2014

Geographic Information System Techniques for Monitoring and Evaluation of HIV/AIDS and Related Programs

John Spencer, Jay Stewart, Becky Wilkes

MS-15-106

2014

Geographic Information System Techniques for Monitoring and Evaluation of HIV/AIDS and Related Programs

John Spencer, Jay Stewart, Becky Wilkes

MS-15-106

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) through Cooperative Agreement GHA-A-00-08-00003-00 and is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with ICF International, John Snow, Inc., Management Sciences for Health, Palladium, and Tulane University. The views expressed in this publication do not necessarily reflect the views of USAID or the United States government. MS-12-106

USAID FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

This course presents a practical guide for using a geographic information system (GIS) to integrate, visualize, and analyze geographically-referenced data extracted from the Demographic and Health Surveys (DHS) and other key data sets to facilitate monitoring and evaluation (M&E) of HIV/AIDS and related health programs. To make the course accessible to learners without commercial GIS software licenses, the course focuses on the use of free and open source software (FOSS).

The course includes a number of practical exercises which will allow you to apply what you have learned, using the FOSS - QGIS. These exercises are estimated to add an additional three hours to completing the course.

Objective

Upon completion of the course, learners will be able to:

- Define a GIS and articulate the value of using a GIS to integrate, visualize, and analyze program data.
- Differentiate the primary uses of several free and open source GIS software options.
- Explain the fundamentals of geographic data, including the shapefile format.
- List the principal geo-referenced sources for HIV/AIDS data and explain how their geographic structure can be leveraged for M&E of HIV/AIDS and related programs.
- Recognize the types of M&E questions that can be answered using the highlighted data sets and GIS software packages.
- Use the free and open source GIS software package, QGIS, to perform essential GIS tasks needed to respond to typical M&E questions.

Target Audience

Although the course is designed primarily for M&E professionals working for or with HIV/AIDS programs in countries receiving Global Health Initiative/PEPFAR support, analysts and decision makers from other public health programs will benefit from the techniques and data sources explained.

Time

2 hours for main course, plus an additional **3 hours** for the (optional) practical exercises.

Instructions for PDF version

Glossary terms are highlighted **like this**. Terms are defined in Appendix 1. Appendix 2 contains the instructions for the Practical Exercises. Quiz questions and answers are included at the end of each Session. Original links are provided in text format where appropriate.

Acknowledgements:

Course Authors:

- John Spencer, UNC/MEASURE Evaluation
- Jay Stewart, UNC/MEASURE Evaluation
- Becky Wilkes, UNC/MEASURE Evaluation

Course Managers:

- Lisa Mwaikambo, JHU·CCP
- Jarret Cassaniti, JHU·CCP

Published/Updated Online

Monday, July 28, 2014

Online access:

http://www.globalhealthlearning.org/course/gis-techniques-m-e-hiv-aids-and-related-programs

Credits

The authors are very grateful to these reviewers and editors who provided their insights and suggestions:

- Phil Bardsley, UNC
- Clara Burgert, ICFI/DHS Program
- Melissa Dunn, UNC/MEASURE Evaluation
- Sivan Goobich, UNC/MEASURE Evaluation
- Nathan Heard, Department of State
- Mark Janko, UNC
- Nandini Jayarajan, JHU/CCP
- Imelda Moise, JSI
- David Perrin, UNC
- Jim Shelton, USAID
- Heather Socha, UNC
- Sam Speller, UNC
- Kristen Wares, USAID

Programming and design support for the GPS simulation Practical Exercise in Session 1 was provided by Weejee Learning, LLC. Team members included: Tracy Bissette, Britta Hansen, Ian Huckabee, Donnie Kohler (Lead Developer), Liz Lusardi, Michelle Thacher, and Stacy Whitenight.

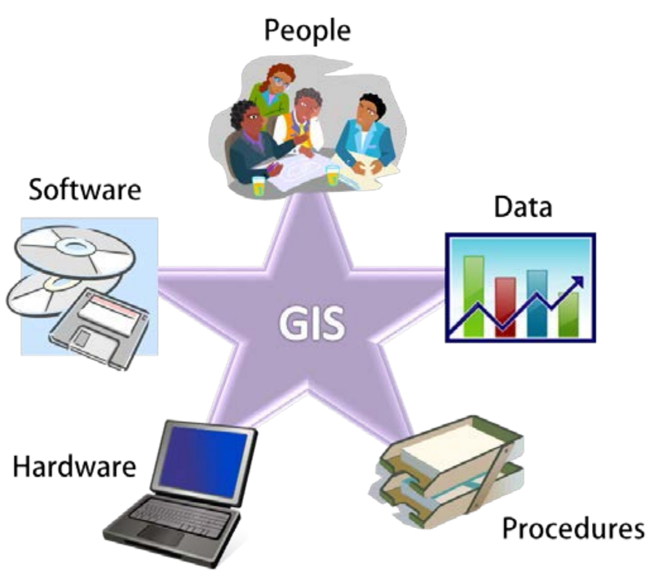
Session 1: The GIS Advantage

After completing this session, learners will be able to:

- Define the meaning of GIS and identify some common uses of a GIS.
- Identify the five components of a GIS: hardware, software, geographically-referenced data, procedures, and people (both users and technical staff).
- Identify the six primary functions of a GIS: collect, store, query, analyze, display, and output geographic data.
- Describe two examples of how a GIS can be used for monitoring and evaluating an HIV/AIDS program.
- Describe a situation in which GIS data can be linked to facilitate program integration.

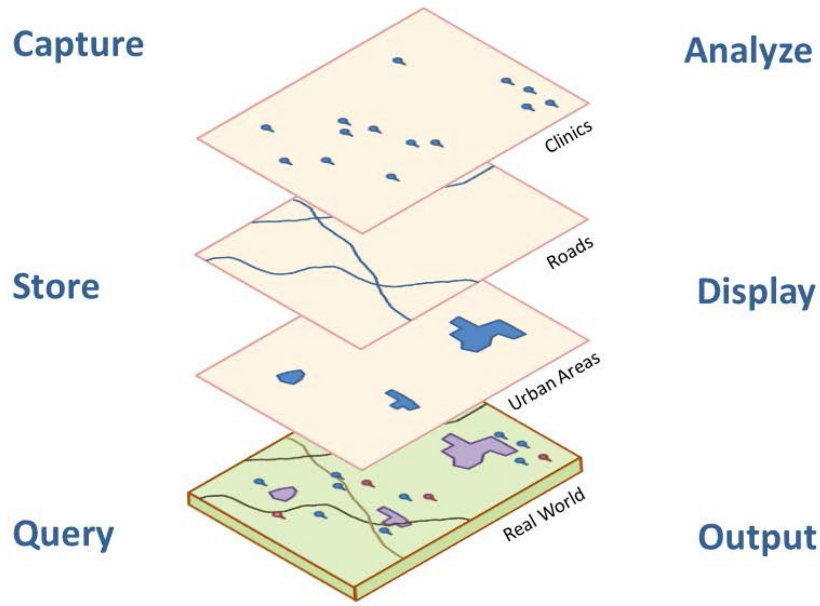
What is a GIS?

For the sake of convenience, a **geographic information system (GIS)** can be thought of as **a database linked to a map**. This is a simplistic definition; in reality a GIS has **five main components**: hardware, software, geographically-referenced data, procedures, and people (both users and personnel).



Six Primary Functions

A GIS can also be defined in terms of its **six primary functions** with respect to geographically-referenced data:

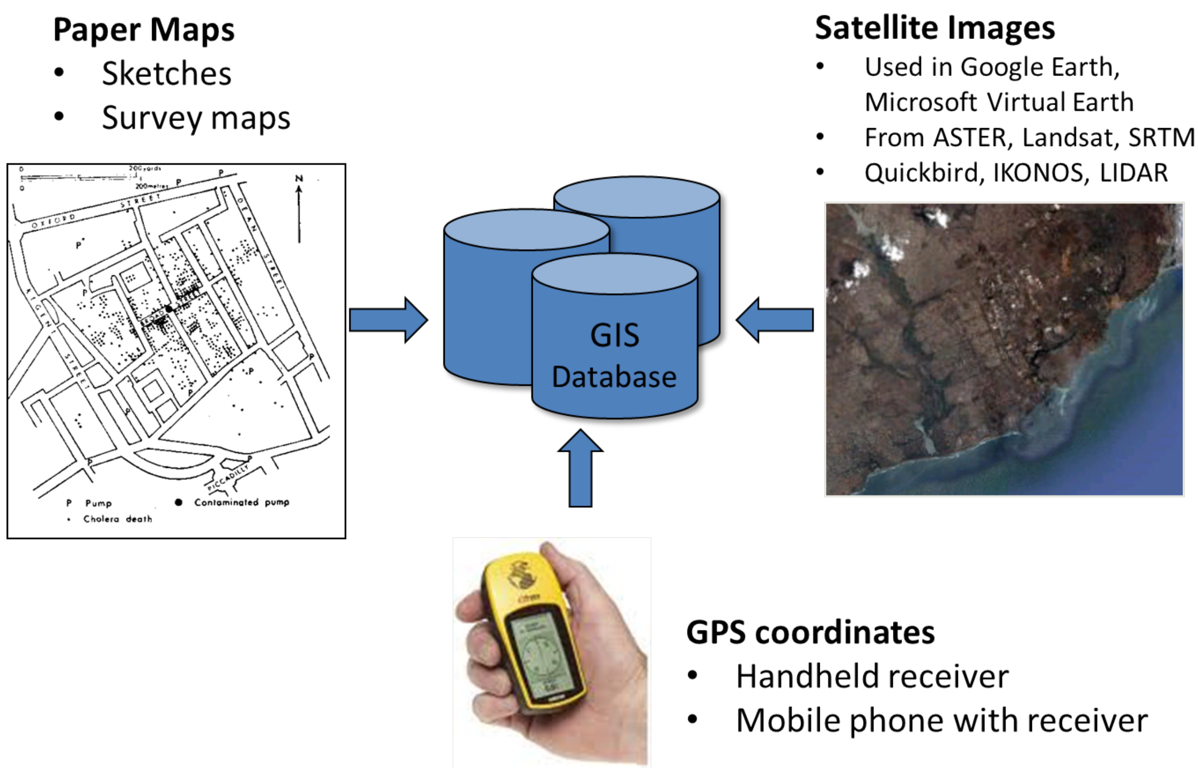


This data can be captured in a series of spatial data files. These files can be overlaid with each other and with base data and imagery to enable further analysis and visualization.

The following sections will provide more details related to each function.

Data Capture

Data can be captured for use in a GIS from a variety of sources, including paper maps, satellite imagery, and GPS (global positioning system) receivers.

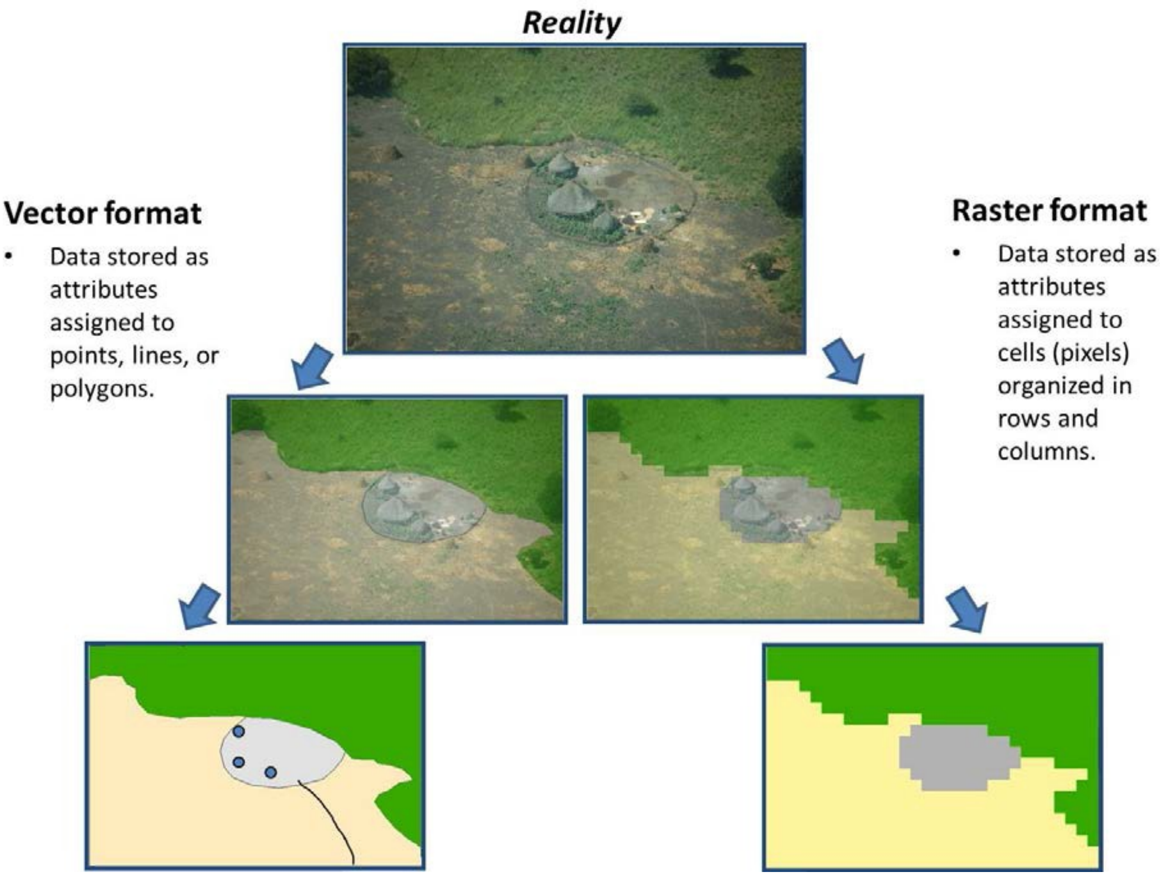


Data Representation

Spatial data can be stored as a **vector-based data file** or a **raster-based data file**. These data storage types represent two distinct ways of abstracting reality for use in a GIS:

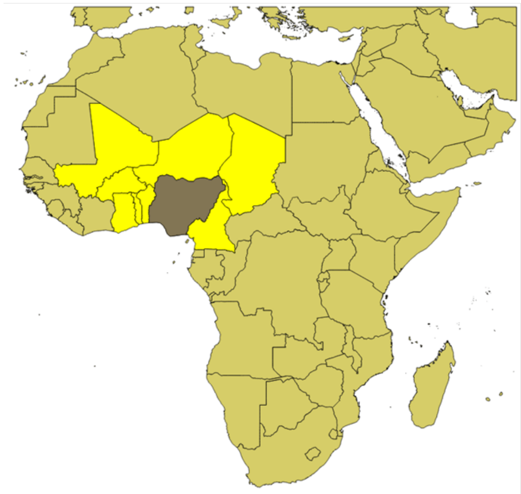
- Vector-based data is a memory-efficient way of storing and representing data, and is useful for measuring distances or areas and for drawing buffers or doing network analysis.
- Raster-based data (the type which comes directly from satellite scans) generally requires more memory, but can be useful for performing advanced geographic which might require zonal statistics or kernel density estimation.

GIS software generally provides two different sets of tools, which require two different user skillsets, for these two types of data analysis. Most of the examples provided in this course concern vector data analysis.



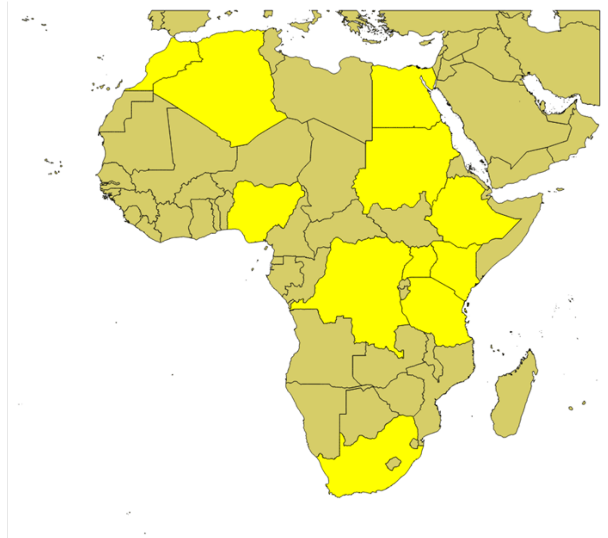
Data Query

Spatial data can be queried either by its location or by its attributes.



Identify geographic entities based on location.

Example: Countries within 250 Kilometers of Nigeria.

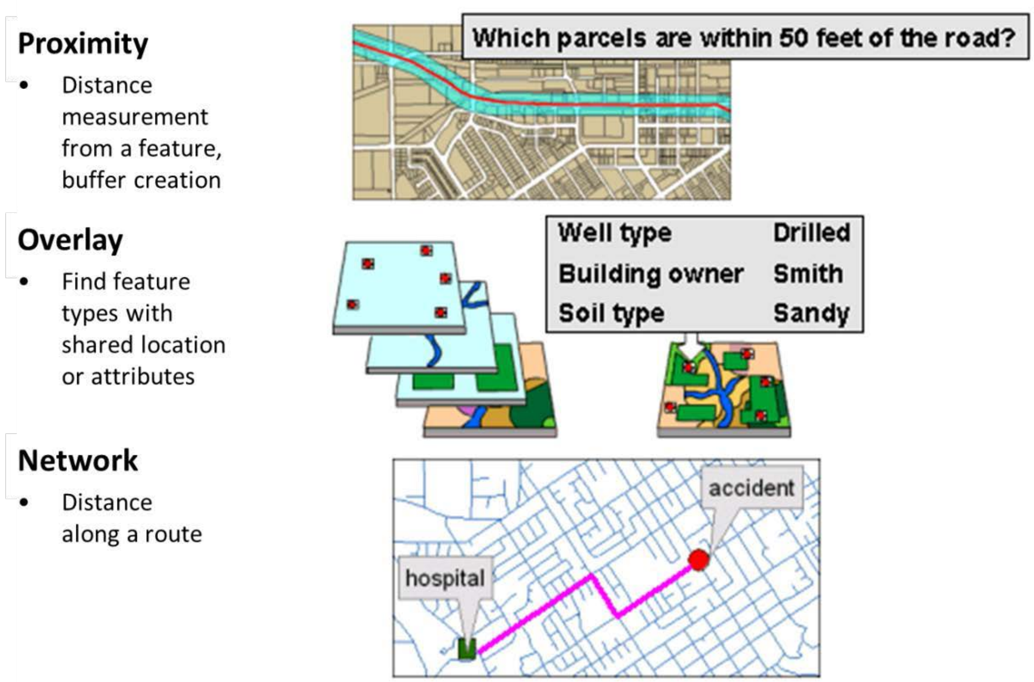


Identify geographic entities based on attributes.

Example: African countries with a population greater than 30,000,000.

Data Analysis

Some examples of spatial data analysis which are possible with a GIS:

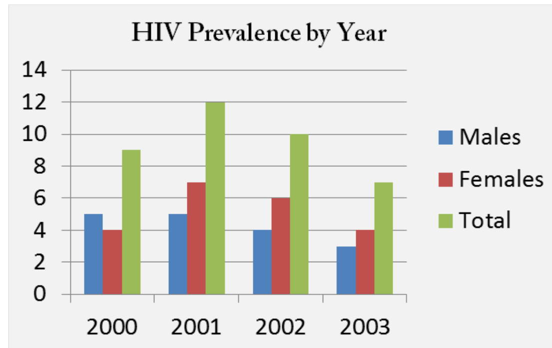


Data Display

Data from a GIS can be displayed in map, graphic, or table form, or often a combination of all three.

Maps

Graphs

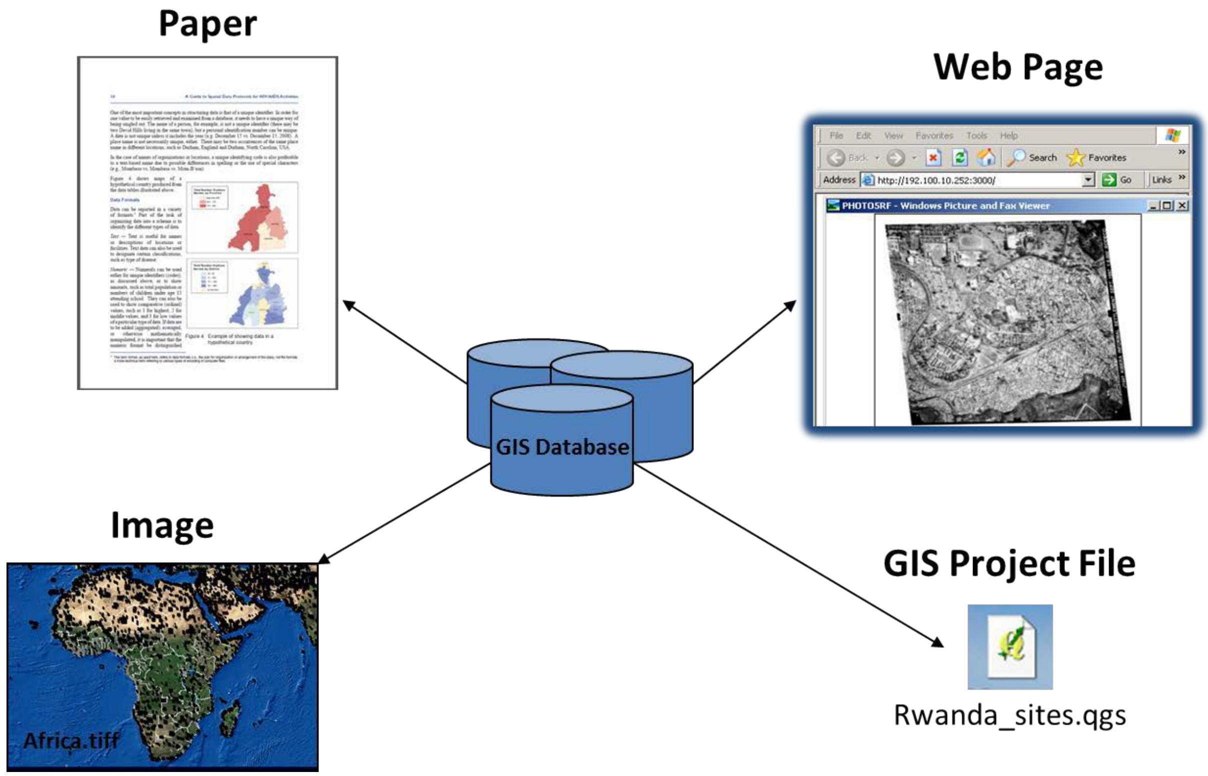


Tables

Prov_ID	Prov_name	Area_km2	HIV_prev
1	Luapula	3,730	11%
2	Central	6,937	13%
3	Eastern	4,180	14%

Data Output

GIS data can be output to paper or electronic formats, and also stored within a formatted project file for further visualization and analysis.

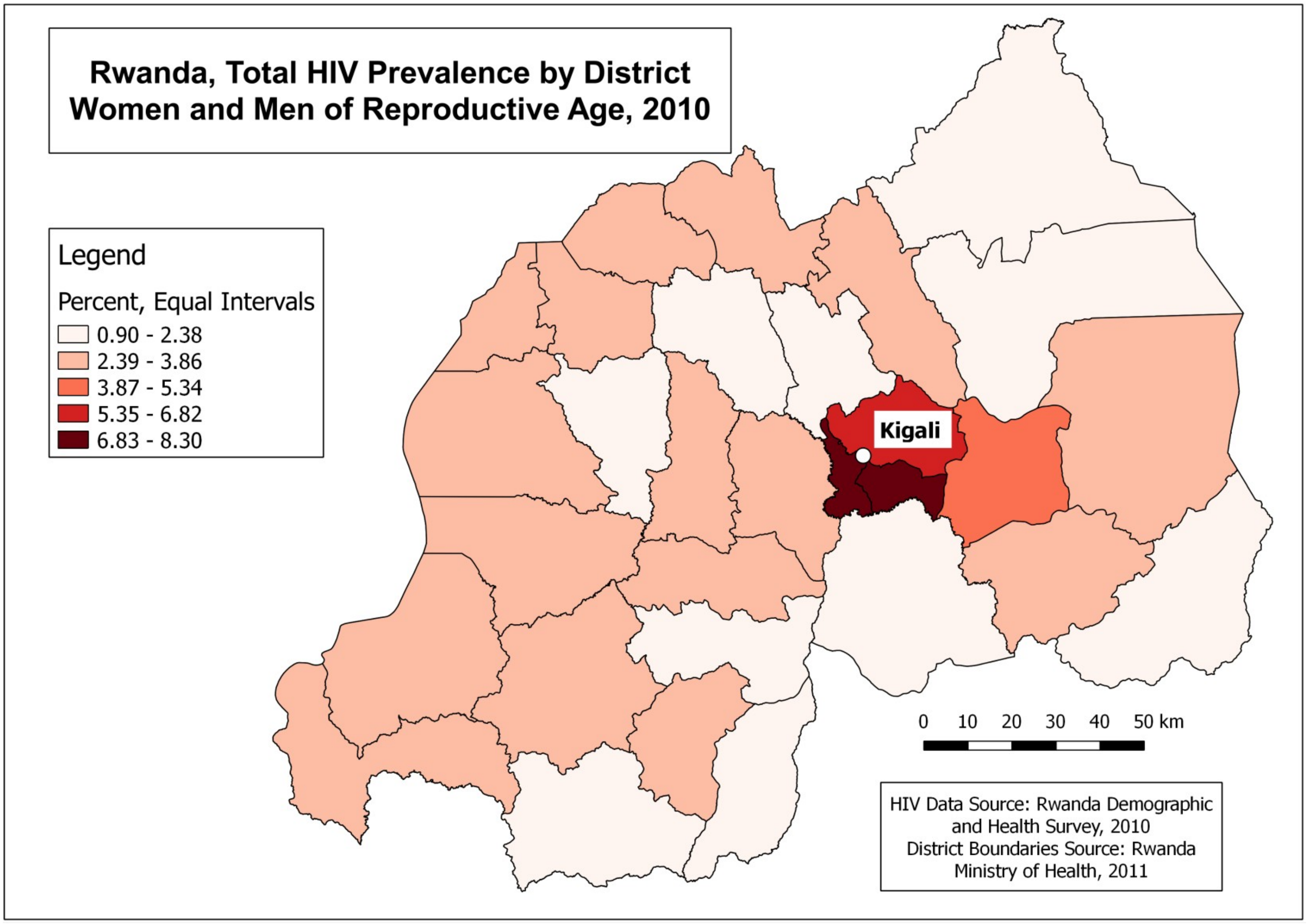


General GIS Advantages

If a picture is worth a thousand words, a map generated using a GIS might be worth a thousand tables. A person looking at a GIS-based map can **quickly detect whether data are clustered, dispersed, or randomly distributed in a given geographic area**.

The ability to use GIS maps to identify clusters provides a distinct advantage in targeting health program interventions. The map below shows clear patterns in HIV prevalence rates that are not immediately evident when viewing the table in list form.

1	DISTRICT	PROVINCE	NISR_CODE	MOH_CODE	HIVPOSF	HIVPOSM	HIVPOST
2	BURERA	NORD	44	404	6	0.6	3.5
3	MUSANZE	NORD	43	403	3.3	2.1	2.7
4	KICUKIRO	VILLE DE KIGALI	13	103	10.1	5.5	7.9
5	KARONGI	OUEST	31	301	3.4	3.3	3.3
6	NYAMAGABE	SUD	25	205	2.9	2.8	2.8
7	RUSIZI	OUEST	36	306	2.8	2.8	2.8
8	NYAGATARE	EST	52	502	2.4	1.4	1.9
9	GATSIBO	EST	53	503	1.2	0.5	0.9
10	NYABIHU	OUEST	34	304	2.1	3.4	2.7
11	RUBAVU	OUEST	33	303	4.3	1.3	2.8
12	GAKENKE	NORD	42	402	0.5	2.5	1.4
13	NGORORERO	OUEST	35	305	2.6	1.4	2.1
14	RUTSIRO	OUEST	32	302	3.7	3	3.4
15	NYARUGENGE	VILLE DE KIGALI	11	101	9.8	6.8	8.3
16	KIREHE	EST	55	505	1.5	0.5	1
17	BUGESERA	EST	57	507	0.8	1.1	1
18	NGOMA	EST	56	506	3.1	2.1	2.6
19	NYAMASHEKE	OUEST	37	307	3.8	3.5	3.6
20	HUYE	SUD	24	204	4.2	2.7	3.5
21	GISAGARA	SUD	22	202	1.4	0.9	1.1
22	RWAMAGANA	EST	51	501	5	4.2	4.6
23	KAYONZA	EST	54	504	4.4	2.9	3.7
24	NYANZA	SUD	21	201	2.1	2.2	2.1
25	RUHANGO	SUD	26	206	3.4	1.6	2.5
26	MUHANGA	SUD	27	207	3.9	1.6	2.9
27	KAMONYI	SUD	28	208	4.4	1.7	3.1
28	GICUMBI	NORD	45	405	3.9	2.9	3.4
29	GASABO	VILLE DE KIGALI	12	102	8.7	4.1	6.4
30	RULINDO	NORD	41	401	2.3	1	1.7
31	NYARUGURU	SUD	23	203	1.3	0.5	0.9

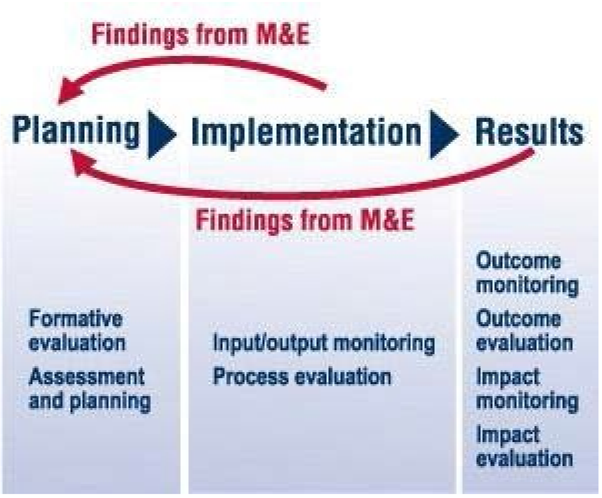


A GIS is capable of much more than simply creating maps. **A GIS has the ability to conduct geographic queries and analysis that can provide information for evidence-based decision-making**. An example might include:

- the identification of populated areas outside a certain distance or travel time from a voluntary counseling and testing (VCT) clinic, or
- the identification of all health facilities within 1 kilometer of a public road.

Most importantly, using a GIS to link and integrate datasets from multiple sectors within a country, such as those pertaining to HIV/AIDS, family planning and reproductive health (FP/RH), tuberculosis (TB), education, demographics, agriculture, energy, etc. can produce fresh **insights for program planning**. Linking and integrating datasets using common **geographic identifiers**, such as district name or district code, has the added benefit of strengthening the **national spatial data infrastructure (NSDI)**.

Cross-Cutting Tool



The HIV/AIDS epidemic exhibits a geographic (also called spatial) distribution or footprint within a given country. As a result, a GIS can be used in every phase of monitoring and evaluation of HIV/AIDS programs—from planning and implementation to results—to visualize and analyze data. This cross-cutting nature of the tool makes it a very powerful resource for M&E and programmatic decision-making.

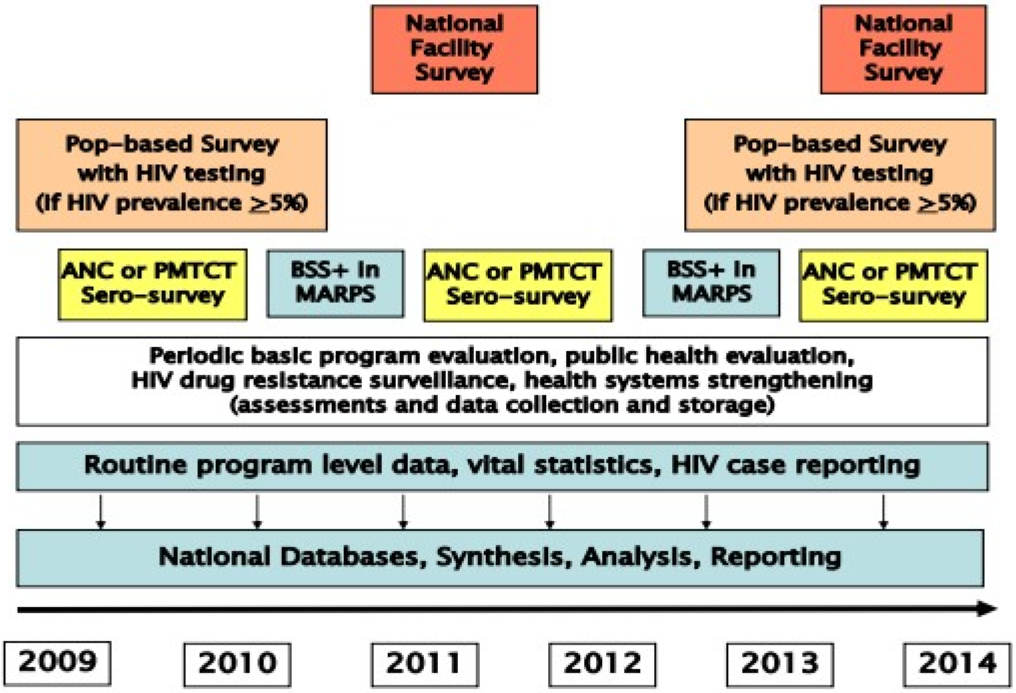
During the **planning phase**, for example, a GIS can be used to map populations at risk in relation to the location of program activities and the resources available to respond to that risk. This GIS-derived information can be critical for effectively targeting interventions.

During the **implementation phase**, a GIS can be used to map and analyze such factors as where services have been delivered, which populations have been served, and where resources are either depleted or sitting idle.

During the **results phase**, a GIS can be instrumental in identifying where observed outcomes and impacts fell short of or exceeded program targets.

Geographic Identifiers as Part of the Data Collection Process

The survey and surveillance data collection process recommended by PEPFAR is illustrated below. In order to be used in a GIS, these databases need to contain geographic identifiers.



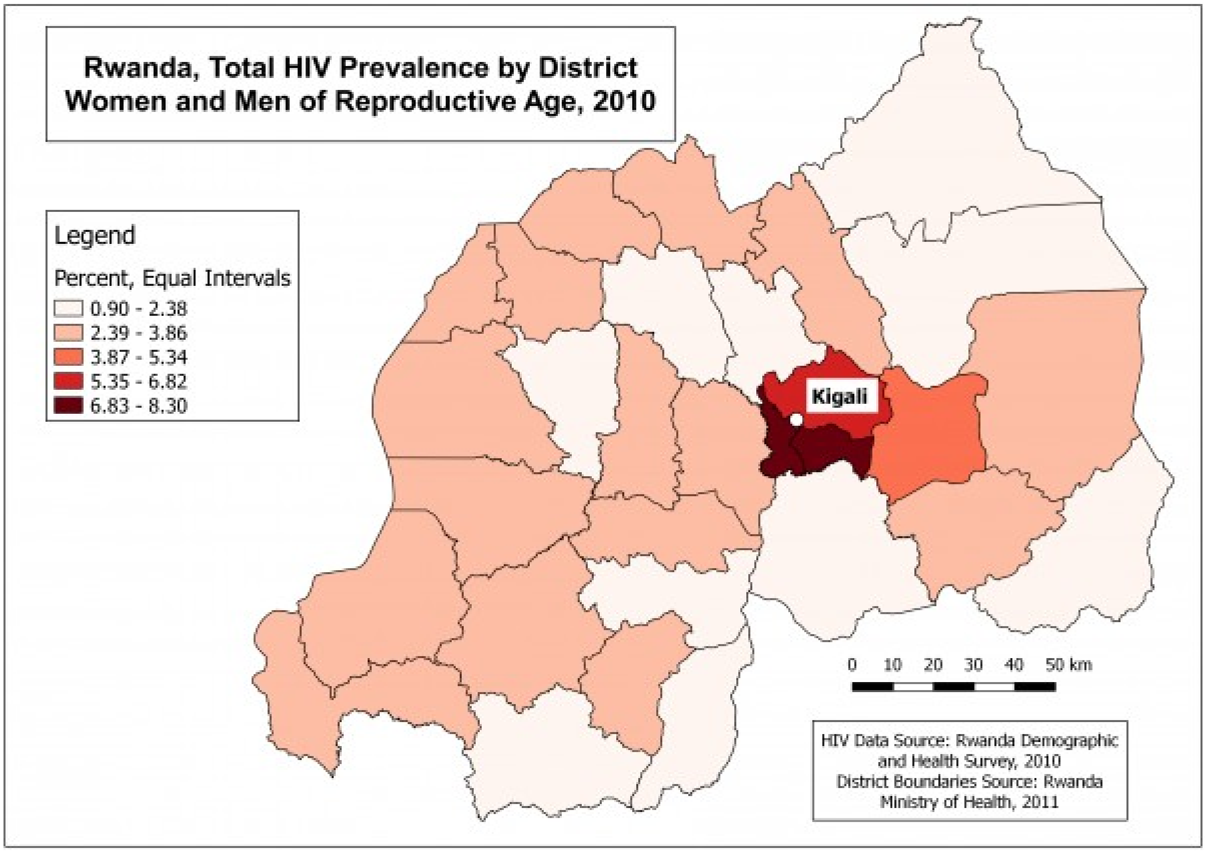
Geographic identifiers can be unique identifiers for:

- Health facilities (such as street addresses or GPS coordinates),
- Administrative divisions (such as district codes or province names),
- Population-based clusters (such as those used by DHS), or
- Venues (social meeting locations) for most-at-risk populations (street addresses or geographic coordinates, or possibly city or neighborhood names).

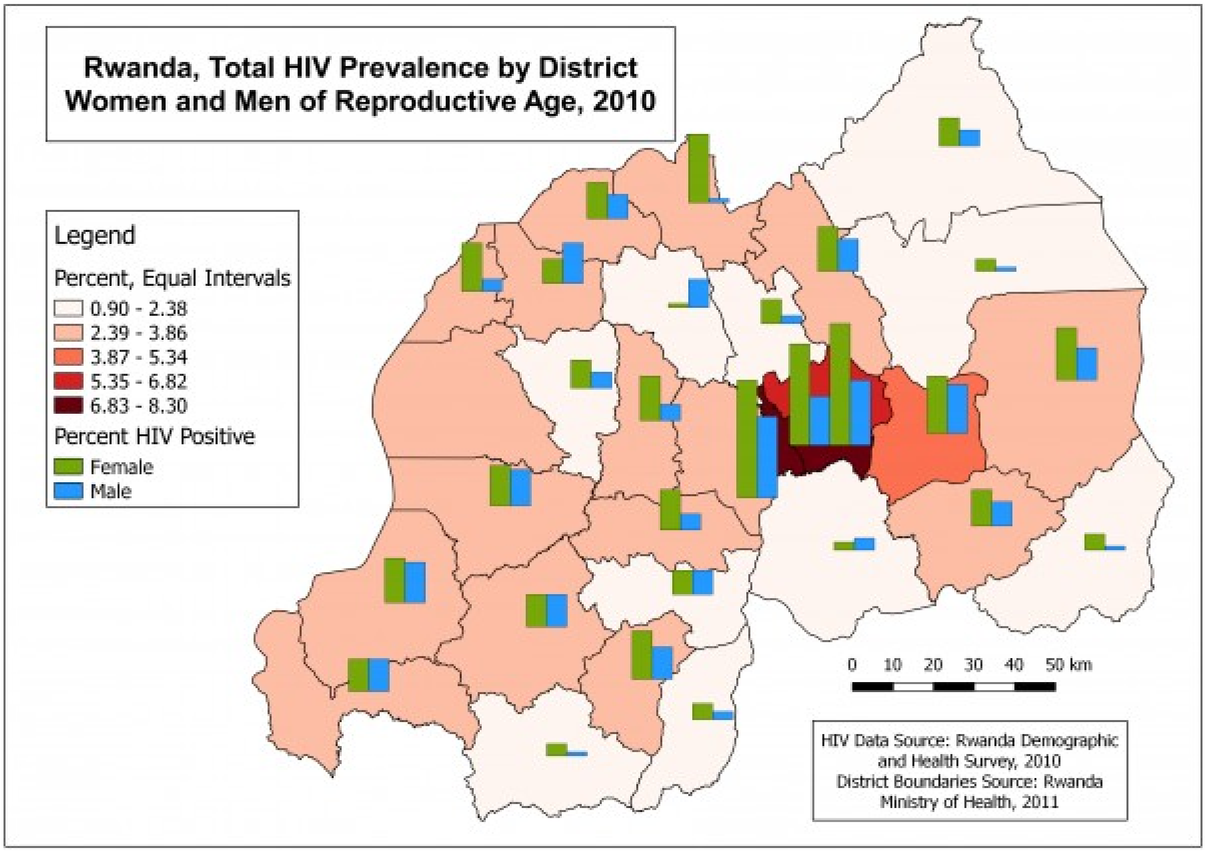
These geographic identifiers provide the foundation for understanding the geographic distribution of HIV/AIDS within a country, which is particularly **useful for HIV/AIDS program planning.**

Targeting of Interventions

If data are available below the national level, a GIS can help target interventions more precisely to make more effective use of finite program resources. In the case of Rwanda, for example, a district-level map of HIV prevalence indicates a greater need for intervention in the urban center of the country, primarily in the three districts that include parts of the City of Kigali.

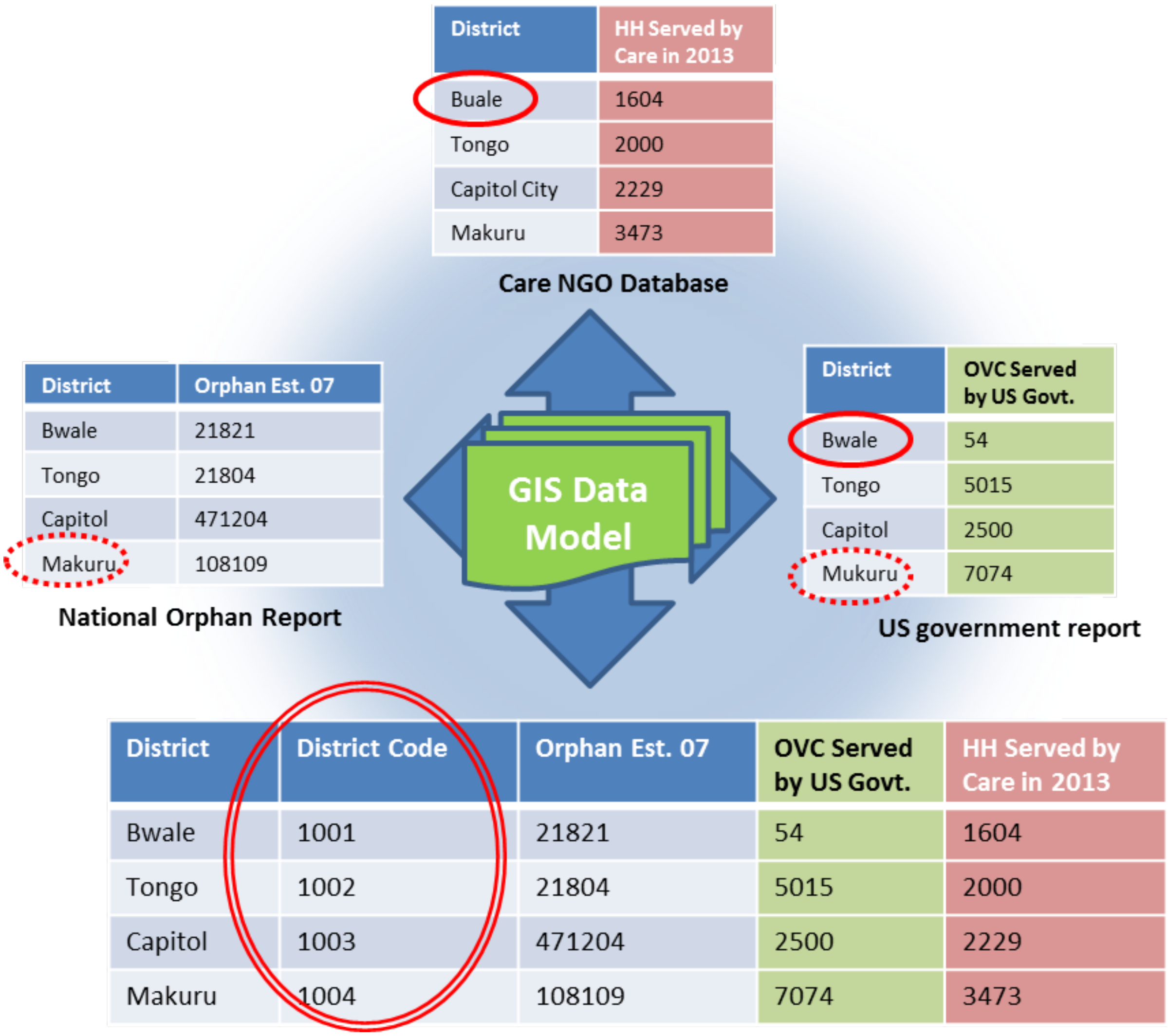


If program data are disaggregated by gender, a GIS can be used to target interventions even more specifically. When HIV prevalence is mapped in this way, it is possible to see the breakdown of HIV by both district (rates for both genders) and gender (rates for males vs. rates for females). For instance, the map shows that in many districts in Rwanda, more women than men are living with HIV. This information might be used to more effectively target HIV prevention efforts.



Data Linking and Program Integration

If databases share common geographic identifiers, such as district names or especially codes, a GIS can be used to link them together. Linking databases using names of geographic entities, such as administrative divisions or health facilities, can lead to mismatches based on slight differences in spelling (as seen in the graphic below, *Makuru* versus *Mukuru* for a district name—spelled one way in the "National Orphan Report," and another way in the "US government report"). For that reason, it is generally **preferable to link databases using unique codes as the geographic identifiers** (*District Code = 1004*).



Linking databases can facilitate program integration, which is a core principle of the U.S. Global Health Initiative (http://www.ghi.gov/) and PEPFAR (http://www.pepfar.gov/.) This is because **linking databases offers opportunities to see patterns in the data that would not be evident if data were not linked.**

For instance, PEPFAR recognizes the significant health benefits that can be achieved for women and children through integration of Prevention of Mother-to-Child Transmission (PMTCT), pediatric HIV, and Maternal and Child Health (MCH) services (PEPFAR 2011). For example, a case study from Kenya published in 2004 has shown that HIV-positive women who receive integrated antenatal care (ANC) and HIV voluntary counseling and testing (VCT) services are more likely to take specific actions that help mitigate the risk of transmitting HIV to their children and partners (Farquhar C et al. 2004).

In addition, since TB and HIV are considered co-epidemics, there is a strong logic for integrating these services. There are many other opportunities for program integration as well, such as integrating certain family planning and reproductive health services with HIV/AIDS services and vice versa. Through data linking, a GIS can play a pivotal role in facilitating the program integration process.

As an additional benefit, linking data using GIS strengthens the national spatial data infrastructure (NSDI) by creating more geographically referenced data. The NSDI is the collective effort to create spatially referenced datasets throughout the public and private sectors. In many countries, the effort is managed through a national body that coordinates data standards. For more information, visit the Global Spatial Data Infrastructure Association (GSDI) (http://www.gsdi.org/.)

In Summary

- A GIS is a database linked to a map. It has 5 main components (hardware, software, geographically-referenced data, procedures, and people).
- The primary functions of a GIS are to capture, store, query, analyze, display, and output geographically-referenced data.
- A GIS can display data on a map, which has the advantage of helping the user quickly visually detect the distribution of data in a particular area.
- The information analyzed by a GIS can help with evidence-based decision-making concerning the targeting of interventions.
- A GIS can also help with data linking and program integration.

Practical Exercises—Session 1

Using a GPS receiver for Data Capture

There are over 24 GPS (Global Positioning System) satellites orbiting the earth every 12 hours, at an altitude of approximately 20,000 Km. Each satellite has an extremely accurate clock on board and is constantly sending its information down to earth, where it can be received with an inexpensive hand-held device called a GPS receiver. As long as 4 or more of these satellite signals are strong enough to be detected (from overhead and from closer to the horizon, in various directions), the receiver unit can pinpoint any position on the earth's surface.

The online version of this course includes a widget which simulates a hand-held GPS receiver (as seen below) and walks the user through a data collection scenario (highlighted areas lead to buttons or explanations of standard unit parts). By following the interactive exercise, you can become familiar with a typical hand-held GPS receiver unit and its capabilities, and you get a chance to practice formatting and collecting geographic coordinates which are tied to a hypothetical data record.

The exercise is estimated to take approximately 5 minutes to complete.

Link to GPS widget: http://www.cpc.unc.edu/measure/resources/training/online-courses-and-resources/certificate-courses/gis-techniques/gps-widget/

Session 1: Knowledge Recap

1. What are the six primary functions of a GIS with respect to geographically-referenced data?

   a) Capture, store, query, analyze, display, and output
   b) Create, edit, transfer, share, input, and output
   c) Input, map, symbolize, revise, print, and distribute
   d) Map, calculate, query, buffer, display, and output

2. Linking and integrating datasets using a GIS, which can produce fresh insights for program planning and strengthen the national spatial data infrastructure (NSDI), require what information to be shared in common by the datasets?

   a) Common attributes (e.g., HIV prevalence by district)
   b) Common geographic coordinate systems
   c) Common geographic identifiers (e.g., district names or codes)
   d) Common number of records

3. What are the five components of a geographic information system (GIS)?

   a) Spheroids, datums, coordinate systems, projections, and maps
   b) Hardware, software, technical personnel, technical documentation, and data
   c) Hardware, software, user manual(s), people (both users and technical personnel), and non-spatial data

d) Hardware, software, procedures, people (both users and technical personnel), and geographically-referenced data

4. For the purpose of monitoring and evaluation (M&E) of HIV/AIDS programs, a GIS is only usefulfor assessing program impacts and outcomes.

   True or False?

5. What are some of the key advantages of using a GIS for M&E of HIV/AIDS and related programs?

   a) A GIS is a cross-cutting tool that can be used throughout the M&E process
   b) GIS data linking promotes program integration, which is one of the core principles of the Global Health Initiative (GHI) and the U.S. President's Emergency Plan for AIDS Relief (PEPFAR)
   c) A GIS facilitates more effective targeting of interventions through identification of clusters and hot spots, especially when geographically-referenced data are available at a more detailed geographic level and/or disaggregated by sex.
   d) All of the above

ANSWERS:

1) (a) Capture, store, query, analyze, display, and output
2) (c) Common geographic identifiers (e.g., district names or codes)
3) (d) Hardware, software, procedures, people (both users and technical personnel), and geographically-referenced data
4) (False) A GIS is a cross-cutting tool that can be used to view and analyze data throughout the M&E process, from the planning phase through the implementation and results phases.
5) (d) All of the above

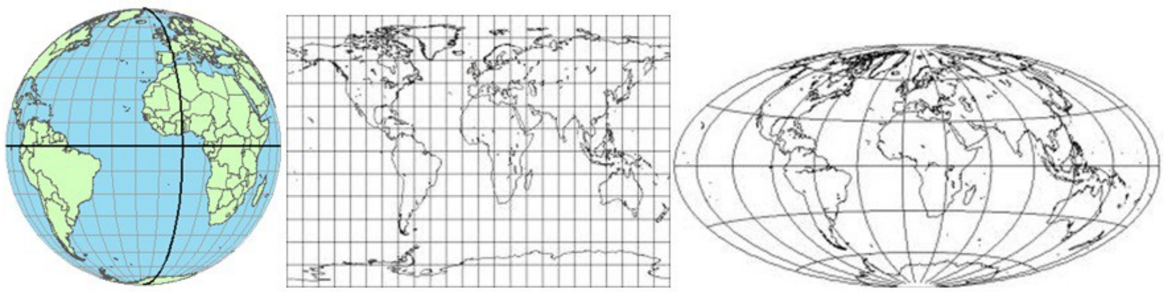
Session 2: Getting Started with GIS Software and Data

Many free and open source software (FOSS) solutions exist for mapping and analyzing geographically-referenced data. This session will explore a variety of the options available. QGIS is the most fully functional FOSS GIS available, and can be used to complement other FOSS solutions for analysis and production of publication-quality maps. The practical exercises at the end of each session are designed specifically for this solution. To use QGIS (or any other GIS) effectively requires an understanding of geographic data fundamentals.

After completing this session, learners will be able to:

- Explain the fundamentals of **geographic data**, including the **shapefile** format.
- Differentiate the primary uses of several free and open source GIS software options.
- Identify QGIS advantages, considerations for use, and key support resources.
- Download, install, and explore QGIS for Windows.

What is Geographic Data?



Before getting started with any GIS software, it is important to review some of the fundamentals of **geographic data** and data storage.

This section and the following sections of this session provide a condensed review of geographic data fundamentals. For more detailed information, please see the Geographic Data session of the *Geographic Approaches to Global Health* course. http://www.globalhealthlearning.org/course/geographic-approaches-global-health

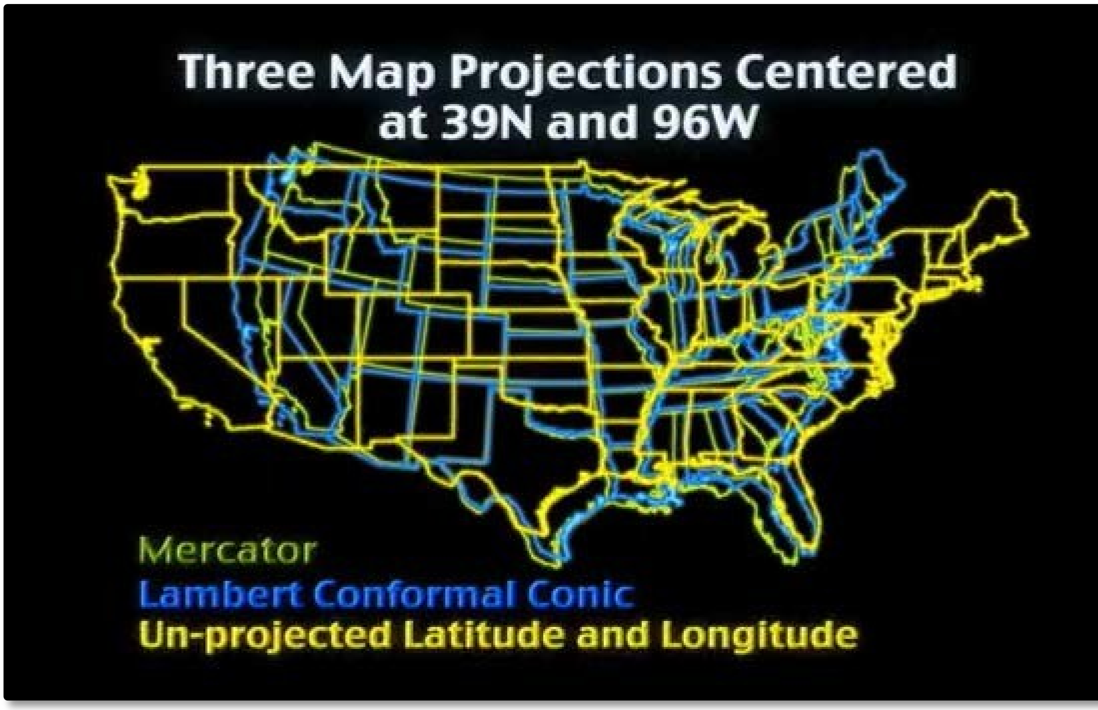
Geographic data can be mapped. Data points become geographic by assigning them locations in relation to the surface of the Earth. In practical terms, this is accomplished using **geographic identifiers**, which are pieces of information that specify the physical location of something. Some examples might be district names, longitude and latitude, or street addresses.

Geographic Data Transformations

Longitude and latitude describe locations on the three-dimensional surface of the Earth (they are one type of **GCS**, or **geographic coordinate system**), and are therefore considered to be "unprojected." **Longitude** and **latitude** are based on "slices" of the globe and are therefore larger at the equator than at the poles. They are excellent for original data capture and storage, but can be difficult to use for accurately measuring surface areas, perimeters, and distances.

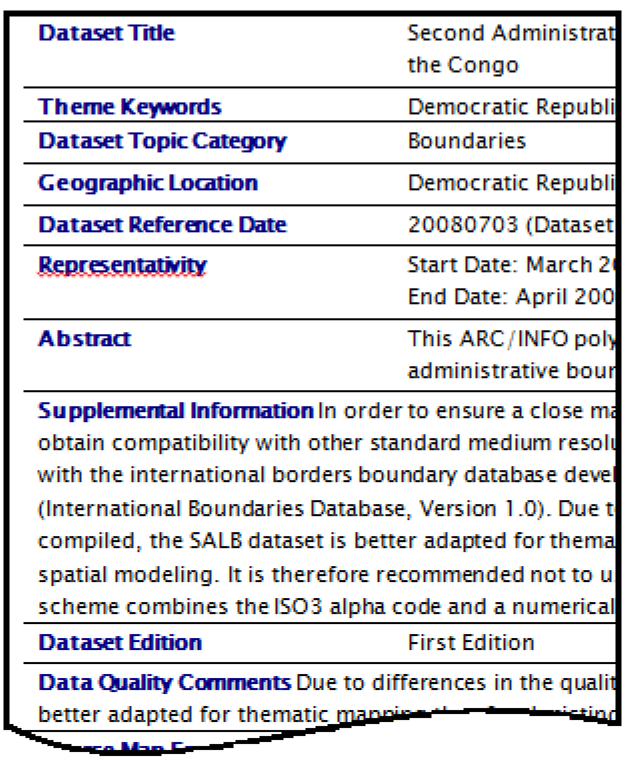
Prior to such measurements, it is recommended that raw latitude/longitude coordinates be transformed or "projected" onto a flat surface. It is important to use a **projection** method that best preserves the characteristic being measured. All projections in some way warp shape, distance, or direction during transformation from a 3D to a 2D surface. The online version of this course links to a one-minute video, courtesy of ESRI, which describes the challenges of different map projections and the map projection decisions that GIS users must make. The web address for the link to the video can be found below the screenshot.

Many texts exist to help with selecting a projection that is suitable for a given geographic area. A GIS can be used to aid this transformation.



Link to Video from ESRI Corporation, "Map Projections and GIS": https://www.youtube.com/watch?v=e2jHvu1sKiI

Geographic Dataset's Metadata



Dataset Title	Second Administrat the Congo
Theme Keywords	Democratic Republi
Dataset Topic Category	Boundaries
Geographic Location	Democratic Republi
Dataset Reference Date	20080703 (Dataset
Representativity	Start Date: March 2 End Date: April 200
Abstract	This ARC/INFO poly administrative bour
Supplemental Information In order to ensure a close ma obtain compatibility with other standard medium resolu with the international borders boundary database devel (International Boundaries Database, Version 1.0). Due t compiled, the SALB dataset is better adapted for thema spatial modeling. It is therefore recommended not to u scheme combines the ISO3 alpha code and a numerical	
Dataset Edition	First Edition
Data Quality Comments Due to differences in the qualit better adapted for thematic mappi	

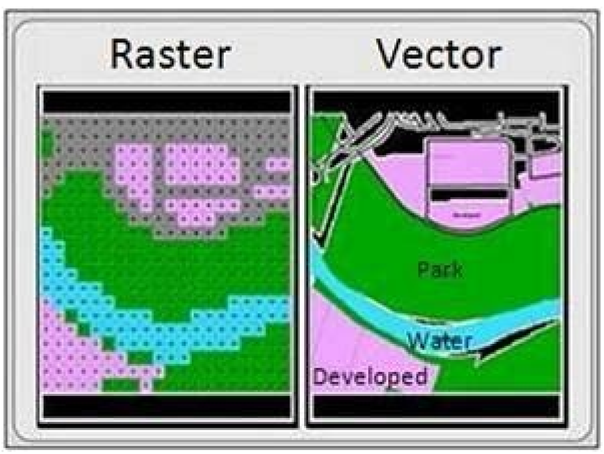
The projection for a dataset should be specified in the dataset's **metadata**, which is "data about data." At a minimum, the metadata should include:

- Data source (organization name/contact information)
- Date data was collected and/or modified
- Coordinate system, projection, and **datum** used to specify the locations and shapes of features represented in the dataset
- **Scale** at which data was collected (and should therefore be used)

A geographic dataset cannot be used with confidence if it is not accompanied by good metadata, so GIS users should insist on complete metadata wherever possible, especially when purchasing a dataset.

It should be noted that there are several key geographic metadata standards, including the ISO 19115 (from the International Standards Organization) and the US FGDC (United States Federal Geographic Data Committee).

Representing Geographic Data



Geographic data can be represented in two primary ways: **vector** and **raster**.

Vector data consists of points (e.g., health facilities), lines (e.g., roads), and polygons (e.g., health districts), whereas **raster data** is arranged in rows and columns in which each grid cell or "pixel" has a set size and a range of attribute values (for example, a satellite image might be composed of grid cells that each represent 30 meters by 30 meters on the Earth's surface and that have attribute values ranging from 1 to 10 based on whether the pixels represent urban, agriculture, forest, etc.).

Another name for the points, lines, and polygons used to represent **vector** data is "geographic primitives." The vector format requires less disk space for storage, which makes it a more efficient format for uses such as network and other types of distance analysis. Remotely imaged data, such as satellite-captured vegetation or elevation data, is captured in **raster** format. The storage size of raster files is highly dependent on the cell (pixel) size, or the "resolution" of the images. In general, high-resolution raster files are **much** larger than vector files, over an identical area. Also, though raster and vector data can be used together in a GIS, it should be noted that users will need differing skill sets and tools to work with and analyze raster data vs. vector data.

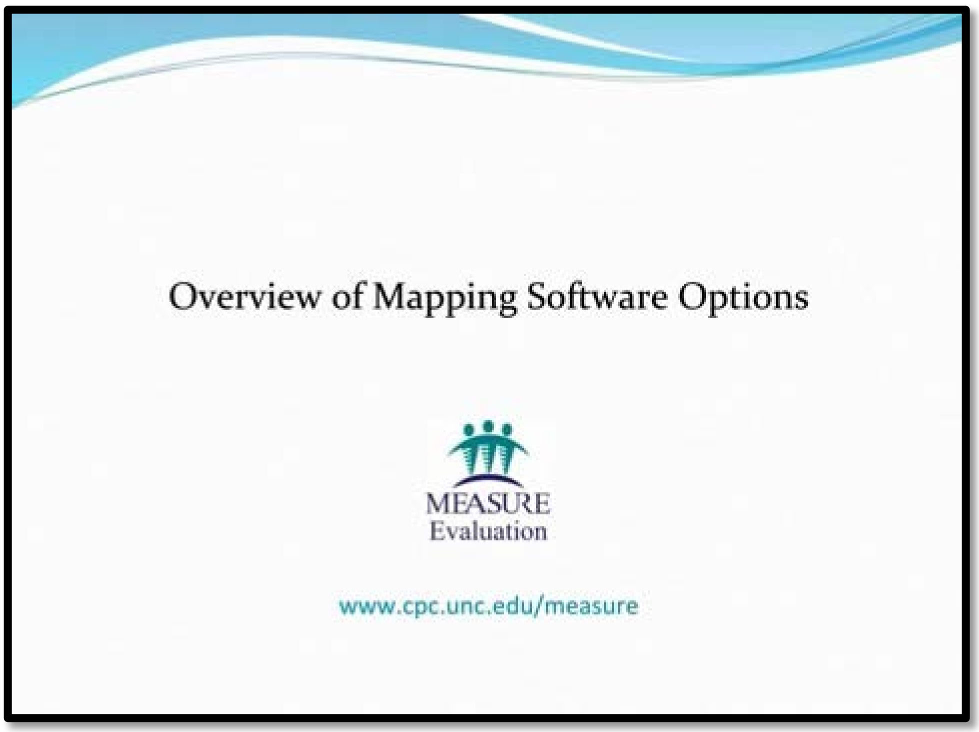
One of the most common file types for storing vector data has historically been the **shapefile**. A shapefile is actually a collection of several files which contain not only geographic information such as an object's coordinates, but also other information such as attributes or projection information. As a result of the trend toward open data formats, however, geographic data are increasingly being made available in file formats that can be read with a simple text editor, such as **KML** and **GeoJSON**.

QGIS, the free and open source software that will be introduced in more detail in the following pages, can import and export geographic data in a variety of vector file formats, including shapefile, KML, and GeoJSON. For this course, we will primarily use spatial and attribute data linked together in shapefile (vector) format.

FOSS GIS and Related Options

GIS expertise was once the exclusive domain of technical specialists using commercial software purchased and licensed from a small community of vendors. Additionally, free and open source software (FOSS) GIS solutions were generally considered not to perform as well as commercial products, which made them less likely to be chosen.

The situation has changed in recent years. **FOSS GIS options are becoming increasingly available, functional, and in demand, with a growing online user presence, and are emerging as comparable to commercial products in functionality.** A 2-minute video from MEASURE Evaluation (link is embedded in online version of course and also given below) shares more about this evolution.



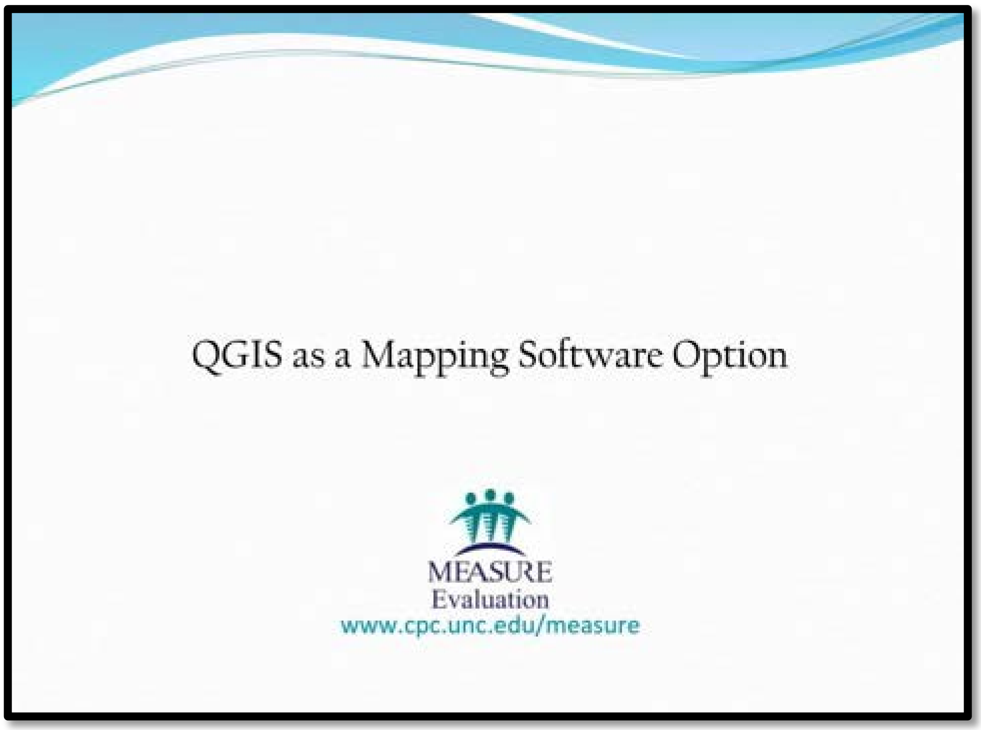
Link to Video from MEASURE Evaluation Geospatial Group, “Mapping Software Overview”:
https://www.youtube.com/watch?feature=player_embedded&v=nyodKVi7abY

NOTE: Though most of the software options reviewed in this session are available to users free of charge, not all of them are open source. The term “open source” refers to the program code being freely shared and contributed to by a user community.

QGIS Advantages

QGIS is a **desktop GIS solution** for Windows, Linux, BSD, and Mac OS X. As a fully functional GIS solution, QGIS allows users to create, store, query, analyze, display, and output spatial data using a variety of input and output formats. QGIS can be used to satisfy the vast majority of day-to-day GIS requirements for public health officers who have some training in the software. This makes QGIS a reasonable FOSS (free and open source) alternative to commercial GIS packages such as ArcGIS, especially in limited-resource settings.

A 2-minute video from MEASURE Evaluation (link is embedded in online version of course and also given below) highlights some of the key advantages of using QGIS:



Link to Video from MEASURE Evaluation Geospatial Group, "QGIS as a Mapping Software Option": https://www.youtube.com/watch?feature=player_embedded&v=nyodKVi7abY

- Available for Windows, Linux, BSD, Mac OS X, and Android.
- Fully functional GIS.
  - Can import and export spatial and attribute data in a variety of formats (**shapefile**, personal **geodatabase**, **KML**, **GeoJSON**, etc.).
  - Can import, map, and export GPS data.
  - Can execute a wide range of GIS queries and analysis on data layers based on either their spatial or non-spatial attributes; this functionality is lacking in many of the leading web-based FOSS solutions such as the District Health Information System 2 (DHIS 2) or DevInfo.
  - Is an excellent complement to other GIS packages, such as DHIS 2 or DevInfo, for the creation of publication-quality maps.

- Possesses good online documentation (http://qgis.org/en/docs/index.html), including a user manual and tutorial.
- Boasts a growing user community that has facilitated an increase in the level of available technical support using web-based FAQs as well as an active forum (http://gis.stackexchange.com/) of both new and experienced users asking questions and trading information.
- Provides an interface that can be displayed in several languages.

QGIS Considerations and Resources

Before selecting QGIS as a potential GIS solution, users should consider the following:

- QGIS is not a commercial product, so technical support is not as readily available as it would be under a maintenance contract from a commercial vendor.
    - The relative lack of technical support makes it more important for QGIS users to have some GIS training and to be able to troubleshoot problems using the more limited resources available.
- The QGIS software package has historically been updated fairly frequently.
    - Although the pace of updates has recently slowed, there is no guarantee of consistency in the release schedule. Frequent functionality and interface changes between versions can be significant, which can require a non-trivial effort to relearn the software.
- QGIS currently lags a bit behind ArcGIS, the leading commercial GIS software, in overall functionality for vector analysis and print output.
    - This is especially true with respect to batch processing of large volumes of data, complex spatial analyses such as network analysis, and manipulation of cartographic elements in map layouts for publication.

Despite these considerations, QGIS is a strong product with a growing user community. In addition to online technical support through mailing lists, chat, forums, and user groups, QGIS offers the following resources, available on their website (http://qgis.org/en/site/forusers/index.html):

- User Guide/Manual for QGIS 2.0
- A Gentle Introduction to GIS
- See the Appendix and "Practical Exercises" sections of this course for additional QGIS training resources.

Other Mapping Options

QGIS is not the only free GIS solution. This page provides a synopsis of a few other mapping options available for spatial data display and analysis.

District Health Information System 2 (DHIS 2) (http://www.dhis2.org/) is a web-based software from the *Health Information Systems Programme (HISP)* (http://www.hisp.org/) that provides a very attractive and easy-to-use dashboard interface that can display maps and statistical graphics (bar charts, pie charts, etc.).

DevInfo (http://www.devinfo.org/) is an open-access, customizable web-based data visualization software from the *United Nations* that was designed for monitoring and reporting progress on the Millennium Development Goals (MDGs), although the software also supports the mapping and analysis of other indicators based on data provided by users.

StatCompiler (http://www.statcompiler.com/) is a web-based data visualization program from *The DHS Program* http://www.dhsprogram.com/ that allows users to map and analyze Demographic Health Survey (DHS) data such as HIV indicators for more than 70 countries.

Epi Info (http://wwwn.cdc.gov/epiinfo/) is a desktop software package from the *U.S. Centers for Disease Control and Prevention (CDC)* that provides a simplified core suite of tools for public health professionals to use for collecting, visualizing, analyzing, and reporting **epidemiologic data**.

Excel to Google Earth (E2G) (http://www.cpc.unc.edu/measure/e2g) is an Excel macro (a mini- program that runs within Excel) available free from MEASURE Evaluation that provides decision- makers and analysts who have no GIS access or training with the ability to create color-shaded maps in Google Earth using data stored in an Excel spreadsheet.

Google Maps (https://maps.google.com/) **and Google Earth** (http://www.google.com/earth) are free mapping tools from Google, both of which can be leveraged for public health studies.

GeoCommons (http://geocommons.com/) is a free online mapping and data-sharing site from FortiusOne that offers a simple and attractive suite of data visualization tools.

OpenGeoDa (https://geodacenter.asu.edu/projects/opengeoda) allows users to conduct **exploratory spatial data analysis (ESDA)** using a variety of statistics-based data visualization techniques.

For more details, view our **Suggested Uses and Considerations** for these mapping options, in the Appendix.

In Summary

- In order to be mapped, data must contain geographic identifiers. Common geographic identifiers will also allow one dataset to be linked to another.
- Metadata (data about the data—where and when and why it was collected) is important.
- Geographic data can be represented in vector or raster format, and each of these formats has advantages for different uses. One common type of file used for vector data is the shapefile.
- Free and/or open source software mapping options are becoming increasingly functional and available. A popular choice (and the one used in this course) is QGIS.

Practical Exercises—Session 2

In this session, you were introduced to the advantages of QGIS in comparison to a number of other FOSS GIS options. Now we invite you to learn more about how to use QGIS by working through two practical exercises. (We estimate these optional exercises will take about 45 minutes to complete.)

See Appendix 2 of this document for the instructions for these exercises. In order to complete these exercises, you will need some accompanying data files which can be accessed online.

1) You will need to **download** the data files from our public DropBox folder (http://bit.ly/Ex2DropBox). Follow the link and click on the blue Download button (they will come together in one .zip folder) and store them on your own computer in a new, easily accessible folder (place it wherever you like and name it whatever you like; the default name is simply "Exercise 2.2").

2) **Unzip** the folder, using your operating system or a program such as WinZip, 7-Zip, or Stuffit Expander. (You'll probably need to use a command like *extract all* and tell your computer where you would like to store the unzipped files.) The unzipped folder will be approximately 8MB in size. **NOTE: Leave the file structure** (folders and files within the main folder) **in place** once the main folder is unzipped. The main QGIS project file, when opened, will look for a folder named *shapefiles*. If it does not find this folder (or the files in it), or if the folder has not been unzipped properly, the exercise will not perform as expected.

By the end of these Session 2 exercises, you will be able to:

- Open a QGIS project already in progress
- Work with the main components of the interface
- Navigate within the map window
- Select, identify, and query geographic features

As mentioned above, we estimate that these supplemental exercises will take you about 45 minutes to complete and are optional for completion of the course.

Session 2: Knowledge Recap

1) QGIS strengths include the following:

A) Provides full GIS capabilities for data capture, storage, query, analysis, display, and output.
B) Provides step-by-step, guided instructions for creating color-shaded maps in Google Earth using data stored in an Excel spreadsheet.
C) Can serve as an excellent complement to other GIS or mapping packages, such as the District Health Information System (DHIS 2) or DevInfo, for the creation of publication-quality maps.
D) A and C

2) What is a geographic identifier?

A) A type of file which contains information about the projection used in a dataset
B) A piece of information which specifies the physical location of something and can be used to join data tables.
C) A file which includes the scale and coordinate system of a set of geographic data
D) A type of file format invented by ESRI that allows files to be shared in other types of mapping programs
E) None of the above

3) At a minimum, a good metadata file should include:

A) The data source
B) The date the data was collected or modified
C) The coordinate system, projection, and datum used to specify locations represented in the data set
D) The scale at which the data was collected (and should therefore be used)
E) All of the above

4) Geographic features stored in a “geographic” or “unprojected” coordinate system, in which coordinates are stored as longitude and latitude in degrees, provide a perfectly suitable basis for measuring distances and surface areas.

True or False

5) One of the most common file types for storing raster data is the shapefile.

True or False

ANSWERS:

1) (d) A and C. QGIS provides full GIS capabilities, and can serve as an excellent complement to other GIS and related software packages that do not provide the ability to create publication-quality maps. QGIS does not provide step-by-step, guided instructions for creating color-shaded maps in Google Earth using Excel data. That can be accomplished using the **Excel to Google Earth** program (http://www.cpc.unc.edu/measure/e2g) from MEASURE Evaluation.
2) (b) A piece of information which specifies the physical location of something and can be used to join data tables.
3) (e) All of the above.
4) False. Before measuring distances or surface areas, geographic features stored using coordinates in longitude and latitude, which are not uniform units of measure around the globe, should first be transformed or "projected" into a flat, two-dimensional coordinate system that best preserves the characteristic being measured.
5) False. Shapefiles are used to store vector data.

Session 3: HIV/AIDS Data Sources

This session highlights the fact that HIV/AIDS data that can be mapped and spatially analyzed for M&E purposes is available from a variety of sources. One needs to understand the geographic data structure of these data sources in order to work with them effectively. Administrative division boundaries in digital format are available from a variety of sources, and are required to satisfy mapping and spatial analysis needs that exceed the capabilities of StatCompiler and similar programs. DHS data is available at different geographic levels, and there are specific considerations for working with the data at each level.

After completing this session, learners will be able to:

- Identify some of the principal sources of HIV/AIDS and related health data that can be mapped and analyzed spatially for program M&E.
- Explain how the geographic structure of these data sources can be leveraged for mapping and spatial analysis.
- Find digital files containing administrative division boundaries needed for mapping or spatial analysis.
- Understand how to work with DHS data at different geographic levels.

Primary HIV/AIDS Data Sources

In a country with a well-developed data infrastructure, much of the data needed by HIV/AIDS programs for evidence-based decision-making may be available from the national routine health information system (RHIS). Augmenting RHIS data with the results of population-based surveys can improve understanding of the HIV/AIDS epidemic.

A good starting point for finding survey data for many countries is the HIV/AIDS Survey Indicators Database (http://hivdata.measuredhs.com/), which is one of the online resources provided by MEASURE DHS.

The indicators are derived primarily from the guidance document: National AIDS Programmes: [A] Guide to Monitoring and Evaluation (http://www.who.int/hiv/pub/me/pubnap/en/), published in 2000 by UNAIDS, which provides standardized indicators for monitoring and evaluating the success of HIV/AIDS programs. Among the indicators included in the database, several are tied to targets set by the UN General Assembly Special Session on HIV/AIDS (UNGASS) or linked to the

Millennium Development Goals (MDGs) and the President's Emergency Plan for AIDS Relief (PEPFAR).

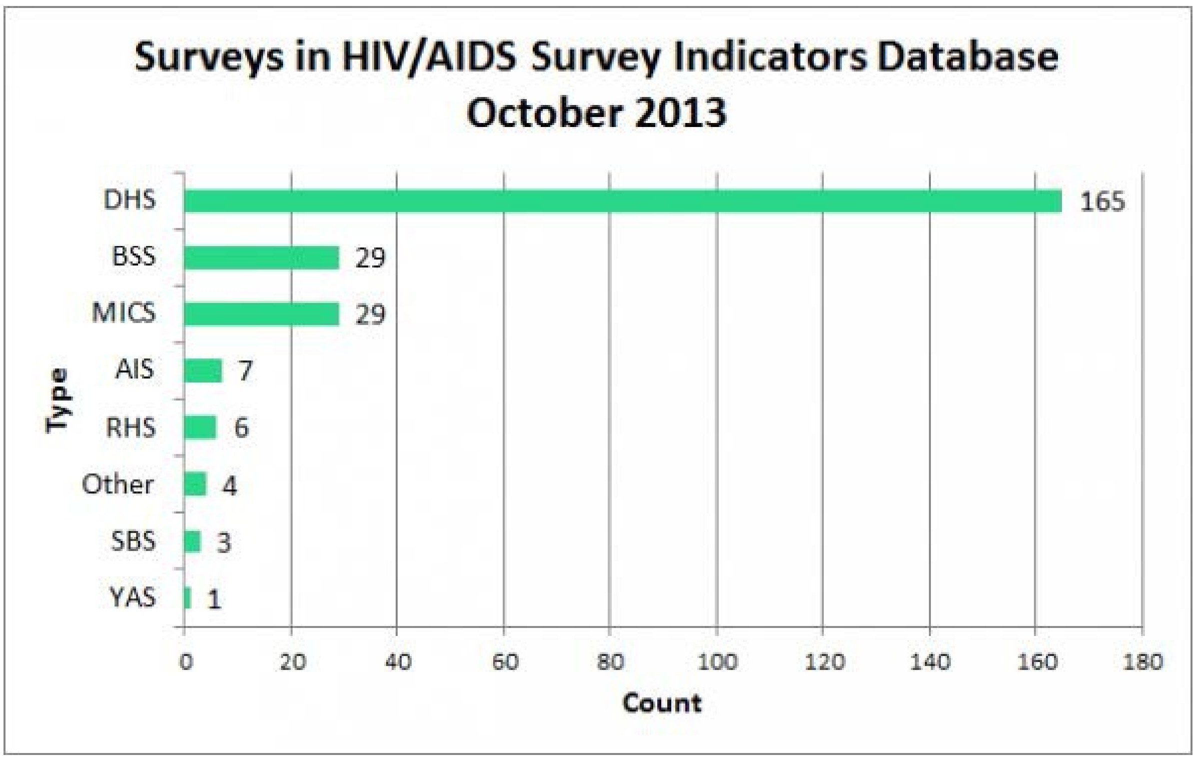
Other Data Sources

The majority of the surveys contained in the HIV/AIDS Survey Indicators Database are **Demographic and Health Surveys (DHS)** from The DHS Program. However, there are also a number of surveys from a variety of other sources, including:

- **AIDS Indicator Surveys (AIS)**
- **Behavioral Surveillance Surveys (BSS)**
- **Multiple Indicator Cluster Surveys (MICS)**
- **Reproductive Health Surveys (RHS)**
- **Sexual Behavior Surveys (SBS)***
- **Young Adult Survey (YAS)***

*limited availability

Below is a histogram that shows the number of surveys by type contained in the database. All of these surveys are listed in the Glossary, for more information about each.



Service Provision Assessment (SPA) Surveys

Another type of survey available from The DHS Program is the Service Provision Assessment (SPA) survey (http://dhsprogram.com/What-We-Do/Survey-Types/SPA.cfm), which is designed to assess the availability and readiness of facility-based health services in developing countries. This is accomplished by asking four main types of questions:

- What is the availability of different health services in a country?
- To what extent are facilities prepared to provide health services?
- To what extent does the service delivery process follow generally accepted standards of care?
- Are clients and service providers satisfied with the service delivery environment?

Source: The DHS Program (http://dhsprogram.com/), October 2013

SPA Indicators

The SPA surveys assess a variety of indicators related to the availability and readiness of HIV/AIDS services, including:

Infrastructure, Resources, and Systems: water; electricity; latrines; infection control; management systems; storage and stock monitoring for vaccines, contraceptives, and medicines

Child Health: availability of vaccines, medicines, and Vitamin A; availability of curative care services and the availability of equipment and supplies for outpatient care; adherence to guidelines for sick child care

Maternal and Newborn Health: availability and appropriate assessment of clients for antenatal care; delivery services; newborn care; emergency obstetric care

Family Planning: availability of contraceptives and supplies; user fees; counseling and client assessment; provision of STI treatment for family planning clients

HIV/AIDS: availability of HIV testing services; HIV/AIDS care and support services; antiretroviral treatment; prevention of mother-to-child-transmission; post-exposure prophylaxis

Sexually Transmitted Infections (STIs)

Malaria: availability of malaria diagnostic and treatment services; guidelines; antimalarial treatments; laboratory diagnostic capacity

Tuberculosis: availability of TB diagnostic services; availability of first-line medicines for treating TB

Basic surgery

Non-communicable diseases: diabetes; cardiovascular diseases; chronic respiratory diseases

SPA final reports for a variety of countries can be accessed in the publications section of The DHS Program site. http://dhsprogram.com/publications/index.cfm

Source: **The DHS Program**, October 2013 (http://dhsprogram.com/What-We-Do/Survey-Types/SPA.cfm)

Geographic Identifiers

As discussed previously, attribute data (such as HIV/AIDS program indicators for districts) must be organized geographically in order to map and analyze them using a GIS.

Geographic organization involves assigning geographic identifiers (such as district names or codes) to data stored in a table and using those geographic identifiers to join the data to the corresponding geographic entities (such as district-level administrative divisions stored in a shapefile).

Joining attribute data to a shapefile will be successful only if the two data sets contain the same **geographic identifiers.** Also, the attribute table and the shapefile table should both contain only a single record for each geographic entity (such as district), and the geographic identifiers in both tables should be unique. For example, neither table should contain more than one occurrence of a particular district name or code, as duplicate geographic identifiers can cause confusion during the joining process.

Geographic Structure of HIV/AIDS Data Sources

The geographic organization of the HIV/AIDS data sources discussed in the previous pages is dependent upon their sampling designs. For example, some surveys are designed to be representative of the nation as a whole, but do not provide information that can be evaluated on a sub-national level. Other surveys target sub-populations in specific geographic locations and therefore cannot be extrapolated to higher administrative regions, including national or sub-national administrative divisions. The table below provides a summary to help you understand the geographic level at which some of the leading HIV/AIDS data sources can be mapped and/or analyzed.

Data Source	Geographic Level(s) Represented	Notes
Demographic and Health Surveys (DHS)	DHS regions, sub-regions, and clusters	• DHS regions generally correspond with existing first-level administrative divisions, such as provinces. • Since 2009, where possible, DHS data are also provided at the second administrative level, such as districts. These data will be provided in the appendices of the final report. • To preserve confidentiality of survey respondents, clusters of households involved in the DHS are represented as a single point. *Cluster locations do not correspond to the exact locations of the populations they represent and should not be used for mapping or spatial analysis at the point level.*
AIDS Indicator Surveys (AIS)	Same as DHS	Same as DHS
Behavioral Surveillance Surveys (BSS)	Sites	Example: Truck stops and truck depots in border and trucking towns in order to understand the behavior of long-distance truck drivers.
Multiple Indicator Cluster Surveys (MICS)	Clusters aggregated to varying levels of administrative divisions	Cluster data are aggregated to administrative divisions. The number of administrative levels represented in the survey can vary by country and sample design.
Reproductive Health Surveys (RHS)	Regions and sub-regions	RHS geographic levels vary depending on sample design and the administrative hierarchy available in a country.
Sexual Behavior Surveys (SBS)	Nation	SBS contained in the MEASURE DHS HIV/AIDS Spatial Data Repository are all for Zambia. *Data are representative at the national level only.*
Young Adult Survey (YAS)	Province	The lone YAS contained in the MEASURE DHS HIV/AIDS Spatial Data Repository is for Indonesia. Data are available at the province level.
Service Provision Assessments (SPAs)	DHS Region	The informed consent agreements used for the SPA surveys guarantee respondents that individual facilities will not be identified. For this reason, SPA data are aggregated to the DHS region for reporting.

Constraints to Working with DHS Data in a GIS

In many low- and middle-income countries, the **DHS** (http://dhsprogram.com/) are the most widely available, reliable, and timely sources of population-based survey data for M&E of HIV/AIDS programs. Before working with DHS data in a GIS, however, there are some key sampling and data collection factors to consider:

Table 14.5 HIV prevalence by socioeconomic characteristics

Percentage HIV positive among women and men age 15-49 who were tested, by socioeconomic characteristics, Rwanda 2010

	Women		Men		Total	
Background characteristic	Percentage HIV positive	Number	Percentage HIV positive	Number	Percentage HIV positive	Number
Religion						
Catholic	3.5	2,947	2.1	2,713	2.8	5,660
Protestant	3.5	2,825	2.4	2,040	3.0	4,865
Adventist	3.7	943	1.9	683	3.0	1,626
Muslim	11.9	77	4.2	107	7.4	184
Traditional/Other/No religion	7.4	111	3.2	147	5.0	258
Missing	*	13	na	na	*	13
Employment (last 12 months)						
Not employed	3.3	1,154	0.7	455	2.5	1,610
Employed	3.8	5,762	2.4	5,235	3.1	10,997
Residence						
Urban	8.7	1,049	5.4	938	7.1	1,987
Rural	2.8	5,867	1.6	4,752	2.3	10,619
Province						
City of Kigali	9.4	808	5.1	741	7.3	1,548
South	3.0	1,593	1.8	1,308	2.4	2,901
West	3.2	1,688	2.0	1,307	2.7	2,995
North	3.1	1,168	1.8	899	2.5	2,067
East	2.5	1,660	1.6	1,435	2.1	3,095
Education						
No education	4.2	1,055	2.9	583	3.7	1,638
Primary	3.4	4,742	2.1	3,922	2.8	8,664
Secondary and higher	4.5	1,023	2.3	1,062	3.4	2,085
Wealth quintile						
Lowest	3.3	1,252	1.9	855	2.7	2,107
Second	3.1	1,392	1.9	986	2.6	2,378
Middle	2.6	1,374	1.5	1,140	2.1	2,514
Fourth	2.5	1,384	2.2	1,236	2.3	2,621
Highest	6.8	1,515	3.3	1,472	5.1	2,987
Total 15-49	**3.7**	**6,917**	**2.2**	**5,690**	**3.0**	**12,607**
50-59	na	na	4.0	641	na	na
Total 15-59	**na**	**na**	**2.4**	**6,331**	**na**	**na**

Note: An asterisk indicates that a figure is based on fewer than 25 unweighted cases and has been suppressed.
na = Not applicable

- Although DHS data are available at the household level, there are no geographic coordinates acquired for individual households during the sampling process. Also, individual households are selected for their ability to represent the DHS region and the nation, not the individual household as a standalone observation. For these reasons, household-level data must have **sampling weights applied** and be **aggregated to the level of the DHS region** prior to analysis.
- DHS surveys use Global Positioning System (GPS) receivers to record the geographic **location of the approximate center of each cluster (or group) of households** surveyed. To preserve the confidentiality of survey respondents, **these cluster locations are displaced** up to two kilometers in urban areas and up to five kilometers in rural areas, with an additional one percent of rural cluster locations displaced up to 10 kilometers. Based on this displacement of cluster locations and the fact that data at the cluster level are aggregated from the household level, it is inappropriate to analyze DHS data at the single point level. (For more information, see *Geographic Displacement Procedure and Georeference Data Release Policy for the Demographic and Health Surveys*: http://dhsprogram.com/publications/publication-SAR7-Spatial-Analysis-Reports.cfm and also *Guidelines on the Use of DHS GPS Data*: http://dhsprogram.com/publications/publication-SAR8-Spatial-Analysis-Reports.cfm.)
- Indicator data provided in DHS final reports are **generally representative at the national and regional level only**.
- Indicator tables are included in the final reports, which are available in PDF format in the publications section of The DHS Program site.
    - Data in tables must be cut from the PDF reports and pasted into a spreadsheet or database to facilitate joining them to a spatial dataset, such as a shapefile at the province or district level. To view a larger image of the data provided in the image on this page, see **Table 14.5**, *HIV prevalence by socioeconomic characteristics* (source: National Institute of Statistics of Rwanda, Rwanda Ministry of Health, ICF International).

    - Extracting data from final reports in PDF format, although a manual process, can be a relatively quick way to produce custom maps that cannot be produced using other tools, such as StatCompiler, and to obtain the data needed for spatial analysis.
- Survey data sets are provided in Stata, SAS, SPSS, and flat-file format.
    - Data must be aggregated at the regional level and exported to a spreadsheet or database format that can be opened in a GIS and joined to a geographic file, such as a shapefile, prior to mapping or spatial analysis.

Working with DHS Data at the Sub-National Administrative Level

If you need to go beyond the capabilities of The DHS Program's **StatCompiler** (http://www.statcompiler.com/) and create custom or publication-quality maps or conduct spatial analysis using DHS data at the sub-national level, in addition to the DHS data, you will need a digital file that contains geographic features corresponding to the administrative level at which the data are available. For instance, if you have DHS data for HIV prevalence at the province level, you will first need a digital file that contains the administrative boundaries for provinces.

Sources for digital files containing administrative boundaries include:

- **Spatial Data Repository** (The DHS Program) http://spatialdata.dhsprogram.com/
- **Second Administrative Level Boundaries** (SALB) http://salbgeonetwork.grid.unep.ch/geonetwork/srv/en/main.home
- **Global Administrative Unit Layers** (GAUL) http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691
- **Global Administrative Areas Database** (GADM) http://www.gadm.org/

Each of these sources will be described in greater detail on the following pages.

If administrative boundary files for your country are not provided through any of these sources, you can consult the SALB site or contact your national statistics organization, census agency, or ministry of health.

Once you have located a geographic file that corresponds to the administrative divisions represented in your DHS dataset, you can use the DHS region names or codes as common geographic identifiers to join the DHS indicators to the geographic file.

You decide

Which dataset is the best? In 2011, the World Bank Mapping for Results team assessed the accuracy and geographic coverage of the SALB, GAUL, and GADM boundary datasets. The team concluded that each dataset had its strengths and weaknesses, and that a dataset should be selected on a case-by-case basis. Overall, however, the team found the GAUL dataset to have "superior accuracy and completeness, in particular in the developing countries of Africa and Latin America." See the complete paper, *Open Geospatial Data: An Assessment of Global Boundary Datasets* online here*:*
(http://www.geos.ed.ac.uk/~gisteac/proceedingsonline/GISRUK2012/Papers/presentation-35.pdf)

The DHS Program Spatial Data Repository

For DHS data, a good starting point for finding a digital file containing administrative boundaries is the **Spatial Data Repository (**http://spatialdata.dhsprogram.com/) as it provides administrative boundary files for many countries with the DHS data already attached. Files from the repository can be downloaded in either **shapefile** or **geodatabase** format, which are both readily opened in QGIS and many other GIS software packages. Take care to examine any metadata and ensure that the administrative boundaries provided match the units in your data—these are subject to change and may need to be updated.

If the HIV Spatial Data Repository does not provide a shapefile or geodatabase corresponding to your country or administrative level of interest, a shapefile or other geographic file can be downloaded from one of the other sources described on the following pages and joined to your DHS indicator data manually using a GIS.

Other Sources for Administrative Level Boundaries

Second Administrative Level Boundaries (SALB)

- Implemented by the United Nations within the context of activities related to the United Nations Geographic Information Working Group (UNGIWG).
- Provides first- and second-level administrative divisions for UN member states, which means geographic coverage is extensive but not worldwide.
- Administrative divisions have been vetted by the corresponding national mapping agencies prior to publication. This vetting process makes SALB boundary files highly reliable in terms of official representation of administrative divisions for a particular country, although

the length of the vetting process means that SALB boundary files may sometimes be out of date.
- Files are available in either shapefile (*.shp) or ArcInfo interchange (*.e00) format, and are accompanied by excellent metadata.
- Registration is required before files can be downloaded.
- Copyright of the SALB dataset is vested in the UN, and use of the dataset is restricted to non-commercial purposes.

Global Administrative Unit Layers (GAUL)

- Maintained by the Food and Agriculture Organization (FAO) of the United Nations.
- Serves as the core dataset for administrative and political boundaries available through the FAO GeoNetwork.
- Provides access to first- and second-level administrative divisions on a worldwide basis. If available, administrative divisions are also provided at the third, fourth, and lower levels.
- Although GAUL data are compiled from the best available data and most reliable sources for a particular country, the boundaries might not have been vetted and approved by authoritative national sources. For this reason, GAUL data are intended for use by the UN community and its authorized partners, and cannot be distributed to the general public. Also, the use of GAUL data should always be accompanied by a disclaimer.

Global Administrative Areas Database (GADM)

- Developed to support a variety of activities (initially the BioGeomancer Project at the University of California at Berkeley) and mapping of census-type data.
- Provides nearly worldwide coverage of first- and second-level administrative boundaries. For some countries, administrative boundaries are available at the third level or below.
- Administrative boundaries are provided in shapefile (*.shp), personal geodatabase (*.mdb), file geodatabase (*.gdb), Google Earth (*.kmz), and R (requires sp package) formats.
- Boundaries are derived from a variety of sources, both official and unofficial (e.g., Wikipedia), and are provided with almost no metadata, so they should be used with caution.
- Boundaries can be used for academic and non-commercial purposes only.

Links to the above websites are provided on page 34 of this document, in the previous section entitled: *Working with DHS Data at the Sub-National Administrative Level.*

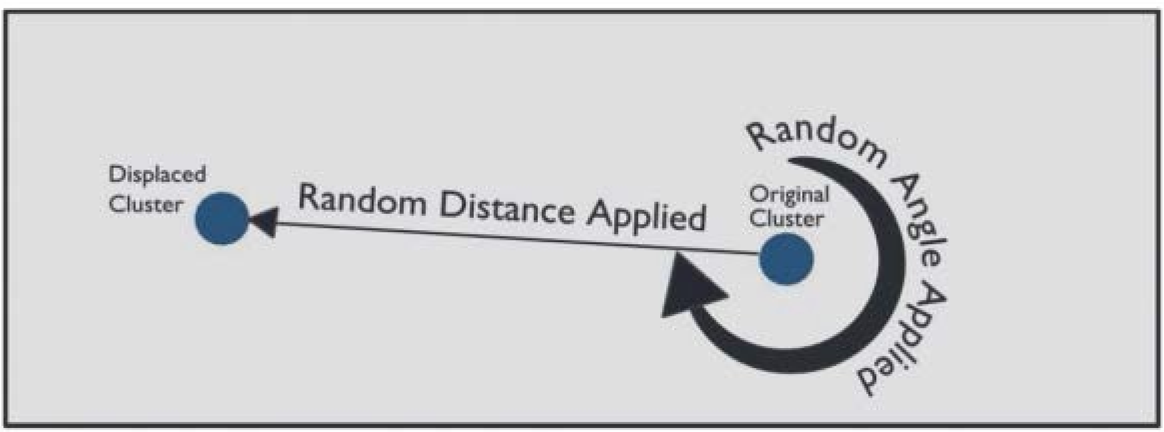
To gain practice joining DHS data to a shapefile and creating a color-shaded map by sub-national administrative areas, please complete exercises 3.1 and 3.2 at the end of this session.

Working with Cluster-Level DHS Data: Cluster Displacement

DHS cluster locations are individual points that each represent the approximate center of a group of households that participated in a particular survey. The original Global Positioning System (GPS) coordinates for clusters are collected during the household listing process as part of the sampling frame development. These raw GPS coordinates are accurate to within approximately 15 to 20 meters.

However, to ensure the confidentiality of the information provided by survey respondents, before publication the raw GPS coordinates are randomly displaced as follows:

- Up to 2 kilometers in urban areas,
- Up to 5 kilometers in rural areas, and
- An additional 1% of rural cluster coordinates are displaced up to 10 km.



The random displacement of cluster locations is controlled to keep points within the national boundary as well as within the DHS survey region boundary. Since 2009, cluster displacement is also constrained to a country's second administrative level boundaries whenever possible.

As a result of cluster displacement, using a GIS to derive measures based on precise distances between DHS clusters and other features is greatly discouraged. Instead, it is recommended to use a GIS to generate catchment areas or buffer zones around clusters and to evaluate the spatial characteristics of interest within distance bands, such as within 0 to 5 km, 5 to 10 km, etc. Because of the bias introduced by the cluster offset, it is also not recommended that DHS be used to analyze relationships between service environment of facilities and health outcomes from DHS data. See *Geographically linking population and facility surveys: methodological considerations* on the **Population Health Metrics** website (http://www.pophealthmetrics.com/content/11/1/14) for a discussion of this topic.

Cluster locations can also be used to link survey data to contextual variables, such as population or land cover. Remember that contextual variables will also need to be aggregated at the DHS region (or sub-region if data at that level are provided in the survey report) in order to match the DHS sampling design.

Highlight

GPS data collection guidance is available from The DHS Program and MEASURE Evaluation and includes:

- Guidelines on the Use of DHS GPS Data (http://dhsprogram.com/publications/publication-SAR8-Spatial-Analysis-Reports.cfm)
- Incorporating Geographic Information Into Demographic and Health Surveys: A Field Guide to GPS Data Collection (http://dhsprogram.com/publications/publication-dhsm9-dhs-questionnaires-and-manuals.cfm)
- Geographic Displacement Procedure and Georeferenced Data Release Policy for the DHS (http://dhsprogram.com/publications/publication-SAR7-Spatial-Analysis-Reports.cfm)
- Common Types of Coordinates Collected with a GPS Receiver **(**http://www.cpc.unc.edu/measure/publications/fs-13-83**)**
- An Overview of Spatial Data Protocols for HIV/AIDS Activities: Why and How to Include the "Where" in Your Data (http://www.cpc.unc.edu/measure/publications/ms-11-41a)

GPS Data File for DHS Surveys

Field Name	Description
DHSID	The 14 character DHS identification code - DHSCC & DHSYEAR & DHSCLUST (with 8 digits) from survey documentation.
DHSCC	The 2 letter DHS country code (http://www.measuredhs.com/data/File-Types-and-Names.cfm).
DHSYEAR	The 4 digit year of data collection from the survey documentation.
DHSCLUST	The integer cluster identification number. This variable will match v001 in the DHS recode file.
CCFIPS	Federal Information Processing Standards (FIPS) 2 letter country code (http://www.itl.nist.gov/fipspubs/fip10-4.htm).
ADM1FIPS	Federal Information Processing Standards (FIPS) 2 letter country code plus 2 letter/digit first sub-national administrative division code (http://www.itl.nist.gov/fipspubs/fip10-4.htm). *NOTE: If this information is not available, this field will be "NULL".
ADM1FIPSNA	Federal Information Processing Standards (FIPS) first sub-national administrative division name (http://www.itl.nist.gov/fipspubs/fip10-4.htm). *NOTE: If this information is not available, this field will be "NULL".
ADM1SALBCO	Second Administrative Level Boundaries (SALB) first sub-national administrative division code (http://www.unsalb.org).

A separate file containing the cluster locations for a DHS survey can be linked to the original survey data (using the field DHSCLUST). This cluster ID field is provided in both the **GPS data file** and the survey dataset (as field V001). The GPS coordinates will have a one-to-many relationship with the records in the survey datasets, as the survey datasets represent multiple households per cluster.

Each DHS cluster is uniquely identified using a 14-character code known as the DHSID. These 14-character codes represent a combination of the 2-letter country code, the 4-digit year of data collection, and the DHS cluster ID number (which matches field V001 in the DHS recode file).

Aside from cluster identifiers, the GPS data file contains the **latitude** and **longitude** of cluster locations. The GPS data file also designates whether a cluster is considered urban or rural, which is important for knowing what the cluster displacement might be, and identifies the name and administrative code of the corresponding DHS region. All of these identifiers can be useful for merging and analyzing datasets.

For more information on how to work with DHS data at the cluster, household, or individual level, visit **The DHS Program** (http://dhsprogram.com/) It is important to be aware of the data use limitations. Several case studies regarding these limitations can be found in *Guidelines on the Use of DHS GPS Data* mentioned earlier in this document.

The final session of this course will provide guidance on turning geographic variables (such as those from DHS data) into maps, which can help effectively communicate stories about the data. But before getting into Session 4, we have provided some optional Practical Exercises. These will give you hands-on experience with working in QGIS, using sub-country DHS data associated with administrative divisions.

In Summary

- There are several primary sources of HIV/AIDS data which have a geographic component.
- Service Provision Assessment (SPA) surveys provide information on services provided at various facilities.
- There are constraints to consider when working with Demographic and Health Survey (DHS) data in a GIS. They are generally representative only at the regional level. Sample cluster locations are displaced to ensure respondent confidentiality.
- The DHS Program's Spatial Data Repository is a good source for this data.
- It is important to take into account methodological considerations before working with the GPS data information on cluster data, due to the data's random displacement and DHS sampling design.

Practical Exercises—Session 3

In this session, you were introduced to joining data from multiple sources and finding digital files containing administrative division boundaries needed for different levels of mapping or spatial analysis. Now, we invite you to practice these skills by working through the following optional exercises using QGIS. (We estimate these exercises will take you about 45 minutes to complete.)

See Appendix 2 of this document for the instructions for these exercises. In order to complete Exercises 3.1 and 3.2, you will need access to some accompanying data files which can be accessed online.

1) You will need to **download** the files from our public DropBox folder: (http://bit.ly/Ex3DropBox). Follow the link and click on the blue Download button (files are bundled as one .zip folder) and store them on your own computer in a new folder that you create (place it wherever you like, as long as it is easily accessible; the default folder name is simply "Exercises 3.1 and 3.2").

2) **Unzip** the folder, using your operating system or a program such as WinZip, 7-Zip, or StuffIt Expander. NOTE: Leave the file structure **in place** once the main folder is unzipped. If subfolders are moved around or if the main folder has not been unzipped properly, the exercises will not perform as expected. The unzipped folder will be approximately 1.5MB in size.

By the end of these Session 3 exercises, you will be able to:

- join a text file containing HIV prevalence data from the Rwanda DHS 2010 final report to a shapefile representing the districts of Rwanda
- use this new shapefile to produce color-shaded maps showing HIV prevalence by district for Rwanda

We estimate that these supplemental exercises will take you about 45 minutes to complete and are optional for completion of the course.

Session 3: Knowledge Recap

Now that you've completed this session, test your knowledge on this subject. Taking this quiz will reinforce key points and identify gaps in learning.

After taking the Knowledge Recap, you'll get to review the correct answers, and in some cases, read an explanation.

1) The HIV/AIDS Survey Indicators Database contains primarily what type of survey?

A) Demographic and Health Survey (DHS)
B) Service Provision Assessment (SPA)
C) Multiple Indicator Cluster Survey (MICS)
D) Young Adult Survey (YAS)

2) What type of survey from MEASURE DHS provides indicators on the readiness of health facilities?

A) Demographic and Health Survey (DHS)
B) Service Provision Assessment (SPA)
C) Multiple Indicator Cluster Survey (MICS)
D) Young Adult Survey (YAS)

3) Which of the following statements are true regarding the geographic structure of some of the primary data sources for HIV/AIDS indicators?

A) DHS are designed to be representative at the national level as well as at the first sub-national level of the DHS region. Since 2009, DHS data are also reported at the second administrative level when possible.
B) Point locations for clusters of households included in a DHS are available from MEASURE DHS, and can be used for measuring precise distances between cluster locations and other phenomena, such as health facilities or schools.
C) The Service Provision Assessment (SPA) surveys from MEASURE DHS provide geographic coordinates for individual health facilities, which allows indicators to be analyzed at the facility level.
D) All of the above

4) Which dataset provides shapefiles containing sub-national administrative boundaries that have been vetted by the national mapping agencies for their respective countries?

A) Global Administrative Unit Layers (GAUL)

B) Global Administrative Areas Database (GADM)

C) Second Administrative Level Boundaries (SALB)

D) None of the above

5) Which statements are true regarding cluster-level data from MEASURE DHS?

A) Cluster locations have been randomly displaced up to 2 kilometers in urban areas and up to 5 kilometers in rural areas, with an additional 1 percent of rural clusters displaced up to 10 kilometers.

B) Cluster locations in the geographic file for a DHS dataset can be linked to the survey data using a unique cluster identifier.

C) Cluster locations can be used to link survey data to broad contextual variables, such as population or land cover.

D) All of the above

ANSWERS:

1) (a) Demographic and Health Survey (DHS)

2) (b) Service Provision Assessment (SPA) FEEDBACK: SPA surveys from MEASURE DHS assess the availability and readiness of facility-based health services in developing countries. SPA indicators cover the following areas: infrastructure, resources, and systems; child health; maternal and newborn health; family planning; HIV/AIDS; sexually transmitted infections (STIs); malaria; tuberculosis; basic surgery; and non-communicable diseases.

3) (a) DHS are designed to be representative at the national level as well as at the first sub-national level of the DHS region. Since 2009, DHS data are also reported at the second administrative level when possible. FEEDBACK: household cluster locations are randomly displaced for privacy purposes and thus do not provide precise distance measurements. SPA surveys do not allow for facility level analysis.

4) (c) SALB -- SALB second-level administrative boundaries from the United Nations have been vetted by the national mapping agencies involved in their creation. As a result, they are of very high quality with respect to other data sources. One potential concern for using SALB boundaries is that they are maintained primarily for UN member states, so their geographic coverage is not worldwide. Another potential concern is that the length of time required for vetting boundaries prior to posting them online can be significant. This means that SALB boundaries can sometimes be out of date.

The Global Administrative Areas Database (GADM) provides boundaries from a variety of sources, both official and unofficial, and provides very little metadata. As a result, although GADM provides boundaries on a worldwide basis, these boundaries should be verified independently and used with some caution.

The Global Administrative Unit Layers (GAUL) dataset provides first- and second-level administrative divisions on a worldwide basis, but the source data for a particular country may not have been vetted by a national mapping agency. As a result, GAUL data are intended for use primarily by UN organizations and their partners and cannot be distributed to the general public. Also, GAUL data should always be used with a disclaimer.

5) (d) All of the above.

Session 4: Essential GIS Techniques

There are some essential GIS techniques that are useful for M&E of HIV/AIDS programs. Using QGIS, new indicators can be calculated within a shapefile. QGIS provides several options for mapping more than one indicator at a time.

There are standard cartographic elements and output considerations that should be included when constructing a publication-quality map.

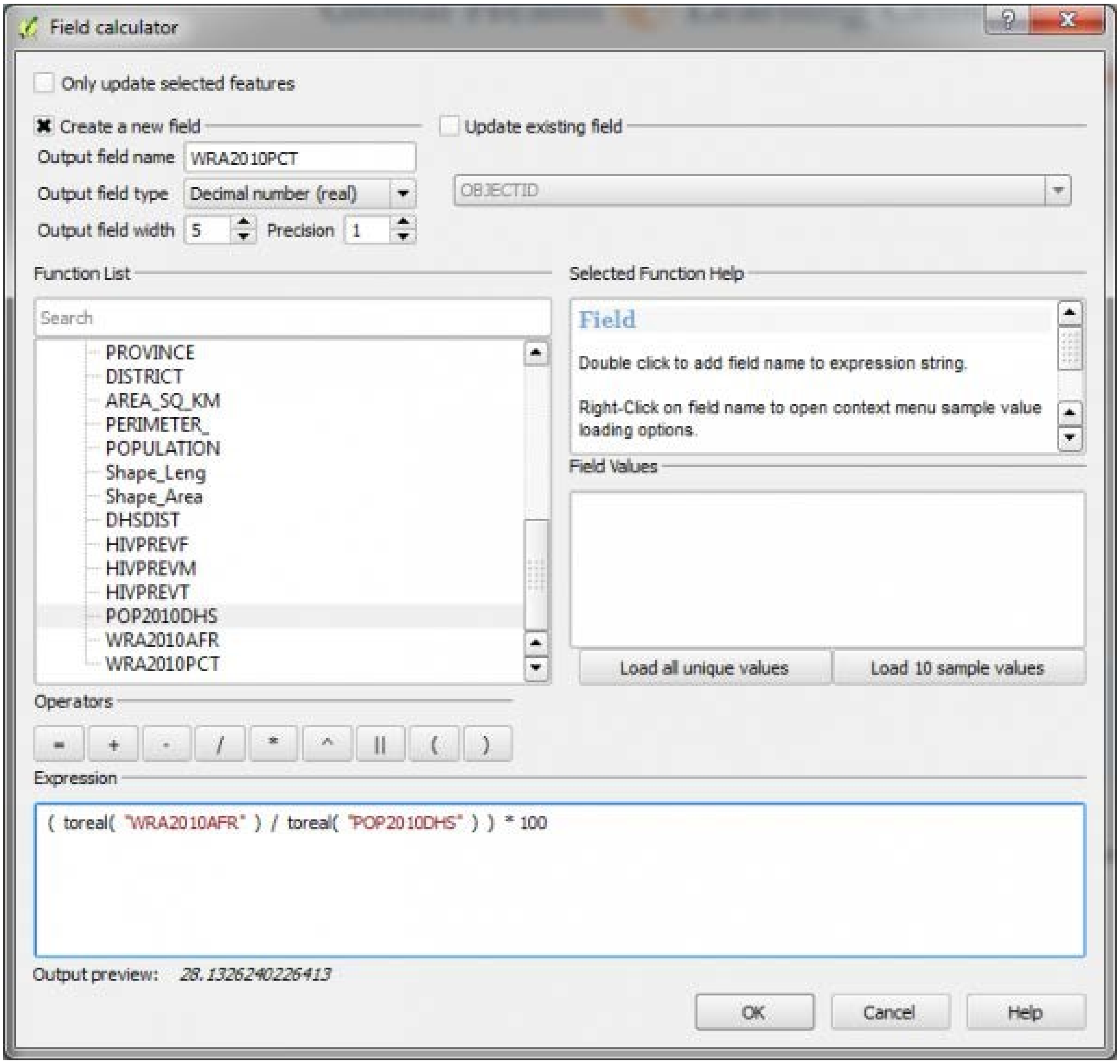
After completing this session - especially the practical exercises at the end of this session, learners will be able to execute a set of GIS tasks that are considered essential for M&E of HIV/AIDS and related programs:

- Calculate a new variable from existing data
- Display two variables simultaneously on a map
- Generate a publication-quality map

Creating a New Variable within a GIS

One of the strengths of GIS is its **data management capabilities**. The software's ability to create new variables is one of its key functions.

There are multiple approaches that users can take to create a new variable. One way is through the use of basic arithmetic operations such as addition, subtraction, and multiplication to create variables. An example of this might be to calculate HIV prevalence by dividing the variable in an attribute table that represents number of HIV positive people in a district by the variable in a table that represents the total population in a district and then multiplying the result by 100.



Another way is to use spatial characteristics to create a new variable. An example of this might be using the GIS to calculate distance from a village to the nearest health facility and then storing that result as a variable.

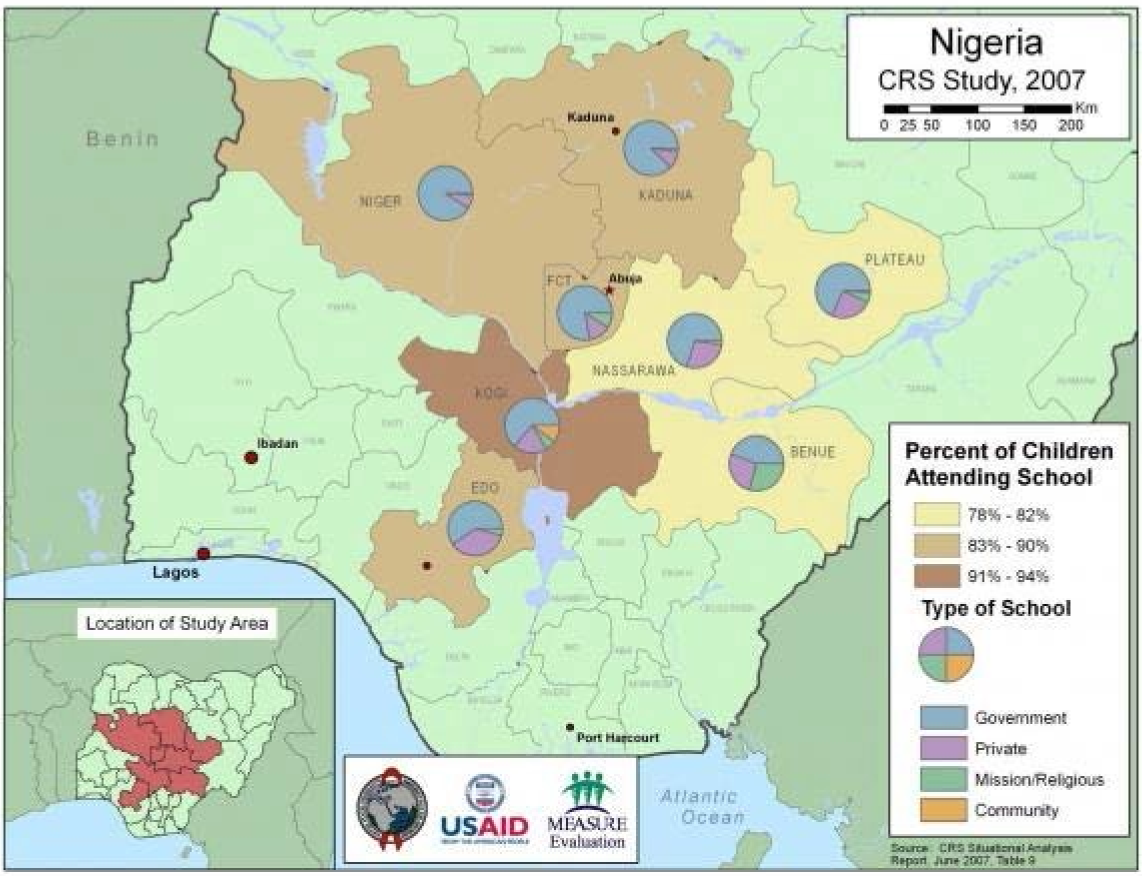
QGIS provides users with a formula builder to help with constructing complex formulas that can employ **Boolean logic** or include characteristics of geographic elements such as area or length. It is possible to then save such a formula for later use.

Maps of Multiple Indicators

Maps are an efficient and effective way to display data. Including multiple variables on a map can increase its usefulness.

Multiple variables can show relationships between multiple datasets. For instance, one map that shows both the presence of ART in facilities and the percentage of persons in associated districts living with HIV can make it easy to see patterns about ART services and whether they are in the areas of greatest need. If that information was displayed in two separate maps, the patterns could be much more difficult to see.

There are several ways to include multiple variables. Two of the most common are discussed on the following pages.



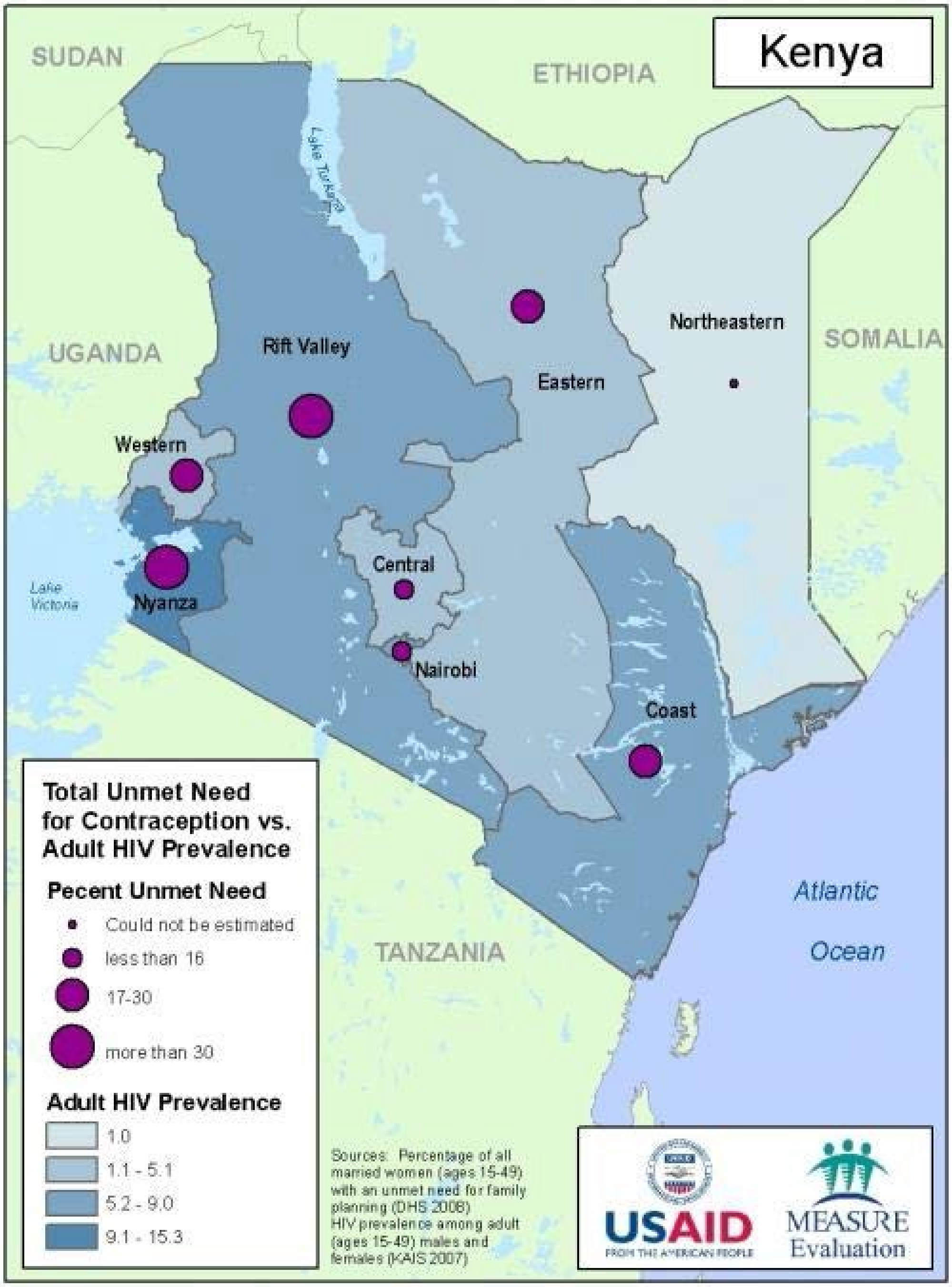
Using Multiple Types of Geographic Primitives

Data can be represented as one of three **geographic primitives**: points, lines, and polygons. Including multiple types of primitives is one of the easiest ways to produce maps of multiple variables.

Common examples of this approach are maps that show data for a facility or a DHS cluster represented by points overlaid on polygons that represent a district or the boundaries of a city. The polygons can then be shaded as in any normal **choropleth map**, and the color or size of the points can vary according to the data values. Lines representing roads or rivers can also be presented with various colors, thicknesses, or line styles (e.g., dashed or dotted lines) to indicate data about the feature, such as the type of road or width of the river.

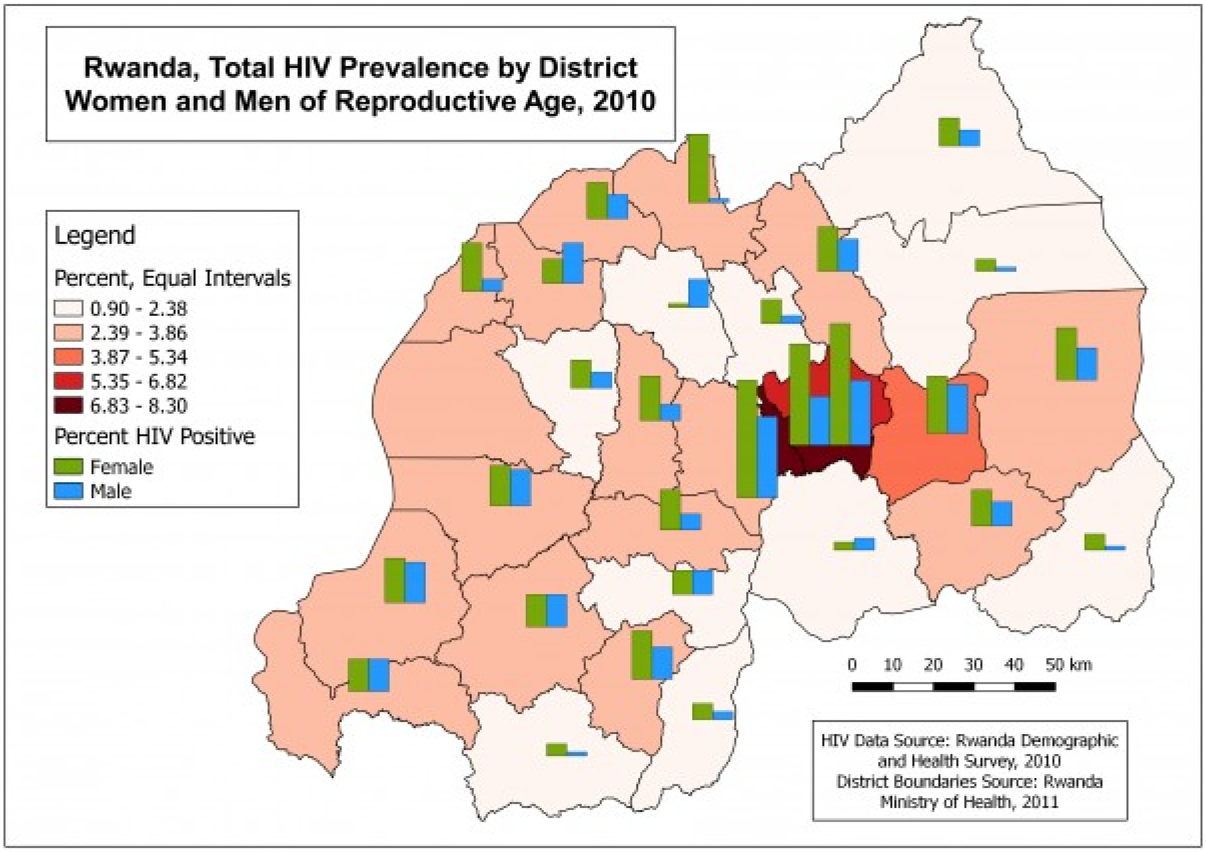
Sometimes, it is desirable to display multiple variables corresponding to polygons. The best way to do this is to create a file of the polygons' centroids and then produce a map that shows both the polygons themselves and the points representing the centroids. The symbology and shading of these centroid symbols can be adjusted as needed based on the desired representation of data.

It is also possible to present points or lines on top of raster grids. An example of this would be a point representing a village overlaid on a raster surface representing facility service areas.



Using Polygons and Charts

QGIS and many other GIS programs allow the user to produce maps that present attribute data for points as pie or bar charts. These can be overlaid on top of polygons. An example of this would be a bar chart showing the OVC population at differing age groups overlaid on polygons shaded to represent HIV prevalence.



It is important to consider several things when producing maps that use charts to represent data:

1. Whether the data can be effectively represented by certain types of charts. Pie charts should be used to show percentages of data and bar charts to show actual numbers.
2. Making sure the charts are readable. A high number of charts in a small area can cause the charts to be overlaid on one another, obscuring each other and also the map beneath. Bar charts and pie charts should have a minimal number of categories and should be shaded distinctly.
3. Many GIS programs do not have the capability to include axis labels for bar charts. This means that readers may not have an idea of the values associated with the bars. It is important to include this information in the map's legend to help readers more effectively interpret the map.

Producing effective multiple indicator maps can be challenging. If the wrong type of multiple indicator map is produced or the data is displayed in an unclear manner, the map can be hard to interpret or possibly be misleading.

Maximizing a Map's Utility

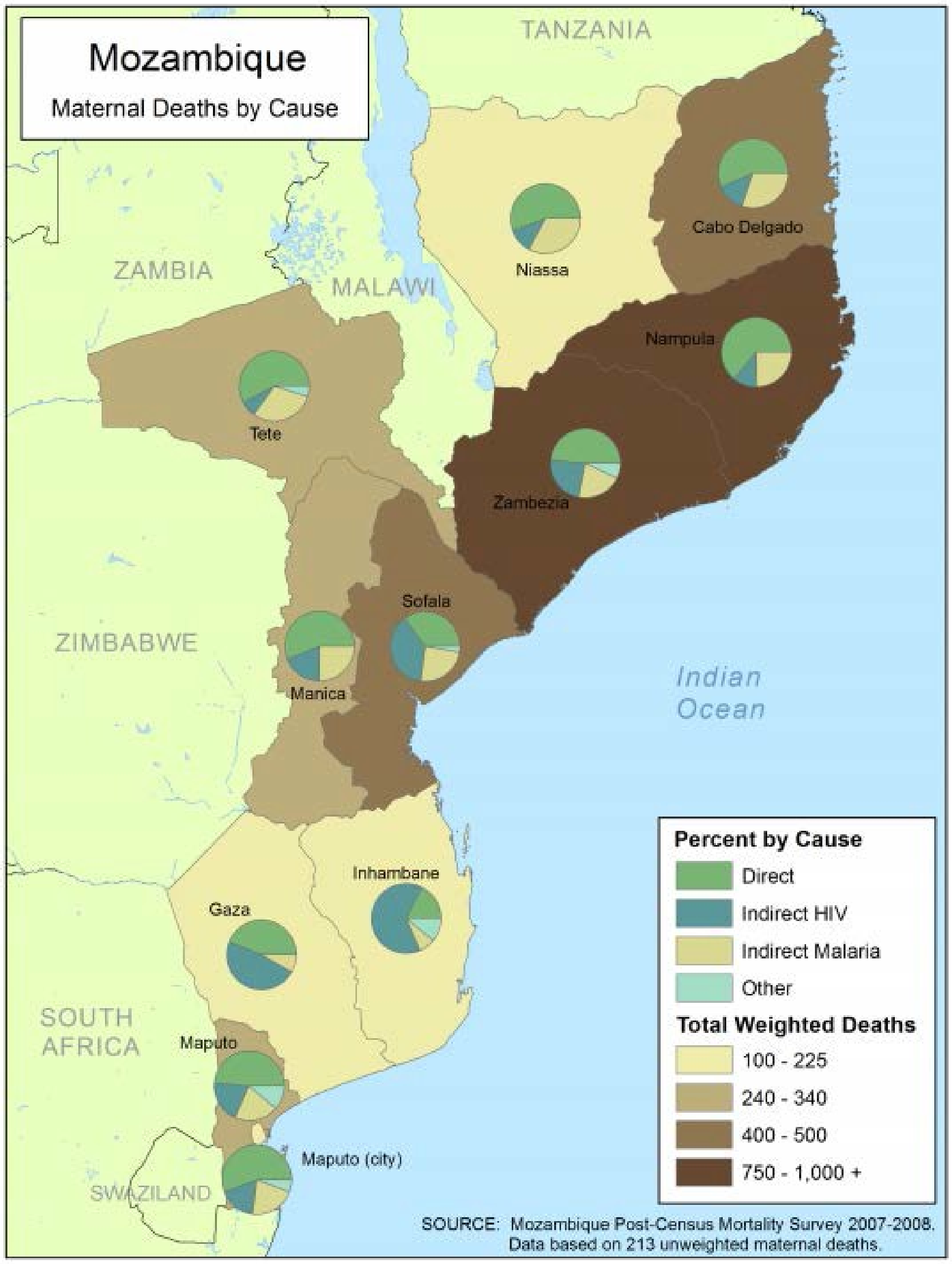
A map should display data in a way that is easy for the reader to understand and interpret. There are steps the map-maker can take to ensure the map can be most useful.

The first is to select classification breakdowns and colors that are easily understood and appropriate for the method used to present the map.

- Selecting an appropriate color or shading for data and features of the map is important to ensure readability and to allow the data to tell a story.
- If a map is to be printed in black and white (as opposed to printed or displayed in color,) the map-maker should either use colors that can be differentiated when printed in black and white, or use shades of gray.

Choosing colors for a map deserves some extra consideration. Readers should be able to easily differentiate between colors on a map—usually not more than seven or eight shades in total. Cartographic tradition dictates that forested areas are often green and water is often blue. Realize that red can often be interpreted as hot, while blue can be interpreted as cold. Increasing the intensity or darkness of the colors can convey increasing values in the data. Do not show a rainbow of colors if you are trying to indicate a simple gradation of values. The Colorbrewer website (http://colorbrewer2.org/) offers more help and tools for choosing map colors.

For more on selecting classification breakdowns, see the following page.



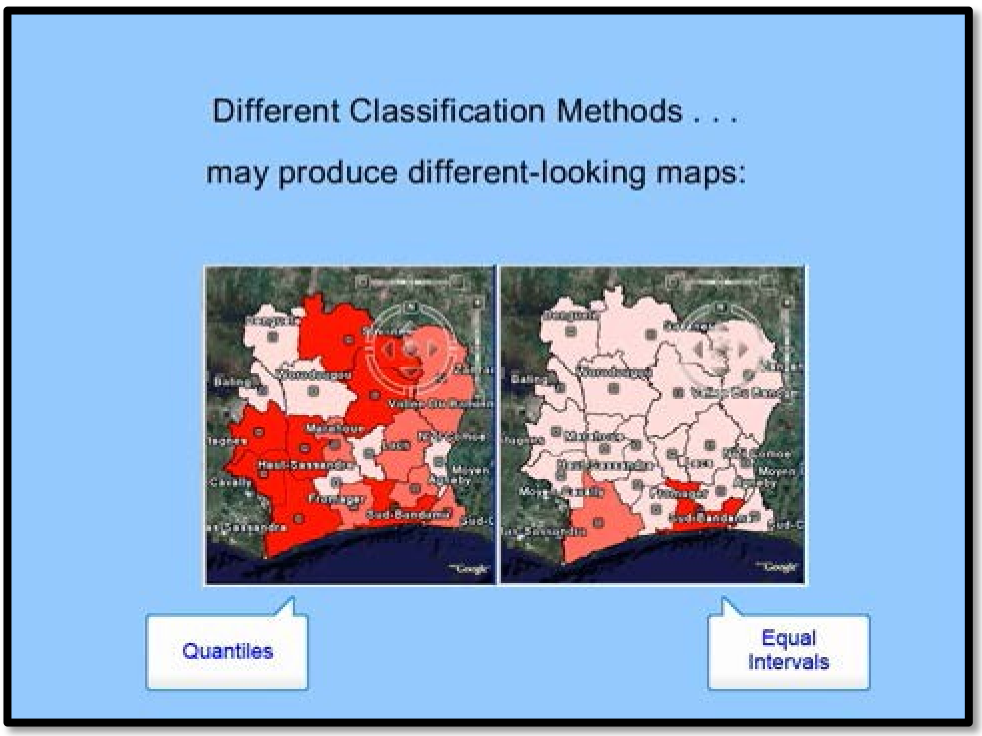
Classification

Data classification refers to the way that the GIS breaks the data into ranges in order to display it on a **choropleth map**. Changing the classification can mask or highlight patterns in the data. GIS typically offers several different ways to classify the data, including the ability to create custom classifications. The most common classifications include:

- **Natural breaks** is a classification approach that breaks the data into categories that minimize variation of each category. A statistical approach known as Jenk's optimization is used by the GIS to create the categories. The natural breaks method is well suited to many different types of data and can be a good first option to try when mapping data (see the following page on the Jenks Natural Breaks Classification: http://wiki.gis.com/wiki/index.php/Jenks_Natural_Breaks_Classification).

- **Quantiles** breaks data into categories with the same number of observations. The number of observations in the category will depend on the number of quantiles the map-maker selects and the total number of records in the dataset. Quantile-based classification is well suited for data with a linear distribution, having a limited number of records with similar values (see video).
- **Equal-interval** breaks the data into equal-sized slices, for instance 1-10, 11-20, 21-30, etc. Equal-interval is useful for emphasizing areas relative to other areas (see video).

A 3-minute video (link is embedded in online version of course and also given below) from MEASURE Evaluation gives more information on quantile vs. equal-interval classifications:

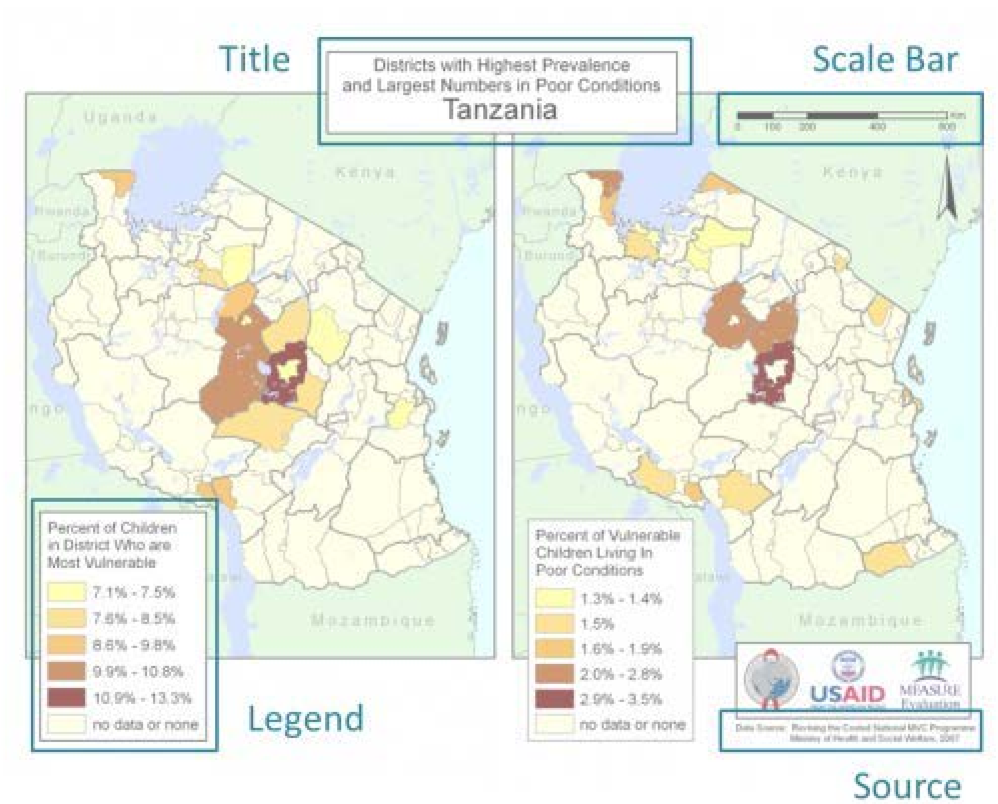


Link to Video from MEASURE Evaluation Geospatial Group, “Comparing Data Classification Methods”: https://www.youtube.com/watch?feature=player_embedded&v=-_M-VUVBGjE

Other Design Aspects to Consider

There are other aspects of the map a map-maker should consider.

- Most maps should include a **legend** that makes it possible for the map reader to understand the symbology and classifications used.
- The map should include relevant information about the **source of the data**.
- The map-maker should provide a **title** and text that clearly identifies the topic of the map and the geographic area being mapped.
- Lastly, some indication of a **map's scale** should be included on the map if distance is an important element of the map or if the reader may not be able to determine distances represented on the map. A map's scale can be represented by a scale bar or a representative fraction. (Recall that you may need to take map projection into account before generating a scale bar.)



In Summary

- Data management and variable creation are key strengths of a GIS.
- Maps are useful for displaying multiple variables at once, which can help in both asking and answering questions about data.

- In designing a map, the author should consider appropriate data classification techniques. Color and pattern choices should be based on cartographic traditions and readability. A legend (description of symbols and colors used) and data source and date are also important.

Practical Exercises—Session 4

Now that you have learned about the essential GIS techniques, we invite you to practice applying what you have learned by working through these optional practical exercises. We estimate that this set of exercises will take approximately 90 minutes.

See Appendix 2 of this document for the instructions for these exercises. In order to complete these exercises, you will need access to some accompanying data files.

1) You will need to **download** the files from our public DropBox folder (http://bit.ly/Ex4Dropbox). Follow the link and click on the blue Download button (they will come together in one .zip folder), and store them on your own computer inside a new folder that you create, in an easily accessible location. You may name this folder whatever you wish; the default name is simply "Exercises 4.1, 4.2, and 4.3".

2) **Unzip** the folder, using your operating system or a program such as WinZip, 7-Zip, or StuffIt Expander. NOTE: Leave the file structure **in place** once the main folder is unzipped. If the folder has not been unzipped properly, or files have been moved, the exercises will not perform as expected. The unzipped folder will be approximately 3MB in size.

By the end of these Session 4 exercises, you will be able to:

- Calculate new variables from existing data stored in a shapefile. New variables can be used to visually and mathematically analyze many types of spatial data, and thus provide highly valuable information for program decision-making.
- Generate multi-indicator maps. You will learn how to use pie charts, bar charts, and proportional symbols for one variable, and superimpose these on a choropleth map showing the geographic distribution of a second variable.
- Use the map composer interface within QGIS to create a publication-quality map.

We estimate that these supplemental exercises will take about 90 minutes to complete and are optional for completion of the course.

Conclusion

GIS allows M&E professionals to manage, display, and analyze data in ways that make public health data come alive visually. Maps often spark new questions about the data and can even reveal some answers as well. A GIS allows new variables to be created that might be used to better analyze the data. A GIS can also make additional methods of data analysis possible, based on geographic attributes such as proximity or density.

Another benefit to putting data into a GIS is that its **data schema** requirements encourage good data practices and allow data in multiple data sets to be joined together. This data-linking capability, based on common geographic identifiers, is an important and useful aspect of GIS. As the NSDI (National Spatial Data Infrastructure) of many countries expands, this capability becomes more significant.

In addition to geographic data becoming more accessible, the tools to work with it are proliferating and becoming easier to use. The use of open source GIS is increasing rapidly, along with a growing online user community and comprehensive and more user-friendly (and multi-lingual) documentation. Like any powerful software, there is a learning curve with GIS. Effective use requires real effort and practice. But the benefits can far outweigh the difficulties, making learning GIS well worth this effort.

Session 4: Knowledge Recap

1) Which of the following is NOT an important consideration when choosing colors on a map?

a) Considering whether the map will be reproduced in color or black-and-white.

b) Choosing shades that reflect which values on the map are highest and which are lowest.

c) Use of traditional cartographic conventions for colors, such as blue to denote water.

d) Using a wide variety of colors and patterns to make the map interesting.

e) Using a limited number of colors or shades which can all easily be differentiated from one another.

2) What is a map called that uses color shading to show higher or lower values for areas such as districts or provinces?

a) Shaded relief map

b) Isopleth map

c) Area map

d) Choropleth map

e) None of the above

3) A choropleth map will be more meaningful if it portrays ratio (percentage or density) values, rather than raw counts.

True

False

4) The term "geographic primitives" refers to:

a) Vector Data

b) Raster Data

c) Points, lines, and polygons

d) Crude symbols used on a map

e) A and C

f) All of the above

5) The most common methods of data classification for mapping are:

a) Equal intervals, quantiles, and natural breaks

b) Standard deviation via histogram

c) Quartiles and quintiles

d) High, medium, low

6) One of the features of a GIS is the ability to create new variables using a formula builder that applies Boolean logic and can perform numeric calculations. Why is this ability important?

a) The user may want to map ratio (percentage) rather than nominal (count) variables and may need to create them.

b) The user may want to create an ID code field rather than use geographic area names, for purposes of joining to another table.
c) The user may want to combine several variables into one summary variable by adding or averaging them.
d) All of the above
e) None of the above

7) One way to study epidemiology in relation to intervention capacity within a country might be to show more than one variable on a map at once, by using pie charts showing percentages of children having several common types of disease, on top of shaded areas showing ratios of hospitals to population by district.

True
False

ANSWERS:

1) (d) Using a wide variety of colors and patterns to make the map interesting. (Too many colors and shades can actually compete for the reader's attention and make a map's message harder to understand. It is important not to choose colors and shades in a random fashion and not to use too large of a variety of shades and types of symbols, which might be hard to discern from one another and ultimately create confusion.)

2) (d) choropleth map

3) True. Raw counts (nominal values) are not the best thing to show as shaded areas. A larger area with a larger population does not tell you much, but will show up as "artificially" important on a map shaded in this manner. It would be much more appropriate, on this type of map, to show a computed ratio value, such as population density (number per square mile) or percentage of the population with AIDS.

4) (e) A (vector data) and C (points, lines, and polygons)

5) (a) Equal intervals, quantiles, and natural breaks

6) (d) All of the above

6) True. Maps are an efficient and effective way to display data, and using multiples variables on a map can increase its usefulness.

GIS TECHNIQUES FOR M&E OF HIV_AIDS AND RELATED PROGRAMS

APPENDIX 1: GLOSSARY

A

AIDS Indicator Surveys

AIS, which are modeled on the larger Demographic and Health Surveys (DHS) from The DHS Program, are focused on providing standardized indicators in a timely and cost-effective manner for M&E of national HIV/AIDS programs. There are two questionnaires administered during an AIS, one for households and one for individuals. Sample design at a minimum ensures representativeness at the national level and for urban and rural areas, although it can sometimes be adjusted to represent regional variations as well. Source: http://dhsprogram.com/.

B

Behavioral Surveillance Surveys

BSS, which were developed by Family Health International (FHI), were designed to identify patterns and trends in HIV-related risk and protective behaviors. BSS sample design can focus on describing HIV risk and response in relation to the full adult population for an entire country or target the understanding of sub-populations, such as youth, female sex workers, men who have sex with men (MSM), and injecting drug users, in key geographic locations within a country. Source: www.popcouncil.org/Horizons/ORToolkit/AIDSQuest/summaries/ssfhibss.html.

Boolean Logic

A type of data-searching statement which uses the operators “and”, “or”, and “not”. It can be used for selecting out certain groups of data, such as all the data which is contained in a certain province, or all the data for women of child-bearing age.

C

Choropleth map

A map that uses colors or shades to represent data on a map. The colors are generally portrayed in such a way as to depict increasing or decreasing values of the data within districts or provinces.

D

Data schema

The structure of a data table or larger group of inter-related tables (database). In other words, how the data fields and tables are arranged. Data schema concerns both the contents and naming and coding conventions of the data and the relationships between data tables.

Datum

A set of control points, which are points on the surface of Earth with known locations, and a corresponding mathematical model used to approximate the shape of Earth and to calculate the location of any given point on that shape. The datum used for the Global Positioning System is known as the World Geodetic System 1984 or WGS 84.

Demographic and Health Surveys (DHS)

Demographic and Health Surveys (DHS) are nationally-representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition. Standard DHS surveys have large sample sizes (usually between 5,000 and 30,000 households) and are typically conducted every five years to allow comparisons over time. The DHS Program captures survey locations based on "clusters", rather than individual households. The DHS Program provides an inventory of available survey datasets at http://dhsprogram.com/data/available-datasets.cfm. Source: http://dhsprogram.com/, accessed June 2010.

DHS recode file

DHS datasets contain thousands of variables with short names and labels. Each recode file provides more detailed descriptions of each variable and identifies the sample weights that must be applied before any analysis can be conducted. For more information on getting started with DHS data analysis and to access the **DHS recode manual**: http://dhsprogram.com/data/Using-Datasets-for-Analysis.cfm. It is strongly recommended that this manual be read before working with any DHS dataset, in order to prevent misinterpretation of the data.

E

Epidemiologic data

Data which is used to study the patterns and causes of health and disease in populations.

Exploratory spatial data analysis (ESDA)

Exploratory Spatial Data Analysis, a method of summarizing and viewing geographic data using a variety of statistics-based data visualization techniques, such as histograms, scatter plots, cluster/hot spot maps, box plots, box maps, and standard deviation maps. Advanced ESDA techniques require an advanced knowledge of statistics and may also require some programming skills.

G

Geodatabase

A newer type of geographic data storage invented by ESRI (Earth Systems Research Institute) in the early 2000s. A geodatabase takes on the appearance of a folder filled with many files, when viewed in Windows, and can thus be easily moved or copied. It can be comprised of various types of geographic data sets which follow specific rules which can be useful for editing the data and for performing advanced analysis. It is the main type of data currently used in ArcGIS, and its contents can be created and viewed by using an ESRI file management system called ArcCatalog.

Geographic coordinate system (GCS)

A coordinate system based on a three-dimensional, spherical surface. As a result of being defined in relation to the more natural, three-dimensional surface of a globe, a GCS is considered to be "unprojected" rather than "projected." For an explanation of geographic versus projected coordinate systems, see Session 3.

Geographic data

Information describing the location and attributes of things, including their shapes and representation. Source: http://resources.arcgis.com/glossary/term/520, May 2010.

Geographic identifier

A geographic identifier is any piece of information that indicates the geographic or spatial location of features on the landscape, such as latitude and longitude, street address, or administrative division name (i.e., province, district, county, etc.). Common geographic identifiers play a critical role in joining data from different sources.

Geographic Information Systems (GIS)

A system that captures, stores, analyzes, manages, and presents data that are linked to a geographic location.

Geographic primitives

A graphic representation of a location; for example, a point to represent the location of a smokestack, or a polygon to represent the location of a toxic plume.

GeoJSON

GeoJSON is an open source format for exchange of geographic data, primarily via the Web, based on JavaScript Object Notation (JSON). The GeoJSON format is supported by a growing number of GIS software packages, including the Geospatial Data Abstraction Library (GDAL), which is integrated into QGIS via a plug-in, and the DHIS 2. For more information on GeoJSON, see http://geojson.org/geojson-spec.html.

GPS data file

GPS stands for "Global Positioning System". A GPS data file is included with each DHS data cluster and contains a cluster identifier, latitude and longitude coordinates, and includes information for altitude and datum (the Earth model the coordinate system is based on: WGS84). For more detail on common types of GPS data, please download this 2-page fact sheet: http://www.cpc.unc.edu/measure/publications/fs-13-83.

K

KML

KML, which originally stood for Keyhole Markup Language, is an XML-based file format that can incorporate descriptive text, image links, and geographic information associated with points, lines, and polygons. It is an open standard officially named the OpenGIS KML Encoding Standard (OGC KML). KML files can be read by Google Earth and several mapping software packages. For a KML tutorial, see https://developers.google.com/kml/documentation/kml_tut.

L

Latitude

Angle between a line connecting the center of the Earth to the equator and a line connecting the center of the Earth to a point on the Earth's surface on, north, or south of the equator along a line of longitude. Latitude ranges from 0 degrees at the equator to 90 degrees at the poles. Latitude is positive north of the equator (0 to 90 degrees) and negative below it (0 to -90 degrees). Lines of constant latitude can be visualized as circles drawn around the Earth horizontally in parallel with the equator.

Longitude

Angle between (a) a line connecting the center of the Earth to the equator at a prime meridian, such as the meridian that passes from pole to pole through Greenwich, England (also known as the Prime Meridian or Greenwich Meridian), and (b) a line connecting the center of the Earth to the equator at its intersection with a meridian that passes through the point of interest. Longitude ranges from 0 degrees at the Prime Meridian to 180 degrees along the meridian on the opposite side of the Earth. The 180th meridian roughly parallels the International Date Line, where the date changes as travelers cross going east or west. Lines of constant longitude can be visualized as half circles drawn on the Earth's surface vertically from pole to pole.

M

Metadata

Data about data. Metadata usually contains information about when a dataset was collected or created and by whom. Geospatial metadata also contains information about projection and datum. Sometimes metadata is stored as part of the main data file, but it can also be in its own separate file. As a best practice, metadata should be provided with any geographic dataset. International standards for geographic metadata are available as ISO 19115.

Multiple Indicator Cluster Surveys

MICS are household-level surveys developed by the United Nations Children's Fund (UNICEF) to fill gaps in the data required for M&E of programs focused on meeting the critical needs of women and children. MICS indicators fall into four main categories: health, education, child protection, and HIV/AIDS. MICS results can be used to guide policy development and target interventions on behalf of women and children. MICS indicators can be mapped online at the country level using the MICS Compiler (see www.micscompiler.org). Source: www.childinfo.org/mics.html.

N

National spatial data infrastructure (NSDI)

A country's NSDI refers to the technology, policies, and people necessary to promote sharing of geospatial data throughout all levels of its government, private and not-for-profit sectors, and academia. It ideally provides a structure that facilitates geographic data sharing and use. (Adapted from http://www.fgdc.gov/nsdi/nsdi.html, accessed November 2013.)

P

Projection

A process in which locations on the three-dimensional surface of the Earth are transformed onto a flat, two-dimensional surface for display, measurement, or other analysis. Any projection method will in some way ultimately compromise either distance, direction, or shape, but is a necessary step if straight-line distances or 2D (flat) areas are to be calculated during spatial analysis. A common type of projection is UTM (universal transverse mercator), in which distances can be measured in meters.

R

Raster data

Spatial data stored in a computer as a series of values in a grid pattern (pixels). This type of data generally requires much more computer storage space than vector data. Larger numbers of pixels over a smaller area provide greater spatial resolution but take up much more memory. This type of data can show continuous change over a surface, such as land cover. Satellites collect data in this format.

Reproductive Health Surveys

RHS are reproductive health surveys conducted around the world with the assistance of the U.S. Centers for Disease Control (CDC). An inventory of RHS can be found at www.cdc.gov/reproductivehealth/Global/surveys.htm. RHS results are summarized on the Global Health Data Exchange (GHDx, http://ghdx.healthdata.org/), which is maintained by the Institute for Health Metrics and Evaluation. As indicated in the RHS inventory available on the CDC site, some RHS indicators can be mapped at a sub-national level using the StatCompiler site from The DHS Program.

S

Scale

The Earth is too large to draw on a map without reducing its size. This reduction is expressed as map scale, which is the ratio of the distance on a map to the actual distance on the surface of the Earth. As a result, a small-scale map displays a small amount of detail, but covers a large geographic area. A large-scale map shows a large amount of detail, but for a small area. Scale can be expressed graphically as a scale bar, or in writing using text or numeric forms: (i) Text: 1 inch = 24,000 inches OR 1 inch = 2,000 feet; or (ii) Numeric: 1:24,000.

Sexual Behavior Surveys

The HIV/AIDS Survey Indicators Database from The DHS Program (http://hivdata.dhsprogram.com/start.cfm) contains three SBS for Zambia, which were designed to provide national level estimates of a number of key indicators for monitoring progress of the national program for HIV/AIDS and sexually-transmitted diseases (STDs). The surveys provide indicators on HIV/AIDS; STD-related knowledge, attitudes, and sexual behavior; orphans and vulnerable children (OVC); and assistance to households and communities affected by the HIV/AIDS pandemic. The national SBS estimates for Zambia can be disaggregated by urban versus rural residence, age group (adolescents, youths, young adults and adults), and sex (male versus female), *but cannot be mapped and analyzed based on sub-national administrative divisions such as provinces or districts*. Source: Central Statistical Office (CSO), Ministry of Health (MOH), University of Zambia, and MEASURE Evaluation, 2010. *Zambia Sexual Behaviour Survey 2009*. Lusaka, Zambia: CSO and MEASURE Evaluation (www.cpc.unc.edu/measure/publications/tr-10-73).

More Zambia SBS are available in the publications section of the MEASURE Evaluation site (see www.cpc.unc.edu/measure/publications, keywords “Zambia sexual survey”).

Shapefile

Shapefiles were invented by ESRI (Earth Systems Research Institute) in the early 1990s, and have become widely used as a type of vector GIS data, especially in the open source world. A shapefile is in reality a collection of at least three separate files, all with the same root name but different file extensions (.shp, .shx, .dbf, .prj), which combine to describe vector (points, lines, and polygons) features of interest. Special GIS software can read these as one “shapefile”.

V

Vector data

A type of data composed of points, lines, and polygons. The most common form of vector data in geographic format is a shapefile. This type of data generally requires less storage than raster data.

Y

Young Adult Survey

The HIV/AIDS Survey Indicators Database from The DHS Program (http://hivdata.dhsprogram.com/start.cfm) contains one YAS for the country of Indonesia. The YAS includes indicators for M&E of HIV/AIDS-related knowledge, attitudes, and behavior. Appendix A contains indicators reported by age and sex that can be mapped and analyzed at the provincial administrative level. Source: Badan Pusat Statistik (BPS)-Statistics Indonesia and Macro International. 2008. *Indonesia Young Adult Reproductive Health Survey 2007*. Calverton, Maryland, USA: BPS and Macro International (http://dhsprogram.com/pubs/pdf/FR219/FR219.pdf).

GIS Techniques for M&E of HIV_AIDS and Related Programs

Appendix 2: Practical Exercises

NOTE: The actual software and data files for each exercise set are available online, and must be downloaded separately for use with these instructions. Links to the data files are provided in the main body of this document, within the introductory explanations for each session's "practical exercises", which occur at the end of each session, just before the quizzes.

The remainder of this appendix, comprising approximately 100 pages in total, includes the instructions for the exercises for Session2, Session 3, and Session 4.

GIS Techniques for Monitoring and Evaluation of HIV/AIDS and Related Programs

Exercise 2.1

Install QGIS

**This training was developed as part of a joint effort between MEASURE Evaluation and MEASURE DHS, with funding provided by USAID and PEPFAR.*

Exercise 2.1: Install QGIS

Summary: In this exercise, you will learn how to install QGIS for Windows and verify the plug-ins for extending the functionality of the software. The instructions can vary based on your operating system.

Objectives:

- Install QGIS.
- Launch the program.
- Verify which plugins are **installed**.
- Verify which plugins are **activated**.

NOTE: These procedures will require internet access.

Step 1: Install QGIS

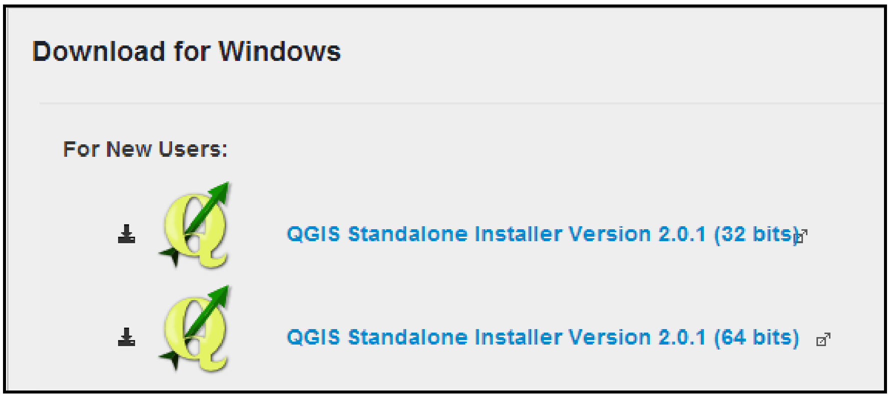
- ☐ <u>NOTE:</u> **You will need administrator rights to install QGIS.** You may need to contact your computer administrator for help with this installation.

- ☐ Begin the installation process by opening a Web browser and visiting the QGIS site http://www.qgis.org/.

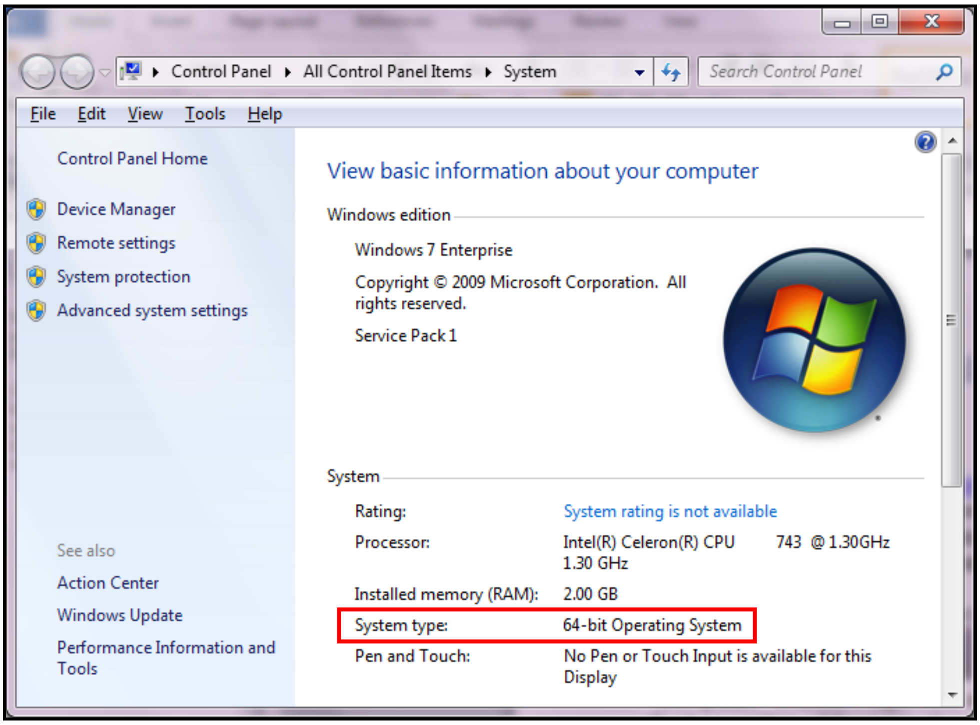
- ☐ On the home page, click on the 

button near the bottom of the page and download the installation file for the most current version of QGIS, which at the time of writing is QGIS 2.0.1 Dufour.

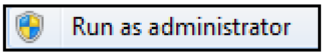
  Note that on the download page there are two different versions of the standalone QGIS installation file for Windows, one for 32-bit operating systems and one for 64-bit operating systems (see below).



- ☐ To verify whether your computer uses a 32- or 64-bit operating system, left-click on the Windows Start button  with your mouse, right-click on the option “Computer” and choose Properties (or select Control Panel and then System), and look for the system type information in the System section of the properties window (see below).

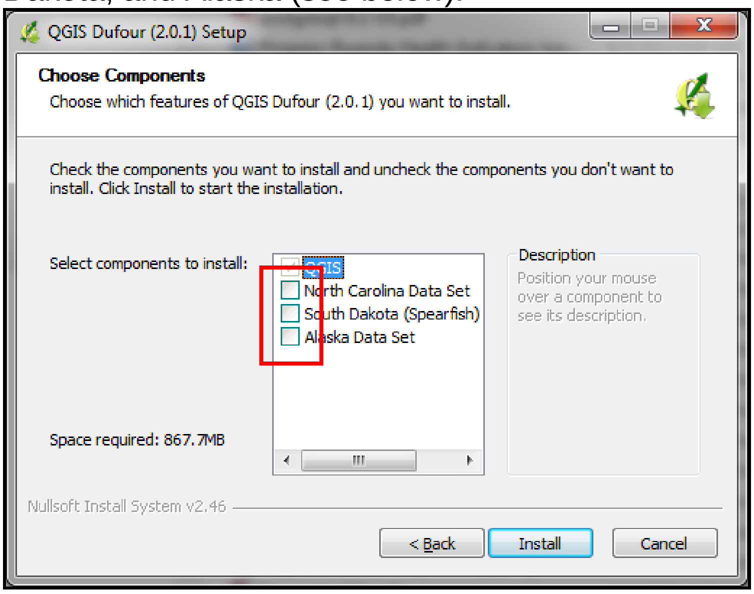


- To begin downloading the installation file for your operating system, click on the appropriate link. Once the file download is complete, you will need to run the installation file.

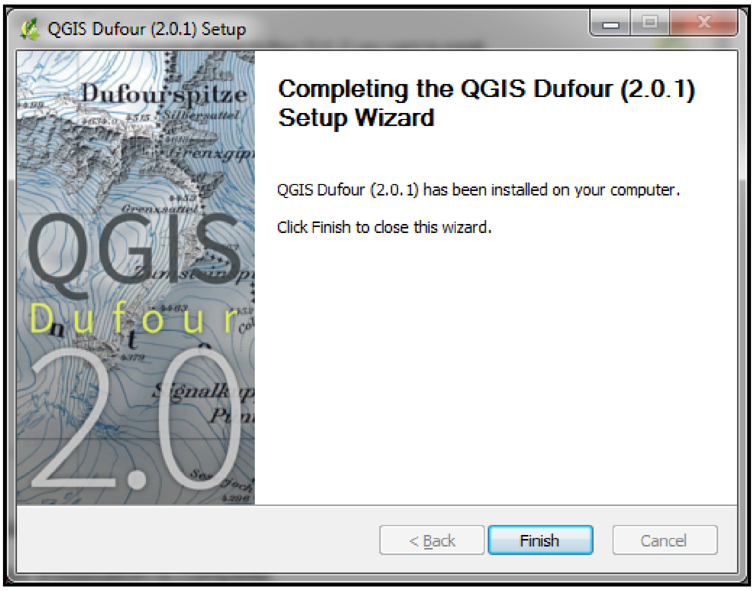
  <u>REMEMBER:</u> **You will need administrator rights to install QGIS.** If you are using Windows Vista or 7, you will need to right-click on the installation file with your mouse and select "Run as administrator." 

  If you do not use this method, the software will be installed but will not function correctly.

- In order to avoid downloading unnecessary data during the installation process, do **not** check the boxes corresponding to test data for North Carolina, South Dakota, and Alaska (see below).



After installation is complete, you will see the following message.

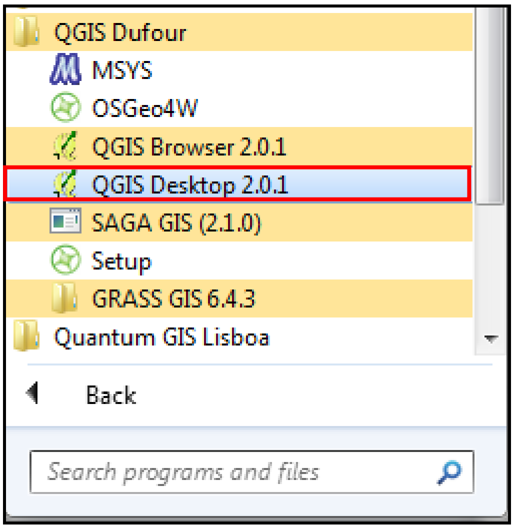


Step 2: Launch the program

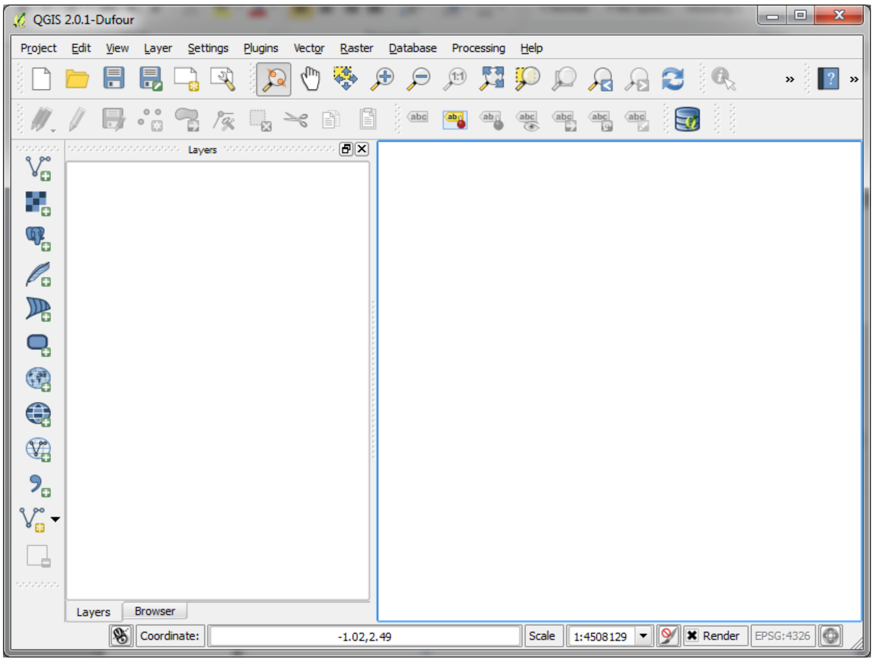
- To launch QGIS:

  Option 1: Double-click the shortcut found on the desktop.

  Option 2: On the desktop, click on the Start button and search for the program in the list of All Programs available on the computer.

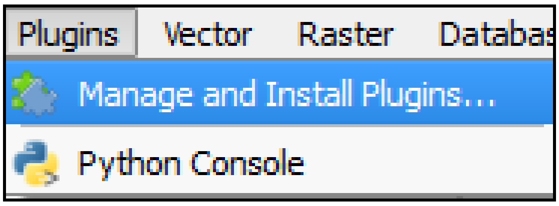


  After launching the QGIS program, you should see a screen that resembles the following.

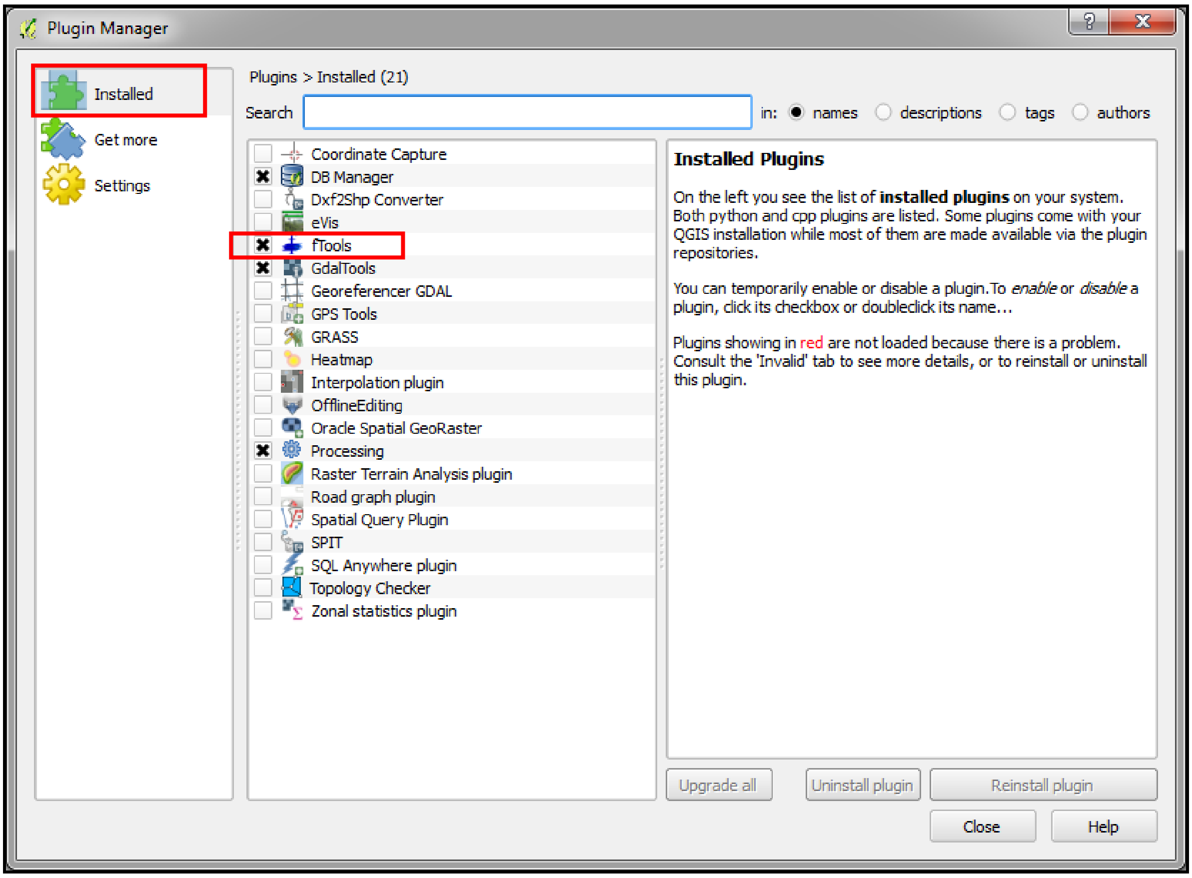


Step 3: Verify which plugins are installed

- On the main QGIS menu at the top of the screen, click on Plugins > Manage and Install Plugins.



You should see a screen like the following:



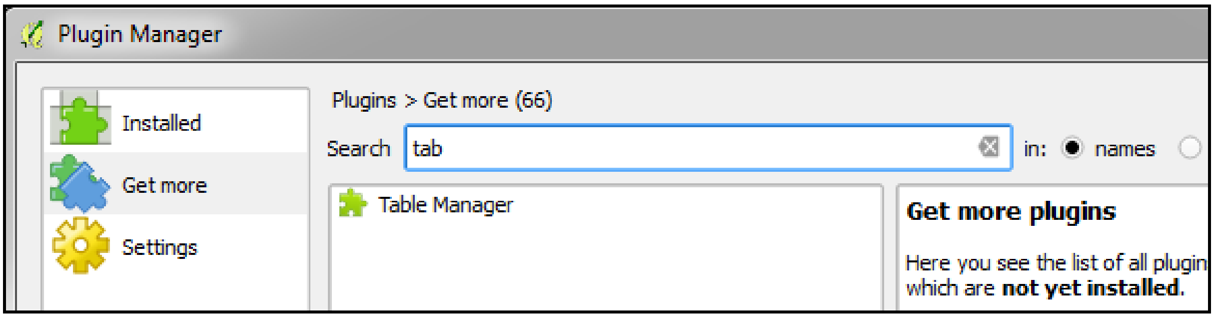
- The dialog box that opens contains the list of plugins that are installed by default. One of these is **fTools**, which provides some core functionality for working with vector data. This dialog box also provides access to external plugin repositories that offer additional plugins for extending the functionality of QGIS.

  (You can verify that you are accessing only the plugins available from the official QGIS plugin repository by clicking on the "Settings" icon in the left-hand navigation pane of the Plugin Manager dialog box. 3rd party plugins are also available, but many of these have not been verified by the QGIS team. For this reason, it is recommended that beginners use only the official QGIS repository plugins.)

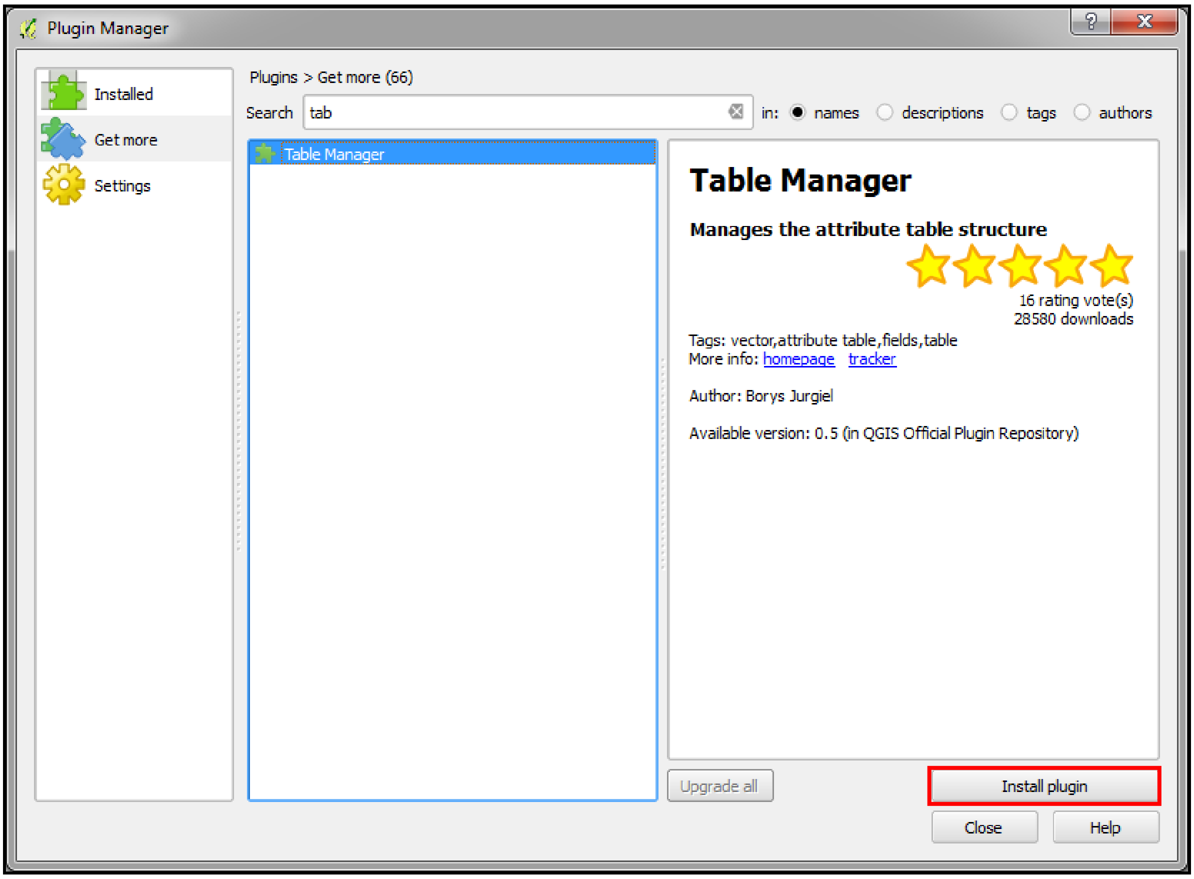
- Install the **Table Manager** plugin

  You will need to install the Table Manager plugin in order to work with attribute data during this course. To install this additional plugin, first click on the "Get more" icon in the Plugin Manager. You will see a list of additional plugins available for download.

  To isolate the Table Manager plugin, in the Search box at the top of the dialog box beginning typing the phrase "table manager." By the time you type "tab" the search box should have filtered out the other plugins so you are left with only the one you are seeking (see below).



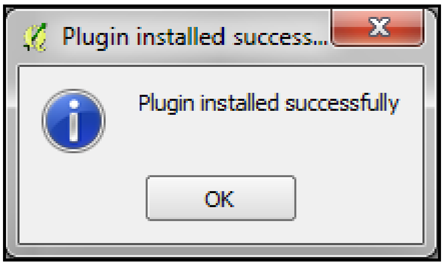
Click on the name of the plugin to select it. You should see a screen that provides additional information about the plugin (see below).



To install the plugin, click on the “Install plugin” button in the lower right corner of the dialog box.

NOTE: Plugins are installed via the Web, so you should make sure you have a stable Internet connection before attempting to download and install them.

After installing the plugin, you should see a message like the following:

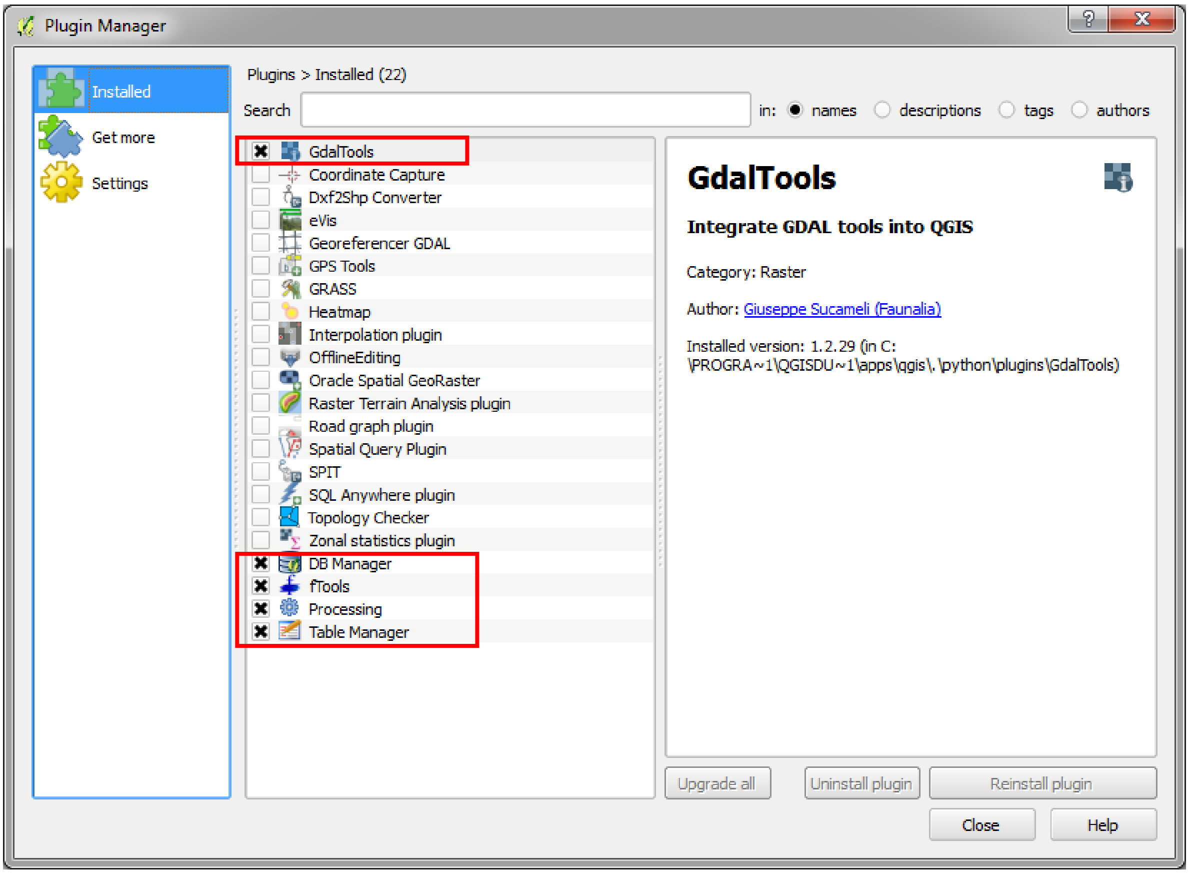


You can verify that the Table Manager plugin was installed by clicking on the Installed icon in the left-hand window of the Plugin Manager dialog box.

Step 4: Verify which plug-ins are activated

After verifying that the plugins you need are installed, you will need to verify that they have also been activated.

- To verify which plugins are activated, on the Installed plugins screen look for those plugins that have an X in the box next to their name.



- To prepare for the exercises developed for this course, make sure that the following plugins have been activated:
    - fTools
    - GdalTools
    - Spatial Query Plugin
    - Table Manager
    - Zonal statistics plugin

END

GIS Techniques for Monitoring and Evaluation of HIV/AIDS and Related Programs

Exercise 2.2

Explore QGIS

This training was developed as part of a joint effort between MEASURE Evaluation and MEASURE DHS, with funding provided by USAID and PEPFAR.

Exercise 2.2: Explore QGIS

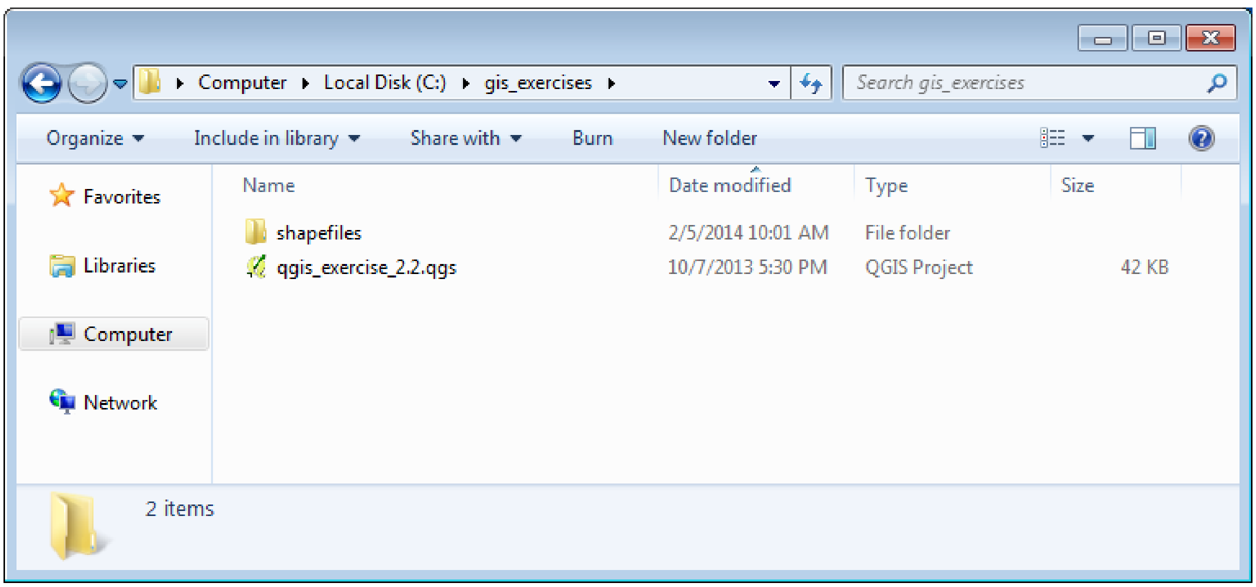
Summary: This exercise will allow you to explore the QGIS interface and gain a familiarity with some of the principal functions of the software. At the end of the exercise, you will be able to perform the following:

Objectives:

- Open a QGIS project already in progress.
- Work with the main components of the interface.
- Navigate within the map window and select, identify, and query geographic features.

<u>NOTE:</u> To complete these exercises, you will need to have QGIS installed on your machine, as instructed in exercise 2.1. You will also need to have downloaded and unzipped the data files associated with the exercise. These files can reside anywhere on your computer, but must be kept together, with the same original file structure.

In the below example, the files have been **unzipped** to a folder called "gis_exercises" on the C: drive. All the **shapefiles** are still housed within their original folder, which is stored at the same level as the **.qgs** file.



Step 1: Open a QGIS project already in progress.

□ Launch QGIS.

You should have a shortcut for QGIS on your desktop. 

If yes, click on it to launch QGIS.

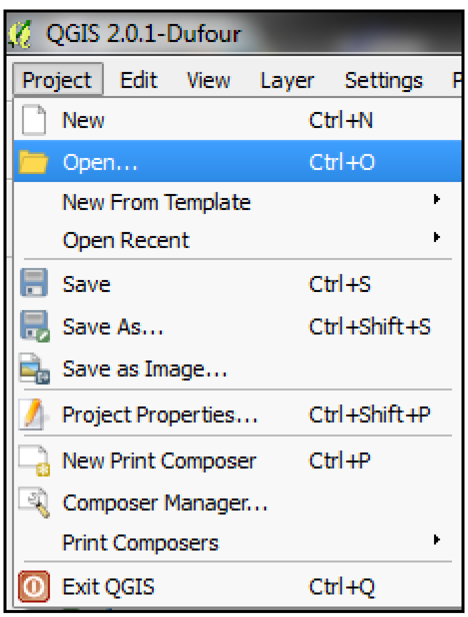
If not, on the Windows desktop click on Start > All programs > QGIS Dufour > QGIS Desktop 2.0.1.

□ Open a QGIS project.

Each time you work in QGIS, you work in a map document that is called a project. A project is a file that organizes the data layers you work with, the map layouts you create, etc., into a single document for easy access. The file name for a QGIS project document will have the extension .qgs.

A QGIS project has already been created as a starting point for this exercise. You will use this file to explore the QGIS interface in order to understand the basic layout of the software as well as some of the key functionality that is provided.

- On the main QGIS menu, click on Project > Open.



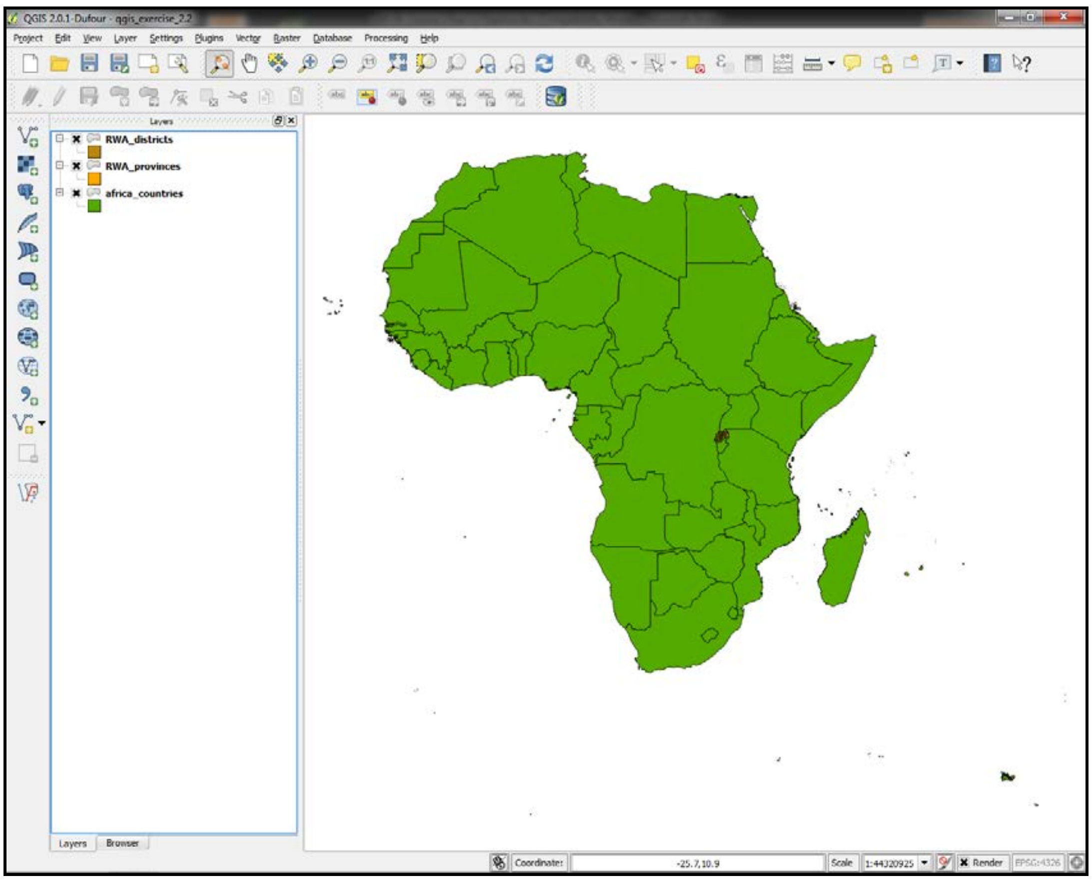
- Navigate to the folder **gis_exercises** and find the QGIS project file **qgis_exercise_2.2.qgs**.

HINT: You should have created a new folder for these exercises on your own computer when you downloaded the files associated with this exercise. If you have any questions on downloading the course exercise files, please refer to the instructions on the "Practical Exercises" page associated with Session 2. See page 1 of this document for an illustration of the downloaded, unzipped files.

- Double-click the name of the file to open the project.

  When the QGIS project opens, you should see a map of Africa with the administrative divisions for Rwanda superimposed. The map should resemble the image below.

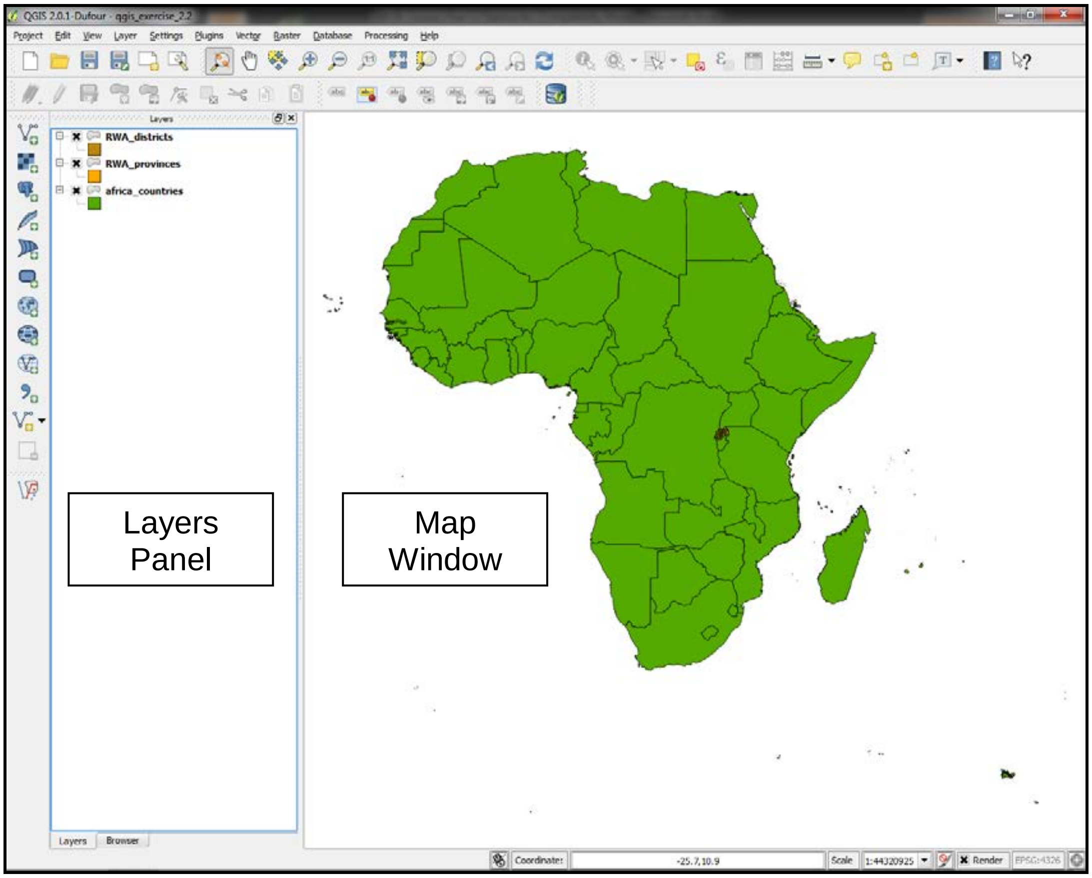
  NOTE: Your screen resolution might be different from that used to create the image below, which might result in a map that is displayed using a different scale and with a different placement of the tool bars in the QGIS interface.



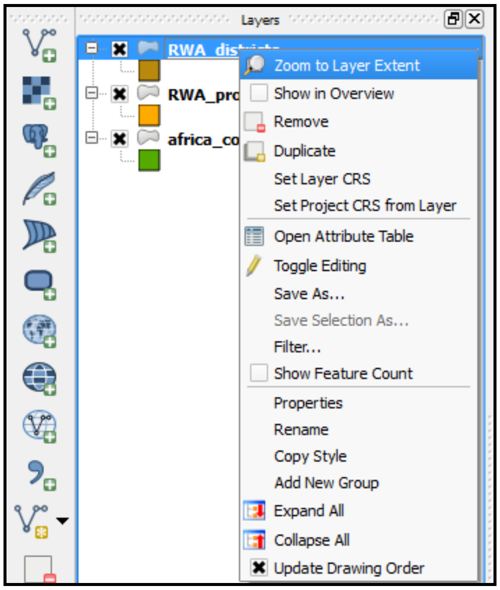
  In the steps that follow, we are going to explore different components of the QGIS user interface.

Step 2: Work with the main components of the interface

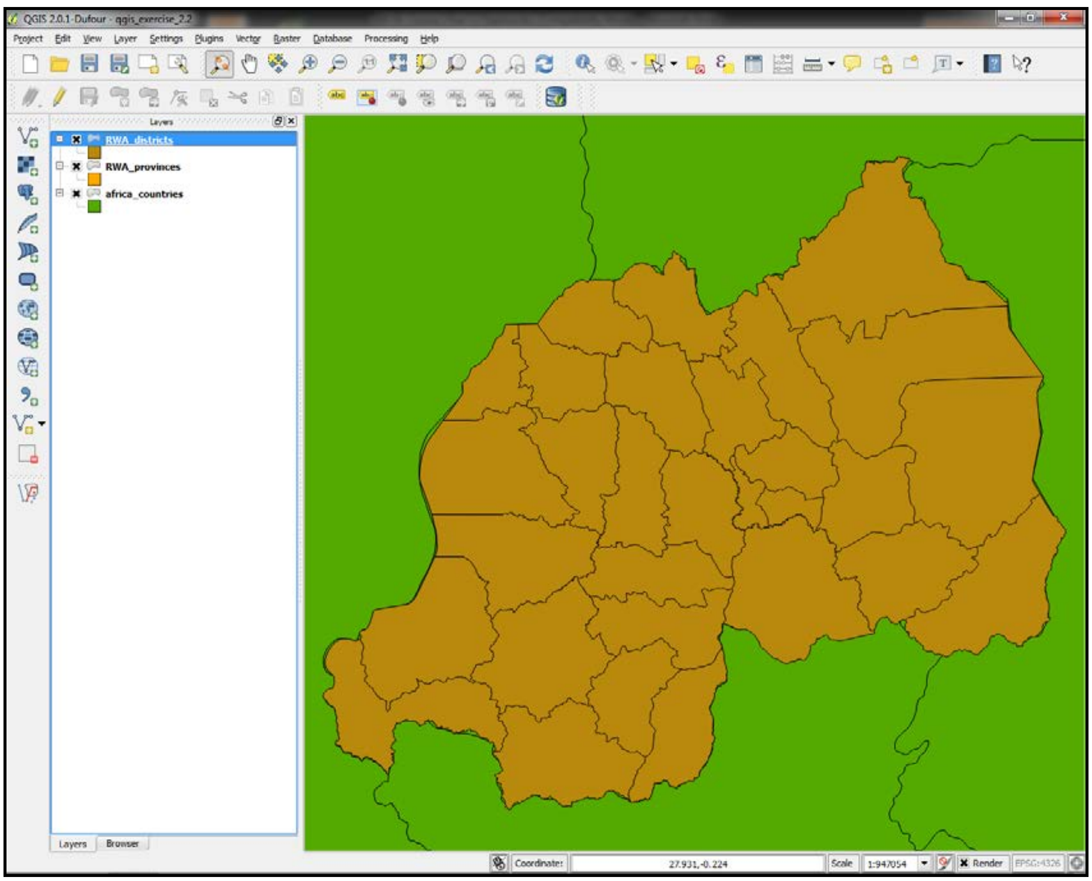
- The QGIS interface provides two main windows, the layers window and the map window (see below).



- To understand how the controls in the Layers Panel work, right-click on the layer RWA_districts, which represents Rwanda districts in 2010, and select "Zoom to Layer Extent" (see below).



After selecting Zoom to Layer Extent you should see a screen that centers on the country of Rwanda and looks like the following:

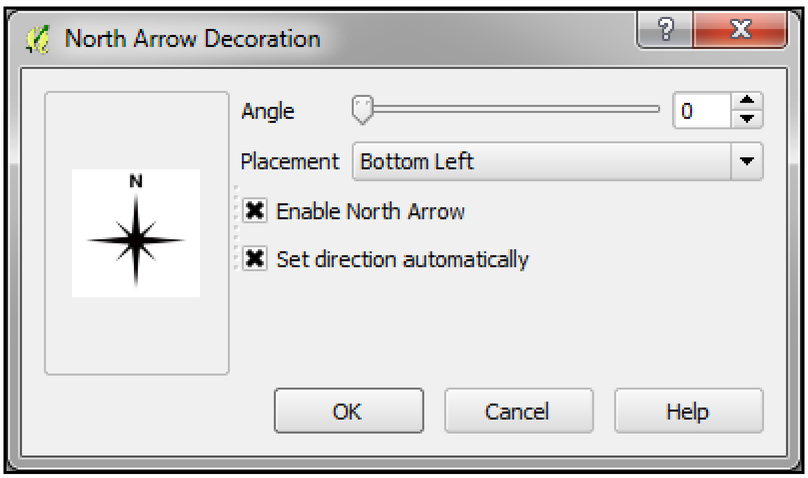


- Click on the boxes to the left of the layer names (with the Xs in them) to turn off and on the different data layers displayed in the layers window.

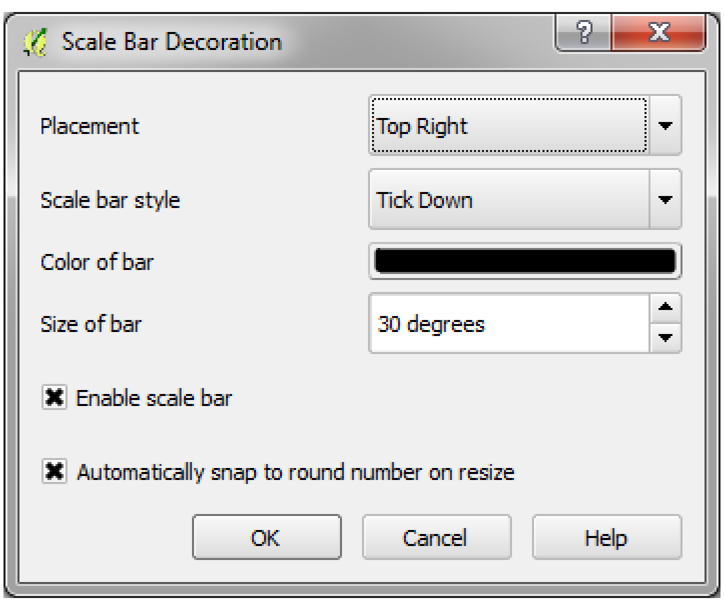
Question: Describe what happened.

Answer: The underlying layers became visible when the overlying layers were turned off. This illustrates that display of individual layers can be turned off and on, and indicates that the order in which the layers are displayed in the navigation window represents the top-to-bottom order in which the layers are displayed in the map window.

- Add a north arrow to the map window by going to the main menu in QGIS and selecting View > Decorations > North Arrow > Enable North Arrow and then clicking on OK.



- Add a scale bar to the map window by going to the main menu and selecting View > Decorations > Scale bar > Enable scale bar and then clicking on OK.

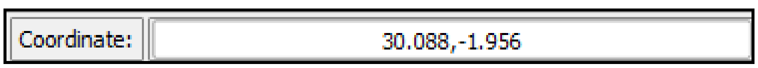


Question: The scale bar is displayed in degrees. What does this mean?

Answer: The units of measure used for the scale bar indicate the coordinate system used for the underlying geographic file. A scale bar displayed in degrees indicates that the geographic files used to display the administrative divisions for Africa and Rwanda are stored in an “unprojected” or “geographic” coordinate system based on longitude and latitude in degrees. This means that the files can be used for mapping, but should be projected to a planar (flat) coordinate system before any spatial analysis is conducted that involves measurement of distances, angles, or surface areas.

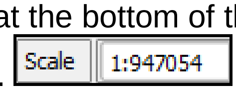
- At the bottom of the map window you should see a box that is labeled “Coordinate.” Move the cursor within the map window and watch what happens to the numbers in the box to the right of this label.

Question: What do the numbers represent, and can you tell whether the coordinates displayed in the box represent degrees or something else?



Answer: It should be possible to see that the numbers displayed in the Coordinate box represent the location of the cursor in the map window based on the underlying coordinate system. Also, with a little knowledge of the ranges of values used for longitude and latitude (longitude ranges from +/- 180 degrees in relation to the Prime Meridian that runs through Greenwich, England, and latitude ranges from +/- 90 degrees from the equator to the poles), it should also be possible to see that the values are reasonable for the location of Rwanda.

- **Question:** Can you identify the scale of the map window when it is zoomed to the extents of the country of Rwanda? Where can this information be found?

Answer: The scale of the map can be found in the box at the bottom of the map window that is located just to the right of the word Scale. 

- In the box in which the scale is displayed, enter a new scale of 1:60,000,000 (60 million, or a 6 with 7 zeroes) and press Enter on the keyboard.

Question: What do you see now?

Answer: You should see the entire continent of Africa in the map window.

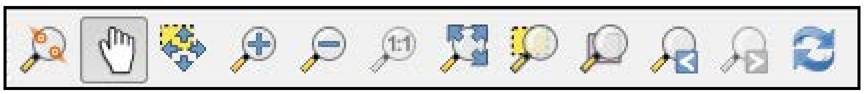
- Enter a new scale of 1:1,260,000.

Question: What do you see in the map window now?

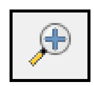
Answer: You should be zoomed in again on the country of Rwanda.

Step 3 : Navigate within the map window and select, identify, and query geographic features.

The preceding step showed you how to work at a high level with the three principal panels or windows within the QGIS interface. In Step 3 you are going to navigate in a more detailed fashion within the map window by using the Map Navigation toolbar (see below).

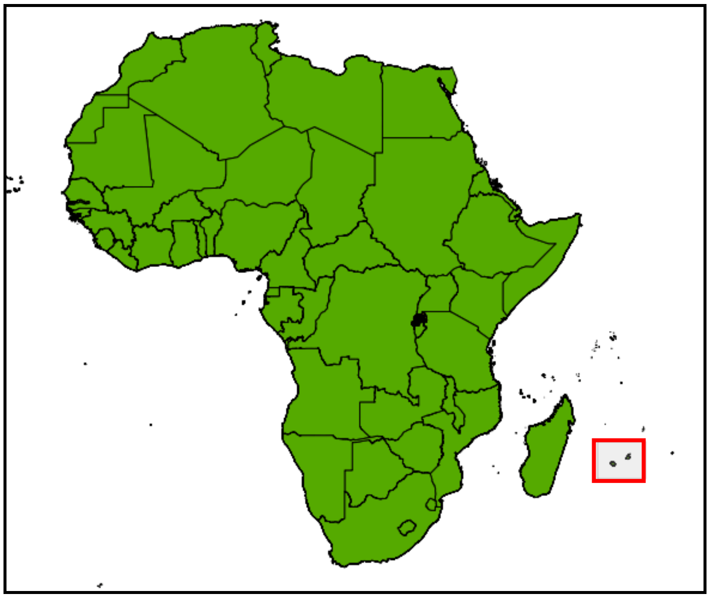


You will also learn how to use some of the key functionality of the Attributes toolbar (see below).

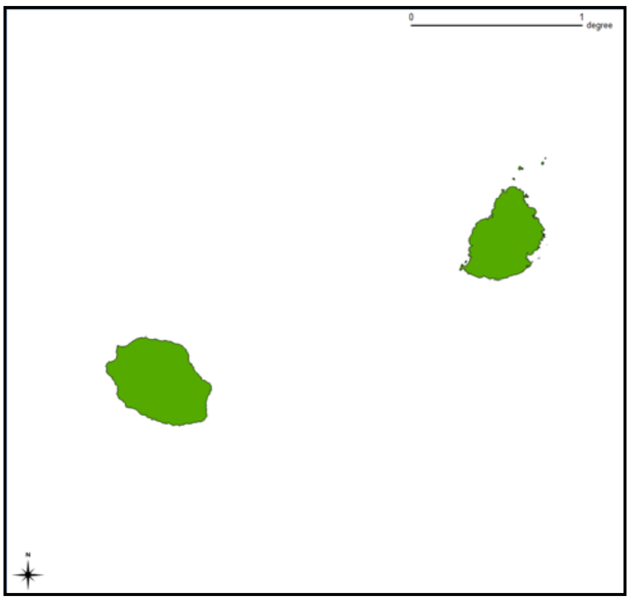
- To see the full geographic extents of all the geographic layers contained in the Layers panel, click on the “Zoom Full” button. 
- To view a smaller geographic area, click on the “Zoom In” button. 

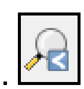
  NOTE: The tool that is currently selected and active appears with a different background color and with a highlight frame around it. For example, after having selected "Zoom In," the button should look like the following: 

  After selecting "Zoom In," position the tip of the cursor to the upper left of the two main islands to the east of Madagascar, then left-click and drag the cursor down and to the right to create a window around the two islands (see below). When you have finished creating the window, release the left-click to finalize the selection.

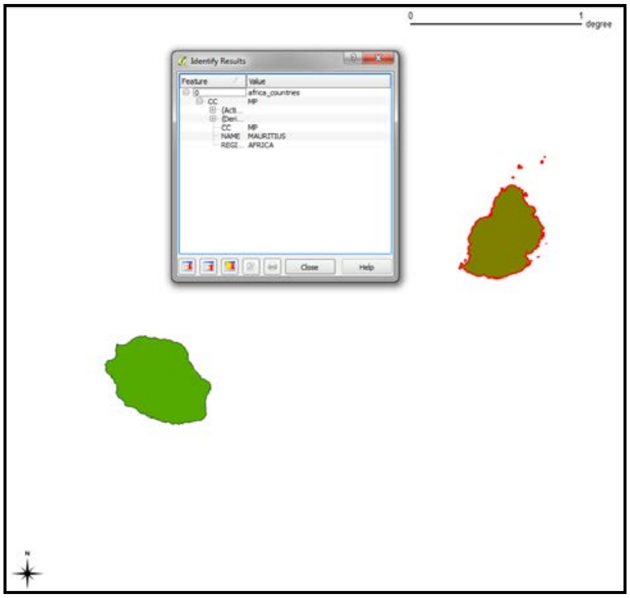


After selecting the two islands to the east of Madagascar, the map window should look similar to the one below.



- To return to the view of the African continent, click on the “Zoom Last” button. 
- To go back to the view of the two islands, click on the “Zoom Next” button. 
- To see the non-spatial attributes of the eastern-most island, click on the “Identify Features” button , left-click on the data layer “africa_countries” in the Layers

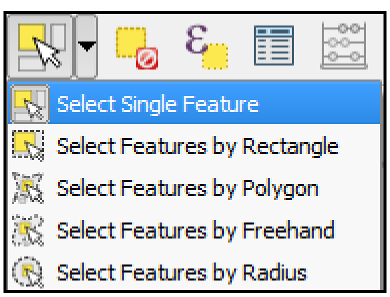
panel, and then left-click on the island of interest." You should see a screen similar to the one below.



Questions: What is the name of the island selected? What is the name of the island that lies to the west of it?

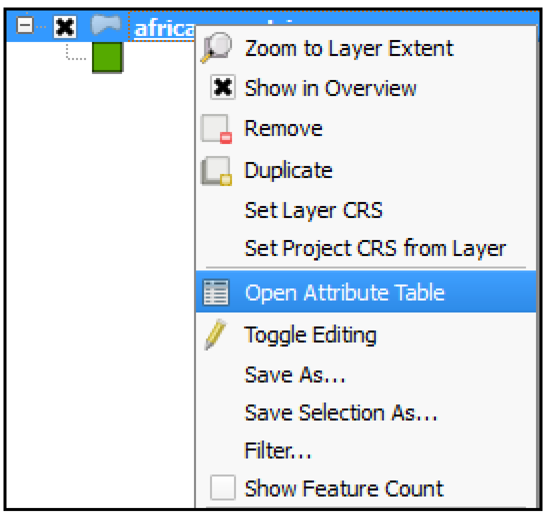
Answers: The island to the east is Mauritius. The island to the west is Reunion.

☐ To learn how to view the non-spatial attributes for features that are stored in QGIS attribute tables, go to the Layers panel and make sure you have left-clicked on the layer "africa_countries" in order to select it. After that, select the "Select Single Feature" tool on the Attributes toolbar (see below).



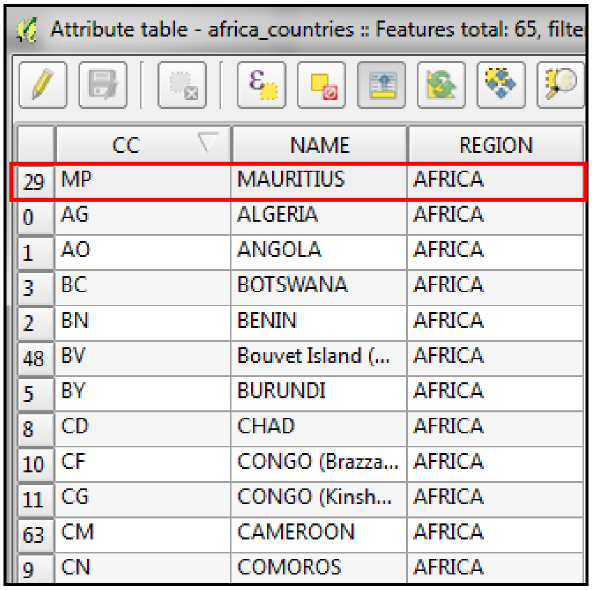
After activating the Select Single Feature tool, left-click on the island of Mauritius. You should see a map window that resembles the one below.

With Mauritius selected, open the attribute table for the "africa_countries" layer by right-clicking on the layer name in the Layers panel and selecting "Open Attribute Table" (see below).



The selected record is not within the portion of the table that is visible. To bring the attribute record for Mauritius to the top of the table, click on the button for "Move selection to top" .

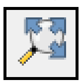
After moving the selection to the top, the table should resemble the following:



Attribute table - africa_countries :: Features total: 65, filte

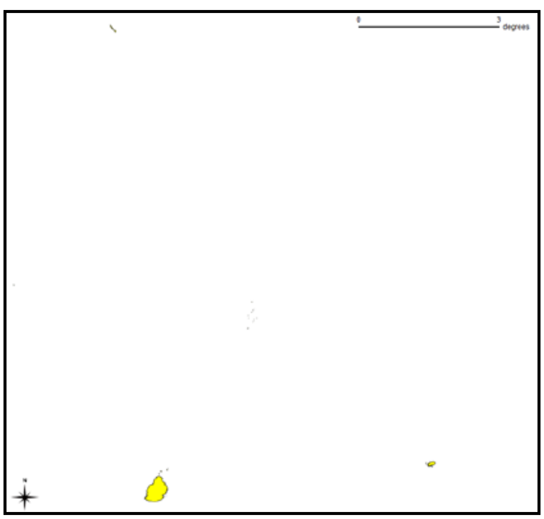
	CC	NAME	REGION
29	MP	MAURITIUS	AFRICA
0	AG	ALGERIA	AFRICA
1	AO	ANGOLA	AFRICA
3	BC	BOTSWANA	AFRICA
2	BN	BENIN	AFRICA
48	BV	Bouvet Island (...	AFRICA
5	BY	BURUNDI	AFRICA
8	CD	CHAD	AFRICA
10	CF	CONGO (Brazza...	AFRICA
11	CG	CONGO (Kinsh...	AFRICA
63	CM	CAMEROON	AFRICA
9	CN	COMOROS	AFRICA

- To return to a view of the continent of Africa, without deselecting the island of Mauritius, close the attribute table for “africa_countries” and click on “Zoom Full.”

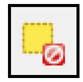


- To set the map window to the geographic extents of just the island of Mauritius, which should still be selected in the attribute table, make sure the “africa_countries” layer is still selected in the Layers panel and click on the “Zoom to Selection” button.

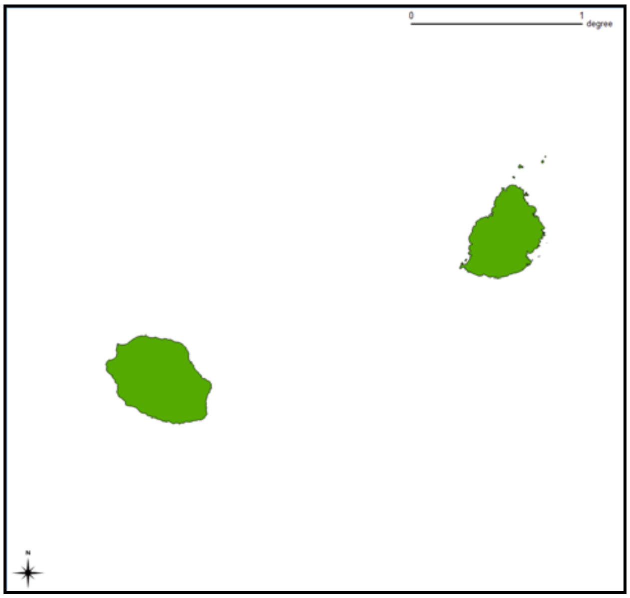
You should now see a map window that resembles the following:



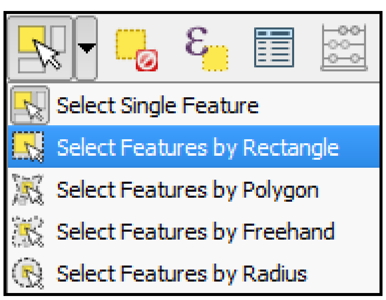
Note that the main island of Mauritius is not in the center of the map window because the group of islands associated with Mauritius extends over a greater area than just the one island.

- To clear the selection of Mauritius in the attribute table for the “africa_countries” layer, click on the “Deselect Features from All Layers” button. 

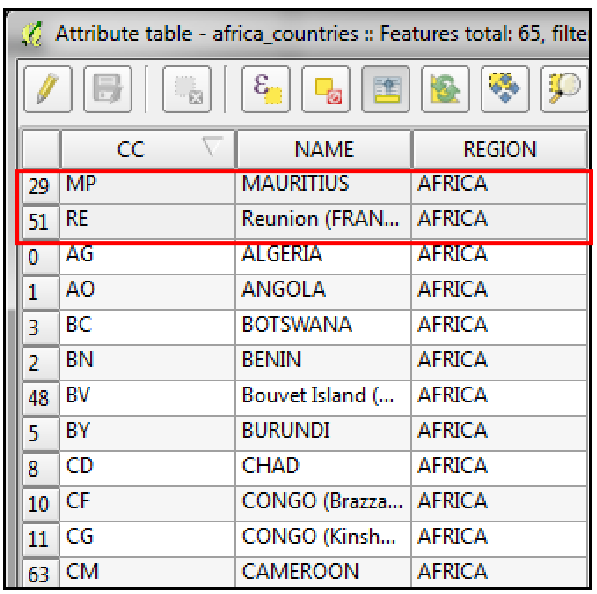
- To create a new shapefile that contains only the two islands of Reunion and Madagascar, begin by clicking on the “Zoom Last” button  until you have re-centered the map window on the two islands (see below).



- To select both islands simultaneously, activate the button "Select Features by Rectangle" and use the cursor to make a rectangle around the two islands. After selection, the two islands should be highlighted in yellow in the map window (see below).



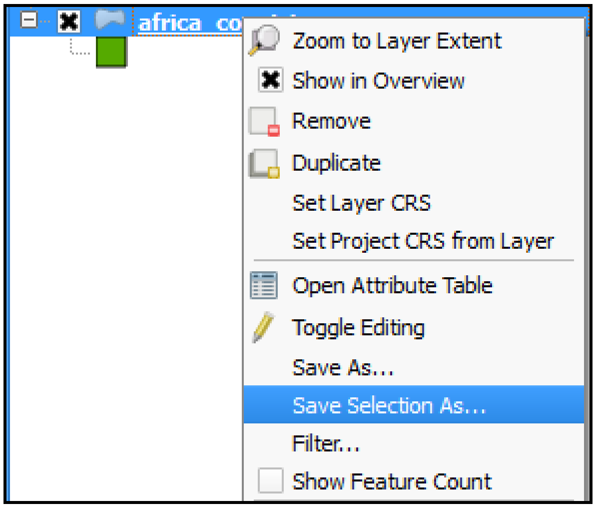
- To verify the selection of the two islands, right-click on the "africa_countries" layer in the Layers panel, select "Open Attribute Table," and move the selected records to the top of the table (see below).



Attribute table - africa_countries :: Features total: 65, filte

	CC	NAME	REGION
29	MP	MAURITIUS	AFRICA
51	RE	Reunion (FRAN...	AFRICA
0	AG	ALGERIA	AFRICA
1	AO	ANGOLA	AFRICA
3	BC	BOTSWANA	AFRICA
2	BN	BENIN	AFRICA
48	BV	Bouvet Island (...	AFRICA
5	BY	BURUNDI	AFRICA
8	CD	CHAD	AFRICA
10	CF	CONGO (Brazza...	AFRICA
11	CG	CONGO (Kinsh...	AFRICA
63	CM	CAMEROON	AFRICA

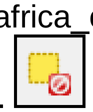
- Save the two selected island groups in a new shapefile by right-clicking on the "africa_countries" layer in the Layers panel and selecting "Save Selection As" (see below).



When the dialog box appears, make sure the output format "ESRI shapefile" is selected in the drop-down menu for Format, click on the Browse button and specify the following output file name (path may vary according to folder you are working in):

C:\gis_exercises\shapefiles\ islands.shp

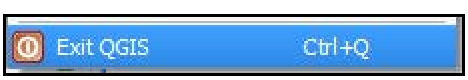
After specifying the output file name, click on OK. You should see a message at the top of the map window that indicates that the file has been saved successfully. A new shapefile containing just the two islands of interest is now available for mapping and analysis.

- Clear the selection of the two islands in the attribute table for the "africa_countries" layer by clicking on the "Deselect Features from All Layers" button. 

- Reset the zoom level to that of the African continent by clicking on the "Zoom Full" button. 

- Save the QGIS project file (you can "SAVE AS" with a new name, to keep your original available again for later, if you like.)

- Close the QGIS project. To do so, select Project > Exit QGIS or use a combination of the Ctrl and Q keys (see below).



END

GIS Techniques for
Monitoring and Evaluation of
HIV/AIDS and Related Programs

Exercise 3.1

Join Attribute Data to a Shapefile

**This training was developed as part of a joint effort between MEASURE Evaluation and MEASURE DHS, with funding provided by USAID and PEPFAR.*

Exercise 3.1 Introduction

The data joining step is often a prerequisite for mapping and spatial analysis. Creating a color-shaded map (also known as a choropleth map), is a good first step in geographic data visualization and analysis.

Most demographic and health data is stored in tables. In order to make a map of these data, they need to be associated with (joined to) spatial data files.

It is important to consider some basic database rules. Since the shapefile contains multiple, unique records, the data to be linked must be in the same form. If a name field is used for linking, ensure that the spelling is exactly the same. Preferably, linking will be done using codes, such as “District ID”, which are stored in both the data table and the shapefile.

Objectives:

- Open and examine a data file containing HIV prevalence data by district for Rwanda.

- Open and examine a shapefile that contains the administrative boundaries for the districts of Rwanda.

- Join the attribute table containing HIV prevalence data to the shapefile, and save the result as a new shapefile.

- Examine the new shapefile and revise the column names to be more meaningful using the Table Manager plug-in.

Requirements:

To complete these exercises, you will need to have QGIS and several plugins installed on your machine, as instructed in exercise 2.1. You will also need to have downloaded and unzipped the data files associated with the exercise. Remember, these files can reside anywhere on your computer, but must be kept together, with the same original file structure.

Exercise 3.1: Join Attribute Data to a Shapefile

In this exercise, you will learn how to join a text file containing HIV prevalence data from the Rwanda DHS 2010 final report to a shapefile representing the districts of Rwanda. You will then learn how to create a color-shaded map to show HIV prevalence by district.

Task Overview: Join an attribute table containing HIV prevalence data to the geometric shapes in a shapefile.

Background:

At some point in time, the majority of indicator data are stored in stand-alone tables. In order to make a map using these data, the data must first be associated with spatial features, such as administrative division boundaries or health facility locations.

As you may recall, a common format for storing spatial data is the shapefile. Before joining attribute data to a shapefile, it is important to take into account certain characteristics of shapefiles:

- The geographic entities described in a shapefile are of type vector-based data, which means they will be collections of points, lines, or polygons.

- Before attribute data, such as HIV indicators, are joined to a shapefile, the shapefile will contain only the physical properties (such as the latitude, longitude, altitude, length, or area) of the geographic entities of interest, along with one or more geographic identifiers, such as administrative name and/or code.

- Before using common geographic identifiers to join attribute data to a shapefile, the attribute data and shapefile data must be referred to at the same geographic level. For example, attribute data for second-level administrative divisions, such as districts, will need to be aggregated prior to joining them to first-level administrative divisions, such as regions or provinces.

- “Shapefiles” are actually a collection of at least three files:

  - Main file, which contains the geographic coordinates that make up the features described (e.g., administrative divisions or health facility locations).
    - Example: regions.shp
  - Index file, which identifies the byte number offset of each main file record from the beginning of the main file.
    - Example: regions.shx
  - dBASE table, which contains attributes for features represented in the main file.

    - Example: regions.dbf

  Source: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

- If a field containing names is used to join attribute data to a shapefile, it is necessary to make sure that the spelling of the names is the exact same in the stand-alone attribute table and the dBASE table that is part of the shapefile.

- To avoid having to identify and correct spelling differences in names when joining attribute tables to shapefiles (for example, “Nyarugenge” vs. “NYARUGENGE” vs. “Nyrau Genge”), it is recommended to base the join on unique administrative codes (for example, district ID = RW0101).

Step 1: Open and examine a file (a data or attribute table) containing HIV prevalence data.

□ Launch QGIS with a blank (new) project document.

You should have a shortcut for QGIS on your desktop. 

If yes, click on it to launch QGIS.

If not, on the Windows desktop click on Start > All programs > QGIS Dufour > QGIS Desktop 2.0.1.

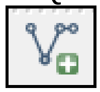
□ Open a comma-delimited file containing HIV prevalence data.

- A comma-delimited text file (extension .csv) has been created for this exercise, as the comma-delimited format is a common file exchange format that can easily be produced using Excel or other spreadsheet or database programs.
  - When opening a comma-delimited file with the .csv extension, QGIS assumes that all fields (columns) contain strings (text) unless the **field types*** are specified in an accompanying CSVT (CSV template) file.
  - A CSVT file name must have the same file name and path as the CSV file it describes, except the file extension is .csvt instead of .csv. For this exercise, the CSV and CSVT file names are as follows:
    - CSV file: gis_exercises\data_tables\ RWA_DHS2010_TableD91.csv
    - CSVT file: gis_exercises\data_tables\ RWA_DHS2010_TableD91.csvt
  - A CSVT file contains a single line of information that specifies the types of columns contained in the CSV file as well as the maximum size of the values contained in those columns. The contents of the CSVT file for this exercise are as follows:

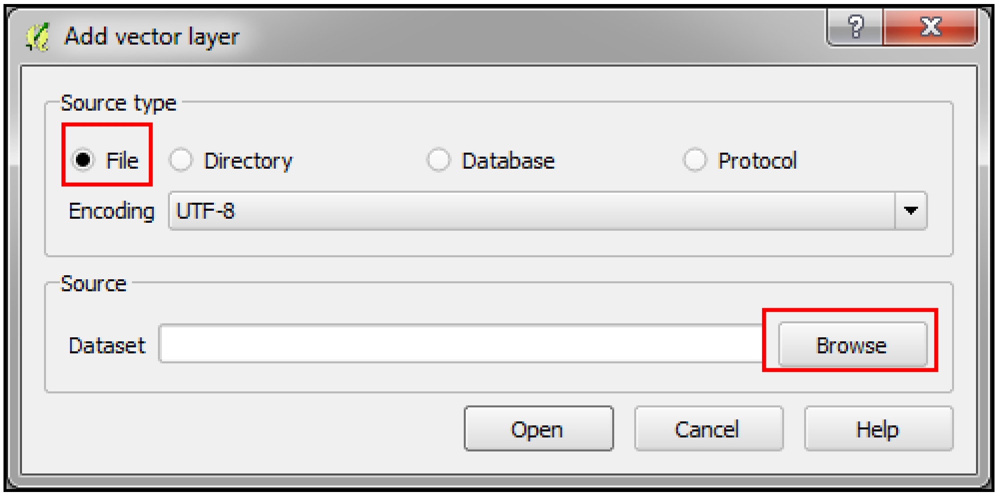
    "string(10)","string(4)","real(4.1)","real(4.1)","real(4.1)"

    This information indicates that the CSV file has five columns. The first two columns contain strings (text), with a length of 10 characters and 4 characters, respectively. The last three columns contain real (floating point) numbers with a total length of 4—including the decimal—and a precision (number of digits after the decimal) of 1 (for example, a value of 10.1).

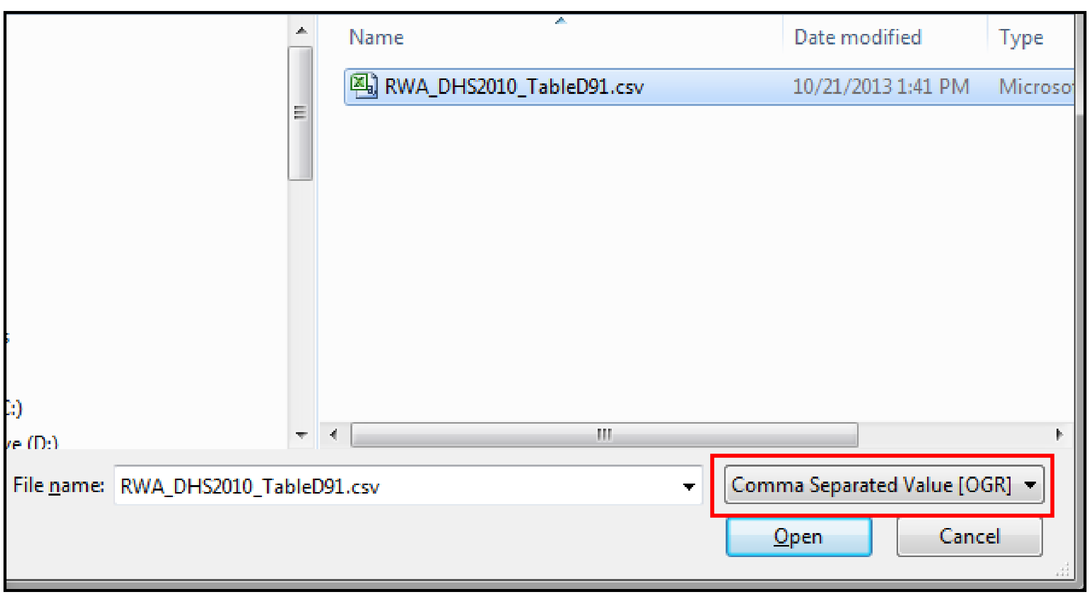
*A “field type” refers to the way a database expects to see data stored in a particular field. The most common field types are “text” (also known as “string”) or integer (for numeric whole numbers) or float (for numeric decimal values).

- To view the CSV file in QGIS, on the toolbar for managing layers click on the Add Vector Layer button. 

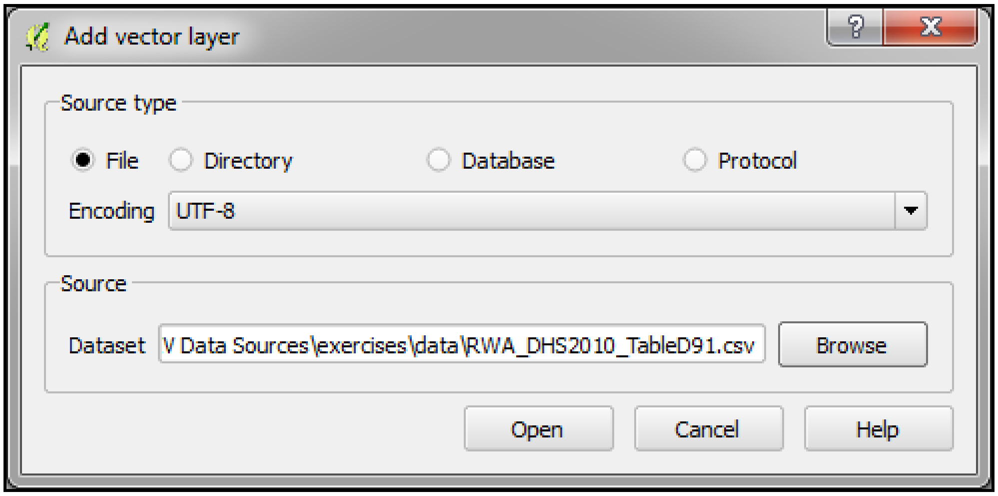
- In the Add vector layer file dialog box that appears (see below), make sure the source type is set to File and click on Browse.



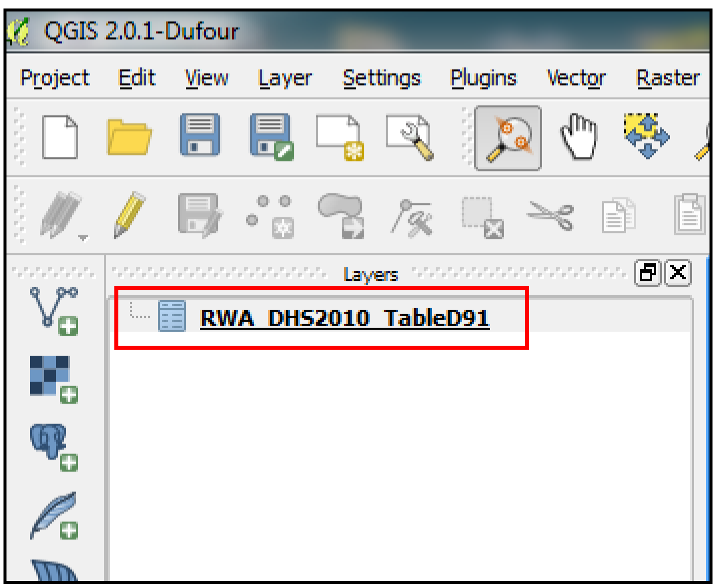
- In the file dialog box that appears (see below):
  - Use the pull-down menu in the lower-right corner to set the file type to “Comma Separated Value [OGR] (*.csv *.CSV).”
  - Navigate to the folder “gis_exercises\data_tables” and find the comma-delimited text file **RWA_DHS2010_TableD91.csv**.
  - Click on the file name RWA_DHS2010_TableD91.csv so that it appears in the file name box.



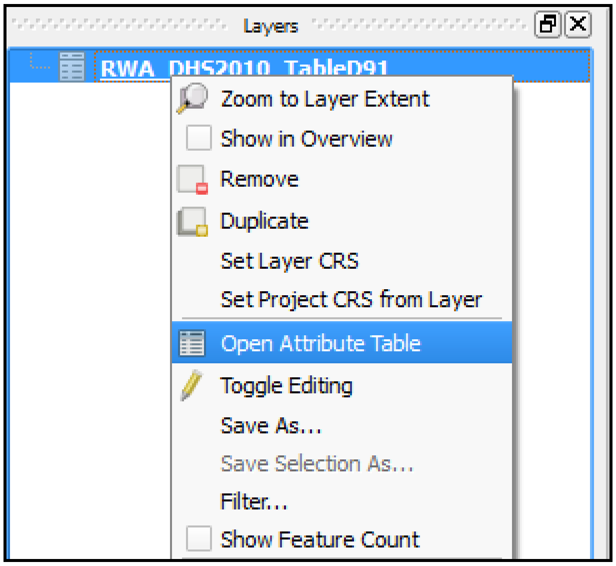
- Click on Open. When the Add vector layer file dialog box reappears with the full path to the CSV file indicated in the Dataset field (see below), click once again on Open.



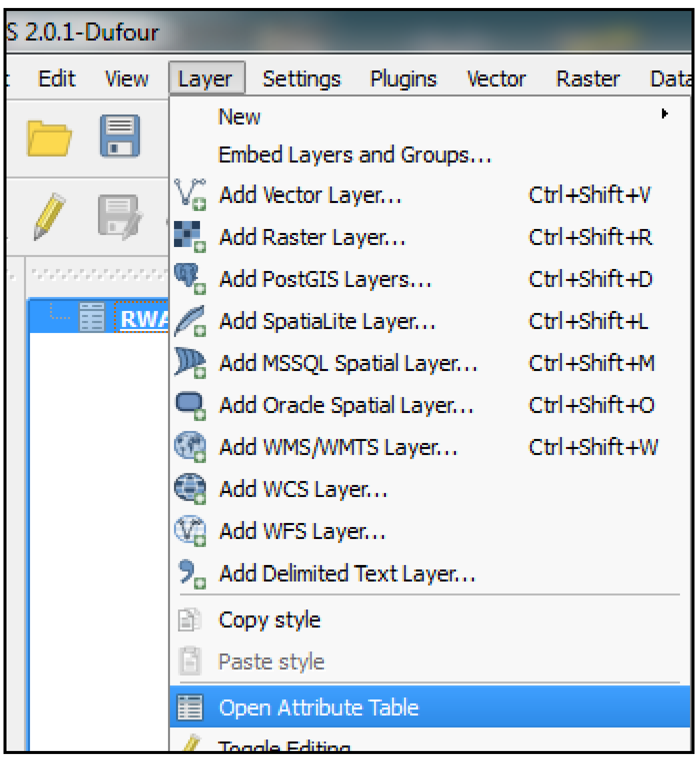
- After clicking on Open for the second time, you should see a new layer in the Layers window corresponding to the comma-delimited text file you just added.



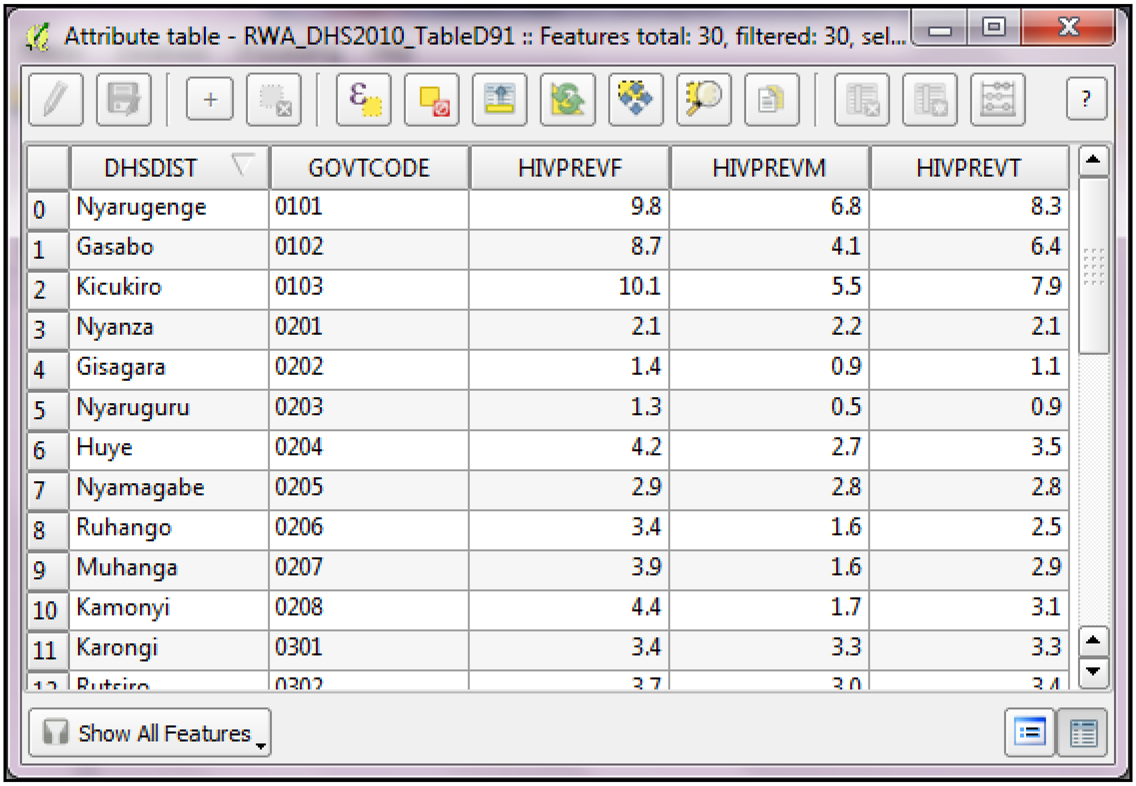
- To view the contents of the new attribute layer, right-click on the layer name and select Open Attribute Table (see below).



An alternative way to view the contents of the new data layer is to highlight the data layer name by single-clicking with the left mouse button and then going to the QGIS main menu and selecting Layer > Open Attribute Table (see below).



After opening the attribute table, you should see a table like the following.



	DHSDIST	GOVTCODE	HIVPREVF	HIVPREVM	HIVPREVT
0	Nyarugenge	0101	9.8	6.8	8.3
1	Gasabo	0102	8.7	4.1	6.4
2	Kicukiro	0103	10.1	5.5	7.9
3	Nyanza	0201	2.1	2.2	2.1
4	Gisagara	0202	1.4	0.9	1.1
5	Nyaruguru	0203	1.3	0.5	0.9
6	Huye	0204	4.2	2.7	3.5
7	Nyamagabe	0205	2.9	2.8	2.8
8	Ruhango	0206	3.4	1.6	2.5
9	Muhanga	0207	3.9	1.6	2.9
10	Kamonyi	0208	4.4	1.7	3.1
11	Karongi	0301	3.4	3.3	3.3

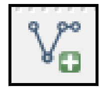
The attribute table you added contains five fields (columns):

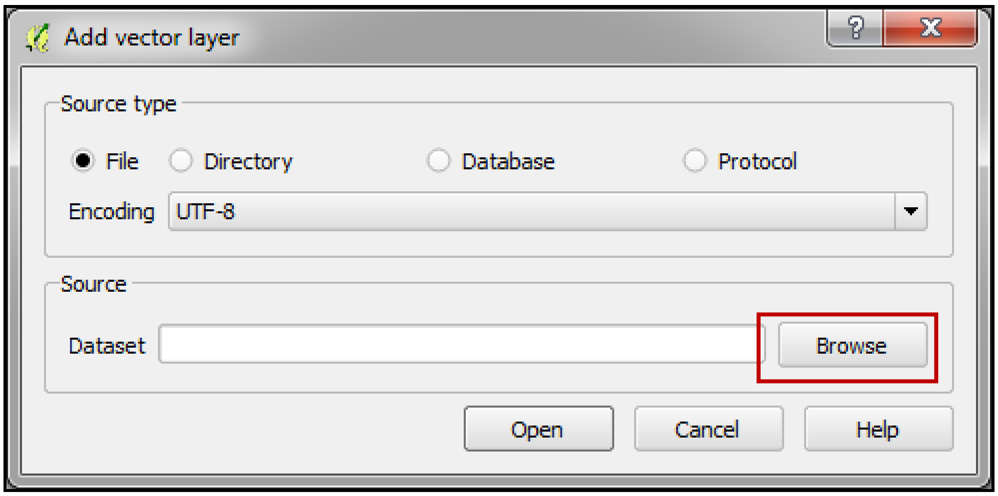
- DHSDIST: The name of the district for which data are reported in Table D.91 in the Rwanda DHS 2010 final report. Data type is string. Maximum length is 10. (refer back to p.5 for a review of the CSVT file which, as you may recall, contains this data type information.)
- GOVTCODE: A unique administrative code assigned to each district by the Rwandan government. Data type is string, NOTE; even though these are numbers, the first value is actually the numeral “0”. Sometimes numerals used as IDs are stored as a “string” data type. No numeric calculations can be perfomed on these data. Maximum length is 4.
- HIVPREVF: HIV prevalence (percent HIV positive) among females of reproductive age (age 15-49) in 2010. Data type is real (floating point). Maximum length is 4. Maximum number of decimal places is 1.
- HIVPREVM: HIV prevalence among males of reproductive age in 2010. Data type is real (floating point). Maximum length is 4. Maximum number of decimal places is 1.
- HIVPREVT: HIV prevalence among the total adult population (females and males) of reproductive age in 2010. Data type is real (floating point). Maximum length is 4. Maximum number of decimal places is 1.

Note that the DHS district names are spelled using initial capital letters. For the government codes, the first two digits represent the province and the last two digits represent the district. For example, districts 0101, 0102, and 0103

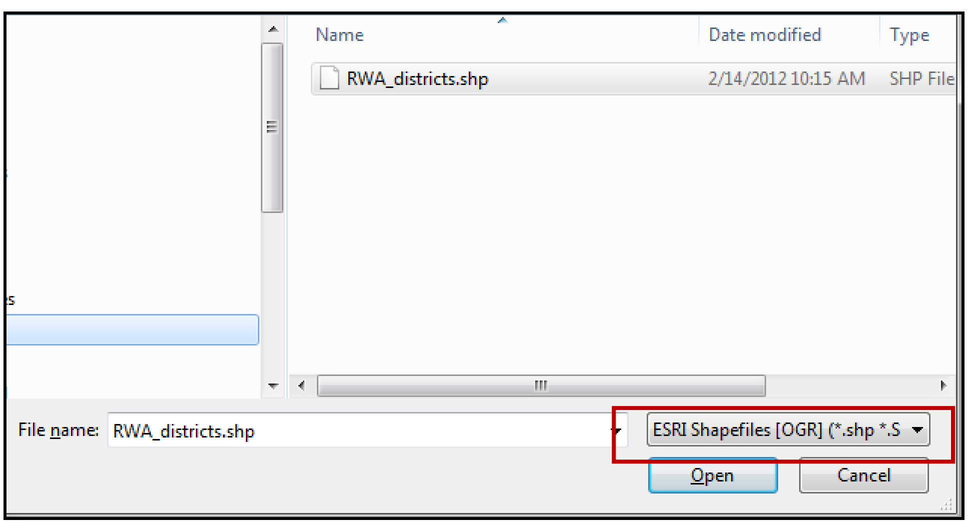
(Nyarugenge, Gasabo, and Kicukiro, respectively), all fall into the province that contains the City of Kigali (province ID = 01).

Step 2: Open and examine a shapefile that contains the administrative boundaries for the districts of Rwanda.

- Once again, in QGIS click on the Add Vector Layer button.  You should see the Add vector layer dialog box.

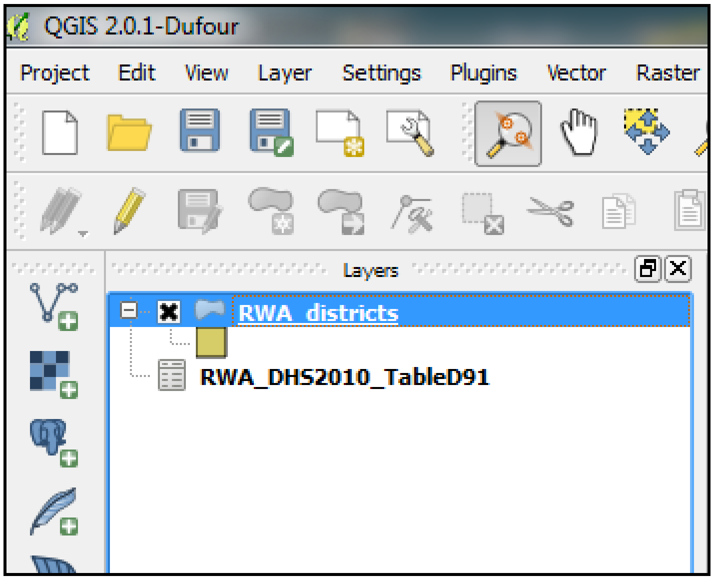


- Click on the Browse button, and in the file dialog window that appears, use the pull-down menu in the lower-right corner to set the file type to "ESRI Shapefiles [OGR] (*.shp *.SHP)." Navigate to the folder "gis_exercises\shapefiles" and find the shapefile **RWA_districts.shp**. Click on the file name RWA_districts.shp so that it appears in the file name box. The file dialog box should resemble the one below.

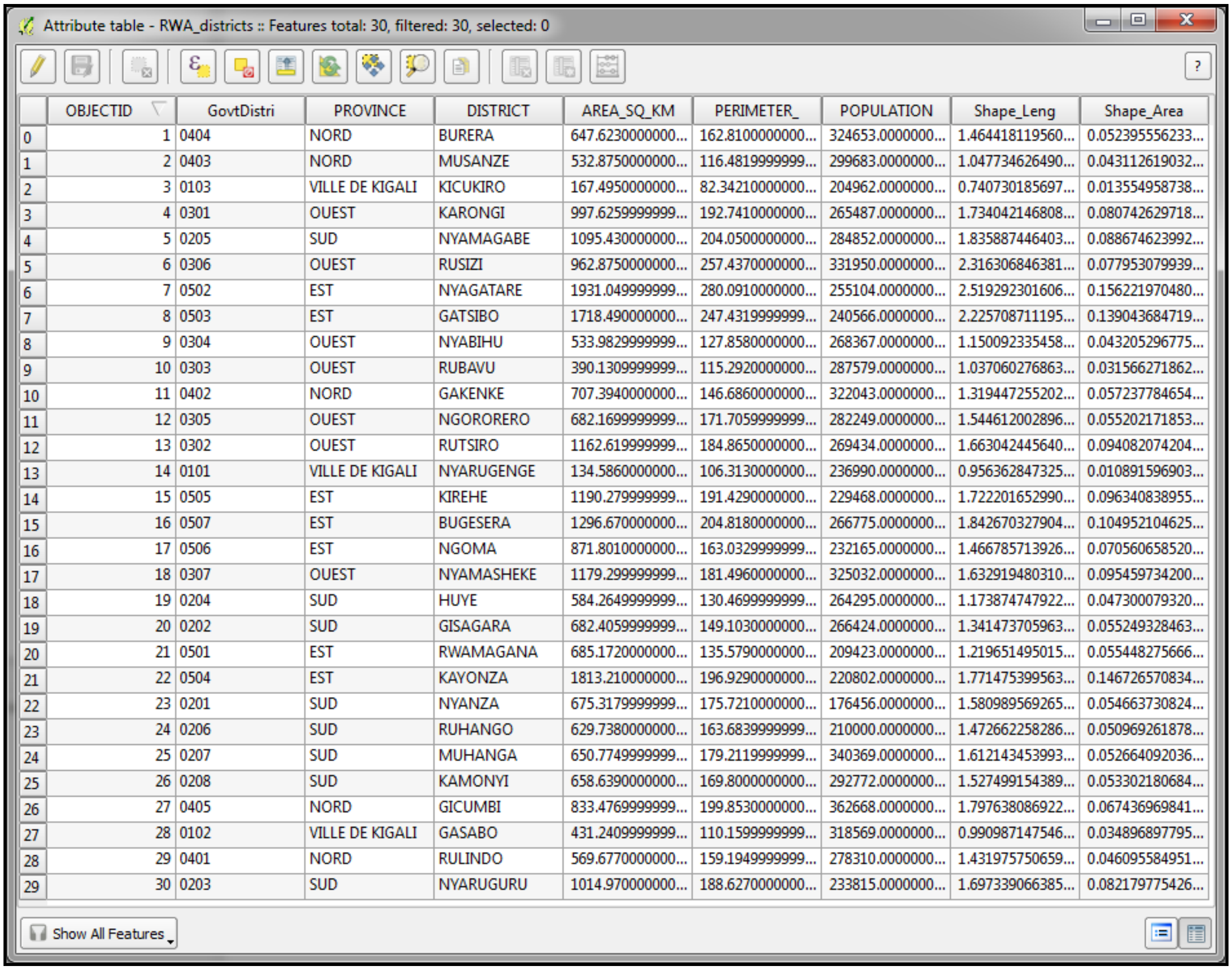


- Click on Open. When the Add vector layer file dialog box reappears with the full path to the shapefile indicated in the Dataset field, click once again on Open.

- You should see a new layer in the Layers window corresponding to the shapefile you just added.



- To view the contents of the attribute table for the shapefile just added, right-click on the layer name and select Open Attribute Table (see below).



Attribute table - RWA_districts :: Features total: 30, filtered: 30, selected: 0

	OBJECTID	GovtDistri	PROVINCE	DISTRICT	AREA_SQ_KM	PERIMETER_	POPULATION	Shape_Leng	Shape_Area
0	1	0404	NORD	BURERA	647.6230000000...	162.8100000000...	324653.0000000...	1.464418119560...	0.052395556233...
1	2	0403	NORD	MUSANZE	532.8750000000...	116.4819999999...	299683.0000000...	1.047734626490...	0.043112619032...
2	3	0103	VILLE DE KIGALI	KICUKIRO	167.4950000000...	82.34210000000...	204962.0000000...	0.740730185697...	0.013554958738...
3	4	0301	OUEST	KARONGI	997.6259999999...	192.7410000000...	265487.0000000...	1.734042146808...	0.080742629718...
4	5	0205	SUD	NYAMAGABE	1095.430000000...	204.0500000000...	284852.0000000...	1.835887446403...	0.088674623992...
5	6	0306	OUEST	RUSIZI	962.8750000000...	257.4370000000...	331950.0000000...	2.316306846381...	0.077953079939...
6	7	0502	EST	NYAGATARE	1931.049999999...	280.0910000000...	255104.0000000...	2.519292301606...	0.156221970480...
7	8	0503	EST	GATSIBO	1718.490000000...	247.4319999999...	240566.0000000...	2.225708711195...	0.139043684719...
8	9	0304	OUEST	NYABIHU	533.9829999999...	127.8580000000...	268367.0000000...	1.150092335458...	0.043205296775...
9	10	0303	OUEST	RUBAVU	390.1309999999...	115.2920000000...	287579.0000000...	1.037060276863...	0.031566271862...
10	11	0402	NORD	GAKENKE	707.3940000000...	146.6860000000...	322043.0000000...	1.319447255202...	0.057237784654...
11	12	0305	OUEST	NGORORERO	682.1699999999...	171.7059999999...	282249.0000000...	1.544612002896...	0.055202171853...
12	13	0302	OUEST	RUTSIRO	1162.619999999...	184.8650000000...	269434.0000000...	1.663042445640...	0.094082074204...
13	14	0101	VILLE DE KIGALI	NYARUGENGE	134.5860000000...	106.3130000000...	236990.0000000...	0.956362847325...	0.010891596903...
14	15	0505	EST	KIREHE	1190.279999999...	191.4290000000...	229468.0000000...	1.722201652990...	0.096340838955...
15	16	0507	EST	BUGESERA	1296.670000000...	204.8180000000...	266775.0000000...	1.842670327904...	0.104952104625...
16	17	0506	EST	NGOMA	871.8010000000...	163.0329999999...	232165.0000000...	1.466785713926...	0.070560658520...
17	18	0307	OUEST	NYAMASHEKE	1179.299999999...	181.4960000000...	325032.0000000...	1.632919480310...	0.095459734200...
18	19	0204	SUD	HUYE	584.2649999999...	130.4699999999...	264295.0000000...	1.173874747922...	0.047300079320...
19	20	0202	SUD	GISAGARA	682.4059999999...	149.1030000000...	266424.0000000...	1.341473705963...	0.055249328463...
20	21	0501	EST	RWAMAGANA	685.1720000000...	135.5790000000...	209423.0000000...	1.219651495015...	0.055448275666...
21	22	0504	EST	KAYONZA	1813.210000000...	196.9290000000...	220802.0000000...	1.771475399563...	0.146726570834...
22	23	0201	SUD	NYANZA	675.3179999999...	175.7210000000...	176456.0000000...	1.580989569265...	0.054663730824...
23	24	0206	SUD	RUHANGO	629.7380000000...	163.6839999999...	210000.0000000...	1.472662258286...	0.050969261878...
24	25	0207	SUD	MUHANGA	650.7749999999...	179.2119999999...	340369.0000000...	1.612143453993...	0.052664092036...
25	26	0208	SUD	KAMONYI	658.6390000000...	169.8000000000...	292772.0000000...	1.527499154389...	0.053302180684...
26	27	0405	NORD	GICUMBI	833.4769999999...	199.8530000000...	362668.0000000...	1.797638086922...	0.067436969841...
27	28	0102	VILLE DE KIGALI	GASABO	431.2409999999...	110.1599999999...	318569.0000000...	0.990987147546...	0.034896897795...
28	29	0401	NORD	RULINDO	569.6770000000...	159.1949999999...	278310.0000000...	1.431975750659...	0.046095584951...
29	30	0203	SUD	NYARUGURU	1014.970000000...	188.6270000000...	233815.0000000...	1.697339066385...	0.082179775426...

Show All Features

The attribute table for the shapefile you added contains nine fields:

- OBJECTID: This is a unique object identifier for each row in the attribute table, which is assigned automatically by ArcGIS when an ESRI feature class is created in a geodatabase. This field exists in the shapefile because the shapefile was derived from an ESRI geodatabase dated April 2011, which was downloaded from the Web site for the Rwanda Ministry of Health (www.moh.gov.rw).
- GovtDistri: A unique administrative code assigned to each district by the Rwandan government. The values in this field are the same as those in the field GOVTCODE contained in the comma-delimited text file added to the QGIS project earlier in this exercise.
- PROVINCE: French province names from the MOH geodatabase.
- DISTRICT: District names in Kinyarwanda, the native language of Rwanda. These are the same district names used in the Rwanda DHS 2010.
- AREA_SQ_KM: Surface area of each district calculated in square kilometers.
- PERIMETER_: The length of the perimeter of each district in kilometers.
- POPULATION: Total district population from the 2000 census.

- Shape_Leng: The length of the perimeter of each district in decimal degrees. Note that the decimal degree, which represents an unprojected coordinate system, is an inconsistent and imprecise unit of measurement for calculating distances. The Shape Length field is created and updated automatically in ArcGIS feature classes, but not in shapefiles.
- Shape_Area: Surface area of each district in decimal degrees. The Shape Area field is created and updated automatically in ArcGIS feature classes, but not in shapefiles.

NOTE: The DISTRICT column in the shapefile table contains the same administrative names as the DHSDIST column in the attribute table containing HIV prevalence data from the Rwanda DHS 2010. The district names in the shapefile table, however, are spelled in all upper case letters. This is different from the district names in the attribute table containing the HIV prevalence data you want to join to the shapefile. As a result, you cannot use the district names as they are currently formatted, to join the HIV prevalence data to the shapefile.

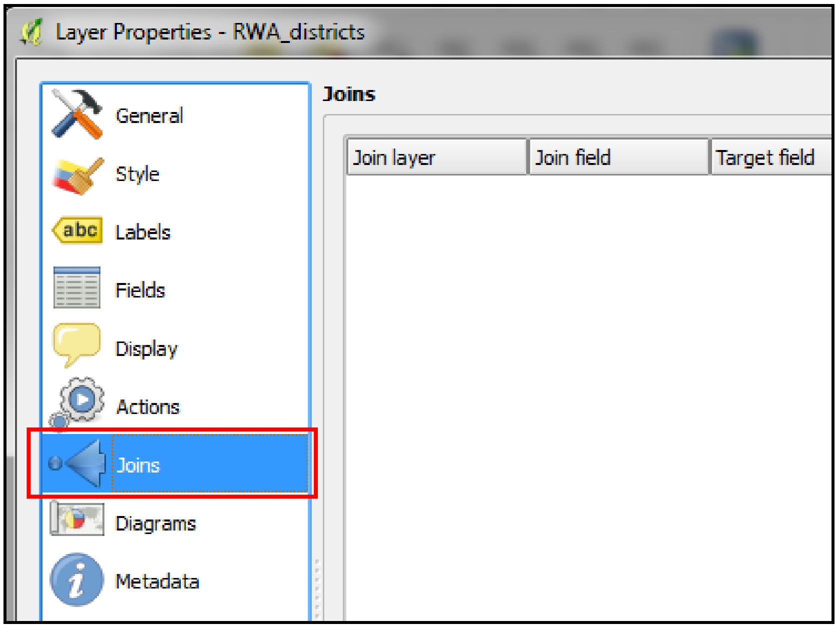
To resolve the problem you have two primary options: (1) create a new column in one of the tables to contain district names with the same capitalization as in the other table or (2) use an administrative code the two tables have in common. Fortunately, you have a common administrative code that can be used to join the two tables, GOVTCODE in the HIV prevalence table and GovtDistri in the shapefile table.

Step 3: Join the attribute table containing HIV prevalence data to the shapefile.

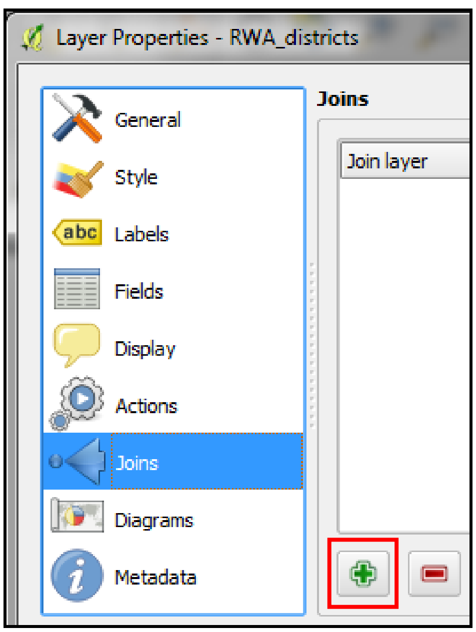
- First, open the Properties dialog box for the shapefile layer.

  - To do that, you have two options: (1) double-click on the layer name for the shapefile (RWA_districts) in the Layers panel, or (2) right-click on the shapefile layer name in the Layers panel and on the popup menu select "Properties."

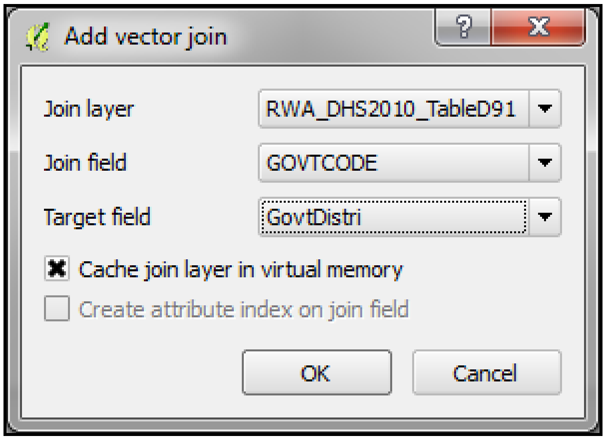
  - When the Properties window opens, highlight the "Joins" icon in the left-hand navigation panel by left-clicking on it (see below).



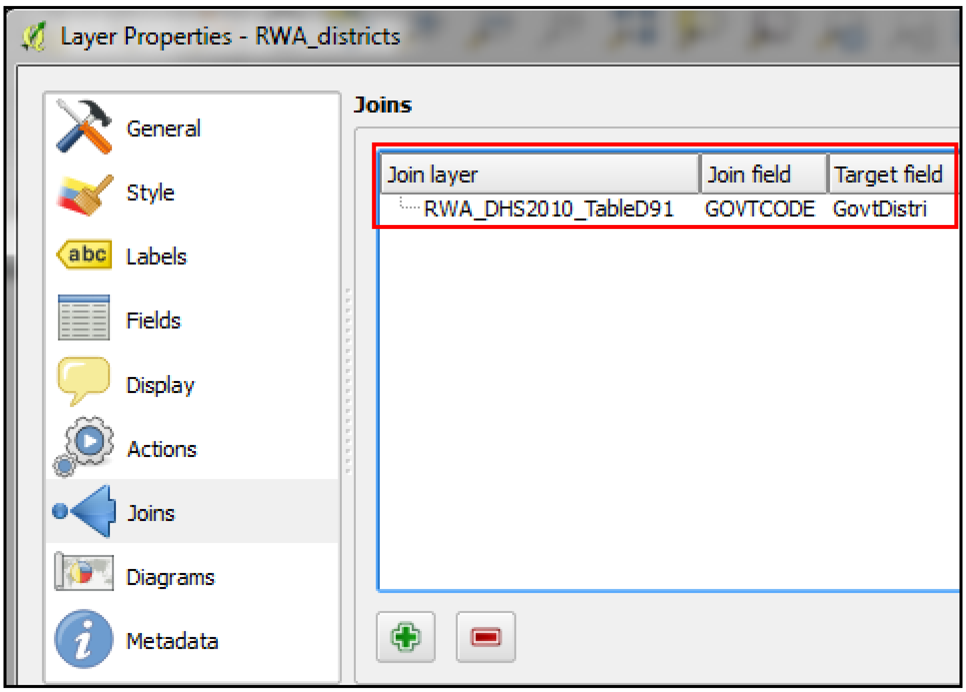
- Then, on the Joins screen within the Layer Properties window, click on the green plus sign (see below).



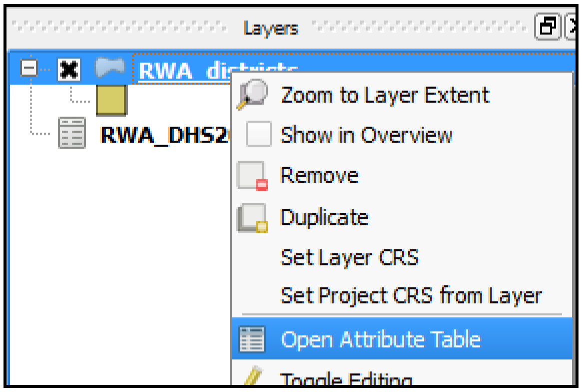
- In the "Add vector join" file dialog box that appears, use the pull-down menus to specify the join layer, join field, and target field as shown below. (NOTE: you will see a box checked that says "Cache join layer in virtual memory". The joined table will only be temporary and will need to be saved separately if it is to be used for other tasks in the future. This will be done at a later step.)



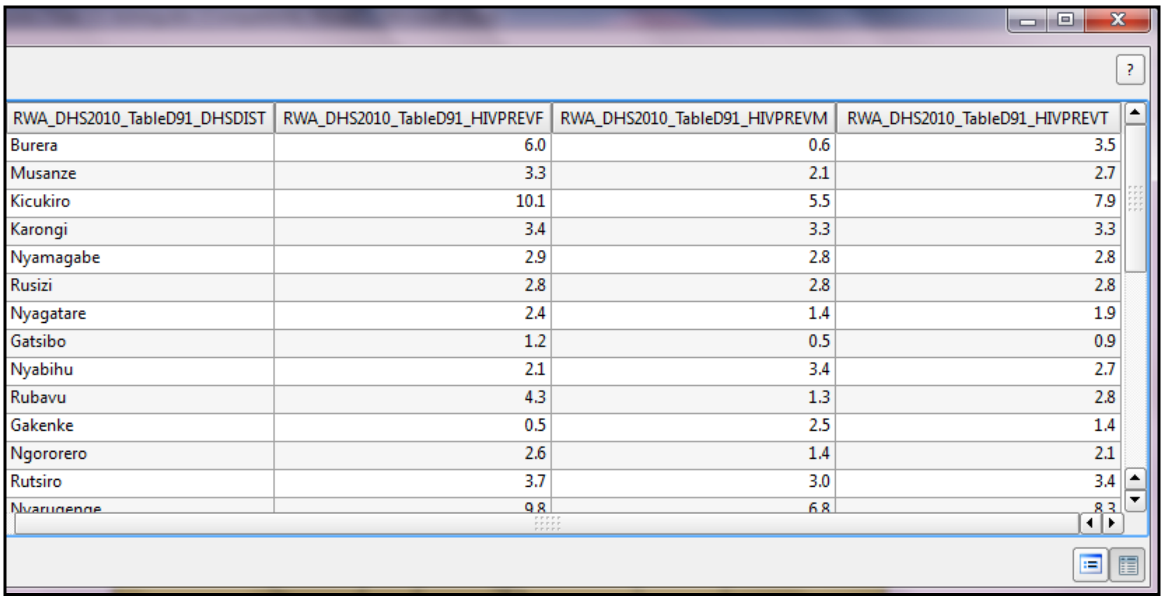
- Click on OK and inspect the Layer Properties window to verify that the join has been accomplished (see below).



- To view the result of the join, click on OK once again to close the Layer Properties window. Next, right-click on the name of the shapefile layer (RWA_districts) in the left-hand Layers panel and choose “Open Attribute Table” (see below).



- When the shapefile's attribute table opens, scroll to the right to view the new columns containing HIV prevalence data (see below).



RWA_DHS2010_TableD91_DHSDIST	RWA_DHS2010_TableD91_HIVPREVF	RWA_DHS2010_TableD91_HIVPREVM	RWA_DHS2010_TableD91_HIVPREVT
Burera	6.0	0.6	3.5
Musanze	3.3	2.1	2.7
Kicukiro	10.1	5.5	7.9
Karongi	3.4	3.3	3.3
Nyamagabe	2.9	2.8	2.8
Rusizi	2.8	2.8	2.8
Nyagatare	2.4	1.4	1.9
Gatsibo	1.2	0.5	0.9
Nyabihu	2.1	3.4	2.7
Rubavu	4.3	1.3	2.8
Gakenke	0.5	2.5	1.4
Ngororero	2.6	1.4	2.1
Rutsiro	3.7	3.0	3.4
Nyarugenge	9.8	6.8	8.3

Note that the column names are very long, as they contain a combination of the input file name (RWA_DHS2010_TableD91) and the column names in the input file (DHSDIST, HIVPREVF, etc.).

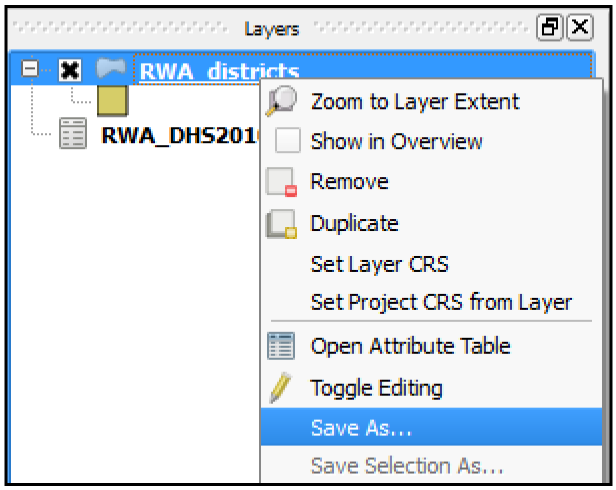
Note also that the join is temporary; the variables for HIV prevalence will not be permanently added to the shapefile table unless you save out a new shapefile using the results of this join.

Step 4: Create a new shapefile from your joined data, and add it to the QGIS project.

- To make the join permanent, save a copy of the shapefile containing the temporarily joined HIV prevalence data. To do so, perform the following tasks:

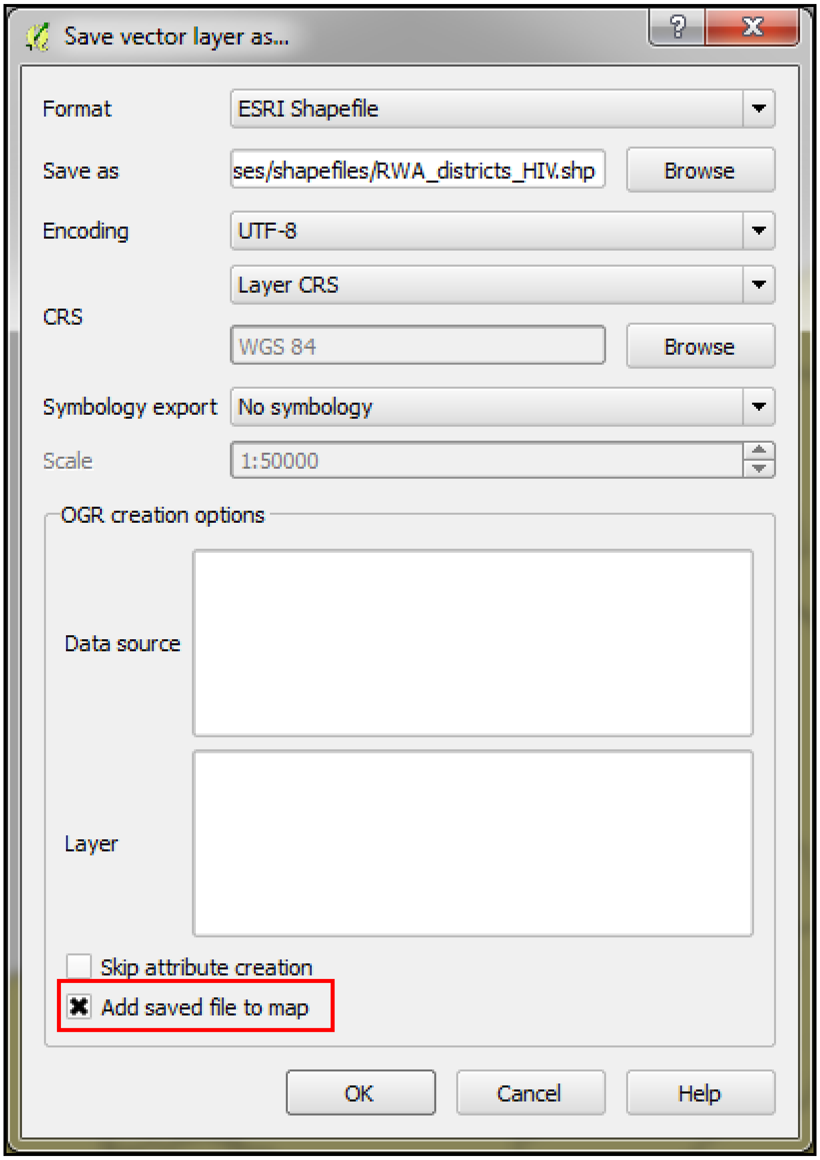
- Close the shapefile attribute table by clicking on the X in the red box in the upper-right corner of the attribute table. 

- Right-click on the name of the shapefile layer in the Layers panel and select "Save As."



- In the Save As dialog box, choose the following settings:
    - Format: ESRI Shapefile
    - Save as: Session gis_exercises\shapefiles\RWA_districts_HIV.shp
    - Add saved file to map: Click on this box so that an X appears in it.

- Leave the remaining settings as they are so that the Save As dialog box looks like the one below.



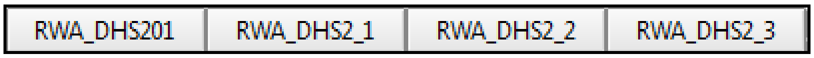
Note: For now it is important to make sure that the choice for Encoding is "UTF-8" and the choice for CRS is "Layer CRS." This will keep the output file in an unprojected coordinate reference system (CRS), which stores features using latitude and longitude in degrees.

- Click on OK and verify that the new shapefile has been added to the Layers panel in QGIS.

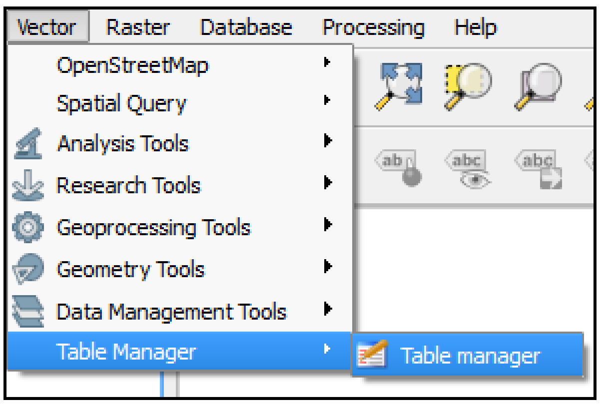
Step 5: Revise the new column names to be more meaningful.

- If you recall, the new column names generated by QGIS automatically in the shapefile table during the join were very long. In the process of creating the new shapefile after the join, however, QGIS automatically shortened these column names to a maximum of 10 characters. This created column names that are not very descriptive and therefore not very useful. In this step we will use Table Manager to create new, more useful names for these columns.

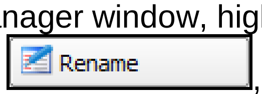
- To verify the new column names corresponding to the HIV prevalence indicators, right-click on the name of the new shapefile layer (RWA_districts_HIV) in the Layers panel and scroll to the right. You should see columns with abbreviated names as shown in the column headings below.



- To change the column names, close the shapefile table, highlight the name of the new shapefile layer in the Layers panel, and start the **Table Manager** plug-in by going to the main menu in QGIS and selecting Vector > Table Manager > Table Manager (see below).



NOTE: If you do not see the Table Manager plug-in on the Vector menu, please refer back to Exercise 2.1, the "installing plugins" section.

- In the Table Manager window, highlight each new variable name, click on the Rename button , and make the following changes:

  - From RWA_DHS201 to DHSDIST. This is the DHS district name.
  - From RWA_DHS2_1 to HIVPREVF. This is female HIV prevalence.
  - From RWA_DHS2_2 to HIVPREVM. This is male HIV prevalence.
  - From RWA_DHS2_3 to HIVPREVT. This is total HIV prevalence.

- Save the changes to the shapefile table by clicking on the Save button at the bottom of the Table Manager dialog box.

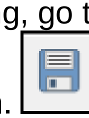
- A new dialog box will open, which will ask you whether you want to keep the layer style currently applied. Answer Yes and then click on the Close button at the bottom of the Table Manager dialog box.

- Verify the changes to the shapefile table by right-clicking on the shapefile layer name in the Layers panel, selecting Open Attribute Table, and scrolling to the right (see below).

DHSDIST	HIVPREVF	HIVPREVM	HIVPREVT
Burera	6.0	0.6	3.5
Musanze	3.3	2.1	2.7
Kicukiro	10.1	5.5	7.9
Karongi	3.4	3.3	3.3
Nyamagabe	2.9	2.8	2.8
Rusizi	2.8	2.8	2.8
Nyagatare	2.4	1.4	1.9
Gatsibo	1.2	0.5	0.9
Nyabihu	2.1	3.4	2.7
Rubavu	4.3	1.3	2.8
Gakenke	0.5	2.5	1.4

- You now have a shapefile containing attributes for each district which you can use for mapping and analysis of HIV data.

Step 6: Save the QGIS project.

- To save the QGIS project with which you have been working, go to the main QGIS menu and select Project > Save or click on the Save button. 

    - In the file dialog box that appears, specify the following output name:

      Session 3 HIV Data Sources\exercises\qgis_exercise_3.1.qgs

- Quit QGIS.

    - To close QGIS, go to the main menu and select Project > Exit QGIS or use a combination of the Ctrl and Q keys (see below).

NOTE on joining other types of data files:

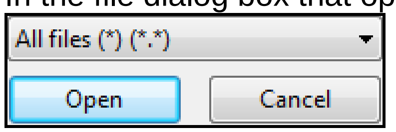
QGIS can also join other types of data files to shapefiles. All that is required is that both files (the attribute data and the shapefile) have a field in common on which to perform the join (such as a geographic identifier code).

- Use Excel files (with the file name extension **.xlsx** or **.xls**), or DBF files, which are dBASE files with the file name extension **.dbf**.

To use these files without losing the intended formatting for each of the data fields, make sure the column types (cell formats) are specified in the source files before attempting to join them.

To add one of these file to QGIS, use the following steps:

- Click on the Add Vector Layer button. 
- In the Add Vector Layer dialog box, click on Browse. 

- In the file dialog box that opens, change the file to “All files (*)(*.*).”


- Specify the file name and click Open twice to open the file. Proceed exactly as you did when joining a shapefile to a CSV file.

END

GIS Techniques for Monitoring and Evaluation of HIV/AIDS and Related Programs

Exercise 3.2

Create a Color-Shaded Map

This training was developed as part of a joint effort between MEASURE Evaluation and MEASURE DHS, with funding provided by USAID and PEPFAR.

Exercise 3.2: Create a Color-Shaded Map

Summary: In this exercise, you will learn how to use the shapefile created during exercise 3.1 to produce color-shaded maps showing HIV prevalence by district for Rwanda. Color-shaded maps are referred to in the geographic lexicon as "choropleth" maps, from the Greek words khōra (region) and plēthos (multitude).

A choropleth, or color-shaded, map is a quick and effective way to visualize data associated with administrative divisions, such as DHS regions or districts.

Objectives:

- Open a QGIS project that is in progress and save the existing QGIS project as a new project, removing unnecessary data layers.
- Learning about classing (categorizing) your data into equal intervals.
- Create a choropleth map showing HIV prevalence by district.

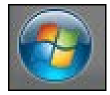
Bonus Work: Explore the different classification methods available for creating choropleth maps.

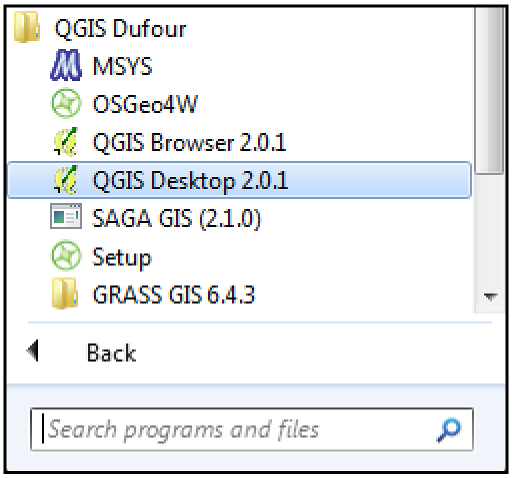
Requirements:
To complete these exercises, you will need to have QGIS and several plugins installed on your machine, as instructed in exercise 2.1. You will also need to have downloaded and unzipped the data files associated with the exercise. Remember, these files can reside anywhere on your computer, but must be kept together, with the same original file structure.

Step 1: Open a QGIS project that is in progress.

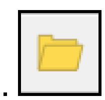
☐ Launch QGIS.

- Option 1: Double-click on the QGIS shortcut on the desktop (or on the Windows taskbar if you have pinned a copy of the QGIS shortcut to there).

- Option 2: In Windows, click on the Start button  and select All Programs > QGIS Dufour > QGIS Desktop 2.01 (see below).



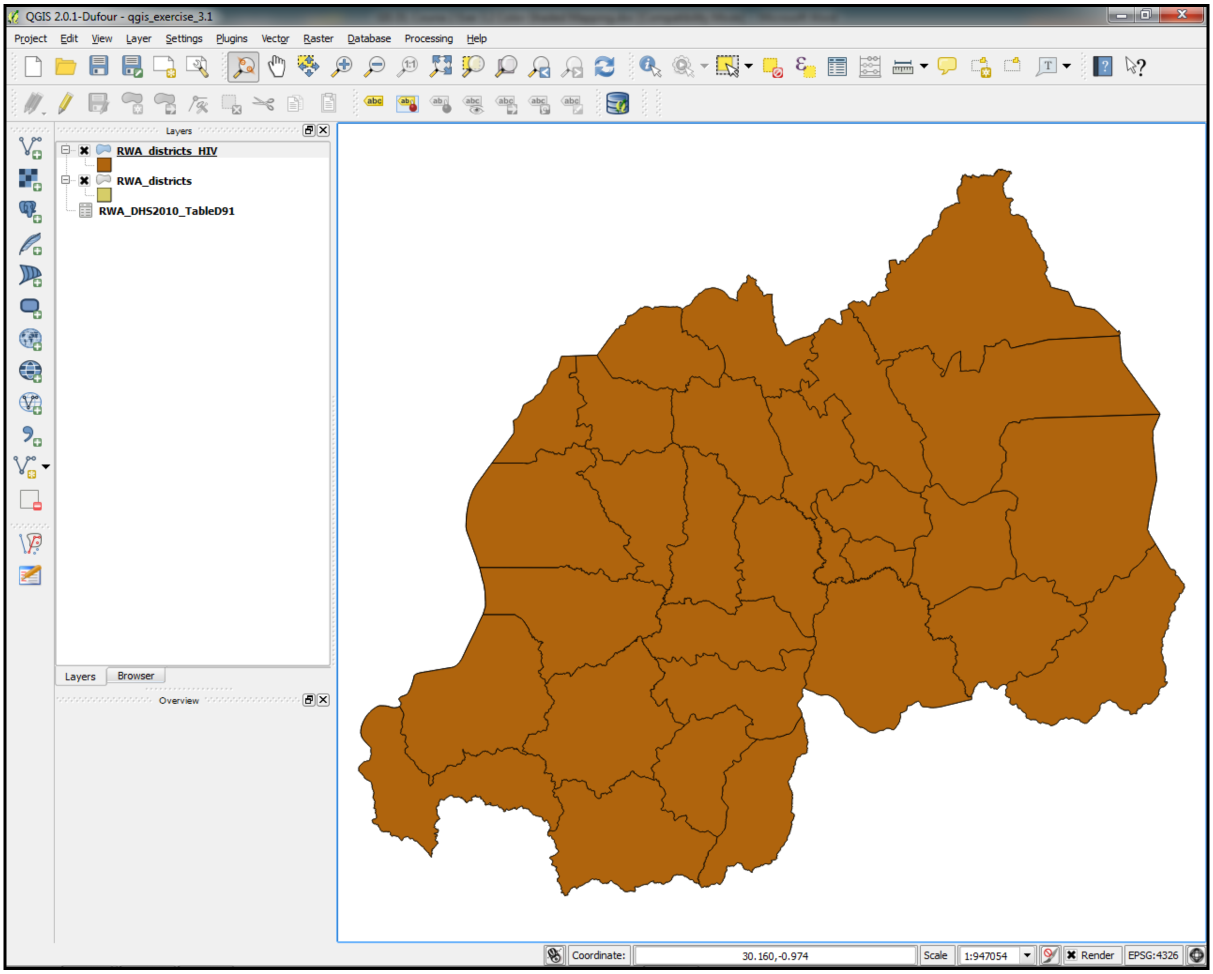
You should see the splash screen for QGIS Desktop 2.01 and then the main QGIS interface.

☐ On the main QGIS menu, select Project > Open or click on the folder icon. 

In the file dialog box that opens, navigate to the folder containing the QGIS project file created in exercise 3.1, highlight the project file name and click on Open. The full path to the QGIS project file should contain the following:

Session 3 HIV Data Sources\exercises\qgis_exercise_3.1.qgs

After opening the QGIS project, you should see a screen that looks like the one that follows.



Step 2: Save the existing QGIS project as a new project.

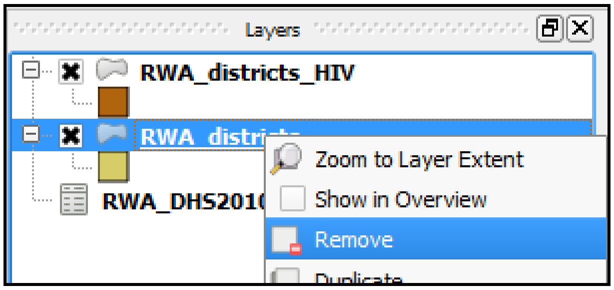
- To save a new version of the QGIS project for exercise 3.2, on the main QGIS menu choose Project > Save As and specify the following output file name:

  Session 3 HIV Data Sources\exercises\qgis_exercise_3.2.qgs

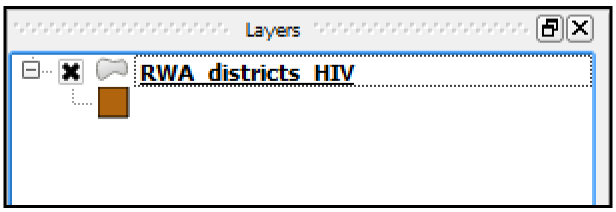
  - After specifying the new output file name, click on the Save button.

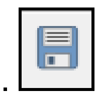
Step 3: Remove unnecessary data layers.

- The new project contains a couple of layers that are now unnecessary: the layers for the comma-delimited text file containing HIV prevalence data and for the original shapefile with no HIV prevalence data yet attached. To remove these unnecessary layers from the current project, right-click on each layer name in the Layers panel and select the Remove option (see below).



- After removing the two layers, you should be left with a single layer corresponding to the district-level shapefile that now contains HIV prevalence data from the Rwanda DHS 2010 (see below).

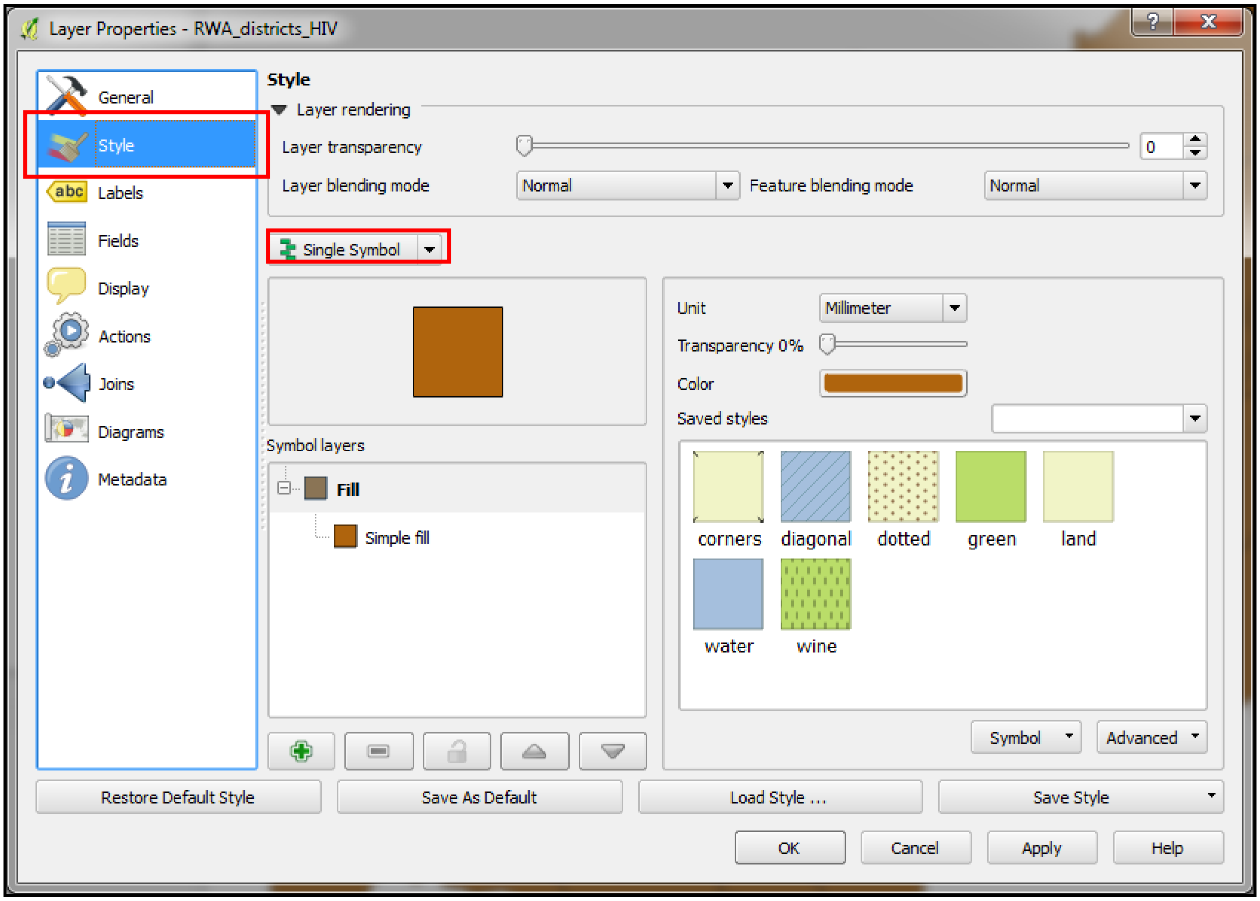


- To save the changes to the project file, on the main menu click on Save. 

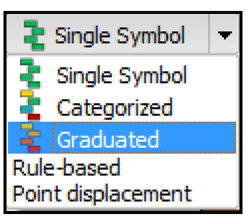
Step 4: Create a choropleth map showing HIV prevalence by district.

- To begin the color-shading process, open the properties window for the shapefile layer by double-clicking on the layer name in the Layers panel or by highlighting the layer name in the Layers panel and on the main QGIS menu selecting Layer > Properties.

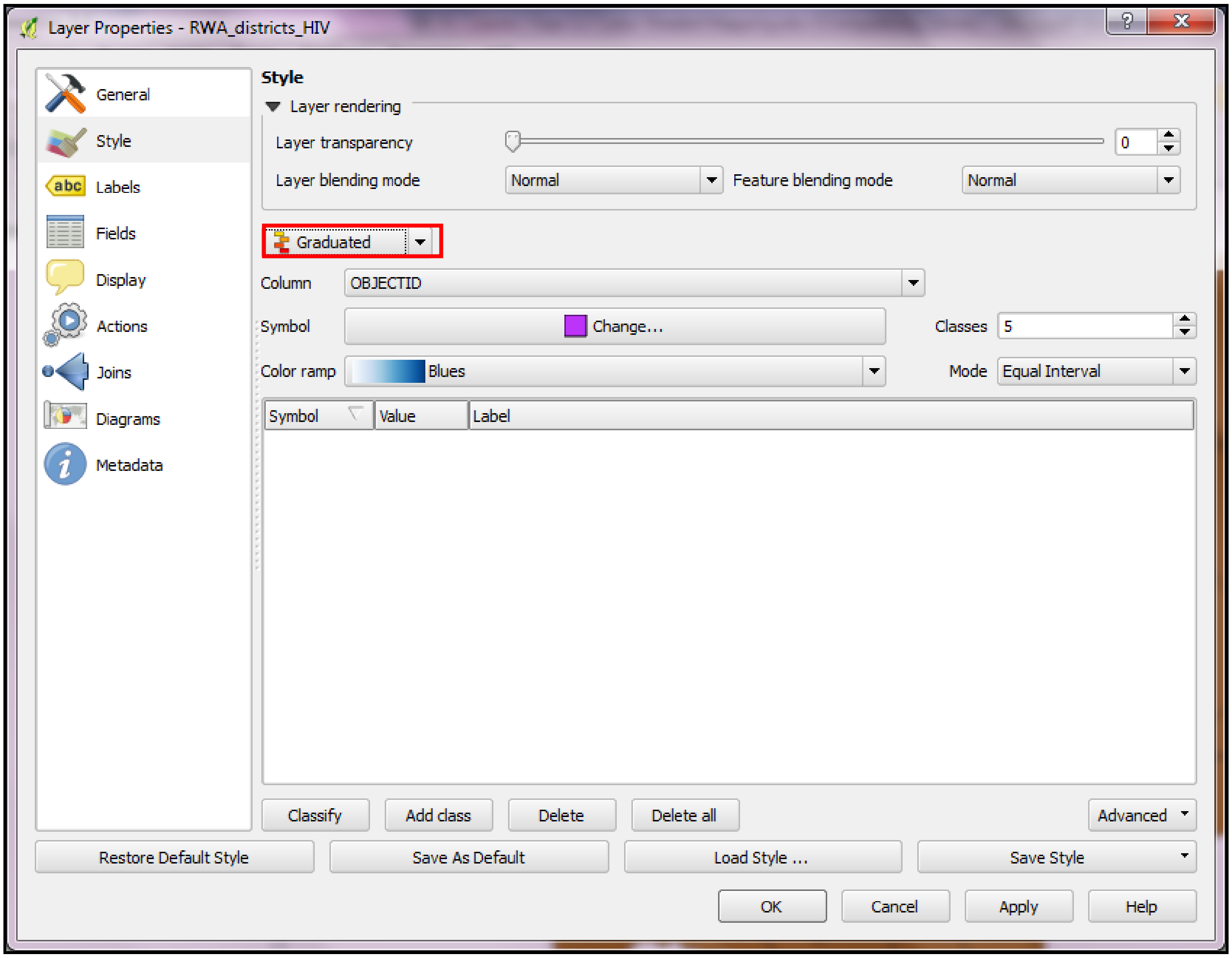
  - After the Layer Properties window opens, left-click on the Style icon in order to highlight it. You should see a screen that looks like the one below.



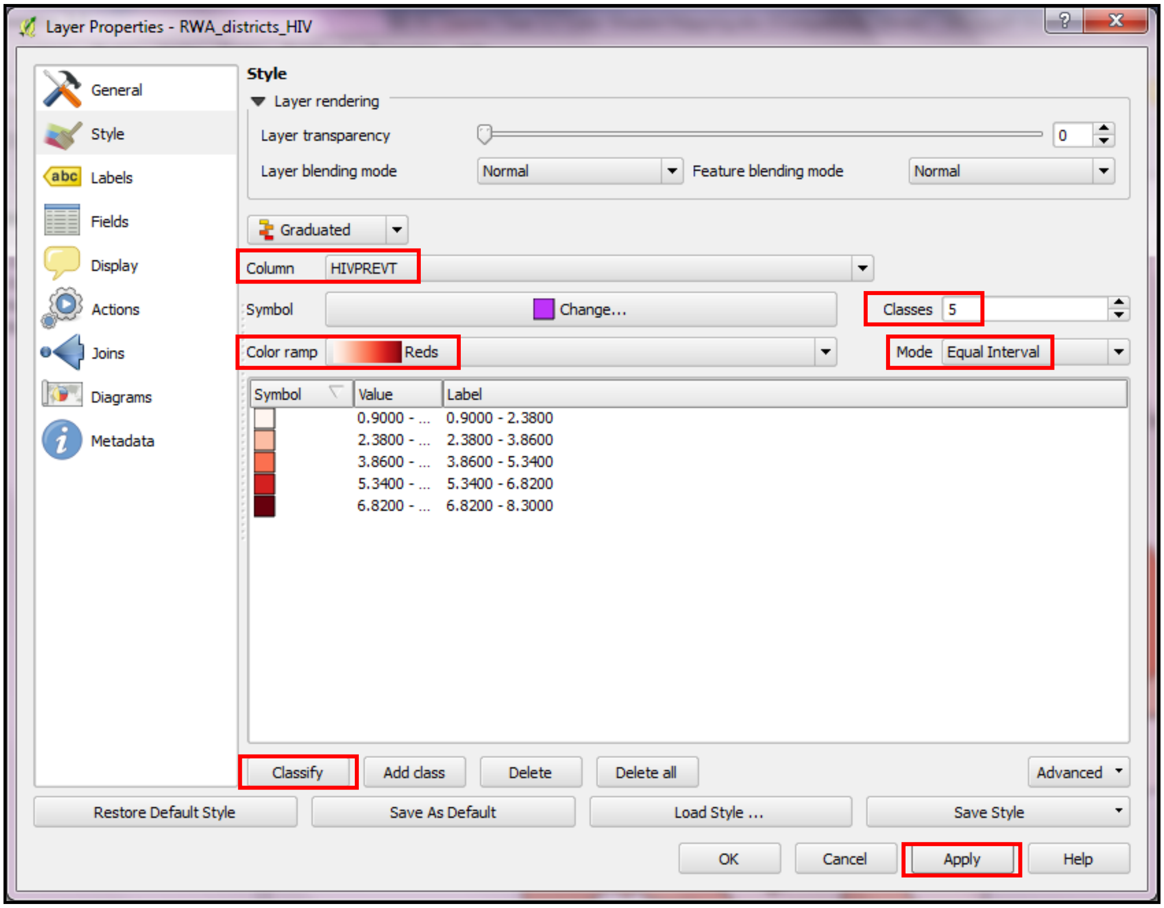
- To be able to use different colors for the different ranges of values for HIV prevalence, change the selection in the pull-down menu above the large colored square from Single Symbol to Graduated (see below).



The Layer Properties window should now look like the one below.



- To map total HIV prevalence by district, in the pull-down menu to the right of the word Column select the column HIVPREVT.
- Leave the number of classes at 5 and keep the classification mode set to Equal Interval.
- Change the color ramp to Reds, then click on the Classify and Apply buttons. The dialog box should resemble the following.



- Click on the OK button to close the dialog box and see the choropleth map of total HIV prevalence by district.

Note that the equal intervals classification method arranged the HIV prevalence data from the DHS final report into classes containing an equal range of values. The formula for calculating the range of values to use for each class is as follows:

Total range of data values / number of classes

Since the range of total HIV prevalence values was 7.4 (the maximum, 8.3, minus the minimum, 0.9) and the number of classes selected was 5, the range of values assigned to each class was calculated as follows:

7.4 / 5 = 1.48

Class 1 = 0.90 to 2.38
Class 2 = 2.38 to 3.86
Class 3 = 3.86 to 5.34
Class 4 = 5.34 to 6.82
Class 5 = 6.82 to 8.30

Note: QGIS correctly calculates the maximum value for class ranges, but allows an overlap in the maximum value for one class and the minimum value for the next class. For example, class 1 ends in a value of 2.38 and class 2 begins with the same value of 2.38. The same value should not be shared by two classes. To correct this problem:

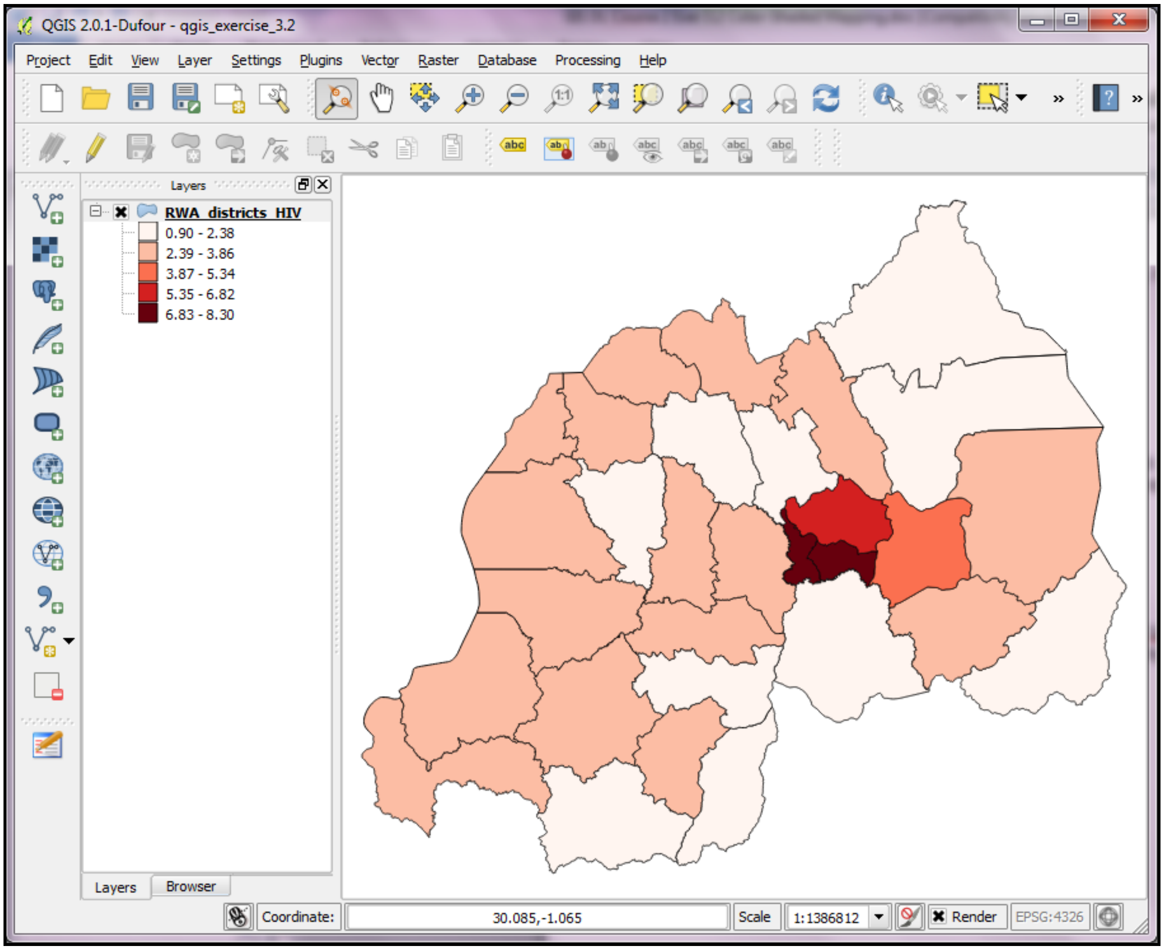
- Open the Layer Properties window, double-click on each range of values in the Values column—beginning with class 2—and increase the lower value in the range by 0.01.

- You will also need to change the label to match the new lower limit. While you are editing the ranges displayed in the labels, you can remove the unnecessary zeroes at the ends of the labels. To do this, double-click on each range of values in the Label column and remove the zeroes.

  The resulting Value and Label columns should look as follows:

Symbol	Value	Label
	0.9000 - 2.3800	0.90 - 2.38
	2.3900 - 3.8600	2.39 - 3.86
	3.8700 - 5.3400	3.87 - 5.34
	5.3500 - 6.8200	5.35 - 6.82
	6.8300 - 8.3000	6.83 - 8.30

- To save the changes, click on Apply, then OK. This will close the Layer Properties window. You should have a map that resembles the one below.



Rwanda, Total HIV Prevalence among Adults Age 15-49 by District, 2010, Equal Intervals.

- The preceding map, in which darker colors represent a higher HIV prevalence, shows that HIV prevalence in 2010 in Rwanda is not evenly distributed at the district level.

  - HIV prevalence is highest in the center of the country, which corresponds to the three urban districts that make up the province of the City of Kigali.

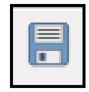
  - There is only one district that corresponds to the middle range of HIV prevalence (class 3, which is 3.87 to 5.34 percent of the total adult population). This district is the one adjacent to the east side of the three districts of the City of Kigali.

  - HIV prevalence is below 4 percent in the majority of districts, and in the majority of Rwanda districts, HIV prevalence in 2010 is between 2.39 and 3.86 percent.

- Equal intervals are easy to interpret, but depending on the data set classified, there can be classes generated with few or no values. Explore the **bonus work** below to find out more about other methods of data classification for choropleth maps.

Step 5: Save the project.

- Save the QGIS project.

    - To save the QGIS project with which you have been working, go to the main QGIS menu and select Project > Save or click on the Save button. 

Bonus Work: Explore the different classification methods available for creating choropleth maps.

- **Hint:** To be able to compare new classification methods for visualizing HIV prevalence by district for Rwanda, add a new layer for each new map using the same shapefile as the one used to create the equal intervals map.

    - Shapefile name: Session 3 HIV Data Sources\exercises\shapefiles\RWA_districts_HIV.shp

- After loading a new layer, open the Layer Properties window for that layer and experiment with the classification methods. There are a total of five classification methods available in QGIS:

    - **Equal Interval:** This is the method used in the above mapping exercise. It divides the total data range into classes (intervals) of equal size. This method works best for data which is evenly distributed, or continuous. It will help outliers show up—there may only be a few observations in one class, and a large number in another.

    - **Quantile (Equal Count):** This method attempts to place the same number of observations into each class. This classification method is good for visualizing a distribution of highly variable values in an ordinal manner, such as low, medium, and high. A danger in using quantiles is that it can hide natural groupings or clusters of values by forcing data into classes with a wide range of values.

    - **Natural Breaks (Jenks):** This classification method seeks to minimize the range of values within a class with respect to its mean (average) value and to maximize the difference between the mean values for each class. This classification method adapts to the natural distribution of a range of values, but can be more difficult to interpret and explain.

    - **Standard Deviation:** Assigns values to classes based on standard deviation. This can be an effective method for identifying outlier (very high or very low) values.

- **Pretty Breaks:** This classification method can use a floating point number to begin the class containing the lowest values and to end the class containing the highest values, but will use integer values (whole numbers) to define the remaining class limits. This is a good method for isolating the outliers in a distribution of values while simultaneously using a simple classification scheme for middle values which is easy to interpret and explain.

- Quit QGIS.

  - To close QGIS, go to the main menu and select Project > Exit QGIS or use a combination of the Ctrl and Q keys.

END

GIS Techniques for Monitoring and Evaluation of HIV/AIDS and Related Programs

Exercise 4.1

Create New Variables in a Shapefile

**This training was developed as part of a joint effort between MEASURE Evaluation and MEASURE DHS, with funding provided by USAID and PEPFAR.*

Exercise 4.1: Create New Variables in a Shapefile

Summary: The creation of new variables from existing geographic data can be used to visually and mathematically analyze many types of spatial data, and thus provide highly valuable information for program decision making. In this exercise you will learn how to calculate new variables from existing data stored in a shapefile.

Objectives:

- The 2 new variables you will calculate in this exercise will be
  - Estimated percent of total population comprised of women of reproductive age (age 15-49) by district in Rwanda in 2010.
  - Estimated count of women of reproductive age who are HIV positive by district in Rwanda in 2010.
- You will learn to start and use an Editing session.
- You will be introduced to the Field Calculator, where you will write an expression to calculate a new field.

Requirements:

To complete these exercises, you will need to have QGIS and several plugins installed on your machine, as instructed in exercise 2.1. You will also need to have downloaded and unzipped the data files associated with the exercise. Remember, these files can reside anywhere on your computer, but must be kept together, with the same original file structure.

Step 1: Launch QGIS.

- Option 1: Click on the desktop shortcut.

- Option 2: Click on Start > All programs > QGIS Dufour > QGIS Desktop 2.0.1.

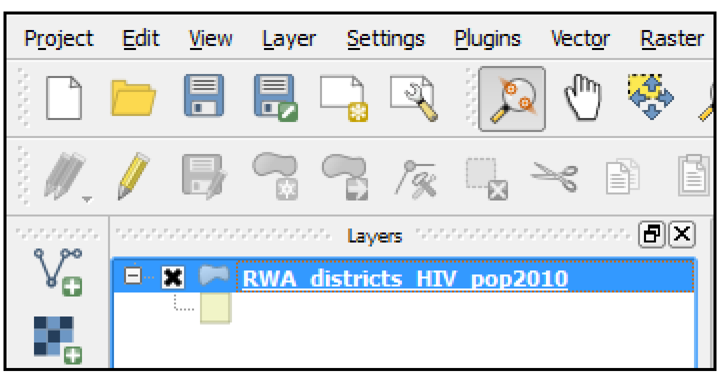
Step 2: Open a shapefile containing source data for calculation of new variables.

- Click on the Add Vector Layer button. 

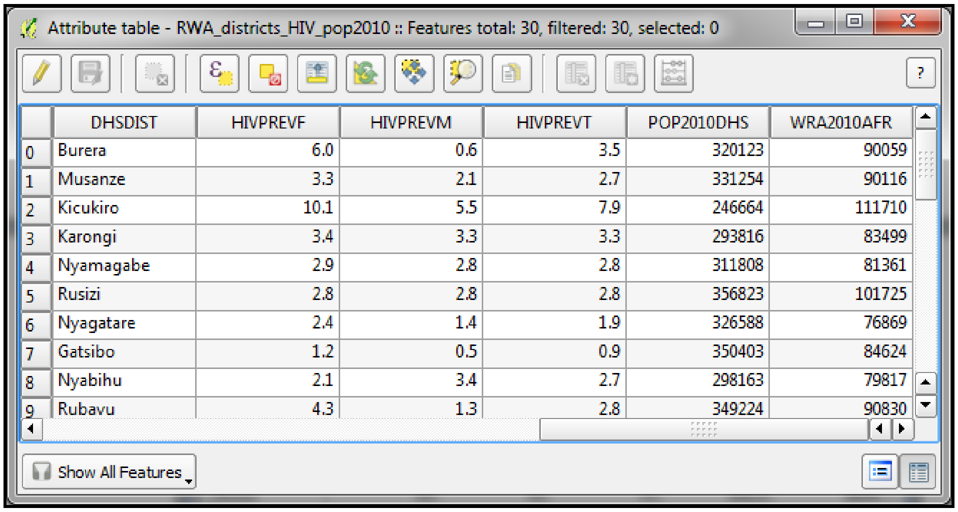
  - Make sure the source type is set to File and click on Browse, and in the file dialog box that appears:

    — Use the pull-down menu in the lower-right corner to set the file type to "ESRI Shapefiles [OGR] (*.shp *.SHP)."
    — Navigate to the folder "Exercises 4.1, 4.2 and 4.3\shapefiles" and find the shapefile **RWA_districts_HIV_pop2010.shp**, select it, and click on Open. NOTE: these shapefiles should be located wherever you unzipped the folder downloaded from Dropbox for this exercise. Its name and location were optional, but the default name was "Exercises 4.1, 4.2 and 4.3".

  - After clicking on Open for a second time, you should see a new layer in the Layers window corresponding to the shapefile you just added.



- View the contents of the attribute table for the new QGIS layer. (Hint: right-click on the layer name and select Open Attribute Table.) It should look something like this:



	DHSDIST	HIVPREVF	HIVPREVM	HIVPREVT	POP2010DHS	WRA2010AFR
0	Burera	6.0	0.6	3.5	320123	90059
1	Musanze	3.3	2.1	2.7	331254	90116
2	Kicukiro	10.1	5.5	7.9	246664	111710
3	Karongi	3.4	3.3	3.3	293816	83499
4	Nyamagabe	2.9	2.8	2.8	311808	81361
5	Rusizi	2.8	2.8	2.8	356823	101725
6	Nyagatare	2.4	1.4	1.9	326588	76869
7	Gatsibo	1.2	0.5	0.9	350403	84624
8	Nyabihu	2.1	3.4	2.7	298163	79817
9	Rubavu	4.3	1.3	2.8	349224	90830

As you can see, the shapefile you opened contains the same fields as the one you created in exercise 3.1, but also contains two additional fields as described below:

- POP2010DHS: Estimated total population by district used for the Rwanda DHS 2010 sampling frame, which was based on the 2012 census preparatory frame for Rwanda. Source: Rwanda DHS 2010, Table A.1.

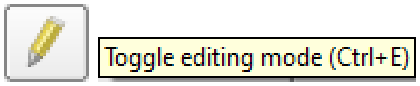
- WRA2010AFR: Estimated number of women of reproductive age (age 15-49) in Rwanda in 2010 by district. Source data were obtained as a 1 km by 1 km raster (grid-based) dataset from AfriPop. For more information on the AfriPop data collection methodology, please see www.afripop.org.

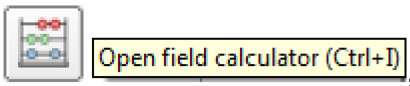
Estimates for women of reproductive age at the district level were extracted from the AfriPop dataset using district boundaries downloaded from the Rwanda Ministry of Health Web site (www.moh.gov.rw) in September 2011 in combination with the zonal statistics plug-in within QGIS (see below).

The AfriPop dataset uses a different kind of data than what we've previously been working with in QGIS. You may recall that there are 2 main ways of representing geographic data: **vector** and **raster**. We've been working mostly with vector data. The AfriPop dataset uses raster data. The QGIS zonal statistics plug-in is an excellent tool for extracting information from a raster (grid-based) dataset based on aggregating data from the raster dataset by administrative divisions, including health districts. More details about this plug-in can be found in the QGIS manual; we do not cover raster data calculations in this exercise.

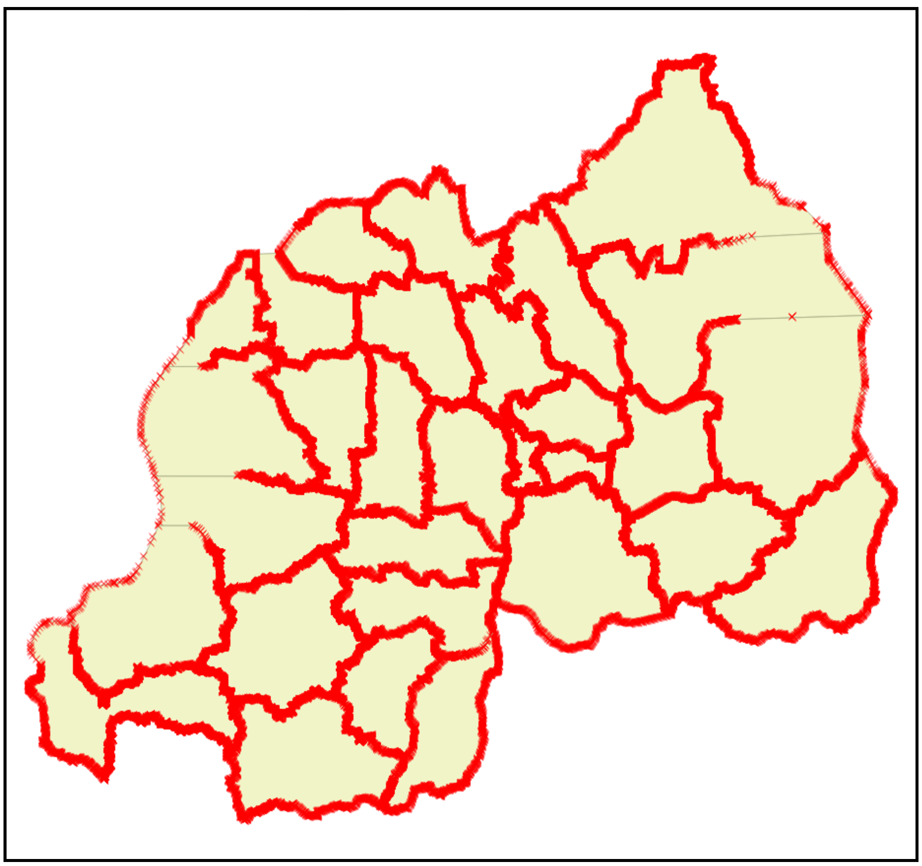
Step 3: Calculate new variables.

☐ To start an editing session within the attribute table for the shapefile, with the shapefile's attribute table still open, click on the button that is decorated with a pencil and that is labeled "Toggle editing mode (Ctrl + E)" (see below). Note: Move your cursor over the button to see the label, which will appear in a pop-up box like an alt tag over an image in a Web browser.



You will know you have successfully entered editing mode when two things happen: (1) the vertices for Rwanda districts, which are pairs of geographic coordinates that mark each point making up the polygons (areas) of the district boundaries, will turn to red Xs in the map window (see below) and (2) several new buttons will be activated in the attribute table window, including one for opening a field calculator 

, which will calculate values in a new or existing column.

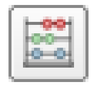


☐ Calculate the estimated percent of the total population that corresponds to women of reproductive age by district in Rwanda in 2010.

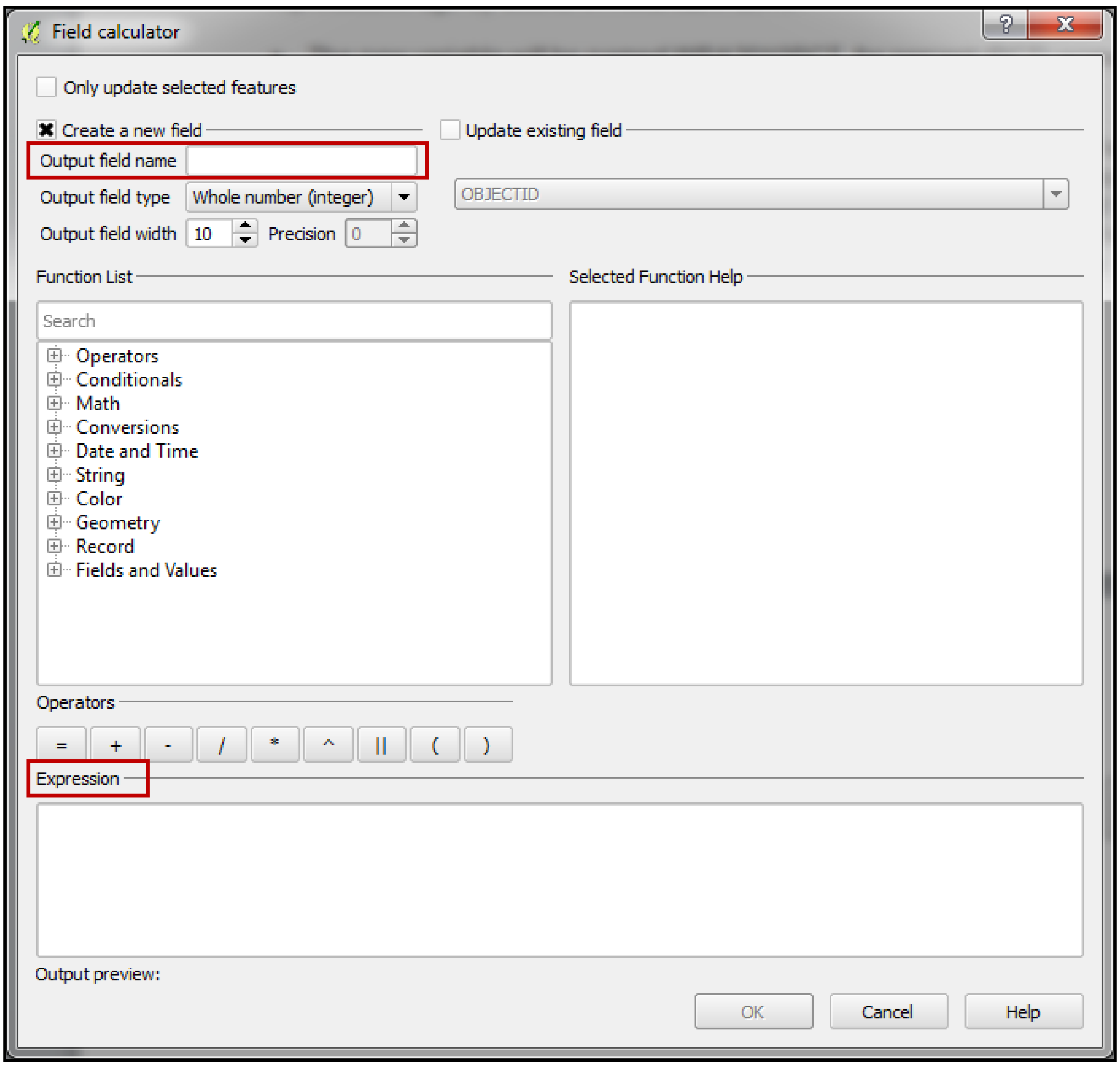
- The new variable will be named WRA2010PCT, for percent (PCT) women of reproductive age (WRA) in 2010.

- To calculate the new variable, you can divide the estimated number of women of reproductive age by district (variable WRA2010AFR) by the estimated total population by district used for the Rwanda DHS 2010 sampling frame (variable POP2010DHS) and multiply that by 100. The general formula will be as follows:

  WRA2010PCT = (WRA2010AFR / POP2010DHS) * 100

- Click on the "Open field calculator" button. 

  You should see a screen that looks like the following:



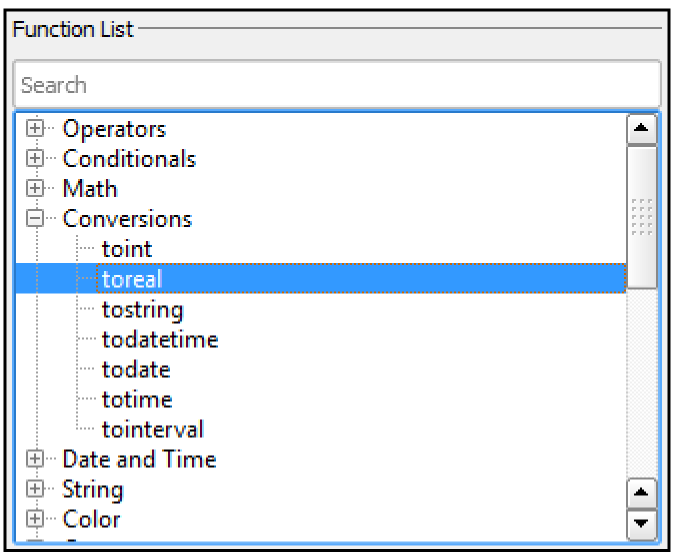
- To enter the name of the new field, make sure the box is selected for the "Create a new field" option.In the box next to the label "Output field name" enter the name WRA2010PCT. This new field name represents the left side of the general formula being used for this calculation. (The right side of the general formula will be used to fill out the Expression window)

- In the pull-down menu for "Output field type" select "Decimal number (real)." This will specify a floating point variable to hold percentages.

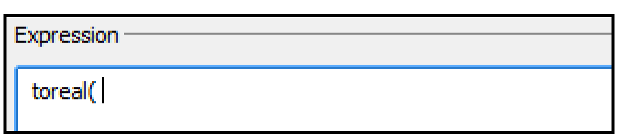
- Change the "Output field width" to 5 and the "Precision" to 1. This field size and precision will accommodate the theoretical maximum value of 100.0 (100 percent). (Note: If you set the output field width to 4, the maximum value that can be stored is 99.9, as the width includes the place allocated for the decimal.)

- Now it is time to perform the calculation. To specify the formula to use for the calculation, move to the Expression window and follow these steps:

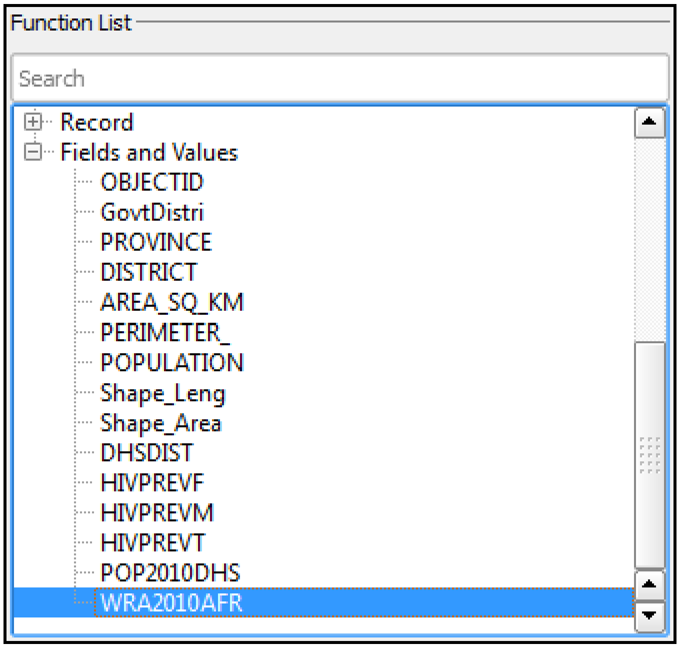
— To convert the values for estimated number of women of reproductive age from integer to decimal, which is required by QGIS 2.0 to produce floating point output, in the Function List box expand the list of Conversions and highlight the option "toreal" (convert to real/floating point number).



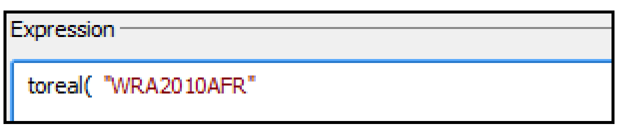
— To have the "toreal" expression appear in the Expression box, double-click on the highlighted option. The new expression should look like the one in the following graphic.



— To select the variable for the estimated number of women of reproductive age, return to the Function List box, expand the list of choices under Fields and Values, and highlight WRA2010AFR (see below).

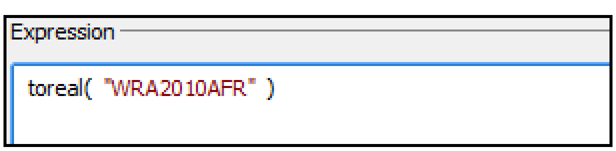


— To have the WRA2010AFR variable appear in the Expression box, double-click on the highlighted option. The new expression should look like the one below.

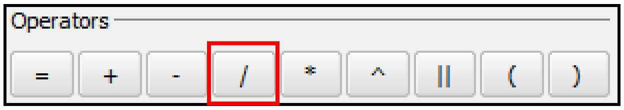


— To finish the expression for converting WRA2010AFR from integer to real, single-click on the operator button for Close Bracket (see below).

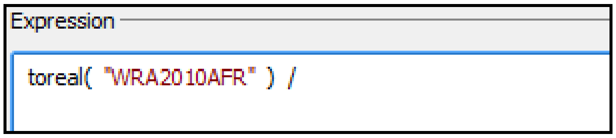
The expression should look like the following.



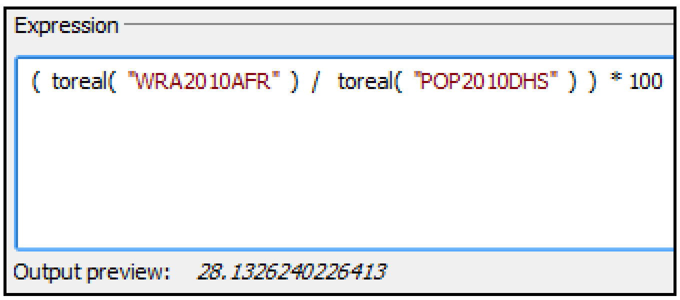
— To insert the division operator, single-click on the corresponding operator button (see below).



The expression should look like the following.

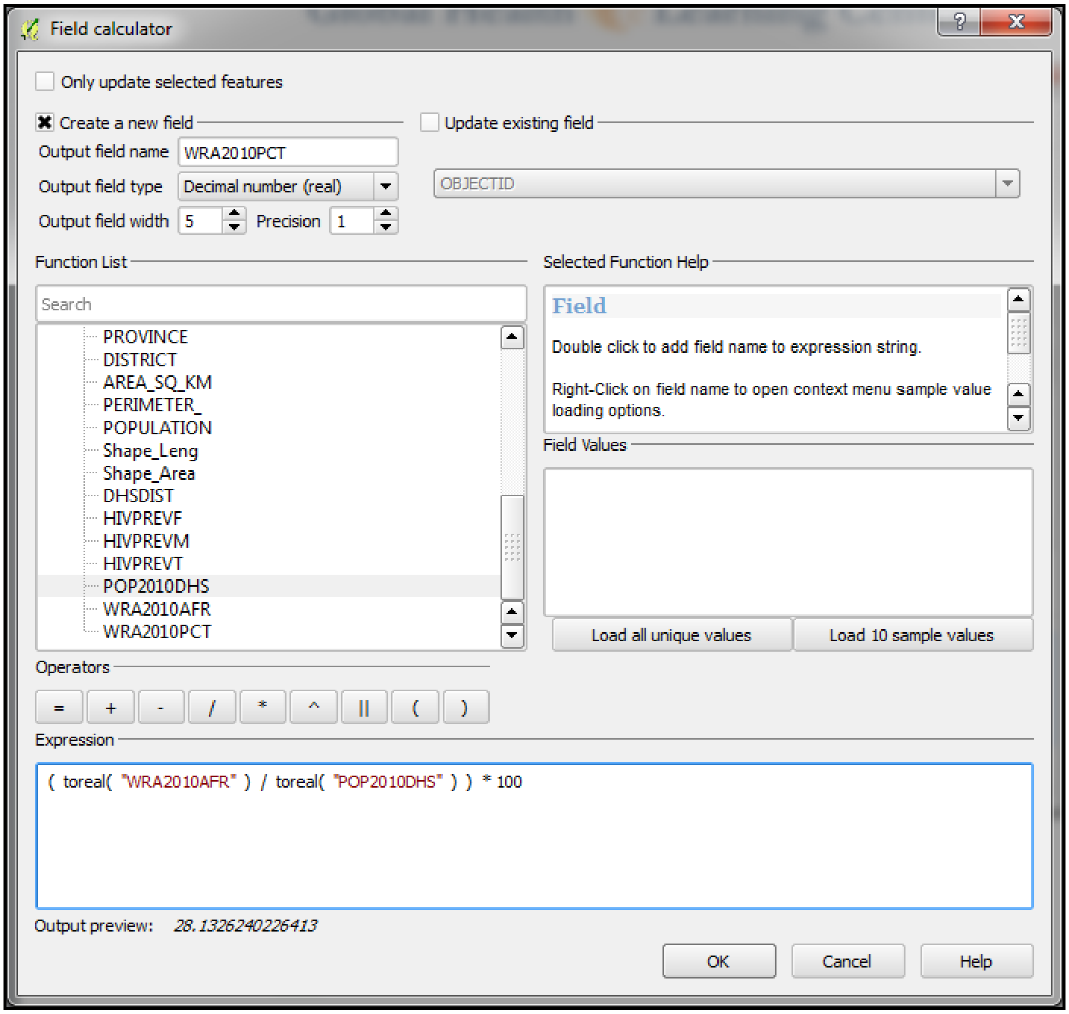


— To divide the estimated number of women of reproductive age by the total estimated population from the sampling frame used for the Rwanda DHS 2010, use the options in the Function List and Operators toolbar to complete the formula in the Expression box so that it appears as follows:



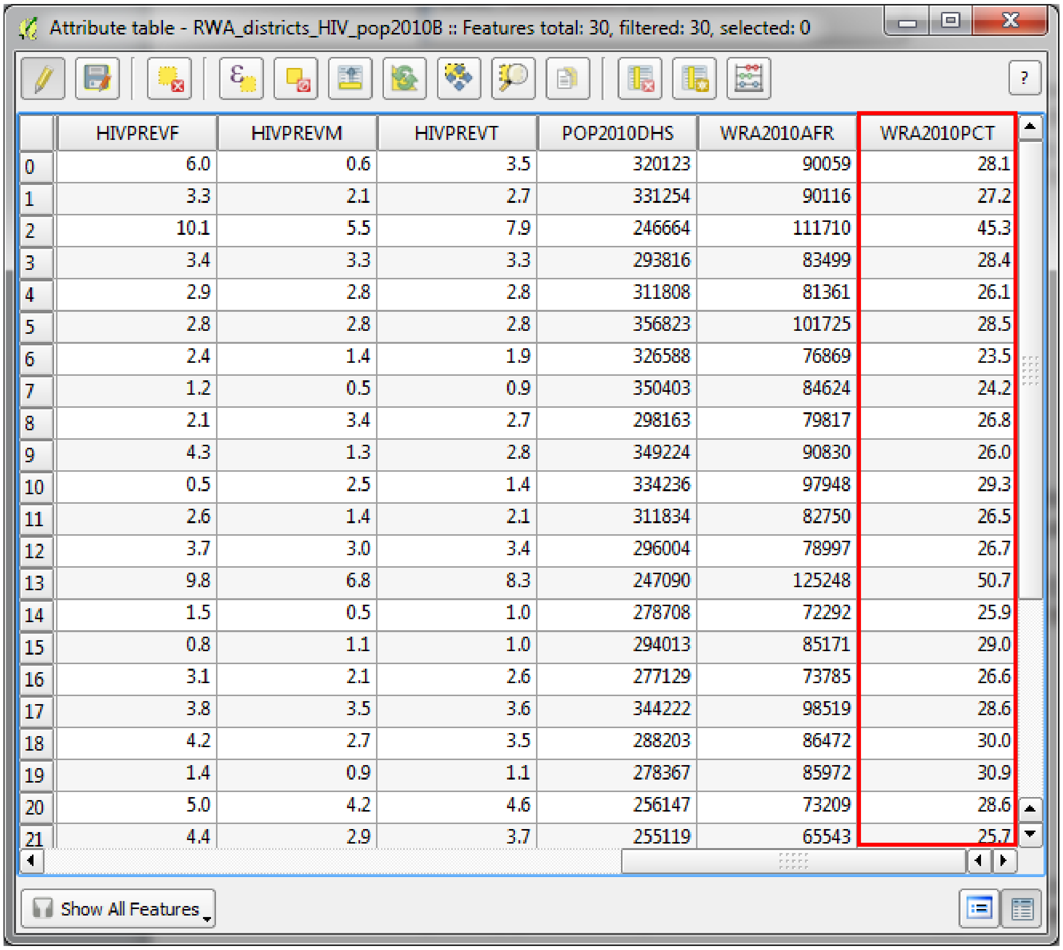
Note that QGIS will display a sample output preview below the Expression window to signify whether the expression is being evaluated correctly.

— After you have entered the formula as it appears in the Expression box depicted above, the field calculator dialog box should resemble the image below.



— To complete the process, click on the OK button at the bottom of the Field Calculator window.

If you open the attribute table for the shapefile and scroll to the right, you should see a new column containing the percent women of reproductive age by district (see below).



	HIVPREVF	HIVPREVM	HIVPREVT	POP2010DHS	WRA2010AFR	WRA2010PCT
0	6.0	0.6	3.5	320123	90059	28.1
1	3.3	2.1	2.7	331254	90116	27.2
2	10.1	5.5	7.9	246664	111710	45.3
3	3.4	3.3	3.3	293816	83499	28.4
4	2.9	2.8	2.8	311808	81361	26.1
5	2.8	2.8	2.8	356823	101725	28.5
6	2.4	1.4	1.9	326588	76869	23.5
7	1.2	0.5	0.9	350403	84624	24.2
8	2.1	3.4	2.7	298163	79817	26.8
9	4.3	1.3	2.8	349224	90830	26.0
10	0.5	2.5	1.4	334236	97948	29.3
11	2.6	1.4	2.1	311834	82750	26.5
12	3.7	3.0	3.4	296004	78997	26.7
13	9.8	6.8	8.3	247090	125248	50.7
14	1.5	0.5	1.0	278708	72292	25.9
15	0.8	1.1	1.0	294013	85171	29.0
16	3.1	2.1	2.6	277129	73785	26.6
17	3.8	3.5	3.6	344222	98519	28.6
18	4.2	2.7	3.5	288203	86472	30.0
19	1.4	0.9	1.1	278367	85972	30.9
20	5.0	4.2	4.6	256147	73209	28.6
21	4.4	2.9	3.7	255119	65543	25.7

- Calculate the estimated count of women of reproductive age who are HIV positive by district in Rwanda in 2010.

  Creating a choropleth (color-shaded) map based on integer values, such as the count of women of reproductive age by district, can be misleading based on differences in the size of the underlying populations. For example, if two districts have the same number of women of reproductive age but one of the districts has 10 times the total population, the concentration or density of women of reproductive age is not the same in the two districts.

  Despite limitations with respect to choropleth mapping for HIV/AIDS programs, it can be quite useful to know the size of a target population in order to understand the logistical and financial requirements for providing services and for monitoring and evaluating the effectiveness of interventions.

  - The new variable will be named NWRAHIVP, for number (N) of women of reproductive age (WRA) who are HIV positive (HIVP).

  - To calculate the new variable, you can multiply the estimated number of women of reproductive age by district extracted from the AfriPop raster dataset using the

QGIS zonal statistics plug-in (variable WRA2010AFR) by the HIV prevalence value for women of reproductive age from the Rwanda DHS 2010 (variable HIVPREVF). Since the variable for female HIV prevalence represents a percentage, for the calculation you will need to divide HIVPREVF by 100 to convert it to a proportion. The formula will be as follows:

NWRAHIVP = WRA2010AFR * (HIVPREVF / 100)

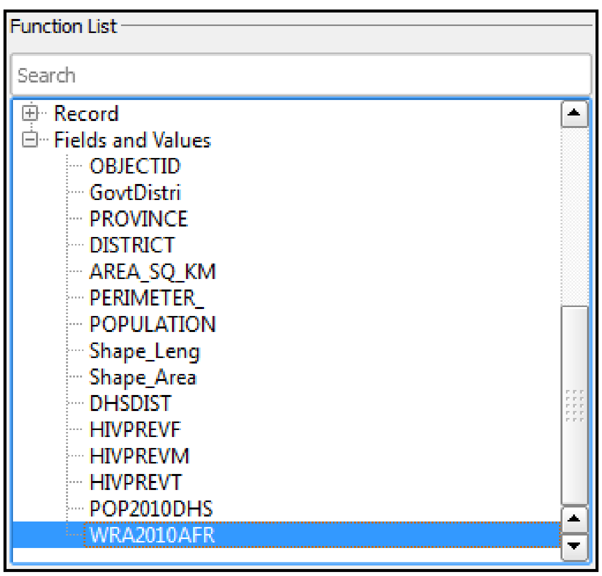
- To enter the name of the new field, make sure the box is selected for the "Create a new field" option and in the box next to the label "Output field name" enter the name NWRAHIVP.

- To specify an integer value for the number of women of reproductive age who are HIV positive, in the pull-down menu for "Output field type" keep the default selection of "Whole number (integer)."

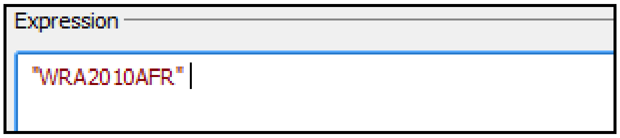
- To set the maximum number of digits to be displayed in the new field, you can change the "Output field width" to 6, which is the number of digits in the largest value in column WRA2010AFR, or you can leave this option set to the default value of 10.

- To specify the formula to use for the calculation, move to the Expression window and follow these steps:

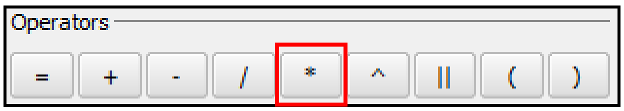
  — To select the variable for the estimated number of women of reproductive age, go to the Function List box, expand the list of choices under Fields and Values, and highlight WRA2010AFR (see below).



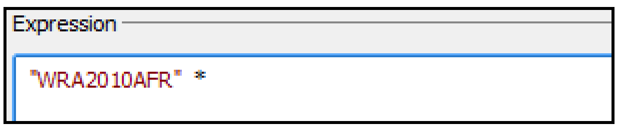
  — To have the WRA2010AFR variable appear in the Expression box, double-click on the highlighted option. The new expression should look like the one below.



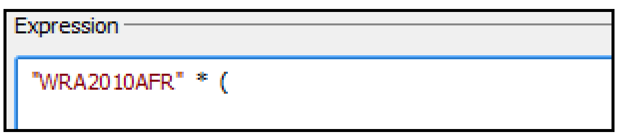
- To insert the multiplication operator, single-click on the corresponding operator button (see below).



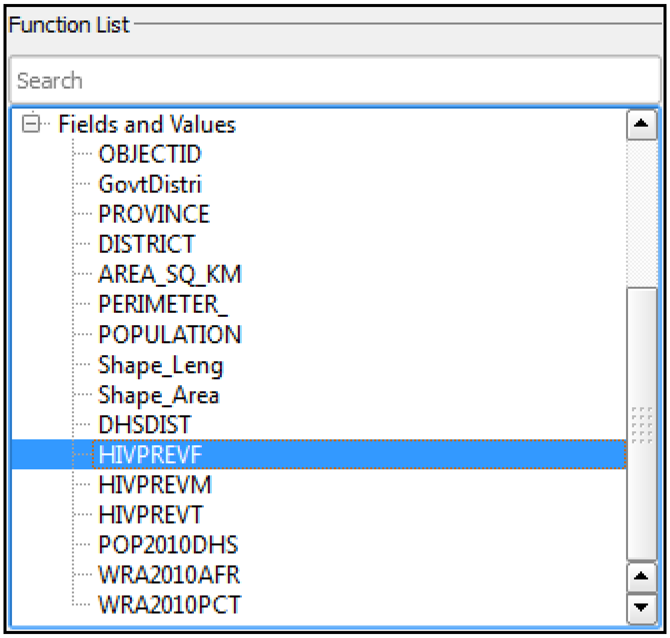
The expression should look like the following.



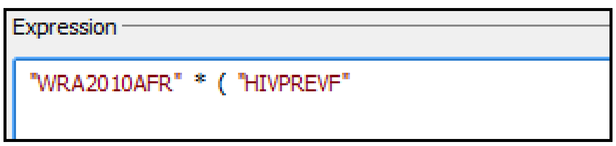
- To begin the portion of the formula pertaining to female HIV prevalence, insert a left parenthesis, which in QGIS is called an "Open Bracket."  The expression should look like the following.



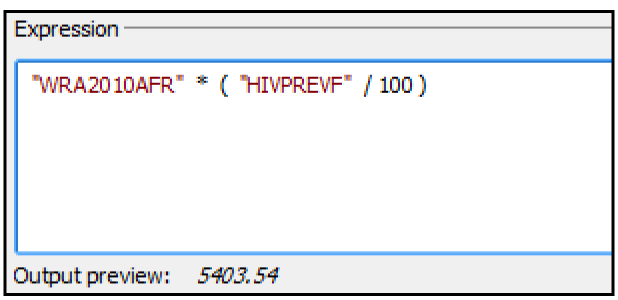
- To insert the variable name for female HIV prevalence from the Rwanda DHS 2010, return to the Function List box, expand the list of choices under Fields and Values, and highlight HIVPREVF (see below).



— To have the HIVPREVF variable appear in the Expression box, double-click on the highlighted option. The new expression should look like the one below.

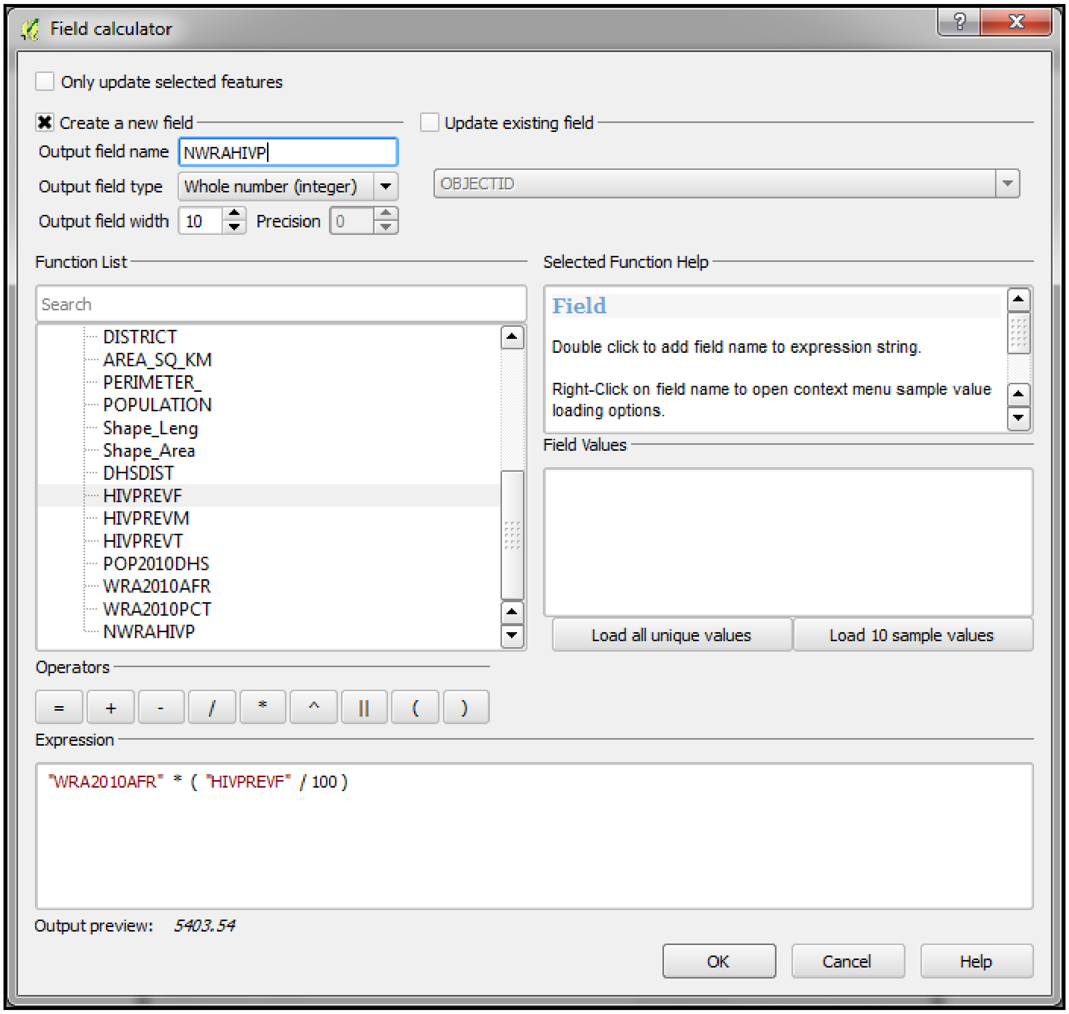


— To finish the expression, divide HIVPREVF by 100 and add a right parenthesis. The expression should look like the following.



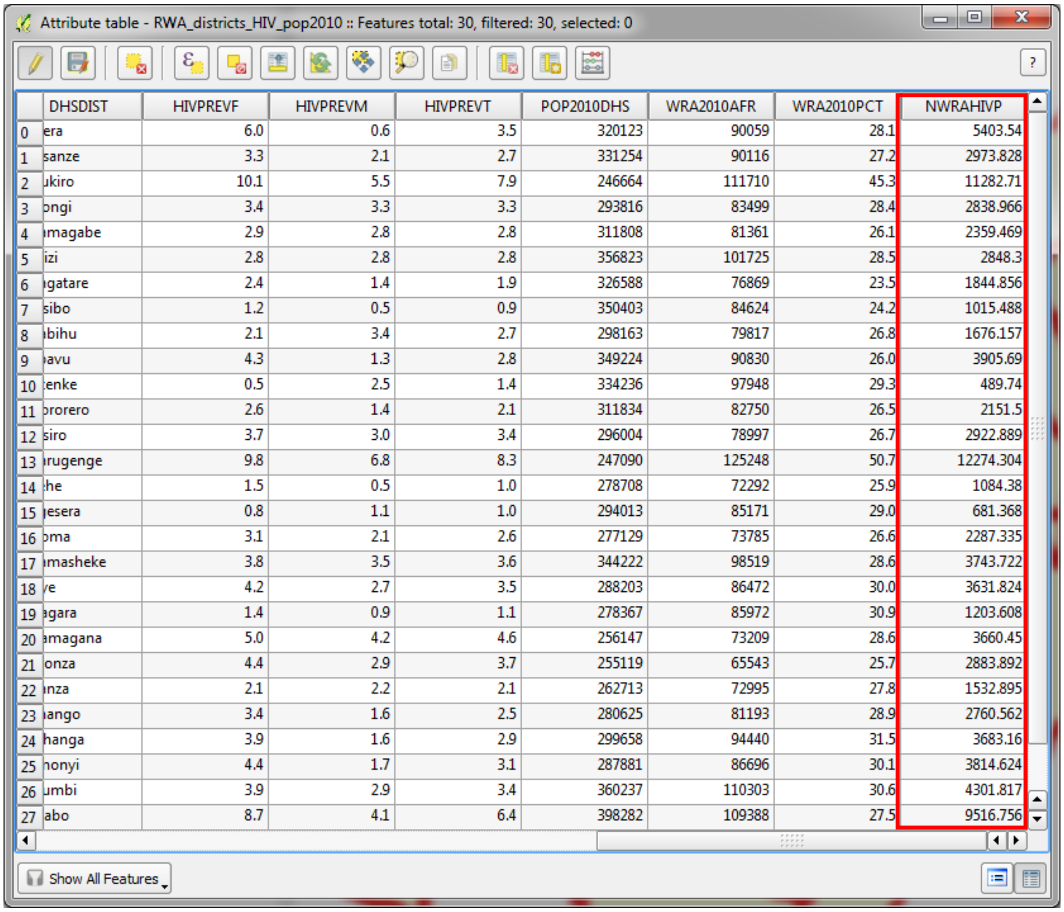
Note that QGIS will display a sample output preview below the Expression window to signify whether the expression is being evaluated correctly.

— After you have entered the formula as it appears in the Expression box depicted above, the field calculator dialog box should resemble the image below.



— To complete the process, click on the OK button at the bottom of the Field Calculator window.

If you scroll to the right in the attribute table for the shapefile, you should now see a new column containing the estimated number of women of reproductive age who are HIV positive by district in 2010 (see below).



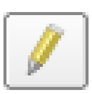
Attribute table - RWA_districts_HIV_pop2010 :: Features total: 30, filtered: 30, selected: 0

	DHSDIST	HIVPREVF	HIVPREVM	HIVPREVT	POP2010DHS	WRA2010AFR	WRA2010PCT	NWRAHIVP
0	era	6.0	0.6	3.5	320123	90059	28.1	5403.54
1	sanze	3.3	2.1	2.7	331254	90116	27.2	2973.828
2	ıkiro	10.1	5.5	7.9	246664	111710	45.3	11282.71
3	ɔngi	3.4	3.3	3.3	293816	83499	28.4	2838.966
4	ımagabe	2.9	2.8	2.8	311808	81361	26.1	2359.469
5	izi	2.8	2.8	2.8	356823	101725	28.5	2848.3
6	ıgatare	2.4	1.4	1.9	326588	76869	23.5	1844.856
7	sibo	1.2	0.5	0.9	350403	84624	24.2	1015.488
8	ıbihu	2.1	3.4	2.7	298163	79817	26.8	1676.157
9	ıavu	4.3	1.3	2.8	349224	90830	26.0	3905.69
10	:enke	0.5	2.5	1.4	334236	97948	29.3	489.74
11	ɔrorero	2.6	1.4	2.1	311834	82750	26.5	2151.5
12	siro	3.7	3.0	3.4	296004	78997	26.7	2922.889
13	ırugenge	9.8	6.8	8.3	247090	125248	50.7	12274.304
14	:he	1.5	0.5	1.0	278708	72292	25.9	1084.38
15	jesera	0.8	1.1	1.0	294013	85171	29.0	681.368
16	ɔma	3.1	2.1	2.6	277129	73785	26.6	2287.335
17	ımasheke	3.8	3.5	3.6	344222	98519	28.6	3743.722
18	/e	4.2	2.7	3.5	288203	86472	30.0	3631.824
19	ıgara	1.4	0.9	1.1	278367	85972	30.9	1203.608
20	ımagana	5.0	4.2	4.6	256147	73209	28.6	3660.45
21	onza	4.4	2.9	3.7	255119	65543	25.7	2883.892
22	ınza	2.1	2.2	2.1	262713	72995	27.8	1532.895
23	ıango	3.4	1.6	2.5	280625	81193	28.9	2760.562
24	hanga	3.9	1.6	2.9	299658	94440	31.5	3683.16
25	ıonyi	4.4	1.7	3.1	287881	86696	30.1	3814.624
26	umbi	3.9	2.9	3.4	360237	110303	30.6	4301.817
27	abo	8.7	4.1	6.4	398282	109388	27.5	9516.756

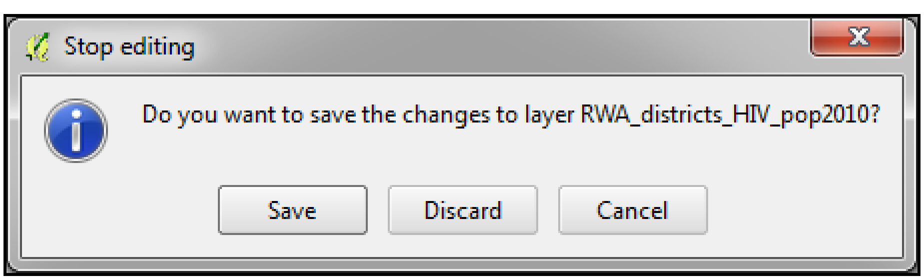
Show All Features

Note that many of the values in the new column contain decimals. This is because the values are temporary.

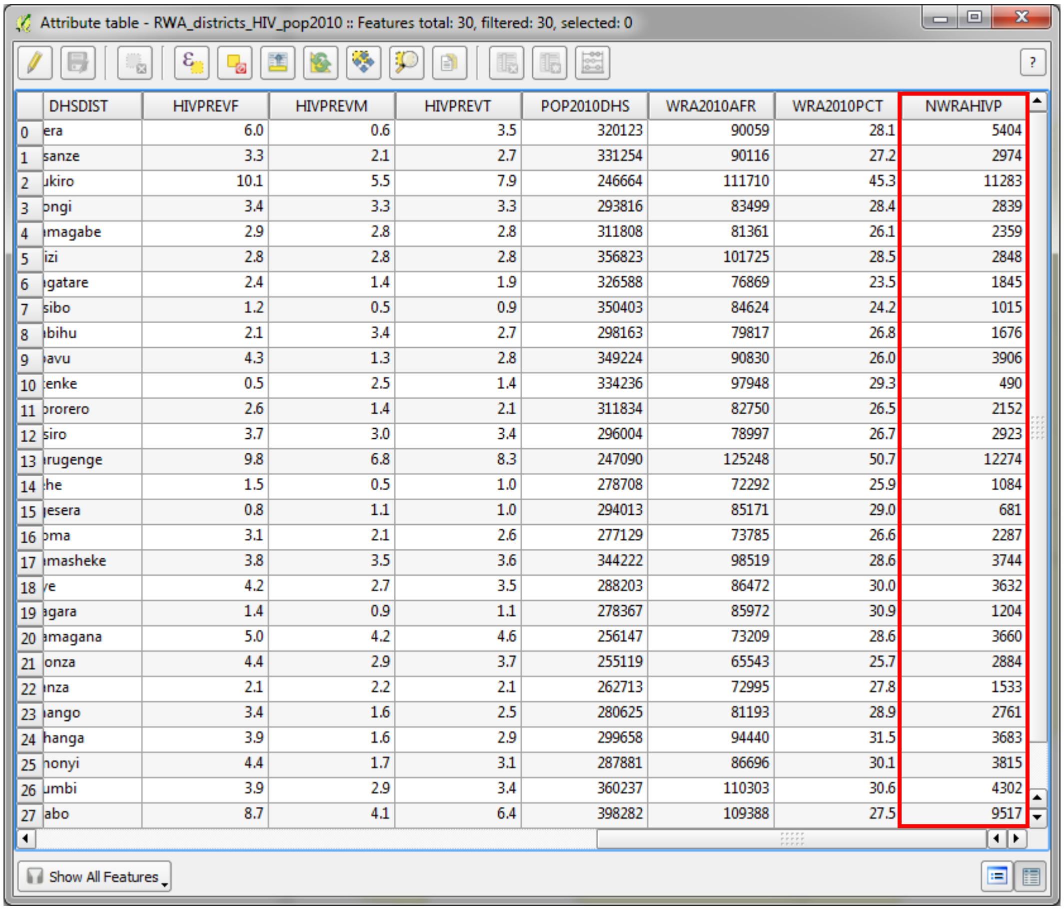
— To save the final values, which will be rounded to the nearest integer, click on the button for “Toggle editing mode.”



You should see the following dialog box.



Click on "Save." If you re-examine the values in the new column named NWRAHIVP (see below), you should see that the final values have been rounded to the nearest integer.



	DHSDIST	HIVPREVF	HIVPREVM	HIVPREVT	POP2010DHS	WRA2010AFR	WRA2010PCT	NWRAHIVP
0	era	6.0	0.6	3.5	320123	90059	28.1	5404
1	sanze	3.3	2.1	2.7	331254	90116	27.2	2974
2	ukiro	10.1	5.5	7.9	246664	111710	45.3	11283
3	ongi	3.4	3.3	3.3	293816	83499	28.4	2839
4	amagabe	2.9	2.8	2.8	311808	81361	26.1	2359
5	izi	2.8	2.8	2.8	356823	101725	28.5	2848
6	gatare	2.4	1.4	1.9	326588	76869	23.5	1845
7	sibo	1.2	0.5	0.9	350403	84624	24.2	1015
8	abihu	2.1	3.4	2.7	298163	79817	26.8	1676
9	avu	4.3	1.3	2.8	349224	90830	26.0	3906
10	enke	0.5	2.5	1.4	334236	97948	29.3	490
11	rorero	2.6	1.4	2.1	311834	82750	26.5	2152
12	siro	3.7	3.0	3.4	296004	78997	26.7	2923
13	arugenge	9.8	6.8	8.3	247090	125248	50.7	12274
14	he	1.5	0.5	1.0	278708	72292	25.9	1084
15	esera	0.8	1.1	1.0	294013	85171	29.0	681
16	oma	3.1	2.1	2.6	277129	73785	26.6	2287
17	amasheke	3.8	3.5	3.6	344222	98519	28.6	3744
18	ye	4.2	2.7	3.5	288203	86472	30.0	3632
19	agara	1.4	0.9	1.1	278367	85972	30.9	1204
20	amagana	5.0	4.2	4.6	256147	73209	28.6	3660
21	onza	4.4	2.9	3.7	255119	65543	25.7	2884
22	nza	2.1	2.2	2.1	262713	72995	27.8	1533
23	ango	3.4	1.6	2.5	280625	81193	28.9	2761
24	hanga	3.9	1.6	2.9	299658	94440	31.5	3683
25	honyi	4.4	1.7	3.1	287881	86696	30.1	3815
26	umbi	3.9	2.9	3.4	360237	110303	30.6	4302
27	abo	8.7	4.1	6.4	398282	109388	27.5	9517

You now have **two new variables in your shapefile**, which can be used for mapping and spatial analysis. In the next exercise you will use these new variables to make a map.

You can now close the attribute table and return to the main QGIS interface.

☐ Save the QGIS project.

- To save your QGIS project, go to the main QGIS menu and select Project > Save or click on the Save button. 
- In the file dialog box that appears, specify the following output name:

  Exercises 4.1, 4.2 and 4.3**qgis_exercise_4.1.qgs**

- Quit QGIS.
  - To close QGIS, go to the main menu and select Project > Exit QGIS or use a combination of the Ctrl and Q keys.

END

GIS Techniques for
Monitoring and Evaluation of
HIV/AIDS and Related Programs

Exercise 4.2

Create Maps of Multiple Indicators

**This training was developed as part of a joint effort between MEASURE Evaluation and MEASURE DHS, with funding provided by USAID and PEPFAR.*

Exercise 4.2: Create Maps of Multiple Indicators

Summary: In Exercise 3.2 you created a choropleth (color-shaded) map of a single indicator. Although maps of single indicators can be quite useful, to see the potential associations between multiple indicators it is important to be able to map more than one variable at a time. To acquire this skill, in this lesson you will learn how to generate multi-indicator maps by using pie charts, bar charts, and proportional symbols for one variable superimposed on a choropleth map showing the geographic distribution of a second variable.

Objectives:

- Create a layer of centroids (center points) from the district boundaries in the shapefile in order to facilitate the automatic placement of data charts (pie charts, bar charts, or proportional symbols) for each district.
- Display pie charts showing one set of data superimposed on a map showing another set of data shaded by district.
- Display proportional symbols for one set of data over a base map showing another set of data shaded by district.
- Display bar charts showing one set of data superimposed on a base map showing another set of data shaded by district.

Requirements*:*

To complete these exercises, you will need to have QGIS and several plugins installed on your machine, as instructed in exercise 2.1. You will also need to have downloaded and unzipped the data files associated with the exercise. Remember, these files can reside anywhere on your computer, but must be kept together, with the same original file structure.

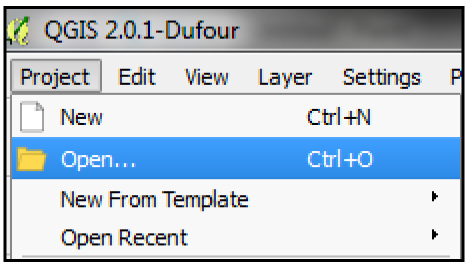
Step 1: Launch QGIS.

- Option 1: Click on the desktop shortcut.

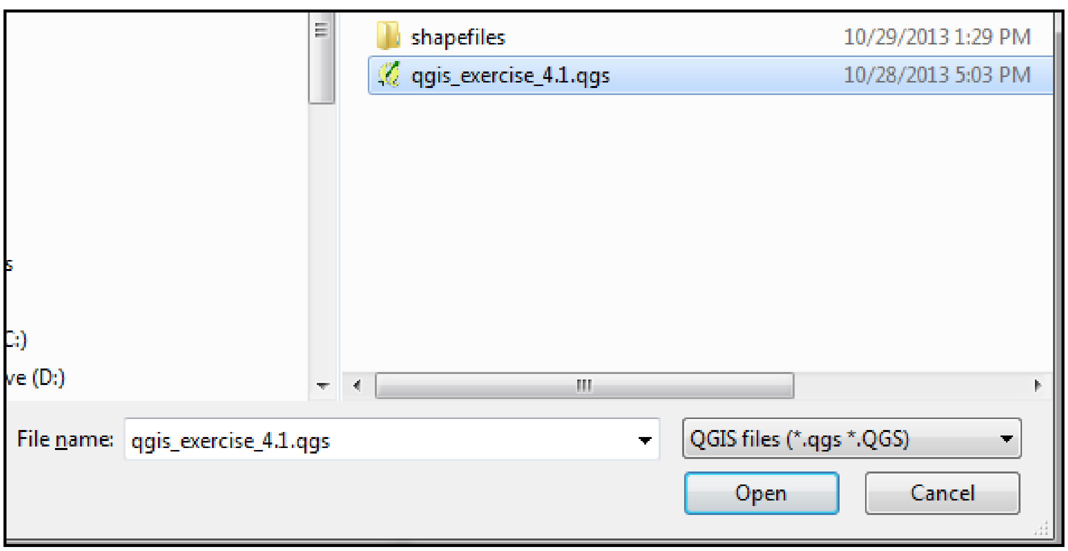
- Option 2: Click on Start > All programs > QGIS Dufour > QGIS Desktop 2.0.1.

Step 2: Create a new project based on the Exercise 4.1 project.

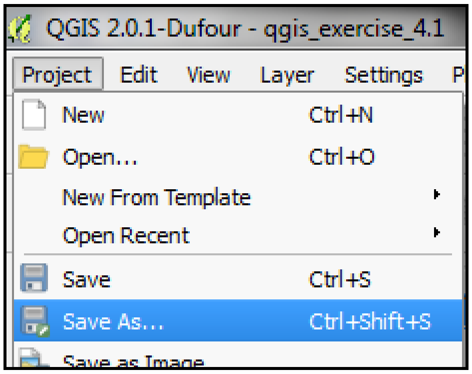
- On the main QGIS menu, select Project > Open.



- Navigate to the folder where you have been working, in this case **Exercises 4.1, 4.2, and 4.3** and open the file **qgis_exercise_4.1.qgs**.



- To create a new project, save the current project with a new name (see below).
    - On the main QGIS menu, select Project > Save As.



- For the output file name, specify the following:

  \Exercises 4.1, 4.2, and 4.3**qgis_exercise_4.2.qgs**

- To complete the process, click on the Save button in the file dialog window.

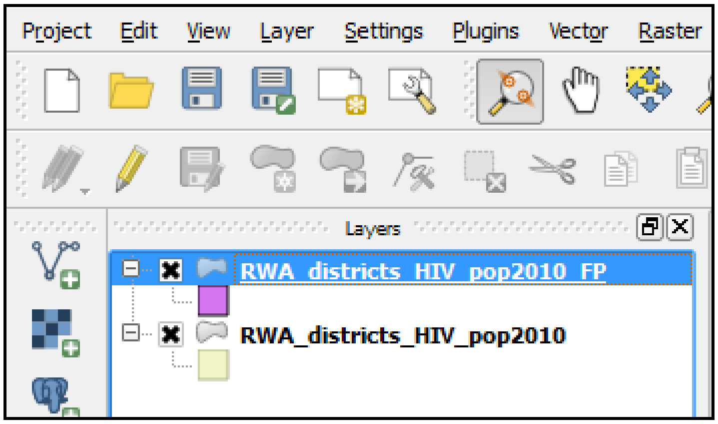
Step 3: Open a shapefile containing data on HIV prevalence and on family planning by district in Rwanda in 2010.

- To add a new shapefile to the project, go to the toolbar for managing layers and click on the Add Vector Layer button. 

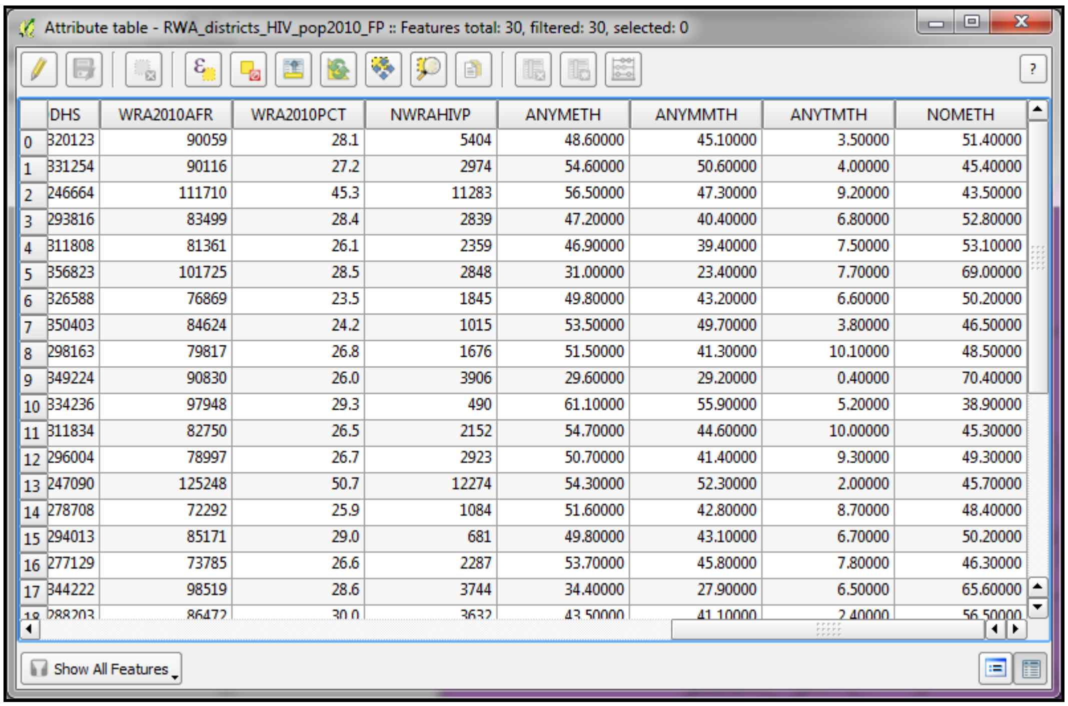
- Make sure the source type is set to File and click on Browse. Set the file type to "ESRI Shapefiles [OGR] (*.shp *.SHP)."

- Navigate to the folder "Exercises 4.1, 4.2 and 4.3\shapefiles" and find the shapefile **RWA_districts_HIV_pop2010_FP.shp**, and open it.

- You should see a new layer in the Layers window corresponding to the shapefile you just added.



- View the contents of the attribute table for the new QGIS layer. (Hint: right-click on the layer name and select Open Attribute Table.) If you scroll to the right you should see a table like the following.



Attribute table - RWA_districts_HIV_pop2010_FP :: Features total: 30, filtered: 30, selected: 0

	DHS	WRA2010AFR	WRA2010PCT	NWRAHIVP	ANYMETH	ANYMMTH	ANYTMTH	NOMETH
0	320123	90059	28.1	5404	48.60000	45.10000	3.50000	51.40000
1	331254	90116	27.2	2974	54.60000	50.60000	4.00000	45.40000
2	246664	111710	45.3	11283	56.50000	47.30000	9.20000	43.50000
3	293816	83499	28.4	2839	47.20000	40.40000	6.80000	52.80000
4	311808	81361	26.1	2359	46.90000	39.40000	7.50000	53.10000
5	356823	101725	28.5	2848	31.00000	23.40000	7.70000	69.00000
6	326588	76869	23.5	1845	49.80000	43.20000	6.60000	50.20000
7	350403	84624	24.2	1015	53.50000	49.70000	3.80000	46.50000
8	298163	79817	26.8	1676	51.50000	41.30000	10.10000	48.50000
9	349224	90830	26.0	3906	29.60000	29.20000	0.40000	70.40000
10	334236	97948	29.3	490	61.10000	55.90000	5.20000	38.90000
11	311834	82750	26.5	2152	54.70000	44.60000	10.00000	45.30000
12	296004	78997	26.7	2923	50.70000	41.40000	9.30000	49.30000
13	247090	125248	50.7	12274	54.30000	52.30000	2.00000	45.70000
14	278708	72292	25.9	1084	51.60000	42.80000	8.70000	48.40000
15	294013	85171	29.0	681	49.80000	43.10000	6.70000	50.20000
16	277129	73785	26.6	2287	53.70000	45.80000	7.80000	46.30000
17	344222	98519	28.6	3744	34.40000	27.90000	6.50000	65.60000
18	288203	86472	30.0	3632	43.50000	41.10000	2.40000	56.50000

Show All Features

The new shapefile you opened contains, among other things, four fields that have been extracted from Table D.32 of the Rwanda DHS 2010, which shows the district-level percent of currently married women age 15-49 according to contraceptive method currently used at the time of the survey. The new fields are as follows:

- ANYMETH: Any method of contraception. This variable indicates the total percentage of married women using any method of contraception, whether it be classified as traditional or modern.

- ANYMMTH: Any modern method of contraception. Modern methods of contraception include female sterilization, male sterilization, pill, intra-uterine device (IUD), injectable, implant, male condom, diaphragm, lactational amenorrhoea method (LAM), and standard days method (SDM).

- ANYTMTH: Any traditional method of contraception. Traditional methods of contraception include the rhythm method, withdrawal, and other.

- NOMETH: No method of contraception used.

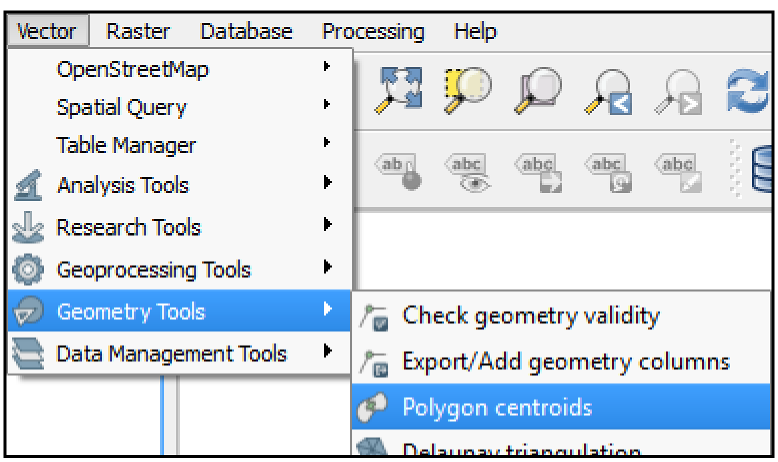
When you have finished examining the attribute table, close it.

Step 4: Generate a new centroid layer.

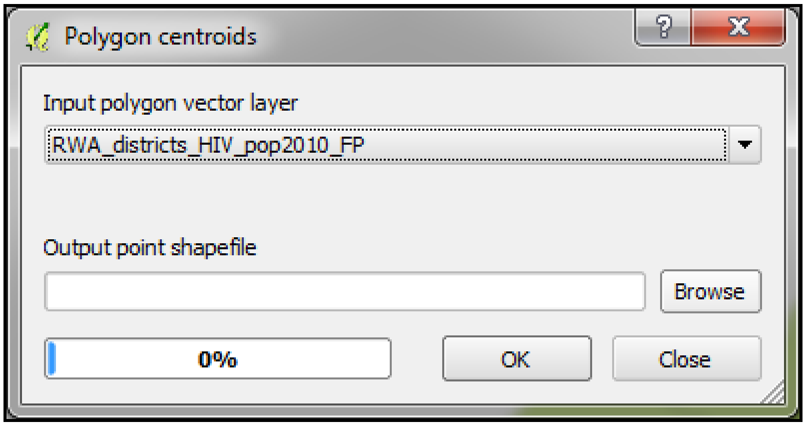
QGIS can display more than one variable at a time, but its automatic placement for display of multiple variables can be somewhat unreliable when the source data are stored in a polygon-based shapefile, such as the one we are using for the districts of Rwanda (depending on the size and shape of the polygons).

So a more reliable method of displaying a second variable is to generate a new layer consisting of centroids (center points) from the district boundaries in the shapefile and use this layer to display the second set of variables.

- To create a layer of centroids (center points) for the districts of Rwanda, go to the main QGIS menu and select Vector > Geometry Tools > Polygon centroids (see below).

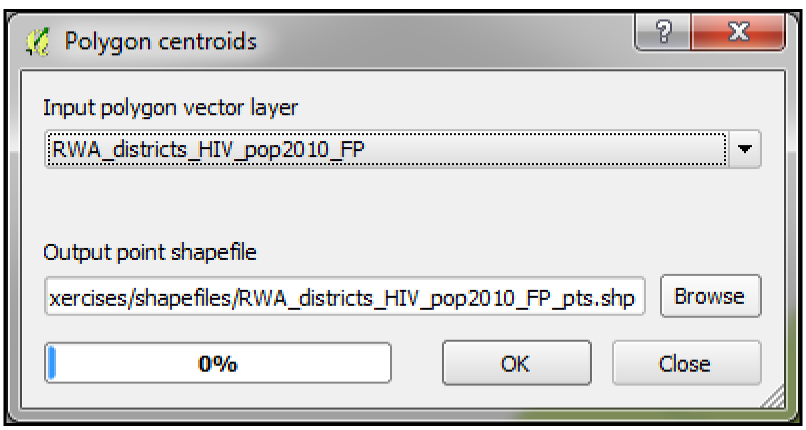


- In the dialog box that appears, go to the pull-down menu and select the layer name RWA_districts_HIV_pop2010_FP (see below).



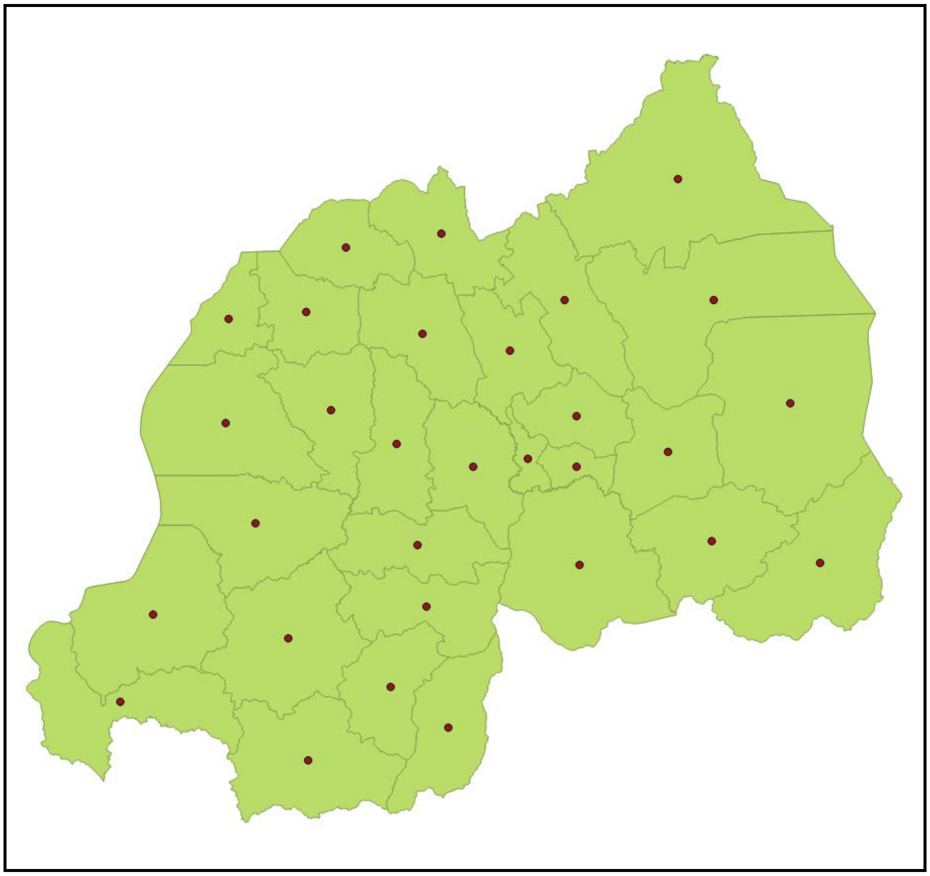
- To specify the output shapefile name, click on Browse.

- In the file dialog window that appears, navigate to the folder **\Exercises 4.1, 4.2 and 4.3\shapefiles**, specify the output file name as **RWA_districts_HIV_pop2010_FP_pts.shp**, and click on the Save button. Note the addition of the letters "_pts" before the shapefile extension (.shp). These letters represent an abbreviation for the word "points." After clicking on Save, you should see a dialog box that looks like the one below.



- Click on OK. You should see a message that tells you the new shapefile was created successfully and that asks you whether you would like to add the new layer to the table of contents (TOC) in the QGIS Layers panel.

- Click on Yes and then click on Close in the "Polygon centroids" dialog box. You should see a map that resembles the following:



You now have a point-based data layer for overlaying diagrams for districts.

- Save the QGIS project.

  - To save your QGIS project, go to the main QGIS menu and select Project > Save or click on the Save button. 

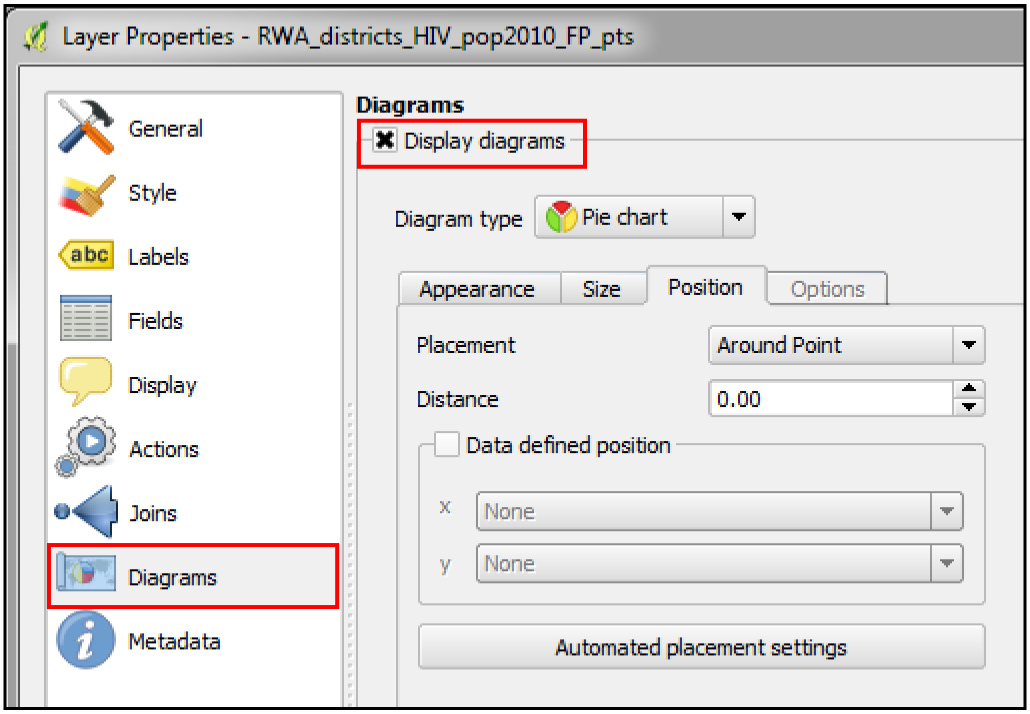
Step 5: Display pie charts showing the percent of currently married women age 15-49 by current contraceptive method superimposed on a base map showing total HIV prevalence by district.

Displaying an indicator in the form of pie charts is an effective way to show data based on sub-categories that together account for 100 percent of the distribution of the indicator. This technique will show each sub-category's contribution to the whole distribution as a proportional slice of the pie.

As noted previously, the data being used for the pie charts was extracted from Table D.32 of the Rwanda DHS 2010, which shows the district-level percent of currently married women age 15-49 according to contraceptive method currently used at the time of the survey.

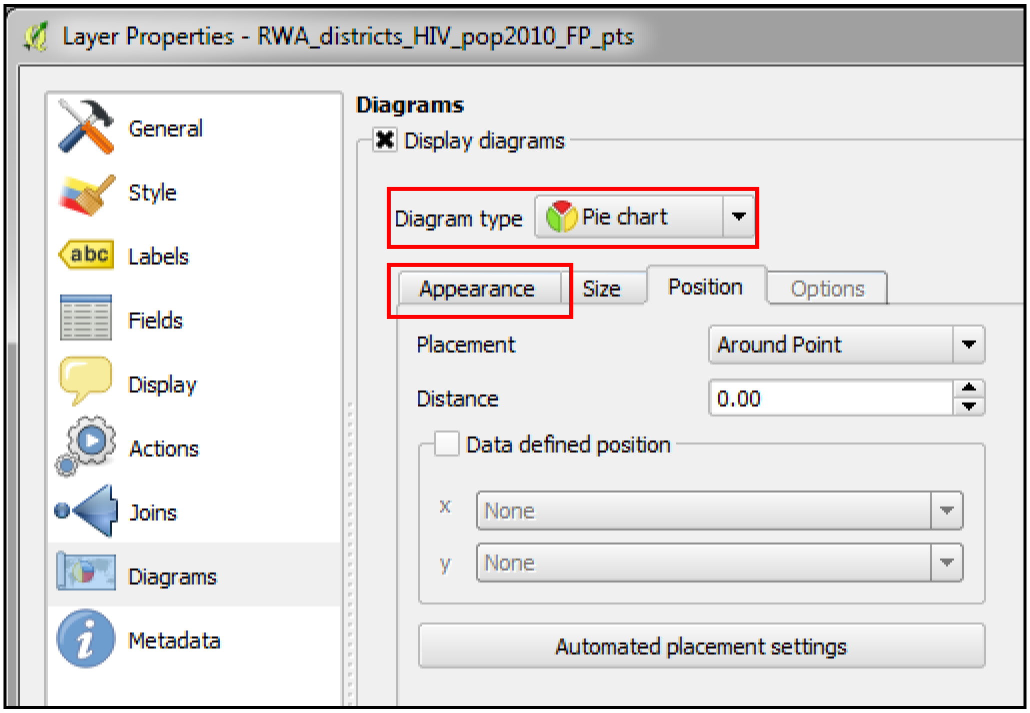
- To begin the creation of pie charts, open the Layer Properties window for the layer named **RWA_districts_HIV_pop2010_FP_pts** by either double-clicking on the name or by highlighting the layer name in the Layers panel and going to the main QGIS menu and selecting Layer > Properties.

- In the Layer Properties window, highlight the **Diagrams** option and activate the display of diagrams by clicking in the box next to **Display diagrams** (see below).

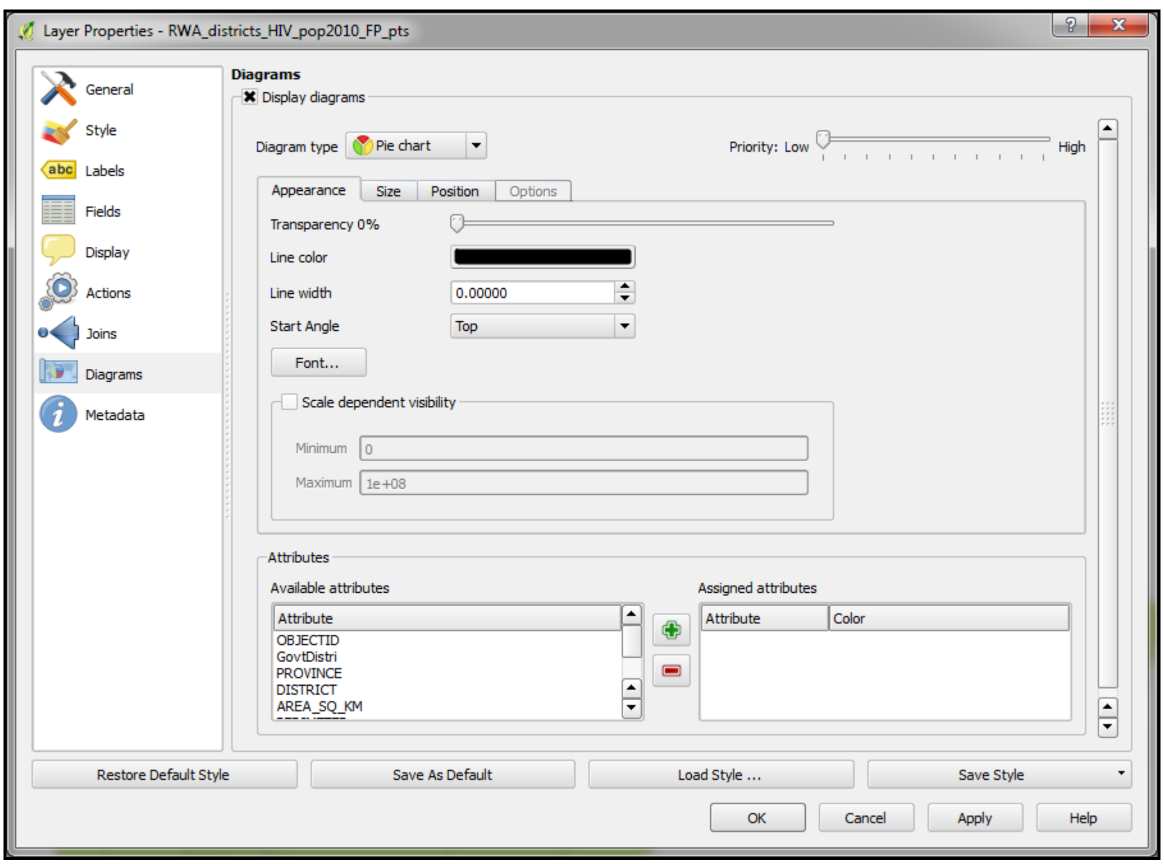


- In the pull-down menu for **Diagram type**, make sure **Pie chart** is selected.

- With **Pie chart** selected as the diagram type, click on the tab named **Appearance**.



You should see a screen that looks like the one below.

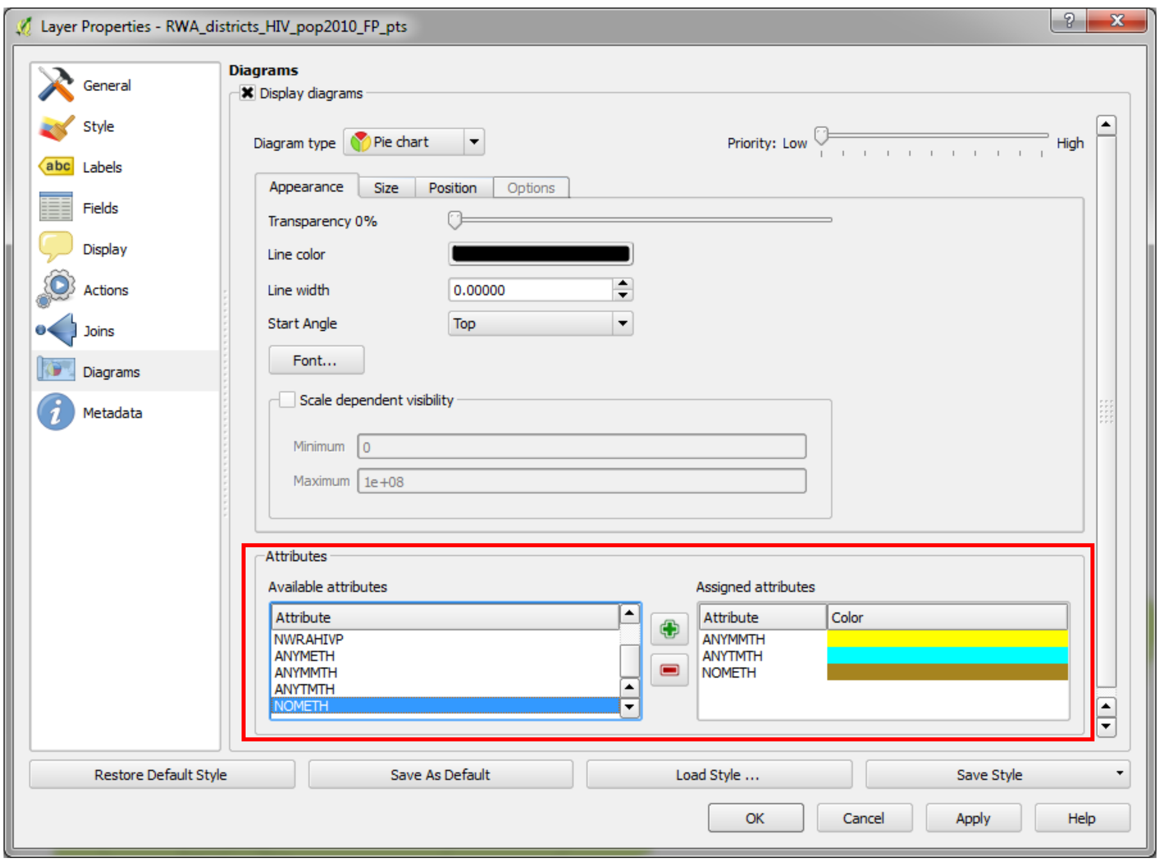


- To select the variables to include in the pie charts, use the scroll bar in the Attributes box in the lower left to scroll down until you see the variables ANYMMETH, ANYTMETH, NOMETH.

  These are the variables that show the percent of married women age 15-49 who are using any modern method of contraception (ANYMMETH), any traditional method (ANYTMETH), or no method at all (NOMETH). The three categories together account for 100 percent of married women age 15-49 by district, so they are well-suited to display in pie charts.

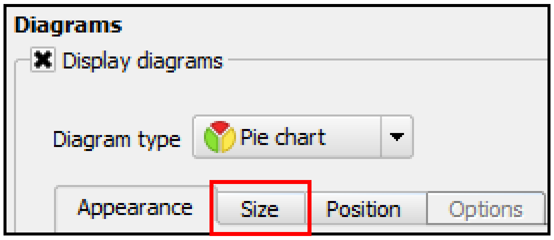
  To select a variable and have it appear in the **Assigned attributes** list, either double-click on the variable name in the **Available attributes** list or highlight the variable name in the **Available attributes** list and click on the green plus sign. 

  After selecting the three variables of interest, the Diagrams window should look like the one below.

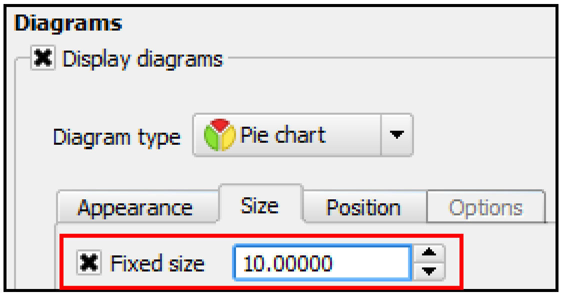


- To change the color that will be used to display the pie slice for a particular variable, double-click on the color bar in the **Assigned attributes** list.

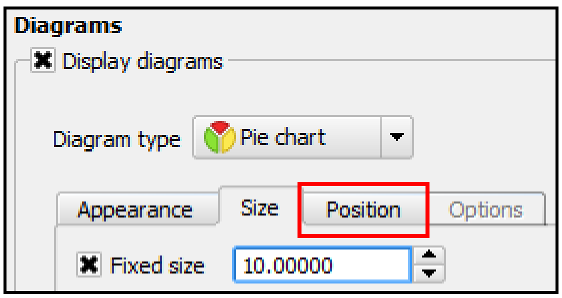
- To set the display size of the pie charts, click on the **Size** tab.



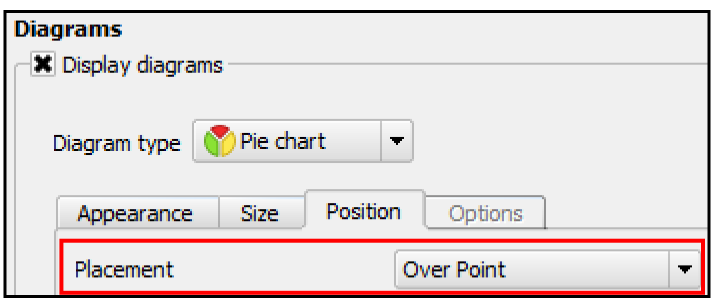
- On the **Size** tab, retain the default option of Fixed size, but reduce the size from 30mm to 10mm (see below).



- To display the pie charts directly on top of the polygon center points created in Step 4 above, click on the tab named **Position**.

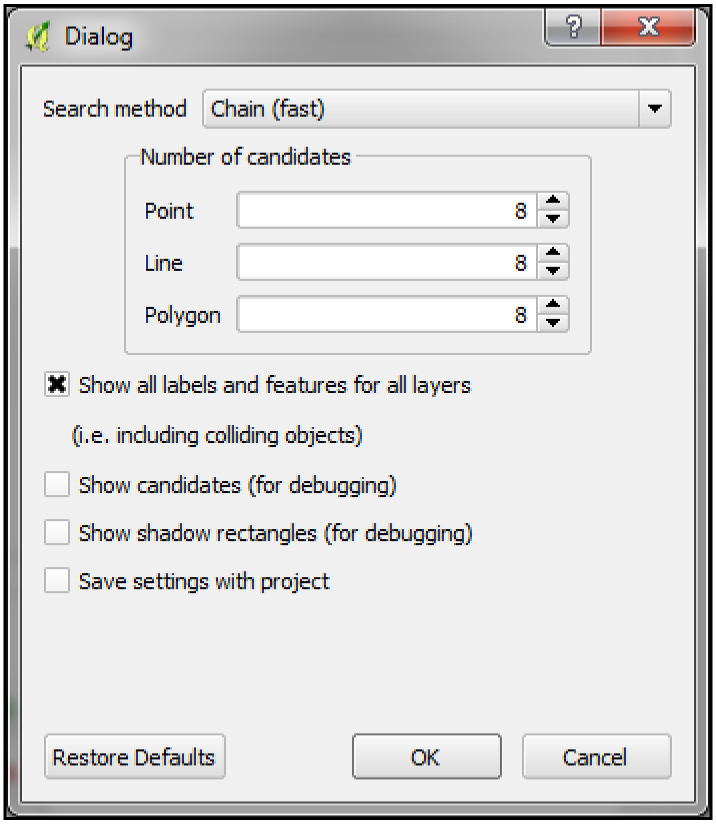


- On the **Position** tab, change the Placement type from **Around Point** to **Over Point** (see below).



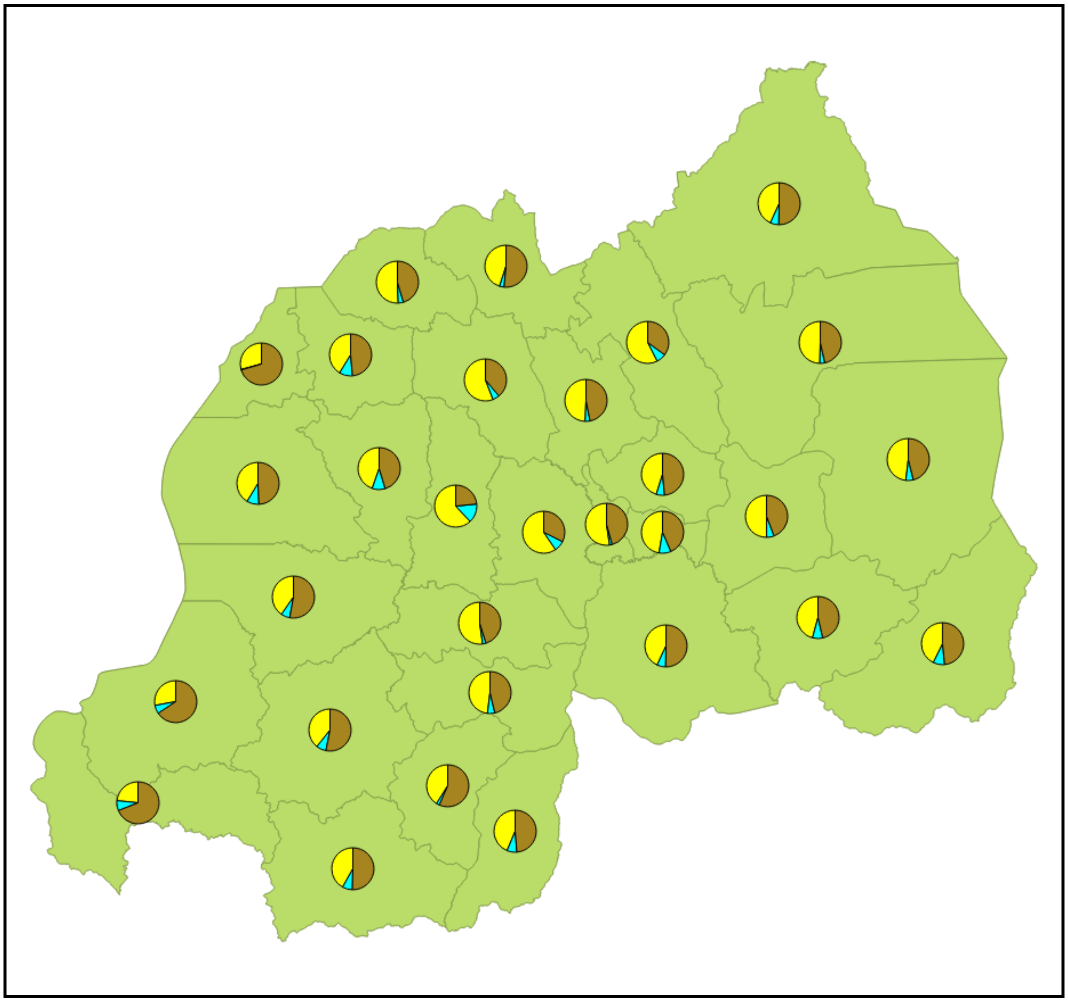
- Still on the **Position** tab, click on the button labeled **Automatic placement settings**.

- In the dialog box that opens, select the option **Show all labels and features for all layers** (see below).



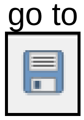
- To view the pie charts in the map window, click on OK to close the dialog box for automated placement settings. Back in the Layer Properties window, click on Apply then OK.

  You should see a map that looks similar to the one below. (You may go back and adjust colors and pie sizes if you wish. A bit of trial and experimentation may be required for good map design.)



The map shows that at the district level in Rwanda, according to the Rwanda DHS 2010, the majority of contraceptive methods used by married women were modern (yellow), although a large proportion of married women were using no method of contraception (brown).

☐ Save the QGIS project.

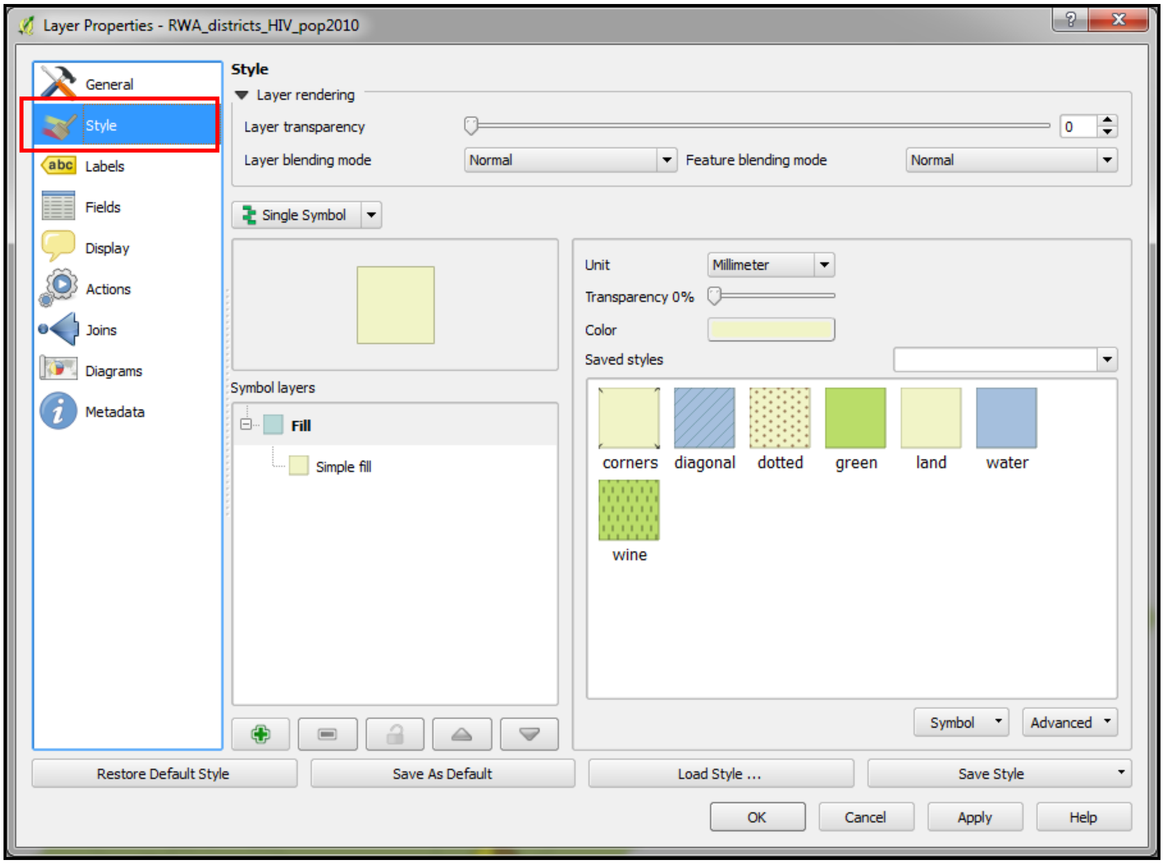
- To save your QGIS project, go to the main QGIS menu and select Project > Save or click on the Save button. 

☐ To see a second variable on the same map and unearth a potential relationship between contraceptive methods chosen by currently married women and female HIV prevalence, repeat the steps followed in exercise 3.2 and create a color-shaded map of **female HIV prevalence** using the layer named **RWA_districts_HIV_pop2010**.

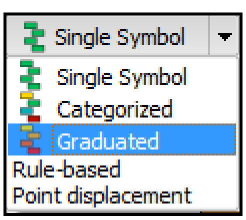
  ☐ To begin the color-shading process, open the properties window for the shapefile layer **RWA_districts_HIV_pop2010** by double-clicking on the layer name in the

Layers panel or by highlighting the layer name in the Layers panel and on the main QGIS menu selecting Layer > Properties.

- After the Layer Properties window opens, left-click on the Style icon in order to highlight it. You should see a screen that looks like the one below.



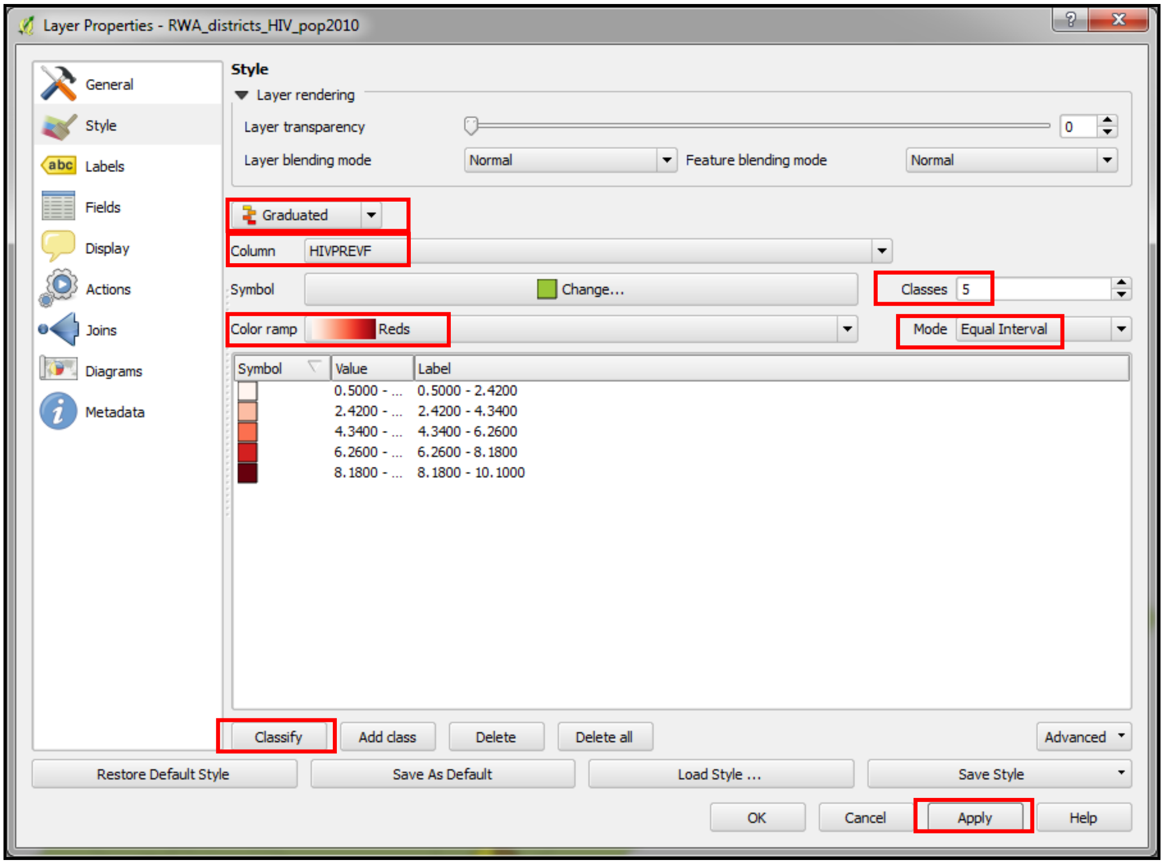
- To be able to use different colors for the different ranges of values for HIV prevalence, change the selection in the pull-down menu above the large colored square from Single Symbol to Graduated (see below).



- To map **female** HIV prevalence by district, in the pull-down menu to the right of the word Column select the column **HIVPREVF**.

- Leave the number of classes at 5 and keep the classification mode set to Equal Interval.

- Change the color ramp to Reds, then click on the Classify and Apply buttons. The dialog box should now resemble the following.



- Click on the **OK** button to close the dialog box and see the choropleth map of female HIV prevalence by district.

  Note that the equal intervals classification method arranged the HIV prevalence data from the DHS final report into classes containing an equal range of values. The formula for calculating the range of values to use for each class is as follows:

  Total range of data values / number of classes

  Since the range of total HIV prevalence values was 9.6 (the maximum, 10.1, minus the minimum, 0.5) and the number of classes selected was 5, the range of values assigned to each class was calculated as follows:

  9.6 / 5 = 1.92

  Class 1 = 0.50 to 2.42
  Class 2 = 2.42 to 4.34

Class 3 = 4.34 to 6.26
Class 4 = 6.26 to 8.18
Class 5 = 8.18 to 10.10

Note: Remember, QGIS correctly calculates the maximum value for class ranges, but allows an overlap in the maximum value for one class and the minimum value for the next class. For example, class 1 ends in a value of 2.42 and class 2 begins with the same value of 2.42. The same value should technically not be shared by two classes. To correct this problem:

- Open the Layer Properties window, double-click on each range of values in the Values column—beginning with class 2—and increase the lower value in the range by 0.01.

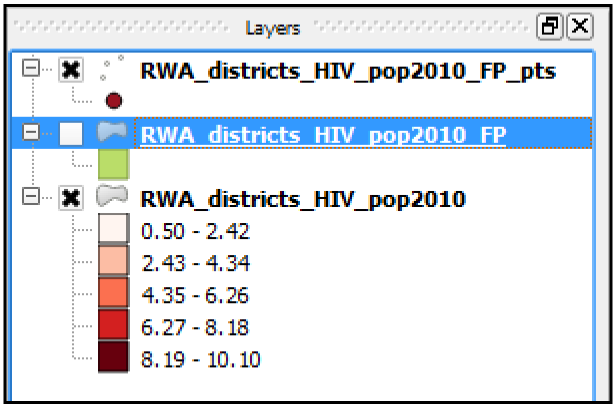
  You will also need to change the label to match the new lower limit. While you are editing the ranges displayed in the labels, you can remove the unnecessary zeroes at the ends of the labels. To do this, double-click on each range of values in the Label column and remove the zeroes.

  The resulting Value and Label columns should look as follows:

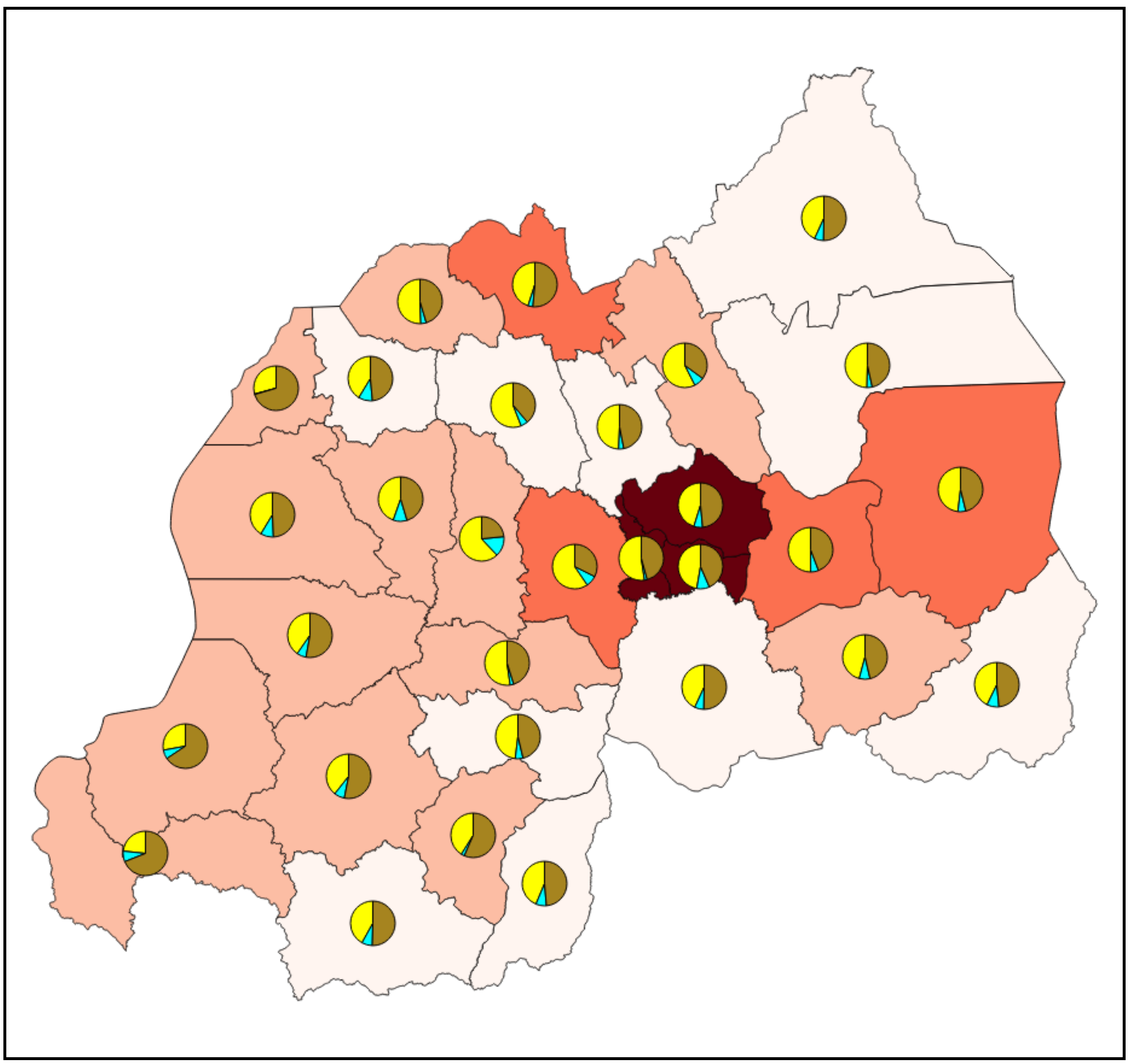
Symbol	Value	Label
	0.5000 - 2.4200	0.50 - 2.42
	2.4300 - 4.3400	2.43 - 4.34
	4.3500 - 6.2600	4.35 - 6.26
	6.2700 - 8.1800	6.27 - 8.18
	8.1900 - 10.1000	8.19 - 10.10

- To save the changes, click on Apply, then OK. This will close the Layer Properties window.

☐ The shading colors for the new map are located at the bottom of the layers list. To make it visible, in the Layers panel turn off the data layer named **RWA_districts_HIV_pop2010_FP** by clicking to uncheck the box next to the layer name (see below).



You should now see a map that resembles the one below.

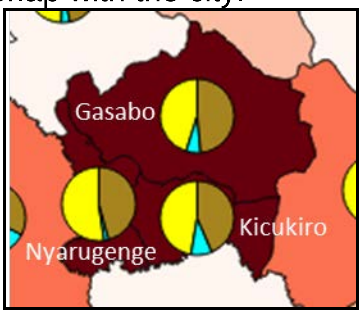


Rwanda by District, 2010, HIV Prevalence among Women Age 15-49 versus Percent of Currently Married Women by Method of Contraception

The map above allows a few observations regarding the relationship between female HIV prevalence and the choice of contraceptive methods among currently married women.

Question: Where is the highest prevalence of HIV?
Answer: The darkest districts, which represent the highest prevalence of HIV among women of reproductive age, encompass the City of Kigali, which is located in the center of the country.

Question: Where are traditional methods of contraception used most?
Answer: Although traditional methods of contraception (blue pie slices) are used much less often in Kigali, they appear to be used more widely in Kigali's southern district of Kicukiru, which is relatively more rural than the other two districts that overlap with the city.



Question: What are the proportions of women using no methods of contraception?
Answer: Somewhat surprisingly, given the high female HIV prevalence in Kigali, the proportion of married women of reproductive age who use no method of contraception is still quite high, though not as high as for outlying, more rural districts.

These are just a few examples of the types of observations that can be made using a combination of a choropleth map for one variable and pie charts for a second variable. Observations such as these can lead to the development of hypotheses concerning the underlying or root causes of an epidemic and can also help lead to a more evidence-based foundation for program planning and decision making.

- [ ] Save the QGIS project.

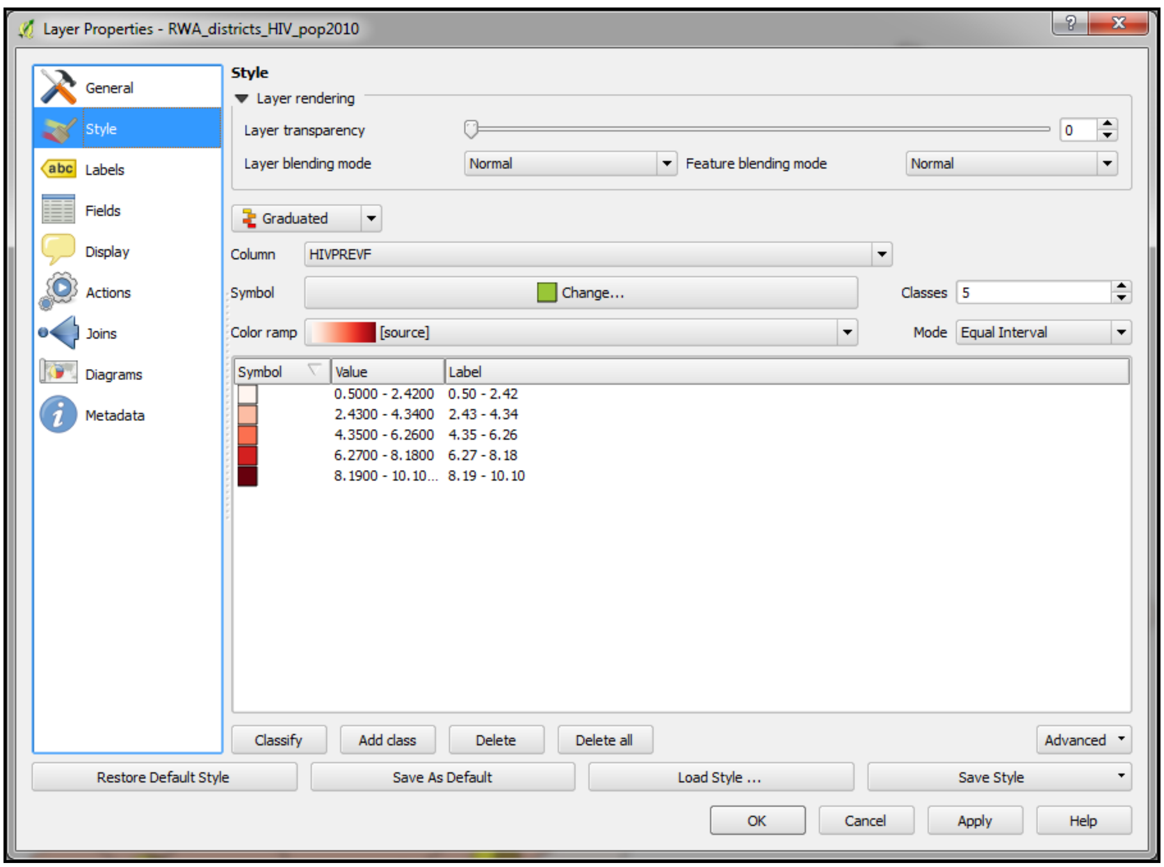
  - To save your QGIS project, go to the main QGIS menu and select Project > Save or click on the Save button. 

Step 6: Display proportional symbols for female HIV prevalence superimposed on a base map showing total HIV prevalence by district.

To accomplish Step 6 you will change the color-shaded map for layer **RWA_districts_HIV_pop2010** to show total HIV prevalence (**HIVPREVT**) and not female HIV prevalence (**HIVPREVF**). You will also change the diagrams displayed for layer **RWA_districts_HIV_pop2010_FP_pts** from pie charts to text diagrams.

- To begin the color-shading process, open the properties window for the shapefile layer **RWA_districts_HIV_pop2010** by double-clicking on the layer name in the Layers panel or by highlighting the layer name in the Layers panel and on the main QGIS menu selecting Layer > Properties.

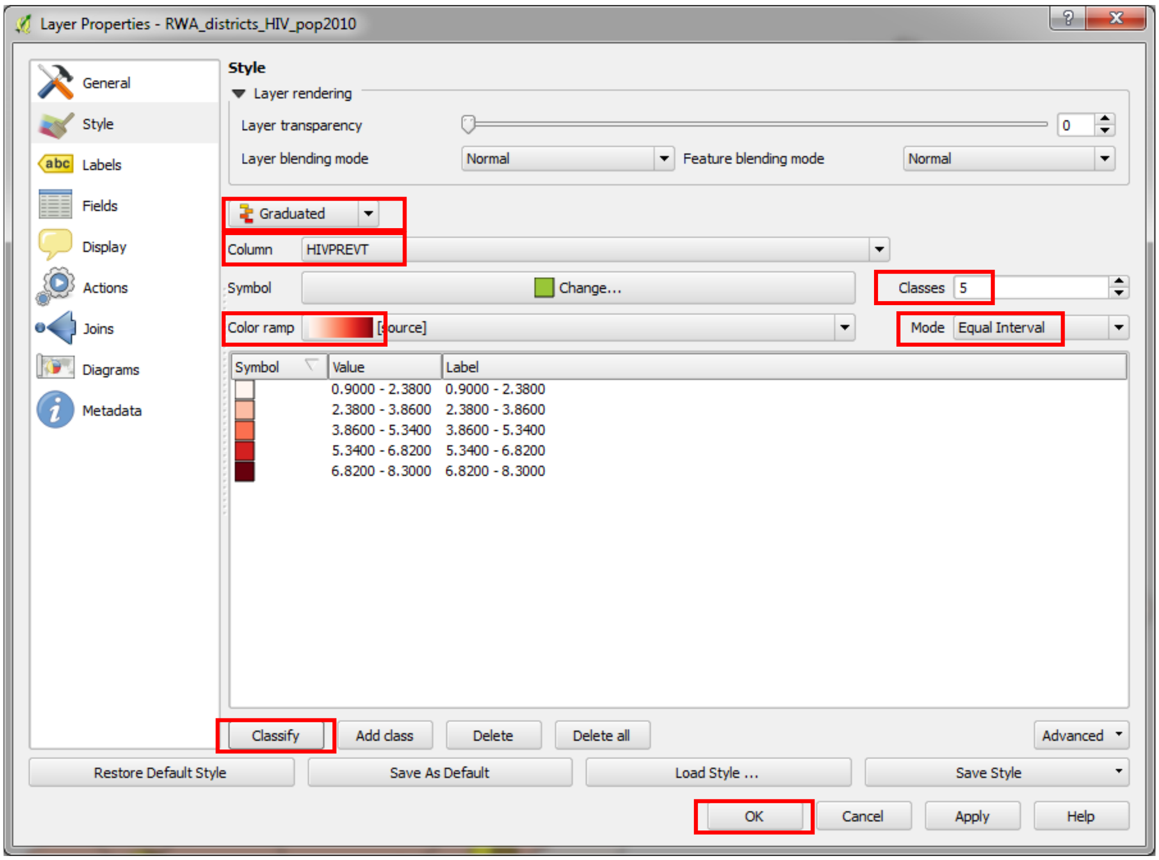
- After the Layer Properties window opens, left-click on the Style icon in order to highlight it. Because this layer is currently being used to display female HIV prevalence as a color-shaded map based on equal intervals, the symbol type should already be set to Graduated (see below).



- To map **total** HIV prevalence by district, in the pull-down menu to the right of the word Column select the column **HIVPREVT**. The ranges of values for the five classes should change automatically.

- Leave the number of classes at 5 and keep the classification mode set to Equal Interval.

- Leave the color ramp set to Reds. The dialog box should resemble the following. (If not, click on the Classify button.)



- Click on the OK button to close the dialog box and see the choropleth map of **total HIV prevalence** by district.

- To remove the overlaps in the ranges of values used to define classes, double-click on each range of values in the Values column—beginning with class 2—and increase the lower value in the range by 0.01.

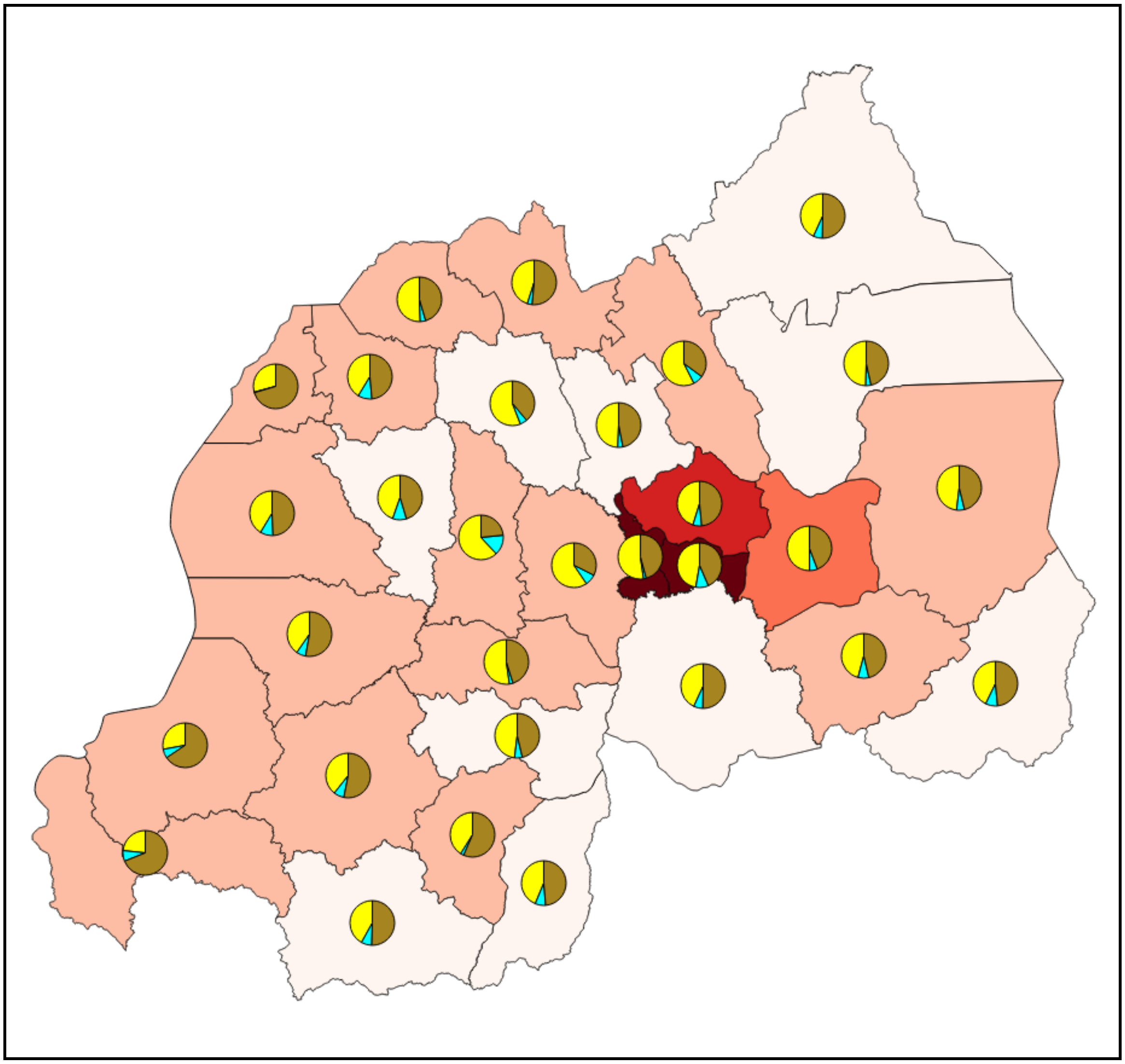
  You will also need to change the label to match the new lower limit. While you are editing the ranges displayed in the labels, you can remove the unnecessary zeroes at the ends of the labels. To do this, double-click on each range of values in the Label column and remove the zeroes.

  The resulting Value and Label columns should look as follows:

Symbol	Value	Label
	0.9000 - 2.3800	0.90 - 2.38
	2.3900 - 3.8600	2.39 - 3.86
	3.8600 - 5.3400	3.87 - 5.34
	5.3400 - 6.8200	5.35 - 6.82
	6.8200 - 8.3000	6.83 - 8.30

To save the changes, click on Apply, then OK. This will close the Layer Properties window.

The new map for total HIV prevalence among men and women age 15-49 should look like the one below. (only slightly different from the previous map)

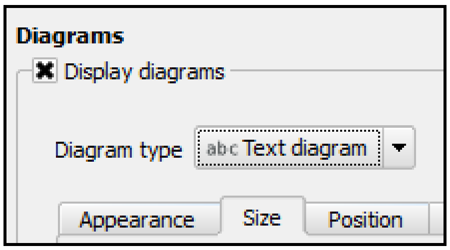


Rwanda by District, 2010, Total HIV Prevalence among Adults Age 15-49 versus Percent of Currently Married Women by Method of Contraception

- To convert the pie charts showing contraceptive methods to proportional symbols showing the relative magnitude of female HIV prevalence by district, open the Layer Properties window for layer **RWA_districts_HIV_pop2010_FP_pts** by double-clicking on the layer name.

- In the Layer Properties window for **RWA_districts_HIV_pop2010_FP_pts**, select the Diagrams option. (make sure “Display diagrams” is still checked.)

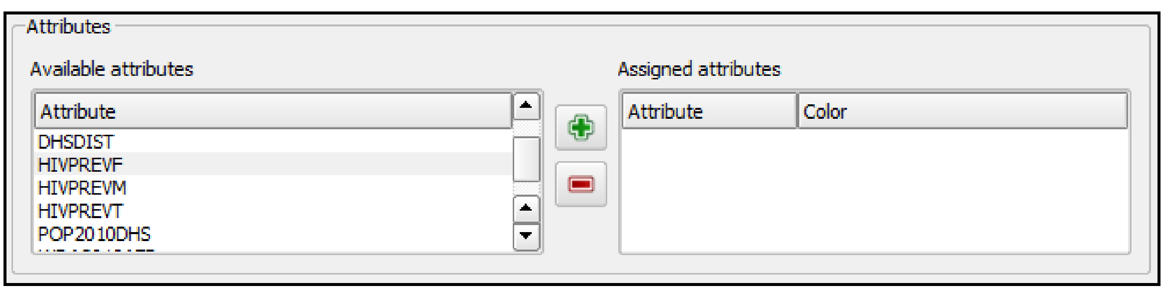
- To change the pie charts to proportional symbols, in the pull-down menu for **Diagram type**, select **Text diagram**.



NOTE: There are other options within QGIS to create proportional symbols, but in QGIS 2.0 the Text diagram option is generally the most straightforward and reliable.

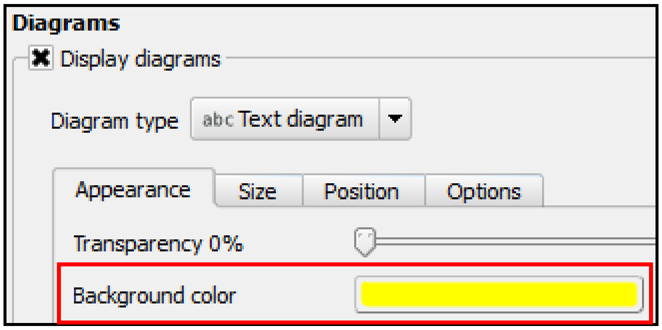
- To select female HIV prevalence as the variable to show as symbols that are proportional in size to the individual values, click on the **Appearance** tab.

- To change the variable being displayed, use the minus sign  to remove the three variables related to contraceptive methods. The **Assigned attributes** section of the Appearance tab should then be empty like the screen below.



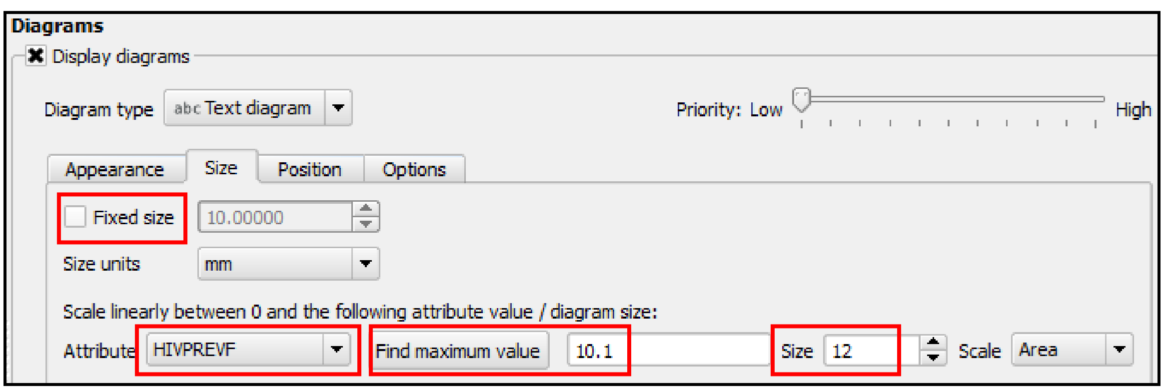
This will result in proportional symbols with no text applied as labels. In the future, if you would like to display the values for an indicator as text labels on top of the proportional symbols, you can use the plus sign to add the variable of interest to the list of assigned attributes.

- To select a new color for the proportional symbols, on the **Appearance** tab click on the colored box next to the option **Background color**. To provide contrast, you can select yellow.

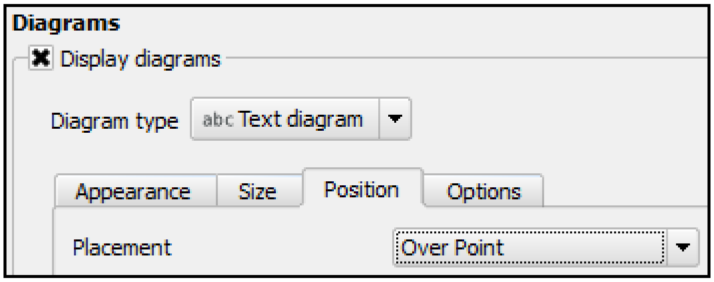


- To make the text diagrams proportional in size to female HIV prevalence:
  - Click on the **Size** tab and deselect the box for **Fixed size**.
  - For the attribute to use for linear scaling of symbol size, select **HIVPREVF**.
  - To set the value of female HIV prevalence to use as the upper threshold for proportional symbols, click on the button labeled **Find maximum value**. The maximum value for female HIV prevalence is 10.1 percent.
  - To set the maximum size to use for the proportional symbols, change the **Size** to 12. (As stated previously, selection of a size for diagrams generally can require a certain amount of trial and error, which has been done for you in this case.)

    The **Size** tab should now look like the one below.

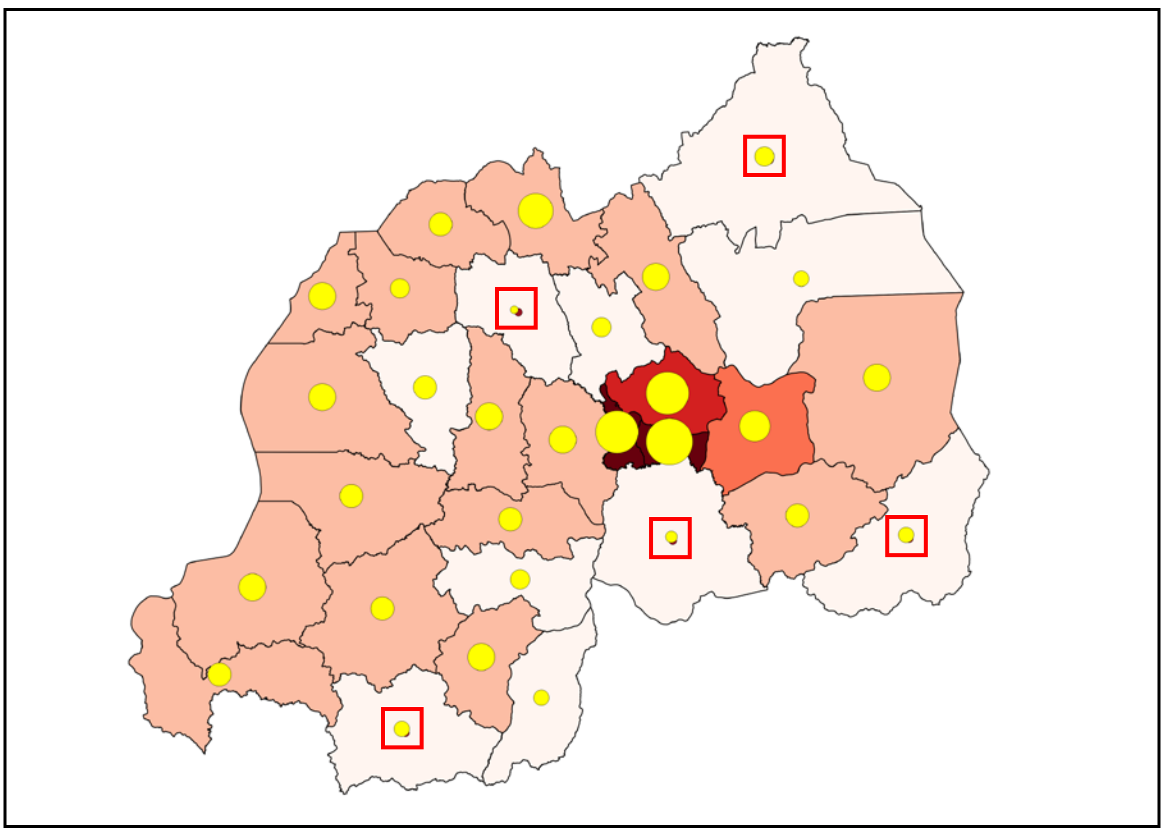


- To make sure the proportional symbols are displayed directly on top of the centroids (center points) of the districts, select the **Position** tab and select **Over Point**.

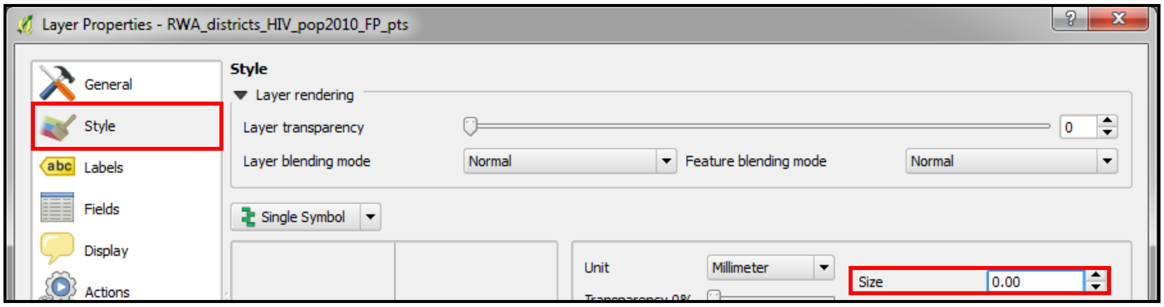


You are almost ready to display female HIV prevalence as proportional symbols. Click **apply**.

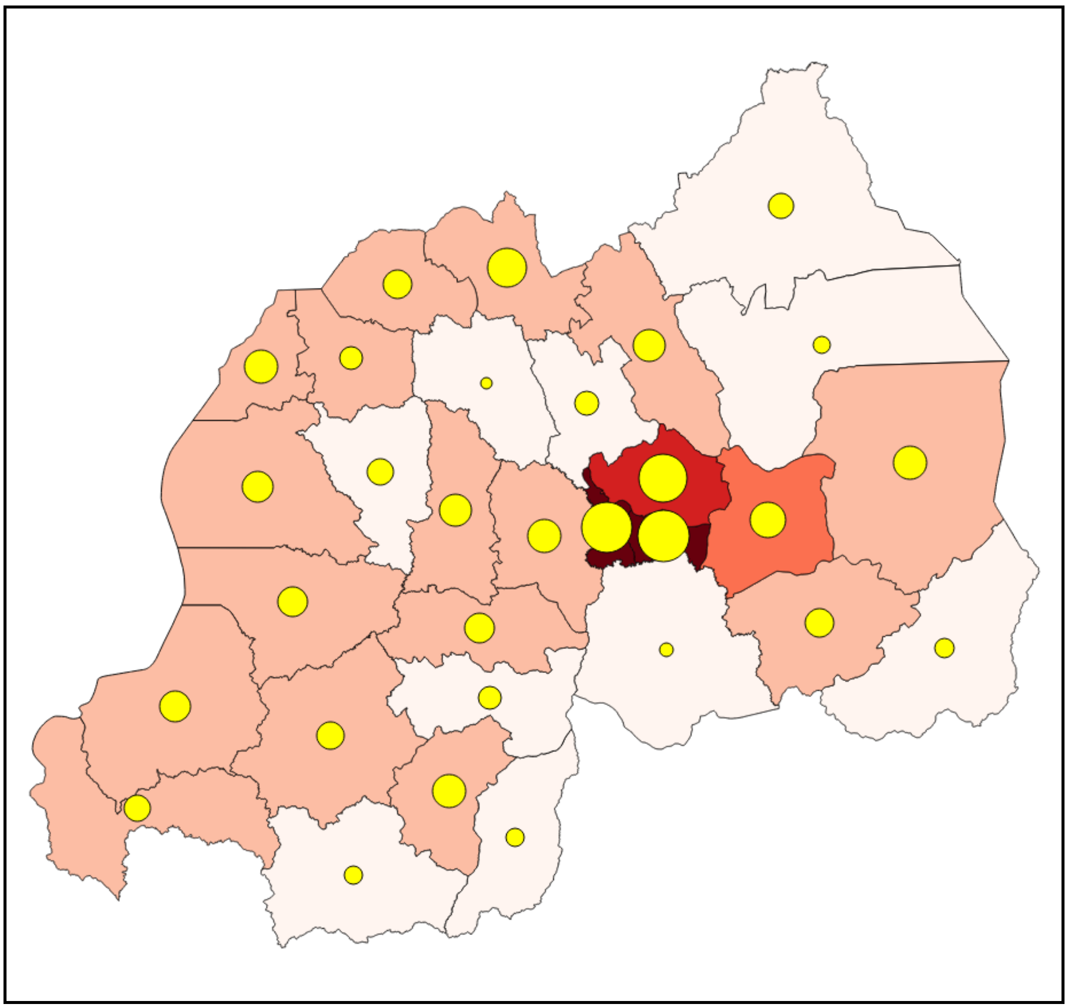
- The layer you are using as the foundation for generating proportional symbols, **RWA_districts_HIV_pop2010_FP_pts**, is displayed as a set of points. These points are separate from and independent of the diagrams you are attempting to superimpose on them as proportional symbols. As a consequence, if you use this layer to create a map in the QGIS map composer interface and export the map as an image, you might see two sets of symbols when you only want one (see below).



To avoid this problem, select the **Style** option in the in the Layer Properties window for the layer **RWA_districts_HIV_pop2010_FP_pts** and set the size of the point symbol used for the layer to zero (see below).



- To finalize the changes and display proportional symbols for female HIV prevalence, click on Apply and OK. You should see a map like the one below.



Rwanda by District, 2010, HIV Prevalence among Women Age 15-49 Superimposed on Total HIV Prevalence among Adults Age 15-49

Observation: As one would expect, given that female HIV prevalence is strongly correlated with total HIV prevalence, the size of the symbols representing female HIV prevalence increases relatively consistently with increases in total HIV prevalence.

Superimposing one variable as a set of proportional symbols on top of another variable displayed as a color-shaded map is a quick and effective way to visualize the geographic distribution of two indicators simultaneously.

- Save the QGIS project.

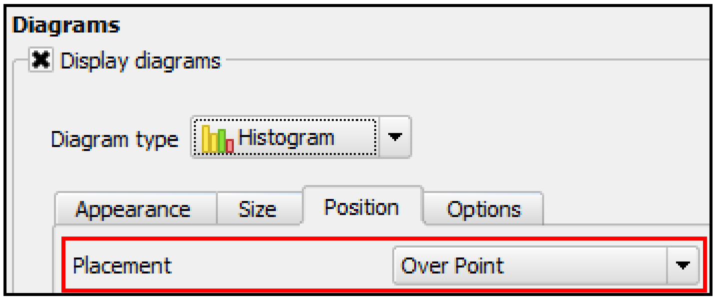
  - To save your QGIS project, go to the main QGIS menu and select Project > Save or click on the Save button. 

Step 7: Display bar charts showing HIV prevalence by sex superimposed on a base map showing total HIV prevalence by district.

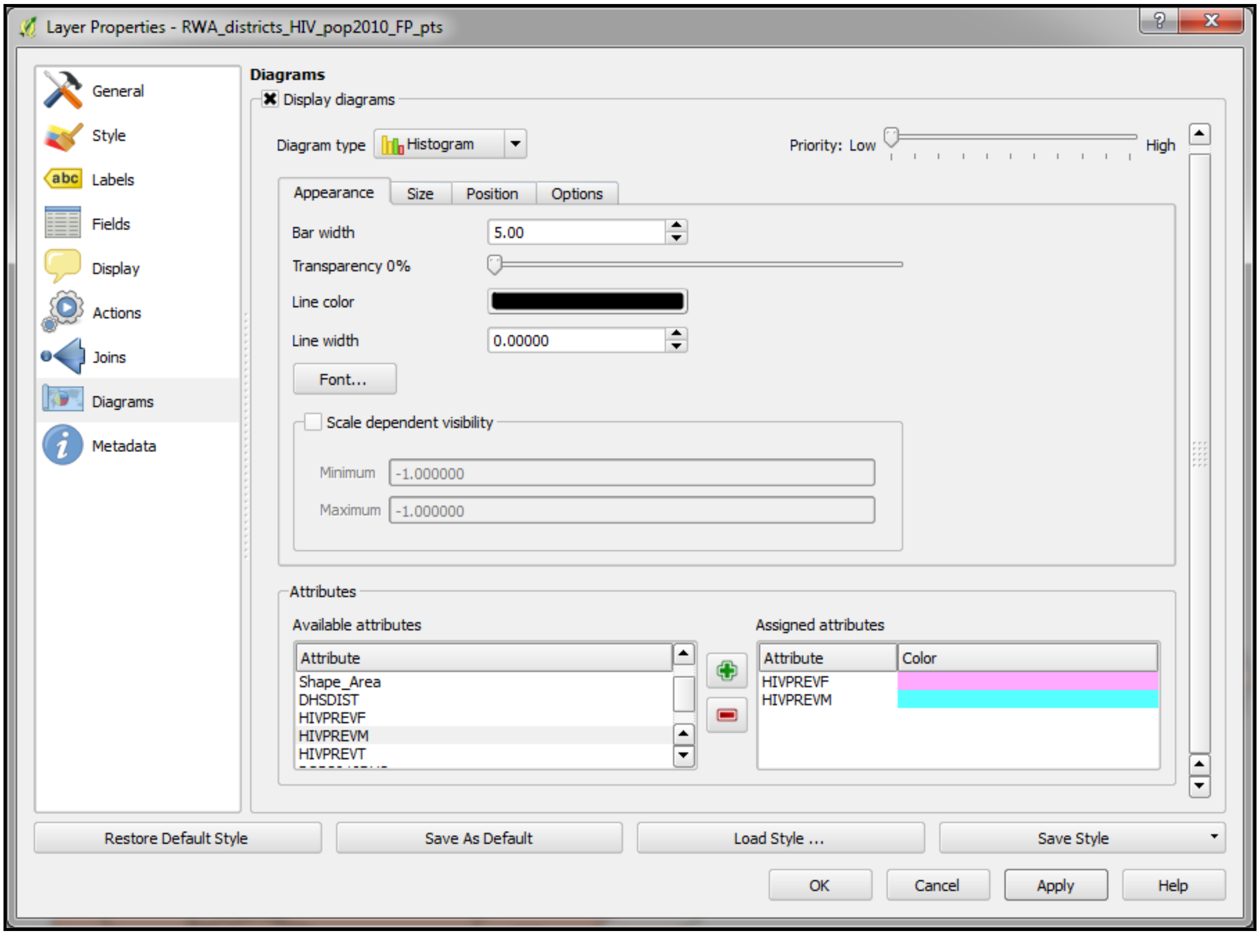
Another method for displaying a second variable on top of a choropleth map is to use bar charts, which are also called histograms. In the context of M&E of HIV/AIDS programs, this cartographic technique can be used to show whether the differences in HIV prevalence among men and women have any kind of geographic pattern.

- To begin the creation of bar charts, open the Layer Properties window for the layer **RWA_districts_HIV_pop2010_FP_pts**, select the **Diagrams** option, and choose the diagram type **Histogram**.

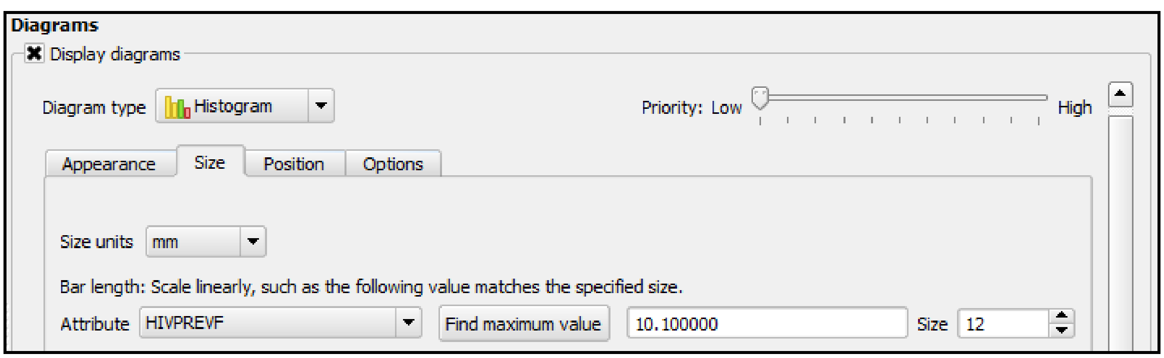
- On the Position tab, leave the Placement choice as **Over Point**.



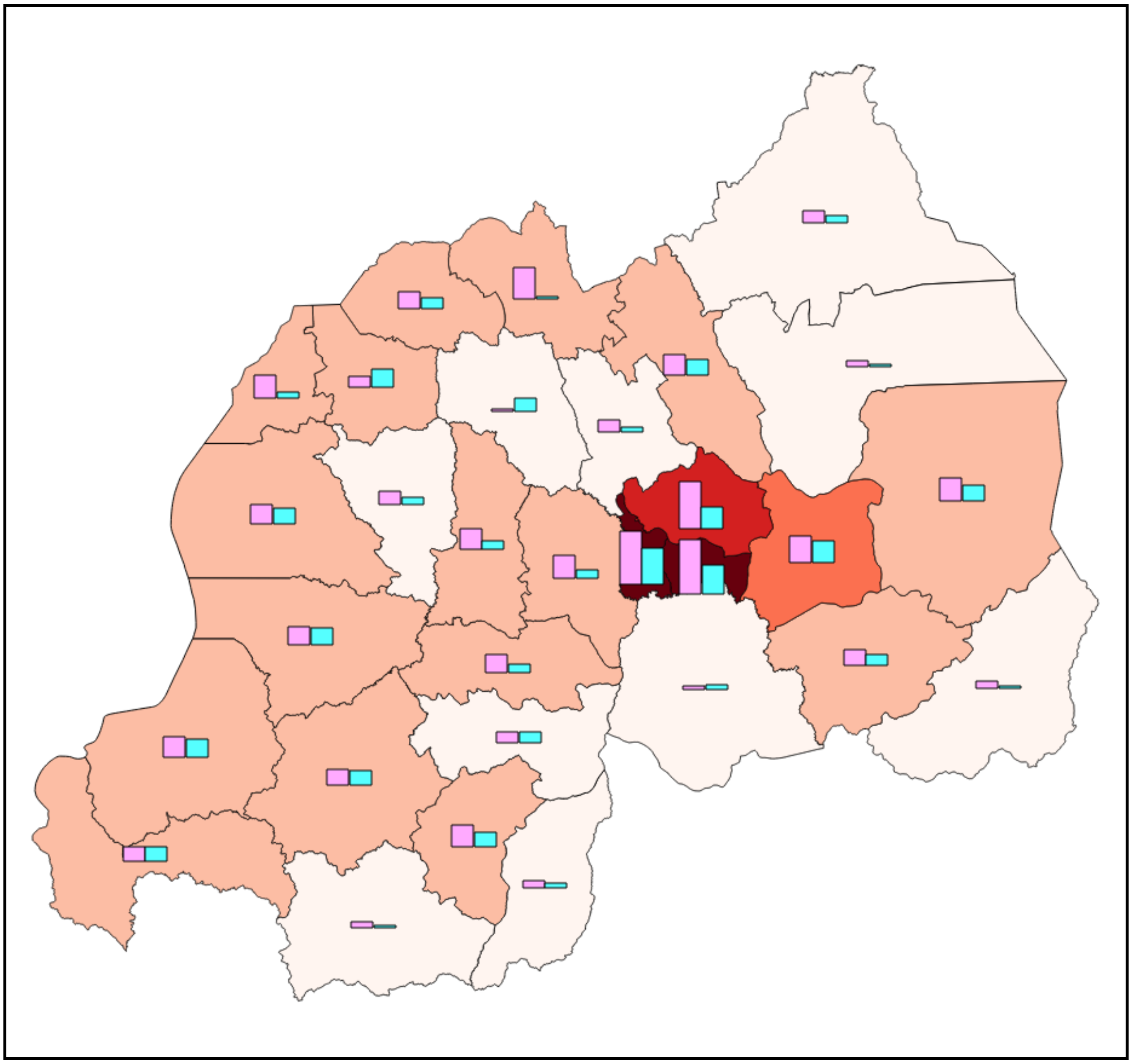
- On the **Appearance** tab, add the variables **HIVPREVF** (female HIV prevalence) and **HIVPREVM** (male HIV prevalence) to the list of assigned attributes and change the color bars to pink and blue for contrast (see below).



- On the Size tab, leave the values set as they were for the preceding step (see below).



- To display the bar charts, in the Layer Properties window click on Apply and OK. You should see a map like the one below.

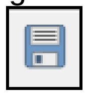


Rwanda by District, 2010, HIV Prevalence by Sex for Adults Age 15-49 Superimposed on Total HIV Prevalence among Adults Age 15-49. Pink bars represent female HIV prevalence and blue bars represent male HIV prevalence.

This map shows that HIV prevalence is highest in the three districts of the City of Kigali and that female HIV prevalence is higher than male HIV prevalence in most districts in Rwanda at the time of the Rwanda DHS 2010. It also indicates that there are a couple of districts where male HIV prevalence is uncharacteristically higher than female HIV prevalence. This information is much more targeted than indicators at the regional or national level and could be quite useful for prioritizing program activities.

NOTE: If the bar charts in your map overlap, you can experiment with the **Bar width** property on the **Appearance** tab and the **Size** value on the **Size** tab (see above).

- ☐ Save the QGIS project.

- To save your QGIS project, go to the main QGIS menu and select Project > Save or click on the Save button. 

- Quit QGIS.

  - To close QGIS, go to the main menu and select Project > Exit QGIS or use a combination of the Ctrl and Q keys.

SUMMARY

In exercise 3.2 you created a choropleth map using a single variable or indicator. In this exercise you produced several different types of maps showing more than one indicator at a time. As you may have noticed, however, none of the maps you have generated include a title, legend, or description of the data sources used. These are basic cartographic elements that are required to make a map ready for formal publication.

In exercise 4.3, you will add these cartographic elements and produce a publication-quality map.

END

GIS Techniques for Monitoring and Evaluation of HIV/AIDS and Related Programs

Exercise 4.3

Create Publication-Quality Maps

**This training was developed as part of a joint effort between MEASURE Evaluation and MEASURE DHS, with funding provided by USAID and PEPFAR.*

Exercise 4.3: Create Publication-Quality Maps

Background: All free and open source GIS software packages can create maps, but not all of them can create publication-quality maps without assistance from third-party software packages such as Adobe Illustrator or Microsoft PowerPoint. For single indicator maps, however, QGIS provides a map composer interface that provides sufficient functionality for generating publication-quality maps.

Summary: In this exercise, you will use the map composer interface within QGIS to create a publication-quality map of HIV prevalence by sex as bar charts superimposed on total HIV prevalence among adults age 15-49, which is the last map produced for exercise 4.2.

To help you overcome the limitations of QGIS with respect to the generation of multi-indicator legends, some additional guidance on that topic will be provided at the end.

Objectives:

- Introduce the Map Composer window in QGIS.
- Compose a map showing HIV prevalence by sex as bar charts superimposed on a base map showing total HIV prevalence by district.
- Export the map for publication.

Requirements*:*
To complete these exercises, you will need to have QGIS and several plugins installed on your machine, as instructed in exercise 2.1. You will also need to have downloaded and unzipped the data files associated with the exercise. Remember, these files can reside anywhere on your computer, but must be kept together, with the same original file structure.

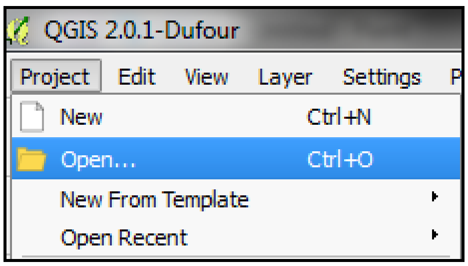
Step 1: Launch QGIS.

- Option 1: Click on the desktop shortcut.

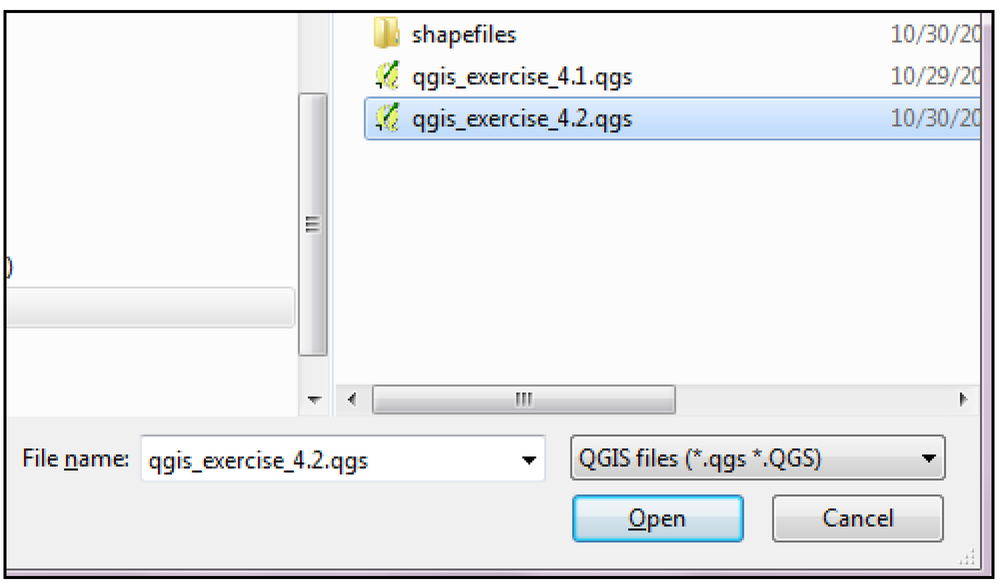
- Option 2: Click on Start > All programs > QGIS Dufour > QGIS Desktop 2.0.1.

Step 2: Create a new project from the one produced for exercise 4.2.

- On the main QGIS menu, select Project > Open.



- Navigate to the folder “Exercises 4.1, 4.2, and 4.3” your folder may have a different name—just use the one you have been working in) and open the file **qgis_exercise_4.2.qgs**.



- To save a new project, save the current project with a new name (see below).
  - On the main QGIS menu, select Project > Save As.

- For the output file name, specify the following:

  Exercises 4.1, 4.2, and 4.3**qgis_exercise_4.3.qgs**

- To complete the process, click on the Save button in the file dialog window.

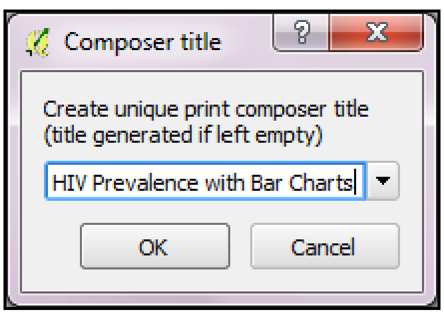
  

Step 3: Compose a publication-quality map showing HIV prevalence by sex as bar charts superimposed on a base map showing total HIV prevalence by district.

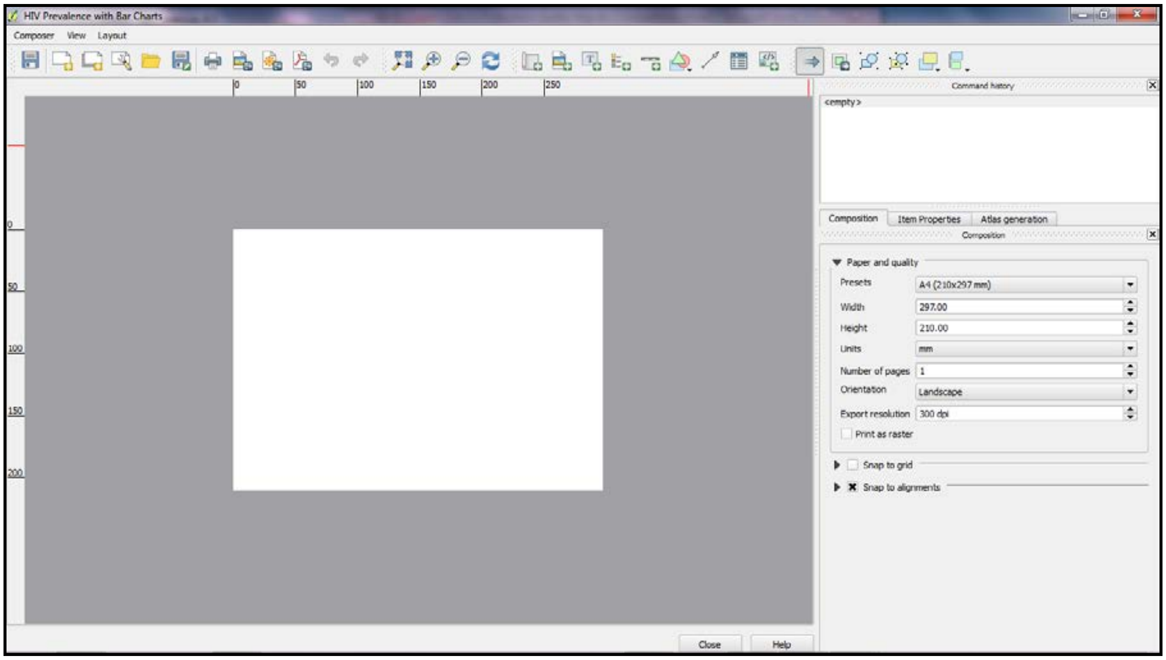
First you will compose a map using the layers currently displayed in the map window.

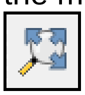
- To begin map composition, go to the main QGIS menu and click on the **New Print Composer** button. 

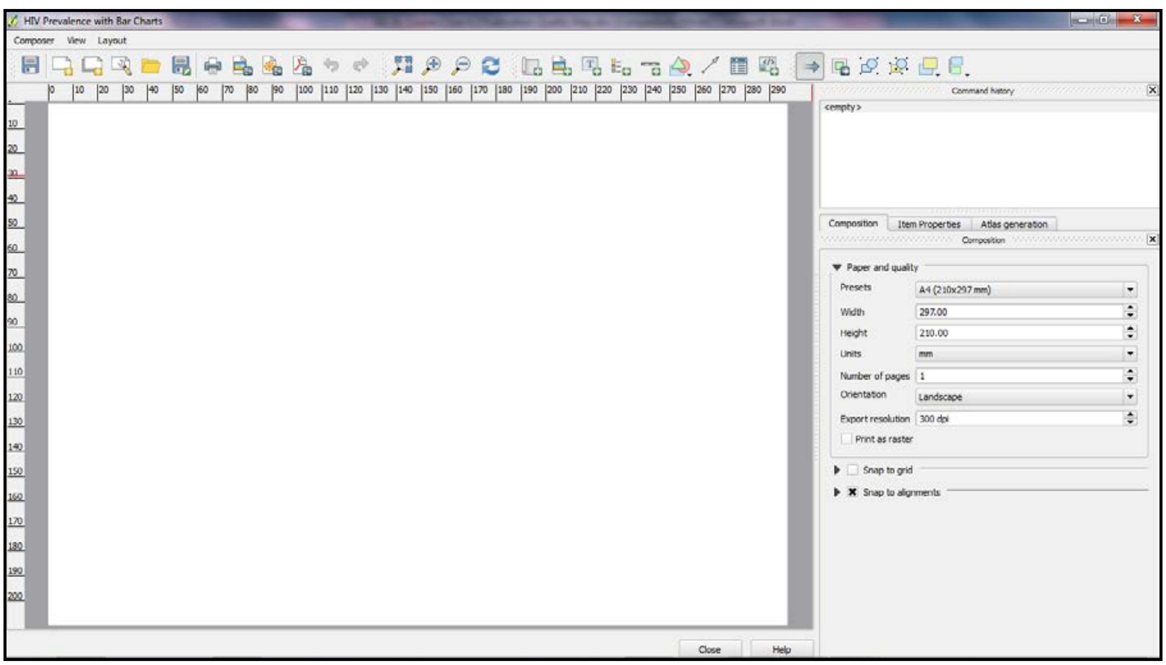
- In the dialog box that appears, enter a title for the map composition. (This is optional, and is not the title of the map, just the title of the map layout so that you can use it again in the future. The default name, if the box is left empty, is “Composer1”.)

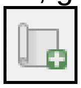
  

- Click on the OK button. You should see a map composer window that looks like the one below.

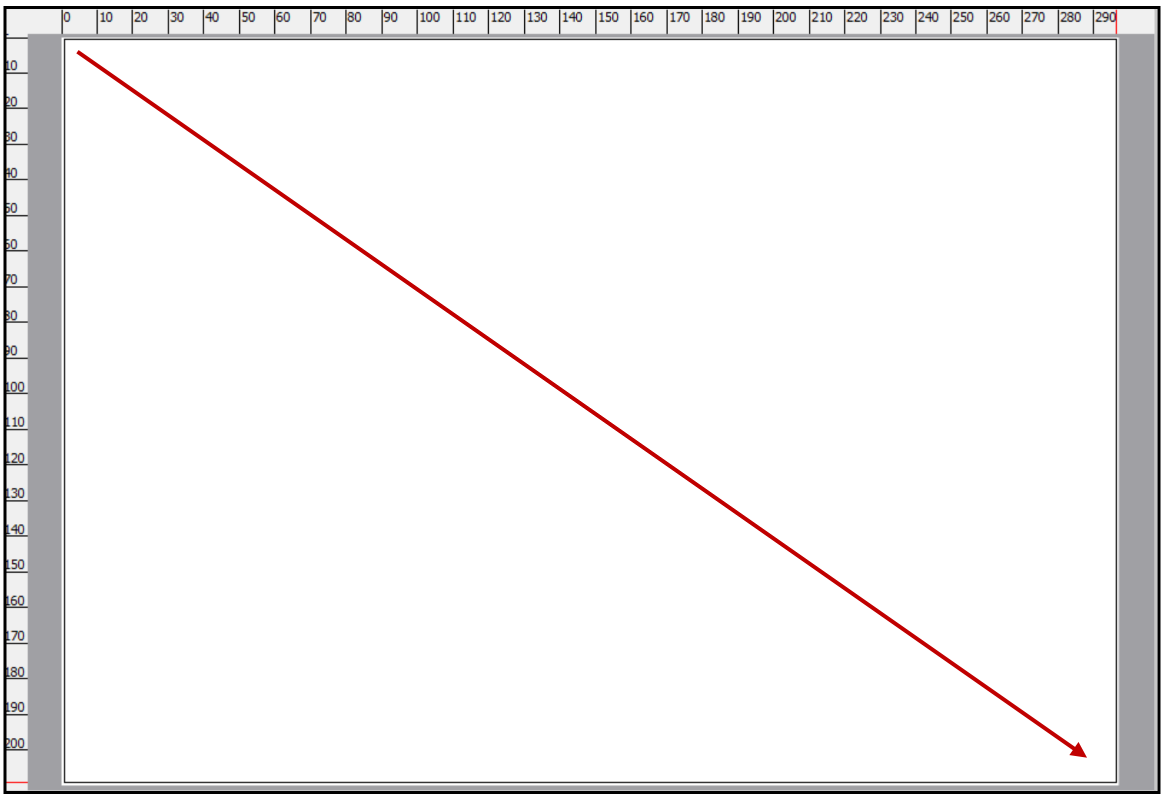


- To enlarge your view of the map composition canvas, which is the white rectangle in the middle of the window, go to the top menu and click on the **Zoom full** button.  You should now see a window that resembles the one below.

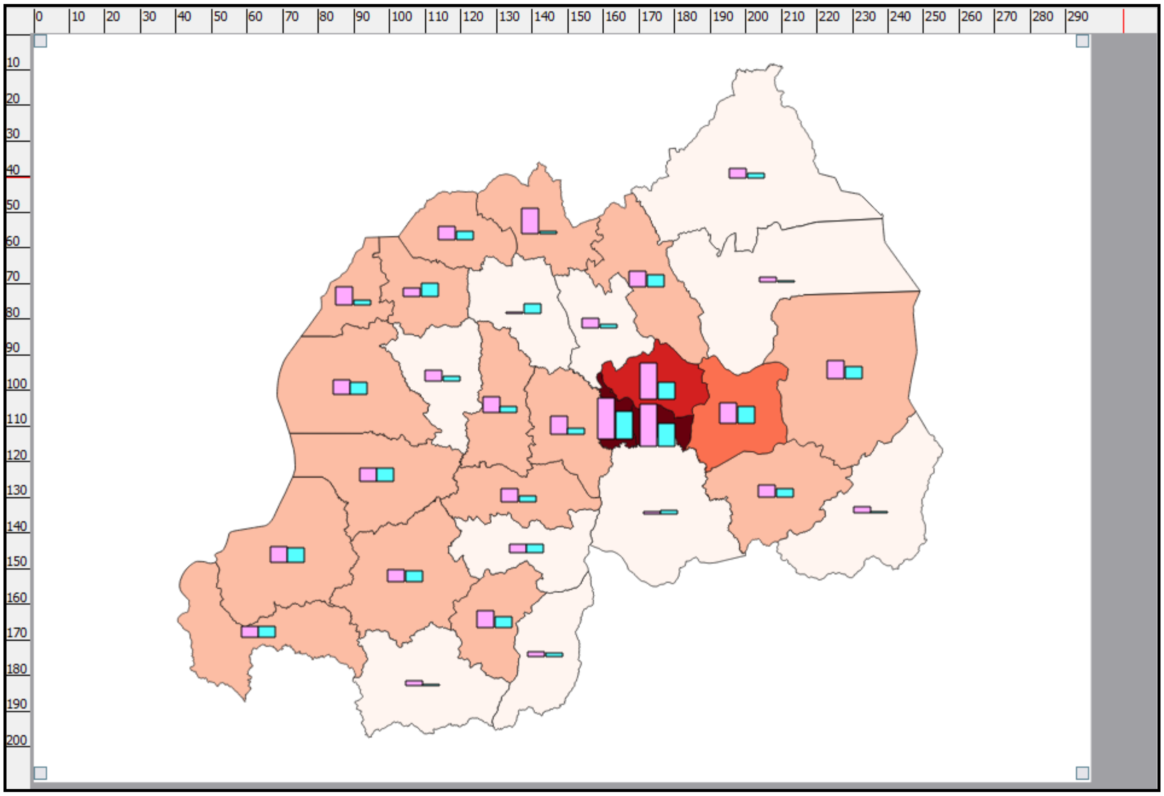


- To add the map in the QGIS map window to the map composition canvas in the map composer, go to the top menu in the map composer and click on the **Add new map** button. 

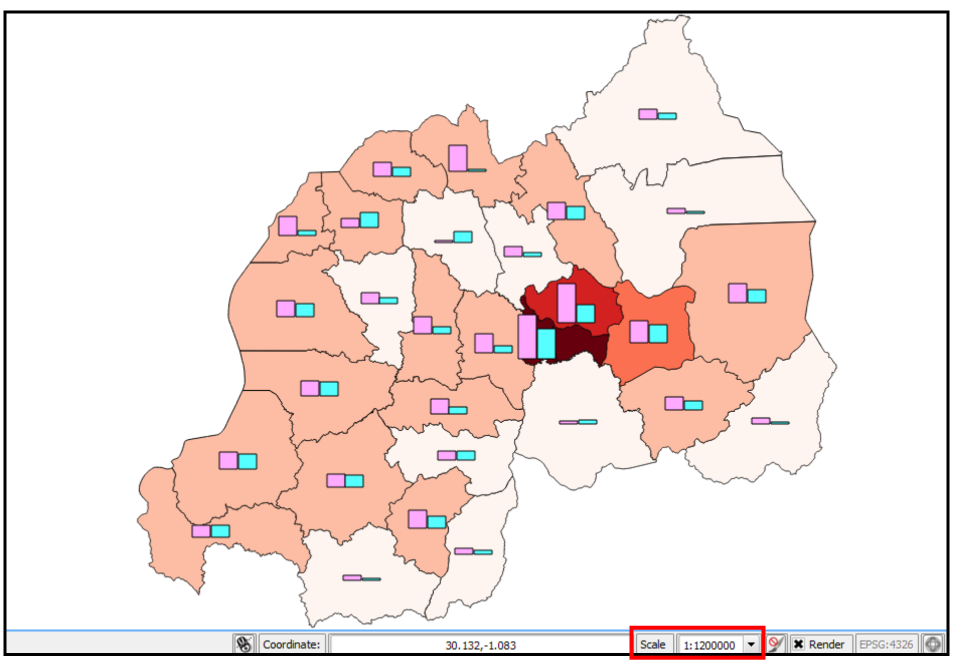
- To indicate the area on the map canvas in which you would like to place the map, left-click in the upper-left corner of the canvas and drag the cursor to the lower-right corner of the canvas in order to draw a box that occupies the entire map canvas. This should give you a black rectangle just within the edges of the map canvas (see below).



- To see the map on the canvas, release the left click on the mouse or keyboard. You should see a map that resembles the one below.



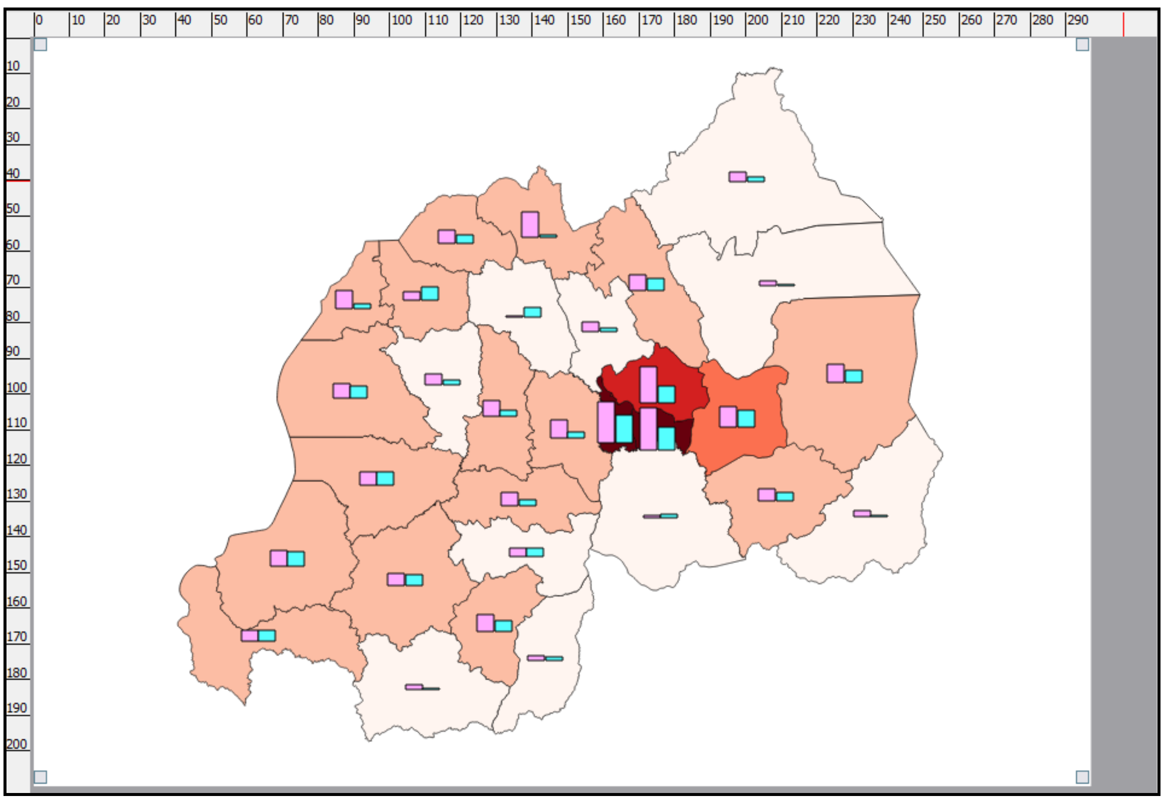
NOTE: If the map is too large to fit on the map canvas, or is too small and creates too much white space, you can click on the map composition window to select the data frame holding the map and then press the Delete key on your keyboard. That will remove the data frame from the canvas and allow you to go back to the QGIS map window and adjust the scale at which the map is displayed (see below).

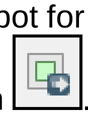


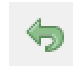
Since the scale represents the size of the map as a fraction of the true size of the corresponding geographic area on the Earth (for example, a scale of 1 to 1, or true size, would be 1:1), to shrink the map increase the denominator (the figure 1200 in the scale 1:1200). To make the map larger, make the scale larger by decreasing the denominator (for example, change 1:1200 to 1:1000). A good scale to try as a starting point for fitting the map within the map composer window with a good balance of white space is 1:1200000 (the denominator has five zeroes).

After you have set a new scale in the QGIS map window (if necessary), return to the map composer window and click again on the **Add new map** button. With the **Add new map** button active, draw a rectangle on the map canvas and release the left-click. Repeat these steps until you are satisfied with the scale at which your map is displayed in the map composer window.

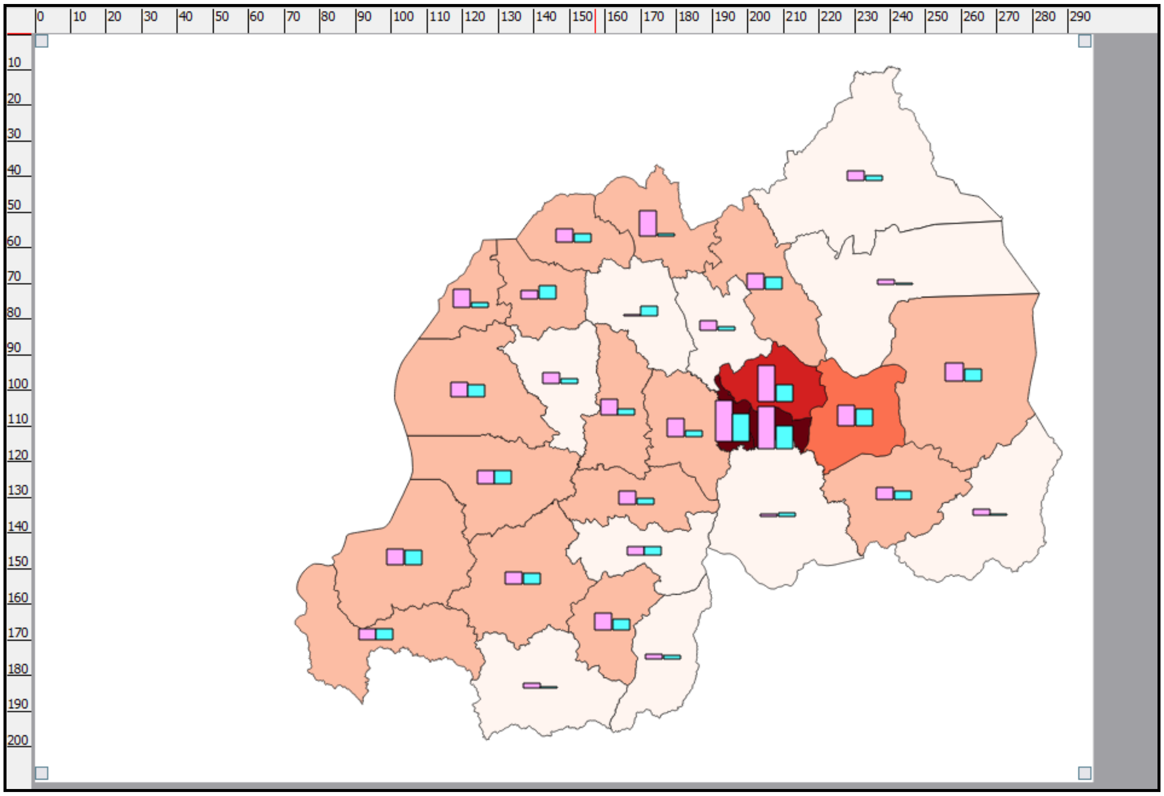
The end result in the map composer window should look like the screen capture below.



- If the amount of white space around the map is reasonable (not overwhelming or razor thin) but the map is not positioned in a good spot for adding a title, legend, etc., you can click on the **Move item content** button . With the **Move item content** button selected, you can left-click on the map window and drag the map where you would like it.

(If you make a mistake, you can click on the **Revert last change**  button.)

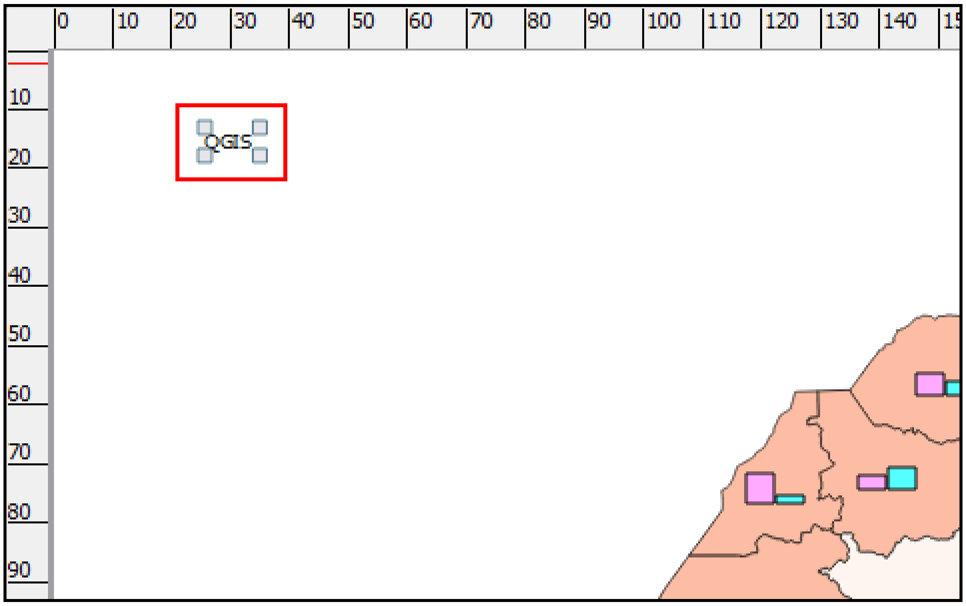
- To accommodate a title and other cartographic elements, move the map content to the center-right in the map composer window (see below).



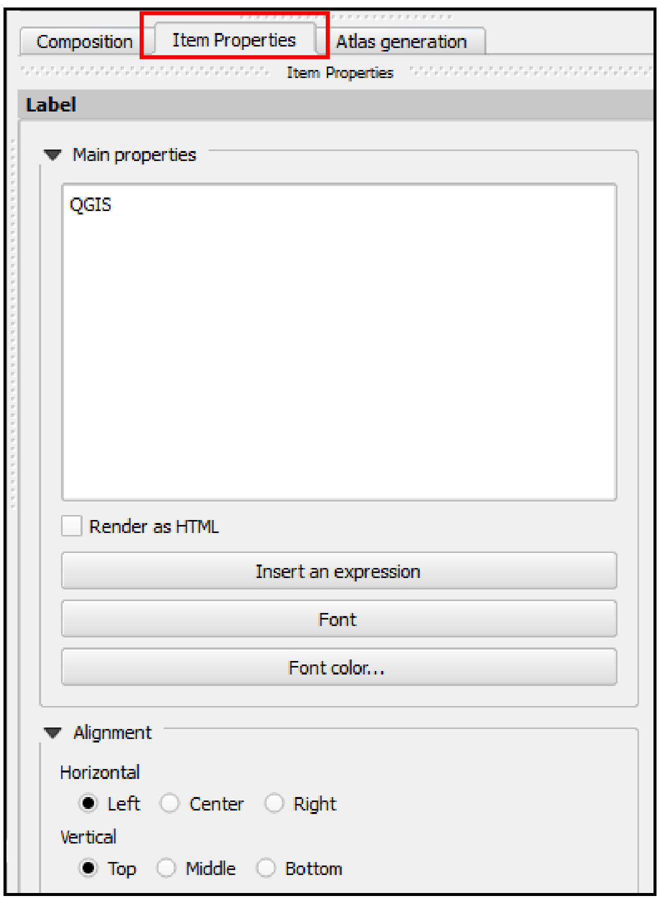
The map content should not be touching the edges of the map canvas, and the three closest sides of the map canvas should be about the same distance from the map content.

When adding the remaining cartographic elements, you should attempt to fill in white space with useful information while avoiding clutter.

- To add a title, on the map composer main menu click on the **Add new label** button  and left-click in the white space near the top of the map composer window. You should see a text box that contains the default label of “QGIS” (see below).



- To change the title to something more meaningful, make sure the new label is active (the corners of the text box will display small blue squares) and click on the **Item Properties** tab on the right side of the map composer window (see below).

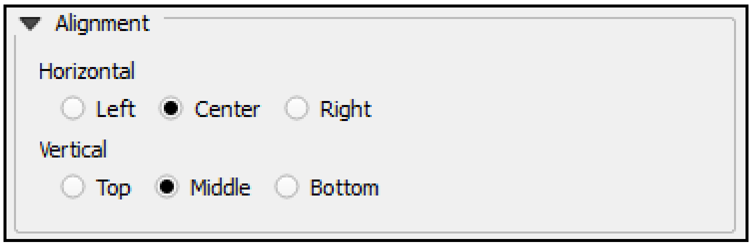


- In the text box under Main properties, replace the word QGIS with the following:

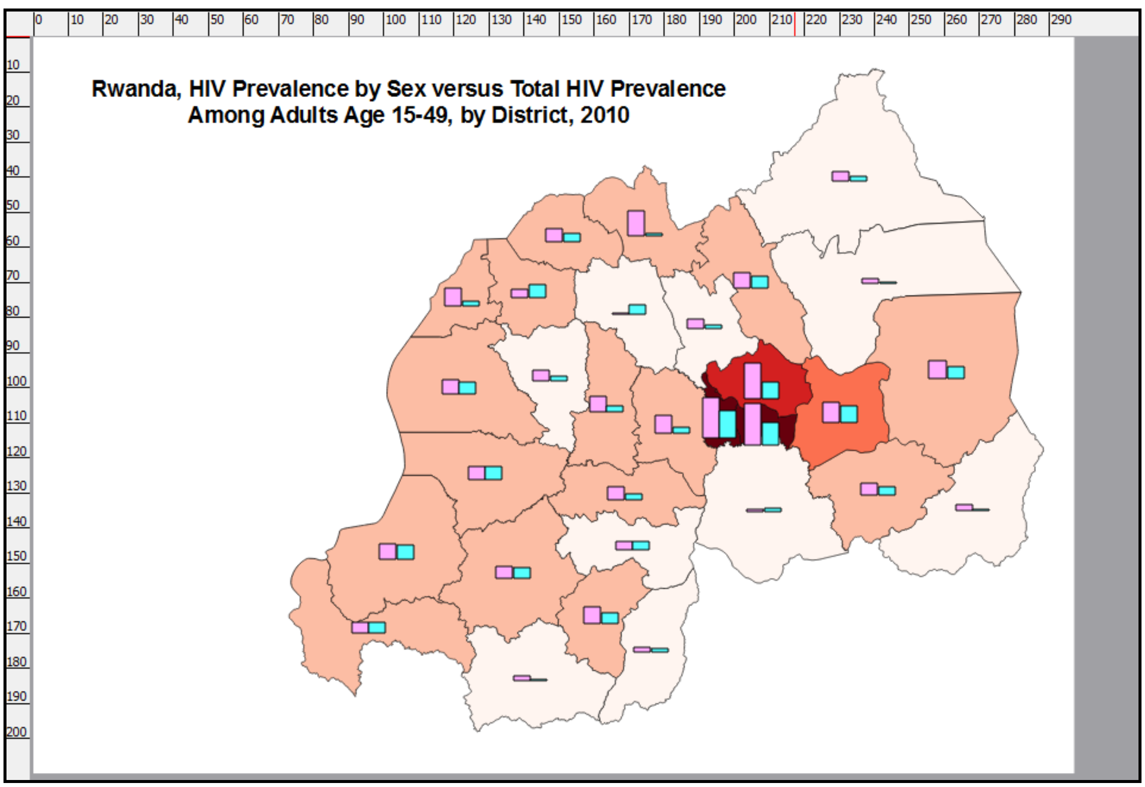
  Rwanda, HIV Prevalence by Sex versus Total HIV Prevalence
  Among Adults Age 15-49, by District, 2010

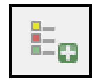
- To change the font, click on the **Font** button underneath the **Main properties** text box and select font = Arial, font style = Bold, size = 18, and click on OK.

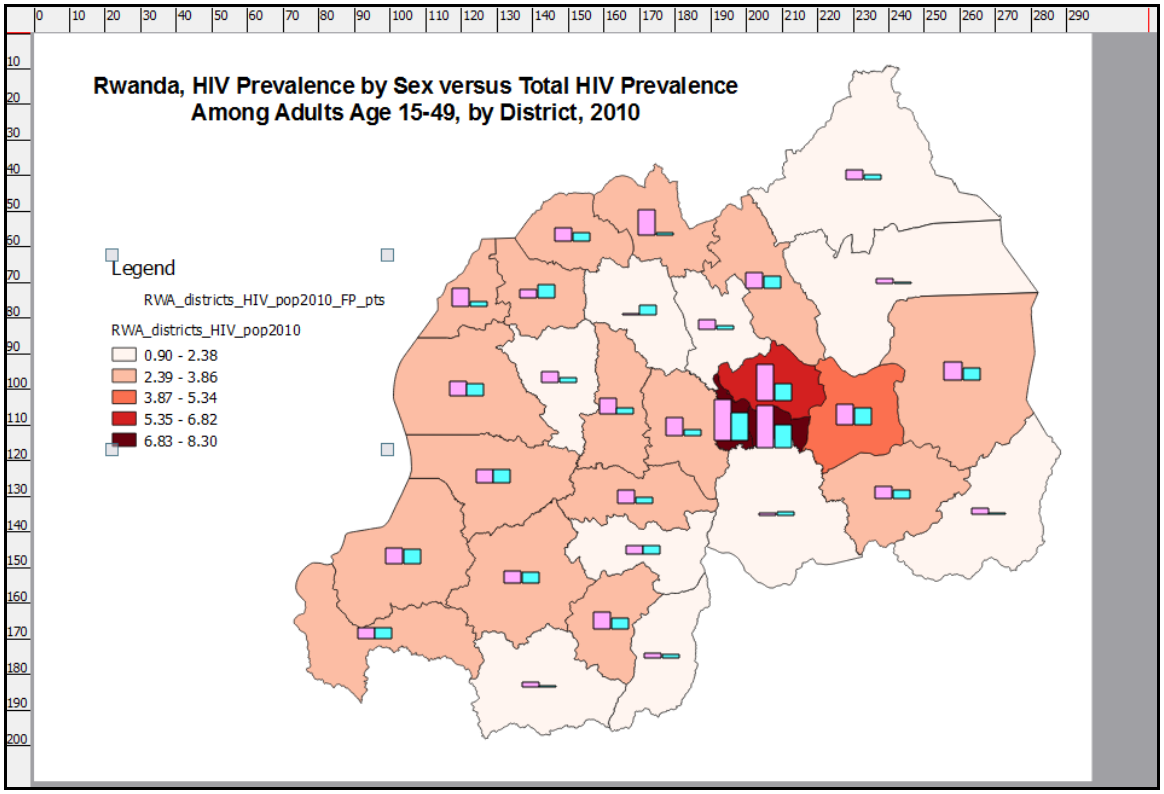
- To center the title in the text box, go to the Alignments section underneath the Main properties section and select Horizontal = Center and Vertical = Middle (see below).



- The default text box size for the new title was too small to display the full text, so you will need to left-click on one of the corners of the text box and resize it. The end result should look like the one below.

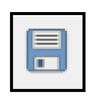


- To add a legend, click on the **Add new legend** button  and left-click in the available white space in the map composer window. You should see a map canvas that looks like the one below.



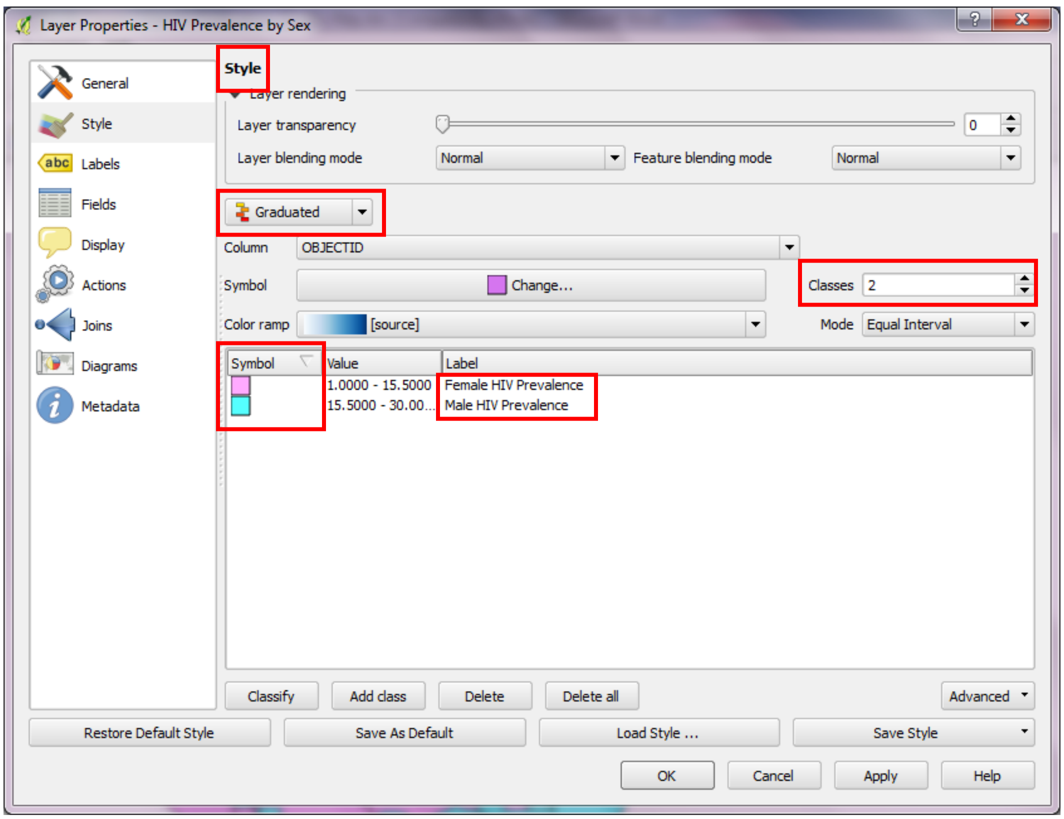
- There are problems with the legend: (1) the labels for the data layers are not descriptive and (2) the layer that corresponds to the bar charts does not show colored bar chart graphics to facilitate map interpretation.

  To correct these issues:

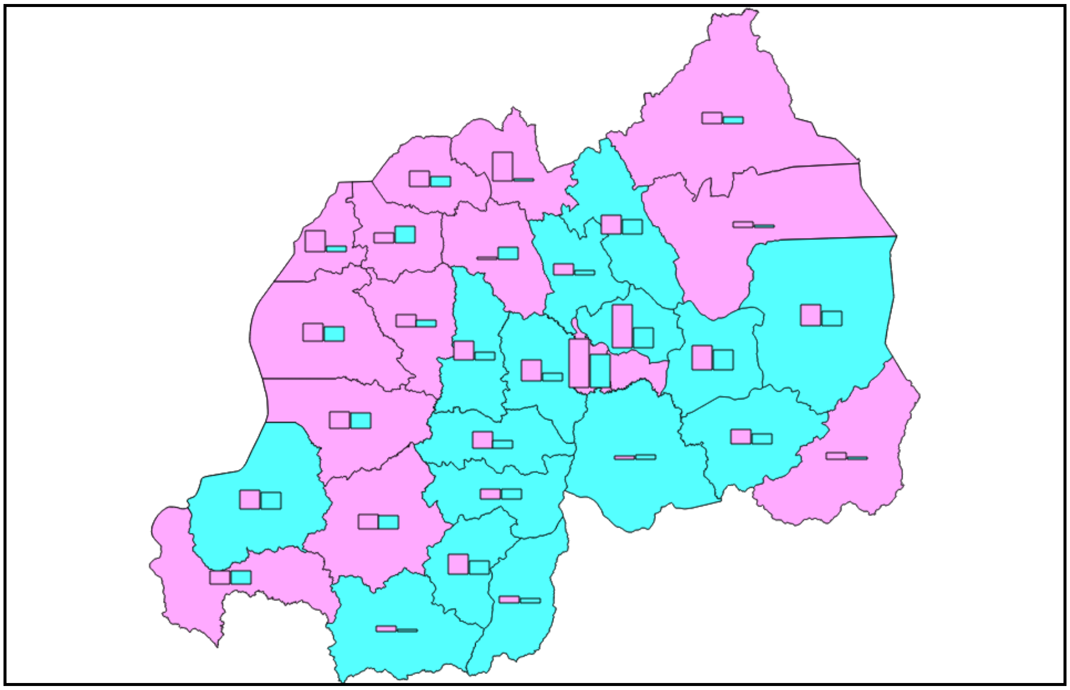
  - First save the project by clicking on the **Save project** button. 
  - *Switch back to the main QGIS interface* and make the following changes to the layer properties:
    - Turn back on the layer **RWA_districts_HIV_pop2010_FP** and rename it “HIV Prevalence by Sex”. To rename the layer, you can highlight the layer name and press the F2 button on the keyboard to go into editing mode, or you can enter a new name by highlighting the layer name in the Layers panel and on the main QGIS menu selecting Layer > Properties > General > Layer name.

- Open the properties window for the layer that is now named **HIV Prevalence by Sex.** To perform this task, double-click on the layer name or highlight the layer name and select Layer > Properties.

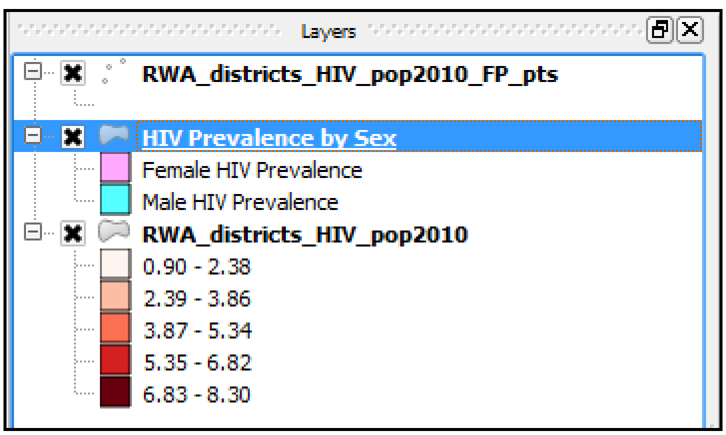
- In the Layer Properties window for the layer **HIV Prevalence by Sex**, select the **Style** option and select Graduated symbols with 2 classes. Then change the labels for the two classes to be Female HIV Prevalence and Male HIV Prevalence. Finally, double-click on the colored symbols to the left of the class values and labels and change the colors to match those used for the bar charts (see below).



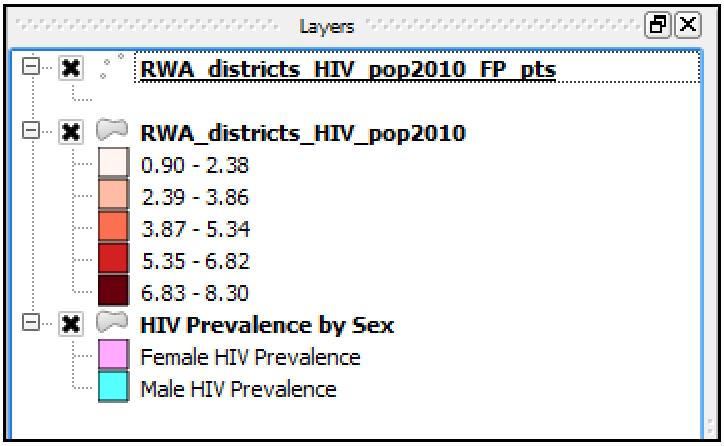
- To save the changes, click on **Apply** and **OK**. You should see a map that looks like the one below.



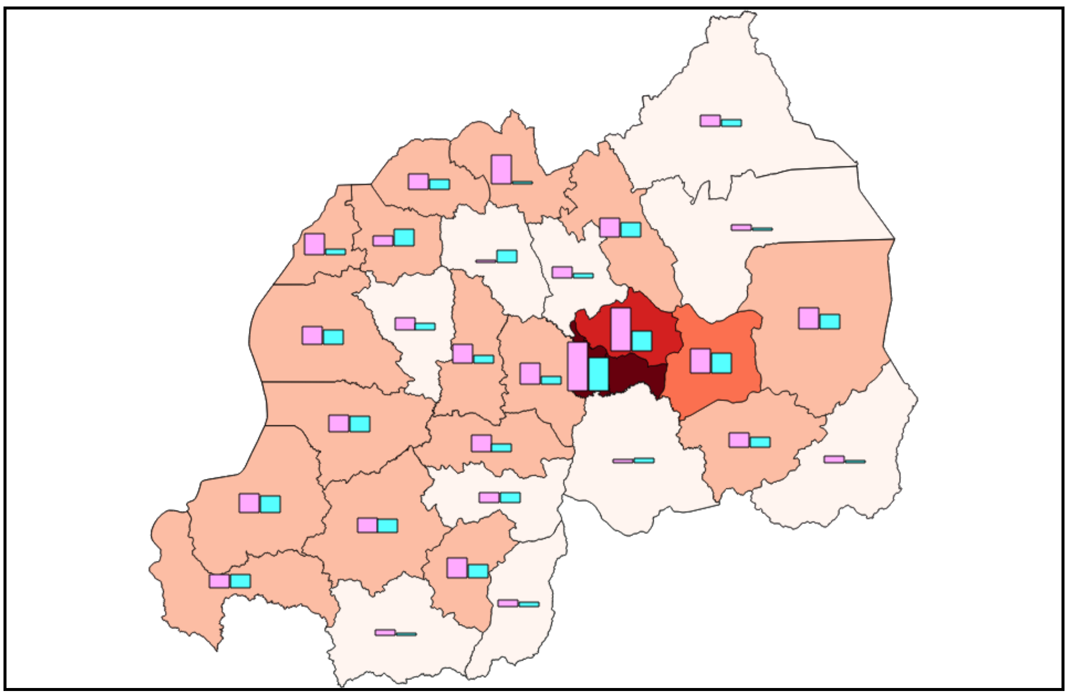
The choropleth map now shows just the two colors used for the bar charts, because the layer in the Layers panel corresponding to the choropleth map is on top of the layer for total HIV prevalence (see below).



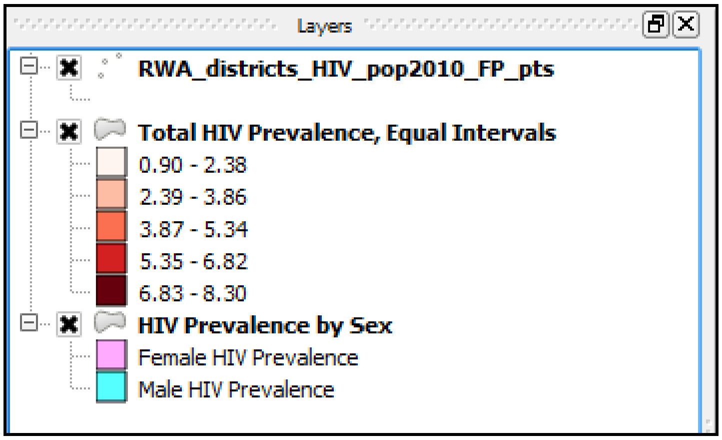
- To retain the layer **HIV Prevalence by Sex** for display in the legend but to remove it from sight in the map window, left-click on the layer name and drag the layer to the bottom of the stack in the Layers panel (see below).

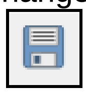


The map window should once again look correct (see below).

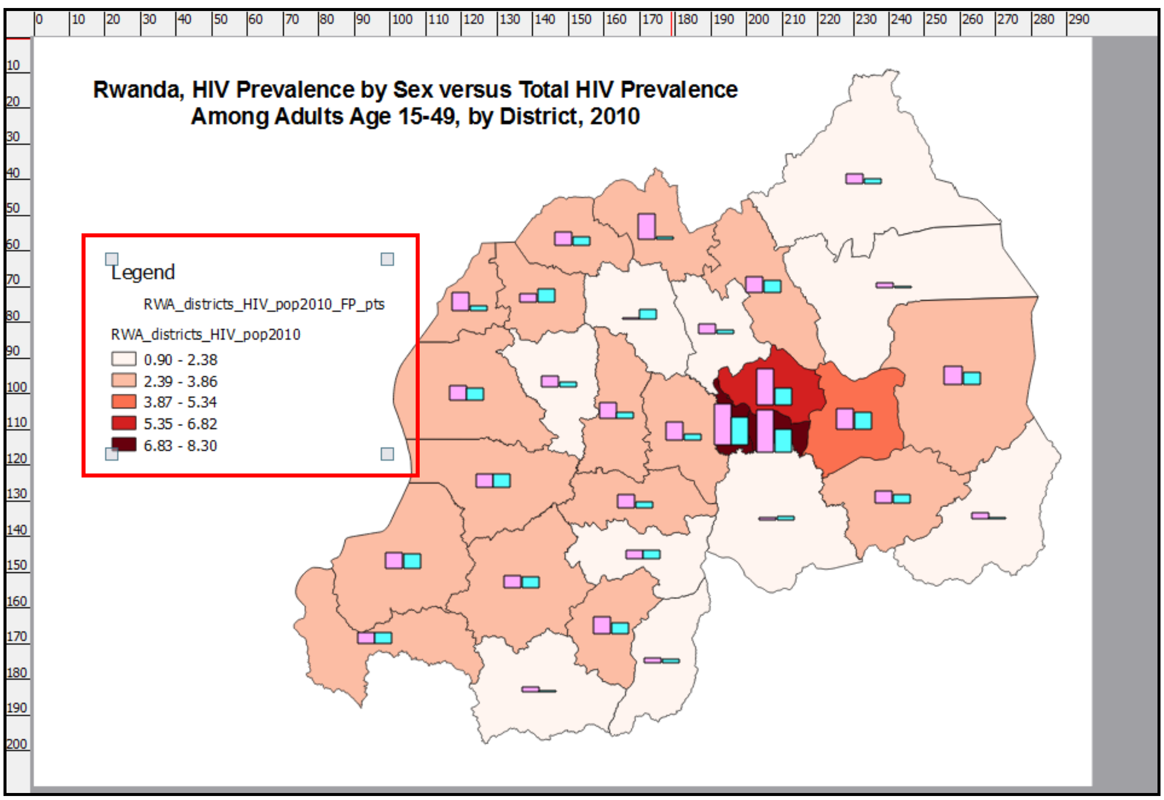


- Rename the layer **RWA_districts_HIV_pop2010** to **Total HIV Prevalence, Equal Intervals** (see below). Note that it is good practice to identify the type of classification scheme used for choropleth maps, in this case equal intervals.

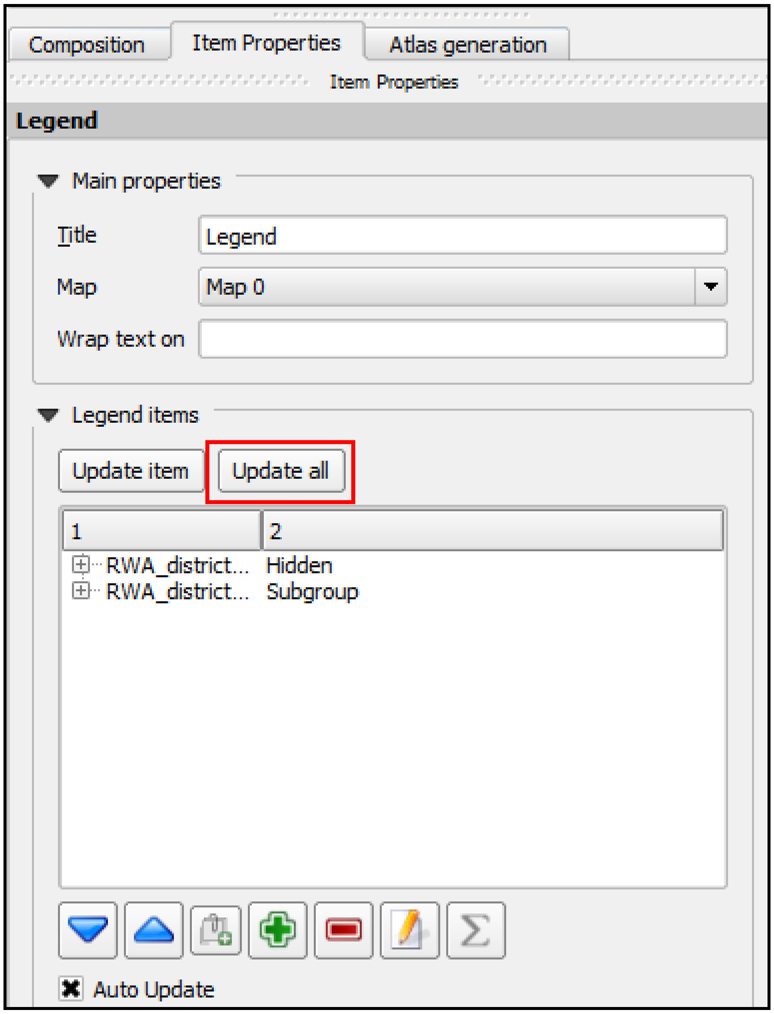


- After making these changes, once again save the project by clicking on the **Save project** button. 

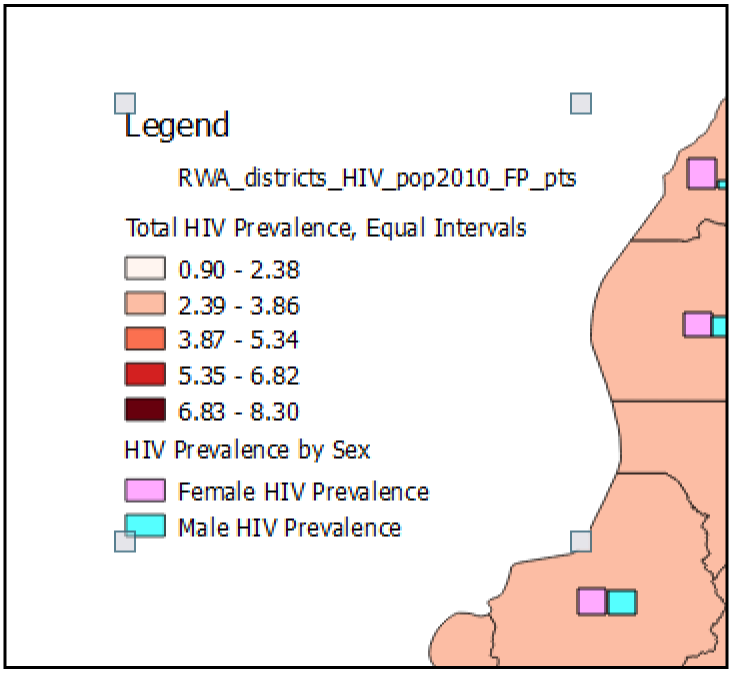
- After saving changes to the project, return to the map composer interface and left-click on the legend on the map canvas in order to make it active (see below).



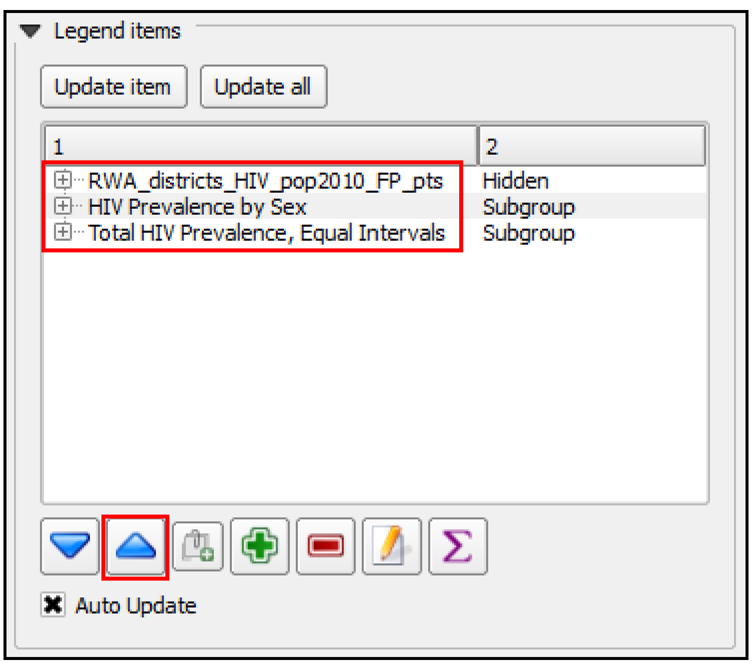
- With the legend selected on the map canvas, go to the **Legend items** section on the **Item properties** tab, which should be located on the right side of the map composer screen, and click on the button **Update all** (see below).



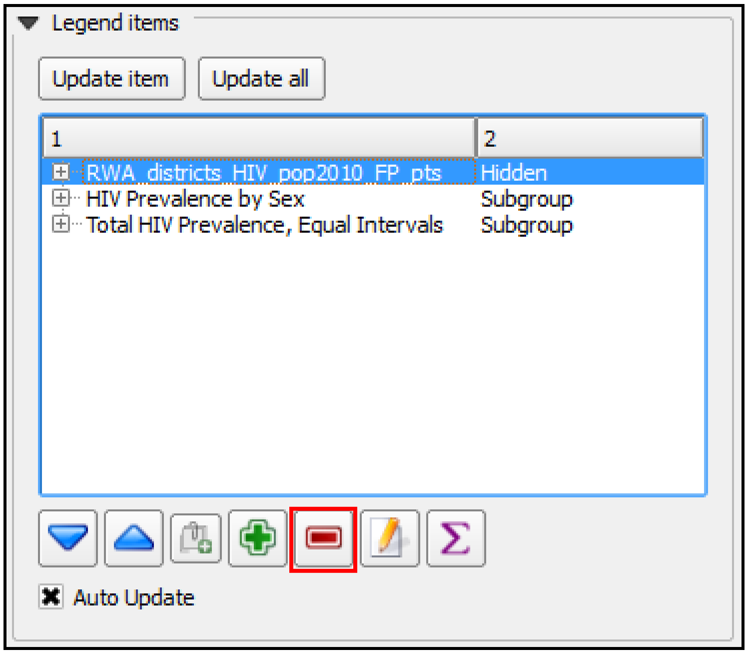
The legend in the map composer window should now look like the one below.



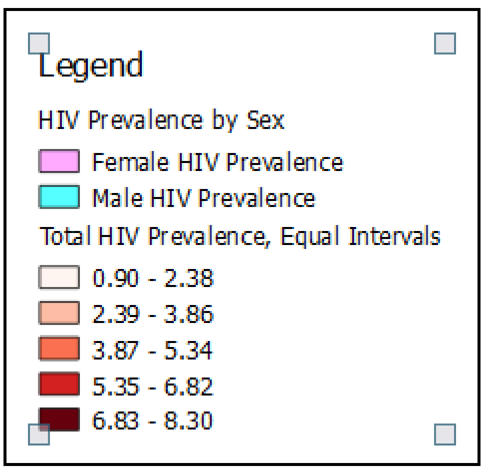
- To clean up the legend, go to the **Legend items** section of the **Item properties** tab, highlight the layer **HIV Prevalence by Sex**, and use the up arrow to move it above the layer **Total HIV Prevalence, Equal Intervals** (see below).



- Since the layer **RWA_districts_HIV_pop2010_FP_pts** is not being used for the legend, it can be removed from the list of legend items. To remove this layer from the list, highlight the layer name in the Legend items panel and click on the minus sign (see below).



- After removing this unnecessary layer from the list of legend items, the legend displayed on the map canvas should now look similar to the one below.



- The legend has been corrected, but to help orient potential users of the map it would be beneficial to add a new layer to show the provincial boundaries.

- To avoid losing your changes, save the project before advancing.

- To add a new layer for the provincial boundaries, return to the main QGIS interface and click on the Add Vector Layer button. 

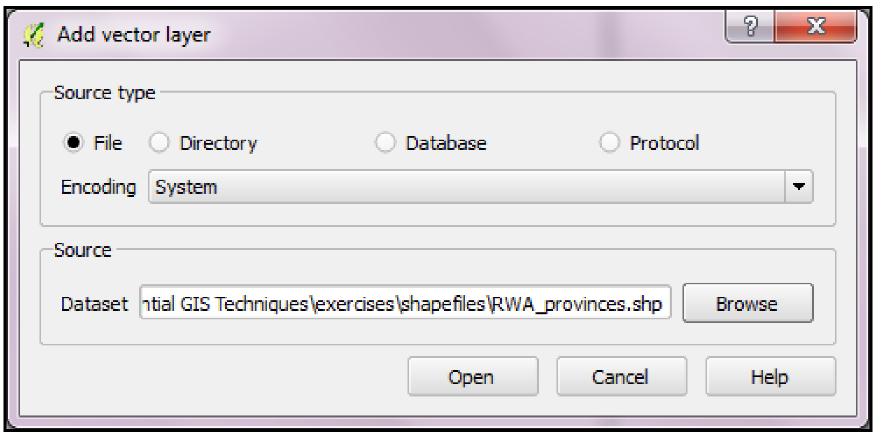
- In the Add vector layer file dialog box that appears (see below), make sure the source type is set to File and click on Browse.

- In the file dialog box that appears:

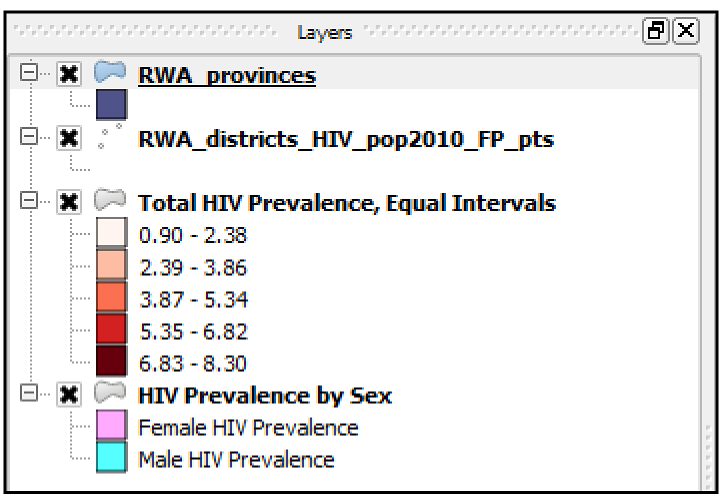
  - Use the pull-down menu in the lower-right corner to set the file type to "ESRI Shapefiles [OGR] (*.shp *.SHP)."

  - Navigate to the folder "\Exercises 4.1, 4.2 and 4.3\shapefiles" and find the shapefile **RWA_provinces.shp** and choose "Open".

- When the Add vector layer file dialog box reappears with the full path to the shapefile indicated in the Dataset field (see below), click once again on Open.

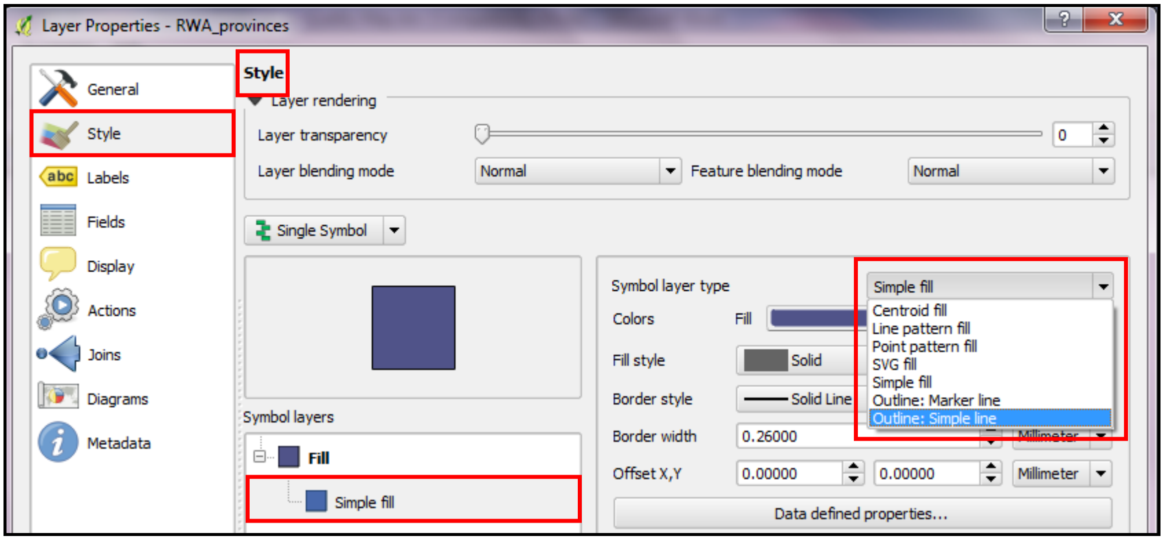


- After clicking on Open for the second time, you should see a new layer in the Layers window corresponding to the shapefile you just added.

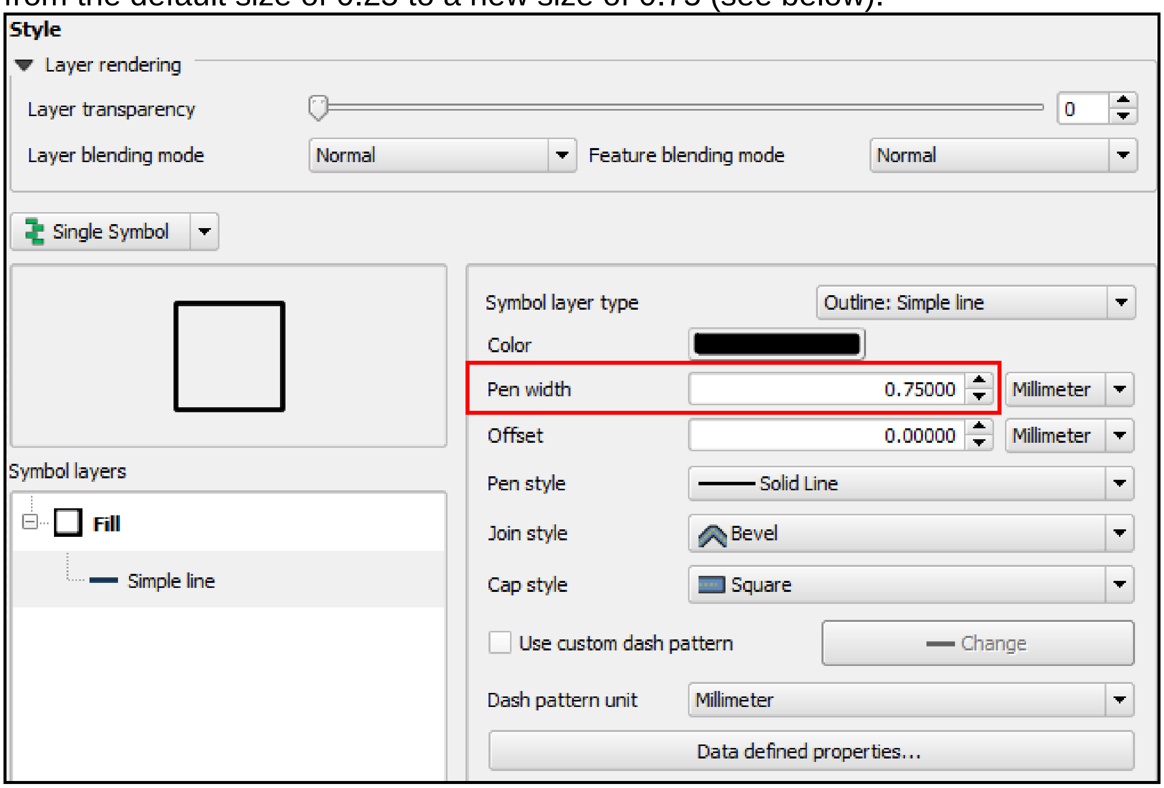


The new provinces layer will be displayed as a solid color on top of the other layers, so you will need to remove the color and just show the provincial boundaries.

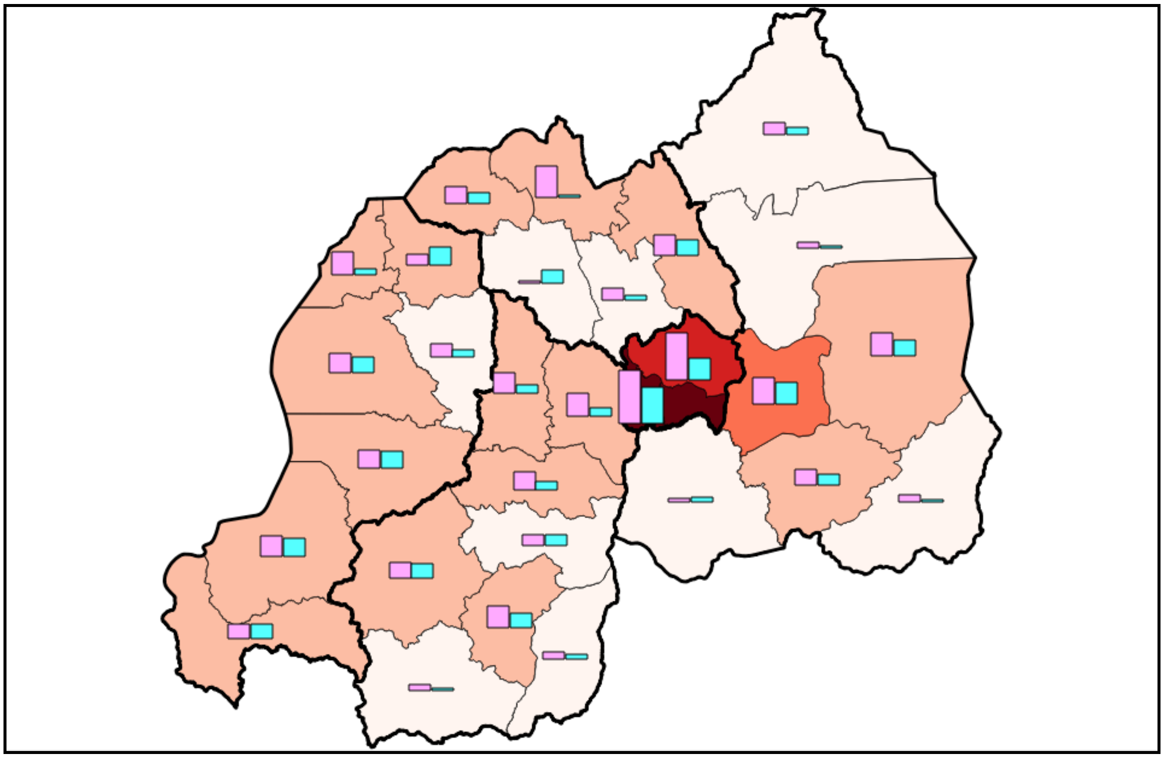
- To change the symbol color for provinces from solid to clear, double-click on the layer name **RWA_provinces** and select the **Style** panel. On the **Style** panel, highlight the symbol layer named **Simple fill** and in the pull-down menu for **Symbol layer type** select **Outline: Simple line** (see below).



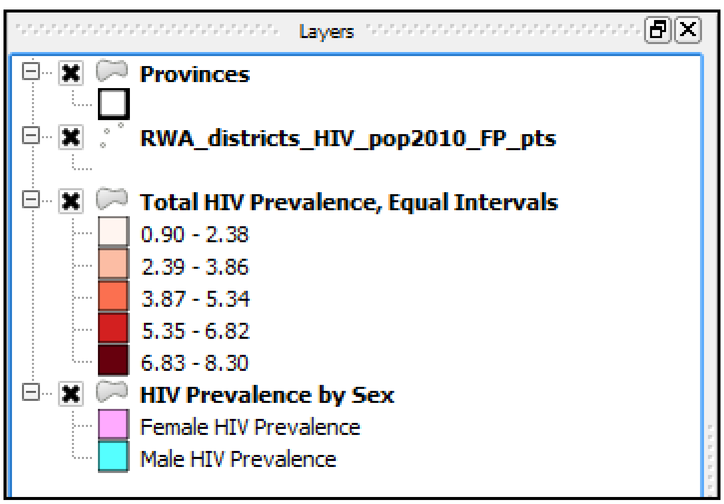
- To thicken the line used to display the province boundary, change the Pen width from the default size of 0.25 to a new size of 0.75 (see below).



- To finalize changes to the symbol used for province boundaries, click on Apply and OK. The map displayed in the QGIS map window should now look like the one below.



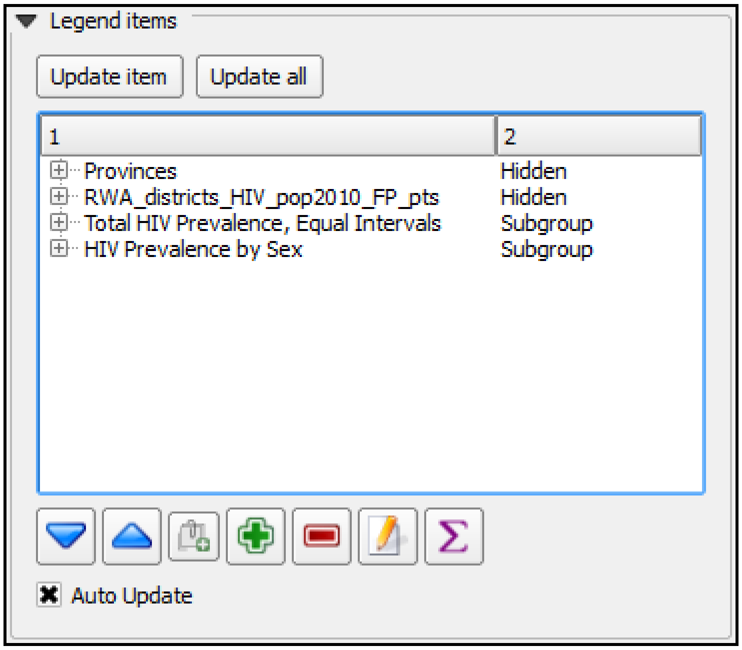
- To rename the province boundaries layer for the legend, highlight the layer name in the Layers panel and hit the F2 button the keyboard to go into editing mode. Rename the layer **Provinces** (see below).



- To avoid losing your changes, save the project before advancing.

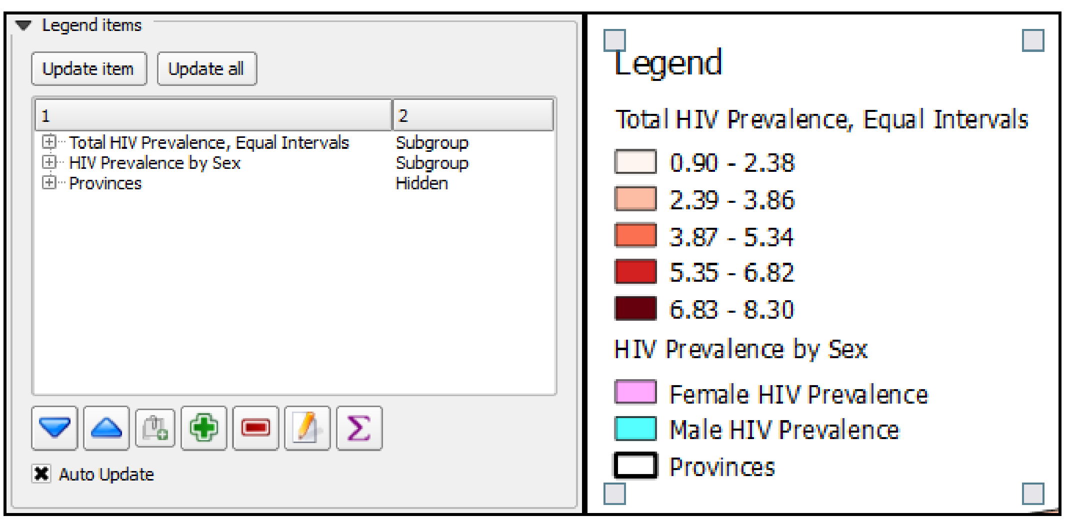
- Return to the map composer window and update the legend.

- Left-click on the legend to activate it and in the **Item properties** panel click on the button **Update all**. You should see a screen like the one below.

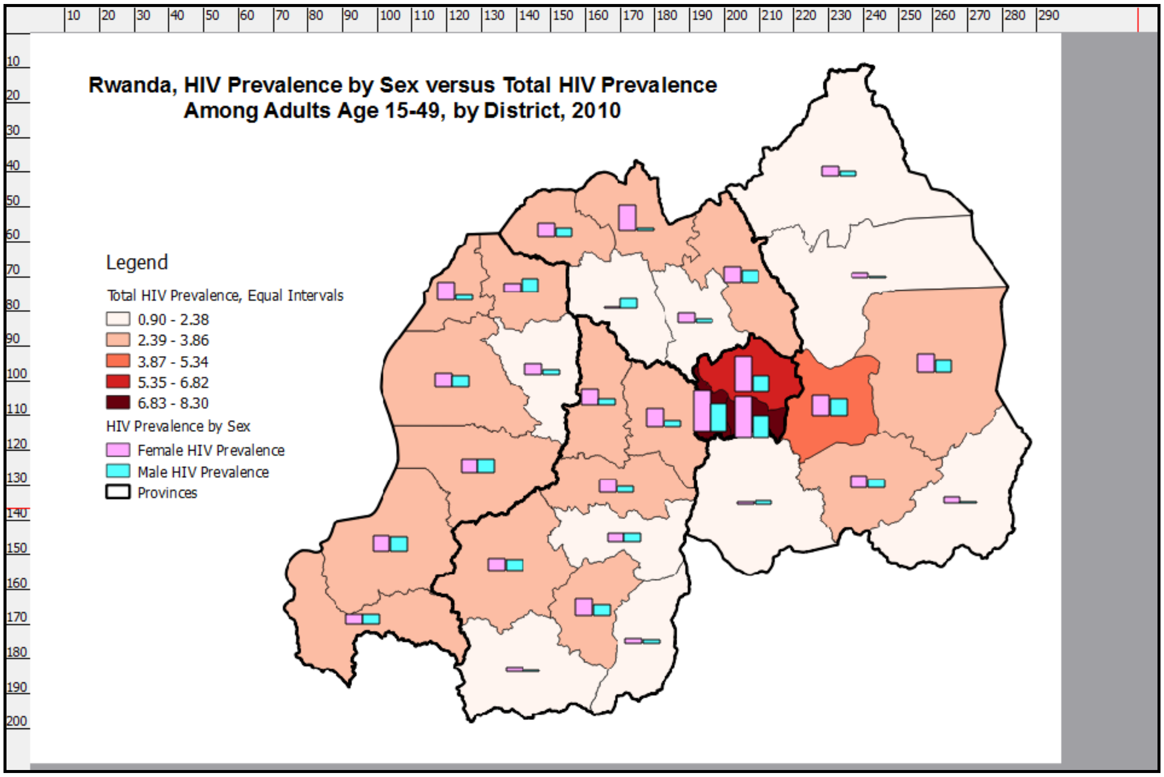


- Remove the layer **RWA_districts_HIV_pop2010_FP_pts** by highlighting it and clicking on the minus sign.

- Move the **Provinces** layer to the bottom of the legend by highlighting it and clicking on the down arrow. The legend items and legend should now look like the ones below.



- Although the legend is now updated, the province boundaries do not appear changed on the map canvas. To update the map canvas to reflect the changes to the province boundaries, click on the **Refresh view** button in the map composer window (see below).



- ☐ To avoid losing your changes, save the project.

- ☐ Now that the relatively time-consuming task of constructing a legend is complete, you should add a description of the data sources used to create the map.

NOTE: Constructing a clear, intelligible legend is a critical step in the map composition process. Legend graphics for proportional symbols and pie charts can be constructed in a similar manner to those created in this exercise for bar charts. The tricks are to use a point-based data layer, such as district-level center points, as the basis for creating the proportional symbol or pie chart diagrams, and to use a copy of that point-based data layer to construct the graphics needed for the legend.

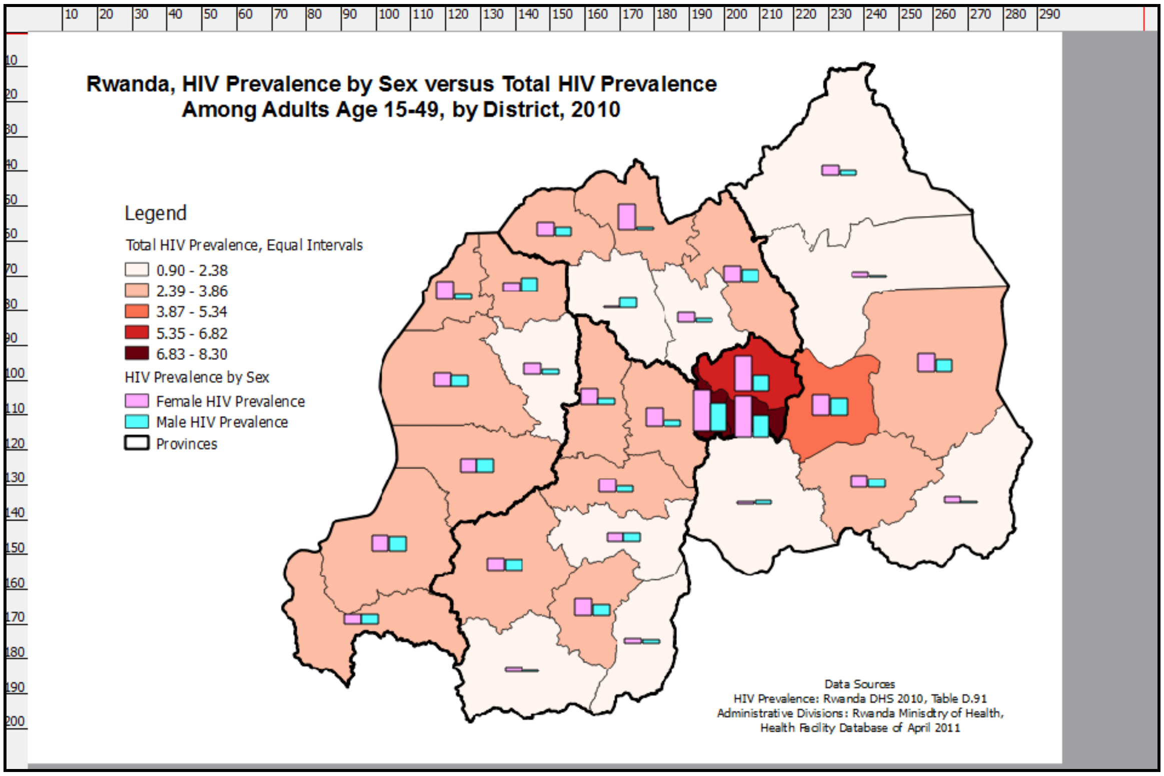
As mentioned previously, you can also create legends using external tools such as LegendSVG from the QGIS community in the Netherlands (see www.qgis.nl).

- ☐ You should also add a text box that describes the data sources used in the construction of the map. For this map, the data sources used were as follows:

  - HIV Prevalence: Rwanda DHS 2010, Table D.91

- Administrative Divisions: Rwanda Ministry of Health, Health Facility Database of April 2011

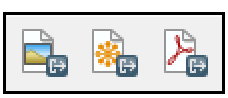
The final layout should look something like the one below.

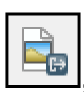


- Save the QGIS project. You are now ready to export the map composition to a publication-quality file format.

Step 4: Export the map for publication.

- To save your new map composition in a publication-ready format, there are three options: Export as image, Export as SVG (scalable vector graphics), and Export as PDF (see below).



Using the **Export as image** option , you have the choice of saving the map composition in three different formats: BMP, JPG, and PNG.

- BMP is a high-resolution, device-independent bitmap (raster) format.

- JPG is a file format designed to store compressed versions of photographs and other raster images. The compression methods employed can result in some loss of information with respect to the original image.

- PNG is also a file format for storing compressed versions of photographs and other raster graphics, but the compression method used for the PNG file format does not result in a loss of data with respect to the original images.

 Save the images using the path and file names specified below.

- BMP image: Session 4 Essential GIS Techniques\exercises\RWA_hivmap.bmp

- JPG image: Session 4 Essential GIS Techniques\exercises\RWA_hivmap.jpg

- PNG image: Session 4 Essential GIS Techniques\exercises\RWA_hivmap.png

Question: Look at the file sizes for the three maps. Which one is the largest and which is the smallest?

Answer: The BMP image is the largest. The JPG image is the smallest.

Question: Is there a significant difference in the clarity of the three images?

Answer: There is not a significant difference in the clarity of the images when viewed on a computer screen, which has a relatively low resolution. For printing purposes, however, the BMP and JPG images were saved by QGIS at 300 dpi (dots per inch), whereas the PNG image was saved at only 96 dpi, and would not be appropriate for printing.

Question: Which map would you send to a publisher and which map would you e-mail to a colleague?

Answer: You would send the BMP or JPG image to a publisher based on their higher print resolution (300 dpi). The BMP image would probably be too large to e-mail, although either the JPG or PNG file could easily be sent to a colleague for on-screen viewing and collaboration.

 Save the QGIS project and quit QGIS.

END

GIS Techniques for M&E of HIV_AIDS and Related Programs

Appendix 3: Key Concepts

Session 1:

• A GIS is a database linked to a map. It has 5 main components (hardware, software, geographically-referenced data, procedures, and people).

• The primary functions of a GIS are to capture, store, query, analyze, display, and output geographically-referenced data.

• A GIS can display data on a map, which has the advantage of helping the user quickly visually detect the distribution of data in a particular area.

•The information analyzed by a GIS can help with evidence-based decision-making concerning the targeting of interventions.

• A GIS can also help with data linking and program integration.

Session 2:

•In order to be mapped, data must contain geographic identifiers. Common geographic identifiers will also allow one data set to be linked to another.

• Metadata (data about the data—where and when and why it was collected) is important

• Geographic data can be represented in vector or raster format, and each of these formats has advantages for different uses. One common type of file used for vector data is the shapefile.

• Free and/or open source software mapping options are becoming increasingly functional and available. A popular choice (and the one used in this course) is QGIS.

Session 3:

• There are several primary sources of HIV/AIDS data which have a geographic component.

• SPA surveys provide information on services provided at various facilities.

• There are constraints to consider when working with DHS data in a GIS. They are generally representative only at the regional level.

Sample cluster locations are displaced to ensure respondent confidentiality.

• The DHS (Demographic and Health Survey) Program's Spatial Data Repository is a good source for this data.

• It is important to take into account methodological considerations before working with the GPS data information on cluster data, due to the data's random displacement and DHS sampling design.

SESSION 4:

• Data management and variable creation are key strengths of a GIS.

• Maps are useful for displaying multiple variables at once, which can help in both asking and answering questions about data.

• In designing a map, the author should consider appropriate data classification techniques. Color and pattern choices should be based on cartographic traditions and readability. A legend (description of symbols and colors used) and data source and date are also important.

GIS Techniques for M&E of HIV_AIDS and Related Programs

Appendix 4: References and Additional Resources

References:

Farquhar C et al. 2004. "Antenatal Couple Counseling Increases Uptake of Interventions to Prevent HIV-1 Transmission" (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3384734/). *Journal of Acquired Immune Deficiency Syndrome* 37(5):1620–1626.

The DHS Program (http://dhsprogram.com/), accessed in October 2013.

Office of the U.S. Global AIDS Coordinator (OGAC). January 2011. "PEPFAR Guidance on Integrating Prevention of Mother to Child Transmission of HIV, Maternal, Neonatal, and Child Health and Pediatric HIV Services". (http://www.pepfar.gov/documents/organization/158963.pdf) Washington, DC: The President's Emergency Plan for AIDS Relief (PEPFAR).

Additional Resources:

*The Center for Geographic Analysis (CGA) at Harvard University developed a series of **14 introductory QGIS videos**, ranging from 3 to 8 minutes in length.*

These videos were produced courtesy of Lex Berman, Center for Geographic Analysis, Harvard University.

The videos can be found at the Harvard CGA non-credit training page:
http://gis.harvard.edu/training/non-credit-training. For trainings on specific features and functionality of QGIS, see the videos below. You were introduced to some of these features and functions as part of the practical exercises of this course.

1) Introduction to the QGIS interface
2) Pan and Zoom controls
3) Handling layers
4) Identify and selection tools
5) Querying by attribute and advanced queries
6) Managing plugins
7) Importing delimited text (x, y coordinates)
8) Importing from Google Maps (KML)
9) Projections and Coordinate Reference Systems
10) Joining tables to shapefiles
11) Georeferencing a scanned image
12) Editing a new vector layer
13) Editing existing features in a vector layer
14) Symbolization

FROM ESRI LEARNING, HERE ARE 5 SHORT EXERCISES THAT CAN BE DONE IN ARCGIS ONLINE (FREE ACCESS—NO LOGIN REQUIRED):

ACTIVITY ONE: See Your World—Map Exploration

1. Go to http://www.arcgis.com and click "Map"

2. Grab, hold, and move the map to pan

3. Test out zooming in and out, using the map's zoom bar, the mouse's scroll-wheel, double-click, and "shift+clickdrag".

4. Zoom all the way out to the world, and all the way in to your home. Notice what happens when you zoom in and out.

5. Use the Search box in the top right to find the address of a friend, a relative, or a place, like your state's capital city.

6. Click the Basemap button and look at each of the different basemaps, from all the way out to all the way in. Notice what happens in each as you zoom in and out.

ACTIVITY TWO: Measure and Mark Your World—Distance Measurements

1. Zoom out to the world. Click "Measure" and choose "line". Measure (click to start, double-click to stop) roughly the distance from western USA to Europe, and western USA to central Asia, and northern Alaska to the southern tip of Africa and then the southern tip of South America. (Extra credit: What's a "great circle"?)

2. Change the measure tool from the ruler to the location tool (looks like a little globe). Click it, wander the map, then zoom in and find the coordinates for your home, Mt. Everest, and the White House. Close the Measure window.

3. Click "Modify Map." Click the "Add" button, choose "Add Map Notes," and use the "Map Notes" template by clicking "Create." Drop a pushpin on Mt. Everest and name it "Mt. Everest." Zoom back to the White House, drop a pin, and name it. This time, click "Change Symbol," and stroll thru the symbol choices, in "Basic" and the other pallets.

4. Zoom back home and use the "Add Features" pallet to add an "Area" for the school grounds, and a line for your route to and from school. Click "Details" (top left) when done creating data.

ACTIVITY THREE: Explore Your World—Searching for Data

1. At the top of the map, click "New Map", and choose "Open". This will give you a new clean map space, with the Topo basemap, and no additional layers.

2. Choose "Add/ Search for Layers." In the "Find" window, type "population". Near the top, you should find "USA Population Density". Click on the name, see a quick thumbnail, and click on "Add to map." Then, at the bottom of the left-hand window, click "Done Adding Layers."

3. Zoom/ pan so you can see all 50 states as states, and then zoom in to your home, one click at a time on the map's zoom bar. As you zoom in, click on the state, and then the county, and then Census Tract, and finally Block Group, and read the text that shows up with each zoom.

4. It sure would be nice to know what the colors mean! At the top of the left-hand window, click the "Legend" button, and see what the colors mean, and see if the colors and meanings change as you zoom in and out.

5. Pan around your state and see if every place looks alike. Wouldn't it help to see thru the population layer to the landmarks below? At the top of the left-hand window, click the "Content" button. Hover the mouse over the name "USA Population Density" until you see a little pull-down menu icon at the right end of the name. Click the pull-down, click "Transparency", and play with the little slider.

ACTIVITY FOUR: Expand Your World—Open a saved map

1. Use a shortcut URL to go directly: http://esriurl.com/recentquakes . Explore briefly, turning the layers' checkmarks off and on.

2. Use a long URL to go directly: http://www.arcgis.com/home/webmap/viewer.html?webmap=79151205f3124c13bc814fda3170e901 Try turning on the "old map" layers.

3. Use ArcGIS Online search. Go to http://www.arcgis.com , then click "Gallery" at the top. In the search box at the top-right of the page (NOT your browser bar's search box), type "USA demographics for schools" and click on the topmost thumbnail. The map should open with 10 layers in it. Zoom to your location, turn off population density, and try the other layers. (Extra credit: If more than one is checked, which one is visible in the map?)

ACTIVITY FIVE: App the World—Try out a focused app

1. Terrain profile = http://esriurl.com/elevation
2. GPX Demo App = http://esriurl.com/gpxdemoapp
3. StoryMaps visit to DC's National Mall = http://storymaps.esri.com/stories/malltour/

Now, what is something about which you would like to make a map?

For more about using ArcGIS Online in education, see http://esriurl.com/mappingwithago .

MEASURE Evaluation
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill NC 27517 USA
P: +1 919-445-9350
F: +1 919-445-9353
E: measure@unc.edu
www.measureevaluation.org

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) through Cooperative Agreement GHA-A-00-08-00003-00 and is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with ICF International, John Snow, Inc., Management Sciences for Health, Palladium, and Tulane University. The views expressed in this publication do not necessarily reflect the views of USAID or the United States government. MS-15-106

eCourse available online:
http://www.globalhealthlearning.org/course/gis-techniques-m-e-hiv-aids-and-related-programs

eCourse available for download:
http://www.measureevaluation.org/resources/publications/ms-15-106

USAID FROM THE AMERICAN PEOPLE

PEPFAR

K4Health Knowledge for Health

MEASURE Evaluation