

Insper Instituto de Ensino e Pesquisa  
Ciências Econômicas

Gustavo Theil

# **Miopia ou controle? Análise da eficiência dos instrumentos regulatórios urbanos para determinação de densidade habitacional**

São Paulo - SP  
2024

Gustavo Theil

**Miopia ou controle? Análise da eficiência dos  
instrumentos regulatórios urbanos para determinação  
de densidade habitacional**

Iniciação Científica

Insper Instituto de Ensino e Pesquisa  
Ciências Econômicas

Orientador: Adriano Borges Costa

São Paulo - SP  
2024

---

Gustavo Theil

Miopia ou controle? Análise da eficiência dos instrumentos regulatórios urbanos para determinação de densidade habitacional/ Gustavo Theil. - São Paulo - SP: 2024

31 p.

: Insper Instituto de Ensino e Pesquisa  
Ciências Econômicas , 2024.

Orientador: Adriano Borges Costa

1. Economia Urbana 2. Política pública I. Gustavo Theil. II. Miopia ou controle?  
Análise da eficiência dos instrumentos regulatórios urbanos para determinação de densidade habitacional.

---

# Resumo

O Plano Diretor Estratégico (PDE 2014) do município de São Paulo tem como um dos principais objetivos gerar adensamento populacional nas áreas mais desejáveis. O escopo desta pesquisa é compreender o funcionamento das forças de mercado e indentificar se os instrumentos de regulação são capazes de gerar os incentivos almejados pelo PDE. As estimativas apontam que cerca de metade da cidade não responde aos incentivos, já que não está registrada no Cadastro Imobiliário Fiscal e portanto se encontra em situação informal. Para a outra metade, identificou-se que entre os instrumentos utilizados para a regulação, a cota parte se demonstrou a mais efetiva para definir a densidade populacional, seguida pelo Coeficiente de Aproveitamento (CA). O número de pavimentos se mostrou irrelevante.

Palavras-chave: Economia Urbana, Política Pública, Plano Diretor

# Abstract

One of the main goals of the Strategic Master Plan (2014) of São Paulo is generate population density in the most desirable areas. The scope of this research is to understand the market forces and to identify whether regulatory instruments are capable of generating the desired outcomes. Estimates indicate that about half of the city does not respond to incentives, as it is not registered in the Fiscal Real Estate Registry and therefore is in an informal status. For the other half, it was identified that among the instruments used for regulation, the “cota parte”, or *shared quota* proved to be the most effective in defining population density, followed by the floor area ratio (FAR). The number of floors was found to be irrelevant.

Keywords: Urban Economics, Public Policy, Master Plan

# Sumário

1	<b>INTRODUÇÃO</b>	6
2	<b>TEORIA</b>	11
3	<b>DADOS</b>	14
4	<b>ANÁLISE</b>	25
5	<b>CONCLUSÃO</b>	30
	<b>REFERÊNCIAS</b>	31

# 1 Introdução

## Por que cidades existem?

Para Bauman, cidades são lugares de oportunidades e são movidas por “newcomers”, pessoas que não são originalmente do lugar e são estranhos à sua realidade. Quando essas pessoas chegam nas cidades, elas apresentam novas perspectivas sobre problemas antigos e suas formas de pensar geralmente conflitam com tranquilidade consequente da familiaridade e convivência dos residentes bem estabelecidos. Apesar disso poder gerar incômodo aos nativos, é para o bem deles e da cidade que suas formas de ser sejam questionadas e desafiadas pelos estranhos. Para o autor, esse estado de inquietude e permanente sentimento de “ser estrangeiro” é o que leva as pessoas e a cidade a buscar reflexão, debate e inovação (BAUMAN, 2003).

Para Jane Jacobs, uma característica que determina o sucesso de uma cidade é sua vitalidade. A autora defendia que a vida na cidade depende fortemente das dinâmicas sociais, das interações cotidianas de seus residentes. Para Jacobs, a cidade proporciona interações nos espaços públicos – onde a vida urbana efetivamente acontece – e possibilita que as pessoas se conectem, encontrem, colaborem e prosperem juntas. Ela valorizava a diversidade, densidade e a mistura de usos do solo urbano, argumentando que a presença de uma variedade de estabelecimentos comerciais, residenciais e culturais em uma mesma área promove a interação entre diferentes grupos sociais e estimulava a criatividade e a inovação (JACOBS, 1961).

Do ponto de vista econômico, o triunfo da cidade se encontra nos benefícios de aglomeração e adensamento. Economias de aglomeração permitem que firmas diferentes escolham ficar geograficamente próximas umas das outras e encontrem benefício econômico ao reduzirem seus custos e aumentarem a produtividade. Jan Brueckner, em *Lectures on Urban Economics*, delimita no capítulo 1 o racional econômico da existência de cidades, que se divide em quatro principais componentes.

A aglomeração tecnológica (*i*) aumenta a produtividade dos trabalhadores, na medida em que os empregos são mais concentrados e há “transbordamento” de conhecimento entre as firmas da região. Além disso, uma oferta maior e mais diversa de trabalho, causa maior competitividade e eficiência na escolha da pessoa certa para cada cargo. Aglomerações pecuniárias (*ii*) reduzem os custos das firmas, sem alterar sua produtividade. Com maior demanda por serviços como segurança, limpeza, contratação e advocacia, estes mercados se desenvolvem, tornam-se mais competitivos, eficientes e baratos. Inclusive, há serviços especializados de nicho, que podem estar disponíveis e acessíveis apenas em grandes centros urbanos. Aglomeração de varejo (*iii*) traz ganhos para os consumidores e comerciantes. Quando o comércio está aglomerado, o consumidor pode escolher entre mais opções e se desloca menos entre seus destinos caso queria comprar mais de um item. Dessa forma, os consumidores ganham e os comerciantes também, visto que

com mais consumidores e maior fluxo, maiores as vendas. Por fim, o custo de transporte (*iv*) é um dos fatores que mais mudam quando há densidade. A redução do custo de transporte, que pode ser considerada uma economia de aglomeração pecuniária, acontece não apenas para os trabalhadores, que se deslocam menos às oportunidades de emprego, mas também às firmas que gastam menos transportando seus bens e serviços (BRUECKNER, 2011).

Nesse sentido, se aglomeração e densidade trazem benefícios, uma cidade bem sucedida é uma cidade que ao longo do tempo tende a se adensar cada vez mais. Entretanto, uma pergunta relevante é se as forças de aglomeração atuam sozinhas ou se precisam de incentivos e regulação. Os modelos de economia urbana demonstram que as próprias forças de mercado agem de forma a incentivar o adensamento, mas há fatores que podem atuar na direção contrária também (BRUECKNER, 2011) – mais detalhes serão discutidos na Seção 2. Nessa perspectiva, a função das instituições seria de intervir de forma a combater essas forças, apoiando o adensamento.

Por outro lado, Edward Glaeser em *Triumph of the City* discute alguns desafios encontrados pela regulamentação e incentivos criados pelas instituições. Segundo o autor, muitas vezes a regulamentação apresenta procedimentos lentos e burocráticos, e acaba gerando decisões focadas em fatores que se sobrepõem o adensamento na escala de prioridades. Inclusive, o que Glaeser propõe é criar impostos à la Coase, de forma que os incentivos se alinhem para uma cidade mais densa, ainda compensando pelos seus possíveis malefícios.

Ronald Coase reconhece natureza recíproca do problema do custo social, no caso em que A causa malefícios a B, mas B também causa malefícios a A. Isso se aplica a um exemplo em que uma empresa (A) quer construir um prédio, mas a comunidade local não quer ser perturbada mudanças na região. Caso a empresa (A) siga com a construção, a comunidade local (B) será prejudicada, mas se a construção for barrada, a empresa (A) também sofrerá e menos oferta habitacional e empregos serão gerados. Coase propõe uma forma de pensar que leva em conta não apenas a ponderação de qual malefício é maior, mas também a compensação da parte que sai prejudicada de um eventual acordo. Dessa forma, é sempre feita a escolha que maximize o bem-estar social, sem causar distorções nas escolhas, ainda compensando a parte prejudicada (COASE, 1960).

“Primeiro, as cidades devem substituir seus longos e incertos procedimentos de licenciamento com um simples sistema de taxação. Se prédios altos criam custos ao bloquear vistas ou luz, então faça uma estimativa razoável desses custos e cobre o construtor de acordo. Se certas atividades são nocivas aos vizinhos, então devemos estimar o custo social e cobrar os construtores por eles, assim como devemos cobrar motoristas pelos custos de gerar engarrafamentos. Esses impostos podem então ser dados às pessoas que estão sendo impactadas, como os vizinhos que perderam a iluminação por um novo prédio que obstrui sua vista.” (GLAESER, 2011)

## Regulamentação em São Paulo

Atualmente em São Paulo, os mecanismos de regulação são principalmente definidos pelo atual Plano Diretor Estratégico de SP (SÃO PAULO, 2014, PDE). No código da lei do PDE, entre os 17 objetivos estabelecidos, ao menos nove estão relacionados a estratégias de adensamento urbano (LIMA, 2021). A ideia principal do plano é direcionar o adensamento para áreas capazes de admitir um grande volume de habitantes, principalmente no entorno de infraestruturas de transporte de alta capacidade.

O PDE institui uma variedade de instrumentos para gerar esse adensamento, alguns atuando sobre empreendimentos já existentes, e outros para os novos também. O artigo 96 do PDE estabelece que “os imóveis não edificados, subutilizados e não utilizados são sujeitos ao parcelamento, edificação e utilização compulsórios”, de forma a passar a exercer sua função social dentro de um prazo estabelecido. Existem, inclusive, dispositivos para lidar especificamente com imóveis com este perfil em áreas que dispõem das características apropriadas para um maior adensamento, como é o caso das ZEIS 3:

“ZEIS 3 são áreas com ocorrência de imóveis ociosos, subutilizados, não utilizados, encortiçados ou deteriorados localizados em regiões dotadas de serviços, equipamentos e infraestruturas urbanas, boa oferta de empregos, onde haja interesse público ou privado em promover Empreendimentos de Habitação de Interesse Social”

Art. 45 do PDE (SÃO PAULO, 2014)

Para a regulação dos novos empreendimentos, há três principais dispositivos. O primeiro deles é o gabarito, que determina a altura máxima, em metros, dos imóveis. Outro instrumento é o coeficiente de aproveitamento (CA), que determina quantas vezes a área do lote pode ser construída. Se o CA é básico ( $CA = 1$ ), e o lote possui  $1.000m^2$ , então pode ser construído um empreendimento que distribui estes  $1.000m^2$  em uma quantidade qualquer de andares que respeite o gabarito. Se o CA for 2, significa que para o mesmo lote, podem ser distribuídos  $2.000m^2$  em  $n$  andares. Por fim, a cota parte é a cota máxima de terreno por unidade habitacional e determina o número mínimo de unidades habitacionais do terreno. Para calcular a cota parte, basta dividir o lote pelo número de unidades, resultando na cota do terreno ocupada por cada unidade habitacional. Dessa forma, o número mínimo de unidades habitacionais é dado pela Equação 1.1, na qual  $A_t$  representa a área do lote e  $Q$  a cota parte.

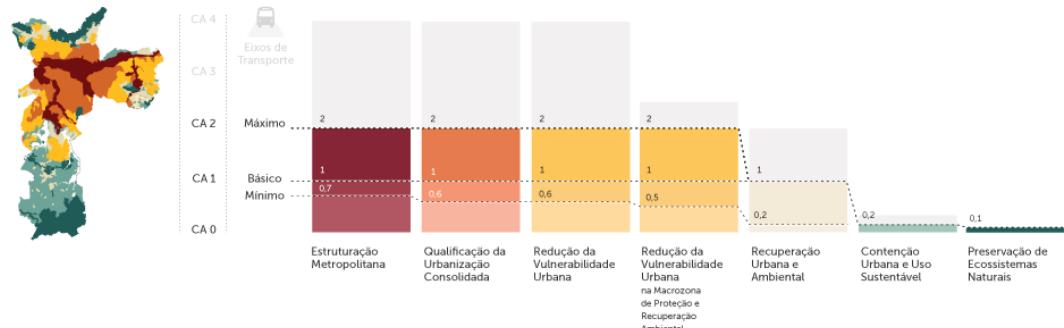
$$N_{min} = \frac{CA_{utilizado}}{CA_{max}} \cdot \frac{A_t}{Q} \quad (1.1)$$

No PDE, foram estabelecidos diferentes níveis de CA na cidade, a depender dos objetivos que se tem em relação à região, como é possível observar na Figura 1. Nas regiões de preservação ambiental, por exemplo, o CA é de 0.1, o menor da cidade, enquanto nas áreas do entorno de

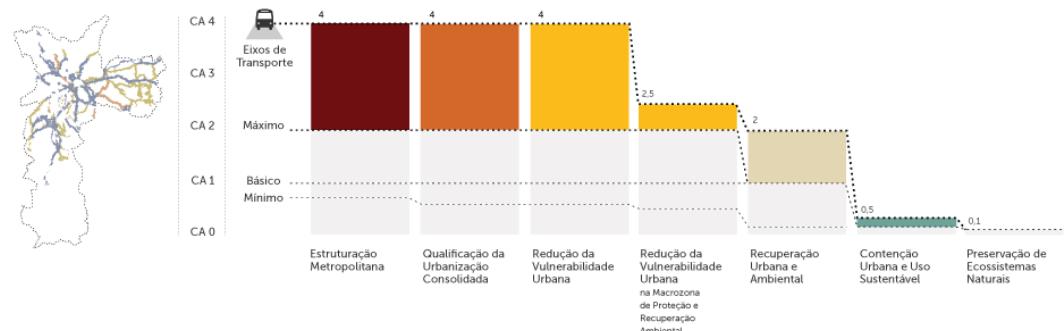
equipamentos de transporte pública de alta capacidade, (EETUs), o CA chega a 4. Com isso, o adensamento é direcionado às áreas que são aptas a receber mais habitantes.

Figura 1 – Coeficientes de aproveitamento (CA) na cidade

### Macroáreas



### Eixos de Estruturação da Transformação Urbana (EETU)



Fonte: <https://gestaourbana.prefeitura.sp.gov.br/marco-regulatorio/plano-diretor/entenda-o-projeto-de-lei-68813/>

Da mesma forma, existem parâmetros para a cota parte na cidade, a depender da localização. Nas Macrozonas de Estruturação e Qualificação Urbana, por exemplo, a cota parte apontada na Equação 1.1 como  $Q$  equivale a 20. Em termos práticos, isso significa que um lote de 1.000m<sup>2</sup> hipotético deve apresentar no mínimo 50 unidades habitacionais. Este valor não é incrementado por um aumento do CA, mas pode decrescer se o CA escolhido pelo projeto seja menor do que o CA máximo da região.

Nesse sentido, não existe exatamente um dispositivo que determine a densidade populacional em si. Isso pode levar a um cenário em que o mercado está fracamente regulado, implicando que a densidade não está sendo efetivamente decidida pelo PDE. Apesar de haver um número mínimo de unidades habitacionais, estas não necessariamente se traduzem em população, dado que ao mesmo tempo que pode abrigar uma única pessoa, também pode abrigar uma família com diversos membros. Além disso, este valor mínimo não é afetado pela verticalização do imóvel, então imóveis com um CA maior não são obrigados a terem mais unidades habitacionais.

## O problema

Levando em consideração que um dos principais objetivos do PDE é estimular o adensamento de determinadas áreas, é importante avaliar se os instrumentos de regulação que estão disponíveis para alcançar este objetivo são eficientes. Em outras palavras, é crucial entender se os 3 components expostos (CA, cota parte e gabarito) são suficientes para determinar a densidade populacional que um novo empreendimento vai gerar.

Caso estes instrumentos apontados não sejam eficientes em determinar a densidade, o nível de adensamento estaria sendo decidido pelo mercado. Entretanto, o que é importante notar é que o mercado sempre busca maximizar seus lucros, o que não necessariamente reflete em maior densidade populacional. Do ponto de vista teórico, é importante compreender quais elementos estão envolvidos nas escolhas das firmas na oferta de unidades de habitação, que será uma discussão da Seção 2.

Na Seção 4, será avaliado se estes instrumentos determinam a densidade demográfica. Para tanto, serão usados os dados do IPTU para calcular os indicadores utilizados na regulação e os dados do Censo de 2022, para identificar a densidade demográfica da região. Caso os indicadores sejam suficientes para explicar a densidade, significa que o PDE consegue definir a densidade usando os instrumentos previstos na lei. Caso contrário, a densidade populacional está sendo efetivamente decidida via mercado, não estando necessariamente alinhada com os objetivos do Plano para a cidade. Na Seção 3 serão discutidos em maior detalhe os dados.

## 2 Teoria

Na literatura microeconômica há diversos modelos que tentam compreender as dinâmicas econômicas da cidade. Utilizando premissas formais e provas matemáticas, é possível resolver o modelo para algumas variáveis endógenas, como preço por metro quadrado na cidade, uso do solo, densidade populacional, entre outros (PAPAGEORGIOU; PINES, 2012; FUJITA, 1989). A seguir será apresentado o modelo simplificado desenvolvido em (BRUECKNER, 2011), com o foco na variável densidade habitacional. O objetivo é compreender como o mercado se comportaria sem a intervenção da regulamentação e qual seria o seu impacto na variável resposta.

O modelo microeconômico começa com premissas simplificadoras fortes, mas que podem ser flexibilizadas na medida que algum componente merece ser discutido. Em sua primeira versão, a cidade é um círculo no qual todos os empregos se encontram no centro, as pessoas se deslocam uma distância ( $x$ ) para trabalhar, o custo de transporte por distância é ( $t$ ) e todos os moradores apresentam a mesma renda ( $y$ ). A renda disponível é definida pela renda subtraída dos custos de transportes ( $T = tx$ ). Por simplificação também, em cada unidade habitacional habita apenas uma pessoa. Nesse sentido, a densidade ( $D$ ) é definida pelo número de unidades habitacionais por área.

Como o sistema está sempre em equilíbrio, todos os moradores apresentam a mesma utilidade em morar em cada ponto da cidade. Caso um lugar fosse melhor de morar, todos iriam querer se mudar para essa localização, aumentando seu preço ( $p$ ). A função de utilidade dos cidadãos é composta por consumo de habitação ( $q$ ) e outros bens, apelidados por pão, ( $c$ )<sup>1</sup>, de forma que toda sua renda disponível será gasta com estes bens, então  $y - tx = c + pq$ . Para o consumidor, sempre é melhor consumir unidades habitacionais maiores e mais pão, ou seja, quanto maior  $p$  e  $c$ , maior sua utilidade. Dessa forma, quanto mais próximo do centro, e, portanto, um menor  $x$ , maior pode ser o gasto com  $q$  e  $c$ .

Analogamente, os produtores também apresentam lucro igual para produção imobiliária em todos os pontos da cidade. Se algum lugar na cidade lucrasse mais do que outro, as firmas buscariam produzir lá e, portanto, seu preço reduziria com a expansão de oferta, equilibrando o sistema. Os produtores apresentam uma função de produção que depende do preço ( $r$ ) da terra ( $l$ ) e do custo ( $i$ ) da verticalização ( $N$ ). Se construir um novo andar em um terreno existente for mais barato que comprar um novo terreno, a incorporadora decidirá por verticalizar, mas a cada novo andar construído, o próximo apresentará um custo maior.

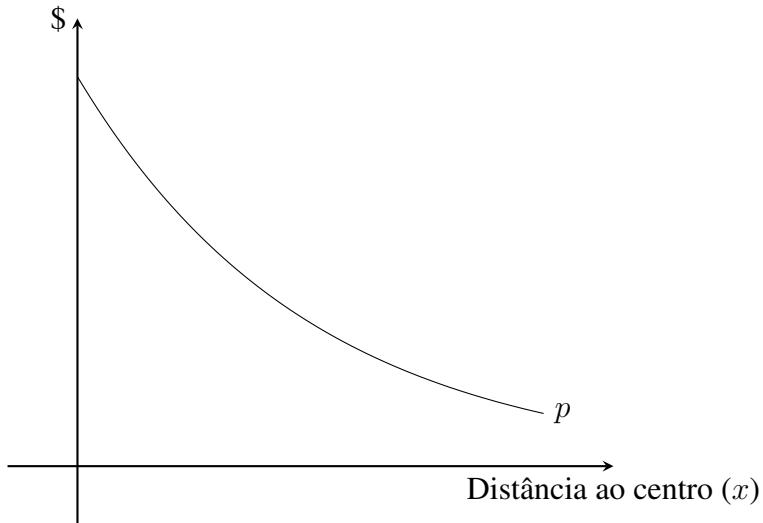
Com o modelo construído, agora é possível resolvê-lo de forma a descobrir o preço do metro quadrado na cidade. Na Figura 2 é possível observar o que acontece com o preço por metro quadrado ( $p$ ) na medida em que aumenta a distância ao centro ( $x$ ). Como a renda

---

<sup>1</sup> Por fins de simplificação, o preço do pão (outros bens) foi definido como unitário.

é constante na cidade e os preços por metro quadrado são maiores no centro, os habitantes no centro conseguem comprar apartamentos menores. Dessa forma a densidade habitacional apresenta o mesmo comportamento do que o preço do metro quadrado: na medida em que aumenta  $x$ ,  $D$  reduz.

Figura 2 – Curva de preço por metro quadrado

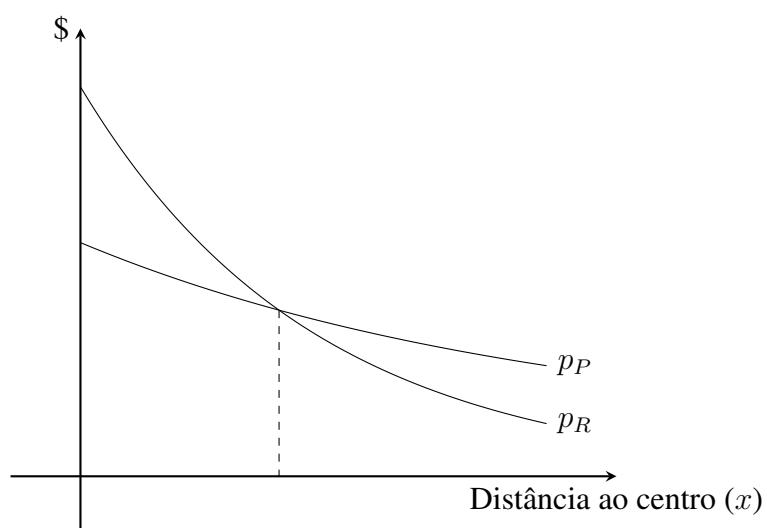


Portanto, com essas premissas e resultados, não é necessária a intervenção do governo para regulamentar a densidade, visto que as próprias forças de mercado adensam as regiões centrais. Dessa forma, regulamentações podem mais gerar peso morto, ineficiência e burocracia. Entretanto, algumas das premissas do modelo são tão fortes, que o tornam descolado da realidade. Inclusive, o efeito da distância pode ser ambíguo na densidade a depender de algumas premissas.

Suponha uma cidade na qual os habitantes possuem mais de uma opção de meio de transporte. Os ricos, por exemplo, se deslocam de carro e os pobres, de ônibus. Como o carro apresenta um custo mais alto, esse grupo vai valorizar mais morar em regiões centrais, visto que o efeito da distância ( $x$ ) em sua renda disponível é maior. Com este novo cenário, o quanto cada grupo está disposto a pagar pelo metro quadrado muda do cenário da Figura 2 para a Figura 3, na qual se observa um padrão de segregação espacial dos dois grupos. Os ricos, que habitam o centro, possuem um maior poder aquisitivo, então vão consumir apartamentos maiores. Como consequência, têm-se que com essas novas premissas, a densidade habitacional no centro já não é mais tão grande.

Com estas novas premissas, um tomador de decisão que tem como objetivo garantir maior densidade habitacional no centro deve intervir neste mercado, visto que a organização *laissez-faire* leva a um equilíbrio menos denso do que o discutido no caso anterior. Nesse sentido, do ponto de vista microeconômico não é determinístico como o mercado se articula para definir a densidade habitacional.

Figura 3 – Curva de preço por metro quadrado com dois grupos



# 3 Dados

## Dados do Censo

Para compreender o perfil geográfico da distribuição populacional de São Paulo foram utilizados os dados preliminares do Censo de 2022, feito pelo IBGE. Segundo o levantamento, a atual população do município de SP se encontra em 11.451.999 de habitantes, divididos em 4.996.529 de domicílios, dos quais apenas 4.316.336 estão ocupados. Na Figura 4 é possível observar quais são as áreas mais densas da cidade. A densidade foi calculada através da população no setor censitário dividida pela sua área. Os dados do censo foram utilizados em sua escala mais granular possível: nível setor censitário.

A metodologia para a delimitação dos setores censitários leva em consideração diversos fatores, como elementos na paisagem que se constituam em barreiras naturais ou artificiais e dificultam o trabalho do agente, pontos de referência estáveis e de fácil identificação no terreno, limites das estruturas territoriais, entre outros (IBGE, 2024). De maneira geral, estes limites são construídos de forma que o agente de coleta consiga cobrir o setor inteiro, minimizando os erros de medição. Estes setores censitários também são georreferenciados através da malha preliminar divulgada pelo IBGE e apresentam um número de identificação que pode ser decomposto segundo o exemplo apresentado a seguir.

21	00873	05	00	0026	P
UF	Município	Distrito	Subdistrito	Setor	Preliminar

É interessante notar na Figura 4, que o clássico modelo, no qual população se concentra principalmente no centro apresenta mérito, dado que perto da região da Sé, centro histórico de São Paulo, há uma enorme concentração habitacional. Entretanto, é surpreendente a densidade habitacional nas favelas de Paraisópolis e Heliópolis, que acabam por ser as áreas mais densas da cidade. Na Figura 5 é possível ver como ainda que haja uma relação entre a densidade habitacional e a distância, ela não é tão perfeita quanto no modelo microeconômico. Paraisópolis, por exemplo, está a 15km do centro e ainda é a região mais densa de São Paulo – mais detalhes serão discutidos adiante.

## Dados do IPTU

Em relação aos dados sobre os empreendimentos imobiliários, a base de dados escolhida foi do IPTU. Para os fins deste artigo, ela é a mais completa, visto que não representa um fluxo de novos imóveis construídos todos os anos como a base da Embraesp, mas apresenta um estoque

Figura 4 – Densidade populacional em São Paulo por setor censitário (Censo 2022)

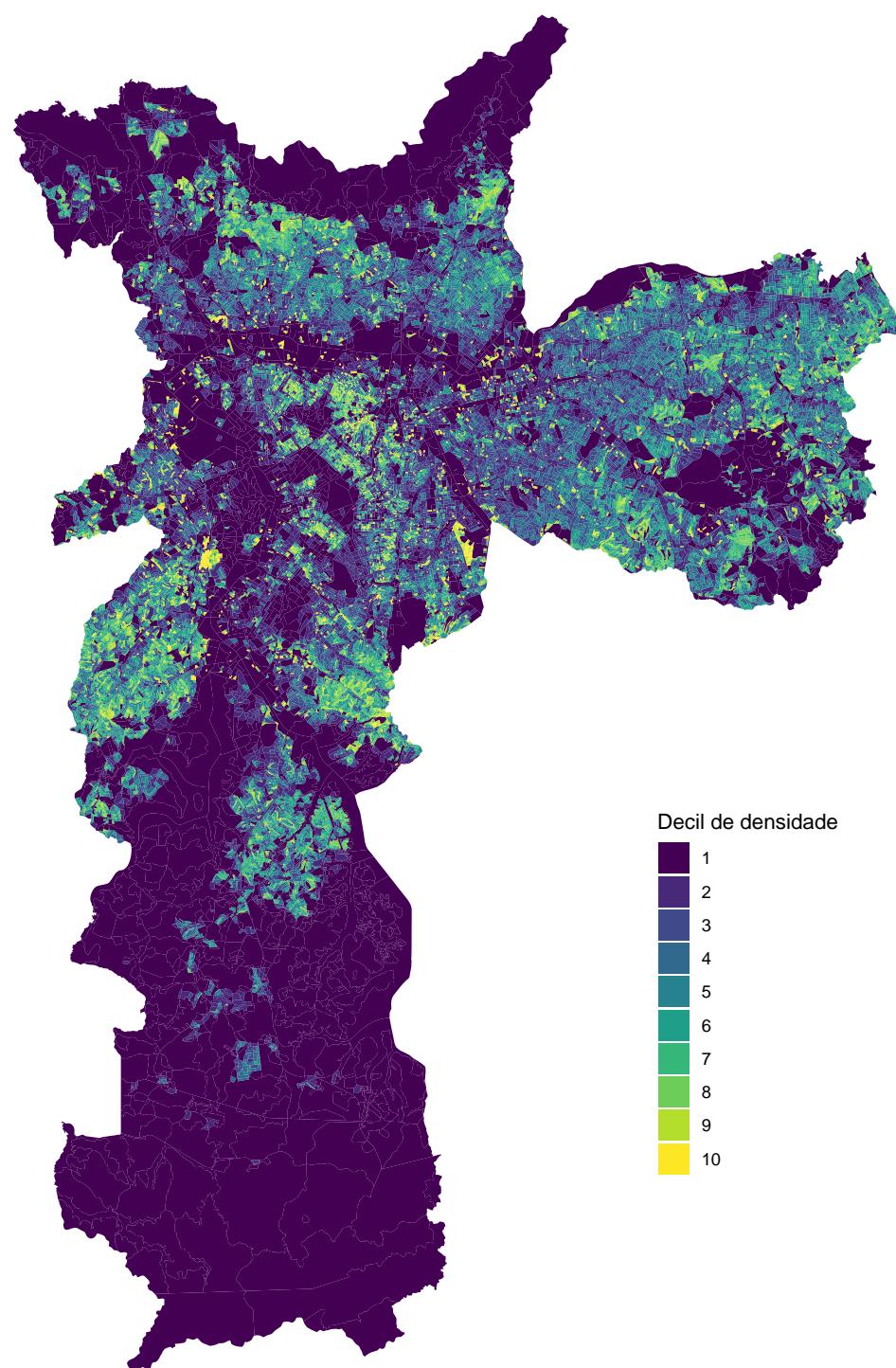
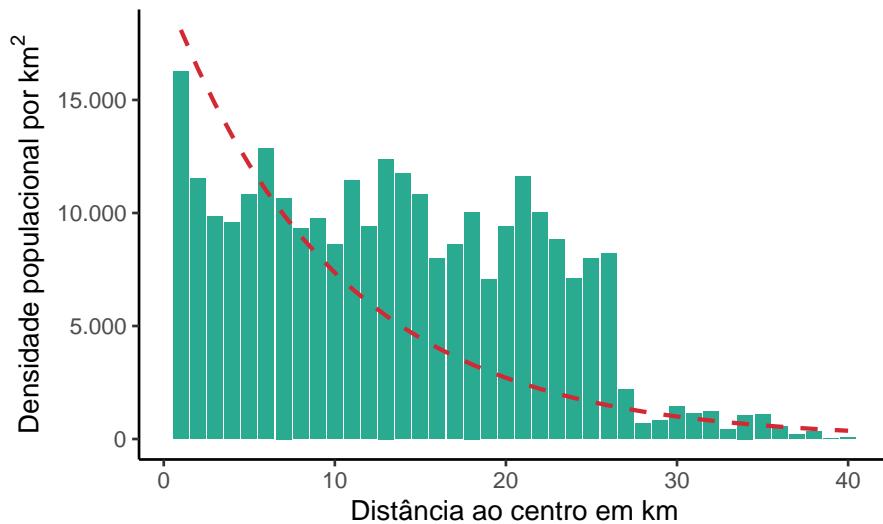


Figura 5 – Densidade habitacional na medida em que se distancia da Sé



imobiliário de tudo que já foi construído legalmente na cidade. Dito isso, estão cadastrados os 3.096.719 números únicos de contribuintes, que, segundo a definição da documentação dos dados no GeoSampa, “A cada imóvel urbano corresponderá um número de inscrição no Cadastro Imobiliário Fiscal, entendendo-se como imóvel: I - a área de terreno, construído ou não, definida em matrícula do competente Serviço de Registro de Imóveis ou em transcrições ainda vigente”. Os únicos dados que não constam nessa base são relativos a lotes irregulares ou não registrados, como em áreas de favelas.

Entre os dados disponíveis do IPTU, se destacam a área do terreno, a área construída, a área ocupada e o número de pavimentos do empreendimento. O tipo de uso do imóvel também está disponível, mas como esta pesquisa foca em densidade habitacional, foram descartados os usos não residenciais. Na figura 6, é possível observar a área construída por tipo e padrão de uso em São Paulo. Logo, com estes dados é possível calcular os indicadores utilizados pela regulamentação do coeficiente de aproveitamento (CA), gabarito e cota parte. A distribuição destes indicadores pode ser vista na Figura 7.

Em São Paulo, 70% da área construída é residencial, sendo que o padrão residencial mais comum é o "C". O padrão é uma divisão feita no cálculo do IPTU (Município de São Paulo, 1986), para criar descontos do imposto para imóveis com características desejáveis do ponto de vista do planejamento urbano, como maior adensamento. Medidas como metragem, pé direito, vaga de garagem, número de pavimentos, elementos arquitetônicos, materiais de construção, etc., são fatores levados em consideração para determinar o padrão. No caso do "C", ele está no meio da escala de densidade, sendo "A" o mais denso, e "E" o menos. O segundo uso mais comum é o comercial, seguido do industrial e entretenimento. Na categoria de entretenimento encontram templos religiosos, clubes, estádios esportivos, cinemas, aeroporto, museu, zoológico, entre outros.

De fato, ao observar os indicadores na Figura 7, é possível notar que o perfil habitacional

Figura 6 – Área construída em São Paulo por tipo e padrão de uso

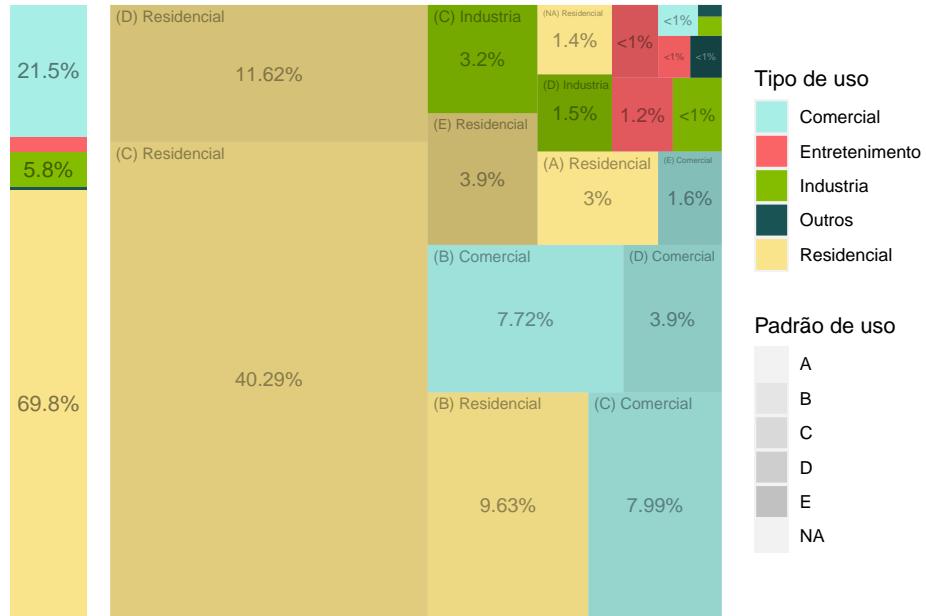
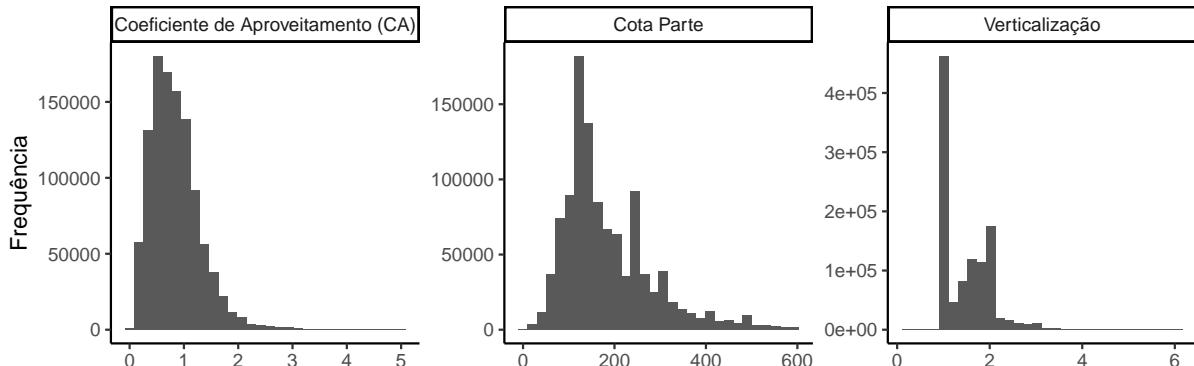


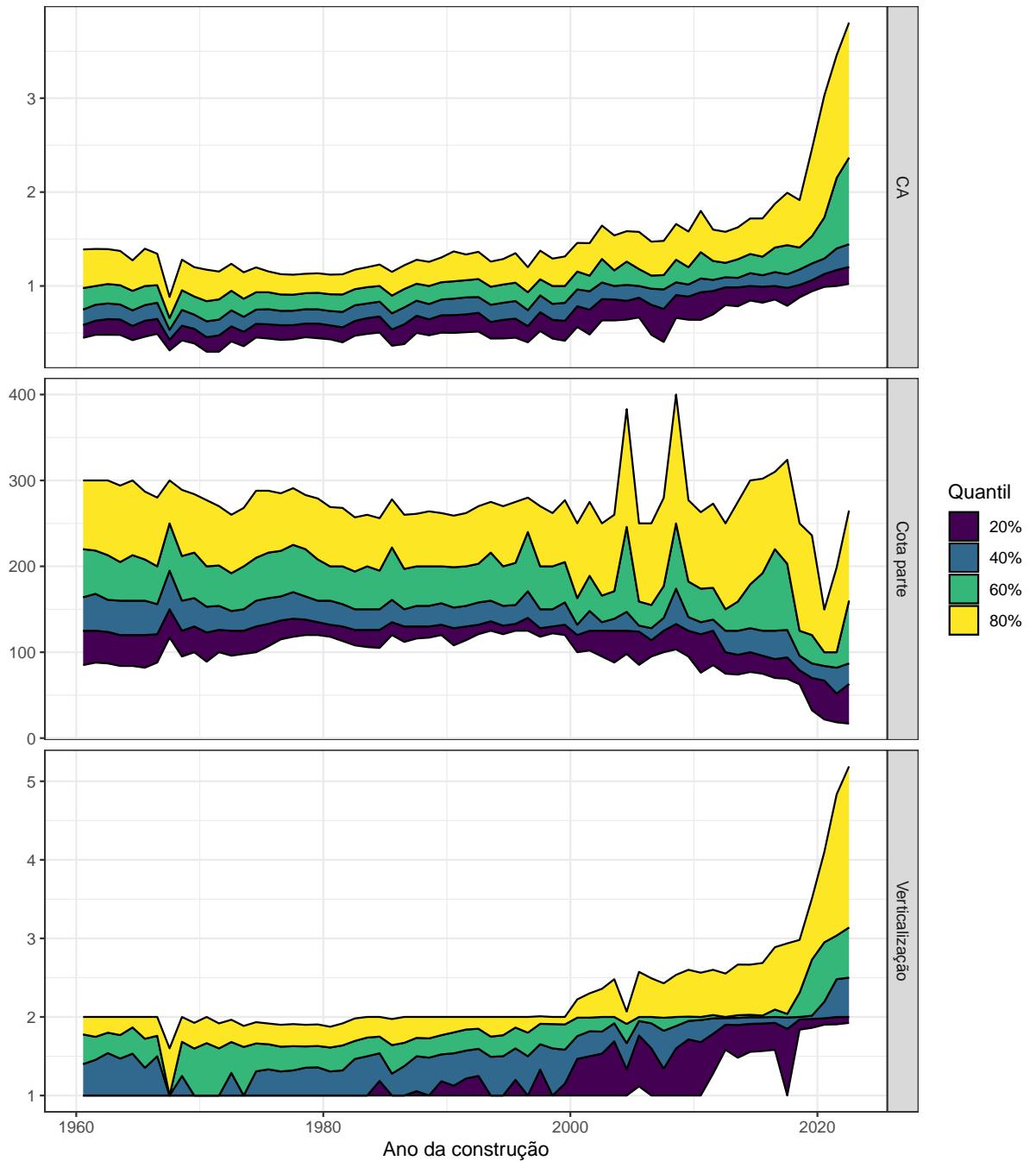
Figura 7 – Distribuição dos indicadores em cada lote de São Paulo



de São Paulo é bastante horizontal. É surpreendente que 95% dos lotes residenciais em São Paulo apresentam 2 ou menos pavimentos. Ainda, a mediana do CA é 0,8, o que indica que em mais da metade dos lotes da cidade foi construído menos do que a área do terreno. Ainda, a cota parte mediana dos lotes da cidade é de  $155m^2$ , o que representa um uso do terreno para poucas unidades habitacionais em média.

Todavia, ao analisar estes indicadores ao longo do tempo na Figura 8, fica evidente a trajetória de adensamento. Na figura, cada linha representa um quantil dos lotes do IPTU. Ao ordenar as construções no ano 2020 de maneira crescente de CA, observa-se que o empreendimento na posição 80% dessa fila apresenta um CA de 3,8. Este mesmo procedimento nos anos 2000 retornaria um empreendimento com CA igual a 1,46. Analogamente, a cota parte que

Figura 8 – Variação no tempo dos indicadores residenciais em São Paulo



em 2020 apresenta o quantil 20% de meros 21,7 metros quadrados, em 2000 apresentaria 100 metros quadrados. Isso evidencia que nos últimos anos São Paulo passou por um forte processo de adensamento.

## Cruzamento dos dados

Os dados do IPTU não são georreferenciados, então desacompanhados de outros dados não é possível cruzá-los com os do Censo. O que possibilita fazer essa junção é que o número único de contribuinte do IPTU, também é o código do Setor, Quadra e Lote (SQL) que o empreendimento se encontra. Nesse sentido, é possível decompor o SQL e cruzar com as bases de lotes, quadras e setores do GeoSampa, que contêm a geometria de cada um dos 1.677.980 lotes, 45.987 quadras e 309 setores da cidade. O que explica a diferença entre o número de lotes e número de números dos contribuintes são os lotes que contém um condomínio, que pode haver diversos números únicos de contribuintes em apenas um lote. Nos dados de IPTU, há 1.314.353 contribuintes em lotes com apenas uma unidade habitacional e 33.129 condomínios, que contêm 1.782.366 unidades. Ao cruzar o SQL do IPTU com a base de lotes, 44.319 contribuintes não encontram um par na outra base, o que representa uma perda de 1,67% das unidades. Caso o join seja feito com a base de quadras, a perda passa a ser de 820 contribuintes, um erro de 0,03%. Já quando são cruzados com base no setor, o erro chega a 0.

Agora, com os dados do censo georreferenciados e os dados do IPTU também georreferenciados, é necessário juntar estas bases. Entretanto, como o recorte dos setores censitários não respeita os recortes dos lotes, algum tipo de critério deve ser adotado para conectar essas geometrias. Duas abordagens foram feitas, a primeira com um join geográfico e a segunda, a partir da rasterização de ambos os dados. Na primeira, para cada um dos 27.592 setores censitários foram identificados todos os lotes que se interseccionam com os setores. A partir disso, foram cortados os lotes que não estão contidos apenas em um setor, de forma a atribuir seus dados a todos os setores que pertence, ponderado pelo percentual de sua área que intersecciona com cada setor. A segunda abordagem envolve dividir a cidade em um quadriculado (*raster*) e converter tanto os dados do IPTU quanto do censo para esta escala. A metodologia é parecida, visto que formado o quadriculado, os lotes são divididos entre cada quadrado, ponderando-se pelo percentual de sua área que intersecciona com cada um.

Dessa forma, para cada setor censitário ou para cada célula do *raster*, há vários lotes, que devem ser agregados segundo algum critério. Para a área construída, área ocupada, área do terreno e unidades, os valores de cada lotes podem ser simplesmente somados. Entretanto, no caso da verticalização fica mais complicado e é necessário estabelecer de maneira muito clara a metodologia de cálculo, que é pouco documentada e unificada na literatura (ÇALIŞKAN; MASHHOODI; AKAY, 2022). Diversas métricas são possíveis para agregar o número de pavimentos de uma região, mas muitas delas podem gerar erros de medição ou acabar representando uma medida muito semelhante ao CA.

Assumindo prédios como paralelepípedos, ou seja, todos os andares apresentam a mesma metragem, é possível traçar relações geométricas nas quais a área ocupada é a superfície que representa a base do paralelepípedo, o número de pavimentos sua altura e a área construída seu volume (Figura 9). Nesse sentido, as relações apresentadas na Equação 3.1, que apresenta

diferentes formas de calcular a verticalização, ficam mais claras. Na equação, AC representa a área construída; AO, a área ocupada; AT, a área do terreno; CA o coeficiente de aproveitamento e tx\_ocupacao, o percentual da área do terreno que é ocupada. Usando o procedimento apresentado, em regiões que possuem o mesmo CA, quanto menor for a taxa de ocupação do terreno, maior será a verticalização. Analogamente, para dois lotes com a mesma ocupação de terreno, quanto maior for o CA, maior é a verticalização.

Figura 9 – Representação do prédio como um paralelepípedo

- Coeficiente de Aproveitamento

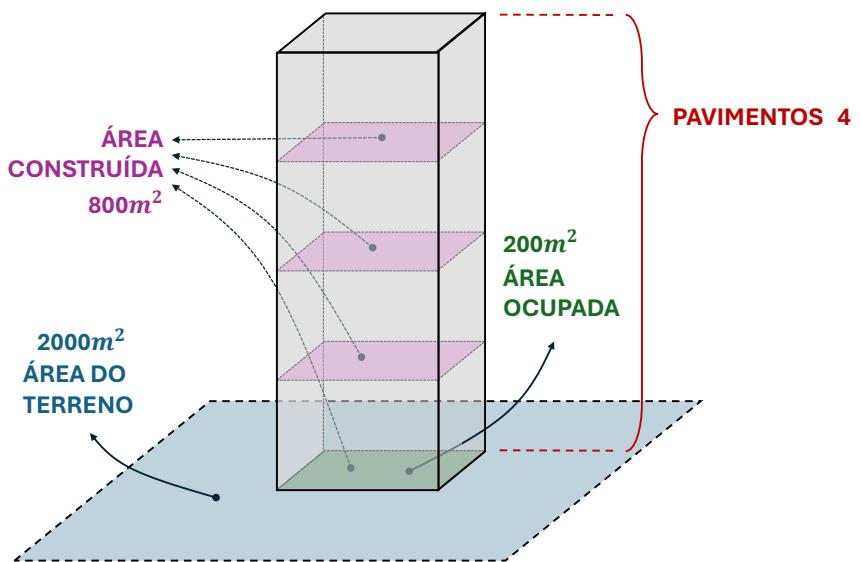
$$CA = \frac{AC}{AT} = \frac{800}{2000} = 0,4$$

- Taxa de ocupação

$$tx_{ocup} = \frac{AO}{AT} = \frac{200}{2000} = 10\%$$

- Verticalização

$$vert = \frac{CA}{tx_{ocup}} = \frac{0,4}{0,1} = 4$$



$$\text{Pavimentos} = \frac{AC}{AO} = \frac{AC}{AT \cdot \underbrace{\frac{AO}{AT}}_{tx_{ocup}}} = \frac{AC}{AT} \div \frac{AO}{AT} = \frac{CA}{tx_{ocup}} \quad (3.1)$$

Com as bases relacionadas, é importante fazer um balanço para analisar se as informações batem. A única informação em comum entre as bases é o número de unidades habitacionais, que nos dados do IPTU são classificadas como unidades, e na do Censo como domicílios. A definição deles não é exatamente a mesma, porém é a única informação possível de ser validada. Caso o número de domicílios no Censo seja maior do que o de unidades no IPTU, é um forte indicativo de que há moradias irregulares. Para medir essa diferença, foi criado um espectro de irregularidade, que é construído a partir do número de unidades dividido pelo número de unidades mais domicílios. Caso o valor calculado seja próximo de 50%, as informações são consistentes, ou seja, o número de unidades do IPTU é igual ao número de domicílios no censo. Na medida em que o balanço se aproxima de 0%, significa que há mais unidades no censo do que no IPTU, enquanto mais próximo de 100% indica o contrário.

“Domicílio é o local estruturalmente separado e independente que se destina a servir de habitação a uma ou mais pessoas, ou que esteja sendo utilizado como tal. A separação fica caracterizada quando o local de habitação for limitado por paredes, muros ou cercas e coberto por um teto, permitindo a uma ou mais pessoas, que nele habitam, isolá-lo das demais, com a finalidade de dormir, preparar e/ou consumir seus alimentos e proteger-se do meio ambiente, arcando, total ou parcialmente, com suas despesas de alimentação ou moradia. A independência fica caracterizada quando o local de habitação tem acesso direto, permitindo a seus moradores entrar e sair sem necessidade de passar por locais de moradia de outras pessoas” (IBGE, 2013).

Para analisar em termos práticos, no histograma da Figura 10 é possível identificar que há diversos setores censitários com domicílios registrados, mas nenhuma unidade habitacional no IPTU. Ademais, são incomuns os casos em que são semelhantes os números de domicílios e unidades. No mapa da Figura 11a, para cada setor censitário foram simulados pontos aleatórios dentro da geometria do setor para representar sua população. Nos setores censitários em que não há nenhum lote presente, ou seja, há uma subnotificação de imóveis na base do IPTU, os pontos foram pintados de vermelho. Em azul, estão as geometrias de lotes irregulares e favelas, disponibilizados no GeoSampa. É possível identificar que a maioria dos setores censitários que não possuem loteamento estão nessas áreas de lotes irregulares e favelas. Na Figura 11b, um procedimento semelhante foi feito. Para cada célula do *raster* foi calculado este indicador de irregularidade para analisar as inconsistências entre as bases. É notável que as células que se encontram em regiões centrais apresentam uma quantidade pequena de erros. Os casos mais críticos são células nas periferias.

Figura 10 – Diferença entre unidades e domicílios por setor censitário

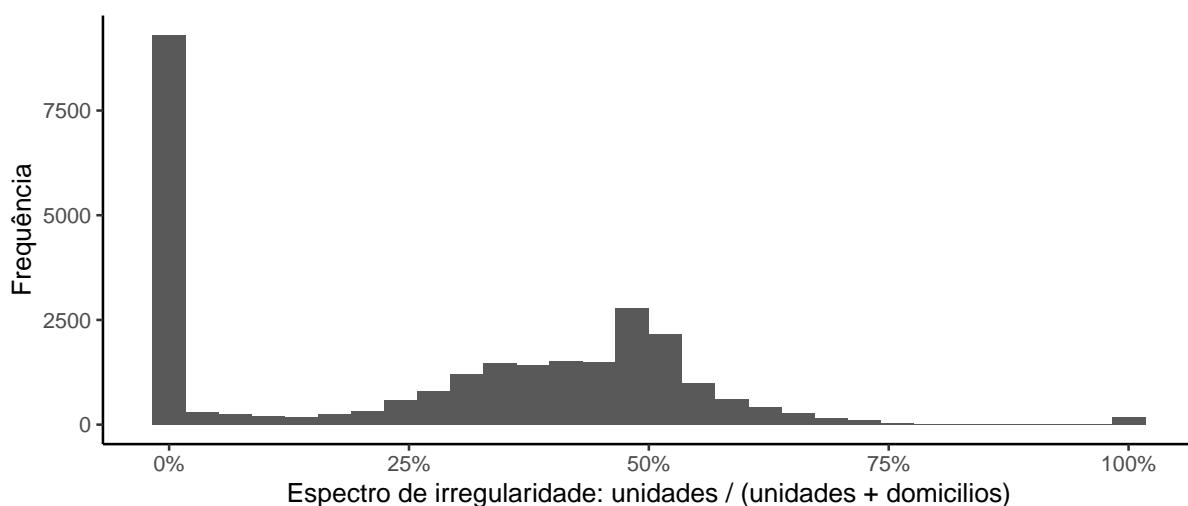
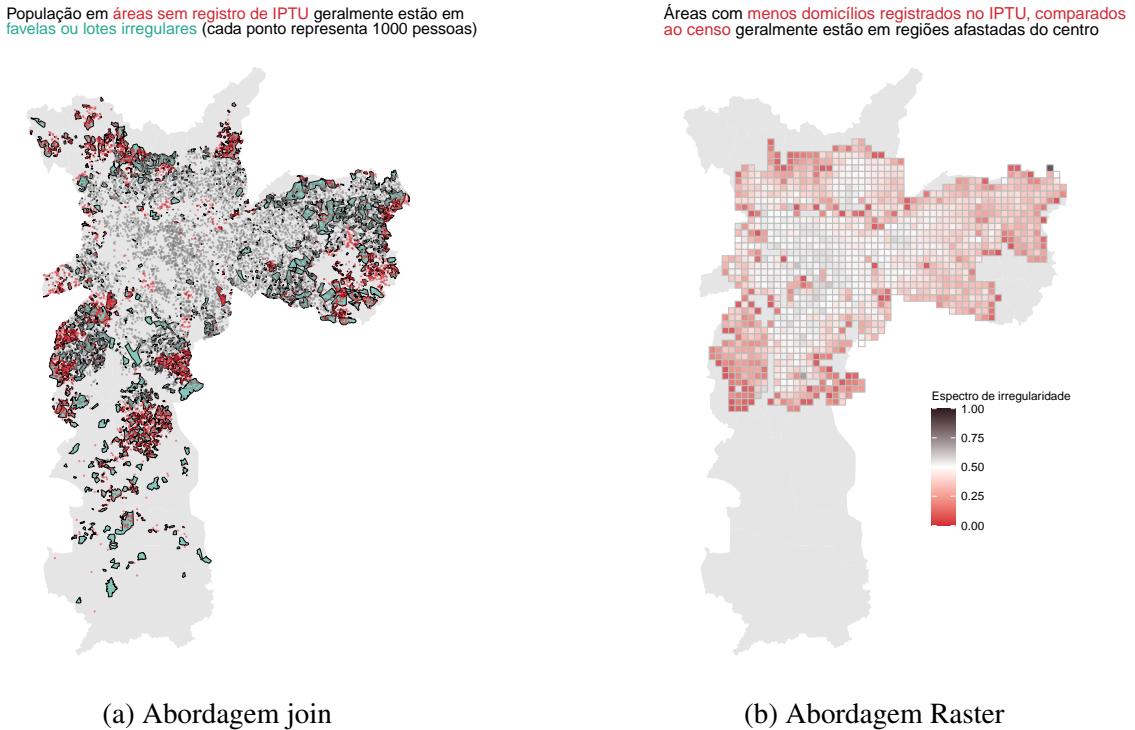


Figura 11 – Inconsistências entre dados do IPTU e do censo



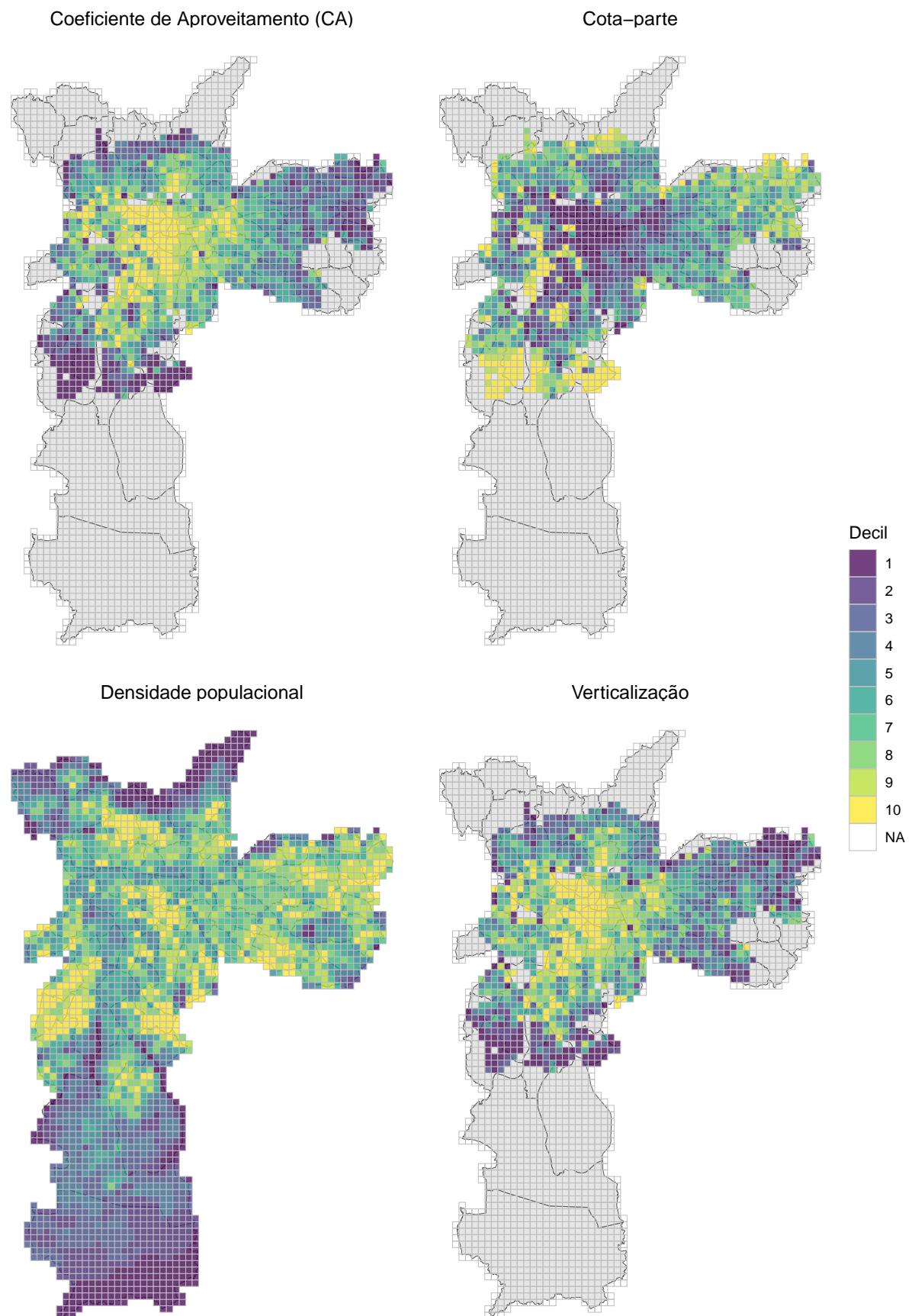
Na tabela 1 estão descritas as 5 células do *raster* que apresentam maior densidade populacional. É interessante observar que 4 das 5 regiões mais densas da cidade são de territórios de favelas e repleta quase completamente de domicílios que não estão registrados no IPTU. A densidade habitacional do município de São Paulo no geral é de 7.529 habitantes por quilômetro quadrado, o que significa que esta célula do *raster* posicionada na região de Paraisópolis apresenta uma densidade quase quatro vezes maior do que o resto da cidade.

Tabela 1 – Células do *raster* que apresentam a maior densidade populacional

Variável	Favelas				
	1. Paraisópolis	2. Heliópolis	4. Paraisópolis	5. Heliópolis	3. Sé (Bela Vista)
População	29.598	25.280	23.824	22.920	24.576
Domicílios (Censo)	11.655	10.178	9.361	9.001	17.875
Unidades (IPTU)	0	1.857	7	3	21.057
Espectro irregularidade	0.00%	15.43%	0.08%	0.03%	54.09%
Densidade habitacional	46.247	39.500	37.225	35.813	38.400
Área	640.000	640.000	640.000	640.000	640.000

Na Figura 12 é possível analisar o perfil geográfico dos instrumentos regulatórios. Como é de se esperar, regiões com maior adensamento também apresentam maior CA, verticalização e menor cota parte. É importante destacar que quanto mais lotes irregulares há em uma região, menos confiáveis são os indicadores calculados de CA, cota parte e verticalização, visto que estes são inteiramente calculados com base nos dados do IPTU, que são disponíveis apenas para lotes regulares.

Figura 12 – Rasters com os indicadores utilizados na regulação



## Alguns resultados preliminares

Apenas através da análise descritiva já é possível concluir que em boa parte da cidade os instrumentos regulatórios não funcionam, visto que as moradias não estão nem registradas nos sistemas públicos. Dessa forma, não importa qual regra for adotada pelo governo, estas áreas não serão impactadas por isso. Em termos numéricos, dos 4.996.529 domicílios registrados pelo Censo, há apenas 2.641.635 unidades habitacionais registradas no IPTU, o que representa um índice de moradia regular (em termos legais) de aproximadamente 52,87%. Além disso, considerando uniforme o número de moradores por domicílio em cada setor censitário, estima-se que 5.944.595 de pessoas habitam em domicílios que não estão registrados no IPTU, o que representa 51,9% da população de São Paulo.

Outra descoberta relevante é de que grande parte das áreas mais densas da cidade estão em regiões de irregularidade. Isso é um indicativo de que há incentivos mais poderosos do que a regulação para definir a densidade de cada região.

## 4 Análise

Para identificar como os indicadores estão relacionados à densidade habitacional, foi elaborada uma regressão linear, especificada na Equação 4.1.

$$densidade_i = \beta_0 + \beta_1 \cdot CA_i + \beta_2 \cdot cota\ parte_i + \beta_3 \cdot verticalização_i + \varepsilon_i \quad (4.1)$$

Na análise conduzida, o nível da observação  $i$  não pode ser o setor censitário, pois como discutido na Seção 3, o setor é construído de forma que um entrevistador sozinho consiga cobri-lo por completo. Como consequência, áreas mais densas apresentarão setores censitários menores e, caso isso não seja controlado na regressão, haverá um grande problema de endogeneidade. Inclusive, na metodologia de separação dos setores censitários, há valores definidos de domicílios que cada setor censitário deve apresentar, de acordo com suas características (IBGE, 2024). Todavia, a área não pode ser um regressor, visto que haveria uma relação de simultaneidade: ao mesmo tempo que a área explica a densidade, a densidade também explica a área. Essa relação viola as suposições necessárias para interpretabilidade dos resultados.

Para contornar este problema, o nível da observação foi definido como cada célula do *raster*, metodologia apresentada na Seção 3. Dessa forma, todas células apresentam o mesmo tamanho e este problema deixa de existir. Note que foram desconsideradas as células em que há dados do Censo, mas não há dados do IPTU, já que não se sabe o valor dos indicadores de regulação urbana nessas regiões. Em áreas em que há uma diferença muito grande entre número de domicílios registrados pelo Censo e unidades do IPTU, há menos confiança de que os indicadores calculados representam de fato a realidade. Portanto, em áreas em que o espectro de irregularidade está muito distante de 50%, haverá viés nas estimativas.

Nesse sentido, foram feitas duas regressões, uma delas irrestrita, e a outra considera apenas valores do espectro entre 40 e 60%. Além disso, em cada uma dessas regressões a densidade, que é a variável dependente, foi computada em nível e em log. Os resultados podem ser observados na Tabela 2.

O primeiro resultado interessante de ser analisado é como a verticalização não influencia a densidade. Do ponto de vista teórico isso faz sentido, visto que o número de andares é um fator principalmente estético, enquanto o CA computa quanto de área está sendo construída em relação ao tamanho do lote. Este resultado é consistente para as regressões (C) e (D), nas quais são considerados apenas os setores em que não há muita irregularidade e os estimadores são mais confiáveis.

Ao analisar a regressão (D), com a densidade em log, pode-se inferir que ao diminuir a cota parte em um metro quadrado, espera-se que com os outros componentes constantes, a

Tabela 2 – Regressão para densidade populacional em São Paulo, 2022

	Todos os setores				Espectro irregularidade: 40 a 60%			
	Nível (A)		Log (B)		Nível (C)		Log (D)	
	Coeficiente	*	Coeficiente	*	Coeficiente	*	Coeficiente	*
(Intercept)	11046.375	***	8.987	***	8632.310	***	9.109	***
cota_parte	-0.004		0.000	***	-6.973	***	-0.002	***
verticalizacao	-249.059		-0.075	**	-48.242		-0.026	
CA	1211.554	***	0.243	***	1672.931	***	0.143	***
R2	0.022		0.052		0.249		0.328	
R2 ajustado	0.020		0.050		0.245		0.325	
Observações	1255		1254		553		553	

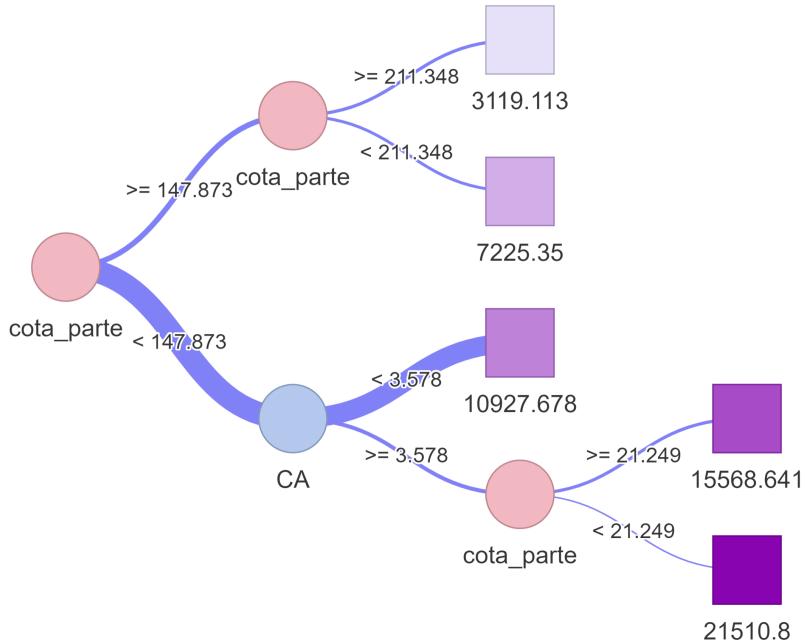
densidade aumente em média 0,2%. O aumento de 1 no CA, segundo os resultados da regressão (D), tem como consequência um aumento esperado médio na densidade de 14%, caso os outros componentes se mantenham iguais. Infelizmente o intercepto não apresenta interpretação, visto que na prática é impossível que uma área tenha cota parte igual a zero, visto que é calculada pela área do terreno sobre o número de unidades, então apresenta o valor máximo igual a área do terreno e se aproxima de zero na medida em que o número de unidades é infinito.

Todavia, há dois pontos que devem ser levantados. Primeiramente, partindo da hipótese de que os instrumentos de regulação apresentados não são eficientes para definir a densidade, os parâmetros da regressão estarão contaminados pelo viés advindo da omissão de regressores relevantes. Em segundo lugar, a forma funcional não foi especificada com base em alguma teoria econômica ou algum tipo de embasamento. Inclusive, ela pode ser reescrita como apresentado na Equação 4.2, na qual se observa o que está por trás dos indicadores.

$$\frac{População_i}{A. Total_i} = \beta_0 + \beta_1 \frac{A. Construída_i}{A. do Terreno_i} + \beta_2 \frac{A. Terreno_i}{N. unidades_i} + \beta_3 \frac{A. Construída_i}{A. Ocupada_i} + \varepsilon_i \quad (4.2)$$

Nesse sentido, foi tomada uma abordagem diferente, na qual foi construída uma *random forest* (WRIGHT; ZIEGLER, 2015). A vantagem dessa abordagem é que não é necessário especificar uma forma funcional, já que a própria metodologia seleciona as variáveis mais importantes. No exemplo de árvore de regressão apresentado na figura 13, cada círculo representa um nó de decisão, então no primeiro à esquerda, por exemplo, para cada observação da amostra gerada via *bootstrap*, é testado se a sua cota parte é menor ou maior que 147 metros quadrados. A depender do resultado, a observação é direcionada para o próximo nó de decisão até que chegue na folha (quadrado na figura). É interessante notar que as folhas cuja predição apresenta a maior densidade são as com cota parte pequena e CA alto.

Figura 13 – Exemplo de uma das árvores de regressão



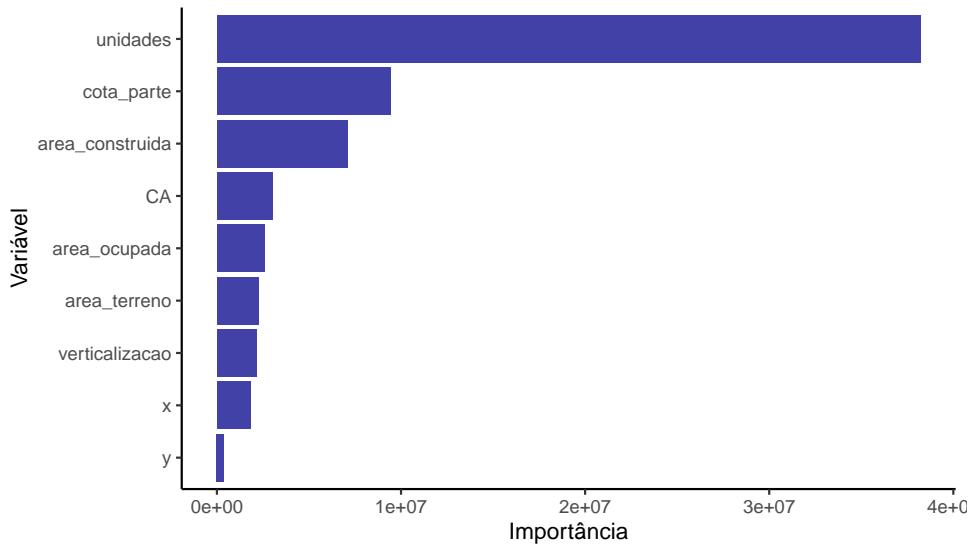
Para construir o modelo de floresta de regressão, é necessário dividir a base de dados em treino e teste, de forma que não haja *overfit*, ou seja, o modelo não “decora” os dados. Com o objetivo de tornar a comparação justa, foi refeita a equação de regressão linear, como especificado na Tabela 2 (C), mas apenas com os dados de treino. Depois, tanto a floresta de regressão, quanto a regressão linear previram os dados de teste, para comparar os modelos. O  $R^2$  na base de teste para o modelo de regressão linear foi de 23,9%, enquanto da floresta de regressão foi de 28,8%.

Depois, foi fornecido à floresta de regressão mais informações que estão disponíveis na base de dados do IPTU, para avaliar se há algum outro componente importante para definir a densidade. Como resultado, o novo  $R^2$  passou a ser de 84,8%, o que é cerca de três vezes maior. Isso é um forte indicativo de que há variáveis muito importantes para definir a densidade que não estavam sendo consideradas. Entre as novas variáveis do modelo, estão o número total de unidades, a latitude e longitude do *raster*, e as áreas do terreno, construída e ocupada.

Para analisar quais informações do modelo são mais importantes, é possível realizar um procedimento de permutação no qual é introduzido ruído nos regressores para testar como isso afeta o erro do modelo (BREIMAN, 2001; NEMBRINI; KÖNIG; WRIGHT, 2018). Esse processo de permutação permite quantificar a importância de cada regressor para explicar a variável dependente. Na Figura 14 é possível observar os resultados.

É interessante perceber que a ferramenta que se demonstra mais importante para definir a densidade é o número de unidades. Na regulação, um instrumento que apresenta uma função parecida é a cota parte, mas esta calcula a área do terreno sobre o número de unidades, então não exercem exatamente a mesma função. A vantagem de usar a cota parte ao invés do número de

Figura 14 – Importância das variáveis do modelo para explicar a densidade



unidades se encontra na dificuldade de definir um número de unidades para cada localização, enquanto a cota parte faz este cálculo automaticamente, com base no tamanho do terreno (Equação 1.1).

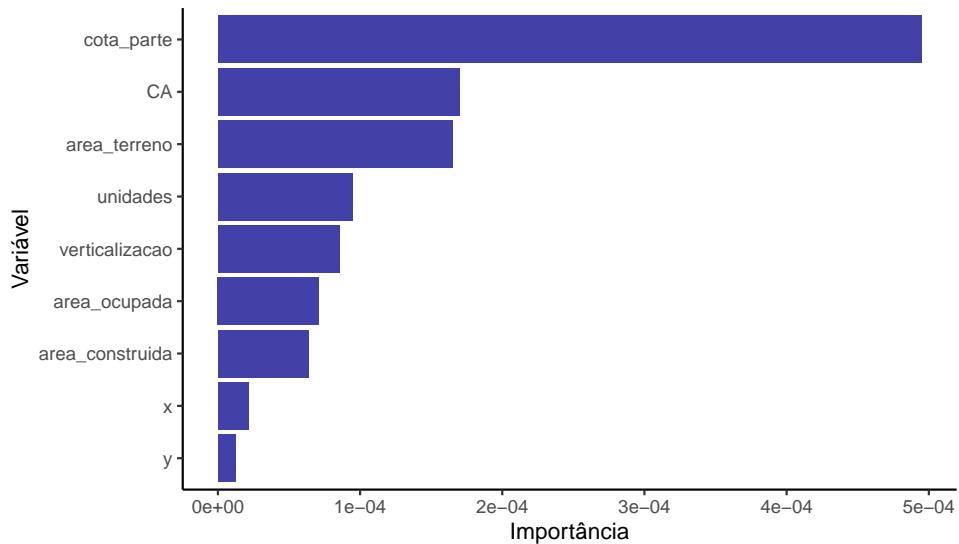
De maneira semelhante, a área construída se apresenta mais relevante em relação ao CA, que é calculado pela área construída sobre a área do terreno. Entretanto, um fator que merece ser destacado é que enquanto estes indicadores utilizam como base a área do terreno, a densidade habitacional é calculada com base na área total da região, que inclui a área de terreno mais lotes não residenciais, áreas públicas como ruas e parques, rios, etc.

Nesse sentido, para criar uma comparação justa, foram realizados novamente os procedimentos, mas agora calculando a densidade populacional como a população sobre a área do terreno, ao invés da área total. Os novos resultados podem ser observados na Figura 15. Agora, a relação se inverteu e a cota parte e o CA se tornam mais importante do que o número de unidades e área construída, respectivamente.

Com a densidade calculada sobre a área do terreno, a performance dos modelos aumentaram bastante, sendo que a regressão linear passa a ter um  $R^2$  na base de teste de 43,6%, a *random forest* de 72,7% e quando incluídas todas as variáveis, o  $R^2$  aumenta marginalmente para 79,6%. Isso significa que as outras variáveis além dos três instrumentos não agregam muitas novas informações para explicar a densidade.

Portanto, os três instrumentos de regulação são mais eficientes ao definir a densidade populacional ao se considerar a área de terreno ao invés da total. Em termos práticos, se a área da região for composta por lotes não residenciais, elementos naturais como rios, ou outros fatores, os instrumentos perdem a eficácia. Entretanto, se um novo lote residencial for construído, apenas com base nos três indicadores é possível ter uma boa ideia de qual densidade ele apresentará,

Figura 15 – Importância das variáveis, mas com a densidade calculada através da população sobre a área do terreno



sendo que outras informações como o número de unidades, não ajudariam muito.

Portanto, os três instrumentos de regulação são mais eficientes para definir a densidade populacional quando se considera a área do terreno, ao invés da área total. Em termos práticos, a eficácia desses instrumentos diminui se a região incluir lotes não residenciais, elementos naturais como rios, ou outros fatores. No entanto, para um novo lote residencial, é possível prever a densidade com base nesses três indicadores, sendo que outras informações, como o número de unidades, não agregariam muita precisão adicional.

Todavia, há outras variáveis que não estão contempladas na base de dados do IPTU ou do censo, que podem ser muito importantes, como proximidade a ofertas de emprego ou meios de transporte de alta capacidade, amenidades, etc. Em outras palavras, não é possível afirmar que os três indicadores utilizados sejam os mais eficientes, mas eles certamente são bons indicativos de densidade populacional.

## 5 Conclusão

Responder categoricamente que os instrumentos são eficientes seria um forte reducionismo, visto que eficiência é um espectro largo e relativamente arbitrário. Ainda assim, os testes de hipótese apresentados ao longo da Seção 4 apontam que o CA e a cota parte são importantes para definir a densidade populacional de uma região, enquanto a verticalização não. Ainda, foi quantificado que a cota parte apresenta uma maior importância, ao comparada com o CA.

Todavia, também foi discutido como outros fatores que não foram medidos podem impactar a densidade. A teoria econômica indica que a proximidade à oferta de emprego é um fator central na fisiologia da cidade, por exemplo. No entanto este fator não é considerado no cálculo do CA ou cota parte permitida em cada região e, portanto, não foi considerada nessa análise. Como uma futura agenda de pesquisa é importante investigar estes outros fatores e como eles se comparam com a eficiência dos atuais instrumentos na determinação da densidade.

Outra descoberta relevante foi sobre a gravidade da miopia que a prefeitura apresenta sobre a moradia na cidade, haja vista os milhões de domicílios que não estão inscritos no Cadastro Imobiliário Fiscal, e, portanto, não pagam imposto e não necessariamente seguem as regulamentações impostas para garantir o cumprimento dos objetivos da cidade. Em termos numéricos, foi estimado que aproximadamente 53% dos domicílios estão cadastrado, enquanto todos os outros estão por fora dos sistemas públicos.

Nesse sentido, mesmo que os instrumentos sejam os mais eficientes possíveis, eles podem impactar no máximo metade da cidade, enquanto a outra metade é regida por instituições informais. Entretanto, é importante destacar que observa-se nessas regiões de moradia informal um adensamento muitas vezes maior do que em regiões regulares. A região mais densa da cidade é a favela de Paraisópolis, o que indica que há forças de adensamento mais fortes que a regulação.

# Referências

- BAUMAN, Z. City of fears, city of hopes. Goldsmiths, University of London, 2003.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.
- BRUECKNER, J. K. **Lectures on urban economics**. [S.l.]: MIT press, 2011.
- ÇALIŞKAN, O.; MASHHOODI, B.; AKAY, M. Morphological indicators of the building fabric: Towards a metric typomorphology. **Journal of Urbanism: International Research on Placemaking and Urban Sustainability**, Taylor & Francis, 2022.
- COASE, R. H. The problem of social cost. **The journal of Law and Economics**, 1960.
- FUJITA, M. **Urban economic theory**. [S.l.]: Cambridge university press, 1989.
- GLAESER, E. **Triumph of the city: How urban spaces make us human**. [S.l.]: Pan Macmillan, 2011.
- IBGE. **Notas técnicas**. Rio de Janeiro, Brasil: Instituto Brasileiro de Geografia e Estatística, 2013. Acesso em: 2024-06-23. Disponível em: <[https://www.ibge.gov.br/apps/snig/v1/download/SNIG\\_notas\\_tecnicas.pdf](https://www.ibge.gov.br/apps/snig/v1/download/SNIG_notas_tecnicas.pdf)>.
- IBGE. **Censo Demográfico 2022: Malha de Setores Censitários Preliminares**. Rio de Janeiro, Brasil: Instituto Brasileiro de Geografia e Estatística, 2024. Acesso em: 2024-06-23. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv102072.pdf>>.
- JACOBS, J. **The death and life of great American cities**. [S.l.]: New York: Random House, 1961.
- LIMA, B. A. A. d. **Para além da forma urbana. Conflitos e contradições socioambientais da cidade compacta proposta para os eixos de adensamento do Plano Diretor Estratégico de São Paulo de 2014**. Tese (Doutorado) — Universidade de São Paulo, 2021.
- Município de São Paulo. **Lei nº 10.235, de 16 de dezembro de 1986**. 1986. Dispõe sobre a forma de apuração do valor venal de imóveis, para efeito de lançamento dos Impostos sobre a Propriedade Predial e Territorial Urbana, concede descontos sobre os valores venais dos imóveis sujeitos à incidência desses impostos, no exercício de 1987, e dá outras providências.
- NEMBRINI, S.; KÖNIG, I. R.; WRIGHT, M. N. The revival of the gini importance? **Bioinformatics**, Oxford University Press, v. 34, p. 3711–3718, 11 2018. ISSN 14602059.
- PAPAGEORGIOU, Y. Y.; PINES, D. **An essay on urban economic theory**. [S.l.]: Springer Science & Business Media, 2012. v. 1.
- SÃO PAULO. **Lei nº 16.050, de 31 de julho de 2014**: Aprova a política de desenvolvimento urbano e o plano diretor estratégico do município de são paulo e revoga a lei nº 13.430/2002. 2014.
- WRIGHT, M. N.; ZIEGLER, A. ranger: A fast implementation of random forests for high dimensional data in c++ and r. **arXiv preprint arXiv:1508.04409**, 2015.