# CLUSTER ANALYSIS AND SPATIAL PATTERN IN KCPE PERFORMANCE: THE CASE OF 2011 IN NAIROBI COUNTY

Charles G. Manyara,Ph.D
Department of Geospatial Science
College of Science & Technology
Radford University, Radford VA
cmanyara@radford.edu

***Abstract***

*In the paper "Spatial and gender inequality in the Kenya Certificate of Primary Education examination result" it has been argued that there is a spatial variation in the Kenya Certificate of Primary Education (KCPE) performance. In the case of urban areas, the premise is that performance deteriorates progressively from the city center to the periphery. This paper provides a counter to the finding that the urban design in Kenya cannot allow for that kind of spatial phenomenon. None of the more common urban design models - the concentric zone, sectors, multiple nuclei, urban realms, or core frame model – can support these finding. Besides, there is nothing unique at the Central Business District (CBD) that can explain improved performance. This study has adopted the city of Nairobi to challenge the methodology and findings in the previous study on Kisumu and Nairobi as being flawed. Using Geographic Information Systems (GIS), the same KCPE results from 2011 for Nairobi County were modeled for clustered analysis. The idea was to establish spatial autocorrelation within Nairobi County using data for public primary schools. Further, attention was paid to the trending in urban settlement and its implications in the spatial display of the KCPE results across the county. The findings indicate that performance is more related to the social economic undertones of the residential estates than the distance from the city. While GIS is a powerful tool for spatial analysis, it is very easy to make wrong conclusions especially if scale is not taken into account.*

**Key Words**: Exam Performance, Spatial autocorrelation, Cluster analysis, Urban structure, Nairobi

## Introduction

In the paper "Spatial and gender inequality in the Kenya certificate of primary education examination results" (Kimosop, Otiso & Ye, 2015) examined nationwide performance on the 2011 Kenya Certificate of Education (KCPE) exam by gender using Geographic Information Systems (GIS) and spatial statistics. They performed hot spot and cold spot analysis to identify concentrations of scores over space in terms of overall performance by gender, and individual subject scores. Among their findings was that …"proximity to major road networks and urban centers positively influences student performance on the KCPE exam probably because these factors influence the availability of educational resources (such as books and teachers), masterly of English (the language of instruction), teacher supervision, and child teach-ability" (Kimosop et al., 2015 pp.44). While this assertion is questionable in a number of ways, one of the more immediate concerns is the rationalization of distance decay based on an assumed availability of learning resources on the roadside and/or city center.

Kimosop et al. (2015) also indicate that their findings based on the buffer distances of 0-2, 2-6 and 6-10 km from the road and from the city center, shows that in fact the overall influence of urban areas on student performance is significantly higher than that of road networks. This argument is based on what they found to be drops in performance. Their results show for instance that the total male pass rate drops from 60% to 47.6% and to 46.5% as the buffer grows from 2 to 6 and to 10 km respectively. It is argued herein that a drop from 47.6% to 46.5% is not statistically significant to begin with. In the case of boys' performance – using Kisumu and Nairobi cities - they found that the respective percent drops from the 2 to the 6 and 10 km buffers are 12.4% and 13.5% respectively. Note that this is not a drop but an increase. The reasonable argument here is that there is no clear relationship between performance and distance from the Central Business District (CBD).

This paper uses cluster analysis and spatial autocorrelation to debunk the aforementioned argument and establish that while certain factors may favor the proximity to the CBD, the distance decay associated with overall performance is highly questionable.

The purpose of this study is to:

1) Find out the relations between exam performance and distance from the CBD in the city of Nairobi,

2) Assess the relationship between neighborhoods (by location) and overall exam performance

3) Determine whether English as an examinable subject and as a language of instruction has a strong correlation with other subjects in particular and exam performance in general.

The study uses the case of Nairobi County to explore performance over space. The idea is to use Geographic Information System (GIS) and run cluster analysis and spatial autocorrelation. The study is pegged on the hypothesis that there is not relationship between the distance from the CBD and exam performance. I argue that some form of concentric buffer zones will emerge from the KCPE exam results across space is to implicitly assume that the urban design in the city takes a ring form.

## Literature Review

### *Urban centers*

Urban planners define urban structure as the arrangement of land use. Among the classical urban structures are the concentric zone, sector and multi-nuclei models (Figure 1). The concentric ring model describes the city as an ecosystem in which residents sort themselves into a series of (five) rings based on class and occupation. This model was developed by Burgess in 1923 to explain the pattern of social areas in Chicago, Illinois (USA). It states that a city grows and expands outwards from its central business district (CBD) to form a series of concentric rings. The CBD becomes the core of the urban environment concentrated with certain retail and business activities and urban open space for parks, green spaces and in some cases industrial parks. The CBD is surrounded by a transitional zone, which contained abandoned factories and warehouses and high-density low income housing. This zone was usually targeted for urban-renewal programs. The third zone was usually dominated by working class families living in modest homes close to their place of work. The fourth zone was predominantly single-family residences interspersed with high-rent apartment complexes, and the fifth zone was a commuter region of low-density residential suburbs and greenbelts.

The sectorial model was proposed by Hoyt (1939). According to Hoyt, patterns of land use are conditioned by major routes that radiate from the city center. In Chicago, an upper class residential sector evolved outward along the Lake Michigan shoreline north of the CBD, while industry extended southward in sectors that followed railroad lines. Therefore the city develops in wedge-shaped sectors instead of rings: Certain areas of a city are more attractive for various activities, which flourish and expand outward in a wedge. The strength of the model is that people that settle in a city would settle in an area near transportation so that they can have easy access to many different places. The weakness with this model is that the theory is based on railroads and does not take into account cars. Besides, other physical features may divert the growth in some areas.

In 1945 Ullman and Harris proposed a multi-nuclei model. The model suggested that land uses develop around secrete nuclei or centers rather than around a single CBD, as suggested by the Burgess and Hoyt. The multiple centers evolve because of agglomerative tendencies and because certain activities repel one another. The multiple nuclei model assumes that car ownership and public transport granted people more mobility and led to the development of specialized regional centers within cities.

The concentric zone model general applicability has been challenged especially in Africa where most cities were designed as replicas of European cities of the colonizing powers. Many towns and nearly all large cities do not grow from around one CBD, but are formed by the progressive integration of a number of separate nuclei or nodes in the urban pattern. These nodes become specialized and differentiated in the growth process and are not located in relation to any distance attribute but are bound by a number of attributes (Boundless, 2015):

- Differential accessibility. - Some activities require specialized facilities such as port and rail terminals. For instance, the retailing sector demands maximum accessibility, which is often different from centrality offered in the CBD.

- Land use compatibility - Similar activities group together since proximity implies improved interactions. Service activities such as banks, insurance companies, shops and institutions are strongly interacting with each other. This can be defined as centripetal forces between activities.

- Land use incompatibility - Some activities are repelling each-other such as high quality residential and heavy industrial. This may be defined as centrifugal forces.

- Location suitability - Some activities cannot afford the rent of the optimal site for their location. They are thus locating at cheaper places, which are not optimal, but suitable for these activities.

It has also been argued that the irregular pattern multi-nuclei model as developed better suits the urban structure in many cities in the developing countries due to its often lack of planning or construction sporadic growth (ibid). The urban model that became the basis of the city plan in

Nairobi was a quasi-sector and multi-nuclei model.

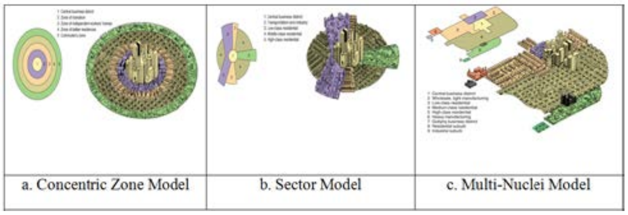


Figure 1. Classic urban structures

## The Origin and Growth of Nairobi

The origin of the settlement at this site dates back to the 19th century. In 1899, when the builders of the Mombasa-Kisumu/Uganda rail line decided to set up camp at "Ewaso Nai´beri"[1] a Maasai reference to the ‘river of cool waters’, this site was attractive because it was well supplied with water. In addition, the cool temperatures were a welcome relief from the hot Mombasa coastal sun the British had to contend with as they built the railway line. Besides, Ewaso Nai’beri offered an escape from the man-eaters of the Tsavo[2] as a swampland reduced the likelihood of an encounter with lions.

After the completion of the railway line, the British moved their administrative headquarters from the hot and humid town of Mombasa to the cooler, swampy town of Nairobi that ultimately became the capital of British East Africa (Kurtz, 1998). The design of the city in 1906 was very much zoning: a zone in the east for Africans’ use and another for Asian’s use. The rest of the city was all for the white Europeans settlers’ use (Figure 2).

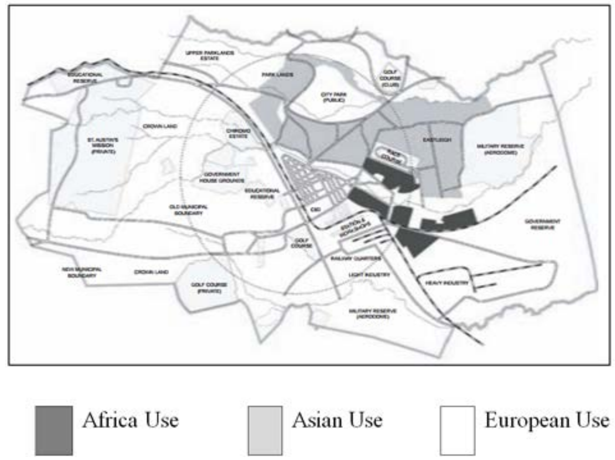


Figure 2. Nairobi in 1906

(Source: Vogel, 2008)

Nairobi was segregated on the basis of race, occupation and residence. Although Europeans never formed much more than ten percent of the Nairobi’s population, the colonial administration maintained tight control of most of the economic and administrative resources, while Asians worked as merchants and artisans, dominating the CBD (Figure 3). Africans, on the other hand, were marginalized and took up menial jobs (Aryeetey-Attoh, 2003). As a result, Nairobi was never considered a permanent home for many Africans (Mehretu, 1993). It can be said that the city was developed primarily for European settlement, consequently, became a replica of European cities. Kurtz (1998) argues that more than most cities, Nairobi offered a perfect opportunity for colonial authorities to experiment with urban planning. Nairobi was the result of two predominant and necessarily conflicting imperatives: on the one hand to create a model of the Garden City, a concept that was becoming important in British urban planning at the end of the nineteenth century; and on the other hand to create an essentially European city

[1] In Northern Kenya there is a major river watershed known as Ewaso Ng’iro (Nyiro) which literary means the river of brown or muddy water.
[2] During the construction of a bridge at Tsavo a year earlier, two very large male lions killed and ate more than 135 of the railway workers.

in the African setting, based on the South African model.

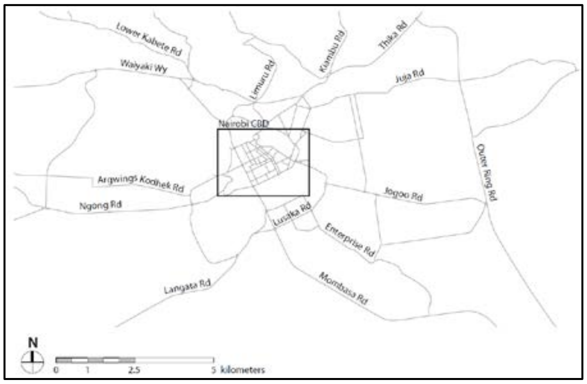


Figure 3. Nairobi major roads and the CBD.

(Source: Gonzales, Chavis, Li, and Daganzo, 2009)

In 1919, the Nairobi municipal community formally became Nairobi City Council. Its boundary was extended to include surrounding peri-urban settlements. The boundary line was revised again in 1927 to cover more ground (48 sq. km) mainly to the southeast – towards present day Machakos County. The official city limits have remained the same since.

Nairobi looks quite different from the colonial city it was. Apart from the very obvious difference in size, there are important structural and functional differences. Formal racial segregation, one of the fundamental principles of the colonial urban design, has been superseded by informal subdivision reflecting the class structure of society. Kenya elites have moved into higher-class neighborhoods where only whites had formerly lived. Class-based segregation also extends to middle and lower classes, who tend to occupy separate neighborhoods. Broadly, the early designation for community use persists; the east is still predominated by Africans and Asians. The colonial African reserves are more or less the present spaces for the low income. The wealthy are mainly associated with the west side of town, Westlands.

Nairobi has also changed substantial with regard to the range of public services and administrative functions. There are far more primary and secondary schools, universities and technical colleges than ever before. There are also far more state bureaucracies that direct and control so many aspects of political, economic and social development. The situation is made even more complex with the promulgation of the New Constitution in 2013. The ordinary citizens are learning to understand the structure and protocol in governance and service delivery.

Nairobi has also undergone immense urbanization. In 1906, the city had a population of 10,512. By 1960 the city population was 310,000. In 1963, when Kenya gained independence from Britain, the City of Nairobi had a population of 350,000. Nairobi continued to experience a high annual population growth rate passing ½ million mark at the 1969 census. The city population was 510,000 in 1970 (Republic of Kenya, 1970), 828,000 in 1979 (Republic of Kenya, 1981), 1,321,000 in 1989 (Republic of Kenya, 1994), 2,137,000 in 1999 (Republic of Kenya, 2000), and 3,138,369 in 2009 (Republic of Kenya, 2010). The projected population in the year 2020 will be almost six million (UN-HABITAT, 2005).

**Spatial variability and the role of English**

In recognition of the nature of urban structure, it is imperative to reassess the case made about the spatial variability of KCPE results in urban areas. In exploring KCPE performance, Kimosop *et al* (2015) resorted to analyze performance based on this assumed progression using buffers of 2 km, 6 km, and 10 km from the urban centroids, which coincided or not with the CBD. They then demonstrated unreliably that in big towns/cities such as Nairobi, Mombasa, Kisumu, Eldoret, and Nakuru, student exam performance varies widely with increasing distance from the city center even though the outer buffer (6-10 km) often falls within the city limits. This is because of the broad internal socioeconomic and school facility differences in the larger cities. The latter is an overly spurious statement from the practical point of view and from the demonstrated results using the same data. In the case of Nairobi, Kimosop *et al* (2015) using the 2011 KCPE results found that the total male pass rate drops from 69.8% to 49.2% and to 47.3%. Similarly the total female pass rate drops from 70.6% to 46.3% and to 40% in the respective buffers; 2 km to 6 km and to 10 km. In a nutshell, this was a fallacy premised on an urban structure that in the first place does not exist.

It becomes even murkier when the findings hinges in English, Kiswahili, and SSR with distance from the CBD and little effect on mathematics and science because "…..the narrow socioeconomic and educational facility differences between urban fringes and rural areas at those distances" (pp.58). That is a speculation, a hunch and by all means not a scientific explanation. To put the argument in geostatistical perspective, Kimosop *et al* used Ordinary Least Squares (OLS)[3] regression analysis to test the significance of the distance from urban centers on student performance. It is therefore in order to explore the relevancy of OLS in the context of data like the KCPE results.

## Ordinary Least Squares (OLS)

The OLS regression attempts to minimize the sum of the squared differences between the values predicted by the model and the values actually observed in the training data. Theoretically the OLS technique uses sample data to estimate the true population relationship between two variables, $\mathbf{X_i}$ **t**he dependent variable(s) and Y the independent variable:

$$\mathbf{Y_{\hat{i}} = b_o + b_1 X_i}$$

OLS allows us to find = $\mathbf{b_o}$ **and** $\mathbf{b_1}$.

The OLS regression is trying to find the constant coefficient(s) **b**1 (and of course, **b**2, **b**3, …, **b**n in the case of more variables) to minimize the quantity, $\mathbf{Y_{\hat{i}}}$ .

In Kimosop *et al* (2015) case the two variables were mean scores of each school ($\mathbf{Y_{\hat{i}}}$) and the distance ($\mathbf{X_i}$) to the centroid of the urban center[4]. The authors found that the signs of the exam performance coefficients ($\mathbf{b_1}$) with distance from the urban center were primarily negative. This meant that the larger the distance from an urban center centroid the lower the 2011 KCPE exam score, except for mathematics whose probability value ($p < 0.05$) was not statistically significant. What this means, implicitly is that the unknown quotient, $\mathbf{b_o}$, and may be a strong factor than a marginal unexplained error. Once again, maintaining that there is a linear relationship between exam scores and distance from the CBD.

There are a number of issues with the application of this model which is, unfortunately, frequently misused and misunderstood. First, without going into the statistical details of the model, the fact that Nairobi urban structure is not concentric their

[3] OLS regression is also known as "Least Squared Errors Regression", "Ordinary Least Squares", or often just "Least Squares". It is one of the most basic and most commonly used predictions used in many fields.

[4] They ran a regression the mean scores against distances from the closest road network.

model is assuming that the progressive distance from the CBD is on an isotropic terrain with uniform environs and amenities. Further, it assumes that, *ceteris paribus*, schools at each distance are the same in all forms that can catapult all students to perform the same. The second big issue is the actual known flaws of the model as used in this context. To appreciate this first let us set the stage for what the OLS is all about.

The OLS regression method produces a line that minimizes the sum of the squared residual ($e_{i(hat)}$) from the line to the observed data points. That is it minimizes:

$\Sigma\, \hat{e}_i^2 = \hat{e}_1^2 + \hat{e}_2^2 + \hat{e}_3^2 + \ldots\ldots\ldots + \hat{e}_n^2$, where n is the sample size.

The sum of the residuals,$\Sigma e_i^2$, is exactly zero. The OLS is also called the residual sum of squares (SSE). Unlike the typical linear regression, this is the amount OLS minimizes. On this basis, we can assume that Kimosop *et al.* picked this for measuring "accuracy" i.e. the sum of squared errors for deriving better predictive stats in the regression. So, the question then is why did *these authors* believe that linear regression (as opposed to non-linear regression) is the best choice of regression for predicting exam performance across the landscape? In other words, if the underlying system we they were modeling is linear with additive noise, and the random variables representing the errors made by the OLS model are uncorrelated from each other, and if the distributions of these random variables all have the same variance and a mean of zero, then the OLS method is the best unbiased linear estimator of the model coefficients (though not necessarily the best biased estimator) in that the coefficients it leads to have the smallest variance.

The problem of selecting the wrong independent variables (i.e. features) for a prediction problem is one that plagues all regression methods, not just OLS regression. The choice of distance from the CBD in this case illustrates that fact. Clearly, using distance to predict the mean score is essentially impossible because there is very little relationship between the independent variables and the dependent variable. It is one of those situations where causality may not be established but instead the regression algorithm discovers a spurious relationship between the independent variable and dependent variable that only happen to be there due to chance (i.e. random fluctuation).

The reality of the data is that in an urban landscape you are dealing with a serious problem of outliers in the data. In such a case the OLS can perform very badly when some values in the sample data have excessively large or small values for the dependent variable compared to the rest of the data. The reason for this is that since the OLS method is concerned with minimizing the sum of the squared error, any point that has a dependent value that differs a lot from the rest of the data will have a disproportionately large effect on the resulting constants that are being solved for. In Nairobi this is the case where within the same distance from the city center, an African public primary school on one side of the road performs contrastingly different from the Asian school across the street. Similarly there are stark differences between a school in Mathare slum and a school in Parklands both being the same distance from city center.

In reality, most systems such as exam performance and location or spatial distance relationship are not linear. By attempting to fit a line through one dimensional data set, a plane through two dimensional data sets, or a generalization of a plane (i.e. a hyper plane) through higher dimensional data sets can be misleading. In practice though, real world relationships tend to be more complicated than simple lines or planes, meaning that even with an infinite number of training points (and hence perfect information about what the optimal choice of plane is) linear methods will often fail to do a good job at making predictions.

**English Language**

The issue of English as an exam subject pervaded Kimosop *et al* paper. The one argument which is common place is the relationship between exam success and the masterly of English language. English is both an exam subject and the language of instruction for all subjects except other languages such as Kiswahili in the case of KCPE. Kimosop *et al* argued that students living in close proximity to urban areas perform better in exams than those in rural areas because of their higher English language proficiency. Quoting Lockheed and Verspoor (1991, p. 153) their study argued that;

> "Children who speak a language other than the language of instruction confront a substantial barrier to learning. In the crucial, early grades when children are trying to acquire basic literacy as well as adjust to the demands of the school setting, not speaking the language of instruction can make the difference between succeeding and failing in school, between remaining in school and dropping out."

In rural areas, English is often the third language; the local native tongue (vernacular) and Kiswahili come first. In urban areas English comes in second or first depending on family status and in part because ethnic diversity of the country's urban environments forces students to communicate regularly in English and Kiswahili. Kimosop *et al* argued that children from urban and rural middle-class families tend to do well in school than their counterparts from poor families because they are more exposed to English and have good nutrition and access to electricity and other infrastructural amenities.

Given, students who are fluent in English language tend to perform better in national exams (Hungi &Thuku, 2010; Glewwe, Kremer, & Moulin, 2009). However, there is also evidence that there exist a difference between communication in English and answering exam questions (*ibid*). In urban areas and in deed across the country there is concern that even the spoken English is increasingly weakened as an exam subject and in answering exam questions by the prevailing use of sheng. Eunice Mlati, the Principal of Moi Avenue Primary School quoted in Dean (2013), sees Sheng as just another obstacle to teaching the next generation of Kenyans. According to Mlati Sheng interferes with performance of students in languages, both English and Swahili. "…Sheng comes in, and test scores go down." It is therefore not all roses in the cities. Nevertheless, urban areas tend to have more of everything; teachers, better supervision, better learning facilities, and more informed and, perhaps, engaged parents than their rural counterparts.

While there may be a concentration of wealthy and highly educated people in the city of Nairobi[5] there are a lot of people who reside in Nairobi who are poor and semi-illiterate. According to the 2009 census (Republic of Kenya, 2009), 50.6% of the adult population in Nairobi have some primary education. Another 18.1% have some secondary education. This

[5] Nairobi was ranked 1st

means that less than 30% have higher or specialized training. It is estimated for instance that 60% of Nairobi's population live in slums (Syagga, Mitullah, & Gitau, 2001). Contrary to Kimosop *et al* blanket assertion, it is only fair to say that some and not all students may be privileged because of their exposure to English earlier in their lives via television, more educated parents, greater access to English reading materials, and also through peers.

**Study Area**

The City of Nairobi (County) is the capital of Kenya. The city covers an estimated 696 sq. km and lies at an average altitude of 1700 meters above sea level. The rapid urbanization in Nairobi has had its consequences including shortage of affordable housing. The scarcity and high cost of housing has forced many poor families to occupy insufficient often poor maintained tenements or congested one or two bedroom units. Others have responded by building their own make shift squatter dwellings. About 60 percent of Nairobi residents live in informal settlements – either slums or squatter settlements (*Ibid*) (Figure 4).

The rapid growth of the city has put immense pressure on the infrastructures; schools, water and sewage system, and transportation. Inadequacy in some of these services can be potentially affecting the hundreds of schools and student performance in national exams. The education sector has experienced a massive explosion of student enrolment in primary schools since the beginning of the Free Primary Education (FPE) program. According to the Ministry of Education's Management Information System, there has been an increase of public and primary schools from 6058 in 1963 to more than 30,000 in 2010 across the country. Between 2000 and 2013 overall national primary school enrollment increased from 6,175,600 to 10,187,500 (Republic of Kenya, 2014). In 2011 the national (primary) school completion was at 74.6% - girls edged boys slightly, 75.2% to 74.0% (*ibid*).


Figure 4. Location of Slums in Nairobi

**Methodology**

This study adopts the 2011 KSCPE data set that was used by Kimosop et al. (2015*).* Nairobi's public primary schools locations were geo-referenced in ArcGIS® software. This data was coupled with Nairobi GIS boundary maps.

The KCPE exam scores were aggregated having originally been obtained from the interactive online Kenya Open Data Initiative (KODI) clearinghouse portal at https://opendata.go.ke/. Unlike Kimosop *et al* who used individual student performance, this study used schools mean subject and overall scores. The rationale for aggregating the score was informed by the objectives of the study. Because the overriding investigation was to determine whether distance and location are factors in performance, the average school score should be representative enough to overcome data redundancy. The boundary base-maps were obtained from miscellaneous online sources.

A total of 190 public primary schools within the City of Nairobi County were used with a total enrollment of 21784, averaging 115 students per school.

**Data Analysis:**

Just like Kimosop et al (2015) study this study did most of the spatial analysis with the ArcGIS® tools. The Kimosop *et al* paper used concentric buffering to demonstrate that KCPE performance progressively got worse with distance from the CBD (Figure 5). The result clearly is more than perfect.

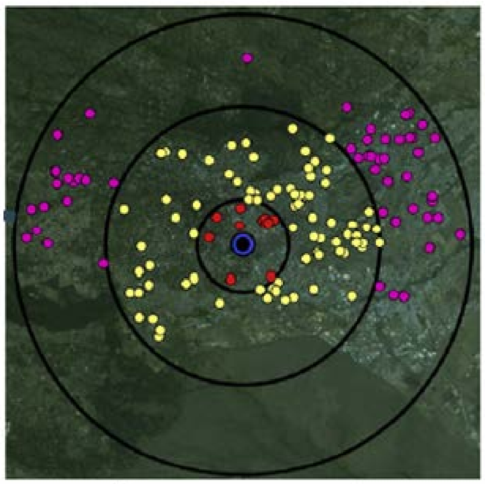

Figure 5. Nairobi, buffers 0-2,2-6 and 6-10

Figure 5 is a reproduction of the Kimosop *et al* figure 12 (p.58). The graph shows that performance change with buffer distances of 2, 6 and 10 Km from the CBD. According to their graph on intra-urban differences, performance decreased with increasing in distance from the city center. For both boys and girls there was significant influence of distance on performance (Figure 6).

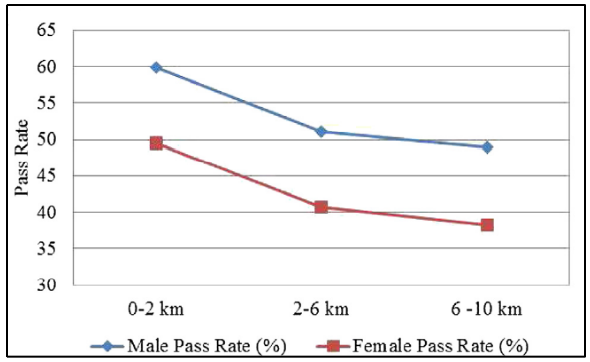


Figure 6. Effect of urban distance on student performance,Nairobi.
( Kimosop *et al* (Fig.13, p.59).

This study used the ArcGIS Calculate Geometry tool to deliver the centroid of the city limits polygon using a center of gravity-based algorithm. The center of gravity (center of mass) calculation uses the geometric center model to output the new point feature class. In this case for those familiar with Nairobi, the center falls in or around Mbotela Estate. This centroid unlike city center or Nairobi GPO[6] was preferred because that gave a better sense of symmetric distances for the spatial spread of the city. Based on this centroid, buffers were generated at 2km, 6km and 10km distance (Figure 7).

A breakdown of the school performance into five groups was overlaid by buffers zones at 2, 6 and 10 kilometers. A visual inspection of the performance based on schools' mean scores does not generate the impression created by Kimosop *et al* as presented in Figure 5. What is evident is a tendency towards the west for the high performing schools in the >252 range.

Kimosop *et al* used a score of ≥250 marks out of 500 points as the pass mark. When those

[6] GPO was the General Post Office located at the junction of Koinange Street and Kenyatta Avenue.

schools with a mean score of ≥250 points are mapped the result is not conclusive as these schools do not conglomerate in the inner core of the city (Figure 8). This is an overlay of the same scores but this time it shows that using cluster analysis, there are two hot spots within the city limits: one in Westlands and another is in the Eastlands. This further proves that the city center to the periphery progressive declining in mean scores is once again disapproved.

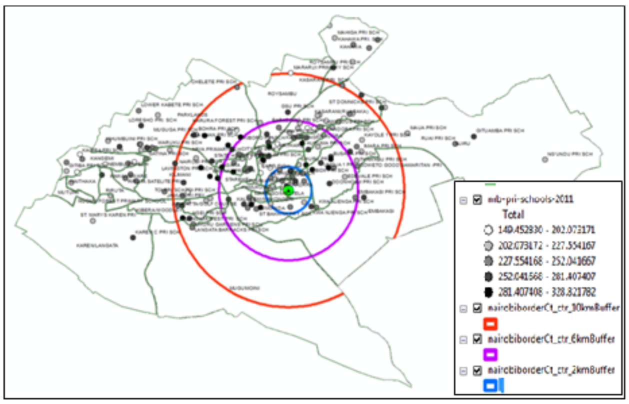

Figure 7. Performance by zones (2, 6, 10 km)

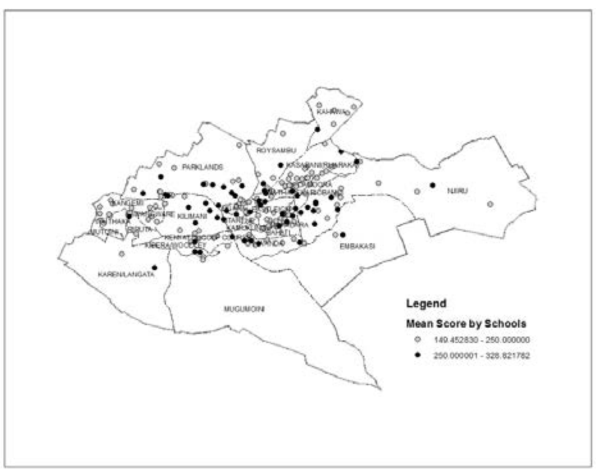


Figure 8. The 250 and higher points score among schools in Nairobi.

**Exploring the Data**

The analysis so far has been based on a 250 points pass-mark. In calculated mean score among Nairobi Public Primary Schools in the 2011 KCPE scores is in fact 237.09 points (Figure 9). The breakdown shows that the 1st quartile is at 214.73, while the 3rd quartile is at 255.67 points.

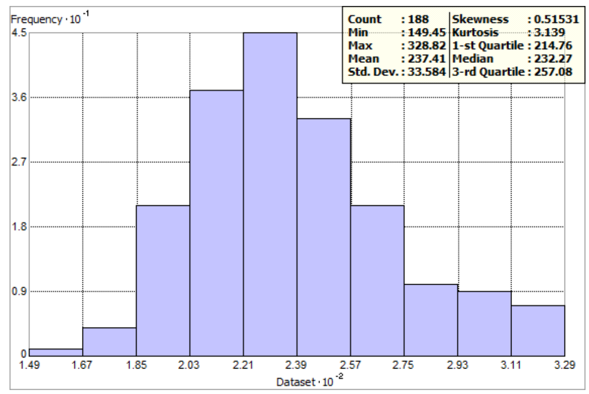


Figure 9. KCPE Histogram

The histogram indicates that the data is unimodal with a slight skew to the right (3.139). The right skew is an indication of more low scores and fewer high exam scores. In fact the highest mean overall score for the city was 329 points. While the distribution is not perfectly normal, but certainly not overly skewed.

A quantile-quantile (QQ) plot to compare the distribution of the data to a standard normal distribution provides a means to measure the normality of the data (Figure 10).

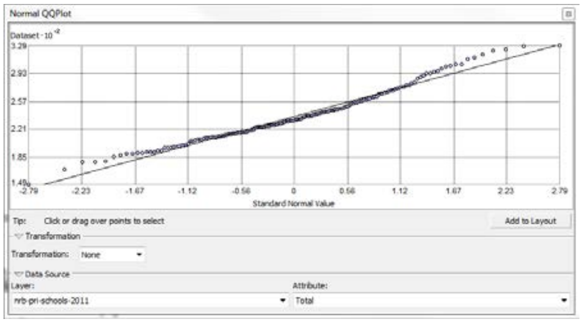


Figure 10. Normal Distribution QQ plot.

The closer the points are to the straight (45 degree) line in the graph, the closer the data follows a normal distribution. The plot is therefore not very close to a straight line. The main departures from this line occur at low values of mean scores and in the high end. Given that the data does not exhibit a normal distribution in the histogram and the QQ plot, it may be necessary to transform the data to make it conform to a normal distribution before using certain kriging interpolation techniques.

*Global Trends in the Data*

The historical characteristic of Nairobi is such that there is an implicit divide of East and West. Even though that is now less clear than in the past, it is reasonable to argue that if family income and lifestyle have an influence on student performance in national exams, there exists a trend in the mean exam scores across the landscape.

If a trend exists in the data, it is nonrandom (deterministic) component of the surface can be represented by a mathematical formula. The trend analysis tools in ArcGIS is useful in identifying the presence/ absence of trends in the input dataset and identify which order of polynomial fits the best trend best.

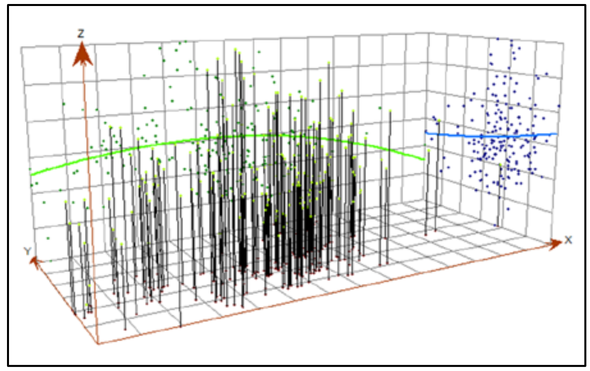


Figure 11. Trend Analysis.

Each of the sticks in the trend analysis plot represents the location and value (Height) of each school's mean score. With the data projected onto the perpendicular planes, east-west and north-south plane, a best-fit (polynomial) was drawn through the projected points, showing trends in specific directions. A flat line (green) indicates that there is no trend while a line that starts with a low values, increases as it moves towards the center of the x-axis, then decreases as show in Figure 11 implies a trend. Similarly the blue line is increasing as it moves north. The data therefore exhibits a west-east trend with a mild strength at some point in the center (green line) while the north trend is ambiguous at least (blue line) with an increasing trend in value northward.

*Spatial autocorrelation and Directional Influence*

The working assumption in the original question was that schools that are close to one another are more alike. To test this assumption a semivariogram/covariance cloud was generated to examine this relationship.

A semivariogram value is the difference squared between the values of each pair of locations, plotted on the y-axis and the distances separating each pair of measurements, plotted on the x-axis (Figure 12).

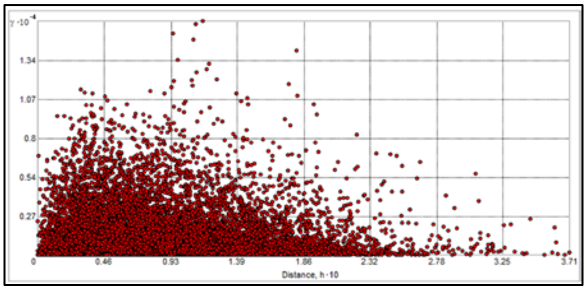


Figure 12. Semivariogram/Covariance Cloud.

Each dot in the semivariogram/covariance cloud represents a pair (bin) of locations. Since locations that are close to each other should be more alike. In the semivariogram plot, the locations that are closest (on the far left on the x-axis) should have small semivariogram values (low values on the y-axis). As the distance between the pairs of locations increases (moving right on the x-axis), the semivariogram values should also increase (move up on the y-axis). However, a certain distance is reached where the cloud flattens out, indicating that the values of the pairs of points separated by more than this distance are no longer correlated.

The plot shows that the distribution of these paired points is positively skewed as more pairs at close distance (near zero on the x-axis) are dissimilar and few have a higher semivariogram value (high on the y-axis) to the right. This is further proof that spatial autocorrelation is weak in this data.

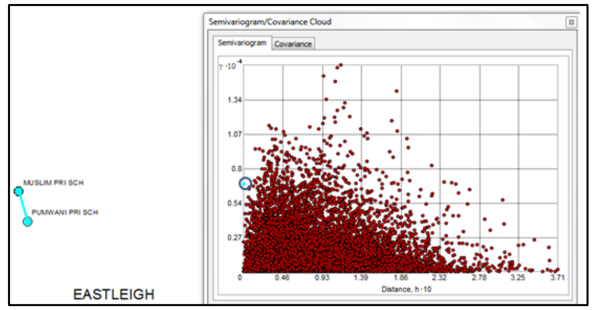


Figure 13. Contrasting Performance and almost zero distance

To illustrate this fact, two schools a few blocks from each other were selected; Pumwani and Muslim Primary Schools from Eastleigh (Figure 13). Pumwani Primary had a mean score of 178.18 points while Muslim Primary School had a mean score of 294.77 points.

*Spatial autocorrelation –surface Interpolation*

The descriptive statistics and general trend analysis show that the data has no outlier (or erroneous) sample points and that there is a trend and an anisotropy. Using the kriging tool in ArcGIS, a geo-statistical interpolation surface was generated to create a variography to create a more accurate surface.

The semivariogram (Figure 14) depicts the spatial autocorrelation of the measured sample points. Once each pair of locations is plotted (i.e. binning the empirical semivariogram), a model is fit through them.

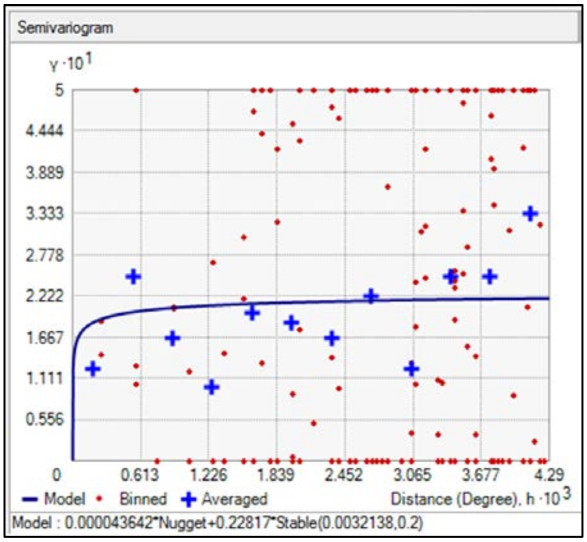


Figure 14. Semivariogram/Covariance Model

The model shows a very short range. The implication is that spatial autocorrelations is almost non-existent. Clearly the points are spatially uncorrelated. This is confirmed by the fact that there is a very small partial sill (sill minus the nugget[7]). Theoretically, whenever lag = 0 therefore the semivariogram value is 0. However, at an infinitesimally small lag, the semivariogram often exhibits a nugget effect, which is some value greater than 0. In this case the partial sill is extremely low. That is the case in this data. School locations are human choices and have very little to do with natural occurrence, the basis of kriging as originally used in mining.

It is therefore understandable that there are lots of schools that are very similar at all distances – note dots along the x-axis (y=0) – while at the same time there are also a lot of schools very different at all distances (y = 5)( Figure 14). Then there are several bins scattered across the plot. The semivariogram as plotted is proof that there is no relationship between school performance and distance as it is evident. It is therefore erroneous to state that schools that are closer together are more alike than schools that are further apart.

*Cross Validation*

To cross-validate the spatial variation a predictive model for values at unknown locations was ran. The kriging cross-validation model graph generated an ideal predictive model (black line). This model takes into consideration the global trends in the data and adjusted for the local directional influence (anisotropy) in the semivariogram.

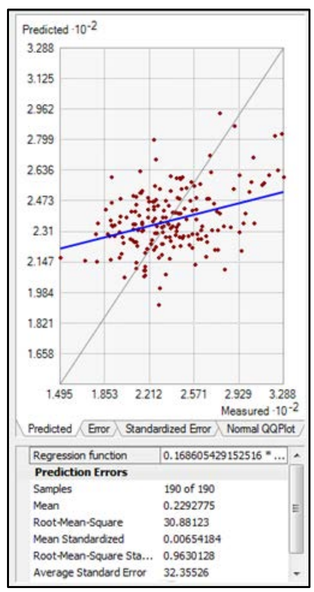


Figure 15. Cross-Validation, Kriging

The data provides the trend that is estimated by the blue line, gentler slope. Clearly the

[7] In naturally occurring data the nugget effect can be attributed to measurement errors or spatial sources of variation at distances smaller than the sampling interval or both. Measurement error occurs because of the error inherent in measuring devices. Natural phenomena can vary spatially over a range of scales. Variation at micro-scales smaller than the sampling distances will appear as part of the nugget effect. Before collecting data, it is important to gain some understanding of the scales of spatial variation. For fixed data as in the case of schools, the effect is human influenced: these are not naturally events but human constructs.

predictive intercept is higher and the trend (slope or constant) is low (Figure 15). The point distribution once again prove that the tendency is towards less relationship between the predicted and the measured

**Cluster Analysis and Probability Mapping**

The cross-validation predictive model had shown high standard error values in the data set. To make a meaningful estimation of values across the study area it was necessary first to point out neighborhoods of high performance and second, generate predictive zones that meet the threshold values beyond which any prediction may be spurious.

An ordinary kriging was used to generate hotspots and cold spots across Nairobi (Figure 16). The map shows that's there are two hot spots in Nairobi, one in Westlands in the vicinity of Kileleshwa, State House, Nairobi, Baraka and Loreto Convent Primary Schools. The second but smaller hotspot is in the east around Umoja and Bidii Primary Schools.

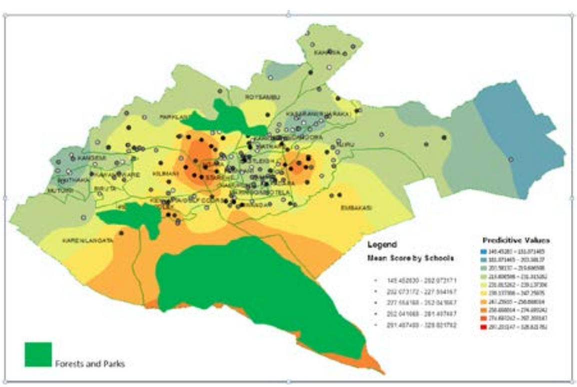


Figure 16. Predictive Values.

There are schools in cold spots (blue area) that actually perform as well as the hotspot schools. Schools like Ruthimitu and Dagoretti Special Schools in Kangemi area (> 220 points score) are doing as well as schools in the vicinity of the hotspots. So is Githuamba Primary in Ruai (Embakasi) that has a mean score of 255.196 points.

To quantify the predicting standard errors for each location the data was ran with the trend removed. The rule of thumb is that 95% of the time , the true zone value will be inside the interval formed by the predicted value ±2 times the prediction standard error, assuming that the data is normally distributed. Notice (Figure 17) that the prediction standard error for locations near sample points tends to be low.

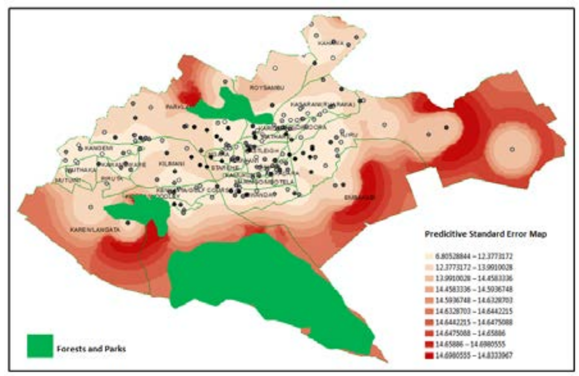


Figure 17. Prediction Standard Error Map

Taking into consideration the global trends in the data and adjusting for anisotropy in the semivariogram, the model is cable of recognizing high scores in areas otherwise classified as cold spots.

**The place of English in KCPE Examination Performance.**

The case was made that proficiency in English is major advantage to candidates because English is both an examinable subject and the language of instruction. Therefore a strong correlation with other subjects –except Kiswahili is expected. In particular for urban children were native languages influence is minimal, are assume to do better in exams based on this postulation. Exceptions were stated such as the use of less constructive use of English in poor neighborhoods such as slums, and the Sub-urban areas of the city where it was argued that English proficiency diminish with distance and therefore exam performance.

As expected a plot of total (mean) score v subjects produced more or less a linear positive relationship (Figure 18).

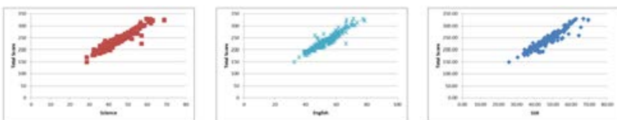

Figure 18. Correlations between total (mean) scores and subject scores

A Pearson's correlation coefficient was ran to test for the strength of the correlation between exam subjects tested with special attention to the correlation between and among KCPE subjects: English (Eng), Kiswahili (Kis), Mathematics (Mat), Science (Sci), and Social Studies and Religion (SSR[8],).

A correlation matric to measures the strength and direction of a linear relationship between subjects is shown in the matrix below (Table 1).

**Table 1. Correlation (*r*) Matrix for KCPE subject scores.**

| | Eng | Kis | Mat | Sci | SSR | Total |
|---|---|---|---|---|---|---|
| **Eng** | | | | | | |
| **Kis** | 0.829309 | | | | | |
| **Mat** | 0.871633 | 0.778706 | | | | |
| **Sci** | 0.872659 | 0.813331 | 0.87537 | | | |
| **SSR** | 0.700308 | 0.699305 | 0.75452 | 0.78413 | | |
| **Total** | 0.935014 | 0.891844 | 0.93478 | 0.94966 | 0.865592 | |

[8] Social Studies and Religion (SSR) includes a bit of Kenyan History, Civil education, current County system of government as well as all the Religious Studies.

The ***r*** across the matrix ranges from 0.699305 (Kis and SSR) to 0.87537 (Sci and Math). However, English explained over 76% in the performance in Math and Science.

Students who did well in science and English and Math in that order are more likely to pass exams: their $r^2$ is 0.901, 0.874 and 0.873 respectively.

From the visual impression of the mapping of schools with a mean score of <53.782676% (Figure 19). It will appear that majority are in Eastlands. This finding support the premise that Eastlands that tend to have more poor neighborhoods are tend to do poorly in English and potentially poorly in the exams.

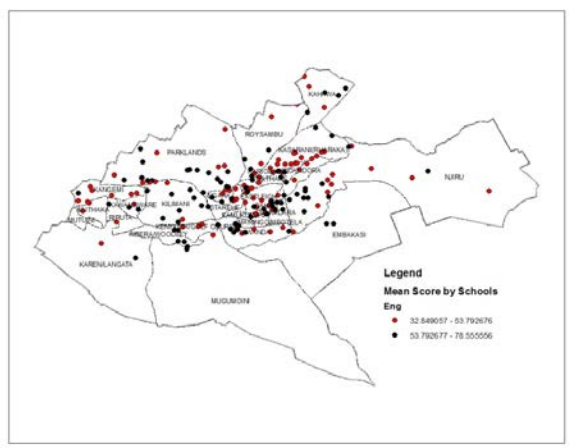


Figure 19.English mean Pass-Fail score**.**

## Conclusions

After examining the data in this study some conclusions can be made. The KCPE exam performance is unimodal but is not very close to a normal distribution as seen in the histogram. The normal QQ plot also shows that the data is not normally distributed, since the points in the plot do not form a straight line. A data transformation may be necessary. Using the Trend Analysis tool, the data exhibited a trend and, once refined, identified that the trend would be best fit by a second-order polynomial.

The semivariogram/covariance cloud illustrated that the unusually high semivariogram values are largely represented by the lines generally east to west reaffirming the early Nairobi settlement zoning and consequent growth pattern. The analysis using this tool indicates that the interpolation model should account for anisotropy. The semivariogram surface indicates there is no spatial autocorrelation in the data.

Going back to the objectives and premise of this study, it can be deduced that proximity to urban centers positively does not influences student performance on the KCPE exam. The findings and conclusions by Kimosop et al. (2015*)* were classic misinformation due to insufficient condition. Items may correlate highly and at the same time may not be related to the theoretical subject: this is the difference between reliability (high correlations between items) and validity (do we measure what we want to measure). Besides, looking at correlations, there are more advanced ways such as reliability analysis and principal factor analysis to ascertain whether events cluster or not. What is evident is that our capital city plan has over time been modified to assume a unique personality. However, it is more about neighborhoods than distance from anything else.

With regard to English as an examination subject and language of instruction, there evidence that good performance in the exam correlate with good English scores. However unless it is confirmed with certainty performance in Math and Science also trend with final mean score. What can be noted here is that because English much more strong with the other subjects than Kiswahili does, that indeed English proficiency is an important factor in overall exam performance.


**Bibliography**

Akin L. Mabogunje (1990). Urban Planning and the Post-Colonial State in Africa: A Research Overview. *African Studies Review*, 33, pp 121-203. Doi: 10.2307/524471.

Akolo, J. (2007). Architects propose transfer of capital city of Nairobi. Retrieved from http:www.kbc.co.ke/story.asp?ID=44432, Nairobi.

Anyamba, T.J.C (2004). Informalisation of a planned neighborhood in Nairobi. Oslo School of Architecture, Oslo.

Aryeetey-Attoh, S. (2003). Geography of Sub-Saharan Africa. 2nd Edition. Upper Saddle River: Prentice Hall.

Boundless ( 2015).The Structure of Cities. Retrieved from https://www.boundless.com/sociology/textbooks/boundless-sociology-textbook/population-and-urbanization-17/urbanization-and-the-development-of-cities-123/the-structure-of-cities-694-10482/

Chun, Y. and Griffith, D.A. (2013). Spatial Statistics & Geostatistics. Washington D.C: SAGE.-

Dean, L. (2013). How the Urban Slang of Nairobi slums is becoming the language of the people. Retrieved from Street Talk, http://www.slate.com/articles/news_and_politics/roads/2013/11/sheng_is_becoming_a_kenyan_language_how_the_urban_slang_of_nairobi_slums.html.

Glewwe, P., Kremer, M., & Moulin, S. (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics*, 1(1), 112-135. Retrieved from http://dx.doi.org/10.1257/app.1.1.112.

Gonzales, E. J., Chavis C., Yuwei Li, & Daganzo, C.F. (2009). Multimodal Transport Modeling for Nairobi, Kenya: Insights and Recommendations with an Evidence-Based Model. Working Paper UCB-ITS-VWP-2009-5, UC Barkley Center for Future Urban Transport. Retrieved from http://www.its.berkeley.edu/sites/default/files/publications/UCB/2009/VWP/UCB-ITS-VWP-2009-5.pdf

Harris, C. D., & Ullman, E. L. (1945). The nature of cities. The Annals of the American Academy of Political and Social Science, 7-17.

Hoyt, H. (1939). The structure and growth of residential neighborhoods in American cities.

Hungi, N., & Thuku, F. W. (2010). Differences in pupil achievement in Kenya: implications for policy and practice. *International Journal of Educational Development*, 30(1), 33-43. Retrieved from http://dx.doi.org/10.1016/j.ijedudev.2009.05.001.

Kimosop, P. K., Otiso, K. M., & Ye, X. (2015). Spatial and gender inequality in the Kenya certificate of primary education examination results. *Applied Geography*, *62*, 44-61.

Kurtz, J. R. (1998). *Urban Obsessions, Urban Fears: The Postcolonial Kenyan Novel*. City, Africa World Press.

Lockheed, M. E., & Verspoor, A. M. (1991). *Improving primary education in developing countries*. Oxford: Oxford University Press for World Bank.

Maksymenko, S., & Tranfaglia, A. (2015). The impact of professional athlete sponsorship on educational attainment in Western Kenya. *International Journal of Educational Development*, 41, 97-103. Retrieved from http://dx.doi.org/10.1016/j.ijedudev.2015.02.004.

Mehretu, A. (1993). "Cities of Sub-Saharan Africa," in S. Brunn, and J. Williams, (Eds), *Cities of the World* ( pp ) New York: Haper Collins.

Mitullah, W. (2007, month date). Urban Slum Report; The case of Nairobi, Kenya. *The Standard Group*, pp.

Mubea, K. and Menz, G. (2014) Spatial Effects of Varying Model Coefficients in Urban Growth Modeling in Nairobi, Kenya. *Journal of Geographic Information System*, **6**, 636-652. doi: 10.4236/jgis.2014.66053.

Myers, G. A. (2003). Verandahs of Power; Colonialism and space in urban Africa, Chapter Three, Colonial Nairobi. Syracuse University Press, New York.( see above Mehretu, 1993).

Nairobi Urban Study Group ( Year). *Metropolitan Growth Strategy* (Nairobi City Council report No. 10973). Retrieved from

Norušis, MJ, *SPSS 13.0 Advanced Statistical Procedures Companion.* Englewood Cliffs: Prentice Hall, 2004.

Ombok, E. (2006). Urbanization gone wrong, retrieved at http://www.businessinafrica.net/opinion/columns/nairobi_notebook/388738.htm, Nairobi.

Overton, J. (1987). The colonial state and spatial differentiation. *Journal of Historical Geography*, 13(3), p.267-282.- not cited

Republic of Kenya (1970) Kenya Population Census 1969. Government Printer, Nairobi.

Republic of Kenya (1981) Kenya Population Census 1979. Government Printer, Nairobi.

Republic of Kenya (1994) Kenya Population Census 1989. Government Printer, Nairobi.

Republic of Kenya (2000) Economic Survey 2000. Government Printer, Nairobi.

Republic of Kenya (2010) Economic Survey 2010. Government Printer, Nairobi.

Republic of Kenya (2014). Education for All 2015 National Review Report: Kenya. A Report prepared the Ministry of Education Science and Technology in view of the World Education Forum in response to UNESCO's invitation to the Member States.

Romesburg, C. (2004). *Cluster analysis for researchers*. Lulu. com.

Syagga, P.M., Mitullah, W.V. and Gitau, S.K. (2001), Nairobi Situation Analysis, Government of Kenya and UNCHS Collaborative Nairobi Slum Upgrading Initiative, Nairobi.

Tryfos, P. (1998). *Methods for business analysis and forecasting: text and cases*. Wiley.

UN-HABITAT (2005) Regional Urban Sector Profile Study. UN-HABITAT, Nairobi.

Vogel, M. (2008). *History of Urban Planning of Nairobi*. Retrieved from http://www.studio-basel.com/assets/files/029_NRB_ATLAS_11_planning_dr_A3.pdf.

White, L.W.T, Siberman L., Anderson, P.R (1948). Nairobi Master Plan for a Colonial Capital. Report prepared for the Nairobi City Council, Nairobi.