Inter-American Development Bank

A User's Guide to the Brazilian Education Panel Databases

Department of Research and Chief Economist

Alyssa Huberts
Fabiana Machado

TECHNICAL NOTE Nº IDB-TN-1264

May 2017

A User's Guide to the Brazilian Education Panel Databases

Alyssa Huberts
Fabiana Machado

Inter-American Development Bank

May 2017

IDB
Inter-American Development Bank

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Huberts, Alyssa.
A user's guide to the Brazilian Education Panel Databases / Alyssa Huberts, Fabiana Machado.
p. cm. — (IDB Technical Note ; 1264)
Includes bibliographic references.
1. Educational indicators-Brazil-Databases. 2. School census-Brazil-Databases. 3. Educational tests and measurements-Brazil-Databases. I. Machado, Fabiana. II. Inter-American Development Bank. Department of Research and Chief Economist. III. Title. IV. Series.
IDB-TN-1264

Abstract

The Brazilian Schools Panel database and Brazilian Municipal Education Panel Database combine and simplify 20 years' worth of data from the Brazilian School Census, educational testing, and educational indicators. This report provides an introduction to the data and serves as a road map to their strengths and limitations. In particular, it draws attention to the points in time at which major changes were made in the format of data collection, as well as the characteristics of different subsamples within the data. The report seeks to provide a practical introduction for researchers interested in using the data to understand and research the Brazilian education system.
Keywords: Education, Brazil, Data, Schools, Municipalities
JEL Codes: H75, I24, I25
Link to Database:
http://www.iadb.org/en/research-and-data/publication-details,3169.html?pub_id=IDB-DB-123

1 Introduction

The Brazilian Schools Panel database and Brazilian Municipal Education Panel Database combine and simplify 20 years' worth of data from the Brazilian School Census, educational testing, and educational indicators. This report provides an introduction to the data and serves as a road map to their strengths and limitations.

2 Data Sources

Table 1: Data Sources

Dataset	Censo Escolar	Prova Brasil	Educational Indicators
Year	1996-2015	2007, 2009, 2011	2007-2014
Types of Schools	All (Private optional)	All public schools with 20 or more students in tested grades	All (Private optional)
Important Notes	Data format changed in 2006		These pass/fail/dropout rates are calculated by the government, so the method of calculating may diverge after this point.

The databases are constructed from three different sources. The first, the *Censo Escolar* aims to cover all schools providing fundamental education (1-8 or 1-9 years). Those that fail to submit the required information for the census risk the withholding of federal funding, and they are not allowed to participate in federally sponsored standardized tests. Thus we consider a school's participation in the census optional, although most private schools do participate.

The second data source, the INEP education indicators portal, provides information at the school level unavailable after 2006 in the census data, including class size and pass, failure, and dropout rates. This database covers the same population as the Censo Escolar.

The third data source, Prova Brasil, Prova Brasil is a national test which falls under Brazil's SAEB system for evaluation education. Up until 2005, the SAEB conducted a sampling of schools along certain strata of special interest. In later years, however, SAEB split, and Prova Brasil took a census-like approach to measuring the schooling conditions and academic performance of students in the 4th and 8th years of public primary education, along with the third year of secondary education. At the point of publication, the Brazilian government had made public the Prova scores for 2007, 2009 and 2011, all of which are

included in the database. For the schools included in the original SAEB sample, 2005 data are also available.

The Brazilian Schools Panel database and its derivative Brazilian Municipal Education Panel Database consist of all 'regular' fundamental education providers included in the census between 1996 and 2015. While the schools database lists both active and inactive schools, all counts in the municipal database are reflective of schools considered 'active' in the year of the census.

The codebooks provide detailed explanations of all variables included in the dataset. In addition, the dofiles used to create the databases are available from the authors upon request.

3 Recommendations Based on Changes in the Data: Time Series Analysis

In 2006, the school census changed its formatting from collecting data at the school level to a much more comprehensive collection of student, teacher, and school information. Due to this shift, the database is significantly different in the period from 1996 to 2005, and the years afterward. We recommend that those hoping to conduct analysis over the entire database carefully consider their choice of variables, or break the analysis into two periods, pre- and post-2006.

4 Missing Observations

In Table 2, we document variables for which there are *no non-missing observations* for active status schools in a given year (i.e., the data for the question was not collected that year). Variables present in all years are not included in the table.

5 School Composition of Database

The entire database includes 4,066,530 year-school observations, from 412,194 different primary schools over time. Figure 1 shows the composition of the database in terms of school status, which has been simplified to 'Active' and 'Inactive' from four categories 'Ativa', 'Paralisada' 'Extinta no ano anteror' and 'Extinta' due to a lack of consistent use in the original data across 'Extinta no ano pasado' and 'Extinta'. The municipal-level database considers only active schools. Figure 2 shows entry into and exit from the census over time, excluding the first and last years of the sample. The data show that a number of schools that are not present in the first 10 years of the census entered after the format change in 2006.

Table 2: Missing Variables for Certain Years

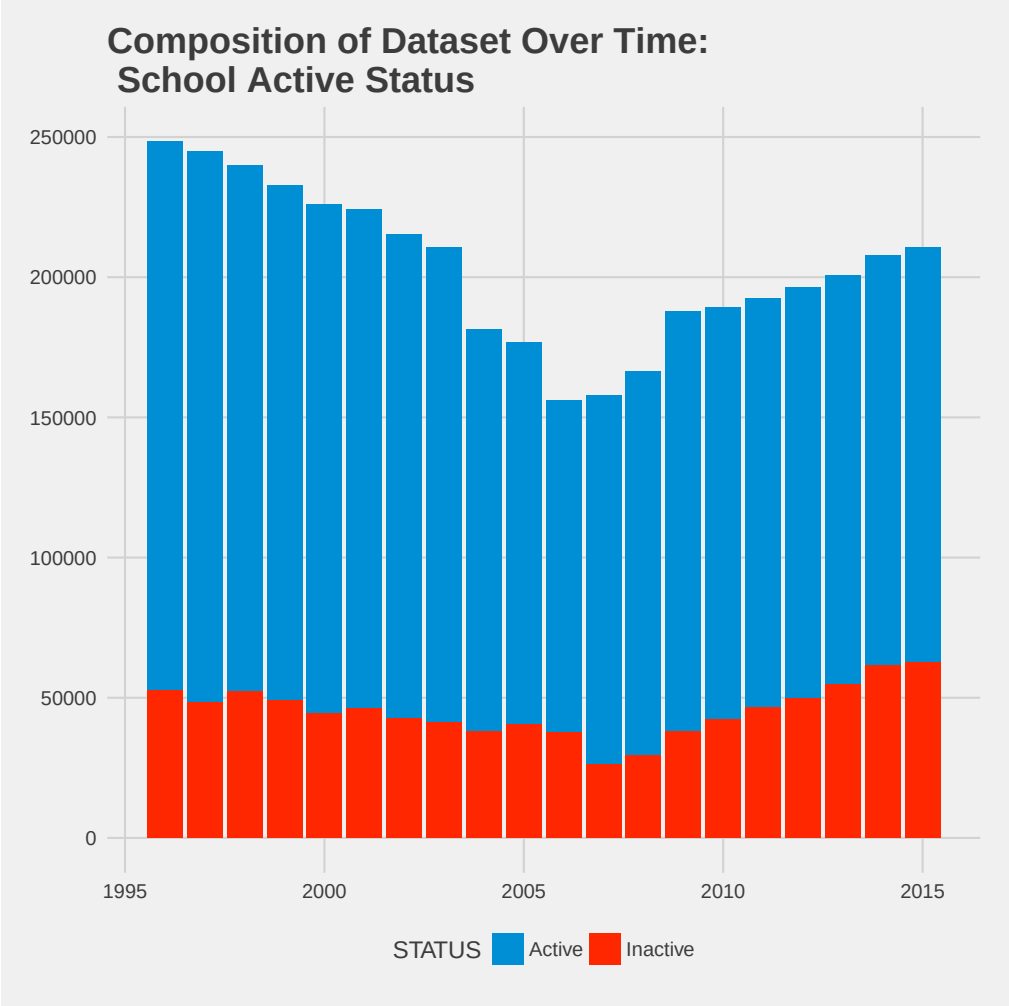
Var	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
TP_CATEGORIA_ESCOLA_PRIVADA	m	m	m	m	m	m	m	m	m	m	m									
EIGHTYEARS	m	m	m	m	m															m
NINEYEARS	m	m	m	m	m												m	m		m
TURMAFUND		m																		
IN_BIBLIOTECA	m																			
IN_LABORATORIO_INFORMATICA	m																			
IN_LABORATORIO_CIENCIAS	m																			
IN_QUADRA_ESPORTES	m																			
IN_EQUIP_TV	m																			
IN_EQUIP_PARABOLICA	m																			
IN_COMPUTADOR	m																			
IN_INTERNET	m	m	m																	
NU_COMPUTADOR	m																			
NU_EQUIP_TV	m											m	m	m	m	m	m	m	m	m
SCHOOL_WATER_PUBLIC	m																			
SCHOOL_WATER	m																			
SCHOOL_WATER	m																			
SCHOOL_ELECTR_PUB	m																			
SCHOOL_ELECTR	m																			
SCHOOL_SEWAGE_PUB	m																			
SCHOOL_SEWAGE	m																			
PROFESS	m										m	m	m	m	m	m	m	m	m	m
CLASSSIZE_I	m	m	m	m	m	m	m	m												m
CLASSSIZE_I	m	m	m	m	m	m	m	m												m
RATE_APROV				m	m	m														m
RATE_APROV_T				m	m	m														m
RATE_APROV_T				m	m	m														m
RATE_FAILURE				m	m	m														m
RATE_FAILURE_I				m	m	m														m
RATE_FAILURE_T				m	m	m														m
RATE_ABANDON				m	m	m														m
RATE_ABANDON_I				m	m	m														m
RATE_ABANDON_T				m	m	m														m
RATE_TRANSFER				m	m	m						m	m	m	m	m	m	m	m	m
RATE_TRANSFER_I				m	m	m						m	m	m	m	m	m	m	m	m
RATE_TRANSFER_T				m	m	m						m	m	m	m	m	m	m	m	m
RATE_OVERAGE	m	m																		m
RATE_OVERAGE_I	m	m																		m
RATE_OVERAGE_T	m	m																		m
PROVA_MEAN_PORT_I	m	m	m	m	m	m	m	m	m		m		m		m		m	m	m	m
PROVA_MEAN_MAT_I	m	m	m	m	m	m	m	m	m		m		m		m		m	m	m	m
PROVA_MEAN_MAT_T	m	m	m	m	m	m	m	m	m		m		m		m		m	m	m	m

6 Providers of Education

Schools in Brazil are administered by either i) the federal government, ii) the state government, iii) the municipal government, or iv) a variety of private education providers. The central government covers higher education, the states upper secondary, and the municipalities early childhood, but the constitution affirms that primary (fundamental) education is shared. In the south and southeast states tend to take greater responsibility for primary education, while municipalities play a larger role in the north and northeast, but many municipalities exhibit both levels of education simultaneously. In the section on municipal-level data we explore this question further. Government spending levels on education are set at 18% of post- transfer revenue for the federal government and 25% for the states and municipalities.

In order to ensure equity in the distribution of funds across different systems, the 1996 *Fundo de Desenvolvimento do Ensino Fundamental* , or FUNDEF, and its 2006 replacement, *Fundo de Manutencao e Desenvolvimento da Educacao Basica*, or FUNDEB, were designed to help establish a national per student funding baseline and tie funding to enrollments. Originally, each state and municipal government was required to contribute 15% of their income, or three fifths of the 25% total dedicated to education, toward the FUNDEF for each state, which then redistributed the funds to the states and municipalities based on enrollment. Thanks to its creation of enrollment incentives, FUNDEF is largely credited

Figure 1: Active vs. Inactive Schools



with Brazil's increase in net enrollments through the 1990s and 2000s. [1] FUNDEB, which revised FUNDEF in 2007, raised the funding requirement to 25% of municipal funds and changed the funding allocation to reflect the number of students within different education levels. FUNDEB's revision also includes early childhood and upper secondary education, and has different spending floors for certain populations including indigenous, youth, and adult education.

Figure 3 shows how the composition of the school census has changed in terms of school type over time. Federal schools, which are mostly military, represent a tiny proportion of the census. Overall, the share of municipal schools in the sample has been decreasing over time, while state schools have remained almost constant and private schools have increased significantly beginning in 2006 with the format change.

Figure 4 shows the percent of schools in urban locations, as designated by the IBGE. Generally, municipal schools tend to feature much more in rural areas, while almost all

[1] Brazil Country Management Unit. "Brazil Municipal Education: Resources, Incentives, and Results. Volume 1: Policy Report" The World Bank. Report No. 24413 BR

Figure 2: First and Last Year in the Dataset

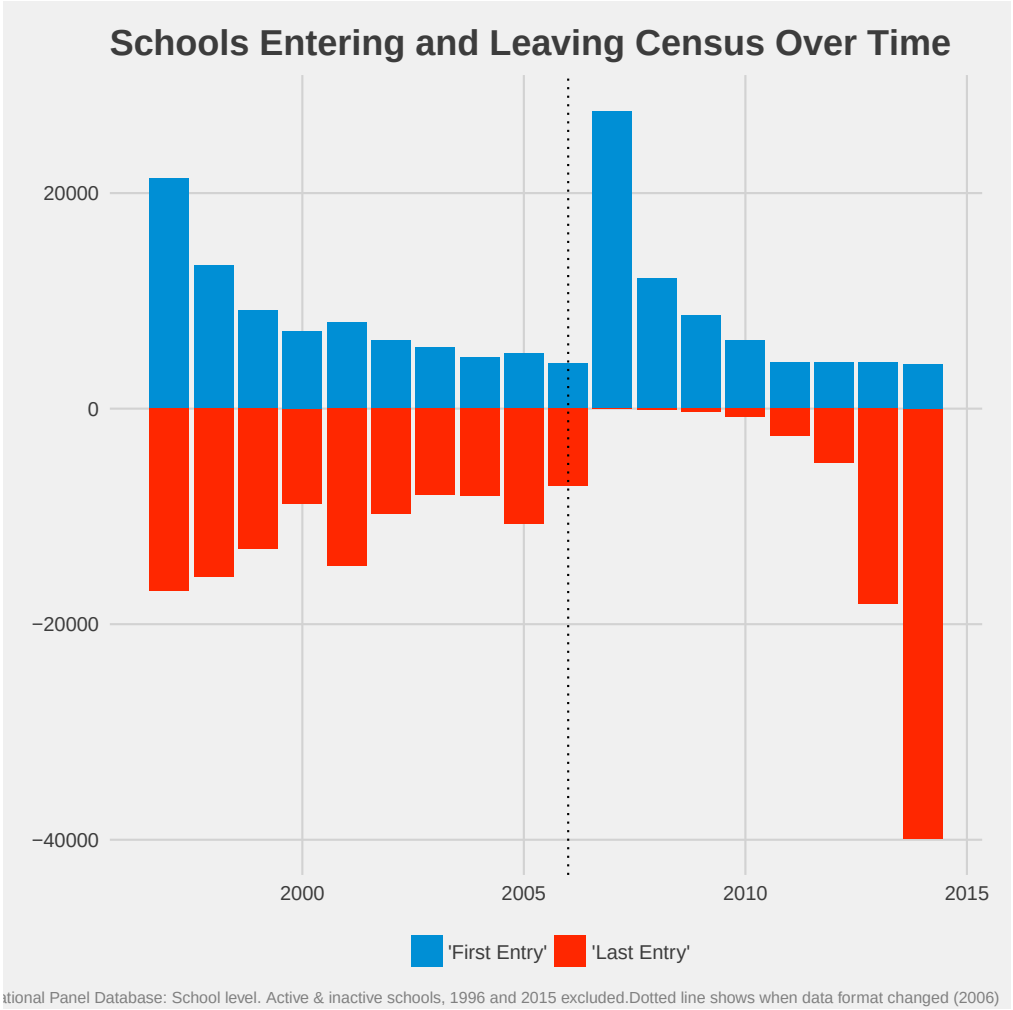
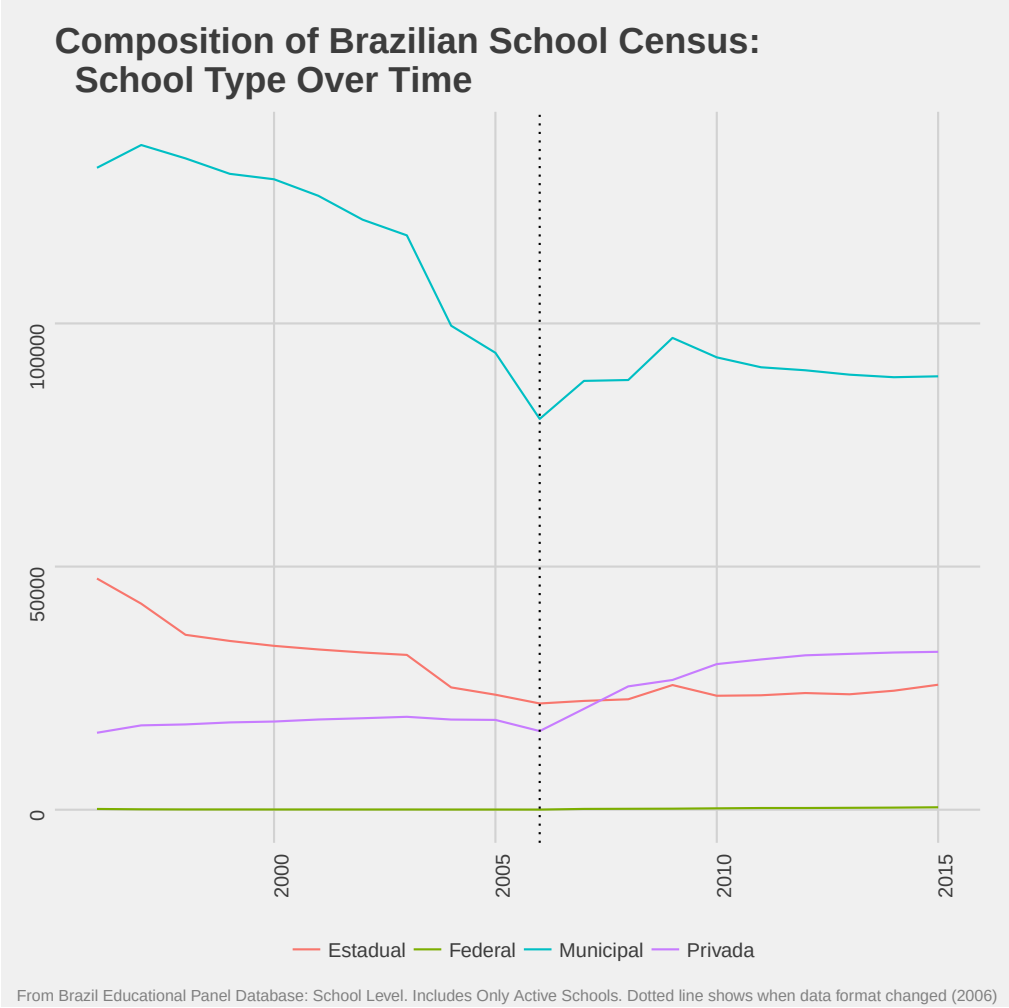
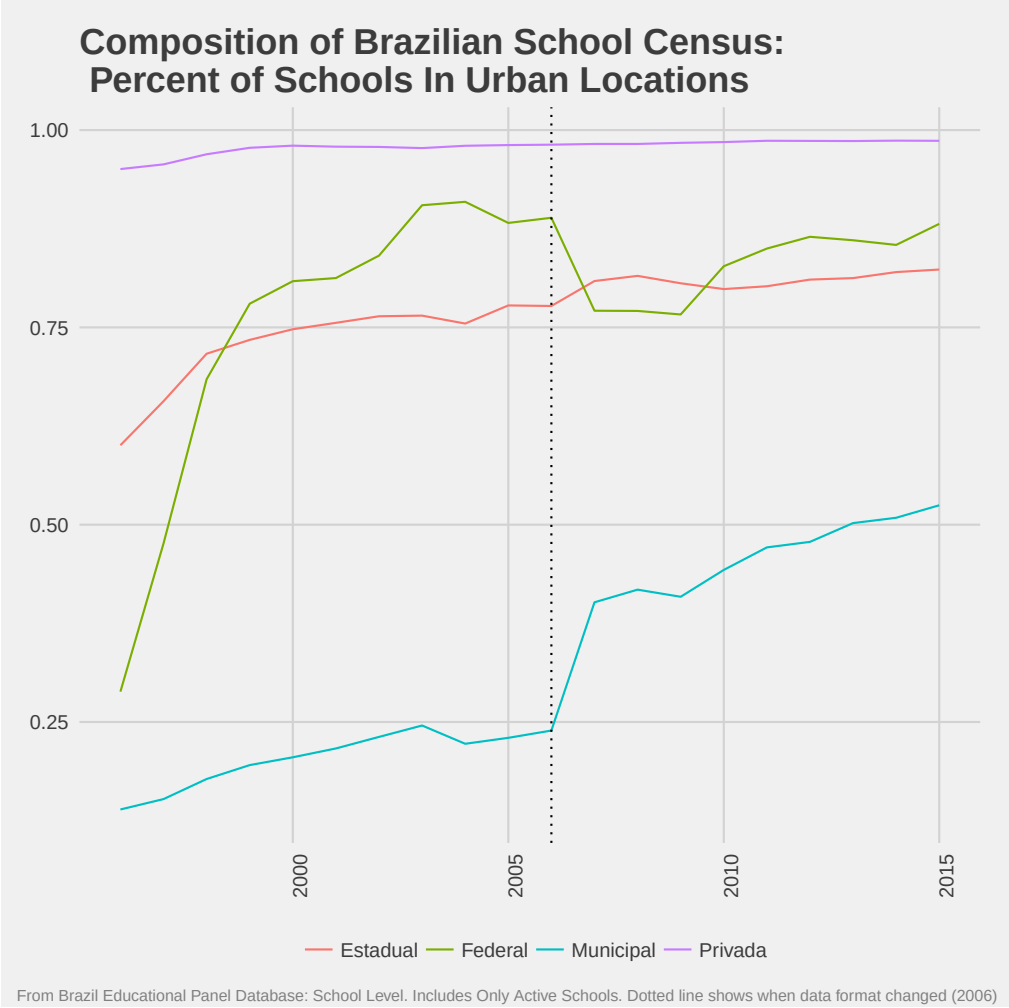


Figure 3: School Type



private schools are urban. It is important to remember that Brazil underwent significant urbanization in the period of the dataset, so a shift in composition toward more urban locations is more representative of changing neighborhoods than changing school priorities.

Figure 4: Percent of Urban Schools by State

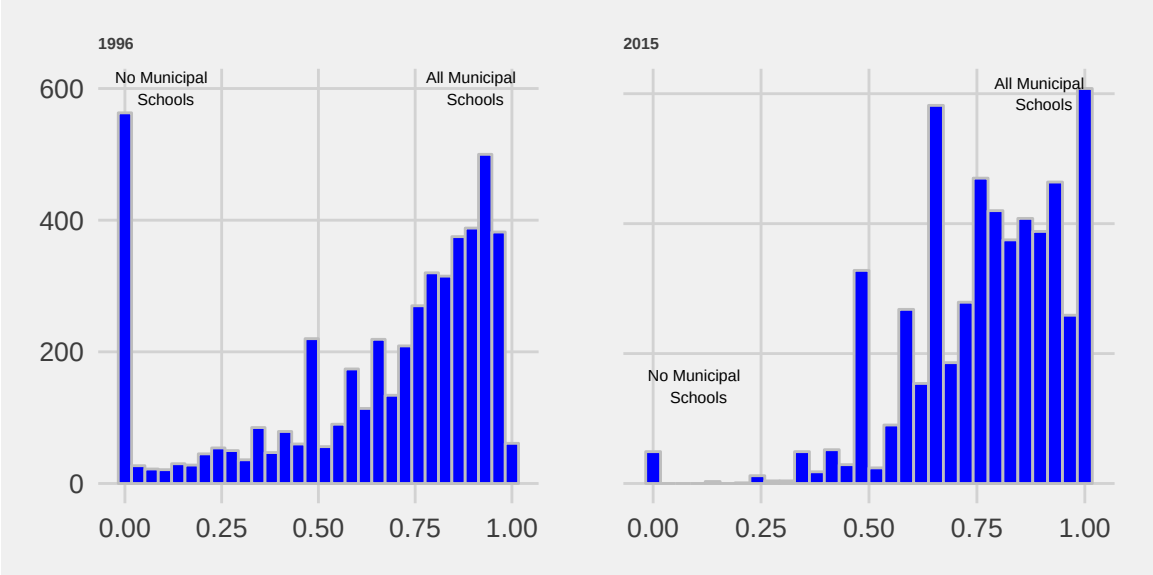


7 Municipal Level Statistics

As mentioned above, different municipalities have divided the provision of education differently. Figure 5 shows histograms with patterns of how different municipalities have provided education, with the percentage of schools run by the municipality on the x axis and the number of municipalities with the given distribution on the y axis. In 1996, the distribution was fairly split, with a significant group of municipalities having only state or federal schools, but most municipalities providing a majority of their public education through municipal schools. It is noteworthy that fewer than 100 municipalities in 1996 had only municipal schools, likely because state schools were earlier to arrive in more rural areas.

By 2015, the trend had shifted. More than 600 municipalities had only municipal schools, while only a small portion (fewer than 100) had only state schools. The maps in Figure 6 shows this pattern in more detail, as whole regions in the northern and northeast regions shift their primary education to only municipal providers, while a more mixed education expands in the southern regions.

Figure 5: Municipal Share of Education. The figures show the distribution of municipalities in 2015. The x axis shows the percent of public schools operated by the municipality and the y axis shows the frequency of each percentage



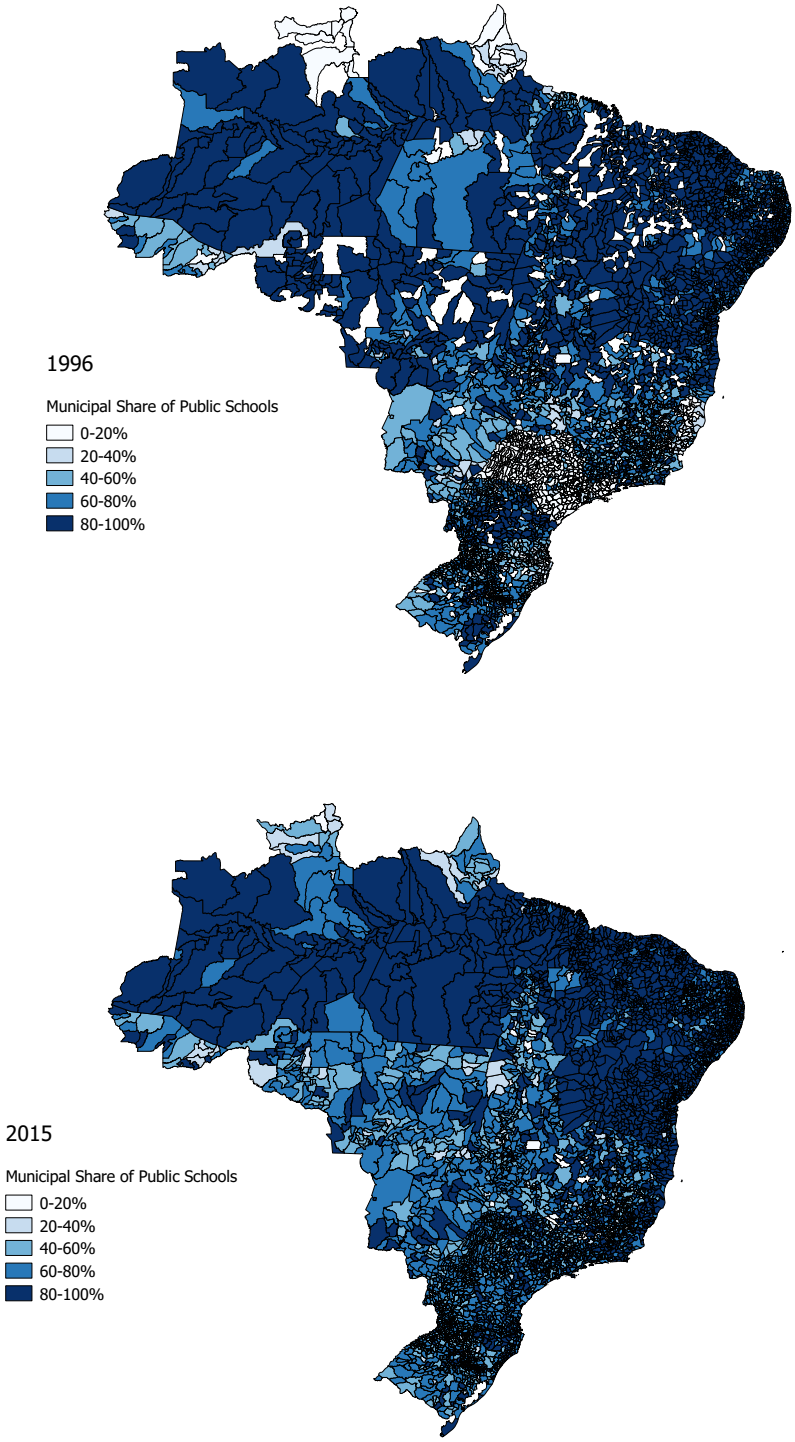
8 Measures of School Quality

The two most commonly used indicators of school quality in Brazil are scores on the Prova Brasil national exam and flow (pass, fail, and dropout) rates. The combination of these two factors is used by the government to calculate the government's most important school performance indicator, the *IDEB* (Indice de Desenvolvimento da Educacao Basica). The IDEB multiplies a school's average pass rate (probability that a student will pass a given year of school) with its average math and Portuguese scores on Prova Brasil, the national performance exam, under a predetermined system of weights. IDEB scores are the most widely considered performance measure, as the government uses them to set education improvement targets, as well as to inform transfers to schools. Neri and Buchman (2008) [2] found in 2008 that, depending on individual school costs, IDEB and the funding mechanisms associated with it can provide incentives for schools to either a) reduce standards for passing thresholds, b) improve investment in human capital for learning improvements or c) some combination of the two.

Those intending to use pass rates for analysis should be aware that Brazilian law allows municipalities to apply social promotion cycles at their own discretion. Because it has some

[2] http://www.cps.fgv.br/ibrecps/discussao/EE2008_QualiEduc_Paper_International_Submission_Final.pdf

Figure 6: Municipal Share of Education. The figures show the map of municipalities in 1996 and 2015. Darker municipalities have a higher share of municipal schools providing public education. Blank spaces are municipalities for which data are missing.



of the highest failure and dropout rates in the world, Brazil in 1996 changed its laws to allow for social promotion cycles, which group students into (usually 4 year) cycles during which they are automatically promoted to the next grade, with only one year in which he or she can be held back. The hope was that this policy would encourage students to remain in school, and also incentivize teachers to work with struggling students so that they could succeed at the next level, rather than just failing them. Because schools and governments had the discretion to decide whether or not to implement social promotion cycles, the policy's implementation is patchwork, and no clear consensus exists on the success of the policy.

Figures 7, 8, and 9 show state average Prova results at 4 testing levels (Portuguese/Math, 4th grade, Portuguese/Math 8th grade) in 2007, 2009, and 2011. The sample of schools taking the Prova should be distinguished from the school census, as it only includes schools which included more than 20 students at the tested grade level, and private schools did not take the Prova exam (with the exception of a few states that were included in the SAEB in 2005). To reflect this discrepancy, Table 3 shows the differences between schools which took at least one of the Prova tests in 2011, and those that did not.

Table 3: Comparison between Prova-Assessed Schools and General Census Population, 2011

Indicator	Prova Sample	Censo Escolar
Number of Schools	41,882	145,774
Percent urban	83.8%	63.5%
Ratio Municipal Schools: State Schools (Approximate)	1.6:1	3.8:1
Average Number of Students (Initial years) in School	117	114
Average Number of students (Final years) in school	113	116
Percent of schools with water from a public network	86%	68%
Percent of schools with electricity from a public network	99%	93%
Average Class Size	19.9	20.1
Teachers' Average Years of Schooling	13.76	13.72

As expected due to the size requirement, schools assessed by Prova tend to be more urban, with increased access to services like water and electricity. The ratio between municipal and state schools in the census is closer to 1:1 than the census, as many smaller, more rural municipal schools do not qualify for Prova Brasil. Within-school characteristics, however,

are not as differentiated, with relatively similar class size, average teacher education level, and number of students.

Figure 10 shows the distributions for the 2011 scores, based on school type, along with the number of schools within the sample. The distributions illuminate some of the patterns which simple means might obscure; such as the higher variance in municipal schools' performance, particularly at the high scoring end. The handful of federal schools score particularly well on the Prova Brasil, especially in the eighth grade.

9 Making Use of Pass/Failure Rates

Due to their use in calculation of the IDEB, we include pass/fail/dropout rates within the dataset. However, readers are advised to exercise their judgment in conducting time series analysis on these variables, as the creation of the IDEB and as a consequence the change in the form of data collection seems to have had a significant effect on the way these rates are reported. Figure 11 shows pass rates for primary education and dropout rates for the second half of primary education as an example. The vertical line in the plot shows 2007, the year that the IDEB was created and that these rates were excluded from the school census, and reported instead within the 'Indicadores Educacaionais'. We suggest that those hoping to conduct time series analysis focus on the period following 2007.

10 Summary

The Brazilian Education Panel's School-Level and Municipal-Level databases include a wealth of information about Brazilian education, spanning more than 20 years and drawing from three different datasets. This note provides a preliminary foundation for exploring the data and understanding the context of the Brazilian education system. In particular, it draws attention to the points in time at which major changes were made in the format of data collection (namely 2006), and different subsamples of the data, such as the differences between Prova-assessed schools and the more general census population. We recommend that those using the data for their own research think carefully about the appropriate subsection of schools for their level and type of analysis.

Figure 7: Prova Average Scores 2007

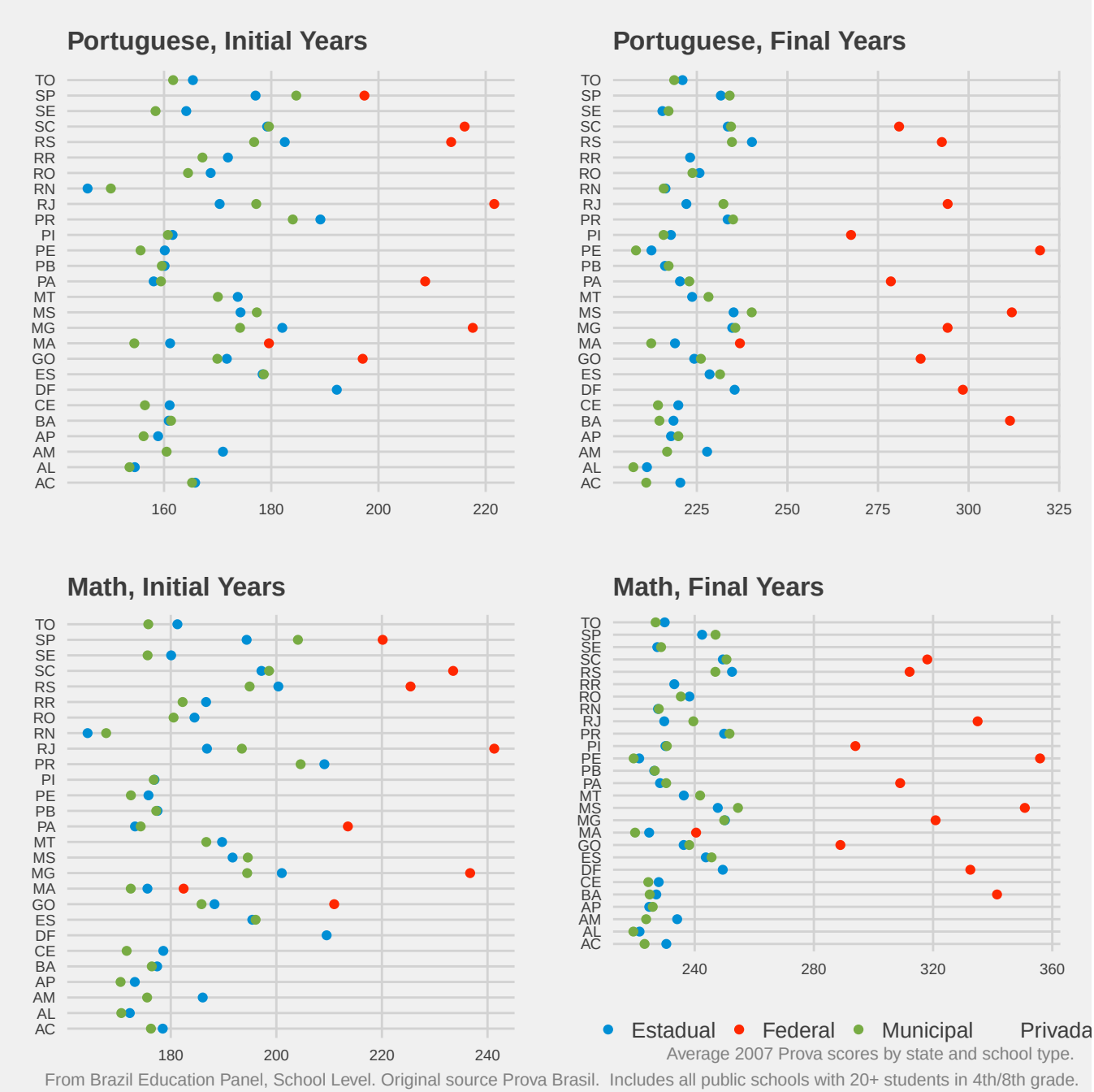


Figure 8: Prova Average Scores 2009

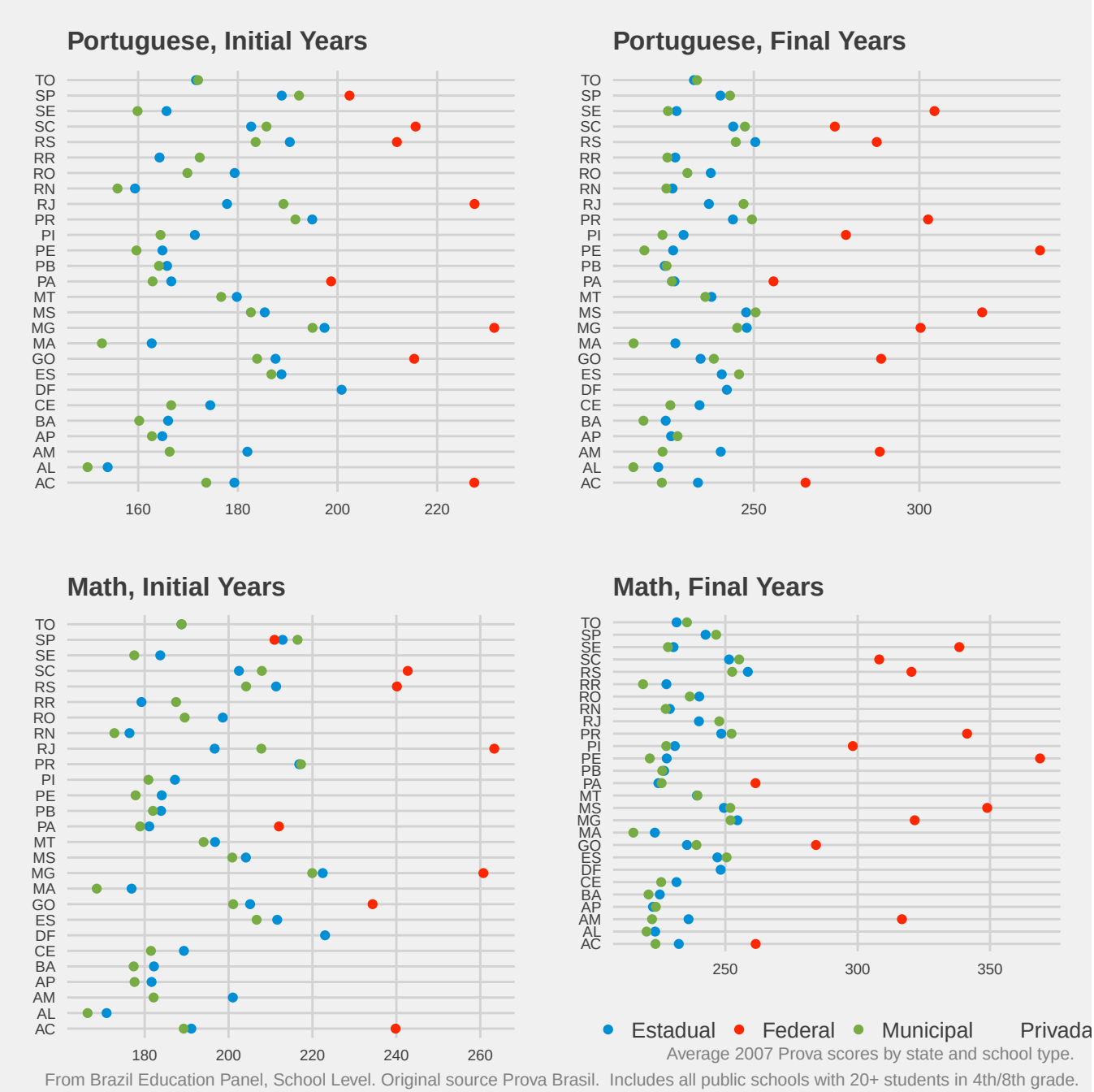


Figure 9: Prova Average Scores 2011

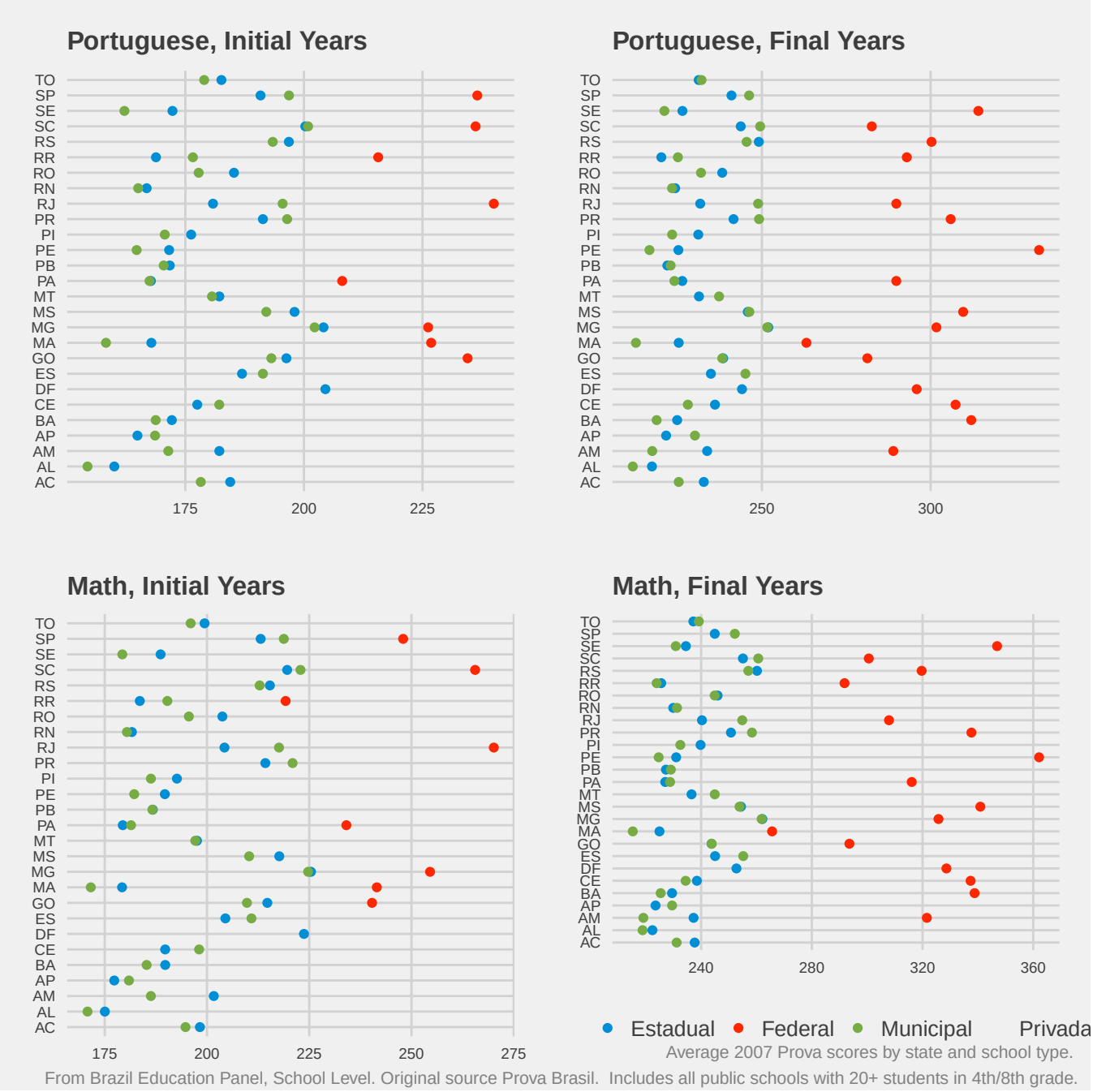


Figure 10: Prova Average Scores 2011: Distribution

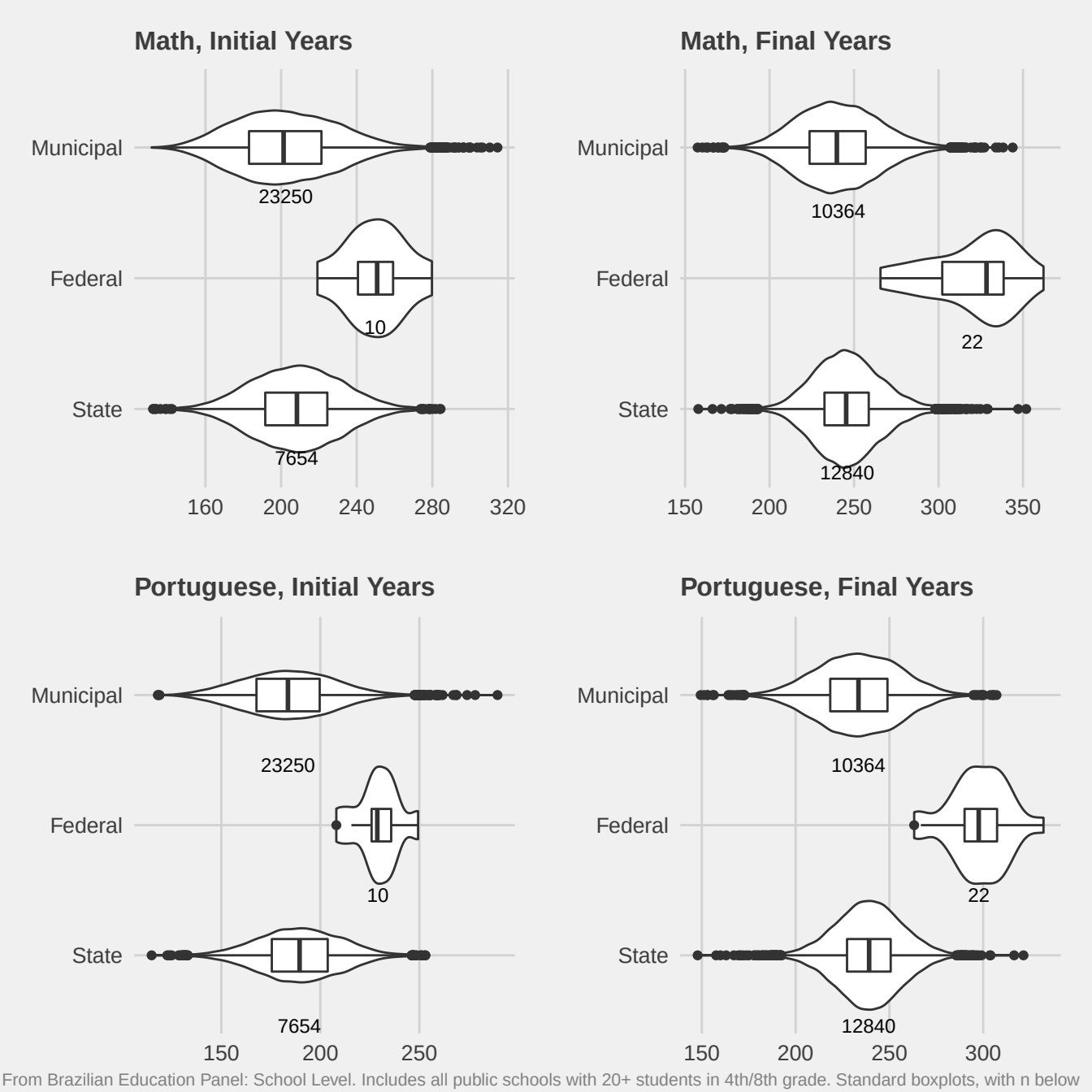
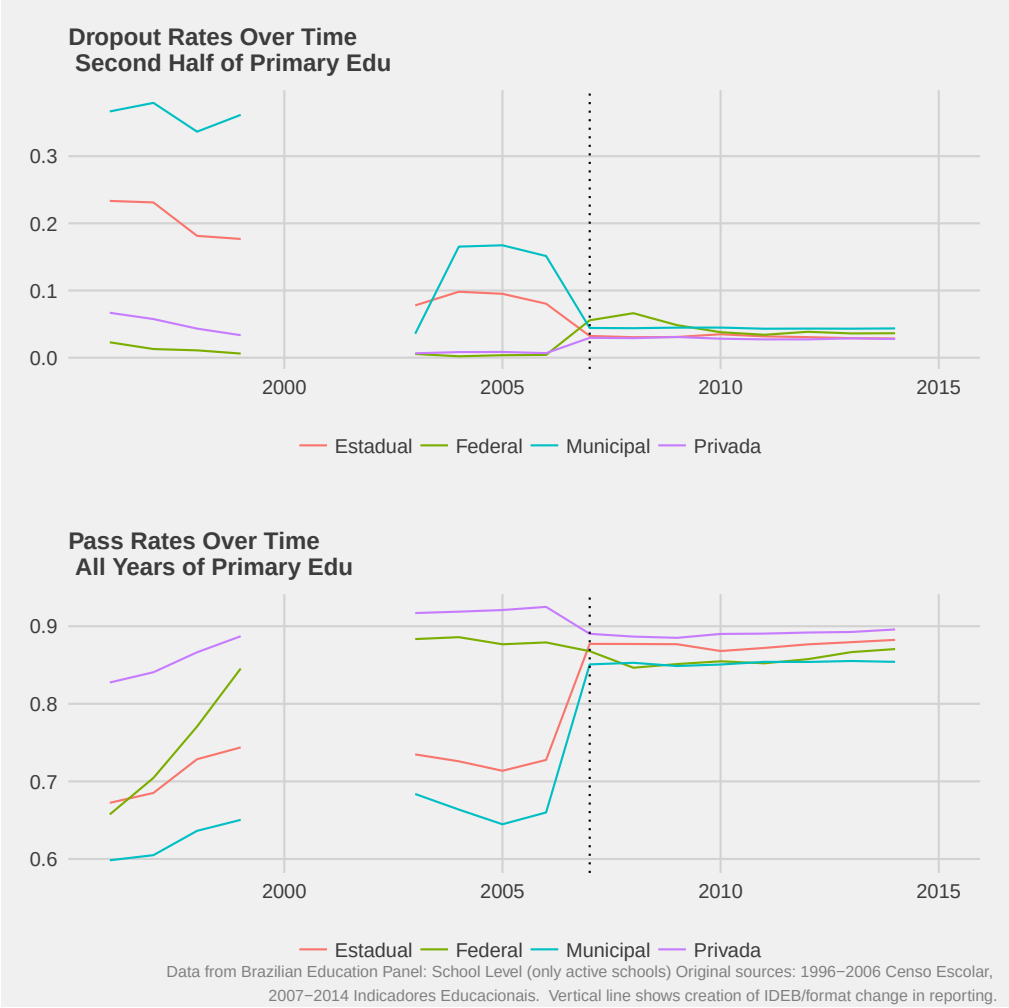


Figure 11: Pass and Dropout Rates over Time



11 References

All data included in the database was downloaded from http://inep.gov.br/microdados between September 1,2015 and June 1, 2017