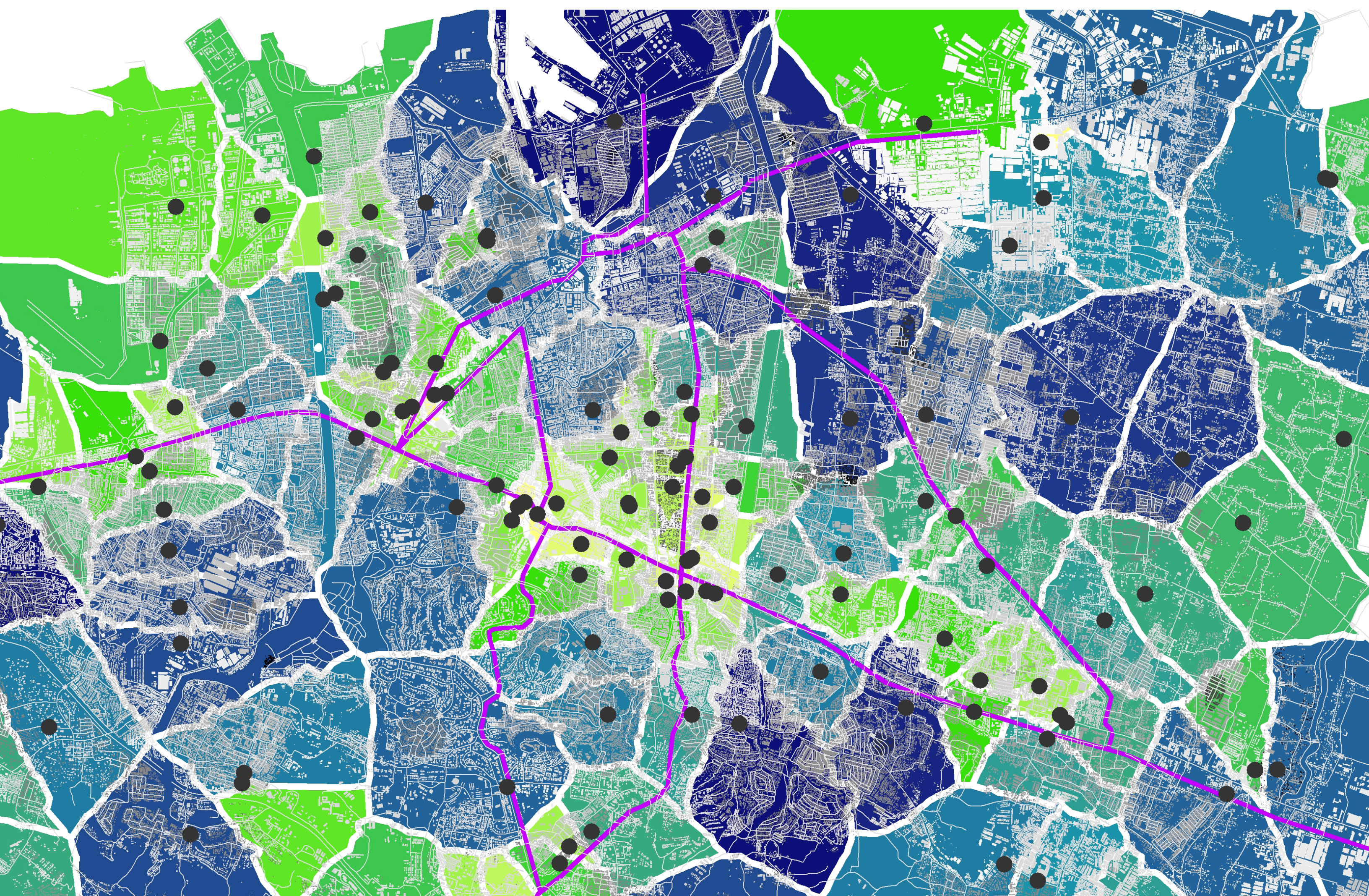
Matt Steele
MUSA Capstone
Spring 2016

Semarang Allocation Techniques

The driving question for this study is, 'does the spatial distribution of schools limit access to education for children in the city of Semarang, Indonesia?' The hypothesis of this report is that distributional access is unequal across the city, and that such limited access is significantly associated with poor school attendance by children.

In answering this question, this study attempts to understand the efficacy of place-based investment strategies. To do so, schools from various grade levels were analyzed making use of demographic data from the national census. The techniques discussed in this report demonstrate an effective means at understanding existing relative levels of access, with clear interpretable outcomes.

The purpose of this study is to inform future the management of school infrastructure and provide a basis and motivation for delineating useful catchment zones.



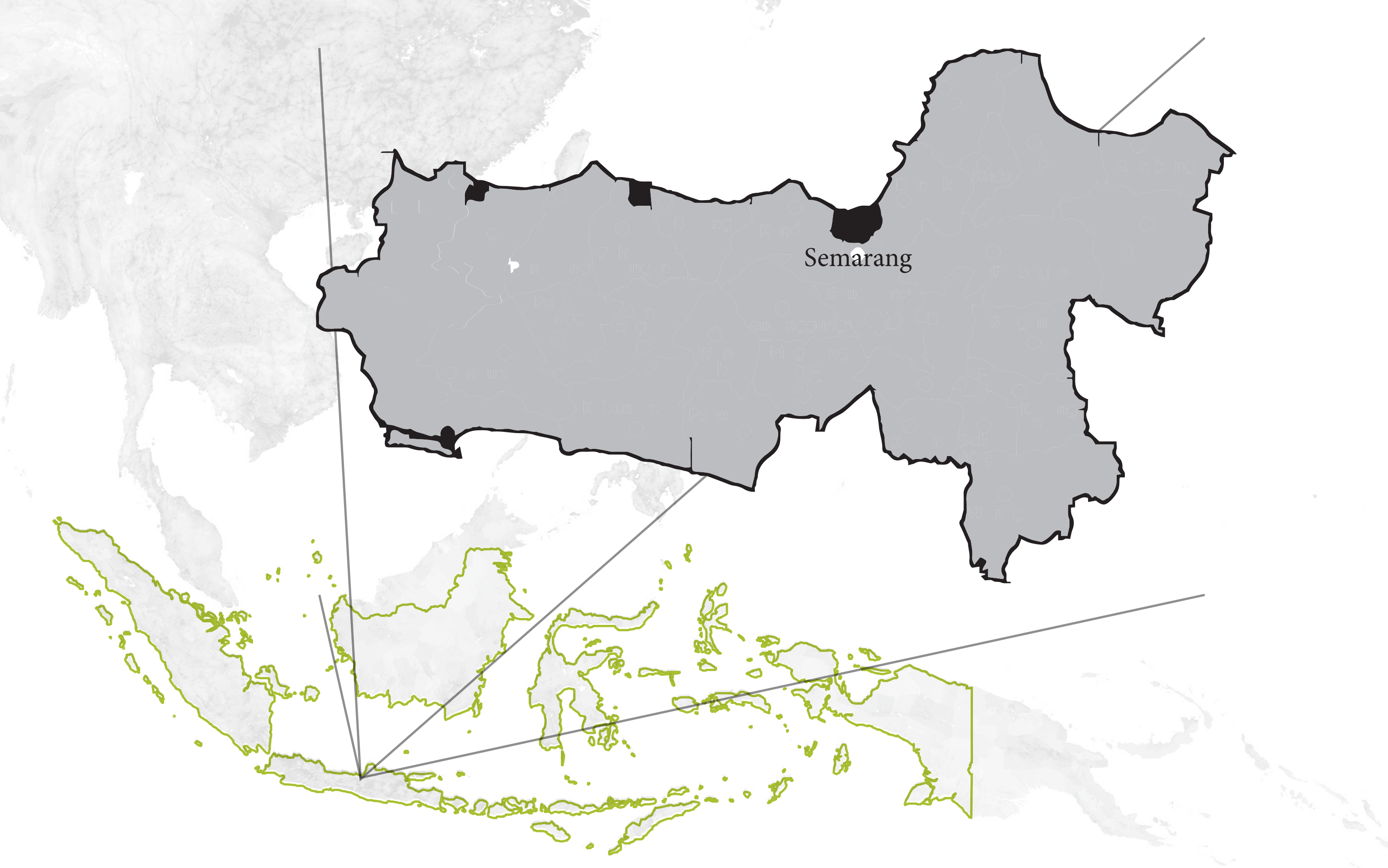
CONTEXT

Background

Indonesia is a major developing country, with the fourth largest population in the world. The country hosts some of the most rapidly urbanizing centers in the world. Indonesian rural migrants have moved to the country's urban centers at an annual rate of 4.4% since 1960, which is a faster average rate than all other East Asian Countries. The country's urban growth has been relatively compact. While such a development pattern means a smaller overall environmental impact, the concentrated growth has put pressures on existing infrastructure.

Indonesia's recent economic growth has been buoyed by the opening of internal markets and the country's rapidly growing consumer class, which will add an estimated 90 million new consumers by 2030. The Country's recent strong growth has been a function of its size and its development context. To sustain increasing well-being and decreasing poverty in the long-term, Indonesia must make investments in education that will ensure it can become competitive and join the ranks of other large developed countries. Unfortunately, education in Indonesia is lagging behind every other major East Asian country.

The Indonesian government began providing 9 years of public education facilities in the 80s. Compulsory education began in the mid-1990s, when 12 years of public education became freely available to students. Primary school spans 6 years. Middle schools provide the next level of educational across three years and high schools reflect the final three years of schooling. Because of the way the educational transfer system is structured, it is

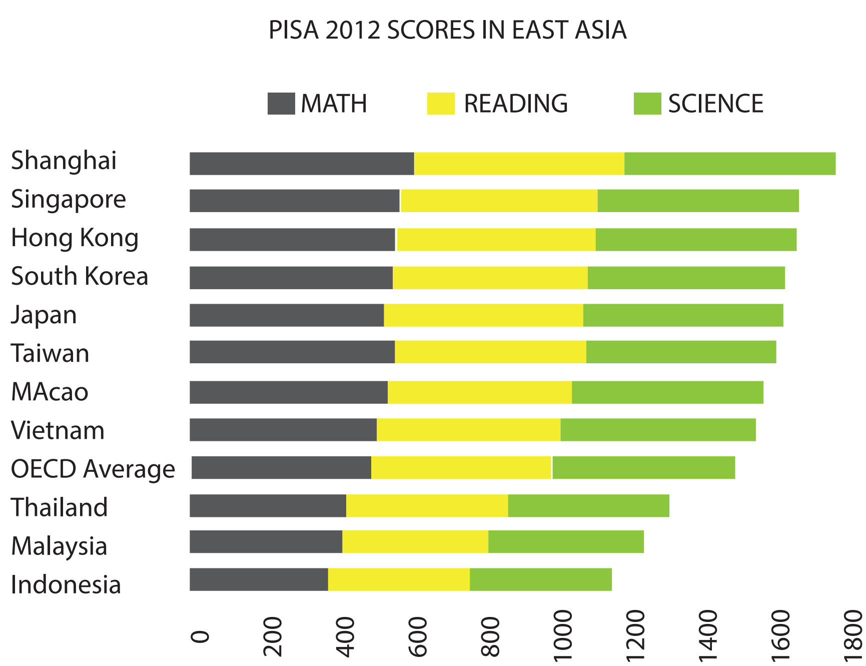


relatively easy for private schools to be set up and funded in part through the city. Particularly primary school certification has minimal requirements, and consequently there has been a proliferation of primary schools through Indonesia. Many of the primary schools are religious in nature, and are often housed in Madrasas. Primary schools exist in a hyper local content and cater to neighborhood children, who live within walking distance of their local schools. High schools, on the other hand, are treated as pre-professional educational programs with higher expected teacher qualifications. These have a regional draw due and limited available seats. As such, its student body commutes across longer distances to attend. Middle schools are in between, offering more advanced instruction but focused more on local neighborhoods.

While educational achievement has been a priority for national studies, few studies have attempted to examine the dynamics at play within the major cities. In 2014, the World Bank published a book titled "Teacher Reform in Indonesia, The Role of Politics and Evidence in Policy Making". This analysis considered data provided by the Indonesia Ministry of Education to understand the impacts of various policy reforms on the school systems and educational achievement that were initiated in 2005. These major policies reforms aimed to increase the quality of education in the schools, but the report indicates that these policies have had mixed results. The reforms were comprehensive but simplistic in their overall conception and their implementation. While there have been significant improvements in achievement, especially by increasing overall enrollment, education in Indonesia appears to be lagging further behind its East Asian peers. The report outlines a number of the barriers still being faced, and provides a nuanced understanding of the failings and successes of the proposed reforms. A significant failing identified is the lack of proper teacher credentials. Only a fourth of teachers throughout Indonesia have formal credentials to provide instruction. The lack of credentials is in part due to the lack of higher education resources available but it is also because teacher hiring and school budget allocations have been politicized, with teacher positions reportedly provided through family networks in exchange for political favors. The lack of proper teacher credentials is enabled by a lack of oversight on educational standards and quality. A key takeaway from this report is that because of the decentralization that occurred after the 1990s, the national government has little ability to make sweeping educational reforms, and the local municipalities shoulder much of the responsibility for improving educational outcomes.

A similar report produced in 2010 by the Indonesia / Australia Basic Education Program examines various barriers from a qualitative perspective. They uncovered that issues surrounding those in poverty remain a major barrier for achieving universal access among Indonesia's youth. While education has been made free in large part by the government, there are still expectations around books, food and uniforms that remain barriers for families with very little income. This report also identifies teacher quality as a major issue and that there is a poor distribution of trained teachers as well as a gap in the perceived quality of schools with their actual quality.

Figure 1: PSA Scores in East Asia



Semarang Context

Due to the historical and political governmental restructuring of 2001, governmental power in Indonesia is highly decentralized, with local municipalities holding the primary responsibility over infrastructure decisions and planning. This includes decisions and planning regarding educational services. The major cities of Indonesia are primarily on the Island of Java, and include Jakarta, Bandung, Semarang and Surabaya. Each of these cities has made improving educational services a key planning priority.

This study attempts to build a local understanding of the dynamics at play for educational achievement using the large coastal city of Semarang as a case study. It is the 6th largest city in Indonesia, with a population of 1.55 million as of 2010. The city has three major wings. The city's southern wing hosts one of the leading universities in the country, the University of Diponogoro, otherwise known as UNDIP. This southern wing has been the focal point for suburban style new development occurring in the city. The northeastern wing is where the grand proportion of development in the city is growing. This area is largely poor with a high ratio of children to total population. The westward wing has grown some, though expansion is limited physically due to the steep terrain that leads to a large mountain on the south-western border. This area is not densely populated, and has high levels of poverty. It remains in a near rural state of development

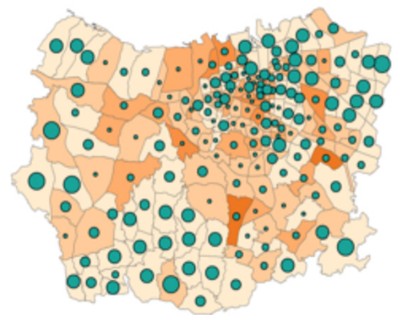


Figure 4: Poverty Levels in Semarang

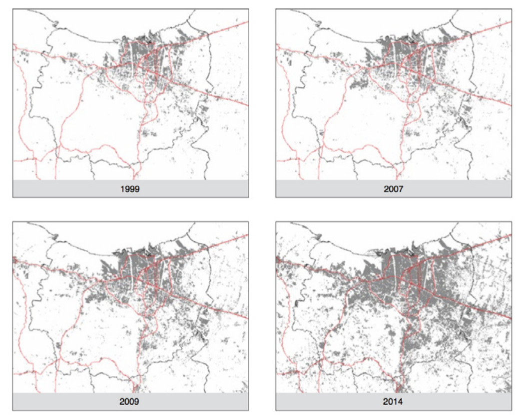


Figure 3: City Growth, Source :CPL

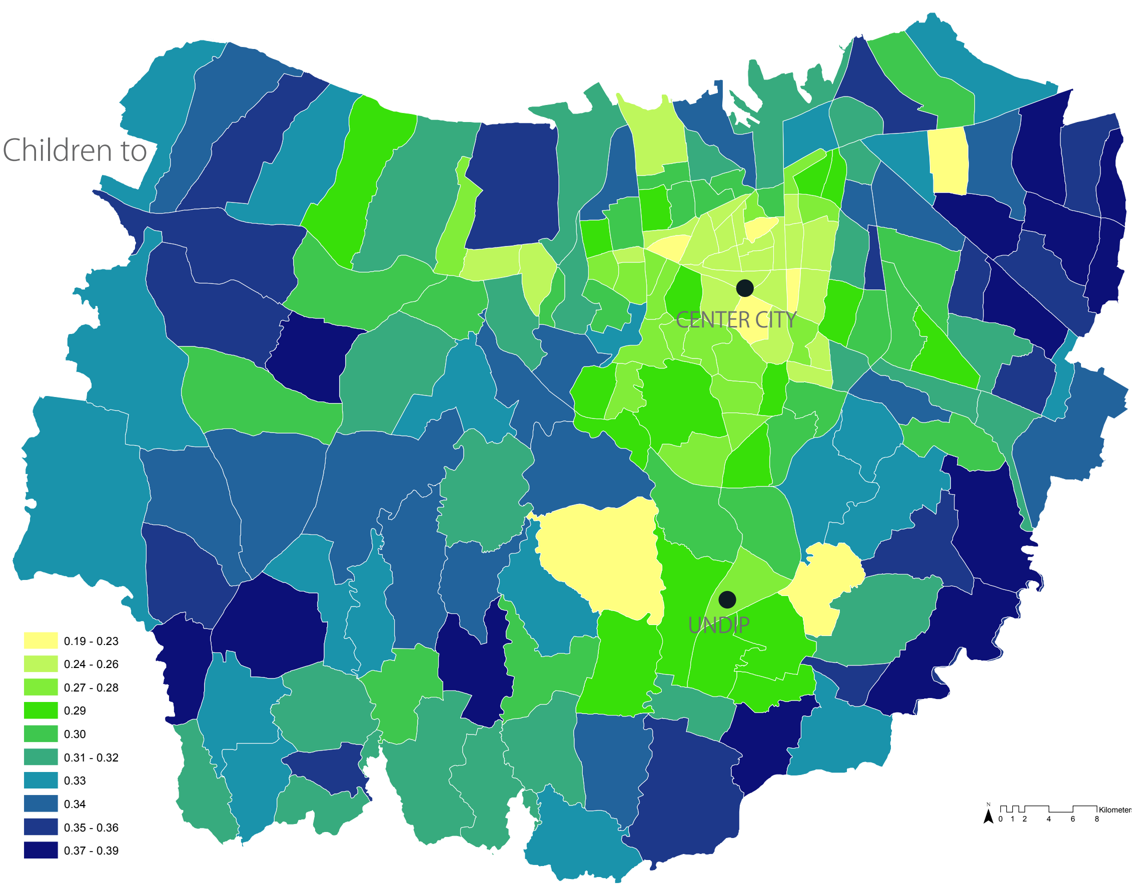


Figure 2: Ratio of Children to Total Population

DATA SOURCES

This study was inspired by engagements in Semarang through the World Bank's City Planning technical assistance (TA) program. The TA program provides support for Indonesia's major cities in developing their capacities for infrastructure planning.

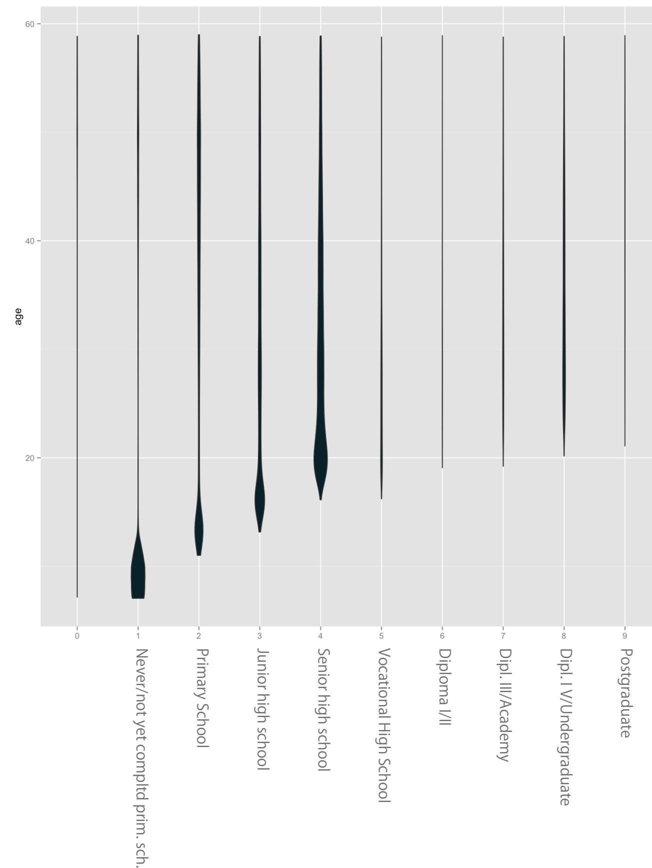
The school's data for this study comes from the Planning Department of the City of Semarang. Population characteristic data was built from the Indonesia 2010 Census micro-data set, made available through the World Bank's Jakarta office. The population figures were gathered from an individual level micro dataset of the census from the national statistics agency. Individual observations from the census database were summarized into their respective geographies in the open source statistical package 'R' and then mapped using ArcGIS.

The shapefiles for schools first provided by the city were not reflective of actual school locations based on satellite and ground imagery available through Google maps. Alternatively, the city provided a table of 1000 schools with addresses that were geo-referenced using Google maps for this report. Geo-locating by address in ArcGIS and other traditional geo-locating tools available online produced few reliable results. The table provided held information for both private and public schools. Many of the private schools were not in the locations specified, or their addresses were off in some way. There was a broad range of private schools in terms of type and size, with a few only reporting a handful of students. In contrast, the data for the public schools provided was more consistent, and the locations for nearly all was identified after ground-truthing through Google maps' street view tool.

Because the data was more reliable and because the city is primarily in control of public schools, public schools became the focus of this study. The data for public high and middle schools is especially consistent and comparable. The city also provided a number of other datasets, including a road layer and the BRT system in place. The road dataset appears to have some inconsistencies due to it being a non-rectified output of a remote sensing exercise. There are disconnections in the road dataset that were caused by visual obstructions in the satellite imagery used to create these datasets.

The administrative boundaries provided proved to complicate spatial analysis. There are a number of competing versions of administrative boundaries for Semarang. Because these boundaries hold political importance, they are contested and often change. The ambiguity of their delineation is a key reason why demographic spatial analysis is limited in Indonesia at the subnational level. The official boundaries provided the city are the not the same boundaries that the national statistical agency uses in building their datasets.

Figure 3, Education Level Achieved by Age



The administrative boundaries for the PODES village survey conducted by the national statistical agency is said to most align with the boundaries used for the 2010 census and were accessed through the World Bank consultant network. They are and are shown in Figure 1. The kecamatan or district level is delineated in green, while village (kelurahan) boundaries are in grey.

The population figures for this analysis were built from anonymized individual-level data from the National Census, conducted by the National Statistics Agency (BPS). The data was acquired through the World Bank's micro-database. While the Census is a comprehensive survey, the data across villages did not appear consistent with expectations.

The data was subsetted by age group to understand access particular to the respective populations.

The census data identified residents based on their academic level achieved as well as their current enrollment status. The adjacent chart in Figure 3 provides a glimpse into the composition of educational attainment throughout Semarang. Most complete high school, but few go on to higher education. Figure 7 shows the proportion of the population who were never enrolled in school. Ac

What is critical to an accessibility analysis is quality data. Much of this project has been coping with the poor quality data available in the development context. While even in the most developed context data issues abound there are many reasons to be concerned that this analysis is subject to data collection bias. While this is a clear limit to interpretation, this analysis can also provides a basis for improving the data systems for the city moving forward. Discrepancies are made obvious in reconciling and comparing datasets such as the building footprint data and the census data. The building footprint dataset suggests that there is some collection bias in the census information, as there are areas with populations inconsistent with the number of buildings present. Further, for a few small areas, the proportion of buildings in some areas seems too few given the population and location of the area. The census data at the sub-district level is of concern, as the distribution of the population by age was not consistent across geographies. The local city governments needs to develop closer relationships with their local census (BPN) agency in order to ensure that geographies don't change and that the census data is reliable consistently over time.

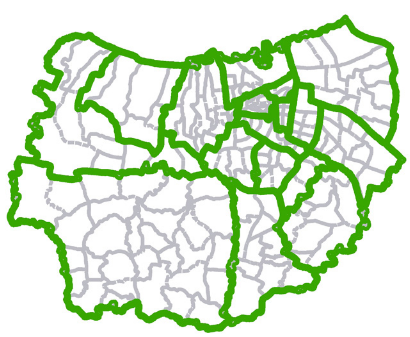
As Semarang develops its analytical capacities it should use each analysis as an opportunity to identify issues and anomalies in the data. The more the data is used, the more these issues are identified and the more data infrastructure is improved. For now, the lack of quality data has motivated a lock-down of access to city data. But data transparency only stands to improve data quality.

Geography / The Unit of Analysis

The primary unit of analysis for this study is the lowest administrative boundaries available, the village, or Kelurahan. There are two key geographies for the city level, the 'village' (Kelurahan) and the 'region' (Kecamatan), which are most closely akin to the census tract and census block group used in the United States, though they cover a generally larger area, and hold political importance.

School catchment areas are zoned by the subdistrict boundaries in a tiered priority system, whereby those living in the priority areas are given preference, and secondary priority is given to another set of sub-districts. Primary school catchment areas in Semarang are defined by the smaller geography, the kelurahan, as a basis. Primary schools have catchment areas that span 2 to sometimes 10 kelurahan areas. Middle and high school catchment areas are more broadly defined. Upper level grade schools have priority catchment areas of 5 to sometimes 10 kecamatan. There are only 16 kecamatan in the city, so that means students have priority for a large number of Semarang's schools, and access to nearly all middle and high schools within the city. More centrally located schools have the broadest availability, but also the greatest proximity to the best schools. The majority of students have the option of attending nearly any middle or high school, assuming there is space,

Figure 5: School Catchment Zones



the schools in the city to some limited degree. For these reasons the catchment area is not a meaningful or useful basis for understanding access. School catchment areas are further complicated by the unavailability of a consistent designation for the subdistrict boundaries. In most formal maps there appears two kelurahan in Semarang with the same name. Further, the boundaries have changed regularly, and different government entities recognize different boundaries, making the jurisdiction for each unclear.

Enrollment in the public schools appears to be a mostly competitive process. This has the benefit of increasing the competitiveness of schools, but it may also mean that there are barriers to quality schools for the average student, who might otherwise have clear geographic priority.

Dependent Variable

This analysis looked at two proxies for academic performance. The first shown in Figure 7 displays the extreme cases of individuals never having attended college. The number of people who were reported as never having attended school was small in number. The other proxy is the number of children not currently enrolled in school, which demonstrates poor attendance overall for school attendance by age group category. There were more numerous observations of individuals being not currently in attendance. For this reason, poor enrollment was better for modeling relationships, and became the key dependent variable. Though it should be said that their patterns are different, with the extreme case of never attending school more clearly following the pattern of rural development in the city, near the mountains in the south east, while those who demonstrated poor attendance was less concentrated across the city, with a large segment being near the center of the city.

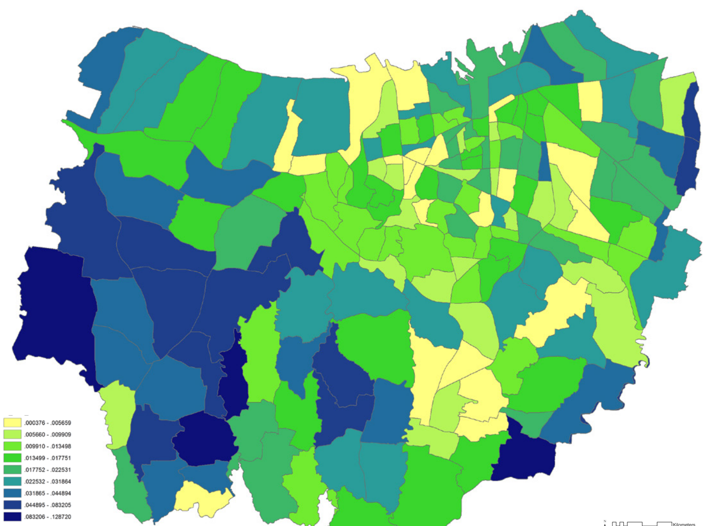
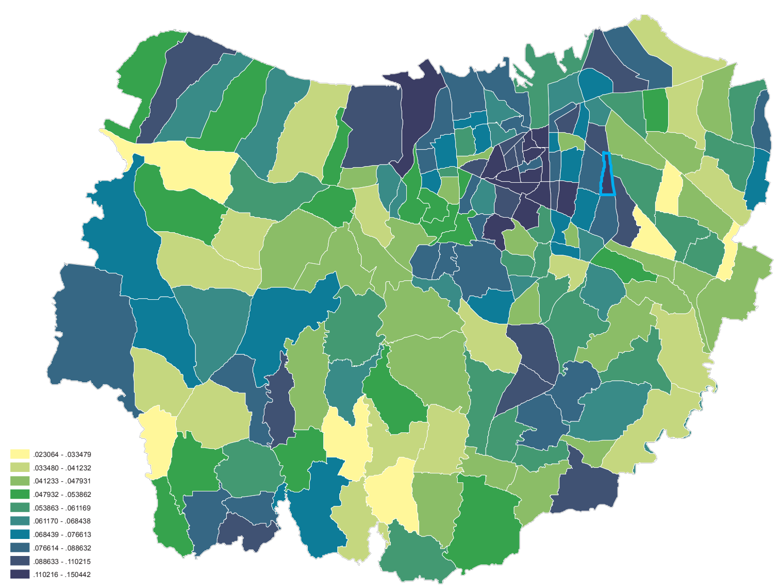


Figure 7: Never Enrolled,As A Percent Of Total

Figure 6: Percent of children who are currently not enrolled



METHODS

Overview

This analysis develops a basis for identifying accessibility to schools, and then tests this metric's significance as it relates to poor attendance of schools by children. To test the hypothesis, a bivariate logistic regression is taken on each set of schools, considering school attendance as the dependent variable.

To create the accessibility metric, schools are associated with the populations in their immediate surroundings. Such a calculation provides a sense of how well the schools are distributed, both in terms of space and population. The proposed metric for is determined based on school points, a friction grid and a granular population distribution. The analysis was applied to primary, middle and high schools in Semarang Indonesia. The technique allows one to understand potential zones of access with respect to the infrastructure and barriers in place as well as with respect to the distribution of the targeted populations for each. The intention of this analysis is to inform a process for determining accessibility based on points of access for various types of infrastructure.

There are three critical steps for this analysis. The first is identifying an appropriate friction layer. The second is creating a population grid that is sufficiently granular for the allocation geographies to avoid the modifiable areal unit problem. Using these layers, the proportion of the population with respect to the access zones of each point of infrastructure can be identified in terms of the population with convenient access. This is a superior method to buffering and network analysis in that it takes into account the presence of other points of access and creates a basis for considering the populations present with respect to each of the access points for infrastructure in place.

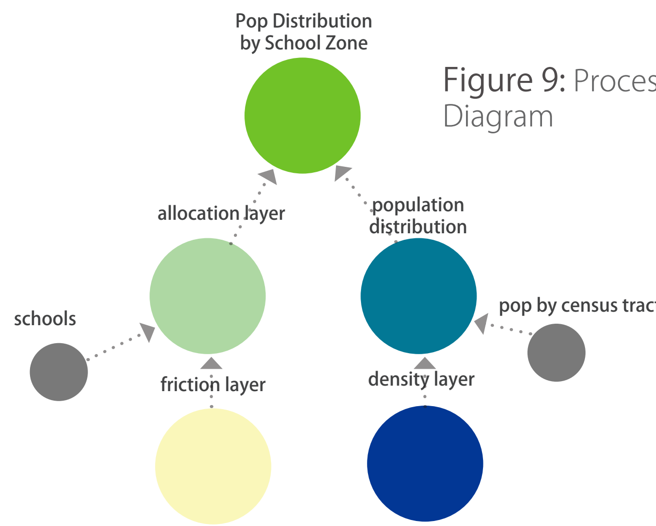


Figure 9: Process Diagram

The goal of the metric is to provide a sense of the number of children, by appropriate age group, within each walking shed of each school, by grade. This gives a clear indication of the level of access, considering the distribution and density of children, as well as the layout of infrastructure.

Friction Layer

The first key step in creating the accessibility metric is using a mobility friction layer to account for barriers and routes. This layer becomes the basis for delineating walking sheds for each school. Conceptually, a friction layer can be used to identify the flow of movement, much like how the slope of a surface directs water. The cells of a friction layer are imbued with values that reflect the abstract costs of traversing them. Major highways, building footprints and rivers were used as a basis for barriers in this analysis, while open space and the existing road network were identified as the primary pathways. For this analysis, open space had a lower score than road pathways, meaning they could be traversed if a roadway path was too cumbersome. Not only is this appropriate for a pedestrian-based friction layer, it is important because the road layer is incomplete in parts. For this analysis, roads were set to 3, making it the preferred path, while open space was set to 10. Barriers such as rivers, the major highways and buildings were set to high values of 100, making them essentially untraversable. The absolute values set in the friction layer are not as important as their relative values for purposes of a

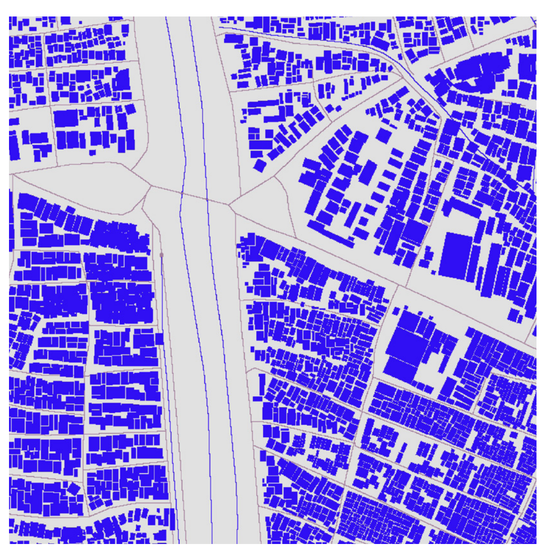


Figure 8: Close Up Friction Layer

cost allocation analysis. The cost allocation tool function requires that the starting point for the distribution not be NoData, which is why it would prove complicated to make the buildings completely untraversable, since the schools are situated on buildings. Road intersections with rivers were assumed to be bridges. The layers were classified by their cost value using the reclassify tool and the final raster was stitched using the mosaic tool in ArcGIS.

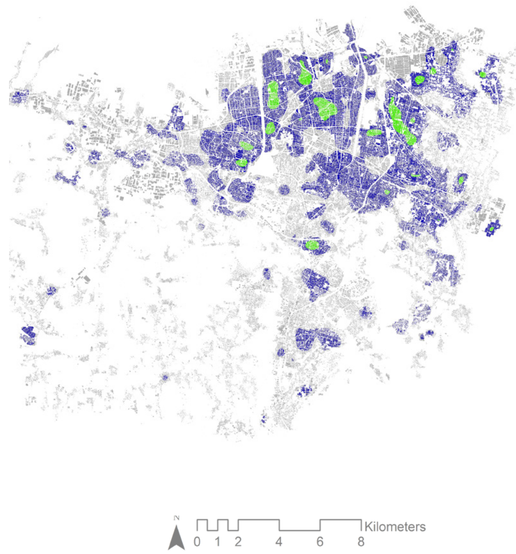


Figure 6: Density Score

Density Score

The intention of this analysis is to understand accessibility with respect to both population and space. For this reason, it is important to consider the distribution of the population. There are a number of techniques for accounting for population in distributional analysis, but where population data is provided in a unit of geography, the analysis must contend with the Modifiable Areal Unit problem by which a geography's spatial definition can bias the interpretation of the data it is reflecting. The best means of mitigating this is to use a unit of analysis that is sufficiently small for the scope of analysis being undertaken. Because school zones are within the context of a walk-shed, the scope of the analysis is at the neighborhood level. Building footprints provide perhaps the most granular unit for purposes of population attribution, and would be appropriate for this scale. For this reason, a dasymetric mapping method was employed to attribute the population data to a recently created dataset of building footprints based on an algorithm of population density.

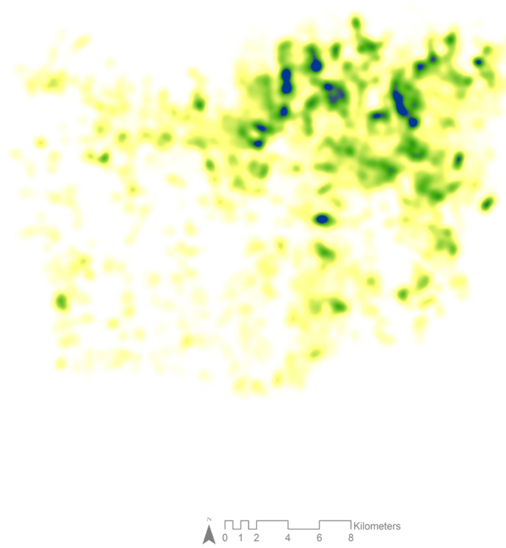


Figure 7: Building Kernel Density

The basis for population density was derived using two considerations, though a number of other factors can be considered. The first was the literal density of the building footprints, computed using the Kernel Density tool. The buildings were categorized by three levels, based on standard deviations, reflecting low, medium and high densities. Secondly the area of the building footprint was used as a basis for density, which was identified using calculate geometry. The buildings were then also classified into low, medium, and high. Buildings were designated based on their area in square meters. The highest density was designated with a score of 3 and reflected 80 square meters. Areas between 80 to 200 square meters were associated with medium density, and buildings with larger footprints than 200 were classified as being more associated with low density. This practice is best paired with a consideration of detailed land use patterns and an understanding of average or allowable floor area ratio by zone (FAR), which is seemingly unavailable in Semarang. The FAR becomes a discounting factor that allows one to consider the use of space with respect to the height of each building in com-

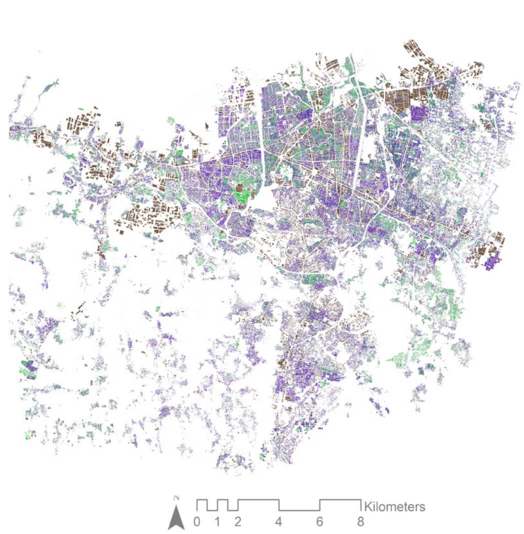


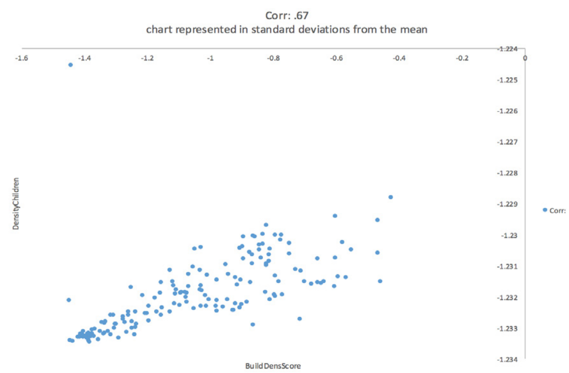
Figure 8: Building Area Score

ing up with density calculations.

A population distribution algorithm can incorporate multiple considerations. Densities throughout Semarang are roughly the same, as the city hosts few tall residential buildings. Most residential buildings are consistently of the same footprint. Large building footprints are associated with wealthier homes, and smaller footprints are associated with slum like housing. Very large footprints are primarily institutional and commercial buildings, though a number are dense residential As evident in the census data, these trends reverse outside the compact areas of the center part of the city. Alone building foot print size is a poor indicator of overall density of the area, but combined with other factors, it can provide a decent basis. These two scores were averaged using the weighted overlay tool. The final output is shown in Figure 10.

One method of evaluating how decent the density score reflects population distribution is to compare the mean aggregate of the density score by village census tract to the actual population density of each geography. As shown in Figure 11, the density score appears to have a .67 correlation with the actual density of population, at the village level. A correlation of 1 would mean that the values perfectly vary together. A correlation above .5 for variables not derived from the same source is considered strong and demonstrates a meaningful relationship. On that basis one can assume this distribution has a generally decent reflection of reality. The scoring basis can be tweaked for improvement but because it is a proxy score, improvement will never reach 1.

Figure 11: Final Density Score



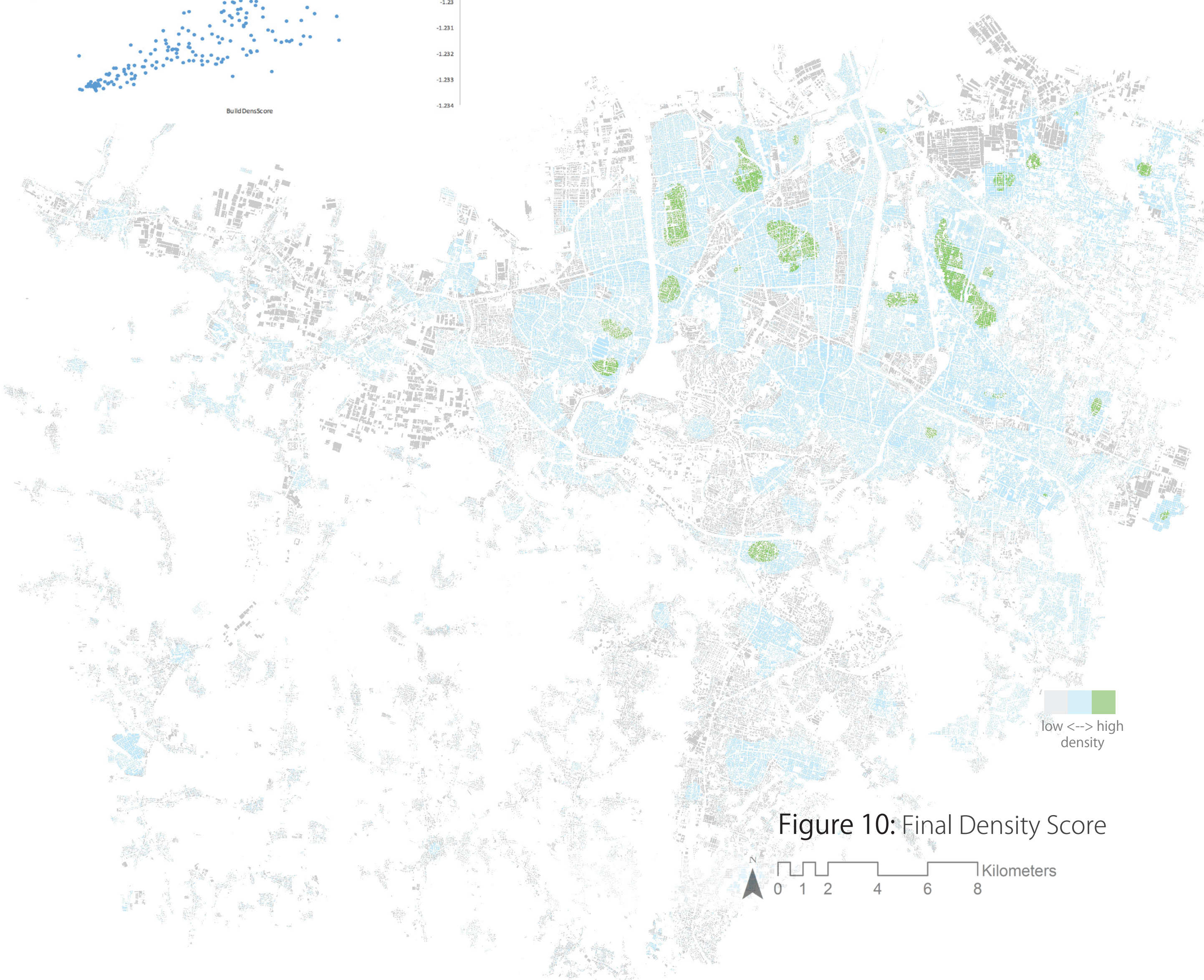


Figure 10: Final Density Score

Dasymetric Calculations

The dasymetric process attributes population figures, in this case the number of children by age group, to a lower geography, in this case the building footpints. The output is an approximate disaggregation that provides a granular basis for understanding access. The process for creating this raster involves identifying for each building type an approximate associated population. The density score (1-3) created was translated into an expected population distribution, with 1 being associated with 5 children, 2 being associated with 15, and 3 being associated with 30. The actual density figure is not as important as their relative value. This means that high density areas are expected to have twice as much population as medium density areas, and low density pixels are expected to have one third the population as medium density areas. The distribution calculation is made by tabulating to table the count of the cells associated with each density level, using the village geography. Using this understanding, an enumerator (e) is derived that reflects the exact expected proportion of the population to attribute to each building based on their score type. The enumerator is the sum of the proportion in each density level, weighted by the expected density for each (5,15,30).

To ensure accuracy, all calculations should be made using the same cell size. The cell size becomes the unit for the model and its dimensions should be incorporated into the final analysis. A cell size of 1 makes an analysis easiest to compute but it may be difficult for less powerful computers to analyze an entire city the scale of Semarang using such a small cell size.

Once an enumerator map is created, this dasymetric map can be used as the basis for attributing any population data from the same census source. This analysis assumes that the differences between the distribution of children relative to total population at the village level is normally distributed. The larger the scope of analysis the less likely this is the case.

$$([DensScore] * [population] * cell\text{-}size(1)^2) / ([e] * [total])$$

To ensure the outcomes of these calculations reflect the distributions of the populations accurately, one should compare the sum of the outputted distribution to the sum of the inputted population village count. There may be some small rounding, but if the calculation worked correctly the populations sum for each kelurahan based on the pixels should be the same as the original population from which it was distributed. In the final output, each pixel has an estimated population associated with it, that appears to range from half a person to over three people. This can be done using zonal statistics, sum tool in ArcGIS.

For purposes of this analysis, this attribution method was applied to three age groups (6-11,12-15,16-18) as well as total population and the total number of children. The population breakdowns by age was derived from the micro census data using R.

Figure 12: Density Score Tools Diagram

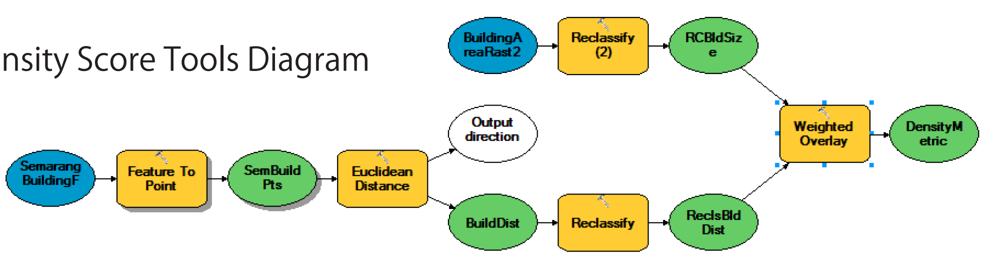


Figure 13: Dasymetric Table Glimpse

o11	twlvto15	sixtnto18	FID_1	VALUE_56	VALUE_1_16	VALUE_3_31	Total	L	M	H	E
	944	1,174	0	106,162.000000000000000	329,195.000000000000000	0.000000000000000	435,357.000000000000000	0.243850449172000	0.756149550828000	0.000000000000000	5,468,735.0
	601	717	1	135,407.000000000000000	85,140.000000000000000	0.000000000000000	220,547.000000000000000	0.613959836225000	0.386040163775000	0.000000000000000	1,954,135.0
88	2,973	2,843	2	10,854.000000000000000	637,356.000000000000000	451,828.000000000000000	1,100,038.000000000000000	0.009866931869630	0.579394530007000	0.410738538123000	23,169,450.
33	2,448	3,308	3	35,589.000000000000000	641,773.000000000000000	36,592.000000000000000	713,954.000000000000000	0.049847749294800	0.898899648997000	0.051252601708200	10,902,300.
	727	829	4	7,948.000000000000000	491,647.000000000000000	0.000000000000000	499,595.000000000000000	0.015908886197800	0.984091113802000	0.000000000000000	7,414,445.0
	228	250	5	48,472.000000000000000	107,008.000000000000000	0.000000000000000	155,480.000000000000000	0.311757139182000	0.688242860818000	0.000000000000000	1,847,480.0
	981	1,368	6	53,204.000000000000000	330,624.000000000000000	7,465.000000000000000	391,293.000000000000000	0.135969720900000	0.844952503623000	0.019077775477700	5,449,330.0
	351	268	7	0.000000000000000	167,334.000000000000000	18,143.000000000000000	185,477.000000000000000	0.000000000000000	0.902181941696000	0.097818058303700	3,054,300.0
	328	368	8	2,760.000000000000000	144,001.000000000000000	5,060.000000000000000	151,821.000000000000000	0.018179303258400	0.948491974101000	0.033328722640500	2,325,615.0
	474	218	9	0.000000000000000	97,274.000000000000000	0.000000000000000	97,274.000000000000000	0.000000000000000	1.000000000000000	0.000000000000000	1,459,110.0
	380	543	10	0.000000000000000	70,721.000000000000000	59,521.000000000000000	130,242.000000000000000	0.000000000000000	0.542996882726000	0.457003117274000	2,846,445.0
	232	233	11	0.000000000000000	44,875.000000000000000	90,482.000000000000000	135,357.000000000000000	0.000000000000000	0.331530692908000	0.668469307092000	3,387,585.0

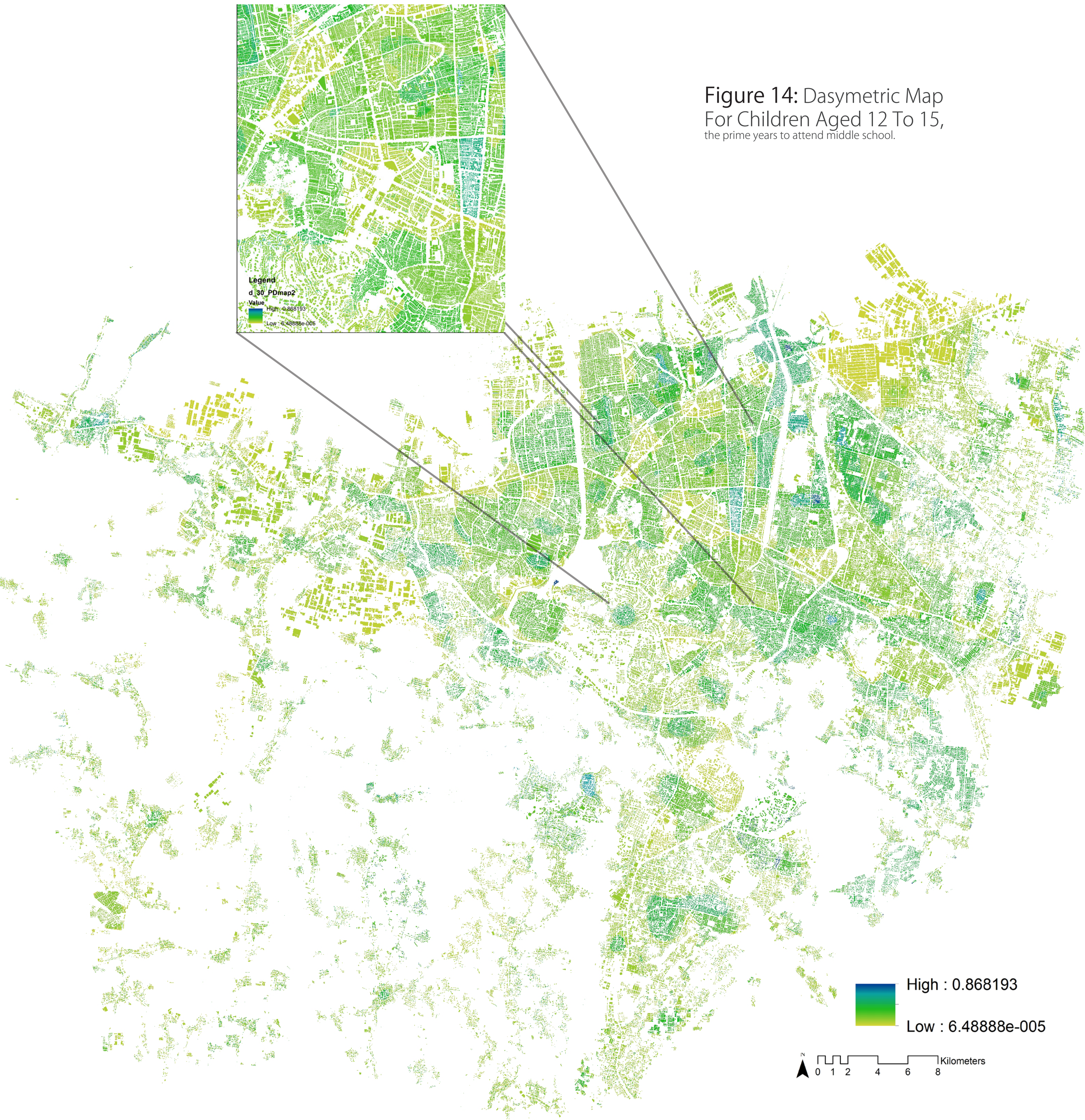


Figure 14: Dasymetric Map For Children Aged 12 To 15, the prime years to attend middle school.

Allocation Mapping of Schools

The population distribution created using the dasymetric method becomes the population basis for understanding levels of access to educational services for each school point. The access zones are created using school points and a friction layer to identify spatial zones using the cost allocation tool. Each of the three school grades was used to build this layer after turning them into 30-cell sized blobs using the focal statistics mean tool. The blobs are made to ensure the school access point reaches beyond the building it is sitting on to ensure the friction layer apportions the access zones correctly. The suggested schools zones created were then turned into polygons and the respective population sum for each was calculated using the zonal statistics tool as map, using the dasymetric population distribution as a basis. High school zones were aggregated to the dasymetric maps of 16 to 18 year olds. Middle school zones were aggregated to dasymetric maps of 12 to 15 year olds and primary schools for the dasymetric maps of children aged 6 to 11. The maps produced indicate the number of students with convenient access to each school. The resulting population sums by zone were rounded down, transformed into integer rasters and then turned into polygons that were then spatially joined with the school points for which they are associated. The attribute tables for the schools were then exported and the schools were ranked based on the size of the population and the area of their allocation they likely conveniently serve.

Segmenting the polygons by total population and by size can give one a sense of the distributional access for each point.

More differentiation between school types would help to provide more nuance to this analysis. In that spirit, the vocational high schools appear to be adjacent to normal high schools and were separated out for purposes of the analysis. Further information about the number of teacher or facilities could provide a better means of understanding the schools ability to meet student demand.

The following rankings depict the suspected proportion of students that each school is likely to be serving. Those to the top are schools with an unusually large student burden. Further information about the facilities for each school could be used to enhance this analysis, as this analysis assumes that all schools are of equivalent capacity. The simplest method to do so would be to identify the number of teachers in each school and calculate the expected teacher to student ratio for each. Some of the school data at student population figures.

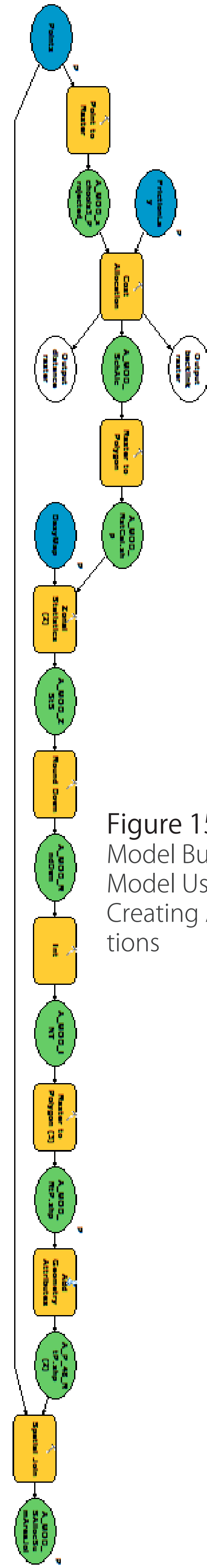


Figure 15: ArcGIS Model Builder Model Used For Creating Allocations

Figure 17: Cost Allocation, High Schools

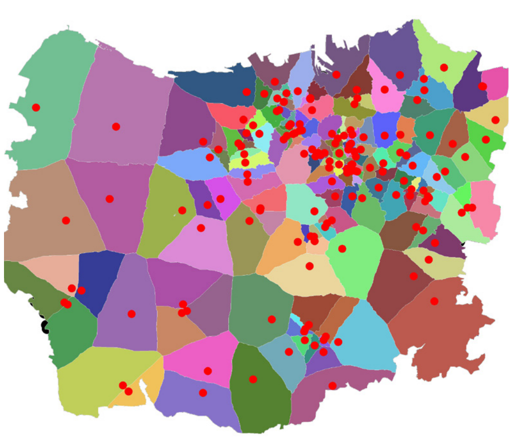
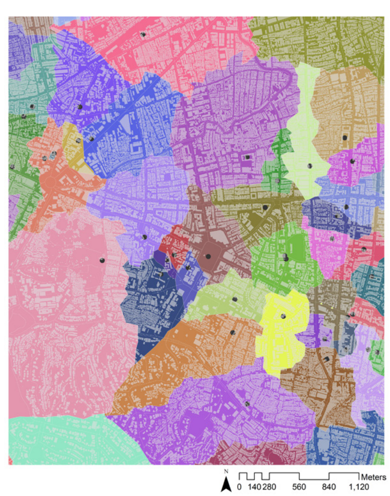


Figure 16: Cost Allocation, High Schools, Close Up



Public Middle and High Schools

Figure 19: Public Middle Schools Service Area

NSS	Grade	Nama	Alamat	Akreditasi	MalePop	FemalePop	PopInZone
201036310004	SMP	SMP Negeri	Jl. Tambak D	1	112	140	6,005
201036308034	SMP	SMP Negeri	Jl. Tlogomuly	1	120	136	5,353
201036316016	SMP	SMP Negeri	Jl. Prof Dr. H	1	129	125	5,344
201036307033	SMP	SMP Negeri	Jl. Kompol R.	1	138	118	5,316
201036308015	SMP	SMP Negeri	Jl. Supriyadi	1	138	150	4,809
201036314019	SMP	SMP Negeri	Jl. Abdulrahr	1	127	129	4,760
201036304013	SMP	SMP Negeri	Jl. Lamongar	1	117	135	4,278
201036309020	SMP	SMP Negeri	Jl. Kapas Uta	1	124	131	4,178
201036303027	SMP	SMP Negeri	Jl. Ngesrep T	1	122	132	3,911
201036312038	SMP	SMP Negeri	Jl. Bubakan I	1	70	87	3,804
201036311006	SMP	SMP Negeri	Jl. Pattimura	1	96	160	3,737
201036306008	SMP	SMP Negeri	Jl. Cinde Ray	1	98	151	3,536
201036308014	SMP	SMP Negeri	Jl. Panda Ray	1	119	136	3,513
201036307017	SMP	SMP Negeri	Jl. Gabeng R.	1	131	123	3,391
201036308009	SMP	SMP Negeri	Jl. Sendang I	1	125	163	3,167
201036314030	SMP	SMP Negeri	Jl. Amarta N	1	113	143	3,132
201036313025	SMP	SMP Negeri	Jl. Kualamas	1	127	114	3,042
201036301023	SMP	SMP Negeri	Jl. RM. Hadi	1	121	135	2,954
201036315028	SMP	SMP Negeri	Jl. Kyai Gilan	1	123	130	2,849
201036312036	SMP	SMP Negeri	Jl. Plampitan	1	135	152	2,797
201036306005	SMP	SMP Negeri	Jl. Sultan Agu	1	122	166	2,417
201036303012	SMP	SMP Negeri	Jl. Ace 42	1	129	126	2,342
201036314001	SMP	SMP Negeri	Jl. Ronggolav	1	120	167	2,297
201036304011	SMP	SMP Negeri	Jl. Karangrej	1	118	137	2,245

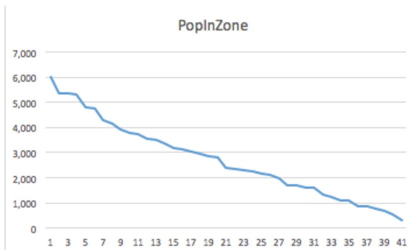
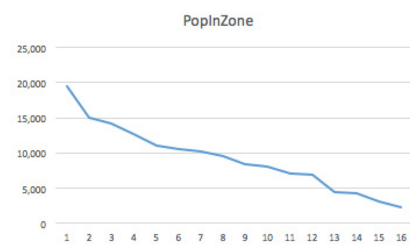


Figure 18: Public High Schools Service Area

NSS	Grade	Nama	Alamat	Akreditasi	MalePop	FemalePop	PopInZone
301,036,309,010.00	SMA	SMA Negeri 10	Jl. Padi Raya No. 16	1	150	169	19,480
301,036,305,011.00	SMA	SMA Negeri 11	Jl. Lamper Tengah XIV	1	172	250	14,999
301,036,307,015.00	SMA	SMA Negeri 15	Jl. Kedungmundu Raya No. 3	1	151	200	14,172
301,036,316,007.00	SMA	SMA Negeri 07	Jl. Untung Suropati	1	131	226	12,637
301,036,305,001.00	SMA	SMA Negeri 01	Jl. Taman Menteri Supeno N	1	200	271	11,091
301,036,307,006.00	SMA	SMA Negeri 06	Jl. Ronggolawe	1	184	245	10,543
301,036,316,008.00	SMA	SMA Negeri 08	Jl. Raya Tugu	1	121	203	10,182
301,036,308,002.00	SMA	SMA Negeri 02	Jl. Sendangguwo Baru No. 1	1	148	268	9,536
301,036,303,009.00	SMA	SMA Negeri 09	Jl. Cemara Raya	1	164	196	8,470
301,036,312,005.00	SMA	SMA Negeri 05	Jl. Pemuda No. 143	1	187	197	8,098
301,036,303,004.00	SMA	SMA Negeri 04	Jl. Karangrejo Raya No. 12 A	1	162	186	7,073
301,036,313,014.00	SMA	SMA Negeri 14	Jl. Kokrosono	1	149	174	6,881
301,036,302,012.00	SMA	SMA Negeri 12	Jl. Raya Gunungpati	1	156	198	4,403
301,036,312,003.00	SMA	SMA Negeri 03	Jl. Pemuda No. 149	4	178	255	4,219
301,036,301,016.00	SMA	SMA Negeri 16	Jl. Ngadirgo Tengah	1	90	123	3,074
301,036,301,013.00	SMA	SMA Negeri 13	Dk Rowosemanding	1	112	171	2,310



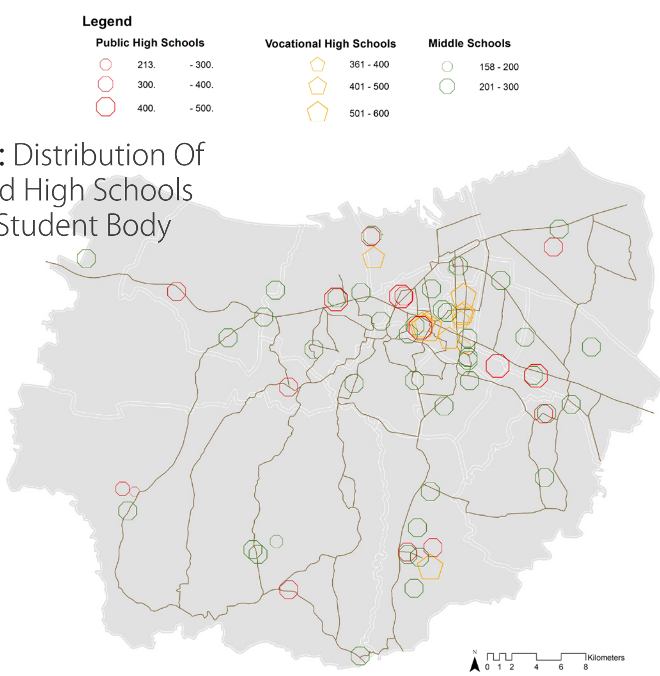
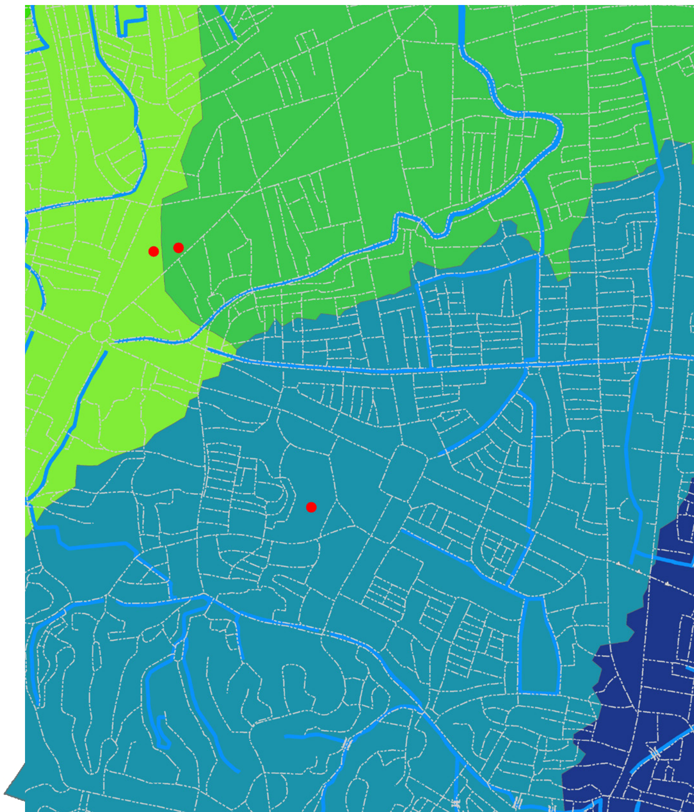


Figure 20: Distribution Of Middle And High Schools Based On Student Body

The vocational high schools tend to be bigger and serve primarily male identified individuals. The middle schools are more standardized in size, while the normal high schools vary a bit. Those in the center are much larger than those on the periphery. This is in part due to a larger regional draw for the more pre-professional high schools. The frequency distributions compares each schools level of access along a ordered continuum. The flatter the distribution, the more spatially even the level of access.

Figure 21: Public High Schools

The distribution of public high schools is spatially even, roughly following the pattern of the population. Heavily populated areas in the northeast are undeserved. This is also the poorest and fastest growing part of the city.



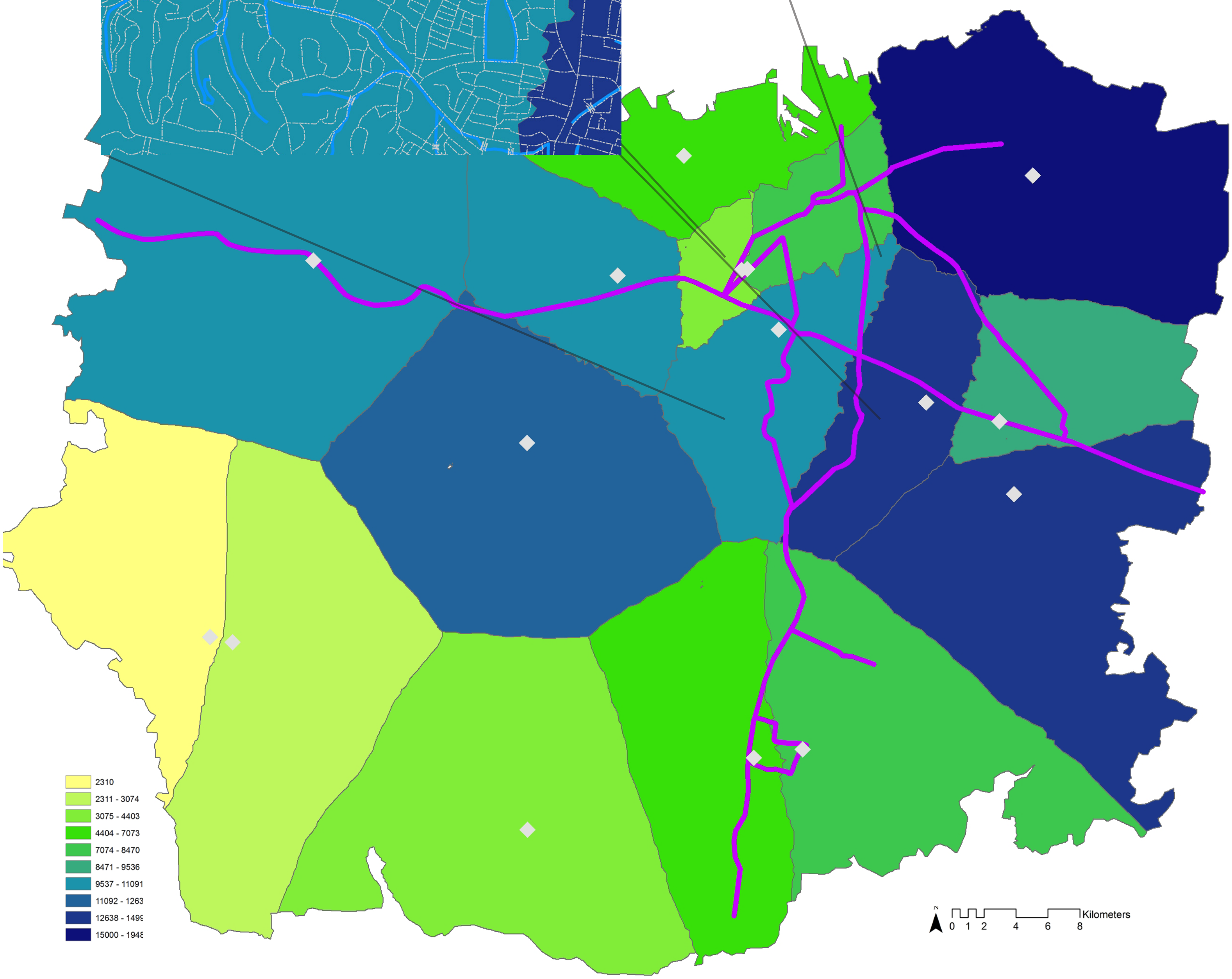
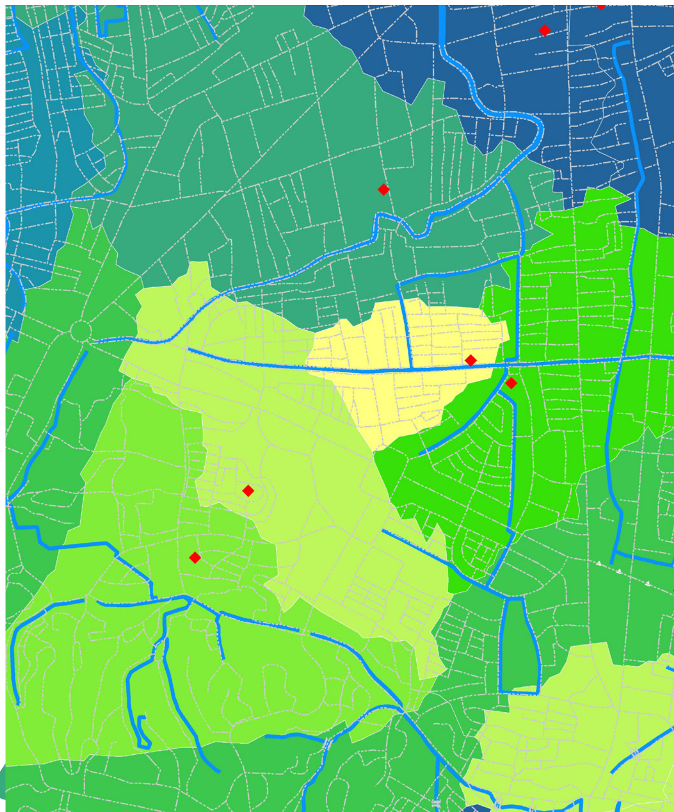
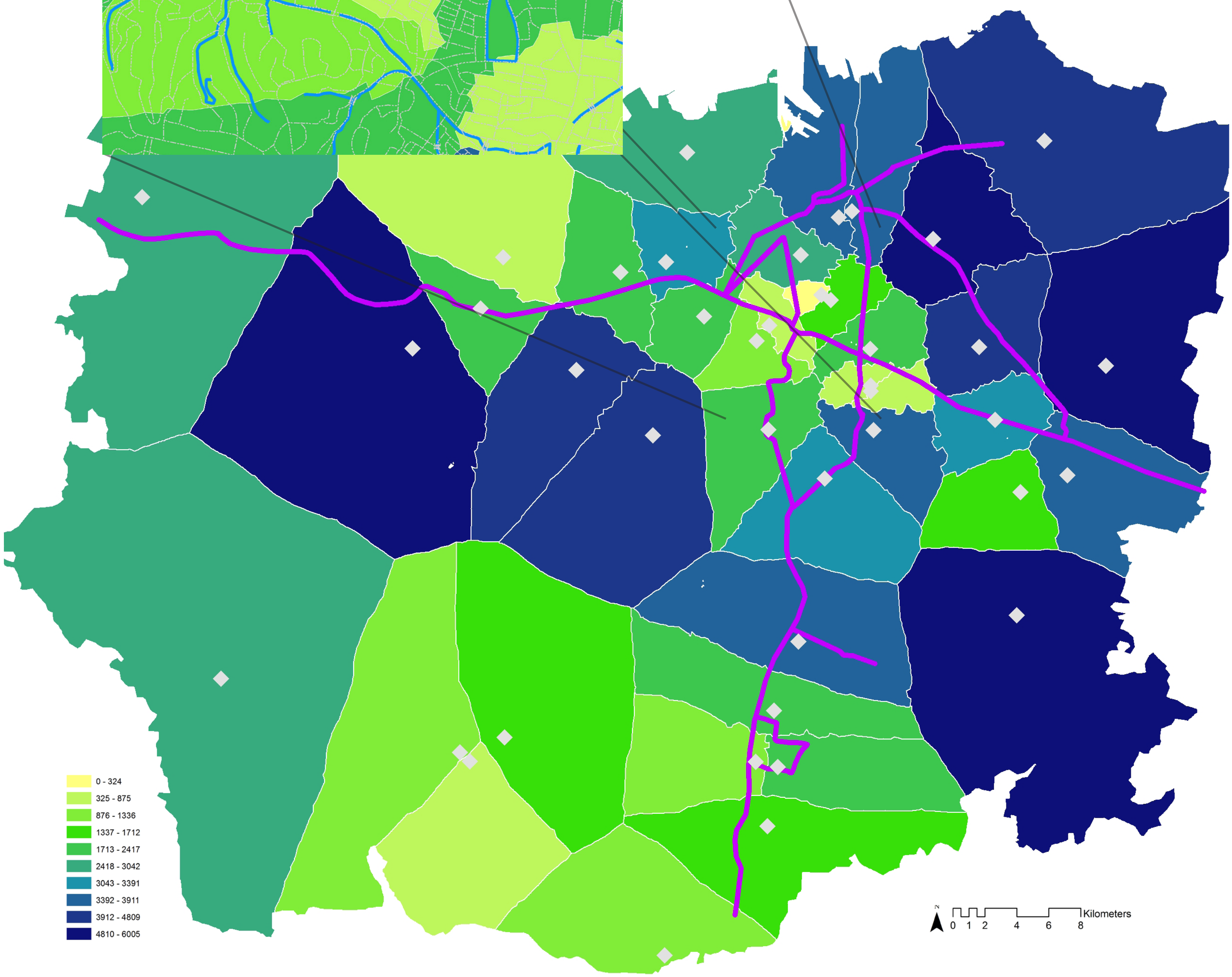


Figure 22: Public Middle Schools

The middle schools also demonstrate similar patterns of access, with the north east being less well serviced. The overall distribution is slightly less even than high schools, but more so than primary schools. The center of the city and the area around the University to the south is well serviced, but an outer ring around the center appears to have the least access.





Primary Schools and Vocational Schools

The primary school data appears to be managed through a separate data system. While the student body population data was not provided there is an accreditation designation used for the schools to indicate how well each school meets the accreditation standards. This is shown in the below map. Four of the six schools designated as underpeforming are clustered in the southern part of the city and the northeastern part of the city.

The population age groups for each respective subset of schools are reflective of the conditions in 2010. While there are many local schools in the central part of the city, the distribution is less in balance with the overall population, with areas on the edge of the city having much less access to a local primary school.

The vocational schools have much larger student bodies and serve the whole region. They are primarily attended by males, and exist as a fast-track academy for those pursuing a professional degree. They are centrally located, and, with one exception, located on the main BRT line for the city.

Figure 23: Primary Schools Service Area and Frequency Chart

NSS	Nama	Alamat	Kelur	A	B	C	PopInZone
101036307012	SDN Sendangmulyo 02	Jl. Klipang Raya No. 2	Sendangmulyo	-	1	-	1,626
101036311022	SDN Kemijen 04	Jl. Cilosari Dalam VI RT.	Kemijen	1	-	-	1,312
101036307011	SDN Sendangguwo 01	Jl. Sawi I	Sendangguwo	-	1	-	1,293
101030110006	SDN Siwalan	Jl. Karangingas	Siwalan	-	1	-	1,280
101036307003	SDN Meteseh	Jl. Bukit Kencana Jaya	Meteseh	-	1	-	1,117
101036315005	SDN Mangkang Wetan 02	Mangunharjo No. 200	Mangkang Weta	-	1	-	1,103
101036314017	SDN Tawang Mas 02	Jl. Tawang Rajekwesi II	Tawang Mas	-	1	-	1,012
101036302021	SDN Sekaran 02	Jl. Taman Siswa No. 33	Sekaran	-	1	-	926
101036312025	SDN Gabahan 01	Jl. Sebandaran No. 17	Gabahan	-	1	-	911
101036308003	SDN Kalicari 01	Jl. Supriyadi	Kalicari	1	-	-	908
101036308005	SDN Kalicari 02	Jl. Sendangsari Utara Ra	Kalicari	1	-	-	908
101036308011	SDN Kalicari 03	Jl. Supriyadi No. 1	Kalicari	-	1	-	908
101036309014	SDN Karangroto 04	Jl. Kudu Raya	Karangroto	-	1	-	871
101036313018	SDN Tanjungmas	Kebonharjo RT. 04 RW.	Tanjungmas	-	1	-	806
101036314055	SDN Kembangarum 03	Jl. Taman Borobudur Tir	Kembangarum	-	1	-	781
101036309006	SDN Muktiharjo Lor	Jl. Muktiharjo Raya No.!	Muktiharjo Lor	-	1	-	768
101036314018	SDN Kalibanteng Kulon 02	Jl. Lebdosari VIII / 1	Kalibanteng Kulo	1	-	-	751
101036310027	SDN Kaligawe	Jl. Tambakan	Kaligawe	-	1	-	748
101036306034	SDN Jatingaleh 04	Jl. Kesatrian No. 1	Jatingaleh	1	-	-	741

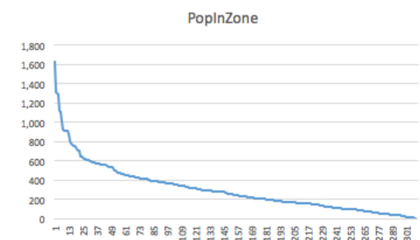


Figure 24: Distribution Of Primary Schools Based on School Quality Metric (1 to 3)

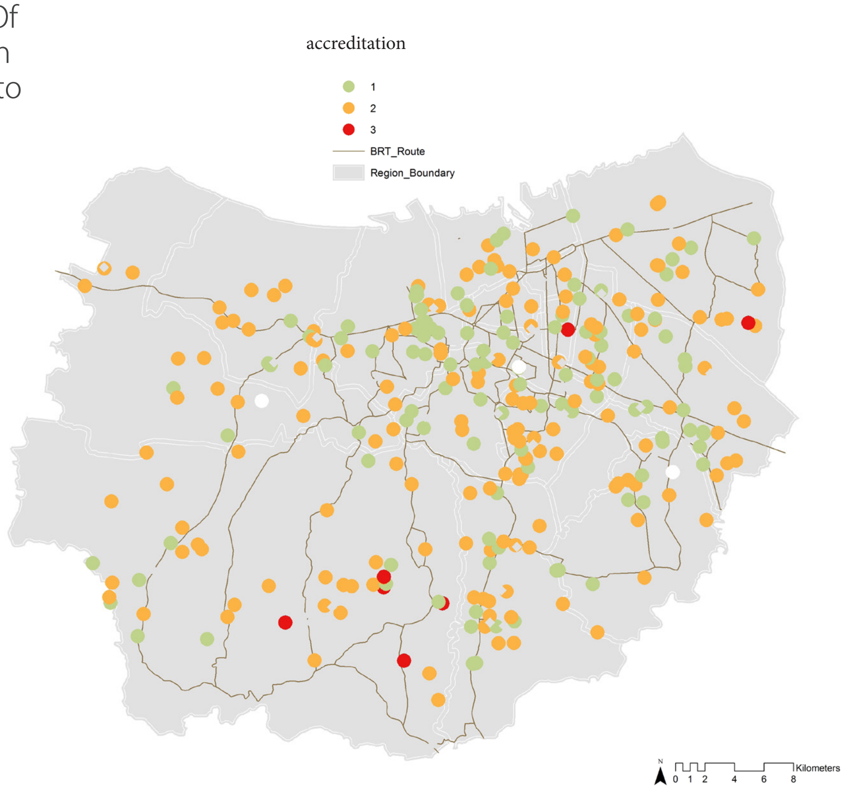


Figure 25: **Public Primary Schools**

Primary schools are distributed throughout the city. Most of the city has decent access to primary schools, with areas along the river and in the south west corner having more limited access, relative to their populations. The quality of schools really matters though, as the primary schools vary dramatically in quality, with many of them being madrasas, with the bulk of teaching revolving around religious doctrine. The primary school system is doubles as a community child care-taking system, and the credentials needed for teacher and school certification are most lax.

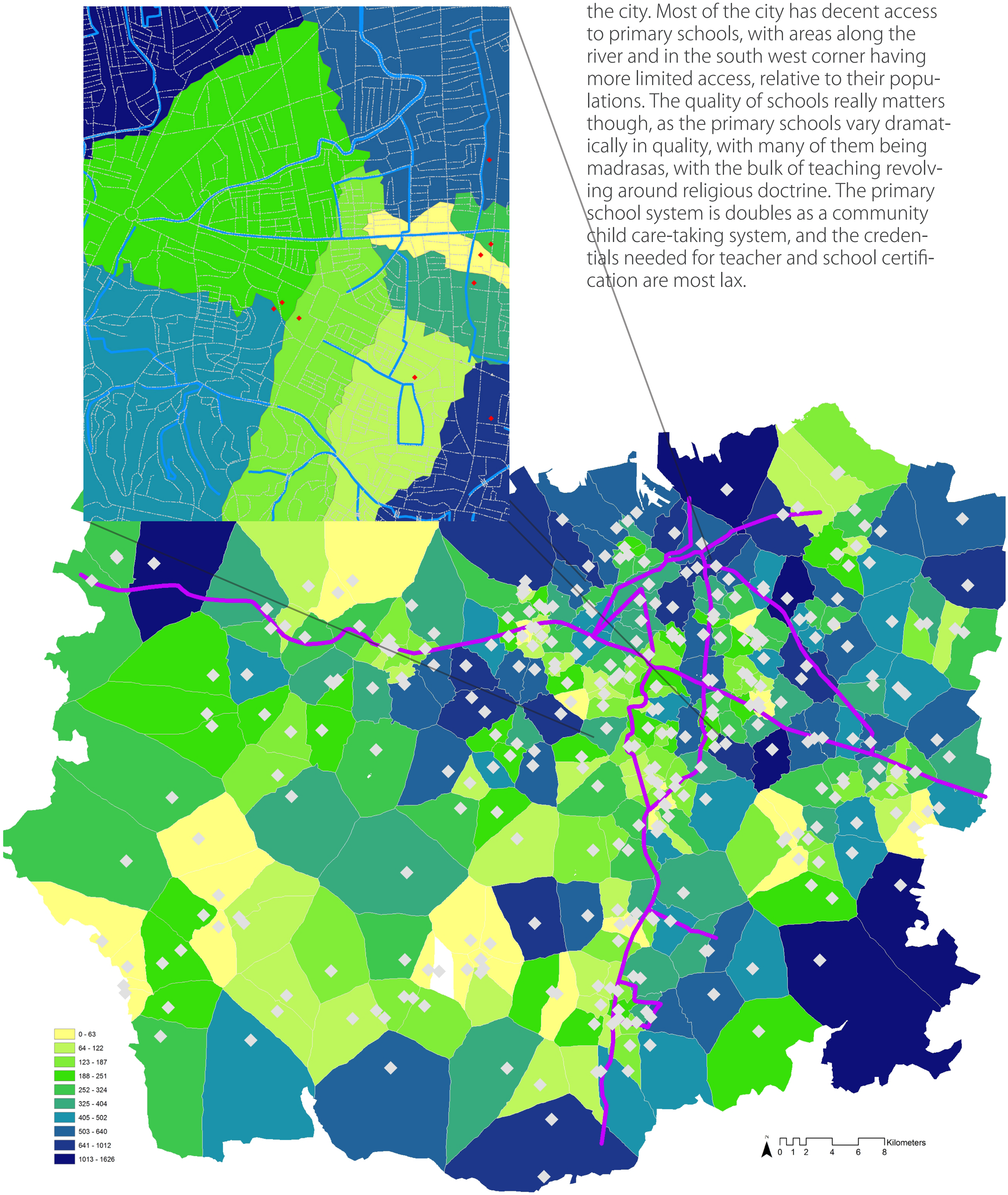
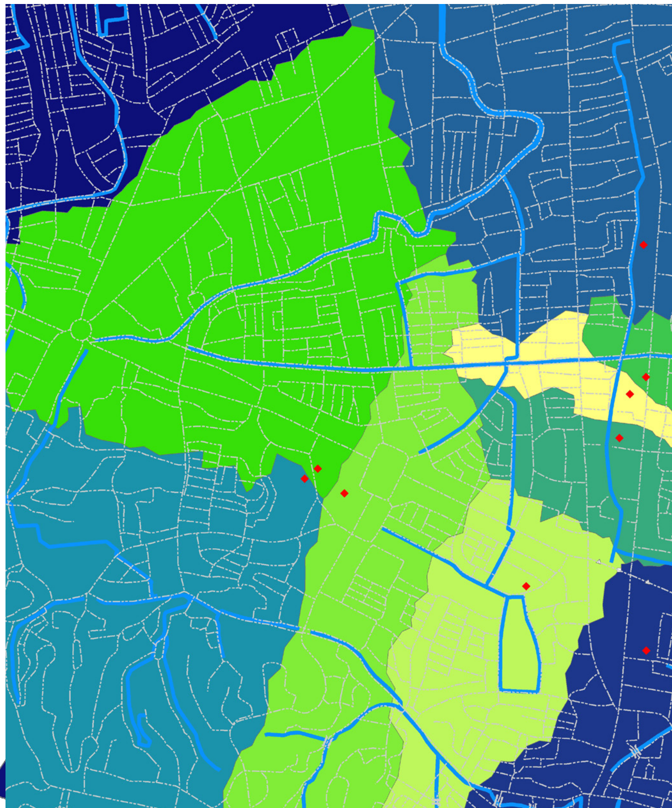
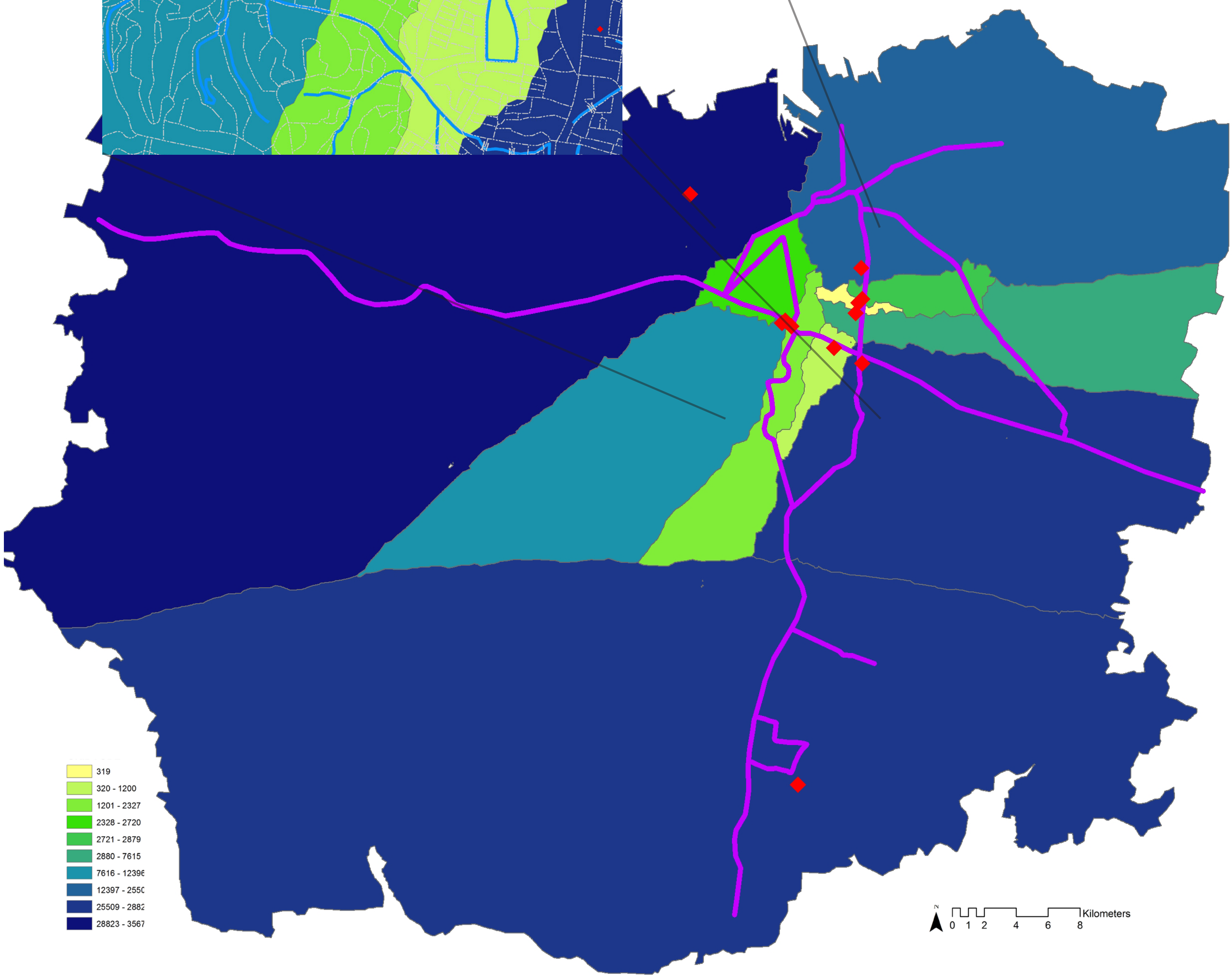


Figure 26: Public Vocational High Schools

The vocational schools are clearly clustered in the center of the city. They have a regional draw, and most commute to them via the bus transit lines, which are conveniently accessible to all but one. These schools were subsetted from the rest of the high schools as they have a different scope. They are primarily pre-professional programs. They are overwhelmingly attended by males.





Inferential Statistics

The distribution of access to schools roughly follows the distribution of the population, but there are some areas with less access than others. To test the significance of access to schools with regards to the actual incidence of low student attendance, 3 bivariate regressions were run. The level of access at each grade level was used as the independent variable for three regressions, using the individual cases of children who were not currently attending school as the dependent variable.

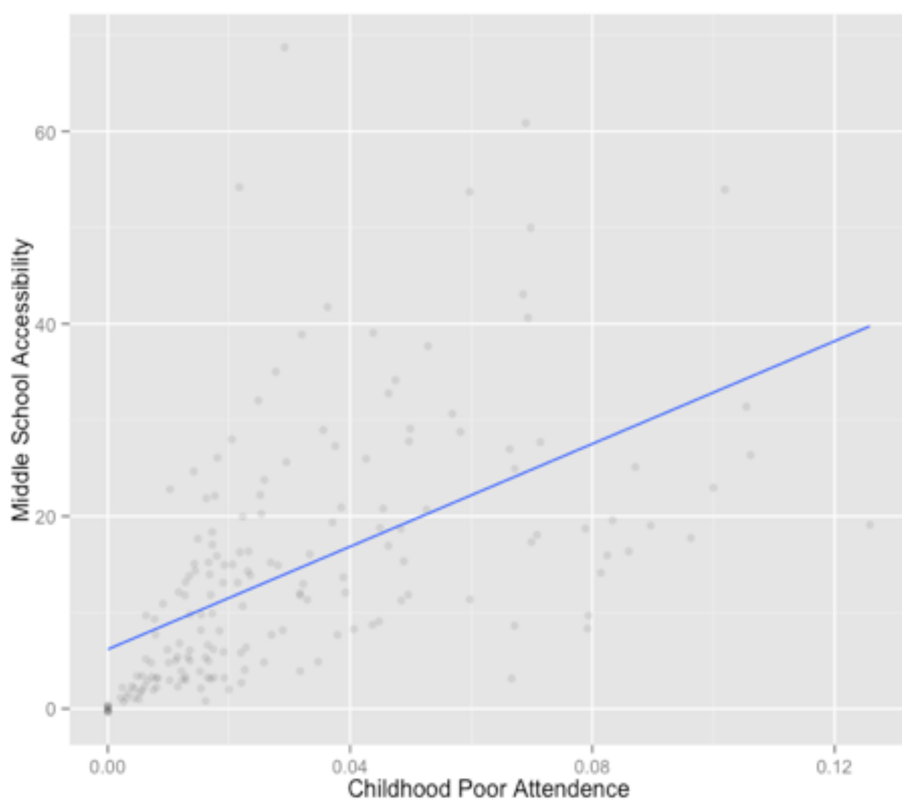
To conduct this analysis the metric created relating to student accessibility was turned into a raster, and the spatial average for each village was taken, using zonal statistics. This village average was then associated with each of the individual census records in the micro dataset. The logistic regression performed used the individual level microdataset as the unit of analysis. The regression provides one a sense of the significance of the relationship between the dependent and independent variable.

The regressions yielded significant results, as shown in Figure 27. Slightly less significance for the distribution by high schools makes sense as high schools are generally more regionally competitive and thus have a broader student base. While Primary schools on the other hand are well distributed, the quality of education by school varies dramatically, so their relative low significance might suggest that their overall presence is not as

Figure 27: Odds Ratios for Logistic Regressions

School Level	OR	X2.5..	X97.5..	Significance	
High Schools	0.999993	0.99998753	0.99999864	0.084	.
Middle Schools	0.999899	0.99986809	0.99993034	2.19E-10	***
Primary Schools	0.999476	0.99886803	1.00005779	0.0147	*

Figure 28: Odds Ratios for Logistic Regressions



meaningful, though more so than high schools. The burden by Middle schools appear to have a strongest level of significance. This makes sense as the Middle Schools are more local in focus but have a higher and more standard quality of education associated with them.

The incident using the micro data created bias in that areas that were closer had a larger number of students overall, as shown in Figure 29, which depicts the frequency of poor attendance observations by distance. Figure 30 visualizes the percentage of children who were not in attendance. This figure was aggregated by village and used in a bivariate OLS regression. As shown in Figure 28, there is a significant positive statistical relationship between accessibility for middle schools and the number of students who are not currently in school. Every increase in 267 meters is associated with an average poor attendance increase of 1%. There are likely many factors contributing to absences, which include that poor areas are farther away from school, and poorer families face more barriers to sending their kids to school, but the relationship is clear.

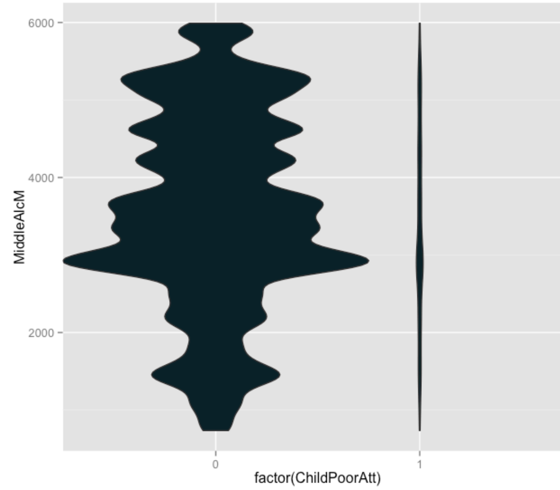


Figure 29: Violin Chart Comparing The Incidence of Poor Attendance

Deviance Residuals:

Min	1Q	Median	3Q	Max
-20.963	-6.177	-3.785	3.691	55.036

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.170	1.256	4.912	2.06e-06 ***
PercMidAccess	266.904	31.338	8.517	7.29e-15 ***

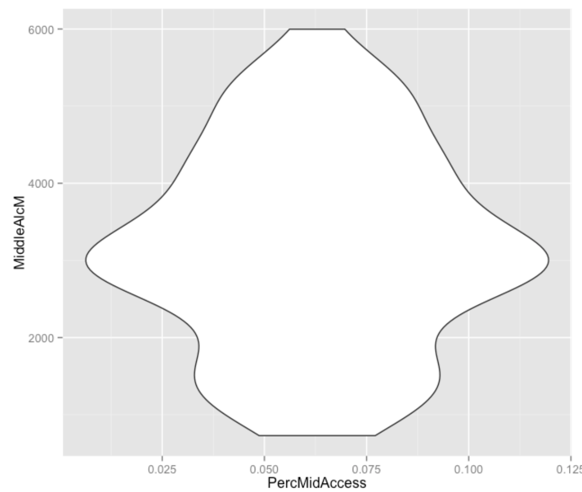


Figure 30: Violin Chart Comparing The Relative Incidence of Poor Attendance

Interestingly, the grand share of the children population exists at a distance from most schools, as shown in Figure 29. When considering the proportion of children who are not in school, the relative proportion, as shown in Figure 30, is also high in this middle distance area from schools, though the clear general overall trend is that children close to schools are more likely to be in attendance. Figure 31 maps the residuals, or the errors, of the bivariate regression. The areas in dark blue demonstrate areas where school attendance is far better than one would expect given the low levels of access. The city center's outer circle of census tracks demonstrate poorer attendance than one might expect, given their greater access to middle schools. This suggests that other factors are more determinant.

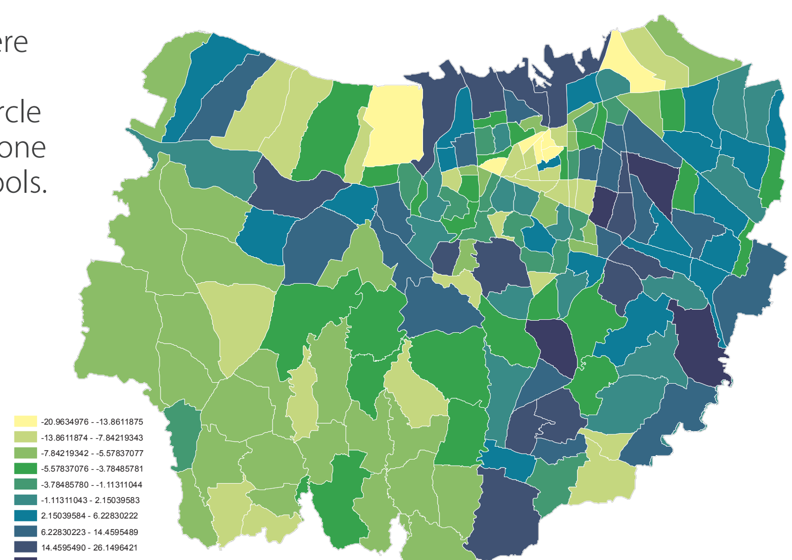


Figure 31: Prediction Error Modeling Access to Middle Schools vs School Attendance

Implications / Recommendations

Access is limited in the fastest growing segments of the city. While the northeastern quadrant of the city received the most population growth over the last decade, especially among children, and has the highest levels of poverty in the city, this area seems the least served by schools at every age level. The expanding eastward frontier of the city should be the focus for infrastructure investments in education and other types of infrastructure, as those wealthier areas to the south of the city's center are oversupplied.

Proximity to middle schools suggest a great improvement for educational outcomes. This suggests that proximity matters when the quality of education is more standard, and the city is oriented toward the local neighborhood. High schools having few available seats and a regional draw has meant that they are not as effectively serving their local neighborhood. There are clear trade-offs to having regionally competitive vs locally based schools. A placed-based focus for schools create more equal outcomes and provides a better means of making targeted and efficient investments to improve social well-fare and decrease inequality.

Broadly the accessibility maps provide the city with an indication of the relative concentration of resources. The current school catchment areas are very broadly defined. The segmentation of the school zones, cloro-plethed by the number of students provides a basis for more meaningful delineation of public school zones. An unsurprising finding of this analysis was that schools better serve the wealthier areas on the west to south periphery of the city.