CHAPTER ONE
INTRODUCTION

1.1 Background

“Cancer is a generic term for a large group of diseases that can affect any part of the body. Other terms used are malignant tumours and neoplasms. One defining feature of cancer is the rapid creation of abnormal cells that grow beyond their usual boundaries, and which can then invade adjoining parts of the body and spread to other organs"(World Health Organization, 2015).It is a major health issue and one of the most common diseases in the World, with more than 11 million cases being diagnosed every year (Yomralioglu et al., 2009) .The World Health Organization (WHO) estimated, that 7.6 million died from cancer in 2005 and 84 million people will die over the next 10 years, if no action is taken. More than 70% of all cancer related deaths occur in countries, where the population has a medium to low level standard of living and limited resources for the prevention, diagnosis, and treatment of cancer(World Health Organization, 2015).Also in Sudan - according to hospitals reports - cancer was the third leading cause of death, after malaria and viral pneumonia accounting for 5% of all death in Sudan(Saeed et al., 2013).

Most cancers are related to exposures to environmental pollutants, lifestyle, age, genetic factors, certain types of infections(Stewart and Wild, 2014).More than 30% of cancer deaths could be prevented by modifying or avoiding key risk factors like environmental pollutants and lifestyle (World Health Organization, 2015).In the health area, physicians observed that certain diseases related to environment tend to occur in some places and not others, thus it is necessary to investigate how the diseases are geographically distributed(Yomralioglu et al., 2009). These studies can be carried out through Geographic Information Systems (GIS), which are computer based systems for the acquisition and update, storage and query, analyses and simulation as well as output and presentation of spatial data (Margaret and Kenneth, 2014). These systems can demonstrate both spatial and associated attribute data. GIS suits best in finding out regional differences and demonstration and interpretation of the relationship of the spatial distribution of any disease with the geographical/environmental factors(Yomralioglu et al., 2009). GIS technology has widely been used for years to explore spatial distribution of diseases and particularly different cancer types to investigate the cause and effecting relationships from different perspectives, in particular to examine, where specific cancer types are more intense (Ellul et al., 2013). Furthermore, it can explored the relationship between the Cancer incidence rates, mortality and environmental risk factors (Yomralioglu et al., 2009). also the relationship between cancer mortality and population exposure to coal mining activities (Hendryx et al., 2010), and the relationships between thyroid cancer incidence and the distance from nuclear power plants (Watase, 2012).

In Sudan, the number of studies on cancer by the use of GIS techniques remains insufficient, As an example, the ministry of health created maps showing cancer cases in an annual report 2007(Federal Ministry of Health ,2007). Another study aiming at mapping cancer density in Gazira State have been published (Elebead et al., 2012).

This study aims at testing correlations between cancer cases and environmental factors in Sudan, for a period in time of about 18 years, in particular between 1995-2013. Hopefully, the results of this study would help to develop and apply prevention and control programs of cancer in Sudan.

1.2 The Research Problem

A first question comes up: Is there any relationship between cancer cases and the environmental factors in Sudan?

In Sudan, according to hospitals reports, cancer was the third leading cause of death after malaria and viral pneumonia accounting for 5% of all death in Sudan (Saeed et al., 2013).As also shown by the annual statistical reports, published by the Ministry of Health at various years, some states have more cancer cases than other states, with regard to all cancer types or specific types (Federal Ministry of Health ,2014).

1.3 Research Motivation:

In Sudan there are few studies using GIS technique to test relationships between cancer cases and environmental factors. Also few mapping systems are available to show cancer cases.

1.4 Hypothesis

Is there any correlation between cancer cases and the environmental factors in Sudan?

1.5 Objectives

1.5 .1 General Objective

The general objective of this study is to find any relationships between cancer cases and the environmental factors in Sudan

1.5 .2 Specific objectives

- To display cancer statistics during the period 1995-2013. This includes all types of cancer.
- To display which region has most cancer cases during the period 2010-2013 for most common cancer types in Sudan (bladder, brain, prostate, leukemia , breast).
- To compare (bladder, Prostate) cancer cases with agriculture density.

- To compare (brain, leukemia, breast) cancer cases with communication towers density.
- To compare cancer cases with population projections and calculate correlations for the period 1998-2013.
- To predict numbers of cancer cases for Blue Nile and Khartoum states at 2014.

1.6 Research Scope

This study was conducted to differentiate in five cancer types (Breast, Leukemia, Brain, Prostate and Bladder) and two environmental factors (communication towers and agriculture) based on county level data, given for 2010- 2013. Here, we examined the correlation between cancer cases and population projections in some states (Northern, Blue Nile, North Darfur, Kassala , AlGedaref ,Red Sea, Khartoum , River Nile, AlGazira, White Nile, and Sennar), for the time period 1998-2013.Further more we predict the number of cases for Blue Nile and Khartoum state at 2014.

1.7 Structure of the Thesis

This thesis is divided into seven chapters:

Chapter one contains an introduction, the hypothesis and research objectives. It also outlines the research problem, the scope and motivation for this thesis. Chapter two deals with related work and literature review. Chapter three discusses the methodology. Chapter four contains the data and the test beds. Chapter five presents the results and Chapter six contains a discussion, the conclusions, some recommendations for future work. The study is concluded by the references chapter.

CHAPTER TWO

LITERATURE REVIEW

2.1 Cancer

Cancer is not a single disease. It is a group of more than 200 different diseases. Cancer is an uncontrolled growth of cells that disrupts body tissues and organs. Cancerous cells are not normal in their structure and function. They grow and multiply to form tumors that invade local tissues and sometimes scatter throughout the body. In the beginning, there are no warning signs to alert us to the disease. Later, the signs of cancer are related to the location of the tumor (World Health Organization, 2015).

2.1.1 Causes of Cancer

Most cancers are related to environmental, lifestyle, or behavioral exposures (Anand et al., 2008).
The term "environmental", as used by cancer researchers, refers to everything outside the body that interacts with humans (Miller , 2014) .Common environmental factors, that contribute to cancer and lead to death include tobacco, obesity, infections, radiation (both ionizing and non-ionizing), and environmental pollutants(Anand et al., 2008).

2.1.1.1 Agriculture and Cancer

Farmers in many countries have lower overall death rates and cancer rates than the general population. This might be true, in particular, due to lower smoking rates, as well as more physically active lifestyles and dietary factors.

However, compared with the general population, the rates for certain diseases, including some types of cancer, appear to be higher among agricultural workers, which may be related to exposures that are common in their work environments. For example, farming communities have higher rates of leukemia, non-Hodgkin lymphoma, multiple myeloma, and soft tissue sarcoma, as well as cancers of the skin, lip, stomach, brain , bladder and prostate.

Farmers, farm workers, and farm family members may be exposed to substances such as pesticides, engine exhausts, solvents, dusts, animal viruses, fertilizers, fuels, and specific microbes that may account for these elevated cancer rates(Freeman et al., 2011).

Many studies showed, that there may be a connection between pesticides and cancer in adults and children. Studies have been done investigating groups of people, who apply pesticides on farms to find out how many of these people get cancer and how many die from it. The results suggest that this group may have a slightly higher risk than the average person of developing non-Hodgkin lymphoma, leukemia, multiple myeloma, and cancers of the prostate and brain.

Also study from the US : Agricultural Health Study found, that men who applied pesticides had higher rates of prostate cancer compared with the average man. Also, another study suggested that people who are exposed to pesticides at work, such as farmers and farm workers, may be at a higher risk of myeloid leukemia, particularly acute myelogenous leukemia(Canadian Cancer Society , 2016).

2.1.1.2 Radiation and Cancer:-

Radiation is energy, that travels through space in the form of waves or particles. There are two types of radiation: ionizing and non-ionizing.

2.1.1.2.1 Ionizing Radiation

Ionizing radiation has enough energy to break chemical bonds between molecules or to form charged molecules (cause ionization). This means, that ionizing radiation is strong enough to damage cells and DNA, and strong enough to increase the chance of developing cancer.

Sources of ionizing radiation exposure include:

- natural background radiation (sources of exposure include cosmic rays from the solar system and radioactive materials in the soil and rocks)
- background radiation from human activities (includes testing and using nuclear weapons as well as generating nuclear power)
- radon
- medical radiation

2.1.1.2.2 Non-ionizing Radiation

Non-ionizing radiation is not as strong as ionizing radiation, and it does not have enough energy to break bonds between molecules. But being exposed to some types of non-ionizing radiation can still harm.

Sources of non-ionizing radiation exposure include:

- radiofrequency fields (including cell phones, cell phone towers and microwave ovens)
- electromagnetic fields (including power lines and household appliances)
- ultraviolet (UV) rays (including the sun and indoor tanning beds) (Canadian Cancer Society , 2016) .

Numerous studies and comprehensive reviews of the scientific literature have evaluated possible associations between exposure to non-ionizing and risk of cancer in children. Most of the researches focused on leukemia and brain tumors, the two most common cancers in children. Studies found no consistent evidence for an association between any non-ionizing source and cancer.

Also some studies have shown no relationship between breast cancer of women and exposure to non-ionizing sources, although a few individual studies have suggested an association (National Cancer Institute, 2016).

2.1.1.2.3 Cell Phone Towers and Cancer

Some people have expressed concern that living, working, or going to school near a cell phone tower might increase the risk of cancer or other health problems. At this time, there is very little evidence to support this idea. In theory, there are some important points that would argue against cellular phone towers being able to cause cancer**.**

In one large study, British researchers compared a group of more than 1,000 families of young children with cancer against a similar group of families having children without cancer. They found no link between a mother's exposure to the towers during pregnancy (based on the distance from the home to the nearest tower and on the amount of energy given off by nearby towers) and the risk of early childhood cancer.

In another study, researchers compared a group of more than 2,600 children with cancer to a group of similar children without cancer. They found, that those who lived in a town ,could have been exposed higher than with average RF radiation of rural areas (American Cancer Society ,2016).

2.3 Cancer in Sudan

The cancer prevalence rate per year is 5,000-7,000 among adults and 300-400 among children, with increasing tendency for adults. Male : female ratios are 1:1.18 for adults and 1.46:1 for children. The five most frequent tumor types are breast cancer, leukemia, prostatic carcinoma, lymphoma and colorectal carcinoma in adults and leukemia, lymphoma, eye tumors, sarcoma and brain tumors in children. Remarkably, the median age of cancer diagnosis is 10-20 years higher in men than in women, mainly due to earlier onset of gender-related tumors in females (cancer of breast, cervix, or ovary) than in men (prostatic carcinoma). Chronic myeloid leukemia is the most frequent haematopoietic malignancy in adults and acute lymphoblastic leukemia in children. Comparing cancer cases with population numbers of Sudanese states, e.g. Northern Sudan, River Nile and Khartoum revealed up to 8-fold higher cancer incidence rates than Al Gedarif, Southern Darfur and Blue Nile. The other states had intermediate incidence rates. Interestingly, esophageal carcinoma occurred proportionally more frequently in Kassala than in the entire Sudan or other states. (Safee et al., 2016)

2.4 Cancer in the World

According to GLOBOCAN 2012, an estimated 14.1 million new cases of cancer in the world occurred: 7.4 million (53%) in males and 6.7 million (47%) in females, giving a male : female ratio of 10:9.About 8.2 million cancer-related deaths occurred in 2012, compared with 7.6 million, in 2008. The World age-standardized incidence rate shows that there are 205 new cancer cases for every 100,000 men in the world, and 165 for every 100,000 females (Carter, 2014).

2. 5 GIS and Health

A GIS can be defined as a computer system with the capacity to capture, store, analyze, and display geographically-referenced information. In other words, it is an informatics tool for storing and managing data, that has been identified according to location. GIS have been used in a wide variety of fields, including the natural, social, engineering, and health.

Health is another focus area that has made increasing use of GIS to help address a number of significant health issues, ranging from diseases management to improved services.

2.5.1 Benefits of Using Geographic Information Systems in Health

GIS can play an important role with regard to the surveillance, management and analysis of diseases. There seem to be important tools for analysis and visualization of data. Furthermore, trends and correlations would be difficult to be understood with traditional ways of processing and imaging of these data (Fradelos et al., 2014) .Public health services, diseases, and any information regarding health can be displayed on a map and correlated amongst many pieces of information such as environmental data, elements of health concern and social information. Thus, it can be also used for monitoring and management of both, diseases and health programs. Therefore, it is necessary that we become able to understand, monitor and emphasize on the reasons, that may be correlated to the development of a disease. Some of these factors could be the environment, conduct and the socioeconomic level of an area. Should the “source” of a disease identifiable, and its development and transmission are known, health administrators will able to deal effectively with pandemic outbreaks (Sears, 2015) .

A GIS is a tool with great potential that might also contribute to the assessment of environmental risks and people’s exposure to them (Fradelos et al., 2014).

GIS software offers healthcare professionals the ability to identify health related trends and more thoroughly target healing efforts based upon those results (Fradelos et al., 2014), For example, a Cancer Surveillance System in Southern California has been established to assess the demographic data, such as home address, workplace, cancer type, and even data collected from wearable health tech of all patients entered into the system. Data is then georeferenced and mapped. Healthcare professionals can visualize the locations of patients and determine, if there are clusters of specific types of cancer associated with similar working conditions or residential areas (Sears, 2015).

The role of GIS should not be limited simply to tracking occurrences of diseases though. One of its most powerful aspects is the ability to use geography and other inputs to identify, where diseases are most likely to spread next. Data such as this can be essential to on-the-ground personnel working to save lives because it enables them to prepare in advance for a disease and can severely limit the impact. Health maps are beginning to play a significant role in the management of disease outbreaks such as Ebola and measles. For instance, during the Disneyland measles outbreak in December of 2014, GIS-based maps were created to help visualize where infected children lived and the potential spread of the disease. Furthermore, it was used to gain a better understanding of vaccination rates and laws in differing counties throughout the US to determine which locations could be hit the hardest given a serious outbreak(Sears, 2015).

GIS can contribute to public health in many ways due to the fact that they can provide information on many issues and support correctly the decision making process. They can provide information regarding the distribution of health services. Thus, any growing disparities might be eliminated. Also, policy-makers would make right decisions. Health professionals can easily identify the difficulties and disparities regarding the accessibility to health services; and so, they are able to cope with the current situation (Fradelos et al., 2014).

2.5.2 Project Examples of GIS and Public Health

One of the earliest studies used maps , conducted by Dr. John Snow in London, when he used maps to show the locations of cholera deaths in the 1850s. He superimposed his maps showing death locations with maps of public water supplies and was able to find an area, where the deaths were clustered near a water pump on the city's Broad Street. The handle for the water pump was later removed and cholera deaths in the area decreased significantly. Dr. Snow's research eventually became part of a field known as disease diffusion mapping (Longley et al., 2015) **.**

In addition, some early studies looked at the prevalence of diseases amongst people living at high versus low elevations and it was found, that people living at low elevations near waterways were more likely to experience malaria **(**Briney, 2014).

Investigating the etiology of asthma prevalence and residential exposure to air pollution in United States – Canada. This study found clustering of asthma cases in close proximity to the high volume traffic areas (Boehmer et al., 2013).

The Division of heart disease and stroke prevention in the United States(US)has developed an interactive map website that presents data for heart disease and stroke mortality rates, county-by-county, for each state (Boyle and Holben , 2012).

2.6 GIS and Cancer

GIS applications have been developed, utilized and continuously refined for the mapping and spatial analysis of cancer incidence, morbidity and mortality data. Four general approaches for utilizing GIS in cancer research have been taken: 1) disease mapping 2) geographic correlation studies 3) risk assessment in relation to a pre-specified point or line source, and 4) cluster detection and disease clustering (Bell et al , 2006).

2.7 Previous Studies

(several studies worldwide are conducted to look for a relationship between Cancer cases and environmental factors).

(Dode et al., 2011): The purpose of this study is to verify the existence of a spatial correlation between communication towers clusters and cases of deaths by specific types of cancer, that may be caused by radiation, such as tumors in the prostate, breast, lung, kidneys and liver in the Belo Horizonte municipality, Minas Gerais state, Brazil, from 1996 to 2006.

After found on the map nearly 300 communication towers in the city, the researcher found, that 81 percent of those victims lived less than 500 meter from one of these towers and breast cancer has high rate compared to another types.

(Yomralioglu et al., 2009)conducted a study to examine whether there is a relationship between cancer types (lung, skin, breast, stomach, prostate, thyroid, and bladder) and geo-environmental factors (land cover, e.g. forest land, agriculture land and residential areas) in Trabzon, Turkey 2004. After the creation of distribution maps of cancer cases, based on land cover ,the number of cancer cases in each land cover class was determined with spatial analysis algorithms. The study found out, that breast cancer occurred more in residential areas than other cancer types, and skin and thyroid cancer occurred more in forestry areas .Moreover, prostate and bladder cancer occurred more in agriculture areas than the other cancer types. Furthermore, a statistical analysis (Pearson's chi-square (x^2)) indicated, that there was a relationship between cancer types and land cover, by x^2=24.391 p = 0.041.

(Wolf and Wolf, 2008) aimed to investigate, whether there is an increased cancer incidence(breast, ovary, lung, osteoid osteoma , and hypernephroma) in populations, living near to communication towers in Netanya, Israel from 1997 to 1998. This study found out an association between increased incidence of cancer and living within 350 meters from towers and breast cancer has high rates compared to another types.

(Eger et al., 2004)examined, whether people living close to communication towers were exposed to a heightened risk of taking ill with cancer in Southern German and Naila city, Germany, from 1994 and 2004. The study found out the proportion of newly developing cancer cases (breast, ovary, prostate, pancreas, bowel, skin, lung, kidney, stomach, bladder) was higher among those patients, who had lived during the past ten years at a distance of up to 400 meters from the towers, compared to those patients living further away .

CHAPTER THREE

METHODOLOGY

3.1 Study Idea

This research aims at examining the correlation between the different kinds of cancer incidence rates(Breast, Leukemia, Brain, Prostate and Bladder) and environmental factors (communication towers, agriculture density) in Sudan from 1995 to 2013. Firstly, various data have to be collected. Secondly, cancer incidence rates at county level are calculated

cancer incidence rate = (cancer cases / population)*100

and processed by a variety of GIS techniques to produce maps displaying the cancer incidence rates in relation to communication towers densities and agriculture density maps.

Thirdly, we examine whether there is any statistical relationship between cancer cases and population. In detail, we are looking for relationships between Breast, Leukemia, and Brain cancer incidence rates and the communication towers density. Furthermore, it is assumed that there is a relationship between Prostate and Bladder cancer incidence rates and the agriculture density. Moreover, we try to predict the number of cancer cases in 2014 for Blue Nile and Khartoum states by using population projections.

3.2 Study Design

The experimental study design for this study was conducted in Sudan using data

from 1995-2013.

3.3 Data Analysis

The data was analyzed using correlation analysis to compare the cancer incidence rates with population projections, agriculture densities and communication towers densities. The statistical probability for significance was set to a P-value of ≤ 0.05.In a regression analysis it is tried to predict numbers of cancer cases for 2014, in combination with spatial analysis techniques (Georeferencing). A heatmap analysis is used to generate a communication towers density map.

3.4 Geographical Information System(GIS)

A Geographical Information System is a computer-based system for the acquisition and update, storage and query, analysis and simulation as well as output and presentation of spatial data. The geographical information system is also called a geographic information system or geospatial information system (Margaret and Kenneth, 2014).

GIS is a" Set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes"(Burrough et al., 2015)

3.4.1 Components of GIS

- Hardware
- Software
- Data
- People (Pandeyand Pathak ,2014).

3.4.2 GIS Sections

- Land Information Systems
- Network Information Systems
- Space Information Systems
- Environmental Information Systems
- Domain Information Systems(Margaret and Kenneth, 2014).

3.4.3 Classification of GIS Functions

A. Data input and encoding
    - Data capture, e.g. digitizing and integration of external data
    - Data validation and editing
    - Data structuring and storage, e.g. construction of different kinds of surfaces and data coding.

B. Data manipulation
    - Structure conversion, e.g. conversion from vector to raster.
    - Geometric conversion, e.g. map registration, scale changes, various transformations, map projection change.
    - Generalization and classification, e.g. reclassifying data, aggregating or disaggregating data.
    - Integration, e.g. combining layers of different surfaces.
    - Enhancement, e.g. image edge enhancement.
    - Abstraction, e.g. calculations of area centroids.

C. Data retrieval
    - Selective retrieval of information based on user-defined themes or criteria, including “browse” facilities

D. Data analysis
    - Spatial analysis, e.g. route allocation, slope and aspect calculations.
    - Statistical analysis, e.g. histograms, frequency analysis, measures of dispersion.

- Measurement, e.g. line length, area and volume calculations, distance and directions

E. Data display

- Graphical display, e.g. maps, graphs.
- Textual display, e.g. report writing, production of tables (Chun ,2010).

3.4.4 GIS Data

GIS data can be separated into two categories: spatially referenced data which is represented by vector and raster formats (including imagery) and attribute tables which is represented in tabular format (Neteler and Mitasova , 2013**).**

3.4.4.1 Vector:

Vector data is split into three types: polygon, line (or arc) and point data. Polygons are used to represent areas such as the boundary of a city**.** Polygon features are two-dimensional and therefore can be used to measure the area and perimeter of a geographic feature.

Line (or arc) data is used to represent linear features. Common examples would be rivers, trails, and streets. Line features only have one dimension and therefore can only be used to measure length. Line features have a starting and ending point**.**

Point data is most commonly used to represent non-adjacent features and to represent discrete data points. Points have zero dimensions (Neteler and Mitasova , 2013**).**

3.4.4.2 Raster:

Raster data (also known as grid data) represents the fourth type of feature: surfaces. Raster data is cell-based and this data category also includes aerial and satellite imagery. There are two types of raster data: continuous and discrete. An example of discrete raster data is population density. Continuous data examples are temperature and elevation measurements. There are also three types of raster datasets: thematic data, spectral data, and pictures (imagery) (Neteler and Mitasova , 2013**).**

3.4.5 Examples of GIS Software

For this thesis the following GIS software can be used: Esri’s ArcGIS suite, MapInfo, QGIS, and GRASS (Margaret and Kenneth, 2014).

3.5 Quantum GIS

Quantum GIS, in short QGIS, is an open source Geographic Information System, that supports most geospatial vector and raster file types and database formats. Unlike many other open source GIS programs, QGIS is available for a number of operating systems(Linux, Unix, Mac OSX and Windows)(Menke et al., 2015).

3.5.1 Features

The major features of QGIS include:

- Direct viewing and exploration of spatial data
  - Advanced symbology (edit rendering styles)
  - QGIS Browser as a simple and fast data viewer
- Support for numerous vector, raster, and database formats, like
  - GRASS, GeoTiff, USGS ASCII DEM, Erdas Imagine
  - ESRI shape file, MapInfo, SDTS and GML.
- Create, edit and export spatial data
  - Work with nodes, lines and polygons
  - Convert between different coordinate systems (re-projection)
  - Down/upload directly to a GPS unit
  - Georeferencing, Digitizing
  - Delimited text import
- Perform spatial analysis
  - Find basic statistics
  - Distance matrix and line intersections
- Publish user map on the internet
- An extensible plug-in architecture(Plugins written in Python, C++)
- Remote control JOSM
- QGIS integrates with other open-source GIS packages, including PostGIS , GRASS, and MapServer to give users extensive functionality (Ruas , 2011).
- Delimited text plugin:

  Not every GIS data comes as a shape file, or in a spatial format. Often the data would come as a table or a spreadsheet. The plugin allows to load this data as attribute tables (link it with existing spatial data called Join) or layer if data include lat/long coordinates.(Gandhi, 2014)
- Visualization

  In GIS, visualization is used to organize spatial data and related information into layers that can be analyzed or displayed as maps, three-dimensional scenes, summary charts, tables (Gregory and Ell,2007) .

3.6 Heatmap Analysis

Heat mapping, from a geographic perspective, is a method of showing the geographic clustering of a phenomenon. Sometimes also referred as hot spot mapping, heat maps show locations of higher densities of geographic entities.

The ‘heat’ in the term refers to the concentration of the geographic entity within any given spot, not to be confused with heat mapping that refers to the mapping of actual temperatures on the earth’s surface(Dempsey, 2012).

The Heatmap plugin uses Kernel Density Estimation to create a density (heatmap) raster of an input point vector layer. The density is calculated based on the number of points in a location, with larger numbers of clustered points resulting in larger values. Heatmaps allow for an easy identification of "hotspots" and clustering of points (Peterson , 2014).

3.7 Digitizing

In GIS data processing quite often geographic data are converted, either from a hardcopy or a scanned image into vector data by tracing the features. During the digitizing process, features from the traced map or image are captured as coordinates in either point, line, or polygon format (Neteler and Mitasova , 2008).

3.7.1 Types of GIS digitization

There are several types of digitizing methods. Manual digitizing involves tracing geographic features from an external digitizing tablet using a puck (a type of mouse specialized for tracing and capturing geographic features from the tablet). Heads up digitizing (also referred to as on-screen digitizing) is the method of tracing geographic features from another dataset (usually an aerial, satellite image, or scanned image of a map) directly on the computer screen. Automated digitizing involves using image processing software that contains pattern recognition technology to generated vectors .Georeferencing is a necessary step in the digitizing process (Neteler and Mitasova , 2008)

3.7.2 Georeferencing

There is a great deal of geographic data available in formats, that cannot be immediately integrated with other GIS data. In order to use these types of data in GIS it is necessary to align it with existing geographically referenced data, this process is called georeferencing(Dobrica et al., 2010).
In detail, georeferencing is the process of assigning real-world coordinates to each pixel of the raster. Many times these coordinates are obtained by doing field surveys - collecting coordinates with a GPS device for a few easily identifiable features in the image or map (Gregory and Ell,2007).

3.8 R Software

R is a free software environment for statistical computing and graphics, supported by the R Foundation for Statistical Computing. It runs on Linux, Windows and MacOS platforms (Menkeet al., 2015).

R was created by Ross Ihaka and Robert Gentleman (Morandat et al., 2012).

What is less well known, is that R also has cutting edge spatial packages that allow for behaving as a fully featured Geographical Information System in the full sense of the word. In fact, some of cutting edge algorithms for image processing and spatial statistics are implemented in R before any other widely available software product (Bivandet al., 2015).

3.8.1 The advantages and drawbacks of R as a GIS:

Despite being able to perform the same operations as dedicated GIS software such as ArcGIS and QGIS, R is fundamentally different in the way, that the user interacts with it. Not only are most operations complete by typing (e.g. you type "plot(map1)" to plot the data contained in the map1 object), the visualization stage is different. There is no dynamic canvas which can be used to pan and zoom - instead R only produces visual or other types of output when commanded to do so, using functions such as plot(Lovelace, 2014).

Table (3.1): A summary of the relative merits of R compared with more traditional GIS software (Lovelace, 2014).

Attribute	**Advantages of R**	**Drawbacks of R**
User interface	Command line interface allows rapid description of workflow and reproducibility	Steep learning curve (eased by RStudio)
Visualizing data	Sophisticated and customisable graphics	No dynamic zoomable canvas
Selecting data	Concise and consistent method using square brackets (e.g. "map1[x > y,]")	Difficult to dynamically select objects from map
Manipulating data	Very wide range of functions through additional packages	Only single core processing
Analyzing/modeling data	Integrated processing, analysis, and modeling framework	Sometimes more than one solution available

3.8.2 Use R Scripts in QGIS

R has grown a lot in the past years and is increasingly used as GIS. For many spatial operations and even spatial statistics certain R packages are simply the best choice up to date.

However R as GIS lacks the simplicity in matters of design and it requires quite an effort to make a good-looking map. Therefore, using QGIS for map-design and output is always the favored choice for developers (Jung, 2013) .

3.9 Statistical Package for the Social Sciences (SPSS)

The Statistical Package for the Social Sciences (SPSS) is a software package used in statistical analysis of data. It was developed by SPSS Inc. and acquired by IBM in 2009. In 2014, the software was officially renamed IBM SPSS Statistics (Asthana and Bhushan, 2016).
SPSS statistical package is one of the most popular statistical packages which can perform highly complex data manipulation and analysis with simple instructions. It was originally meant for the social sciences, but has become popular in other fields. SPSS has four windows - Data editor; Output viewer; Syntax editor and the Script window (Gerber andVoelkl ,2012).

3.9.1 IBM SPSS Statistics Features

The IBM SPSS Statistics is available in a base version and three editions:

- SPSS Statistics Base provides a range of statistical procedures suitable for many problems, including crosstabs, linear regression, Monte Carlo simulation, geospatial analytics, and the ability to extend built-in capabilities with Python, R, or Java code.
- SPSS Statistics Standard includes techniques such as logistic and non-linear regression and presentation quality custom tables to help business managers and analysts.
- SPSS Statistics Professional addresses non-standard analyses and issues such as data quality and automation with features such as automatic data preparation, decision trees and forecasting.
- SPSS Statistics Premium supports enterprise businesses needing advanced techniques such as structural equation modeling (SEM), sampling assessment and testing, and small sample and rare occurrence analysis (International Business Machines , 2016).

3.9.2 Benefits

- Quality decision-making: Quickly gain understanding and insights from datasets in any format using advanced statistical procedures, ensuring high accuracy to drive quality decision-making.

- More power for advanced users: Dramatically increase analytical power, flexibility and productivity through programming options such as command syntax and external programming languages like R, Python, and others.

- Easily communicate results: Reveal deeper insights and provide better confidence intervals via presentation-ready reports, visualizations and geographic spatial analysis (International Business Machines , 2016).

3.10 Correlation Analysis

Correlation analysis is a statistical method to measure the degree of relationship between two variables.

Correlation coefficients are expressed as values between +1 and -1. The sign of the correlation coefficient (+ , -) defines the direction of the relationship, either positive or negative. A positive correlation coefficient means change in the value of one variable will predict a change in the same direction in the second variable. A negative correlation coefficient indicates, that a change in the value of one variable predicts a change in the opposite direction in the second variable. Taking the absolute value of the correlation coefficient measures the strength of the relationship. A coefficient of +1 indicates a perfect positive correlation , -1 indicates a perfect negative correlation and zero indicates there is no relationship at all(Sharma , 2005).

3.10.1CorrelationEquations

3.10.1.1 Pearson correlation coefficient

$$r_p = \frac{N\sum xy - \sum (x)(y)}{\sqrt{N\sum x^2 - \sum (x^2))][N\sum y^2 - \sum (y^2)]}}$$

Where:
r = Pearson r_p correlation coefficient
N = number of value in each data set
$\sum xy$ = sum of the products of paired scores
$\sum x$ = sum of x scores
$\sum y$ = sum of y scores

$\sum x^2$= sum of squared x scores
$\sum y^2$= sum of squared y scores(Sharma , 2005).

3.10.1.2 Spearman rank correlation

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Where:
P= Spearman rank correlation
d_i= the difference between the ranks of corresponding values x_i and y_i
n= number of value in each data set(Sharma , 2005).

3.11 Regression Analysis (RA)

The term regression analysis used to describe data and to explain the relationship between one dependent variable and one or more independent variables (Retherford andChoe , 2011).

3.11.1 Major Uses for Regression Analysis

(1) causal analysis

(2) forecasting an effect

(3) trend forecasting(Retherford and Choe , 2011).

3.11.2 Regression Analysis Method

- linear regression:
  is using a model-dependent variable as linear function of the independent variable .
  - simple linear regression uses one independent variable with one dependent variable in the analysis.
  - multiple linear regression uses two or more independent variables with one dependent variable in the analysis (Retherford and Choe , 2011).
- non-linear regression:
  is using a model-dependent variable as a polynomial function of the independent variable .When plotted as a function of the independent variable it forms a curve(Retherford and Choe , 2011).

3.11.3 Simple Linear Regression Equation

If we have a simple linear relation between an independent variable x and a dependent variable y the following linear equation is valid:

$y = a + bx$

with

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$a = \frac{\sum y - b\sum x}{n}$$

Where :

y = dependent variable

x = independent variable

b = The slope of the regression line

a = The intercept point of the regression line along the y axis.

N = Number of values or elements

Σxy = Sum of the product of x and y Scores

Σx = Sum of x Scores

Σy = Sum of y Scores

Σx^2 = Sum of square x Scores(Retherford and Choe , 2011).

The relation above is called a linear univariate relation, in contradiction to linear multivariate relations, to be defined as

$y = a + bx_1 + cx_2 + dx_3 + \ldots$

with b, c and d as regression parameters and x_1, x_2, x_3 as independent variables.

If we define the regression as an univariate polynomial function this looks as follows:

$y= a + bx + cx^2 +dx^3 + ex^4 + \ldots$

with a, b, c, d and e as regression parameters.

3.12 Ethical Considerations

For this study we obtained approvals from the College of Graduate Studies ,Sudan University of Science and Technology, the Department of Statistics in the Federal Ministry of Health, the central bureau of statistics and the National Telecommunication Corporation, which is gratefully acknowledged.

CHAPTER FOUR

DATA AND TESTBED

4.1 Data Categories

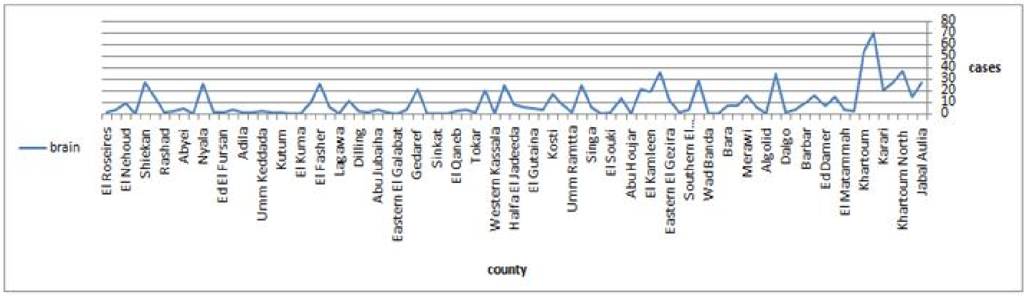
Four data categories were used in this study: Firstly, we have got cancer cases data (1995-2013) from the Department of Statistics in the Federal Ministry of Health as shown in Table (4.1) , Fig (4.1) ,Fig (4.2), Fig (4.3), Fig (4.4) and Fig (4.5). Secondly, for the environmental factors we have got communication towers data from The National Telecommunication Corporation as shown in Fig (4.6). Thirdly, agriculture heat map data were downloaded from internet (Eltoum et al. , 2015) as shown in Fig (4.7), and fourthly, population projections (1995-2013) as shown in Table (4.2). we got from the Central Bureau of Statistics, together with baseline data (counties and states boundaries as shown in Fig (4.8)).

Table (4.1): Cancer cases of Sudan States for the Period 1995 to 2013

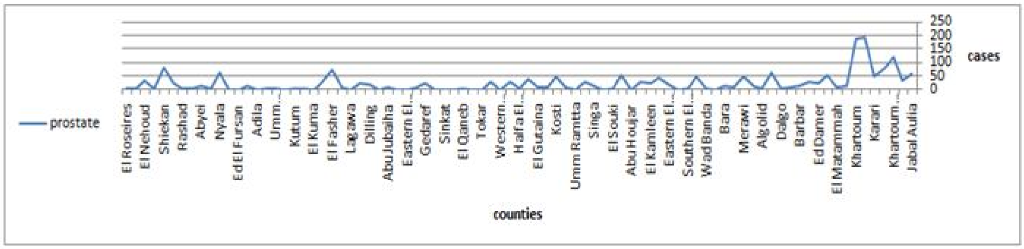
State	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Khartoum	1044	1517	1528	1554	1883	1913	1983	2171	2445	3287	3477	3877	4404	4740	7756	8907	10213	10377	13905
Sinnar	0	3	5	12	18	25	32	41	53	68	89	112	137	151	191	209	215	217	223
Northern	5	7	12	19	33	56	78	94	128	147	152	164	189	203	237	345	368	371	437
Blue Nile	0	3	4	5	6	8	11	13	14	17	21	22	23	24	28	60	62	65	98
North Darfur	0	2	5	18	27	48	65	89	101	122	138	142	149	157	160	220	291	297	304
Kassala	11	12	16	17	19	24	34	42	43	54	59	65	66	94	99	109	180	475	828
AlGedarif	9	10	11	16	17	26	44	45	51	56	60	64	68	74	116	136	156	158	291
Red Sea	5	12	17	28	32	37	58	62	67	72	79	83	87	93	98	175	186	189	190
River Nile	0	6	23	49	84	107	148	165	198	223	254	281	307	362	383	426	427	435	472
Al Gezira	22	70	83	90	100	102	112	121	131	196	305	370	398	546	750	925	997	1196	1273
White Nile	0	5	9	17	23	138	153	179	195	206	249	262	288	301	342	487	499	504	513

Table (4.2): Population projections (in thousand) of Sudan States for the Period 1998 to 2013

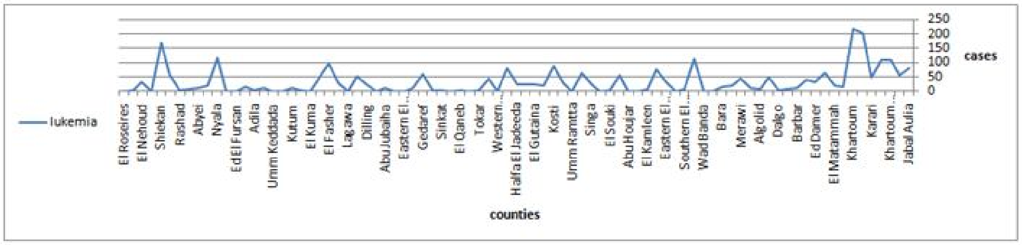
State	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Khartoum	4372	4568	4740	4936	5139	5274	5325	5515	5553	5758	5761	5974	6007	6203	6268	6535
Sinnar	1140	1145	1173	1204	1236	1268	1285	1301	1334	1340	1368	1396	1404	1454	1518	1580
Northern	562	573	582	593	603	614	624	634	644	654	699	721	741	764	789	814
Blue Nile	599	618	636	655	675	696	716	738	759	783	832	858	882	911	936	966
North Darfur	1364	1412	1455	1503	1552	1603	1655	1709	1763	1821	2112	2181	2237	2295	2354	2414
Kassala	1398	1435	1469	1507	1545	1584	1625	1666	1708	1752	1790	1821	1910	1988	2061	2134
AlGedarif	1348	1370	1419	1420	1465	1495	1515	1567	1573	1621	1655	1674	1728	1739	1784	1843
Red Sea	713	717	721	724	728	732	734	737	739	740	1308	1317	1328	1346	1367	1396
River Nile	866	883	900	918	936	954	972	996	1008	1026	1120	1158	1192	1230	1269	1309
Al Gezira	3177	3280	3374	3477	3575	3583	3692	3710	3797	3845	3905	3989	4014	4130	4133	4285
White Nile	1402	1441	1476	1515	1555	1595	1636	1676	1718	1731	1762	1796	1867	1937	2012	2087



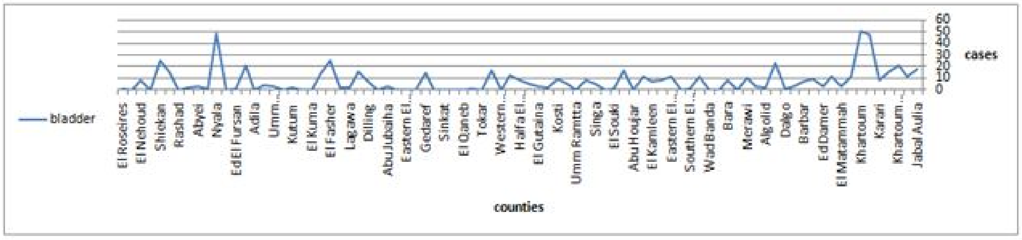
Fig(4.1): Brain Cancer cases of Sudan counties for the Period 2010-2013



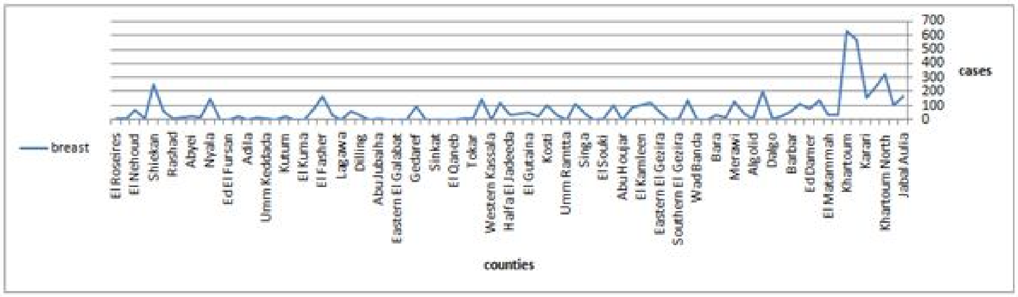
Fig(4.2): Prostate Cancer cases of Sudan counties for the Period 2010-2013



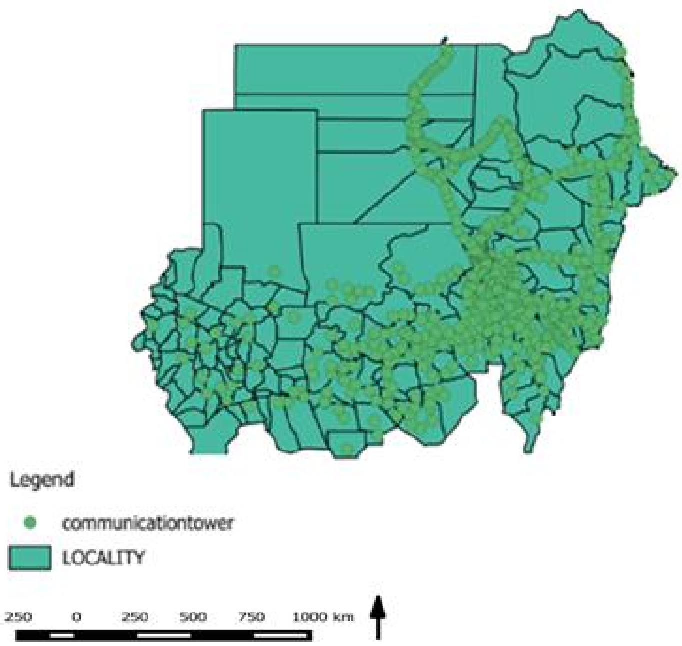
Fig(4.3): Lukemia Cancer cases of Sudan counties for the Period 2010-2013



Fig(4.4): Bladder Cancer cases of Sudan counties for the Period 2010-2013



Fig(4.5): Breast Cancer cases of Sudan counties for the Period 2010-2013



Fig(4.1): Map showing Distribution of communication towers in Sudan

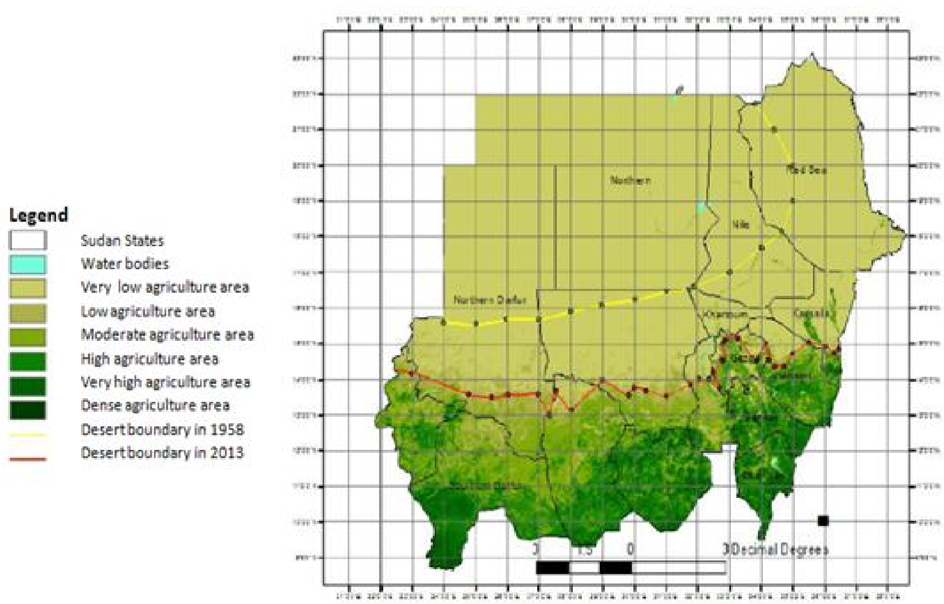


Fig (4.2): Map showing agriculture in Sudan

4.2 Study Area

Sudan is a country located in the northeastern part of Africa between the 22.4 latitude northern equator and 38.22 longitude, with an area of 1,886,068 km^2 and divided into 18 states containing 172 counties (see Fig 4.3). Its capital city is Khartoum, which is located in the northeast central part. The population in the beginning of the year 2011 is estimated to be about 33.419.625 persons at a growth rate of 2,53% annually. Thus the population density reaches 14 persons for one km square. The population of rural areas constitutes about 24,6% of the total population (National Information Centre , 2015).

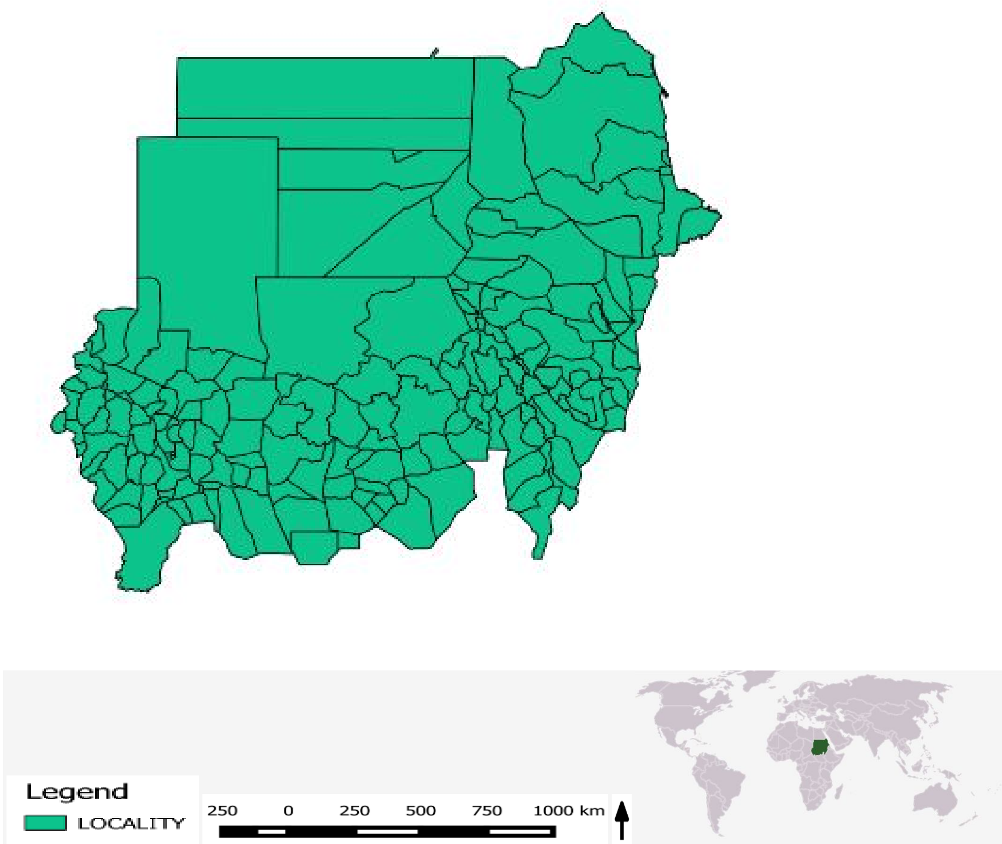


Fig 4.3: Map showing the counties of Sudan.

CHAPTER FIVE
RESULTS

5.1 Results

The following figures demonstrate the application of the proposed methods.

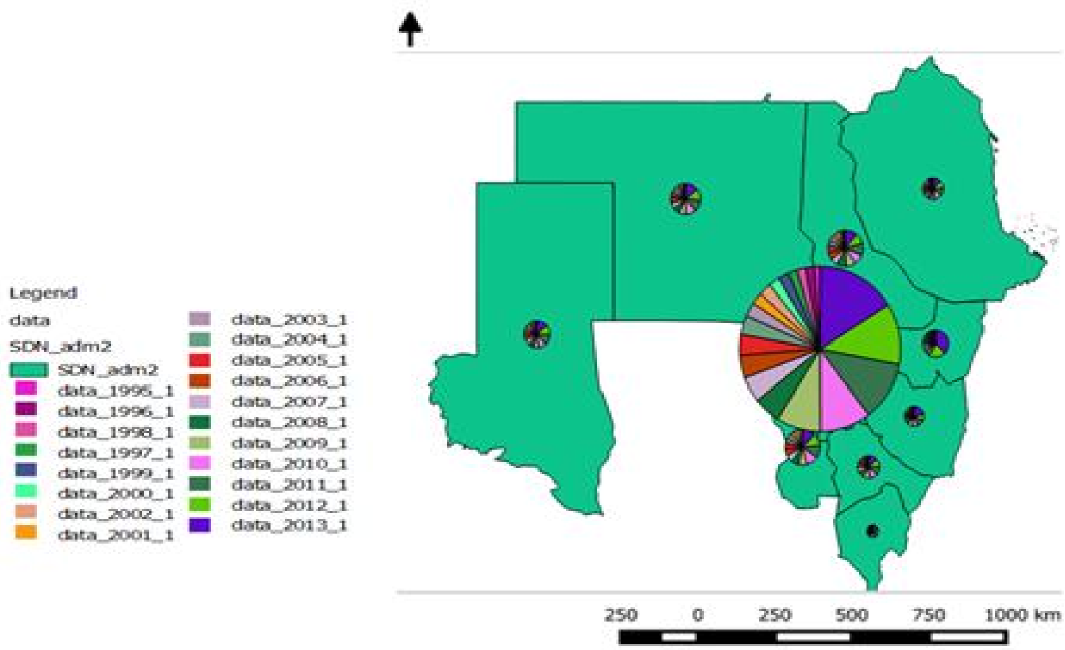


Fig (5.1): Map showing the cancer incidence statistics in Sudan by states at period 1995 –2013.

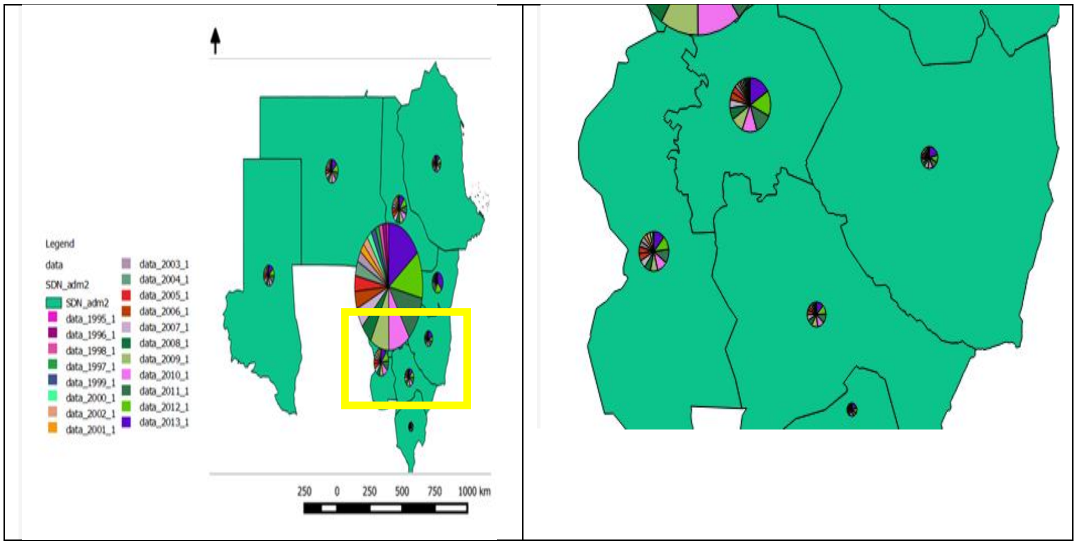


Fig (5.2): Map showing the cancer incidence statistics in Algedaref, White Nile, Sennar, Algazira states– Sudan at period 1995 –2013.

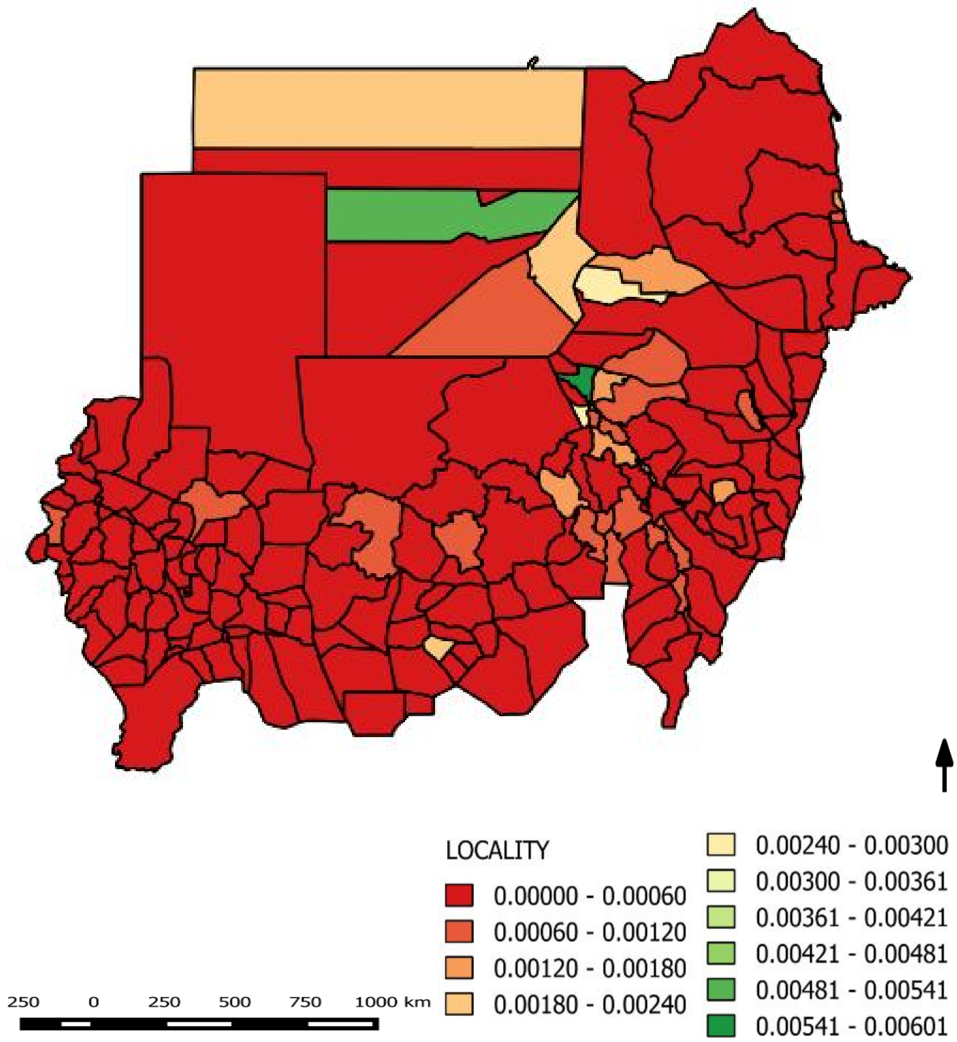
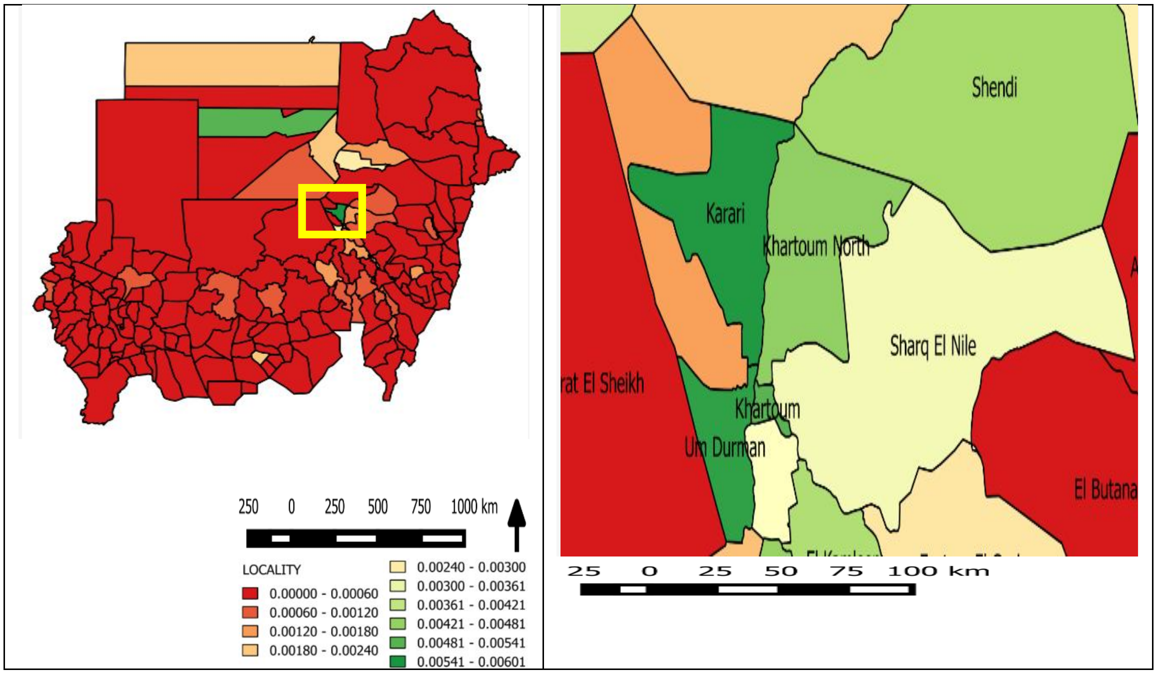


Fig (5.3): Map showing the brain cancer incidence rate in Sudan by county at period 2010 –2013.



Fig(5.4): Map showing the county has very high density of brain cancer incidence in the Sudan at period 2010–2013.

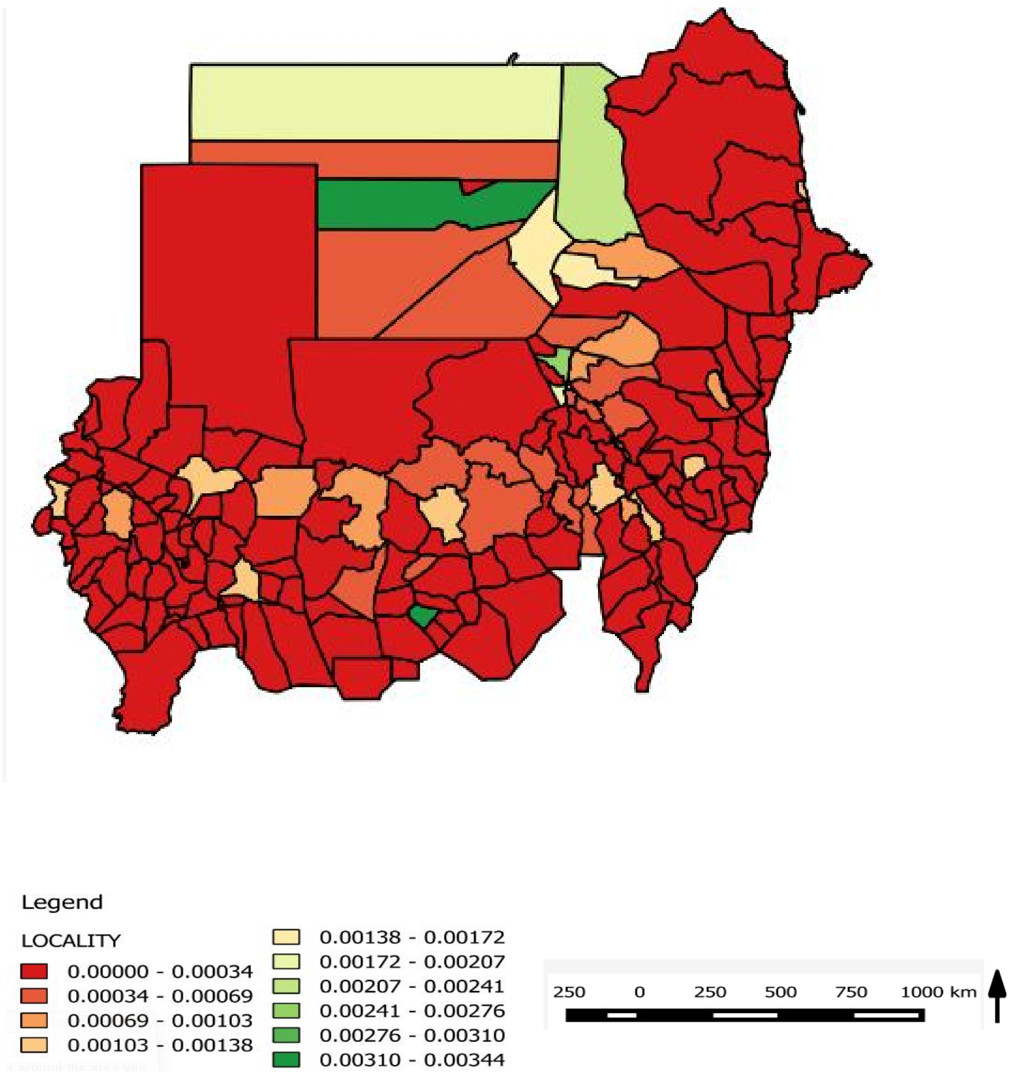


Fig (5.5): Map showing the bladder cancer incidence rate in the Sudan by county at period 2010 –2013.

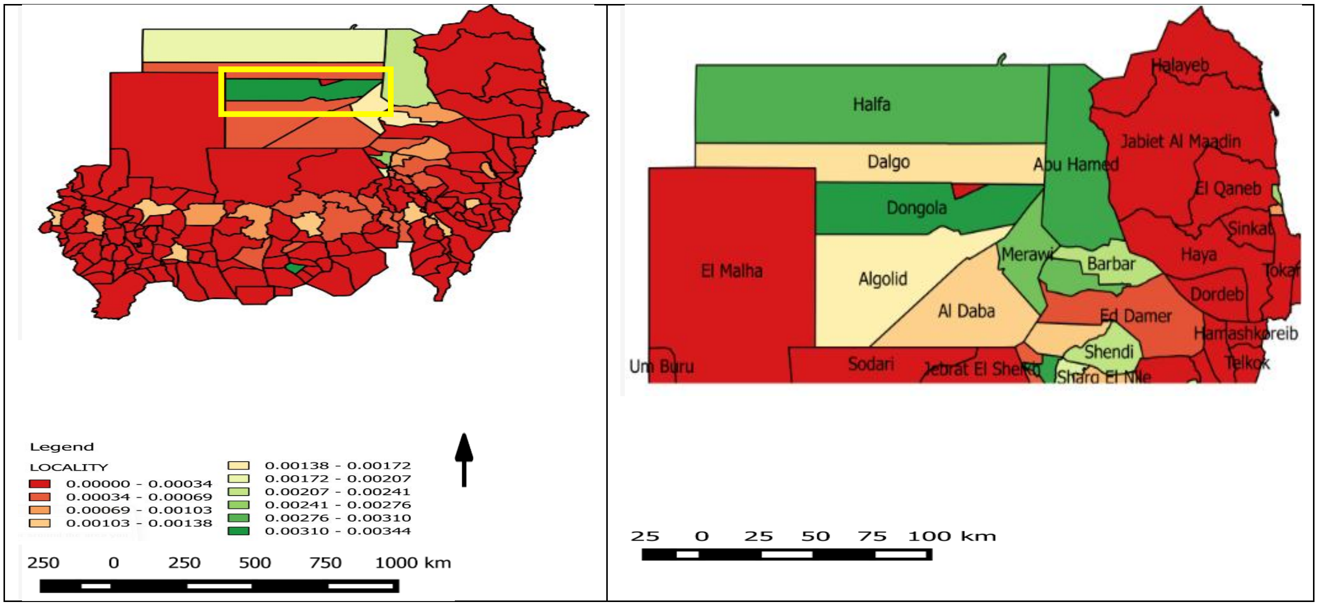


Fig (5.6): Map showing the county has very high density of bladder cancer incidence in the Sudan at period 2010–2013.

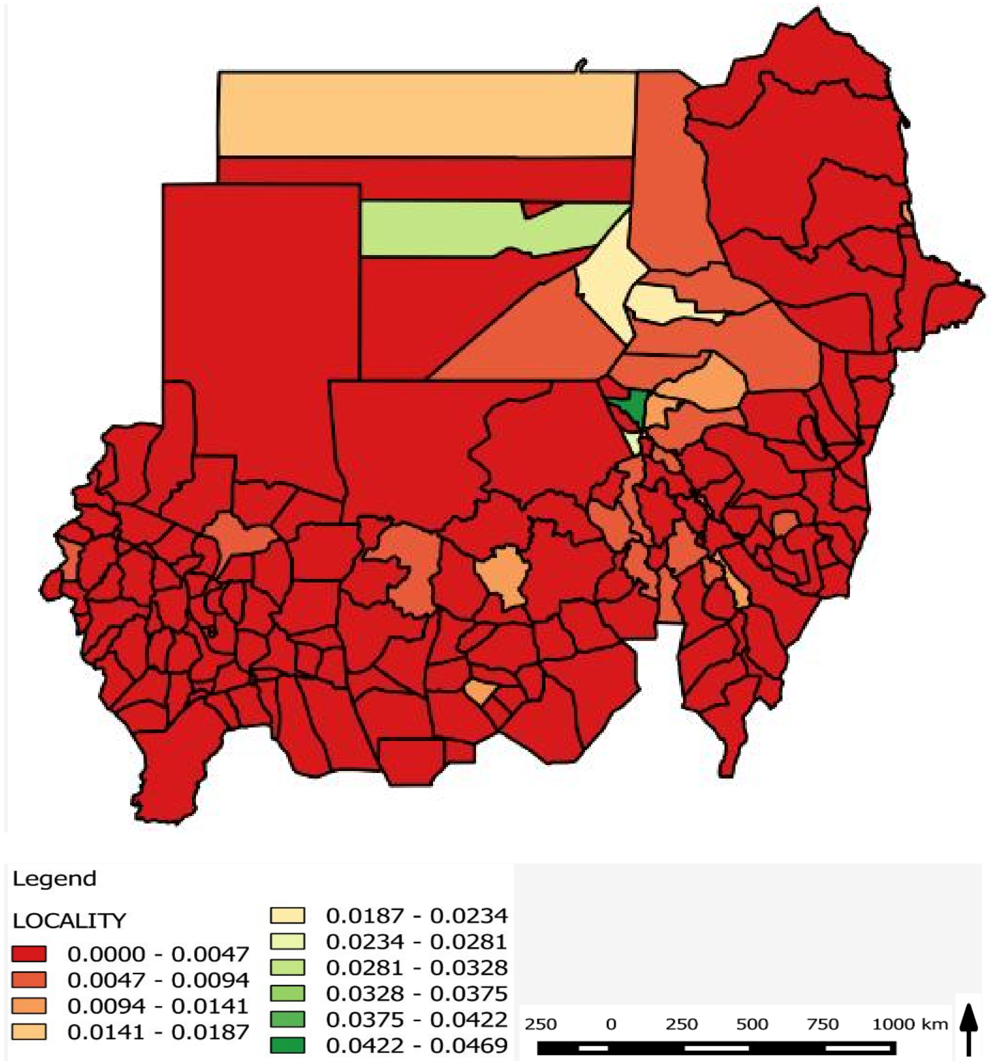


Fig (5.7): Map showing the breast cancer incidence rate in the Sudan by county at period 2010 –2013.

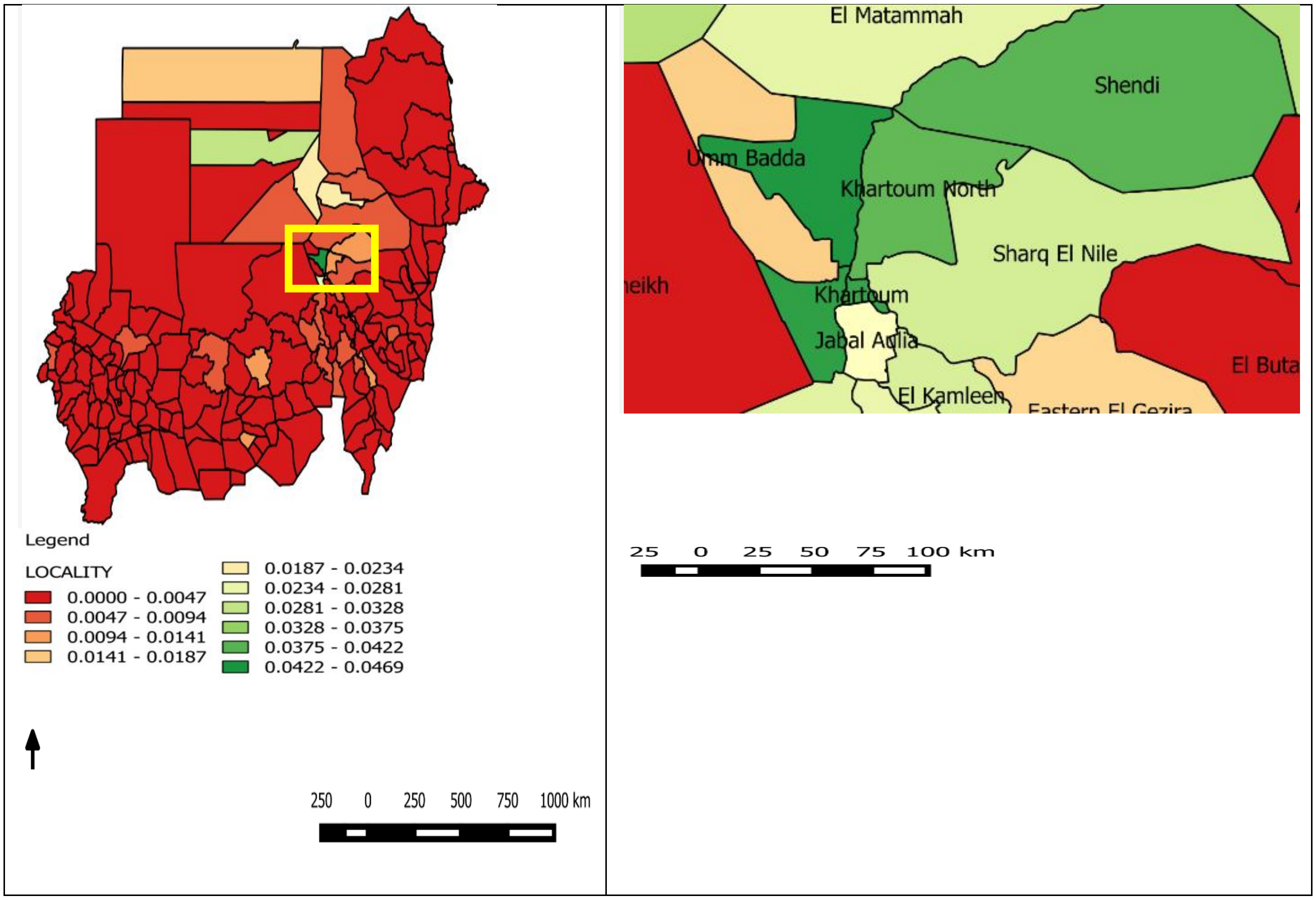


Fig (5.8): Map showing the county has very high density of breast cancer incidence in the Sudan at period 2010–2013.

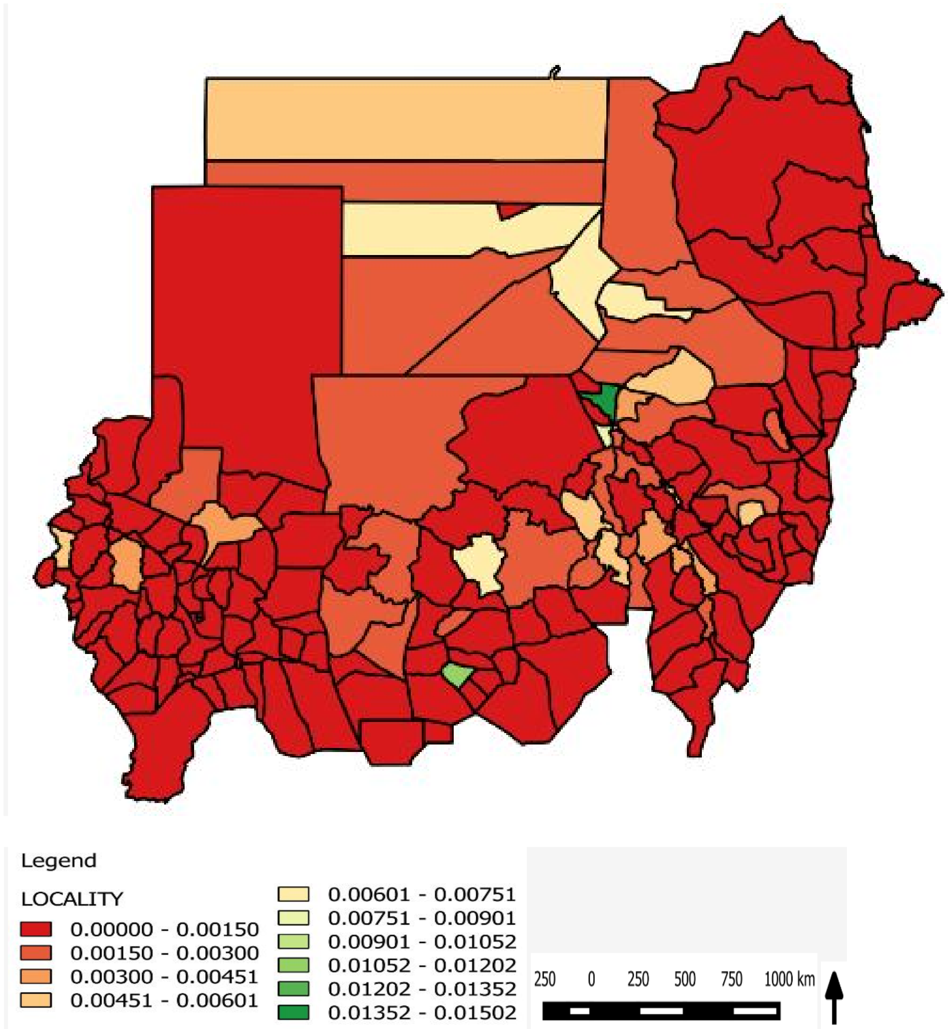


Fig (5.9): Map showing the leukemia cancer incidence rate in the Sudan by county at period 2010 –2013.

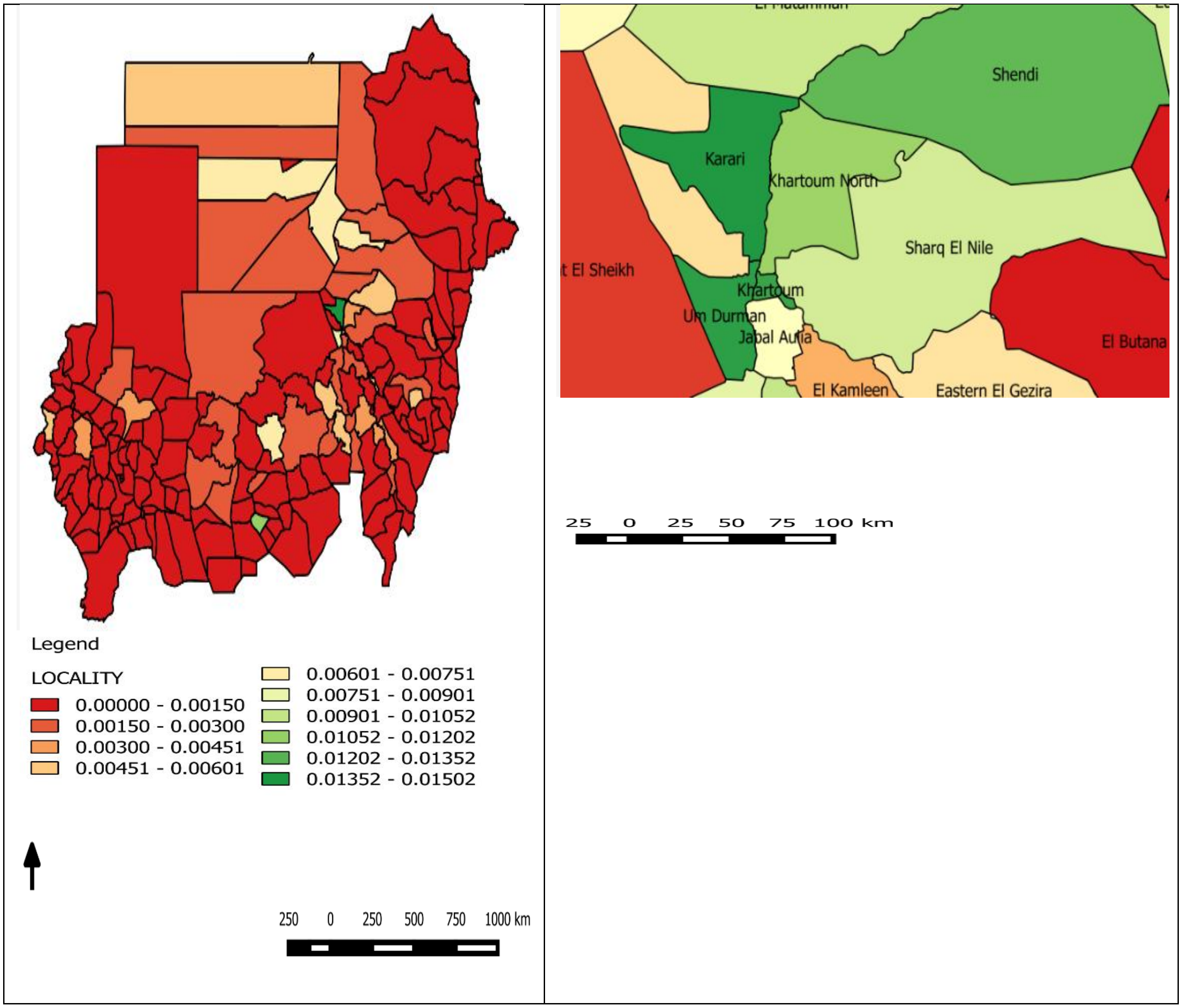


Fig (5.10): Map showing the county has very high density of leukemia cancer incidence in the Sudan at period 2010–2013.

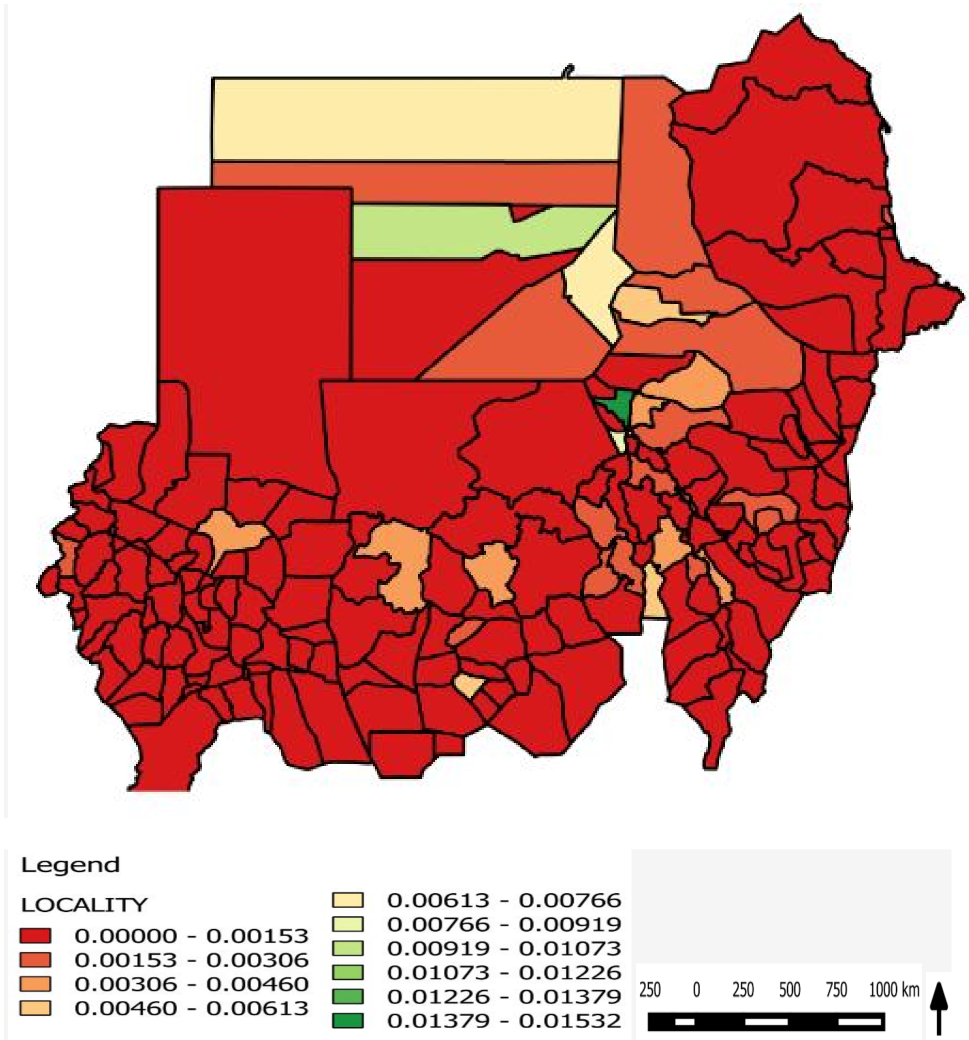


Fig (5.11): Map showing the Prostate cancer incidence rate in the Sudan by county at period 2010 –2013.

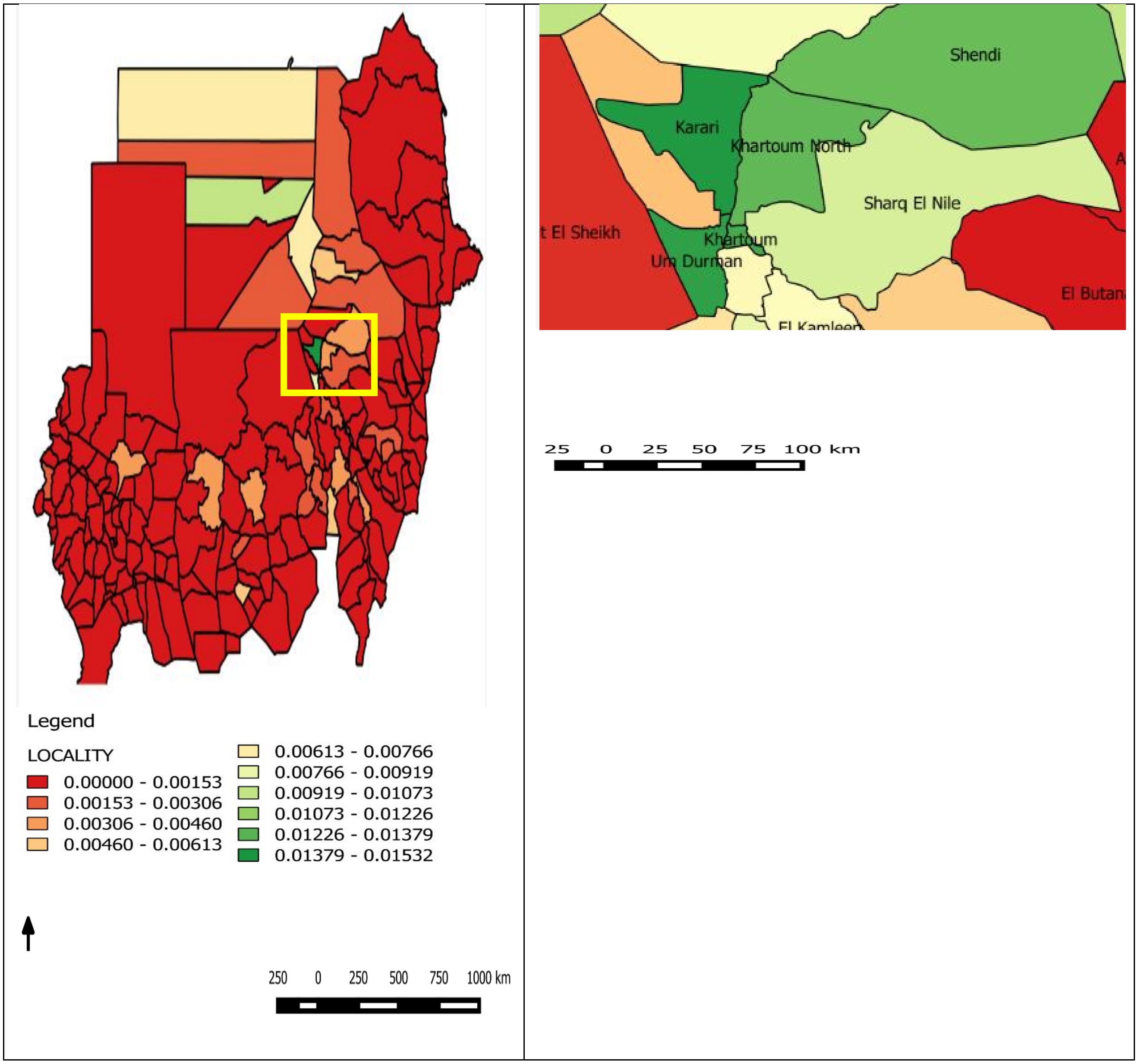


Fig (5.12): Map showing the county has very high density of prostate cancer incidence in the Sudan at period 2010–2013.

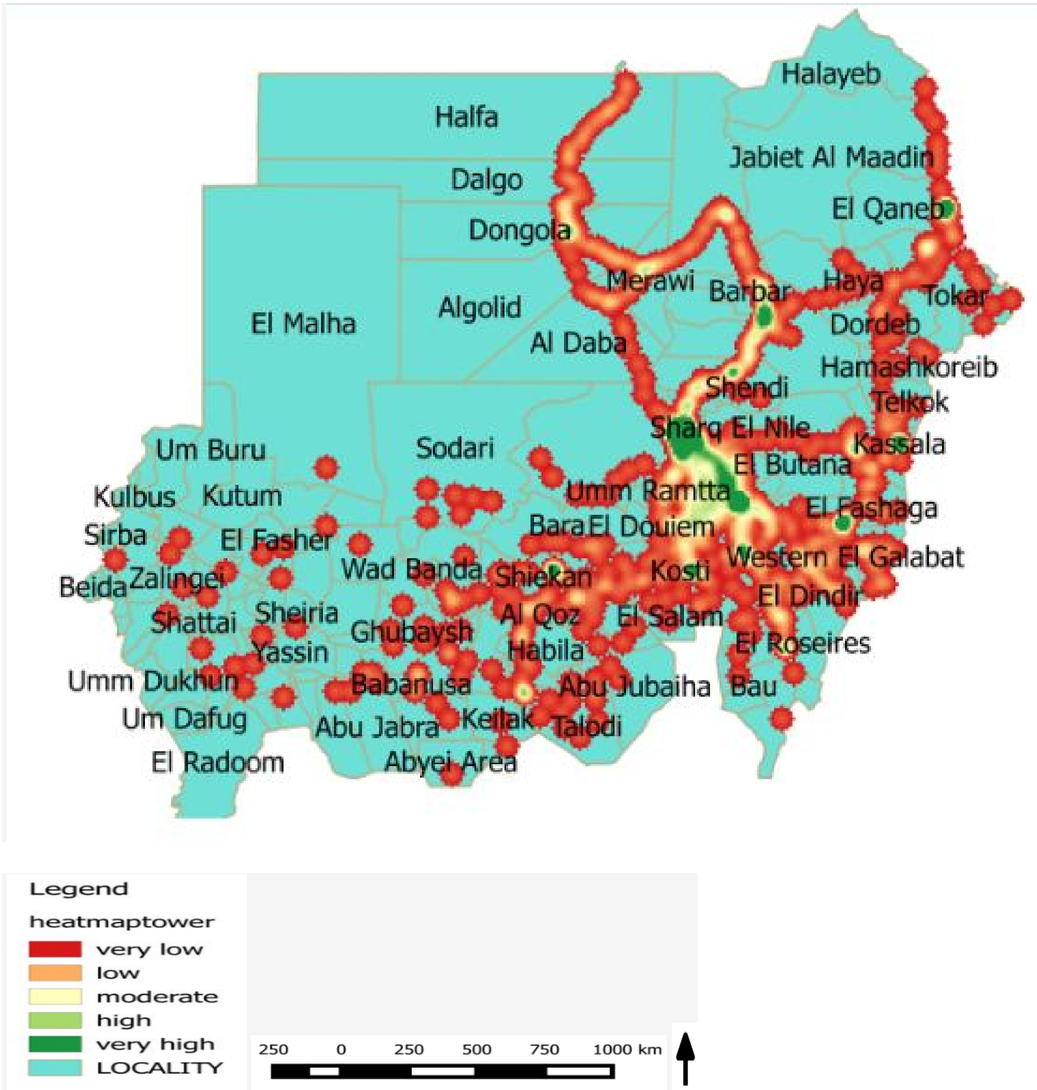


Fig (5.13): Map showing heatmap analysis of communication towers in Sudan

Regression analysis results:

The prediction of cancer incidence in Khartoum state at 2014 is as follows:

***number of cases in 2013=13905 cases - Number of population 6,809,000

Table (5.1): cancer cases in Khartoum state at 2014

Fit	Lower	Upper
11282.32 cases	4985.86 cases	17578.78 cases

The prediction of cancer incidence in Blue Nile state at 2014 is as follows:

***number of cases in 2013=98 cases - Number of population 992,000

Table (5.2): cancer cases in Blue Nile state at 2014

Fit	Lower		Upper
75.90946cases	47.19295cases		104.626cases

Table (5.3): The correlation coefficients of cancer incidences with population by states in Sudan , 1998–2013.

State	correlation coefficient
Khartoum	0.7696405
Northern	0.9863779
Blue Nile	0.9061369
North Darfur	0.9497851
Algedaref	0.3061685
Algazira	0.7206473
Kassala	0.9476851
RedSea	0.831527
River Nile	0.9706226
Sennar	0.9009
White Nile	0.974095

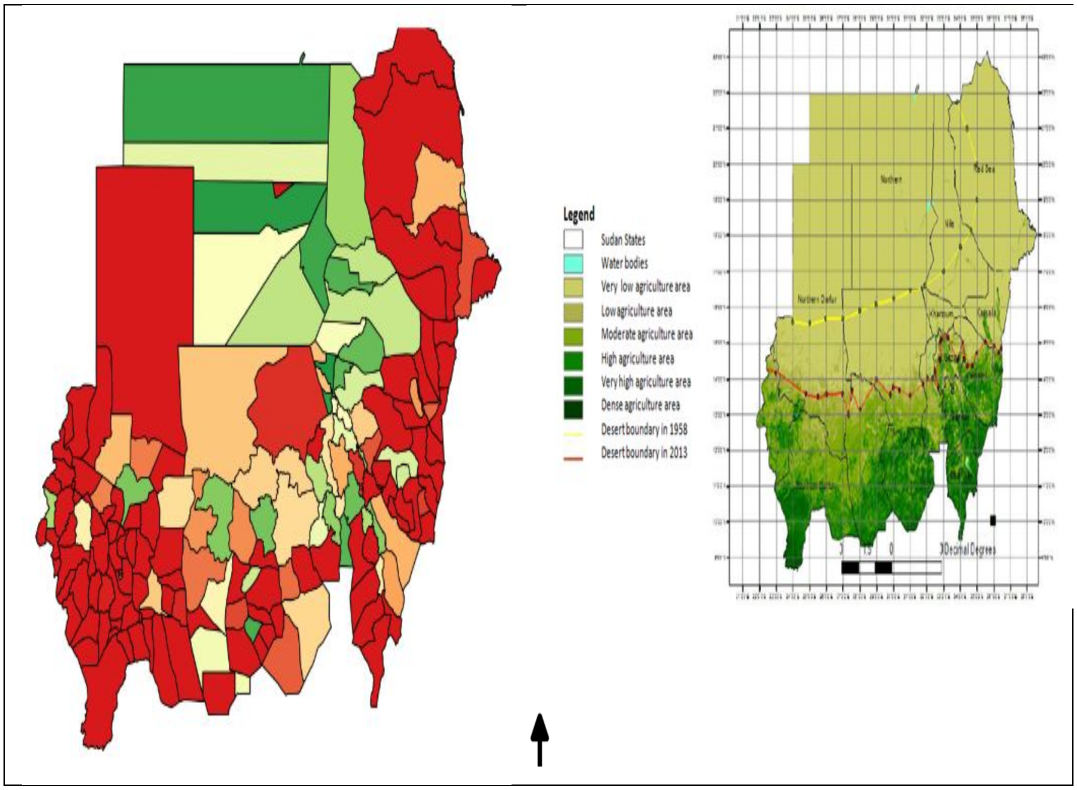


Fig (5.14): Map showing heatmap analysis of agriculture and Prostate cancer incidence rates in Sudan, 2010-2013

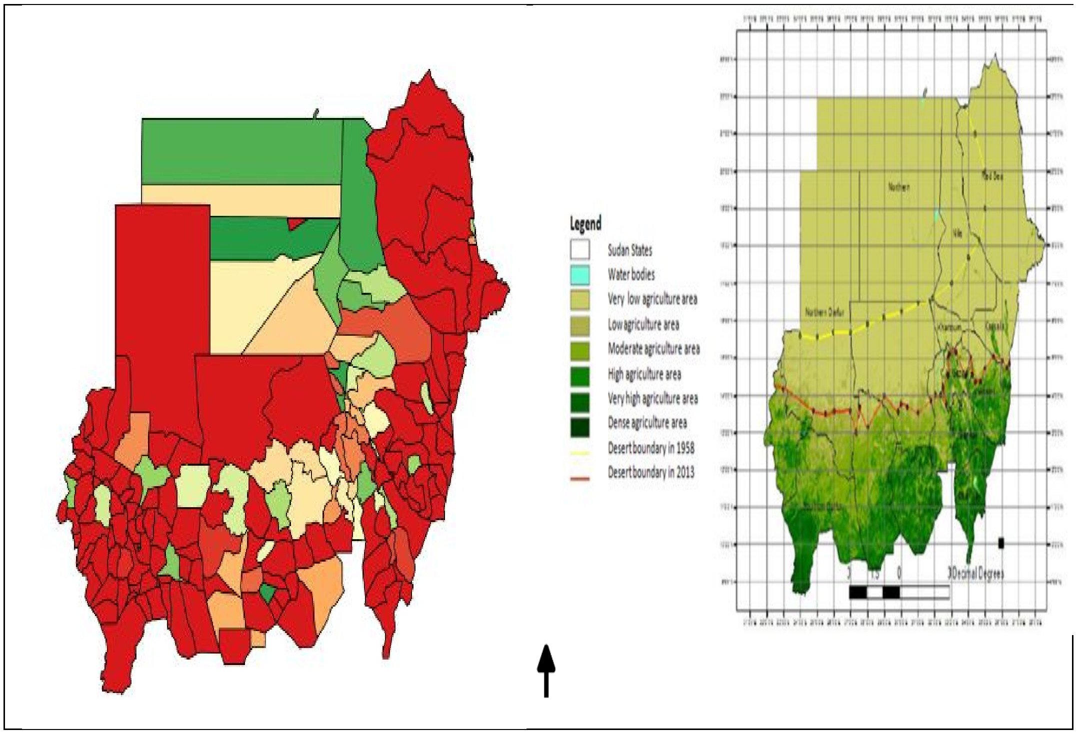


Fig (5.15): Map showing heatmap analysis of agriculture and bladder cancer incidence rates in Sudan, 2010-2013

Table (5.4)The correlation coefficient and P – value of bladder cancer incidence rates *with agriculture density in Sudan , 2010–2013*

Correlation Coefficient	-.160
P – value	.003

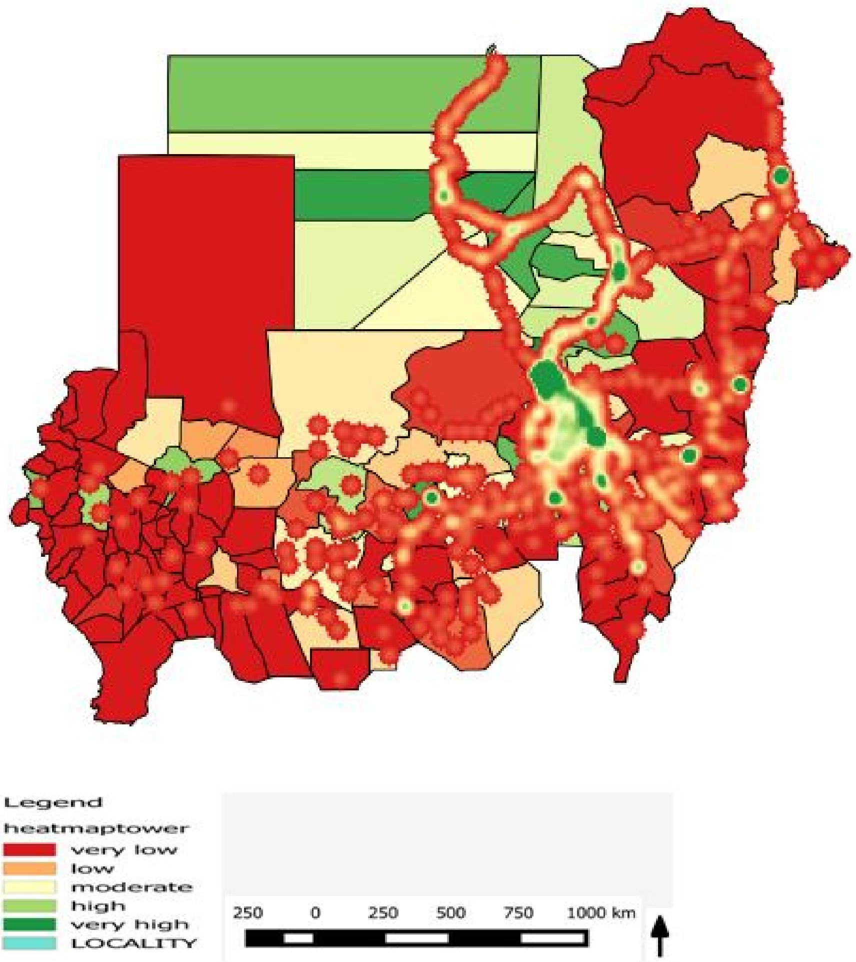


Fig (5.16): Map showing heatmap analysis of communication towers and leukemia cancer incidence rates in Sudan, 2010-2013

Table (5.5) The correlation coefficient and P – value of communication towers density with leukemia cancer incidence rates in Sudan , 2010–2013

Correlation Coefficient	.403
P – value	.000

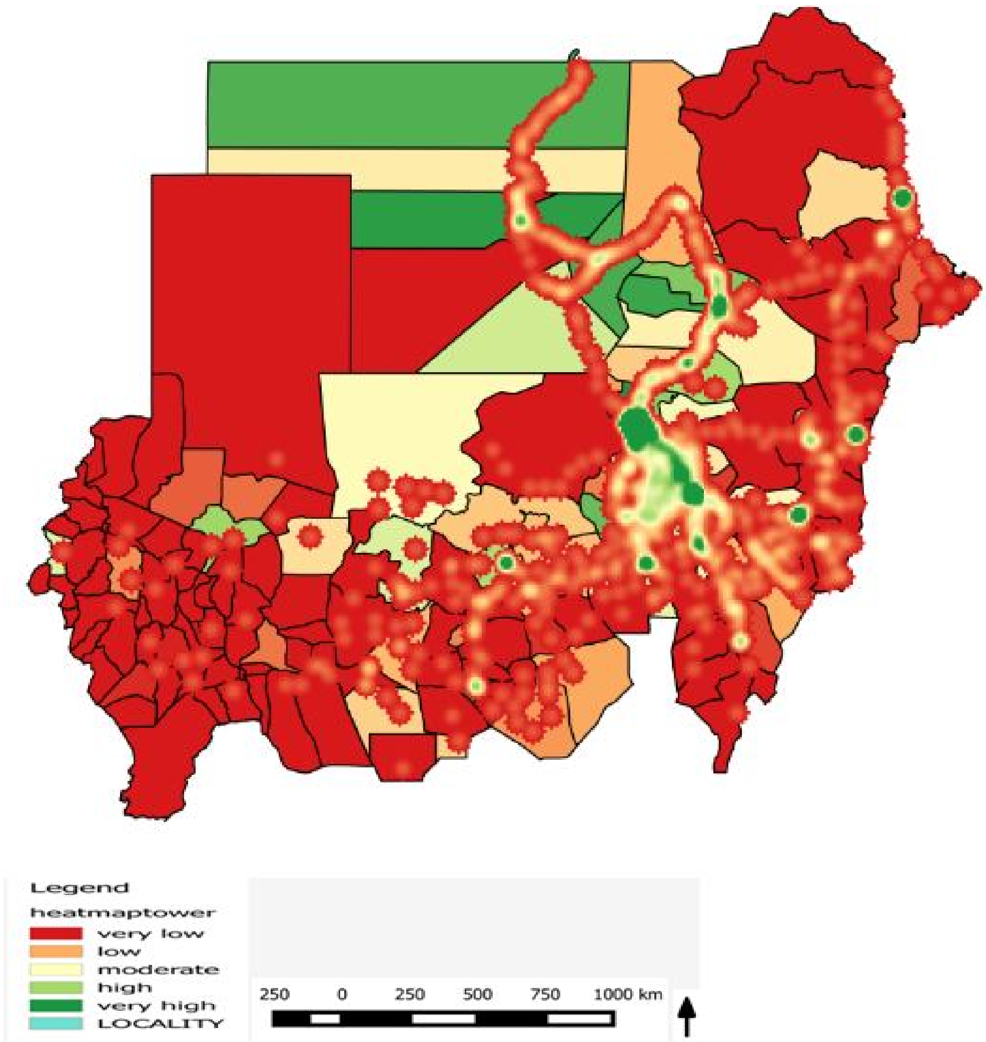
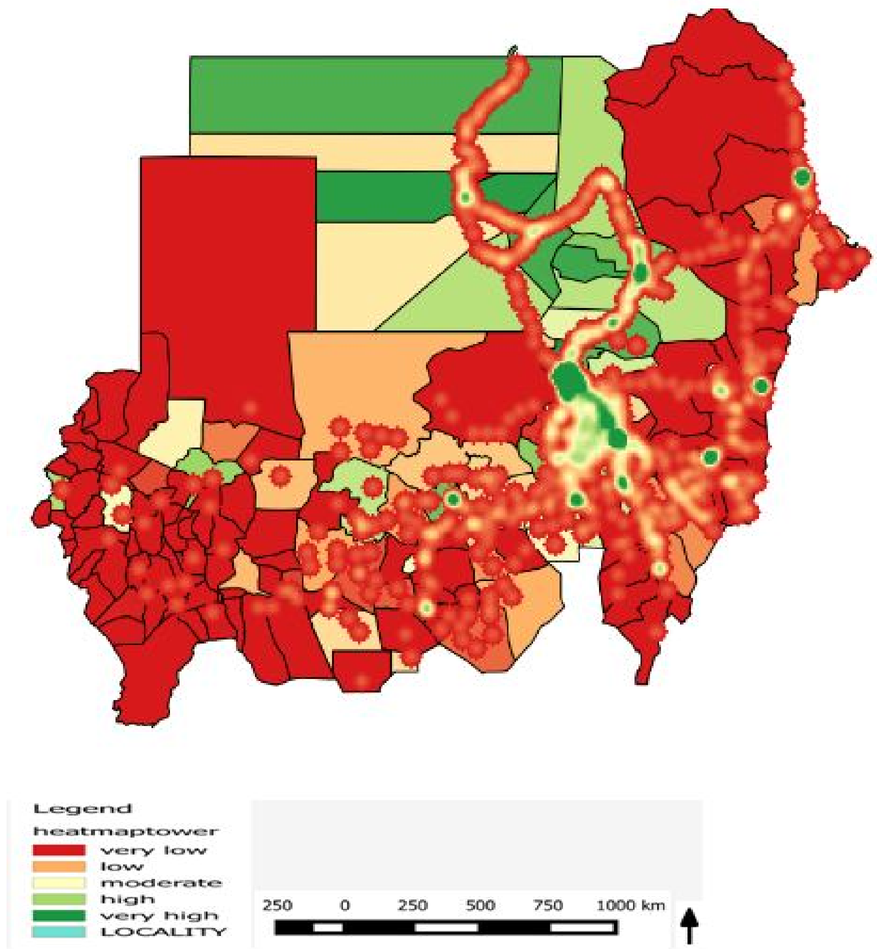


Fig (5.17): Map showing heatmap analysis of communication towers and brain cancer incidence rates in Sudan, 2010-2013

Table (5.6) The correlation coefficient and P – value of communication towers density with brain cancer incidence rates in Sudan , 2010–2013

Correlation Coefficient	.419
P – value	.000



Fig(5.18): Map showing heatmap analysis of communication towers and breast cancer incidence rates in Sudan, 2010-2013

Table (5.7)The correlation coefficient and P – value of communication towers density with breast cancer incidence rates in Sudan , 2010–2013

Correlation Coefficient	.378
P – value	.000

5.2 Discussion

Cancer is a major health issue and one of the most common diseases in the World. In Sudan according to hospitals reports, cancer was the third leading cause of death after malaria and viral pneumonia accounting for 5% of all death in Sudan (Saeed et al., 2013).

This study examined whether there is a correlation between environmental factors and incidences of breast, leukemia, brain, prostate and bladder cancer in Sudan. The major findings of this study is, that the distribution of cancer incidence statistic is non homogeneous between states. Khartoum state has a high incidence at 1995-2013 as shown in Fig (5.1). Dongola county in Northern state has high bladder cancer incidence rate (0.0034) as shown in Fig (5.6) and Karari county in Khartoum state has high breast (0.0468), leukemia (0.0150), brain (0.0060), prostate (0.0153) cancer incidences rates as shown in Fig (5.4), (5.8), (5.10), (5.12)). All states have a significant strong positive association between population and cancer incidence except Algedaref state, which has a significant weak positive association at 1998-2013 as shown in table (5.3). A significant weak (negative) association between Agriculture density and bladder (r_s =-0.160, p – value = .003), prostate (r_s =-0.190, p – value = .001) cancer incidence rates is found as shown in table (5.5), table (5.4), Fig (5.14), and Fig (5.15). Moreover, a significant weak (positive) association between communication towers density and breast (r_s =0.374, p – value = .000) cancer incidence rates is found as shown in table (5.8) and Fig (5.18). A significant moderate (positive) association between communication towers and leukemia (r_s =0.403 , p – value = .000), brain (r_s =0.419, p – value = .000) cancer incidence rates is found as shown in table (5.6), table (5.7), Fig (5.16) and Fig(5.17) . Absolute numbers of cancer cases in 2014 for Blue Nile are 105, Khartoum state is about 17,579 - as shown in tables (5.1) and (5.2).

These findings are consistent with results of previous studies [(Dode et al., 2011),(Wolf and Wolf, 2008),(Eger et al., 2004)], which suggests that communication towers is one of the risk factors of breast cancer. But proved the opposite previous study [(Yomralioglu et al., 2009)], which suggests there is a positive association between Agriculture density and bladder and prostate cancer. Previous studies [(Dode et al., 2011),(Wolf and Wolf, 2008),(Eger et al., 2004)] used an individual data level to determine the location of patients. In the contrary, the previous studies [(Yomralioglu et al., 2009)] and this study use data at county level to calculate incidence rates.

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

6.1 Conclusion

Findings from this study support the hypothesis that there is a correlation between cancer incidence rates and environmental factors (e.g. communication towers, and agriculture). Positive values are obtained between Breast, Leukemia and Brain cancer and communication towers. Obviously, it is negative between Prostate and Bladder cancer and agriculture. The absolute number of cases in 2014 for Blue Nile are 105 and for Khartoum state 17,579. These findings are useful to researchers and governmental agencies for risk assessment, regulations, and control of environmental contamination.

6.2 Recommendation on further research

Further research should be done to assess cancer cases in more details in counties which have high incidence rates. Furthermore, relationships between cancer types and other environmental factors (such as pollution from factories, roads, etc) are to be examined. It is recommended to use an individual data level rather than at county level.

REFERENCES

Anand, P., Kunnumakara, A., Sundaram, C., Harikumar, K., Tharakan, S., Lai, O., Sung, B. and Aggarwal, B. (2008). 'Cancer is a Preventable Disease that Requires Major Lifestyle Changes'. *Pharmaceutical Research*, 25(9), pp. 2097-2116. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles /PMC25155 69/ >[Accessed 21 June. 2016].

American Cancer Society.(2016). *Cellular Phone Towers*. [online] Available at: <http:// ww w.cancer.org/cancer/cancercauses/othercarcinogens/athome/cellular-phone - towers .> [Accessed 15Feb.2016].

ASTHANA , H and BHUSHAN, B. 2016. *Statistics For Social Sciences (With Spss Applications)*. 2 nd ed .[ebook] . New Delhi: PHI Learning Pvt. Ltd. p.6. Available at: < https://books.google.com/books?id=bh-JDAAAQBAJ&pg= PA 10&dq=Statistical+ Package +for+the +Social+Sciences+(SPSS)&hl=ar&sa= X&ved =0ahUKE wid _f6Jp6LOAhVGVhQKHYVoCkIQ6AEISTAG#v=onepage&q=Statistical%20Packag e%20for%20the%20Social%20Sciences%20(SPSS)&f=false>[Accessed 10 Feb. 2016].

Bell, S., Hoskins, R., Pickle, L. and Wartenberg, D. 2006 'Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public',*International Journal of Health Geographics*, 5(1), p. 49. doi: 10.1186/1476-072x-5-49. Available at:<http://www.ncbi.nlm.nih.gov/pmc /articles/PMC1647272/ > . [Ac cessed 21 June. 2016].

Bivand , R., Pebesma ,E., and Gómez-Rubio ,V**.** 2013 . *Applied Spatial Data Analysis with R*. 2nd ed.[ebook] . New York : Springer Science & Business Media. p.278.Available at: < https://books.google.com/books ?id=v0eIU9Ob JXg C&pg =PA362&dq=Applied+spatial+data+analysis+with+R&hl=ar&sa=X&ved= 0ahUKE wihir3bjbbOAhXE2xoKHaRLAxIQ6AEIHDAA#v=onepage&q=Applied%20spatial %20data%20analysis%20with%20R&f=false> [Accessed 22 Oct**.**2015].

Boehmer ,T., Foster, S., Henry, J. , Woghiren, E. and Yip, F. (2013). 'Residential proximity to major highways - United States, 2010'.*MMWR supplements*, 62(3), p.46-50. Available at: <https://www.cdc.gov/mmwr/preview/mmwrhtml/su6203a8.htm>. [Accessed 5 Apr. 2016].

Boyle ,M and Holben , D. 2012. *Community Nutrition in Action: An Entrepreneurial Approach*. 6th .[ebook] . New York: Cengage Learning. p.54. Available at:< https: //books.google.com/books/about/ Community_Nutrition_in _Action_An_Entrepr.html ?id=gbEKAAAAQBAJ > [Accessed 10 Feb. 2016].

Briney, A. (2014). *Overview of Public Health and GIS*. [online]. Available at:<https://www.gislounge.com/overview-public-health-gis/>[Accessed 8 Dec. 2015].

Burrough ,P., McDonnell, R., Mcdonnell, R. and Lloyd , C.2015. *Principles of Geographical Information Systems*. 3rd ed .[ebook] . Oxford: Oxford University Press. p.3. Available at: < https://books.google.com/books?id=kvoJCAAA QBA J&pg=PA321&dq =Principles+of+ Geographical+Information+Systems&hl= ar&sa =X&ved=0ahUKEwis96ORubvOAhXFVhQKHWNzB0IQ6AEIIzAB#v=onepage&q =Principles%20of%20Geographical%20Information%20Systems&f=false > [Accessed 7 Apr. 2016].

Canadian Cancer Society (2016). *Pesticides*. [online] Available at: <http://www .cancer.ca/en/prevention-and-screening/be-aware/harmful-substances-and-environmental-risks/pesticides/?region=on> [Accessed 5 Jan. 2016].

Canadian Cancer Society (2016). *Radiation*.[online]Available at: <http://www. cancer.ca/en/ prevention-and-screening/be-aware/harmful-substances-and-environmental-risks/radiation/?region =on#ixzz4ESfCXF1l>[Accessed 7Jan. 2016] .

Carter, D. (2014).'New Global Survey Shows an Increasing Cancer Burden'. *AJN, American Journal of Nursing*, 114(3), p.17. Available at: <http://journals.lww.com/ ajnonline/ Fulltext /2014/03000/New_Global_Survey_Shows_an_Increasing_Cancer . 14.aspx >. [Accessed 26 June. 2016].

Chun , B . (2010).'Effect of GIS-integrated Lessons on Spatial Thinking Abilities and Geographical Skills'. *Journal of the Korean Geographical Society*, 45(6), pp. 820-844. Available at: <http://www.kgeography.or.kr/homepage/kgeography/ ww w/old/publishing/journal/45/06/10.PDF>. [Accessed 21 June. 2016].

Dempsey, C. (2012). *Heat Maps in GIS*. [online] Available at: <http://www. gislounge .com/heat-maps-in-gis/> [Accessed 4Nov. 2015].

Dobrica , L.,Ionescu , T.,Dobrica , L. and Elena , S. (2010).' Spatial data acquisition, management and visualization in geographic information systems'. *University Politehnica of Bucharest Scientific Bulletin , Series C*, 72(3), pp.93-107. Available at:< http://www .scientific bulletin.upb.ro/rev_docs_arhiva/full8411.pdf>. [Accessed 9 Mar. 2016].

Dode, A., Leão, M., Tejo, F., Gomes, A., Dode, D., Dode, M., Moreira, C., Condessa, V., Albinatti, C. and Caiaffa, W. (2011). 'Mortality by neoplasia and cellular telephone base stations in the Belo Horizonte municipality, Minas Gerais state, Brazil' . *Science of The Total Environment*, 409(19), pp.3649-3665. Available at:< http:// ww w . emfandhealth.com /DodeLetterOnline.pdf> [Accessed 15 Jan. 2016].

Eger, H., Hagen, K., Lucas, B., Vogel, P. and Voit, H. (2004). 'The Influence of Being Physically Near to a Cell Phone Transmission Mast on the Incidence of Cancer'. *Environment - Health Corporation*, 17 (4),PP: 326-332. Available at: <http://avaate .org/IMG/pdf/20041118_naila.pdf [Accessed 28 Nov. 2015].

Elebead, F., Hamid, A., Hilmi, H. and Galal, H. (2012). 'Mapping Cancer Disease Using Geographical Information System (GIS) in Gezira State-Sudan'. *Journal of Community Health*, 37(4), pp.830-839. Available at: <http://www.ncbi.nlm.nih. gov/pubmed/22227773>.[Accessed 22 June. 2016].

Ellul ,C., Zlatanova ,S., Rumor , M.,and Laurini R . 2013 **.** *Urban and Regional Data Management: UDMS Annual 2013***.[**ebook] **.** New York**:** CRC Press. p.78. Available at: **<** https://books.google.com/books?id= BtDLBQAAQBAJ &printsec =frontcover&dq=Urban+and+Regional+Data+Management:+UDMS+Annual+2013& hl=ar&sa=X&ved=0ahUKEwigy_mErLHOAhUGXhQKHU0FBNMQ6AEIHDAA#v =onepage&q=Urban%20and%20Regional%20Data%20Management%3A%20UDMS %20Annual%202013&f=false> [Accessed 22 Oct.2015].

Eltoum, M., Dafalla, M. and Ibrahim, I., 2015, 'The Role of Ecological Factors in Causing Land Surface Desertification, the Case of Sudan'. *Journal of Agriculture and Ecology Research International* , 4(3) pp.105-116. Available at:< https://www .resear chgate.net/publication/282603016_The_Role_of_Ecological_Factors_in_Causing_La nd_Surface_Desertification_the_Case_of_Sudan.> [Accessed 21 Nov. 2015].

Federal Ministry of Health ,2007, *annual health statistical report 2007*. [online] Availableat:<http://www.sho.gov.sd/controller/knb_documents.php?id=155&lid=1&q uery=Annual%20Statistical%20Report%202007# >[Accessed 18 Sept. 2015].

Federal Ministry of Health ,2014, *annual health statistical report 2014*. [online] Available at: <http://www.sho.gov.sd/controller/dwn_hub_files.php?id=750> [Accessed 18 Sept. 2015].

Fradelos, E., Papathanasiou, I., Mitsi, D., Tsaras, K., Kleisiaris, C. and Kourkouta, L. (2014). 'Health Based Geographic Information Systems (GIS) and their Applications'. *ActaInformaticaMedica*, 22(6), p.402. Available at: <http://www.ncbi.nlm.nih .gov/pmc/ articles/PMC4315644/ .> [Accessed 15 Apr. 2016].

Freeman, L., Rusiecki, J., Hoppin, J., Lubin, J., Koutros, S., Andreotti, G., Zahm, S., Hines, C., Coble, J., Barone-Adesi, F., Sloan, J., Sandler, D., Blair, A. and Alavanja, M. (2011). 'Atrazine and Cancer Incidence Among Pesticide Applicators in the Agricultural Health Study (1994–2007)'. *Environ Health Perspect*, 119(9), pp.1253-1259. Available at:<http:// www.ncbi.nlm.nih.gov/pubmed/21622085>. [Accessed 15 Apr. 2016].

Gerber, S and Voelkl , K. 2012. *The SPSS Guide to the New Statistical Analysis of Data* . [ebook] . New York: Springer Science & Business Media. p.4. Available at: < https://books.google.com/books?id=9eALBwAAQBAJ&pg=PA24&dq=spss&hl=ar&sa=X&ved=0ahUKEwizsraOi6LOAhVKORQKHbxkCjEQ6AEISTAF#v=onepage&q=spss&f=false> [Accessed 10 Feb. 2016].

Gregory, I and Ell , P. 2007. *Historical GIS: Technologies, Methodologies, And Scholarship*. [ebook] .Cambridge: Cambridge University Press. p.48. Available at: <https://books.google.com/books?id=I_HhhJF5Un4C&printsec=frontcover&hl=ar&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false>[Accessed 11 Oct. 2015].

Hendryx, M., Fedorko, E. and Anesetti-Rothermel, A. (2010).'A geographical information system-based analysis of cancer mortality and population exposure to coal mining activities in West Virginia, United States of America'. *Geospat Health*, 4(2), p.243. Available at: http://www.ncbi.nlm.nih.gov/pubmed /20503192 .[Accessed 21 June. 2016].

International Business Machines. (2016). *IBM SPSS Statistics - United States*. [online] Available at:< https://www.ibm.com/marketplace /cloud /statistical-analysis-and-reporting/us/en-us> [Accessed 10 June. 2016].

Jung, M. (2013). *Port your R scripts to QGIS using SEXTANTE*. [online] Conservation Ecology. Available at:< https://conservationecology. wordpress.com / 2013/08/14/port-your-r-scripts-to-qgis-using-sextante/> [Accessed 15Oct. 2015].

Longley, P., Goodchild ,M., Maguire, D., and Rhind , D**.** 2015. *Geographic Information Science and Systems.* [ebook] . West Sussex : John Wiley & Sons**.** p.292 . Available at:< https://books.google.com/books ?id= C_EwBgA AQBAJ &printsec= frontcover&hl=ar&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false> [Accessed 10 Feb. 2016].

Lovelace, R. (2014). *R, an Integrated Statistical Programming Environment and GIS | R-bloggers*. [online] R-bloggers. Available at: <http://www.r-bloggers.com/r-an-integrated-statistical-programming-environment-and-gis/> [Accessed 22 Oct.2015].

Margaret, L. and Kenneth, E. (2014). *Geographic Information Systems as an Integrating Technology*. [online] Colorado.edu. Available at: <http://www.colorado .edu/geography/gcraft/notes/intro/intro.html > [Accessed 15 Oct. 2015].

Menke, K., Smith ,R., Pirelli ,L., and Van ,J. **.** 2015. *Mastering QGIS*. [ebook] . birmingham: Packt Publishing Ltd**.** p.2. Available at:< https://books.google .com /books?id=D9i6BwAAQBAJ&printsec=frontcover&hl=ar&source=gbs_ge_summary _r&cad=0#v=onepage&q&f=false> [Accessed 10 Mar. 2016].

Miller , A. 2014 . *What Causes Cancer?: What We Know and What it Means.*[ebook] . Canada: FriesenPress. p.18.Available at: < https://books.google.com/books ?id=K NaZBAAAQBAJ&pg=PA13&dq=causes+of+cancer&hl=ar&sa=X&ved=0ahUKEwj coPHj9aTOAhWLaxQKHRb0DScQ6AEIWjAI#v=onepage&q=causes%20of%20can cer&f=false> [Accessed 22 Dec.2015].

Morandat, F., Hill, B., Vitek, J. and Osvald, L. (2012). *Evaluating the Design of the R Language Objects and Functions For Data Analysis*. [ebook] Purdue University, p.1. Available at: <http://r.cs.purdue.edu/pub/ecoop12.pdf> [Accessed 21 Oct.2015].

National Cancer Institute. (2016). *Electromagnetic Fields and Cancer*. [online] Available at: <http://www.cancer.gov/about-cancer/causes-prevention/risk /radiation / electromagnetic-fields-fact-sheet#q3> [Accessed 1 May. 2016].

National Information Centre.**(**2015**).** *REPUBLIC OF SUDAN*. [online]: Available at < http://esudan. gov.sd/en/sudan/index.htm l >[Accessed 18 Feb.2016].

Neteler, M and Mitasova , H. 2008. *Open Source GIS: A GRASS GIS Approach*

.[ebook] . New York: Springer Science & Business Media. p.131-133.Available at: <https://books.google.com/books?id=cjX0BwAAQBAJ&printsec=frontcover&hl = ar &source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false> [Accessed 25 Oct . 2015].

Neteler, M and Mitasova , H. 2013. *Open Source GIS: A GRASS GIS Approach*

. [ebook] . New York:Springer Science & Business Media. p.21-24.Available at: < https://books.google.com/books?id=G20ECAAAQBAJ&printsec=frontcover&hl=ar& source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false> [Accessed 22 May 2015].

Pandey **, J** and Pathak , D.(2014). *Geographic Information System*.[ebook] . New Delhi: The Energy and Resources Institute. p.8-9.Available at: < https:// books. google.com/books?id=MS44AwAAQBAJ&pg=PR4&dq=Geographic+Information+S ystem+2014&hl=ar&sa=X&ved=0ahUKEwjztJD60cXOAhXBOxQKHSHcBYQQ6A

EIHDAA#v=onepage&q=Geographic%20Information%20System%202014&f=false> [Accessed 1 Oct.2015].

Peterson , G. 2014. *GIS Cartography: A Guide to Effective Map Design*. 2nd ed .[ebook] . New York: CRC Press. p.107. Available at:< https:// books .google .com /books?id=yM6SAwAAQBAJ&pg=PA107&dq=heat+map+in+GIS&hl=ar&sa=X&ved=0ahUKEwjovrSkzqDOAhWKchQKHfj0Bp4Q6AEIJTAA#v=onepage&q=heat%20map%20in%20GIS&f=false > [Accessed 17 Nov. 2015].

Retherford , R and Choe ,M. 2011. *Statistical Models for Causal Analysis* . [ebook] . west Sussex: John Wiley & Sons. p.1-2.Available at: < https://books.google.com /books?id=Z5NdR0ttM5cC&pg=PA1&dq=major+uses+for+regression+analysis+is+causal+analysis&hl=ar&sa=X&ved=0ahUKEwiu4Jqn5qLOAhWL8RQKHWVBD-oQ6AEIWzAI#v=onepage&q=major%20uses%20for% 20regression%20analysis % 20 is%20causal%20analysis&f=false> [Accessed 22 Oct.2015].

Ruas , A. 2011. *Advances in Cartography and GIScience*. [ebook] . New York: Springer Science & Business Media. p.299. Available at: < https:// books. google. com/books?id=vF-QrlnqC8wC&printsec=frontcover&hl=ar&source=gbs_ge_ summary _ r&cad=0#v=onepage&q&f=false> [Accessed 10 Feb. 2016].

Saeed , I., Abuidris , D., Mohamed ,K. and Mohammed ,S. 2013 ' Cancer registry in Sudan: A brief overview'. *The Internet Journal of Epidemiology*, 11(2), p. 1-5. Available at: <http://ispub.com/IJE/11/2/14574> .[Accessed 15May. 2016].

Safee, D., Cao, J., Fadul, B., Kadiouglu, O., Khalid, H., Yassin, Z., Mustafa, S., Saeed, E. and Efferth, T. (2016). 'A Five-year Survey of Cancer Prevalence in Sudan'. *Anticancer Research*, [online] 36(1), pp.279-286. Available at:< https ://www.researchgate.net/ publication/ 289125463_A_Five-year_Survey_of_ Cancer_ Prevalence_in_Sudan. >[Accessed 15May. 2016].

Sears, J. (2015). *HEALTH: Maps offer a picture of how diseases spread*. [online]. Available at:< http://www.pe.com/articles/people-761095-geraghty-health.html> [Accessed 15 Apr. 2016].

Sharma , K. 2005. *Text Book Of Correlations And Regression*.[ebook] . New Delhi: Discovery Publishing House. p.1-8.Available at: < https://books.google. com/books ?id=obb7_6k6XDQC&printsec=frontcover&dq=Correlation+analysis&hl=ar&sa=X&ved=0ahUKEwi-7oXi-aLOAhXI7xQKHRCVBlgQ6AEITzAG#v=onepage&q =Correlation%20analysis&f=false > [Accessed 1 Oct.2015].

Stewart, B. and Wild, C. (2014). *World Cancer Reports*. [ebook] who, p.3. Available at:< http://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014 [Accessed 13 Oct.2015].

Watase, H. (2012). *Association between thyroid cancer incidence and the distance from nuclear power plants in the U.S.*. master. Oregon Health & Science University. Available at: <http://digitalcommons .ohsu .edu/cgi/viewcontent.cgi ?article =17 83 &context=etd>.[Accessed 13 Oct.2015].

Wolf, R. and Wolf, D. (2008).' Increased incidence of cancer near a cell-phone transmitter station'. *International Journal of Cancer Prevention*, 1(2),pp.123-128. Available at:< http://www.powerwatch.org.uk/news/ 200502 07_israel.pdf>. [Accessed 17 Jan. 2016].

World Health Organization. (2015). *Cancer*. [online] Available at: <http://www. who. int/mediacentre/factsheets/fs297/en/> [Accessed 8Mar. 2016].

Yomralioglu, T., Colak, E. and Aydinoglu, A. (2009). 'Geo-Relationship between Cancer Cases and the Environment by GIS: A Case Study of Trabzon in Turkey'. *International Journal of Environmental Research and Public Health*, 6(12), pp.3190-3204. Available at : <http://www.ncbi. nlm.nih.gov /pmc/articles/PMC2800344/>**.** [Accessed 28 Nov.2015].

A. Correlation analysis code:-

```
// this statements explain the  input is a vector layers

    ##Layer=vector

    ##Layer2=vector

// specify the input field of the first vector layer

    ##population=Field Layer

// specify the input field of the second  vector layer

    ##Cases=Field Layer2

//the function  cor  calculate the Correlation Coefficients

    >cor(Layer[[population]],Layer2[[Cases]])
```

B. Regression analysis code:-

// this statements explain the input is a vector layers

```
##Layer=vector
##Layer2=vector
```

// specify the input field of the first vector layer

```
##population=Field Layer
```

// specify the input field of the second vector layer

```
##Cases=Field Layer2
```

// The function **lm** create a linear regression model .

```
mod<-lm(Layer2[[Cases]]~Layer[[population]])
```

//store value of independent variable to using in prediction.

```
newdata <-data.frame(Layer[[population]]<-value of independent variable )
```

// the function predict value of dependent variable

```
>predict(mod, newdata, interval="predict")
```