Geospatial Analysis in Global Health M&E

A Process Guide to Monitoring and Evaluation for Informed Decision Making

MEASURE Evaluation

USAID FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

Geospatial Analysis in Global Health M&E

by
Imelda K. Moise
Marc Cunningham
Andrew Inglis

A Process Guide to Monitoring and Evaluation for Informed Decision Making

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) under terms of Cooperative Agreement GHA-A-00-08-00003-00 and implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with Futures Group, ICF International, John Snow, Inc., Management Sciences for Health, and Tulane University. The views expressed in this publication do not necessarily reflect the views of USAID or the United States government.
MS-14-98. January 2015. Cover photo courtesy of Yohana Mapala, Futures Group.

ACKNOWLEDGEMENTS

The development of this process guide was made possible by a grant from the U.S. Agency for International Development (USAID). We greatly appreciate USAID's continued support in this arena. The team of 18 experts deserves particular thanks for providing the authors with guidance, sharing their experience, and reviewing the process guide. In addition, the 141 M&E practitioners who responded to the GIS Capacity Survey provided invalu able information used in developing a user profile for a "typical" MEASURE Evaluation-trained GIS user. At John Snow Inc., special thanks to Kola Oyediran, Tariq Azim and Moussa Ly for their many hours of initial review support. Our thanks also go to Gary Steele, who contributed his expertise in facilitation and training. Many JSI staff and partners, as well as the outside experts who participated in the meeting (listed below) offered us support and guidance as we developed this guide.

Alec Moore
John Snow, Inc.

Beth Sutherland, PhD
MEASURE Evaluation

Clara R. Burgert
ICF International

Dana Thomson
Harvard Medical School

Geoffrey M. Jacquez, PhD
SUNY Buffalo & BioMedware

John Spencer
MEASURE Evaluation

Josh Colton
ICF International

Katie Meyer
John Snow, Inc.

Kristen Wares
USAID

Livia Montana, PhD
University of North Carolina at Chapel Hill

Maria Muñiz
CIESIN

Micah Winograd
John Snow, Inc.

Michael Emch, PhD
University of North Carolina at Chapel Hill

Marcia Castro, PhD
Harvard University

Mark R. Montgomery
Stony Brook University & Population Council

Nathan Heard, PhD
U.S. Department of State

Patrick Gault
USAID

ACRONYMS

DHS	Demographic and Health Survey
GIS	geographic information system
GPS	global positioning system
HMIS	health management information system
HIV	human immunodeficiency virus
LISA	Local Indicators of Spatial Association
LGA	local government administration
LMIS	logistic management information system
MAUP	Modifiable Area Unit Problem
M&E	monitoring and evaluation
MICS	Multiple Indicator Cluster Survey
NGO	nongovernmental organization
PMTCT	prevention of mother-to-child transmission
QGIS	previously known as "Quantum GIS"
RHIS	routine health information system
SARA	Service Availability and Readiness Assessment
SPA	Service Provision Assessment
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
USG	United States Government
WHO	World Health Organization

TABLE OF CONTENTS

Preface	**6**	
Overview	**7**	
Section 1	**14**	***Informed Decision-Making Process***
Section 2	**20**	***Geospatial Analysis Techniques for M&E***
Conclusion	**36**	
References	**38**	
Glossary	**41**	
Appendix 1	**44**	***Checklist for Geospatial Analysis***
Appendix 2	**47**	***Example Scenarios for Geospatial Analysis for M&E***
Appendix 3	**65**	***Commonly Used Geolocated Routine Data Sources for M&E***
Appendix 4	**67**	***Prerequisites for Geospatial Analysis***
Appendix 5	**70**	***Spatial Considerations: Modifiable Area Unit Problem***
Appendix 6	**71**	***Recommended Reading***
Appendix 7	**75**	***Step-by-Step GIS Exercises***

PREFACE

Geospatial Analysis in Global Health: A Monitoring and Evaluation Guide for Making Informed Decisions provides monitoring and evaluation (M&E) practitioners an overview of geospatial analysis techniques applicable to their work. This guide shows how geospatial analysis can be used to support public health program decision-making along with routine planning and M&E.

The use of geographic information systems (GIS) for M&E of health programs is expanding. As a result of this expansion, a growing number of users are seeking to move beyond basic GIS techniques (such as facility mapping), into more advanced GIS applications that combine various GIS techniques, outputs, and routine M&E datasets to conduct geospatial analysis. However, knowing which advanced analysis approaches are most relevant for M&E can be challenging for M&E professionals with limited formal GIS training.

To identify the most appropriate spatial analysis techniques and help M&E professionals understand how to incorporate them into M&E, MEASURE Evaluation convened an experts meeting on Spatial Analytical Methods for M&E in December 2013 in Rosslyn, Virginia.

Participating in the meeting were 18 GIS and global health experts with experience in either spatial analysis or M&E. The meeting's objective was to identify key decision points where M&E practitioners might include spatial analysis techniques in their work. The participants identified several key challenges that M&E practitioners faced when including geospatial analysis in their work:

- Mixed skill levels—basic to advanced—among GIS practitioners in many settings.
- Limited knowledge of GIS among public health decision makers.
- Limited understanding among M&E practitioners about how to use GIS in M&E.
- Limitations and incompleteness exist in many of the commonly available routine public health and programmatic datasets.

To address these challenges, meeting participants recommended the development of a guide to give M&E and GIS practitioners an overview of how to select appropriate geospatial analysis techniques to help overcome the drawbacks of commonly used M&E data. This guide provides examples of ways to apply geospatial analysis within the context of M&E, along with resources for additional information if needed.

OVERVIEW

The value of GIS for visualizing data through mapping has gained increasing support. Health program planners can use GIS techniques to bring geographic perspectives to program planning and M&E (e.g., to track trends in implementation, identify service gaps, and justify additional funding).

Continuing support from international funders, implementing partners, and government agencies has greatly strengthened the availability of various sources of reliable data. While some challenges persist, country-level spatial data infrastructure has improved considerably, especially in the past five years. Furthermore, open-source mapping software is becoming increasingly accessible. Thus, global health and M&E have moved from a data-poor to a data-rich environment which offers opportunities for much more nuanced use of GIS capacity and analysis in the M&E context.

However, understanding of the potential uses of advanced geospatial analysis is less widespread within the M&E community. Most GIS training has focused on developing basic GIS skills.[1] However as data becomes more available and the number of M&E professionals using GIS grows, there is a need for more complex GIS skills and analysis. While basic GIS techniques can produce useful maps depicting the geographic locations of health facilities, the more advanced techniques of geospatial analysis can provide much more nuanced descriptions of health system dynamics. By integrating multiple routine M&E datasets, employing spatial analysis, and applying a systematic approach to location-based data, geospatial analysis yields more complex descriptions of a given service area (e.g., identifying underlying features such as access, travel distance, disease prevalence, and other factors that can affect the impact of health services and interventions).

M&E professionals, practitioners, and staff as well as funding organizations, governments, and implementing agencies often lack a comprehensive understanding of how to best approach geospatial analysis by utilizing a systematic approach. Additionally, GIS users and M&E practitioners, including those who routinely create and use basic GIS maps, may lack specific and relevant training on using geospatial analysis techniques,

1) The basic concepts of GIS and basic skills include learning how to use ESRI's ArcGIS or QGIS software to create a thematic map. Topics may include: what is GIS?, GIS data formats, projections, applications of GIS to specific subject areas, and basic functions of a GIS (e.g., adding data, creating new layers, joining tables, calculating attribute values, classifying features by quantity, and labeling data).

and they may be without guidance on matching geospatial techniques with a specific task. Documented case examples are needed to show the benefit provided by geospatial analytic methods for M&E.

This guide provides an overview of geospatial analysis techniques appropriate for M&E. The document is intended for M&E practitioners who want to use GIS to answer more complex questions using more advanced geospatial analysis. It will:

- detail key decision points along the M&E continuum that can be appropriate for geospatial analysis,
- outline the steps needed to develop and complete exploratory data analysis and geospatial analysis (also determining whether such an analysis is appropriate),
- describe the most commonly used geospatial techniques for specific M&E evaluation and management questions, and
- provide examples showing how geospatial analysis can be used to support public health programmatic decision-making along with routine M&E.

Geospatial Analysis

Geospatial analysis is a method of combining, examining, and mapping various types of geographically referenced data (demographic, statistical, topographic, epidemiological, and other data) to turn geographic data into information that can be used for program planning and decision-making.

While basic GIS techniques can produce simple and useful maps (identifying, for example, locations of health services or HIV prevalence), a more advanced geospatial analysis enables examination of relationships, trends, patterns, and influences that may not be visible in maps generated from a single data source. By layering data from multiple sources, filtering, and applying statistical methods, geospatial analysis can provide deeper insights about the conditions and factors that influence health services and interventions. With these results, M&E practitioners and program managers can tailor interventions to specific local needs, plan future programs and activities, identify service gaps, and justify programmatic investments.

Effective geospatial analysis for M&E depends on the availability of high-quality geographically referenced M&E data. The growing availability of spatially referenced country- and regional-level data on programs, the populations they serve, and the

context in which they operate make it increasingly feasible to use geographic tools to target interventions more accurately. Simultaneously, the availability of specialized software, including open-source programs,[2] reduces barriers to conducting geospatial analysis. Not only does this type of software create significant advantages for basic mapping; it also allows users to characterize spatial relationships by layering M&E datasets (e.g., program indicators, survey information, target population characteristics and service environment).

Effective geospatial analysis requires an integrated use of existing M&E data, project-specific knowledge, and analytical skills. A geospatial analysis may start with a generated map in which the GIS user asks a series of spatial-related questions such as:

- "Does the map suggest a pattern that means something?"
- "What might be causing this pattern?"
- "Can trends in the pattern be projected or estimated?"

These types of questions can lead to two key types of geospatial analysis:

1. spatial pattern description
2. spatial pattern relationships

This guide is meant to introduce users to these terms and to the types of questions to ask when considering a geospatial analysis for M&E.

Purpose of the Guide

This guide provides information on key decision-making points for selecting appropriate geospatial analysis techniques, resources, and outputs that will support informed programmatic decision-making. It is designed to address the increasing complexity of the program environment and help users make the best use of today's richer data environment. It is also meant to help address the increasingly complex programmatic environment, which requires M&E practitioners to develop a more sophisticated understanding of the potential of geospatial analysis for program monitoring and planning.

2) Examples of open-source programs include QGIS, SaTScan and GeoDa, among others. A description of several open-source programs can be found in Section 2.

The guidance in this document is built around a framework (illustrated in the *Framework for the Guide* which will be described in the beginning of Section 1) that incorporates M&E questions, decision points, and GIS questions and identifies techniques for addressing these questions.

Intended Audience

The concepts presented in this guide have wide applicability. Some examples of people who might benefit from this guide include M&E professionals, GIS users, M&E managers within national governments, U.S. government (USG) staff, and implementing partners.

Because this document focuses on M&E data and its use in geospatial analysis, the "ideal" user is expected to already know how to perform basic geospatial analyses to inform operational decisions and track ongoing interventions. Basic geospatial analyses include queries and reasoning, descriptive summaries, optimization, measurements, and transformation. (A description of these basic geospatial analysis techniques is presented in the Glossary.) In addition, while the guide provides a description of each technique and presents examples of how the techniques can be used for M&E, GIS training beyond basic mapping is required. Once these skills are acquired, they must then be reinforced through practice and continuous use of these methods.

The guide can be helpful for M&E practitioners who want to use GIS to answer more complex questions using more advanced geospatial analysis.

An Acquired Skill

This guide provides an overview of how to apply geospatial analysis in M&E. While detailed guidance on specific analytic techniques aren't presented in this document, MEASURE Evaluation has developed a series of documents on GIS and use of spatial data (see Box 1).

BOX 1 **GIS AND SPATIAL DATA GUIDANCE**

An Overview of Spatial Data Protocols for HIV/AID S Activities: Why and How to Include the "Where" in Your Data.
http://www.cpc.unc.edu/measure/publications/ms-11-41a

The Joint MEASURE Evaluation and The DHS Program QGIS Curriculum.
http://spatialdata.dhsprogram.com/resources

Step-by-Step GIS Exercises (Appendix 7)

Using Geospatial Analysis to Inform Decision-Making in Targeting Health-Facility-Based Programs.
http://www.cpc.unc.edu/measure/publications/ms-14-88

Training and continuous use of GIS will build M&E practitioners' GIS skills. It is important for M&E practitioners to understand clearly the strengths and limitations of their skills. For complex analyses that lie beyond the skills a practitioner has developed, it is acceptable to seek outside help. In some instances, experts from local universities, GIS centers, and other local agencies and organizations can provide help with projects or activities involving GIS analysis. When GIS users rely on the larger GIS community, it minimizes the risk of faulty assumptions that lead to inaccurate and unreliable use of M&E data, which can lead in turn to inaccurate analytical conclusions and, in worst-case scenarios, to faulty projections and misdirected programmatic planning.

We recommend that M&E specialists who are conducting geospatial analysis use several guidance documents concurrently with this guide. These resources can be accessed on the Spatial Data Repository website at http://spatialdata.dhsprogram.com/resources, under the header labeled "resources."

Structure of the Guide

Geospatial analysis is an iterative process entailing specific steps, each of which leads to the next step. For the purposes of this guide, we have developed a framework (depicted on page 11) that brings together:

- essential steps for M&E and GIS,
- spatial questions applicable to M&E, and
- geospatial techniques which can be used to analyze M&E data..

Section 1 of the guide provides more detail on each step of the spatially informed decision-making process illustrated in the framework on page 11.

This process brings together analysis of program-specific M&E decision points, available M&E data, and appropriate GIS techniques. Note that one essential step of preparing for a geospatial analysis is ensuring that this type of analysis is indeed appropriate for the type of questions being asked, the type of data being used, and the representation of the data. This guide emphasizes how to select appropriate geospatial analysis techniques once M&E spatial questions have determined that GIS is appropriate as an alternative tool for analysis and visualization.

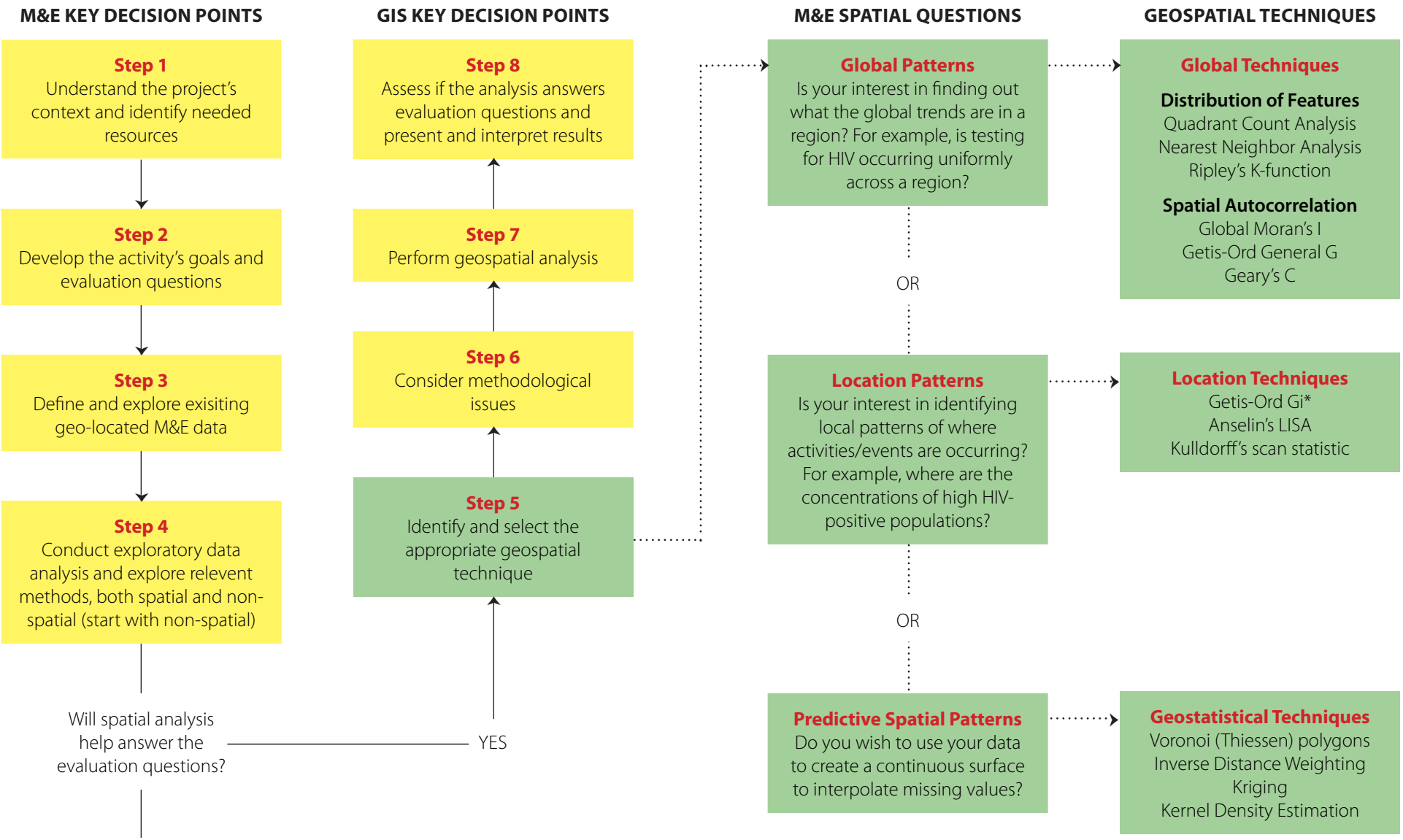
Section 2 provides a brief summary of the range of available geospatial analysis techniques that can be used with M&E datasets. It lists each technique with a description and example.

The **glossary** defines common terms used in GIS and geospatial analysis.

The **appendices** provide further information and reference pre-requests for geospatial analysis, existing geo-located M&E data and exercises for geospatial analysis techniques. Specifically, each appendix provides the following tools:

- **Appendix 1**—Checklist of the GIS decision-making analysis process, including key decision points for selecting an appropriate technique for analysis.
- **Appendix 2**—Scenarios illustrating practical application of the principles described in the guide (most useful for those looking to apply geospatial analysis for assessing coverage, for resource allocation, or for linking datasets derived from different sources).
- **Appendix 3**—Commonly used geo-located data sources for M&E.
- **Appendix 4**—Essential steps for conducting geospatial analysis.
- **Appendix 5**—Spatial considerations (Modifiable Area Unit Problem).
- **Appendix 6**—Recommended reading.
- **Appendix 7**—Exercises for geospatial analysis techniques (e.g., exploring GeoDa, mapping clusters, delineating catchment area and hot spot analysis).

Framework for the Guide



NOTE

It might be necessary to collect more data and revisit one or more of the previous steps if the question has not been adequately answered upon completion of Step 8.

SECTION 1

Informed Decision-Making Process

When M&E professionals use data to support decision making there is a well-established process whereby analysis methods are employed to identify patterns in data. When spatial data is used, the process is similar; however, there are some important differences. This section presents a decision-making process framework developed for conducting an appropriate geospatial analysis for M&E. The process outlined is summarized in the Framework for the Guide. In essence, it describes a sequence of linked processes which entail:

- determining the goal of the analysis and whether a spatial approach will be appropriate, and
- developing and analyzing the geospatial questions.

This spatially informed decision-making guide serves as a starting point for M&E practitioners with basic GIS skills interested in progressing to more advanced geospatial analysis. It is also ideal for those M&E practitioners with advanced GIS skills who are interested in using GIS to support M&E efforts.The eight steps for undertaking a geospatial analysis in M&E are:

1. Understand the project's context and identify needed resources.
2. Develop the activity's goals and evaluation questions.
3. Define and explore the existing geo-located M&E data.
4. Conduct exploratory data analysis and explore relevant methods (both spatial and non-spatial).
5. Identify and select the appropriate geospatial technique.
6. Consider methodological issues for the technique selected.
7. Perform the geospatial analysis.
8. Assess whether the analysis answers evaluation questions; present or refine results.

M&E Steps When Considering Spatial Analysis

The following steps should be carried out during the development of any geospatial analysis for M&E. Note that it may be necessary to perform more than one analysis to obtain results that answer the M&E questions sufficiently to be useful for the project or program.

Step 1: Understand the Project's Context and Identify Needed Resources

Listed below are a number of considerations M&E practitioners must take into account as part of their M&E work, and before conducting any geospatial analysis (see Rugg's M&E Staircase Steps[3]). Each step is the foundation for the next step in the analytic process:

- Are the evaluation questions and objectives clearly defined, as well as linked to important program or policy decisions? Is the outcome measurable?
- Does the M&E team have the necessary analytical skills to perform the analysis?
- Do data exist that are appropriate for geospatial analysis?
- Can data from multiple sources be linked based on a unique identification number?[4]
- Is denominator data available (e.g., census, number of HIV patients) and will it be used?
- Will non-spatial and spatial data issues (e.g., scale, data quality, and availability) affect geospatial analysis?
- Are data appropriately aggregated to answer the evaluation question, and to avoid statistical bias and scale and measurement issues?
- Are geographical data used correctly, with legal and ethical concerns rigorously addressed in the handling of data?
- Are appropriate spatial methods available for existing spatial M&E data?
- Are relevant spatial techniques available (e.g., open source software such as QGIS)?
- Are the right program or policy experts involved in the program evaluation to ensure that the results of the analysis are interpreted honestly and accurately?

NOTE: Thinking through the above questions ahead of time will simplify the geospatial analysis. Ensure that any response to any of the above questions is unresolved before proceeding. If the questions remain unresolved, geospatial analysis may not be appropriate to the current situation.

Step 2: Develop the Activity's Goals and Evaluation Questions

To be appropriate for geospatial analysis, the questions asked in this step (as in Step 1) must try to answer geographic questions about the "where" and "why there" of an M&E issue, question or problem. Key geographic questions ask: Where is it located?

3) Rugg's M&E Staircase Steps provide guidance on hypothesis generation, programmatic activities, final outcomes, and overall effects. A logic framework endorsed by UNAIDS MERG: accessible at http://www.unaids.org/en/media/unaids/contentassets/documents/document/2010/7_1-Basic-Terminology-and-Frameworks-MEF.pdf.

4) See MEASURE Evaluation's *Using Geospatial Analysis to Inform Decision Making in Targeting Health Facility-based Programs: A Guidance Document*. Accessible at: http://www.cpc.unc.edu/measure/publications/ms-14-88.

Why is it there? What is the significance of the location? For example:

- Where were most of the positive pulmonary tuberculosis (PTB) cases last month?
- What is the closest CD4 lab along a street or delivery route?
- What percentage of HIV-positive pregnant women resides within a square kilometer of a sentinel antenatal clinic?
- What factors or determinants exist in specific regions that may have contributed to higher TB cases in this region vis-à-vis the other region?

Step 3: Define and Explore the Existing Geo-located M&E Data

An effective geospatial analysis must have a valid set of M&E data. (See Appendix 3 for a detailed description of spatial M&E data sources.) Generally, three types of data, or data models, represent the way in which computers record geographic features:

Events or point pattern data are discrete features. For example:

- Anything that can be described by its geographical location or using its exact coordinates such as a discrete latitude and longitude location. This may include health facilities, locations of health facilities providing a specific product or service, or child immunizations provided over a given time period at a series of locations.
- Lines comprise any feature with a length (e.g., rivers, roads, or rural paths).
- Polygons comprise any feature having boundaries (e.g., administrative units such as districts, provinces, wards, or local government areas [LGAs]).

Counts and aggregated rates are typically aggregated to administrative units (such as births per district or HIV prevalence in a ward) and collected across groups.

Continuous surfaces comprise continuous variables and/or densities of point data.

- Examples of continuous variables include: digital elevation models, triangulated irregular network (TIN) models, etc. (Goodchild, Haining, and Wise, 1992).
- Examples of densities of point pattern data may include densities of HIV cases or health facilities over a fixed area or region.

Step 4: Conduct Exploratory Data Analysis and Explore Relevant Methods

Geospatial analysis is often preceded by a phase of data cleaning and exploratory data analysis associated with the visual presentation of the data in the form of graphs and maps. Both spatial and non-spatial aspects of the data and data characteristics should be considered (see Box 2). It is also important to consider other decision points for which M&E data can be defined and explored:

Data Transformation

Some data may need to be transformed before a valid analysis can be performed in the target software, such as transforming a spreadsheet of global positioning system (GPS) coordinates or a list of address-level data into a point boundary format (shapefile). Notably, this process may include a change in the coordinate system, data attributes, or geographic feature types.

Data Linking and Exploratory Analysis

Data from different data sets need to be linked. Triangulating these different datasets can provide a better picture of program activities and conditions in a given area.

BOX 2 DATA CONSIDERATIONS

Non-spatial
- Data source and characteristics
- Data representativeness, e.g., sample or population
- Survey measurement
- Frequency of collected data
- Outliers

Spatial
- Ecological fallacy
- Scale
- Modifiable Area Unit Problem (MAUP)
- Outlier Spatial Units
- Spatial autocorrelation

Explore Methods (Spatial and Non-spatial)

Always analyze datasets to summarize their main characteristics (basic descriptive statistics) using various visual methods such as charts and graphs. When possible, formulate hypotheses that could lead to the selection of appropriate analysis techniques.

Additionally, after going through the M&E key decision points, it may become clear that non-spatial questions are more important. This determination implies the need for a different approach to analysis that may involve non-geospatial methods. If the data available is non-spatial, but has some kind of geographic reference, a combination of spatial and non-spatial techniques may be necessary to display and visualize data. However, these techniques are beyond the scope of this document. The next sub-section focuses on applying the GIS steps of geospatial analysis, assuming that the steps above led to a finding that this is the appropriate method for answering evaluation questions.

GIS Steps

Step 5: Identify and Select the Appropriate Geospatial Technique

As noted in Step 3, geospatial analysis in M&E deal with point data and aggregated data from routine information systems and from facility and population-based survey data. These data sources may be analyzed for location patterns, global patterns, or predictive spatial patterns. It is important to choose the appropriate procedures and techniques. For example:

- Analyze point pattern data using techniques developed for this type of data (e.g., Anselin's Local Indicators of Association [LISA], SaTScan).
- Do not use point pattern analysis to analyze continuous data or perform frequency analysis or crosstabs to analyze continuous data. Always implement non-spatial analysis techniques before considering any spatial analysis technique.
- To select appropriate geospatial techniques, consider evaluation and management questions and how the results of the analysis are to be used.
- The three data types noted in Step 3 can often be mapped onto each other. For example, facility-level data (point data) and roads (line data) can be aggregated to administrative units (polygons), while point data, such as prevalence of a disease, may be converted into a continuous surface.
- Keep in mind that data aggregation can be virtually irreversible, because it can result in the loss of the original information or add noise (meaningless data) (Gatrell, 1991). Data should be stored at the smallest geographic unit possible, with analysis performed at the same level.

Step 6: Consider Methodological Issues for the Technique Selected

Methodological issues or errors need to be considered when designing a GIS project. Error and uncertainty in a GIS typically arise in any of four stages:

- When conceptualizing the spatial problem (e.g., scale of analysis, vagueness, or ambiguity).
- When measuring and representing the spatial problem (e.g., the relative error might be introduced during data integration).
- When the quality and accuracy of data is unclear.
- During the analysis phase, methodological issues such as scale and aggregation may lead to the modifiable area unit problem (MAUP) that may distort results (see Appendix 5).

Step 7: Perform the Geospatial Analysis

The next step is to perform the analysis. For example, if trying to see a quick view of the areas with the lowest or highest number of children who received a measles vaccine, simply mapping the individual vaccination cases and looking at the maps may yield the intended results. However, if the information will be used to assess the project's impact on measles vaccination, the analysis may require a more precise measure of the location, along with other information, such as the vaccine supply stock-out for a given health facility over a given time period.

Step 8: Assess Whether the Analysis Answers Evaluation Questions; Present or Refine Results

The interpretation of the output or map generated by geospatial analysis is a critical step for a valid geospatial analysis. Interpretation requires an understanding of the GIS task, program, and the geospatial techniques used, as well as an understanding of how to interpret the output in the context of the GIS task. It is essential that the M&E team is:

- able to interpret the output accurately, and
- able to summarize it sufficiently to meet program or project expectations.

Often during geospatial analysis—after generating the output or map—the analysis chosen may not answer the research questions. If this occurs, refine the results or choose an alternative method. This situation is not surprising; the GIS user often has to conduct several analyses before identifying an appropriate analysis that can answer evaluation questions. It may also be that additional data are needed. Thus, M&E and GIS practitioners need to take into account the time that will be expended to conduct a geospatial analysis. Geospatial analysis is never a one-step process. Implementing a proper analysis may entail many back-and-forth analyses and key decision points. More details for each step in the guide are provided in the Checklist of the GIS Decision-making Analysis Process (Appendix 1).

SECTION 2

Geospatial Analysis Techniques for M&E

This section provides further insight into the range of geospatial analysis techniques that can be used for M&E. Often the objectives of geospatial analysis for M&E are to detect, characterize, and then make inferences about spatial patterns. This might include identifying areas of locally increased risk or unmet need (such as need for family planning); pinpointing areas where interventions are working well; and determining factors that lead to or affect spatial interactions.

Geospatial analyses are primarily used to:

- describe spatial patterns (exploration),
- test hypothesis about patterns (inference), and
- predict patterns (based on mapping, or interpolation)

The outputs of geospatial analysis for M&E can vary from simple choropleth maps[5] for highlighting patterns and locations of interest and giving a visual representation of counts within administrative units, to maps displaying estimates (for example projections of TB incidence) derived from complex and multi-sourced models developed to describe the structure of interventions.

In addition, there are differences among analytical techniques, the most important of which pertain to:

- computational effort,
- resolution/visual interpretation, and
- quantitative interpretations/inferences that can be made from the analyses.

To date, basic mapping strengthened with exploratory analyses has been sufficient for M&E practitioners. However, geospatial analysis is needed to predict and estimate factors (such as accessibility) that could be contributing to the findings depicted on the maps.

To deepen understanding of geographically referenced program data, and to make optimal use of available geospatial technology, programs need to build GIS capacity

5) Choropleth maps are maps that use color coding or patterning to indicate the measurement of the statistical characteristic or variable being displayed, such as population density or regional contraceptive prevalence rate per district.

further along from simple visualization to encompass the understanding and use of advanced geospatial skills among GIS users and M&E practitioners. These skills are attainable through training, practice, and use to the level that one is comfortable implementing geospatial analysis without assistance.

How to Perform Geospatial Analysis in a GIS

The majority of the techniques outlined below can be performed through ArcGIS's Spatial Analyst extension,[6] a commercial program developed by the Environmental Systems Research Institute (ESRI). The Spatial Analyst extension in ArcGIS provides a rich suite of techniques and capabilities for performing comprehensive, raster-based spatial analysis.

In ArcGIS, these analytical techniques are grouped into four categories: 1) measuring geographic distributions, 2) analyzing patterns, 3) mapping clusters, and 4) modeling spatial relationships. Sometimes, using these techniques may require Python scripting (a source code language). Note that often in geospatial analysis, several techniques are used in combination to take advantage of the specific strengths of each approach.

Other popular open-source GIS software can also be used to create maps or charts, or to perform other types of geospatial analysis. These include the following:

Kulldorff SaTScan Statistics

This free software developed by Martin Kulldorff can be used to analyze spatial, temporal and space-time data using the spatial, temporal, or space-time scan statistics. This statistics tool requires the user to have a number of types of files: SaTScan ASCII file format, dBase, comma-delimited or space-delimited files. The essential files include: 1) a geographic coordinate file, 2) a case file (needed for all probability models except the continuous Poisson model), and 3) either a population file (the Poisson model) or a control file (the Bernoulli model). The results must then be visualized using ArcGIS. For details on how to use SaTScan, see Kulldorff M. 2014, "SaTScanTM User Guide 9.3". Document can be downloaded at http://www.satscan.org/.

6) This software is available for download for ArcGIS for Windows users. For instructions on how to install GIS software for Mac users, see: http://serc.carleton.edu/eyesinthesky2/installgis/mac_users.html.

QGIS

In QGIS, advanced geospatial analysis can be performed by using Vector Analysis Tools or with a third-party plugin such as the "ftools-qgis." The ftools-qgis plugin can be used to generate proximity polygons, Voronoi polygons, or Thiessen regions—all useful for delineating catchment area. For details on how to use these and other Vector Analysis Tools in QGIS, download The Free Quantum QGIS Training Manual, which is available at http://manual.linfiniti.com/en/vector_analysis/basic_analysis.html or see the online QGIS Documentation, which is accessible at http://www.qgis.org/en/docs/index.html#documentation-for-qgis-2-0.

Luc Anselin's GeoDa

GeoDa is designed to implement techniques enabling exploratory spatial data analysis. It provides a user-friendly and graphical interface for performing descriptive spatial data analysis, such as global and local autocorrelation statistics and indicators of spatial outliers, and some more advanced regression analysis techniques. Spatial data can be displayed on maps, in histograms, and in scatterplots. For details on how to use GeoDa, tutorials, and examples, see *Exploring Spatial Data with GeoDaTM: A Workbook* at https://geodacenter.asu.edu/system/files/geodaworkbook.pdf.

Types of Geospatial Analysis

This sub-section provides an overview of approaches for performing three types of geospatial analysis commonly performed in the M&E context:

Global Patterns

Global pattern analysis can be used with M&E data to measure the uniformity (or lack of uniformity) of data across a region. Global measures include determination of the distribution of features (dispersed, random, or clustered) and also measurements of spatial autocorrelation (similarity between neighboring points).

Describing the Distribution of Features

Global geospatial data analysis applied to point data typically entails analysis of point distributions in space and of the relationship between point distributions and other spatial features. The objective of the analysis would be to determine if the point pattern is dispersed, random, or clustered. Common techniques for analysis for distribution of features include:

- Quadrant count analysis
- Nearest neighbor analysis
- Ripley's K-function

Measuring Spatial Autocorrelation

Global pattern analysis is important to help answer questions such as, "Is there statistically significant similarity or dissimilarity of neighbors (spatial autocorrelation) for the whole study region?" Alternatively, this technique may be used to help summarize the overall degree to which similar observations tend to occur near each other. Techniques for understanding spatial patterns and trends beyond a local context are called "global calculations" or "global measures of spatial autocorrelation."

Global calculations or statistics can be very important for M&E to monitor whether the target populations receiving an intervention exhibit similar outcomes or if disease prevalence exhibits similar patterns. The most common calculations include:

- Global Moran's I
- Getis-Ord general G
- Geary's C[7]

Location Patterns

An equivalent local measure can be calculated for most global measures. This discrete measurement can be used along with a global measure. Local measures are useful for identifying local patterns or clusters associated with outcomes, because health outcomes can have both local and regional impact. A cluster (also known as a local calculation) is a region that has unusually high counts or rates.

Note that local statistics can not only indicate local spatial clusters, but also diagnose outliers in global spatial patterns. This type of data representation can be applied to both point and aggregate data. Local measure can be used to answer questions such as "Where does spatial clustering occur?" or "Where is the location of spatial outliers?" The most common techniques for mapping clusters, or local calculations, include:

- Getis-Ord Gi* (Hot Spot Analysis)[8]
- Anselin's local indicators of spatial association (LISA)[9]
- Kulldorff's Scan Statistic

7) Geary's C statistic is named for its creator, Roy Geary, and the Getis-Ord G statistic is named for its creators, Arthur Getis and Keith Ord.
8) The Getis-Ord Gi* statistic is named for its creators, Arthur Getis and Keith Ord, and its pronounced "G-I-Star."
9) Anselin's LISA is also named for its creator, Luc Anselin.

Predictive Spatial Patterns

Calculations for predicting spatial patterns are often based on sample data, such as data collected from selected antenatal HIV sentinel clinics or DHS clusters. Such data, through capabilities of a GIS, can be converted into geostatistical data using spatial interpolation techniques where the variable value at unknown locations is predicted based on the values of sample points. The output will be a continuous smooth surface showing the variation of the feature attribute in space or the density of program indicators or outputs modeled as a continuous field. For M&E, these techniques can be quite powerful: sample data can be used to determine possible hot spots for intervention activity, identify unmet need, and prioritize areas for interventions.

Many methods are associated with geostatistics, each having unique qualities. They range from techniques that draw boundaries around the sample points, to techniques which use statistical models to estimate the values between the sample points. For example, the methods for creating Voronoi polygons, or Thiessen polygons, divides a region up into polygons, or catchment areas, around each sample point (assuming that values for unsampled points within a particular polygon are likely to be similar to the value at the nearest sample point—see Appendix 7 for a tutorial). If the goal is to estimate a continuous surface, it is preferable to use geostatistical spatial interpolation techniques which include various kinds of estimates such as kernel density estimation (KDE).

The example that follows the descriptions of these methods illustrates an application of KDE.

- Voronoi (Thiessen) polygons
- Inverse Distance Weighting (IDW)
- Kriging
- Kernel Density Estimation (KDE)

Explanation of Techniques

Quadrant Count Analysis

The quadrant count analysis is useful for summarizing the frequency of points per grid square (e.g., facilities per unit area). A mean/variance ratio is calculated to indicate whether the pattern of the case of interest (e.g., intervention) is regular, random, or clustered.

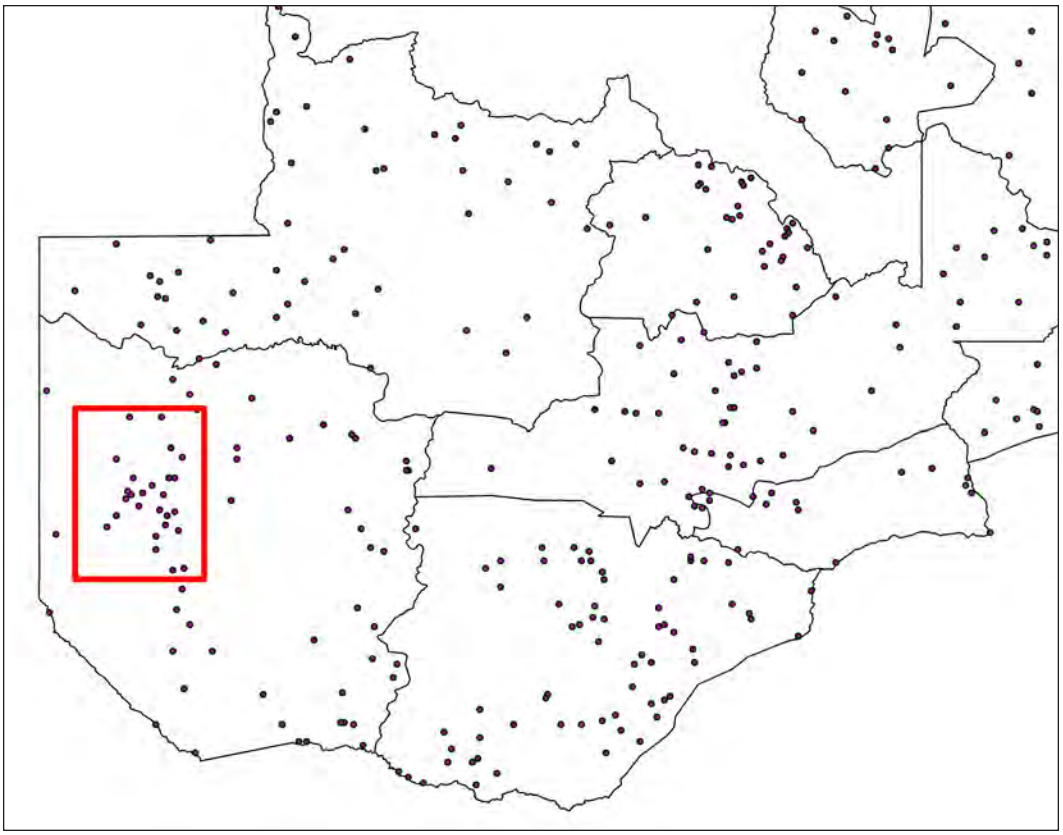


FIGURE 1
Health facilities in southern Zambia, 2006.

Figure 1 gives a visual impression of a clustered pattern or of health facilities within a specific quadrant of Zambia's Southern Province in 2006, examined using data downloaded from the Southern African Human-development Information Management Network (SAHIMS).

The M&E practitioner looking at this map might ask a simpler but more fundamental question:

- "Is there evidence in the pattern which indicates that the health facilities are very dissimilar?" In other words, "Is the pattern just a random distribution of points?" or
- "Are health facility locations in some way dependent on one another?"

Generally, health facilities need to be dispersed throughout the region to ensure provision of health services to a widely dispersed population.

NOTE: Three considerations are essential for conducting both successful quadrant count analysis and nearest neighbor analysis: 1) proper selection of the geographic scale, 2) extent, and 3) projection of the study area.

Nearest Neighbor Analysis

The nearest neighbor method is based on comparing the distribution of the distances from each data point to its nearest neighbor in a given dataset with a randomly distributed dataset. This analysis can be used to compare distribution of groups such as TB-infected and uninfected individuals, analyze the location distribution of health care facilities, or calculate the distances and areas of communities relative to health services. Consider the difficulty of trying to objectively measure or to describe scattered patterns of the location of health facilities or TB cases in a region. Certainly, a disease such as TB may occur in spatially clustered locations, suggesting a localized environmental factor may be responsible.

Sometimes, the observed patterns may be obvious, such as when facilities are extremely concentrated or located far apart. One way that a pattern can be measured objectively is with the nearest neighbor index. This statistic calculates an index ratio of an observed distance to a single nearest feature divided by distance from randomly distributed data.

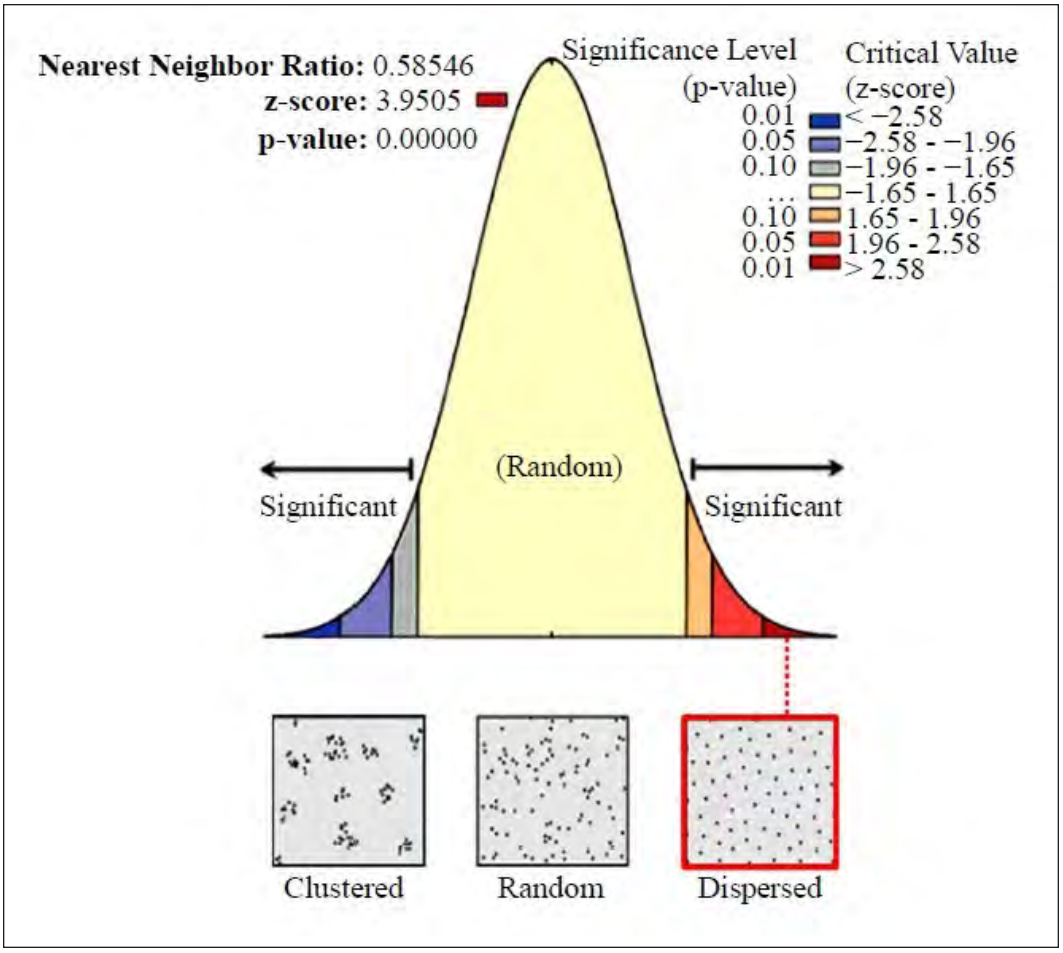


FIGURE 2 **Nearest neighbor findings on public clinics in Malaysia.**

Figure 2 shows the high density of public health clinics in relation to the population size in Malaysia, using the nearest neighbor analysis and data on health clinics conducted by Hazrin and colleagues (Hazrin et al., 2013). The distance between each feature centroid and its nearest neighbor's centroid location was measured, with all nearest neighbor distances averaged. The nearest neighbor ratio indicated in the chart (0.5854) shows that the expected mean distance to public health clinics is indeed dispersed. This distribution may reflect the efforts of the Malaysian Ministry of Health to ensure equitable distribution of health facilities in the country.

Generally, values of the index ratio greater than 1 will exhibit a trend toward dispersion; an index of 1 will reflect features that occur very randomly, while values less than 1 exhibit a trend toward clustering. In this case, the z-score is higher at 3.9505, suggesting dispersal.

Ripley's K-function

Ripley's K-function is a distinctive statistic in that it detects clustering and relationships between points at a series of distances or spatial scales. An index is calculated by measuring the distance from each feature to all the other features in the data set, generating a K value. The peak value that exceeds the confidence envelope by the greatest margin will be the distance of most significant clustering. The output from the K-function is a line graph.

For example, Figure 3 shows how the Ripley's K-function is applied to survey data on health outcomes and GPS coordinates of survey respondents to evaluate clustering of family planning methods utilized by females located in the rural areas of four districts (Chibuto, Chokwe, Guija and Mandlakaze) of Gaza province in southern Mozambique.

In this study, Ripley's K-function is used to detect global clustering as compared to a random distribution. The graph shows the K-function for patterns of family planning use. The difference between the calculated K-function values is plotted against distance (d) to compare patterns of clustering (in women reporting family planning use). The solid line represents the underlying pattern of the respondents' distribution. Dashed lines reflect the confidence envelope around the difference of expected distributions (zero line).

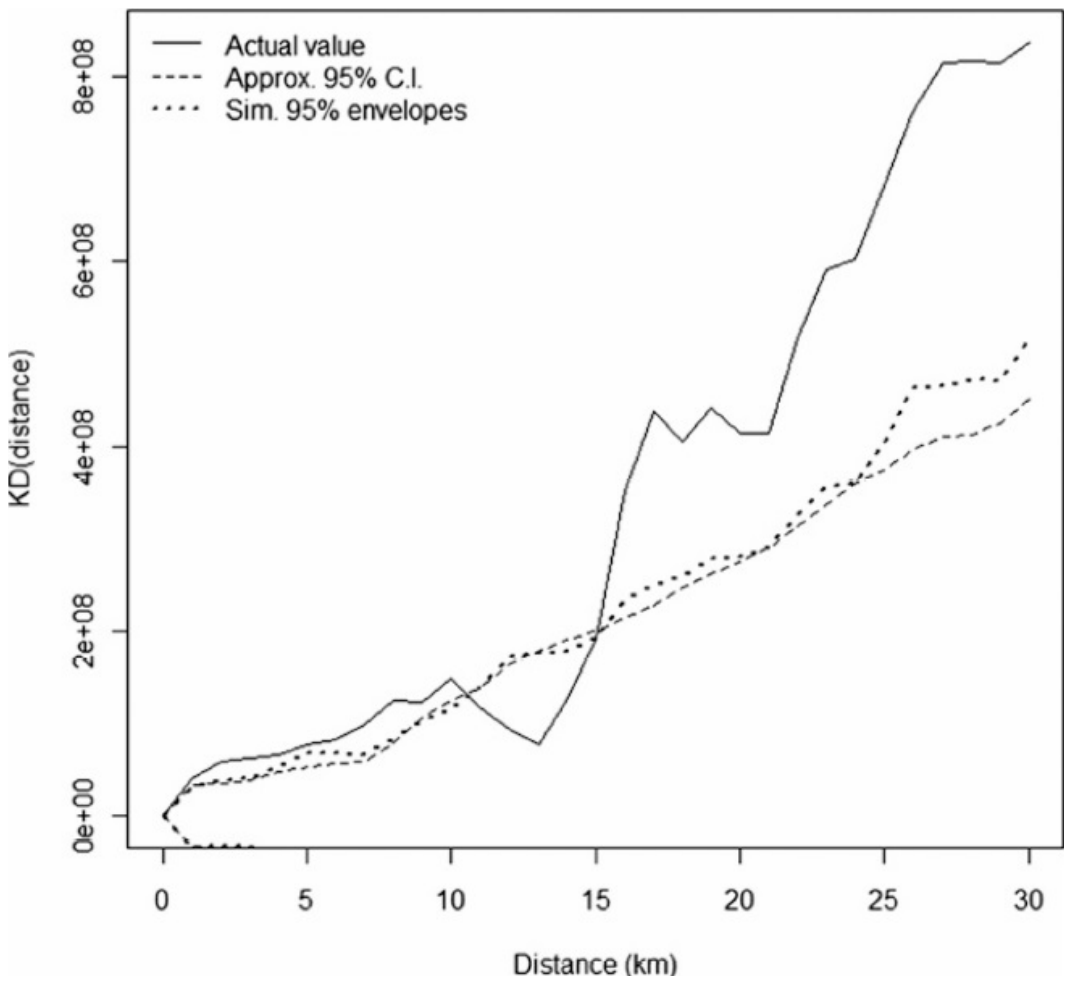


FIGURE 3
Cluster analysis of family planning use in four districts of Gaza Province southern Mozambique, 2006.

On this chart, significant clustering occurs at smaller (<10 km) or larger distances (>15 km), where estimated values are larger compared to those from a random distribution, as evidenced by the curve of actual values, which falls within the confidence bands. No significant clustering is found in intermediate distances (10 km, 15 km). This chart indicates spatial variability in the utilization of family planning services. Program managers might use such a chart, for example, as evidence to show that utilization of family planning services in Gaza province is widely accepted or is well established—since utilization of family planning services is observed both at shorter (less than 10km) and longer distances (larger than 15 km).

Global Moran's I

Investigation of global clustering patterns across regions can be very important for monitoring changes in disease prevalence over time. Global Moran's I examines the extent of clustering that exists in a dataset observed across the entire region, which is the similarity or dissimilarity of features, based on both their locations

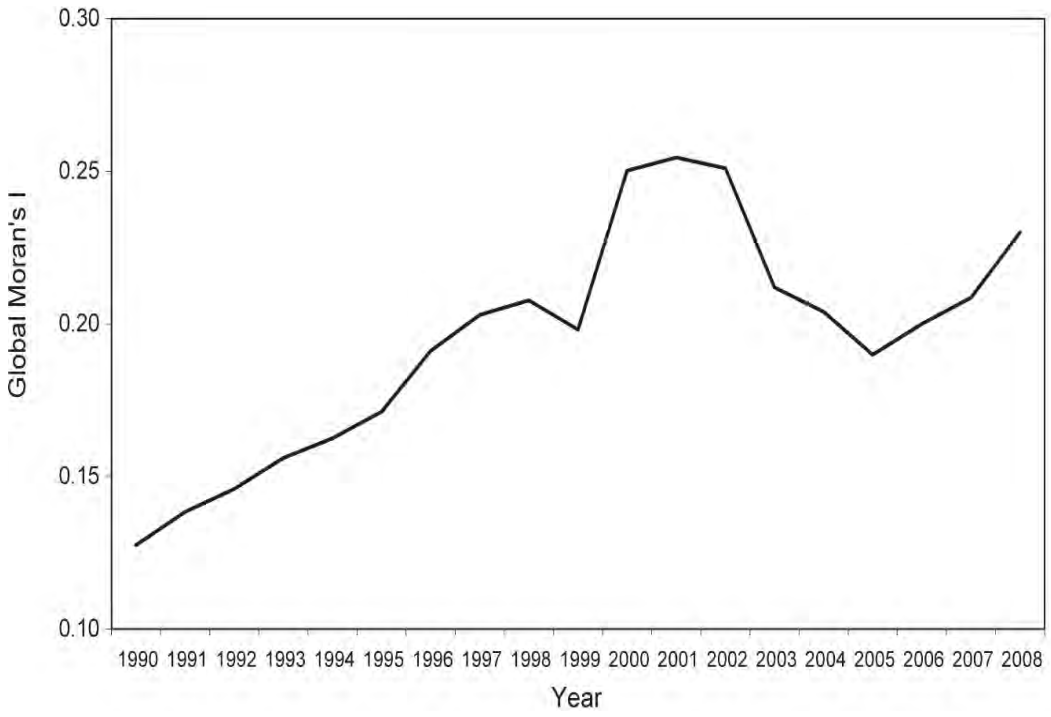


FIGURE 4
Spatial autocorrelation of health care levels (number of hospital beds per 10,000 persons) among provinces, 1990–2008.

and values. If neighboring areas are more similar, we would obtain a positive spatial autocorrelation, which might violate the assumption that the observation values in each sample are independent of each other. A Moran's I value near +1.0 indicates clustering; 0 indicates randomness; and a value near –1.0 indicates dispersion.

Figure 4 shows a chart obtained from analyzing spatial autocorrelation of health care levels in Chinese provinces (number of hospital beds per 10,000 persons from 1990 through 2008). In this chart, Moran's I shows a generally upward trend of spatial autocorrelation, interrupted by a sharp decline from 2001 to 2005, demonstrating that spatial association exerts fairly heavy influence on the observed regional inequality in the health care system in China. An increasing global (approaching +1) Moran's I indicates that the spatial distribution of health care has become more uneven or that spatial clustering of high and low values is becoming more prominent in the country.

Geary's C Statistic

Geary's C statistic, like Moran's I, is also useful for detecting spatial clustering occurring anywhere in a study area (Anselin, 1992). These statistics do not identify where the cluster(s) occur, nor do they identify differences in spatial patterns within the area. If the interest is in identifying potential local clustering within smaller areas inside a study area, local clustering statistics such as Local Indicators of Spatial

Autocorrelation (LISA) or Getis-Ord Gi* (described in next section) need to be used. Often, global techniques are used first to identify potential clustering; then, local methods are used to pinpoint the clusters in the sample area.

NOTE: While Moran's I and Geary's C result in similar conclusions, it's important to know that Moran's I is a more global measurement and is sensitive to extreme values of x, whereas Geary's C is more sensitive to differences in small neighborhoods. It is beyond the scope of this guide to describe more than a few of the most commonly used methods, and even then, these methods are described only briefly.

Getis-Ord General G

This tool (also referred to as a high/low clustering tool) is similar to average nearest neighbor, except that general G determines the level of clustering across the entire study area. The tool is used to show distance at which values are clustering at a rate higher than expected by chance (Getis & Ord, 1992).

In contrast to the average nearest neighbor, a high z-score obtained from Getis-Ord general G identifies clustering (rather than dispersing). As an example, this statistic can be used to monitor the number of emergency room visits, and to look for unexpected spikes in certain disease admissions that might indicate an outbreak of a local or regional health problem.

In summary, the techniques described above may be applied to similar problems, but each reveals different nuances. For example, the average nearest neighbor tool may be used to identify TB incidents near congregated facilities, such as schools and shelters, which are known to cause exposure. The Getis-Ord general G (high/low clustering) tool can then be applied to determine if clustering of TB incidents near these schools and shelters are unusually high and thus need to be investigated.

Getis-Ord Gi*statistic (Hot Spot Analysis)

The Getis-Ord Gi* statistic is useful for finding locations of clusters (e.g., hot spots of high or low values) of events or cases. This type of technique may also be applied to area data.

In Figure 5, the Hot Spot Analysis (Getis-Ord Gi*) is applied to the violence dataset aggregated at the state level, using the attribute field containing women who reported "any form of violence" for Nigeria. These data were obtained from the 2008 Nigeria

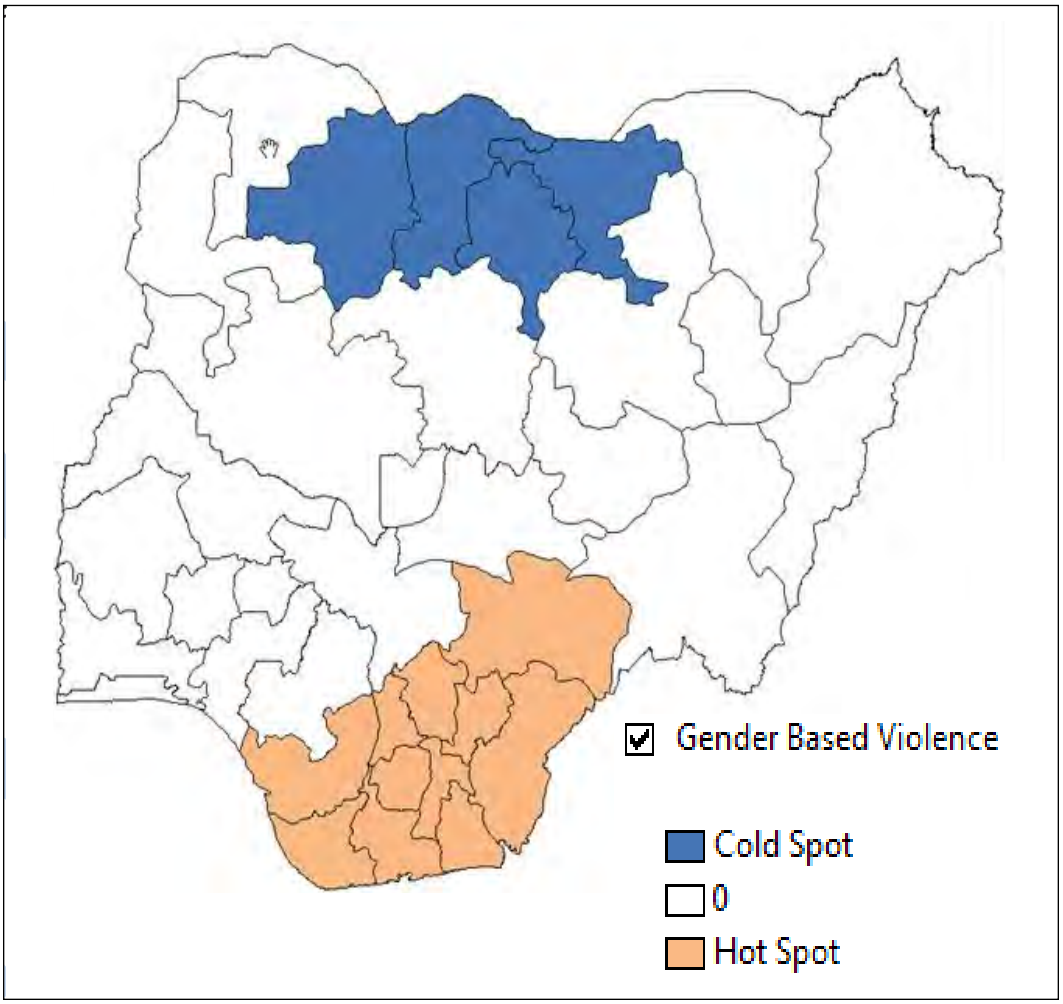


FIGURE 5
Cold and hot spots for any form of gender-based violence, 2008, Nigeria.

Demographic Health Survey (NDHS), conducted by the National Population Commission from June to October 2008 and the 2010 ANC Sentinel Surveillance Survey. The picture that emerges will not surprise those working in gender-based violence or HIV prevention programs. Most violence is found in southern states, which also have high HIV prevalence rates.

Anselin's Local Indicators of Spatial Association (LISA)

This statistic, like the Getis-Ord Gi*statistic, is based on a set of weighted features. For each administrative unit, the index is calculated based on neighboring administrative units with which it shares a border. This statistic assumes that different patterns or processes may occur in different parts of the region (Anselin, 1995).

LISA statistics are useful in M&E to describe the nature of the local distribution of the case of interest (e.g., where are areas of high or low prevalence of HIV, including spatial outliers). The high-high and low-low locations (positive local spatial

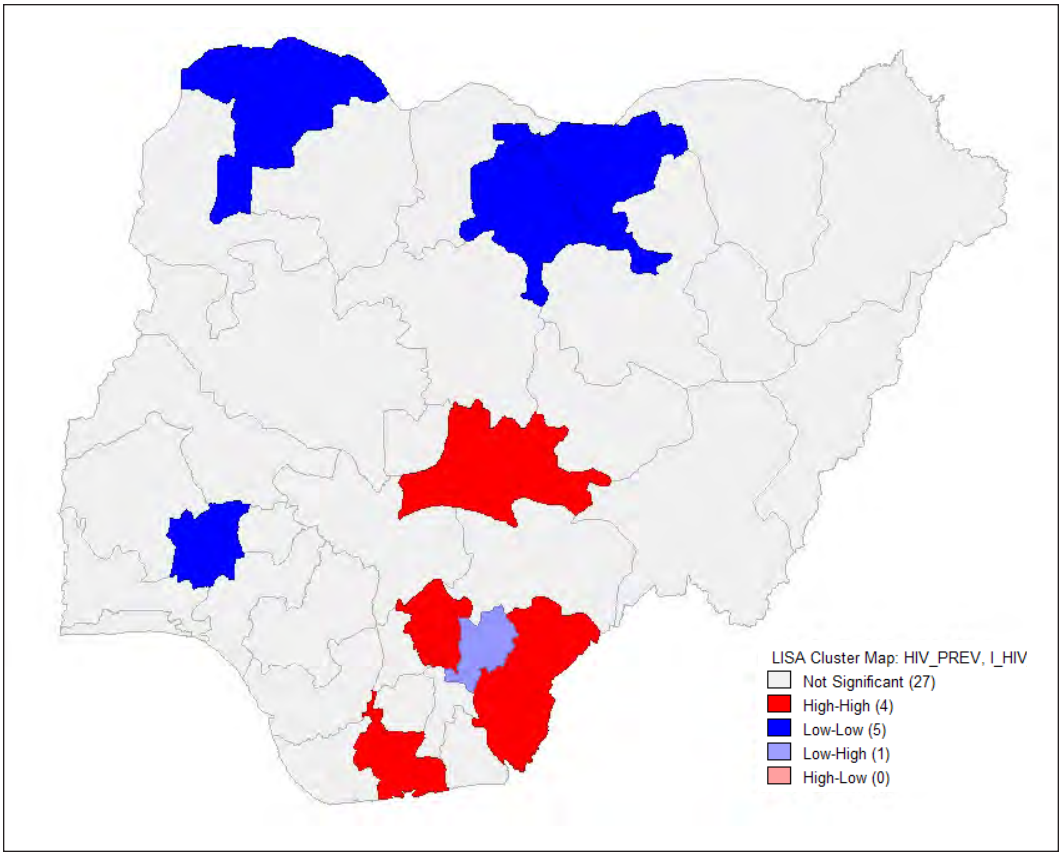


FIGURE 6
HIV prevalence rates, ANC sentinel sites, Nigeria, 2010.

autocorrelation) are typically referred to as spatial clusters, while the high-low and low-high locations (negative local spatial autocorrelation) are termed spatial outliers. While outliers are single locations by definition, this is not the case for clusters.

In Figure 6, the Local Moran's I statistic (the local indicator of spatial association) is used to analyze HIV prevalence rates in Nigeria using HIV prevalence rates aggregated to a state level. This map shows that clusters with high-prevalence states are located in the southern states of Nigeria, low-prevalence states are located in the northern parts of the country, and one cluster is located in the southwest. The map might be used by policy-makers and program managers to adjust allocation of HIV resources and target evidence-based preventive strategies to reduce HIV incidence.

Kulldorff's Scan Statistic (available in SaTScan)

Kulldorff's scan statistic is a spatial scan statistics method for detecting and evaluating statistically significant spatial clusters, such as HIV, TB incidence, or risk factors for a specific disease. For M&E, such a statistic can be applied effectively to small- or medium-sized sets of geographically referenced data (e.g., point data for cases in an administrative unit, towns, or health facilities, or provinces with one or a few

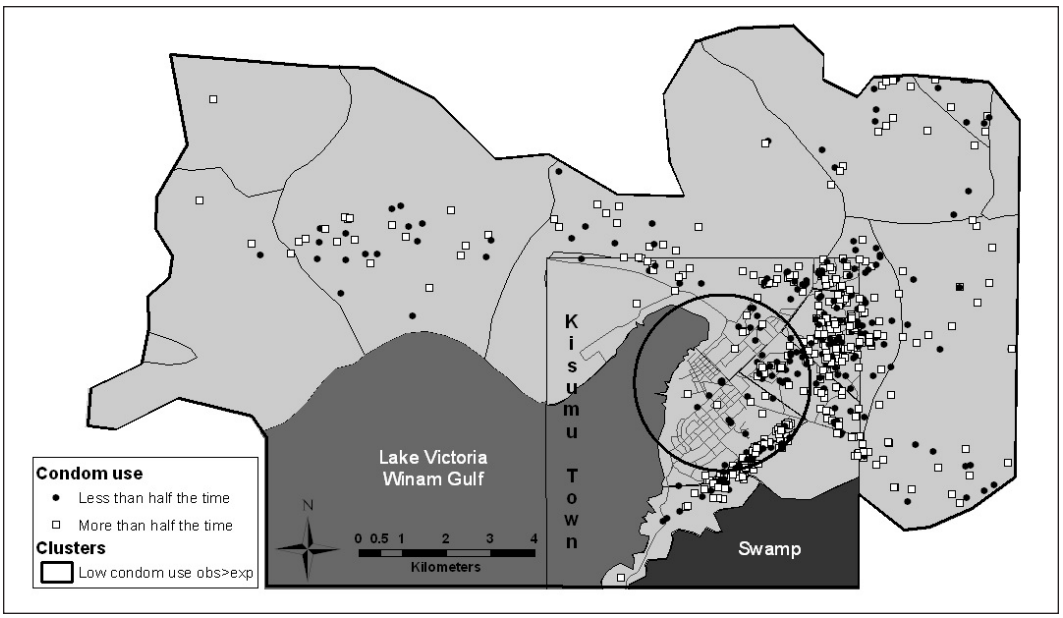


FIGURE 7
Clusters showing variations in condom use (variable less than half the time compared to more than half the time), Kisumu, Kenya, 2005–2006.

districts). The detection of emerging geographical clusters due to some sudden occurring risk factor is integral part of M&E.

In Figure 7, the Kulldorff's scan statistic (Bernoulli model) is applied to survey data collected from participants of a controlled trial of male circumcision to reduce HIV incidence in young men in Kisumu, Kenya to explore the determinants risk behavior. This m ap shows a statistically significant cluster of men who used condoms less than half the time in the past 6 months.

This cluster included 134 men, of whom 73 reported using condoms less than half the time, with 49.55 expected, resulting in a cluster rate ratio of 1.47. This geographically large cluster with a radius of 2 km covers a great part of central Kisumu, where a high concentration of participants resides—a sprawling low-income neighborhood. Program managers might use this map to help inform service delivery and infrastructure development programs for HIV services in Kisumu.

Voronoi Polygons

Voronoi polygons, or Thiessen polygons, divide a region into catchment areas around a sample of point locations. Any point which falls within one of these polygons is assumed to have a value similar to the original (central) point. In other words, any location within a particular polygon (catchment area) is closer to its central point than to any other sample point location which falls into a different polygon.

Inverse Distance Weighting (IDW)

IDW is a technique which can be used to interpolate a raster surface from a set of points. It uses a linearly weighted function between sample points; the weight is a function of inverse distance. The best results from IDW can be obtained only when sampling is sufficiently dense with regard to the local variation you are attempting to simulate. This method tends not to create "ridges" or "valleys" unless point samples come from such locations. (Watson, D.F. and G.M. Philip. 1985. A refinement of inverse distance weighted interpolation. *Geo-Processing.* 2:15–327.)

Kriging

Kriging is another tool which estimates a surface from a given set of points. It varies from the other methods by offering a more "interactive investigation of the spatial behavior of the phenomenon" being studied. Mathematical formulas determine the smoothness of the resulting surface. Thus not only is a surface produced, but also a measure of certainty of the accuracy of the predictions resulting from that surface. Kriging is a multistep process including exploratory spatial analysis of the original data, creating a variogram to serve as a model, and creating and exploring the resulting generated surface. It is mostly used in geologic applications. (ESRI, 2011 online help document for ArcGIS 10 Spatial Analyst Extension)

Kernel Density Estimation (KDE)

KDE techniques in geospatial analysis may be applied to line or point datasets with spatially extensive attributes, to show areas of estimated high or low concentration. KDE assumes that while implementation of program interventions may vary (countrywide vs. regional), the risk of ill health or unmet need exists everywhere. Thus, the risk of a negative health outcome can be viewed as being distributed continuously over space, with high scores representing areas of good service or health and low scores signifying gaps in service, unmet needs, or poor health.

For M&E, KDE techniques are particularly effective for mapping the service environment, because people are expected to seek services at their nearest service facility; KDE is useful for showing the decaying effect of distance on a facility's service area. In addition, KDE may be used to establish a relationship between available services and the characteristics of the individuals needing those services. For these types of uses of KDE, parameters associated with demand are used to allocate the service across an area of potential supply.

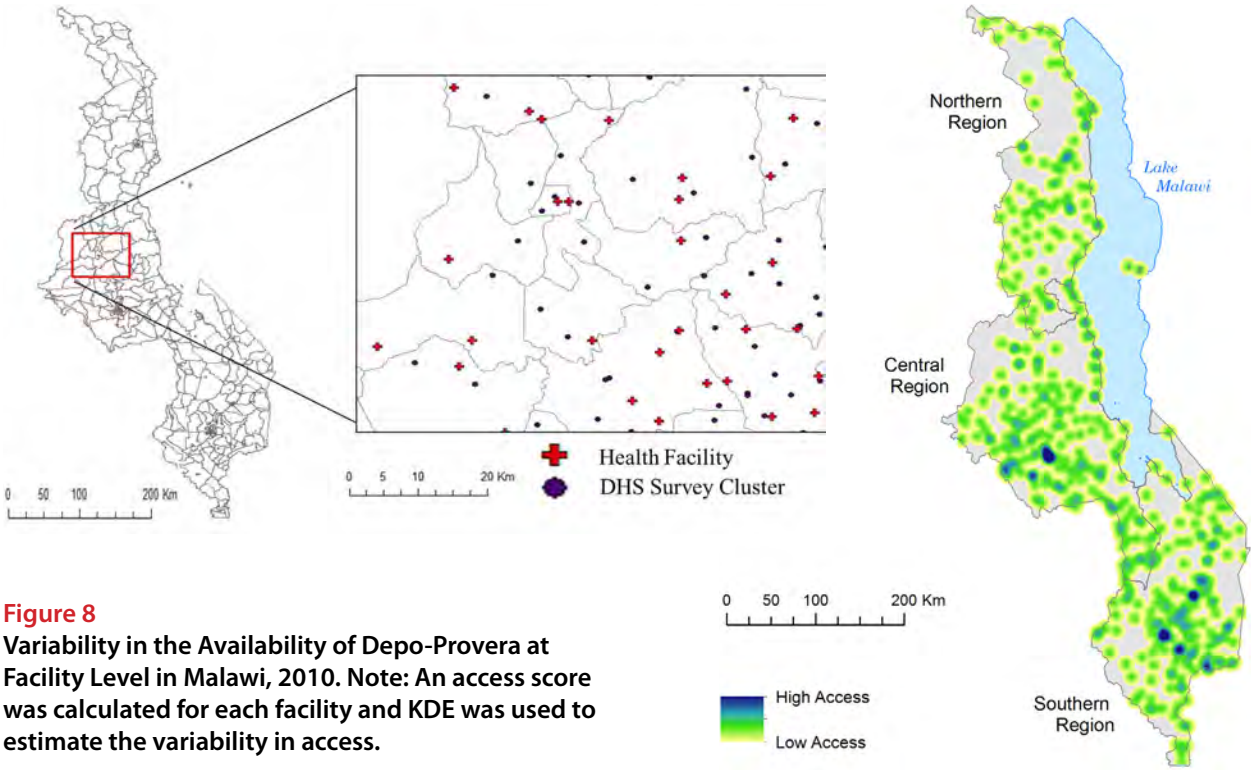


Figure 8
Variability in the Availability of Depo-Provera at Facility Level in Malawi, 2010. Note: An access score was calculated for each facility and KDE was used to estimate the variability in access.

Figure 8 shows variability in the availability of Depo-Provera at a service delivery point relative to distance from a DHS cluster in Malawi, using data from the MEASURE DHS 2010 Malawi Survey and management information from the USAID|DELIVER PROJECT for 2008–2013. The surface of the estimated service intensity is represented on the right, which could be thought of as "areas with less reliable contraceptive supply."

The interpolated surface shows patterns of point distribution with strong gaps in service in the central region of the country. This finding is not surprising for those familiar with Malawi, since this region is mostly a game reserve with few inhabitants.

Notably, and as seen on this map, KDE's strength is its ability to provide an estimate of density at any location in the area, irrespective of pre-defined administrative boundaries.

More in-depth discussions on the use of KDE in public health can be found in Gatrell, 1991; Gatrell, Bailey, Diggle, & Rowlingson, 1996; Shi, 2009, 2010; Spencer & Angeles, 2007.

CONCLUSION

GIS and related geospatial analysis methods allow M&E practitioners to use the geographic context to examine the relationship of program interventions to health outcomes and access, and to explore how service delivery can be improved. This guide provides a description of available geospatial techniques that can be used for M&E. However, while the guide illustrates how these techniques may be applied for M&E, it is still essential to build the capacity of global M&E practitioners (as well as GIS specialists). This can be done through formal training, so that all users understand these techniques and know how to apply them in their work. Programs need to build GIS capacity among evaluation teams and, where necessary, seek GIS consultation.

Regardless of the M&E situation/context, it is important to consider, from the outset, the key decision points or questions in the M&E decision-making process. The framework used in this guide describes decision points for both M&E and GIS, from the view point of the GIS end user, and summarizes common techniques for these decision points. M&E questions may be developed based on project requirements, as a way of measuring the effectiveness of a project, or as a way of assessing the project's impact. These questions are examined through the selection of appropriate data, which can be used to select appropriate geospatial analysis techniques through the given key decision. Often, the mapped evidence (visualization) and analysis of the spatial relationships underlying program-specific M&E indicators can provide the key to understanding the processes and effects of that program's interventions.

Performing geospatial analysis is not a linear process; it is a recursive and iterative process with many components, each building on the preceding step. These components include clearly defining a problem; identifying existing data sources and data acquisition; performing data preprocessing and database management; manipulating data and executing the analysis; generating a product; and interpreting and presenting the results. Most of the techniques discussed here should be used in combination with other spatial and non-spatial techniques, because each technique illustrates a specific aspect of the data under consideration. For example, it may already be known that a specific method is needed to answer the M&E spatial question, but additional data are needed before the analysis can begin; or perhaps the data have already been cleaned and manipulated, and all that is needed is to perform analysis.

The point of this guide is to provide an overview of geospatial analysis for use in M&E. The conceptual framework, description of techniques, and exercises provided in this document are intended to help M&E practitioners to understand the best geospatial analysis technique supported by the available M&E data, and to move towards conducting the complex analyses required for making the best use of scarce health care resources. The guide is not a substitute for training and experience—those remain essential—but is intended to support the use of the ever-improving availability of high-quality data and specialized software to enhance health care monitoring and ultimately, improve health care outcomes.

REFERENCES

Anselin L (1992). *Spatial data analysis with GIS: an introduction to application in the social science. Technical Report*. National Center for Geographic Informatin and Analysis, University of California.

Anselin L (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*. 27(2): 93–115. doi: 10.1111/j.1538-4632.1995.tb00338.x

Anselin L (1999). Interactive Techniques and Exploratory Spatial Data Analysis. In Longley, MF Goodchild PA, Maguire DJ, and Rhind DW (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications* (pp. 251–264). New York: Wiley.

Anselin L and Getis A (1992). Spatial statistical analysis and geographic information systems. *The Annals of Regional Science*. 26(1): 19–33. doi: 10.1007/BF01581478

Aronoff S (1989). *Geographic Information Systems: A Management Perspective*. Ottawa: WDL Publications.

Cliff AD and Ord JK (1975). The choice of a test for spatial autocorrelation. In Davies JC and McCullagh MJ (Eds.), *Display and Analysis of Spatial Data*. London: John Wiley and Sons.

Cliff AD and Ord JK (1981). *Spatial processes—models and applications.* London: Pion.

Fotheringham SA and Charlton M (1994). *GIS and Exploratory Spatial Data Analysis: An Overview of Some Research Issues*. 1: 315–327.

Gatrell A (1991). Concepts of Space and geographical data. In Maguire DJ and Rhind DW (Ed.), *Geographical Information Systems: principles and applications* (Vol. 1, pp. 119–134). London: Longman.

Gatrell A, Bailey T, Diggle P, and Rowlingson B (1996). Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. *Transactions of the Institute of British Geographers.* (21): 256–274.

Getis A and Ord JK (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*. 24(3): 189–206. doi: 10.1111/j.1538-4632.1992.tb00261.x

Goodchild M, Haining R, and Wise S (1992). Integrating GIS and spatial data analysis: problems and possibilities. *International Journal of Geographical Information Systems*, 6(5): 407–423. doi: 10.1080/02693799208901923

Hazrin H, Fadhli Y, Tahir A, Safurah J, Kamaliah M, and Noraini M (2013). Spatial patterns of health clinic in Malaysia. *Health*. 5(12), 2104-2109.

Li Y and Wei DYH (2010). Spatial-Temporal Analysis of Health Care and Mortality Inequalities in China. *Eurasian Geography and Economics*. 51(6):767–787.

Persaud, et al., (2013). "Using GIS to effectively monitor resources for managing TB in West Arsi Zone, Ethiopia," poster presented at the 44th Union World Conference on Lung Health, Oct 30–Nov 2013, Polais des Congres, Paris, France.

Rugg D, Peersman G, Carael M. Global advances in HIV/AIDS monitoring and evaluation. In *New Directions for Evaluation*. Hoboken, NJ, Wiley Periodicals, Inc.; 2004.

Shi X (2009). A Geocomputational Process for Characterizing the Spatial Pattern of Lung Cancer Incidence in New Hampshire. *Annals of the Association of American Geographers*. 99(3): 521–533. doi: 10.1080/00045600902931801

Shi X (2010). Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds. I*nternational Journal of Geographical Information Science*. 24(5): 643–660. doi: 10.1080/13658810902950625

Skiles, et al. (submitted to International Perspectives on Sexual and Reproductive Health). The effect of access on demand and use of injectable contraceptives in Malawi.

Skiles MP, Burgert CR, Curtis SL, and Spencer J (2013). Geographically linking population and facility surveys: methodological considerations. *Population Health Metrics*. 11(1): 14. doi: 10.1186/1478-7954-11-14

Spencer J and Angeles G (2007). Kernel density estimation as a technique for assessing availability of health services in Nicaragua. *Health Services and Outcomes Research Methodology*.

Tao, et al. (2012). Geographic influences on sexual and reproductive health service utilization in rural Mozambique. *Applied Geography.* 32:601–607.

Westercamp, et al., (2010). Geographic influences on sexual and reproductive health service utilization in rural Mozambique. *International Health Geographics.* 9(24):1–9.

World Health Organization. 2011. *World Health Organization (WHO) Report 2013: Global Tuberculosis Control*. Geneva: WHO.

GLOSSARY

ArcGIS
Collection of software products developed by ESRI. This includes ArcView, ArcEditor, and ArcInfo levels of functionality as well as the main applications of ArcMap, ArcCatalog, and ArcToolbox.

Attributes
Descriptive information about a geographic feature or location (e.g., the population of a neighborhood, disease information, or the name of a road) usually noted in a table.

Basemap
Map containing geographic features used for locational reference. For example, roads are commonly found on basemaps.

Buffer
Zone of a specified distance around coverage features, useful for proximity analysis.

Continuous data
Interpolatable data with an infinite number of possible values—usually a gradient of values such as elevation or slope (as contrasted with discrete data).

Coordinate
Set of numbers (x, y values) that designate location in a given reference system (coordinate system). Coordinates represent locations on the Earth's surface.

Descriptive summaries
Data condensed to capture the essence of a data set in one or two numbers. These summaries are the spatial equivalent of the descriptive statistics commonly used in statistical analysis, including the mean and standard deviation.

Geographic Information System (GIS)
Organized system comprising computer hardware, software, geographic data, and personnel whose purpose to collect, store, and update, manipulate, analyze, and display all forms of geographically referenced information.

Geospatial analysis

The process of examining the locations, attributes, and relationships of features in spatial data through overlay and other analytical techniques to address a question or gain useful knowledge.

Global Positioning System (GPS)

System of global navigation satellites used for determining specific locations on the Earth. A GPS-derived coordinate can be very accurate, which makes it a useful tool for geospatial analysis.

Line

Geographic shape that has length and direction, but no area, and connects at least two X,Y coordinates.

Map projection or projection

Mathematical model that transforms the locations of features on the Earth's surface (sphere) to locations on a two-dimensional surface (flat map).

Open-source Software

Software that can be freely used, changed, and shared (in modified or unmodified form) by anyone. Open-source software is made by many people, and distributed under licenses that comply with the Open Source Definition. QGIS is an example of open-source software.

Point

A single X (e.g. longitude), Y (e.g. latitude) coordinate pair that represents a geographic feature.

Polygon

A shape used to represent areas. A polygon is defined by the lines that make up its boundary and a point inside its boundary for identification. Polygons have attributes that describe the geographic feature they represent.

Polyline

A line made up of a sequence of line segments.

Quantile

Data classification method that distributes a set of observations into groups that each contains an equal number of observations.

Quartile

Type of quantile in which each class, or dataset, is divided into four equal groups.

Raster data (commonly referred to as grids)

Term for continuous spatial distributions represented as surfaces. A matrix of cells (or pixels) organized into rows and columns (or a grid) where each cell contains a value representing information, such as elevation. Raster data comprise digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps. Files of this data type can be quite large if the cells in the grid are small (of high resolution).

Spatial statistic or spatial analysis

Includes any of the formal techniques that study entities using their topological, geometric, or geographic properties.

Transformations

Simple methods of spatial analysis that change datasets by combining or comparing them to obtain new datasets and eventually, new insights. Transformations use simple geometric, arithmetic, or logical rules, and they can include operations that convert raster data to vector data or vice versa. They may also create fields from collections of objects or detect collections of objects in fields.

Vector data

Data type to visualize and store spatial data. Vector data consists of lines or arcs, defined by beginning and end points, which meet at nodes. The locations of these nodes and the topological structure are usually specified when stored. Features are defined by their boundaries only, and curved lines are represented as a series of connecting arcs.

Queries

Requests that examine feature or tabular attributes based on user-selected criteria and display only those features or records that satisfy the criteria.

APPENDIX 1

Checklist for Geospatial Analysis

This appendix summarizes the steps of the guide that are discussed in Section 1. This is not an exhaustive list, but a starting point. The purpose is to assure that the M&E practitioner has the right type of data to answer the evaluation questions, select a relevant spatial technique, and save time.

Step 1: Understand the Project's Context and Identify Needed Resources

- ❒ Determine the GIS task/activity's context.
- ❒ Be prepared to explain the spatial problems and interests stimulating GIS analysis work.
- ❒ Determine the broader purpose of the GIS task.
- ❒ Determine the extent to which the GIS task or tasks are relevant to the program.

Step 2: Develop the Activity's Goals and Evaluation Questions

- ❒ Formulate clearly and concisely defined location-based M&E evaluation questions.
- ❒ Ensure that the questions developed are meaningful.
- ❒ Determine if the questions can be answered based on available M&E data.
- ❒ Determine if the evaluation questions developed are specific to the project or current task.

Step 3: Define and Explore the Existing Geo-located M&E Data

- ❒ Define the available data types for spatial analysis. Below are the three basic data types that can be used with geospatial analysis.
  - Points can be described as any discrete location with latitude and longitude coordinates.
  - Lines comprise anything having a length, e.g., roads, rivers.
  - Areas or polygons comprise anything having boundaries, e.g., administrative units.
- ❒ Determine if relevant measurable indicators are available for spatial analyses. If not, identify additional sources of data to find the relevant indicators.
- ❒ No matter what the evaluation questions are, it is always best to use the smallest level of geographic analysis possible; data can be aggregated later for analysis at larger geographic units.

Step 4: Conduct Exploratory Data Analysis and Explore Relevant Methods

- ❒ Be aware of both non-spatial and spatial data sources and associated characteristics.
- ❒ Determine which, if any, denominator or data source will be used, and what the denominator is going to be. See MEASURE Evaluation's *Using Geospatial Analysis to Inform Decision Making in Targeting Health Facility-Based Programs: A Guidance Document* accessible at http://www.cpc.unc.edu/measure/publications/ms-14-88 for potential denominator data sources. Examples include:
  - Modeled population data; if using modeled data, ensure that the user has skills or guidance on how to generate quality denominator data.
  - Total population of urban and rural targets by gender differentials (male vs. female), or by age group.
  - Service utilization data on key populations, such as the population of pregnant women or the population of people living with HIV.
  - Point data obtained from Gazetteer.[1]
- ❒ If no denominator data exist, consider using other forms of a denominator, such as the stages of diagnosis or estimates-based comparisons, such as areas with high proportions of early stage of a disease versus areas with high proportion of late stage disease.
- ❒ If the GIS task entails assessing the coverage area for provided services or delineating reach of provided services, first define a catchment boundary. There are many ways to do this including;
  - Manually delineating the coverage area of the contours (i.e., creating a polygon shapefile, and carefully drawing the catchment area).
  - Using buffers to delineate the reach of the program, e.g., within 5km of a health facility or a community-based facility.
  - Using key informants or exit interviews to identify residences of those receiving services.
- ❒ Design a proper format for data entry and/or coding protocol.
- ❒ Conduct data cleaning and manipulation. Look for any out-of-range data values and check results to see if any indicator has such an issue by running basic statistics.
- ❒ Determine the resources available to support geospatial analysis, including:
  - Software such ArcGIS or open source tools, e.g., QGIS

1) Gazetteer data sets are tables or point shapefiles giving the coordinates of settlements and populated places. A low-level administrative unit polygon layer with population data in its attribute table (such as census enumeration areas) can be converted to a point dataset by taking the coordinates of the polygon centroids

- GIS skills
- Time and human resources

❒ Determine if available data will need transformation before a valid geospatial analysis can be performed in a GIS.

❒ Determine if the boundary files have a coordinate system already defined. A best practice is to project all datasets into a common coordinate system before doing analysis.

Step 5: Identify and Select the Appropriate Geospatial Technique

❒ Start with non-spatial techniques before implementing spatial approaches.

❒ Determine if the selected technique is associated with the available data types.

❒ Consider whether the selected technique is appropriately associated with the evaluation questions.

❒ Document the techniques and statistics that will help answer each of the evaluation questions.

Step 6: Consider Methodological Issues for the Technique Selected

❒ Almost every geospatial or statistical technique has assumptions behind it. Document them.

❒ Verify violations or problems with the technique selected. A minor violation usually does not create serious problems. However, if there is a serious violation, appropriate data transformation or selection of different geospatial procedures may be necessary.

Step 7: Perform the Geospatial Analysis

❒ Analyze each essential step in sequence.

❒ Determine if the analysis performed answers the evaluation questions.

❒ For complicated analyses that are beyond the user's GIS skill level, seek help.

Step 8: Assess if the Analysis Answers Evaluation Questions; Present or Refine Results

❒ Determine if the end-user understands the project, the techniques to use to perform the analysis, and how to bring the analysis results into the context of the project.

❒ Determine if the end-user understands how to interpret and summarize the output results sufficiently so that the results can be explained to a supervisor.

APPENDIX 2

Example Scenarios for Geospatial Analysis for M&E

This section presents three scenarios drawn from MEASURE Evaluation's work in Africa. The scenario topics were chosen to cover a wide variety of country, regional, and intervention examples, data, and data quality and analysis techniques, and varied levels of funding. The analysis performed in each of the scenarios involved a range of attribute data functions (e.g., queries, selecting features based on attributes), and/or integrated analysis of spatial and attribute data (e.g., overlay, interpolation, and so on), and GIS map production, such as labeling features. The scenarios address the following key M&E issues:

- patterns of service coverage for a provided service
- allocation of human resources across facilities
- linking population based outcomes with survey data

For the purpose of this guide, participants in the December 2013 meeting identified key decision points and GIS techniques that can be implemented to solve the M&E spatial problem identified in each scenario.

Scenario 1: Assessing Service Coverage, Iringa, Tanzania

Step 1: Understand the Project's Context and Identify Needed Resources

USAID/Tanzania has been supporting the scale up of HIV prevention activities across Iringa's seven districts. The funded HIV/AIDS services include (1) voluntary testing and counseling (VCT), (2) prevention of mother to child transmission (PMTCT), and (3) care and treatment (CT). USAID/Tanzania needs to evaluate whether 1) funded HIV programs have been effective at reaching the target population and 2) facilities providing HIV services are accessible by the greatest number of the target population.

While implementation of such an evaluation requires a detailed understanding of the existing HIV services and prevention activities in the region, the locations and coverage of the funded HIV prevention activities in the country were unknown. In 2012, USAID/Tanzania asked the USAID-funded MEASURE Evaluation Project to undertake this evaluation. The key decision points are outlined in Box 3.

BOX 3

KEY DECISION POINTS IN AN ASSESSMENT OF SERVICE COVERAGE IN IRINGA, TANZANIA

Step 1: Understand the Project's Context and Identify Needed Resources
USAID/Tanzania needs to evaluate whether:
1) funded HIV programs have been effective at reaching the target population and
2) facilities provided HIV services accessible by the greatest number of the target population.

Step 2: Develop the Activity's Goals and Evaluation Questions

Goals
1. Define the "coverage" area for each PMTCT service
2. Estimate the target population within each PMTCT service's reach
3. Generate a final map for PMTCT depicting facility-level service coverage across the region to be presented to USAID/Tanzania

Evaluation Questions
1. Where are the facilities?
2. What are the facility-level service reach and coverage patterns for PMTCT?

Step 3: Define and Explore the Existing Geo-located M&E Data
- Point data: Site data, service statistics for each site.
- Aggregated data: Regional HIV prevalence estimates, gridded total population data, LANDScan (2010).

Step 4: Conduct Exploratory Data Analysis and Explore Relevant Methods
- Defined the coverage or catchment area
- Determined the relevant denominator
- Estimated the total population
- Recorded the geographic coordinates
- Developed a standardized coding protocol
- Estimated the targeted population
- Extracted the service data on utilization of PMTCT services
- Performed ground truthing

Step 5: Identify and Select the Appropriate Geospatial Technique
- Survey and key informant interviews
- Document review
- GPS and digital mapping/camera
- Basic statistics
- ArcGIS-georeferencing, zonal statistics and mapping

Step 6: Consider Methodological Issues for the Technique Selected
Buffers may overestimate or underestimate the population inside the defined buffers.

Step 7: Perform the Geospatial Analysis
GIS was used to generate catchment area boundaries for each facility based on key informant interviews and generate a PMTCT coverage map for Iringa.

Step 8: Assess Whether the Analysis Answers Evaluation Questions; Present or Refine Results
The PMTCT coverage map in Figure A shows that coverage is uniform across Iringa with more facilities offering services. Between October 2009 and March 2010, coverage ranged from 30% to 80% across the middle band (northeast to the south central portion).

NOTE: For simplicity, this scenario focuses on PMTCT.

Step 2: Develop the Activity's Goals and Evaluation Questions

Since the locations and coverage of funded HIV prevention activities in the country were unknown, MEASURE Evaluation proposed using local mapping and data-gathering techniques in the evaluation. Once the evaluation methods were decided upon, the next step required deciding on what the evaluation was going to accomplish (e.g., the goals of the evaluation).

In this evaluation, the goals were threefold:

1. Define the "coverage" area for each PMTCT service.
2. Estimate the target population within each PMTCT service's reach.
3. Generate a final map for PMTCT depicting facility service coverage across the region to be presented to USAID/Tanzania.

Based on these evaluation goals, two questions to be answered were identified. The identified evaluation questions included:

1. Where are the facilities?
2. What are the facility level service reach and coverage patterns for PMTCT?

A good practice for developing evaluation questions is to create a logic model first and then use that model to derive the types of questions that the evaluation can answer. For example, the questions may ask about a single step in the logic model (e.g., How well is the output being implemented?) or may focus on the relationship between two or more steps of the logic model, such as, to what extent do the outputs of the project achieve the desired outcomes and impacts?

In the current scenario, the process of developing questions was relatively easy, because it was known beforehand that the questions would have to focus on processes, exploring how PMTCT services are operating in Iringa, such as the number and types of people served by each funded facility, and how individuals gain access to provided services. These questions helped to identify service gaps where the effectiveness of PMTCT services could be improved or adjusted.

In addition, the following considerations are critical to the success of the evaluation:

- Enough time to collect the data.

- Amount and type of data available to answer key evaluation questions
- Available resources to support and complete each data collection effort (e.g., local expertise, time, access and connections)
- Analysis techniques to be used.

Step 3: Define and Explore the Existing Geo-located M&E Data

The data for this evaluation consisted of information from routine systems and surveys. Data collection entailed several collection methods and analysis: paper-based mapping, digital data capture, surveys, and GPS technology. The GPS coordinates obtained were linked with program service statistics from semi-annual and annual reports from 2011 to 2012. It is important to note here that data collection was pre-designed to facilitate geospatial analysis, since we had already anticipated the analysis and techniques to use. The units of analysis area or catchment area were also pre-defined.

The catchment area was defined based on a survey question that asked key informants to mark on a paper map "Which locations or villages on this map do most of on the site's PMTCT clients come from?" A summary of the data types, preparatory data key decision points, and anticipated geospatial analysis techniques appears in the table below.

Data	Type	Preparatory Data Key Decision Points	Geospatial Analysis Technique
Point Data	Analysis unit	• Downloaded images and maps • Printed a paper map used as proxy for determining catchment area	• Paper-based mapping
	Site data	• Recorded geographic coordinates (GPS coordinates) of the location of each key informant interviewed • Recorded informant information on a survey, with picture of the completed field coverage map taken • Digitized paper map-based catchment area around each site on key informants' responses to the survey question that asked; "Which locations or villages on this map (mark) do most of your site's PMTCT clients come from?"	• GPS • Key informant interviews • Digital mapping/ camera • Survey • Geo-referencing
	Service statistics for each site	• Obtained program service statistics from semi-annual and annual reports, 2011 to 2012	• Spreadsheets (Excel)

Data	Type	Preparatory Data Key Decision Points	Geospatial Analysis Technique
Lattice or Region Data	Regional HIV prevalence estimates	• Calculated estimates of the sex and age distributions of the target population for the different HIV prevention programs based on data from the U.S. Census Bureau for Iringa	• Basic statistics • Spreadsheets
Geostatistical data	Gridded total population data LANDScan (2010)	• Overlaid population data to calculate the total population within the reach or catchment of the facility.	• Zonal statistics

STEP 4

Conduct Exploratory Data Analysis and Explore Relevant Methods

Preliminary data review was carried out concurrently with data collection and analysis. Below are key decision points and phases that were implemented to generate our analysis. We proceeded in the following manner:

- Defined the catchment area based on key informants' responses to a survey question that asked; "Which locations or villages on this map (mark) do most of the site's PMTCT clients come from?"
- Reviewed the created catchment area map with local experts to determine if there were any recent changes to the catchment area.
- Determined the relevant denominator (population) data source as LANDScan (2010) based on the empirical catchment area maps.
- Estimated total population (denominator) for each coverage area using population gridded data extracted from LANDScan (2010). Total population was aggregated to catchment areas using zonal statistics in ArcGIS.
- Recorded the geographic coordinates (GPS coordinates) of the location of each key informant interviewed and matched facilities with collected GPS coordinates.
- Developed a standardized coding protocol to link facility/central business office information recorded in the survey to facility "reach," which had been digitized from the maps for each service, using the district name, facility name, service examined, and map grid for the facility.
- Estimated the proportion of that total population to whom each PMTCT service is targeted using district demographic breakdown percentages and regional HIV prevalence estimates.
- Used semi-annual and annual reports to extract service data on utilization of PMTCT services from 2011 through 2012.
- Performed ground truthing (verification) on the data collected.

Step 5: Identify and Select the Appropriate Geospatial Technique

As mentioned in Step 1, the analysis for this evaluation was designed with sufficient lead-time. The evaluation's data collection was designed to facilitate geospatial analysis. For example, the available data types for answering the evaluation questions were already identified.

Step 6: Consider Methodological Issues for the Technique Selected

The key problem with the 5km buffers used to examine service reach and coverage for PMTCT services was that this method could overestimate or underestimate the population inside the defined buffer. In the current evaluation, PMTCT catchment areas with coverage estimates greater than 100 percent were observed to be smaller around the five-kilometer diameter. It is possible that we underestimated the catchment area for these facilities, or the population in these areas (see Figure A on the next page).

Step 7: Perform the Geospatial Analysis

A facility-level map of PMTCT coverage across Iringa was generated by combining service statistics for each site with target populations. GIS was used to generate catchment area boundaries for each facility based on key informant interviews and to generate a PMTCT coverage map for Iringa.

Step 8: Assess if the Analysis Answers Evaluation Questions; Present or Refine Results

It can be seen from the map that between October 2009 and March 2010, PMTCT coverage ranged from 30% to 80% across the middle band from the northeast to the south central portion of Iringa.

Further analysis showed that there are more facilities offering PMTCT services in urban areas than in rural areas, which suggests urban/rural gaps in PMTCT services. This finding requires further analysis of the data and further evaluation.

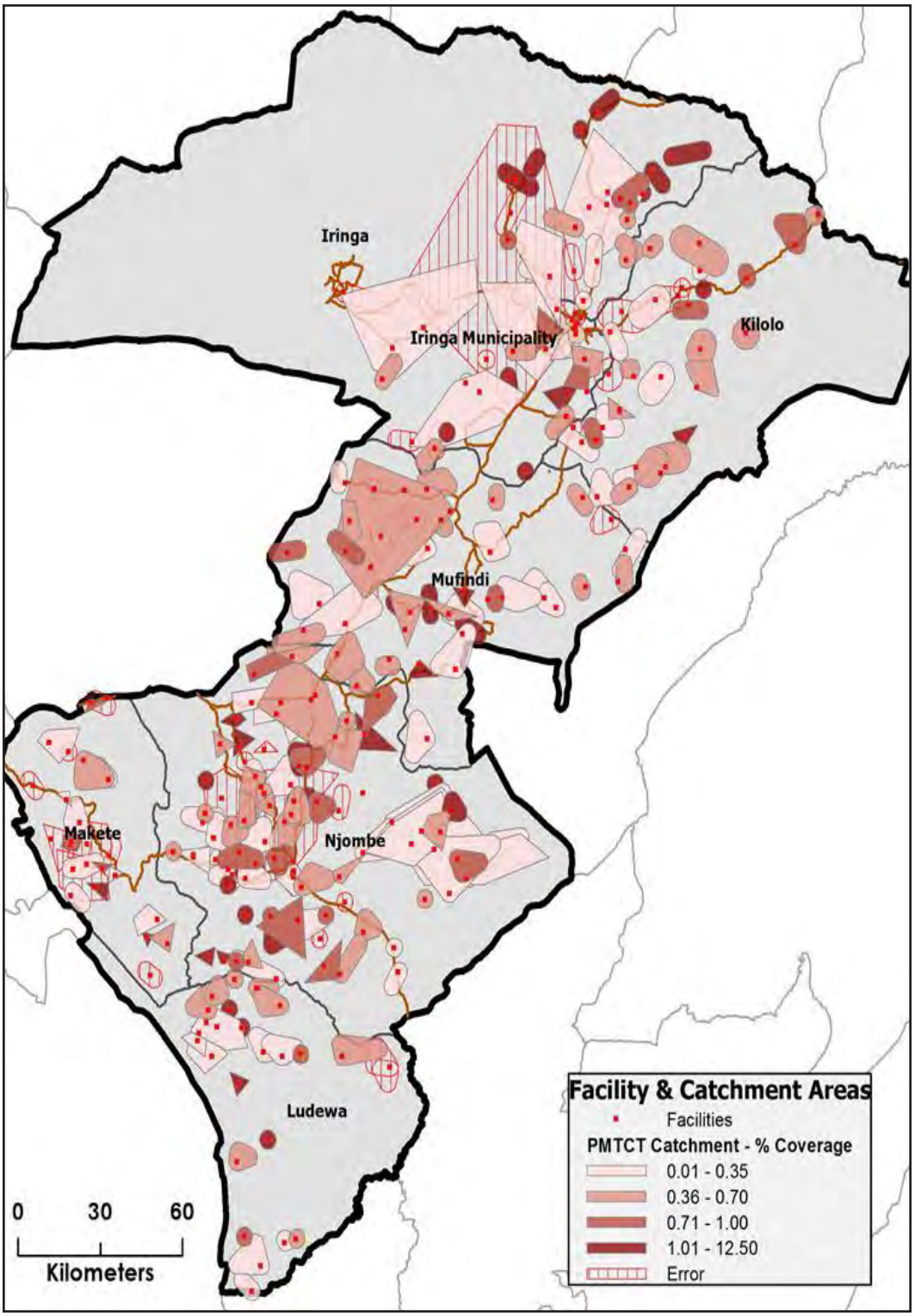


FIGURE A
PMTCT services: percent coverage (October 2009–March 2010), Iringa, Tanzania.

Scenario 2: Human Resource Allocation, Oromiya, Ethiopia

Step 1: Understand the Project's Context and Identify Needed Resources

According to the World Health Organization (WHO), Ethiopia has the seventh highest Tuberculosis (TB) burden in the world, with estimated TB cases at 261 per 100,000 in 2009 (World Health Organization, 2011). This means that monitoring and evaluation of TB screening, diagnosis, treatment, and Isoniazid preventive therapy (IPT) initiation are critical for controlling the existing disease burden.

To this effect, the USAID Bureau for Global Health's Tuberculosis TB CARE I project has been supporting TB staff in the region of Oromiya, through the provision of TB trainings, TB diagnosis and treatment standard operating procedures, and supportive supervision for improved record keeping and data use. Much of their work is piloted in the high-burden West Arsi Zone.

This scenario highlights the work that TB CARE I and MEASURE Evaluation have supported with the health staff of West Arsi Zone. This work focuses on helping staff map and assess the availability of TB services, human resources, case distribution, and case outcomes. Box 4 summarizes the decision points identified for evaluation in this context.

Step 2: Develop the Activity's Goals and Evaluation Questions

The goal for this work is to gain a better understanding of the current TB environment in Ethiopia, and to enhance policy development and improve the allocation of staff and training resources for managing TB.

Evaluation Questions

- What is the current TB disease distribution and service environment?
- How should staff and training resources be allocated?

Notice that the two questions ask about disease distribution, service environment and human resources. These questions enable evaluators to anticipate how to approach the analysis, to know what techniques to use, and to know how to present the results. For example, it was known TB case data were needed to assess disease distribution, population data to assess service environment, and program data (e.g., human resources) to assess staffing issues. The question at this point is how to determine the service environment or catchment area of each facility.

BOX 4

KEY DECISION POINTS USED IN THE ASSESSMENT OF HR ALLOCATION USING TUBERCULOSIS HMIS AND HR DATA IN OROMIYA, ETHIOPIA

Step 1: Understand the Project's Context and Identify Needed Resources
Tuberculosis TB CARE I project has been supporting TB staff in the region of Oromiya, through the provision of TB trainings, diagnosis and treatment standard operating procedures, and supportive supervision for improved record keeping and data use.

Step 2: Develop the Activity's Goals and Evaluation Questions

Goals
- Map current staff against staffing norms at each facility.
- Map the availability of health staff based on nearby population.
- Map staff case load as a function of outpatient department (OPD) patients.
- Map the proportion of staff trained for TB management compared with TB case load.

Evaluation Questions
- What is the current TB disease distribution and service environment?
- How should staff and training resources be allocated?

Step 3: Define and Explore the Existing Geo-located M&E Data
- Point data: Facility survey data; TB cases, outcomes, and other information; and GPS coordinates.
- Aggregated data: Gridded and projected total population data from Afripop and Census 2007.

Step 4: Conduct Exploratory Data Analysis and Explore Relevant Methods
- Performed and examined descriptive statistics for each indicator of interest.
- Delineated catchment areas using a 5km buffer around each facility.
- Estimated the gridded population data to target population within buffers.
- Identified "hard to reach areas."
- Identified rural facilities and compared their outcomes.
- Divided human resources based on buffers.

Step 5: Identify and Select the Appropriate Geospatial Technique
- Descriptive statistics.
- Spatial methods, GIS techniques and functions, e.g., buffers, overlay, join, spatial joins, queries and GPS.

Step 6: Methodological Issues
- Modelled total population.
- Buffers may over estimate or under estimate.
- Staffing varies across the region.

Step 7: Perform the Geospatial Analysis
- Examined descriptive statistics for TB indicators relating to the TB care pathway.
- Compared these indicators using bivariate regression.
- Examined spatial patterns in program coverage across Oromiya region.

Step 8: Assess Whether the Analysis Answers Evaluation Questions; Present or Refine Results
- TB case notification rates vary significantly between woredas
- The distribution of facilities offering TB diagnostic and treatment services correlate with the TB case distribution and population density.

There are many ways to define coverage or catchment area (see the checklist in Appendix 1). In this instance, catchment area was defined through use of buffers. We define catchment area as a 5-kilometer radius around each clinic, because this represents a reasonable walking distance for someone who might be trying to access TB services.

GIS Activity Goals

The next step requires thinking about the intended output and how to present results. Essentially, in this scenario, we wanted to:

- Map current staff against staffing norms at each facility
- Map the availability of health staff based on nearby population
- Map staff case load as a function of outxpatient department patients
- Map the proportion of staff trained in TB management compared with TB caseload.

Step 3: Define and Explore the Existing Geo-located M&E Data

After framing the evaluation questions above, we knew that case data, program data, and population data were needed to answer the evaluation questions. Below is a list of the types and sources of the data obtained. Data are listed based on the data types because the type of data and features helped determine the techniques used.

Events or Point Pattern Data

- Facility survey data with human resources data.
- TB cases and outcomes.
- GPS coordinates for all health centers and hospitals offering TB services within 5 of 11 districts of West Arsi Zone (coordinates were collected by West Arsi Zone TB focal staff).
- Road network and hydrology base data (in vector data format).

Lattice Data or Areas with Counts and Aggregated Rates (raster data format)

- Gridded population data by Kebele (smallest administrative unit used) (AfriPop).
- Projected population data (from 2007 Census) by Kebele (smallest administrative unit).

As to selecting analysis methods, our goal in this project was to generate maps showing TB cases, human resource capacity, laboratory services, and service reach. Thus, a facility-level survey was developed and was administered using a structured

questionnaire. Survey information was then exported into a GIS for mapping, comparison, and analysis. Note that the lowest administrative unit of analysis that could be used in this scenario is the Kebele. The population at Kebele administrative level was used because this was the smallest unit for which we had projected and gridded population data.

Step 4: Conduct Exploratory Data Analysis and Explore Relevant Methods

In Step 3, we established that GIS was needed for mapping. The next step was to process data in a way that made sense for GIS task goals listed in step 2. To attain project goals, we carried out the following tasks:

- Performed and examined descriptive statistics for each indictor of interest.
- Delineated catchment areas using a 5-kilometer buffer around each facility based on facility GPS coordinates that were collected by TB CARE I staff.
- Estimated gridded population for the target population within each 5-kilometer buffer of a facility.
- Identified hard-to-reach areas by overlaying facility, population, hydrology, and a road network, using help from experts in the local context.
- Identified rural facilities and compared their outcomes with those from urban areas. This was done by overlaying population density and an existing road network.
- Examined gaps and inequity in staffing by dividing human resources by the population within each buffer.

Step 5: Identify and Select the Appropriate Geospatial Technique

Both non-spatial and spatial methods were used. Non-spatial methods included descriptive statistics; spatial methods involved GIS techniques and functions such as GPS, buffers, overlay, joins, spatial joins, and queries to generate the maps we needed.

Step 6: Consider Methodological Issues for the Technique Selected

It is important to bear in mind the possible methodological issues and assumptions inherent in the techniques applied to the data. First, we modeled our total population data, which meant that our target population estimates at the facility level might include inaccuracies.

Second, while several methods of calculating catchment areas for a facility exist, each has advantages and flaws. For example, the 5-kilometer buffer used in this scenario

is simple to implement, but it may not represent the actual distance individuals are willing or able to travel to access health services. As a result, it is possible that the number of people per health care worker results either was under- or over-estimated.

Third, actual health staff workload depends on several factors, which may vary substantially across a region. These include a population that is dependent on a facility for services, incidence of disease in that population, cases presenting at the facility, time to diagnose and treat those cases, and the availability of trained staff to diagnose and treat patients at each facility.

Finally, in cases with nearby facilities, population was allocated to both. Thus, estimates of the actual population per staff may be higher than calculated. This will be especially true in urban areas where the density of facilities is high. In addition, sites not offering TB services were not included in the analysis.

Step 7: Perform the Geospatial Analysis

For analysis, descriptive statistics for TB indicators relating to the TB care pathway were examined and compared to indicators using bivariate regression. We also examined spatial patterns in program coverage.

Step 8: Assess if the Analysis Answers Evaluation Questions; Present or Refine Results

Maps 1 and 2 show the current TB disease distribution and service environment. This information enables health authorities to develop spatially targeted TB action plans aimed at addressing the gaps observed and improving the quality of TB services. Specifically, the maps show that:

- TB case notification rates vary significantly between woredas, ranging from 70 cases per 100,000 inhabitants to 180 cases per 100,000 inhabitants (Map 1).
- The distribution of facilities offering TB diagnostic and treatment services correlates with the TB case distribution and population density (Figures B and C).
- Microscopes at 25% (7/28) of the facilities were defective or partially functional at time of survey (Figure C).

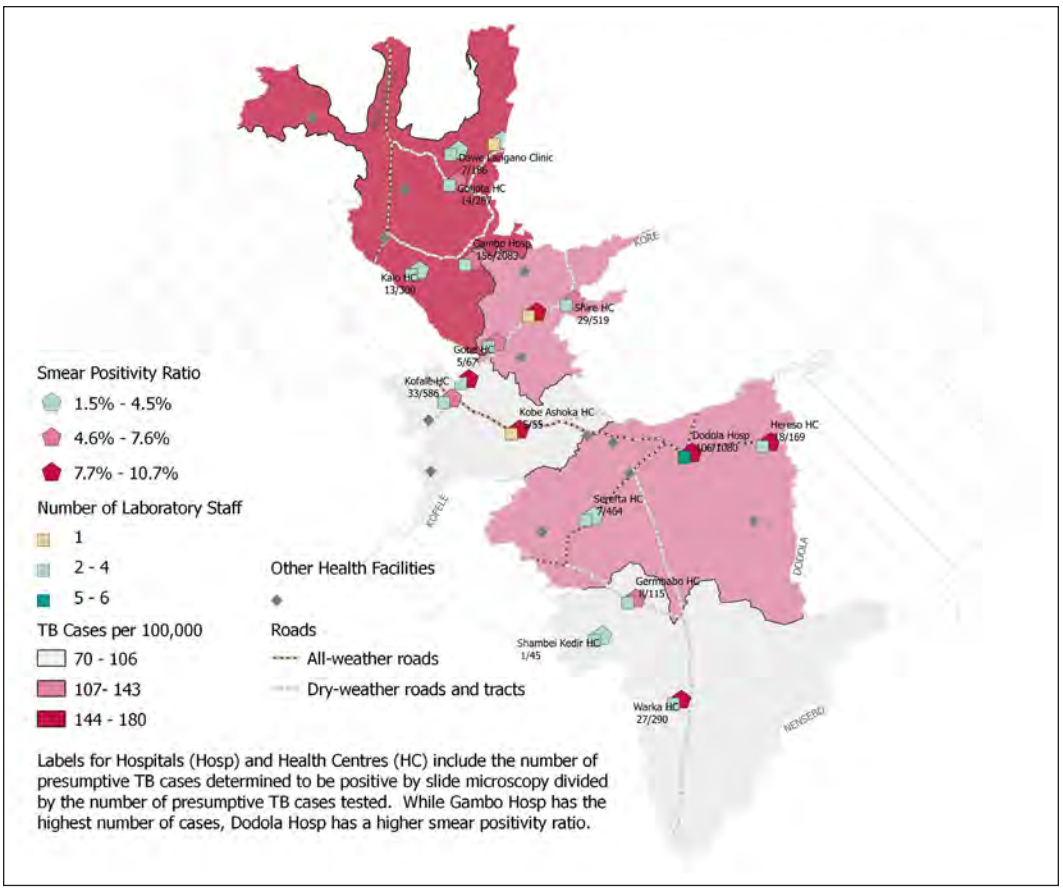


FIGURE B
Case distribution, presumptive TB screening and lab staffing.

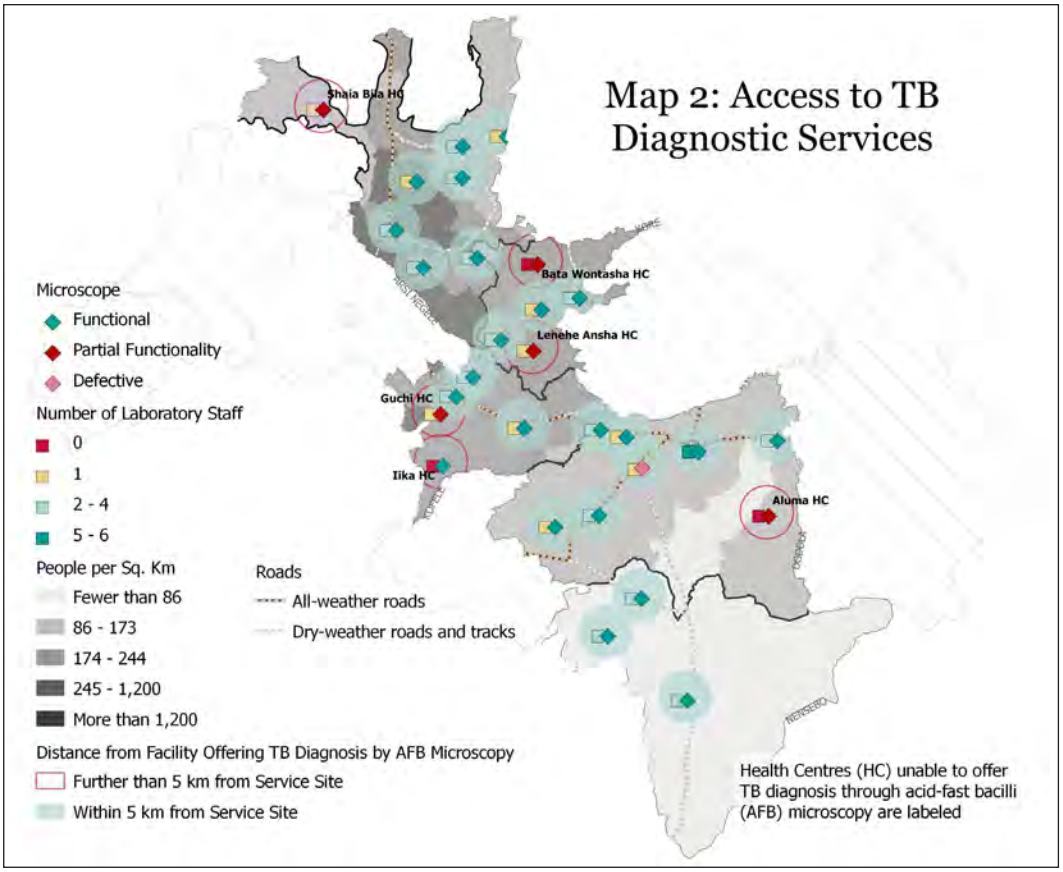


FIGURE C
Access to TB diagnostic services.

Scenario 3: Linking Population-Based Surveys to Facility Data

Step 1: Understand the Project's Context and Identify Needed Resources

The relationship between health services and population outcomes is an important area of public health impact evaluation that requires bringing together data on outcomes and the relevant service environment. Household surveys such as the DHS are a leading source of data on population health status and health care-seeking behavior, whereas health facility surveys are an increasingly accessible source of data on the availability and quality of health services. Separately, these data provide information on the population demand and service supply-side environments; but there are limited data sources that implicitly link the two independent surveys and data sources.

Step 2: Develop the Activity's Goals and Evaluation Questions

This scenario requires improved understanding of the opportunities and limitations of linking two independent surveys to understand the health impact of health services on health outcomes.

Evaluation Question

- How can population and facility data be linked to analyze the relationship between health outcomes and access to or quality of health services?

After the evaluation question was determined, the next action was to develop the critical steps necessary to answer the evaluation questions:

- Define appropriate health outcomes that could be measured in this way from the household survey.
- Define facility-based information relevant to the health outcome identified (simple access or provision of services, or more complex quality of care measures).
- Link population and facility data to analyze the relationship between outcomes, access to services, and the quality of provided health services at each facility.

Step 3: Define and Explore the Existing Geo-located M&E Data

The data used are from a population-based survey, Interim DHS, and a facility-based country Service Provision Assessment Survey (SPA). Below is a description of each data source for each data type. Note that some SPA surveys are a sample of facilities and not a complete census.

Events or Point Pattern Data: Health Facility Inventory or Survey: SPA

- 2007 Rwanda SPA, which included all public and all private facilities with at least five staff members employed, along with a subset of smaller private facilities: 538.
- SPA facilities were geo-coded (GPS location) and not displaced.
- Some SPA surveys were a sample of facilities and not a complete census.

Household Survey Data DHS

- DHS geo-codes (records GPS location for) the cluster locations, but these actual locations are then displaced by random distances to maintain confidentiality, 0–2 km for urban points and 0–5 km for rural points (1% of rural up to 10 km).
- 2007–2008 Rwanda Interim DHS, collected between December 2007 and April 2008 totaling 249 survey clusters, 185 rural and 64 urban.

Step 4: Conduct Exploratory Data Analysis and Explore Relevant Methods

The available data must be processed in a way that makes sense for the GIS task goals listed in step 2. To attain project goals, the following key data review and processing decision points were conducted:

- Determine if DHS data could be linked with SPA data.
- Develop a clear theoretical link between the service environment being examined and the population-based individual indicator.
- Abstract data from an equivalent time for analysis.
- Define catchment area and population.
- Aggregate DHS cluster data to an administrative level.
- Aggregate SPA administration (level 1 or 2 data).

Step 5: Identify and Select the Appropriate Geospatial Technique

The four methods examined for linking clusters with facilities were:

1. administrative boundary,
2. Euclidean buffer,
3. road network, and
4. kernel density estimation (KDE).

The latter two are more advanced techniques. The road network link behaved similarly to the Euclidean buffer link. Again, the appropriate geospatial linking technique depends on the question, the quality of the available data, and the GIS skill level available. Each has its own advantages and disadvantages, and each is affected differently by random displacement of survey clusters, as shown in Figure D.

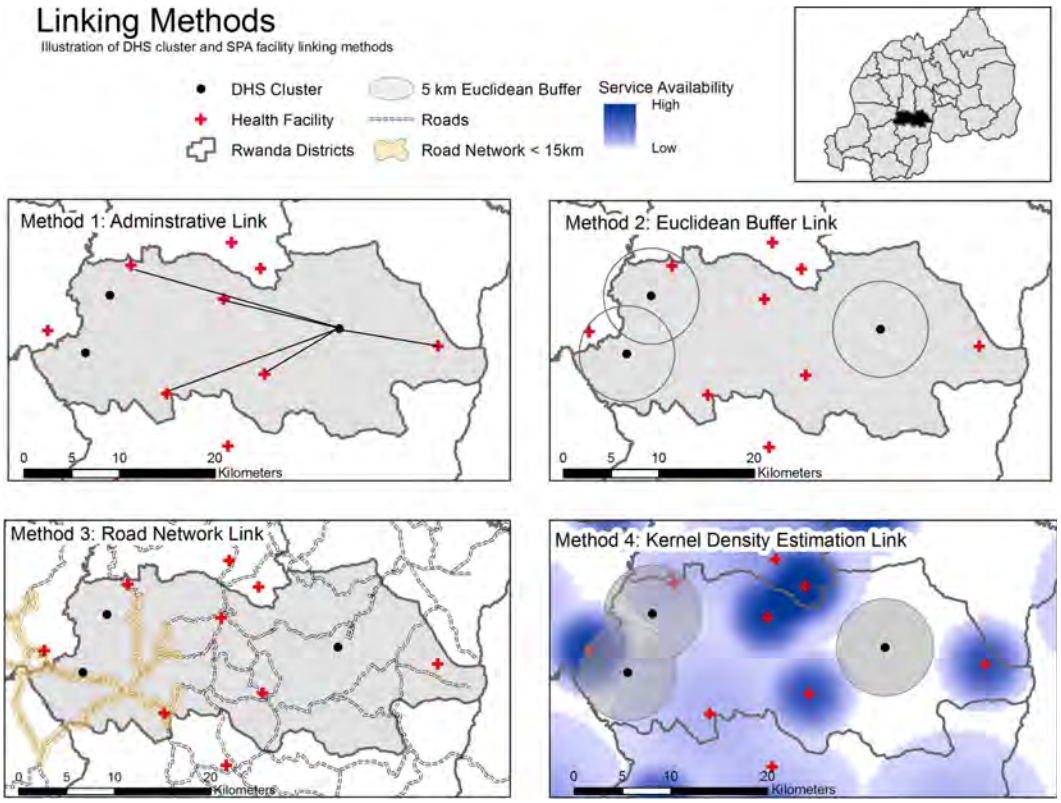


FIGURE D
Illustration of DHS cluster and SPA facility linking methods used, Rwanda.

STEP 6

Consider Methodological Issues for the Technique Selected

We considered possible misclassification bias that could occur because of using a sample of facilities in the analysis. In addition, SPA and DHS timelines are usually different, and this can lead to questions on how best to summarize the data. Furthermore, we considered the issue of ecological fallacy when linking data at different levels.

STEP 7

Perform the Geospatial Analysis

The table below highlights the various linking techniques used for the analysis, also describing the limitations and assumptions that were considered, as well as the level of skill required for each technique.

Linking Method	Assumptions	Limitations	Skill Level	Technique
Link with nearest facility	Assumes that individuals go to the nearest facility when seeking services. This method is discouraged.	High limitations	Intermediate	Requires creation of Theissen polygons for each facility, then matching the survey cluster to the Theissen polygon it falls in.

Linking Method	Assumptions	Limitations	Skill Level	Technique
Administrative link	Assumes that the overall service environment in the administrative area is representative of the service environment for any particular cluster in that area. This service environment is rarely meaningful for analysis.	Low limitations	Low	Requires applying the district/ regional average (which can be calculated inside or outside of the GIS) to the facilities in that administrative area.
Euclidean buffers	A 5-kilometer Euclidean buffer is often used to estimate a 1-hour walking period. Distances can be varied based on how far it is assumed people would be willing to travel for services.	Low limitations for descriptive analysis of common services. Higher limitations for regression.	Intermediate	Requires creating Euclidean buffers and applying these buffers to facilities that fall within them, and then determining what to do with the overlapping buffers covering the same survey cluster.
KDE	Examines an additive service environment, assuming that a high service environment from multiple facilities has greater effect than a high service environment from a single facility.	Low limitations for descriptive analysis. Higher limitations for regressions.	Fairly high	Requires creating KDE and determining the appropriate service environment statistic.

Step 8: Assess if the Analysis Answers Evaluation Questions; Present or Refine Results

A significant relationship was found between the facility readiness score for family planning KDE and an individual's use of a modern family planning method. Results show the following:

- 69% of clusters were within a 5-kilometer radius of a facility offering the pill.
- 65% of clusters were within a 5-kilometer radius of a facility offering injectables.
- For facilities offering implants, 27% of clusters were within a 5km radius.

The corresponding percentages for clusters within the same district as a facility offering a service are 96%, 100%, and 75%, respectively. In conclusion, each of the four ways to link population surveyed data with facility service environments using routine or typically collected data has limitations. For example, displaced clusters, while increasing respondent confidentiality, introduce bias into analyses when linking the clusters with nearby facilities.

Creating these individual-level links requires a census of facilities. This could be from a national survey, or from a master facility list combined with routine health data. The method of envisioning a service environment and the GIS skills available will determine the analytic method.

For more information on the above scenario, see

- Skiles MP, Burgert CR, Curtis SL, and Spencer J (2013). Geographically linking population and facility surveys: methodological considerations. *Population Health Metrics*. 11(1): pp.14. doi: 10.1186/1478-7954-11-14.

APPENDIX 3

Commonly Used Geolocated Routine Data Sources for M&E

The Master Facility List should provide the basis for a common unique code across the HMIS, LMIS, and IDSR.

Health Systems Registry Database (HSRD) / Master Facility List

Master facility list typically attempts to document all health facilities in a country. It can include information on facility location (admin area or GSP coordinates), services, or human resources.

- **Source:** country MOH.
- **Frequency:** annually.
- **Scale:** facility locations might be specified by administrative areas or GPS coordinates.
- **GIS challenges:** an authoritative master facility list is not always available or complete. Validation of the facility list can require extensive work.Are all facilities—public and private—present? Are the GPS coordinates correct? Is data up-to-date?

Health Management Information System (HMIS)

An HMIS contains routine statistics for facility-based services.

- **Source:** country MOH.
- **Frequency:** monthly or quarterly.
- **Scale:** the system may be paper-based (data typically aggregated before being entered into a computer) or electronic (with data at the facility level).
- **GIS challenges:** data quality, identifying a facility's catchment population, and lack of private sector engagement.

Logistics Management Information System (LMIS)

A LMIS contains current and dispensed information for medicines and other commodities for facilities or pharmacies.

- **Sources:** USAID|DELIVER Project, country MOH.
- **Frequency:** monthly or quarterly.
- **Scale:** the system may be paper-based (data typically aggregated before being entered into a computer) or electronic (with data at the facility level).

- **GIS challenges:** data quality is often a concern. Further, when mapped at the facility level, the catchment population needs to be estimated for coverage indicators.

Integrated Disease Surveillance and Response System (IDSR)

Cases for 40 priority diseases are collected through a surveillance system. This might or might not be integrated with the HMIS.

- **Source:** WHO often supports the country MOH.
- **Frequency:** weekly.
- **Scale:** the system may be paper-based (data typically aggregated before being entered into a computer) or electronic (with data at the facility level).
- **GIS challenges:** over-reporting in areas with good surveillance. Maintaining confidentiality for cases mapped to the household level.

Community Health Information System (CHIS)

A CHIS is typically a system designed to allow community health workers to document and track cases seen and services offered. In most cases, the data is aggregated to a nearby health facility and submitted through the HMIS.

- **Source:** country MOH.
- **Frequency:** monthly or quarterly.
- **Scale:** though community-level data is collected, it is typically aggregated to the level of a near-by health facility.
- **GIS challenges:** data in an electronic format with GPS coordinates at the lowest level of disaggregation is rarely available.

Vital Statistics

Vital statistics provide routine information on new births, number of deaths by age and sex, and cause of death by age and sex. When unavailable, some of these can be estimated from population based surveys.

- **Sources:** central statistics agency, country MOH.
- **Frequency:** monthly (rarely available).
- **Scale:** data can be collected or aggregated at either the community or the facility level, depending on the data collection process. GPS coordinates might not be available at the scale of data collection, and the data might need to be aggregated to higher scales where boundary files are available.
- **GIS challenges:** data is rarely available. Rates may be unstable for uncommon causes of death if the population is low and/or the geographic scale is high.

APPENDIX 4

Prerequisites for Geospatial Analysis

To carry out an effective and accurate geospatial analysis that is useful for M&E assessments and decision-making, both M&E and GIS practitioners must make realistic assessments of available resources: both the quality and appropriateness of available data, and the skills and understanding of those who will develop the analysis and use the findings. The following points cover the essential steps for judging whether these are adequate.

Before Starting: Prerequisite Steps for Geospatial Analysis

There are three prerequisite steps to consider before starting a geospatial analysis:

- Understand the context of the management and/or evaluation question; otherwise, geospatial analysis will be a desk exercise decoupled from decision-making.
- Understand the limitations of the data and the implications for available geospatial techniques before investing time and effort in data analysis.
- Acknowledge the GIS skills, strengths and limitations of the GIS user(s), and seek assistance before starting a spatial analysis beyond one's GIS skill level.

To assess your own core geospatial abilities and knowledge visit http://www.careeronestop.org/competencymodel.[2] The Geospatial Technology Competency Model covers a whole spectrum of expected skills and growth patterns. Additionally, evaluators and program managers play interdependent roles in the process of developing a geospatial analysis and using the results:

- The evaluation team's ability to properly interpret and summarize the results of the analysis is vital to help decision-makers understand the findings of the spatial analysis and the relevance of the findings to the program or project.
- Policy-makers and program managers must be clear about two essential steps: identify appropriate data and provide honest and accurate interpretation of results.

Spatial Analysis is an Iterative Process

Geospatial analysis is a complex procedure. Like other types of analysis, it is often an iterative process: solving one spatial problem can give rise to new questions.

---

2) If using a hard copy of this guide, check your GIS skills against this list, "Core Geospatial Abilities and Knowledge," found on the CareerOneStop website: http://www.careeronestop.org/competencymodel/blockModel.aspx?tier_id=4&block_id=708&GEO=Y (accessed July 30, 2014).

Practitioners will need to clearly redefine the spatial problem, identify the right data and/or indicators, perform and evaluate the analysis, and interpret and present the findings. In addition, the map generated may not completely meet the needs of the evaluation questions; or additional data may be needed to enhance the findings. In either case, and at any point in the analysis process, it is appropriate to refine the output by repeating one or more steps, to generate the analysis that best fits the intended result or addresses management/evaluation questions.

Data Are Fundamental

Before initiating a decision to perform geospatial analysis, the M&E practitioner needs to undertake evaluation to assess the availability of any existing public health and programmatic data and consider its usefulness to the current project or evaluation questions. When considering the usefulness of the existing geo-located data, it is essential to know the data source and its frequency, scale, and limitations, as well as why and how the data was collected.

Appendix 3 provides a list of commonly used geo-located data sources for M&E, including the known limitations. In addition, local governments, health ministries, universities, and other evaluation agencies are rich sources of published and unpublished statistics.

Sometimes, existing surveys or administrative data do not entirely satisfy the needs of current analysis. In that case, it is necessary to combine data from a variety of sources. However, note that the integration of data from multiple data sources and formats (e.g., points, lines, and areas), data at different scales, or data that possesses innate errors may lead to uncertainties in the accuracy of data and this should be taken into account (Skiles, Burgert, Curtis, & Spencer, 2013). Further, the scale of available geographical routine data and the various population proxies employed often make it difficult to link data and draw general conclusions.

The Importance of Continued GIS Training and Guidance for M&E

While this guide provides information on the key starting points for selecting appropriate geospatial analysis techniques as well as provides links and references to a number of excellent resources, it is not a substitute for GIS training, courses, and/or systematic guides that focus on specific analysis functions, methods, and statistics to use to solve a spatial M&E problem.

To answer a spatial M&E evaluation question, it may be necessary to perform individual analysis functions in a GIS, but the analysis must be performed within the context of a more complete analysis strategy (Aronoff, 1989). This means that the analysis may involve a range of attribute data functions (e.g., performing queries, selecting features based on attributes), and/or integrated analysis of spatial and attribute data (e.g. overlay, interpolation, and so on).

For free GIS training opportunities, visit ESRI's training section at http://www.esri.com/training/main/training-catalog. In addition, the following basic GIS training materials are available online.

- Getting Started with GIS (for ArcGIS 10) accessed at: http://training.esri.com/gateway/index.cfm?fa=catalog.webCourseDetail&courseid=1911 (accessed July 30, 2014).
- QGIS training materials accessed at www.qgis.org/en/documentation/manuals.html.
- Geographic Approaches to Global Health, an eLearning course available from MEASURE Evaluation (https://training.measureevaluation.org/certificate-courses/geo-global-health-en) or take the follow-up course which includes hands-on practical exercises—*GIS Techniques for HIV/AIDS Programs*—available at https://training.measureevaluation.org/node/90

If you would like to learn the value of linking a GIS and a LMIS, and learn the steps you need to take when you link the two systems, see the USAID | DELIVER PROJECT's "A Guide for Preparing, Mapping, and Linking Logistics Data to a Geographic Information System," accessible at: http://deliver.jsi.com/dlvr_content/resources/allpubs/guidelines/GuidLinkLogDat_GIS.pdf.

APPENDIX 5

Spatial Considerations: Modifiable Area Unit Problem

The modifiable area unit problem (MAUP) is a methodological error that needs to be considered in Step 6. It can occur during the analysis phase and potentially distort the results of certain spatial analysis techniques. MAUP consists of two problems—zone and scale—and isolating the effects of one from the other is often difficult. The scale problem is most likely to occur when data from one scale of geographical units (such as a village) is aggregated into units of greater or lesser areal units (e.g., wards, districts, provinces). The zonal problem[3] arises when keeping the same scale of evaluation (e.g., at the regional level) but changing the actual shape and size of those areas. Figures E and F illustrate the scale problem.

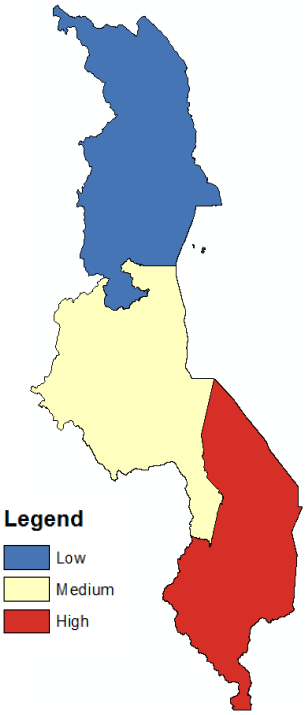


FIGURE E
Regional-level HIV presence, Malawi.

Figure E shows the distribution of mean HIV prevalence across Demographic and Health Survey (DHS) regions in Malawi, using regional data. Based on this map, there is distribution patterns are greatly simplified. Higher levels of HIV prevalence appear to be concentrated in the southern region, while lower prevalence is shown in the northern region of the country. However, the district mean HIV prevalence map on right depicts a more complex situation, showing that districts in the central portion of the country actually have lower rates of HIV. This map illustrates how a district-level analysis yields a much clearer picture of the variation in mean HIV prevalence across the country.

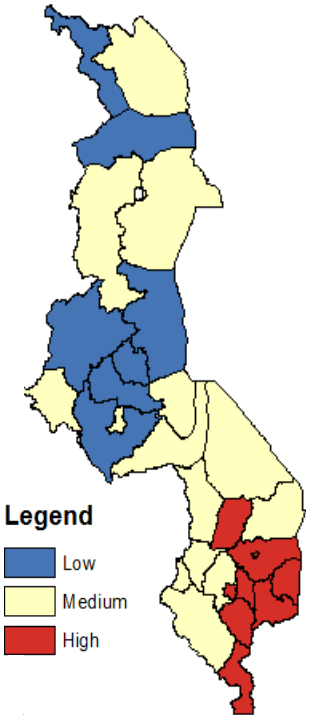


FIGURE F
District-level HIV prevalence, Malawi.

This example demonstrates that the two different geographic units used for analysis yielded very different results from the same data. Maps such as these may have very important implications for allocating HIV resources or developing HIV interventions to target priority areas. Therefore, it is important to consider the choice of geographical aggregation units to use in any geospatial analysis, to ensure that the spatial variation of the data is displayed in a comprehensible manner.

3) The zonal problem arises from the way in which administrative units of similar size and number, but with different boundaries, are defined.

APPENDIX 6

Recommended Reading

General

- The USAID | DELIVER PROJECT (2012). A *Guide for Preparing, Mapping, and Linking Logistics Data to a Geographic Information System*. http://www.jsi.com/JSIInternet/Inc/Common/_download_pub.cfm?id=12241&lid=3.
- Stewart J, Sutherland E, Wilkes B, Spencer J, and Herndon N (2011). *An Overview of Spatial Data Protocols for Family Planning Activities: Why and How to Include the "Where" in Your Data*. Chapel Hill, NC: MEASURE Evaluation, 2011.
- Stewart J, Wilkes B, Spencer J, and Herndon N. *An Overview of Spatial Data Protocols for HIV/AIDS Activities: Why and How to Include the "Where" in Your Data*. Chapel Hill, NC: MEASURE Evaluation.

Interactive and Exploratory Data Analysis

- Anselin L (1999). Interactive techniques and exploratory spatial data analysis, in P. Longley, M. Goodchild, D. Maguire and D. Rhind (Eds.), *Geographical Information Systems*. 2nd Ed: 25–264. New York, Wiley.
- Buja A, Cook D, and Swayne D (1996). Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics*. 5:78–99.
- Fotheringham et al., (2000). *Chapter 4: Exploring Spatial Data Visually, in Quantitative Geography: Perspectives on Spatial Data Analysis*, pp. 65–92.
- MacDougall EB (1992). *Exploratory Analysis, Dynamic Statistical Visualization, and Geographic Information Systems, Cartography and Geographic Information Systems*. 19(4):237–246.
- Messner S, Anselin L, Baller R, Hawkins D, Deane G, Tolnay S (1999). The spatial patterning of county homicide rates: an application of exploratory spatial data analysis. *Journal of Quantitative Criminology.* 15:423–450.
- O'Loughlin J (2002). The electoral geography of Weimar Germany: exploratory spatial data analysis (ESDA) of Protestant support for the Nazi party. *Political Analysis*. 10:217–243.
- Wegman E, Solka J, and Martinez W (Eds.) (2003). Data Visualization. *Special Issue of Computational Statistics and Data Analysis.* 28:421–649.

Quadrant Count Analysis for M&E and Public Health Data

- Chaikaew N, Tripathi NK, and Souris M (2009). Exploring Spatia l Patterns and Hotspots of Diarrhea in Chiang Mai, Thailand. *International Journal of Health Geographics*. 8(36):1–10.
- *Spatial Analysis Applied to Epidemiology Seminar Final Report.* http://aulavirtual.ig.conae.gov.ar/moodle/pluginfile.php/513/mod_page/content/72/seminario_Cu%C3%A9llar.pdf

Distance Techniques

- Getis A and Ord JK (1992). The Analysis of Spatial Association by Use of Distance Statistics, *Geographical Analysis*. 24(3): pp.189–206.
- Alabi MO (2011). Towards Sustainable Distribution Of Health Centers Using GIS: A Case Study from Nigeria. *American Journal of Tropical Medicine & Public Health*. 24(3):130–136.
- Lohela TJ, Campbell MRO, and Gabrysch S (2012). Distance to Care, Facility Delivery and Early Neonatal Mortality in Malawi and Zambia. *PLOS One*. 7(12):e52110.

Point Pattern Analysis

- Cressie NA (1993). *Statistics for Spatial Data, revised edition.* John Wiley & Sons, New York.
- Dale MRT (1999). *Spatial Pattern Analysis in Plant Ecology.* Cambridge University Press, Cambridge.
- Diggle PJ (2003). *Spatial Analysis of Spatial Point Patterns, 2nd edition.* Arnold Publishers, London, United Kingdom.
- Stoyan D and Helga S (1994). *Fractals, Random Shapes and Point Fields (Methods of Geometrical Statistics)*. John Wiley & Sons, New York.
- Upton GJG and Bernard F (1985). *Spatial Data Analysis by Example (Volume 1: Point Pattern and Quantitative Data).* John Wiley & Sons, Chichester/New York.

Applying the indexes of spatial autocorrelation with programmatic data

- Bailey TC and Gatrell AC (1995). *Interactive Spatial Data Analysis*. John Wiley and Sons, New York, NY.
- Haining R (2003). *Spatial Data Analysis. Theory and Practice*. Cambridge University Press, Cambridge.
- Longley, et al., (2001). *Geographic Information Systems and Science*. John Wiley and Sons, New York, NY.

- Moise I and Kalipeni E (2012). Applications of geospatial analysis to surveillance data: a spatial examination of HIV/AIDS prevalence in Zambia. *GeoJournal.* 77(4):525–540. doi: 10.1007/s10708-010-9349-7.
- Pui-Jen et al., (2009). Spatial Autocorrelation Analysis of Health Care Hotspots in Taiwan in 2006. *BMC Public Health*. 77(464):1–13.
- Upton, GJG and Fingleton B (1985). *Spatial Data Analysis by Example (Volume 1: Point Pattern and Quantitative Data)*. John Wiley & Sons, Chichester/New York.
- Waller AL and Gotway CA (2004). *Applied Spatial Statistics for Public Health Data*, New Jersey, John Wiley and Sons, Chapter 7. *Spatial Clustering of Health Events: Regional and Count Data*. New York, NY: John Wiley and Sons.

Applying Getis and Ord's Gi*statistic for M&E data

- Cromley EK and McLafferty SL (2002). GIS and Public Health. New York: The Guilford Press.
- Getis A and Ord JK (1995). Local Spatial Autocorrelation Statistics Distribution Issues and an Application. *Geographical Analysis.* 27(4).
- Haining R. *Spatial Statistics and Analysis of Health Data*. In: Gatrell A, Loytonen M, eds. GIS and Health. London: Taylor & Francis, 1998, pp. 29–47.
- Ord JK and Getis A (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*. 27(4):286–306.
- Mitchell A (2005). *The ESRI Guide to GIS Analysis, 2*. ESRI Press, Redlands, CA.

Applying LISA for Geospatial Analysis and M&E

- Anselin L (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*. 27:93–115.
- Ord JK and Getis A (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*. 27: 286–306. doi: 10.1111/j.1538-4632.1995.tb00912.x.
- Anselin L (1996). The Moran scatterplot as an ESDA tool to Assess Local Instability in Spatial Association. *Spatial Analytical Perspectives on GIS*. 4:111–127.

Applying SaTScan for M&E

- Jacquez GM and Greiling DA (2003). Local clustering in breast, lung and colorectal cancer in Long Island, New York. *International Journal of Health Geographics*. 2(1):3.
- Jacquez GM and Oden N (1996). Statistical Software for the Clustering of Health Events. *Stat Med.* 15(15):951–952.

Applying Geostatistical Techniques for M&E

- Cressie, N.A.C. (1993). *Statistics for spatial data.* New York, NY: John Wiley and Sons.
- Goodchild, M.F, Haining R and Wise S. (1992). Integrating GIS and geospatial analysis: problems and possibilities. *International Journal of geographic Information Systems*. 6(5):407–423.

Introductory texts on geostatiscs

- Isaaks, E. and Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*. New York: Oxford University Press.
- Chilès, J and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty.* New York: John Wiley and Sons.

APPENDIX 7

Step-By-Step GIS Exercises

In previous sections, this document outlined eight key essential steps which guide the process for using GIS for M&E. The following four exercises provide hands-on step-by-step guidance for step seven—performing the geospatial analysis—for an illustrative set of geospatial tools and open source software platforms. They were developed as an add-on to the joint MEASURE Evaluation and The DHS Program QGIS curriculum accessible at: http://spatialdata.dhsprogram.com/resources/.

Data used in these exercises can be accessed at: http://www.measureevaluation.org/publications/ms-14-98.

Once you have successfully downloaded the data files, you will need to Unzip them to a location on your local drive, such as C:\MyWork\ExerciseData. You will also need to download and install the free software package GeoDa, located here: http://geodacenter.asu.edu/software/downloads.

And the software package QGIS, located here: http://www.qgis.org/en/site/. The QGIS installation is quite large (around 300Mb) and will require a stable and fast internet connection to download.

Exercise list:
1. Exploring GeoDa (requires GeoDa and the Exercise Data folder)
2. Mapping Clusters (requires GeoDa and the Exercise Data folder)
3. Delineating Catchment Area (requires QGIS and the Exercise Data folder)
4. Hot Spot Analysis (requires GeoDa and the Exercise Data folder)

Exercise 1: Exploring GeoDa

NOTE: In order to perform this exercise, you must already have GeoDa installed on your computer, and you must also have downloaded a ZIP file called "Exercise Data". Please see the explanation at the beginning of Appendix 7 for more instructions.

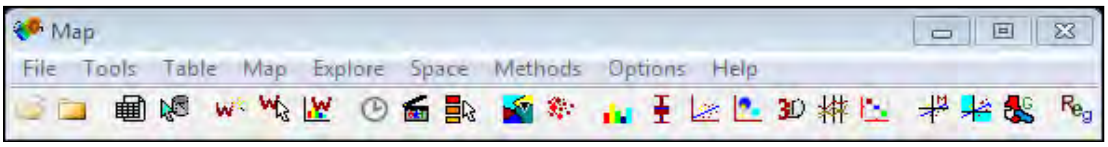
This exercise will introduce you to GeoDa and some of the basic functions of the program. By the end of this exercise, you should be able to:

1. Launch GeoDa
2. Open a shapefile
3. Explore the attribute table
4. Change map view
5. Create a thematic map

1. Launch GeoDa

If you have a shortcut for GeoDa on your desktop, double-click it to start GeoDa. Otherwise, navigate to C:\Program Files\GeoDa and then double-click on GeoDa.exe

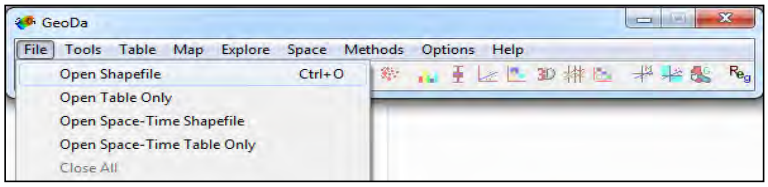
You are now looking at the interface of the GeoDa application. This looks different from many of the other GIS software types, as the only thing that appears is the navigation pane. The navigation pane provides access to all the different functions that GeoDa provides. Take a moment to explore the navigation pane to identify and locate functions that may be of use to you.



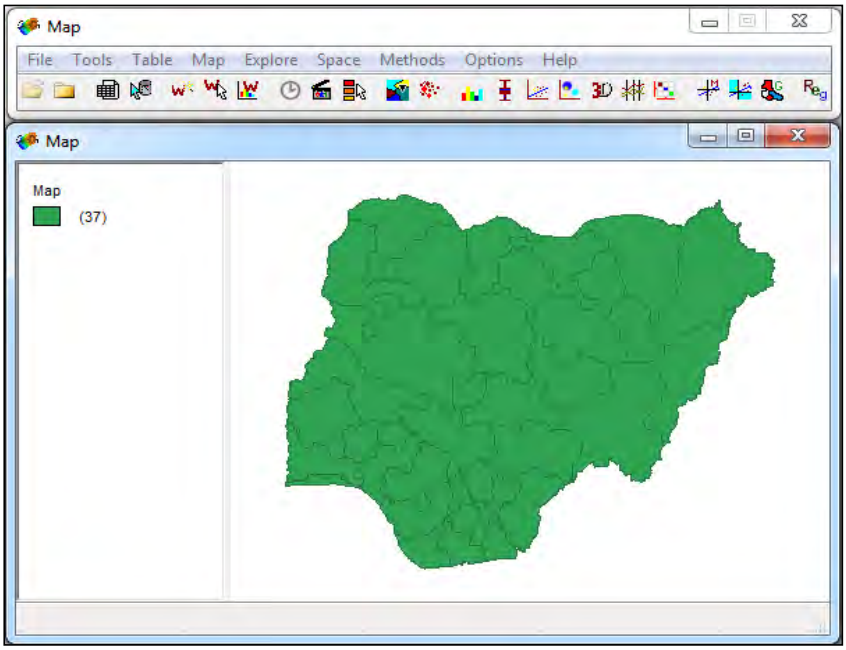
2. Open a shapefile

Unlike other GIS programs that work with map documents that may contain multiple shapefiles, GeoDa only allows the user to work on one shapefile at a time. You will use the shapefile provided in the exercise data zipfile to examine some of the features of GeoDa. (NOTE: These files need to be unzipped and installed on your local drive before they can be used. What we are calling here a "shapefile", will actually be seen by Windows as a collection of files, all having the same name but different extensions such as .shp, .shx, and .dbf. These should all be kept together in one location. It is fine to keep them all in the folder called "Exercise Data".)

2a. Click File > Open Shapefile



2b. Navigate to the data Folder, select Violence_and_HIV.shp, and click OK. When the Violence_and_HIV.shp opens, you will see a window appear that displays a map of Nigeria. It should look similar to the image below.



NOTE: The map of Nigeria and the navigation pane are two separate windows and are not automatically lined up or arranged in a way that is preferable to you. The navigation pane can be moved but not resized. Using the same methods as you would with any other window, you can rearrange and resize the map window until you are satisfied with the layout.

3. Explore the Attribute Table

It is good practice be familiar with and understand the data before doing any geospatial analysis. The best way to do this is to explore the attribute table.

3a. Click the Open Table icon on the navigation pane.

3b. Examine the table until you understand the values and field names of your data.

Table

	OBJECTID	ADM1NAME	Shape_Leng	Shape_Area	State_imp	State	HIV_Prev	CI	Any_V_ever	Phys_V_Eve	Sex_V_Ever
1	1	Abia	4.19898421317	0.38548504348	Abia	Abia	7.300000	5.6 - 9.0	36.100000	33.100000	10.500000
2	2	Abuja	3.42954727096	0.60426039987	Abuja	FCT	8.600000	7.0 - 10.2	8.300000	8.100000	1.400000
3	3	Adamawa	10.93622880800	2.81727148228	Adamawa	Adamawa	3.800000	2.6 - 5.1	22.300000	20.600000	7.400000
4	4	Akwa Ibom	3.55365829392	0.54982193969	Akwa Ibom	Akwa-Ibom	10.900000	8.9 - 12.9	49.900000	46.800000	19.800000
5	5	Anambra	3.46566411170	0.37505982119	Anambra	Anambra	8.700000	7.1 - 10.2	18.500000	18.500000	1.500000
6	6	Bauchi	12.90332236720	4.05567911900	Bauchi	Bauchi	2.000000	1.1 - 2.9	11.700000	7.400000	6.700000
7	7	Bayelsa	4.98969118107	0.82780830273	Bayelsa	Bayelsa	9.100000	7.2 - 11.1	22.200000	21.900000	3.600000
8	8	Benue	8.98155367092	2.56560812410	Benue	Benue	12.700000	10.8 - 14.6	47.900000	46.200000	8.100000
9	9	Borno	14.18090686040	5.48543720903	Borno	Borno	5.600000	4.1 - 7.2	20.000000	18.600000	6.100000
10	10	Cross River	7.46766588846	1.72861372597	Cross River	Cross-River	7.100000	5.5 - 8.8	29.300000	28.000000	6.700000

3c. You can sort the table by a specific attribute by double-clicking on the field name.

- The first time you double-click; the column will be sorted in ascending order.
- The second time you double-click, the column will be sorted in descending order.
- The third time you double-click, the column will be sorted in the original order in which the data were entered.

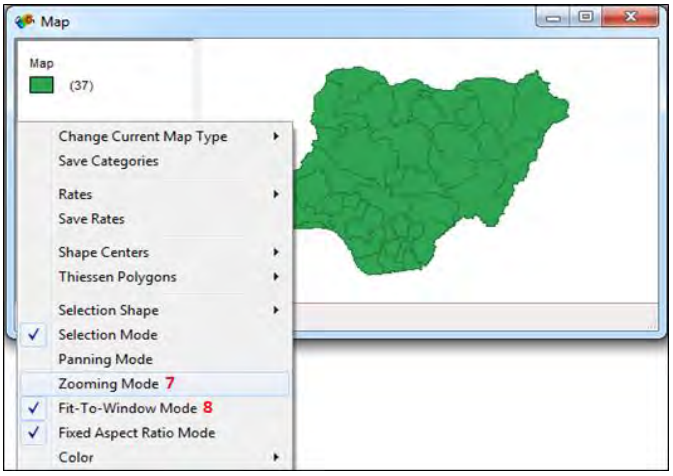
4. Change Map View

The map view can be changed in a few ways to improve visualization. One option is to increase the window size.

4a. Click and drag the bottom right corner of the window until it is at the desired size. Another way of changing the map view is to zoom.

4b. Right-Click on the map and select zooming mode. This changes the cursor to a zoom tool. To zoom in, click and drag a box around the area of interest.

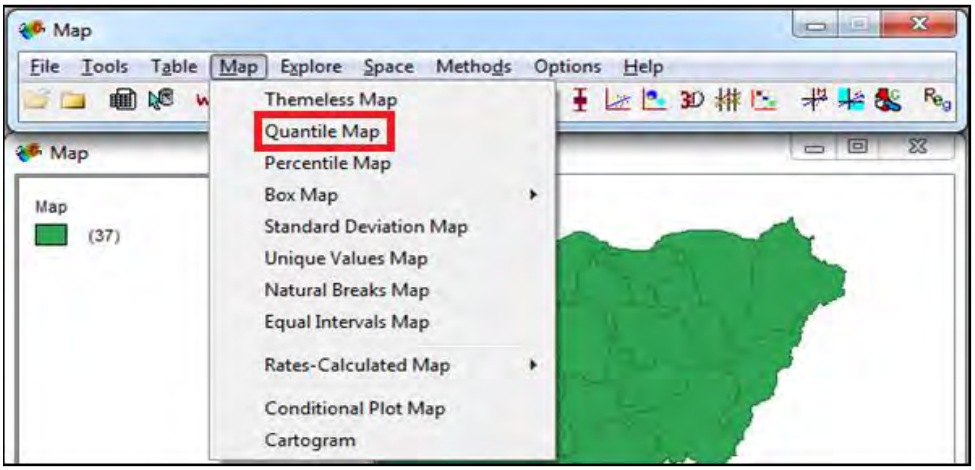
4c. To reset the view, right-click and select fit-to-window mode.



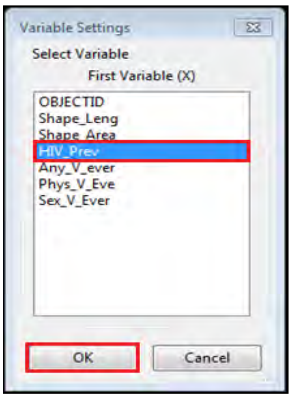
5. Create a Thematic Map

GeoDa allows the user to create univariate thematic maps very quickly. If you click on Map, a dropdown menu appears that offers a variety of thematic mapping methods and classification options. For this exercise, you will create a quantile map.

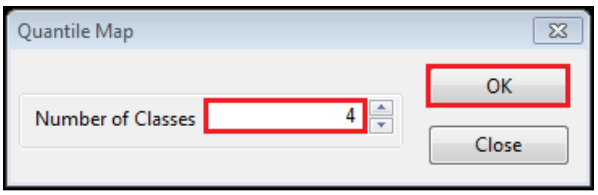
5a. Click Map > Quantile Map.



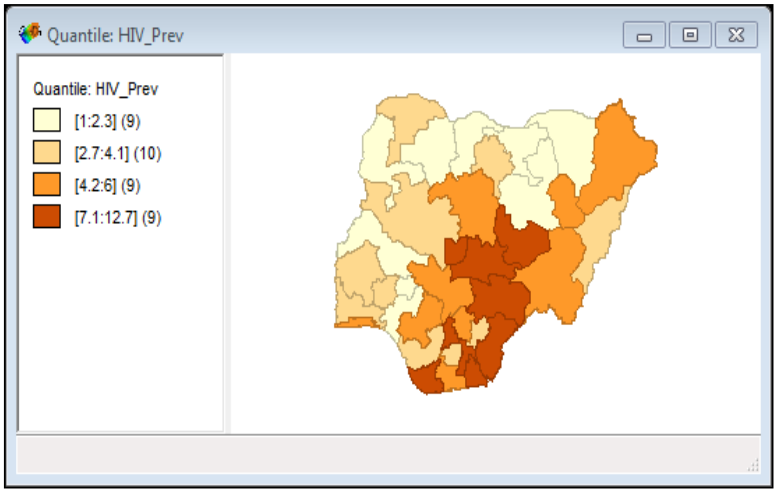
5b. When the variable settings window appears, select HIV_Prev and click OK. This variable contains HIV prevalence values for each province in Nigeria.



5c. Enter the number 4 for the number of classes. Click OK.



A quantile map showing the HIV prevalence rate by province will appear. The map should look like the image below. Note that the data has been divided into 4 classes of roughly equal size.



You may now exit GeoDa without saving and proceed to Exercise 2.

EXERCISE END

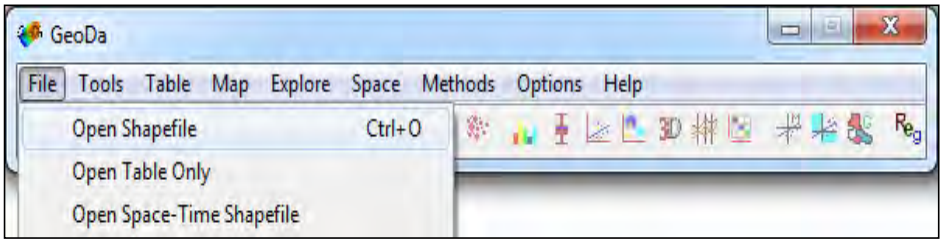
Exercise 2: Mapping Clusters

In this exercise, you will be exposed to mapping clusters using the Local Indicators of Spatial Association (LISA) analysis. LISA measures spatial autocorrelation, a measure of the degree to which features are clustered or dispersed, and can be used as a method for cluster analysis. Cluster analysis is a statistical method that groups objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups.

You will use GeoDa to conduct a LISA analysis. GeoDa is a free, open-source program that is designed to provide users with statistical methods for analyzing geographic data by combining maps and statistical graphs. NOTE: you should already have this software and the associated data files for this Exercise, installed from Exercise 1. For this exercise you will use Local Moran's I to map clusters of HIV prevalence in Nigeria. You will use Local Moran's I because it identifies clusters and produces a cluster map.

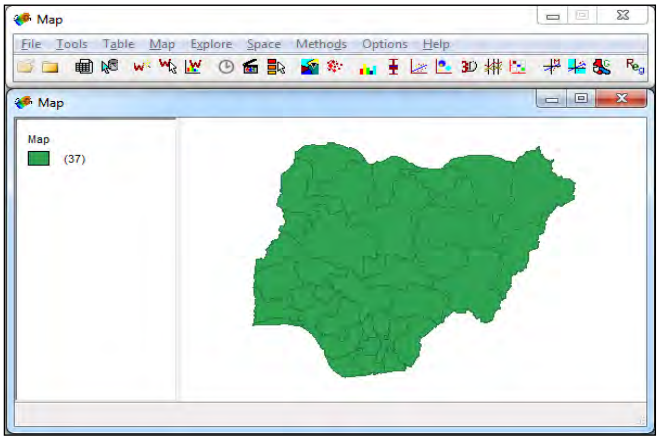
1. Open the project

1a. Launch the GeoDa application.
1b. Click File > Open Shapefile.



1c. Navigate to the data folder, select Violence_and_HIV, and click Open.

1d. A window containing a map of Nigeria will appear which should look similar to the image below.



1e. You should now explore the data to identify the variable that will be used for mapping clusters of HIV prevalence in Nigeria.

1f. On the navigation pane, click the open table icon.

1g. Take a look at all the attributes the table contains and identify the variable to map HIV prevalence.

1h. This table has a variable named HIV_Prev, which you will use for this exercise.

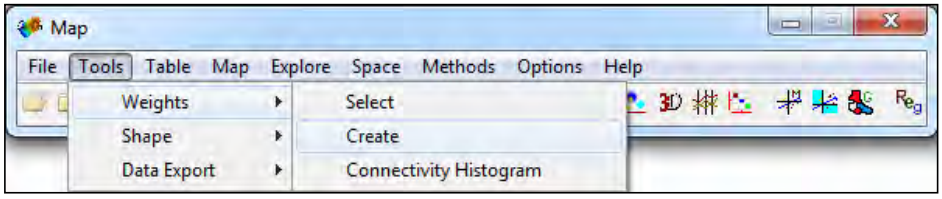
OBJECTID	ADM1NAME	Shape_Leng	Shape_Area	State_imp	State	HIV_Prev	CI	Any_V_ever	Phys_V_Eve	Sex_V_Ever
1	Abia	4.19898421317	0.38548504348	Abia	Abia	7.300000	5.6 - 9.0	36.100000	33.100000	10.500000
2	Abuja	3.42954727096	0.60426039987	Abuja	FCT	8.600000	7.0 - 10.2	8.300000	8.100000	1.400000
3	Adamawa	0.93622880800	2.81727148228	Adamawa	Adamawa	3.800000	2.6 - 5.1	22.300000	20.600000	7.400000
4	Akwa Ibom	3.55365829392	0.54982193969	Akwa Ibom	Akwa-Ibom	10.900000	8.9 - 12.9	49.900000	46.800000	19.800000
5	Anambra	3.46566411170	0.37505982119	Anambra	Anambra	8.700000	7.1 - 10.2	18.500000	18.500000	1.500000
6	Bauchi	2.90332236720	4.05567911900	Bauchi	Bauchi	2.000000	1.1 - 2.9	11.700000	7.400000	6.700000
7	Bayelsa	4.98969118107	0.82780830273	Bayelsa	Bayelsa	9.100000	7.2 - 11.1	22.200000	21.900000	3.600000
8	Benue	8.98155367092	2.56560812410	Benue	Benue	12.700000	10.8 - 14.6	47.900000	46.200000	8.100000
9	Borno	4.18090686040	5.48543720903	Borno	Borno	5.600000	4.1 - 7.2	20.000000	18.600000	6.100000
10	Cross River	7.46766588846	1.72861372597	Cross River	Cross-River	7.100000	5.5 - 8.8	29.300000	28.000000	6.700000

2. Create weight file

Spatial statistics like Moran's I require the creation of a spatial weight matrix. A spatial weight matrix reflects the intensity of the geographic relationship between observations in a neighborhood. Spatial weights can be created based on distance or contiguity. Objects that are points may only use distance-based weights while

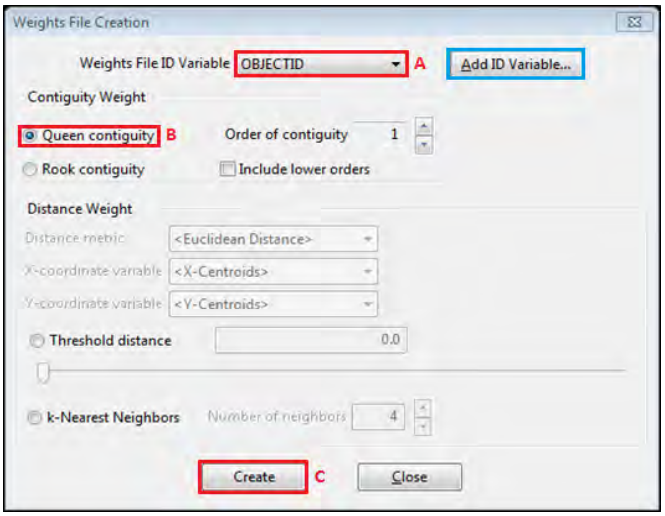
polygons can use both distance and contiguity weights. For this exercise, you will create a weight file that is defined by Queen Contiguity. The Queen Contiguity defines objects as neighbors if they share one or more common point, while Rook Contiguity only defines objects as neighbors if they share a common border. Because of these rules, Queen Contiguity will also have more neighbors.

2a. Click Tools > Weights > Create.



2b. Use the image on the next page as a reference to complete the following steps:

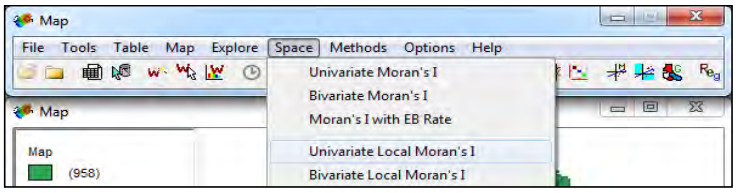
- Select ObjectID for the Weights File ID Variable. The Weights File ID Variable must be a variable that has a unique value for each object. Many shapefiles and datasets have an object ID field, or field that has a unique numerical value. If your dataset does not have a unique variable, GeoDa will create one for you. If needed, you just click on Add ID Variable.
- Select Queen Contiguity for the Contiguity Weight.
- Click Create, navigate to your Participant_Work folder, enter Violence_and_HIV for the file name, click Save, and then click OK.



3. LISA analysis

You are now ready to conduct the LISA analysis. You will select univariate Local Moran's I as the statistical test, because this test identifies clusters and produces a cluster map; and you will select univariate because you are only observing the spatial autocorrelation of a single variable, HIV prevalence. If you were observing the spatial autocorrelation of two variables, such as HIV prevalence and domestic violence, you would select bivariate local Moran's' I.

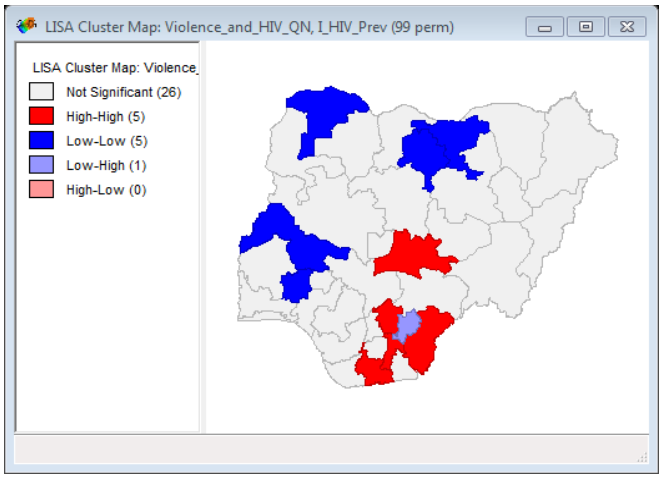
3a. Click Space > Univariate Local Moran's I.



3b. Select HIV_Prev as the variable.
3c. Click OK.
3d. Check the box for Cluster Map.
3e. Click OK.
3f. A cluster map should appear that has significant clusters. Your final product should look similar to the image below. This map has five (5) classes. The following is a brief description of each class, and what it tells you about the spatial clustering of HIV prevalence in Nigeria. Note that he high-high and low-low locations indicate positive spatial autocorrelation and are referred to as clusters, while the high-low and low-high locations indicate negative spatial autocorrelation and are referred to as outliers.

- **Not Significant**—no significant relationship between HIV prevalence rates with neighboring states.
- **High-High**—identifies states that have high HIV prevalence rates whose neighbors also have high HIV prevalence rates.
- **Low-Low**—identifies states that have low HIV prevalence rates whose neighbors also have low HIV prevalence rates.
- **Low-High**—identifies states that have low HIV prevalence rates that are neighbored by states with high HIV prevalence rates.
- **High-Low**—identifies states that have high HIV prevalence rates that are neighbored by states with low HIV prevalence rates.

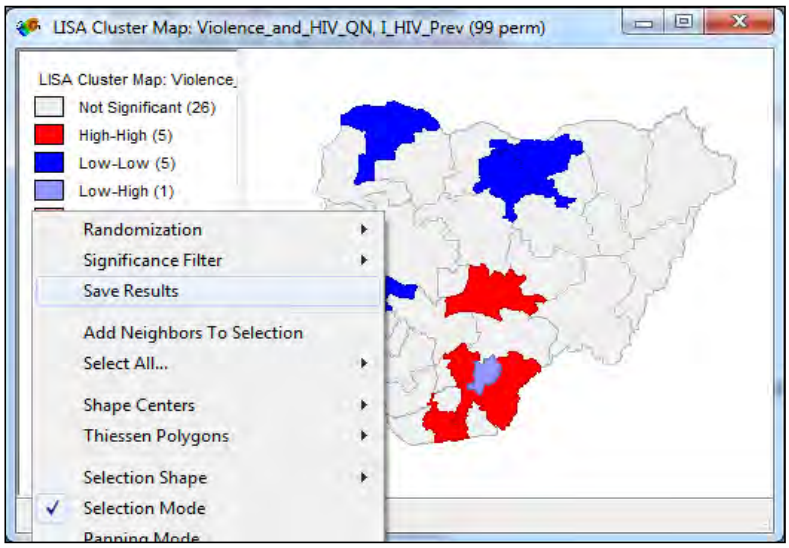
The map below locates clusters of high HIV prevalence rates in the south of Nigeria and clusters of low HIV prevalence rates in the north of Nigeria.



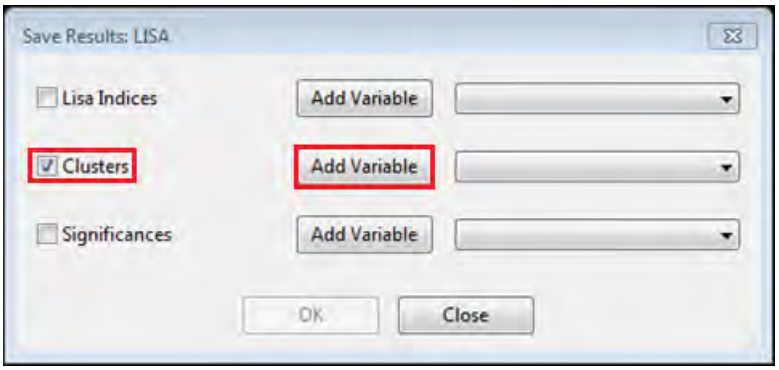
4. Save your results

The next step is very important: to save your results to a shapefile. When you save your results you have the option to save three (3) values: LISA indices (Local Moran statistics), clusters (five values following the order of the legend categories from 0 to 4), and significances (P-values).

4a. Right-Click on your cluster map and select the Save Results button.

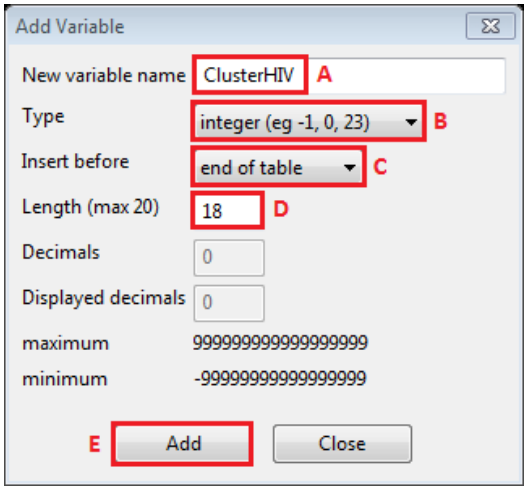


4b. Check the box next to clusters and then click Add Variable.



4c. Use the image below as a reference to complete the following steps:

- Enter ClusterHIV as the new variable name.
- Select integer for type.
- Select end of table for insert before.
- Accept the automatic value of 18 for length.
- Click Add.



4d. Click OK.

4e. Exit GeoDa without saving. This completes the exercises associated with the GeoDa software.

EXERCISE END

Exercise 3: Delineating Catchment Areas

Suppose your supervisor wants to know the catchment area for each of the facilities that are affected by your program's interventions (e.g., the area that the health facility's users come from). If you have a facility with a very large catchment area, you might want to consider building an additional health facility.

There are several different ways to calculate this, depending on how much you know about each health facility and how much time and budget are available. You could simply say that the catchment area for each facility is a circle with a radius of 5km, centered on each facility, a method completed with the buffer tool. However, if you have two facilities closer than 5km to each other, you have overlap, which may be accurate, but does not give you an exclusive zone for each health facility.

If the health facilities are spread out, you will have areas that are not in the catchment area for any health facility. In fact, if you cannot overlap the circles, there will always be areas that are not in any catchment area. So, while it is easy to make circles (and you do not even need a computer to make them), the results will be poor, at best. However, you could create Voronoi polygons for each health facility. This could reveal facilities where you might want to run a survey to better define the catchment areas for those facilities in particular.

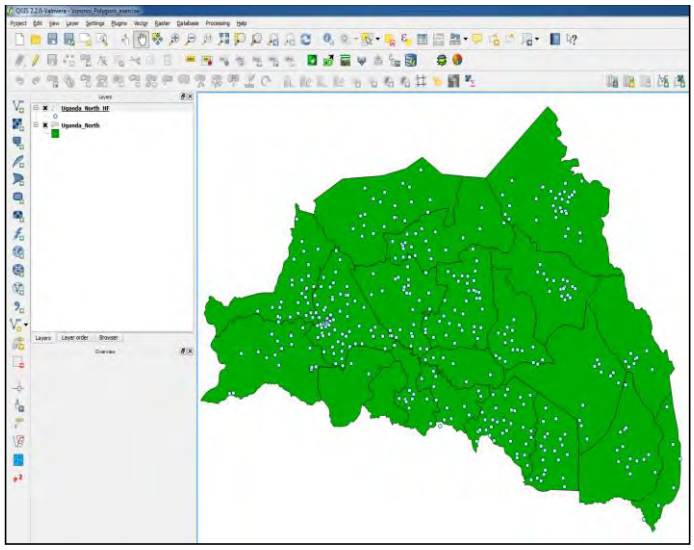
In this exercise, you will use QGIS. You will be exposed to how to delineate catchment areas using Voronoi (Thiessen) polygons based on health facility locations in northern Uganda. Voronoi polygons are generated from a set of geographically referenced points. In this situation, any area inside the polygon is closer to that health facility than any other health facility.

NOTE: In order to complete this exercise you need to have QGIS installed on your computer, and you need to have access to the associated Exercise Data files, mentioned at the beginning of Appendix 7, as well.

1. Open project file

1a. Launch the QGIS Desktop application.
1b. Click Project > Open.
1c. Navigate to your Exercise Data folder and select Voronoi_Polygons_Exercise.qgs.

1d. Click Open. You should see a map, similar to the one below, which displays the district boundaries and location of health facilities in northern Uganda.

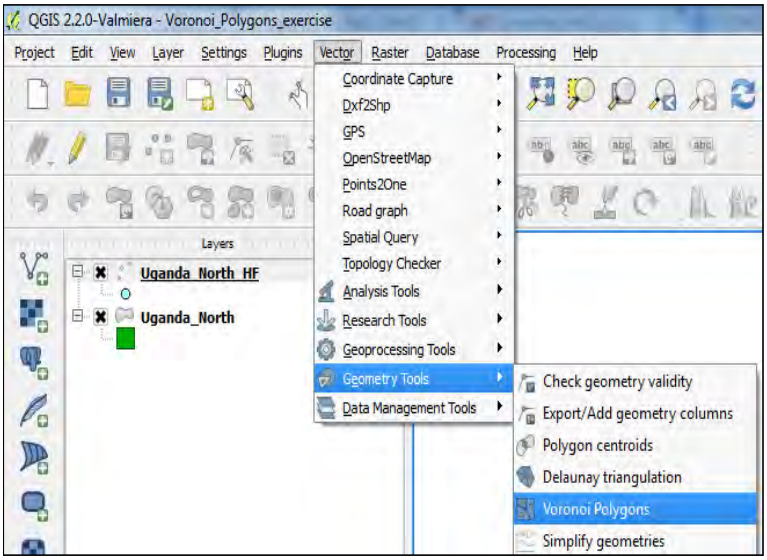


1e. Save the project as a new file by clicking Project > Save As…
1f. Navigate to your Exercise Data folder, enter Voronoi_Polygons_Appendix_6 for the name.
1g. Click Save.

2. Voronoi Polygons

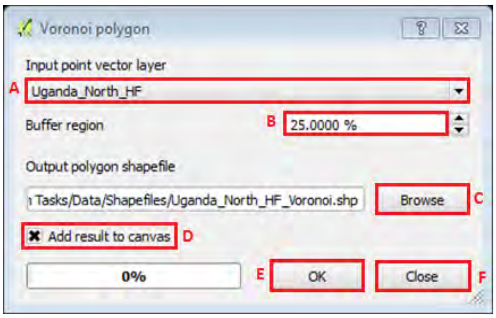
You will now create the Voronoi polygons. This will show which health facility is located within the shortest distance for locations in northern Uganda.

2a. Click Vector > Geometry > Voronoi Polygons.

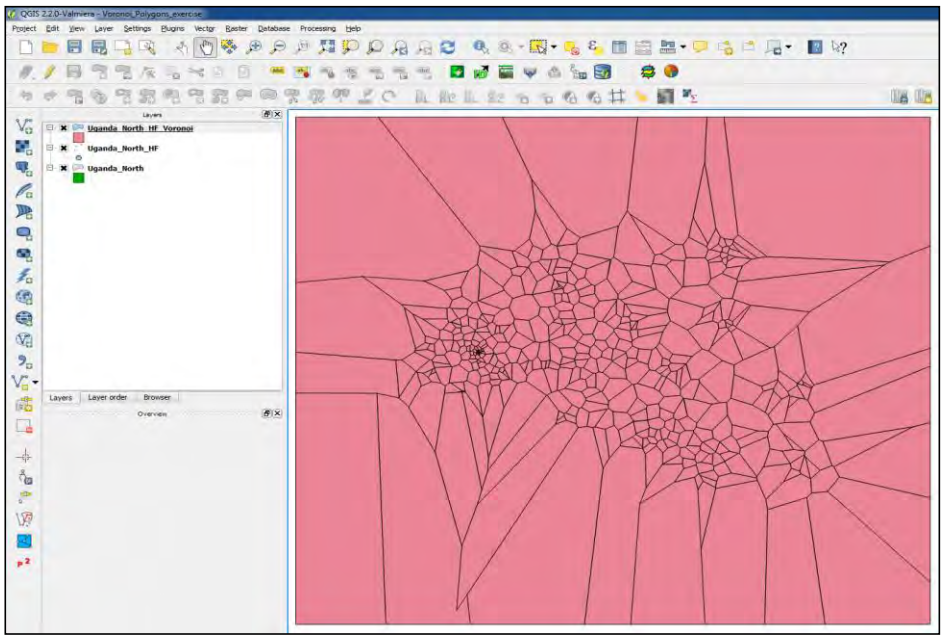


2b. The Voronoi polygons dialog box will appear. Use the image below as a reference to complete the following steps:

- Select Uganda_North_HF as the input point vector layer.
- Set the buffer region to 25%. This will ensure that the polygons cover the entire region. The areas the extend beyond the boundary of northern Uganda will be removed later in the exercise by using the Clip Tool.
- Click Browse under output polygon shapefile. Navigate to your Exercise Data folder and name the file Uganda_North_HF_Voronoi. Then click Save. Check the box to add results to canvas.
- Click OK.
- Click Close.



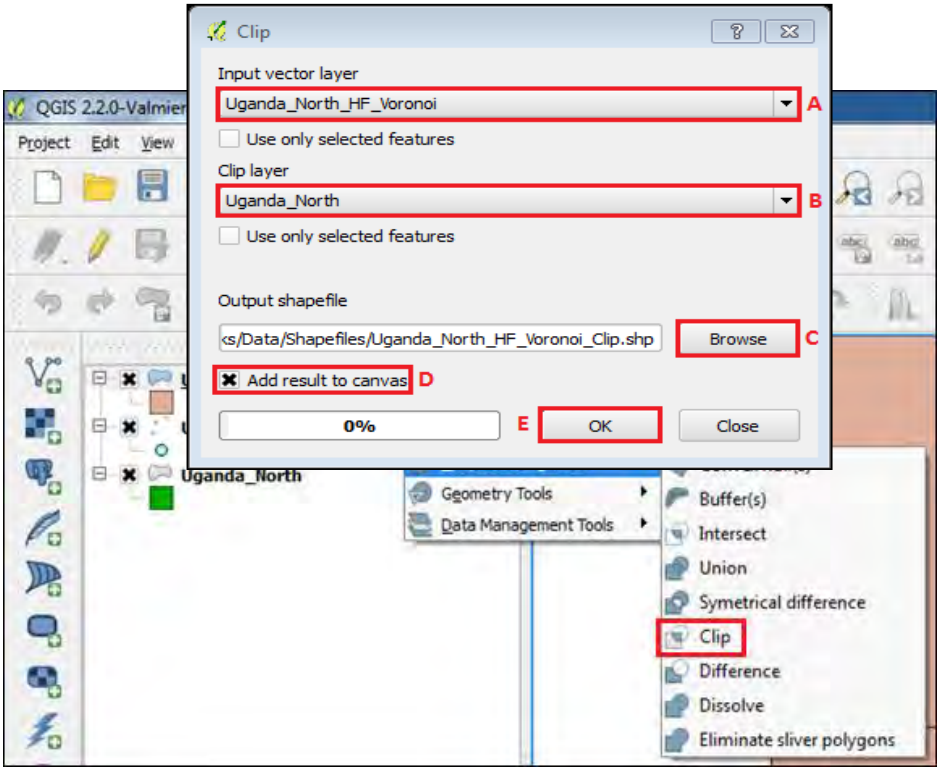
2c. Your results should look similar to the image below.



3. Clipping

Currently the Voronoi polygons extend beyond the northern Uganda boundaries. To focus the Voronoi polygons to the area of interest you will use the clip tool. Note that the clip tool allows users to cut out a piece of one feature class using one or more of the features in another feature class as a cookie cutter. This process will result in removing the areas of the Voronoi polygons that are not within northern Uganda.

3a. Click Vector > Geoprocessing Tools > Clip.



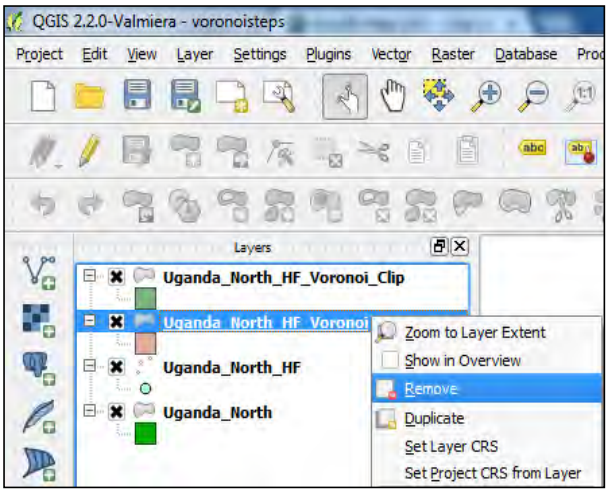
3b. The Clip dialog box will appear. Use the image on the next page as a reference to complete the following steps:

- Select Uganda_North_HF_Voronoi as the input vector layer.
- Select Uganda_North as the Clip layer.
- Click Browse and then navigate to your Exercise Data folder. Enter Uganda_North_HF_Voronoi_Clip as the file name and click Save.
- Check the box to add results to the canvas.
- Click OK

3c. Click Close.

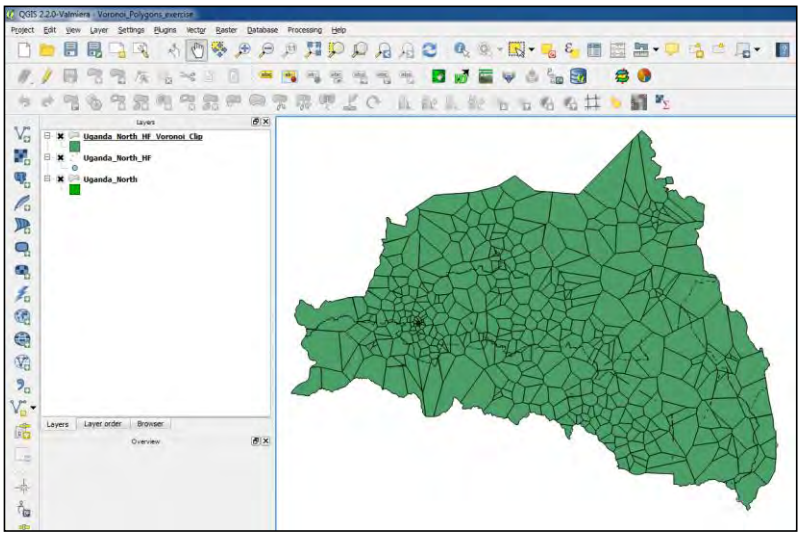
3d. Right-click on the Uganda_North_HF_Voronoi shapefile in the Layer Panel.

3e. Select Remove.



3f. Click OK.

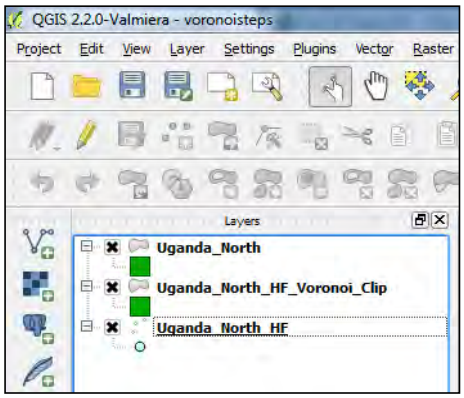
3g. Your results should look similar to the image below.



4. Visualization

You now have a Voronoi polygon shapefile that is clipped to the area of northern Uganda. However, it is obstructing the view of the other data layers. You need to reorder the layers in the Layer Panel and adjust the style of layers to visualize the data properly.

4a. Click and drag the Uganda_North shapefile in the Layer Panel to the top of the list.



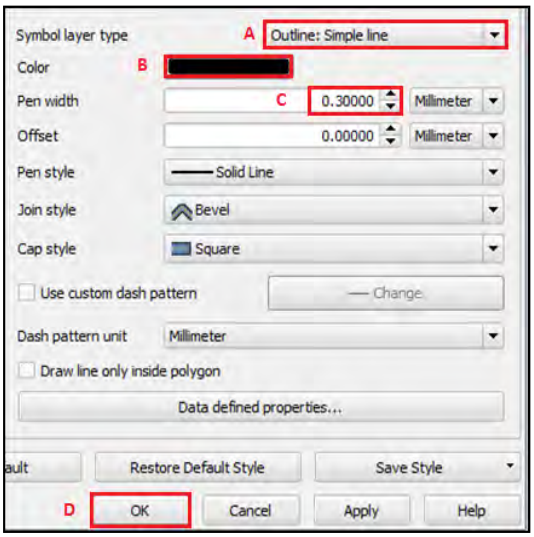
4b. Double-click on the Uganda_North shapefile in the Layer Panel to open the Layer Properties window.

4c. Click the Style tab.

4d. Select Simple fill under Symbol Layers.

4e. Use the image below as reference to complete the following steps:

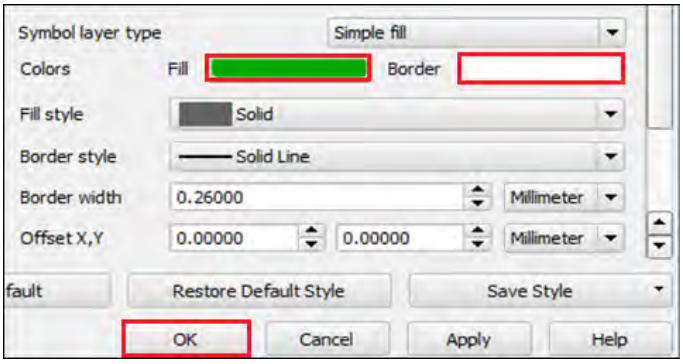
- Change the Symbol layer type to Outline: Simple line.
- Set Color to black.
- Adjust Pen width to 0.3
- Click OK.



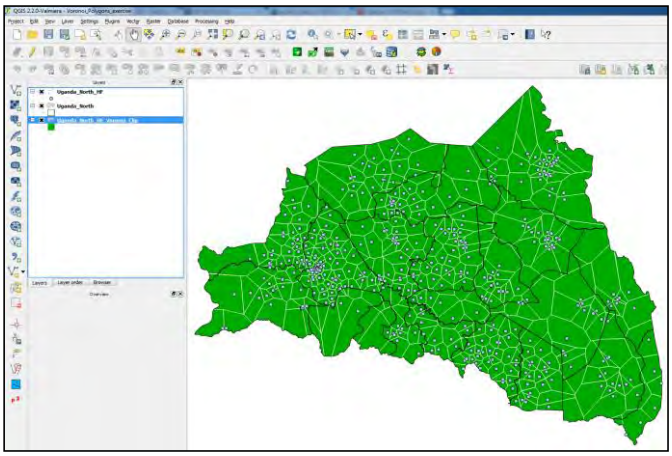
4f. Double-Click on the Uganda_North_Voronoi_Clip shapefile in the Layer Panel to open the layer properties window.

4g. Click the Style tab.

4h. Select Simple fill under Symbol Layers.
4i. You may now select any colors of your choice for fill and border colors. The example in the image below uses green fill and white borders.
4j. Click OK.



4k. Click and drag the Uganda_North_HF shapefile to the top of the Layer Panel. You have now delineated catchment area of health facility locations in northern Uganda using Voronoi polygons. Any location that is within a Voronoi polygon is closer to that health facility than any other health facility. You final product should look similar to the image below.



EXERCISE END

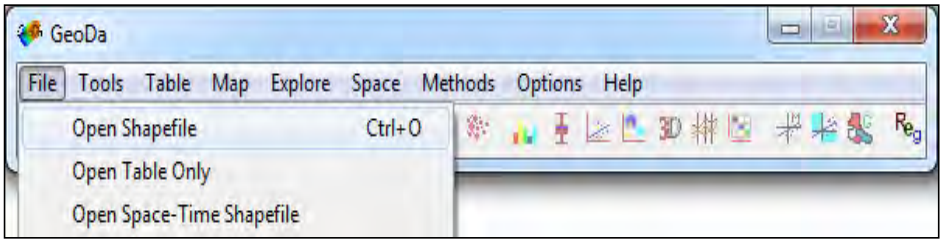
Exercise 4: Hot Spot Analysis

In this exercise you will be exposed to mapping hot spots using Getis-Ord Gi*, or local G statistic. The local G statistic measures spatial autocorrelation, and tells you where features with either high or low values cluster spatially. The clusters composed of features with high values are called "hot spots."

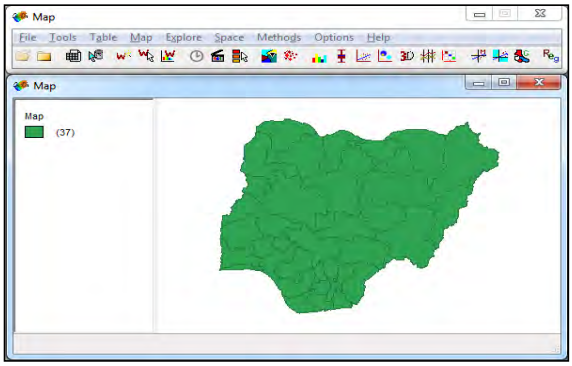
You will use GeoDa to conduct a hot spot analysis. GeoDa is a free, open source program that is designed to provide users with statistical methods to analyze geographic data by combining maps and statistical graphs. This is the software you have already used for Exercises 1 and 2 in this Appendix. For this exercise, you will use local G statistic to map clusters of HIV prevalence in Nigeria. You will use local G statistic because it identifies hot and cold spots and produces a map.

1. Open the project

1a. Launch the GeoDa application.
1b. Click File > Open Shapefile.



1c. Navigate to your training folder, select Violence_and_HIV, and click Open.
1d. A window containing a map of Nigeria will appear which should look similar to the image below.



1e. You should now explore the data to identify the variable that will be used for mapping hot spots of HIV prevalence in Nigeria.

1f. On the navigation pane, click the open table icon.

1g. Take a look at all the attributes the table contains and identify the variable to map HIV prevalence.

OBJECTID	ADM1NAME	Shape_Leng	Shape_Area	State_imp	State	HIV_Prev	CI	Any_V_ever	Phys_V_Eve	Sex_V_Ever
1	Abia	4.19898421317	0.38548504348	Abia	Abia	7.300000	5.6 - 9.0	36.100000	33.100000	10.500000
2	Abuja	3.42954727096	0.60426039987	Abuja	FCT	8.600000	7.0 - 10.2	8.300000	8.100000	1.400000
3	Adamawa	0.93622880800	2.81727148228	Adamawa	Adamawa	3.800000	2.6 - 5.1	22.300000	20.600000	7.400000
4	Akwa Ibom	3.55365829392	0.54982193969	Akwa Ibom	Akwa-Ibom	10.900000	8.9 - 12.9	49.900000	46.800000	19.800000
5	Anambra	3.46566411170	0.37505982119	Anambra	Anambra	8.700000	7.1 - 10.2	18.500000	18.500000	1.500000
6	Bauchi	2.90332236720	4.05567911900	Bauchi	Bauchi	2.000000	1.1 - 2.9	11.700000	7.400000	6.700000
7	Bayelsa	4.98969118107	0.82780830273	Bayelsa	Bayelsa	9.100000	7.2 - 11.1	22.200000	21.900000	3.600000
8	Benue	8.98155367092	2.56560812410	Benue	Benue	12.700000	10.8 - 14.6	47.900000	46.200000	8.100000
9	Borno	4.18090686040	5.48543720903	Borno	Borno	5.600000	4.1 - 7.2	20.000000	18.600000	6.100000
10	Cross River	7.46766588846	1.72861372597	Cross River	Cross-River	7.100000	5.5 - 8.8	29.300000	28.000000	6.700000

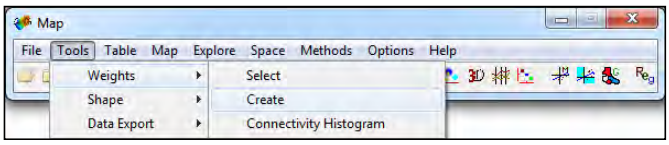
1h. This table has a variable named HIV_Prev, which you will use for this exercise.

2. Create a weight file

Spatial statistics like local G statistic require the creation of a spatial weight matrix. A spatial weight matrix reflects the intensity of the geographic relationship between observations in a neighborhood. Spatial weights can be created based on distance or contiguity. Objects that are points may only use distance-based weights, while polygons can use both distance and contiguity weights.

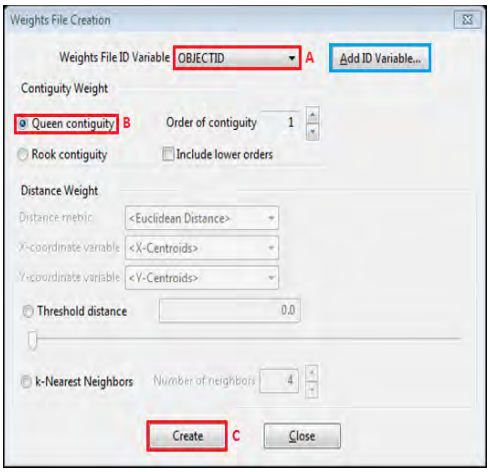
For this exercise, you will create a weight file that is defined by Queen Contiguity. You may remember from exercise 2 that the Queen Contiguity defines objects as neighbors if they share one or more common point, while Rook Contiguity only defines objects as neighbors if they share a common border. Because of these rules, Queen Contiguity will always have more neighbors.

2a. Click Tools > Weights > Create.



2b. Use the image on the next page as a reference to complete the following steps:

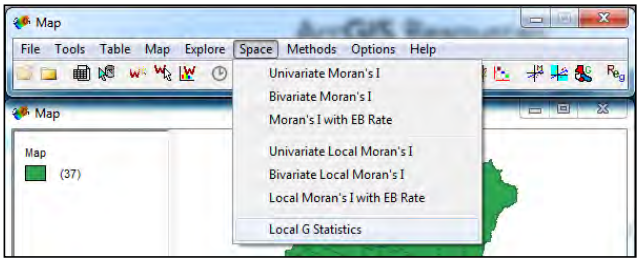
- Select ObjectID for the Weights File ID Variable. Note that the Weights File ID Variable must be a variable that has a unique value for each object. Many shapefiles and datasets have an object ID field, or field that has a unique numerical value. If your dataset does not have a unique variable, GeoDa will create one for you. If needed, you just click on Add ID Variable.
- Select Queen Contiguity for the Contiguity Weight.
- Click Create, navigate to your Participant_Work folder, enter Violence_and_HIV for the file name, click Save, and then click OK.



3. Hot Spot Analysis

You are now ready to conduct the hot spot analysis.

3a. Click Space > Local G Statistics.



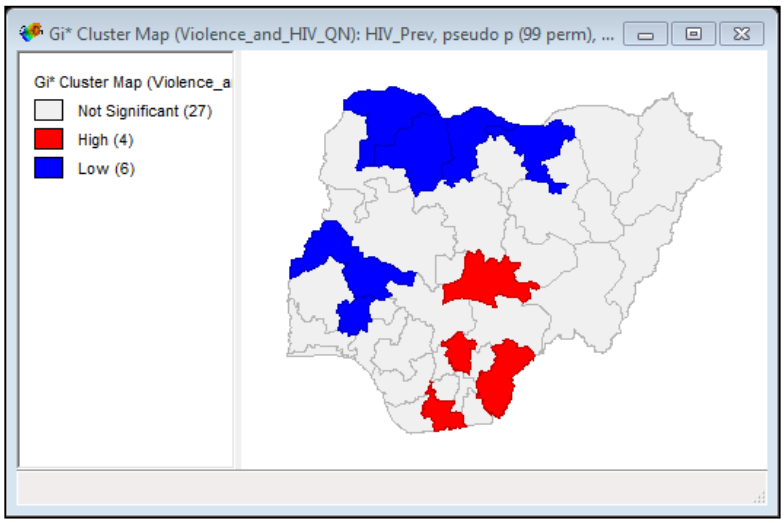
3b. Select HIV_Prev as the variable.

3c. Click OK.

3d. Check the box for Gi* cluster map, pseudo p-val.

3e. Click OK.

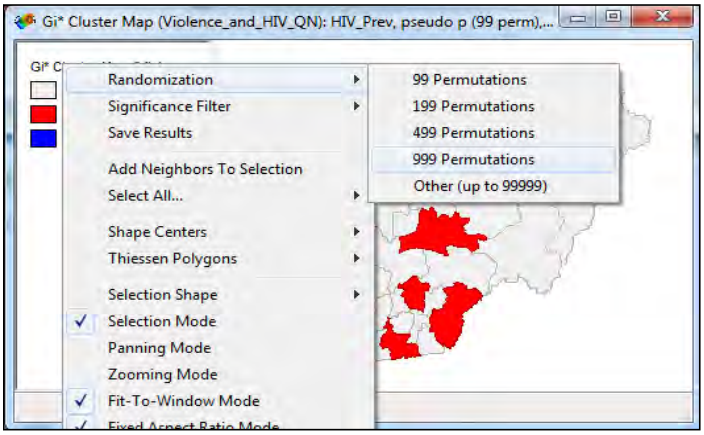
3f. A map should appear that identifies clusters of areas with high HIV prevalence rates and areas with low HIV prevalence rates. Your final product should look similar to the image below. This map has two classes, high and low. The map below locates hot spots of high HIV prevalence rates in the south of Nigeria and cold spots of low HIV prevalence rates in the north of Nigeria.



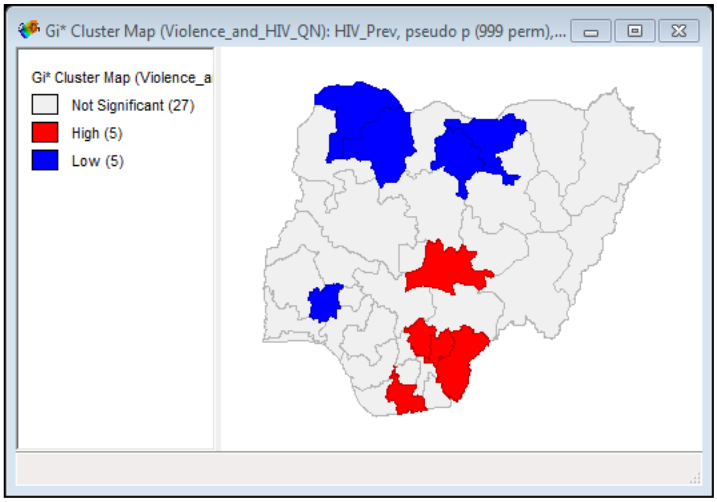
4. Significance assessment

Because hot spot analysis is done through a statistical test, it is important to assess the significance of the test. This can be done by increasing the permutations. By default, the test is automatically done with 99 permutations. Increasing the number of permutations will increase the statistical significance of the results.

4a. Right-Click on your hot spot map and select Randomization and then 999 Permutations.



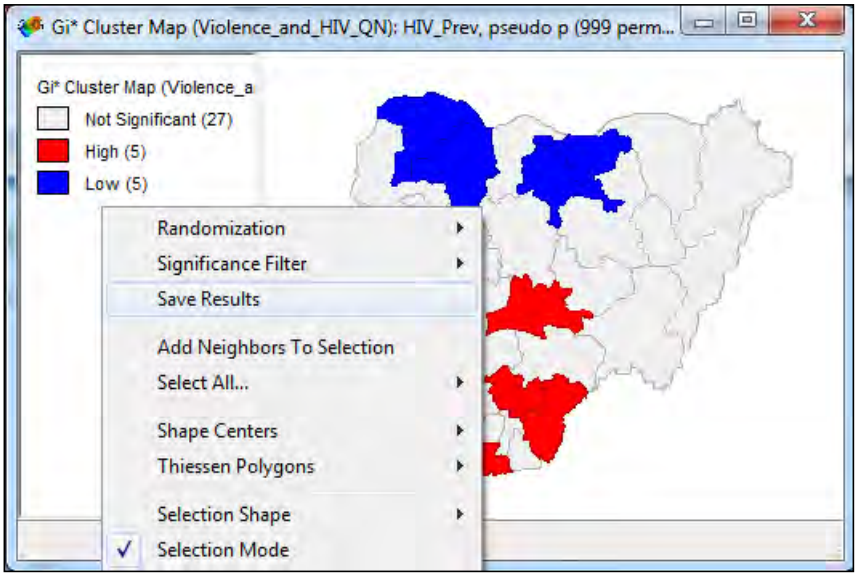
This tells the GIS to run the statistical test 999 times which eliminates bias by producing random results. You may notice a slight change in the clusters on your map.



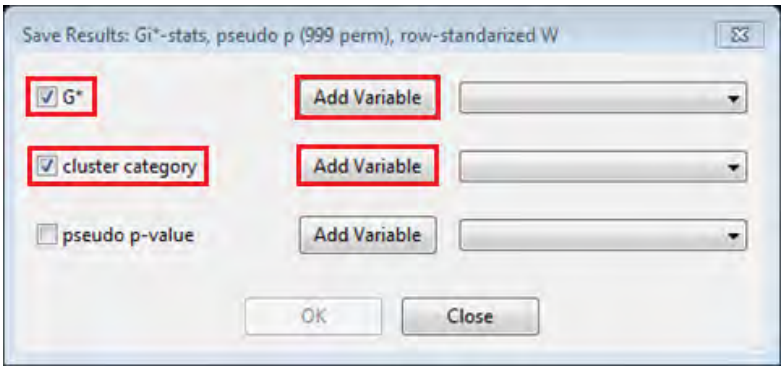
5. Save your results

The next step is very important: to save your results to a shapefile. When you save your results, you have the option to save 3 values: G*, cluster category (hot or cold), and pseudo P-value.

5a. Right-Click on your hot spot maps and select Save Results.

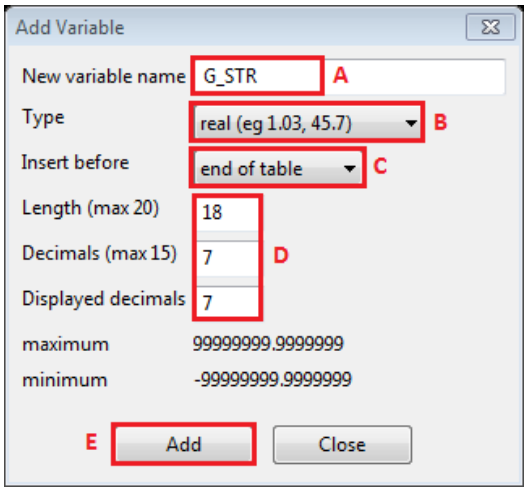


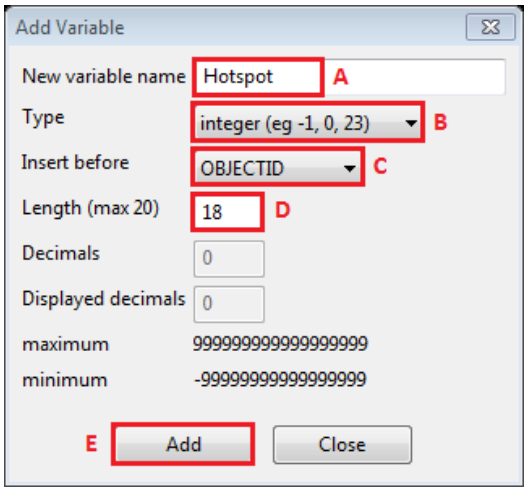
5b. Check the box next to G* and cluster category, then click Add Variable.



5c. Use the image below as a reference to complete the following steps:

- Enter G_STR as the new variable name.
- Select real for type.
- Select end of table for insert before.
- Accept the automatic values for length and decimal settings.
- Click Add.
- Enter Hotspot as the new variable name.
- Select integer for type.
- Select end of table for insert before.
- Accept the automatic values for length and decimal settings.
- Click Add.





5d. Click OK.

5e. Exit GeoDa without saving.

EXERCISE END

MEASURE Evaluation

University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, NC 27517

http://www.measureevaluation.org