Department of Education Analytics

A data-driven dive into the Philippine education system

Troy James R Palanca | Jumbo Dumbo Thoughts

August 12, 2015

Contents

Executive Summary

Context

Key Findings

Recommendations

Data Management

Data sources

Data was sourced from public sources, mostly from the Department of Education's open data portal. The complete list of data inputs is as follows:

Table 1: Data descriptions and sources

Dataset	Observation Unit	Applicable Year(s)	Description	Data Source
Enrollment	Schools	S.Y. 2011-2012 up to S.Y. 2014-2015	Enrollment counts for males and females	Department of Education
Rooms	Schools	S.Y. 2012-2013	Room counts per (academic, nonstandard, unused)	Department of Education
MOOE	Schools	S.Y. 2014-2015	MOOE allocated from the national budget	Department of Education
Teachers	Schools	S.Y. 2013-2014	Teacher counts (mobile, regular)	Department of Education
School locations	Schools	Recent	Latitude-Longitude pairs for various schools	Department of Education
Cities / Municipalities	City / Municipality	Recent	Coordinate pairs and other descriptive information for cities / municipalities	Bangko Sentral ng Pilipinas, PH Open Data Portal
PH Shapefile	Country	Accessed 08/11/2015	Shapefile for the entire country - to be used for mapping	PhilGIS.org
PH Provinces Shapefile	Country	Accessed 08/11/2015	Shapefile for the country divided into provinces - for mapping province-specific information	PhilGIS.org

Data processing

Before analysis, we process the various datasets as follows:

- Grade levels in the enrollment data were harmonized across the years (e.g. Grade 7 and Year 1 are combined, and so on).
- Locations were integrated into the masterlist of schools, when available.
- Teachers, rooms, and budget data were integrated into the masterlist of schools, when available.
- Data were saved into RData images for easy access.

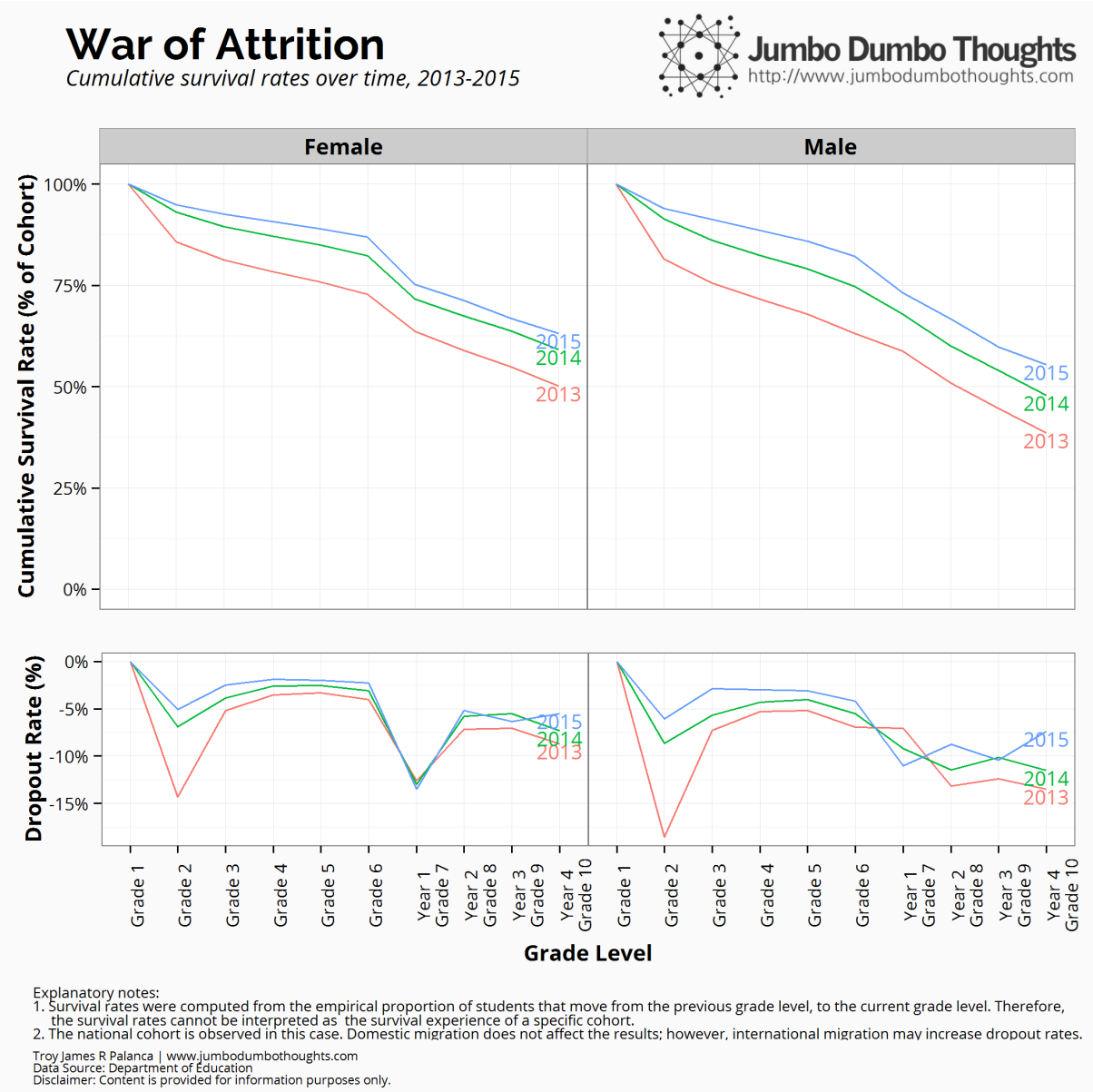
Survival Analysis

We perform a survival analysis to determine the retention performance of the Philippine education system. Analysis of survival rates allow us to determine the relative performance across genders, schools, years, and grade levels.

National-level cohort

We first perform a survival analysis of the nationwide cohort.

Survival rates over time



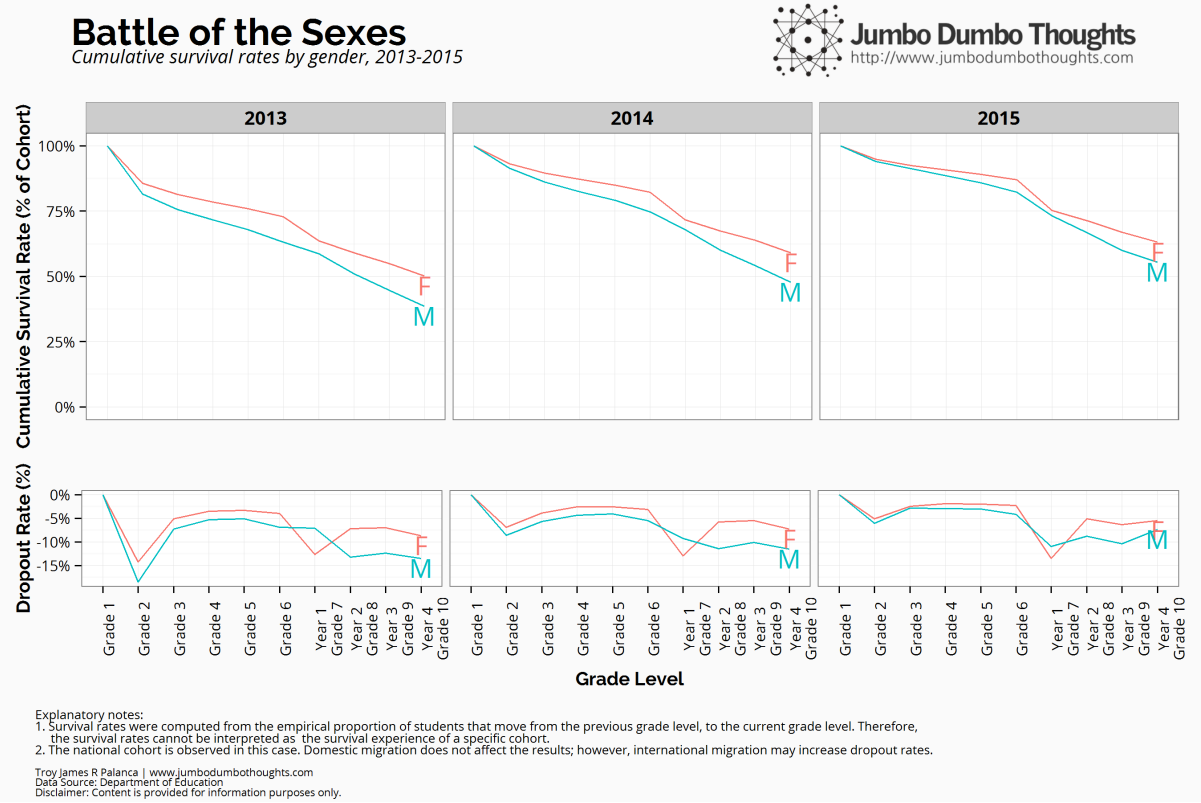
Cumulative survival rates have been increasing over time, indicating an improvement in overall retention rates for the public school system. On the bottom panel, we can see that the improvement can be attributed to reduction in the dropout rate for Grade 2 students. The transition rate to secondary school does not seem to have improved over time.

This analysis, however, comes with a few caveats:

- Since we only have raw enrollment numbers for the public school system, we cannot determine whether drops in the cohort size are due to dropout, transfer to private school, or international migration. Despite this, domestic migration will not affect the dropout numbers, since we sum all enrollment figures nationwide.
- In computing cumulative survival rates, we compute the inverse of the shrinkage in cohort size for each grade level for that year, and then compute the cumulative product of the survival rates. What this means is that the survival curve cannot be interpreted as the experience of a single cohort, but all cohorts present during that particular year. I think this is reasonable because school system performance is affected not by the particular cohorts in play, but the policies, procedures, and circumstances in place during that year.

Survival rates by gender

What are the differences between male and female survival rates? We flip the faceting to determine the answer to that question.



This chart yields two key observations about the gender differences in survival rates:

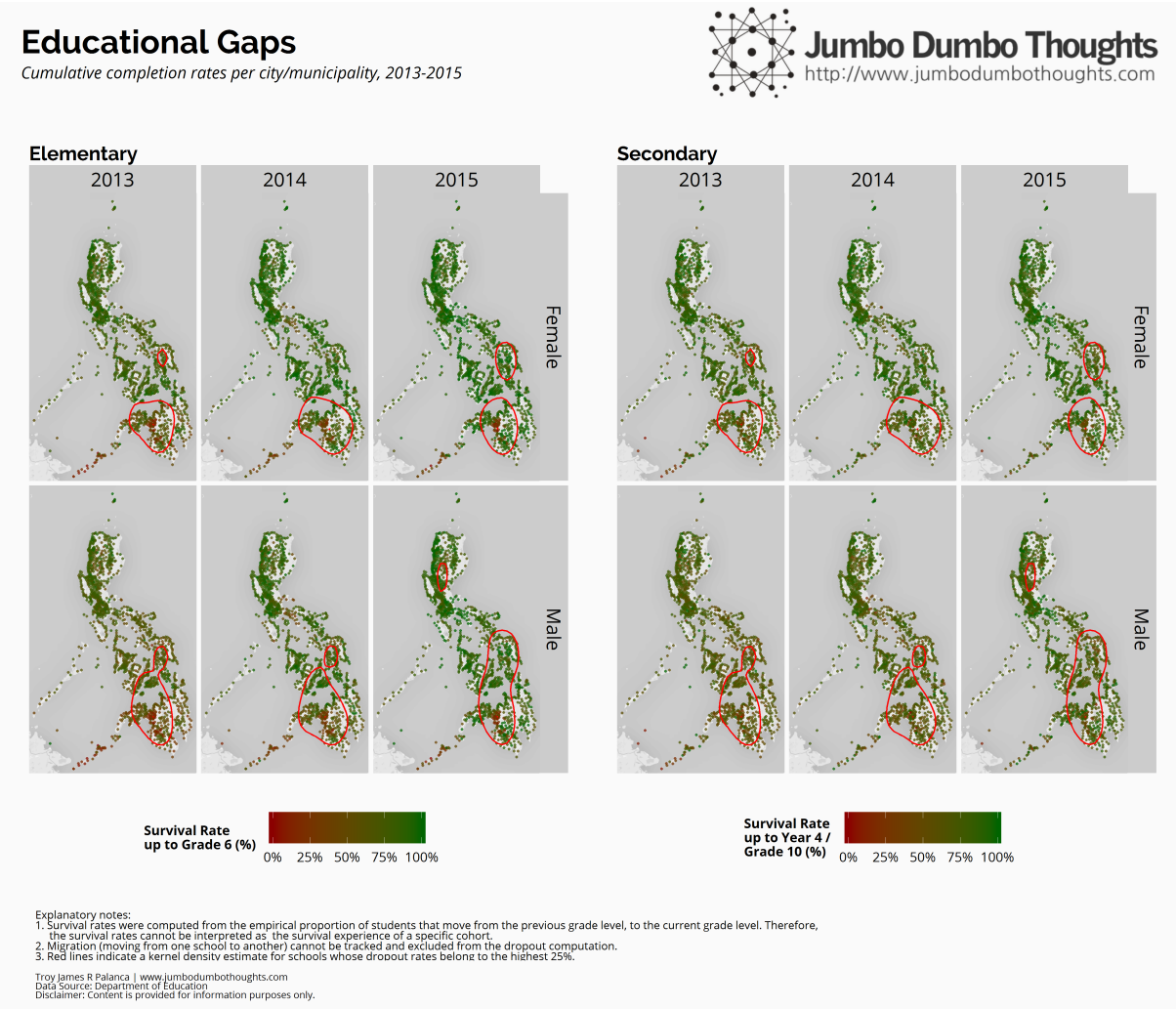
- Females are better at staying in school, but are given less chances to do so, especially during the transition from elementary to high school.
- Overall, females are still more likely to stay in school than males. Demographics, family needs, and will come into play.

City/Municipality Level Cohort

Next, we want to analyze the differences across cities and municipalities in terms of survival rates.

Survival maps

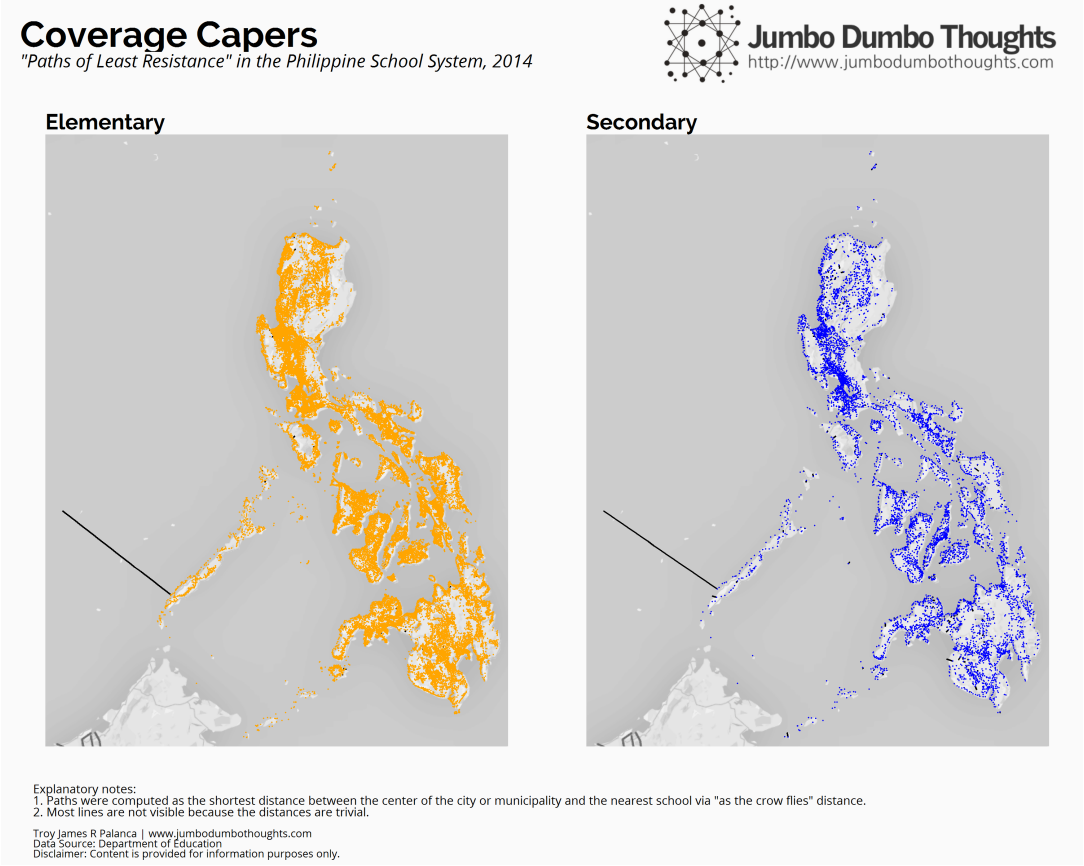
We first plot the cities and municipalites coded with the cumulative survival rate to determine whether there are any "dropout hotspots." We also add a density line that isolates concentrations of schools whose dropout rates belong to the worst 25%.



A constant hotspot of high dropout rates in the ARMM region, especially in Maguindanao. Recently, however, especially for males and for elementary schools, Eastern Visayas has also experienced drastic dropout rates. This may be the result of Supertyphoon Yolanda, a strong storm that rampaged across the region in November 2013.

Path of Least Resistance

We will assess the coverage of the Philippine educational system by computing, for each city or municipality, a "path of least resistance," which is defined as the shortest distance that one would have to travel from his/her city or municipality to the nearest elementary or secondary school.



Except for Kalayaan (part of the disputed Spratly Islands territories), all cities/municipalities seem to be covered by schools[1]. This seems to suggest that beefing up existing schools should be the priority over building new schools. However, this does not mean that connectivity (easy access) is not a concern. Lack of bridges or other transport infrastructure may hamper access to schools.

Capacity Analysis

We perform an analysis of school capacity[2] along three dimensions:

- teacher capacity, as defined by the student-to-teacher ratio[3],
- room capacity, as defined by the student-to-room ratio[4], and
- budgetary capacity, as defined by the amoung of maintenance and other operating expense (MOOE) budget per student[5].

Exploratory data analysis

We first determine centrality and dispersion by visually inspecting histograms of the computed capacity metrics. Take note the following definitions:

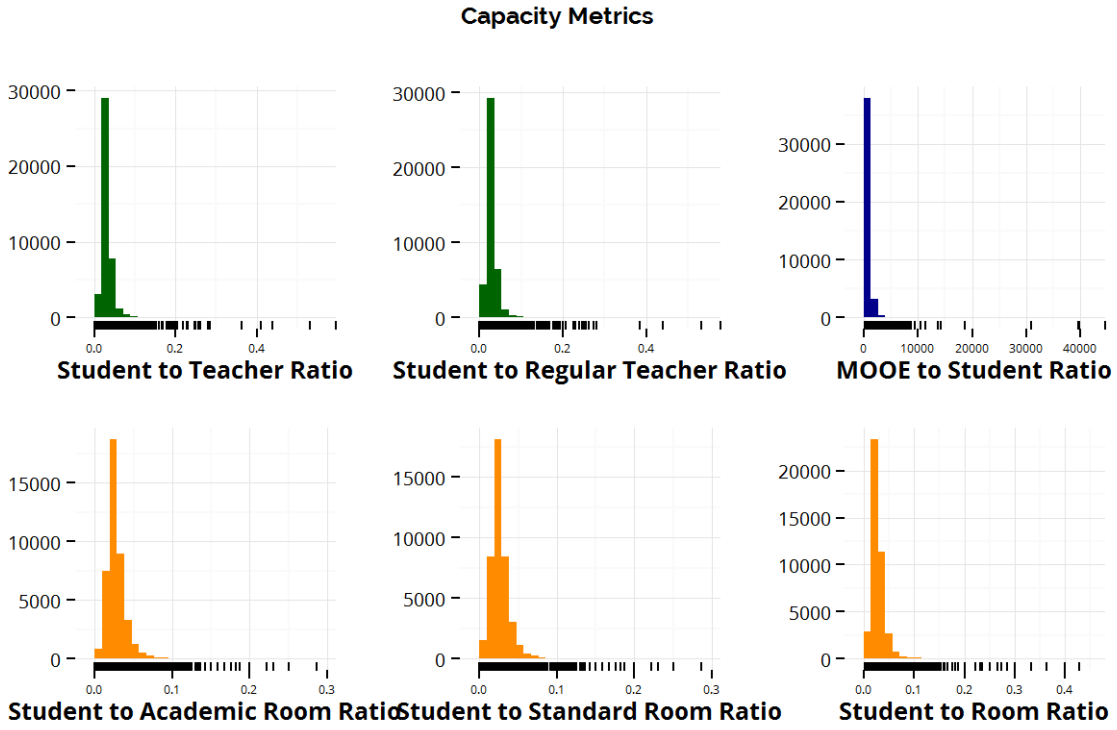
[1] We remove schools with zero teachers, schools, or rooms (4,896 or 10.5%). Schools whose capacity figures are unavailable are likewise ignored (4 or <0.01%).

[2] We remove schools with zero teachers, schools, or rooms (4,896 or 10.5%). Schools whose capacity figures are unavailable are likewise ignored (4 or <0.01%).

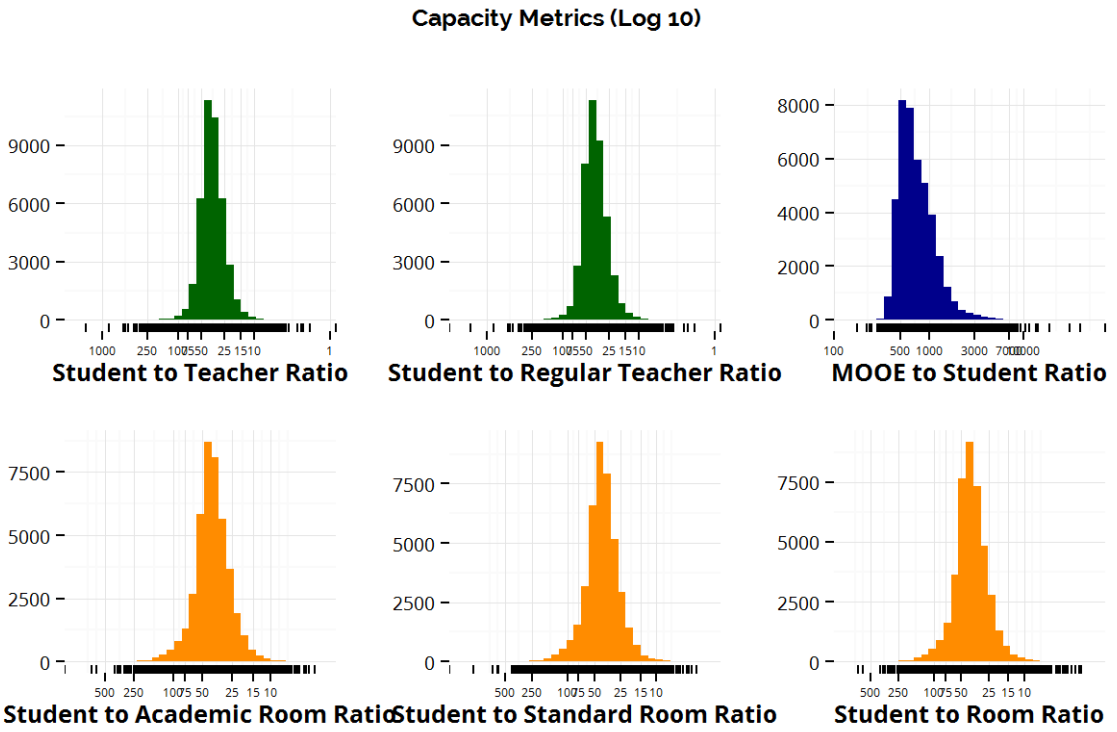
[3] Computed for 2014 only, due to data availability. For this initial section, we take all kinds of teacher (mobile, SPED, regular, and instructor).

[4] Computed for 2013 only, due to data availability. For this initial section, we assume full utilization of nonstandard and unused rooms.

[5] Computed for 2015 only, due to data availability. We assume full disbursement of these amounts.



The metrics in their raw form are skewed to the right, i.e. there are some extreme positive outlier schools in the dataset. Skewed distributions are difficult to inspect and may adversely impact the result of further statistical methods, such as cluster analysis. We remedy the variables by taking the base 10 logarithms of each of the variables - transforming the scale from additive to multiplicative.

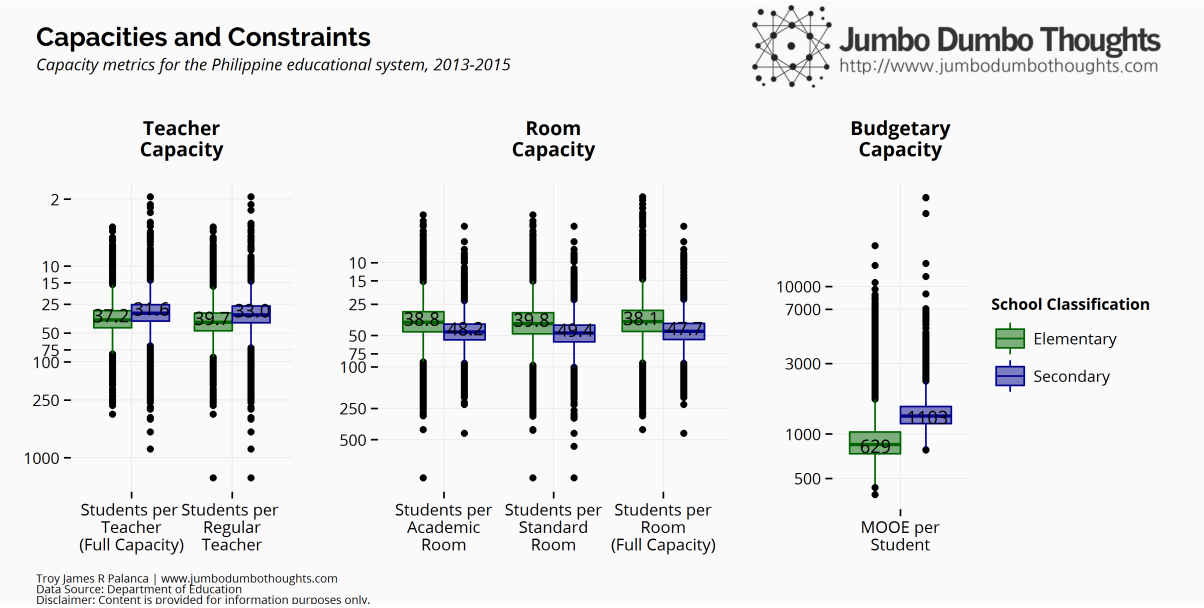


The logged capacity metrics are more symmetrical, with the exception of the MOOE Ratio.

From the histograms, we can already observe large disparities in school capacity. Student-to-teacher ratios can range from under 10 students per teacher to up to 100 students per teacher. Student-to-room ratios are similarly dispersed. While most schools budget lower than P1,000 per student, some schools

can have up to P10,000 per student.

We can further explore the average values and relative dispersion of capacity metrics through boxplots.



The average student-to-teacher ratio hovers at around 37:1 for elementary and 32:1 for secondary. Compared to the OECD standard of 18:1, teacher capacity still pales in comparison, especially when smaller class sizes are linked better academic performance[6].

Room capacity is similar for elementary schools, at an average of 39 students per room. However, high school rooms are more packed, with an average of 48 students per rooom. Whereas teacher capacity is relatively better for high schools, room capacity is drastically worse. Still, one could argue that high school students may not require smaller class sizes as study habits may have already been inculcated.

On average, elementary schools budget P639 per student, while secondary schools budget P1,103 per student. No direct comparisons can be made since differences in curricula and teaching talent required may be substantially different between elementary and secondary schools.

It should be noted that introducing an additional shift in the class schedule effectively doubles teacher and room capacity. However, having different shifts, especially for students in the early stages of education, may not be ideal. Strategies to explore using the current capacity more efficiently are explored in later sections.

Capacity clustering

Cluster analysis, or clustering, is a statistical method of grouping objects with others of similar characteristics. In this case, we use cluster analysis to group schools along with others of similar capacity metrics. We can then determine a capacity building strategy for each specific cluster.

Exploratory Analysis

We first determine the feasibility of clustering by examining the shape of the input space.

[6] Although the question of whether smaller class sizes are better for students overall is still debated, the argument for larger class sizes generally emphasizes the impact on social skills, not academic performance.

Capacity metrics for elementary (left) and secondary (right) schools.

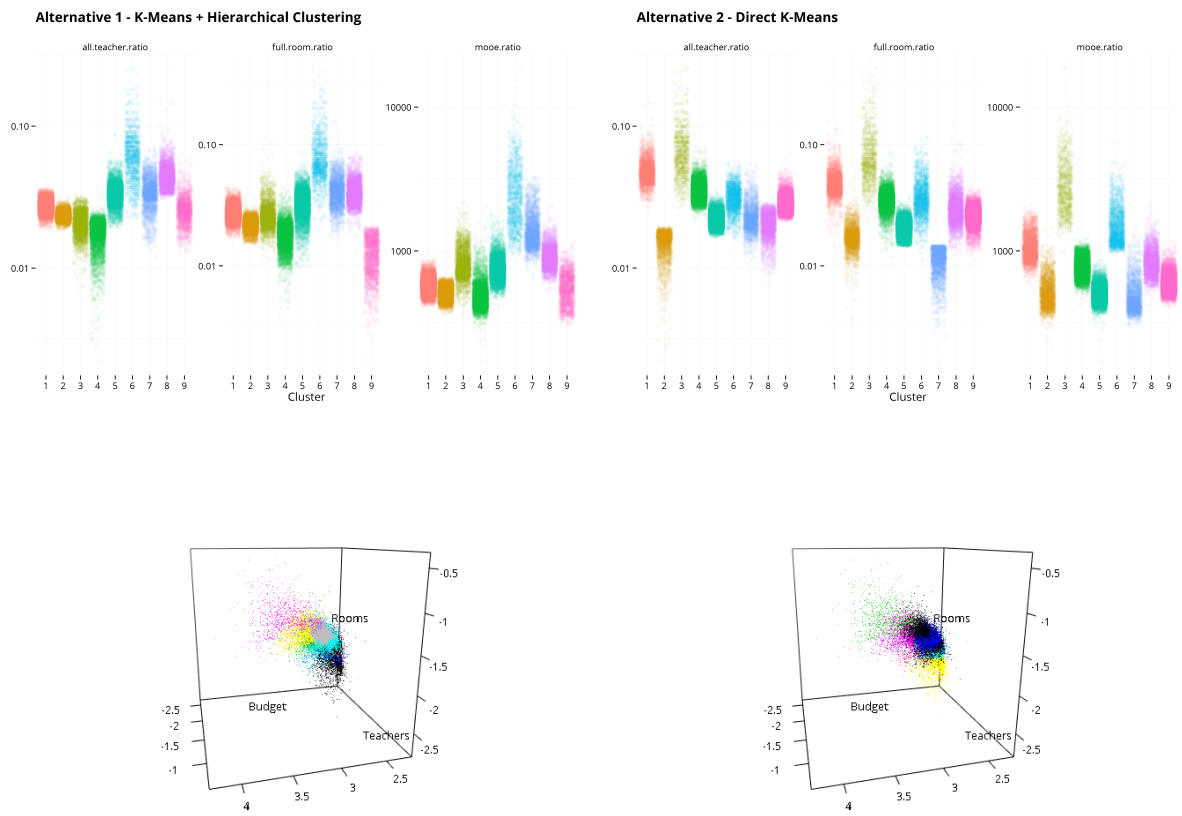
For both elementary and secondary schools, the capacity metrics are gradiented, and there are no distinct clusters. However, this does not mean we cannot perform clustering in order to both size the areas of concern and produce a specific "capacity strategy"" per cluster.

Alogrithm Selection

For this clustering exercise we consider two methods - agglomerative hierarchical clustering and divisive k-means clustering:

- **Hierarchical clustering** is a method of clustering that "builds" up individual units based on their nearest neighbors. As the algorithm progresses, the clusters become larger and larger. Hierarchical clustering has the benefit of capturing more nuances in the input space.
- **K-means clustering** is a method of clustering that divides up the space into k clusters, each defined by a centroid and the units assigned based on the nearest centroid. K-means clustering has the advantage of creating more definite boundaries between the clusters.

After clustering, the results of the cluster analysis are as follows:

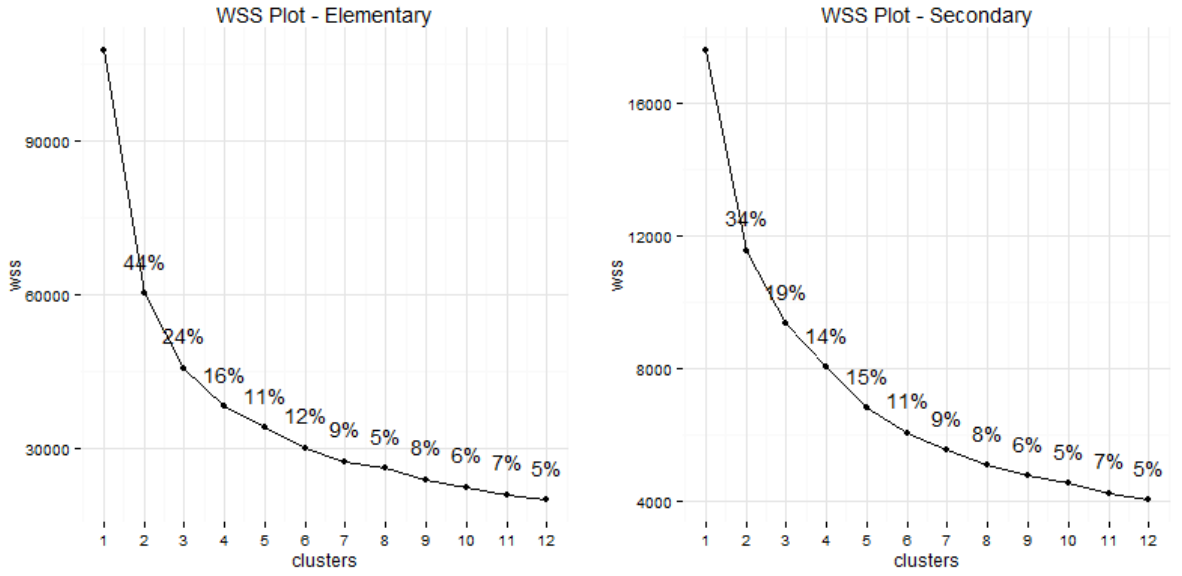


Capacity clusters for hierarchical (left) and k−means (right) clustering algorithms.

Because the shape of the input space is relatively simple, we prefer direct k-means, as it produces clearer and more distinct clusters. This would not be the case if the shape of the input space was more complex.

Cluster Profiling

We then perform the direct k-means algorithm separately on the elementary and secondary schools. In order to determine the ideal number of clusters, we plot the weighted sum of squares of the clusters that result at each level of k.



We set a cutoff of 10% or more in marginal reduction in the weighted sum of squares. Therefore for

both elementary and secondary schools, the selected number of clusters is 6. We perform the k-means clustering at $k = 6$ for both elementary and secondary schools with seed `721992`.

Table 2: Elementary Schools Cluster Profiles

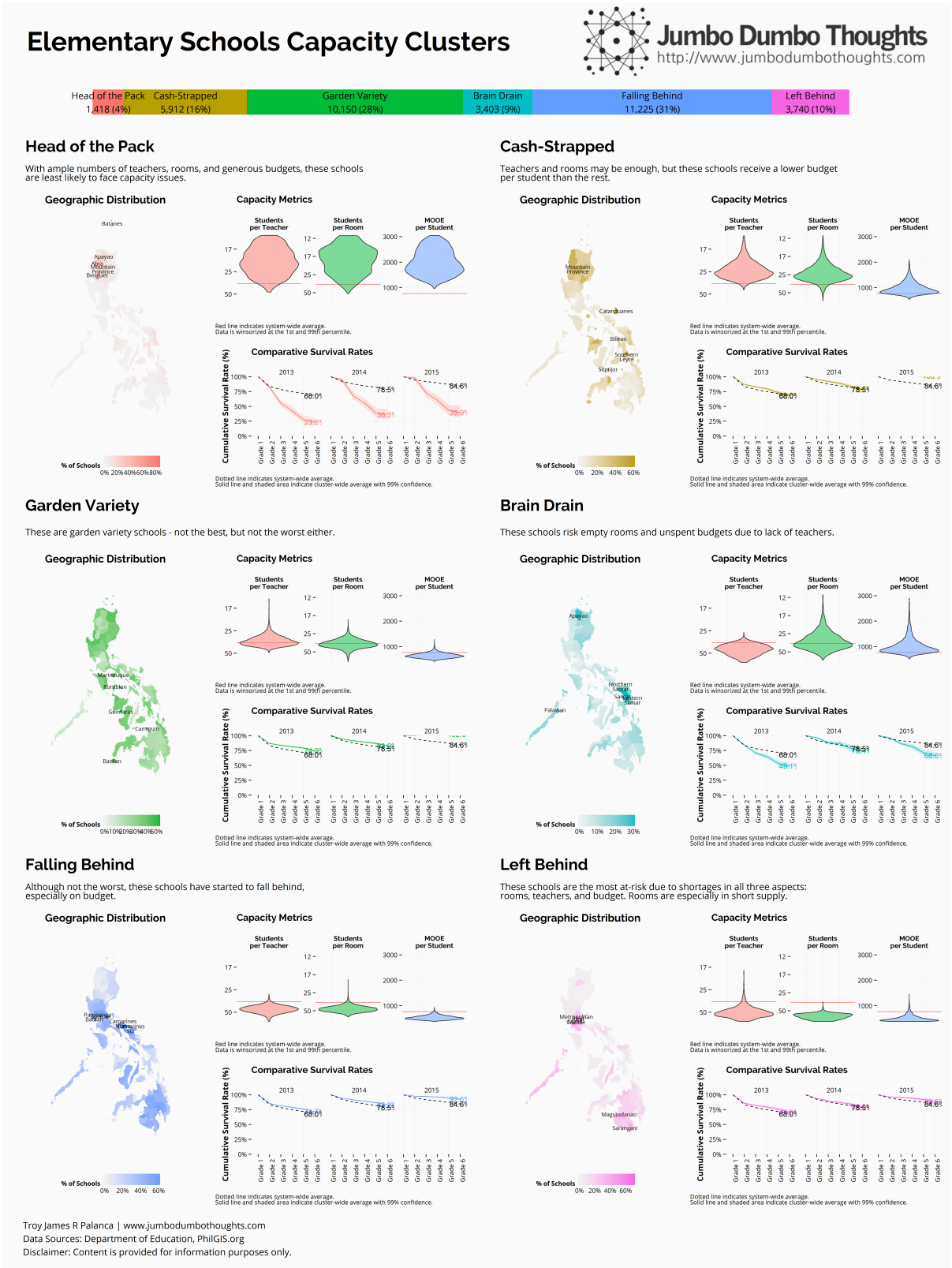
Cluster	Description	Students per Teacher	Students per Room	MOOE per Student
Head of the Pack	With ample numbers of teachers, rooms, and generous budgets, these schools are least likely to face capacity issues.	17	14	2582
Cash-Strapped	Teachers and rooms may be enough, but these schools receive a lower budget per student than the rest.	24	24	973
Garden Variety	These are garden variety schools - not the best, but not the worst either.	32	35	649
Brain Drain	These schools risk empty rooms and unspent budgets due to lack of teachers.	44	31	1032
Falling Behind	Although not the worst, these schools have started to fall behind, especially on budget.	44	45	531
Left Behind	These schools are the most at-risk due to shortages in all three aspects: rooms, teachers, and budget. Rooms are especially in short supply.	52	75	492

Table 3: Secondary Schools Cluster Profiles

Cluster	Description	Students per Teacher	Students per Room	MOOE per Student
Head of the Pack	With ample numbers of teachers, rooms, and generous budgets, these schools are least likely to face capacity issues.	8	18	4939
Close Second	These schools aren't far from the top. Capacity is unlikely to be a pressing issue.	18	33	1700
Brain Drain	These schools risk empty rooms and unspent budgets due to lack of teachers.	92	44	1105

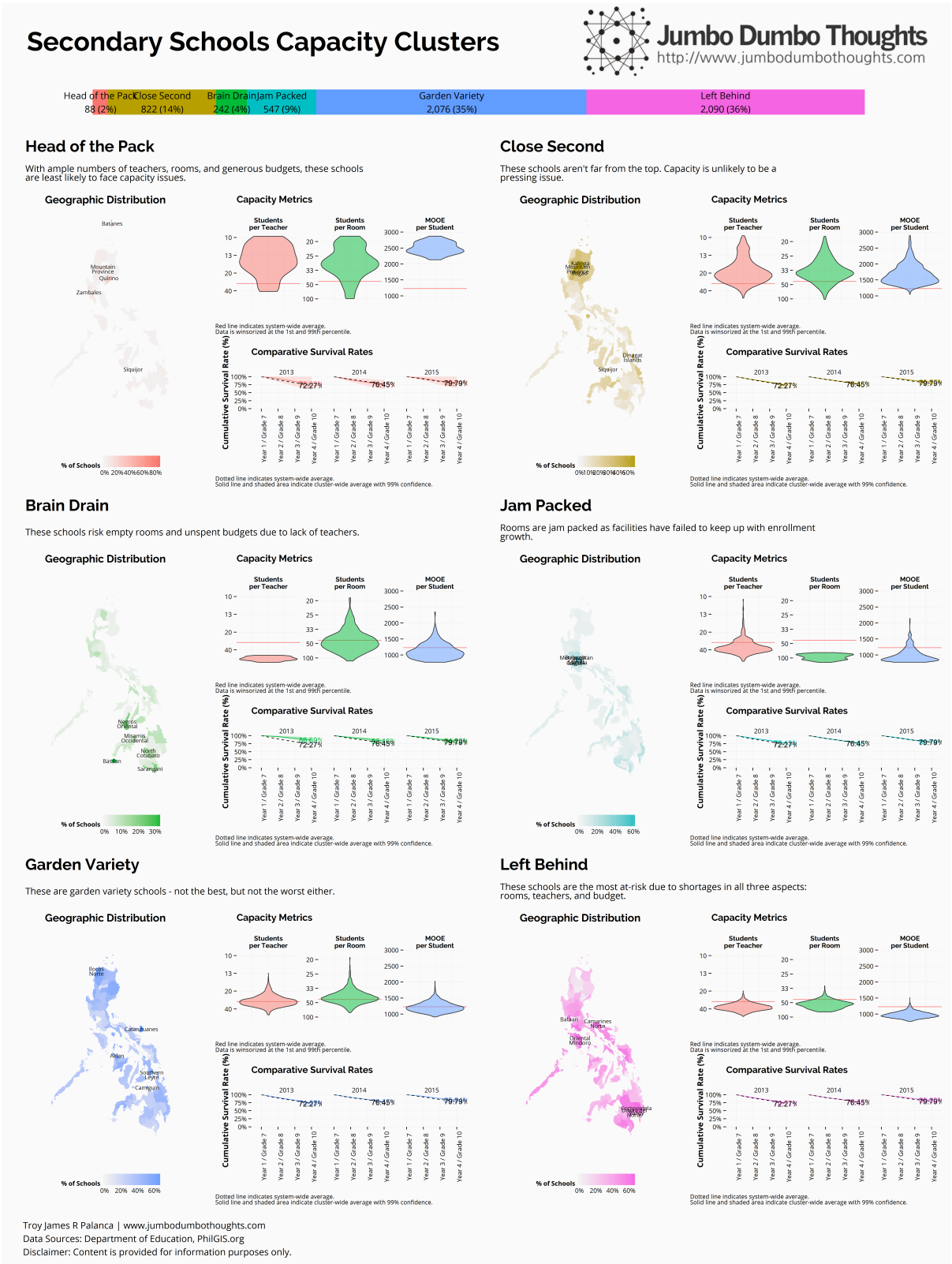
Cluster	Description	Students per Teacher	Students per Room	MOOE per Student
Jam Packed	Rooms are jam packed as facilities have failed to keep up with enrollment growth.	34	99	983
Garden Variety	These are garden variety schools - not the best, but not the worst either.	28	42	1220
Left Behind	These schools are the most at-risk due to shortages in all three aspects: rooms, teachers, and budget.	36	53	973

We then analyze the clusters by geographic distribution and comparative dropout rates.



Elementary schools with the best capacity metrics are those found in the upper part of Luzon. However, it is also to be noted that these are also schools with very high dropout rates compared to the rest of the public school system. An analysis of dropout rates and capacities is performed in the next section.

Cash-strapped schools are located near the eastern seaboard of Luzon. Teachers are in short supply in Eastern Visayas and Palawan. Most surprisingly, however, schools in the "Left Behind" and "Falling Behind" category are within Metro Manila and its environs.



While one might expect that the secondary school capacity situation mirrors that of elementary schools, the situation is quite different. Schools with the best and second capacity metrics, primarily in Central Luzon, Batanes, Dinagat Islands, and Siquijor, have dropout rates that are consistent with the national average. High school teachers are in short supply in Negros and Western Mindanao. Of course, Metro Manila, Bulacan, and Cavite schools round out the bottom, with severe shortages in rooms, teachers, and budget.