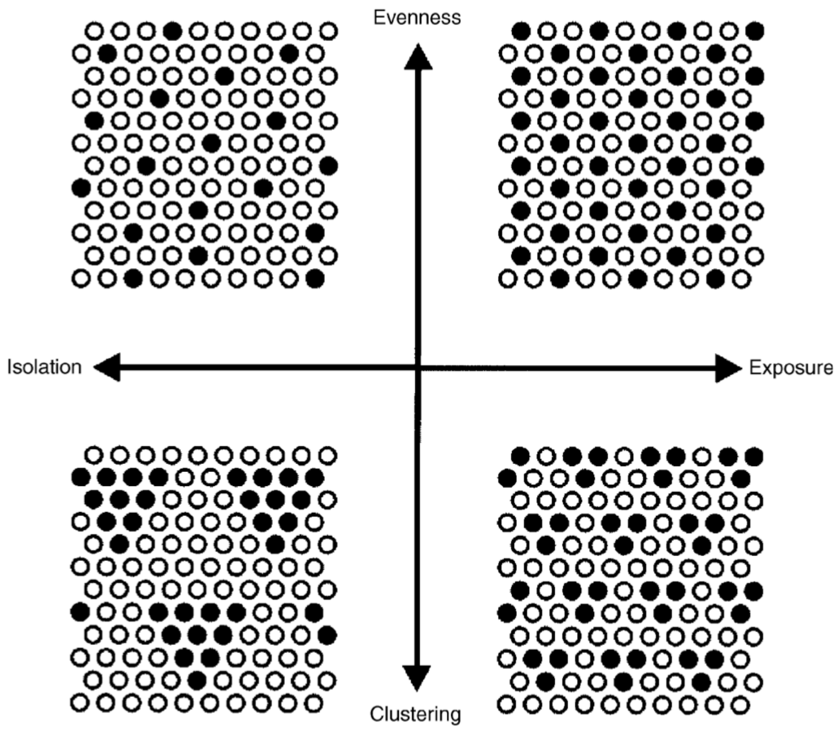
Since the 1950s, many studies have been focused on segregation measurement. Multiple measures were developed ever since, as presented in section 3.2. Thus, a systematic classification of the main indices is helpful to justify the choice of the indices here adopted.

Massey and Denton (1988) propose five different dimensions of segregation: evenness, exposure, clustering, centralization and concentration:

- *evenness:* the degree of under- or over-representation of individuals of a group in a territorial unit (*e.g.* a district) in comparison to the proportion of that group in the overall population in the territory (*e.g.* the entire city). Measurement examples: Index of Dissimilarity, Gini Index, Entropy Index or Information Index.
- *clustering:* is a measure of the proximity of territorial units with similar group compositions. (*e.g.* a city where districts with higher percentage of blacks are all placed in the north of a city, neighboring each other, should have a high clustering index). Measurement examples: Absolute Clustering Index, Spacial Proximity Index, Relative Clustering Index.
- *exposure:* the degree to each a member of a group is likely to encounter members of another group. The exposure of a group to itself would be *isolation*. Measurement examples: Interaction Index, Isolation Index.
- *concentration:* refers to the amount of physical space occupied by a group in the city. Measurement examples: Duncan's Delta Index, Absolute Concentration Index, Relative Concentration Index.
- *centralization:* refers to the placement of a group near to the center of the city. Measurement examples: Proportion in Central City, Absolute Centralization Index, Relative Centralization Index.

In a theoretical and methodological contribution, Reardon and O'Sullivan (2004) propose that three of the Massey and Denton's dimensions can be reorganized in two. Evenness and clustering would be two extremities of the same dimension, while isolation and exposure, two extremities of the second. Figure 24 illustrates the two-dimension approach.

Figure 24 – The dimensions of spatial segregation



Source: Reardon & O'Sullivan (2004).

Isolation (exposure) measures are more sensitive to variation in population proportions, because the probability of encountering a member of the same (other) group depends on the proportion of groups in the overall population of the region of study. On the other hand, evenness (clustering) measures are not as sensitive to variation of the racial composition in the area of interest. The perfectly even distribution do not necessarily depend on group proportions in the overall population of the region of study. The existence of the same group proportion in all sub-units of the territory and in the entire region of study suffices to make a certain area racially even.

Even though, variations in this sensibility to racial composition can be found in evenness indices as well (BARROS; FEITOSA, 2018). Their results suggest the spatial Dissimilarity Index ($\tilde{D}$) is more sensitive to the choice of the region of study, while the spatial Information Theory Index ($\tilde{H}$) is more sensitive to variation in grouping system. This happens because $\tilde{H}$ takes into account an ideal racial composition according to the number of groups (*e.g.* 50% of the population pertaining to each group, for a 2-group system) as a benchmark. This benchmark is not altered by the chosen data aggregation level, but is susceptible to changes in the number of groups. This might be taken into account in further analysis.

The lack of consensus in literature about evenness/clustering measures in comparison to isolation/exposure (BARROS; FEITOSA, 2018) and D, an evenness/clustering measure, being one of the most commonly applied index – Brazilian literature included, as shown by Table 1 – are the reasons why this dimension is the focus of this research.

Reardon and O'Sullivan (2004) review the most used segregation indices for each dimension in both spatial and aspatial versions and propose an evaluation framework, with eight criteria according to which they assess each index. Although the authors judge $\tilde{H}$ as being the more consistent evenness/clustering index, D is still largely adopted.

In its both spatial and aspatial versions, D fails to fully meet the exchange criteria and the decomposition criteria (REARDON & O'SULLIVAN, 2004, p. 141). If segregation is reduced when two individuals of different groups living in different points in the region of interest change places in a way that, after the exchange, they are in points where they are less likely to be encountered **and** "some territorial units become more similar to the overall proportion of the population, leaving proportions unchanged in all other units and for all other groups everywhere" (REARDON & O'SULLIVAN, 2004, p. 135), the exchange criteria are met.

D also fails to meet the decomposition criteria:

§ "**Additive spatial decomposability.** *If X subareas are aggregated into Y larger spatial areas, then a segregation measure should be decomposable into a sum of within- and between-area components.*

§ **Additive grouping decomposability**: *if M groups are clustered in N supergroups, then a segregation measure should be decomposable into a sum of independent within- and between-supergroup components.*"
(REARDON & O'SULLIVAN, 2004, p. 136).

Hence, D has been proven not to be the best way of measuring segregation and has been largely criticized. Other than the before mentioned checkerboard and the grid problem, D in its aspatial (and most commonly used) version as well as the spatial autocorrelation approach – the Moran's I (Frank, 2003) – rely on tract definitions. On relying on pre-established tract boundaries, they are susceptible to the Modifiable Areal Unit Problem (MAUP), a frequent issue with spatial data, first highlighted by Yule and Kendal (1950) . This problem relies on the fact that different data aggregation engenders different results. Since political boundaries (*e.g.* districts of a city) are usually defined in an arbitrary way and do not reflect social separation or discontinuity, measures that rely on those boundaries might engender questionable results (OPENSHAW & TAYLOR, 1979).

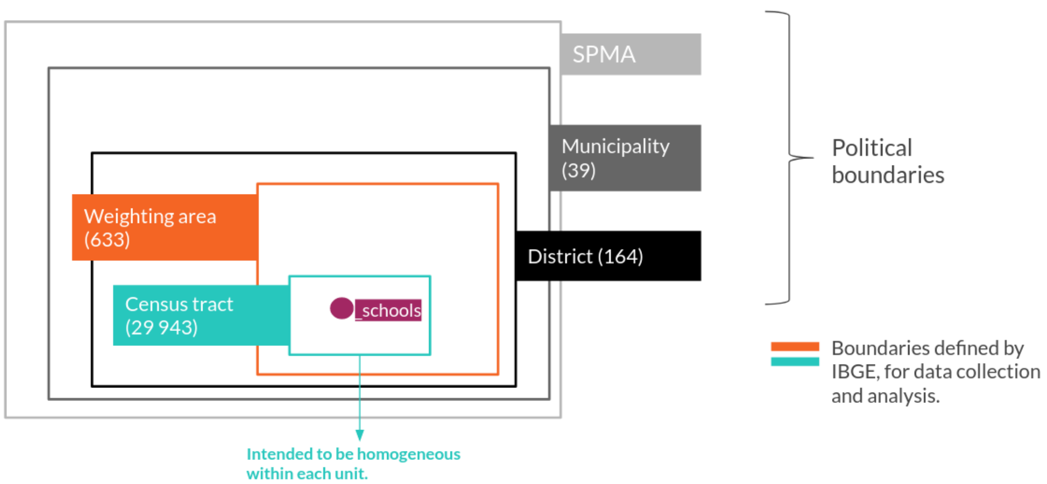
Thus, two measures of evenness/clustering are adopted – the aspatial Dissimilarity Index (D), because it is the most popular, and the spatial Information Theory Index, considered the most conceptually and mathematically satisfactory index according to Reardon and O'Sullivan's (2004) tests.

The main analysis of this research consists on the comparison of residential segregation indices throughout the urban territory. The spatial distribution of local indices is the focus of this research, to the detriment of global indices. This is also a wise choice since the calculations couldn't be made to the entire SPMA, for computational reasons. The choice of the specific area of interest will be detailed later in this chapter. Hence, global indices could not be compared with previous researches.

5.1.1 Aspatial measure: the Dissimilarity Index (D)

D is calculated following the definition presented in section 3.2 using demographic and school census data. Demographic census data is available at the census tract (*setor censitário*) level, and school census data, at the school level. As D cannot be computed for the most disaggregated data level because it requires aggregating original data in larger units, both school and residential D are computed for the weighting area (*área de ponderação*). Figure 25 illustrates the geographic levels in SPMA.

Figure 25 – Geographic levels in SPMA



Source: Elaborated by the author.

School segregation could be calculated for the census tract level with D, even though data is available at the school level, because there are too many tracts without any school or with one single school, resulting in too many observations to which the index would not be calculated.[1] In addition, schools in urban areas doesn't only serve students residing in the same census tract where the school is located. Census tracts are small in terms of area and could be

[1] Section 5.3 describes these numbers in detail.

the size of a block or even one single building. Hence, it is more reasonable to choose a slightly larger area, the reason why the weighting area seems to be a better fit.

5.1.2 Spatial measure: Information Theory Index ($\tilde{H}$)

In order to facilitate reading, let's clarify some notation used in this section. Following Reardon and O'Sullivan's notation (REARDON; O'SULLIVAN, 2004):

- R: region to be studied.
- T: total population of R.
- T_m: total population of group m living in R (same for group n).
- p: local environment, a sub-unit of R.
- q: all other sub-units of R other than p.
- Q: total number of subunits of R (ideally, points where each individual is placed).
- M: total number of population subgroups (in this case, racial groups).
- τ_p: population density in point p, given by a fraction of inhabitants per area unit.
- τ_{pm} or τ_{pn} : group m or group n population density in point p.
- $\phi(p, q)$: proximity function – must increase when proximity increases.
- Φ_p: proximity of a point $p \in R$ to all other q points in R:

$$\Phi_p = \int\limits_{q \in R} \phi(p, q) dq$$

Integral will be used to denote summations over all points of R.

- $\tilde{\tau_p}$: spatial version of τ_p. Spatially weighted average density of population in p.

$$\tilde{\tau_p} = \frac{1}{\Phi_p} \int\limits_{q \in R} \tau_q \phi(p, q) dq$$

This definition is also valid for the calculation of the spatially weighted density of a subpopulation, for instance group m ($\tilde{\tau_{pm}}$). Note that the geographic distribution of $\tilde{\tau_{pm}}$ is a smoothed surface of group m in R.

- π_m: proportion of group m in total population.
- π_{pm}: proportion of group m in p, given by: $\pi_{pm} = \frac{\tau_{pm}}{\tau_p}$.
- $\tilde{\pi_{pm}}$: spatially weighted proportion of group m in p, given by:

$$\tilde{\pi_p} = \frac{\tilde{\tau_{pm}}}{\tilde{\tau_p}}$$

Thus, the Information Theory Index can be defined as:

$$\tilde{H} = \frac{1}{TE} \int_{p \in R} \tau_p \tilde{E_p} dp$$

Where E and $\tilde{E_p}$ are aspatial and spatially weighted entropy indices, respectively – measures of population diversity, defined as follows (Theil, 1972 *apud* Reardon and O'Sullivan, 2004):

$$E = -\sum_{m=1}^{M} (\pi_m) log_M(\pi_m)$$

$$\tilde{E_p} = -\sum_{m=1}^{M} (\tilde{\pi_m}) log_M(\tilde{\pi_m})$$

Thus, $\tilde{H}$ measures the difference of diversity between the overall studied region R and an individual's local environment p. The more diverse p is, the higher $\tilde{H}$ will be for the entire region of interest, being 0 complete racial homogeneity and 1 complete racial integration (no difference between diversity in R and p). However, as before mentioned, this investigation focus on local measures, to detriment of global indices. The local measure applied by this research is the Spatial Entropy $\tilde{E_p}$ calculated for each local environment p (census tract in this case). Thus, the map of $\tilde{E_p}$ should represent a surface of values indicating how diverse is that geographic unit, taking into account racial composition in p itself, as well as all other q sub-units of R.

$\tilde{E_p}$ also ranges from 0 to 1, but unlike $\tilde{H}$, $\tilde{E_p}$ assumes higher values when local environment p is more diverse or less segregated. Thus, further comparisons of $\tilde{E_p}$ with D will require an adjustment of scale: D will be compared with $(1 - \tilde{E_p})$. Thus, the closer to 1, the higher the segregation in a certain geographic unit for both D and $(1 - \tilde{E_p})$.

In this research, $\tilde{E_p}$ is computed at the census tract level, the smallest scale possible. Relying on tract boundaries engenders previously mentioned issues, such as the MAUP. However, as census tracts are usually defined taking into account real physical barriers (*e.g.* streets, avenues, rivers), their boundaries might actually be an important source for guessing physical urban structure (GARCÍA-LÓPEZ; MORENO-MONROY, 2016). In addition, taking into account data in neighboring units – neighborhood being defined by the distance-decay function – reduces the MAUP.

5.2 Data

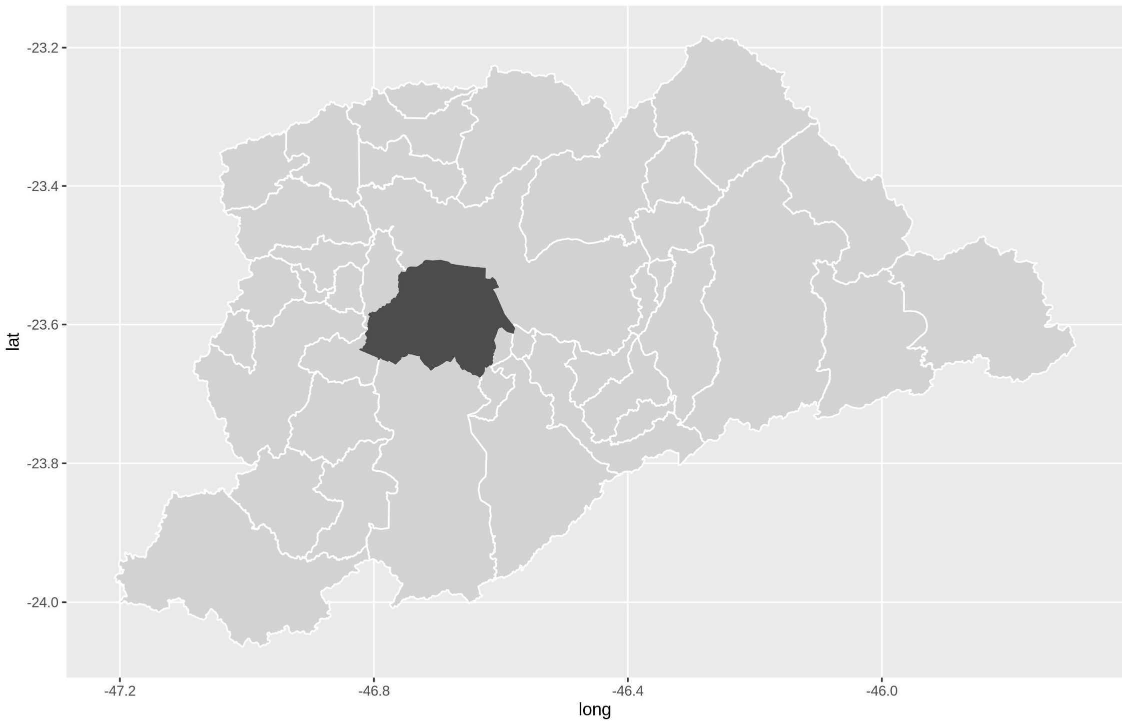
2010 Brazilian census data is the main data source of this research. Census data on racial composition of population in each census tract is used in order to calculate the indices above. Open Street Maps' Open Trip Planner is the source of travel time and Brazilian 2010 school census (INEP) as well as the SPMA 2012 schools shapefile by Centro de Estudos da Metrópole (CEM) are used in order to calculate school segregation.

5.2.1 Region of interest

Because of computational limitations in calculating multiple spatial indices, this research is restricted to a smaller area. Center-Western SPMA (CW SPMA) was chosen since it comprises most of the population living in high-high clusters (analysis in section 4.4): 140,258 people out of 211,432 with racial data in the entire SPMA, corresponding to 66.34%. This is the best choice considering the objective of exploring measures to capture micro-segregation. Figure 26 highlights (in dark grey) the location of the chosen area of interest in SPMA, and Figure 27 presents the census tracts boundaries in CW SPMA, color-filled by district.

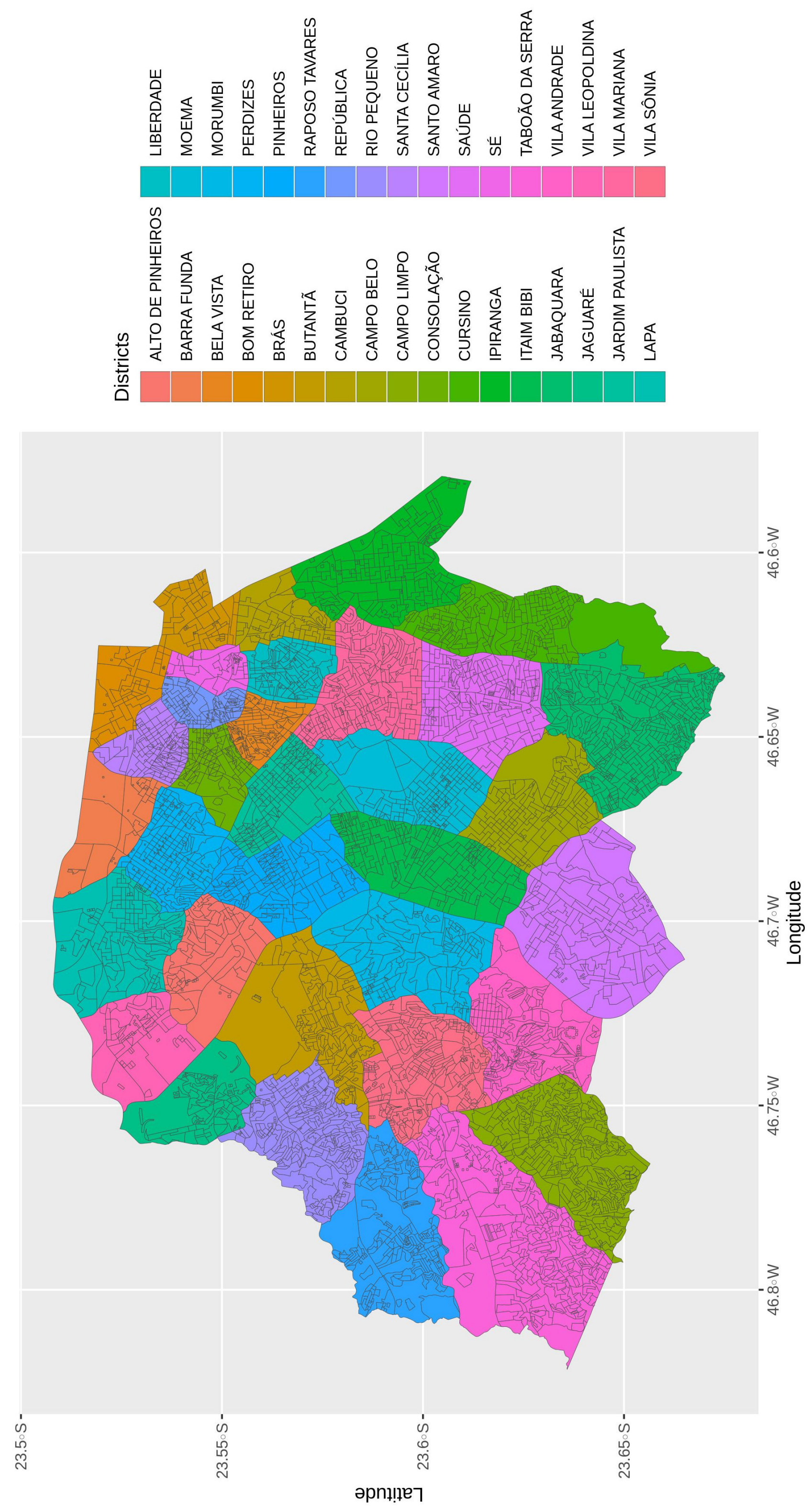
This area is composed by 5,717 census tracts, 5,585 of which have racial data available, corresponding to 180 weighting areas and 34 districts: 33 within the municipality of São Paulo and one district in Taboão da Serra. According to 2010 census data, 2,963,543 people live in this area and there is racial data available to 2,962,575 of them – the equivalent to 99.97%. Table 7 presents the statistical summary of racial data in SPMA.

Figure 26 – CW SPMA – Location in SPMA
Municipalities' political boundaries



Source: IBGE (2010). Elaborated by the author.

Figure 27 – Area of Interest – census tracts' boundaries, colored by district



Source: IBGE (2010). Elaborated by the author.

Table 7 – Statistical summary of Center-Western SPMA residents by racial group – sum by census tract

	Min	Median	Mean	Max	NAs	SD	N
Blacks (Pretos)	0.00	13.00	24.68	386.00	132	32.46	5717
Browns (Pardos)	0.00	50.00	110.16	2131.00	132	158.77	5717
Whites (Brancos)	1.00	351.00	370.60	1834.00	132	179.60	5717
Yellows (Amarelos)	0.00	14.00	24.36	480.00	132	35.24	5717
Native Brazilians (Indígenas)	0.00	0.00	0.65	109.00	132	2.95	5717

Souce: IBGE (2010). Elaborated by the author.

5.2.2 Defining proximity function Φ_p

In order to establish relationship between census tracts (proximity function), three variables were used: euclidean distance between census tracts' centroids and time of travel in two modals: walking and by car. Public transportation could not be used because two municipalities were involved and not only the city of São Paulo, which would require specific files with public transportation time tables of both municipalities and São Paulo state public transportation system, as required by the source here adopted: Open Street Planner. Travel time matrices were obtained from Open Street Planner API, following Pereira and others (2010).[2] IBGE 2010 census tracts shapefile[3] was used in order to calculate census tracts' area and centroid. Centroids were defined as the mass centroid for most of the polygons. However, a random point inside the polygon replaced its centroid in cases in which the mass centroid fell outside the polygon (concave polygons) – 1326 out of 29646 polygons.

Two Gaussian distance decay functions were adopted: with a 500-m and with a 4-km bandwidth, for micro- and macro-segregation, respectively. Gollini and others' formula for the Gaussian curve was adopted (GOLLINI et al., 2013), and bandwidths ranges were chosen following Lee et al. (2008). As references display functions considering distance in space and not in time (*e.g.* travel time), those functions were adapted. Time decay functions were calculated as Gaussian, but with bandwidths of 720 seconds (equivalent to 12 minutes) for short range segregation and 2700 seconds (equivalent to 45 minutes) for long range – taking into account that a person hardly ever walks further than that in a single travel on a daily basis in an urban area such as SPMA.[4] The same bandwidths were adopted for car travels to guarantee more consistency in the analysis of how segregation changes in an urban area depending on

[2] Pereira, R. H. M.; Grégoire, L.; Wessel, N.; Martins, J. (2019). Tutorial with reproducible example to estimate a travel time matrix using OpenTripPlanner and Python. Retrieved from https://github.com/rafapereirabr/otp-travel-time-matrix. doi:10.5281/zenodo.3242134.

[3] Retrieved from IBGE's geoftp webpage. Downloaded in Dec. 07, 2018.

[4] One would rather change the modal (public transportation, car or bicycle) unless (s)he has not the financial conditions.

the transportation means that are made available. Table 8 presents the statistical summary of euclidean distance, travel time and area of census tracts in the chosen area of interest.

The number of observations in euclidean distance data corresponds to the square of the number of polygons in the area of interest (5717^2), because it is composed by the distances of all polygons to all polygons. Travel time matrices follow the same rationale, but there are some missing polygons that could not be found using Open Street Planner API. Those were 203,476 missing values in walking travel time matrix and 607,771 in travel time by car matrix – corresponding to 0.62% and 1.86% of the 32,684,089 (or 5717^2) total observations that would compose those matrices if there were no missing travels, respectively. Those are the only missing data in the geographic components of the Spatial Entropy.

Table 8 – Geographic components of spatial entropy – statistical summary

	Min	Median	Mean	Max	SD	N
Euclidean distance (km)	0.00	8.53	8.73	23.93	4.38	32684089
Walking travel time (s)	0.00	8125.00	8256.27	20323.00	3969.73	32480613
Travel time by car (s)	0.00	1627.00	1635.20	4408.00	648.33	32076318
Tracts' area (sq. meters)	395.75	23572.16	44167.88	3877591.74	112349.54	5717

Data source: IBGE (2010) and Open Street Planner (2019). Elaboration by the author.

5.3 Testing robustness

In order to assess which indices better measure social distance, residential indices will be compared with school segregation because it is able to translate real social contact between individuals, at least in a part of their lifes. Following Fernandes (2017) , school segregation will be used as a proxy for social distance. Schools really are an environment of sociability, which does not necessarily apply to the streets or other public spaces of the neighborhood where one resides. Thus, school segregation will be triangulated with the other residential segregation indices and comparison will allow checking which index might be a better choice for measuring social separation.

School segregation is calculated using D. Race of students is available at the school level (number of enrollments in each school by racial group) and schools are geocoded as points, allowing D to be computed for any other upper level. Thus, school D was calculated for the weighting areas. The choice of computing school D for the census tract level (lower than the weighting area – see Figure 25) was rejected because only 1,943 out of 29,943 census tracts had at least two georeferenced schools in 2010, corresponding to 6.49% of them. It makes the analysis of school segregation at the tract level impossible. In contrast, there is only one weighting area with less than two schools out of 633 in the entire SPMA, making it possible to calculate school D in 99.8% of these geographic units.

D was chosen because one hypothesis of this work is the resemblance between school segregation and residential segregation calculated with $\tilde{E}_p$, in contrast to a smaller correlation between school and residential segregation calculated with D. Thus, if this hypothesis is confirmed, one cannot argue this resemblance is due to the chosen measure – because measure here is favouring resemblance of residential and school segregation both calculated with D. Finally, varying proximity functions in the calculation of $\tilde{E}_p$ is also a way of testing robustness and to avoid the grid problem.

5.3.1 School data

There were 5,278,051 school enrollments in SPMA in 10,603 schools in 2010. There are 10,937 schools in SPMA in CEM 2013 georeferenced school data base, and 9,842 of them are in 2010 school census.[5] Thus, this is the number of schools in SPMA in 2010 that could be georeferenced and were considered in segregation analysis, corresponding to 92.82% of the schools and 97.79% of the enrollments in SPMA in 2010.

There were 699,295 enrollments in schools located within CW SPMA in 2010, 13.55% of SPMA georeferenced schools enrollments. Number of public and private schools in the area of interest (CW SPMA) and in other census tracts of SPMA is displayed in Table 9.

Table 9 – Number of private and public schools in CW SPMA and other areas

	Public	Private
Area of interest	589	960
Other	4,930	3,363

Data source: IPEA (2010) and CEM (2013). Elaboration by the author.

Table 10 – Number of enrollments by racial group in CW SPMA and other areas

	Black	White	Brown	Yellow	Native Brazilian	Not declared	Total
Other	3.2%	37.8%	21.93%	0.47%	0.24%	36.37%	100% or 4462271
Area of interest	2.73%	41.04%	16.48%	1.17%	0.22%	38.36%	100% or 699295

Data source: IPEA (2010) and CEM (2013). Elaboration by the author.

Relative frequency of enrollments with no race declared is slightly higher in CW SPMA in comparison with the entire metropolitan area. Even though, the 478,635 enrollments with no race declared in the regions of interest correspond to less than 0.01% of all enrollments in the area – a negligible amount.

[5] This is expected with the creation of new schools from 2010 to 2013.

6 RESULTS

6.1 Statistical and geographic description of segregation indices

Residential segregation was calculated using both Dissimilarity Index at the district and weighting area level, and Spatial Entropy – the local component of the Information Theory Index – with varying proximity functions. Table 11 presents the statistical summary of residential segregation indices in the area of interest.

School segregation was calculated using the Dissimilarity Index (D) for district and weighting area level. D was calculated for the area of interest using all schools data, and private and public schools data, separately. Statistical summary is displayed in Table 12.

In both tables, short-range indices are more disperse than their long-range counterparts. For instance, Spatial Entropy calculated with euclidean distance and a 500-m bandwidth and 4-km bandwidth, respectively. This variation cannot be observed in residential and public schools D, although this could be expected for this measure for one more reason: smaller scale in this case implies smaller geographic units for which D is calculated and thus, a larger number of observations: 180 weighting areas *versus* 34 districts.

In addition, higher levels of segregation – higher D and lower $\tilde{E}_p$ – were expected for the short-range indices, in comparison to their long-range counterparts. This is true for all $\tilde{E}_p$ pairs means and medians of the long- and short-range version of the same index, although a very small difference can be noticed in Spatial Entropy means calculated with travel time by car. It is also worth mentioning the very low variation in the last Spatial Entropy displayed in Table 11. This is due to the large bandwidth adopted regarding the size of the chosen area of interest: the time-decay function with a 45-minute bandwidth (in travel time by car) makes weights of census tract very similar in the calculation of Entropies. Hence, the concentration of Entropy values in this case at higher levels (low segregation) in comparison with Entropy calculated with walking travel time and the same bandwidth suggests the importance of transportation in connecting or separating people in urban space.

In terms of school segregation, D presents higher means and medians for the district level in comparison with weighting area level, since a district comprises higher diversity of schools without changing within school diversity, the chances of having higher segregation levels increase (Table 12).

Table 11 – Statistical summary of residential segregation indices – Center-Western SPMA

	Min	Median	Mean	Max	NAs	SD	N
Dissimilarity Index – Weighting area level	0.09	0.28	0.29	0.58	0	0.10	180
Dissimilarity Index – District level	0.15	0.30	0.32	0.51	0	0.10	34
Spatial Entropy – Euclidean distance (500-m band.)	0.28	0.68	0.68	1.00	0	0.24	5585
Spatial Entropy – Euclidean distance (4-km band.)	0.58	0.72	0.75	0.99	0	0.11	5585
Spatial Entropy – Walking travel time (12-min band.)	0.29	0.71	0.69	1.00	0	0.23	5585
Spatial Entropy – Walking travel time (45-min band.)	0.42	0.73	0.75	1.00	0	0.14	5585
Spatial Entropy – Travel time by car (12-min band.)	0.23	0.73	0.76	1.00	0	0.11	5585
Spatial Entropy – Travel time by car (45-min band.)	0.23	0.77	0.77	1.00	0	0.03	5585

Data source: IBGE (2010) and OSP (2019). Elaborated by the author.

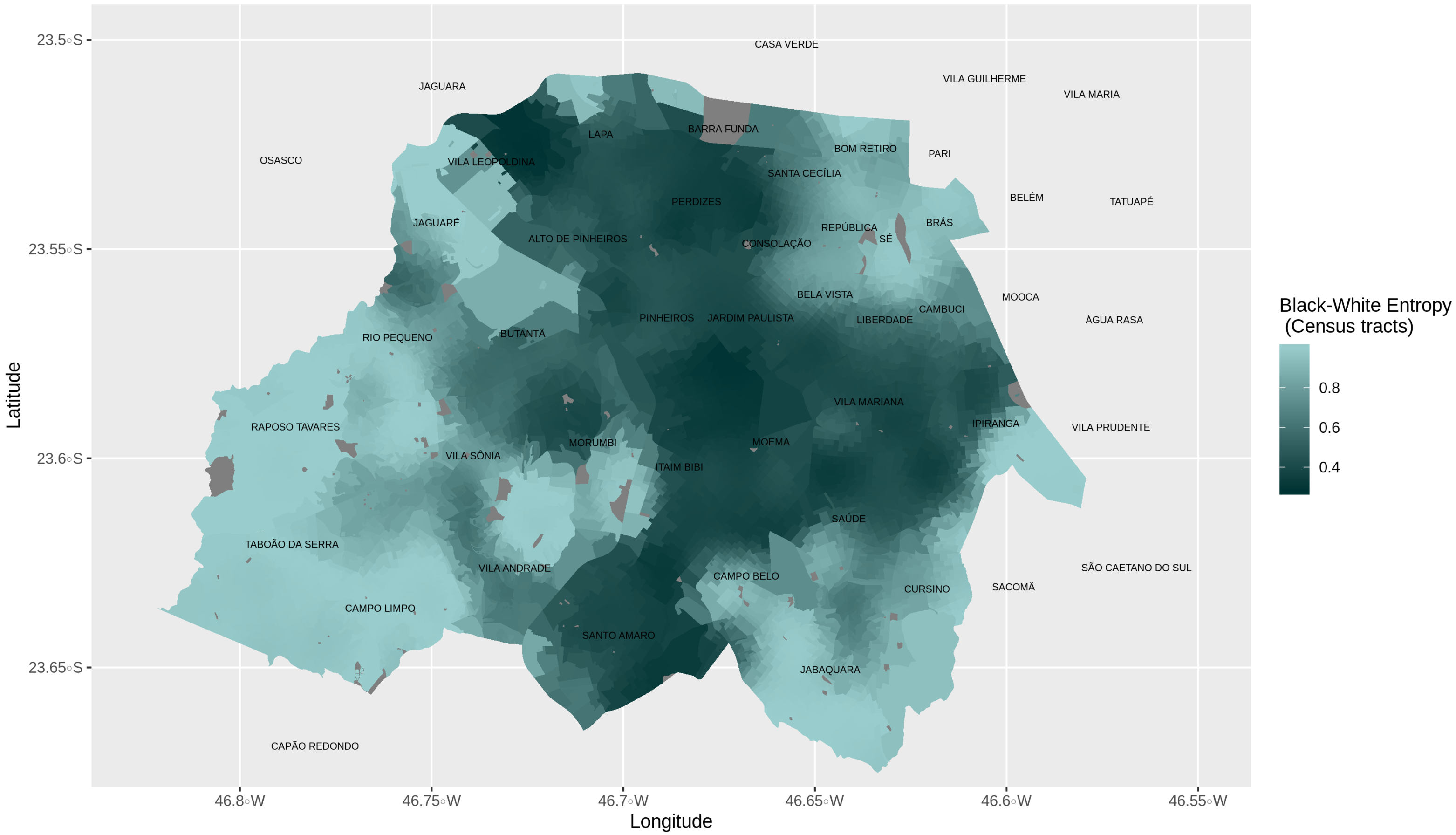
Table 12 – Statistical summary of school segregation – Center-Western SPMA

	Min	Median	Mean	Max	NAs	SD	N
All schools – District	0.13	0.44	0.42	0.70	0	0.15	34
All schools – Weighting area	0.05	0.23	0.29	0.75	0	0.18	180
Public schools – District	0.04	0.15	0.15	0.30	0	0.05	34
Public schools – Weighting area	0.00	0.10	0.10	0.25	5	0.05	180
Private schools – District	0.09	0.48	0.51	0.76	0	0.16	34
Private schools – Weighting area	0.00	0.36	0.35	0.77	0	0.18	180

Data source: INEP (2010) and CEM (2013). Elaborated by the author.

Regarding the geographic distribution of the spatial indices, short-range indices display more intense discontinuities in comparison to their long-range counter parts. Also, those areas of sharp discontinuities present an interesting pattern when it consists on a clear boundary between a rich and a poor district: the richer areas present low levels of Entropy (high segregation) and poorer areas present high Entropy levels (low segregation). For instance, Pinheiros (high income) and Jaguaré (low income); or Morumbi (high income) and Vila Andrade (low income), as shown in Figure 28. Other spatial entropy maps are presented in the Appendix B.

Figure 28 – Spatial entropy in CW SPMA – Neighborhood relationship between census tracts established with the Gaussian function of euclidean distance and a 500-meter bandwidth



Data source: IBGE (2010). Elaboration by the author.

6.2 Confronting micro-segregation measures

As previously mentioned, micro-segregation will be the focus of analysis from now on. Thus, short-range indices will be compared. In order to facilitate comparison, Entropy is adjusted in a way that D is compared with $(1 - Entropy)$. Thus, the higher the value, the more segregated is the geographic unit for both measures. In addition, D – computed to the weighting area level – was inputted to all census tracts composing the weighting area in all following analysis.

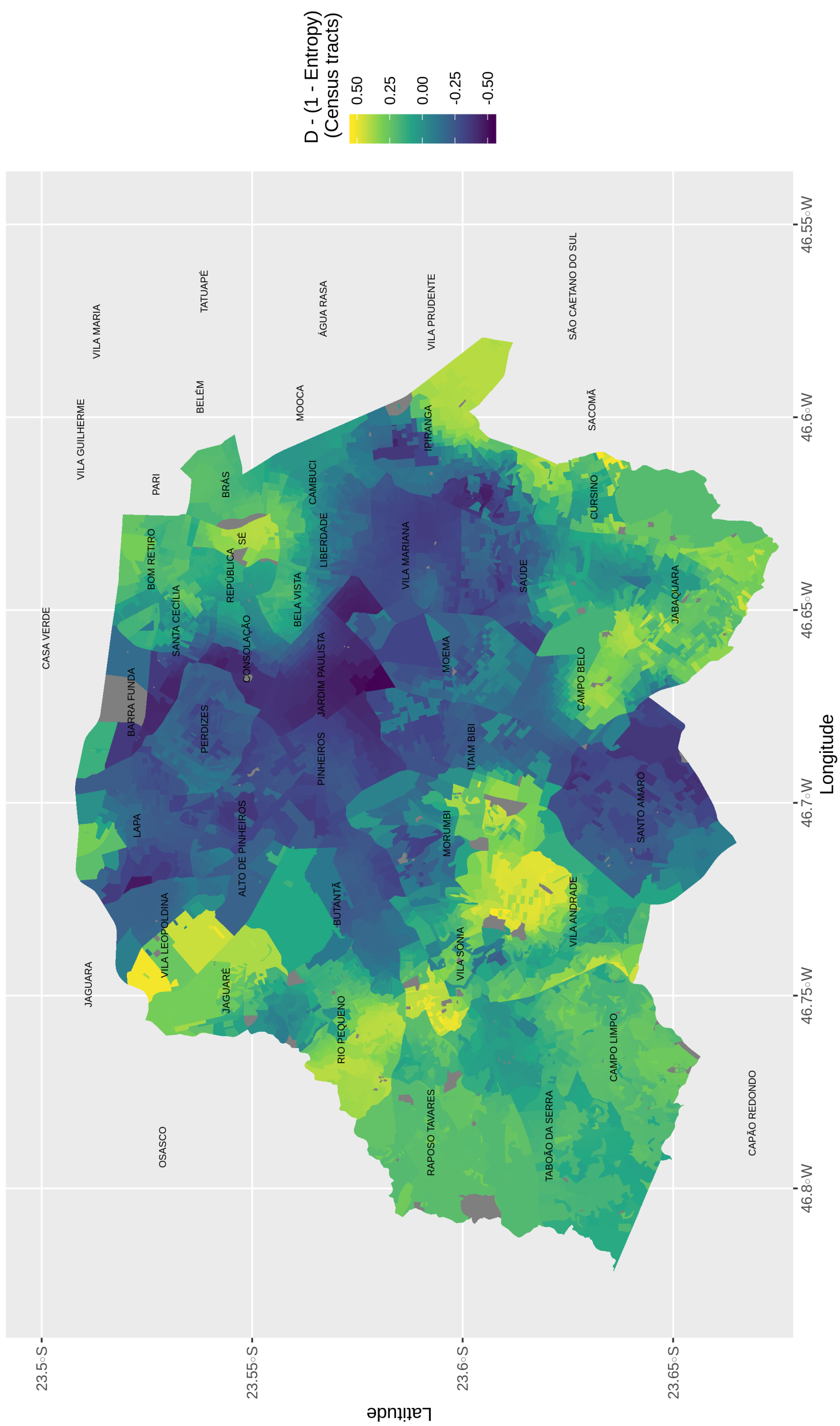
In order to compare the sensibilities of each index and how they differ from the most used measure – the residential D –, three maps were elaborated with the difference between residential D and the three short-range entropy indices – calculated with euclidean distance (500-m bandwith), walking travel time (12-min bandwith) and travel time by car (12-min bandwith). Maps in Figures 29, 30 and 31 display $D - (1 - Entropy)$, with entropies calculated with the above mentioned proximity functions, respectively. Thus, higher values in those maps indicate that D is higher than $(1 - Entropy)$; inversely, negative values indicate that $(1 - Entropy)$ is higher than D; zero indicates that those measures have equal values in that area.

Differences are sharper in the first two maps – that are actually very similar. However, natural barriers are more marked in the second map than the first, such as Pinheiros river, that cross the map from North to South, contrasting the dark purple right side with the greener left side. This was expected since Entropy was computed using walking travel time in the second map, more realistic about the real cost or distance between two individuals in the urban space. This marks were already present in the original Entropy map (Figure 32 in the Appendix).

In addition, difference in both first two maps is negative in rich neighborhoods, indicating $(1 - Entropy)$ is relatively higher in rich neighborhoods than is D. Analogously, difference is positive in poorer areas, thus D is relatively higher than $(1 - Entropy)$ in those areas. This happens because the same value of D represents the same level of segregation within each geographic unit, which doesn't give any insights on between-units segregation, or about the racial composition of each unit. Entropies, on the other hand, by considering all neighboring tracts in the calculation of a single tract entropy, allow a better comparison of segregation levels between geographic units. Thus, D as a segregation measure jeopardizes further insights on segregation, considering the urban micro-segregation pattern in CW SPMA, marked by some very white and rich neighborhoods close to racially-mixed poorer areas. Indeed, D being able to capture this pattern will depend on whether those census tracts fall in the same weighting area or not.

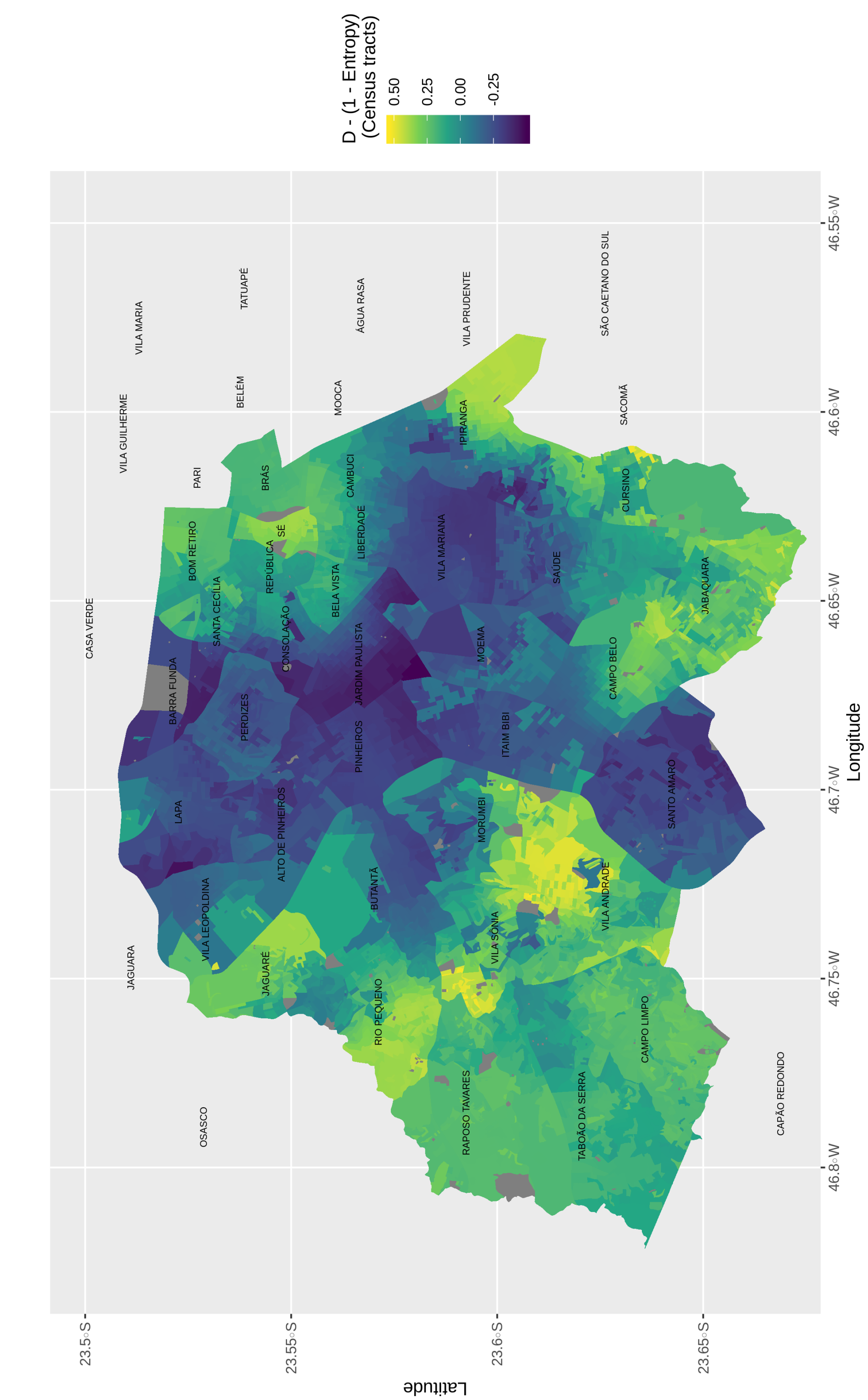
Differences in the third map are smoother than the other two because the original distribution of Entropy calculated with travel time by car is smoother (Figure 34). Again, this shows the potential of transportation in reducing distances and thus, segregation in the metropolis.

Figure 29 – Difference between residential dissimilarity index and spatial entropy – Euclidean distance with a 500-meter bandwidth



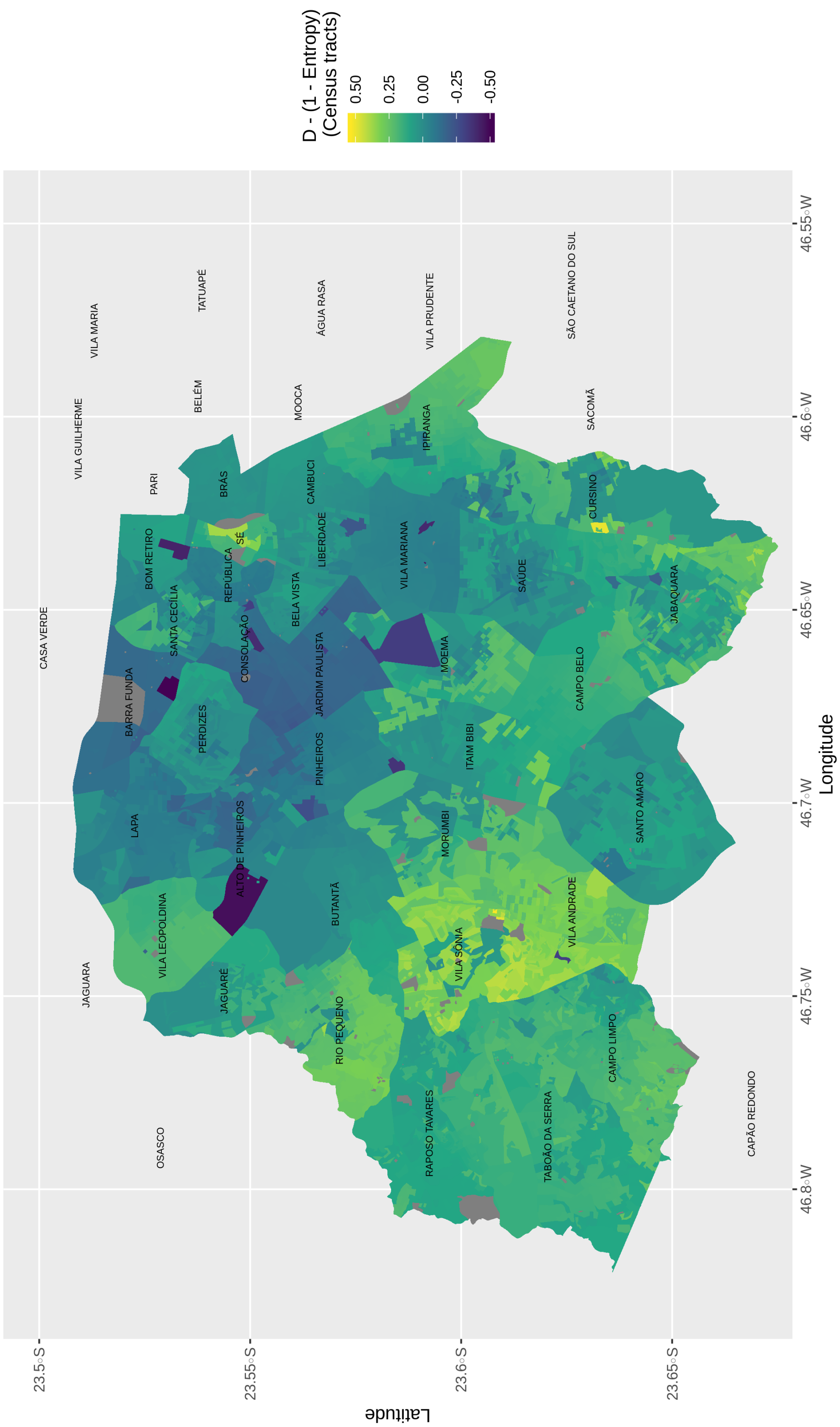
Data source: IBGE (2010). Elaboration by the author.

Figure 30 – Difference between residential dissimilarity index and spatial entropy – Walking travel time with a 12-min bandwidth



Data source: IBGE (2010) and OSP (2019). Elaboration by the author.

Figure 31 – Difference between residential dissimilarity index and spatial entropy – Travel time by car with a 12-min bandwidth



Data source: IBGE (2010) and OSP (2019). Elaboration by the author.

6.3 Determinants of micro residential segregation

In order to better describe the measures here discussed, linear models predicting the residential micro-segregation measures were analysed. Partial correlations between different measures of residential segregation (Y) and other relevant variables (X) are tested by a linear regression as follows:

$$Y = \alpha + \beta X + \epsilon$$

Total residents is a control for the size of the census tract. Income variables were considered because of the large literature on the conterminous nature of race and class in Brazil and the previous characterization of SPMA urban space. Percentage of households with a *per capita* income of less than 1 minimum wage was omitted to avoid multicollinearity. Finally, racial composition of census tract, noticeably the percentage of residents in the census tract that are white (*brancos* and *amarelos*) was considered, since the previous results qualitatively indicate a pattern of segregation that isolates one specific racial and income group (Chapter 4). Table 13 shows the linear models of residential segregation measures through the four short-range indices before mentioned.

Results show the coefficient of Percentage of households 1-3 m.w. is consistently negative, although it is not significant in models number (2) for the three entropies' subtables. In those subtables, it becomes significant in model (3) when the percentage of white and yellow residents in census tract is included. An increase of 1 percentage point in the census tract percentage of households with a monthly *per capita* income ranging from 1 to 3 minimum wages reduces $(1 - Entropy)$ in 0.0023 to 0.0025, on average. Thus, the presence of this medium-low income stratum is positively correlated with racial diversity.

In contrast, higher income strata are consistently correlated with segregation when it is calculated using Spatial Entropy. Coefficients of the Percentage of households with *per capita* income of 10 minimum wages or more and Percentage of households with *per capita* income of 3-10 minimum wages have their values reduced from model (2) to (3) in the three Entropies subtables. Percentage of white and yellow residents seems to capture the effect of tracts' racial composition that is conterminous to income and was being captured by the income variables in models (2).

This is not the case when segregation is calculated using D. Percentage of households with a *per capita* income of 10 minimum wages or more looses significance in the 3$^{\text{rd}}$ model of the first subtable in Table 13. Indeed, this pattern is less clear when segregation is calculated using D, because D can assume low values if a geographic unit population is evenly distributed within it, although it is very uneven in comparison with its neighbors, as discussed in section 6.2. Thus, an entirely black or white weighting area would have low D, even though it is neighboring a racially diverse weighting area.

Finally, the percentage of white and yellow residents is consistently and positively correlated with segregation, corroborating with previous findings. A marginal increase in the percentage of white and yellow population of the census tract results in an increase of approximately 0.006 points in Entropy (in the three cases presented in Table 13) and 0.001 in residential Dissimilarity Index.

Table 13 – Determinants of residential segregation

Residential D						
Term	(1)	sig.	(2)	sig.	(3)	sig.
Total residents	-0.00	***	-0.00	***	-0.00	***
Perc. households 1-3 m.w.			-0.06	***	-0.11	***
Perc. households 3-10 m.w.			0.05	***	-0.04	***
Perc. households 10 m.w. or more			0.07	***	-0.01	.
Perc. white and yellow residents					0.13	***
$(1 - Entropy)$ **with euclidean distance**						
Term	(1)	sig.	(2)	sig.	(3)	sig.
Total residents	-0.00	***	0.00	.	0.00	***
Perc. households 1-3 m.w.			-0.03	.	-0.25	***
Perc. households 3-10 m.w.			0.60	***	0.15	***
Perc. households 10 m.w. or more			0.61	***	0.24	***
Perc. white and yellow residents					0.61	***
$(1 - Entropy)$ **with walking travel time**						
Term	(1)	sig.	(2)	sig.	(3)	sig.
Total residents	-0.00	***	-0.00	.	0.00	**
Perc. households 1-3 m.w.			-0.02	.	-0.23	***
Perc. households 3-10 m.w.			0.56	***	0.12	***
Perc. households 10 m.w. or more			0.59	***	0.23	***
Perc. white and yellow residents					0.58	***
$(1 - Entropy)$ **with travel time by car**						
Term	(1)	sig.	(2)	sig.	(3)	sig.
Total residents	-0.00	***	-0.00	.	0.00	**
Perc. households 1-3 m.w.			-0.02	.	-0.23	***
Perc. households 3-10 m.w.			0.56	***	0.12	***
Perc. households 10 m.w. or more			0.59	***	0.23	***
Perc. white and yellow residents					0.58	***

Significance levels: *** P $\leq$ 0.01, ** P $\leq$ 0.05.

6.4 Comparing school and residential segregation indices

School segregation calculated for the weighting area – a measure that seems to reflect social contact between groups – is here compared to all residential indices under analysis. Table 14 presents the correlation between the short-range indices. School D at the weighting area is less correlated to residential

D than it is to all Spatial Entropy measures. Even though, school segregation correlations range within what is considered a moderate correlation interval.

Table 14 – Correlation matrix of short-range segregation indices

	School D	Entropy – euclid. dist.	Entropy – walking	Entropy – car	Residential D
School D	1.00	-	-	-	-
Entropy – euclid. dist.	0.45	1.00	-	-	-
Entropy – walking	0.46	0.98	1.00	-	-
Entropy – car	0.44	0.76	0.78	1.00	-
Residential D	0.34	0.21	0.22	0.18	1.00

Source: Elaboration by the author.

Table 15 presents the linear models of school segregation varying the independent variable of interest in each model: the four residential micro-segregation indices explored by this research. Linear models can be represented by the following equation:

$$SchoolSeg = \alpha + \beta_1 ResidentialSeg + \beta_2 X + \epsilon$$

Where X is the set of control variables. Total residents and students are controls of geographic unit size and have coefficients that are consistently very close to zero. The number of private and public schools are controls for size and school composition of the weighting area. These variables also have coefficients with values that are very close to zero, although they assume positive and negative values in all models, respectively.

Percentage of households with *per capita* household income of 1-3 and 3-10 minimum wages are also negatively correlated with school segregation in all models, except model (2) in the first subtable in Table 15. Percentage of households 10 m.w. or more is significant and positively correlated to school segregation in all models (2) and (3).

The Percentage of households in the lowest income strata (1-3 m.w.) coefficient has absolute values that consistently increase when control for racial composition is included in model (3) in all four subtables. In contrast, the partial correlation of the percentage of households in the highest strata with school segregation looses strength when Percentage of white students is included in model (3) in all four subtables. This strengthening and weakening of partial correlations when race is controlled, respectively, indicates how important is income for the choice for segregation. When race is controlled, the presence of lower income families in a geographic unit reduces segregation more intensely, while higher income households contributes less intensely to the increase of segregation if compared to the previous model (2). This might be related to what has been already described about the Brazilian public school system rules and the revealed preference towards private schools. Meanwhile, the Percentage of households with 3-10 m.w. *per capita* income coefficient remains unchanged when control for tract racial composition is added in all three entropies subtables (models (3)).

Finally, the Percentage of white students is the variable with the strongest partial correlation with school segregation (approximately 0.70 in all four models (3)). A marginal increase – of one percentage point – in this variable, corresponds to an expected increase of 0.007 in school segregation, corroborating with the previously described in Chapter 4.

All residential segregation measures are positively correlated with school segregation in models (1) and (2), but Entropies loose significance or change signal in the third model, when racial composition of student body in the weighting area is included in the model, while residential D is consistently significant and positively correlated with school D. At first glance, this partial correlation analysis contrasts with the matrix correlations analysis presented in Table 14, where residential D presents the lowest correlation with school D among all short-range residential indices.

However, results in Tables 15 and 14 are complementary. Spatial Entropy is a measure that extrapolates the census tract racial composition, and computes difference in a census tract diversity in comparison to its neighbors. Since areas with less diversity compared to its surroundings tend to be areas mostly inhabited by whites – as shown in previous descriptions and in partial correlations analysis in Table 13 – it is not surprising that variation in Entropies is dropped from the model when the Percentage of white students is included.

Thus, residential $\tilde{E_p}$ seems to better measure social distance than residential D in the SPMA context. It was expected because of all limitations regarding an aspatial index, in comparison to a spatial measure. Even though, these results can serve as an important argument in favour of better measures. In addition, although school is an important part of an individual's life and socialization, it is not all of it. The place of residence, on the other hand, is part of an individual's life throughout all of it. It can reveal the magnitude of social contact between racial groups in domestic settings. The most important theoretical limitation of this approach, however, is the increasing importance of other spaces of sociability – work and leisure – in a metropolitan context. Thus, a more comprehensive investigation should take into account other data, such as workplace racial composition, public and private leisure spaces, families and social network.

Table 15 – Determinants of school segregation

	(1)	sig.	(2)	sig.	(3)	sig.
Residential D	0.28	***	0.22	***	0.26	***
Number of private schools	0.01	***	0.01	***	0.00	**
Number of public schools	-0.03	***	-0.03	***	-0.02	***
Total residents	-0.00	***	-0.00	***	-0.00	***
Total students	0.00	***	0.00	***	0.00	***
Perc. households 1-3 m.w.			-0.05	***	-0.08	***
Perc. households 3-10 m.w.			0.01	.	-0.05	***
Perc. households 10 m.w. or more			0.25	***	0.11	***
Perc. white students					0.72	***

	(1)	sig.	(2)	sig.	(3)	sig.
$(1 - Entropy)$ **with euclidean distance**	0.17	***	0.12	***	0.01	.
Number of private schools	0.01	***	0.01	***	0.00	***
Number of public schools	-0.03	***	-0.03	***	-0.02	***
Total residents	-0.00	***	-0.00	***	-0.00	***
Total students	0.00	***	0.00	***	0.00	***
Perc. households 1-3 m.w.			-0.06	***	-0.10	***
Perc. households 3-10 m.w.			-0.05	***	-0.05	***
Perc. households 10 m.w. or more			0.19	***	0.12	***
Perc. white students					0.70	***

	(1)	sig.	(2)	sig.	(3)	sig.
$(1 - Entropy)$ **with walking travel time**	0.19	***	0.13	***	0.02	.
Number of private schools	0.01	***	0.01	***	0.00	***
Number of public schools	-0.03	***	-0.03	***	-0.02	***
Total residents	-0.00	***	-0.00	***	-0.00	***
Total students	0.00	***	0.00	***	0.00	***
Perc. households 1-3 m.w.			-0.06	***	-0.10	***
Perc. households 3-10 m.w.			-0.05	***	-0.05	***
Perc. households 10 m.w. or more			0.19	***	0.11	***
Perc. white students					0.69	***

	(1)	sig.	(2)	sig.	(3)	sig.
$(1 - Entropy)$ **with travel time by car**	0.35	***	0.22	***	-0.09	***
Number of private schools	0.01	***	0.01	***	0.00	***
Number of public schools	-0.03	***	-0.03	***	-0.02	***
Total residents	-0.00	***	-0.00	***	-0.00	***
Total students	0.00	***	0.00	***	0.00	***
Perc. households 1-3 m.w.			-0.07	***	-0.09	***
Perc. households 3-10 m.w.			-0.03	**	-0.03	**
Perc. households 10 m.w. or more			0.21	***	0.13	***
Perc. white students					0.73	***

Significance levels: *** P $\leq$ 0.01, ** P $\leq$ 0.05.

7 CONCLUSION

This work aims to methodologically contribute to the measurement of micro-segregation in metropolitan areas. The most popular index – the Dissimilarity Index (D) – is compared with Spatial Entropy, calculated with varying proximity functions in order to identify racial diversity in census tracts. Proximity was computed with both euclidean distance and travel time. Thus, one innovation here proposed is the incorporation of travel time – a more accurate proxy for the distance between two points or individuals in an urban area – in the calculation of micro-segregation indices.

In this regard, one limitation of this research is the difference in dates of demographic census (2010) and travel time data (2019). A more complete analysis would also take into account multi-modal travels including public transportation. Another payed API can provide such data with high quality and precision, but was too expensive for the scope of this research.

Results show that Spatial Entropy calculated with travel time – specially walking travel time, in this case where area of interest was relatively small – is more sensible to urban barriers or facilitators, such as rivers, avenues and bridges. In addition, transportation means that are made available can reduce segregation.

Computational limitations made it necessary to reduce the scale of the analysis to the chosen Center-Western area of São Paulo Metropolitan Area (SPMA). The entire SPMA contains approximately 30,000 census tracts and would require this number squared in mathematical operations in order to compute the neighborhood matrix. This would translate in more than one week worth of computer processing for one single matrix in a personal computer. A deeper analysis, comparing macro- and micro-segregation measures and its components (LEE et al., 2008) would be interesting and possible with a more powerful hardware or a more efficient coding strategy, which could be developed with the support of a programmer.

All spatial indices applied are more strongly correlated with school segregation (proxy for social distance) than is residential D, even though school segregation itself was computed using D (Table 14). In addition, Spatial Entropies are affected by racial composition of neighboring census tracts, which doesn't happen with the residential D (Table 13). D can assume low values if a geographic unit population is evenly distributed within it, although it is very uneven in comparison with its neighbors. This can jeopardize segregation analysis in a metropolitan context where some neighborhoods or blocks can be very homogeneous within it, but very different from the neighboring blocks, which has been shown not to be an exception in SPMA (Chapter 4).

Micro-segregation in SPMA seems to present itself as an one-tailed segregation pattern resulting from the self-segregation of medium-high- and high-income whites (*amarelos* and *brancos*) – which D might be failing to capture. Spatial Entropy distributions in the area of study are more consistent with the description of social and urban contrasts in SPMA described in Chapter 4. Higher education and income levels of whites allow choices in terms of service provision. Regarding education, the racially homogeneous private schools are a choice of white middle- and upper-class families, since public schools are free of charge and the Brazilian state guarantees universal access to basic education (K-12). Those choices are extended to the place of residence as well. Areas with low Entropy levels (high segregation) are areas with mostly high-income and white population. Those rich white neighborhoods are also marked by the existence of private condominiums well equipped with security infrastructure.

The descriptive analysis developed by this research might be a first step towards testing a white flight *à la brésilienne*. In a country of deep social inequalities, social separation from undesirable groups can be built through market mechanisms or state neglect, with no need for explicitly racist policies. Not randomly, Latin American studies on segregation focus on socioeconomic, rather than racial segregation. Hence, the description of the micro-segregation patterns in the largest Latin American Metropolitan Area is also a theoretical contribution to the characterization of this phenomenon that has already been well described in a qualitative manner.

BIBLIOGRAPHY

ABRAMO, P. A cidade com-fusa: A mão inoxidável do mercado e a produção da estrutura urbana das grandes metrópoles latino-americanas. *Revista Brasileira de Estudos Urbanos e Regionais (RBEUR)*, Associação Nacional de Pós Graduação e Pesquisa em Planejamento Urbano e Regional, v. 9, n. 2, p. 25–54, 2007.

ANSELIN, L. Local indicators of spatial association—LISA. *Geographical analysis*, Wiley Online Library, v. 27, n. 2, p. 93–115, 1995.

BALLION, R. Les familles et le choix du collège. *L'orientation scolaire et professionnelle*, v. 15, n. 3, p. 183–202, 1986.

BARBUJANI, G. Human races: classifying people vs understanding diversity. *Current Genomics*, Bentham Science Publishers, v. 6, n. 4, p. 215–226, 2005.

BARROS, J.; FEITOSA, F. F. Uneven geographies: Exploring the sensitivity of spatial indices of residential segregation. *Environment and Planning B: Urban Analytics and City Science*, SAGE Publications Sage UK: London, England, v. 45, n. 6, p. 1073–1089, 2018.

BARTH, F. Introduction. In: BARTH, F. (Ed.). *Ethnic groups and boundaries: The social organization of culture difference*. Long Grove, Illinois: Waveland Press, 1998. p. 9–38.

BELL, W. A probability model for the measurement of ecological segregation. *Social Forces*, The University of North Carolina Press, v. 32, n. 4, p. 357–364, 1954.

CALDEIRA, T. P. *City of walls: crime, segregation, and citizenship in São Paulo*. [S.l.]: Univ. of California Press, 2000.

CALDEIRA, T. P. Qual a novidade dos rolezinhos? Espaço público, desigualdade e mudança em São Paulo. *Novos estudos CEBRAP*, scielo, p. 13 – 20, 03 2014. ISSN 0101-3300.

CARVALHO, I. M. M. de; BARRETO, V. S. Segregação residencial, condição social e raça em Salvador. *Cadernos Metrópole*, Pontifícia Universidade Católica de São Paulo, n. 18, p. 251–273, 2007.

CASTELLS, M. *The urban question: a Marxist approach*. [S.l.]: Edward Arnold, 1972.

DUNCAN, O. D.; DUNCAN, B. A methodological analysis of segregation indexes. *American sociological review*, JSTOR, v. 20, n. 2, p. 210–217, 1955.

FACINA, A. "não me bate doutor": funk e criminalização da pobreza. *V ENECULT-Encontro de Estudos Multidisciplinares em Cultura, Salvador. Anais do V ENECULT-Encontro de Estudos Multidisciplinares em Cultura*, 2009.

FARBER, S. et al. Measuring segregation using patterns of daily travel behavior: A social interaction based model of exposure. *Journal of Transport Geography*, Elsevier, v. 49, p. 26–38, 2015.

FEITOSA, F. F. *Índices espaciais para mensurar a segregação residencial: o caso de São José dos Campos (SP)*. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2005.

FEITOSA, F. F. Urban segregation as a complex system: an agent-based simulation approach. *Ecology and Development Series*, n. 70, 2010.

FERNANDES, G. Is the brazilian tale of peaceful racial coexistence true? some evidence from school segregation and the huge racial gap in the largest brazilian city. *World Development*, Elsevier, v. 98, p. 179–194, 2017.

FRANÇA, D. *Segregação racial em São Paulo: Residências, redes pessoais e trajetórias urbanas de negros e brancos no século XXI*. Tese (Doutorado) — Universidade de São Paulo, 2017.

FRANK, A. I. *Using measures of spatial autocorrelation to describe socio-economic and racial residential patterns in US urban areas*. [S.l.]: Taylor & Francis New York, 2003. v. 13.

FRÚGOLIJR., H. *São Paulo: espaços públicos e interação social*. São Paulo: Marco Zero, 1995.

GARCÍA-LÓPEZ, M. Á.; MORENO-MONROY, A. I. Income segregation and urban spatial structure: evidence from brazil. CAF, 2016.

GOLLINI, I. et al. Gwmodel: an r package for exploring spatial heterogeneity using geographically weighted models. *arXiv preprint arXiv:1306.0413*, 2013.

GRAFMEYER, Y. *Regards sociologiques sur la ségrégation*. [S.l.]: L'Harmattan, 1994.

GRIGORYEVA, A.; RUEF, M. The historical demography of racial segregation. *American Sociological Review*, SAGE Publications Sage CA: Los Angeles, CA, v. 80, n. 4, p. 814–842, 2015.

GUERRA, M. F. *Vende-se qualidade de vida: Alphaville Barueri-implantação e consolidação de uma cidade privada*. Tese (Doutorado) — Universidade de São Paulo, 2013.

HACKING, I. Kinds of people: Moving targets. In: OXFORD UNIVERSITY PRESS. *Proceedings-British Academy*. [S.l.], 2007. v. 151, p. 285.

HARVEY, D. *Rebel cities: From the right to the city to the urban revolution*. [S.l.]: Verso books, 2012.

HERMANN, B. M.; HADDAD, E. A. Mercado imobiliário e amenidades urbanas: a view through the window. *Estudos Econômicos (São Paulo)*, scielo, v. 35, p. 237 – 269, 06 2005. ISSN 0101-4161.

JAMES, F. J. A new generalized "exposure-based" segregation index: Demonstration in Denver and Houston. *Sociological Methods & Research*, Sage Publications, v. 14, n. 3, p. 301–316, 1986.

JARGOWSKY, P. A. Take the money and run: Economic segregation in us metropolitan areas. *American sociological review*, JSTOR, p. 984–998, 1996.

KOWARICK, L. *A espoliação urbana*. [S.l.]: Editora Paz e Terra, 1979. v. 44.

LEE, B. A. et al. Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review*, Sage Publications Sage CA: Los Angeles, CA, v. 73, n. 5, p. 766–791, 2008.

LESSER, J. Negotiating national identity: Middle Eastern and Asian immigrants and the struggle for ethnicity in Brazil. *UC San Diego Working Papers*, n. 8, 2000.

LIEBERSON, S. Measuring population diversity. *American Sociological Review*, JSTOR, v. 34, p. 850–862, 1969.

LOPES, A.; FACINA, A. Cidade do funk: expressões da diáspora negra nas favelas cariocas. *Revista do Arquivo Geral da Cidade do Rio de Janeiro*, Garamond Universitária, v. 6, p. 193–206, 2012.

LOURY, G. *The anatomy of racial inequality*. [S.l.]: Harvard University Press, 2002.

MARCUSE, P. Enclaves yes, ghettos no. *Desegregating the city: Ghettos, enclaves, and inequality*, State University of New York Press Albany^ eNY NY, p. 15–30, 2005.

MARICATO, E. *Metrópole na periferia do capitalismo*. [S.l.]: HUCITEC, 1996.

MARQUES, E. Estrutura social e segregação em são paulo: transformações na década de 2000. *DADOS-Revista de Ciências Sociais*, Universidade do Estado do Rio de Janeiro, v. 57, n. 3, p. 675–710, 2014.

MARQUES, E.; SCALON, C.; OLIVEIRA, M. A. Comparando estruturas sociais no Rio de janeiro e em São Paulo. *Dados-Revista de Ciências Sociais*, Universidade do Estado do Rio de Janeiro, v. 51, n. 1, p. 215–238, 2008.

MARQUES, E.; TORRES, H. *São Paulo: Segregação, pobreza e desigualdades sociais*. [S.l.]: Senac, 2005. 1–336 p.

MASSEY, D. S.; DENTON, N. A. The dimensions of residential segregation. *Social forces*, Oxford University Press, v. 67, n. 2, p. 281–315, 1988.

MORGAN, B. S. The segregation of socio-economic groups in urban areas: a comparative analysis. *Urban Studies*, Sage Publications Sage UK: London, England, v. 12, n. 1, p. 47–60, 1975.

MORGAN, B. S. A distance-decay based interaction index to measure residential segregation. *Area*, JSTOR, p. 211–217, 1983.

NOBLES, M. *Shades of citizenship: Race and the census in modern politics*. [S.l.]: Stanford University Press, 2000.

OPENSHOW, S.; PJ, T. A million or so correlation coefficients, three experiments on the modifiable areal unit problem. *Statistical applications in the spatial science*, Pion, p. 127–144, 1979.

OSÓRIO, R. G. O sistema classificatório de cor ou raça do IBGE. Instituto de Pesquisa Econômica Aplicada (Ipea), 2003.

RAMOS, F. R. *Análise espacial de Estruturas Intra-Urbanas: O caso de São Paulo*. Dissertação (Mestrado) — INPE, São José dos Campos, 2004.

RAMOS, F. R. *Três ensaios sobre a estrutura espacial urbana em cidades do Brasil contemporâneo: economia urbana e geoinformação na construção de novos olhares*. Tese (Doutorado) — Fundação Getulio Vargas, São Paulo, 2014.

RANGEL, M. A. Is parental love colorblind? allocation of resources within mixed families. *XXX Meeting of the Brazilian Econometric Society*, 2008.

REARDON, S. F. A conceptual framework for measuring segregation and its association with population outcomes. *Methods in social epidemiology*, Jossey-Bass San Francisco, CA, v. 1, n. 169, p. 169–192, 2006.

REARDON, S. F.; O'SULLIVAN, D. Measures of spatial segregation. *Sociological methodology*, Wiley Online Library, v. 34, n. 1, p. 121–162, 2004.

RESENDE, T. d. F.; NOGUEIRA, C. M. M.; NOGUEIRA, M. A. Escolha do estabelecimento de ensino e perfis familiares: uma faceta a mais das desigualdades escolares. *Educação & Sociedade*, Centro de Estudos Educação e Sociedade, v. 32, n. 117, p. 953–970, 2011.

ROLNIK, R. *A cidade e a lei: legislação, política urbana e territórios na cidade de São Paulo*. [S.l.]: Studio Nobel, 1997.

SABATINI, F. The social spatial segregation in the cities of latin america. Inter-American Development Bank, p. 1–40, 2003.

SAMUELSON, P. A. A note on the pure theory of consumer's behaviour. *Economica*, JSTOR, v. 5, n. 17, p. 61–71, 1938.

SCANDIUCCI, G. Cultura hip hop: um lugar psíquico para a juventude negro-descendente das periferias de são paulo. *Imaginário*, scielopepsic, v. 12, p. 225 – 249, 06 2006. ISSN 1413-666X.

SCHELLING, T. C. *Neighborhood tipping*. [S.l.]: Harvard Institute of Economic Research, Harvard University, 1969.

STEPAN, N. L. Eugenia no brasil, 1917-1940. *Cuidar, controlar, curar: ensaios históricos sobre saúde e doença na América Latina e Caribe. Rio de Janeiro: Fiocruz*, v. 567, p. 331–391, 2004.

TELLES, E. E. Residential segregation by skin color in brazil. *American Sociological Review*, JSTOR, p. 186–197, 1992.

TELLES, E. E. *Race in another America: The significance of skin color in Brazil*. [S.l.]: Princeton University Press, 2004.

THEIL, H. *Statistical decomposition analysis*. [S.l.]: North-Holland Publishing Co., 1972.

THEIL, H.; FINIZZA, A. J. A note on the measurement of racial integration of schools by means of informational concepts. Taylor & Francis, 1971.

TORRES, H. de G. et al. Poverty and space: pattern of segregation in São Paulo. In: *Workshop on Spatial Segregation and Urban Inequality in Latin America, Austin, Texas*. [S.l.: s.n.], 2002. p. 15–16.

TORRES, H. G. Segregação residencial e políticas públicas: São Paulo na década de 1990. *Revista Brasileira de Ciências Sociais*, Associação Nacional de Pós-Graduação e Pesquisa em Ciências Sociais, v. 19, n. 54, 2004.

TORRES, H. G.; MARQUES, E. C. Reflexões sobre a hiperperiferia: novas e velhas faces da pobreza no entorno municipal. *Revista Brasileira de Estudos Urbanos e Regionais*, Associação Nacional de Pós Graduação e Pesquisa em Planejamento Urbano, n. 4, p. 49–70, 2001.

VILLAÇA, F. *Espaço intra-urbano no Brasil*. [S.l.]: Studio nobel, 1998.

VILLAÇA, F. São Paulo: Urban segregation and inequality. *Estudos Avançados*, SciELO Brasil, v. 25, n. 71, p. 37–58, 2011.

WELLER, W. O hip hop como possibilidade de inclusão e de enfrentamento da discriminação e da segregação na periferia de São Paulo. 2004.

WHITE, M. J. The measurement of spatial segregation. *American journal of sociology*, University of Chicago Press, v. 88, n. 5, p. 1008–1018, 1983.

WILSON, W. J. The truly disadvantaged: The inner city. *The Underclass, and Public Policy*, v. 8, 1987.

WONG, D. W. Spatial indices of segregation. *Urban studies*, Sage Publications Sage UK: London, England, v. 30, n. 3, p. 559–572, 1993.

YULE, G. U.; KENDALL, M. An introduction to the theory of statistics. london. *Griffin Co*, p. 271, 1950.

ZHANG, J. Tipping and residential segregation: a unified schelling model. *Journal of Regional Science*, Wiley Online Library, v. 51, n. 1, p. 167–193, 2011.

Appendix

APPENDIX A – OTHER STATISTICAL SUMMARIES

Table 16 – Statistical summary of household income in SPMA (2010) – Number or percentage of households per census tract

	min	median	mean	max	sd	N
# households w/ no income	0.00	0.00	0.03	1.00	0.09	29598
# households under 1/8 pcmw	0.00	0.48	0.46	1.00	0.24	29598
# households w/ 1/8-1/4 pcmw	0.00	0.00	6.66	227.00	17.52	29598
# households w/ 1/4-1/2 pcmw	0.00	3.00	13.13	271.00	22.04	29598
# households w/ 1/2-1 pcmw	0.00	10.00	16.74	188.00	18.57	29598
# households w/ 1-2 pcmw	0.00	17.00	20.12	158.00	15.96	29598
# households w/ 2-3 pcmw	0.00	52.00	54.76	316.00	34.56	29598
# households w/ 3-5 pcmw	0.00	46.00	52.14	349.00	38.80	29598
# households w/ 5-10 pcmw	0.00	18.00	23.88	328.00	23.51	29598
# households w/ 10 pcmw or more	0.00	3.00	5.84	116.00	7.56	29598
perc. low-inc households	0.00	0.00	1.22	71.00	2.49	29598
perc. high-inc households	0.00	6.00	11.29	336.00	14.84	29598

Source: Data from IBGE (2010). Calculations by the author.

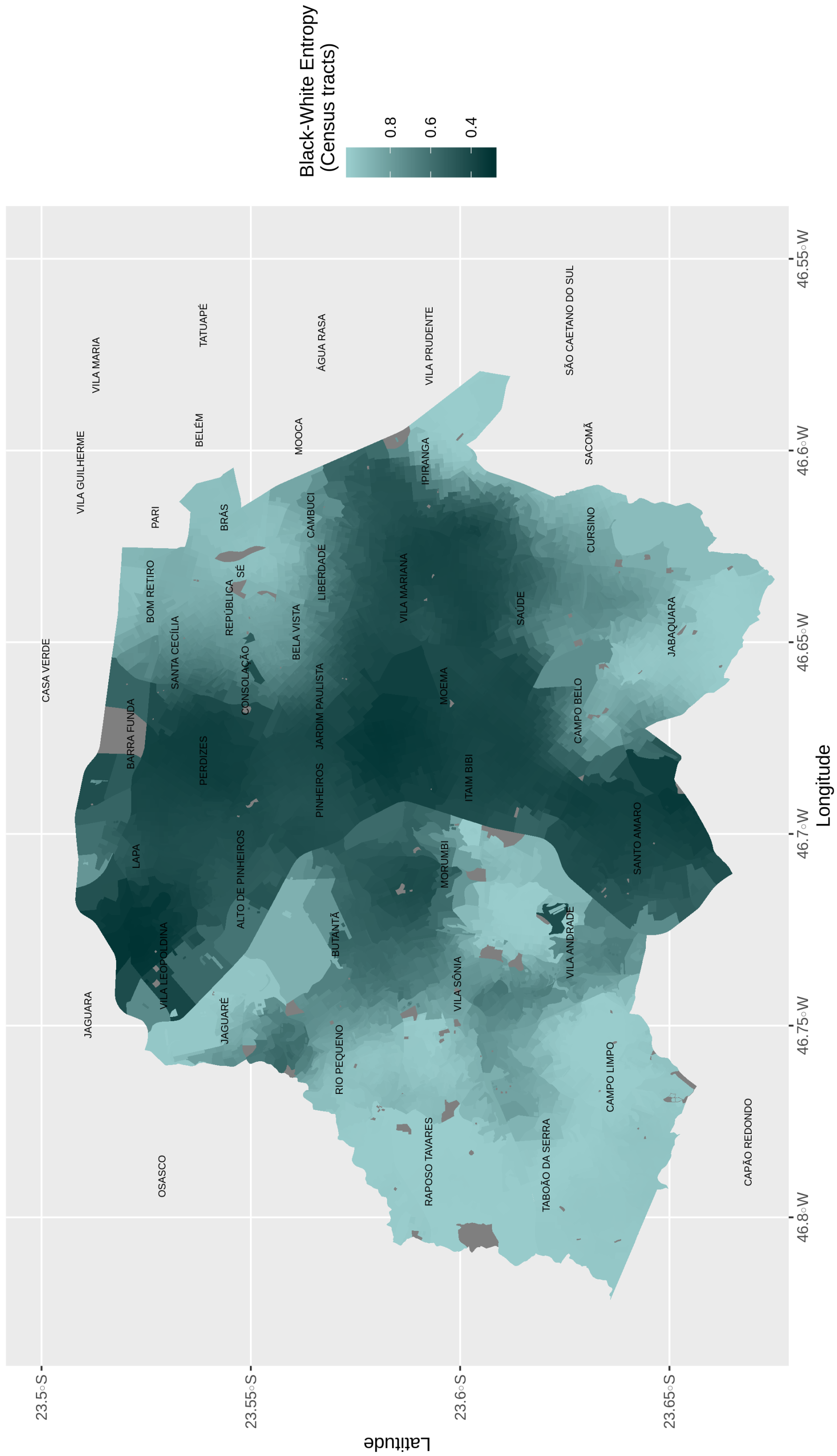
Table 17 – Statistical summary of 2010 Demographic Census–Sample Survey in SPMA

Age	Race	Literacy	High school	College	Wage
Min.: 0.00	white: 697250	yes: 1072885	no: 883847	no: 1032103	Min.: 0
1st Qu.: 16.00	black: 78522	no: 59494	yes: 332764	yes: 184508	1st Qu.: 600
Median: 30.00	yellow: 20711	NA’s: 84232			Median: 950
Mean: 31.94	brown: 418029				Mean: 1710
3rd Qu.: 46.00	native brazilian: 1405				3rd Qu.: 1600
Max.:138.00	not declared: 694				Max.: 800000
					NA’s: 647459
sd: 19.92					sd: 4133.97

Source: Data from IBGE (2010). Calculations by the author.

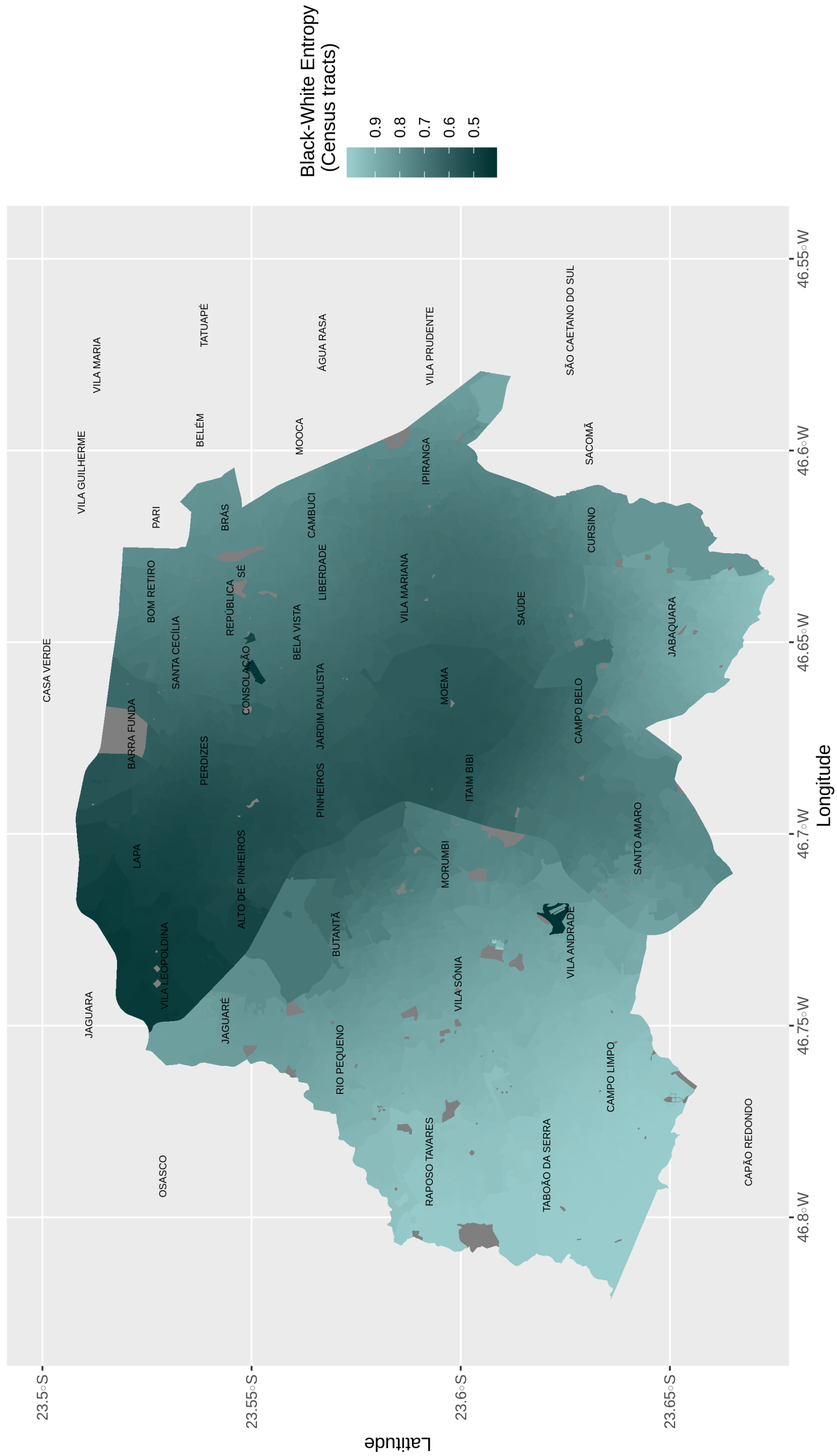
APPENDIX B – OTHER MAPS

Figure 32 – Spatial entropy in CW SPMA – Walking travel time with a 12-min bandwidth



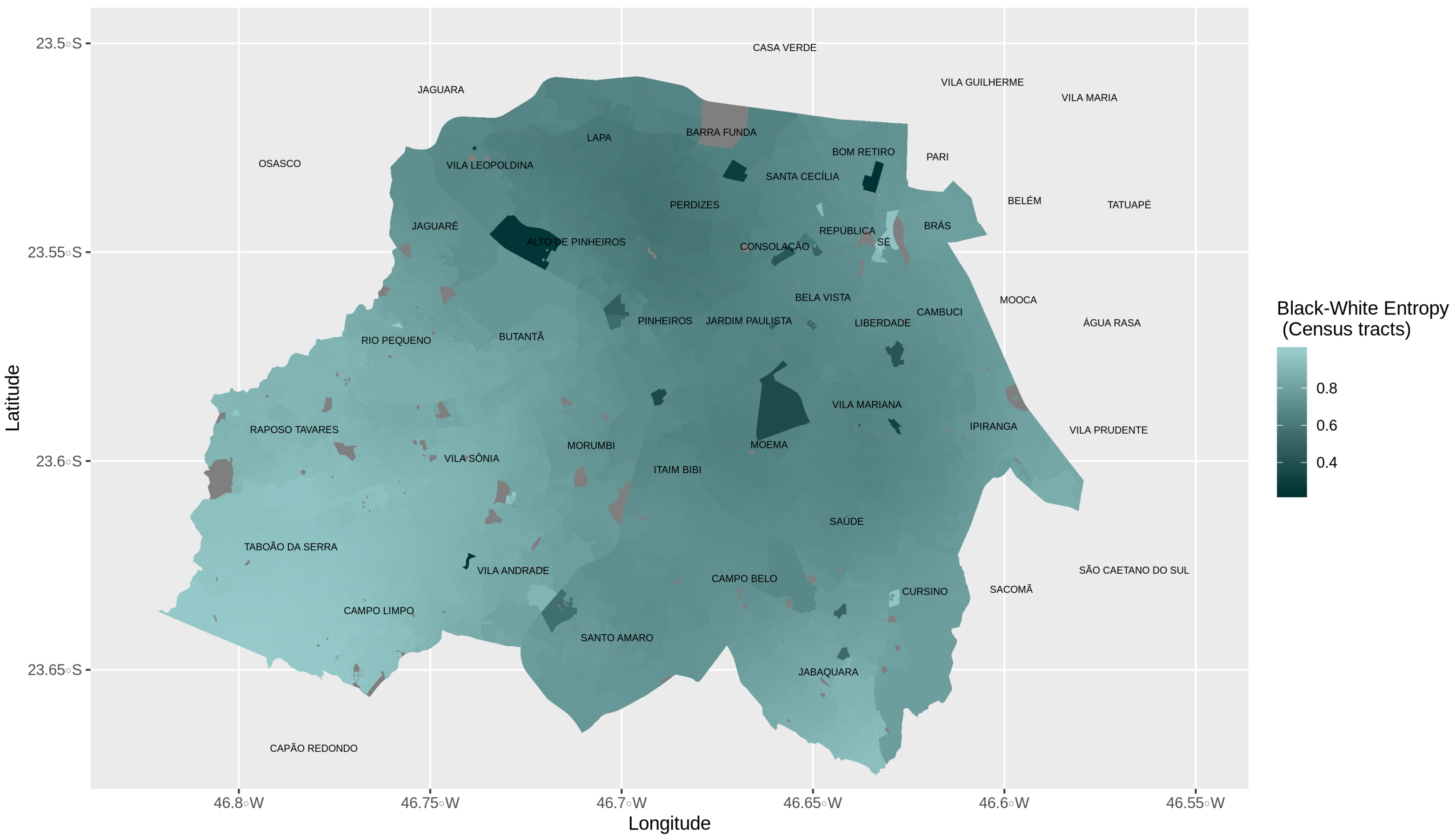
Data source: IBGE (2010) and OSP (2019). Elaboration by the author.

Figure 33 – Spatial entropy in CW SPMA – Walking travel time with a 45-min bandwidth



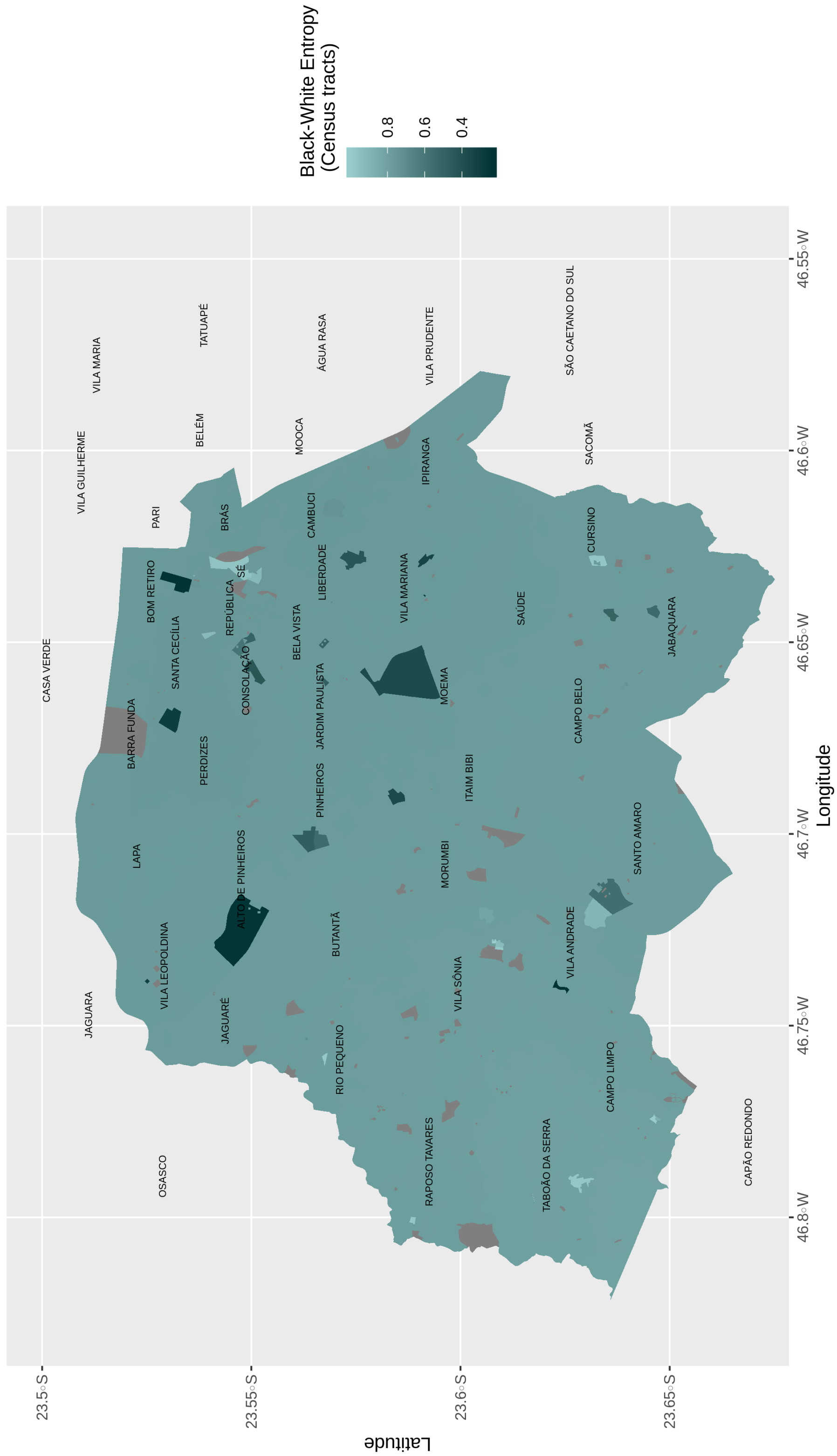
Data source: IBGE (2010) and OSP (2019). Elaboration by the author.

Figure 34 – Spatial entropy in CW SPMA – Travel time by car with a 12-min bandwidth



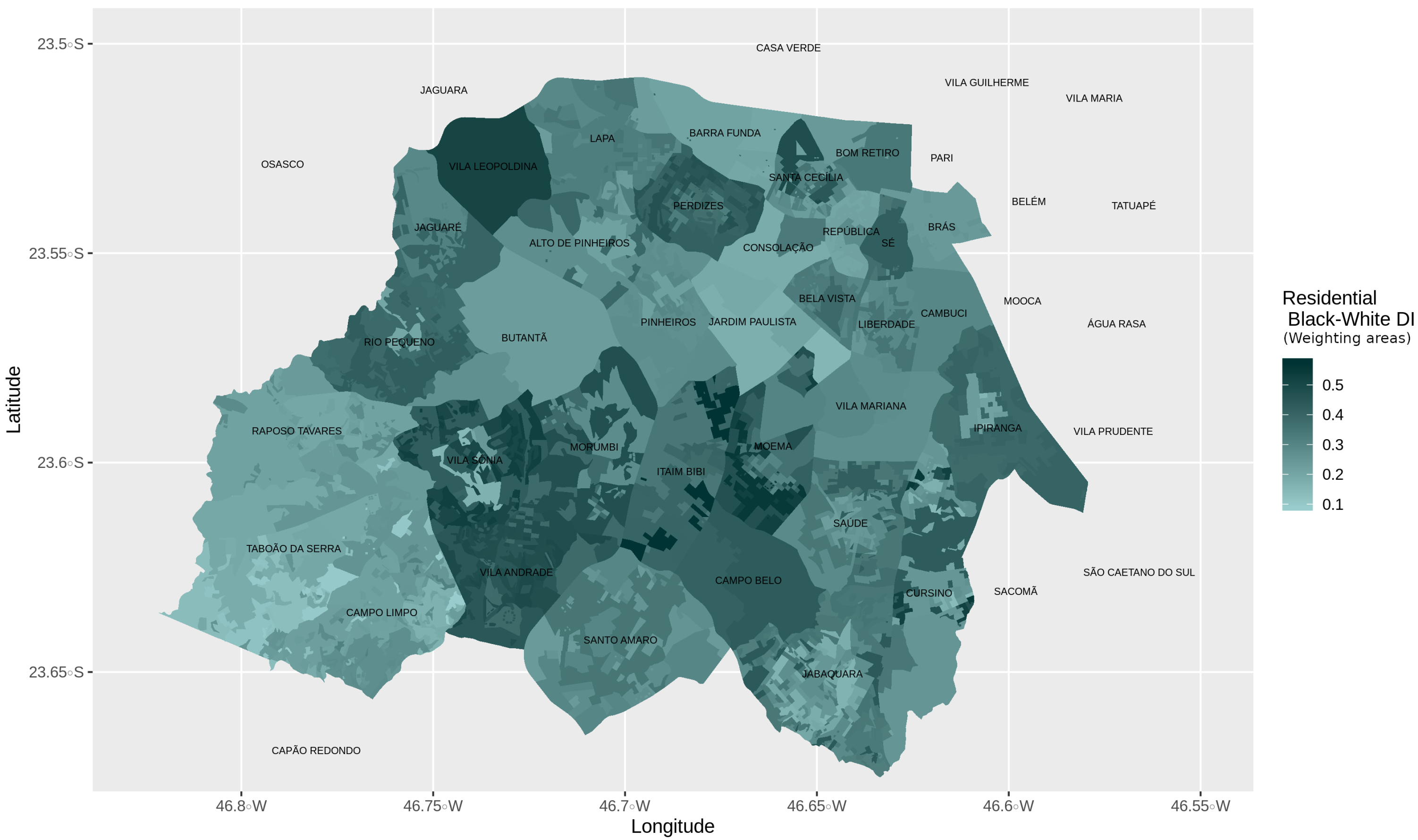
Data source: IBGE (2010) and OSP (2019). Elaboration by the author.

Figure 35 – Spatial entropy in CW SPMA – Travel time by car with a 45-min bandwidth



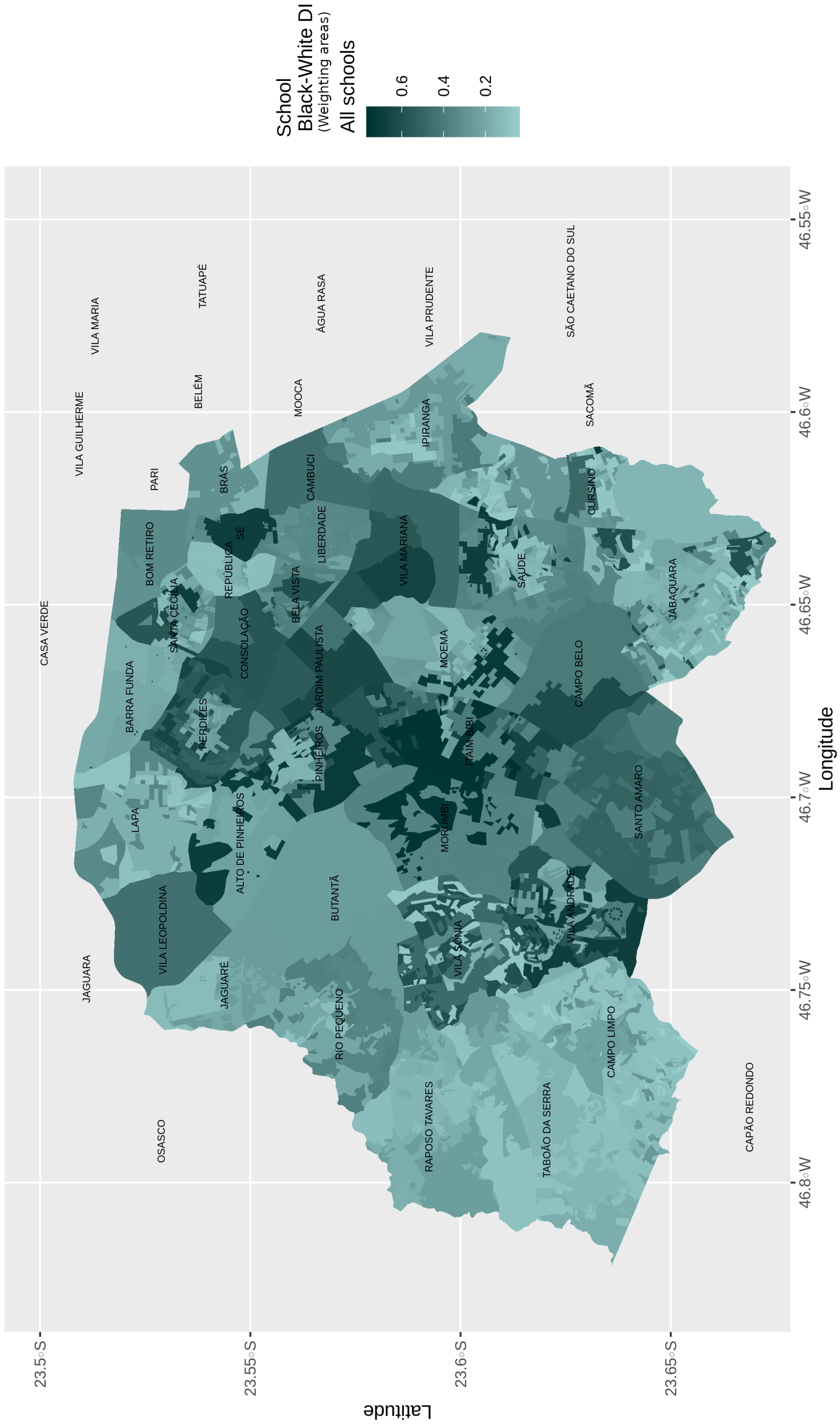
Data source: IBGE (2010) and OSP (2019). Elaboration by the author.

Figure 36 – Residential Dissimilarity Index in CW SPMA – weighting area



Data source: IBGE (2010). Elaboration by the author.

Figure 37 – School Dissimilarity Index in CW SPMA – all schools – weighting area



Data source: INEP (2010) and CEM (2013). Elaboration by the author.