Methodology

See this paper for a detailed explanation of the methodology and research.

We calculate two metrics for access to public high schools:

The **Access Metric** A_j combines the demand and supply for high schools in a region. It represents high school capacity adjusted for distance, nearby demand, and competition between nearby schools. It is calculated using the 3-Step Floating Catchment Area (3SFCA) method for each census tract. Here's the process:

1. Measure the *education supply using each high school's location and capacity*.
2. Divide the *supply* by the number of children aged 15 to 17 years in the area (*demand*), resulting in a measure for *demand-adjusted supply* R_j.
3. Adjust for *distance* from the school to the students W_{ij} and *competition* from nearby schools G_{ij}.
4. Sum up all access levels from each census tract to all other schools within 16 km to get the overall access for a census tract.

$$A_i = \sum_{j \in \{Dist(i,j) < 16km\}} R_j W_{ij} G_{ij}$$

The **Quality-Access Metric** H_i is measured similarly to *Access Metric* but also accounts for the quality of each school. It is calculated as follows:

1. Calculate the access-weighted average quality of schools in the region (using the Primary Education Development Index - IDEB) Q_i.
2. Multiply the Access-Metric by the quality measure.

$$H_i = A_i Q_i$$

Both metrics are scaled to each state, reflecting how public high school policy and management is done at the state level in Brazil.

The final metrics are a composite of several aspects of access and do not have a direct translation to a countable value such as 'teachers per student.' It is best thought of as a *relative* indicator of access - it does not tell you how much access a neighborhood has, but rather which ones have the most access. *It makes assumptions about how people choose schools and value quality and distance costs.* As such, the absolute value of the metric can be quite sensitive to changes in these assumptions, but the relative comparisons, which are the main interest of this research, are robust.

Map Representation: The dots on the map represent the estimated locations of students, each dot represents a group of ten students. Demographic data such as population, income, and racial distribution are gathered from the 2010 Demographic Census. School information is collected from the 2020 School Census.