

Measuring Spatial Accessibility to Public High Schools in Brazil: New Methods and Tools

Felipe Horta Lemos Vieira de Oliveira

College of Social Sciences, College of Computer Sciences, Minerva University

Economics and Society, Data Science and Statistics

February 2023

Executive Summary

Aims and Background

We propose a new method for measuring accessibility to education and apply it to measure accessibility to **public high schools in Brazil**. We also create a **web app to visualize such access**. Despite high government spending on education in Brazil, low educational outcomes and high dropout rates persist. Public high schools host most students (87,4% of total enrollment) but perform drastically worse than their private counterparts (INEP, 2022). People from lower socioeconomic status and people of color face higher dropout and test failure rates, an inequality that helps the perseverance of poverty. A significant component of the meager achievement is the lack of access to quality schools, which are unevenly distributed geographically. The proposed tool aims to **identify areas with low access to public high schools so policymakers can redirect funds accordingly**.

Method

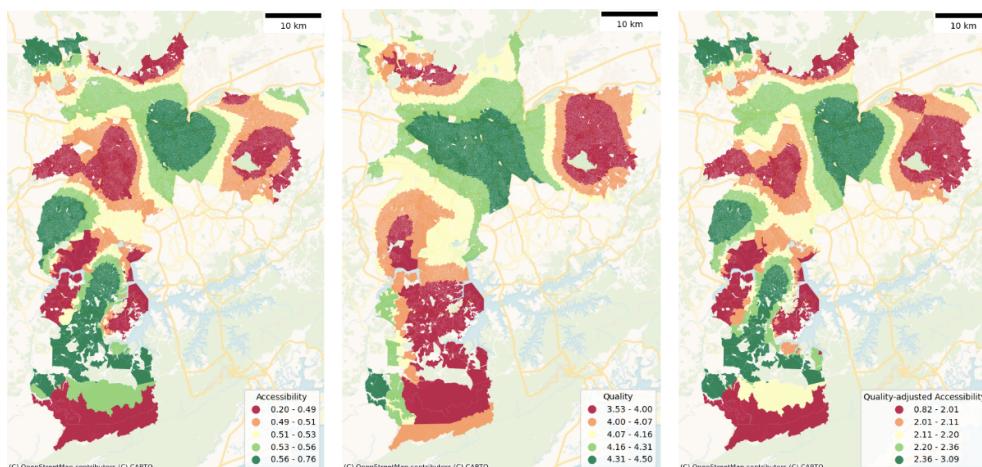
We conduct a literature review of existing spatial accessibility methods, concluding that there is a gap in the literature for a method that accurately reflects how families select schools. Crucially, **existing methods ignore the quality of education**, a fundamental aspect of access that determines student performance (Banerjee & Duflo, 2011).

The proposed method extends the existing models in three ways:

- **Adds a quality component to the selection weight**, representing that families will prefer the best schools in their region.
- **Calculates an average quality metric for each region**, defined as the average quality of surrounding schools weighted by the access to each school.
- Finally, it **combines the two metrics to calculate a measure of quality-adjusted supply**, which provides a unified view of access to education in a region.

Results

We find inequalities between racial, income, and regional groups in Brazil. In general, rich, white regions which are located disproportionately in the South and Southeast have higher quality and higher supply of public high schools. The results indicate that better targeting of investments in education could reduce inequalities and accelerate national economic development.



Abstract

This paper proposes a new measure of access to education. We use the proposed measure to assess access to public high schools in Brazil and create a tool that policymakers can use to identify low-access regions. The proposed method and its application improve upon existing measures of access by ignoring arbitrary geographical boundaries, accounting for competition among schools, and adding a more accurate model of families' preferences regarding schools, which involves both distance costs and their assessment of the school's quality. The study uses data from the Brazilian Demographic Census and the School Census to create a map of relative accessibility at the census tract level. The information can be accessed through a web application and visualized as an interactive map. The results highlight the unequal distribution of educational resources in Brazil: rich, white regions in the South and Southeast have a significantly higher supply and quality of public high school schools. Enhancing policymakers' understanding of the educational landscape in their region of influence is the first step towards a fairer distribution of resources and an important contribution to the reduction of inequality in Brazil.

Table of Contents

| | |
|---|-----------|
| Executive Summary | 2 |
| Aims and Background | 2 |
| Method | 2 |
| Results | 2 |
| Abstract | 3 |
| Table of Contents | 4 |
| Introduction | 6 |
| Public High Schools in Brazil | 7 |
| Literature Review | 10 |
| Interactive tools on education | 13 |
| Data | 16 |
| Schools | 16 |
| School Quality | 18 |
| Population | 18 |
| Distance | 20 |
| Method | 22 |
| An Overview of Metrics of Spatial Accessibility | 22 |
| Quality | 25 |
| Calculating the accessibility metric to public high schools in Brazil | 28 |
| Distance weight | 32 |
| Interpretation of the accessibility metric | 35 |
| Creating the Web Application | 36 |
| Results | 36 |
| Region | 37 |
| Urban/Rural | 38 |
| Poverty | 38 |
| Race | 39 |
| Case Study: São Paulo | 40 |
| The Tool | 46 |
| Limitations and Next Steps | 47 |
| Conclusion | 48 |
| References | 50 |
| Appendix | 58 |
| Supplemental Materials | 58 |

Introduction

Education is unarguably a fundamental component of the sustainable development of a country (United Nations, 2022). In addition, if guaranteed to all citizens, it can be an essential agent for social equity, giving the entire population the tools to prosper. As Nobel prize economist Amartya Sen put it, "education makes us the human beings we are. It has major impacts on economic development, on social equity, gender equity. In all kinds of ways, our lives are transformed by education and security" (White, 2004).

This paper analyzes the spatial accessibility of public high schools in Brazil. We propose a method to measure the spatial accessibility of public high schools in Brazil and examine the extent to which disparities in access to educational facilities are associated with regional inequality and socioeconomic status. We build a tool that policymakers can use to identify areas with low access with high precision across the country. This study's findings will help inform policymakers and stakeholders on where to improve access to education in Brazil by highlighting areas with limited access to educational facilities. In addition, this research will contribute to the existing literature on spatial accessibility to education by providing an in-depth analysis of the spatial distribution of educational facilities in Brazil.

According to federal regulation in Brazil, it is the State's duty to provide free public education to everyone between the ages of 4 and 17. Furthermore, every child should be guaranteed a spot in the closest school to their household (L9394, 1996). In practice, the guarantee of a spot means that schools are often overpopulated, and families are denied a place in their school of preference. In addition, for many children, the closest school is too far to be considered accessible in any practical sense (Goulart et al., 2019). Policymakers need to ensure adequate access to schools, especially for rural and low-income neighborhoods, which first

requires an accurate measure of accessibility to direct resources to locations needing them the most.

It is important to note that there are other components to this inequality in access. There is an economic aspect, such as the high cost of private schools and low quality of public ones or the high-dropout rates of lower-income students because of a necessity to work (FGV, 2022). There is also a social component, with many students not being interested in the content of the schools (FGV, 2022). Although important, these components are beyond the scope of this paper.

This paper is organized as follows: Section 1 introduces the scope of the paper and discusses the selection of our subject of analysis (public high schools in Brazil). I discuss how a lack of information on educational accessibility and the structure of the education system call for a tool to identify regions of low school access in the country. Section 2 discusses the data sources, details about each variable, how some values are estimated, and data limitations. In Section 3, I propose a new method for measuring educational accessibility in Brazil, one that considers quality as well as the supply of schools. I discuss the evolution of methods for calculating spatial accessibility and the limitations of each one. In Section 4, I present the study's results, showing statistics about the accessibility to public high schools across Brazil. Using the proposed method, I use São Paulo as a case study for a regional analysis. Finally, in Section 5, I discuss the implications of our findings and the limitations and suggest the next steps for future research.

Public High Schools in Brazil

Brazil is an adequate case study for the accessibility of public education for several reasons. It is a geographically large country, with a high contrast between regions in terms of geography, development, and educational outcomes. Furthermore, despite being one of the

countries that spent the most on education relative to their GDP (World Bank, 2023), it still suffers from low educational outcomes and high school dropout rates. The reason for the contrast between high spending and low performance is that money is directed to low-efficacy programs or low-demand regions. There are reports of schools being built in areas with no demand, while families in regions, usually more peripheral, must travel hours to reach the closest school (Parajara, 2008). There needs to be more understanding of the current educational landscape to direct funds where they are needed. Despite having one of the world's richest demographic and school datasets, no existing system in Brazil can quickly utilize this information to indicate areas of high need.

Additionally, Brazil has a problem with educational inequality. The OCDE considered it one of the greatest challenges to be overcome in educational policy (2021). People from lower socioeconomic status and people of color have consistently higher dropout and test failure rates and get accepted less into the best universities (Neri, 2009). Critically, there needs to be more supply of quality education. In some regions, over 55% of students that are not at school reported a lack of transportation as one of the reasons for evading (MEC, 2015). Furthermore, long travel times directly influence school attainment and cause physical and emotional wear (Goulart et al., 2019).

Poorer communities tend to have lower access to quality education, resulting in lower educational attainments and lower future earnings, mortality, and disease rates. Each of these factors, conversely, is associated with lower school attendance rates. Kids who are sick will miss more days of school, and kids who cannot afford to live in central locations tend to need to drop out altogether more often (Banerjee & Duflo, 2011). Furthermore, kids from poorer households are more likely to suffer pressure from their families to be an income providers, which is one of

the main reasons cited for dropping out of school in Brazil (IBGE, 2019). This feedback loop of low access to education results in lower development for the individual and less human capital for the country, which means lower labor productivity and less development for the nation as a whole (Barro, 1996).

Although Brazil has both public and private schools, I have decided to focus our study only on public high schools. Public schools host the majority of students - 88% of students in high schools were in public schools in 2019 (DEED, 2019). Additionally, the private sector already has much better educational outcomes, suggesting that accessing it is not a matter of spatial distance and low supply, but high costs (DEED, 2019). The method and tool I propose are better utilized by a central planner who has the leverage to redirect resources at a city, state, or nationwide level, as is the case for public schools. In Brazil, public high schools have three main types: state schools (97% of total enrollment), federal, and municipal. The Union is responsible for creating the guidelines for education in the country, but the implementation of public high schools is mainly the responsibility of individual states, which should coordinate with each of its municipalities how they want to organize the high school education system. With the exception of a few federal public high schools and some municipal schools in large cities (mostly São Paulo), the vast majority of students enroll in state-administered schools. The funding of these schools comes mainly from state and municipal taxes and is complemented by the Union, whose primary goal is to reduce the inequality between states¹. After receiving the funds, states have considerable autonomy in where to spend it (within some constraints) (L9394, 1996). Our tool's foremost purpose is to identify areas with low access to public high schools so that the organization in charge can redirect funds accordingly. Given the structure of public education

¹ There is also a contribution from company taxes, but that money is used only for administrative costs like salaries and food provisions, not for maintenance or development of new schools.

administration and resource management in Brazil, the target audience of such a tool is state education administrators.

High school is the level of education with the highest dropout and absenteeism rates. For comparison, nearly every student in primary education attends school (PNAD, 2020). Attending high school was not mandatory until 2013, partially explaining the low attendance. Furthermore, the proportion of children that need to work to earn a living, which is the biggest reason for school evasion, increases with age and is the highest for teenagers in high-school age (Neri, 2009). Focusing on one school level is necessary to ensure that our representation of demand is accurate. If we analyzed more than one educational level simultaneously, students would be double-counted since each person can have a demand for either high school or primary school at a given time, not both. In the analysis, we can select a specific age range (14-17), ensuring that those people only have a demand for high schools.

Literature Review

Few studies have considered spatial accessibility to education in Brazil. Most researchers study access from a sociological, historical, or economic perspective. A spatial analysis of access has often been disregarded, even though spatial accessibility is a critical component when evaluating equity in a community (Allen & Farber, 2020; Dadashpoor et al., 2016; Taleai et al., 2014). Among the few studies that attempted a geographical analysis of the school supply in Brazil, nearly all of them utilized the p-median technique, which consists of determining the area of influence of each school, given by all census tracts for which a given school is the closest. Then, by dividing the school capacity by the number of school-aged people in the influence area, one can see if there is enough supply (Pegoretti, 2005; Pizzolato et al., 2004, 2012). These

studies are examples of area-based techniques, which measure access as the supply-to-demand ratio in a given area (total supply in an area divided by total demand in the same area). This type of technique ignores how distance may affect the accessibility of different families that live within a given area. Furthermore, the technique assumes that families always choose the school closest to them, which ignores other factors necessary for school selection, such as school quality.

Another study analyzed potential access in the city of São Paulo by looking at the average time to a student's three closest schools and found that families living in the periphery of the city have less access to high-quality schools (Pizzol et al., 2021). Again, this method only measures access to the closest schools and ignores other variables influencing school choice. Furthermore, its units of observation are large, being extensive geographical areas themselves, so people that live far away are grouped into one location (the centroid of the spatial unit), lowering the accuracy of the results.

Abroad, however, especially in the United States and China, this field of research has made tremendous strides. Since the proposal of a supply and demand analysis of accessibility by Luo and Wang (2003), at first in the context of healthcare, many researchers have applied these methods to education. Some studies have looked at educational access by analyzing differences between regions at the national level, and some have looked at differences between neighborhoods at the city level (Walsh et al., 2015; Wang et al., 2021). Some even looked at the family level (Davis et al., 2019). Studies of spatial access to education resources have consistently found unequal levels of access to schools among different races, income levels, professions, and locations (urban/rural areas, inner-city/suburb) (Davis et al., 2019; Ogryzek et

al., 2022; Raju et al., 2020; Wang et al., 2021; Xu et al., 2018). However, studies have yet to make a comprehensive, multi-level analysis at the national level, using small observation units.

Most studies have used a simple formulation of access which considers only the distance between schools and families, oversimplifying how families choose their school and the differences between schools. It is well-researched that students prefer higher-quality schools when resources are unevenly distributed (Xiang et al., 2018). Educational equity should consider not only the possibility of a student attending a school but the quality of education they receive. There is a lack of a method with a realistic model of families' choices and preferences and a holistic view of access, which considers quality as well as quantity, especially in the Latin American context (Vecchio et al., 2020).

Interactive tools on education

Figure 1
Catálogo de Escolas

The screenshot shows the 'Catálogo de Escolas' search interface from the 'Censo Educação Básica'. At the top, there are dropdown filters for Region (Sudeste), State (SP), Municipality (Todos os Valores de Colunas), School Name (Escola Waldorf), Location (Selecionar Valor), Localização Diferenciada (Selecionar Valor), Dependência Administrativa (Selecionar Valor), Etapa e Modalidade de Ensino (Selecionar Valor), and Porte da Escola (Matrículas) (Selecionar Valor). Below the filters is a button labeled 'Aplicar'. The main area displays three tables of school data:

| Restrição de Atendimento: ESCOLA EM FUNCIONAMENTO E SEM RESTRIÇÃO DE ATENDIMENTO | | | |
|--|---|--------------------------|---|
| Código Escola | Nome da Escola | UF | Município |
| 35004877 | ESCOLA WALDORF CASA DO BOSQUE | SP | Itapeiranga |
| CEP | Endereço | Categoria Administrativa | Ela pa de Ensino |
| 18200-970 | VERA MARTA DE LARA, 71B JARDIM NOVO AEROPORTO. 18200-970 Itapeiranga - SP | Privada | Creche, Pré-Escola, Anos Iniciais do Ensino Fundamental |

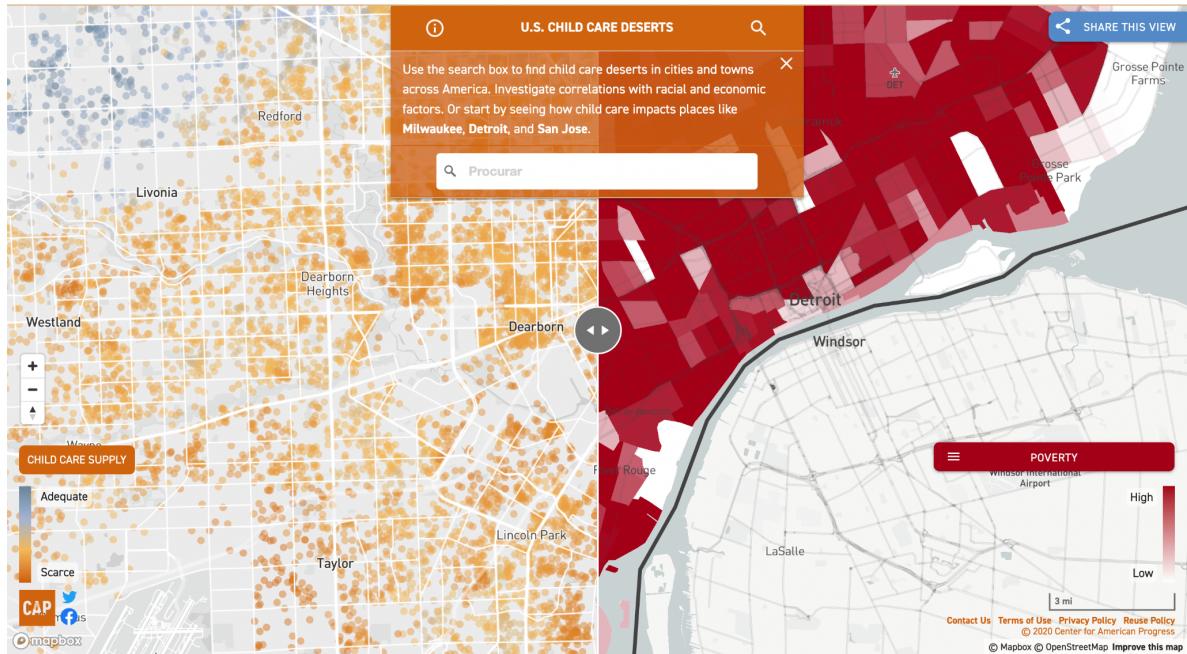
| Restrição de Atendimento: ESCOLA EM FUNCIONAMENTO E SEM RESTRIÇÃO DE ATENDIMENTO | | | |
|--|--|--------------------------|---|
| Código Escola | Nome da Escola | UF | Município |
| 35006143 | ESCOLA WALDORF FLAMBOYANT | SP | Avaré |
| CEP | Endereço | Categoria Administrativa | Ela pa de Ensino |
| 18706-000 | CARLOS RAMIRES, 651 VILA SANTA IZABEL. 18706-000 Avaré - SP. | Privada | Creche, Pré-Escola, Anos Iniciais do Ensino Fundamental, Atividade Complementar |

| Restrição de Atendimento: ESCOLA EM FUNCIONAMENTO E SEM RESTRIÇÃO DE ATENDIMENTO | | | |
|--|--|--------------------------|-------------------------------------|
| Código Escola | Nome da Escola | UF | Município |
| 35007194 | ESCOLA WALDORF JASMIN | SP | Santo André |
| CEP | Endereço | Categoria Administrativa | Ela pa de Ensino |
| 09190-640 | ITAPORANGA, 104 PARAISO. 09190-640 Santo André - SP. | Privada | Anos Iniciais do Ensino Fundamental |

No publicly available tools measure educational access on a national scale in Brazil. The closest platform that uses school data is the *Catálogo de Escolas*, a school catalog hosted online by INEP, the governmental research institute responsible for running the School Census. Users of this tool can search for schools in a specific region, filtering for some characteristics (INEP, 2023a). Despite pulling from a large dataset, the tool is simple, and only displays results in a table format. It does not measure accessibility; it merely provides detailed information on specific schools upon request.

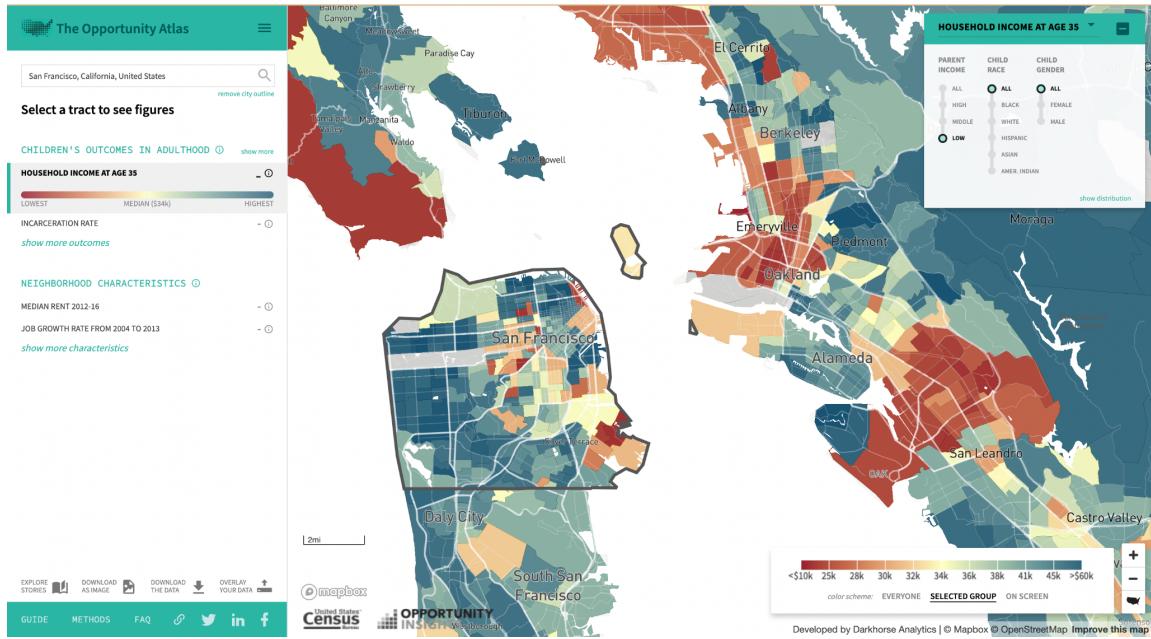
Figure 2

Childcare Deserts



In the United States, the best example of making an accessibility map was done by the organization American Progress, which built a map of childcare deserts using the 2SFCA technique and census data (Center for American Progress, 2020). Their tool has two views separated by a slider, which allows the user to compare, for the same region, the accessibility to primary childcare centers and some demographic feature, such as poverty rates. The tool is intuitive to use and responsive. It encompasses all of the United States at the census tract level.

Figure 3
The Opportunity Atlas



Another prominent example of a public tool using geographic data is the Opportunity Atlas, which shows the life outcomes of people born in different neighborhoods across the US (United States Census Bureau & Opportunity Insights, 2018). Also encompassing all of the US at the census tract level, the tool lets users filter for a specific subpopulation and see various outcomes for them and how they differed by neighborhood. Another feature that makes this map engaging and provides a deeper understanding of the data is responsive charts which update as the screen moves across the map.

In conclusion, although several researchers have studied spatial accessibility, there is a gap in the literature for a method that includes the quality of schools in the accessibility metric. Furthermore, no researchers have used a method to calculate accessibility for an entire country while keeping the observation units small for high precision. The gap is even more prominent in

Brazil, where few studies have looked at spatial accessibility. In addition, there is no tool that accesses and displays this information for policymakers and the general public. The most similar tool, the Childcare Deserts website, measures a different education level in a different country. Still, previous contributions' features are essential in a new solution. In terms of methods, we will build on previous spatial accessibility models, how researchers have measured quality for schools, and how to integrate quality into the accessibility metric (never done in the education context). For the tool, we want to keep the design, ease of use, and way of presenting information of both the Childcare Desert and the Opportunity Atlas.

Data

Schools

The data on the schools was gathered from the 2019 Censo Escolar (INEP, 2022), a yearly survey of every school in Brazil. The data includes several features of the school, including the administration type (public or private), the number and education of teachers, the number of students, some data on the infrastructure, and the school's address.

In our analysis, I only looked at public high schools, which include Municipal, State, and Federal schools, each managed by their respective jurisdiction. I considered all schools that had at least one active high-school class. The dataset includes data on 20,500 public high schools in the country. On average, each High School has 10 classes, 323 students, and 8 teachers². This represents an average of 31 students per class and 38 students per teacher.

² There is a potential underreporting of teachers and classes in this survey, but exploring that is beyond the scope of this capstone.

The geocoded location of the schools is downloaded from the *geobr* package in Python (Pereira, 2022), which contains the official spatial datasets of Brazil. The package uses OpenStreetMap to geocode the addresses provided by the schools. 2,023 schools (10%) in the dataset do not have geolocation, so we cannot use them in our estimates. If there is a systematic relationship between the location of the school and the availability of data, our results could be biased. If, for example, OpenStreetMap does not work as well in rural areas, the accessibility of these regions would be underestimated since the model would undercount schools. Unfortunately, there is little we can do to avoid this bias except to be aware that it might be present.

Our access metric requires a measure of the supply of each school. The ideal variable to measure would be the total capacity (maximum number of students it can host). This number, however, is not available in our dataset, so we use the number of classrooms as a measure of how many students the infrastructure of the school can host.³ Although there is no limit set by law on the number of students per classroom nationally, there have been proposals to set a hard limit of 35 students per classroom in high schools (PL 4731/2012, 2012). Using 35 as our best guess for the maximum number a classroom can hold, our estimate for school capacity becomes $35 \times \text{number of classrooms}$.

This is not a perfect proxy, as the classroom size can vary depending on the structure of the school. Some schools cannot host 35 students, so this metric is likely an upper bound. We should consider the possibility that schools with less demand will likely reduce the number of classes, perhaps repurposing a classroom. If that is the case, we would potentially confuse low demand for low supply and underestimate access levels in regions with low demand.

³ Another candidate for a proxy was the number of teachers. However, from a conversation with an educator in a public school, I discovered that there is a lot of variance in the hours each teacher works since a single teacher can lead one or multiple classes a week.

School Quality

The measure of school quality used is the Primary Education Development Index (Índice de Desenvolvimento da Educação Básica - IDEB), which is an indicator that combines information from the SAEB (an evaluation conducted every two years with every public school in the country and is the primary metric used by the government to track educational achievement) and approval rates of students. This gives each school a holistic score that combines educational achievement and completeness (INEP, 2023b). This metric is the standard for measuring the quality of education in Brazil by the government. Importantly, we assume that parents and students also use this metric (or other metrics that correlate with it) as their quality measure.

The original IDEB score goes from 0 to 10, but in our analysis, we change the range to 0-1 to be consistent with the literature and give more easily interpretable results. The average school in Brazil scores 0.4 in our quality measure, with a standard deviation of 0.07.

Table 1

Characteristics of the schools

| | Mean | Standard Deviation |
|--------------------|-------------|---------------------------|
| Number of teachers | 8.40 | 13.445 |
| Number of students | 323.16 | 289.47 |
| Estimated capacity | 367.28 | 289.08 |
| Quality | 0.40 | 0.07 |

Population

The data on students and their location is taken from the 2010 Census, conducted by the Brazilian Institute of Geography and Statistics (IBGE). The dataset was downloaded from the

Base dos Dados database (Base dos Dados, 2022), which pre-cleaned the data using Data Zoom, a statistical package developed by the Economics Department of PUC University (PUC-Rio, 2022). The census has extensive data on the population of Brazil at a census tract level. Census tracts are the smallest unit of observation for the census and comprise an average of 613 people. For each census tract, we get the number of people aged 15 to 17 as our population (prospective high school students). In Brazil, high school is mandatory for students in that age group. We also get the racial distribution of the census tract as a whole and of our age subset, average monthly income, number of total households, total population, and gender distribution. Since we do not have the precise address of each family, we set the population's location at their census tract's centroid. The geometry of the centroids is also downloaded from *geobr*.

The data encompasses 310,120 census tracts, each with an average number of 613 people. We have an average of 34 high-school-aged students per census tract, a total of 10,336,610 potential high-schoolers. The racial categories in the census are white (average of 47% in each census tract), Black (7%), pardo⁴ (44%), Asian (1%), and indigenous (0.7%).

Table 2

Characteristics of the Census Tracts

| | Mean | Standard Deviation |
|--------------------------------|---------|--------------------|
| Number of households | 185.31 | 102.34 |
| Number of people | 613.52 | 352.41 |
| Average Monthly Earnings | 1229.58 | 1375.38 |
| Number of people aged 15 to 17 | 34.02 | 22.65 |
| Percentage Black | 6.7 | 9.30 |
| Percentage White | 47.42 | 27.41 |
| Percentage Indigenous | 0.78 | 7.35 |
| Percentage Pardo | 44.04 | 25.37 |

⁴ Pardo is a broad term used by IBGE to refer to mixed-race individuals, usually a mix between Black and white, but indigenous are sometimes also considered).

| | | |
|------------------|------|-------|
| Percentage Asian | 1.02 | 3.049 |
|------------------|------|-------|

Distance

The distance between each family and the school is calculated as the Euclidean distance between the centroid of the census tract and the location of the school. A better measure of distance would be travel time or distance using the road networks, which accounts for differences in the transport infrastructure of each location. However, calculating a distance matrix of this form proved computationally unfeasible. The Euclidean distance is calculated on a plane, so the first step is to project the latitude and longitude points, which effectively squishes the surface area of Brazil onto a plane. The projection chosen was the SIRGAS 2000 / Brazil Mercator⁵, which is the official projection for Latin America.

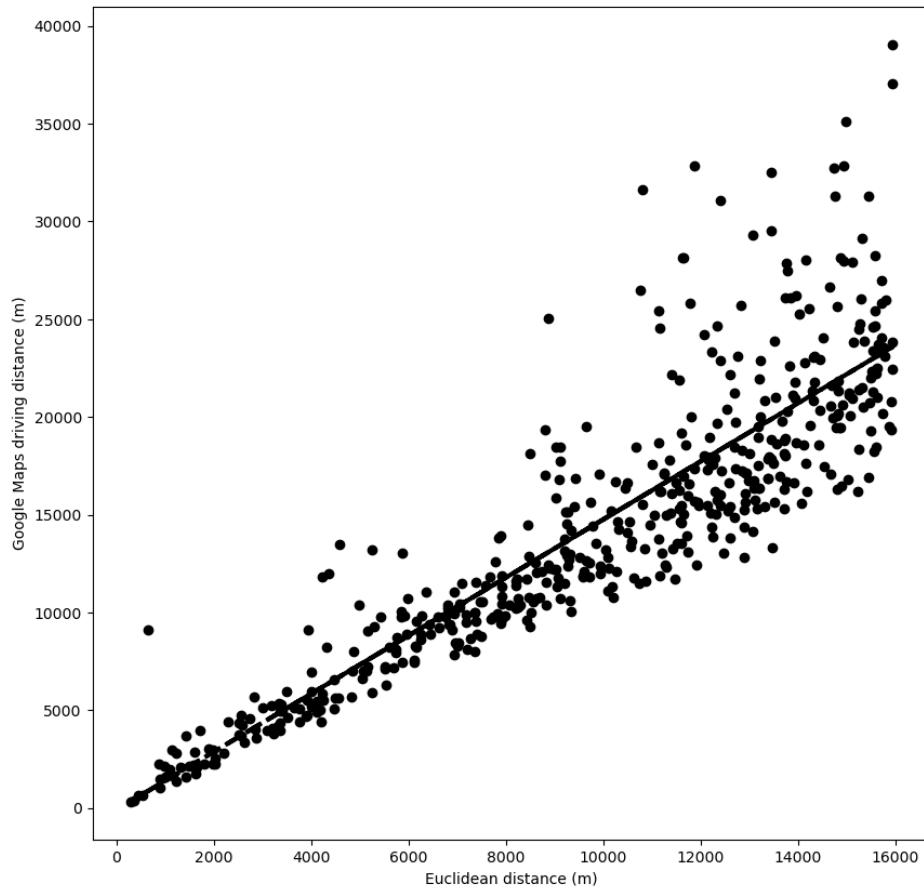
As a sanity check, I computed the driving distance of 1,000 random census tract-school pairs using the Google Maps API and compared the results with the Euclidean distance. We can be confident that the Euclidean distance is a good approximation of the real driving distances if we see a linear relationship between the two variables. We find that there is in fact a linear relationship, but it is not homoskedastic. The driving and euclidean distance are tightly correlated for short distances (except for one outlier), but the correlation weakens as distances increase, resulting in a funnel shape. This analysis highlights that although the Euclidean distance can be a reasonable approximation, we should be less confident in our results when distances are large. This is especially true for rural regions since, in addition to having larger a smaller school density and consequently larger average distances from census tracts to schools,

⁵ <https://epsg.io/5641>

they have a less developed road network, which increases the variation in driving distances between points.

Figure 4

Comparison between Google Maps distance and Euclidean distance



With over 300,000 census tracts and 18,000 schools, calculating every pair distance would result in a matrix of over 5 billion entries, which is computationally inviable. However, as our model sets the maximum distance people are willing to travel to reach their school, we can ignore distances bigger than this threshold, which we have set to be 16km (see Methods section). This drastically reduces the size of the resulting matrix since we only need to write the distances

below this threshold. However, to calculate the distances, we would still need to check every pair point, which is computationally expensive. My solution is to split the calculation into geographical chunks. In practice, we divide our dataset based on the 138 mesoregions of Brazil and calculate the distance matrix of each one separately. That way, we only have to check the distance for pairs of points that can reasonably be close to each other. The downside is that using this method, we assume that people cannot cross mesoregions to go to school, which is unrealistic since there is no official border between them. However, this should happen rarely due to the large geographical size of the mesoregions.

Method

I create a measure of access to high-school education by combining the supply and demand of schools in each area. I base our approach on the enhanced three-step floating catchment area method (E3SFCA), proposed by Wan et. al. ([2012](#)), to measure spatial access to health services. Besides adapting the method for the education context, I improve upon it by including a measure of school quality.

An Overview of Metrics of Spatial Accessibility

Understanding the educational landscape of a region requires creating a simplified model within the complexity and computational constraints of researchers while still accurately representing the overall state and behavior of the system. Over time, models measuring spatial accessibility to education have evolved in accuracy and complexity. The first and most basic way to measure accessibility in a region was to calculate the provider-to-population ratio in each area. In practice, this means dividing the number of spots available in each high school by the number of children of high school age. The resulting value is interpretable but has a significant

limitation: it assumes that children can only access schools in the same area as them. This is flawed because the area borders are often arbitrary, and there are no real barriers to crossing them.

Attempting to solve this problem, Joseph & Bladock (1982) proposed using a gravity model, which assumes that the closer a population gets to a provider, the more "attraction" there is between them, that is, the more likely and more able they are to use their services. An easier-to-interpret reformulation of the gravity model was made by Luo & Wang (2003), who combined it with a floating catchment model. It works in two stages: first, determine a catchment area, which is the maximum distance between a provider and a population. We can think of this area and the maximum distance a person would be willing to travel to access a provider. Calculate the distance-adjusted supply-to-demand ratio for each provider:

$$R_j = \frac{S_j}{\sum_{k \in \{Dist(k,j) \leq d_0\}} P_k}$$

where S_j is the number of providers in j , and P_k is the population in the location k , which must be within the catchment (the distance between k and j must be smaller than d_0).

Then, sum up the supply-to-demand ratio of each provider around an area to get the access level of that area.

$$A_i = \sum_{j \in \{Dist(i,j) \leq d_0\}} R_j,$$

where A_i is the accessibility index of location i , $Dist(i, j)$ is the distance between i and j , and d_0 is the catchment area.

This model has been widely used, but it has two significant limitations.

1. It does not differentiate between people living in the same catchment area, so it does not weigh in the difference in distances within each catchment.
2. It does not consider how the demand for a provider might change depending on the availability of other suppliers near a population.
3. It only considers distance as the cost of accessing schools.

One proposed solution to fix the first problem is adding a weight W representing the cost of the distance between two locations. This method is called the enhanced two-step floating catchment area method (E2SFCA) (Luo & Qi, 2009). They proposed that the weight be determined by Gaussian decay function:

$$W(d) = e^{\frac{-d^2}{\beta}},$$

where d is the distance between two locations and β is the impedance weight.

This weight exponentially decreases as the distance increases. This weight is then added in equation 1 by multiplying P_k with W_{kj} (weight of the distance between k and j). Now, before dividing the supply by the demand, we weigh each demand point by their distance to the provider. We also add W in step 2, multiplying R_j (the supple-to-demand ratio of each provider) by W_{ij} , representing how people have less access to schools further away.

Although the model fixes the problem of not considering differences in distance within a catchment, it does not address the other two. For our measure of access, we need a metric that

addresses all the issues above. A good candidate is the three-step-floating-catchment-area 3SFCA, proposed by Wan et. al. (2012). This model assumes that the demand for a provider is influenced by the availability of other providers nearby, so it assigns a travel-time-based competition weight to each pair of populations and providers.

That is done in an extra step before the ones described in the previous method. For each population, calculate the selection weight as

$$G_{ij} = \frac{W_{ij}}{\sum_{k \in \{Dist(i,k) \leq d_0\}} W_{ik}},$$

where G_{ij} is the selection weight between location i and service site j , $Dist(i, k)$ is the distance from i to any provider k within the catchment, and d_0 is the catchment size. W_{ij} and W_{ik} are the assigned distance weights for j and k , respectively.

The selection weight is included in the previous equations just like the distance weight: we include it as a denominator when calculating R_j and as a multiplier when calculating the total accessibility of a region A_i .

Quality

The three-step method is considered superior to the others because it fixes the problem of overestimation of demand (Wan et al., 2012). However, it still leaves out a major component of education: the school's quality. Families do not pick schools based solely on the distance. They also likely consider a variety of other factors (Holmes Erickson, 2017), the most prominent and

easily measurable of which is the quality of education they will have at a given location. Furthermore, it is not sufficient to look at spatial accessibility when analyzing educational equity. Equity should not be measured only based on who can access *a* school but on what *kind* of school people can access and how that varies depending on their location. In the book Poor Economics, Esther Duflo and Abhijit Banerjee show that having enough supply of schools is not enough to foster learning. An equally important component is the quality of schools. As they put it, "getting children into school is a very important first step: this is where learning starts. But it isn't very useful if they learn little or nothing once they're there" (Banerjee & Duflo, 2011, p. 144). Having both quantity and quality dimensions of access for a region will provide a holistic view of each region's education availability.

In previous studies, researchers have included quality into accessibility metrics by either multiplying a quality score by the supply when calculating the final accessibility score, thus imbuing one in the other, resulting in a unified metric (Hu et al., 2020), or by calculating a separate metric for average quality (Wang et al., 2021). Mathematically, the two methods are one and the same since multiplying the average quality metric by the accessibility metric will result in the same value as combining quality from the start. They are just different ways of presenting the results. In the first approach, the resulting metric will consider quantity and quality of supply, providing in a single number the adequacy of education in a region. However, with this method, we cannot differentiate areas that need more schools (low quantity) from sites that need more investment for the existing schools (low quality). By calculating average quality for regions separately, we can get a more detailed understanding of problems in each region, but we might not get the big picture. In this paper, I will use both strategies to isolate variations in

quality from variations in supply but also get a unified number that can be used to judge the adequacy of a region's educational access quickly.

I modify the 3SFCA method by including the quality of high schools in two ways. First, we include a measure of quality as a component in the selection weight - thus representing the idea that people consider quality when choosing schools and higher-quality schools have higher demand. Second, we calculate the average quality of education for each census tract in addition to the accessibility metric. Later, we combine the quality and quantity metrics to get one unified "quality-and-distance-adjusted-supply" metric.

To include school quality as a criterion for school selection, we rewrite the selection weight G_{ij} , which previously only had a measure of distance (meaning that people considered how far a given school j is compared to other schools k available). We add a measure of quality to the selection weight G_{ij} , which becomes:

$$G_{ij} = \frac{W_{ij}C_j}{\sum_{k \in \{Dist(i,k) \leq d_0\}} W_{ik}C_k},$$

where C_j is the quality of school j . Now, the population considers not only the distance of a given school compared to other schools in the area but the quality of a school compared to others in the area. Another way of thinking about G_{ij} is the probability that one would choose school j over all other schools k within the maximum distance radius.

To calculate the quality of education in a census tract, we compute the average quality of schools within a tract's catchment area, weighted by the accessibility A_{ij} between that census tract and each school.

$$Q_i = \frac{\sum_{j \in \{Dist(i,j) \leq d_0\}} C_j A_{ij}}{\sum_{j \in \{Dist(i,j) \leq d_0\}} A_{ij}}.$$

Calculating the accessibility metric to public high schools in Brazil

Step 1: Assign a selection weight to each census tract-school pair. For each census tract, we select a catchment area, find all providers within that area, and determine a distance weight W using a Gaussian function (see Distance Weight section for a discussion on the choice for β).

$$W(d) = e^{\frac{-d^2}{10000}}.$$

We also assign a selection weight G for each high-school-census tract pair, which in the original 3SFCA method is their distance weight divided by the sum of all the distance weights from nearby providers. We modify it by adding a measure of quality to G .

Another way of thinking about $G_{i,j}$ is the probability that one would choose school j over all other schools k within the maximum distance radius. It models how people have a higher probability of going to a close-by or higher-quality provider if they have the option.

$$G_{ij} = \frac{W_{ij}C_j}{\sum_{k \in \{Dist(i,k) \leq d_0\}} W_{ik}C_k},$$

where G_{ij} is the selection weight between the census tract i and high-school j , $Dist(i, k)$ is the distance from i to any school k within the catchment area and d_0 is the maximum distance that determines the catchment size. W_{ij} and W_{ik} are the assigned distance weights for j and k , respectively. C_j and C_k are the quality of school j and schools k , respectively. The selection weight equals 1 when only one school is available within the catchment of a census tract but decreases as the number of alternative schools increases.

Step 2: Calculate the distance-adjusted capacity-to-population ratio for each school. For each school, we calculate the capacity-to-population ratio weighted by the distance from the supplier to the census tract. We first define a catchment area for each school using the same method as before and find all census tracts within that catchment. We define the population of each tract P_i as the number of children from 15 to 17 years old in that tract. The supply of the school S is defined as the number of classrooms in a given school. We then calculate the distance-adjusted capacity to population ratio for a given school as

$$R_j = \frac{S_j}{\sum_{k \in \{Dist(k,j) \leq d_0\}} P_k W_{jk} G_{jk}},$$

where S_j is the capacity of school j , P_k is the population in the census tract k , d_{jk} is the distance between the census tract k 's centroid and the school (in our model, we measure the euclidean distance between locations), W_{jk} is the distance weight function that represents the assumption that more distant schools are less accessible, and G_{jk} is the selection weight for the tract-school pair. If two providers have the same capacity, but one has more young children nearby, that provider's R_j will be smaller. This represents how nearby children will compete for the same spots in school.

Step 3: Sum up the supply value for all schools within their catchment area for each census tract, weighting each school by the distance. We start by drawing the catchment area around each centroid. We then sum up the provider-to-population ratios we calculated in the previous step of each school in the area, weighing each by the distance and the selection weights.

$$A_{ij} = R_j W_{ij} G_{ij},$$

$$A_i = \sum_{j \in \{Dist(i,j) \leq d_0\}} A_{ij},$$

where A_{ij} is the accessibility between a census tract i and a school j , while A_i is the accessibility index for each census tract i .

Step 4: Calculate the average school quality in each census tract, weighted by the accessibility index between the census tract and each school.

In the previous step, we calculated A_{ij} , which is essentially the access a location i has to a school j . To calculate the average quality, we will weigh each school in the catchment area by their accessibility score since the higher the score, the more of that school's resources are received (or accessed) by the location.

$$Q_i = \frac{\sum_{j \in \{Dist(i,j) \leq d_0\}} C_j A_{ij}}{\sum_{j \in \{Dist(i,j) \leq d_0\}} A_{ij}},$$

where Q_i is the average education quality of location i , C_j is the quality of school j , and A_{ij} is the accessibility score between location i and school j .

Finally, we calculate an integrated, quality-adjusted access metric

$$H_i = A_i Q_i$$

Where H_i is a quality-adjusted accessibility metric for location i , which combines both the quantity of supply, the distance, and the quality of schools in the area.

I used the *access* package for Python (Saxon et al., 2021), which has a built-in method for applying the 3SFCA method, as the primary reference for calculating accessibility metrics. The package simply applies most of the mathematical calculations described above efficiently. See their GitHub repository for more details. However, this existing solution can only implement the basic 3SFCA method; it does not have the capability of including quality in the calculation, as described above. Therefore, we modified the source code to adapt its functions to consider quality in the selection weights and calculate the average quality of each census tract.

Distance weight

Most of the literature uses a Gaussian distribution to determine the distance weights, but they disagree on what value to use for the impedance weight β . The problem is that its choice can significantly impact the final access metric results (although the relative access values stay relatively the same) (Wan et al., 2012). Ideally, the distance weight should represent the actual preferences of families concerning distance. Since we do not have direct information on people's preferences, we can consider the recommended commute to school times derived from academic literature and government regulation.

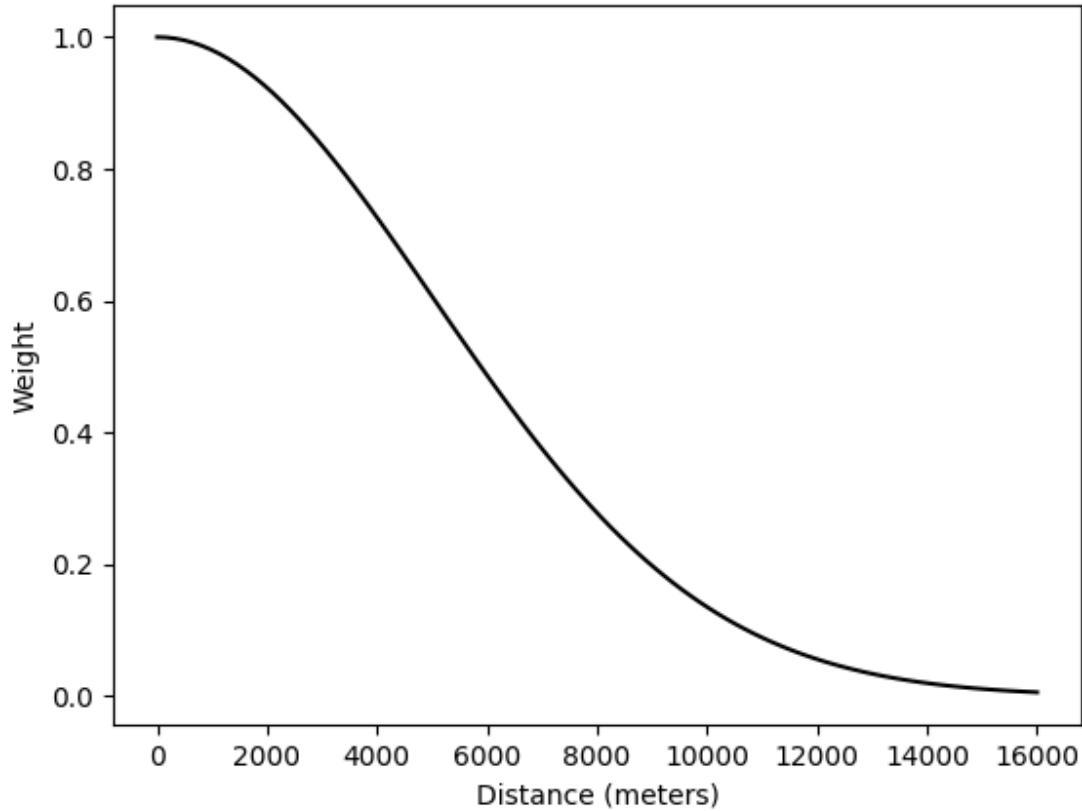
There is no official guideline for the maximum time a student can travel to reach their school (L9394, 1996), as this number will depend highly on contextual factors. Thus, we must turn to the relevant academic literature. The maximum recommended time in Brazilian literature varies between 30 and 60 minutes (Goulart et al., 2019; Pegoretti, 2005) each way. I try to be consistent with most of the literature (Raju et al., 2020; Wang et al., 2021; Xu et al., 2018) (although it seems like many authors set this value arbitrarily) and attempt to approximate a catchment size of 60 minutes, which is in the upper range of what has been done in previous papers. For this estimation, I will use point estimates when possible but upper bounds when there is a range. That is because legislation and academic literature will tend to underestimate how far people are willing (or are required) to travel. For example, the same paper that suggested a maximum of 30 minutes in transport also recorded that some students may spend way longer than this, with some recorded transport times taking 4:30 hours both ways (Goulart et al., 2019). Additionally, having an upper range is appropriate because we want to include rural communities that often do not have schools nearby.

Since we only have the Euclidean distance between schools and census tracts, we need to convert the maximum time to a maximum distance, using the average transportation speed to make a best guess. Most people in Brazil travel to school or work by bus (G1 & Brasília, 2015), so we can use these transportation methods in our calculation. We have data for the average speed of a bus in São Paulo, which was 16km/h in 2019 (Pinho, 2019). Assuming that every student uses a bus (again, an unreasonable assumption that provides an upper bound), buses drive at 16km/h (São Paulo likely has higher average speeds than other cities because of its better infrastructure) and the distance between schools and census tracts is a straight line, we can calculate that 16km is an upper bound for the maximum distance such that the transport time is not over 60 minutes each way.

Thus, we want to set β such that the preference for a school over 16km is close to 0. We set it as $\beta = 10000$, such that when the distance is up to 4km, the preference is relatively high (72% at 4km), low preference for 8km and above (28% at 8km), and almost zero when the distance is over 16km (<0.05%).

Figure 5

Weights according to distance, following a gaussian function



As a sanity check, we can look at data on the average time spent going to work in Brazil. Recent research found the state capital city with the largest commute time was Rio de Janeiro, with an average of 47 minutes, and the smallest average commute was Porto Alegre, with an average of 30 minutes. Additionally, 18% of people had a commute of over an hour (G1, 2013). Considering that our estimates are an upper bound and that commute to work tends to be longer than the commute to school (Pizzol et al., 2021), the data suggests that our estimates are an appropriate reflection of the population's preferences regarding school distances and maximum commuting time.

We assume that the travel times for a given distance are the same regardless of location, which is unrealistic. Rural areas, with smaller road networks, would have longer travel times

than urban areas, and therefore access to these areas is underestimated in the present model. Future research should consider these differences and utilize a more robust method for calculating the maximum distance and distance weights.

Interpretation of the accessibility metric

The accessibility statistic generated by the method is a combination of the demand and supply for high-school spots in a given region. It can be thought of as the supply of high-school spots adjusted by distance and nearby demand and considering the competition between schools. However, there is no "unit" with a clear interpretation. Some authors interpret it at the adjusted slots-to-student ratio in a region, but this interpretation is highly sensitive to the arbitrary selection for the Gaussian weights and the catchment area size. Furthermore, it breaks apart when including the selection weight, which necessarily reduces the accessibility value for all observations, but that reduction does not represent more limited access, it simply reflects the preference of students for closer-by schools over further-away schools when given the option.

The only reasonable interpretation of the raw accessibility metric is as a relative indicator of access. Therefore, it only makes sense to analyze a unit's metric in comparison to others. In other words, we can only make statements in the form "census tract X has twice as much access to public high schools than census tract Y" or "region X has high access relative to other regions in the country".

It's worth remembering that the metric is a composite of several aspects of accessibility, and does not have a direct translation to a simple countable value such as 'teachers per student'. In broad terms, it is a comparison of the supply and demand of schools in a region, accounting for both the distance and quality of schools. It makes several assumptions about how people choose schools and how they value quality and distance costs. As such, the absolute value of the

metric can be quite sensitive to changes in the assumptions, but the relative comparisons, which are the main interest of this research, are robust.

Creating the Web Application

I built the website using the Streamlit Python library, which also takes care of the hosting. The dataset with the accessibility metrics was hosted in Google Cloud and accessed through an API. For the visualization, we used the Plotly library, which allows the creation of highly customizable and interactive maps.

Results

I calculated three main metrics for each census track in Brazil. The first measure is the spatial accessibility to schools, which can be interpreted as the relative accessibility level of a region to public high schools. Note that this is not simply an area-based measure of total supply divided by total demand. This is a distance-based measure, meaning values are adjusted by the distance from students to the schools and competition between schools. The second measure is the average quality of schools around a neighborhood, adjusted by the accessibility between each given school and the neighborhood. The third metric is a combination of the previous two: quality-adjusted spatial accessibility, which is the accessibility multiplied by the average quality, thus resulting in a unified metric that measures the adequacy of schooling resources in a region, both in terms of quantity and quality.

Previous studies in Brazil have calculated area-based accessibility levels for individual cities (Maller & Gandolpho, 2014), travel times to the closest schools (Pizzol et al., 2021), and average quality of education (Alves & Silva, 2013). There has never been, to my knowledge, a neighborhood-level, nationwide, holistic metric of accessibility. In other locations, researchers

have calculated distance-based accessibility metrics, but never on a national level. Additionally, few studies have considered school quality as an important factor in accessibility, and, to my knowledge, no studies have created a unified metric for education accessibility that includes quality.

Inequality in education can take many shapes and forms and appear at different levels of analysis. For a better understanding of the current educational landscape in Brazil it is necessary to analyze different levels of scale. With a geographical lens, we will look first at the macro level by analyzing the regional differences in access. We also look at how different types of geography might have unequal access to schools by analyzing the contrast between rural and urban areas. Then, we focus on one city to look at the micro, and neighborhood level to look at how the dynamic of accessibility inequality presents itself in one city and how it interacts with the other components. We also analyze differences from a demographic lens, looking at how different races and income strata differ in access to education. It is only by looking at these different levels of analysis that we can create a big picture of the dynamics of educational inequality in Brazil while simultaneously having a good understanding of the local distribution of school resources.

Region

Table 3 shows the average accessibility by region. There is a large regional variation in both quality and quantity of school supply across Brazil. Interestingly, the North region performs the best in terms of quantity (0.731) and the worst in quality (0.273). The South and Southeast region have high relative values of both quantity and quality. These regional differences are a reflection of the different levels of investment and development in these regions. The South and Southeast regions concentrate the majority of the wealth in the country, while other regions tend to be neglected.

Table 3*Outcome variables for regions in Brazil*

| | Northeas t | Nort h | Sout h | Southe ast | Midwe st |
|-------------------------|---------------|-----------|-----------|---------------|-------------|
| Supply | 0.513 | 0.731 | 0.717 | 0.68 | 0.664 |
| Average Quality | 0.338 | 0.273 | 0.393 | 0.414 | 0.359 |
| Quality-Adjusted Supply | 0.195 | 0.254 | 0.295 | 0.285 | 0.276 |

Urban/Rural

There is a huge contrast between rural and urban regions, with urban areas having nearly double the adjusted supply (0.727 vs 0.372). The urban areas also have much higher average quality, at 0.399 versus 0.304 for rural areas. As a country that has experienced an explosion in urbanization in recent years, there is a tendency to ignore the needs of the countryside areas.

Table 4*Outcome variables for urban and rural regions*

| | Urban | Rural |
|-------------------------|-------|-------|
| Supply | 0.727 | 0.372 |
| Average Quality | 0.399 | 0.304 |
| Quality-Adjusted Supply | 0.295 | 0.148 |

Poverty

We create two income groups, one of the census tracts where the earning of the main household earner is below the Brazilian poverty line, which is R\$ 406 (IBGE, 2020), and one where it is above the poverty line. We find that census tracts below the poverty line have much lower accessibility levels (0.405) and quality levels (0.285). Note that this is a national comparison; within cities, it is likely that this trend is reversed since there is a bigger supply of public high schools near low-income families who cannot afford private schools.

Table 5

Outcome variables of people below and above the Federal Poverty Line (R\$406)

| | Below FPL | Above FPL |
|-------------------------|------------------|------------------|
| Supply | 0.405 | 0.678 |
| Average Quality | 0.285 | 0.39 |
| Quality-Adjusted Supply | 0.151 | 0.276 |

Race

Supply is similar across racial demographics, but it is slightly higher for indigenous people (0.76) and slightly lower for black people (0.63). Indigenous people have, on average, access to lower quality schools, at 0.259 average quality against 3.7 for the other groups. White people have the highest average school quality, at 0.398.

Table 6

Outcome variables across racial categories

| | Black | White | Indigenous | Pardo | Asian |
|-------------------------|--------------|--------------|-------------------|--------------|--------------|
| Supply | 0.631 | 0.691 | 0.76 | 0.646 | 0.663 |
| Average Quality | 0.363 | 0.398 | 0.259 | 0.364 | 0.375 |
| Quality-Adjusted Supply | 0.243 | 0.285 | 0.27 | 0.252 | 0.263 |

The big picture is quite clear. Urban, rich, white regions (which are predominantly located in the South and Southeast) have a high level of educational access, while rural, poor regions with the majority of people of color systematically have lower access to quality schools. Naturally, these individual inequalities are intertwined. White people earned 74% more than people of color in 2019. They are also more likely to live in central areas with more schools (and more likely to put their children in private schools, which have better educational outcomes). The

result of over 300 years of slavery, unequal urbanization, and lack of inclusive policies⁶ is a country with deep inequality, reflected in inequitable access to education. Although this trend does not come as a surprise, it highlights a great barrier to equity and development in Brazil. By keeping the population that was historically marginalized without access to education, the cycles of underdevelopment are perpetuated.

Case Study: São Paulo

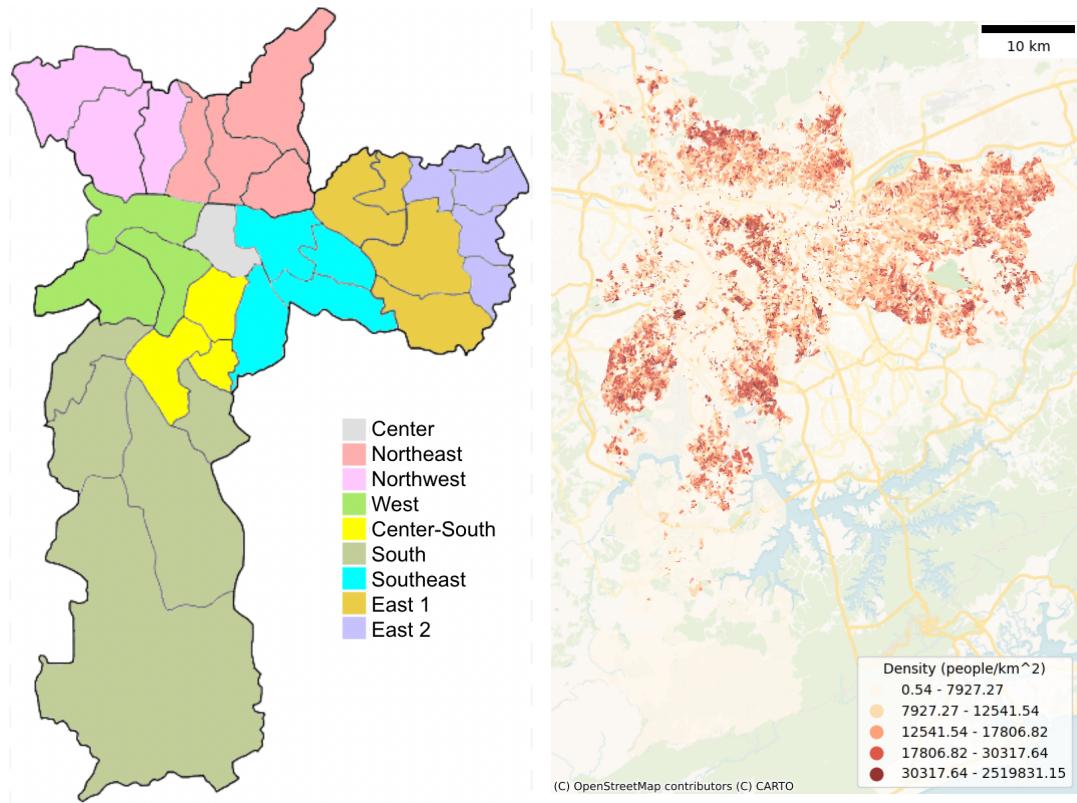
Although looking at nationwide results can generate some interesting insights about how education varies across the country, it omits a crucial dynamic of education inequality, which is local by nature. With neighborhood-level data, the biggest power of the dataset is seen at a city-wide level. Visually, one can clearly see how school accessibility is related to neighborhood demographics.

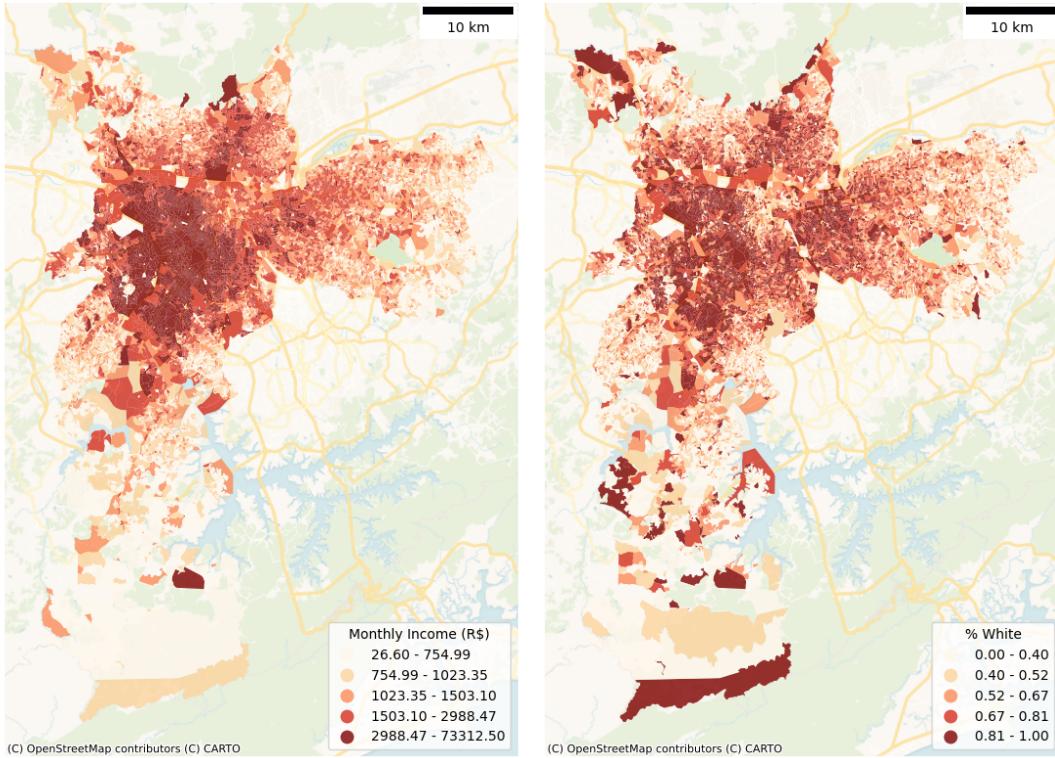
São Paulo is a good example for our analysis as the largest and one of the most unequal cities in the country. With over 12 million inhabitants and a Gini index value of 0.65 (IBGE, 2010), São Paulo has deep segregation and social disparities. Its large geographical size implies regional differences in demographics and educational attainment between its neighborhoods (Moraes & Belluzzo, 2014).

⁶ That has been changing recently in education. For example, with the widespread adoption of income and racial quotas to public universities.

Figure 6

Administrative regions of São Paulo and density, income, and racial composition maps





Maps of density (top-right), monthly income (bottom-left), and percentage white (bottom-right) in the city of São Paulo. Rich, white neighborhoods are mainly located in the central and West regions, while the peripheral neighborhoods are poorer and more densely populated

Most of the population in São Paulo lives in the East, North, and Southern regions (note that the Southern region is the biggest and most people live in the northernmost part of it) (Fig 6). The most populous areas are also the poorest and the ones with higher concentrations of Black and Pardo people. Wealth tends to be concentrated in the Central and West areas (Fig 6).

Figure 7

Accessibility, Average Quality, and Quality-Adjusted Accessibility in São Paulo

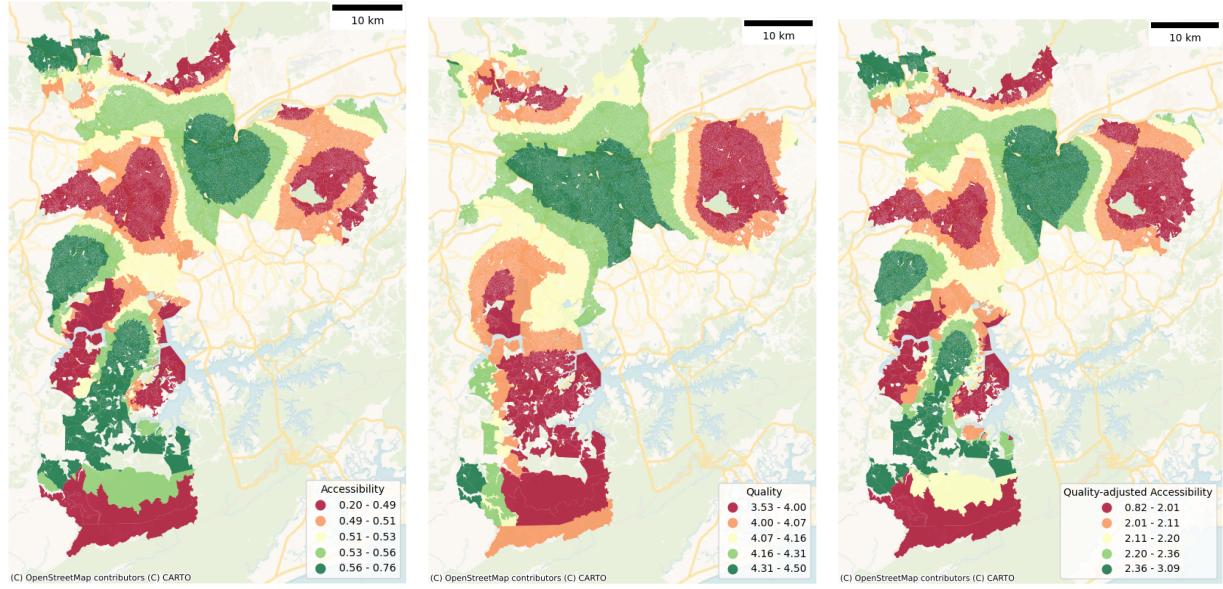


Figure 7 plots the spatial accessibility, average quality, and quality-adjusted accessibility of schools. We see that there are three clusters of high-access regions. The southernmost one is in a region with very few people, so not many students are competing for the same spots, thus increasing the slots per student. The middle cluster, on the other hand, is in a highly-populated region, which implies the presence of many schools in the area. Both of these regions, despite having high spatial accessibility, have lower-quality schools. High-quality schools are mainly in the third high-access cluster in the central region, where the wealthiest people are. There are four clusters of low-accessibility regions. One in the West zone, one in the East zone, one in the northern periphery of the North zone, and one around the South Zone.

To check the validity of our results and ensure our metric of accessibility accurately reflects the accessibility in the real world, we can compare our findings with evidence from other sources. Ideally, regions that are shown to have low accessibility based on our map of São Paulo

should be the same zones where people experience a lack of access. We look at multiple sources to paint a clearer picture of regions that experience inadequate accessibility.

The reason for the low accessibility to public schools for each region is very different. The West Zone contains the richest neighborhoods in the city, with an average income of R\$ 2.174,55 (IBGE, 2010). The neighborhood follows a general trend in Brazil where richer families tend to prefer private schools over public ones (Adrião, 2009). The West Zone contains 5 out of the 10 neighborhoods with the highest percentage of students in private schools, and all of them have less than 31% of students in public schools, compared to an average of 75% for the city (Rede Nossa São Paulo, 2021). We can deduce that the demand for public schools is very low in this region, which explains the low supply and consequently low accessibility levels. In fact, the low interest in public schools more than compensates for the low supply. Schools from some of these neighborhoods have to "import" students from other neighborhoods to keep schools from closing, forcing students to long and often exhausting commutes (Parajara, 2008). The other two low-accessibility regions, however, have a high demand for public schools. They are the regions with the lowest infrastructure, access to public transportation, and income.

There is some evidence that these regions are, in fact, suffering the most from low access to schools. A recent journalistic report found that 8 out of the 9 neighborhoods where people have reported a need for more slots in public high schools were in the South Zone, and all of them were in neighborhoods with low access in the model (Moreno & Ianelli, 2022). The report also shows that many families are being forced to move their children to private schools, increasing the financial burden on an already low-income population.

In 2022 there were over five thousand children without a spot in any schools (SP1, 2022). They can wait for several years on a waitlist until being able to enroll. Of these children, a vast

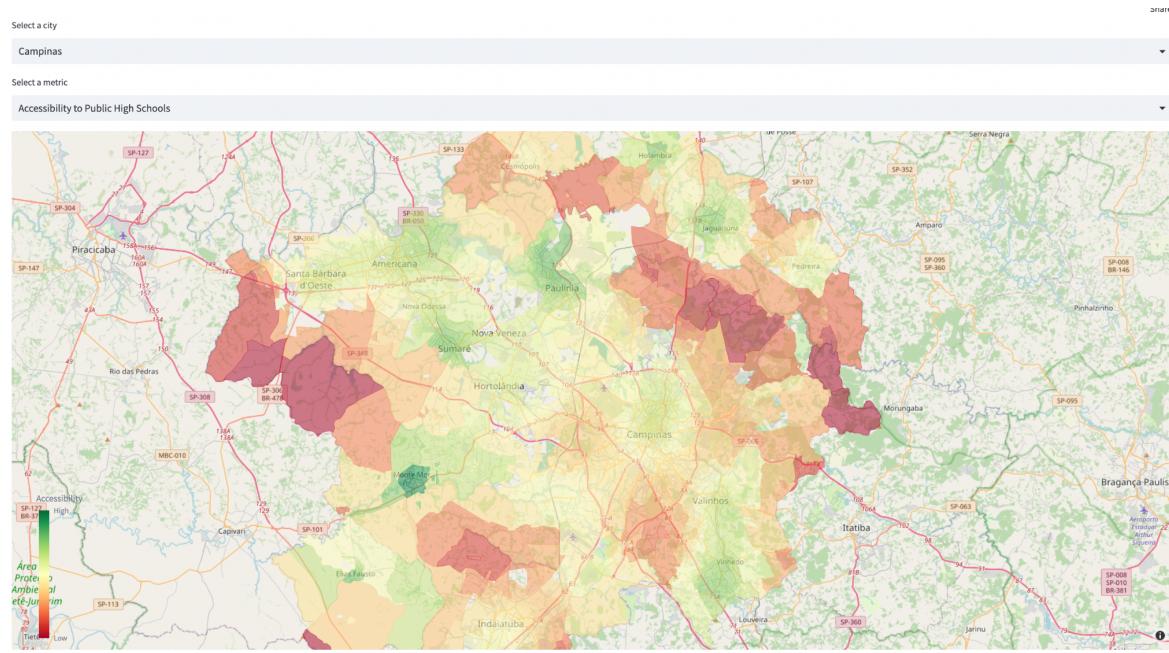
majority of them live in the South and East Zones, corroborating the findings of the model. Daniel Cara, a Professor of Education at São Paulo University summarizes the landscape of educational access in the city: "the demand for school enrollment in a city like São Paulo is already irregular. In the expanded center, the demand is much lower, there are even several vacant spaces. And on the periphery, it's the opposite." (SP1, 2022)

Although no single piece of evidence proves that the proposed metric has internal validity, the combination of news texts, expert commentary, and reports conducted by the city indicate that the regions of low access in the model correspond to the same regions that do in fact suffer from low access to public high schools.

The Tool

Figure 8

Screenshot of Mapa do Ensino Médio website



I built a website that allows policymakers and the general public to quickly visualize the public high school education access distribution of any city in Brazil. The tool will help policymakers, educational professionals, and other members of civil society better plan educational infrastructure and development in the country. The tool, called Mapa do Ensino Médio (High School Map) allows users to select a city and a metric (average quality, distance-adjusted supply, or quality-and-distance-adjusted supply) or a demographic variable (percentage white or income) and visualizes them as an interactive map.

The map color-codes the metric or variable for easy interpretation; users can click on specific census tracts to see more detailed information. The tool is currently hosted here: <https://felipehlvo-access-to-education-map-app-u0pb18.streamlit.app/>. Despite being a simple

prototype, the website already contains the core functionality, which allows users to quickly visualize educational accessibility at a micro-level and investigate how it might correlate with demographic factors. I plan to continue refining the tool over the next months, ideally with the help of web development professionals (I am in communication with several news media companies to discuss a potential collaboration).

Limitations and Next Steps

While this paper provides a new, improved method for measuring educational accessibility and implements it for the first time at a national scale in Brazil, it still has several limitations that should be addressed in future research. A significant limitation is that we do not have data on all the schools (around 10% are missing geolocation). A related issue is that the location of families is estimated based on the centroid of the census tracts, which could lead to measurement errors in the distances between them and the schools, especially when census tracts are large. Another important limitation regarding distance is that we are currently measuring the cost of access as the Euclidean distance. A better method would be to use travel times since the same distance can correspond to vastly different commutes depending on the region. For this to work, we need to have access to at least the road network of Brazil, but ideally also the transport mode of students and the public and private transportation schedules and routes.

We make several assumptions regarding how families choose schools. We ignore potentially important components such as the presence of friends and family members in the school body, cultural barriers, and subtle forms of corruption that often happens in these systems (Lopes & Toyoshima, 2013). We make estimates given the available information regarding school capacity and school quality. For the former, we use classrooms as a proxy. For the latter,

we rely on government estimates which might not be the same as the quality perceived by parents and students. Finally, there are data quality issues. We based our demographic information on the 2010 Census, which is 13 years out of date⁷. There is some anecdotal evidence that false reporting of students is widespread in the School Census because school funding increases based on the number of students registered.

Regarding the tool, several improvements can be made, especially in computational efficiency and UI design. Currently, only one city can be loaded at a time, whereas ideally, the entire map would be available at once, and the user could freely zoom, scroll, and pan. However, due to the size of the dataset, that is not feasible using the computational resources available.

Despite these limitations, the proposed accessibility metric uncovers salient features of the educational landscape of Brazil and how resources are distributed. The combination of a realistic method to measure education access, its application into an existing dataset at a national level with census tract precision, and a tool that facilitates the investigation of the data is a meaningful contribution to the educational sector in Brazil. It can influence the redistribution of resources to make the country more equitable and accelerate development.

Conclusion

Brazil has a long history of inequality and segregation, which is clearly reflected in its education sector. Still, there had not been a rigorous analysis of the accessibility to schools at a national level. In this paper, we proposed a new methodology for measuring access to education, which includes a measure of the quality of schools, thus providing a more accurate measure of how families actually make decisions about where to enroll. The measure is distance rather than area-based, models competition between schools, and two variables that influence preference

⁷ The 2020 Census was delayed because of the Covid-19 pandemic. It should come out later this year.

(quality and distance). We applied the methodology in Brazil using nationwide data at the census tract level, thus creating a comprehensive yet detailed metric of accessibility that can be easily compared across areas at regional and local levels. We built a tool to visualize the data, making it easy for policymakers and the general public to understand access in their region and make more informed decisions about where to live or invest.

We found that education resources are unevenly distributed in Brazil. Urban, rich, white regions (predominantly located in the South and Southeast) have a high level of educational access, while rural, poor regions predominantly comprised of people of color systematically have lower access to quality schools. We also investigate the local dynamics of education distribution, finding that in Sao Paulo, the South and East regions have a low supply and quality of schools, despite housing most of the population. Wealthier regions like the Center and the Center-West have higher quality and a disproportionately high supply compared to their demand. The analyses highlight that Brazil needs to focus on making education more equitable by understanding which regions have lower access and redirecting resources to them.

References

- Adrião, T. (2009). Indicações e Reflexões sobre as Relações entre Esferas Públicas e Privadas para a Oferta Educacional no Brasil. *Políticas Educativas – PoLED*, 3(1), Article 1.
<https://seer.ufrgs.br/index.php/Poled/article/view/22531>
- Allen, J., & Farber, S. (2020). Planning transport for social inclusion: An accessibility-activity participation approach. *Transportation Research Part D: Transport and Environment*, 78, 102212. <https://doi.org/10.1016/j.trd.2019.102212>
- Alves, T., & Silva, R. M. da. (2013). Estratificação das oportunidades educacionais no Brasil: Contextos e desafios para a oferta de ensino em condições de qualidade para todos. *Educação & Sociedade*, 34, 851–879.
<https://doi.org/10.1590/S0101-73302013000300011>
- Banerjee, A. V., & Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight.... PublicAffairs*.
<https://www.goodreads.com/book/show/10245602-poor-economics>
- Barro, R. J. (1996). *Determinants of Economic Growth: A Cross-Country Empirical Study* (Working Paper No. 5698). National Bureau of Economic Research.
<https://doi.org/10.3386/w5698>
- Base dos Dados. (2022). *Censo Demográfico*. Base Dos Dados.
https://basedosdados.org/dataset/br_ibge-censo-demografico?bdm_table=microdados_domicilio_1970
- Center for American Progress. (2020). *Do you live in a Child Care Desert?* Do You Live in a Child Care Desert? <https://childcaredeserts.org/>
- L9394, no. 9.394, National Congress (1996).

https://www.planalto.gov.br/ccivil_03/leis/l9394.htm

PL 4731/2012, no. PL 4731/2012 (2012).

<https://www.camara.leg.br/propostas-legislativas/560047>

Dadashpoor, H., Rostami, F., & Alizadeh, B. (2016). Is inequality in the distribution of urban facilities inequitable? Exploring a method for identifying spatial inequity in an Iranian city. *Cities*, 52, 159–172. <https://doi.org/10.1016/j.cities.2015.12.007>

Davis, E. E., Lee, W. F., & Sojourner, A. (2019). Family-centered measures of access to early care and education. *Early Childhood Research Quarterly*, 47, 472–486.

<https://doi.org/10.1016/j.ecresq.2018.08.001>

DEED. (2019). *Resumo Técnico IDEB 2019*.

https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/resultados indice_desenvolvimento_educacao_basica_2019_resumo_tecnico.pdf

FGV. (2022). *Retorno para Escola, Jornada e Pandemia*. <https://cps.fgv.br/RetornoParaEscola>

G1, D. (2013, October 24). *Quase 20% levam mais de uma hora para chegar ao trabalho, diz Ipea*. Brasil.

<http://g1.globo.com/brasil/noticia/2013/10/nas-grandes-cidades-186-levam-mais-de-1h-para-chegar-ao-trabalho.html>

G1, D., & Brasília, em. (2015, October 14). *Principal meio de locomoção dos brasileiros é andar de ônibus ou a pé*. Economia.

<http://g1.globo.com/economia/noticia/2015/10/principal-meio-de-locomocao-dos-brasileiros-e-andar-de-onibus-ou-pe.html>

Goulart, L. M. L., Morais, A. A. de, & Vieira Jr, N. (2019). Tempo de permanência no transporte escolar sobre o desempenho estudantil. *INTERRITÓRIOS*, 5(9), 244.

<https://doi.org/10.33052/inter.v5i9.243594>

Holmes Erickson, H. (2017). How do parents choose schools, and what schools do they choose?

A literature review of private school choice programs in the United States. *Journal of School Choice*, 11(4), 491–506. <https://doi.org/10.1080/15582159.2017.1395618>

Hu, S., Song, W., Li, C., & Lu, J. (2020). A multi-mode Gaussian-based two-step floating catchment area method for measuring accessibility of urban parks. *Cities*, 105, 102815. <https://doi.org/10.1016/j.cities.2020.102815>

IBGE. (2019). *PNAD Educação 2019*.

https://biblioteca.ibge.gov.br/visualizacao/livros/liv101736_informativo.pdf

IBGE. (2020, November 12). *Síntese de Indicadores Sociais: Em 2019, proporção de pobres cai para 24,7% e extrema pobreza se mantém em 6,5% da população* | Agência de Notícias. Agência de Notícias - IBGE.

<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/29431-sintese-de-indicadores-sociais-em-2019-proporcao-de-pobres-cai-para-24-7-e-extrema-pobreza-se-mantem-em-6-5-da-populacao>

INEP. (2022). *Censo Escolar*. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep.

<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/censo-escolar>

INEP. (2023a). *Catálogo de Escolas*. <https://inepdata.inep.gov.br/analytics/saw.dll?dashboard>

INEP. (2023b). *Índice de Desenvolvimento da Educação Básica (Ideb)*. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep.

<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/ideb/i>

ndice-de-desenvolvimento-da-educacao-basica

Joseph, A. E., & Bantock, P. R. (1982). Measuring potential physical accessibility to general practitioners in rural areas: A method and case study. *Social Science & Medicine*, 16(1), 85–90. [https://doi.org/10.1016/0277-9536\(82\)90428-2](https://doi.org/10.1016/0277-9536(82)90428-2)

Lopes, L. S., & Toyoshima, S. H. (2013). EVIDÊNCIAS DO IMPACTO DA CORRUPÇÃO SOBRE A EFICIÊNCIA DAS POLÍTICAS DE SAÚDE E EDUCAÇÃO NOS ESTADOS BRASILEIROS. *Planejamento e Políticas Públicas*, 41, Article 41. [//www.ipea.gov.br/ppp/index.php/PPP/article/view/265](http://www.ipea.gov.br/ppp/index.php/PPP/article/view/265)

Luo, W., & Qi, Y. (2009). An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & Place*, 15(4), 1100–1107. <https://doi.org/10.1016/j.healthplace.2009.06.002>

Luo, W., & Wang, F. (2003). Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a Case Study in the Chicago Region. *Environment and Planning B: Planning and Design*, 30(6), 865–884. <https://doi.org/10.1068/b29120>

Maller, R., & Gandolpho, A. A. (2014). LOCALIZAÇÃO DE ESCOLAS DO ENSINO FUNDAMENTAL: CASO DE ITAIPAVA/RJ. *Revista de Engenharia da Universidade Católica de Petrópolis*, 8(2), Article 2.

MEC. (2015, April 19). *Falta de transporte dificulta acesso à escola*.

<http://portal.mec.gov.br/ultimas-noticias/201-266094987/2567-sp-787759183>

Moraes, A. G. E. de, & Belluzzo, W. (2014). O diferencial de desempenho escolar entre escolas públicas e privadas no Brasil. *Nova Economia*, 24, 409–430.

<https://doi.org/10.1590/0103-6351/1564>

Moreno, A. C., & Ianelli, C. (2022, May 16). *Demand por vagas para alunos em escolas*

estaduais é maior na Zona Sul de SP; famílias migram para rede particular. G1.

<https://g1.globo.com/sp/sao-paulo/noticia/2022/05/16/demandas-por-vagas-para-alunos-e-m-escolas-estaduais-e-maior-na-zona-sul-de-sp-familias-migram-para-rede-particular.ghtml>

Neri, M. (2009). *Motivos da Evasão Escolar 4 Texto Principal*. Centro de Políticas Sociais, FGV.

Ogryzek, M., Podawca, K., & Cienciała, A. (2022). Geospatial tools in the analyses of land use in the perspective of the accessibility of selected educational services in Poland. *Land Use Policy*, 122, 106373. <https://doi.org/10.1016/j.landusepol.2022.106373>

Parajara, F. (2008, October 10). *Escolas estaduais de bairros nobres de SP “importam” alunos da periferia para não fechar*. O Globo.

<https://oglobo.globo.com/brasil/educacao/escolas-estaduais-de-bairros-nobres-de-sp-importam-alunos-da-periferia-para-nao-fechar-3825712>

Pegoretti, M. S. (2005). *Definição de um indicador para avaliar a acessibilidade dos alunos da zona rural às escolas da zona urbana*. <https://repositorio.ufscar.br/handle/ufscar/4401>

Pereira, G. (2022). *Download Official Spatial Data Sets of Brazil*.

<https://ipeagit.github.io/geobr/index.html>

Pinho, M. (2019, September 21). *Velocidade dos ônibus em SP em 2019 é a mais baixa em 4 anos*. R7.com.

<http://noticias.r7.com/sao-paulo/velocidade-dos-onibus-em-sp-em-2019-e-a-mais-baixa-em-4-anos-21092019>

Pizzol, B., Giannotti, M., & Tomasiello, D. B. (2021). Qualifying accessibility to education to investigate spatial equity. *Journal of Transport Geography*, 96, 103199.

<https://doi.org/10.1016/j.jtrangeo.2021.103199>

Pizzolato, N. D., Barros, A. G., Barcelos, F. B., & Canen, A. G. (2004). Localização de escolas públicas: Síntese de algumas linhas de experiências no Brasil. *Pesquisa Operacional*, 24, 111–131. <https://doi.org/10.1590/S0101-74382004000100006>

Pizzolato, N. D., Raupp, F. M. P., & Alzamora, G. S. (2012). REVISÃO DE DESAFIOS APLICADOS EM LOCALIZAÇÃO COM BASE EM MODELOS DA p-MEDIANA E SUAS VARIANTES. *Pesquisa Operacional para o Desenvolvimento*, 4(1), Article 1.

PNAD. (2020, July 15). *PNAD Educação 2019: Mais da metade das pessoas de 25 anos ou mais não completaram o ensino médio* | Agência de Notícias. Agência de Notícias - IBGE. <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/28285-pnad-educacao-2019-mais-da-metade-das-pessoas-de-25-anos-ou-mais-nao-completaram-o-ensino-medio>

PUC-Rio. (2022). *Censo—Sobre—Data Zoom*. <http://www.econ.puc-rio.br/datazoom/censo.html>

Raju, S., Radhakrishnan, N., & Mathew, S. (2020). Spatial accessibility analysis of schools using geospatial techniques. *Spatial Information Research*, 28.

<https://doi.org/10.1007/s41324-020-00326-w>

Rede Nossa São Paulo. (2021, October 21). Mapa da Desigualdade 2021 é lançado. *Rede Nossa São Paulo*.

<https://www.nossasaopaulo.org.br/2021/10/21/mapa-da-desigualdade-2021-e-lancado/>

Saxon, J., Koschinsky, J., Acosta, K., Anguiano, V., Anselin, L., & Rey, S. (2021). *An Open Software Environment to Make Spatial Access Metrics More Accessible*.

<https://doi.org/10.13140/RG.2.2.12396.28807>

SP1. (2022, February 7). *Sobe para 5.020 o número de crianças sem vaga em escola pública no estado de SP*. G1.

<https://g1.globo.com/sp/sao-paulo/noticia/2022/02/07/sobe-para-5020-o-numero-de-crianças-sem-vaga-em-escola-publica-no-estado-de-sp.ghtml>

Taleai, M., Sliuzas, R., & Flacke, J. (2014). An integrated framework to evaluate the equity of urban public facilities using spatial multi-criteria analysis. *Cities*, 40, 56–69.
<https://doi.org/10.1016/j.cities.2014.04.006>

United Nations. (2022). *Goal 4 | Department of Economic and Social Affairs.*

<https://sdgs.un.org/goals/goal4>

United States Census Bureau, & Opportunity Insights. (2018). *The Opportunity Atlas.*

<https://opportunityatlas.org/>

Vecchio, G., Tiznado-Aitken, I., & Hurtubia, R. (2020). Transport and equity in Latin America: A critical review of socially oriented accessibility assessments. *Transport Reviews*, 40(3), 354–381. <https://doi.org/10.1080/01441647.2020.1711828>

Walsh, S., Flannery, D., & Cullinan, J. (2015). Geographic accessibility to higher education on the island of Ireland. *Irish Educational Studies*, 34(1), 5–23.

<https://doi.org/10.1080/03323315.2015.1010302>

Wan, N., Zou, B., & Sternberg, T. (2012). A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science*, 26(6), 1073–1089. <https://doi.org/10.1080/13658816.2011.624987>

Wang, Y., Liu, Y., Xing, L., & Zhang, Z. (2021). An Improved Accessibility-Based Model to Evaluate Educational Equity: A Case Study in the City of Wuhan. *ISPRS International Journal of Geo-Information*, 10(7), Article 7. <https://doi.org/10.3390/ijgi10070458>

White, A. (2004, September 2). Time for School ~ Interview: Amartya Sen | Wide Angle | PBS. *Wide Angle.*

<https://www.pbs.org/wnet/wideangle/interactives-extras/interviews/time-for-school-interview-amartya-sen/1477/>

World Bank. (2023). *Government expenditure on education, total (% of GDP) | Data.*

https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS?most_recent_value_desc=true

Xiang, L., Stillwell, J., Burns, L., Heppenstall, A., & Norman, P. (2018). A geodemographic classification of sub-districts to identify education inequality in Central Beijing.

Computers, Environment and Urban Systems, 70, 59–70.

<https://doi.org/10.1016/j.compenvurbsys.2018.02.002>

Xu, Y., Song, W., & Liu, C. (2018). Social-Spatial Accessibility to Urban Educational Resources under the School District System: A Case Study of Public Primary Schools in Nanjing, China. *Sustainability, 10, 2305.* <https://doi.org/10.3390/su10072305>

Appendix

Supplemental Materials

Github Repository: https://github.com/felipehlvo/access_to_education_map

Web App: <https://felipehlvo-access-to-education-map-app-u0pb18.streamlit.app/>