DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

THE USE OF SPATIAL STATISTICS FOR ANALYZING HEALTH AND DEMOGRAPHIC DATA IN TANZANIA

by

Ally BAKARI

June, 2015

İZMİR

THE USE OF SPATIAL STATISTICS FOR ANALYZING HEALTH AND DEMOGRAPHIC DATA IN TANZANIA.

A Thesis Submitted to the Graduate School of Natural and Applied Sciences at Dokuz Eylül University In Partial Fulfilment of the Requirements for the Master of Science in Geographical Informational System

by

Ally BAKARI

June, 2015

İZMİR

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"THE USE OF SPATIAL STATISTICS FOR ANALYZING HEALTH AND DEMOGRAPHIC DATA IN TANZANIA"** completed by **ALLY BAKARI** under supervision of **PROF. DR. YALÇIN ARISOY** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Yalçın Arısoy

Supervisor

Assoc. Prof. Dr. Orhan GÜNDÜZ

Jury Member

Prof. Dr. Talip GÜNGÖR

Jury Member

Prof. Dr. Ayşe OKUR

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENT

First and foremost I thank the Almighty God for the life given throughout my studies. My sincere appreciation and deep gratitude go to Professor Yalçın Arısoy for his excellent guidance, supervision and encouragement.

I also very grateful to Dr. Mert Çubukçu and Mr. Mahmoud Usama (Ms.) for their constructive suggestions.

We are greatly indebted to management of The Demographic and Health Surveys Program of ICF International in Rockville (USA) and National institute bureau of Statistics (Tanzania) for granting us the data for this research. I also appreciate the strong support, cooperation, encouragement and guidance received from the Management and the staff of Dokuz Eylül University.

Special appreciation goes to my family for their support, encouragement and love extended during the entire period of my study. Not forgetting my dear friends for helping me throughout the research.

Lastly, we are indebted to University of Dokuz Eylül for contributing in one way or another to the success of our academic pursuit.

God bless them all.

Ally BAKARI

THE USE OF SPATIAL STATISTICS FOR ANALYZING HEALTH AND DEMOGRAPHIC DATA IN TANZANIA.

ABSTRACT

In most of African countries, the demographic Health data is obtained by conducting surveys. The interviewers and field facilitators are always responsible for collecting data at field level. The papers work (questionnaires) or Person Digital Assistant (PDA) are always used as a tools during the data collection stage. After that, the data is submitted to the cleaning department before the analysis performed. Generally, different organizations have different ways of cleaning and organizing their data.

In this research, the data used was officially obtained from The Demographic and Health Surveys Program of ICF International in Rockville (USA) and National institute for Medical Research (Tanzania). The GIS software (Arc GIS 10.2 and Map Info Version 11.5) were used to analyze and visualize a spatial pattern relationship and autocorrelation between the data.

Among of the statistical methods for analyzing spatial Demographic Health Survey Data (DHSD), is the use of the GIS spatial statistics tools. In this tools, the data can be captured and processed so as to visualize their distribution pattern and its spatial autocorrelation. These tools include, the high or low clustering (Getis-Ord General G), cluster and outlier analysis, spatial autocorrelation (Moran I) and Hot Spots analysis.

Before running these tools, the data should be cleaned and projected into an appropriate coordinate system. These data were extracted based on questions from a research questionnaire. MapInfo version 11.5 was used as a tool to extract these data by using query commands and recorded on a Microsoft excel spread sheet. The sheet will be attached at the end of this research for reference purpose.

The spatial and cartographic capabilities of the Arc GIS helped in generating different maps which used as the basis of the discussion in the analysis stage. Since the research is based on analysis of Health data, only malaria, anemia and HIV/AIDS diseases were detailed analyzed and other data were also analyzed so as to build additional recommendation based on findings.

Keywords : Demographic Health Survey, Geographical Information Systems, Spatial statistic, Spatial autocorrelation, Hot and Cold-spot analysis, High-Low values, Negative autocorrelations, Positive autocorrelations, High/Low-value clustering.

TANZANYA SAĞLIK VE DEMOGRAFİK VERİLERİNİN ANALİZİNDE KONUMSAL İSTATİSTİK TEKNİKLERİNİN KULLANIMI.

ÖZ

Demografik sağlık verileri Afrika ülkelerinin çoğunda anketlerle belirlenmektedir. Anket görevlileri ve yardımcıları verileri doğrudan sahada toplamaktan sorumludurlar. Veri toplama aşamasında anket formları veya el bilgisayarları kullanılmaktadır. Daha sonra bu veriler analiz edilmeden önce hataların temizlenmesi için kontrol bölümüne sunulmuştur. Doğal olarak analiz için farklı organizasyonlar amaçları doğrultusunda verileri farklı biçimlerde düzenlemektedirler.

Bu çalışmada kullanılan veriler “The Demographic and Health Surveys Program of ICF International in Rockville (USA)” ve “National Institute for Medical Research (Tanzania)” tarafından elde edilmiştir. Verilerin analizinde, veriler arasındaki ilişkilerin araştırılmasında ve verilerin görüntülenmesinde CBS programları “Arc GIS 10.2” ve “MapInfo-Ver.11.5” kullanılmıştır.

Demografik sağlık verilerinin istatistiksel analizinde kullanılan yöntemlerden biri CBS programlarının konumsal istatistik ara yüzeyidir. Bu yöntemde (araçta), veriler toplanmakta ve işlenmekte, dağılım paternleri ve konumsal içsel korelasyonları görüntülenmektedir. Sözedilen araç (arayüzey) düşük ve yüksek düzey kümelenme (gruplaşma) (Getis-Ord General G), kümelenme ve aykırı değer analizi, konumsal oto korelasyon (içsel bağıntı) (Moran I) ve “Hot Spots” analizlerini içermektedir.

Yukarıda sözedilen yöntemler uygulanmadan önce veriler kontrol ve organize edilerek uygun koordinat sistemine dönüştürülmüştür. Öncelikle araştırmaya yönelik sorular belirlenmiş; MapInfo kullanılarak gerçekleştirilen sorgulamalarda, sorgu ve sorguya ait veri alanları, sorgularda kullanılacak ve sorgu sonucunda oluşacak tablolar önceden tespit edilerek bir Excel dosyası oluşturulmuştur. İlgili dosya çalışma sonunda ekler bölümünde verilmiştir.

Arc GIS programının konumsal ve kartografik yeteneği yardımıyla farklı tematik haritalar türetilmiştir. Bu haritalar çalışmada değerlendirilen hastalıklara ilişkin tartışmalara esas teşkil etmektedir. Çalışmada araştırılan hastalıklar malarya, anemi ve HIV/AIDS'dir. Diğer bazı veriler bulgulara ilişkin tartışmaları aydınlatacak ve öneriler oluşturacak şekilde analiz edilmiştir.

Anahtar kelimeler : Demografik Sağlık Anketi, Coğrafi Bilgi Sistemleri, Mekansal istatistik, Mekansal otokorelasyon, Sıcak ve Soğuk nokta analizi,Yüksek ve Düşük değerler, Negatif otokorelasyon, Pozitif otokorelasyon, Yüksek ve Düşük değerli kümeleme.

CONTENTS

LIST OF FIGURES

Pages

LIST OF TABLES

Pages

CHAPTER ONE
INTRODUCTION

Demographic and Health Surveys (DHS) are nationally-representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, Health and nutrition. Usually, the DHS questionnaires are designed to probe both individual and household information in the DSA. The individual information could be, education background, religion, birth and death, while the altitude, longitude and latitude is among of household information.

The collected data is then statistically analyzed to get reliable information which will be used for decision making. Usually, the data is analyzed by statistical software including SPSS, STARTA or SAS. However, all of these software has no capability of joining the data with its spatial relationship.

Thus, the emerge of Geographical Information System (GIS), come to solve this challenge, whereby the data can be analyzed while coordinating to its spatial properties. So, the decision of what type of software to use during analysis will depend with the scope and objectives of the research.

Since the DHS involve the collection of household's information and members within the households at specific location, the GIS is the best software to analyze them because of its spatial and cartographic capability.

In Tanzania many organizations conduct the demographic Surveillance surveys which finally shared their data and prepare the final reports. In this research, the data was officially requested from The Demographic and Health Surveys Program of ICF International in Rockville (USA) and The National Bureau of Statistics-Tanzania.

The DHS is responsible for collecting and disseminating nationally representative data on Health and population in many countries. These data are the ones which typically used to monitor and evaluate the child health and population program in the many countries including Tanzania. Since it is more often repeated twice a year, it is a good source of data for providing a specific trend of the data necessary to gauge progress.

In this research, the goal was based in monitoring, evaluating and forecasting the demographic Health data in Tanzania where by the common diseases were spatially analyzed and presented on a map so that the government and Health stake holders can be on a good position to understand it for the further measures. Also, the research aimed to analyze these data to investigate the spatial distribution of some diseases and the relationship between these diseases and census data.

The results of analysis could be used by the government to provide an intermediate action for more severe regions or used to shape the policies and change the intervention program objectives as well as the health and population planning. Example, it is simply worthy to provide an iron supplementation in the place where is more affected by anemia.

In this study, Arc GIS 10.2 and MapInfo 11.5 were used as GIS software. Also, Microsoft Excel was used extensively especially in data analysis and creation of tables and some figures.

CHAPTER TWO

TANZANIA

2.1 Geography, History and Economy

2.1.1 Geography

The United Republic of Tanzania is the largest country in East Africa, covering 940,000 square kilometers, 60,000 of which are inland water. Tanzania lies south of the equator and shares borders with eight countries: Kenya and Uganda to the north; Rwanda, Burundi, the Democratic Republic of Congo, and Zambia to the west; and Malawi and Mozambique to the south.

Tanzania has an abundance of inland water, with several lakes and rivers. Lake Tanganyika runs along the western border and is Africa's deepest and longest freshwater lake and the world's second deepest lake. Lake Victoria, the world's second largest lake, drains into the Nile River, which flows to the Mediterranean Sea.

The Rufiji River is Tanzania's largest river, and it drains into the Indian Ocean south of Dar- es- Salaam. Although there are many rivers, only the Rufiji and the Kagera, which empties into Lake Victoria, are navigable by anything larger than a canoe.

One of Tanzania's most distinctive geological features is the Great Rift Valley, which was caused by geological faulting throughout eastern Africa and is associated with volcanic activity in the northeastern regions of the country. Two branches of the Great Rift Valley run through Tanzania.

The western branch holds Lakes Tanganyika, Rukwa, and Nyasa, while the eastern branch, which ends in northern Tanzania, includes Lakes Natron, Manyara, and Eyasi. Except for a narrow belt of 900 square kilometres along the coast, most of Tanzania lies 200 metres or more above sea level, and much of the country is higher than 1,000 metres. In the north, Mount Kilimanjaro rises to 5,895 metres—the highest point in Africa.

The main climatic feature for most of the country is the long dry spell from May to October, followed by a period of rainfall between November and April. The main rainy season is from March to May, along the coast and around Mount Kilimanjaro, with short periods of rain arriving between October and December. In the western part of the country, around Lake Victoria, rainfall is well distributed throughout the year, with the peak period falling between March and May. The map below shows the geographic boundaries and water bodies of Tanzania (Figure 2.1).

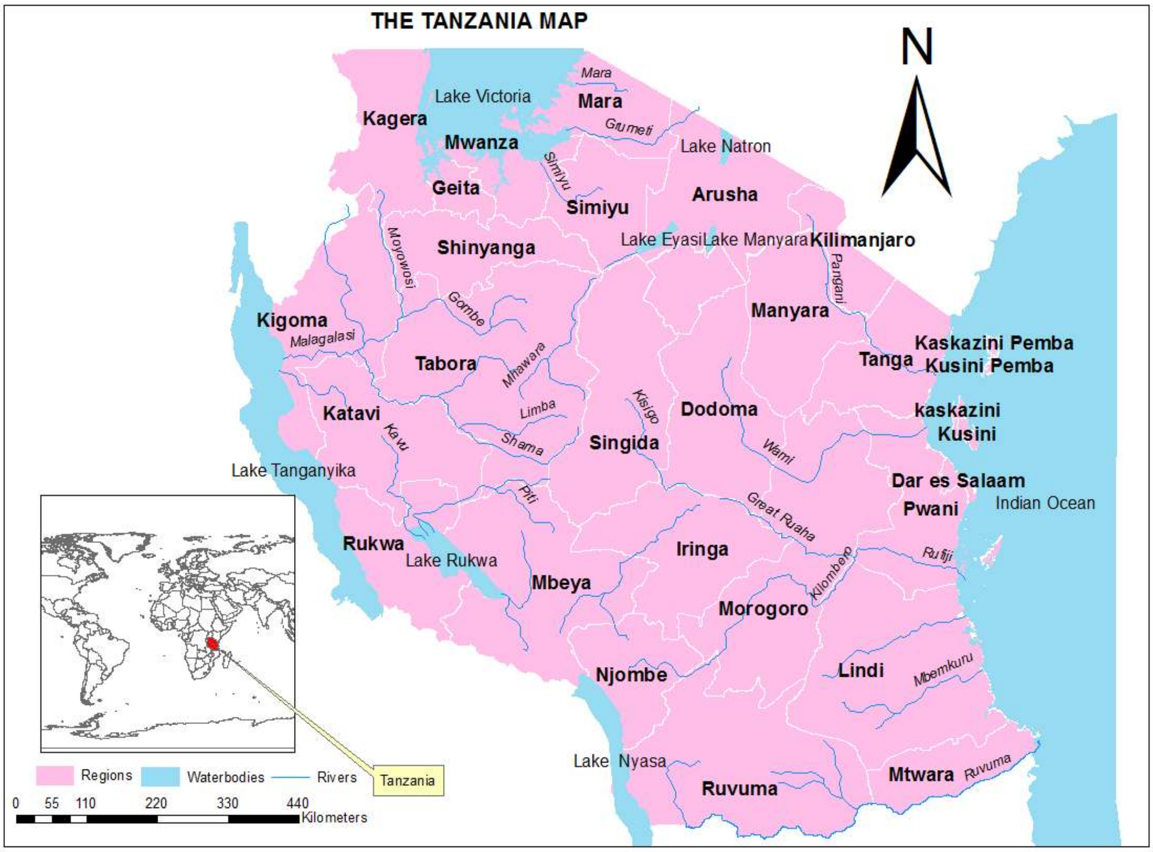


Figure 2.1 The map of Tanzania

2.1.2 History

Tanzania (formerly Tanganyika) became independent of British colonial rule in December 1961. One year later, on 9 December 1962, it became a republic, severing all links with the British crown except for its membership in the Commonwealth. The offshore island of Zanzibar became independent on 12 January 1964, after the overthrow of the rule of the sultanate. On 26 April 1964, Tanganyika and Zanzibar joined to form the United Republic of Tanzania.

Tanzania is currently operating under a multiparty democratic system of government, with the president and the national assembly members being elected every five years. Tanzania's president can hold office for a maximum of two five-year terms. For administrative purposes, mainland Tanzania is divided into 25 regions, and Zanzibar is divided into 5 regions. Each region is subdivided into several districts.

2.1.3 Economy

Tanzania has a mixed economy. Agriculture, comprised of crop growing, animal husbandry, forestry, fishery, and hunting, played a key role in past years. In the current economy, activities in the service industry account for 42 percent of the Gross Domestic Product (GDP). In 2009 the agricultural sector grew by 3.2 percent compared with growth of 4.6 percent in 2008 (NBS, 2009).

During the same period, the growth rate of crops decreased from 5.1 percent to 3.4 percent and that of livestock decreased from 2.6 percent in 2008 to 2.3 percent in 2009. The drought during the 2008-2009 planting season caused these decreases in growth, particularly in the northern part of Tanzania, where there was inadequate pasture and water for livestock.

In 2009, the GDP grew by 6.0 percent, which compared poorly with 7.4 percent growth in 2008. The slowdown in growth for 2009 was attributed to the impact of the global financial crisis as well as the drought in 2008-2009, which affected agricultural production, hydro power generation, and industrial production. They all contribute significantly to total GDP (MFEA, 2009).

The 2009 GDP at current prices is $21,702.03538 million, which is equivalent to $16,548.73789 million at 2001 constant prices. With an estimated population on the Tanzania Mainland of $31,307.69 in 2009, the per capita income is $534 at current prices, compared with $546.31 in 2008, indicating an increase of 10.3 percent (NBS, 2010).

2.2 Health Services System in Tanzania

The Government, Parastatal Organization, voluntary organization, Religions Organization, Private Practitioners and Traditional Medicine, provides Health Services in Tanzania. The referral system (structure) starts from the community level (village) up to the treatment abroad.

The Health system and especially the Governments referral system assumes a pyramidal pattern of a referral system recommended by Health planners, that is from dispensary to Consultant Hospital (World Bank, 1993).The structure of Health services at various levels in the country is as follows:

1. Village Health Service

This is the lowest level of Health care delivery in the country. They essentially provide preventive services which can be offered in homes. Usually each village Health post have two village Health workers chosen by the village government amongst the villagers and be given a short training before they start providing services.

2. Dispensary Services

This is the second stage of Health services. The dispensary cater for between 6,000 to 10,000 people and supervise all the village Health posts in its ward.

3. Health Centre Services

A Health Centre is expected to cater for 50,000 people which is approximately the pop. of one administrative division.

4. District Hospitals

The district is a very important level in the provision of Health services in the country each district is supposed to have a district hospital. For those districts which donate have Government normally negotiates with religious organizations to designate voluntary hospitals get subventions from the Government to contract terms.

5. Regional Hospitals

Every region is supposed to have a hospital. A regional Hospital offer similar services like those agreed at district level, however regional hospitals have specialists in various fields and offer additional services which are not provided at district hospitals.

6. Referral/Consultant Hospitals

This is the highest level of hospital services in the country presently there are four referral hospitals namely, the Muhimbili National Hospital which cater the eastern zone; Kilimanjaro Christian Medical Centre (KCMC) which cater for the northern zone, Bugando Hospital which cater for the western zone; and Mbeya Hospital which serves the southern Highlands.

2.3 National Health Policies and Plans

The aim of the National Health Policy is to improve the Health and well-being of all Tanzanians, with a focus on the most at risk groups of people and to address the needs of the population. The main priorities of the policy and plan of 1990 were as follows:

i. Reduce infant and maternal mortality and morbidity and to raise the population's life expectancy.

   a. The plan will attempt to reach these goals through the provision of better maternal and child Health services, promotion of nutrition, control of communicable diseases and the integrated treatment of common conditions.

ii. Increase the availability and accessibility of Health services to all people at no cost.
iii. Training cadres at all levels, from community to national level, to allow greater self-sufficiency in trained personnel.
iv. Mobilize the community on how to prevent common Health problems and increase the capacity for communities to assess and analyze problems so that effective action may be implemented through greater community participation.
v. Promote awareness among at the community level that Health problems can be managed successfully through multi-sector cooperation, such as education, water and sanitation, community development, women's organization, agriculture, Non-Government Organization (NGO) and the Government.

However, there has since been developments in this policy. Private medical practice was realized one year later and a cost-sharing system was introduced in the public Health facilities in July15th 1993. A common framework for a comprehensive Health Sector Reform was formulated, focusing on an "Inter-Ministerial Technical Committee" and a "Working Group" in office of the Ministry of Health (MOH).

The working group included technical representation from other ministries, which acted by coordinating sector studies, developing a "Health Sector Strategy" and preparing a "Health Sector Plan". In 1994 the Government prepared the Social Sector Strategy which, with respect to Health, proposed a restructuring of the ministry with a focus on the provision of basic and preventative Health services, decentralization of services with greater independence given to district Health authorities, greater attention to policy formulation, quality control and promotion of private investment in the sector.

In the Joint Ministry of Health (JMH) and Donors Health Sector Reform (DHSR) mission of 1995, a donor statement was produced underlying the principles of the reform:

- Adjustment of the Government's function in the Health sector, due to liberalization of the economy, including its interface with the private sector and changes in services such as logistics, pharmaceuticals and human resources.
- The new role for the Ministry of Health in policy, sponsorship and regulation of the Health sector.
- Financing of provision of essential cost-effective Health packages.
- Decentralization of the public sector Health system, with an emphasis on the production unit on a district level.
- Reallocation of Government funds in favor of the essential Health packages, with donor support, and development of the financing services not included in the Health packages.
- Distribution of resources to areas on the basis of population, weighted for socio-economic factors.
- Continual improvement in the central drugs and supplies system.

The joint mission allowed the realization that to implement the government's plan required regular review the three-year "Government Action Plan" (within the "Strategic Health Plan"), which had the following features.

Firstly, to review basic governing principles to allow continued relevance to a changing environment, secondly, to allow agreement on the next steps in implementing reforms and on the tasks to be undertaken during the next period, such as sequencing and costs, and finally agreement on the implementation mechanisms as well as on the monitoring and evaluation of the action plan for the next review.

In 1996, the Health Sector Reform Plan of Action (HSRPA) formed the basis for implementation of the Health sector/donor support in the reform process. Improved Health of the family became a major strategy for reform in the Health sector. The MOH coordinated this initiative, and provided increased emphasis on the provision of Health, mainly through nutrition and family planning 16 intervention. Greater community-based Health care was also highlighted as an integral part of the reform plan.

The National Health Policy was reviewed in 2002 and is now subject to government approval. The Proposal for Health Sector Reform of 1994, the Tanzanian Development Vision 2025, Poverty Reduction Strategy 2000, gender mainstreaming, the importance of vulnerable groups and HIV/AIDS have been incorporated in the updated Health Policy.

The main focus points of the Health Sector Reform Strategic Plan 2002-2007 (Zanzibar) and the Medium-Term Strategic Plan 2002-2004 (Mainland) are as follows:

- To improve access, quality and efficiency of district primary Health services.
- To improve the management, quality, efficiency and financial viability to support primary level services, teaching and research, in Level Two and Level Three hospital services.
- To improve the capacity for the MOH in sector-wide management, policy development, analysis and national planning; development of guidelines; performance monitoring, evaluation and regulation of service delivery and practice.
- To develop human resources for Health by improving skills and building management capacity at all levels of the Health system.
- To strengthen the national support systems for drugs and other medical supplies and equipment, physical infrastructure, Health management information systems and transport.
- To improve Health financing and financial management.
- To promote the participation of the private sector in the provision of Health services.
- To develop and implement a system for donor and government involvement, coordination, monitoring and evaluation.

However, reforms have not responded fully to the Health needs of the Mainland and Zanzibar population. Access to quality Health services, especially among poor and rural communities, remains low. In fact, according to the World Health Report 2000, Tanzania is performing poorly, specifically in attaining fairness of financial contribution to Health costs.

Floods and drought account for 38% and 33% of all emergencies respectively. Policies and strategies on the issue of national disasters and emergency situations are weak judging from recent experiences. Regional and Local Authorities are required to have “Disaster Management Committees” and an “Emergency Preparedness Unit” has been set up at the MOH to cater for management of emergencies. In the case of a large humanitarian emergency, the country has limited capacity to manage the event independently and greatly depends on the international community for assistance. The Table 2.1 below shows the Health facilities in Tanzania mainland according to ownership.

Table 2.1 Health facilities in Tanzania mainland.

Facility Ownership	**Consultant Specialized**	**Hospitals**			**Health Centre**	**Dispensary**
		Regional	**District**	**Other**		
Government	6	15	39	4	211	2019
Parastatal	0	0	1	3	3	110
Private	1	0	0	33	11	396
Total	**7**	**15**	**40**	**40**	**225**	**2525**

CHAPTER THREE

METHODOLOGY

3.1 Geographical Information Systems (GIS)

GIS is a computer based system that provides the following three sets of capabilities to handle geo processing;

- Data capture and preparation
- Data management including storage and maintenance
- Data manipulation and analysis

This implies that a GIS user can expect support from the system to enter a georeferenced data to analyze it in various ways, and to produce a map presentation data. This would include support for various kind of coordinate systems and transformation between them, options for analysis of the georeferenced data. And obviously a large degree of freedom of choice in the way this information is presented (such as color scheme and symbology).

Alternatively;

The GIS could be defined as a system of hardware and software that links mapped objects with text information that describes them and provides tools for storage, retrieval and manipulation of both types of data. Generally the GIS data has two parts;

- Spatial data, this is the data with location values (mapped data)
- Attribute data, the data which explain the features of the entity (tabular data)

Since the GIS data contains two types of data as described above, also its database must be able to store this type. Previously, the generation of software was equipped with relatively rudimentary facilities. Since 1990's there has been an increasing trend in GIS applications that used a GIS for spatial analysis, and use a database for storage. In more recent years, spatial databases have emerged. Besides traditional administrative data, they can store representations of a real word geographic phenomenon for GIS usage. These databases are special because they use additional techniques different from tables to store these spatial representations.

Some properties of GIS is as described below;

- Data can be stored, retrieved, manipulated and analyzed
- It is a combination of a picture image and associated data
- It contains cartographic features like points, lines polygons and their attributes
- New information is generated through system operations
- Querying attribute and/or spatial data is possible
- Many procedures can be automated within the system

The GIS capabilities include;

- Cartographic capabilities
- Database management
- Analytical

a) The GIS Cartographic Capabilities

Cartography is the art and science of making maps. To create modern, high-quality cartography requires the use of appropriate technology. This provides efficient processes that amplify our human creative and expressive skills in order to communicate the essential spatial message. The GIS is among of the software which has a strong cartographic abilities including;

- Coordinate geometry and translation routines
- Edge matching routines
- Projection routines
- Zooming and scaling routines
- Length and area computation routines
- Curve fitting routines

b) Database Management

GIS software packages provide support for both spatial and attribute data, i.e. they accommodate spatial data storage using a vector approach and attribute data using tables. Historically, the database management systems (DBMS) have been based on the notation of tables for data storage. For some time, substantial GIS applications have been able to link an external database to store attribute data and make use of its superior data management functions. Currently all major GIS packages provide facilities to link with DBMS and exchange the attribute data with it. Spatial (vector) and attribute data are still some times stored in separate structures, although they can now be stored directly in a spatial database (Huisman & Deby 2009).

Below are the database management capabilities of GIS;

- Organization of data
- Separation of graphic and non-graphic data
- Linking attribute data to geographic data
- Attribute processing
- Data retrieval
- Data manipulation
- Data storage
- Maintenance of integrity of data

c) Analytical Capability

Most GIS provide the capability to build complex models by combining primitive analytical functions. Systems vary as to the complexity provided for spatial modelling, and the specific functions that are available. However, most systems provide a standard set of primitive analytical functions that are accessible to the user in some logical manner. The following are the analytical function of GIS;

- Data processing
- Computational operations
- Modeling
- Thematic mapping
- Spatial analysis
- Graphic and non-graphic editing
- Generation of new data

There are many statistical models, which can be used for demographic estimation and projection. Most of these models although they can account for change in demography, they lack the spatial aspect (it is not possible to geographically view and analyze the patterns) (Klosterman, 1993); thus, the use of the statistical models into GIS, so that the spatial analytical power of GIS is utilized.

Geographic information systems plays a large role, including measuring the geographic displacement, monitoring the spatial change over time and point pattern analysis of household/individual locations. Some of the inherent advantages of GIS are its efficient and flexible storage, display, exchange of certain kinds of spatial data, ability to locate interactive spatial features and perform spatial analysis at variable scale, and to integrate information from a variety of sources, as well as opportunities for a variety of quantitative analysis models into one user interface. In turn, this allows for spatial analyses that would either not have been possible or at a minimum far more difficult prior to the advent of GIS.

The stages for GIS analysis methodology to carry out demographic spatial analysis followed with the help of the flow chart diagram below (Figure 3.1).

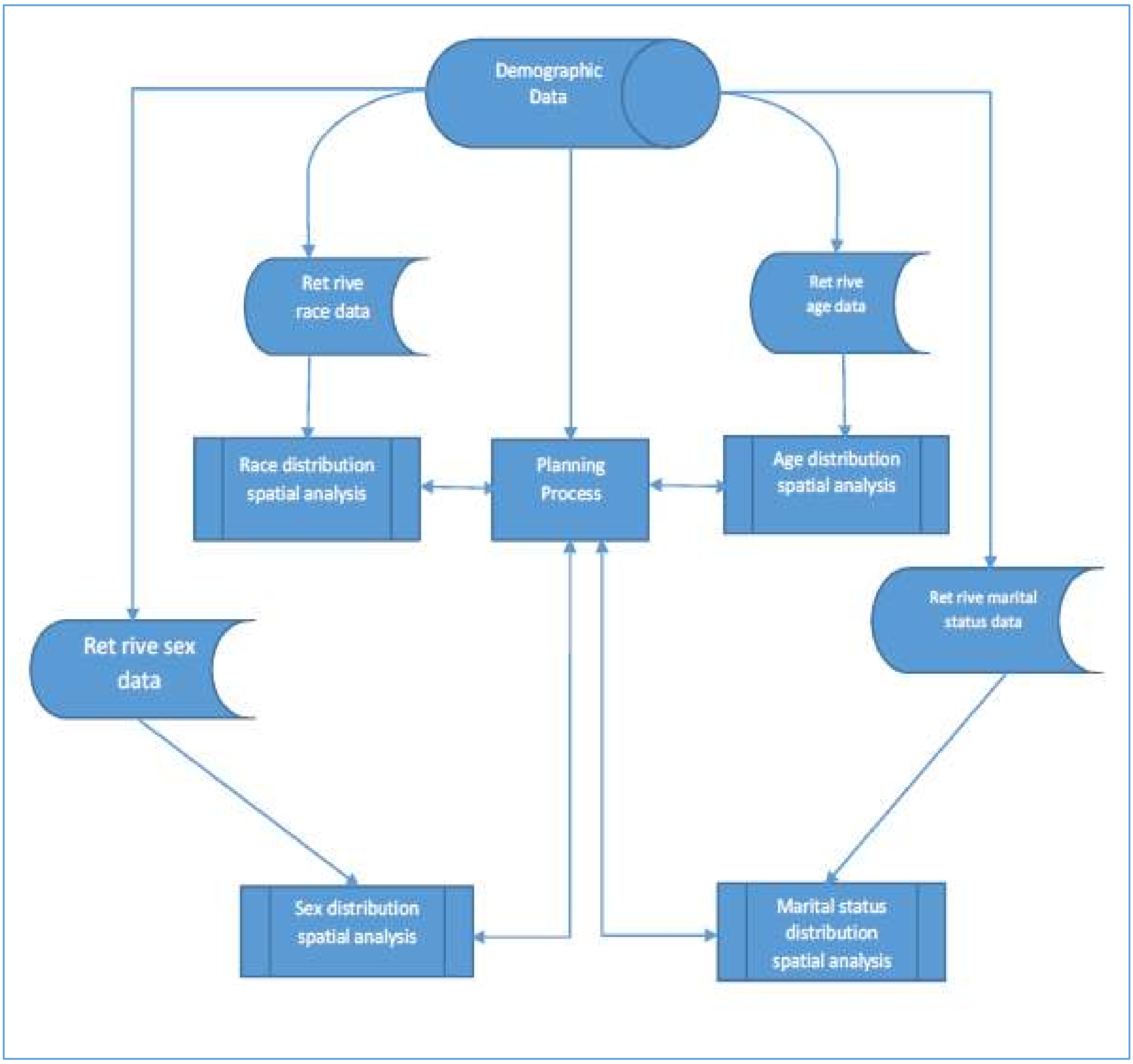


Figure 3.1 Flow chart shows the steps necessary for Demographic spatial analysis

3.2 Spatial Autocorrelation

3.2.1 The Global Moran's I

Spatial autocorrelation statistics include the space and properties of spatial objects, namely, the spatial relationships as well as characteristics of property values among objects with adjacent to each object. Spatial autocorrelation statistics includes many spatial autocorrelation indicators and methods, in the applications of correlation analysis research, spatial statistics are used to test the existence of spatial correlation by MORAN'S I.

Global Moran's I Index is used to identify the spatial pattern which may be clustered, dispersed or random based on feature locations and attribute values simultaneously. This statistical test compares the values of neighboring locations and a strong positive spatial autocorrelation (clustered) is indicated when the neighboring units over the whole study area have similar values. If neighboring units have very dissimilar values, then the statistics indicates strong negative spatial autocorrelation (dispersed) (Zhang et al, 2014)

The value of Moran's I Index is between -1 and 1. 'Z' value is calculated to assess whether the observed clustering/dispersing is statistically significant or not. When the Z score indicates statistical significance, a positive Moran's I index value indicates tendency toward clustering while a negative Moran's I index value indicates tendency toward dispersion. When the Moran's I Index is not significantly different from 0, there is no spatial autocorrelation and the spatial pattern is considered to be random (Jung & Park 2014; Amaral & Anselin 2013; Gwarda, 2014).

Generally, by running the Global Moran I, apart from identifying whether the feature is clustered, dispersed or random, the tool also calculate the values of "Z-Score " and "P- Value". These values always helps in evaluating the Moran I index value.

Most of the statistical analysis starts by Null hypothesis. The null hypothesis for pattern analysis tools is completely spatial Randomness. Therefore it is a task for the P- value and Z- Score to tell whether to accept or reject the Null hypothesis.

The p-value is a probability. For the pattern analysis tools; it is the probability that the observed spatial pattern was created by some random process. When the p-value is very small, it means it is very unlikely (small probability) that the observed spatial pattern is the result of random processes, so you can reject the null hypothesis (Lin, 2004; Ord & Getis, 2001).

To reject the null hypothesis, you must make a subjective judgment regarding the degree of risk you are willing to accept for being wrong (for falsely rejecting the null hypothesis).Consequently, before you run the spatial statistic, you select a confidence level. Typical confidence levels are 90, 95, or 99 percent. A confidence level of 99 percent would be the most conservative in this case, indicating that you are unwilling to reject the null hypothesis unless the probability that the pattern was created by random chance is really small (less than a 1 percent probability) (Zhang & Zhu 2011).

The spatial autocorrelation (GLOBAL MORAN'S I) tool is an inferential statistic, which mean that the results of the analysis are always interpreted within the context of the null hypothesis (H_o).

Let assume the;

P* is a critical value, P*- Value, and

P- Value is the result obtained

Therefore,

a) If [P > P*], then the P- Value is not spatial statistically significant and you cannot reject the null hypothesis, it is quite possible that, the spatial distribution of the feature value is a results of the random spatial process.
b) If [P < P*], then the P-Value is spatial statistically significant. But;

i. If [P < P*] and the Z-SCORE is Positive then,
The null hypothesis may be rejected. The spatial distribution of the high values or low values is more spatially clustered than would be expected if the underlying spatial processes were random.

ii. If [P < P*] and the Z-SCORE is Negative.
You may reject the null hypothesis. The spatial distribution of high values and low values in the dataset is more spatially dispersed than would be expected if underlying spatial processes were random. A dispersed spatial pattern often reflects some type of competitive process—a feature with a high value repels other features with high values; similarly, a feature with a low value repels other features with low values (Yu & Chun, 2010; Ma, 2014).

The Moran's I formula is described below;

$$\mathbf{I}= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} W_{ij}}{2\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij} S^2} \qquad (3.1)$$

S^2 – Sample variance for attribute x (Region attribute)

$\bar{x}$ - Sample mean for attribute x

W_{ij} - is a measure of spatial relation between the feature point “i’ and “j”

x_i -Attribute value for point feature “i”

$W_{ii} = 0$

n - Number of point features,

$S_{ij=} \left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)$

The Interpretation is as follow;

$I > \frac{-1}{n-1}$ The cluster similar attribute value

$I \cong \frac{-1}{n-1}$ Close, Random Distribution.

$I < \frac{-1}{n-1}$ Dispersed, Uniform, pattern with close point shows different attribute values

$$S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} \tag{3.2}$$

$$W_{ij} = \frac{1}{d_{ij}}, \frac{1}{d_{ij^2}} \tag{3.3}$$

3.2.2 The Local Moran's I

The basic calculation principle of Local Moran I tool is described below, the most commonly used measure of local autocorrelation index is Local Moran I, which is evolved by autocorrelation index of Global Moran I, namely, the sum of all the Local Moran I index is proportional to the global spatial relationship index, reflecting that the similar positive correlation is transiting to the opposed one. The tool is also used to identify the cluster feature of high or low values, and the spatial Outliers (Zhang & Zhu 2011; Yang et al, 2011)

For a particular geographic entity unit, the calculation formula of Local Moran I as follows:

$$I_i = \frac{x_i - \bar{x}}{S_{i^2}} \sum_{j=1, j \neq i}^{n} W_{ij}(x_j - \bar{x}) \tag{3.4}$$

Where;

x_i is an attribute for feature i,

$\bar{x}$ is the mean of corresponding attribute.

W_{ij} is the spatial weight between feature i and j.

$$S_{i^2} = \sum_{j=1, j\neq i}^{n} (x_j - \bar{x}) - (\bar{x})^2 \tag{3.5}$$

n is equating to the number of features.

The zI_i - Score for statistics are computed as;

$$\mathbf{zI_i} = \frac{\mathbf{I_i} - \mathbf{E}[\mathbf{I_i}]}{\sqrt{\mathbf{V}[\mathbf{I_i}]}} \tag{3.6}$$

Where;

$$\mathbf{E}[\mathbf{I_i}] = -\frac{\sum_{j=1, j\neq i}^{n} \mathbf{W_{ij}}}{\mathbf{n-1}} \tag{3.7}$$

$$\mathbf{V}[\mathbf{I_i}] = \mathbf{E}[\mathbf{I_i}^2] - \mathbf{E}[\mathbf{I_i}]^2 \tag{3.8}$$

3.2.3 The Hot Spots Analysis (Getis – Ord Gi)*

The Moran's I has well- established statistical properties to describe the spatial auto correlation globally. However it is not effective in identifying the different type of clustering spatial patterns. These patterns are sometime described as the Hot Spots and Cold spots.

The G- Statistics (Getis & Ord, 1992, 1995), has the advantage of detecting the presence of Hot and Cold Spots over the entire area of study. A local measure of spatial autocorrelation is the local version the General G statistics.

Generally, the Hot spots analysis tool calculate the Getis – Ord G* for each feature in the datasets. The results of Z-SCORE and P-VALUE tells the feature if it is either High or Low values cluster. The tool works by looking at each features within the statistically significant hot spots. A feature with high values is interesting but may not be statistically significant Hot Spot. To be statistically significant hot spot; A feature with high spots must be surrounded by the features with the high values also (Stopka et al, 2014). The local G statistics is defined as;

$$G_i(d) = \frac{\sum_j W_{ij}(d) x_j}{\sum_j x_j} ; i \neq j \tag{3.9}$$

The G is usually defined by distance "d" within each aerial can be regarded as a neighbors of i. The weight w_{ij} (d) is 1, if an aerial unit "j" is within "d"; or 0 otherwise, thus the weight matrix is essentially a binary symmetrical matrix but the neighboring relationship is defined by distance (d).

Thus, the sum of weight is;

$$W_i = \sum_j W_{ij}(d) ; j \neq i \tag{3.10}$$

The recommended conceptualization for calculating the Hot Spots analysis is the "The fixed Distance Band" because during analysis each feature is analyzed within the context of neighboring features. Neighboring features inside the specified critical distance receive a weight of one (1) and exert influence on computations for the target feature. Neighboring features outside the critical distance receive a weight of zero (0) and have no influence on a target feature's computations (Peeters et al, 2014; Handique et al, 2011).

CHAPTER FOUR
DATABASE OF THE STUDY

4.1 Metadata

In this study the data was obtained from two different sources, The Demographic and Health Surveys Program of ICF International in Rockville (USA) and National institute bureau of Statistics (Tanzania). The survey was conducted between the years 2011/2012. From the ICF, the DHS excel dataset provides two tables; Member Table and Household Table. The original database structure of the table is given in the appendices A and B. The metadata of these tables are given in Table 4.1.

On the other hand, the National Bureau of Statistics of Tanzania provides the shape file for each region in Tanzania. This data contains some census information such as population, age distribution, areas for each region in kilometer square (km^2), sex and voting population.

Table 4.1 The Metadata of Households and members table

Table Name	**Years**	**Organization**	**Records**	**Fields**	**Size**
Households	2011/12	ICF	10040	93	2.80MB
Members	2011/12	ICF	54020	102	27.4MB

Although, the original database has a good amount of information about the respondents, there are many empty (null) fields and/or some special codes. The general code system for the given data was, 4 digit, 3 digit, 2 digit and 1 digit variables respectively. Depending on the responses the code might decreasing from these special code accordingly; for example, 9996, 996, 96, 6; 9995, 995, 95, 5. Below is Table 4.2 which shows the special codes used during the interview processes.

Table 4.2 The special codes and its meaning used during the survey

CODES	EXPLANATION
Blanks	Represent, the variable is not applicable
9999, 999, 99, 9	The given questions should have been answered by respondents but the interviewer has no information about that variable
9998, 998, 98, 8	The Question is unknown to the Respondent (DON'T KNOW)
9997, 997, 97, 7	Implies the answer was inconsistency to the respondents.
"0"	Was coded for negative responses (For example, No education or No problem) though some countries coded "2" for negative responses to their questionnaires.

4.2 Organizing Data

The variables used in original questionnaires are always convenience for data collection but not always for analysis. Often the same questions are always asked in several places in questionnaire but with different respondents. Therefore it is important after data collection to have other form of combination and create a form that will be easily used during analysis.

In general, the data structure depends much with the software and hardware requirements used to analyze them. In this research, the data was formatted in a specific standard and form so as to simplify its identification and accelerate processing during the analysis stage.

For respondent identification each record of data starts with case identification (CASE ID) for both household and individual information. The requested data for demographic Health survey (DHS) were composed many fields for analysis. So, it is a duty of the researcher to select an appropriate data fields which suit to his or her research; same to this research only fields which suits to the research title were selected.

Therefore, 10040 households (HH) records and 54020 members (HM) were selected.

Based on these data, the research questions were used to write a queries. The sheet composed 38 questions and their fields are described in the Table 4.3 below. The research question sheet will be attached at the end of this research and the CD which contain all of the queries.

Table 4.3 Fields used on the research questions sheet.

ID	FIELD	EXPLANATION
1	Q_ID	The unique question Identification number.
2	Question	The explanation of the aim of the query.
3	Input Table(s)	The input tables for extracting required information. Since the original tables are HH and HM, it doesn't mean that, only these one will always be used as the input tables, however the output table for any question could be used as the input for the other question.
4	Fields	The field's name (codes) related to the analyzed questions.
5	Output Tables	The generated tables after running the query for the specific question. The results from these tables will further be spatially distributed on a map.
6	Query	The query documentation for each question.
7	Query File	The saved query name for each question. It is very important to save the query for each question because, it will be easily to trace the mistakes or to modify the query at the time you observe an errors.
8	Results	This field shows the output, if it is a table or a value.

Among of the interested part was how to decide and select which fields should represent the economic activities. This is because economic indicators for rural and urban area were different. Finally the conclusion ends by selecting appropriate fields for both rural and urban which represents the economic status. The selected fields for rural areas are described in Table 4.4 below:

Table 4.4 Fields used for calculating rural economic status.

FIELDS	TYPE OF DATA	MEANING
HV009	VALUE	Numbers of HH members
HV025	0,1	Type of Residence
HV206	YES,NO	Has Electricity
HV211	YES,NO	Has Motorcycle
HV212	YES,NO	Has a Car
HV216	VALUE	Number rooms for Sleeping
HV227	YES,NO	Has a mosquito bed Net
HV245	VALUE	Has hectares For Agriculture Land
HV246A	VALUE	Own Cattle
HV246B	VALUE	Own cows/Bulls
HV246C	VALUE	Own Donkey/Horses
HV246D	VALUE	Own Goats
HV246E	VALUE	Own Sheep
HV246G	VALUE	Own Pigs
HV246F	VALUE	Own chicken/Other Poultry

Then, the average members of household slept per one room was calculated as a ration of number of rooms for sleeping per numbers of HH members, (HV216/HV009) and add as the new field. By doing so, the new field economic status “ECOSTAT” was created by running a query. The logic and the methodology behind this query is given in the appendix C.

CHAPTER FIVE

ANALYSIS

5.1 General Overview

In this section, the organized data was used to demonstrate the general picture of the Tanzanian's Health related data. Refining data and other manipulation was given in the previous chapter. In this chapter, some thematic maps regarding general overview of Tanzania is given first; then Health related results were followed. Since Anemia, Malaria and HIV was the base of analysis, they were investigated and separately presented under different titles.

The GIS software was used to derive the outputs from each question and analyzing the data. The output results obtained when running the queries using MapInfo were then used as the basis for analyzing the data. The results obtained were in form of tables and some specific values for some questions. These tables were imported into Arc Map for spatial distribution and production of cartographic maps.

The analysis performed was based on the DHS data, economic status and the population density. Different tables, maps and figures were produced to sharpening and visualizing the results so as to accelerate the decision making.

Population density can be defined as the number of people per occupied land. It usually explains the average number of people living in a specific piece of land. By doing so, Population density measure the intensity of land use with respect to the number of people within that land.

Though the population density is a good indicator of aerial population distribution, it conceals many of the internal disparities in population concentration and its spatial distribution. People are normally very selective with regards to the location settlements. Similarly, certain human activities are selective in nature in term of location and suitability. Other factor that influence the population concentration includes accessibility to the reliable transport, Health and water services, climatic condition, availability of fertile land and reliable economic activities.

Generally in Tanzania, the population density vary from one region to another, and there is a number of reasons which course this changes, for example the level of urbanization, availability of social services, job opportunities (include opportunity for the self-employment like mining and fishing) and the fertility of land for agriculture activities (Madulu,2002).

Normally, people are very selective when it comes to the establishment of settlements and economic activities. In most cases the most potential areas are utilized first before expanding to the marginal areas. The following population density map, Figure 5.1 was produced from given data using the Arc Map;

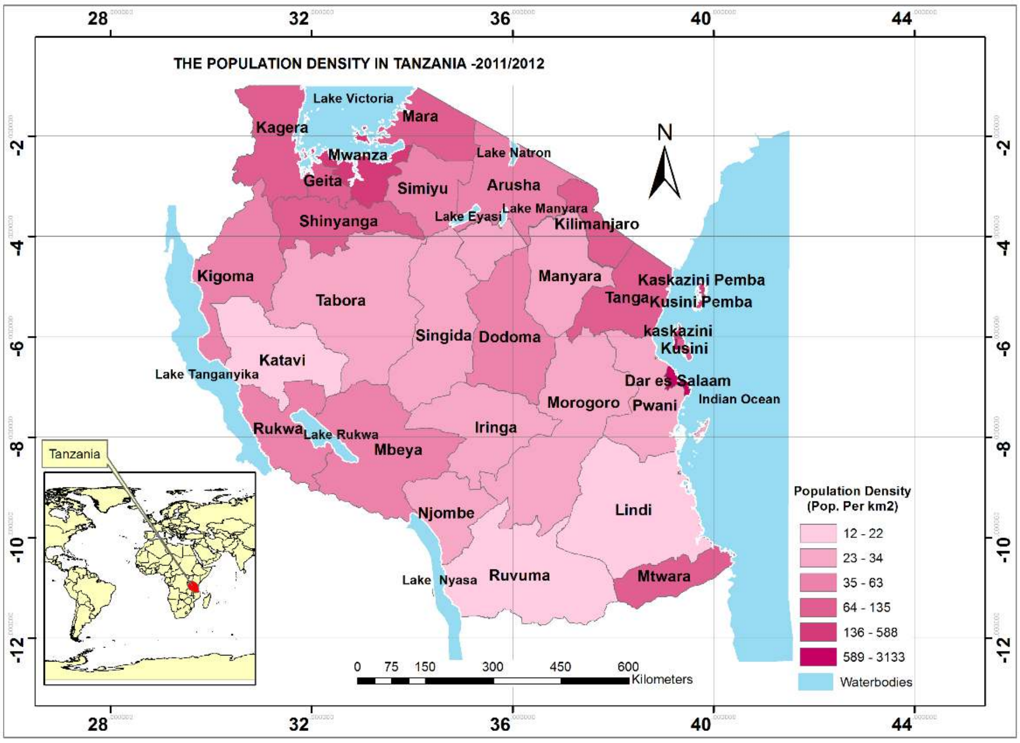


Figure 5.1 The spatial distribution for population density in Tanzania.

The spatial distribution of population density on Tanzania map showed that, the northern part, where the most populated regions are Dar-es- Salaam and Mwanza, is highly populated than the southern part.

There could be a number of reasons for this result, in this research, among of the reason which could be the cause of the northern part to be more populated is the existence of the natural resources, People tend to shift from their home to seek opportunity. This was proved when the natural resources such as a gold, iron and steel, mica, petrol and coal were spatially distributed on the same map the result showed that, most of them were located in south- west, east and northern part of Tanzania where minerals are available.

The other reasons could be the availability of good infrastructure, Health services and land fertility. Below is the map, Figure 5.2 from the given data which shows the natural resources with population density;

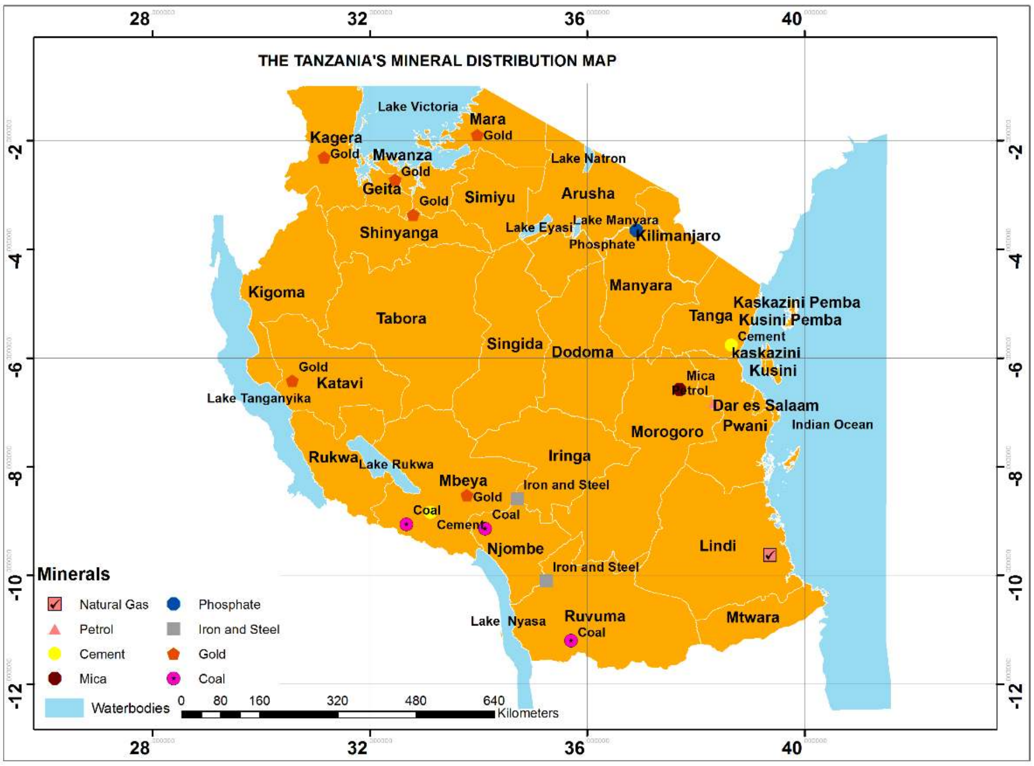


Figure 5.2 The minerals distribution in Tanzania

Infrastructure is among of the most important factor in accelerating the economic growth for any country. The government of Tanzania is also trying to connect all of the cities to the main roads so as to enable the flow of raw materials within the country. Since the eighty percent (80%) of the Tanzania economic depends on Agriculture, the government has to ensure also these connection is not just end across the cities but to extend the program until to the farmer's plantations. In this way they will be in a good position of transporting their products to the industries easily and safely for further processing or selling. Figure 5.3 below shows the major and other roads connected to the cities all over the country. Figure 5.4, shows the spatial distribution of rural economic status in Tanzania- 2011/2012. Figure 5.5 shows the distribution of urban economic status. The map, Figure 5.6 shows the spatial distribution for an altitude level in Tanzania regionally.

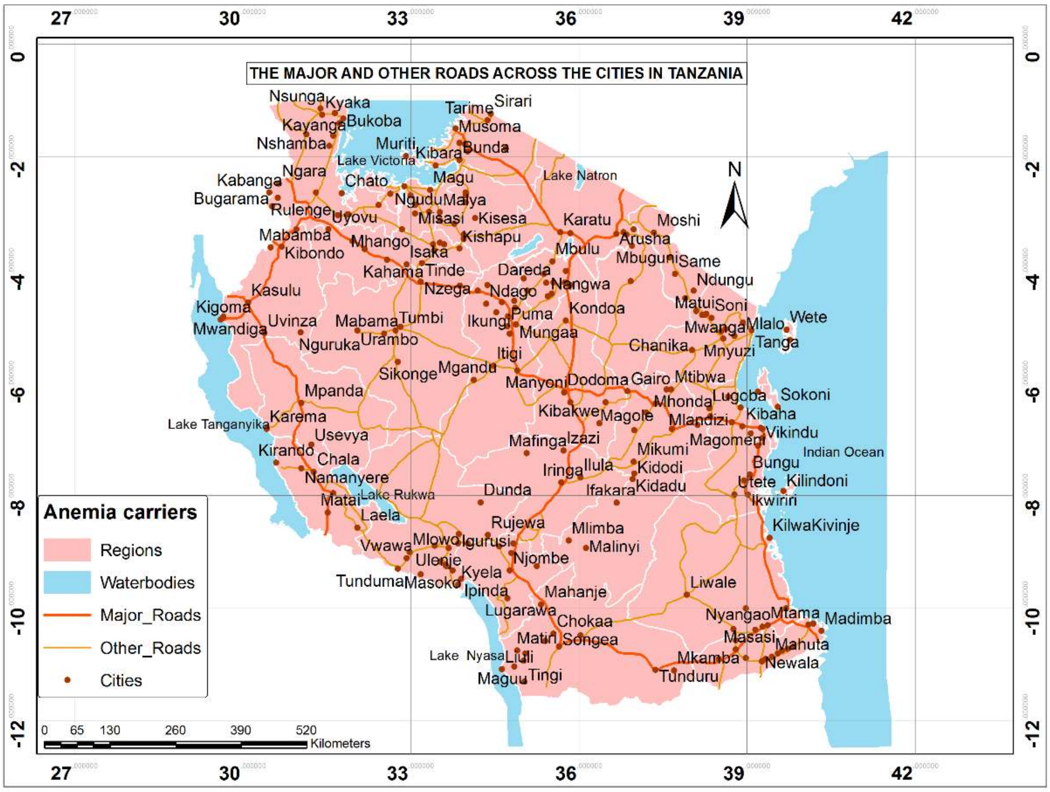


Figure 5.3 The major road and other roads in Tanzania.

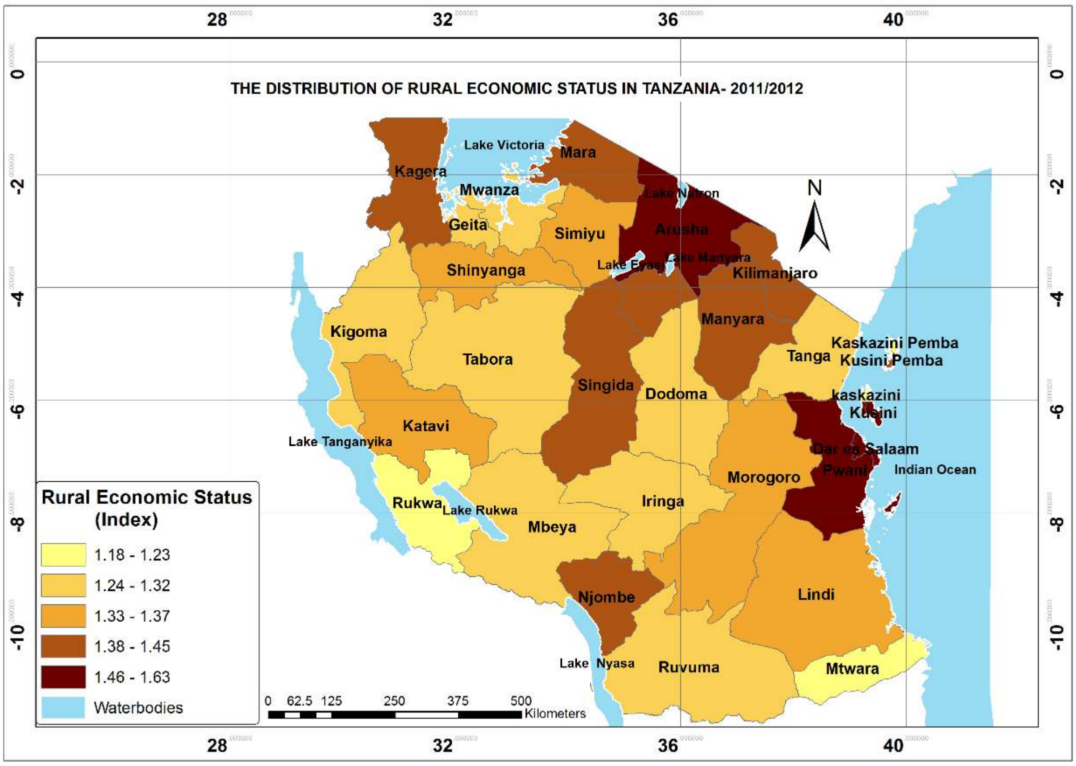


Figure 5.4 The spatial distribution of the rural economic status

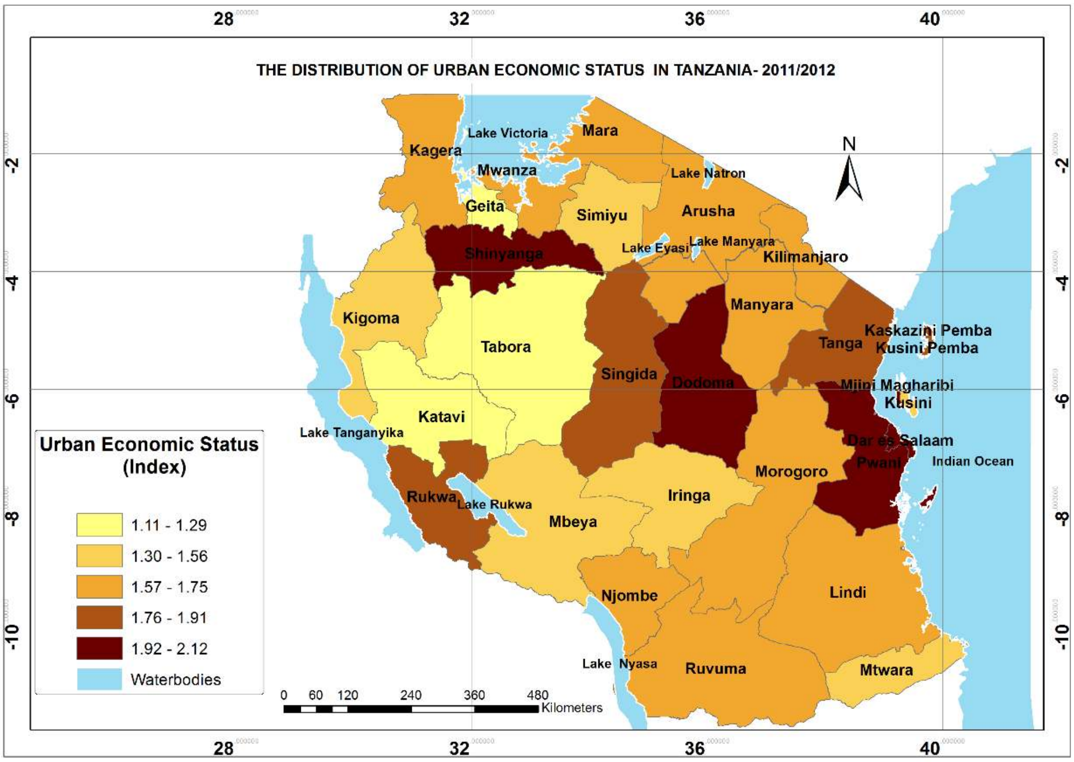
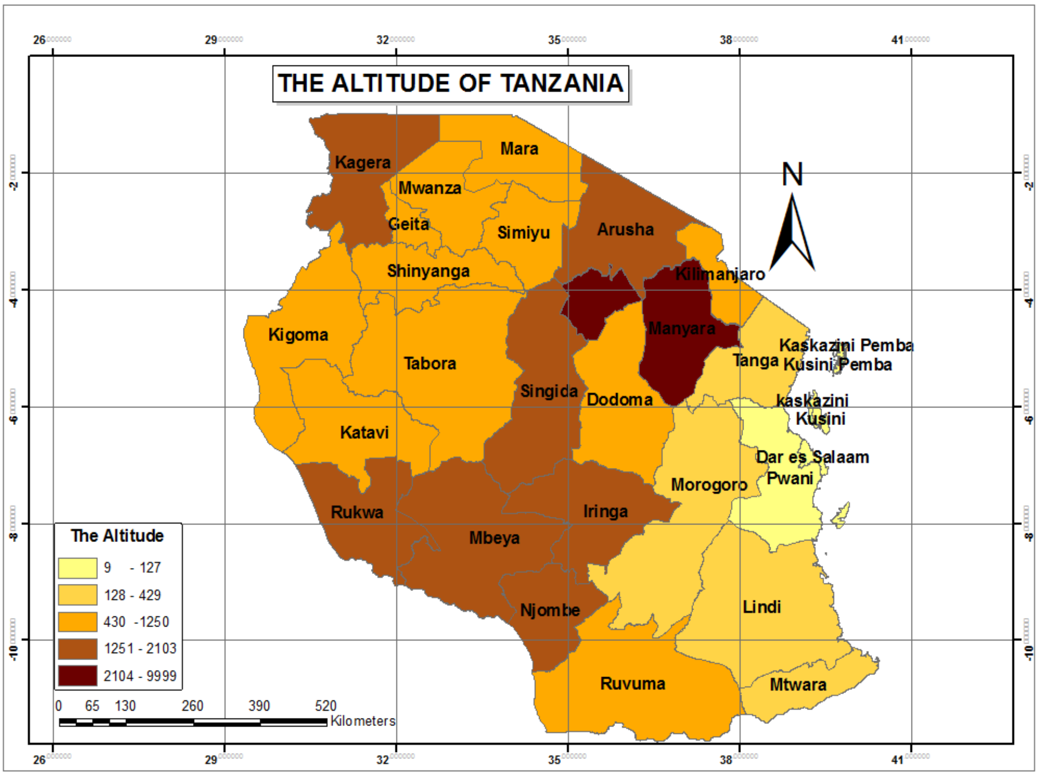


Figure 5.5 The spatial distribution for urban economic status.



The Figure 5.6 The spatial distribution for the altitude level in Tanzania regionally.

The general description of the datasets used in this thesis is described in Table 4.1, Moreover, apart from the health information generated from the datasets, the sex ration, average age and education of the population were calculated as the important factor in the demographic phenomena. The sex ration female to male (Female/Male) was found to be 1.067. By just looking the sex ration, it tells that, the number of female is more compared to the number of male. But, statistically, the female and male were found to have the percentage of 51.62% and 48.37% respectively where by the female population is greater than the male population by 3.25%.

Also the population was found to have the following education categories, 0.38%- no education, 50.65% - primary education, 10.66% - secondary education and 0.29% -high education. Since the Tanzanian population found to have a “youth population” (Average of 22age’s).the government must train them further in the courses of technical, technological and scientific knowledge so as to accelerate the development.

5.2 Analysis of Health Related Data

5.2.1 Malaria

Malaria is a major public Health concern for all Tanzanians, especially for pregnant women and children under age 5. The disease is a leading cause of morbidity and mortality among outpatient and inpatient admissions. It accounts for up to 40 percent of all outpatient attendance (MOHSW, 2006). Many parts of the country, including the uplands, report malaria transmission throughout the year, although it occurs most frequently during and after the raining season from April to May.

Malaria is caused by four species of plasmodia parasites that are transmitted by Anopheles mosquitoes. In Tanzania, *Plasmodium falciparum (P. Falciparum)* is the most common parasite. It causes severe malaria, and is fatal if not recognized promptly and properly managed. The most severe cases occur among persons who have not yet developed sufficient immunity to malaria through previous exposure. Children under age 5 are at highest risk, followed by pregnant women, because of their reduced natural immunity.

Pregnant women are four times as likely to experience the complications of malaria as non-pregnant women, and malaria is a major cause of pregnancy loss, low birth weight, and neonatal mortality. The map, (Figure 5.7), shows the spatial distribution of malaria for the children under five years old (<6 years old).

Malaria poses many societal and economic burdens in Tanzania, ranging from school absenteeism to low productivity in the workplace. In the short term, widespread malaria reduces agricultural production and other economic outputs. In this research, the rural and urban economic status was distributed differently on the maps because the economic indicators for the economic stability are always different from household to household or individual to individual in both urban and rural areas.

The distribution shows that some regions in which malaria is high have low economic status. In case of rural areas (Figure 5.4), both Mtwara, Rukwa, Geita, Ruvuma, Kigoma and Tabora regions were mostly affected with malaria and the reason could be their low level of economic status while other regions with a low level of malaria infection found to have a good economic status. Examples of these regions are Arusha, Kilimanjaro, Dar-es-salaam, Singida and Njombe.

For the urban areas, the distribution shows that, there is also close relationship between the economic status (Figure 5.5) and the rate of malaria infection (Figure 5.7), some of regions found with high infection of malaria observed to have a low level of economic status, examples of these regions are Tabora, Kigoma, Mtwara, Morogoro, Lindi and Ruvuma.

The vice versa is true for the areas with low infection, seems to have high economic status. Examples of these regions are Dodoma, Dar-es-salaam, Shinyanga, Singida and Tanga. The conclusion to that is, the regions with high economic status are more likely to afford the expenses associated with the prevention of malaria compared to the regions with low economic status.

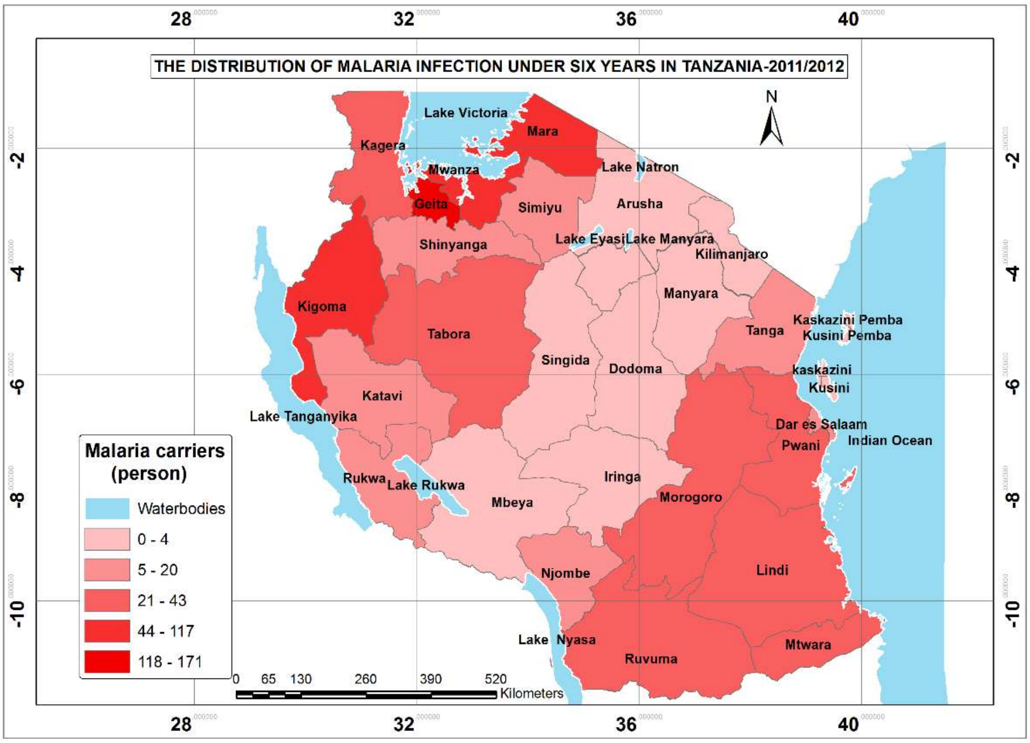


Figure 5.7 The spatial distribution of malaria infection children under six years.

Additionally, the cumulative effect in the long term may be a decrease in national economic capacity and development. The international Roll Back Malaria (RBM) Initiative works to reduce the malaria burden. The primary objective of RBM is to increase access to the most effective and affordable protective measures. These measures include use of Insecticide-Treated mosquito Nets (ITNs) for sleeping and increased coverage of prompt and effective treatment.

The RBM Initiative also promotes the use of Intermittent Preventive Treatment (IPT) of malaria among pregnant women in Tanzania, the recommendations of the RBM Initiative are implemented through the National Malaria Control Strategy (MOHSW, 2002). The National Malaria Control Strategy also includes other vector control measures such as Indoor Residual Spraying (IRS) and epidemic prevention and control.

The Government of Tanzania, primarily through the Ministry of Health and Social Welfare (MOHSW), is committed to the control and prevention of malaria. A considerable amount of the Health budget is allocated to address malaria and malaria-related illnesses. Household expenditures related to malaria are high and are mainly for treatment. These costs are expected to rise substantially because of the recent introduction of Artemisinin-based Combination Therapy (ACT). ACT is a response to the emerging resistance of malaria parasites to monotherapy antimalarial drugs, for example, Sulphadoxine Pyrimethamine (SP) and chloroquine, which used to be first-line antimalarial drugs in Tanzania.

The malaria component of the 2010 TDHS measures malaria prevention and treatment outcomes. At the household level, it includes coverage of malaria interventions, possession and use of ITNs, and indoor residual spraying (IRS). At the individual level, information is collected about prompt antimalarial treatment of children under age 5 with fever and use of IPT among pregnant women. Many of the indicators were assessed in the 2004-05 TDHS and the 2007-08 THMIS, which allows for trend analysis (NBS, 2005; TACAIDS et al., 2008).

The use of insecticide-treated mosquito nets (ITNs) is a primary Health intervention designed to reduce malaria transmission in Tanzania. Distribution of Long-Lasting

Insecticide Nets (LLINs) to children under 5 in Mainland started in late 2008 with a pilot project in Tanga. It was expanded region by region in May 2009.

It is anticipated that widespread use of ITNs will reduce mosquito density and biting intensity. ITNs are being promoted through three main channels:

- In the public sector as community-based projects.
- In public/private partnerships implemented by nongovernmental organizations directly in the community.
- In the private sector as social marketing initiatives, such as those assisted by Population Services International (PSI).Table 5.1 below shows the number of mosquito nets owned by households in Tanzania.

Table 5.1 The numbers of mosquito nets owned by Household

Number of Nets	HH Owned	Percentage (%)
0	660	6.57
1	1742	17.35
2	3094	30.82
3	2491	24.81
4	1173	11.68
5	476	4.74
6	340	3.39
7	64	0.64
TOTAL	**10040**	**100**

The spatial distribution of the LLIN's on a map shows that there is a positive output to the regions which use these LLN's. This can be shown on map below, Figure 5. 8 produced by given data which show the distribution of the LLN's .

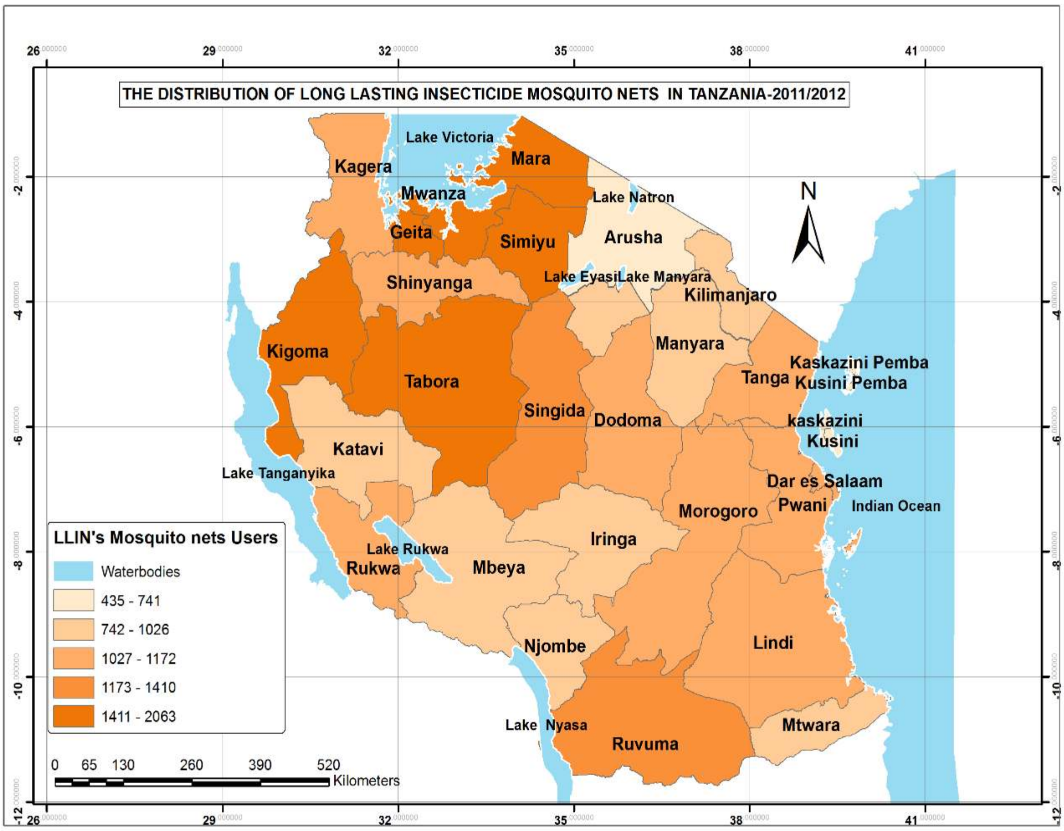


Figure 5.8 The distribution of long lasting insecticides nets in Tanzania

a) Spatial Statistic Analysis for Malaria in Tanzania

i. The Global Moran I Analysis for the Malaria Distribution

The spatial analysis tool from an Arc Map within the Arc GIS 10.2 was used to run the analysis, where by “the inverse distance squared” was used as the conceptualized spatial relationship. Finally the processing extent was defined before generating the final report.

The following observation was determined after analysis; The Global Moran Index, I = 0.335661, Z- Score = 3.327491 and the P- Value was 0.000876 (Table 5.2).This output will be used as the basis for the discussion below.

The followings below are the two reports which generated after running the analysis; Table 5.2 shows the Global Moran I index as a summary for the children under six years whose affected with malaria while Figure 5.9 show the Moran’s I spatial autocorrelation for the same children while the Table 5.3 shows the input dataset information where by "SMALE" is a (sum of malaria carriers) and "local_malaria" represents point features for each region. The Global Moran index I summary for the malaria infected children under five years old.

Table 5.2 The global Moran I for the children below six years old who's affected by Malaria.

Global Moran's I Summary	
Moran's Index:	0.335661
Expected Index:	-0.034483
Variance:	0.012374
z-score:	3.327491
p-value:	0.000876

Table 5.3 The input data information for the malaria affected children under six years

Dataset Information	
Input Feature Class:	Local_malaria
Input Field:	SMALE
Conceptualization:	INVERSE_DISTANCE
Distance Method:	EUCLIDEAN
Row Standardization:	False
Distance Threshold:	2.0652 Degrees
Weights Matrix File:	None
Selection Set:	False

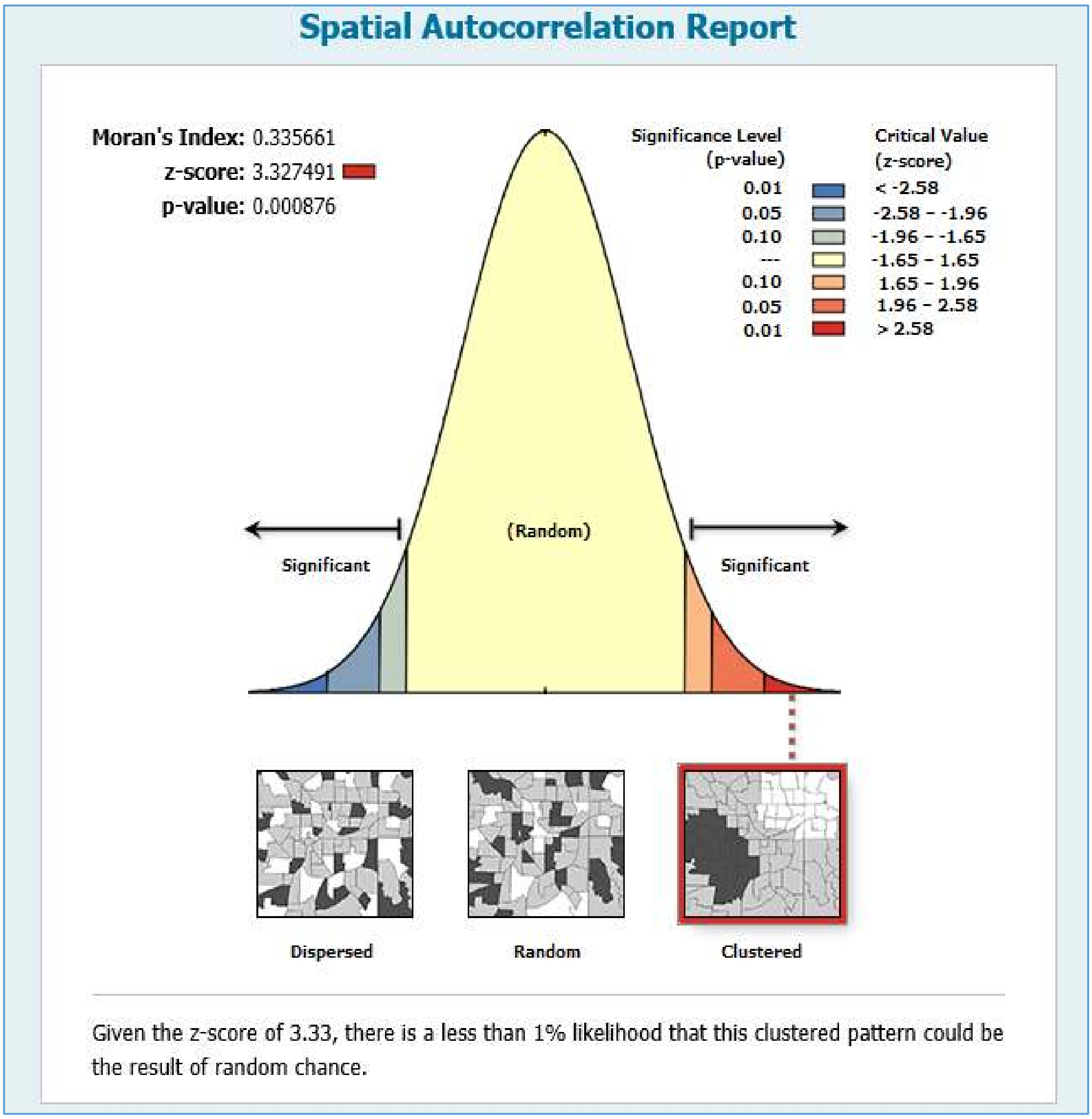


Figure 5.9 The spatial autocorrelation for the children affected with malaria.

The analysis of result shows automatically that, the distribution is clustered. Alternatively, if the test is taken from mathematical interpretation still it shows the same outcome is clustered.

That's means,

$$I \;>\; \frac{-1}{n-1} \;=\; 0.335 \;>\; 0.034$$

Generally, When P- value is very small it means that, the observed spatial pattern of children affected with malaria does not represent the theoretical pattern of randomness concept. Therefore, the P – Value is statistically significant and the low or high numbers of spatial distribution of malaria affected people are not randomly but they are spatially clustered. The result was extracted with the 99% level of confident level which insure the accuracy of the output.

It also show that, the spatial distribution for the children under five year whose affected with malaria in all of the 30 regions has significant similar with neighboring region with the perfect spatial agglomeration structure, the infection rate of malaria has almost similar with the adjacent Regions in property, that means, the high-rate infection regions aggregates with high ones, while the low rate infection regions aggregates with low infected region also, the conclusion could be the following, the rate of malaria infection in the region during the year has aggregated phenomena while its spatial relationship is that, the regions with high rate of malaria infections tend to be neighboring to each other and the lower regions inclined to be adjacent with the same character.

ii. The Local Moran I Analysis for the Malaria Distribution

The local Moran I analysis, usually perform analysis of the specified field from the input data future class. The tool identifies the spatial cluster with high or low values as well as spatial outlier. The tool calculates a local Moran's I value, a z-score, a p-value, and a code representing the cluster type for each statistically significant feature. The z-scores and p-values represent the statistical significance of the computed index values (Zhanga & Lin, 2006).

Below is the Table 5.4 which shows the individual output for Z-SCORE and P-value generated by the Local Moran' I analysis tool.

From the Table 5.7, it shows that twenty two (22) regions have the positive Local Moran' I Index values which show that, the region with high rate of malaria infection has surrounded by the regions with high infections too or, if the region has a low infection rate, the surrounded regions have low infection rate also. Therefore, the spatial distribution for the malaria infected children is spatially clustered.

Also; only eight (8) regions shows the negative Local Moran' I index which means that, they have no similar values with their neighboring regions in the sense that the rate of malaria infections for the surrounding regions are quite different to each other and therefore they are called outlier features. These regions are, Kagera, Kigoma, Katavi, Tabora, Shinyanga, Simiyu, Morogoro and Ruvuma.

Table 5.4 The P-Values, Z-Scores and Moran I generated by the Local Moran's I.

Anselin Local Moran I Values

Region_Codes	LMiIndex	LMiZScore	LMiPValue
1	0.220609	0.562545	0.573745
2	0.817083	0.730811	0.464894
3	0.652744	0.712291	0.476285
4	0.448466	0.478516	0.632283
5	-0.015374	0.088644	0.929364
6	0.098918	0.172322	0.863184
7	1.067567	0.729518	0.465685
8	0.014915	0.074495	0.940617
9	0.014915	0.074495	0.940617
10	-0.016978	-0.042484	0.966113
11	0.214745	0.562065	0.574072
12	0.210744	0.461422	0.644496
13	0.165924	0.343053	0.731558
14	-0.014651	0.030244	0.975873
15	0.078137	0.306629	0.759126
16	-0.049755	-0.204735	0.837779
17	-0.449001	-0.319986	0.748979
18	-0.03665	0.006284	0.994986
19	5.62911	4.877983	0.000001
20	0.795283	1.049599	0.293903
21	0.654767	0.705693	0.480379
22	0.196051	0.430674	0.666706
23	-0.049755	-0.204735	0.837779
24	-0.737894	-0.621968	0.533963
25	4.016409	3.445749	0.000569
26	2.081839	1.174041	0.240379
27	1.926505	1.162755	0.244929
28	1.89758	1.120449	0.262522
29	1.598115	1.157162	0.247206
30	1.864052	1.177379	0.239044

b) The Hot Spots Analysis (Getis Gi*) For Malaria

The Hot Spots Analysis usually produces a G-i- star statistic for each feature. It is abbreviated by Gi*.It also generate the Z-values and P- values where by its resultant tells if the features with high or low values are clustered spatially. Below is a table 5.5 generate by Arc Map 10.2 which show the Z-scores and P- Values obtained after running the Hot Spots Analysis.

Table 5.5 The Z and P values generated by hot spots analysis

Hot Spot Analysis (Getis & Ord Gi*)

Region_Codes	GiZScore	GiPValue	Gi_Bin
1	-0.959753	0.33718	0
2	-1.159387	0.246299	0
3	-1.159387	0.246299	0
4	-1.740163	0.08183	-1
5	-1.295334	0.195205	0
6	-1.117158	0.263927	0
7	-1.550726	0.120967	0
8	0.379598	0.704244	0
9	0.379598	0.704244	0
10	-0.184162	0.853887	0
11	-0.971885	0.331108	0
12	-0.862509	0.388408	0
13	-0.75468	0.450441	0
14	-0.456255	0.648207	0
15	-0.729129	0.465923	0
16	1.018526	0.308428	0
17	2.728613	0.00636	3
18	3.01253	0.002591	3
19	3.147688	0.001646	3
20	1.701631	0.088825	1
21	-1.125228	0.260493	0
22	-0.808068	0.419052	0
23	1.018526	0.308428	0
24	2.866435	0.004151	3
25	2.226307	0.025994	2
26	-1.550726	0.120967	0
27	-1.550726	0.120967	0
28	-1.550726	0.120967	0
29	-1.66022	0.09687	-1
30	-1.66022	0.09687	-1

c) Results of The Analysis

To be more consistent, the use of malaria sickness density as the input dataset could be more realistic for the distribution of malaria infection throughout the entire region because it easily even to control the place with an expected results. The formula used to calculate sickness density is as given below;

$$\text{Sickness Density (Sd)} = \frac{Number\ of\ Sickness\ people\ in\ the\ Region(Ns)}{Population\ of\ the\ respected\ Region(Pr)} \qquad (5.1)$$

But, since each region has different in population size, in fact the population for each region is more large compared to the total number of infected people, so it is significantly to consider the total number of infections for each disease during the analysis.

From the table 5.4 above, it shows that, only ten (10) regions are spatially statistically significant and have the high intense of clustering of high values (HOT SPOTS). Since these regions have the positive Z –Scores; it means that, they have a spatial dependence to the most affected people (high values) in the sense that, the regions with high rate of malaria infections has a high spatial clusters.

On the other hand, the remained regions are not spatially statistically significant. They have a negative values which indicate that, the regions have spatial cluster of low value (COLD SPOT) which means the rate of malaria infection within these regions are low. The highness or lowness of malaria depend much on the people knowledge about the disease (Saxena et al, 2011)

The people living in the regions with low infection rate could have been more knowledgeable about different methods of combating the malaria compared to the other regions. They might be aware on the use of insecticides mosquito bed net and destroying the stagnant water which is the most common way of combating and preventing the spread of malaria. The map, figure 5.9 below generated from the Hot Spots analysis output values (Z- values) shows the malaria sickness density map in Tanzania. For clear visualization the Figure 5.10 shows the hot and cold interpolated surface for malariaavi infections Regions (Tatem et al, 2012).

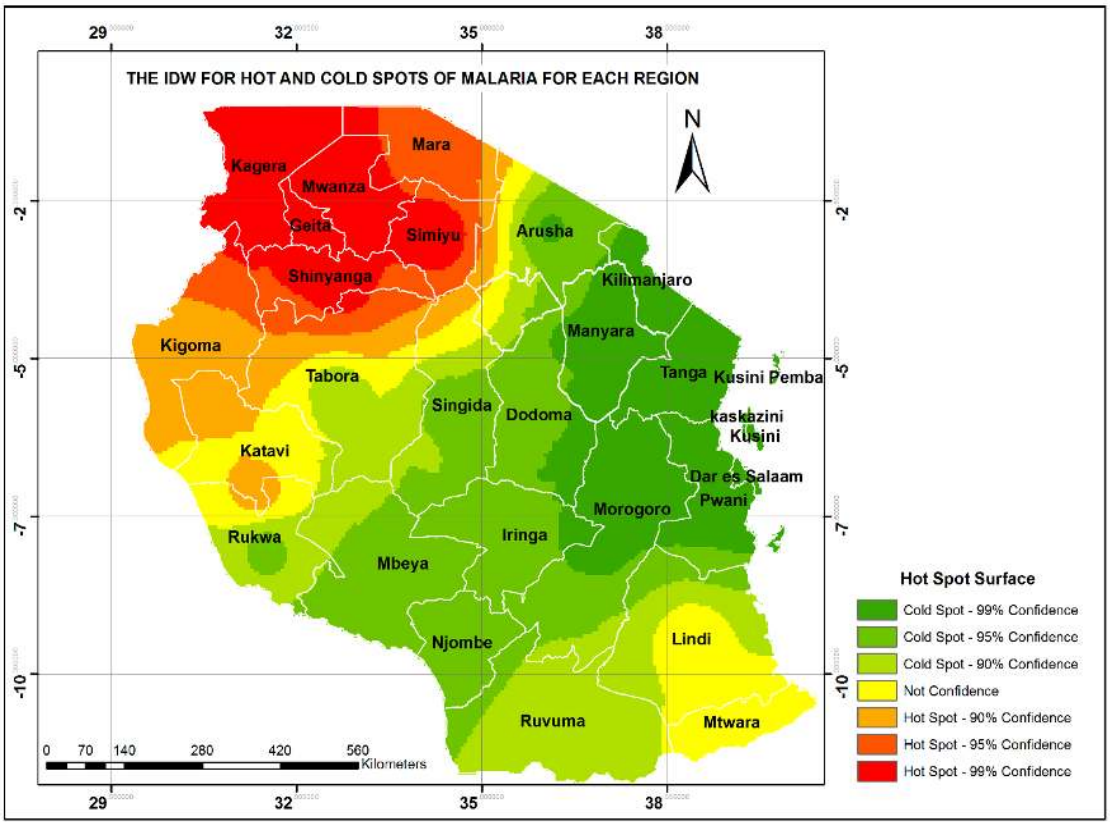


Figure 5.10 The IDW for the malaria hot and cold spots.

5.2.2 Anemia

Anemia is a serious Health problem that usually results from poor nutrition, infection, or chronic disease. When the number of red blood cells or concentrations of hemoglobin are low a person is said to have anemia. Hemoglobin is a protein (metalloprotein) inside the red blood cell that contains iron and transports oxygen.

Anemia is the most common disorder of the blood. Approximately 3.5 million Americans are affected by it. It is much more common in developing countries, especially in very poor areas where people suffer from malnutrition. In many parts of Africa severe anemia is also caused by Malaria. As many people who become infected with Malaria already have pre-existing anemia, most commonly due to malnutrition and helminthiasis (a disease caused by a parasitic worm), the problem is compounded.

The table 5.6 below shows the rate of anemia infection for the children under five years in Tanzania;

Table 5.6 The rate of anemia infection for the children below six ages

Codes	Rate	Number of children affected (Age≤5)
1	Severe	142
2	Moderate	2118
3	Mild	2156
4	Not Anemic	3377
Total number of anemia infected children		7793

But, also the rate of infection differs from age to age as shown in the table 5.7 below;

Table 5.7 The rate of anemia infection with age groups.

Ages (Years)	Number of Children Affected (Age≤5)
< 1	704
1	1297
2	966
3	792
4	634
5	23

For more visualization the anemia infection data in Tanzania can be spatially distributed on a map as shown in Figure 5.11. The map was generated from the output data analysis using Arc Map as a tool,

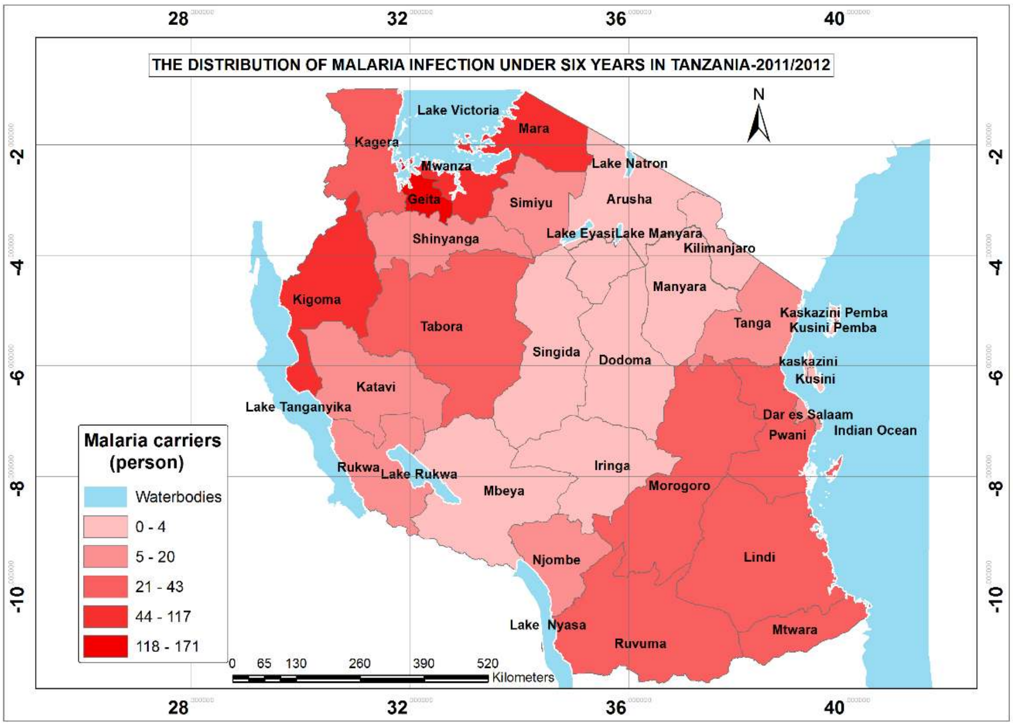


Figure 5.11 The spatial distribution of anemia infection in Tanzania

The spatial distribution on a map shows that, the level of anemia is different from one region to another, the most affected regions are, Geita, Simiyu, Mara, Tabora, Mwanza and Kigoma, while the least affected regions are, Iringa, and Kilimanjaro**.**

There is good evidence that the anemia seen in this distribution map, Figure 5.10 is caused largely by malaria, though the other cause could be by iron deficiency, however improving malaria control and may be the iron supplementation programs could have dramatic anti-anemia effects. Although, the values in the field which show the iron deficiency was not sufficiency for distribution in the given data for this research, but the malaria distribution on the Figure 5.3 shows there is closest relationship between anemia and malaria in the sense that, some of the regions with high infection of anemia have also high infection of Malaria. Among of these regions are Geita, Mara, Mwanza and Kigoma.

The other regions have a partial relationship, and this can be caused by the different sample taken for measurement. Otherwise the two-distribution has a positive relationship.

One approach to preventing anemia in children is by reducing anemia and improving iron status in pregnancy and pre-pregnancy. Iron supplementation and intermittent preventive malaria treatment are recommended for women attending antenatal clinics, although the transient nature of pregnancy may result in incomplete correction of anemia or iron deficiency before delivery. Furthermore, women tend to present to clinic relatively late in pregnancy and there is evidence that anemia in the first trimester reduces fetal growth and birth weight.

The global guidelines for iron supplementation in childhood may need to be refined if the benefits from iron prophylaxis are to be optimized. With the exception of children of low birth weight, for whom iron supplements are recommended from 2 months of age, WHO/UNICEF/INACG recommend that supplementation should commence only at 6 months; however, by this age anemia has already affected a large proportion of children.

Anemia was less of a problem at altitudes greater than 1250 m, which probably reflects the reduced burden of malaria at such heights. Interestingly, this effect was more marked for life-threatening anemia than for less severe anemia. This suggests that malaria may be a relatively important cause of the most severe anemia but that other causes of less severe anemia, such as iron deficiency, are still prevalent at altitude. An element of altitude-induced secondary polycythemia may also be present, although this is unlikely at the altitudes studied (Figure 5.6).

a) The Global Moran I Analysis for the Anemia Distribution

The total number of anemia sickness was calculated so as to find the effect of infection children under five years old per respective region's population. After that, sum of infections was used as an input dataset in Moran's I tool so as to determine the existence of patterns. Here is the Table 5.8 which show the Global Moran index I for anemia infected children under five years old while Table 5.9 shows the detail of the input dataset information.

Table 5.8 Global Moran I analysis for the anemia carriers (under six ages)

Global Moran's I Summary	
Moran's Index:	0.349909
Expected Index:	-0.034483
Variance:	0.013063
z-score:	3.363193
p-value:	0.000770

Table 5.9 The input dataset used by Global Moran I analysis for anemia carriers.

Dataset Information	
Input Feature Class:	POINT_AMH_LAST_point
Input Field:	SUM_A
Conceptualization:	INVERSE_DISTANCE
Distance Method:	EUCLIDEAN
Row Standardization:	False
Distance Threshold:	2.0652 Degrees
Weights Matrix File:	None
Selection Set:	False

The summary of distribution for anemia carriers is clearly shown in the Figure 5.12 below.

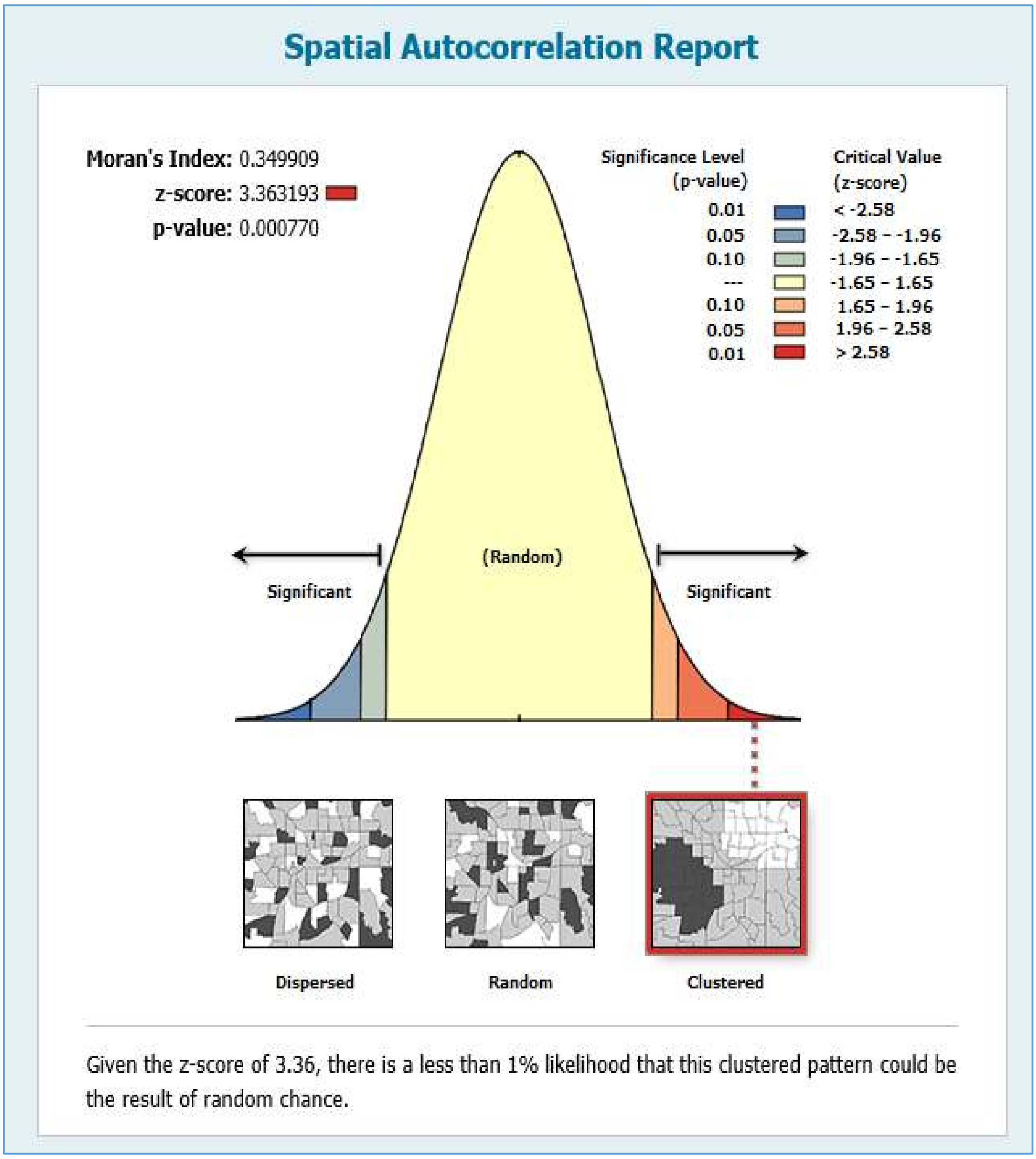


Figure 5.12 The out of Global Moran I for anemia sickness density.

From the above, it show that, there is a powerfully strong positive spatial autocorrelation .This means that the region in which the rate of anemia infection is high is surrounded also by the region with the high rate of infection and vice versa. This could be caused but many factors including, the parents might not aware on which kind of food (rich in iron) should feed their children, absence of folic acid and vitamin B12 in the children meal, since Tanzania is among of the tropical Country the children under five years is always attacked by different diseases, example malaria, this attack could reduce the rate of production of red blood cell resulting to chronic anemia infection (Magalhaes & Clements, 2011).

b) The Local Moran I Analysis for the Anemia Distribution

The Local Moran's I tool analyze the existence of spatial clusters independently. The index also can identify the local instability in the form of typical localizations and spatial outlier. Aselin (1995) proposed a local Moran index or local indicator of spatial association (LISA) to capture local pockets of instability or local clusters. The local Moran for a region "i" measures the association between a value at "I" and the value of its nearby area. The tool was used to analyze the sickness density data for anemia infection people, the Table 5.10 below shows the output Z- Score and P-value after running the Local Moran I analysis. The nineteen (19) and nineteen (25) Region codes, which is Mwanza and Geita respectively shows that, they have a high - high values. That means the rate of anemia infection in these regions is high.

By identifying that, the government shall be in a good position of intervene and provide an appropriate measures of monitoring the situation.

Table 5.10 The Z-scores and P- values for anemia carriers after running the Local Moran I analysis

Anselin Local Morans I			
Region_Codes	LMiIndex	LMiZScore	LMiPValue
1	0.35585	0.84069	0.400522
2	0.79991	0.69535	0.486836
3	0.87688	0.909968	0.362839
4	-0.046629	0.083586	0.933386
5	0.107384	0.231295	0.817086
6	0.368783	0.369367	0.711854
7	0.28726	0.270374	0.786873
8	0.28351	0.678647	0.497362
9	0.28351	0.678647	0.497362
10	0.114611	0.576218	0.564468
11	0.828851	1.948245	0.051386
12	0.217837	0.460904	0.644867
13	0.092655	0.220443	0.825526
14	0.078727	0.194061	0.846128
15	-0.0926	-0.267941	0.788745
16	-0.148226	-0.667117	0.504697
17	0.145147	0.227922	0.819707
18	0.865479	1.121149	0.262224
19	6.234466	5.253421	0
20	1.913532	2.364942	0.018033
21	0.261492	0.310953	0.755836
22	0.89428	1.676825	0.093577
23	-0.148226	-0.667117	0.504697
24	2.84468	2.731797	0.006299
25	6.387185	5.285178	0
26	0.025195	0.126412	0.899406
27	0.306243	0.276177	0.782412
28	0.31266	0.277244	0.781593
29	0.063407	0.135536	0.892188
30	-0.230603	-0.03409	0.972806

c) The IDW derived from the Anemia Local Moran's I Values.

For better visual, the inverse distance weigh (IDW) was used to connect the infection rate of anemia to rendering point into a raster. The IDW was used to visual the surface because it always direct interpolates the values to determine the smoothness of the surface. The map, Figure 5.13 shows the interpolated area with the high and low infection rate of anemia infections.

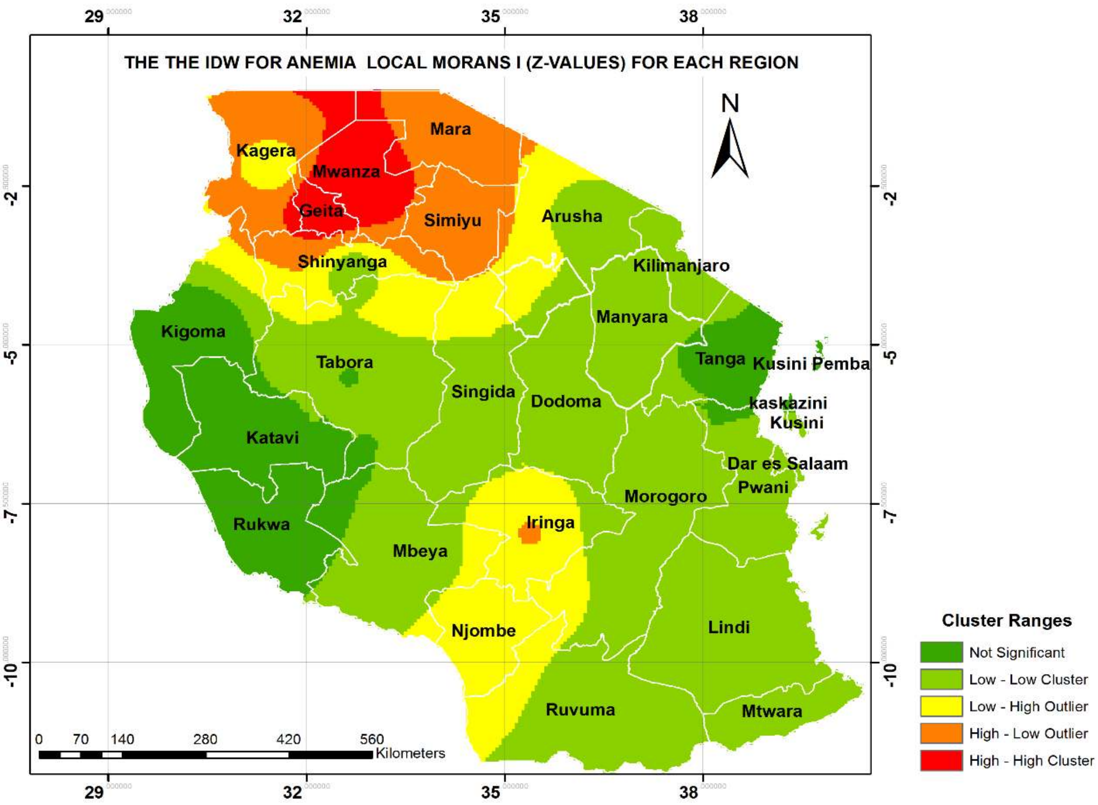


Figure 5.13 The IDW for anemia infection in Tanzania.

d) The Hot Spots Analysis

The rate of anemia infected people for each regions were used as input to run the Hot Spot analysis, the results was found to be highly spatially statistically significant in the northern regions. This means that, the rate of anemia infection people in the northern part is high compared to the other regions. The reasons behind could be the existence of more number of malaria infection people.

The highest frequency of sickle cell disease word wide is found in malaria regions of equatorial in Africa "That's means the regions with high infection of malaria is more likely to have high rate of anemia infection also (Anthro,2012). The above relationship was found to be matches when the both malaria and anemia infection were spatially distributed on map, the regions with high infection rate of malaria were also found to have a high rate of anemia infection. The following report, Table 5.11 shows the Z-scores and P-values for anemia density which generated by the Hot spots analysis.

e) The IDW derived from the Anemia Hot Spots values.

For better visualization the rate of anemia infection was interpolated using the IDW, this interpolation helps to visualize the area with high rate of anemia (Hot Spots) closely so that the government may be on a good position to set the immediate reaction and policy implementation to rescue the regions. The surface with high (Hot Spots) and low (Cold Spots) rate infections for anemia was described in Figure 5.14 below.

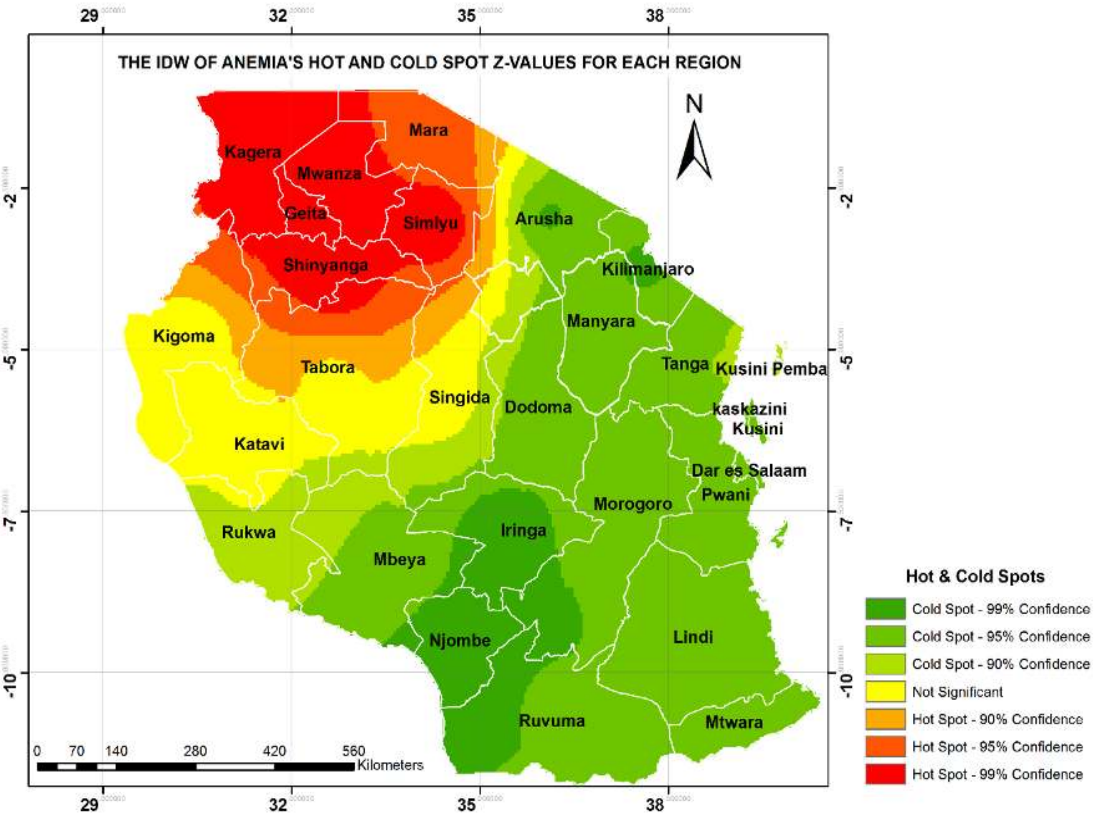


Figure 5.14 The interpolated surface by IDW for anemia sickness density in Tanzania

Table 5.11 The results for hot spots analysis surface for anemia disease

Hot Spot Analysis (Getis - Ord Gi*)

Region_Codes	GiZScore	GiPValue	Gi_Bin
1	-1.171845	0.241259	0
2	-1.31432	0.188739	0
3	-1.31432	0.188739	0
4	-0.653206	0.513624	0
5	-0.872605	0.382878	0
6	-0.66943	0.503221	0
7	-0.592982	0.553193	0
8	-1.118206	0.263479	0
9	-1.118206	0.263479	0
10	-1.063001	0.287781	0
11	-1.911043	0.055999	-1
12	-0.717597	0.473006	0
13	0.786218	0.43174	0
14	0.651805	0.514527	0
15	-0.175122	0.860984	0
16	0.330923	0.740703	0
17	3.611957	0.000304	3
18	3.130499	0.001745	3
19	3.45927	0.000542	3
20	1.994634	0.046083	2
21	-1.12796	0.259337	0
22	-2.012413	0.044176	-2
23	0.330923	0.740703	0
24	3.421975	0.000622	3
25	3.074051	0.002112	3
26	-0.592982	0.553193	0
27	-0.592982	0.553193	0
28	-0.592982	0.553193	0
29	-0.381834	0.702585	0
30	-0.381834	0.702585	0

5.2.3 HIV

HIV is a lot like other viruses, including those that cause the "flu" or the common cold. But there is an important difference – over time, the immune_system can clear most viruses out of the body. That isn't the case with HIV AIDS – the human immune system can't seem to get rid of it. That means, when a person is affected with HIV is for life. The map, Figure 5.15 shows the spatial distribution of HIV infections in each region in Tanzania

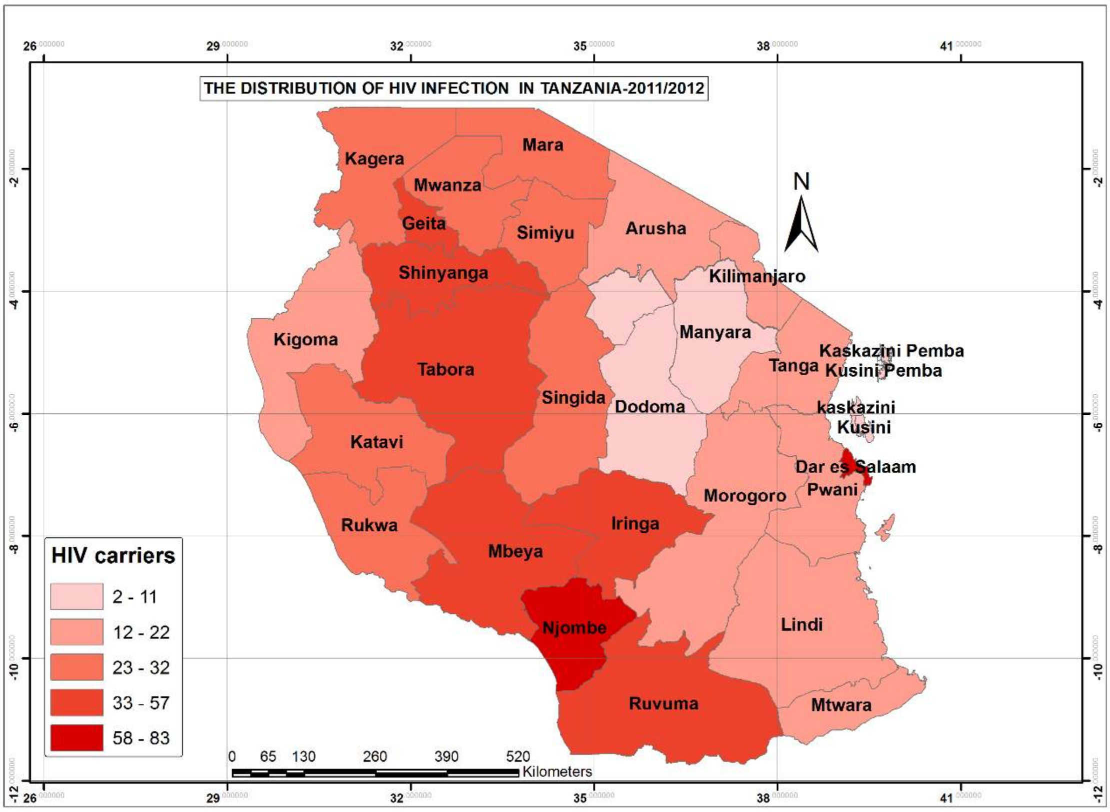


Figure 5.15 The spatial distribution of HIV for each region in Tanzania

a) The Global Moran‘s I

The total number for HIV carriers for each region was used as the input feature to perform the global autocorrelation (Moran ‘I) so as to find out the existence of clustering distribution and its type. The Table 5.12 below shows global Moran I for

HIV infections while Table 5.13 shows the summary of input dataset information used during analysis. Then the Figure 5.16 shows output of its spatial auto correlation.

Table 5.12 The output results for HIV infection after running global Moran I.

Global Moran's I Summary	
Moran's Index:	0.226503
Expected Index:	-0.034483
Variance:	0.012663
z-score:	2.319226
p-value:	0.020383

Table 5.13 The input dataset information for HIV carriers used by Global Moran I.

Dataset Information	
Input Feature Class:	Regions
Input Field:	SUHIV
Conceptualization:	INVERSE_DISTANCE
Distance Method:	EUCLIDEAN
Row Standardization:	False
Distance Threshold:	2.0652 Degrees
Weights Matrix File:	None
Selection Set:	False

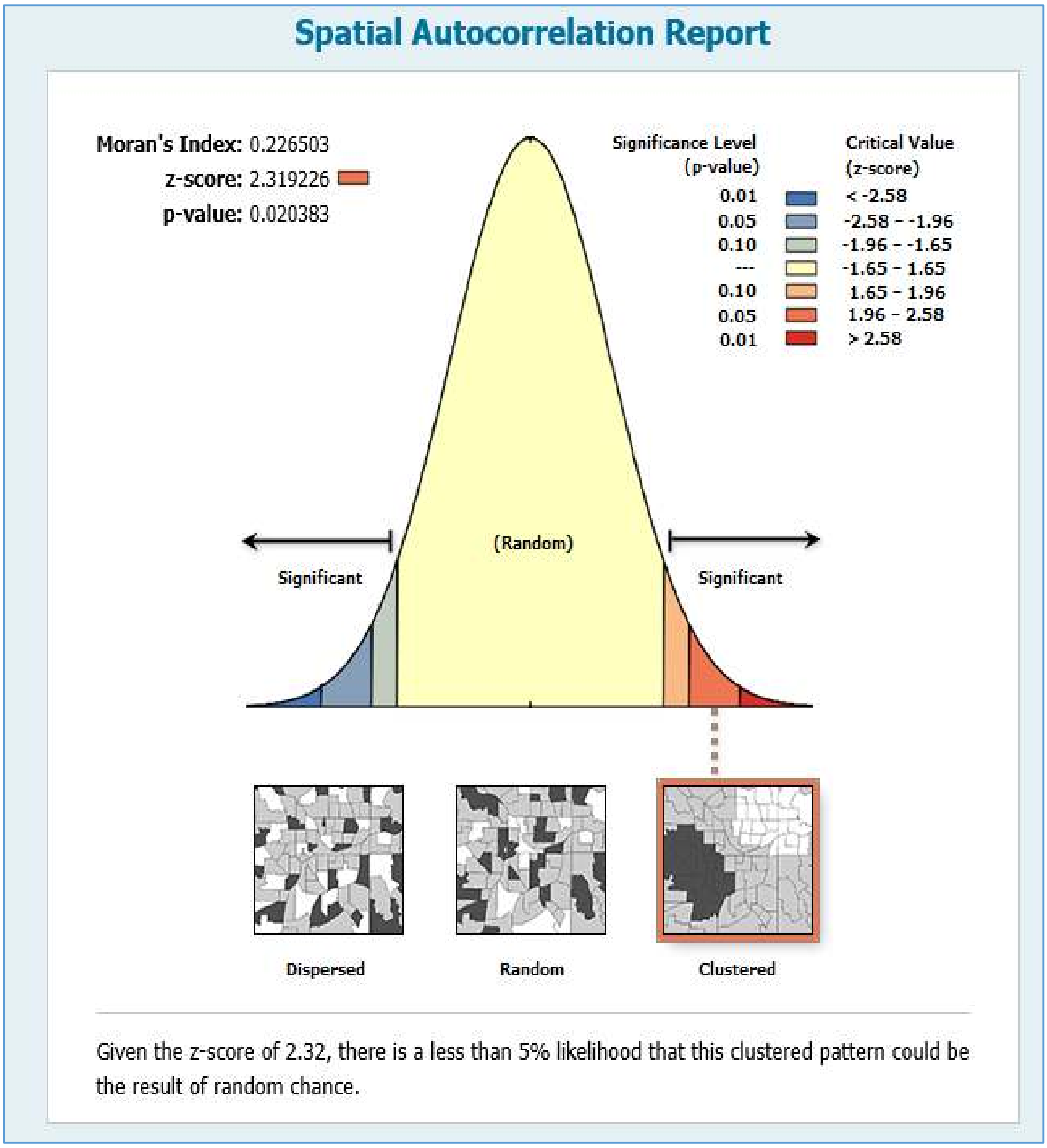


Figure 5.16 The global Moran I distribution for HIV infections

From the Figure 5.16, the results showed that, the calculated P-Values is 0.02 and the uncorrected P-Value is 0.05 at the 95% level of confidence. Therefore, the calculated P-Value is less than the uncorrected P-Value which indicate that the spatial distribution for the HIV infection people across the regions is statistically significant, that is to say, the infections is not the results of the random spatial process (The infection observed to be clustered distributed).

b) The Local Moran's I

In this analysis, the spatial cluster for HIV infection people in each region is calculate where by the both local Moran 'I values, Z-scores and P-Values was also identified. From the Table 5.14, only five (5) regions with the region codes (1, 6, 7, 13 and 18) have the negative Local Moran I which means that their neighboring regions have dissimilar rate of HIV infections. The rest have positive values which indicate that the neighboring regions have the same values whether is high or low infection rate (Buehler, et al, 2014; Chimoyi & Musenge, 2011).

On the other hand the Table shows the region with the cluster/outlier type (COType) of HIV infection rate which distinguishes between a statistically significant cluster of high values (HH), cluster of low values (LL), outlier in which a high value is surrounded primarily by low values (HL), and outlier in which a low value is surrounded primarily by high values (LH).

The Mbeya and Njombe regions were observed with a statistically significant cluster of high values (HH) while the Kaskazini Pemba and Kusini Pemba observed with a cluster of low values (LL).Only one region was an outlier in the sense that, its high infection rate of HIV infections is surrounded primarily by the regions with low infection rate (Qian et al, 2014)

This analysis helps to identify the region with high and low infection rate. The identification could be very helpfully for the government to apply any necessary measure to rescue the patients and provide the immediate service in time. The Table 5.14 below shows the values of Local Moran 'I values, Z-Scores and P-Values as the output of Local Moran's I analysis.

Table 5.14 The Z-scores and P-values as the HIV output for the Local Moran I analysis

Anselin Local Morans I			
SOURCE_ID	LMiIndex	LMiZScore	LMiPValue
1	-0.301508	-0.555795	0.578351
2	0.399365	0.352818	0.724225
3	0.232373	0.263744	0.791977
4	2.057772	1.65493	0.097939
5	0.284633	0.420171	0.67436
6	-0.040085	0.061326	0.951099
7	-12.334034	-6.590603	0
8	0.086896	0.220339	0.825607
9	0.086896	0.220339	0.825607
10	0.430328	1.931553	0.053415
11	0.644618	1.437673	0.150527
12	1.777906	3.175742	0.001495
13	-0.103167	-0.078018	0.937814
14	0.227144	0.429127	0.667831
15	0.022124	0.104873	0.916477
16	0.022762	0.137863	0.890349
17	0.686917	0.685656	0.49293
18	-0.102851	-0.071239	0.943207
19	0.002347	0.082106	0.934562
20	0.052996	0.118679	0.90553
21	0.585073	0.583441	0.559597
22	3.250928	5.588447	0
23	0.022762	0.137863	0.890349
24	0.09073	0.157584	0.874784
25	0.289539	0.291559	0.770624
26	3.425397	1.74219	0.081475
27	2.680106	1.474769	0.140275
28	2.662978	1.432201	0.152086
29	4.194126	2.672351	0.007532
30	4.614403	2.572148	0.010107

c) The Hot Spots Analysis (Getis & Ord Gi*) For HIV

The HIV carriers were further analyzed to determine the Hot spots (Most affected Regions with HIV) and the Cold Spots(Less affected Regions with HIV).The output generates the new feature with the Z-Score values, P-Values and confidence level Bin_Gi. The table 5.15 shows the output results for HIV Z-scores and P- values after running the Hot Spot Analysis.

Table 5.15 shows the Z-scores values and P-values for HIV after running the Hot spots analysis

Hot Spot Analysis (Getis - Gi*)

Region_Codes	GiZScore	GiPValue	Gi_Bin
1	-0.021763	0.982637	0
2	-0.879204	0.379291	0
3	-0.879204	0.379291	0
4	-2.076328	0.037864	-2
5	-1.210921	0.225926	0
6	-0.923656	0.355666	0
7	-1.819858	0.068781	-1
8	-0.613678	0.539428	0
9	-0.613678	0.539428	0
10	2.39768	0.016499	2
11	1.909211	0.056235	1
12	2.421368	0.015462	2
13	0.347383	0.728303	0
14	0.856442	0.391754	0
15	1.028881	0.303536	0
16	-0.613678	0.539428	0
17	1.114711	0.264974	0
18	0.458097	0.646883	0
19	0.857864	0.390967	0
20	0.606003	0.544513	0
21	-1.126629	0.259899	0
22	3.217505	0.001293	3
23	-0.613678	0.539428	0
24	0.836524	0.40286	0
25	0.828376	0.407458	0
26	-1.819858	0.068781	-1
27	-1.819858	0.068781	-1
28	-1.819858	0.068781	-1
29	-1.670384	0.094843	-1
30	-1.670384	0.094843	-1

Table 5.15 shows that, there is a high infection rate for HIV disease for six regions. The rate of infection is statistically significant with the different level of confidence. These region are Njombe (Hot Spots with 99 percent (99%) level of confidence), Ruvuma (Hot Spots with 95 percent (95%) level of confidence), Mbeya (Hot Spots with 95 percent (95%) level of confidence), Iringa (Hot Spots with 95 percent (95%)

level of confidence), Shinyanga (Hot Spots with 90 percent (90%) level of confidence and Rukwa (Hot Spots with 90 percent (90%) level of confidence).

Lack of the knowledge on how the disease can be transmitted and its preventive measures, practicing unsafe sexual intercourse and inadequate of HIV.AIDS preventive programs which could have a constant tasks for providing Health education to the societies could be the reasons for the high infections for these regions.

d) The IDW for HIV Infections

The IDW as one of the interpolation method was used to visualize the area with high infections (Hot Spots) closely. This could be helpful in facilitate the decision making for the necessary measure to rescue the transmission of the disease. The following Figure 5.17 shows the IDW representing the clusters of Hot and Cold values for the HIV infections (Zhang et al, 2014).

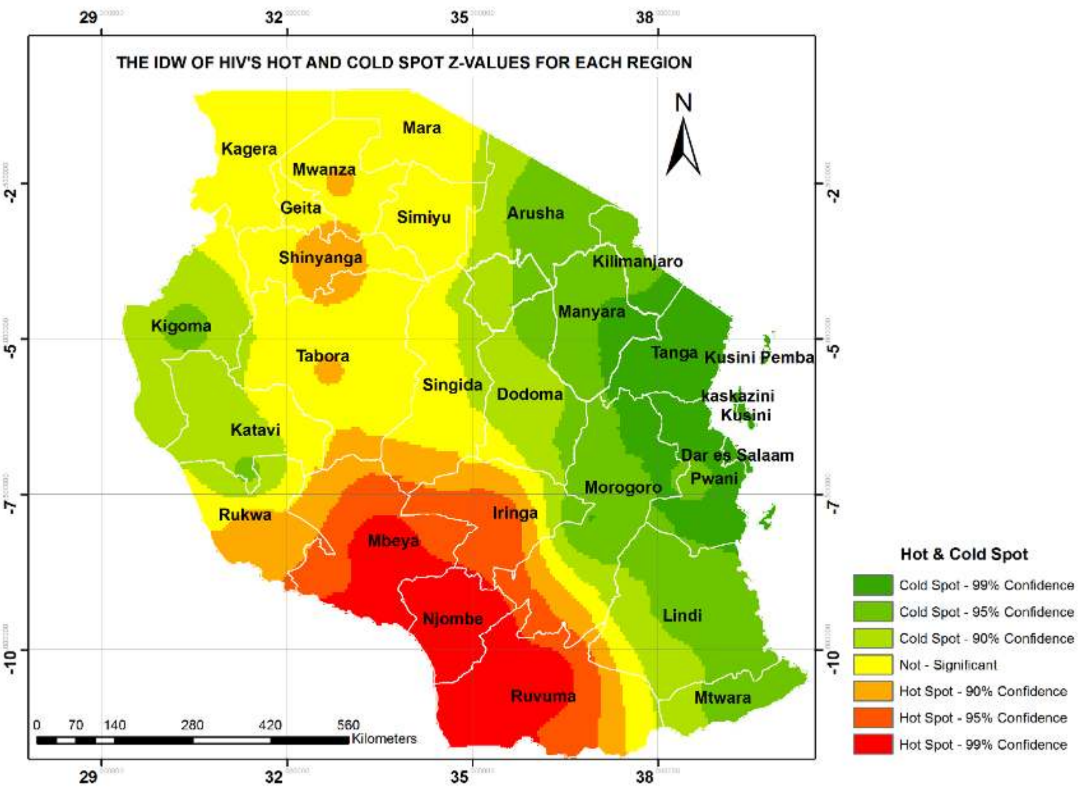


Figure 5.17 The interpolated surface for HIV using IDW

CHAPTER SIX
CONCLUSION AND RECOMMENDATION

One of the reason for the lack of success in the anemia control intervention is designed on the assumption that nutritional iron deficiency is a major course of anemia. Information on infectious causes of anemia is not routinely collected for the children under 5 years old, however, the collection of urine and blood samples should also be included during the DHS surveys so as to gather the important information that shall help to trace the major causes this disease.

This could be very helpful to the government on initiating the policies and implementation of the necessary measures to deal with these infections. Since Tanzania is one of the tropical country, it has always a constant attack of parasitic infection nutritional iron infection which causes anemia. So, it is important to collect these samples routinely so as to track and examine the sources.

As the map described above, the low level of economic status in some of regions were observed to be among of a factor for spreading of anemia disease in the sense that, those families cannot afford the nutritional food for their children or to protect them against malaria. So the government via the ministry of Health should have to provide a constant channel capable of delivering effective anti-anemia interventions to these regions with infections. The spatial distribution of anemia disease on map could be helpful to them so as to accelerate a decision making and implements the policies for a targeted groups.

Based on malaria, it is true that most of Africa countries have been suffering for a long time now with this disease, As the maps shows, the most affected regions seems to have low number of mosquito nets and low level of economic status, so the government should ensure that these nets should be provided with the insect sides to reduce the effects while investing much on the malaria researches. Sometime it is hard to believe, if all of these years there is no solutions for malaria disease! The government must be seriously collaborate with scientists, clinicians, policy-makers, Health managers and Health-care providers word wide, working together as a team to reduce the burden of malaria on the continent.

Finally, the rate of HIV/AIDS seems to affect more youth of 45-50 ages who are the engine for development, so the government should also provide a constant control in facilitating the vaccination and classes for awareness of this disease. Poverty could be one of the reason for the young people to participate in unsafe sex intercourse, for this reason the government should always keep thinking on how to create opportunities for their people or provide a good environment for them to initiate their own job.

By the aid of spatial analysis, the hot and cold spots clusters for HIV/AIDS were identified, all of these clusters were clearly shown on maps. This will simplify for the government to implement the policies specifically for these regions to combating and finally eradicate to it.

REFERENCES

Amaral, P. V. & Anselin, L. (2014). Finite sample properties of Moran's I test for spatial autocorrelation in tobit models. *Paper in Regional Science, 93* (4), 774-781.

Buehler, C. P., Blevins, M., Ossemane, E. B., Calvo, L. G., Ndatimana, E., Vermund, S. H., et al. (2014). Assessing spatial patterns of HIV knowledge in rural Mozambique using geographic information system. *Tropical Medicine and International Health, 20* (3), 353-364.

Chimoyi, L.A. & Musenge, E. (2011). Spatial analysis of factors associated with HIV infection among young people in Uganda. *BMC Public Health, 14* (555), 2-14.

Getis, A. & Ord J. K. (1992). The analysis of spatial association by use of distance statistics. *Ohio State University Press, 24* (3), 190-195.

Getis Arthur & Ord J. K. (1995). local spatial autocorrelation statistics: distributional issues and an application. *Ohio State University Press, 27* (4), 287-294.

Gwarda, K. L. (2014). Spatial analysis of foreign migration in Poland in 2012 using geographically weighted regression. *Comparative Economic Research, 17* (4), 137-154.

Handique B.K., Khan S.A, Chakraborty K, Goswami J and Sarma K.K (2011). Spatial statistical analysis of Japanese encephalitis occurrence and identification of disease hot spots- case studies in a je endemic district of north east India. *International Journal of Geoinformatics, 7* (1), 7-14.

Huisman, O. & Deby, R. A. (2009). Principle of geographic information systems. ITC, Enschede- Netherlands.

Jung, I. & Park, G. (2014). P-value approximations for spatial scan statistics using extreme value distributions. *Statistics in Medicine*, *34,* 504-514.

Klosterman – Hardcover (1993). *Spreadsheet models for urban and regional analysis.* Retrieved 27th, January 2015, from http://product.half.ebay.com.

Lin, G. (2004). Comparing spatial clustering tests based on rare to common spatial events. *Computers, Environment and Urban Systems, 28,* 691-699.

Magalhaes, R. J. S. & Clements, A. C. A. (2011). Mapping the risk of anaemia in preschool-age children: the contribution of malnutrition, malaria, and helminth infections in West Africa. *Plos Medicine, 8* (6), 1-17.

Madulu, N. F. (2002). Population Distribution and Density in Tanzania: Experiences From 2002 Population and Housing Census. Institute of Resource Assessment. University of Dar-Es-Salaam-Tanzania.

Ma, G. and Zhang, Z. (2014). *The change of urban population spatial structure of nanjing*, 1982-2007. Retrieved 18th, February 2015, from http://ieeexplore.ieee.org/Xplore/home.jsp.

Ministry of Health and Social Welfare - MOHSW (2002). *Tanzania National Health Accounts* (NHA). Dar-Es-Salaam - Tanzania.

Ministry of Finance and Economic Affairs - MFEA (2009). *Tanzania Economic Survey*. Dar-Es- Salaam, Tanzania.

National Bureau of Statistics - NBS (2010). *Tanzania Demographic and Health Survey*. Icf Macro, Calverton, Maryland U.S.A.

National Bureau of Statistics - NBS (2009). *National Accounts of Tanzania Mainland*. Dar- Es-Salaam -Tanzania.

National Bureau of Statistics-NBS and Orc Macro (2005). *Tanzania Demographic and Health Survey*. Dar Es Salaam-Tanzania.

Ord, K., & Getis, A. (2001). Testing for local spatial autocorrelation in the presence of global autocorrelation. *Journal of Regional Science, 41*, 411–432.

Peeters, A., Zude, M., Kathner, J., Unlu, M., Kanber, R.,Hetzroni, A., et al., (2015). Getis–Ord's hot- and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data. *Institute of Soil, Computers and Electronics in Agriculture, 111,* 140-150.

Qian, S., Guo, W., Xing, J., Qin, Q., Ding, Z., Chen, F., et al., (2014). *Diversity of HIV/AIDS epidemic in China: a result from hierarchical clustering analysis and spatial autocorrelation analysis.* Retrieved 18th, February 2015, from http://ieeexplore.ieee.org/Xplore/home.jsp.

Saxena, R.; Nagpal, B.N.; Das, M.K.; Srivastava, A.; Gupta, S.K.; Kumar, A.; Jeyaseelan, A.T. and Baraik, V.K. (2011). *A Spatial Statistical Approach to Analyze Malaria situation at micro level for priority control in Ranchi district, Jharkhand,* India, *Indian J Med Res, 136,* 776-782.

Stopka, J. T.; Krawczyk, C.; Gradziel, P. And Geraghty, E. M. (2014). Use of spatial epidemiology and hot spot analysis to target women eligible for prenatal women, infants, and children services. *American Journal of Public Health, 104* (S1), 183-189

Tanzania Commission for Aids [TACAIDS], Zanzibar Aid Commission [ZAC], National Bureau of Statistics [NBS], Office of the Chief Government Statistician [OCGS], and Macro International, Inc (2008). Tanzania Hiv/Aids and Malaria Indicator Survey. Dar- Es Salaam – Tanzania.

Tonglin, Z. and Lin, G., (2006). A decomposition of Moran's I for clustering detection. *Econ Papers, 51*, 6123-6137.

Tatem, A. J., Adamo, S., Bharti, N., Burgert, C. R., Castro, M., Dorelien, A., et al., (2012*)*. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. *Population Health Metrics, 10* (8), 2-12.

World Bank (1993). World development report investing in Health. New York: *Oxford University Press.*

Yang, M., Ma, J., Jia, P.,Pu, Y. and Chen, G. (2011). *The use of spatial autocorrelation to analyze changes in spatial distribution patterns of population density in jiangsu province.*Retrieved18th,February2015,fromhttp://ieeexplore.ieee.org/Xplore/home.jsp.

Yu, Z. and Chun, D. (2010). Spatial Autocorrelation Analysis of Population in Main Years, *Computer Research and Development, International Conference on*, 844 – 848.

Zhang, J. And Zhu, Y. (2011). *The population spatial distribution model based on the spatial statistics in shandong province.* Retrieved 18th February 2015, from http://ieee.org/Xplore/home.jsp.

Zhang, L.; Qin, Y.; Zhang, J. and Lu, C. (2014). *Spatial differentiation of urban carbon emissions.* Retrieved 18th, February 2015, from http://ieeexplore.ieee.org/Xplore/home.jsp.

Zhang, X., Bao, S., Zhu, X. And Su, K. (2014). *Spatial intelligence with spatial statistics.* Retrieved 18th, February 2015, from http://ieeexplore.ieee.org/Xplore/home.jsp.

APPENDICES

Appendix A: THE HOUSEHOLDS TABLE

ID	CODE	CODE MEANING
1	HHID	Case Identification
2	HV000	Country code and phase
3	HV001	Cluster number
4	HV002	Household number
5	HV003	Respondent's line number (answering Household questionnaire)
6	HV004	Ultimate area unit
7	HV005	Household sample weight (6 decimals)
8	HV006	Month of interview
9	HV009	Number of household members
10	HV010	Number of eligible women in household
11	HV011	Number of eligible men in household
12	HV012	Number of de jure members
13	HV013	Number of de facto members
14	HV014	Number of children 5 and under (de jure)
15	HV015	Result of household interview
16	HV016	Day of interview
17	HV017	Number of visits
18	HV024	Region
19	HV025	Type of place of residence
20	HV026	NA - Place of residence
21	HV027	Household selected for male interview
22	HV035	Number of eligible children for height and weight
23	HV040	Cluster altitude in meters
24	HV041	Total adults measured

25	HV042	Household selected for hemoglobin
26	HV201	Source of drinking water
27	HV202	NA - Source of non-drinking water
28	HV204	Time to get to water source (minutes)
29	HV205	Type of toilet facility
30	HV206	Has electricity
31	HV207	Has radio
32	HV208	Has television
33	HV209	Has refrigerator
34	HV210	Has bicycle
35	HV211	Has motorcycle/scooter
36	HV212	Has car/truck
37	HV213	Main floor material
38	HV214	Main wall material
39	HV215	Main roof material
40	HV216	Number of rooms used for sleeping
41	HV217	Relationship structure
42	HV218	Line number of head of household
43	HV219	Sex of head of household
44	HV220	Age of head of household
45	HV221	Has telephone (land-line)
46	HV225	Share toilet with other households
47	HV226	Type of cooking fuel
48	HV227	Has mosquito bed net for sleeping
49	HV228	Children under 5 slept under mosquito bed net last night
50	HV235	Location of source for water
51	HV238	Number of households sharing toilet
52	HV243A	Has mobile telephone
53	HV245	Hectares of agricultural land (1 decimal)
54	HV246	Owns livestock, herds or farm animals

55	HV246A	Owns cattle
56	HV246B	Owns cows/ bulls
57	HV246C	Owns horses/ donkeys/ mules
58	HV246D	Owns goats
59	HV246E	Owns sheep
60	HV246F	Owns chickens/other poultry
61	HV246G	Owns pigs
62	HV252	NA - Frequency household members smoke inside the house
63	HV253	Has dwelling been sprayed against mosquitoes in last 12 months
64	HV253A	Dwelling sprayed by: government worker/program
65	HV253B	Dwelling sprayed by: private company
66	HV253C	Dwelling sprayed by: NGO
67	HV253X	Dwelling sprayed by: other
68	HV253Z	Dwelling sprayed by: don't know
69	HV270	Wealth index
70	HV271	Wealth index factor score (5 decimals)
71	HML1	Number of mosquito bed nets
72	HML1A	Number of mosquito bed nets with specific information
73	HML2	Number of children under mosquito bed net previous night
74	SHREG1	Zone
75	SHREG2	Mainland/Zanzibar
76	SH101A	Provider of drinking water
77	SH107AA	Battery or generator for power
78	SH107AB	Paraffin lamp
79	SH107EA	Iron
80	SH108A	Source of energy for lighting in the household
81	SH112A	Number of sleeping spaces used in the household
82	SH115A	Acres of land for farming owned by the household (1 decimal)
83	SH115B	Acres of land for grazing owned by the household (1 decimal)

84	SH115AA	Use land for farming or grazing that the household doesn't own
85	SH115BA	Acres of land used for farming not owned by the household (1 decimal)
86	SH115BB	Acres of land used for grazing not owned by the household (1 decimal)
87	SH115BC	Hectares of agricultural land not owned by the household (1 decimal)
88	SH119	Distance to the nearest marketplace (kms)
89	SH120	Number of meals the household has per day
90	SH121	Number of days the household ate meat/fish in the past week
91	SH122	Frequency having problems in satisfying the food needs of the hh last year
92	SH123	Distance to the nearest Health facility (kms)
93	SH123A	Means of transportation to the Health facility

Appendix B: THE MEMBERS TABLE

ID	CODE	CODE DEFINITION
1	HHID	Case Identification
2	HVIDX	Line number
3	HV000	Country code and phase
4	HV001	Cluster Number
5	HV002	Household number
6	HV003	Respondent's line number (answering Household questionnaire)
7	HV006	Month of interview
8	HV007	Year of interview
9	HV016	Day of interview
10	HV020	Ever-married sample
11	HV210	Has bicycle
12	HV211	Has motorcycle/scooter
13	HV212	Has car/truck
14	HV217	Relationship structure
15	HV218	Line number of head of household
16	HV227	Has mosquito bed net for sleeping
17	HV228	Children under 5 slept under mosquito bed net last night
18	HV243A	Has mobile telephone
19	HV243B	Has watch
20	HV243C	Has animal-drawn cart
21	HV243D	Has boat with a motor
22	HV244	Owns land usable for agriculture
23	HV245	Hectares of agricultural land (1 decimal)
24	HV246	Owns livestock, herds or farm animals
25	HV246A	Owns cattle
26	HV246B	Owns cows/ bulls
27	HV246C	Owns horses/ donkeys/ mules

28	HV246D	Owns goats
29	HV246E	Owns sheep
30	HV246F	Owns chickens/other poultry
31	HV246G	Owns pigs
32	HV247	Has bank account
33	HV270	Wealth index
34	HV101	Relationship to head
35	HV102	Usual resident
36	HV103	Slept last night
37	HV104	Sex of household member
38	HV105	Age of household members
39	HV106	Highest educational level attained
40	HV107	Highest year of education completed
41	HV108	Education completed in single years
42	HV109	Educational attainment
43	HV110	NA - Member still in school
44	HV111	Mother alive
45	HV112	Mother's line number
46	HV113	Father alive
47	HV114	Father's line number
48	HV115	Current marital status
49	HV116	Currently, formerly, never married
50	HV118	Eligibility for male interview
51	HV120	Children eligibility for height/weight and hemoglobin
52	HV121	Member attended school during current school year
53	HV122	Educational level during current school year
54	HV123	Grade of education during current school year
55	HV124	Education in single years - current school year
56	HV140	Member has a birth certificate
57	IDXH4	Index to Household Schedule

58	SH15A	Level of education attended
59	SH15B	Highest grade completed
60	SH17A	Level of education attending
61	SH17B	Grade attending
62	SHEDUC	Education as defined in report
63	HA0	Index to Household Schedule
64	HA51	Line number of parent/caretaker
65	HA54	Currently pregnant
66	HA60	Marital status
67	HA61	Read consent statement - HIV
68	HA62	Bar code for HIV blood sample
69	HA63	Result of measurement - HIV
70	HA64	Consent for additional tests
71	HA65	Result of woman's individual interview
72	HA66	Woman's highest educational level
73	HA67	Woman's highest year of education
74	HA68	Woman's highest educational level (questionnaire configuration)
75	HA69	HIV weight (6 decimals)
76	HC52	Read consent statement - hemoglobin
77	HC53	Hemoglobin level (g/dl - 1 decimal)
78	HC55	Result of measurement - hemoglobin
79	HC56	Hemoglobin level adjusted for altitude (g/dl - 1 decimal)
80	HC57	Anemia level
81	HC60	Mother's line number (woman's questionnaire)
82	HC61	Mother's highest educational level
83	HC62	Mother's highest year of education
84	HC63	Preceding birth interval (months)
85	HC64	Birth order number
86	HC68	Mother's highest educational level (questionnaire configuration)

87	HB60	Marital status
88	HB62	Bar code for HIV blood sample
89	HB63	Result of measurement - HIV
90	HB64	Consent for additional tests
91	HB65	Result of man's individual interview
92	HB67	Man's highest year of education
93	HB68	Man's highest educational level (questionnaire configuration)
94	HML4	Months ago net obtained
95	HML18	Pregnancy status from Individual file
96	HML19	Person slept under an ever-treated net
97	HML20	Person slept under an LLIN net
98	HML30	Line number of parent/caretaker (for malaria testing)
99	HML31	Read consent statement for malaria
100	HML32	Final result of malaria from blood smear test
101	HML33	Result of malaria measurement
102	HML35	Result of malaria rapid test

Appendix C: ECONOMIC STATUS QUERY AND CONSIDERATION

The query for calculating the economic status was as follow;

ID, HHID, HV025, HV206, HV211, HV212, HV227, HV245, HV246A, HV246B, HV246C, HV246D, HV246E, HV246F, HV246G, HV216, HV009, int (1000 * (HV206 + HV211 + HV212 + HV227 + HV245N + HV246AN + HV246BN + HV246CN + HV246DN + HV246EN + HV246FN + HV246GN + RoomINDN)) / 1000 "**ECOSTAT**"

Whereby;

$$RoomIND = \frac{HV216}{HV009}$$

To be more consistence, the fields from HV245, HV246A to HV245G were normalized in form of "0" and "1" so as to match with other fields, this normalization helped in getting the total value after addition which indicate the rural economic status.

The normalized formula was as follow;

$$Z_i = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

Where by X_i is the actual value in the field

X_{min} is the minimum value within column

X_{max} is the maximum value within the column

Z_i is the normalized value.

On the other hand the urban economic status was calculate the same way as we calculate the economic status for rural areas, though there was a slightly less field compared to the rural area due to the fact that the animals were not included in calculating the urban economic status.

The selected fields for urban economic status were as follow, has electricity (HV206), has motorcycles (HV211), has a car (HV212), has mosquito bed net for sleeping

(HV227) and (HV216/HV009). The complete fields used for analysis would be included among the appendixes because it contains many value fields.

Appendix D

A CRONYMS

AIDS	Acquired immune deficiency syndrome
ACT	Artemisinin based Combination Therapy.
DHS	Demographic Health Survey
DBMS	Database management systems
DHSD	Demographic Health survey data
DHSR	Donors Health Sector Reform
DSA	Demographic Surveillance Area
GIS	Geographical Information System
HIV	Human Immuno Deficiency Virus
HSRPA	Health Sector Reform Plan of Action
HH	Household
IDW	Inverse Distance Weight
IPT	Intermittent Preventative Treatment
IRS	Indoor Residual Spraying
JMH	Joint Ministry of Health
KCMC	Kilimanjaro College Medical Centre
LLIN	Long Lasting Insecticide Nets
LISA	Local Indicators of Spatial Association
MOH	Ministry of Health
MOHSW	Ministry of Health and Social Welfare
NGO	Non-Government Organization
PDA	Personal Digital Assistant
PSI	Population Service International
Sd	Sickness Density
SPSS	Statistical Package for the Social Sciences.
SAS	Statistical Analysis System
SP	Sulphadoxine Pyrimethamine
USA	United States of America