USAID
FROM THE AMERICAN PEOPLE

The Demographic and Health Surveys Program

Health Data Mapping Online Workshop Participant's Guide

September 16, 2022

Health Data Mapping Online Workshop
2022 Participant's Guide

The Demographic and Health Surveys (DHS) Program
ICF
Rockville, Maryland, USA

September 16, 2022

The DHS Program is funded by the U.S. Agency for International Development (USAID). The information provided in this guide is not official U.S. government information and does not necessarily represent the views of USAID or the U.S. government.

CONTENTS

1.INTRODUCTION TO THE GUIDE

1.1 PURPOSE

This guide is meant to summarize the content covered in The Demographic and Health Surveys (DHS) Program's annual Health Data Mapping Virtual Workshop. Specifically, this version of the guide corresponds to the 2022 implementation. Participants in this workshop are early- to mid-career professionals who work in the fields of Public Health or Demography, preferably for National Ministries of Health or Bureaus of Statistics. They also live and work in countries where DHS Program surveys are conducted. The workshop is conducted in a semi-synchronous format, meaning that participants complete some work independently but meet several times with the facilitators. Each week, participants are expected to spend about two to four hours a week learning new skills and completing assignments. They engage with facilitators and each other on The DHS Program's Learning Hub (learning.dhsprogram.com) and meet for live virtual sessions four times during the course.

This guide covers the content that participants complete on their own. It includes both the activities and conceptual underpinnings. However, it is not meant to replace the workshop itself. Some of the content covered during the workshop's synchronous sections is essential to understanding how to map demographic and health data. While this guide was not designed as a standalone course or workbook, some professionals who have not participated in the Health Data Mapping Workshop will find it useful.

1.2 OVERVIEW

By the end of the Health Data Mapping Workshop, participants should be able to (1) create maps for decision-making and planning programs, (2) present and explain maps for policymaking and decision-making, and (3) develop a plan for future application and transference of knowledge to colleagues. To accomplish these goals, the workshop and this guide primarily use two software packages: QGIS 3.22.4 and Microsoft Excel.

This guide is designed to be used alongside the exercise files distributed with it. The guide is broken into 10 sections. Section 1 and 10 cover introductory and additional information respectively. So, sections 2 through 9 make up the core content. These eight sections correspond to the core eight modules of the Health Data Mapping Workshop. These are described below.

1. **Introduction to the Guide** provides an overview of the Health Data Mapping Virtual Workshop and this guide.

2. **Overview of Maps and GIS** describes how to define a geographic information system (GIS), identify applications for GIS in public health, and explain the difference between vector data sources (points, lines, and polygons) and raster data sources.

3. **Introduction to QGIS** covers how to install QGIS, add panels/toolbars/plugins to the application, add vector data (points, lines, and polygons) to a map, and use navigation tools and identification tools to explore a layer's geography.

4. **Cartography Basics** covers how to identify the key components of a map, use print layouts in QGIS, and create and export basic maps from QGIS.

5. **Queries and Selections** covers how to query data in QGIS by adding filters to a layer's attributes, select data from a layer based upon their geographic location compared to other data, and summarize the number of points within a polygon.

6. **Cleaning GPS Data in Excel** introduces how to clean GPS data using proper formatting and tools in Excel, use crosswalk tables in Excel to examine data, create pivot tables in Excel, and add GPS data into QGIS and display its coordinates.

7. **Joins** covers how to join data by attributes, join data by their spatial location, and calculate the sum of points within a polygon.

8. **Labeling Map Features** introduces how to create labels for features in a QGIS map, change the style of map labels in QGIS, manipulate the positioning of labels in QGIS, and create a label key in QGIS.

9. **Advanced Map Layouts** covers how to create two advanced map layouts in QGIS: inset maps and extent indicators.

10. **Additional Information** contains a list of geospatial data resources and a glossary of all terms used in this guide.

2. OVERVIEW OF MAPS AND GIS

2.1 CONCEPTS

To begin, we will explore how maps and location data are used in the field of public health. These ideas will serve as a foundation upon which the other skills will be built. By the end of this first section, you will be able to:

1. Define a geographic information system (GIS),
2. Identify applications for GIS in public health, and
3. Explain the difference between vector data sources (points, lines, and polygons) and raster data sources.

2.1.1 Definitions

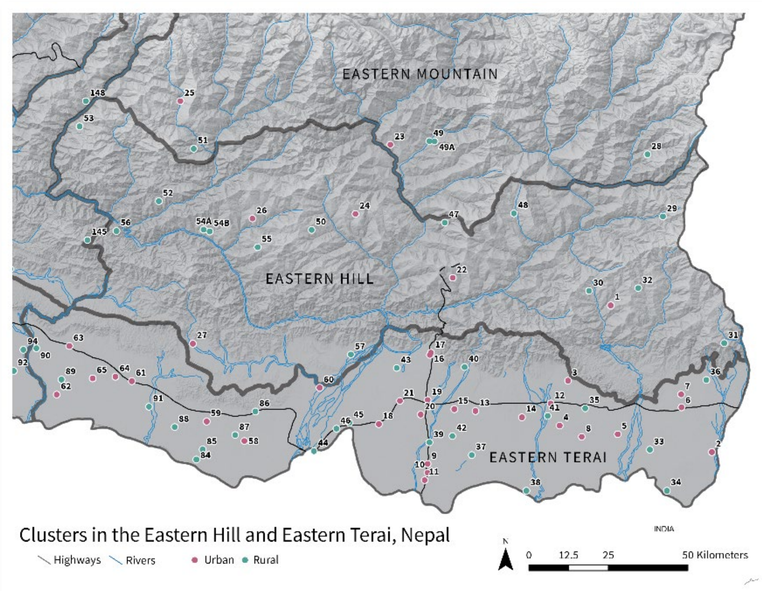
The following terms are used in this section.

- **Geographic information systems (GIS)**: software used to organize, display, and analyze different geographic data
- **Geographic/geospatial data**: any data that contains information about a location on Earth's surface (examples of location information include X/Y coordinates or geographic names)
- **Vector data**: geographic data in the form of points, lines, or polygons that represent places or phenomena on Earth's surface
- **Raster data**: geographic data in the form of a collection of pixels where each pixel represents a square piece of the Earth's surface or atmosphere
- **QGIS (formerly Quantum GIS)**: the open source/free GIS software used in this guide

2.1.2 Value of Maps

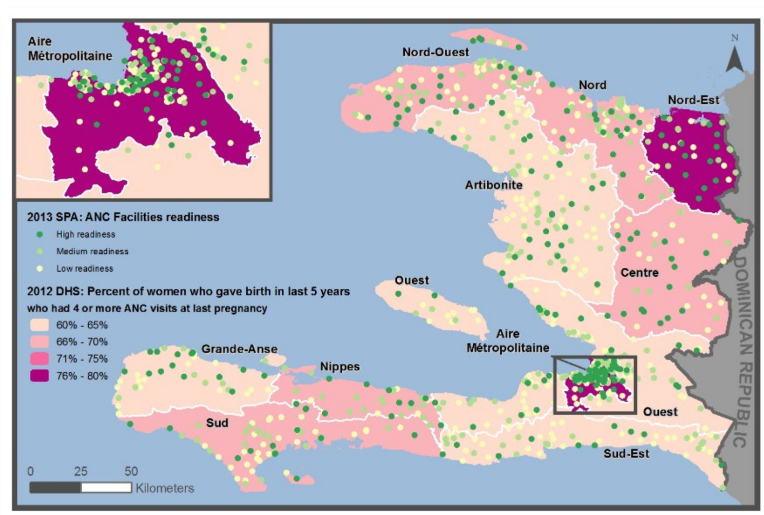
Maps can do much more than to simply help someone find a location. Maps are tools used to visually illustrate relationships, comparisons, and pattern representations of data. They can be used to tell stories and inform decision-making. The list below includes examples of how maps are used in the field of Public Health.

- **Context**: Maps can add context to presentations and reports. There are many circumstances in which a good map can help tell a story about a given place. For example, maps can highlight the bounds of administrative units and help data users identify where different health outcomes occurred in the country.
- **Environmental Health**: Maps help us to understand how geography, such as environment, population density, water, and location of services, can affect health. For example, Maps can show the relationship between communities and the natural environment. In particular, the map below highlights the locations of rivers and depicts the elevation of the terrain to visualize the impact of geography on individual communities' health outcomes.

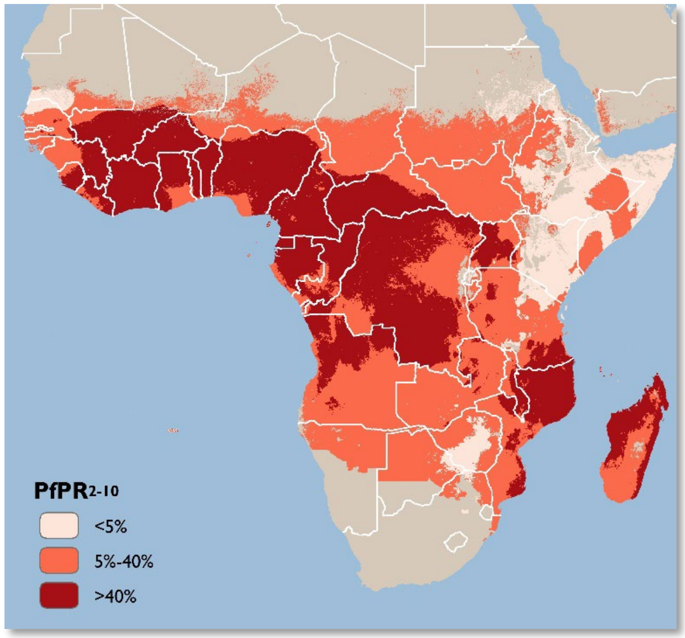


- **Demography**: Maps also help to highlight group behaviors and demographic characteristics. Patterns in human geography and demographics are important to take into account when analyzing public health outcomes. Thematic maps allow us to compare poverty levels. When paired with health information, these maps could help inform policy decisions regarding program implementation.

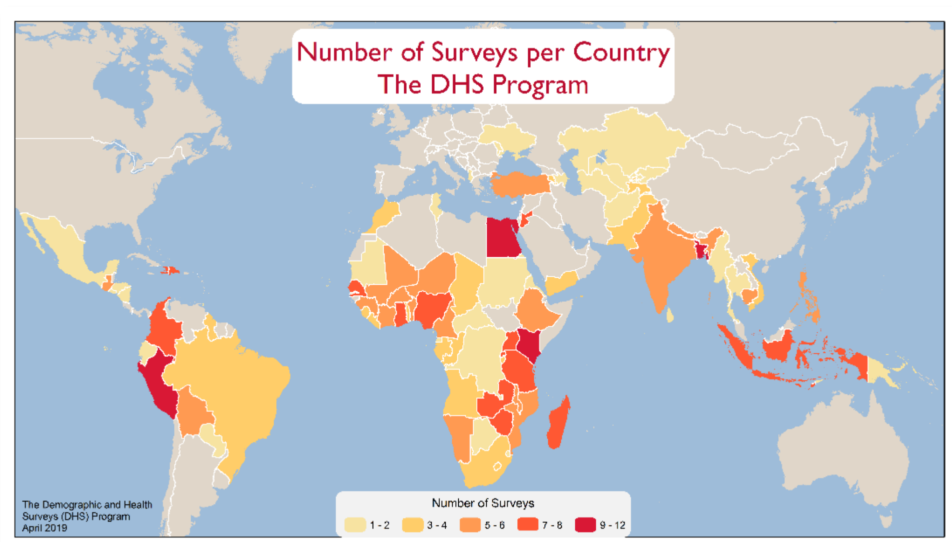
- **Health Infrastructure**: Maps can help to show the distribution of resources and services. Infrastructure is essential to the distribution of health care and services. Maps can reveal where additional infrastructure is needed, and which facilities may be performing well or underperforming. The map below shows us how the readiness of health facilities relates to the percentage of women who had four or more antenatal care visits.



- **Epidemiology**: Maps are also good at showing the spatial spread of a given disease. Epidemiological studies often include a spatial component, and maps can bring that data to life. The map below shows the distribution of malaria based on the prevalence rate among children. A map can help teams of professionals, researchers, and international donors make decisions about where to target and prioritize programming., and how to stop the disease from spreading further.



- **Planning**: Maps can show where activities take place and where there is a need for future activities. The map below shows where The DHS Program has worked over the years, as well as the type of survey it has done. Maps like this can help an organization evaluate the work it has done so far and plan for the future.



2.2 INSTRUCTIONS

The primary activity for this chapter is to explore DHS Program indicators on STATcompiler, our online data visualization tool. It comes from Activity 2.2 in the Health Data Mapping Workshop.

1. Go to STATcompiler (statcompiler.com).
2. Select ‘Choose Country’ and choose your country.
3. Select an indicator that interests you.
4. Click the Map icon above the data.
5. Click Subnational.
6. Explore the subnational indicator values.

3. INTRODUCTION TO QGIS

3.1 CONCEPTS

In the following seven sections of this guide, you will be introduced to many tools and functions used in GIS software packages. Here, we will focus specifically on QGIS, a cross-platform, free, and open-source desktop GIS application. With this software, you can view, edit, analyze, and make maps with geospatial data. This section covers how to:

1. Install QGIS.
2. Add panels, toolbars, and plugins.
3. Add vector data (points, lines, and polygons) to a map in QGIS.
4. Use navigation tools and identification tools to explore a layer's geography.

3.1.1 QGIS Definitions

The following terms are used in this section.

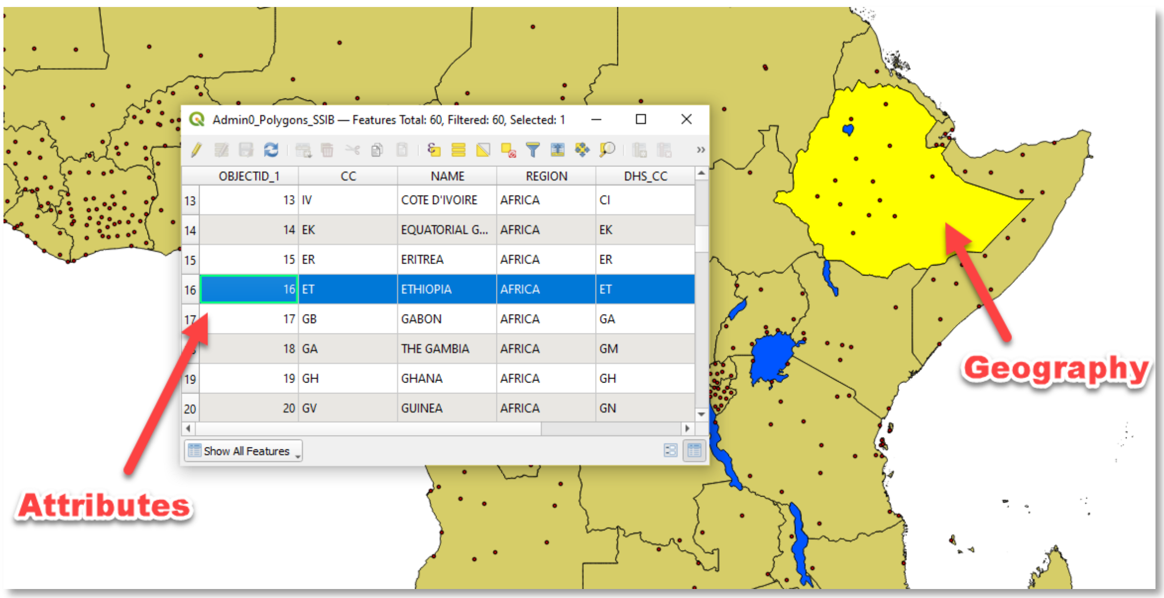
- **Attributes (of a GIS layer)**: the variables and values for each feature in a GIS layer
- **Attribute table**: a tool used in QGIS to explore a layer's attributes
- **Feature**: individual data points in a GIS layer are called features (e.g., in a file of GPS locations, each location is considered an individual feature; in GIS layer that contains administrative boundary, each individual boundary polygon is considered the feature)
- **Geography (of a GIS layer)**: the graphic component of a GIS layer that depicts the features' physical location on Earth's surface
- **Layer**: dataset displayed in a GIS are called layers, because the user visually stacks these datasets on top of each other to create maps and conduct analyses
- **Panels (QGIS)**: the QGIS user interface includes small windows called 'panels' that allow the user to change content in the project and conduct analyses; by default, panels appear on the left- or right-hand side of the QGIS map area; to open panels in QGIS, go to View → Panels (e.g., the Layers Panel is where you see your map layers listed and the Layers Styling Panel is where you change the symbology of those layers)
- **Shapefile**: vector datasets that can be used in GIS software; they are comprised of up to 11 separate files, but only require three: *.shp, *.shx, and *.dbf
- **Symbology**: how the geography of a GIS layer is stylized (e.g., the colors, shape, patterns, and transparency used to depict different features)
- **Toolbars (QGIS)**: QGIS tools are organized into various toolbars, each with icons representing the different tools; by default, the toolbars are docked above the QGIS map area; to turn on

toolbars, go to View → Toolbars (e.g., the Map Navigation Toolbar contains the tools used to zoom and pan around the map)

- **Vector data**: geographic data in the form of points, lines, or polygons that represent places or phenomena on Earth's surface

3.1.2 Geography and Attributes in GIS Files

The concepts of a map layer's 'geography' and 'attributes' are introduced above but require additional attention. It's important to understand these two components of a GIS file. The graphic below shows the difference between a single GIS layer's geography and attributes. Note that Ethiopia's geography (highlighted in yellow) corresponds to the highlighted row in the attribute table.



3.2 QGIS INSTRUCTIONS

3.2.1 Installing QGIS

Since QGIS is open-source software, changes are continually being made to the code. This course was created using an earlier version of QGIS. Icons and menus may appear different in other versions. Version 3.22.4, code named *Białowieża*, is the latest long-term release of the software as of January 2022. The newest version of QGIS can be found at this URL: qgis.org/en/site/forusers/download.html.

This section covers the installation instructions found in Activity 3.1 of the Health Data Mapping Workshop. Follow the steps below to download and install QGIS.

1. Download the installation file for 3.22.4.

    a. PC users: qgis.org/downloads/QGIS-OSGeo4W-3.22.4-1.msi

    b. MAC users: qgis.org/downloads/macos/qgis-macos-ltr.dmg

2. Once the download is complete, save the *.exe file to your computer, right-click on it and select 'Run as Administrator'.

3. When the QGIS *Białowieża* (3.22.4) Setup Wizard window opens, click Next and then, 'I Agree,' to agree to the license agreement. Select a destination folder in which to install QGIS, or click Next to install it in the default location. Click Install and wait while the program installs. Do not select any of the datasets on the 'Choose Components' screen. Click Finish when the installation is complete.

4. If you have a shortcut for QGIS on your desktop (or within a folder named QGIS 3.22.4), double click it to start QGIS. Otherwise, click Start → All Programs → QGIS 3.22.4 → QGIS Desktop 3.22.4.

5. Opening QGIS Desktop for the first time will create all of the folders necessary for installing plugins.

3.2.2 Exploring QGIS

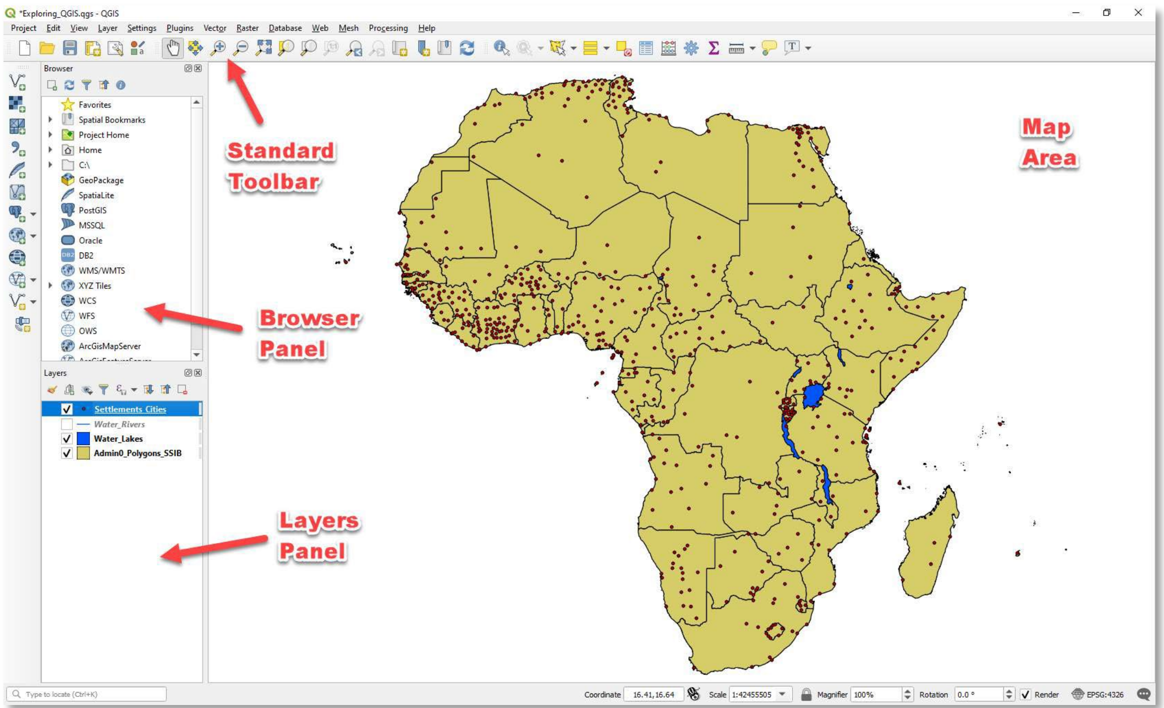
This section will introduce you to QGIS and some of the basic functions of the software. Whenever you work in QGIS Desktop, you are working with a map document. This document can contain various layers, which are populated by geographic/geospatial datasets. A QGIS map document has a *.qgz file extension. A map document was created for this exercise. Follow the steps below to open the exercise file for this section.

The content below corresponds to Activity 3.2 of the Health Data Mapping Workshop. You will use the exercise files found in the \DHS_QGIS_2022\Exercise_Files\3.2.2_ExploringQGIS\ folder.

1. If you have a shortcut for QGIS on your desktop, double click it to start QGIS. Otherwise, click Start → All Programs → QGIS 3.22.4 → QGIS Desktop 3.22.4.

2. Click Project → Open. Navigate to the \DHS_QGIS_2022\Exercise_Files\3.2.2_ExploringQGIS\ folder. Double-click Exploring_QGIS.qgz to open and display the map.

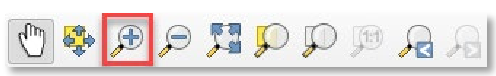
When the Exploring_QGIS.qgz opens, you will see an image depicting Africa with the location of various countries. It should look similar to the image below.



3. Click View → Panels in the menu. Uncheck the box next to the Layers Panel. What does your QGIS window look like now? You'll see that the Layers Panel on the lefthand side of the screen disappears. Turn the Layers Panel and Browser Panel on by checking the boxes next to their names.

4. Next, click View → Toolbars and explore some of the toolbars available.

5. On the Standard Toolbar, click the Zoom In button (highlighted in red below). By clicking the button, the Zoom In tool has now been activated.



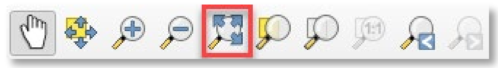
6. Position the magnifying glass pointer to the upper left of a country of your choice. Click and hold the left mouse button while dragging your mouse to the bottom right of the country of your choice, making a square around the chosen country. (The selection you make will be filled in with a light gray color.) Release the mouse to zoom into the selected area.

7. To return to the previous view click the Zoom Last button (highlighted below). This will take you back to the previous zoom level that shows all of Africa.

8. Again, use the Zoom In tool to zoom in on Southern Africa. You should now see Southern Africa in the center of the map. Now, click the Pan Map (highlighted below) button on the standard toolbar. Move the cursor hand to the top of your map area. Click and hold the left mouse button while moving the map so that the top half of Africa is centered on your map. Release the mouse

button. You will see that the map has been repositioned. You should now see Northern Africa in the center of the map.

9. To return to the whole view of Africa, click the Zoom Full button (highlighted below).



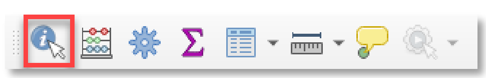
Now that you explored how to zoom in and out, move, and navigate around your map, you will now learn to turn a layer on or off using the Layers Panel. The Layers Panel in QGIS shows you the layers of data that are available to be shown on your map. Think of layers as pieces of data that can be added one on top of the other and displayed together.

10. In the Layers Panel, uncheck the Settlements_Cities check box. The red dots on the map that show world cities will disappear. Then, check the Water_Rivers check box. You will see blue lines on your map that show rivers.

11. Check the Settlements_Cities check box so that the cities are shown again. Click the Settlements_Cities name so that it is highlighted in blue. Click and hold the left mouse button while moving that layer to the bottom of the Layers Panel below the other layers. Now, the order of the layers from top to bottom is: Water_Rivers, Water_Lakes, Admin0_Polygons_SSIB, Settlements_Cities. As you can see, most of the dots disappear except for those on small islands and those next to the ocean.

12. Let's return the settlements layer to the top of our Layer Panel. Click, hold, and drag the Settlements_Cities name to move that layer above the others.

We'll now explore the Identify Features tool. To access this tool (highlighted below), you need to have the Attributes Toolbar activated. This tool will always provide you with information from whichever layer is highlighted in the Layers Panel. Therefore, you must check to make sure that the layer of interest is highlighted.



13. Click on the Water_Lakes layer in the Layers Panel to highlight it.

14. On the toolbar click the Identify Features button (highlighted below).

15. Click on a lake on your map. A dialog box will open that lists all the attributes for that region.

16. Click on the Cities_Settlement layer in the Layers Panel and use the Identify Features tool to identify one of the cities on the map.

Now, that you've explored the basic components of QGIS, you can close the project without saving.

3.2.3 Creating a New QGIS Project

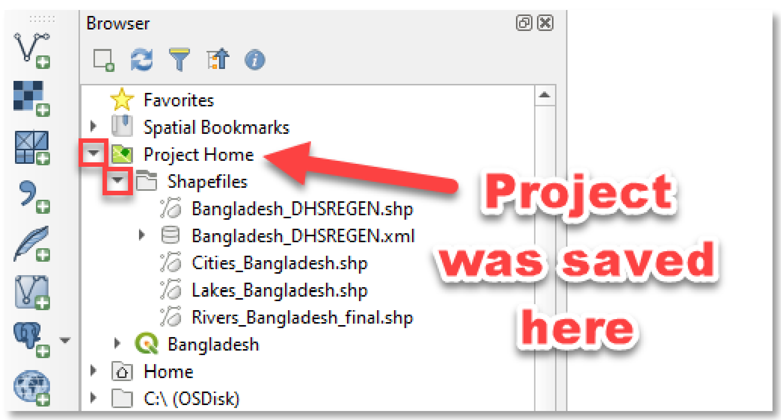
This exercise will introduce you to how to add data to a QGIS project. Vector data is a type of spatial data that is stored as a point made of two coordinates. GPS points are a common type of vector data. Vector data can also be lines and polygons, which are made up of a series of connected points.

This section corresponds to Activity 3.5 in the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\3.2.3_CreateQGISProject\ folder. However, for this exercise, the folder does not include a QGIS project. You'll have to create your own by following the steps below.

1. Start QGIS Desktop and choose the 'New Empty Project' from the Project Templates.

2. Let's save our project next in the \3.2.3_CreateQGISProject\ folder. By saving the project next to the \Shapefiles\ folder, we'll be able to easily access those shapefiles from the project. So, click Project → Save As. Navigate the to the \DHS_QGIS_2022\Exercise_Files\3.2.3_ CreateQGISProject\ folder and save the project as 'Bangladesh.qgz'.

3. The folder where you saved your QGIS project is called the Project Home. By expanding the Project Home folder in the Browser Panel, you can access the other files stored in that folder. To add shapefiles to this project, drag them from your \Shapefiles\ folder into the QGIS map area. Add all the shapefiles (Bangladesh_DHSREGEN.shp, Cities_Bangladesh.shp, Lakes_Bangladesh.shp, and Rivers_Bangladesh_final.shp) to your map.



It's best practice to save all your project files in your Project Home, but you can also add layers to your project from other locations on your computer. To add a shapefile stored somewhere else, go to Layer → Add Layer → Add Vector Layer and then browse your computer for the shapefile to add.

4. Now that we've added the shapefiles to our project, save the project.

5. When a layer is added, QGIS assigns it a layer name. Sometimes this name is not very descriptive because it inherits the dataset or the feature name. To rename a layer, right-click on the layer in the Layers Panel and select Rename Layer. Go ahead and rename the layers in your project. Also, be sure to order your layers in the order they're listed below.

    a. Cities_Bangladesh.shp → 'Cities'

    b. Lakes_Bangladesh.shp → 'Lakes'

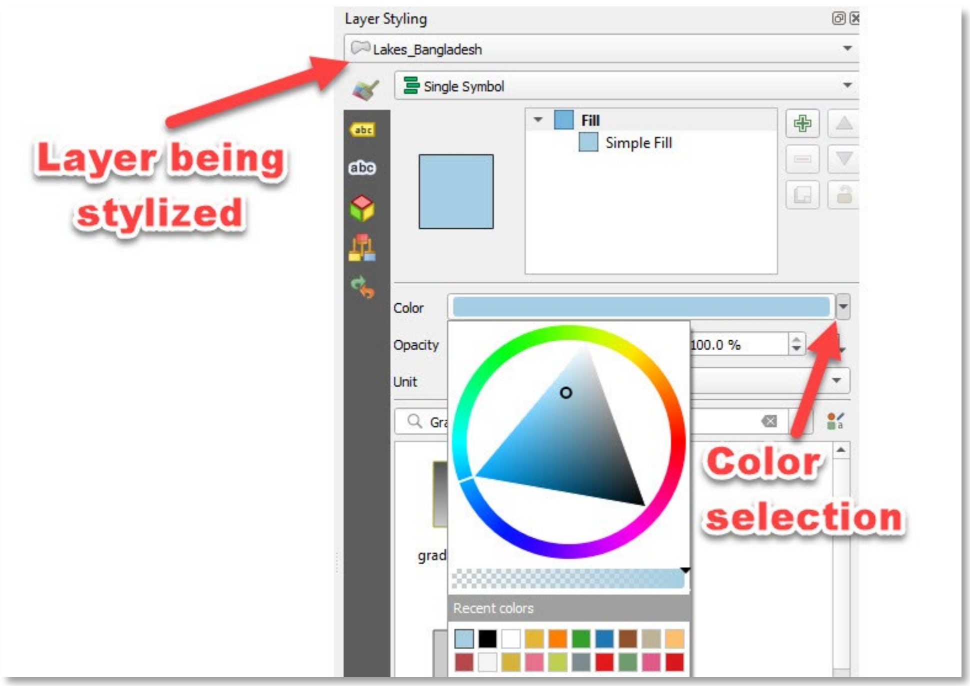
    c. Rivers_Bangladesh_final.shp → 'Rivers'

d. Bangladesh_DHSREGEN.shp → 'Regions'

We'll now change the symbology of our layers. We'll make the rivers and lakes blue, the cities red, and the regions light brown or gray.

6. To change the symbology of a layer in QGIS, we need to make sure that the Layer Styling Panel is activated. To open the Layer Styling Panel, click on View → Panels and check the box next to Layers Styling Panel.

7. In the Layers Styling Panel, select the layer you want to stylize from the first drop-down box (see the screenshot below).



8. Then, change the color of that feature. Make the rivers and lakes blue, the cities red, and the regions light gray or brown.

Now, let's explore the attribute table for one of our layers. Remember, the attribute table is the tool in QGIS uses to organize and display the variables and values for GIS features in our project.

9. In the Layers Panel, right-click on the Regions layer and select Open Attribute Table.

10. Scroll through the attribute table to view the data associated with each feature within the shapefile. Consider the different pieces of information contained in the attribute table. Also, think about other information you would like to be included in such a table.

11. Save and close your project.

4. CARTOGRAPHY BASICS

4.1 CONCEPTS

Before making maps with QGIS, it is important to understand basic map design. In this section, we will identify the components of a map and introduce best design practices. After completing this section, you will be able to:

1. Identify the key components of a map,
2. Use print layouts in QGIS, and
3. Create and export basic maps from QGIS.

4.1.1 Definitions

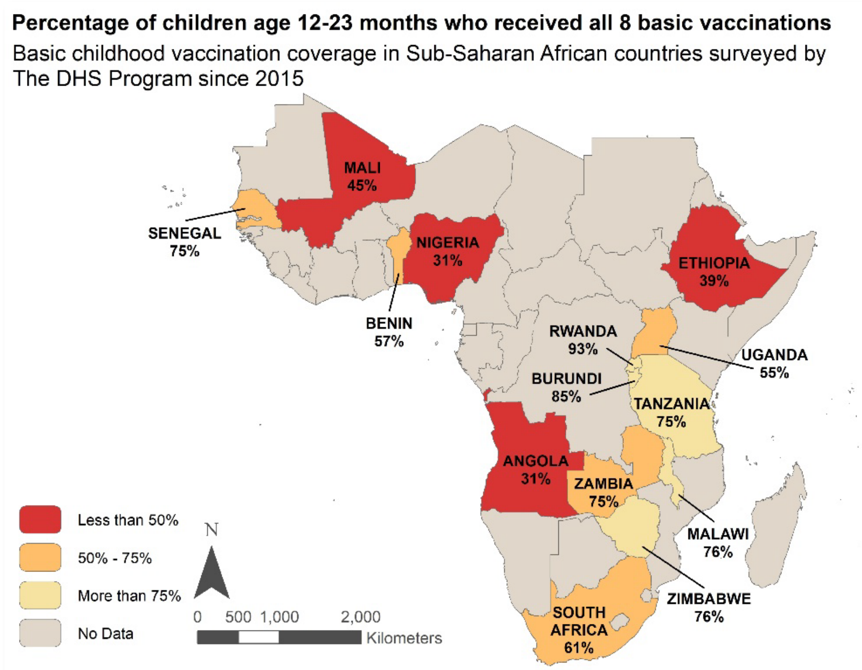
The following terms will be used in this section.

- **Balance (map design)**: the positioning of elements on a map layout so that they are spread evenly across the page and white space is limited
- **Binary data**: data consisting of two possible values, often 1 and 0 or TRUE and FALSE
- **Cartography**: the science and art of drawing maps
- **Categorical data**: qualitative or quantitative data divided into groups (e.g., countries broken into income groups or communities categorized as urban or rural)
- **Choropleth map**: a thematic map that uses colors or shading within areas to indicate the value of an indicator in those areas (e.g., a choropleth map of stunting in a country may use increasingly darker shades of red to indicate a higher prevalence of stunting in some administrative areas compare to others)
- **Continuous data**: quantitative data with a theoretically infinite number of possible values between a given high and low value (e.g., elevation data or precipitation data where each data point is equal to any value between a given maximum and minimum value)
- **Coordinate systems**: reference frameworks with a set of rules used to determine where places are located on or near Earth's surface
- **Equator**: an imaginary line on Earth's surface that is equal distance from the North and South poles and divides the planet into the Northern and Southern hemispheres
- **Figure-ground organization (map design)**: the use of contrasting colors to differentiate map elements in the foreground from those in the background (i.e., the use of contrasting colors to focus the attention of the map user)
- **Geoid**: the spherical shape of Earth

- **GPS**: the Global Positioning System, developed and maintained by the United States government, uses a satellite constellation for navigation on Earth's surface
- **Latitude**: measurement north or south of Earth's equator
- **Legend**: a visual guide of the symbology used in a map (e.g., if a capital city is represented with a star and all other cities are represented with a circle, the legend will show a star and circle labeled according to what they represent)
- **Longitude**: measurement east or west of the prime meridian
- **Map elements**: different components that together comprise the map layout (e.g., the map frame, legend, title, scalebar, etc.)
- **Map layout**: what is depicted on a virtual or printed map page
- **Prime meridian**: the imaginary line chosen to represent 0° longitude that passes through Gao in Mali and Lake Volta in Ghana
- **Print layout (QGIS)**: the user interface in QGIS used to create map layouts
- **Projections**: the geometric transformations used to flatten the spherical Earth so that it can be depicted in a two-dimensional map or GIS
- **Quantitative data**: numerical data
- **Qualitative data**: non-numerical data
- **Reference maps**: map designed to show where places or other geographic features are located in relation to each other (e.g., a city map used by tourists or a topographic map used by geologists)
- **Scale**: the relationship between the dimensions of our map and the actual world
- **Scalebar**: a graphical representation of a map's scale
- **Symbology**: the use of symbols, shapes, and colors to represent physical places and geographic phenomena in a map
- **Thematic maps**: maps designed to convey information about a single topic or theme (e.g., percent of households with electricity or population density)
- **Visual hierarchy (map design)**: the prominent placement of items in the map layout to emphasize their importance
- **WGS 1984**: the coordinate system used by the GPS and commonly used in cartography and GIS application

4.1.2 Introduction to Cartography

Visual design can make or break your map's message. With map visual design in mind, you can direct your audience's eye to the core message on your map. By labeling and coloring only the relevant countries on this map, the audience focuses on the story you want to tell. Consider the map below. When looking at the map, where are your eyes drawn? Many people will first look at the countries colored red with the bold, black labels. The dark colors and contrast will draw their attention. Next, their eyes may shift to the title which is bold and at the top of the page.



Cartography involves several elements. When creating a map, it is essential to keep the following concepts in mind: layout, type, medium, design, and scale. Every map should have the following five key elements:

- **Map frame**: The map frame is the most important part of the map. It covers the main content of the map and all the symbols and elements that communicate the map's meaning to the map reader. As we will see later in **Subsection 4.1.4**, a map can have more than one map body.

- **Descriptive title**: The title should succinctly express the information your map is reflecting. The map's title needs to define the 'what', 'where', and 'when' of the map. The 'what' should include the indicator or value definition. What is the map about? Examples include malaria prevalence, population density, and unemployment rates. The 'where' may be an entire country, a province, a district, or a smaller area. The 'when' tells the map reader the timeframe of data collection for the map.

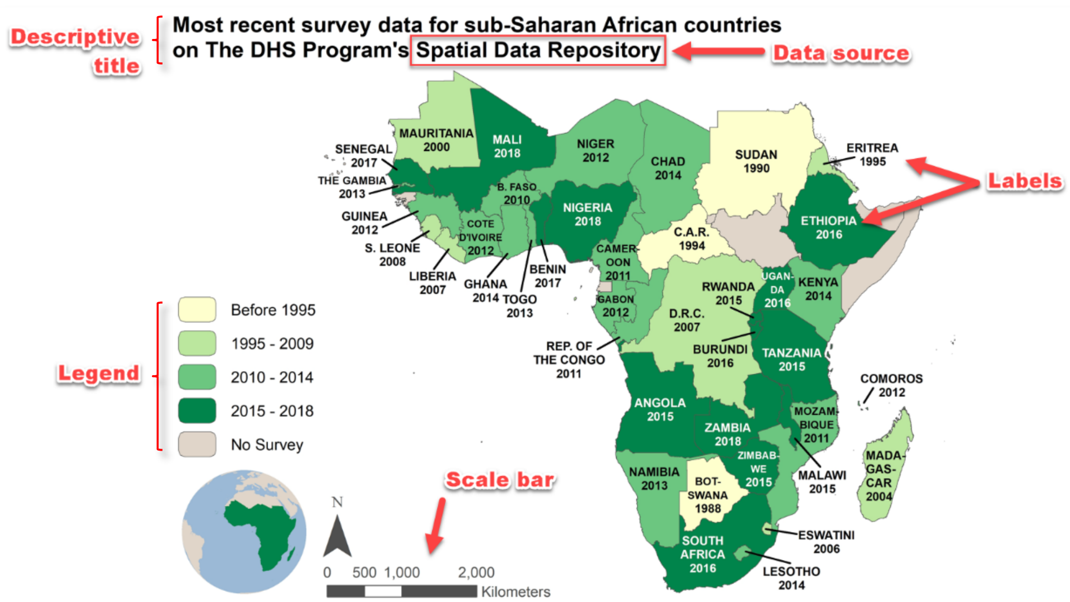
- **Data source**: The data source of the map tells the map reader where the information presented in the map came from. It is important to include the year in the data source information. This way, the map reader knows if the data are current and can make comparisons between maps to observe changes over time. The data source should include the name of the agency or survey that collected the data, or the report in which the data are found.

- **Legend**: A map's legend explains how map readers should interpret the map. Legends may explain symbols, shading, or other elements found on the map body that need explanation. Without a legend, the map's audience may not understand what the map is showing. For instance, if we look only at the body of this map, we see different colors, but we would not know that the colors indicate "the mean number of people per sleeping room in a household" without a legend. With the legend, we can tell that the darker colors indicate a higher average number of people per sleeping room in each province of Ethiopia. A map's legend needs to give a short descriptor of the data shown. It should also include the unit of measure, such as if the data is a percentage or a count. As you can see in these examples, legends can look very different from map to map.

- **Labels**: While labels won't be explored until [Subsection 8.1.2](), it's important to understand the importance of labels. Labeling is an essential part of map making. It can mean the difference between maps that are busy and difficult to read and maps with clear labels that direct your eyes.

- **Scale bar (or scale text)**: A scale bar is typically a line marked like a ruler in units proportional to the map's scale. A scale bar helps the map reader estimate distances between places and the size of features. This map uses kilometers in its scale bar. A more detailed map of a smaller area, such as a village or town, might use meters instead. Scale bars should use well-rounded numbers such as 1, 5, 10, 100, or 1000, that can easily be understood by the map reader.

The diagram below highlights these five essential map elements. Some maps also include a north arrow and inset map as well.



There are two main types of maps that we'll be covering in this course: reference maps and thematic maps.

- **Thematic maps** (like the one above) are designed to convey information about a single topic or theme, such as percent of households with electricity or population density, as in this example. It's important to know what kind of map you will be producing because it will affect your design choices and use of symbols.

- **Reference maps** (like the one below) are designed to show where geographic features are located, in relation to each other. Some may emphasize a particular feature, such as this map which emphasizes the administrative divisions of Uganda. Other reference maps are made without a specific emphasis on one feature over another.

Maps may be classified into two main mediums. Print maps, or hard copy maps, can be found in brochures, magazines, books, and The Demographic and Health Surveys themselves. For print maps, you must know if the product will be produced in color or black and white.

Digital maps, on the other hand, are displayed on a cell phone, tablet, laptop, or computer. A great example of a digital map is included in The DHS Program's online application, STATcompiler (statcompiler.com).

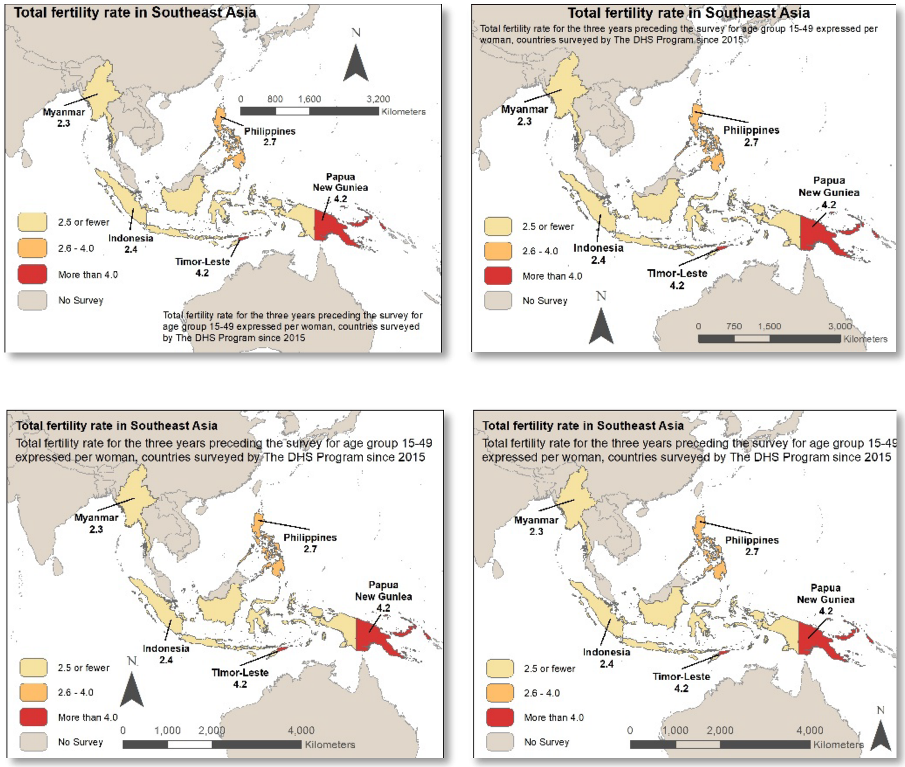
While print maps are static, digital maps can be dynamic or, in other words, interactive. Many online mapping applications, such as STATcompiler, allow the user to select and query different map features, as well as manipulate the data that is displayed. While this secetion primarily pertains to static map design, many of the design principles can be applied to interactive maps as well.

4.1.3 Map Design Basics

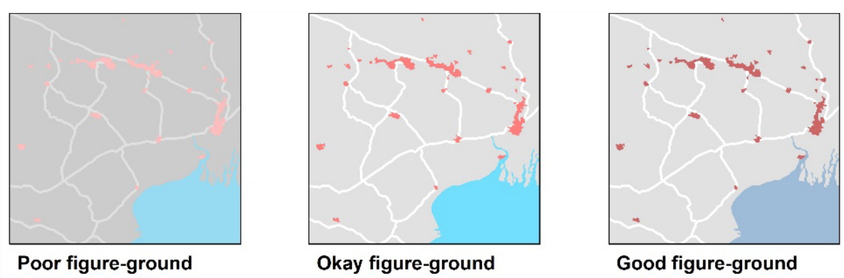
Map design consists of several elements itself. It includes balance, visual hierarchy, figure-ground organization, color, and scale. We'll discuss the first three elements below.

- **Visual hierarchy** in the layout helps direct map readers to the most important information first. Visual hierarchy enables map readers to identify and interpret patterns. Elements that are placed near the top and towards the center of the page will appear higher in the visual hierarchy than those that are towards the bottom or are at the edges of the map.

- **Balance** involves the organization of the map as well as any extra elements on the page. Positioning heavier elements together can make the page look top-heavy or bottom heavy. Centering the map slightly above center ensures that it is in the most prominent position on the page. The position of elements can also cause the eye to move in a desired direction. In the example below, the top two maps are unbalanced while the bottom two are well balanced. In the top two, elements of the map layout are spread unevenly across the page. In the bottom two, the map elements are spread evenly around the map and are placed flushed with the map's edges. This allows for a more balanced map.

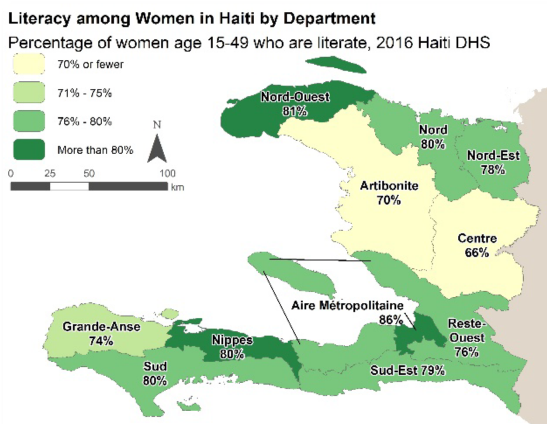


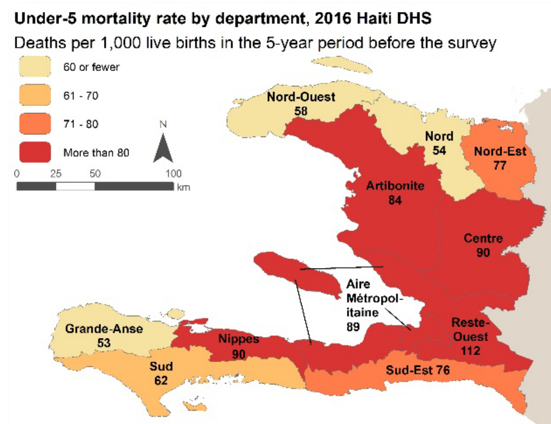
- **Figure-ground organization** is the separation of the figure in the foreground from the background. In terms of cartography, we consider the figure-ground organization of features within a map, such as boundaries and roads, as well as between elements of the map layout, such as the title and legend. The higher the contrast between features and elements, the more something will stand out, usually the feature that is darker or brighter. When there is no variation in visual contrast, the map reader has a hard time distinguishing important features from the less important background. See the image below for examples of figure-ground organization.



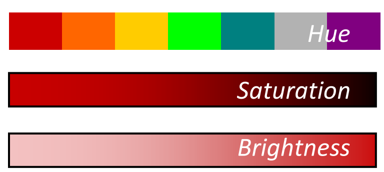
Now let's discuss **colors**. Colors convey a lot in mapping. Red often denotes something bad or negative while green is often used to denote something good or positive. Blue, on the other hand, is most often associated with water features. Darker colors can be interpreted as more of something and lighter colors as less of something, though this is not always the case.

In the maps below, we see an example of appropriate color use in maps. Green is used to show a positive variable and a darker shade means more of something – in this case, literacy rates. The other map shows red to map a negative variable and darker shades of red mean more of something negative – in this case, under-5 mortality rates. While red is often associated with negative data, this is not always the case. For example, all indicator maps in DHS reports use a red-orange-yellow scale, regardless of what the indicator is. This consistent style makes DHS indicator maps instantly recognizable.





Digital colors can be mixed and made through a combination of three elements: hue, saturation, and brightness (see the graphic below). Hue is the type of color, numerated by where it falls on the color wheel. These include primary, secondary, and tertiary colors.



With a hue selected, a color can be further tweaked by changing the level of saturation and brightness. Saturation is the intensity of the hue while brightness is the level of lightness or darkness.

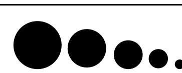
Different hues are used to show different categories or types of locations. In other words, different hues are often used to represent different types of categorical data. However, the level of saturation and brightness can be used to show different quantities of things.

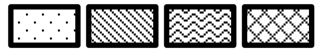
An important consideration to think about is colorblindness. About 5 to 8% of men and 0.5% of women have some type of colorblindness. Even though it's a small percentage of the overall population, poorly chosen colors can make the map unreadable for people with colorblindness. There are color schemes that are optimal for people who are colorblind, such as the ones shown here. You can find these and other colorblind safe schemes on the website, Color Brewer 2.0 (colorbrewer2.org).

Cartographers have come up with a series of guidelines about how to match different visual characteristics used to represent continuous and categorical data. The use of graphics to represent data and physical space in maps is referred to as symbology.

Visual characteristics that effectively represent continuous data are color saturation, size, and texture.

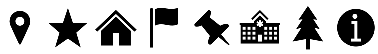
Representing Continuous Data

Color saturation	
Size	

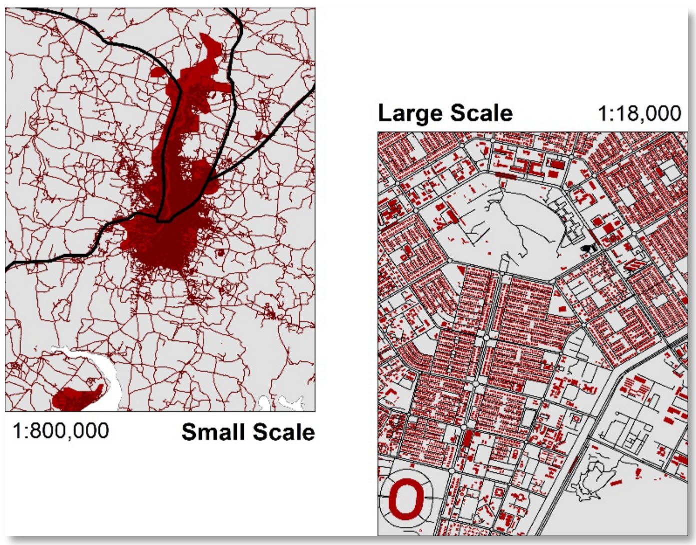
Texture	

Visual characteristics that work best for showing categorical data are color hue and shape.

Representing Categorical Data

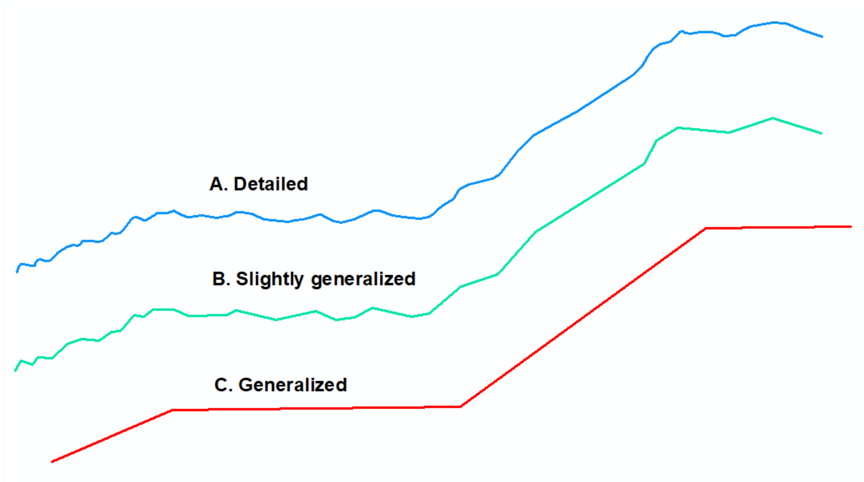
Color hue	
Shape	

The final element in map design is **scale**. Whenever we contort the globe to map a flat map, we have to choose a scale. The scale tells us the relationship between the dimensions of our map and the actual world. Let’s look at an example.



You might be surprised that the small-scale map actually covers a larger area. The large-scale map, however, shows a higher level of detail but of a smaller area. Small-scale maps earn their name because they reduce a large area of land to a relatively small map unit. Large-scale maps reduce an area of land to a relatively larger space. Scale is important to understand as it affects the level of detail a map and the GIS data set can show.

Generalization (depicted below) is the reduction and simplification of features for change of scale or resolution. The idea is that if you are zoomed way out, the features on your map are going to look less detailed and intricate compared to when you are zoomed in.



Map scale can be communicated three different ways on a map: as a scale bar as a representative fraction or stated textually.

Scale examples

Scale bar	
Representative fraction	1:500,000
Text	1 cm = 5 km

In sum, if you understand layout, type, medium, design, and scale, you are now ready to make a map that your audience will remember.

4.1.4 Additional Map Layout Elements

As explained in [Subsection 4.1.2](), the essential elements include the map frame, title, data source, legend, labels and scale bar. These elements help the map reader understand what the map is depicting. They provide important contextual information. Without them, the map reader may not be able to understand what the map is designed to communicate. However, there are some additional elements that may add value to your map. Some examples include north arrows, projection name, additional map frames, explanatory text, and graphics.

- **North arrow**: North arrows help orient the map reader and are most useful on small-scale maps when direction may be difficult to discern. Use a north arrow when the map reader might not be able to orient the map, such as when the map isn't oriented with north at the top.

- **Projection name**: As we'll explore in [Subsection 4.1.5](), map projections are different representations used to depict our round Earth on a flat surface. In cases where your map uses a unique projection or dramatically contorts the view, you should include the map projection as a footnote.

- **Additional map frames**: Map layouts sometimes include more than one map frames. Sometimes these map frames might compare two maps (e.g., two maps in one layout comparing two different health indicators). In other cases, a smaller map, called an extent indicator map, may be used to orient the map reader to the location of the map in relation to other places (e.g., including an extent indicator in a map of the Gambia showing where the country is in West Africa). Another example is what is called an inset map, which are smaller maps that enlarge a portion of the primary map frame (e.g., a map of Bangladesh might include an inset map of Dhaka, showing the city at a larger scale). You will learn how to create maps with multiple map frames in **Subsection 9.1.2**.

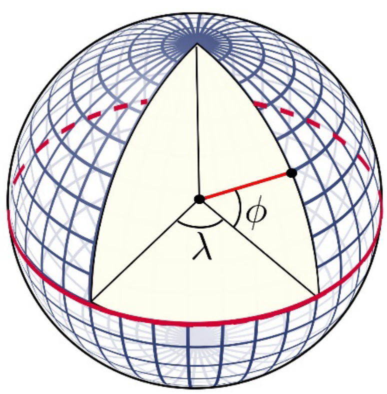
- **Explanatory text**: Maps may often include additional text that gives context to the map. This may include the name of the map maker, the funding organization, information about the map's subject or purpose, a graph or table based on the data in the map, and the date of publication.

- **Graphics**: A map may also include images, such as a company or organization's logo.

4.1.5 Coordinate Systems and Map Projections

When we create a map in QGIS, we are distorting the shape of our planet. Earth is an imperfect sphere with high peaks and low valleys. To transform the complex shape of Earth into flat maps, we use coordinate systems in GIS software. There are two primary types of coordinate systems: geographic coordinate systems and projected coordinate systems.

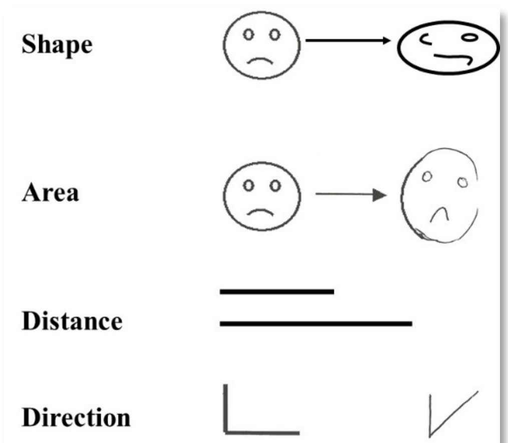
- **Geographic coordinate system**: This coordinate system uses a three-dimensional surface to define locations on Earth. It uses latitude and longitude to identify positions on the Earth's surface, which are based on angular measurements to fit the nearly spherical shape of the Earth. In other words, location is measured in relation to the curved surface of Earth. An example is shown below.



- **Projected coordinate system**: In this coordinate system, locations are defined on a flat surface. By turning the spherical world flat, the Earth is distorted. Rather than latitude and longitude, projected coordinate systems measure locations using units such as meters and feet. An example is shown below.

Three common shapes, used to turn our spherical Earth into a flat projection, are cylinders, circles, and cones. These three shapes, in turn, become cylindrical, azimuthal, and conic projections. As a result, we have many different map projections with many different shapes.

Since map projections cannot provide absolutely accurate representations of Earth, we need to prioritize which data properties we want to preserve. Converting features from a 3D surface onto a flat surface always leads to some type of distortion in shape, area, distance, or direction. A good way to remember this is that projections make geographers SADD. SADD is an acronym for Shape, Area, Distance, and Direction.

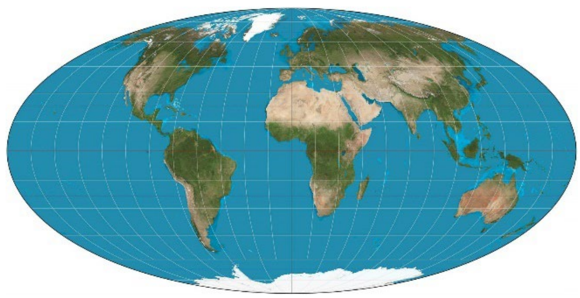


The graphic above illustrates these four types of distortions. In the first example, the shape is distorted and it's hard to tell that the second image is a frowning face. In the second example, the second face has an area distortion, because it is much larger than the original image. The last two examples illustrate distance and direction distortions.

- **Conformal**: Conformal map projections preserve the angles (direction) between objects on Earth's surface. These are useful for navigational charts and weather maps. While shape is preserved for small areas, the shape of a large area such as a continent will be significantly distorted. This example map specifically uses a Lambert conformal conic projection.

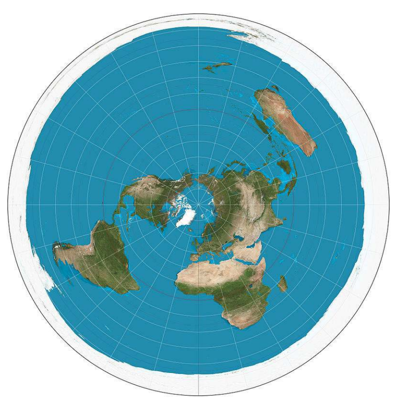
© Daniel R. Strebe / Wikimedia Commons / CC-BY-SA-3.0

- **Equal-area**: Equal-area map projections preserve the area of map features. In order to do so, they distort the shape of some features. These projections are most often used for thematic maps that illustrate population or environmental data. This example map specifically uses an equal-area Mollweide projection.



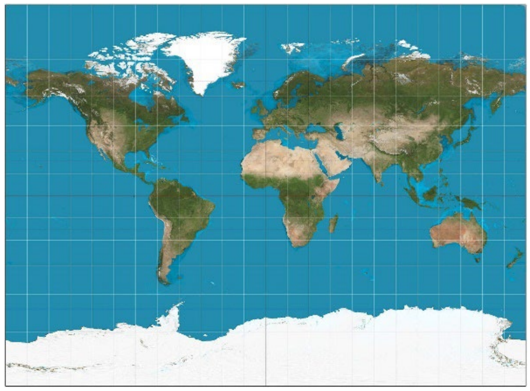
© Daniel R. Strebe / Wikimedia Commons / CC-BY-SA-3.0

- **Equidistant**: Equidistant projections preserve the distance between one or two points and all other points on a map. Given the nature of geometry, no equidistant projection can preserve distances from all points to all other points. Equidistant projections are useful for telecommunication or radio maps to display point-to-point communication. This example map uses an azimuthal equidistant projection. The United Nations emblem is based upon a map using this specific projection.



© Daniel R. Strebe / Wikimedia Commons / CC-BY-SA-3.0

- **Compromise**: Compromise map projections attempt to balance all four SADD distortions – shape, area, distance, and direction. These projections are therefore popular for general-purpose world maps. A famous user of compromise projections is the National Geographic Society. This example is a Miller cylindrical projection, a common compromise map projection.



In this course, we'll primarily be using a geographic coordinate system called WGS 84 (sometimes written as WGS84). Built in 1984, this is the geographic coordinate system used by the United States Global Positioning System, or GPS. It is also commonly used in online and desktop GIS applications.

4.2 QGIS INSTRUCTIONS

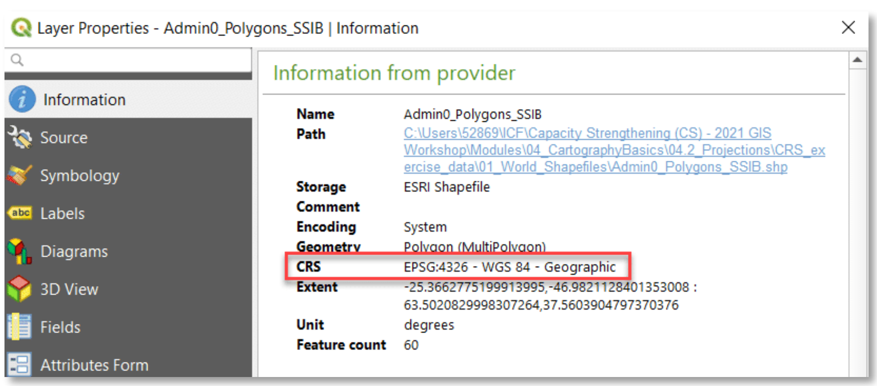
4.2.1 Coordinate Reference Systems in QGIS

Before creating map layouts, let's explore how to identify and change the coordinate reference system you are using in your QGIS project.

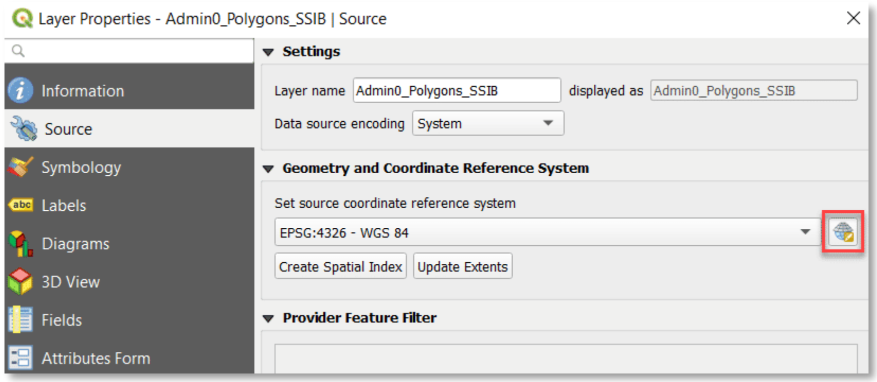
This section corresponds to Activity 4.2.2 in the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\ Exercise_Files\4.2.1_CoordinateSystems\ folder.

1. Open the Coordinate_Systems.qgz project file in the above-mentioned folder.

2. In the Layers Panel, right-click on the Admin0_Polygons_SSIB layer and select Properties. Open the Information tab. The coordinate system of the layer is noted below in red. It shows that the layer is in the WGS84 coordinate reference system.



3. For more information or to change the coordinate reference system, click on the Source tab (the second tab). Click the Select Coordinate System button, which is a globe icon button to view more coordinate system information. This is identified in the image below.



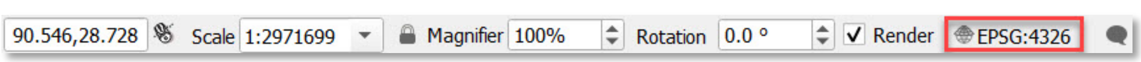
4. When the Coordinate Reference System Selector window opens, the WGS84 row is highlighted. Scroll all the way to the top of the Coordinate Reference System Selector sub-window in the Predefined Coordinate System Reference Systems box to determine which type of coordinate system WGS84 is. You will see that WGS84 falls under the Geographic Coordinate Systems classification.

It's important to note here that units of measurement for geographic and project coordinate systems are different. In geographic coordinate systems, the units of the layer will be in angular units, such as decimal degrees or degrees, minutes, seconds. There will be only 1-3 digits to the left of the decimal point. For projected coordinate systems, the units of the layer will be in either meters, feet, or some other linear unit of measurement. There will be 6-8 digits to the left of the decimal point. These units are visible on the bottom of your map project window. Usually bottom left of your screen.

5. Click Cancel to exit the Layer Properties window.

Now, let’s look at how different coordinate reference systems impact measuring distance.

6. Click on the CRS Status button (shown below) at the bottom-right corner of your QGIS window.



7. Type 3395 into the Filter section. Click on the WGS84/World Mercator Projected Coordinate System to select it. Click OK.

8. Right-click on the Admin0_Polygons_SSIB layer in the Layers Panel. Select Zoom to Layer.

9. Click the Measure Line button on the Attributes Toolbar to activate the tool.

10. Let’s measure the distance between Tunis, Tunisia and Cape Town, South Africa. With the Measure Line tool activated, click on Tunis (the green point) and then Cape Town (the blue point). The distance in meters between the two cities will display as the first segment in the Measure window. Write this distance down and make a note that it was measured using the World Mercator projected coordinate system. Close the Measure window.

We’ll now compare this measurement with the same measurement but using a different coordinate reference system.

11. Again, click on the CRS Status button at the bottom of your QGIS window. Type 3035 into the Filter section. Select the Lambert Azimuthal Equal Area → ETRS89 / ETRS-LAEA Projected Coordinate System. Click OK.

12. Right-click on the Admin0_Polygons_SSIB layer in the Layers Panel and select Zoom to Layer.

13. Use the Measure Line tool to measure the distance between Tunis and Cape Town. Again, write down the distance between the two cities, but label this measurement as Lambert Azimuthal Equal Area.

Consider the difference between these two measurements.

14. Close your project without saving.

4.2.2 Creating Print Layouts in QGIS

In this section, you will create a map of rivers and lakes in the Democratic Republic of the Congo (DRC). Specifically, you will create a map layout that includes a descriptive title, the data source, and a legend.

This section corresponds to Activity 4.4.2 of the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\4.2.2_PrintLayouts\ folder.

1. Open the DRC_Rivers.qgz project.

2. Before we create our Print Layout, let's check the CRS assigned to this map. To do this, click on Project → Properties.

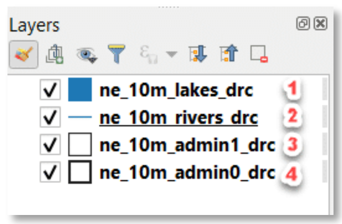
3. In the Project Properties window, select the CRS tab. You'll see that the CRS for this map is set to WGS 84. This is the preferred CRS for viewing global data in a GIS. For this activity, however, we want to use a CRS that is a little more precise for the DRC.

4. In the Filter box, type "Africa". From the Predefined Coordinate Reference Systems list, select the CRS called Africa Albers Equal Area Conic. With this CRS selected, click OK at the bottom of the window.

What happened to the DRC's shape? Does it look different using the equal area projection? You’ll notice that the DRC now appears more narrow and taller, as if it were stretched. The equal area projection preserves the area measure but distorts shapes.

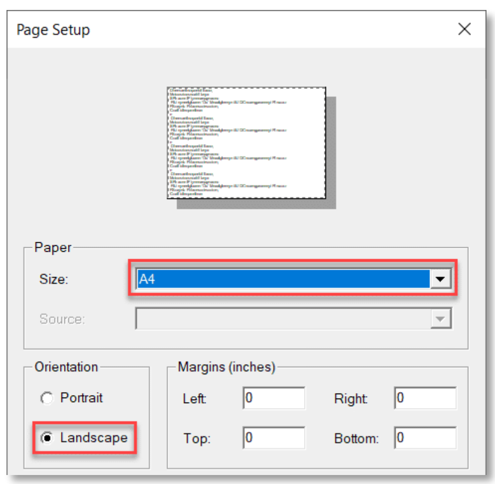
5. The next thing we want to do to prepare our map, is to make sure our layers are properly displayed and ordered. Luckily, a cartographer already did a lot of work on this map. There are even labels added to it. (You'll learn about adding labels in a future section).

6. The next thing we want to do to prepare our map, is to make sure our layers are properly displayed and ordered. Luckily, a cartographer already did a lot of work on this map. There are even labels added to it. (You'll learn about adding labels in [Subsection 8.1.2](#)). However, the layers are not ordered correctly in the Layers Panel. Please adjust the layers in the Layers Panel so that they are in the order below.



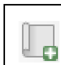
7. With the layers now in the correct order, let’s create a print layout. Click Project → New Print Layout. Give your new Print Layout an appropriate title such as “DRC Rivers and Lakes”.

8. With a new Print Layout created, let's select the size of our page. Click Layout → Page Setup. In the Page Setup dialog box, select the standard page size used in the country from the Size dropdown. For this map, keep the Orientation set to Landscape. Once you select the appropriate page size for your map, click OK.



9. The Print Layout window will display a blank canvas on which you can add different map elements. To add map elements, you will select the elements from the list of icons along the left-hand side of the screen, click-hold-and-drag your mouse on the canvas, and release your mouse. With the following five layout tools, you can add the essential components of a map layout.

5 Layout Tools for Adding Elements

	Add map to the Print Layout
	Create title/text
	Add legend

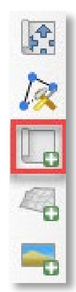
	Add scalebar
	Add north arrow

10. To navigate around your layout, map, or to move elements you created, you can use the following five tools.

5 Layout Tools to Manipulate the Layout

	Pan around the Print Layout
	Zoom in on the Print Layout
	Zoom out from the Print Layout
	Select and move elements in the Print Layout
	Pan around and change the extent of the map itself

11. The first element we're going to add to our layout is the map itself. To do this, click on Add Item → Add Map. Then, click on your blank page, hold, drag, and release your cursor to create your map. You can also use the Add New Map button in the toolbar on the lefthand side of the screen by clicking the icon identified below.



12. Once you create your map on the layout, you may want to change its extent (zoom level) or move the map's frame (pan around the map). The easiest way to do this, is to use the Move Item Content tool (identified below). This is the third tool down on the left-hand toolbar.

13. Once you have this tool selected, you can click-hold-drag your cursor to pan around the map. If you have a computer mouse, you can use the mouse wheel to zoom in and out. Hold down the Crtl button on your keyboard while zooming to increase zoom precision. To change the map's frame, you can also use the element's properties. In the Items menu on the right side of QGIS, you can select the Item Properties tab and tweak the settings under the Extents heading. This is not usually recommended as it's more difficult than manually panning and zooming.

14. To add a title to the layout, you can use the menu at the top of the Print Layout window. Click on Add Item → Add Label. You can also use the Add New Label tool on the left-hand toolbar. It's the ninth tool from the top.

15. With this tool selected, click-hold-drag on your layout to create a new text box. You'll use this first box as your title.

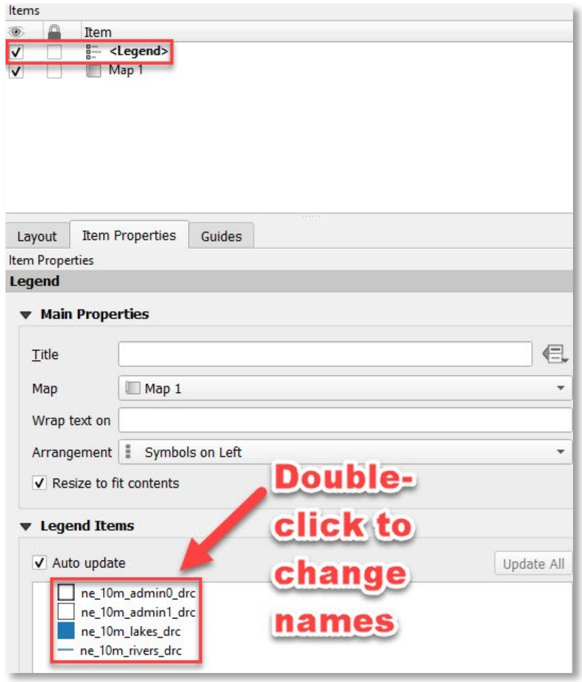
16. When you create a text box, you'll see that it contains the dummy text, "Lorem ipsum". We can change the contents of this text box and its style in the Item Properties menu on the right-hand side of the QGIS print layout window. You can also change other properties here like the font, text size, and paragraph alignment.

17. To create the citation, repeat **Step 2** to create an additional text box. This second text box should contain the following: "Source: Natural Earth Data 2021".

18. As demonstrated with other elements in our layout, there are two ways to add a legend. The first way is to use the menu: Add Item → Add Legend. We can also use the Add New Legend tool on the left-hand toolbar. It's the tenth from the top, located just below the Add New Label tool.

19. Using one of the mentioned methods, click-hold-drag-release to create a new legend in your layout. As you can see, this new legend contains the following four items: (1) ne_10m_lakes_drc, (2) ne_10m_rivers_drc, (3) ne_10m_admin1_drc, and (4) ne_10m_admin0_drc.

20. These names will not be readable for those unfamiliar with your data. So, we need to change the properties of our legend to make it clearer. In the Item Properties for the legend, expand the section called Legend Items (shown below). Double-click on a layer name to change it. Change these layer names to "Lakes", "Rivers", "Regions", and "DRC" respectively.



21. Before we finish with our legend, let's give it a name. It's common for cartographers to name their legends, "Legend". To do this, enter the text, "Legend", in the Title box of the legend's Item Properties.

22. We can add a scale bar with the Add New Scale Bar tool or from the menu (Add Item → Add Scale Bar).

23. Once you've added the scale bar to your map, let's explore the element's properties. We can change a lot of properties here, including the number of segments in our scale bar as well as its overall style. However, the two most important properties to check are the Map and Scalebar Units properties. When we start adding advanced elements to our map in later sections (e.g., extent indicators and locator maps), we need to make sure that our scale bar is reflecting the scale of the proper map in our layout. As displayed in the image, the Map dropdown menu is correctly set to 'Map 1'. Also, we want to make sure our scale bar is using the correct units. In this case, the Scalebar Units property is correctly set to kilometers.

24. Once you are done editing your map layout, let's export an image of the map. To do this click on Layout → Export as Image. Be sure to export your image to a folder you will remember.

4.2.3 Displaying Categorical Data in QGIS

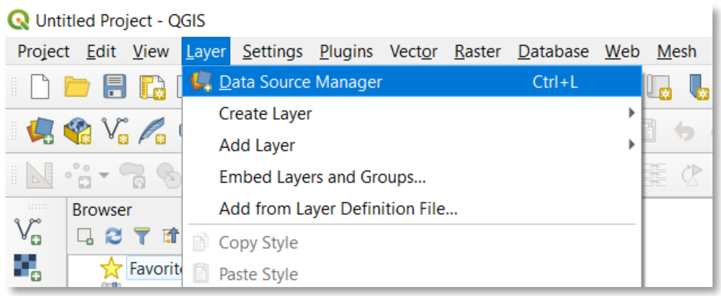
In this section we will explore categorical symbology in QGIS. Specifically, we're going to create a map of African countries, colored by their income group. This dataset is derived from Natural Earth Data.

The instructions below correspond to those included in Activity 4.5 of the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\ 4.2.3_ CategoricalData\ folder.

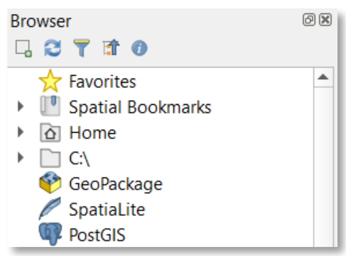
1. Open QGIS and create a new project by clicking the text, New Empty Project, under the heading, Project Templates.

2. With the new project open, add the shapefile you downloaded in this activity's exercise files; to do this, follow one of the two steps below.

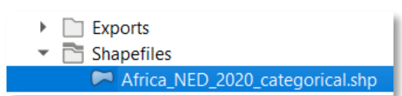
    a. The first way to add a layer is to use the dropdown menus at the top of QGIS. As illustrated in the screenshot below, click on the Layer menu and Data Source Manager from the dropdown menu. This will open the Data Source Manager in a popup window. In the Data Source Manager, navigate to the folder where you unzipped your exercise files be expanding the folders with the carrot buttons.



    b. You can also add data directly from the Browser Panel. In the Browser Panel, navigate to your unzipped exercise folder by clicking the carrot buttons.

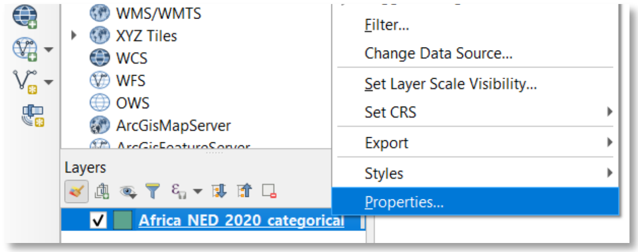


3. Once you have expanded the folders in either the Data Source Manager or Brower panel, navigate to the Shapefiles folder and double-click on the Africa_NED_2020_categorical.shp shapefile to add it to your map. See the screenshot below.

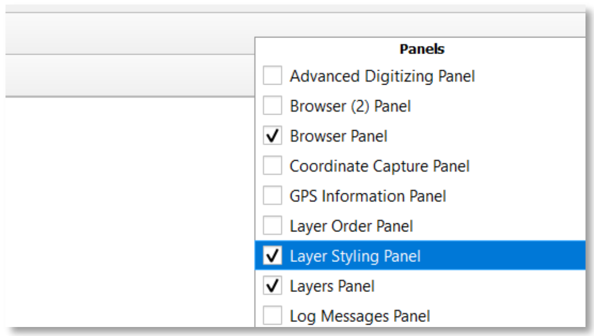


4. With the shapefile added to our project, we can now begin styling it. As with adding a layer, there are two primary ways to style a layer. Styling a layer can be done in either a popup window or a panel.

    a. To style a layer in a popup window, right-click on the layer in the Layers Panel and click Properties. In the Layer Properties popup window, select Symbology tab from the left-hand menu.

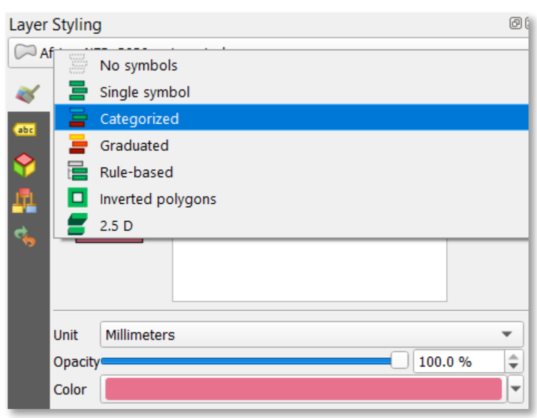


b. The other way to style a layer is to use the Layer Styling Panel. To open the Layer Styling Panel, either click on the Layer Styling icon ( ) in the Layers Panel or right click on the QGIS menu bar and check the box next to the Layer Styling Panel.



If you ever accidentally close a panel or do not see a panel you need displayed, you can right-click on the QGIS menu bar to toggle on and off panels.

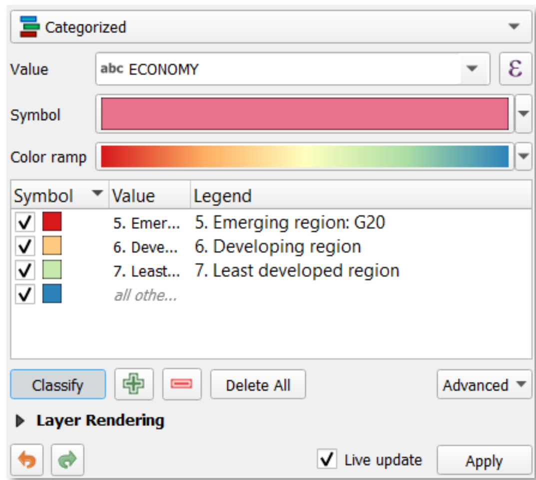
5. In either the Symbology tab of the Lay Properties popup window or the Layer Styling Panel, you can change the data's symbology so that qualitative data is displayed in different colors. To do this, you need to change the style from Single Symbol to Categorized. This can be done in the dropdown menu at the top of the symbology window/panel. See the screenshot below.



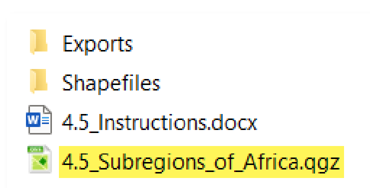
6. Once you select Categorized from the dropdown menu, your layer will disappear. To display our data with categorical symbology, we need to select an attribute (column) in the layer's attribute table that contains the data we want displayed. Under the first dropdown menu (in which you selected Categorized), there is a second dropdown menu next to the text, Value. In this dropdown menu, the Value dropdown menu, select one of the following attributes: ECONOMY, INCOME_GRP, REGION_UN, SUBREGION, or REGION_WB.

In the Value dropdown menu, you will see that most attributes in this layer have the symbol 'abc' next to them. Only the NE_ID attribute has a '123' next to it. The abc denotes attributes that contain text while the 123 denotes attributes containing numerals.

7. With an attribute selected, we can now choose our symbology. Click the dropdown menu next to the text, Color Ramp, and select the color ramp you want to use for your layer. Once you select a color, click the Classify button at the bottom of the menu.



8. With our qualitative data displayed with categorical symbology, we can now create a simple map layout in QGIS. Before we do so, we should save the work we have already done. It is important that we save our map next to the folder containing our data. In the case of this exercise, our data is stored in a folder called Shapefiles which you unzipped at the beginning of this activity. To save the QGIS project, click on the Project menu and select Save As. In the popup window, navigate to the folder where you unzipped your exercise files. Save the QGIS project next to the folder containing the shapefiles. In the example below, the filename includes the text, “Subregions_of_Africa”, so that the reader knows what the map displays.



9. First, we need to create a new Print Layout. In the QGIS menu, click on the Project menu and then New Print Layout from the dropdown. In the popup box, enter a title for your Print Layout. Once you enter a title and click OK, your new Print Layout will appear in a separate window.

10. The Print Layout window will display a blank canvas on which you can add different map elements. To add map elements, you will select the elements from the list of icons along the left-hand side of the screen, click-hold-and-drag your mouse on the canvas, and release your mouse. Be sure your map includes all five of the elements listed below.

    a.  Add map to the Print Layout

b. Create title/text

c. Add legend

d. Add scalebar

e. Add north arrow

11. With the five elements added to your map and organized to your liking, you are now ready to export your map. Before we export the map, let's save our Print Layout. To do that, click on the Layout menu and select Save Project.

If for some reason you close your Print Layout or QGIS Project, you can reopen the Print Layout by clicking the 'Project' menu in QGIS and selecting 'Layouts' from the dropdown menu and then the layout you created from the list there.

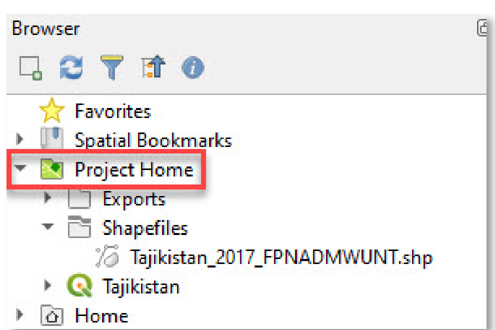
12. To export the map, click on the Layout menu in your Print Layout window and select Export as Image.

4.2.4 Displaying Continuous Data in QGIS

Here, you will create a map of unmet need for family planning according to the 2017 Tajikistan DHS. You will create what is called a choropleth map, a thematic map that uses colors to show how an indicator varies by geography.

This section corresponds to Activity 4.6 of the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\ 4.2.4_ContinuousData\ folder.

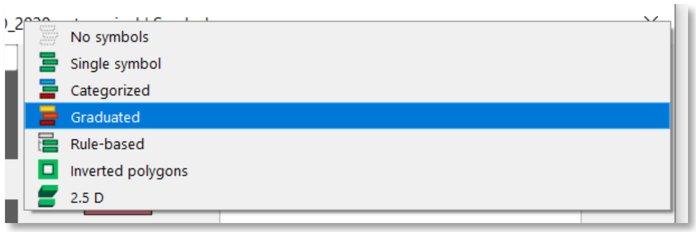
1. Open QGIS and create a new project by clicking the text, 'New Empty Project', under the heading, 'Project Templates'.
2. For this exercise, let's practice adding layers using the Browser Panel. Before you navigate to the exercise files, save your project in the exercise folder. Go to Project → Save As. Navigate to the 4.2.4_ContinuousData folder and save your QGIS project inside it. Call your project "Tajikistan". After saving the project, you will now see a green folder in the Browser Panel called Project Home (shown below). Now, you can easily access your exercise files from this folder.



3. Expand the Shapefiles folder in the Project Home section of the Browser Panel. Drag the Tajikistan_2017_FPNADMWUNT.shp into your QGIS project. It should then appear in your map frame and in the Layers Panel.

4. With the shapefile added to our project, we can now begin styling it. Open the Layer Styling Panel by clicking on View → Panels → Layer Styling.

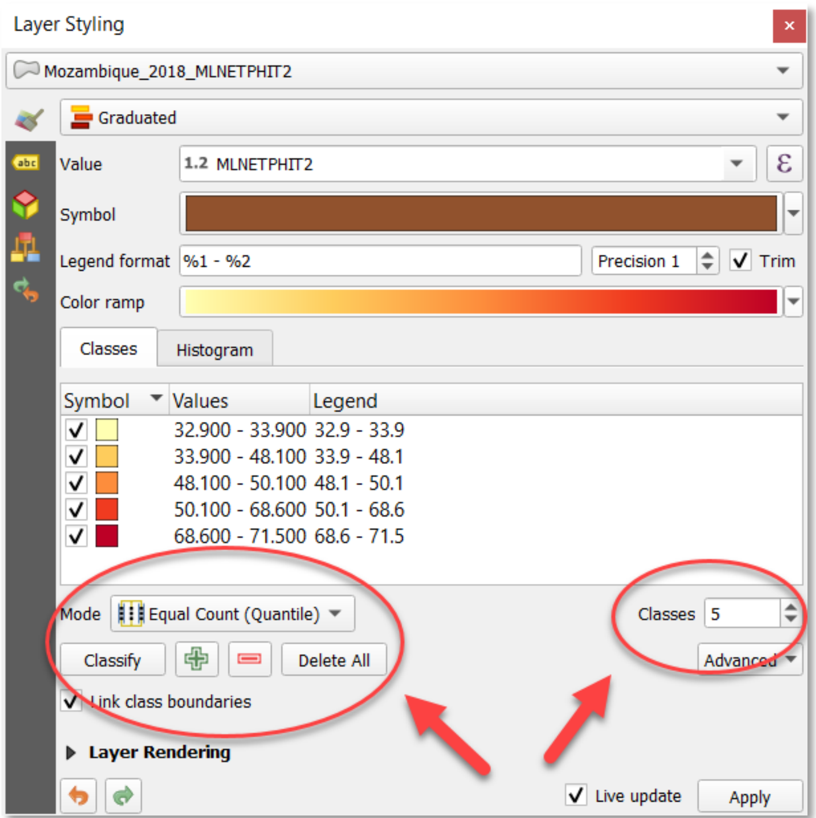
5. Change the style from Single Symbol to Graduated. This can be done in the dropdown menu at the top of the symbology window/panel. See the screenshot below.



6. Once you select Graduated from the dropdown menu, your layer will disappear. To display our data with continuous symbology, we need to select an attribute (column) in the layer's attribute table that contains the data we want displayed. Under the first dropdown menu (in which you selected Graduated), there is a second dropdown menu next to the text, Value. In this dropdown menu, the Value dropdown menu, select FPNADMWUNT.

7. With an attribute selected, we can now choose our symbology. Click the dropdown menu next to the text, Color Ramp, and select the color ramp you want to use for your layer. Once you select a color, click the Classify button at the bottom of the menu.

8. The current default classification for graduated symbology is quintiles: 5 classes of equal intervals. You can change the classification mode and number of classes with the tools noted in the screenshot below. In order to activate your changes to the classification, you must click the Classify button.



9. Now, let's create a print layout. In the QGIS menu, click on 'Project' and then 'New Print Layout' from the dropdown. In the popup box, enter a title for your Print Layout. Once you enter a title and click OK, your new Print Layout will appear in a separate window.

10. The Print Layout window will display a blank canvas on which you can add different map elements. To add map elements, you will select the elements from the list of icons along the left-hand side of the screen, click-hold-and-drag your mouse on the canvas, and release your mouse. Be sure your map includes all three of the elements listed below.

    a.  Add map to the Print Layout

    b.  Create title/text

    c.  Add legend

Your map's title should include the country name, indicator name, and year the data was collected. You should also include a second text block with the source of your data.

11. Once your map layout is organized to your liking, save and export your print layout as an image.

5. QUERIES AND SELECTIONS

5.1 CONCEPTS

When making a map or analyzing geospatial datasets, you may need to filter out unnecessary datapoints. If, for example, you have a dataset of every health facility in a country but only want to show health posts, you will need to filter out the hospitals and clinics. Maybe there is also an attribute in the data that says if the facility is open or shut down. In that case, you can filter the data so only open health posts are included. In this section, we will explore how to create such subsets. By the end of the section, you will be able to:

1. Query data in QGIS by adding filters to a layer's attributes,
2. Select data from a layer based upon their geographic location compared to other data, and
3. Summarize the number of points within a polygon.

5.1.1 Definitions

The following terms are used in this section.

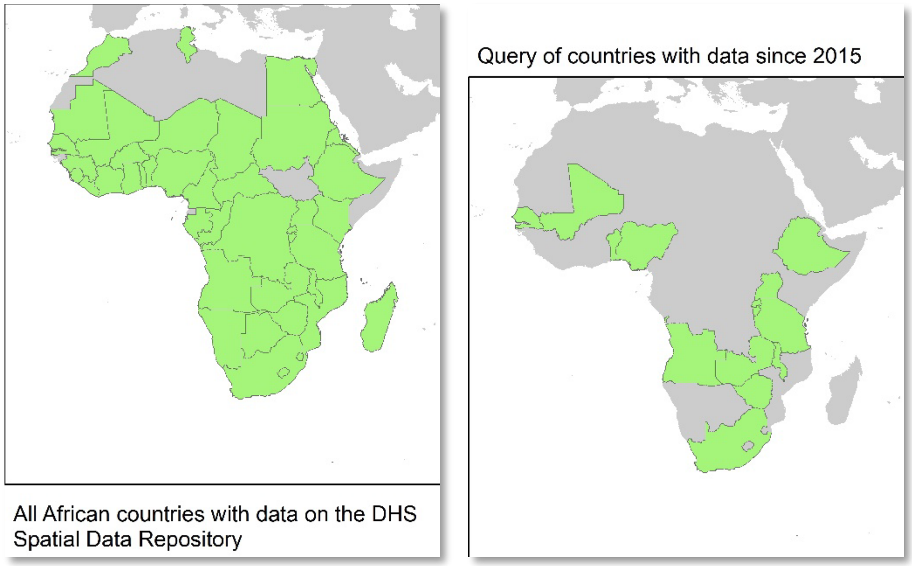
- **Selection**: a subset that is highlighted in a GIS dataset based on its attributes or location
- **Subset**: a smaller dataset created by querying out, selecting, or removing features based on given criteria
- **Query**: the process of filtering features out of a dataset by certain criteria

5.1.2 Querying Data

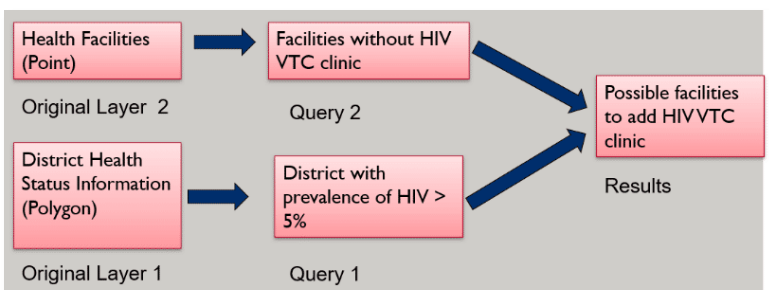
To query data is to create subsets of data based upon patterns or attributes within a dataset. From this subset we can:

- Create a new layer in our GIS that only contains that subset,
- Conduct a focused analysis on that subset or calculate statistics on it,
- Edit the content of that subset without affecting the original dataset, and
- Use that subset to select features from other layers.

In this first example (below), we have a layer of all African countries with data on the DHS Spatial Data Repository. Through a query, we then isolate the countries that have data as recent as 2015.



Here is another example (below). If a dataset of health facility points contains an attribute telling us whether or not it's an HIV clinic, we can create a query that displays only the facilities that are not HIV clinics. We can then compare this queried dataset to a second queried dataset.

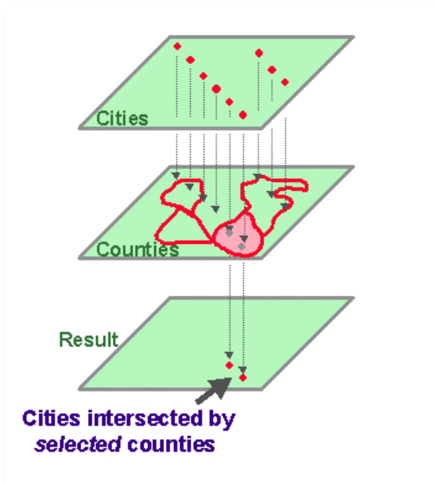


In this example, the second dataset is a polygon layer with the administrative districts of a country. As you can see, the districts with an HIV prevalence rate of more than 5% are queried. Districts with a lower HIV prevalence are not included in the query.

Through a spatial selection, we can use the queried district layer (showing districts with HIV prevalence of more than 5%) to select facilities without HIV clinics. After the query and spatial selection, we are left with a dataset of possible facilities to which an HIV clinic could be added.

5.1.3 Spatial Selections

Spatial selection is a type of querying used in GIS. It is the process of choosing one or more features in a dataset given their geographic location in relation to another dataset, instead of based on their attributes as in the last example.



For example, we can use a spatial selection to extract cities that fall within the boundaries of a queried county dataset (see the example above).

5.2 QGIS INSTRUCTIONS

5.2.1 Querying Data in QGIS

In this activity, you will create a map of total fertility rate (TFR) in DHS Program countries. However, the dataset provided includes TFR numbers from old surveys. So, you will use filters in QGIS to query out older surveys to create a map comparing TFR in countries with relatively recent surveys.

This section is based on Activity 5.3 of the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\5.2.1_ QueryingData\ folder.

1. In the unzipped folder, open the QGIS project file, Total_Fertility_Rate.qgz.

2. Explore the QGIS project. As you can see, there are two layers in this project. The first layer includes data from The DHS Program's Spatial Data Repository (SDR) and the second layer's data comes from Natural Earth Data (NED, naturalearthdata.com).

3. Review the attributes of these two layers by opening the attribute table. Right-click on the layer in the Layers Panel and select Open Attribute Table.

In the table, you will see that the NED layer contains very basic information about the countries in our map and that the SDR layer contains an attribute called FEFRTRWTFR. This attribute contains the TFR for each country included in the layer. The TFR is calculated for the three years preceding the most recent DHS Program survey for women age 15-49.

Also, note the two attributes, Survey and YearLabel, in the SDR attribute table. While both of these columns contain data on the year in which the survey was conducted, they store the data in two different forms. The Survey attribute stores the survey year as a number while the YearLabel stores the survey year as text.

The SDR layer contains TFR rates calculated from DHS Program survey data. Not all of the surveys are recent. For example, the data for El Salvador is based on data collected in 1985. While the data in Ghana was collected in 2019. For this exercise, we want to create a map that displays data no older than 2015. We want to query/filter out countries whose most recent survey was conducted before that year. To do this, we need to apply a filter to the SDR layer.

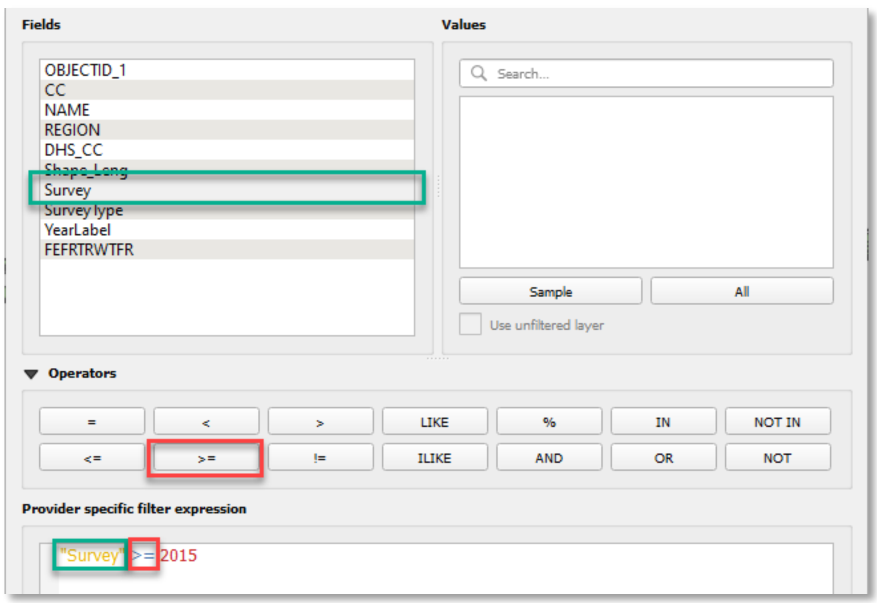
4. Select the SDR layer in the Layers Panel by clicking on it.

5. Right-click on the SDR layer and select Filter from the popup menu. This will open the Query Builder. (The Query Builder can also be opened by clicking the Layer dropdown menu at the top of QGIS and selecting Filter.)

6. To filter out data from surveys before 2015, we need to write the following expression in the expression box at the bottom of the Query Builder.

"Survey" >= 2015

You can either write this expression manually or select Survey from the 'Fields' list, click the '>=' operator, and type 2015 afterwards. See the screenshot below.



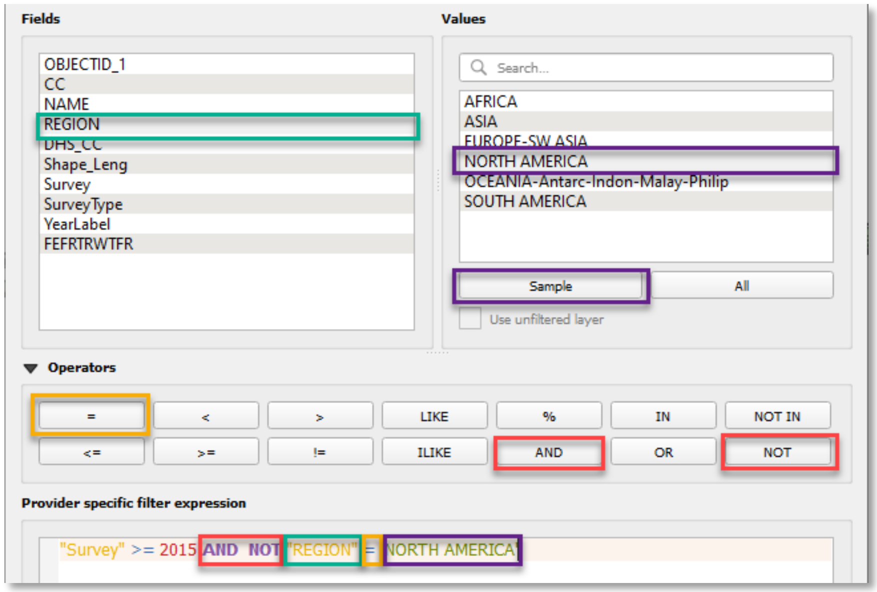
7. Click OK and the Query Builder will close. As you can see, the number of countries in the SDR layer has decreased. If we open the attribute table, we will see that only 41 countries remain (out of 87 countries originally displayed).

8. Since the majority of the remaining countries are in Africa and Eurasia, lets filter out those in the Americas. To do this, open the Query Builder again by right-clicking on the SDR layer and selecting 'Filter'.

9. We are now going to modify the expression so that it reads as follows.

**"Survey" >= 2015 AND NOT "REGION" = 'NORTH AMERICA'
AND NOT "REGION" = 'SOUTH AMERICA'**

You can either write this expression manually or build the query by clicking on the fields and operators. To do this, see the screenshot below.



10. With the full expression written, click OK to close the Query Builder.

11. Now you should only see SDR countries in Africa and Eurasia. To center this region in your map, right-click on the SDR layer in your Layers Panel and select Zoom to Layer.

12. Open the Layer Styling Panel and create graduated symbology to depict the TFR indicator (called FEFRTRWTFR in the attribute table). For a reminder of how to create graduated symbology in QGIS, see **Subsection 4.2.4** of this document.

5.2.2 Conducting Spatial Selections in QGIS

In this activity, you will use filters and selections to identify which health facilities in a dataset are located in Dhaka, Bangladesh. This exercise includes a shapefile of health facilities downloaded published by healthsites.io on the Humanitarian Data Exchange (data.humdata.org).

This section is based on Activity 5.4 in the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\5.2.2_SpatialSelections\ folder.

1. Open the Bangladesh.qgz project.

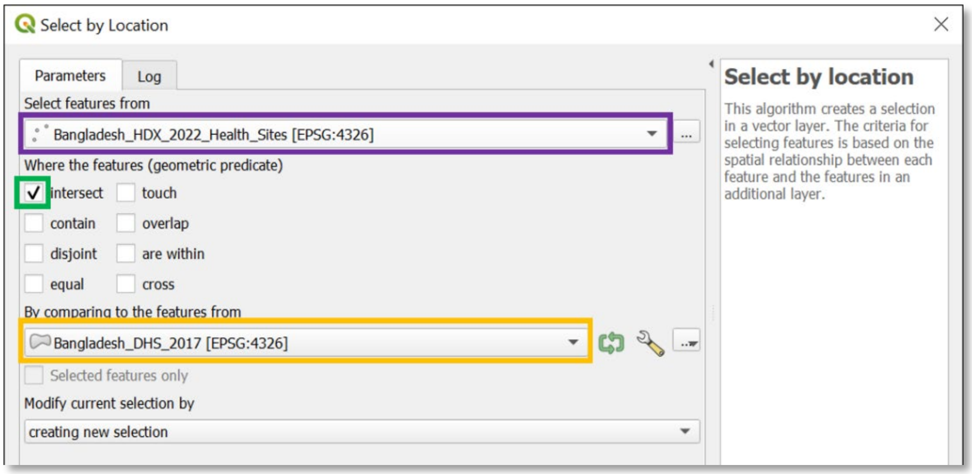
2. Create an attribute query on the administrative boundary of your case study so that only one region is displayed. In the Layers Panel, right-click on your case study's admin boundary and select Filter. In the Query Builder, build a query to display one admin boundary. The query should read:

```
"DHSREGEN" = 'Name of region'
```

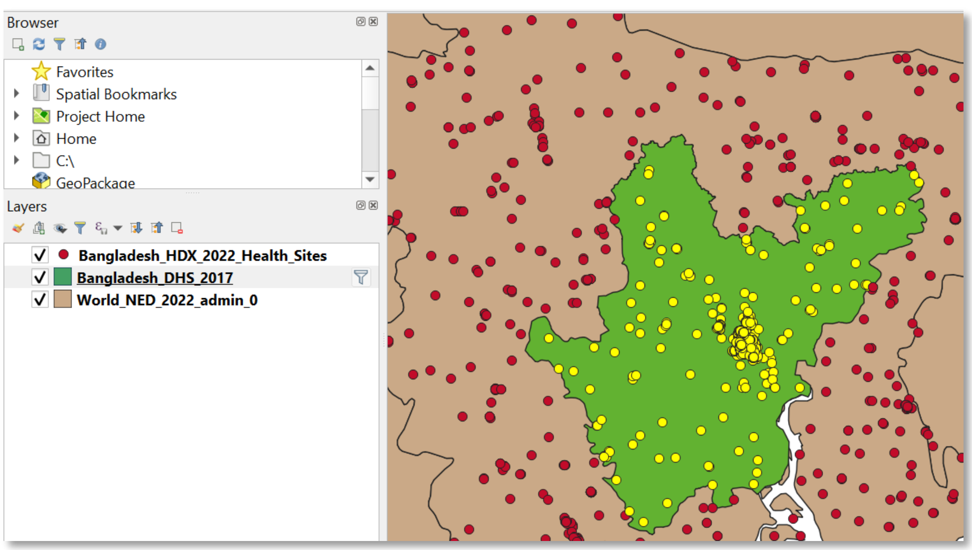
Replace `'Name of region'` with a region name, such as `'Dhaka'`.

3. Now that you have queried that layer, we can use it to spatially select health facilities. To do this, click on the Vector menu at the top of QGIS, then select Research, and click on the tool, Select by Location.

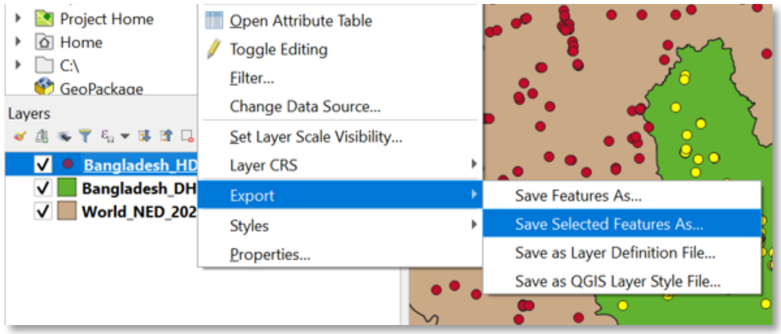
4. In the Select by Location tool window, you can choose which layer you want to select from, the selection method, and the layer which you are using to make the selection. For this exercise, we want to select features from the health facilities layer that intersect with the queried case study admin layer. Once you have selected the correct layers in the tool window (see screenshot below), click Run.



5. After you have run the process, close the window. In your map, you should see the health facilities within your queried region colored yellow. This change in color indicates that the health facilities have been selected. The screenshot below shows health facilities selected in the chosen region of Bangladesh. Note that the health facilities in Dhaka are colored differently than those outside.



6. We now want to export the selected health facilities as a new layer. Once this selection is exported, we can analyze them and give them different symbology. To do this, right-click on the health facilities layer in your Layers Panel, select Export, and choose Save Selected Features As. This is shown in the screenshot below.



7. In the Save Vector Layer As popup window, we need to choose where to save the layer. First, leave the Format as ESRI Shapefile. This is the correct format. Click the '…' button next to File Name. In the new popup window, navigate to your \5.2.2_SpatialSelections\ exercise folder and open the \Shapefiles\ folder. Save your selection in the \Shapefiles\ folder as Selected_Facilities.shp. Once you have selected where to save your new shapefile, click Save. And then click OK in the main popup window.

8. Once you click OK, the window will disappear, and the new layer will be added to your Layers Panel and map window.

6. CLEANING GPS DATA IN EXCEL

6.1 CONCEPTS

The GPS coordinates collected during the implementation of DHS Program surveys are essential to the datasets we produce. Location information allows us to verify that the correct households were visited, link our survey data to other datasets, and conduct advanced geospatial analyses on the results. In this section, you will be introduced to GPS data, the technology used to collect it, and processes to clean GPS data in Microsoft Excel. By the end of the section, you will be able to:

1. Clean GPS data using proper formatting and tools in Excel,
2. Use crosswalk tables in Excel to examine data,
3. Create pivot tables in Excel, and
4. Add GPS data into QGIS and display its coordinates.

6.1.1 Definitions

The following terms are used in this section.

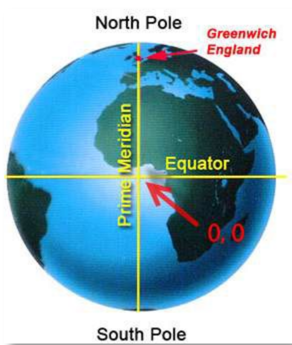
- **Altitude**: the height of a point or location above sea level; to collect altitude data with a GPS receiver, you need at least four satellites
- **Cells (Microsoft Excel)**: the boxes in a spreadsheet that contain values
- **Clean data (for QGIS)**: a cleaned dataset that (1) has a single header row with variable names that do not use special characters, (2) has no merged cells, (3) is formatted so that every row is a single observation (e.g., each row is a household cluster or health facility), and (4) uses the same format for any geographic information in a column
- **Crosswalk tables**: a table comprised of values pulled from multiple datasets; use the =VLOOKUP() function to create a crosswalk table in Excel
- **Decimal degrees**: coordinates in DDD.DDDDD format (e.g., 34.73459), this is the format preferred by many GIS software including QGIS
- **Degrees, minutes, seconds**: coordinates written in DD° MM’ SS.S” format (e.g., 34° 44’ 04.5”)
- **Equator**: an imaginary line on Earth’s surface that is equal distance from the North and South poles and divides the planet into the Northern and Southern hemispheres
- **GPS**: the Global Positioning System, developed and maintained by the United States government, uses a satellite constellation for navigation on Earth’s surface
- **GPS receivers**: a processor that uses radio signals from satellites to calculate the user’s location (e.g., these receivers are found in handheld GPS devices and cellphones)
- **Interference**: when the radio signal from the satellite gets distorted by an object (e.g., a building, hard ground, mountains, heavy tree cover, etc.) before it reaches the GPS receiver

- **Latitude**: measurement north or south of Earth's equator
- **Longitude**: measurement east or west of the prime meridian
- **Pivot tables**: a table that aggregates values from a larger dataset (e.g., if you have a dataset of every health facility in a country with information about the administrative areas, you can create a pivot table of the number of health facilities in each administrative area)
- **Prime meridian**: the imaginary line chosen to represent 0° longitude that passes through Gao in Mali and Lake Volta in Ghana
- **Triangulation**: the process of using three intersecting circles to determine location
- **Workbook (Microsoft Excel)**: a Microsoft Excel file, often ending in the *.xlsx or *.xls file extension
- **Worksheet (Microsoft Excel)**: an individual page or sheet in a Microsoft Excel workbook

6.1.2 Introduction to GPS Data

Today, GPS receivers are quite common. Many of us even have them in our pockets right now. However, before discussing the GPS capabilities built into cellphones, we will begin by learning about standalone GPS receivers and the technology that powers them.

To learn how a GPS receiver works, it is important that you understand some of the science that make GPS possible.

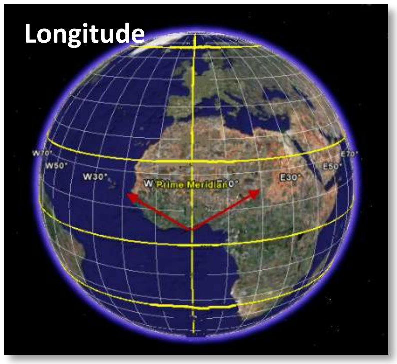


When measuring the location of objects on the surface of the Earth, we need a starting point. That starting point is where the equator and the prime meridian intersect. The equator is the line that runs horizontally across the earth dividing the planet into northern and southern hemispheres. The prime meridian is an imaginary line that runs vertically from the North Pole through Greenwich, England, and all the way down to the South Pole. The intersection of these two lines is the origin or starting point of the geographic coordinate system: 0, 0.

With 0, 0 in mind, let's review latitude and longitude.

Latitude (depicted in the lefthand image below) runs from 0 to 90 degrees from the equator to the North pole and 0 to -90 degrees from the equator to the South pole. Latitude lines measure how far north or south you are from the equator

Longitude (depicted in the righthand image below) runs 0 to 180° going East or 0 to -180° going west from the prime meridian. Longitude lines measure how east and west you are from the prime meridian.



© [Google Earth]

We record latitude and longitude in angular units because the Earth is a sphere or shaped like a ball. Combined, these two measurements provide a unique location on the surface of the Earth.

Latitude and longitude may be expressed in three main formats: (1) degrees, minutes, seconds, (2) degree and decimal minute, and (3) decimal degrees. This is similar to measuring distance, mass, or volume using different measurement systems. For example, we can say that a CubeSat, a miniature cube-shaped satellite, has a mass of 1.33 kilograms. We can also say it has a mass of 1,330 grams or even 2.9 pounds.

Below is a breakdown of all three latitude and longitude formats.

- DD MM SS.S – Degrees, minutes, seconds
- DD MM.MMM – Degree and decimal minute
- DDD.DDDDD – Decimal degree

An example of each is written below respectively.

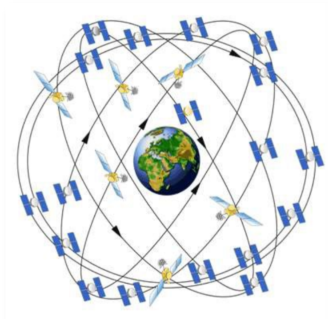
- 34° 44’ 04.5” – Degree, minutes, seconds
- 34° 44.0754’ – Degree and decimal minute
- 34.73459 – Decimal degrees

A best practice is to use the decimal degrees format unless otherwise instructed. It is the simplest format and easy to use. While decimal degrees is the preferred format, it is important that you recognize all three, as you may encounter the other two formats at some point.

6.1.3 How GPS Works

The Global Positioning System, or GPS, was originally developed by the US military as an aid to navigation and for defense purposes. The system took 16 years to develop and was fully operational in 1994.

The space segment of the GPS (depicted below) is made up of the 24 satellites and a few substitute satellites that are constantly orbiting Earth every 12 hours. The satellites are about 20,000 km above Earth’s surface. The satellites are organized in a way that ensures there are always at least 4 satellites above the horizon anywhere on Earth.

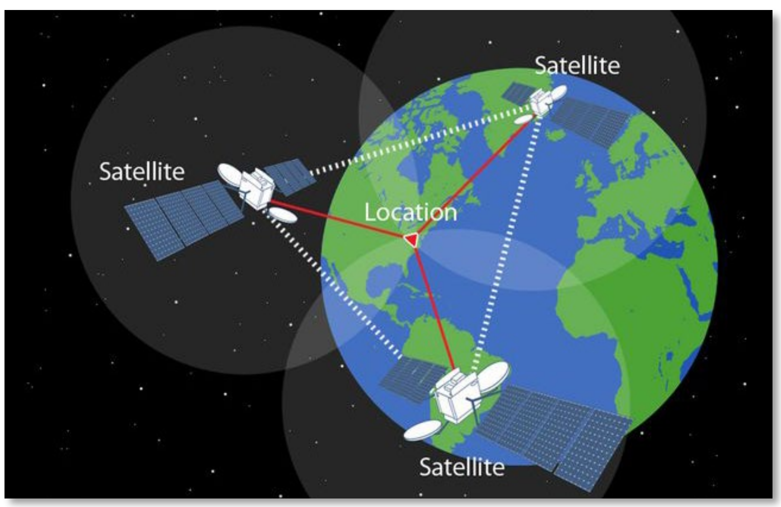


© Raghu N. / Research Gate

Today, the GPS orbits Earth alongside other satellite systems. The global constellation of satellites includes India's NavIC, Japan's Michibiki, China's BeiDou, the European Space Agency's Galileo, and Russia's GLONASS. These systems are used for many different purposes, such as agriculture, aviation, addressing natural disasters, ground transportation, recreation, surveying, defense, and even for improving health systems.

Modern GPS receivers are programmed to use signals from the space segment of the GPS and the other constellations of satellites, such as India's NavIC.

GPS works using a process called triangulation (depicted below). Each satellite continually transmits radio signals with its exact position and the exact time. Back here on Earth, the GPS receiver will compare the time when the signal from the satellite was sent with the time it was received. The location and timestamp are used to calculate distance from the satellite.



© National Geographic Society

The receiver uses the signals from 3 or more satellites to calculate its position. This is called triangulation. While 3 satellites are the minimum to calculate latitude and longitude, a 4th satellite is required to add elevation.

After the GPS unit calculates the intersection of the imaginary spheres it displays that data as latitude, longitude, and altitude which is collectively called a waypoint.

6.1.4 Types of GPS Receivers

There are several types of GPS receivers, each with their own strengths and weaknesses.

High-grade GPS receivers used for land surveys and engineering projects have sub-centimeter accuracy. While these tools are much more accurate than simpler tools, such as handheld GPS receivers, surveying equipment is expensive and cumbersome. It also requires advanced training and, in some cases, teams of people to operate. Such equipment might be necessary for construction projects, but it is not appropriate for collecting GPS data for household health surveys. (Pictured on the left is the [Topcon]() DT-305L theodolite.)

Traditionally, The DHS Program trained fieldworkers how to collect coordinates using handheld GPS receivers. These units are more compact than high-performance surveying equipment and are available at varying prices and levels of accuracy, though they are not as accurate as high-performance surveying equipment. While handheld GPS units are useful for collecting household or cluster locations, they are, in some cases, being replaced by GPS dongles. (Pictured on the left is a Garmin Etrex 10 handheld GPS.)

GPS dongles are increasingly useful because they can be attached to tablets and laptop computers. This can allow data collectors to geo-enable these other devices, as well as applications on those devices. For example, we can use a GPS dongle to provide coordinates for survey data that is being recorded directly into a data collection app. Due to this, health survey fieldworkers do not need to carry around GPS units, which weigh more and are more expensive. (Pictured on the left is a GlobalSat G-Star IV USB GPS dongle.)

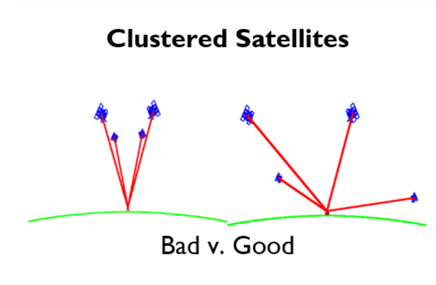
As mentioned in the beginning of this lesson, many of us have GPS receivers in our pockets. Cellphones can use a combination of GPS and cellphone towers to pinpoint locations on Earth and provide its user with coordinates. To access this information, there are some free smartphone applications that a user can download. We'll explore some of these a bit later in this lesson.

While different GPS receivers have different levels of accuracy, there are also several general factors that can impact the accuracy when collecting a waypoint. Some of these factors are outside of the user's control, but others are controllable.

Satellite clock errors occur when the GPS receiver's timing and the timing of the atomic clock on the satellite are different. This is an uncontrollable factor. The satellites have extremely accurate atomic clocks, but the receiver does not. If technicians equipped each receiver with an atomic clock, the receiver would cost over 100,000 USD. Orbital errors are another uncontrollable factor and can occur if a satellite is off any amount in its orbit or when there are different rounding systems being used between the satellite and the GPS unit.

There are two controllable factors that should be considered when measuring GPS coordinates. The first controllable factor is interference. Interference is when the radio signal from the satellite gets distorted by an object, such as a building, hard ground, mountains, heavy tree cover, or even a human body before it reaches the GPS receiver. The solution to this problem is to choose wisely when selecting the location to take a GPS waypoint. GPS waypoints should be taken in an area with a clear view of the sky. GPS receivers should be kept at least 5 meters away from any roofed structure and 10 meters away from any buildings taller than two stories when taking waypoints. The best location in an urban area is a street corner, nonreflective roof top, or other open area. In rural areas the location should be in a clearing outside of the tree canopy and far away from tall buildings.

When taking GPS waypoints, it is also important to be aware of the distribution of satellites (depicted below). This is the second controllable factor. Your unit can display the position in the sky of the satellites it sees. Usually, the satellites will be represented symbolically on a graphic user interface. If you notice that all the satellites are clustered in one portion of the screen, then there can be a significant error introduced to your coordinates. If this is the case, you should wait a few minutes for the satellites to reposition and the signal to improve.



As mentioned earlier, latitude and longitude can be expressed in three different formats: decimal degrees, degree and decimal minute, and degree, minute, seconds. Best practice is to use the decimal degrees format unless otherwise instructed.

This format may not be the default setting on a GPS receiver. If it is not, the operator will need to change the settings so that GPS coordinates are recorded and stored using this format.

6.1.5 GPS Receivers in Mobile Phones

What about the GPS receivers in our phones? Many mobile phones today have built-in GPS receivers. These GPS receivers work regardless of an internet or cellular connection. So long as you have your location services turned on, your phone should be able to tell you your location.

Like handheld GPS receivers, mobile phones use trilateration to provide the user with their location. However, some phones can also use an internet and/or cellular connection to improve their location services.

There are many mobile applications that use a mobile phone's location. A common example are social media apps that use the phone's location to enable certain content, tailor advertisements, and connect users to other people around them. Mobile phone camera apps often use GPS location. Many phones add geographic information (also called geotags) to photographs taken by the user.

While many of these mobile applications use a phone's location in the 'background'. There are some applications whose primary purpose is to use your location. The most common of these are navigation applications, such as Google Maps and Waze. But there are also mobiles apps built to collect GPS coordinates for use in GIS.

Some of the popular applications used to collect GPS coordinates include Geopaparazzi, Input, ODK, and QField. All four applications are free to use and available for Android phones.

- **Geopaparazzi** (osgeo.org/projects/geopaparazzi) is extremely lightweight, so it won't take up much space on your phone. It was built to conduct fast GIS data collection. However, Geopaparazzi was designed to work with gvSIG (a different GIS software), so you will need a plugin (such as IO Geopaparazzi) to export your waypoints into QGIS.

- **Input** (inputapp.io) is available for both Android and Apple iOS. Like QField, it was built to work with QGIS.

- **ODK** (getodk.org) is a robust survey form builder that can also collect GPS coordinates.

- **QField** (qfield.org) is built on top of QGIS, allowing users to setup maps and forms in QGIS on their workstation, and deploy those in the field.

While the use of these mobile GPS collection apps are not covered in this workshop, we highly encourage you to explore them on your own.

6.1.6 Data Entry Errors

Data entry can be challenging in a variety of ways. Although there are many ways to organize data entry, the process is time consuming. For projects with small budgets, this is one process that can prove to quite costly

Data entry is prone to human error. Therefore, professionals responsible for entering data must be properly trained. They should know how to format their entries, ensure quality, and navigate the software of choice. Which brings us to our final challenge: technology. Between software licenses and hardware limitations, technological disparities can pose many challenges to a team. Continuity and affordability should be taken into consideration whenever determining what technology can and will be used.

No matter how hard you try to prevent them, teams will ultimately face data entry errors. Understanding what potential errors could be, however, can help you mitigate, find, and address them.

- **Formatting errors**: The most preventable data entry errors are those related to entry form setup (depicted below). Proper databases are made up of a single row of data categories followed additional rows containing the data entered. To avoid errors here, a team should use a standard data entry form to ensure uniformity.

	A	B	C	D	E	F	G
1	ID	1114940	1114942	1114957	1114958	1114965	1115006
2	Name	Ravi River	Punjab Plains	Jhelum River	Hindustan	Basantar River	Sutlej River
3	Class	H	T	H	L	H	H
4	Code	STM	PLN	STM	RGN	STM	STM
5	Country	IN	IN	IN	IN	IN	IN
6	Elevation	133	206	147	344	[illegible]	100
7	Timezone	Asia/Kolkata	Asia/Kolkata	Asia/Kolkata	Asia/Kolkata	Asia/Kolkata	Asia/Kolkata
8	Date	6/6/2017	1/16/2012	2/3/2017	6/23/2006	3/9/2016	9/8/2016
9	Y	30.62123	30	31.16853	28	32.47452	29.34806
10	X	71.82683	75	72.15066	76	75.01449	71.02032

	A	B	C	D	E	F	G	H	I	J
1										
2	Places in India									
3	Identifiers		Group Information				GPS Data			
4	ID	Name	Class	Code	Country	Timezone	Date	Elevation	Y	X
5	1114940	Ravi River	H	STM	IN	Asia/Kolkata	6/6/2017	133	30.62123	71.82683
6	1114942	Punjab Plains	T	PLN	IN	Asia/Kolkata	1/16/2012	206	30	75
7	1114957	Jhelum River	H	STM	IN	Asia/Kolkata	2/3/2017	147	31.16853	72.15066
8	1114958	Hindustan	L	RGN	IN	Asia/Kolkata	6/23/2006	344	28	76
9	1114965	Basantar River	H	STM	IN	Asia/Kolkata	3/9/2016	300	32.47452	75.01449
10	1115006	Sutlej River	H	STM	IN	Asia/Kolkata	9/8/2016	100	29.34806	71.02032
11	1163293	Tithwal	P	PPL	IN	Asia/Kolkata	12/5/2015	1080	34.39351	73.77416

	A	B	C	D	E	F	G	H	I	J
1	ID	Name	Class	Code	Country	Timezone	Date	Elevation	Y	X
2	1114940	Ravi River	H	STM	IN	Asia/Kolkata	6/6/2017	133	30.62123	71.82683
3	1114942	Punjab Plains	T	PLN	IN	Asia/Kolkata	1/16/2012	206	30	75
4	1114957	Jhelum River	H	STM	IN	Asia/Kolkata	2/3/2017	147	[illegible]	72.15066
5	1114958	Hindustan	L	RGN	IN	Asia/Kolkata	6/23/2006	[illegible]	28	76
6	1114965	Basantar River	H	STM	IN	Asia/Kolkata	[illegible]	[illegible]	32.47452	75.01449
7	1115006	Sutlej River	H	STM	IN	Asia/Kolkata	[illegible]	100	29.34806	71.02032
8	1163293	Tithwal	P	PPL	IN	Asia/Kolkata	12/5/2015	1080	34.39351	73.77416
9	1163420	Thruti	P	PPL	IN	Asia/Kolkata	8/7/2018	1383	33.52682	74.15939
10	1163626	Thang	P	PPL	IN	Asia/Kolkata	9/6/2012	2726	34.9274	76.79336

- **Missing or incorrect data**: The most common error occurs when someone accidentally enters the wrong information. When searching for errors in data, it is important to look for typos, nonsensical values, missing data, and entries in the wrong format. If there is text where numbers should be and numbers where text should be, there is a problem. Some values, however, may look incorrect initially but are actually meant to stand out. For example, the number 9999 is often assigned where there is missing numerical data.

D	Name	Elevation
1114940	Ravi River	133
1114942	Punjab Plains	206
1114957	Jhelum River	
1114958	Hindustan	
1114965	Basantar River	300
1115006	Sutlej River	100
1163293	Tithwal	1080
1163420		1383
1163626	Thang	2726
1164949	Siachen Glacier	3625
1166907	Richhmar Gali	2383
1167718	Punch	
1169325	Dhoro Puran River	4
1171555	Mahmud Khaneke	184
1178984	Ganga Sagar	4
1179242	Ferozepore Headworks	185
1180957	Chumik	5625
1181619	Chalunka	3105

→

D	Name	Elevation
1114940	Ravi River	133
1114942	Punjab Plains	206
1114957	Jhelum River	9999
1114958	Hindustan	9999
1114965	Basantar River	300
1115006	Sutlej River	100
1163293	Tithwal	1080
1163420	NULL	1383
1163626	Thang	2726
1164949	Siachen Glacier	3625
1166907	Richhmar Gali	2383
1167718	Punch	9999
1169325	Dhoro Puran River	4
1171555	Mahmud Khaneke	184
1178984	Ganga Sagar	4
1179242	Ferozepore Headworks	185
1180957	Chumik	5625
1181619	Chalunka	3105

- **Accidental duplication**: When cleaning up data, it is also important you look for duplicate values and non-unique codes. Sometimes certain data can have duplicate values, but it is important you know if your dataset should. Often duplicate values are double counts that can throw off your results when the time comes to analyze the data. In most datasets, it is often important that each entry has its own unique identifier. These unique identifiers should have a standardized format across the dataset and will be used to set data entries apart from each other. Identifiers can also be used to join separate datasets together.

ID	Name
1114940	Ravi River
1114942	Punjab Plains
1114957	Jhelum River
1114958	Hindustan
1114965	Basantar River
1114965	Sutlej River
1163293	Tithwal
1163420	Thruti
1163626	Thang
1252653	Zunheboto
1252654	Zunheboto
1164949	Siachen Glacier

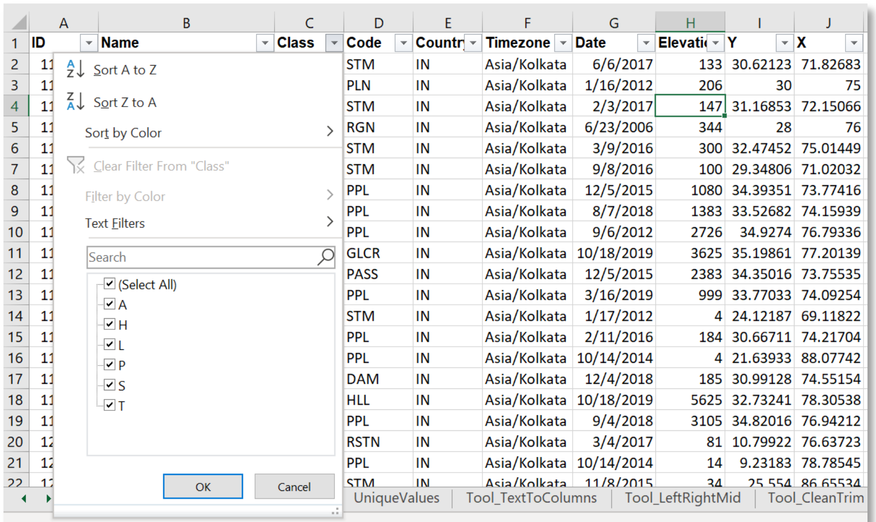
→

ID	Name
1114940	Ravi River
1114942	Punjab Plains
1114957	Jhelum River
1114958	Hindustan
1114965	Basantar River
1114967	Sutlej River
1163293	Tithwal
1163420	Thruti
1163626	Thang
1252653	Zunheboto
1252654	Zunheboto
1164949	Siachen Glacier

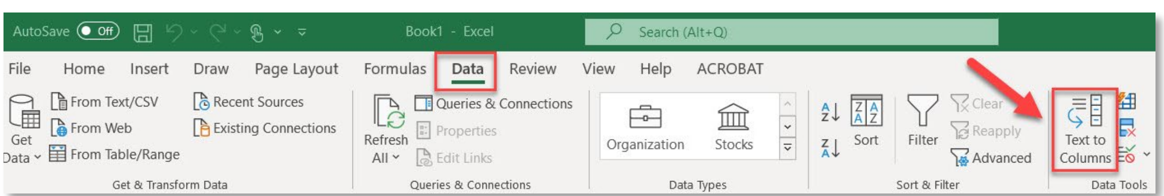
6.1.7 Tools for Cleaning Data in Excel

Before we start cleaning data in Excel, let's explore some of the tools we'll be using.

- **Filter and Sort**: Filtering and sorting data, especially numbers, can help identify outliers. Filter is best applied to data with only one column header and no blank rows between the data. When sorting, be sure that the sorting is applied to all the data so that the sorting does not change only some of the information in a row.



- **Text to Columns**: There are several helpful formulas that can be used in Excel to assist in the data cleaning process. Text to Columns can be used to split one column into several. This tool is found in the Data ribbon in the Data Tools section (shown below). This tool can also be used to change the format of numbers stored as text. If converting a number column stored as text you can click finish in the first step. To use Text to Columns, select one column at a time and follow the wizard steps.



- **Left, Right, and Mid**: Sometimes one column will contain multiple pieces of information without a delimiter. It is best to store these data in separate columns. The =LEFT(), =RIGHT(), and =MID() formulas allow you to extract these data systematically. As demonstrated in the screenshot below, these formulas allow you extract only the information you need while keeping the original data.

	A	B
1	Timezone	Name
2	Asia/Kolkata	=RIGHT(A2,7)
3	Asia/Kolkata	=RIGHT(A3,7)
4	Asia/Kolkata	=RIGHT(A4,7)
5	Asia/Kolkata	=RIGHT(A5,7)
6	Asia/Kolkata	=RIGHT(A6,7)
7	Asia/Kolkata	=RIGHT(A7,7)
8	Asia/Kolkata	=RIGHT(A8,7)
9	Asia/Kolkata	=RIGHT(A9,7)
10	Asia/Kolkata	=RIGHT(A10,7)
11	Asia/Kolkata	=RIGHT(A11,7)
12	Asia/Kolkata	=RIGHT(A12,7)
13	Asia/Kolkata	=RIGHT(A13,7)
14	Asia/Kolkata	=RIGHT(A14,7)
15	Asia/Kolkata	=RIGHT(A15,7)

	A	B
1	Timezone	Name
2	Asia/Kolkata	Kolkata
3	Asia/Kolkata	Kolkata
4	Asia/Kolkata	Kolkata
5	Asia/Kolkata	Kolkata
6	Asia/Kolkata	Kolkata
7	Asia/Kolkata	Kolkata
8	Asia/Kolkata	Kolkata
9	Asia/Kolkata	Kolkata
10	Asia/Kolkata	Kolkata
11	Asia/Kolkata	Kolkata
12	Asia/Kolkata	Kolkata
13	Asia/Kolkata	Kolkata
14	Asia/Kolkata	Kolkata
15	Asia/Kolkata	Kolkata

- **Clean and Trim**: If data are copied from another source, like Word, there could be special, non-printable characters like bullets (•). The =CLEAN() formula removes these characters leaving only printable characters. Text in cells should not contain "extra" spaces. This means that there should be one space between words and no spaces before the first word or after that last word. The =TRIM() formula removes these "extra" spaces. As depicted in the screenshot below, the =CLEAN() formula should be run first because it may leave extra spaces.

	A	B	C	D
1	Raw	Cleaned	Trimmed	
2	Ravi River	=CLEAN(A2)	=TRIM(B2)	
3	Punjab Plains	=CLEAN(A3)	=TRIM(B3)	
4	Jhelum River	=CLEAN(A4)	=TRIM(B4)	
5	▯Hindustan	=CLEAN(A5)	=TRIM(B5)	
6	Basantar River	=CLEAN(A6)	=TRIM(B6)	
7	Sutlej River	=CLEAN(A7)	=TRIM(B7)	
8	Tithwal	=CLEAN(A8)	=TRIM(B8)	
9	Thang	=CLEAN(A9)	=TRIM(B9)	
10	Siachen Glacier	=CLEAN(A10)	=TRIM(B10)	
11	Richhmar Gali	=CLEAN(A11)	=TRIM(B11)	
12	Punch	=CLEAN(A12)	=TRIM(B12)	
13	Dhoro Puran River	=CLEAN(A13)	=TRIM(B13)	
14	Olavakod Junction	=CLEAN(A14)	=TRIM(B14)	
15	▯Keelakarai	=CLEAN(A15)	=TRIM(B15)	
16	▯Kamla Nadi	=CLEAN(A16)	=TRIM(B16)	
17	Dhauliganga River	=CLEAN(A17)	=TRIM(B17)	
18	▯Kauriala Ghat	=CLEAN(A18)	=TRIM(B18)	
19	Baranai Nadi	=CLEAN(A19)	=TRIM(B19)	

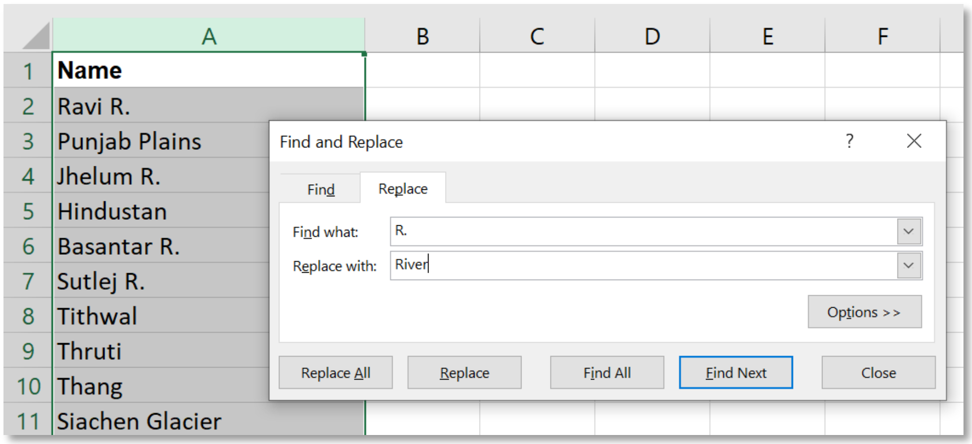
	A	B	C
1	Raw	Cleaned	Trimmed
2	Ravi River	Ravi River	Ravi River
3	Punjab Plains	Punjab Plains	Punjab Plains
4	Jhelum River	Jhelum River	Jhelum River
5	▯Hindustan	Hindustan	Hindustan
6	Basantar River	Basantar River	Basantar River
7	Sutlej River	Sutlej River	Sutlej River
8	Tithwal	Tithwal	Tithwal
9	Thang	Thang	Thang
10	Siachen Glacier	Siachen Glacier	Siachen Glacier
11	Richhmar Gali	Richhmar Gali	Richhmar Gali
12	Punch	Punch	Punch
13	Dhoro Puran River	Dhoro Puran River	Dhoro Puran River
14	Olavakod Junction	Olavakod Junction	Olavakod Junction
15	▯Keelakarai	Keelakarai	Keelakarai
16	▯Kamla Nadi	Kamla Nadi	Kamla Nadi
17	Dhauliganga River	Dhauliganga River	Dhauliganga River
18	▯Kauriala Ghat	Kauriala Ghat	Kauriala Ghat
19	Baranai Nadi	Baranai Nadi	Baranai Nadi

- **Concatenate**: The Concatenate function takes text strings and joins them together into one text string. You can specify strings by their cell reference, or by enclosing a bit of text within quotes ("text"). This function will accept number values, but it is often better to use the =TEXT function first on numbers in order to control the text length that will be outputted.

	A	B	C	D	E
1	Country	Class	ID_Old	ID_Text	ID_New
2	IN	H	1114940	=TEXT(C2,"00000000")	=CONCAT(A2,B2,D2)
3	IN	T	1114942	=TEXT(C3,"00000000")	=CONCAT(A3,B3,D3)
4	IN	H	1114957	=TEXT(C4,"00000000")	=CONCAT(A4,B4,D4)
5	IN	L	1114958	=TEXT(C5,"00000000")	=CONCAT(A5,B5,D5)
6	IN	H	1114965	=TEXT(C6,"00000000")	=CONCAT(A6,B6,D6)
7	IN	H	1115006	=TEXT(C7,"00000000")	=CONCAT(A7,B7,D7)
8	IN	P	1163293	=TEXT(C8,"00000000")	=CONCAT(A8,B8,D8)
9	IN	P	1163420	=TEXT(C9,"00000000")	=CONCAT(A9,B9,D9)

	A	B	C	D	E
1	Country	Class	ID_Old	ID_Text	ID_New
2	IN	H	1114940	01114940	INH01114940
3	IN	T	1114942	01114942	INT01114942
4	IN	H	1114957	01114957	INH01114957
5	IN	L	1114958	01114958	INL01114958
6	IN	H	1114965	01114965	INH01114965
7	IN	H	1115006	01115006	INH01115006
8	IN	P	1163293	01163293	INP01163293
9	IN	P	1163420	01163420	INP01163420

- **Find and Replace**: Spreadsheets often contain data that are not necessary. An easy way to remove extraneous information is through find and replace. Use this tool if you want to remove or replace entire words, phrases, or characters. In the example below, the Excel user is replacing the text “R.” with the word “River” in every cell of column A.



6.2 EXCEL AND QGIS INSTRUCTIONS

6.2.1 Cleaning Data in Excel

Here, you will learn the fundamental tools and techniques for cleaning data in Microsoft Excel. As this is a longer activity, the content is broken into 8 parts: (1) reformatting a workbook, (2) converting text to columns, (3) extracting characters from text, (4) replacing text, (5) cleaning and trimming text, (6) concatenating values, (7) comparing cells using formulas, and (8) filtering and sorting.

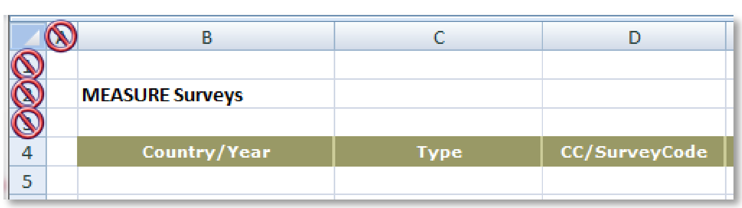
This section corresponds to Activity 6.3 of the Health Data Mapping Workshop. The exercise file for this section can be found in the \DHS_QGIS_2022\Exercise_Files\6.2.1 _CleaningData\ folder. It’s called PracticeData.xlsx.

Please note, the screenshots in this section are from an older version of Microsoft Excel. Your software will likely appear different, though the same options should be present.

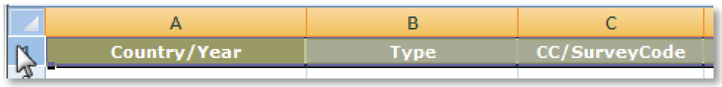
Part 1. Reformatting the workbook

To begin, we will remove the extra rows and columns from our spreadsheet. We’ll also adjust the column names so that each column has only one cell in Row 1 that will serve as the column name.

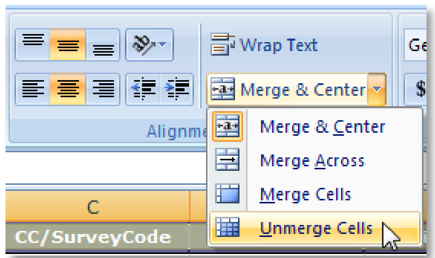
1. Save a copy by clicking File → Save a Copy. Save the file as PracticeData_tidy.xlsx
2. Delete Rows 1-3 and Column A by right clicking in the row number and column header and selecting Delete.



3. Select Row 1 by left clicking in the row number and highlighting the row.



4. Click the arrow next to Merge and Center in the Home menu. Select Unmerge Cells from the drop-down list.



5. Copy cells H2, I2, and J2 (SurveyData, GPSData, Biomarkers) by selecting the cells and pressing CTRL + C on your keyboard.

6. Click on cell H1 and press CTRL + V to paste the cells in H1, I1, and J1.

Remember, clean data consists of one row per record and a single column name. Extra rows and columns create problems for sorting, filtering, and pivoting. These tables may look nice when printed or viewed by a human, but a computer will not understand data in this format.

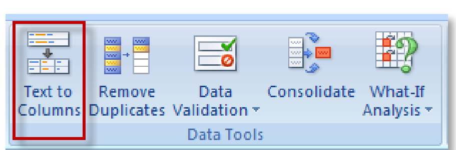
Part 2. Converting text to columns

As mentioned in Section 6.1.7, the Text to Columns tool can split one column into several using a delimiter (e.g., comma, semicolon, tab, etc.). This tool will also store the data with the proper format. For example, numbers stored as text will be converted to numbers. In this example, the country name and year(s) of the survey are stored in one column (Country/Year) and years are numbers stored as text. Numbers stored as text will restrict the ability to sort and filter properly but will not affect mathematical formulas in Excel. QGIS will not recognize numbers that are stored as text. Split the column by following the steps below.

7. Insert a new column to the right of column A by right clicking on the column heading of Type (Column B) and selecting Insert.

8. Select the entire Country/Year column by left clicking the header.

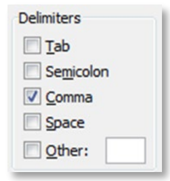
9. Click the Text to Columns tool in the Data ribbon.



10. Select Delimited in the first window in the Convert Text to Columns Wizard.

11. Click Next.

12. Check Comma and uncheck Tab in the Delimiters section.



13. Click Next.

14. Select General under Column data format section.

15. Click Finish.

16. If you receive a warning "Do you want to replace the contents of the destination cells?" click OK.

17. Enter new column names – "Country" for Column A and "Year" for Column B.

Text to Columns can also convert data stored in one format into another format. For example, numbers could be stored as text. Examine Columns C and D (Type and CC/SurveyCode). You may see the green arrow in the upper left corner of each cell. Clicking on one of these green arrows should give you a warning message of "Number Stored as Text".

Part 3. Extracting characters from text

Sometimes one column will contain multiple pieces of information. For the purposes of sorting or filtering, you may need these pieces of data in separate, distinct columns. The LEFT and RIGHT formulas allow you to extract these data systematically. In this example, the survey code is combined with the year of the survey. You will separate the information and create a new column for the survey code.

18. Insert a new blank column by right clicking on CC/SurveyCode (Column D) and selecting Insert.

19. In D2 type the formula =RIGHT(E2,3).

Within the parentheses, the first variable specifies the cell that will be shortened. The second variable is the number of characters from the right. In this case, we are extracting the first three characters starting on the right from CC/SurveyCode. These three characters make up the Survey Code.

20. Fill this formula down the entire column either by dragging the bottom right corner of D2 or double clicking on the bottom right corner once the cursor becomes a black cross.

21. Name Column D SurveyCode.

22. Delete the entire CC/SurveyCode column.

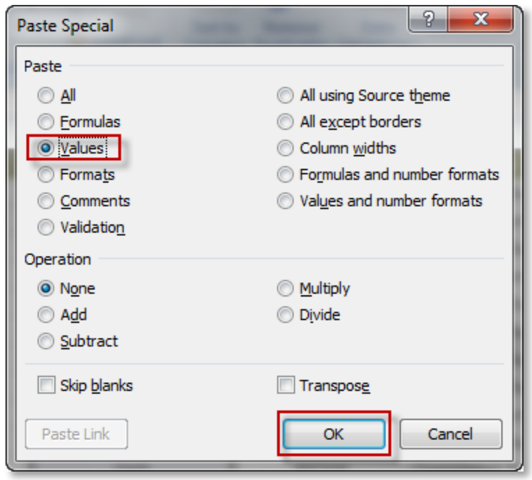
What happened? You should see a "#REF!" error in the new column. This is because the formula was referencing the column you deleted. To fix this, you must copy and paste the formula results using Paste Special.

23. Click press CTRL + Z to undo the previous step. The CC/SurveyCode column should now reappear.

24. Click the column D heading to select the entire column and press CTRL + C to copy.

25. Right-click on the column D heading and select Paste Special.

26. Select Values (highlighted in the screenshot below) to paste the results of the formula and click OK.



27. Insert a new column in between SurveyCode and CC/SurveyCode columns.

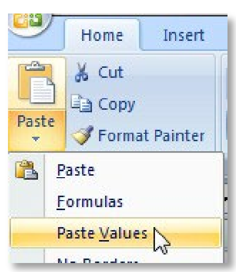
28. In E2 type the formula =LEFT(F2,2).

Within the parentheses, the first variable specifies the cell that will be shortened. The second variable is the number of characters from the left. In this case, we are extracting the two characters starting on the left from CC/SurveyCode. These two characters make up the Country Code.

29. Fill this formula down the entire column either by dragging the bottom right corner of E2 or double clicking on the bottom right corner once the cursor becomes a black cross.

30. Click the column E heading to select the entire column and press CTRL + C to copy.

31. Click the arrow below the Paste button within the Home Ribbon and select Paste Values.



32. Enter a name for this new column in E1 as CtryCode.

33. Delete the CC/SurveyCode column.

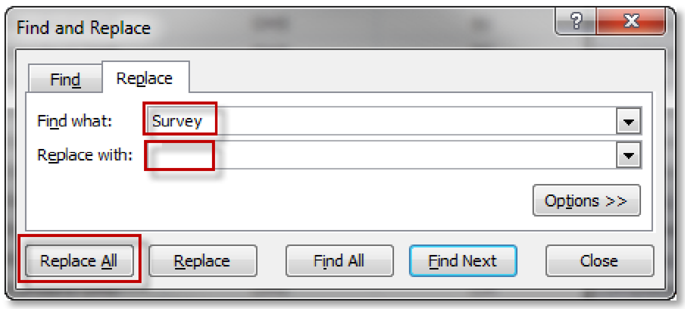
Part 4. Replacing text

Spreadsheets often contain data that are not necessary. An easy way to remove extraneous information is through find and replace. In this example, we will remove the suffix "Survey" from the Status column.

34. Left click on the Status column header to select the entire column.

35. Press CTRL + F to open the Find and Replace tool.

36. Enter Survey in the 'Find what:' box. Leave the 'Replace with:' box blank and click Replace All. This is shown in the screenshot below.



By leaving the replace with box empty we are essentially deleting the word "Survey" from the column as well as the trailing space.

37. Close the Find and Replace window and examine the Status column now. What do you see? You'll see that the word "Survey" has been removed from all cells.

Part 5. Cleaning and trimming text

If data is copied from another source, like Word, there could be special, non-printable characters like bullets (•). The =CLEAN() formula removes these characters leaving only printable characters. Trailing, leading or extra spaces cause problems when sorting data or when matching on an entire cell's content. The =TRIM() formula removes these spaces leaving a single space between words.

38. Insert a new, blank column to the right of Type.

39. In D2 type the formula =CLEAN(C2).

40. Fill this formula down the entire column either by dragging the bottom right corner of D2 or double clicking on the bottom right corner once the cursor becomes a black cross.

If there are extra spaces, they will remain. The =TRIM() formula will remove them.

41. Click the column D heading to select the entire column and press CTRL + C to copy.

42. In the same column, right click on the column heading and select Paste Special.

43. Select Values to paste the results of the formula.

44. Delete the Type column (Column C).

45. Enter Type as the name for the new column in C1.

46. Insert a new, blank column to the left of Column B.
47. In B2 type the formula =TRIM(A2)
48. Fill this formula down the entire column either by dragging the bottom right corner of C2 or double clicking on the bottom right corner once the cursor becomes a black cross.
49. Click the new column heading to select the entire column and press CTRL + C to copy.
50. In the same column, right-click on the column heading and select Paste Special.
51. Select Values to paste the results of the formula.
52. Now, rename Column B as Country and delete Column A.

Part 6. Concatenating

We can create a unique survey ID by combining Country Code, Survey Code, and Year using the concatenate formula. The concatenate formula separates each cell reference with a comma (,).

53. Insert a new, blank column between the CtryCode and Status columns.
54. In F2 type the formula =CONCATENATE(E2,"-",D2,"-",B2).

The comma (,) separates cell references that will be combined in the new cell. Insert text by using quotes (") before and after the text. In this example we are combining cells E2, D2, and B2 and inserting a dash (-) between each.

55. Fill this formula down the entire column either by dragging the bottom right corner of F2 or double clicking on the bottom right corner once the cursor becomes a black cross.
56. Name column F SurveyID.

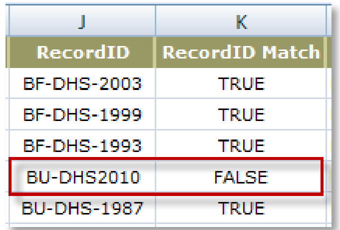
Part 7. Comparing using formulas

Comparing data within your file is a useful way to examine the quality of the data. For example, the newly created SurveyID should match the RecordID from the original data. If they don't match, this could indicate a problem with the data or with our data cleaning process.

The formula =[CELL REFERENCE 1]=[CELL REFERENCE 2] checks for an exact match between two cells.

57. Insert a new, blank column between RecordID (Column J) and SurveyData (Column K).
58. In K2, type =F2=J2.
59. Fill this formula down the entire column either by dragging the bottom right corner of R2 or double clicking on the bottom right corner once the cursor becomes a black cross.
60. Name the new column "RecordID Match".
61. Look down RecordID Match. Do any display “FALSE”? You’ll see that the record for the 2010 Burundi DHS has a value of “FALSE”.

62. Check the record for the Burundi 2010 DHS. There is a dash missing from the BU-DHS-2010 RecordID. Select the Burundi BU-DHS-2010 Record ID cell and add the dash between the “S” and the “2”.



J	K
RecordID	RecordID Match
BF-DHS-2003	TRUE
BF-DHS-1999	TRUE
BF-DHS-1993	TRUE
BU-DHS2010	FALSE
BU-DHS-1987	TRUE

The text in the cell for the Burundi 2010 DHS RecordID Match column should change to TRUE once the dash has been added.

63. Delete the columns RecordID (Column J) and RecordID Match (Column K).

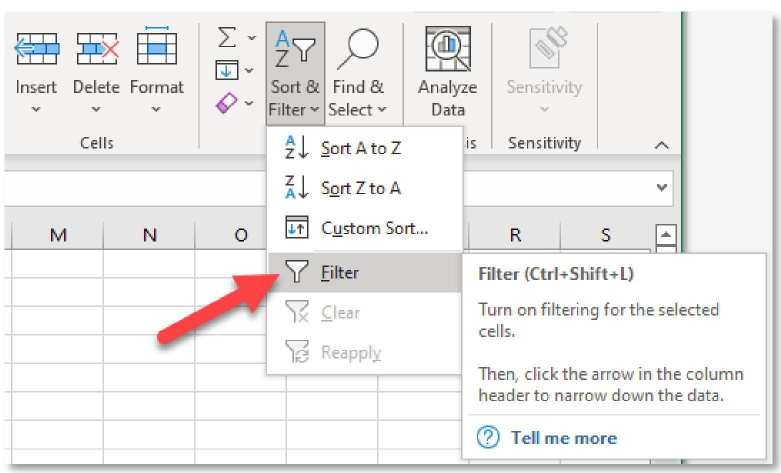
Part 8. Filtering and sorting

Once the obvious errors have been cleaned, we can begin to dig deeper and examine the internal consistencies of the data. This will help us spot other errors that may not be visible on the first look. For this, we will use the filter and sort functions. The data must be properly formatted for these functions to produce usable results. For example, numbers must be stored as numbers, not text.

64. Select Columns A-L by left clicking Column A's header, holding the left mouse button, and dragging to the right until you get to Column L. Or click anywhere in the table and press CTRL + A to select all cells in the table.

You must select all data within the worksheet before sorting. If you do not select all data, the unselected cells will be become separated from the proper record during the sorting process, and cells will be assigned to the wrong record.

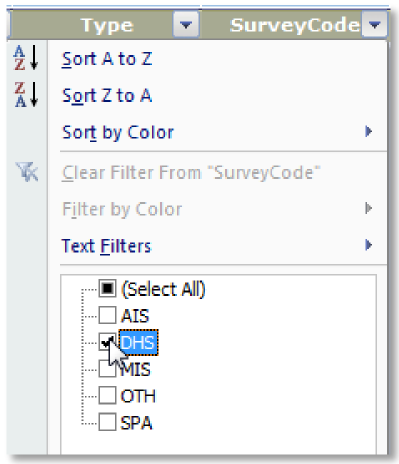
65. On the Home tab of Excel’s menu bar, click Sort & Filter → Filter (shown below).



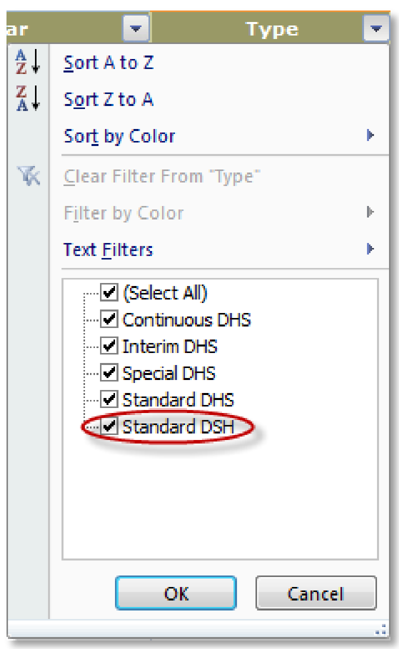
66. Click on the grey down arrow in the SurveyCode column (Column D).

67. Uncheck the Select All box.

68. All of the boxes should now be unchecked. Click on the box next to DHS to recheck it. Click OK.



69. Now you should only see different types of DHS surveys. Click on the grey down arrow in the Type column to verify that all of the records are a type of DHS survey. Do you notice anything wrong? You should see all DHS survey types, but the last entry on the list is misspelled.



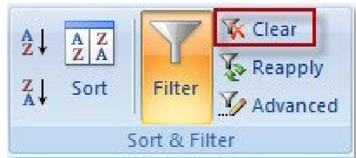
70. Uncheck the Select All box.

71. Recheck the misspelled "Standard DSH" box.

72. Click OK.

73. Correct the spelling of this record to "Standard DHS".

74. Click the Clear button in the Sort & Filter section of the Data tab to clear the filters and show all the survey records.



75. Save your work and close Excel.

6.2.2 Crosswalk Tables in Excel

Crosswalk tables are used to bring together different data files that have data at the same level of aggregation. In this activity, we will be enriching one worksheet, Survey_Totals, with data from two other worksheets.

This section corresponds to Activity 6.4 in the Health Data Mapping Workshop. The exercise file for this section can be found in the \DHS_QGIS_2022\Exercise_Files\6.2.2 _CrosswalkTables\ folder. It's called CrosswalkData.xlsx.

1. Open CrosswalkData.xlsx in Excel.

2. Look at the data in each worksheet GIS_data, Survey_Totals, and CountryCodes. What kind of data is in each worksheet? Do the country codes in each worksheet match those in the other worksheets? The Survey_Totals worksheet contains the number of DHS Program surveys implemented in each country. The CountryCodes worksheet contains the different country codes used by organizations to abbreviate national names. Lastly, the GIS_data worksheet has some basic variables about different countries, including the region each country is in.

We want to enrich our Survey_Totals worksheet with the DHS country codes found in the CountryCodes worksheet and the region names from the GIS_data worksheet. To do this, we will create a crosswalk table with the =VLOOKUP() function. The Excel function, =VLOOKUP(), is used to connect two worksheets within the same workbook. The formula for this function is written below.

`=VLOOKUP(lookup_value, table array, col_index_num, [range_look_up])`

`lookup_value` is the cell reference that contains the information you are looking up. **`table array`** is the place where the information you are matching to is located. Make sure the first column of the table array contains the same information as the lookup_value. **`col_index_num`** is the column number (starting with the lookup_value/first column) that you want returned to your table. **`range_look_up`**, enter FALSE or 0 to indicate that you want the match to be made exactly letter for letter/number for number.

3. Open the Survey_Totals worksheet by clicking the tab.

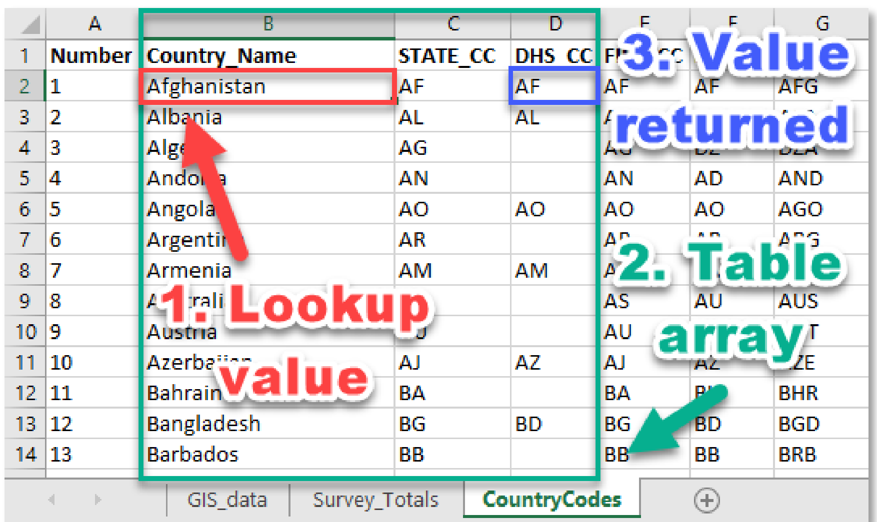
4. Insert a new column between the Country (Column A) and AIS (Column B) columns by right-clicking on Column B and selecting Insert.

5. Column B will contain the country codes used by The DHS Program. So, enter the text, “DHS_CC”, in cell B1.

6. In cell B2, we’re going to use the =VLOOKUP() function to pull a country code from the CountryCode worksheet using the country’s name in Column A. So, enter the formula written below in cell B2. Then hit the Enter button.

=VLOOKUP(A2,CountryCodes!B:D,3,FALSE)

This formula commands Excel to look up the value from cell A2 (Afghanistan) within the Survey_Totals worksheet. Excel will look for Afghanistan in column B of the Country Codes sheet, because column B is the first column of your specified Table Array. Once Excel finds the value Afghanistan in column B of the Country Codes sheet, it will return the value listed in the 3rd column (Column D) of the table array (columns B to D) from the Afghanistan record. The range lookup value of FALSE specifies that Excel will only return values for exact matches between the lookup value and the first column of the table array.

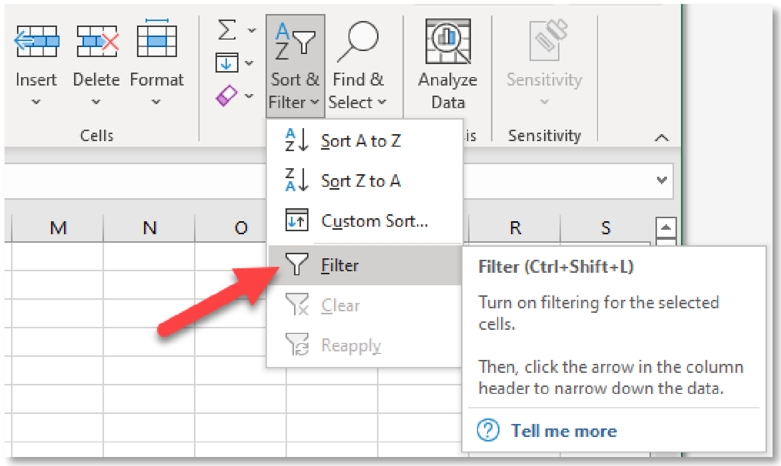


This formula should return the value, “AF”, in cell B2.

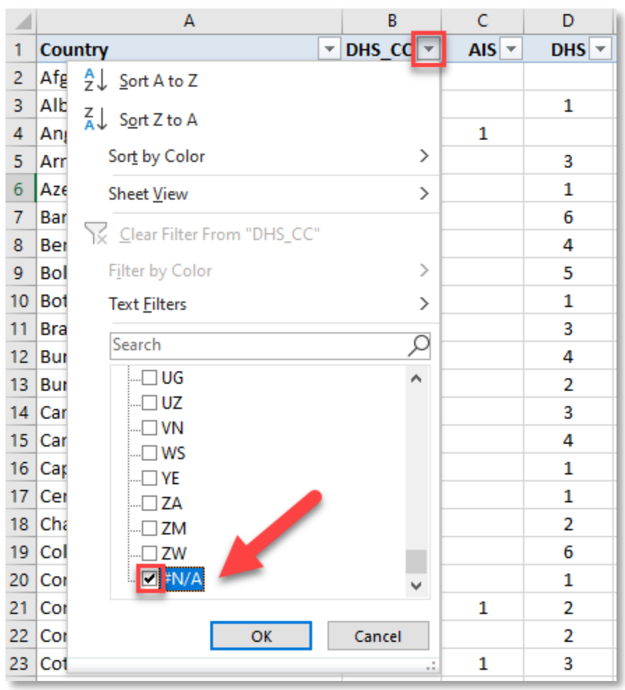
7. Click B2, then move the mouse to the bottom right corner, when the mouse changes to a small **+** double click. This will fill in the formula down to the bottom of your table (as long there are no cells to the immediate left that contain no data).

After you have copied down the formula, most of the cells in Column B should now contain a DHS country code. However, you will see some cells where the formula didn’t work. To easily identify which cells don’t have a DHS country code, we can use filters. To add a filter to the dataset,

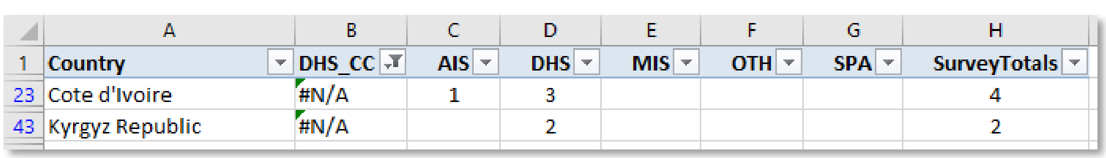
1. To add a filter, select the first row of the worksheet. Then, on the Home Tab of Excel’s menu bar, click Sort & Filter → Filter (shown below).



8. Click on the grey arrow in the DHS_CC column. Uncheck the box next to ‘Select All’, scroll to the bottom of the list, and check the box next to ‘#N/A’ (shown below). Click OK.



This filters the results and shows there are 2 countries that did not match.



	A	B	C	D	E	F	G	H
1	Country	DHS_CC	AIS	DHS	MIS	OTH	SPA	SurveyTotals
23	Cote d'Ivoire	#N/A	1	3				4
43	Kyrgyz Republic	#N/A		2				2

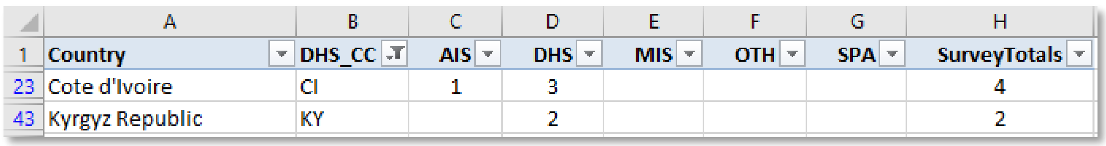
9. Examine the “#N/A” rows that did not match.

10. Look at cell A23 (Cote d’Ivoire) in the Survey_Totals worksheet, and then look at cell A42 (Côte d'Ivoire) in the CountryCodes worksheet.

This is an example of why making a match on a name is difficult, since slight difference in spelling, accents, or spaces can make the match invalid.

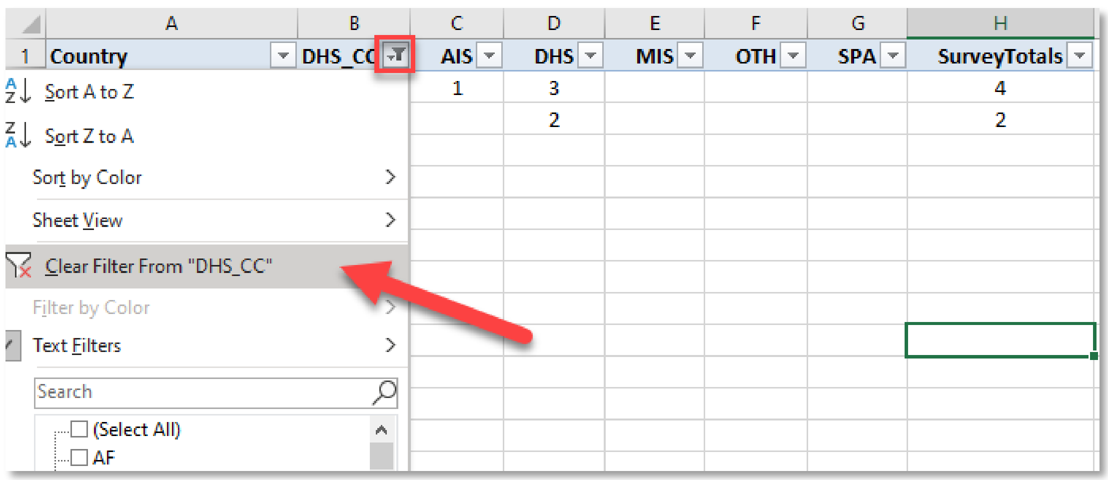
11. Find the name equivalent in the CountryCodes sheet for each non-matched country and copy the DHS_CC for that country and insert it into the DHS_CC column in the Survey_Totals sheet.

Your results should look like this.



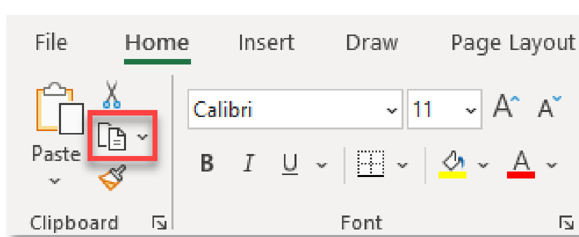
	A	B	C	D	E	F	G	H
1	Country	DHS_CC	AIS	DHS	MIS	OTH	SPA	SurveyTotals
23	Cote d'Ivoire	CI	1	3				4
43	Kyrgyz Republic	KY		2				2

12. Clear your filters by clicking on the filter button in Column B and clicking Clear Filter from “DHS_CC” (shown below).



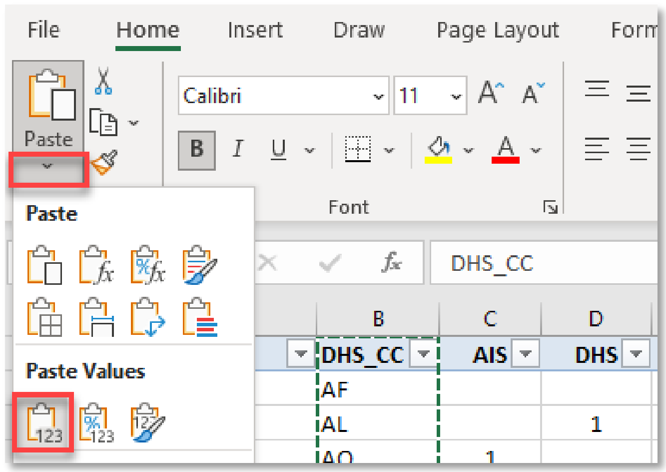
Now, Column B contains all the DHS country codes. However, some of the cells in this column actually contain formulas instead of the actual values. It’s best practice to replace formulas with actual values if we plan on using this dataset in QGIS. So, we’re going to copy the content in Column B and paste it in the same column but as values.

13. Select all of the contents of Column B by clicking on B above the column. Copy the content by clicking the copy button in the Home Tab (shown below) or by using the keyboard shortcut, Ctrl+C.

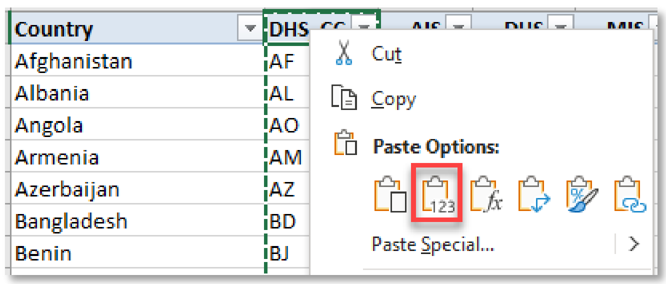


14. Now, paste the content as values (instead of formulas). This can be done in three ways:

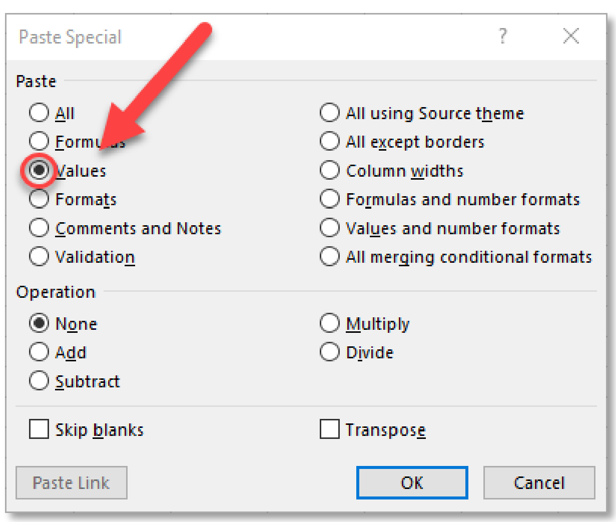
a. The first method is to select column B and click the Paste → Paste Values in the Home Tab (shown below).



b. The second method is to right-click in cell B1 and click the Paste Values button (shown below).



c. The final method is to select cell B1 and use the keyboard shortcut, Ctrl+Alt+V, to access the Paste Special menu. Then, select the button next to 'Values' and click OK (shown below).



We now want to use the =VLOOKUP() function to pull values from the GIS_data worksheet. Specifically, we want the region names for the countries added to the Survey_Totals worksheet.

15. In the Survey_Totals worksheet sheet, insert a column between DHS_CC (Column B) and AIS (Column C). Enter the text, "Region", in cell C1.

16. In cell C2, use the =VLOOKUP() function to extract the region for the country based on the DHS country code in cell B2.

=VLOOKUP(B2,GIS_data!C:D,2,FALSE)

This formula will look up the value in Column B of the Survey_Totals sheet, find the match in Column C of the GIS_data sheet, and return the corresponding value from Column D (the second column from the table array) in the GIS_data worksheet. The function will only return values where the lookup value (Column B of the Survey Totals sheet) is an exact match to a record within the first column of the Table Array (Column C of the GIS_data sheet).

17. Click C2, then move the mouse to the bottom right corner, when the mouse changes to a small **+** double click or click and drag down to the bottom of the table.

18. Examine Column C. All of the countries should have properly matched.

19. Copy Column C and paste the values in Column C using one of the methods describes in **Step 14** above.

20. Click Save.

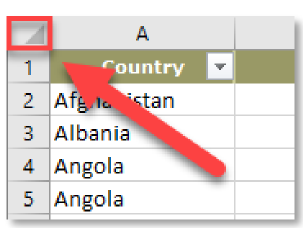
6.2.3 Pivot Tables in Excel

A pivot table is one way to summarize data. It allows us to add together data for similar groups. For example, you may have survey data but you would like to see the country totals. Similarly, you may have multiple rows per survey code but you would like to see the data at the code level. In this example we will aggregate survey records for each country.

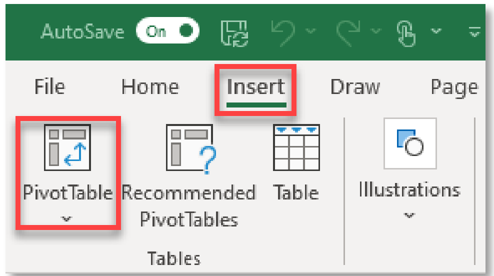
This section corresponds to Activity 6.5 in the Health Data Mapping workshop. The exercise file for this section can be found in the \DHS_QGIS_2022\Exercise_Files\6.2.3 _PivotTables\ folder. It's called PivotData.xlsx.

1. Open the Excel file, PivotData.xlsx.

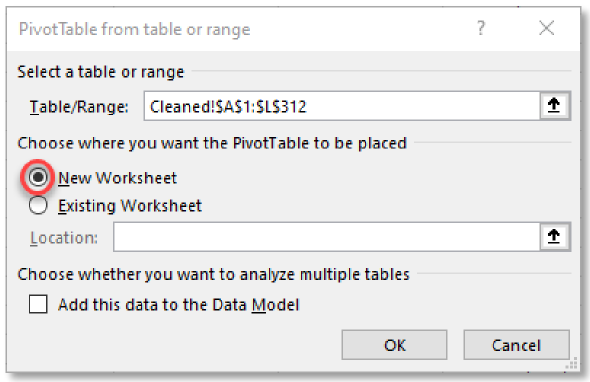
2. Select all the data in the worksheet by clicking on the upper left, unnumbered, un-lettered square (shown below). You can also select all of the cells by clicking on one populated cell in the worksheet and using the keyboard shortcut Ctrl+A.



3. Open the Insert Tab and click the PivotTable button (shown below).

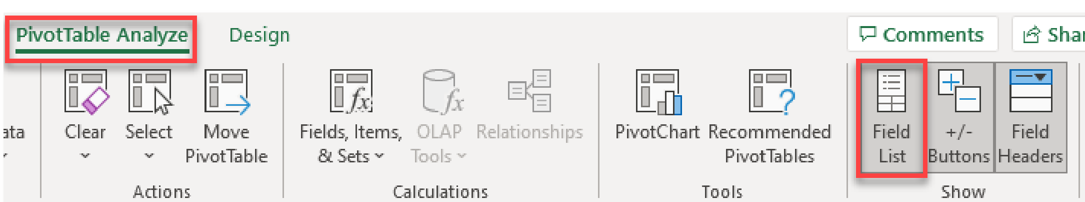


4. Make sure the button next to 'New Worksheet' is selected and his OK.

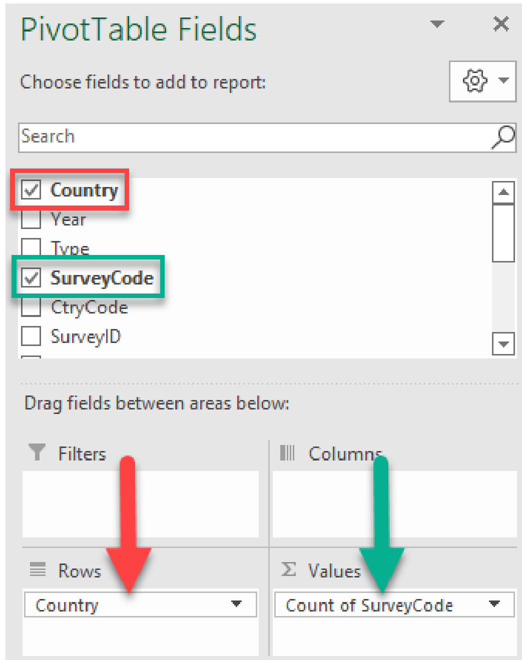


This will create a new worksheet containing the pivot table. The pivot table will reference data in the original worksheet so that any changes to the original data will be reflected in the pivot table. If the pivot table was already created, you will have to click Refresh in the PivotTable Tools, Data menu or press Alt+F5 for the pivot table to update.

The view will automatically switch to the new sheet called Sheet1 with the pivot table. On the right side of the page will be the PivotTable Field List that will allow you to customize your pivot table. If you do not see the Pivot Table Field List, go to the PivotTable Analyze Ribbon and select the Field List icon from the Show Section (shown below).



5. In the PivotTable Field List, click and drag the Country field to the Row Box and SurveyCode field to the Values Box (shown below).



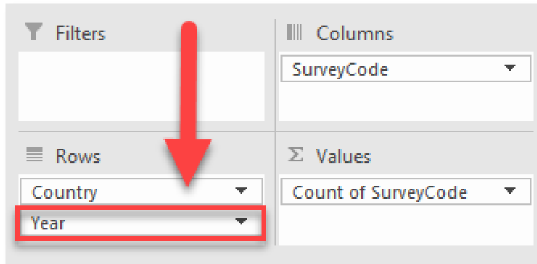
Notice that the data in Column B are aggregated as a count. There are other options, but this is the best choice for these data.

6. Drag the SurveyCode field to the Columns Box.

You should see a table like the one below. Each row is one country. Each column is one survey code.

	A	B	C	D	E	F	G
1							
2							
3	**Count of SurveyCode**	**Column Labels**					
4	**Row Labels**	**AIS**	**DHS**	**MIS**	**OTH**	**SPA**	**Grand Total**
5	Afghanistan				1		1
6	Albania		1				1
7	Angola	1		2			3
8	Armenia		3				3
9	Azerbaijan		1				1
10	Bangladesh		6		1	1	8
11	Benin		4				4
12	Bolivia		5				5
13	Botswana		1				1

7. Drag the field Year to the Rows Box. Be sure to place the field below Country (shown below).

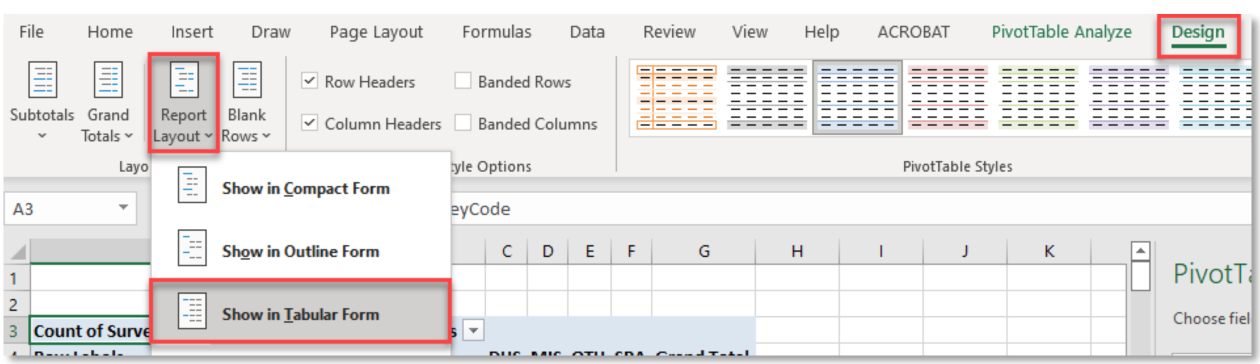


You should see a table like the one below. Now each survey is grouped by country in the rows and the survey code is displayed in each cell by year.

Count of SurveyCode	Column Labels					
Row Labels	AIS	DHS	MIS	OTH	SPA	Grand Total
⊟Afghanistan				1		1
2010				1		1
⊟Albania		1				1
2008		1				1
⊟Angola	1		2			3
2006			1			1
2010	1					1
2011			1			1

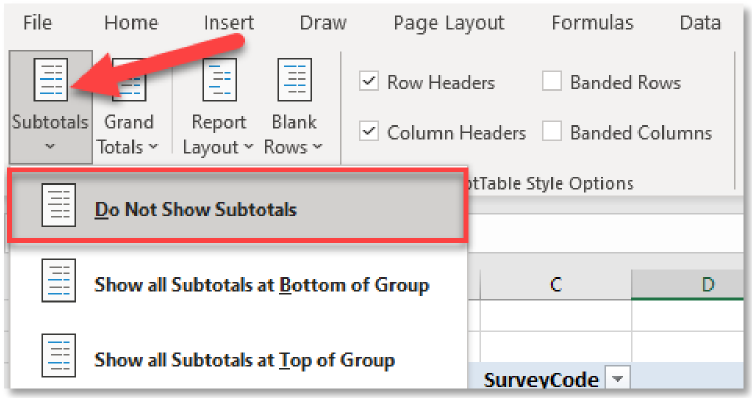
The default pivot table layout is to nest field information within rows. In the table you created, the survey years for a country are nested within the same column as the country name. This formatting is easy for people to read but is not considered to be a clean Excel sheet since the type of information contained in Column A varies by row (country name versus survey year). We will reformat the pivot table, so each column contains only one type of information.

8. Click on any cell within the pivot table. Then, click the Design Tab and select Report Layout → Show in Tabular Form (identified below).

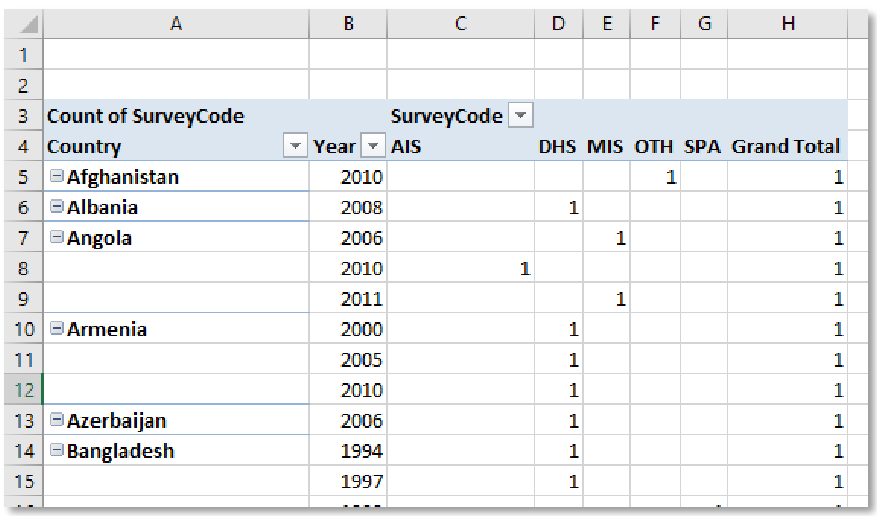


Your table should now have a separate column for the Country name (Column A) and the Survey Year (Column B). However, a subtotal for each country is still displaying within Column A.

9. Click the Subtotals icon and select Do Not Show Subtotals.



Your table should look similar to this example, with a separate column for each data type, but not displaying subtotals. You may change the layout of your pivot table to your preferred style by exploring the other layout menu options.

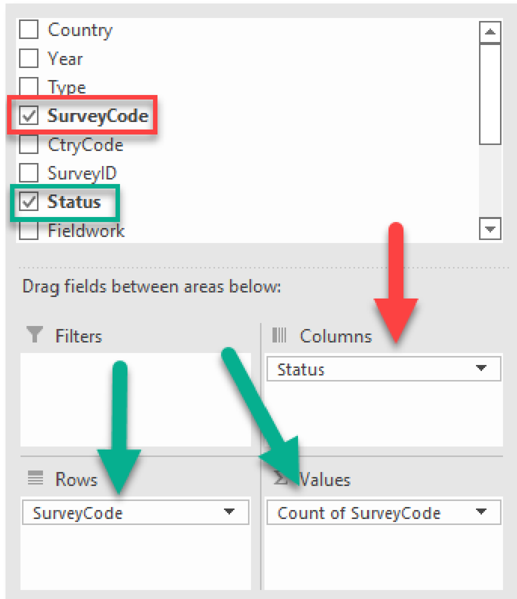


	A	B	C	D	E	F	G	H
1								
2								
3	Count of SurveyCode		SurveyCode					
4	Country	Year	AIS	DHS	MIS	OTH	SPA	Grand Total
5	⊟ Afghanistan	2010				1		1
6	⊟ Albania	2008		1				1
7	⊟ Angola	2006			1			1
8		2010	1					1
9		2011			1			1
10	⊟ Armenia	2000		1				1
11		2005		1				1
12		2010		1				1
13	⊟ Azerbaijan	2006		1				1
14	⊟ Bangladesh	1994		1				1
15		1997		1				1

Now that you've created your first pivot table, we will now practice creating a variation with the same dataset. First, we'll create a table showing the number of complete and ongoing surveys, aggregated by survey type (e.g., DHS, AIS, MIS).

10. Remove all the fields from the PivotTable by un-checking the box next to their name in the Choose Fields to Add to Report Box.

11. Drag the Status field to the Columns Box, the SurveyCode field to the Rows Box, and the SurveyCode field to the Values Box (shown below).



You should see a table like the screenshot below. Each row is one survey category. The columns contain the counts for the number of ongoing and completed surveys within each survey category.

	A	B	C	D
1				
2				
3	**Count of SurveyCode**	**Status**		
4	**SurveyCode**	**Completed**	**Ongoing**	**Grand Total**
5	AIS	8	3	11
6	DHS	222	29	251
7	MIS	8	3	11
8	OTH	20	3	23
9	SPA	15		15
10	**Grand Total**	**273**	**38**	**311**

12. Click Save.

6.2.4 Adding GPS Data to QGIS

This lesson will teach you how to add GPS data from a CSV file to your QGIS project. By the end of this lesson, you will be able to (1) prepare and export a spreadsheet in Excel as a CSV file, (2) display CSV data with geographic coordinates in QGIS, and (3) export a point vector file from QGIS using data from the CSV file.

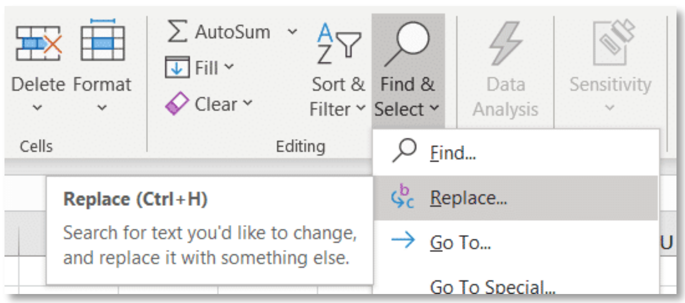
You may already know that Excel can save files in many formats including CSV files. CSV files, which stands for Comma Separated Values, are a type of delimited file. When software reads CSV files, they read anything separated by commas as separate values. QGIS can also read CSV files. In fact, many programs can read this particular file type. So, you will likely find CSV files to be quite useful. In this lesson, we will open an Excel file, modify it, and save it as a CSV. Then, the data in QGIS. To begin, follow the instructions below.

This section corresponds to Activity 6.6 of the Health Data Mapping Workshop. The exercise file for this section can be found in the \DHS_QGIS_2022\Exercise_Files\6.2.4_AddGPSDataToQGIS\ folder. It's called NigeriaModelData.xlsx. Please note that the coordinates in this file were randomly generated. The locations were not collected from GPS units. They are modeled data made specifically for this lesson.

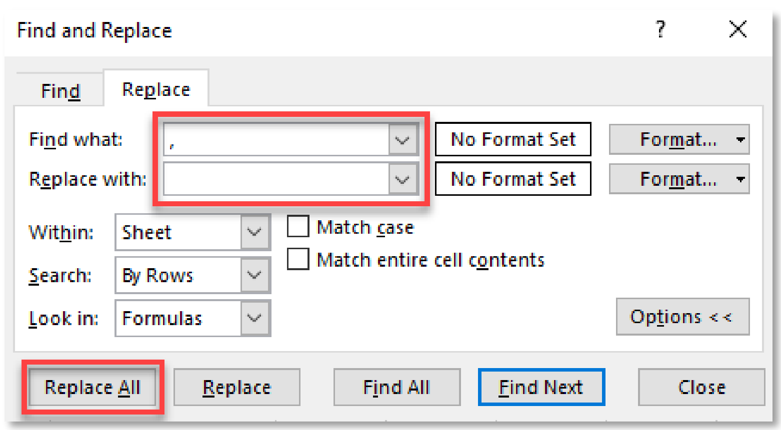
1. Open the Nigeria_modeldata.xls document and review the data it contains. Note, the two columns, Latitude and Longitude. These two columns contain the geographic data we need to display this data in a GIS. The coordinates for this data are displayed in decimal degrees, the format you learned about at the beginning of this section.

2. Notice how data in Column D, Notes, contains commas between the text "Cluster" and "U"? If we want to export this data as a CSV, we have to remove all commas from the data. Otherwise, software may not handle our data properly.

3. As depicted in this image, click the Find & Select button and select Replace. This will open the Find and Replace popup window.



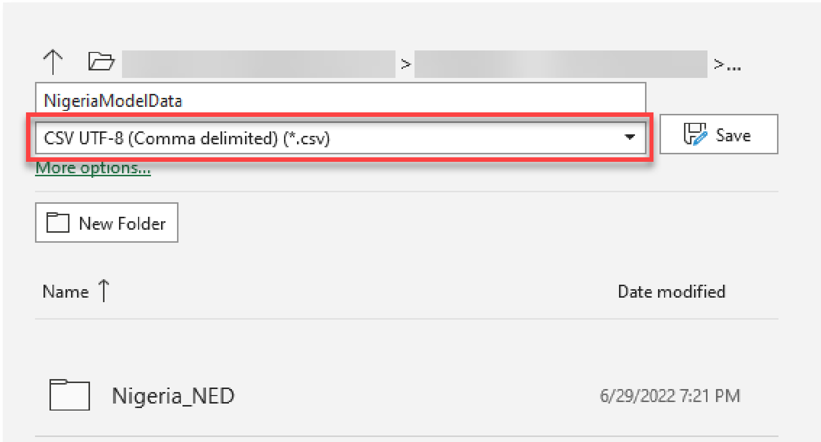
4. As shown below, enter a comma (,) into the Find What Box and leave the Replace With Box blank. Click the Replace All button. This will delete all the commas (,) in the spreadsheet. If you wanted, you could also use this tool to replace the commas with another symbol instead of deleting them. Click OK to close the window that alerts you that there is no data to replace. This means that there are no commas within the file. Then, close the Find and Replace window.



5. Click File → Save A Copy.

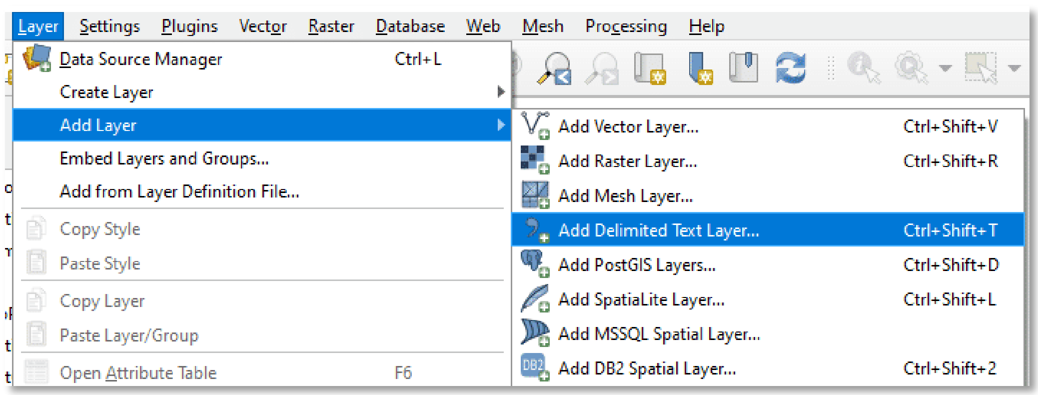
6. Navigate to the \DHS_QGIS_2022\Exercise_Files\6.2.4_AddGPSDataToQGIS\ folder. Select CSV Comma Delimited (*.csv) as the document type. This option can be seen in the image

below. Click Save. Click Yes in the window that appears to warn you that the file may contain features that are not compatible with CSV.

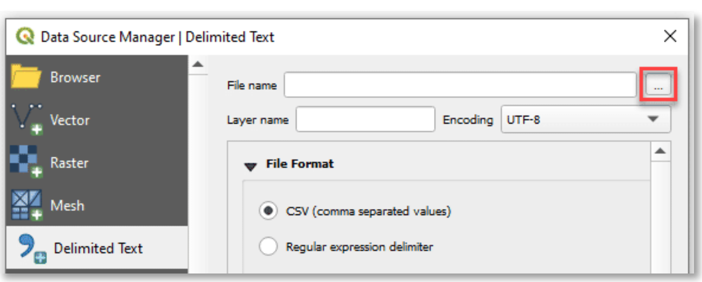


7. Close Excel (you might need to click Yes several times to do this as Excel will confirm that you want to save and close the file in a different format).

8. Using a delimited file you can add your point data to QGIS. Start QGIS and open a new project. Select Layer → Add Layer → Add Delimited Text Layer. You can also click the Add Delimited Text Layer on the toolbar if displayed.



9. Click the Browse button next to File name.

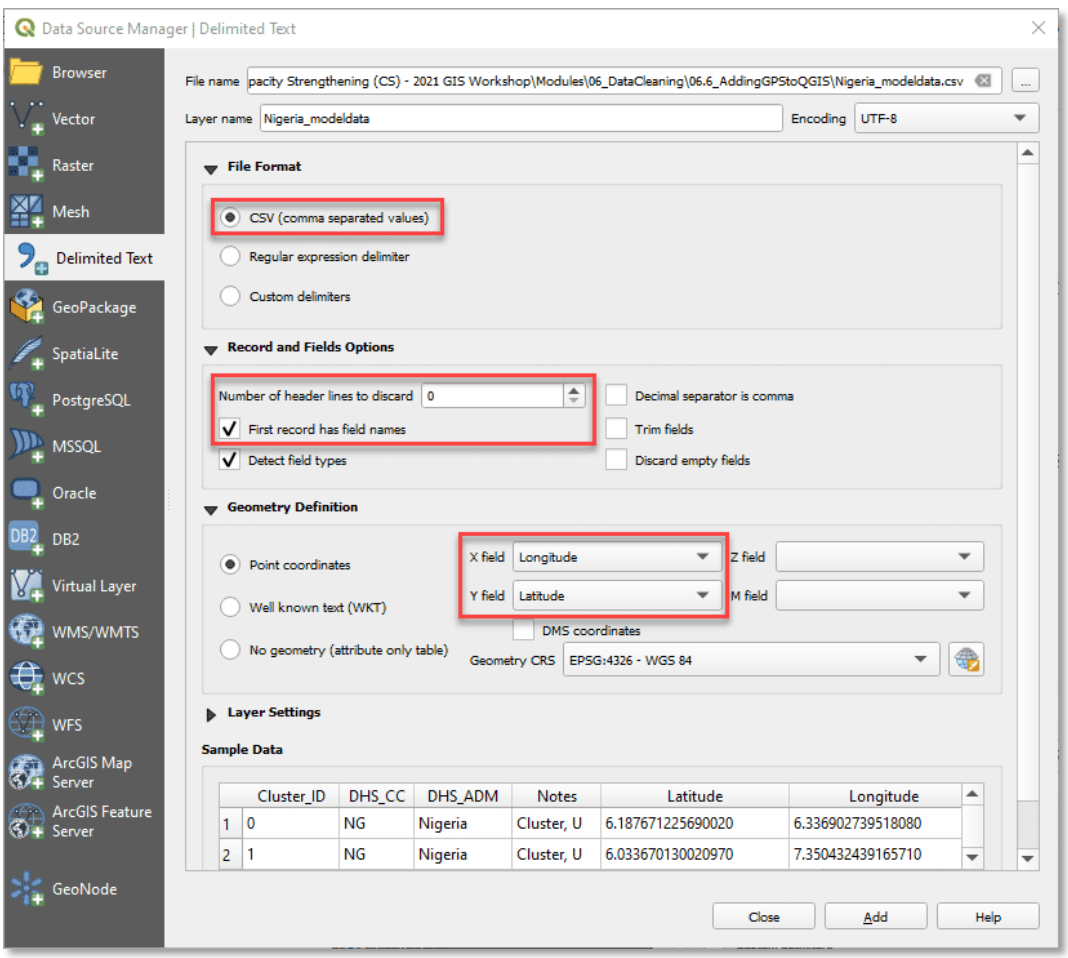


10. Navigate to the CSV you created. In some cases, the rest of the dialog will automatically fill itself in or QGIS might tell you that you need to parse the data and the fields in the window will remain blank.

11. In the *Data Source Manager* window, make sure the following conditions are set.

    - The Layer name box keep the text that appears: "Nigeria_modeldata".
    - Make sure that the CSV radio button under File format is selected.
    - Under Record and Field Options, the Number of Header Lines to Discard is set to 0 and the First record has field names radio button is selected.
    - Make sure that the X field says "Longitude" and the Y field says "Latitude" within the Geometry Definition section.
    - Ensure that the Geometry CRS is set to "EPSG:4326 - WGS 84".

Please refer to the image below as an example of how your window should appear.



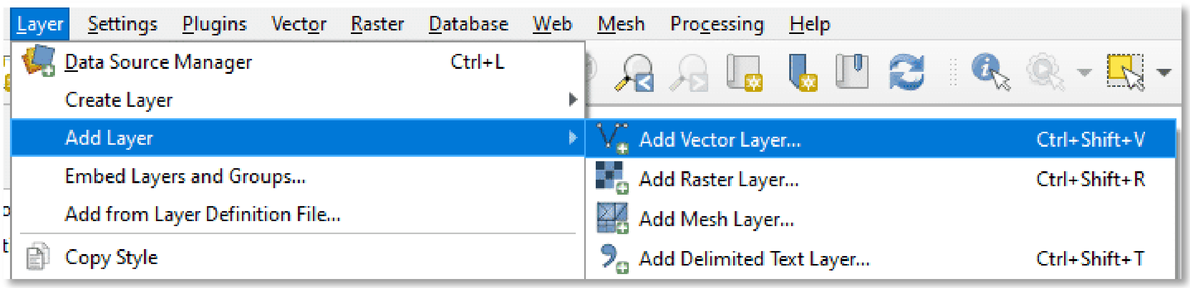
12. With the parameters above set in your Data Source Manager Window, click Add.

13. Right-click on the Nigeria_modeldata in the Layers Panel and select Open Attribute Table.

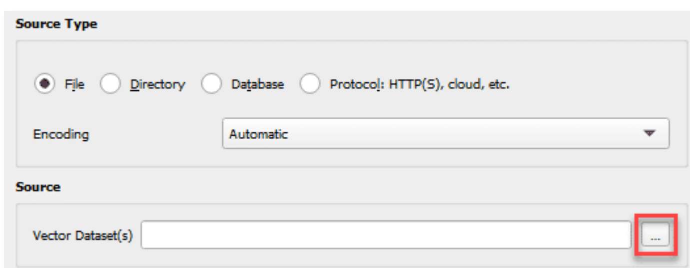
14. Check that the attributes of the layer properly imported into QGIS by looking through the attribute table. Then, close the attribute table.

Before we save our new layer, we should check the layer's geography. To do this, we can compare our layer to a shapefile. The clusters you added to your map should fall within the administrative boundaries of Nigeria. In this part, you will use the Nigeria_NED shapefile which can be found in the \DHS_QGIS_2022\Exercise_Files\6.2.4_AddGPSDataToQGIS\Nigeria_NED folder.

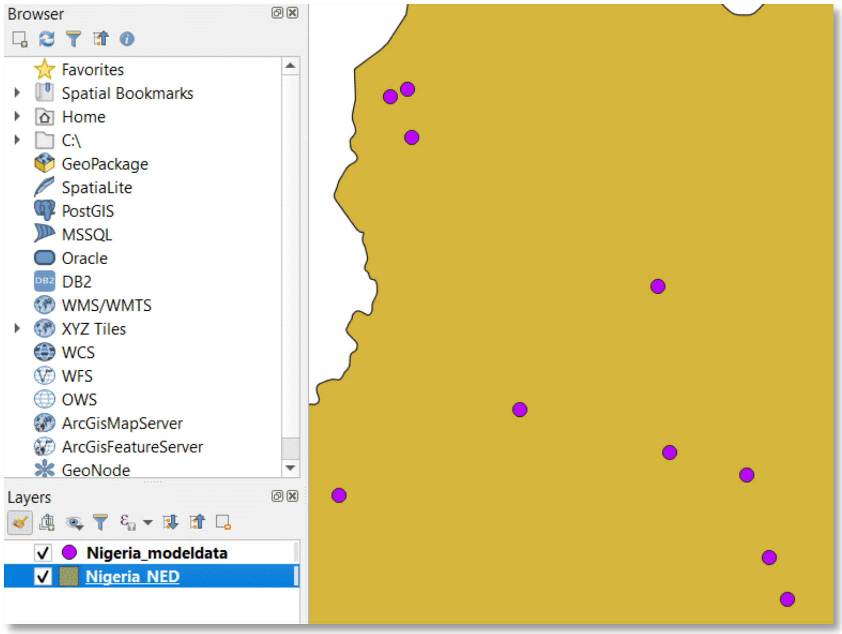
15. Add the Nigeria_NED.shp to your QGIS project by clicking Layers → Add Layer → Add Vector Layer (shown below).



16. In the Data Source Manager Window, click the Browse Button (...) which is identified below. In the new navigation window, navigate to and add the shapefile (Nigeria_NED.shp) to your QGIS project. Once you have selected the *.shp file, you need to click Add in the bottom-right corner of the Data Source Manager.



17. Once you add the Nigeria_NED.shp file to your QGIS project, it may cover the cluster data. To fix this, so that the clusters are visible, we need to click, hold, and drag the Nigeria_NED file below the Nigeria_modeldata file in the Layers Panel. See the image below.



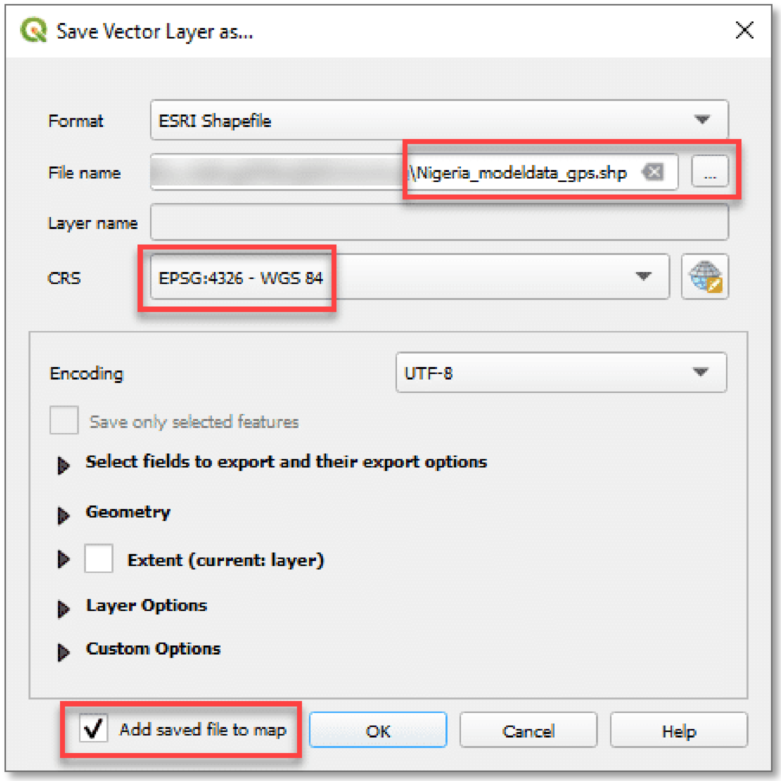
Now that we have checked both the attributes and geography of our layer, we can now save it. When you first add a delimited file to QGIS, it does not save the information. It needs to be manually saved, otherwise it will be a temporary file that cannot be used in multiple projects. To save the Nigeria_modeldata as a shapefile, we need to export it from our project.

18. Highlight the Nigeria_modeldata layer in the Layers Panel. Right-click on the layer and select Export → Save Features As.

19. In the Window, make sure that the Format is set to ESRI Shapefile. Click the Browse Button (...) and navigate to the folder where you unzipped the Nigeria_NED shapefile. Name the new file, Nigeria_modeldata_gps, and click Save. For the CRS (Coordinate Reference System) make sure it is set to "ESPSG:4326 - WGS 84".

The most common GPS format is WGS84 but verify this with the GPS unit or the data collection agency. Often this will already be in the GPS data file so make sure to check all the fields before importing.

20. Check the box next to ‘Add saved file to map’ to automatically add the newly created shapefile to your map document. Your window should look like the example below. Then, click OK.



21. Click OK in the pop-up window that confirms that creation of your shapefile.

Your new shapefile will appear in the Layers Panel.

22. Save your QGIS project in the folder where you downloaded your data.

6.2.5 Cleaning and Using GPS Data

As mentioned at the beginning of this chapter, GPS data can come in different forms. You may encounter GPS data in Decimal Degrees, but you may also find GPS data that uses Degrees, Minutes, Seconds. In this exercise, you will follow a series of instructions to process and use GPS data in different formats. This exercise will help you practice the various techniques you learned in this section.

This section corresponds to Activity 6.7 of the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\6.2.5_CleanAndUseGPSData\ folder.

1. Let's start by opening the GPSPoints.xls document in Microsoft Excel. Once you have the Excel file open, you will see there are three different tabs, each with GPS data for a different country. If you look at the coordinate information in each tab, you will notice that each country has its coordinates written differently.

    a. Uganda's coordinates are written in decimal degrees (e.g. 0.29739, 32.6553).

    b. Tanzania's coordinates might appear to be written in degrees and decimal minutes (e.g. -6° 82.097', 39° 2.7745'), but they are actually written in decimal degrees with the whole numbers separated from the decimals.

c. Kenya’s coordinates are written in degrees, minutes, and seconds (e.g. -1° 18’ 48”, 36° 46’ 46”).

Before we add this data to QGIS, we want to make sure that the GPS coordinates are all written in the same format. Specifically, it’s best practice to convert them into the decimal degrees format.

2. We'll begin with Uganda's data. Since this data is already in decimal degrees, we will not need to make any conversions. All we need to do is copy the data over to a new Excel sheet.

	A	B	C	D	E	F
1	ID	DHSCC	ISO3	URBAN_RURA	LATNUM	LONGNUM
2	1	UG	UGA	U	0.29739	32.6553
3	2	UG	UGA	U	0.36219	32.61547
4	3	UG	UGA	U	0.32141	32.57544
5	4	UG	UGA	U	0.24698	32.6265
6	5	UG	UGA	U	0.30167	32.56819
7	6	UG	UGA	U	0.37905	32.59282
8	7	UG	UGA	U	0.37905	32.51077
9	8	UG	UGA	R	0.26604	32.56228
10	9	UG	UGA	R	0.3606	32.70452
11	10	UG	UGA	R	0.3896	32.66095

3. To copy the data, select all the data in the Uganda tab (Ctrl+A) copy the data (Ctrl+C). Then open a new Excel document (Crtl+N) and paste in the data (Ctrl+V). Now, all we have to do is save the new Excel document as a CSV. Make sure to save the document in your exercise folder, or a folder you will remember. We suggest saving this first file with the following name: "6.7_UG.csv".

4. We will need to do a bit of editing to prepare Tanzania's data. Tanzania's coordinates are formatted as decimal degrees with the whole numbers separated from the decimals. We will need to convert this data into regular decimal degrees.

5. Copy all the data in the Tanzania tab (Ctrl+A then Ctrl+C) and paste it in a new Excel document (Ctrl+N then Ctrl+V).

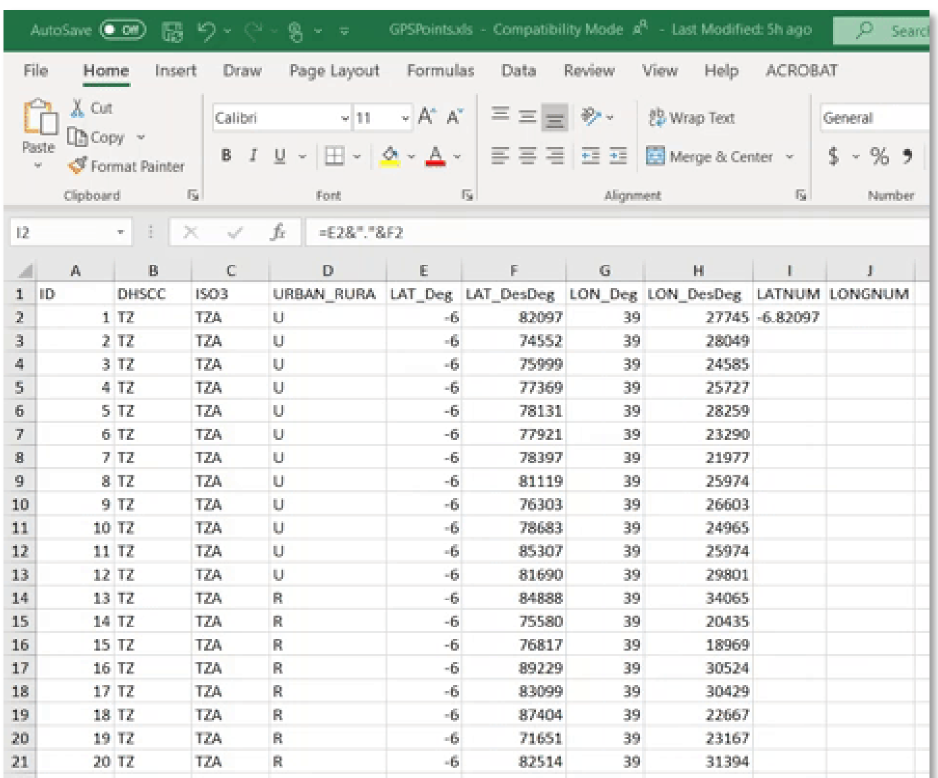
6. Once the data is in a new file, we need to make two new columns, one for the latitude and the other for the longitude in decimal degrees. Add these two new columns to the end of the table, and name them "LATNUM and LONGNUM" respectively.

	A	B	C	D	E	F	G	H	I	J
1	ID	DHSCC	ISO3	URBAN_RURA	LAT_Deg	LAT_DesDeg	LON_Deg	LON_DesDeg	LATNUM	LONGNUM
2	1	TZ	TZA	U	-6	82097	39	27745		
3	2	TZ	TZA	U	-6	74552	39	28049		
4	3	TZ	TZA	U	-6	75999	39	24585		
5	4	TZ	TZA	U	-6	77369	39	25727		
6	5	TZ	TZA	U	-6	78131	39	28259		
7	6	TZ	TZA	U	-6	77921	39	23290		
8	7	TZ	TZA	U	-6	78397	39	21977		
9	8	TZ	TZA	U	-6	81119	39	25974		
10	9	TZ	TZA	U	-6	76303	39	26603		
11	10	TZ	TZA	U	-6	78683	39	24965		
12	11	TZ	TZA	U	-6	85307	39	25974		
13	12	TZ	TZA	U	-6	81690	39	29801		
14	13	TZ	TZA	R	-6	84888	39	34065		
15	14	TZ	TZA	R	-6	75580	39	20435		
16	15	TZ	TZA	R	-6	76817	39	18969		
17	16	TZ	TZA	R	-6	89229	39	30524		
18	17	TZ	TZA	R	-6	83099	39	30429		
19	18	TZ	TZA	R	-6	87404	39	22667		
20	19	TZ	TZA	R	-6	71651	39	23167		
21	20	TZ	TZA	R	-6	82514	39	31394		

7. We'll now convert the latitude values. In the first cell under the "LATNUM" heading, we're going to combine the whole number in Column E with the decimal in Column F. In that cell, type the following formula:

=E2&"."&F2

8. Once you type that formula into the cell under LATNUM, click Enter on your keyboard.



	A	B	C	D	E	F	G	H	I	J
1	ID	DHSCC	ISO3	URBAN_RURA	LAT_Deg	LAT_DesDeg	LON_Deg	LON_DesDeg	LATNUM	LONGNUM
2	1	TZ	TZA	U	-6	82097	39	27745	-6.82097	
3	2	TZ	TZA	U	-6	74552	39	28049		
4	3	TZ	TZA	U	-6	75999	39	24585		
5	4	TZ	TZA	U	-6	77369	39	25727		
6	5	TZ	TZA	U	-6	78131	39	28259		
7	6	TZ	TZA	U	-6	77921	39	23290		
8	7	TZ	TZA	U	-6	78397	39	21977		
9	8	TZ	TZA	U	-6	81119	39	25974		
10	9	TZ	TZA	U	-6	76303	39	26603		
11	10	TZ	TZA	U	-6	78683	39	24965		
12	11	TZ	TZA	U	-6	85307	39	25974		
13	12	TZ	TZA	U	-6	81690	39	29801		
14	13	TZ	TZA	R	-6	84888	39	34065		
15	14	TZ	TZA	R	-6	75580	39	20435		
16	15	TZ	TZA	R	-6	76817	39	18969		
17	16	TZ	TZA	R	-6	89229	39	30524		
18	17	TZ	TZA	R	-6	83099	39	30429		
19	18	TZ	TZA	R	-6	87404	39	22667		
20	19	TZ	TZA	R	-6	71651	39	23167		
21	20	TZ	TZA	R	-6	82514	39	31394		

9. You now need to copy it down to the bottom of the table. To do that, select the cell with your formula by clicking on it. Once it's selected, hover your cursor over the bottom right-hand corner of the cell. When you do so, your cursor will turn into a little black +. Once you see that +, double-click and the formula in that cell will be copied down to the bottom of the table.

10. Now, let's convert the longitude values. In the cell under the "LONGNUM" heading, cell J2, write the following formula:

    =G2&"."&H2

11. After you type this formula, we're going to copy it down the table like we did in the previous step. To do that, select the cell with your formula by clicking on it. Once it's selected, hover your cursor over the bottom right-hand corner of the cell. When you do so, your cursor will turn into a little black **+**. Once you see that **+**, double-click and the formula in that cell will be copied down to the bottom of the table.

12. Now, your Tanzania data should have columns with coordinate data in decimal degrees. Let's save our new data as a CSV. Be sure to save this data wherever you decided to save your Uganda data. We suggest saving this Tanzania file as a CSV with the following name: "6.7_TZ.csv".

13. Finally, let's prepare the Kenya data. As we did with Uganda and Tanzania, let's copy the Kenya data from our "GPSPoints.xls" document and into a new Excel workbook.

14. Once the data has been copied to a new workbook, let's set up our new coordinate columns. Add two additional columns to your Kenya data table, called "LATNUM" and "LONGNUM" respectively. We're going to populate these two columns with decimal degree coordinates, but since we will be converting coordinates from degrees, minutes, and seconds (DMS) to DD, the formulas will be a bit more complicated.

15. In the first cell under the "LATNUM" heading, cell K2, type the following formula:

    =E2+(F2/60)+(G2/3600)

16. Now, let's copy this formula down the column to the bottom of the table. To do this, select the cell with your formula by clicking on it. Once it's selected, hover your cursor over the bottom righthand corner of the cell. When you do so, your cursor will turn into a little black +. Once you see that +, double-click and the formula in that cell will be copied down to the bottom of the table.

17. Now, we're going to repeat steps 13 and 14 for the longitude. In the first cell under the "LONGNUM" heading, cell L2, type the following formula:

    =H2+(I2/60)+(J2/3600)

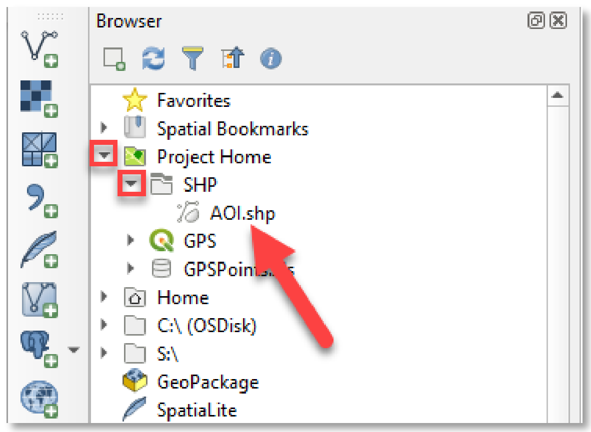
18. Now, let's copy this formula down the column to the bottom of the table. Repeat the instruction in **Step 14** for column L2.

19. Let's save our new Kenya data. Be sure to save it as a CSV in the same place your saved your Uganda and Tanzania data. We recommend using the following name for this new spreadsheet: "6.7_KY.csv".

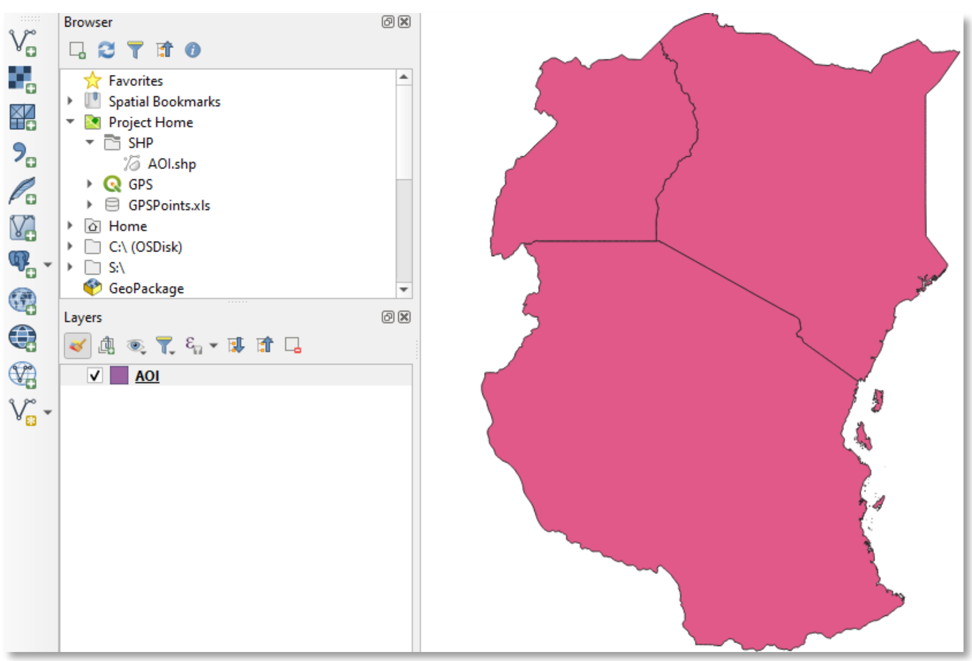
With our CSV data prepared, let's add the data to QGIS to evaluate our conversion work. Before we add the CSV data to a QGIS project, however, let's set up a project.

20. First, open a new QGIS project. Then, save the project in the exercise folder, \DHS_QGIS_2022\ Exercise_Files\6.2.5_CleanAndUseGPSData\, so you can easily access the data therein. To do this, click on Project → Save As and navigate to this folder. Save the project as GPS.qgz.

21. With the project saved, we can now easily access the shapefiles in the exercise folder by expanding the Project Home folder in the Browser Panel. Expand the Project Home folder and the \SHP\ folder (shown below). Then drag the area of interest shapefile (called AOI.shp) into the map frame.



Your result should look similar to that in the image below.

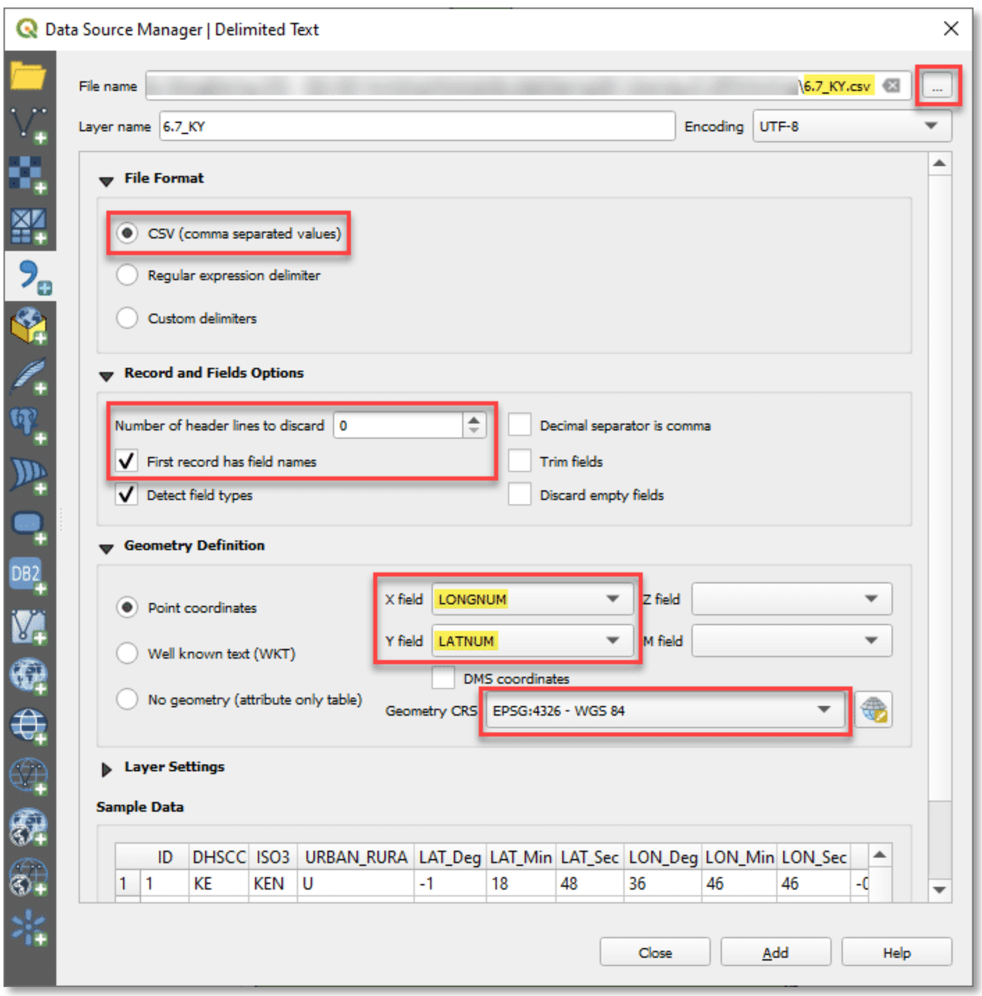


22. Now, we'll add the CSV data to our project. We'll begin with our Kenya data. On the QGIS menu bar, select the Layer → Add Layer → Add Delimited Text Layer.

23. When you select this option, the Data Source Manager popup menu will appear with the Delimited Text tab already selected. Next to the File Name Textbox, you'll see a Browse Button (…). Click the button and navigate to your "6.7_KY.csv" document.

24. With the "6.7_KY.csv" selected, we need to configure a few settings. Be sure that the CSV radio button is selected under the File Format section. Under the Record and Fields Options, make sure the Number of Header Lines to Discard is set to 0 and that the box next to 'First record has field names' is checked.

25. We also need to define which attributes contain coordinates. Under the Geometry Definition heading, you'll see some dropdown menus. In the X Field dropdown, we want to select the column in our data with longitude coordinates. So, select the "LONGNUM" attribute from the list. In the Y Field dropdown, we want to select the column in our data with latitude coordinates. So, select the "LATNUM" attribute from the list.

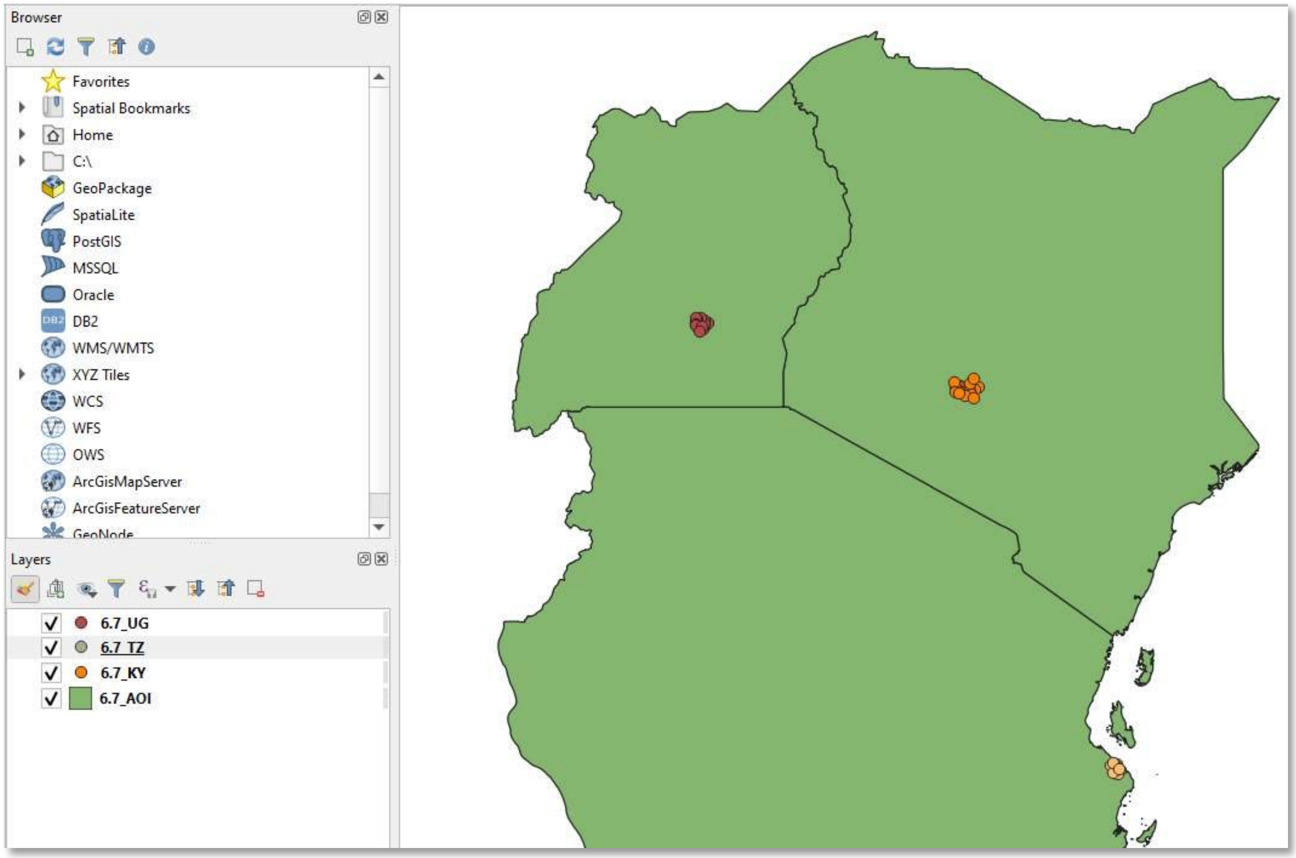


26. Click Add at the bottom of the popup menu. This will add the Kenya data to our Layers Panel and the map.

27. We'll now repeat steps 4 through 7 for the Tanzania and Uganda data which were saved as "6.7_TZ.csv" and "6.7_UG.csv" respectively. Once you have added the Tanzania and Uganda

GPS clusters to the map, you can evaluate the coordinate conversion. Check to make sure the GPS clusters are within the proper country boundaries.

Your map should now look similar to the map below.



The GPS clusters in your map are visualizations of the CSV data. As was explained in previous lessons, this visualization is temporary. If we wanted to use this GPS data in other maps, we would need to export the layers as shapefiles from QGIS. For this exercise, however, we don't need to do that. For now, we'll create a simple map in QGIS.

28. To create a map using this data, follow the steps below.

29. First, be sure that the three GPS cluster layers have different colors. Feel free to change the symbology of the layers in the map as you see fit.

30. Then, use the magnifying glass icon to zoom in on the GPS clusters. Configure your map so that all the clusters can be seen.

31. With our data visible, we should create a new Print Layout. In the QGIS menu, click on Project → New Print Layout from the dropdown. In the popup box, enter the title, "GPS Points". Once you enter a title and click OK, your new Print Layout will appear in a separate window.

32. The Print Layout window will display a blank canvas on which you can add different map elements. To add map elements, you will select the elements from the list of icons along the left-hand side of the screen, click-hold-and-drag your mouse on the canvas, and release your mouse.

Be sure your map includes all five of the elements listed below. For this exercise, be sure your map includes the following elements.

- Title
- Legend
- Scale bar

33. With these three elements added to your map and organized to your liking, you are now ready to export your map. Before we export the map, let's save our Print Layout. To do that, click on the Layout → Save Project.

34. To export the map, click on the Layout Menu in your Print Layout window and select Export as Image, choosing your desired format.

7. JOINS

7.1 CONCEPTS

When using geospatial data, we often need to merge data from multiple data sources. In GIS, we can combine data sources by joining them together. The datasets can be joined based upon their attributes or a relationship in their geographies. By the end of this section, you will be able to:

1. Join data by attributes,
2. Join data by their spatial location, and
3. Calculate the sum of points within a polygon.

7.1.1 Definitions

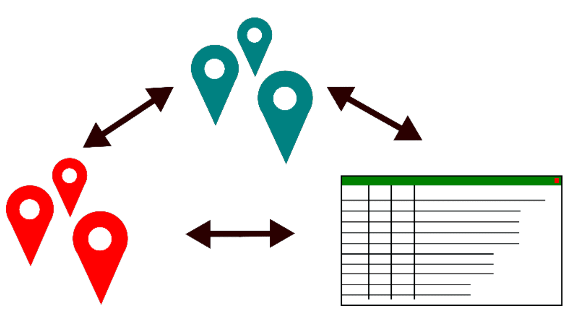
The following terms are used in this section.

- **Attribute join**: merging two datasets based on a common variable called a join key
- **Attribute table**: a tool used in QGIS to explore a layer's attributes
- **Join data**: when joining two datasets, this is the dataset that will be added to the target data
- **Join key**: a common variable used to merge two datasets, often a unique identifier
- **Spatial join**: merging two datasets based upon the geographic location of their features
- **Target data**: when joining two datasets, this is the dataset to which the join data will be added
- **Unique identifier**: a variable in which each observation is unique (e.g., a cluster or facility ID)

7.1.2 Joining Data

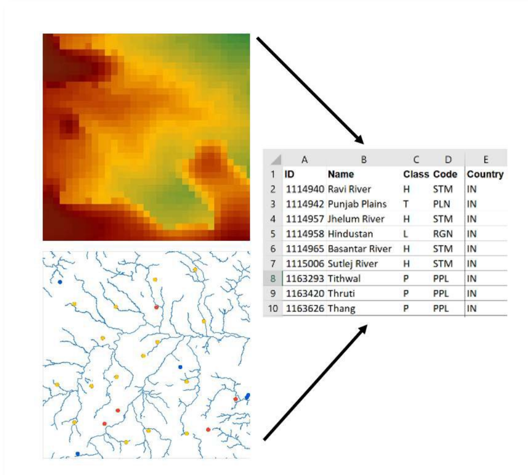
This section corresponds to Activity 7.1 in the Health Data Mapping Workshop. By the end of the section, you will be able to:

1. Explain the difference between spatial joins and attribute joins,
2. Define a join key (unique identifier), and
3. Differentiate a target layer from join data.



As a reminder, all layers in a GIS contain two primary components: geography and attributes. A layer's geography is a representation of its location. The geography can be in the form of vectors, specifically points, lines, and polygons. It could also be in the form of raster images, which are pixelated images in which each pixel represents a square unit of space.

Both vector and raster data layers have attribute data associated with them, which can provide a wide range of information. When joining two data sources, one layer will be the target data and the other will be the join data. The target layer is the GIS layer to which the other data source will be joined. The join data, as we will see, can either be in the form of another GIS layer or an outside table. In either case, the join data is the data which is added to the original target layer.



	A	B	C	D	E
1	ID	Name	Class	Code	Country
2	1114940	Ravi River	H	STM	IN
3	1114942	Punjab Plains	T	PLN	IN
4	1114957	Jhelum River	H	STM	IN
5	1114958	Hindustan	L	RGN	IN
6	1114965	Basantar River	H	STM	IN
7	1115006	Sutlej River	H	STM	IN
8	1163293	Tithwal	P	PPL	IN
9	1163420	Thruti	P	PPL	IN
10	1163626	Thang	P	PPL	IN

Joining a GIS layer to an outside table by its attributes is called an attribute join. In addition to attribute joins, a user can also combine data in a GIS via what is called a spatial join. A spatial join combines data from one GIS layer to another GIS layer based upon the geographic locations of both.

Attribute Joins

Attribute data are descriptive information about a geographic feature in a GIS. For example, attributes of a river might include its name, length, and associated basin. All attribute data are stored in what is called an attribute table, and every layer in a GIS should have one.

Within each attribute table there should be at least one unique identifier. A unique identifier acts as an attribute that allows a layer in a GIS to join with an outside table. Both the GIS data table and the outside table must have a column that has the same unique identifier. When we use a unique identifier to join two datasets, we often call it a join key.

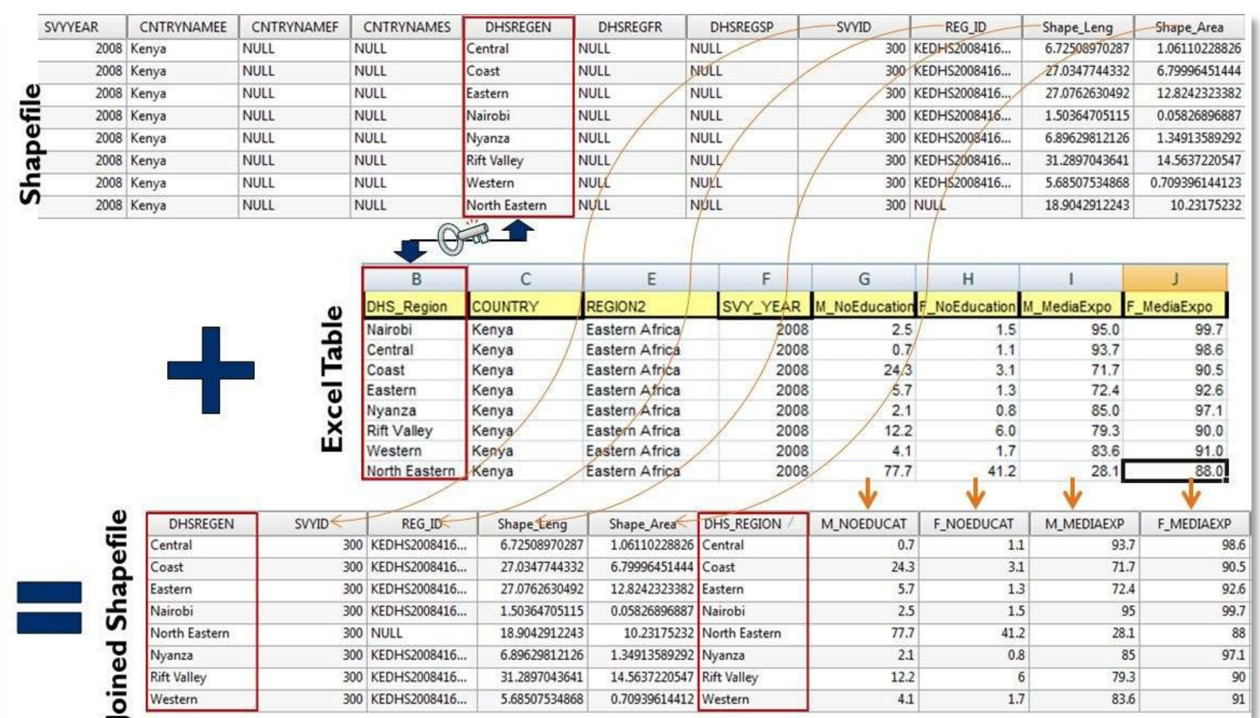
In the example below, the unique identifier is the ID value. While the River_name attribute may also be unique, there is a chance that multiple rivers in the attribute table have the same name. If we know each river name is unique, this attribute could also be considered a unique identifier and used as the join key.

ID	River_name	Length_km	Basin_km2
1	Mississippi	3,730	2,981,076
2	Amazon	6,937	7,050,000
3	Niger	4,180	2,117,700

To join data from an outside table, the join key must be written exactly the same way. If its values are text and not numerical, they need to have the same spelling, capitalization, and same spacing. Before joining a GIS layer to join data, such as an outside table, a user may want to make edits to the target layer's attribute table. While Excel has more tools for editing data in a table, QGIS does have some helpful attribute table editing capabilities. In QGIS, the user can:

- Link data from an outside table (Excel, DBF, csv, etc.) or another layer,
- Add columns and enter data by hand,
- Change the data format of columns (e.g. from a numerical value to text), and
- Create new columns using two or more existing columns through calculation or concatenation.

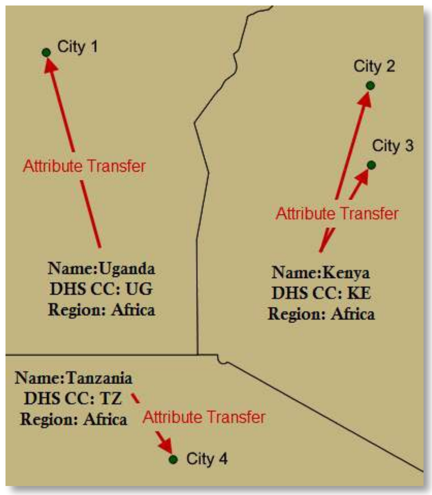
Once you are ready to conduct the join, you can add a table to your QGIS project and use an attribute join to link it to your target layer. The diagram below illustrates how this process works. Note that the join key is called DHSREGEN in the shapefile (the target layer) and DHS_Region in the Excel table (the join data). Both of these variables contain the same names.



Spatial Joins

A spatial join copies the attributes from an existing GIS layer and joins them to the attribute table of another layer based upon the geographic locations of features within each GIS layer. A spatial join may join different types of vector data (e.g. Points, Lines, Polygons) and raster data. In QGIS, a user can spatially join two GIS layers that intersect or overlap.

In the example below, the target layer is a shapefile of cities, represented as points on the map. The join data is a shapefile of countries, the polygons in the map. The attributes of each country – country name, country code, and region – are transferred to all of the cities that are located within the country boundaries.



In some GIS software, such as ArcGIS Pro, spatial joins also allow you to count the number of points within a polygon. In QGIS, this is accomplished with a separate tool. You will learn how to conduct a standard spatial join in Section 7.2.2 and learn how to count points within a polygon in Section 7.2.3.

7.2 QGIS INSTRUCTIONS

The following two sections will use the exercise files provided in the \DHS_QGIS_2022\Exercise_Files\7.2.1_Joins\ folder.

7.2.1 Attribute Joins

This section corresponds to Activity 7.3 of the Health Data Mapping Workshop. In this activity, you will be enriching a shapefile of countries from SDR with variables from a dataset acquired from Natural Earth Data. To do this, you will conduct an attribute join in QGIS.

1. Open a new QGIS project. Save the project in the \DHS_QGIS_2022\Exercise_Files\7.2.1_Joins\ folder.

2. In the Browser Panel, expand the Project Home folder. In it, you will find two subfolders: \shp_osm_2021_facility\ and \shp_sdr_2021\. Expand the \shp_sdr_2021\ folder and drag the shapefile into the map viewer. The sdr_national_data shapefile should now appear in the Layers Panel.

3. In the Project Home folder of the Browser Panel, you will also see an Excel file called custom_data_ne.xlsx. Expand this Excel file in the Browser Panel and drag the ne_custom worksheet into your Layers Panel.

In the Layers Panel, you should now see two different items: the ne_custom worksheet and the sdr_national_data shapefile. If you can't find the Layers Panel, you can add it into your project by right-clicking on the QGIS menu bar and checking the box next to the item called Layers Panel.

4. In the Layers Panel, right-click on the sdr_national_data and, in the popup menu, select Properties.

5. In the Layer Properties Window, select the tab on the left-hand side called Joins.

6. In the bottom left-hand corner of the Joins tab, you'll see a green **+** button. Click that button and the Add Vector Join popup will appear. Next to the text, Join layer, there will be a dropdown menu. In this first dropdown menu, select the Excel file, ne_custom. Next to the text, Join field, select 'DHSCC' from the dropdown menu. In the third dropdown menu, next to Target field, select 'DHS_CC'. Now that the join is set up, click the button, OK. When the first popup disappears, press OK on the Layer Properties popup.

7. Now, let's test the join. Right click in the QGIS menu bar at the top of the program and check the box next to the Layer Styling Panel text. This will open the Layer Styling Panel.

8. In the Layer Styling Panel, select the 'sdr_national_data' layer from the top dropdown menu. Then, in the second dropdown menu, select the option Categorized. Below the Categorized dropdown menu, you'll see the text Value next to another dropdown menu. Open this dropdown menu, scroll down, and select the attribute that ends with the text '.._INCOME_GRP'. With the '.._INCOME_GRP' attribute selected, click the button, Classify, near the bottom of the Layer Styling Panel. When you select this button, the countries in your map will be classified by income group.

9. In the Layers Panel, right-click on the 'sdr_national_data' layer and select the option at the top of the menu, Zoom to Layer.

7.2.2 Spatial Joins

This section will use the exercise files from the previous section, those in the \DHS_QGIS_2022\Exercise_Files\7.2.1_Joins\ folder. The instructions below come from Activity 7.4 in the Health Data Mapping Workshop.

In this activity, you will be using a spatial join to add information from a shapefile from SDR to a later of health facilities acquired from OpenStreetMap.

1. For this activity, you are going to use the QGIS project and exercise data you used in the previous activity. Begin by navigating to the location where you saved that project and opening it.

2. With your project open, let's add one of the layers we hope to use in our spatial join. To do this, expand the Project Home folder in the Browser Panel and then expand the \shp_osm_2021_facility\ folder. Drag the osm_health_facilities shapefile into your map frame. It should now appear in your Layers Panel.

3. For this activity, we want to extract information from the sdr_national_data layer and add it to the osm_health_facilities. Before we do so, open the attribute tables for these two layers one-by-one by right-clicking on each layer in the Layers Panel and selecting Open Attribute Table. Once you open up the attribute table of the health facility layer, you'll see that our health facilities don't

have information about in which country they reside. We want to join the information from the polygon layer, the sdr_national_data shapefile, to the osm_health_facilities layer.

4. Click on the Vector menu → Data Management Tools → Join Attributes by Location.

5. In the Join Attributes by Location popup, select osm_health_facilities point shapefile as the Input layer. And then select the sdr_national_data layer as the Join layer. Make sure only the Intersects box is checked under the Geometric predicate text. Under the Join type text, select the option Take attributes of the first located feature file (one-to-one) option from the dropdown menu. Click Run to execute the join.

6. Close the popup menu, and you'll see a new layer called 'Joined layer' has been added to our Layers Panel. This is a temporary layer. In order to use it elsewhere, we need to save it. To do this, right-click on the 'Joined layer' in the Layers Panel and select Export → Save Layer As. Export the layer as osm_facilities_join as a shapefile in the same folder where you unzipped your exercise files.

7. Let's add symbology to this new layer so that the points are colored by country. Open up the Layer Styling Panel by right-clicking on the QGIS menu bar and checking the box next to Layer Styling Panel. Under the first dropdown menu, select the new osm_facilities_join layer. Select Categorized from the second dropdown menu. Next to the text that says Value, select the attribute, 'DHS_CC', from the dropdown menu. Then, click the Classify button at the bottom of the panel. This will classify the points by their DHS country code.

8. Your sdr_national_data shapefile most likely has the categorized symbology you used in the previous exercise. For this map, we only want our new layer, osm_facilities_join, to have categorized symbology. In the Layer Styling Panel, select sdr_national_data and change the symbology to Single symbol. Feel free to use whatever color you like for its symbology.

9. Right-click on osm_facilities_join in the Layers Panel and select Zoom to Layer.

7.2.3 Counting Points in Polygons

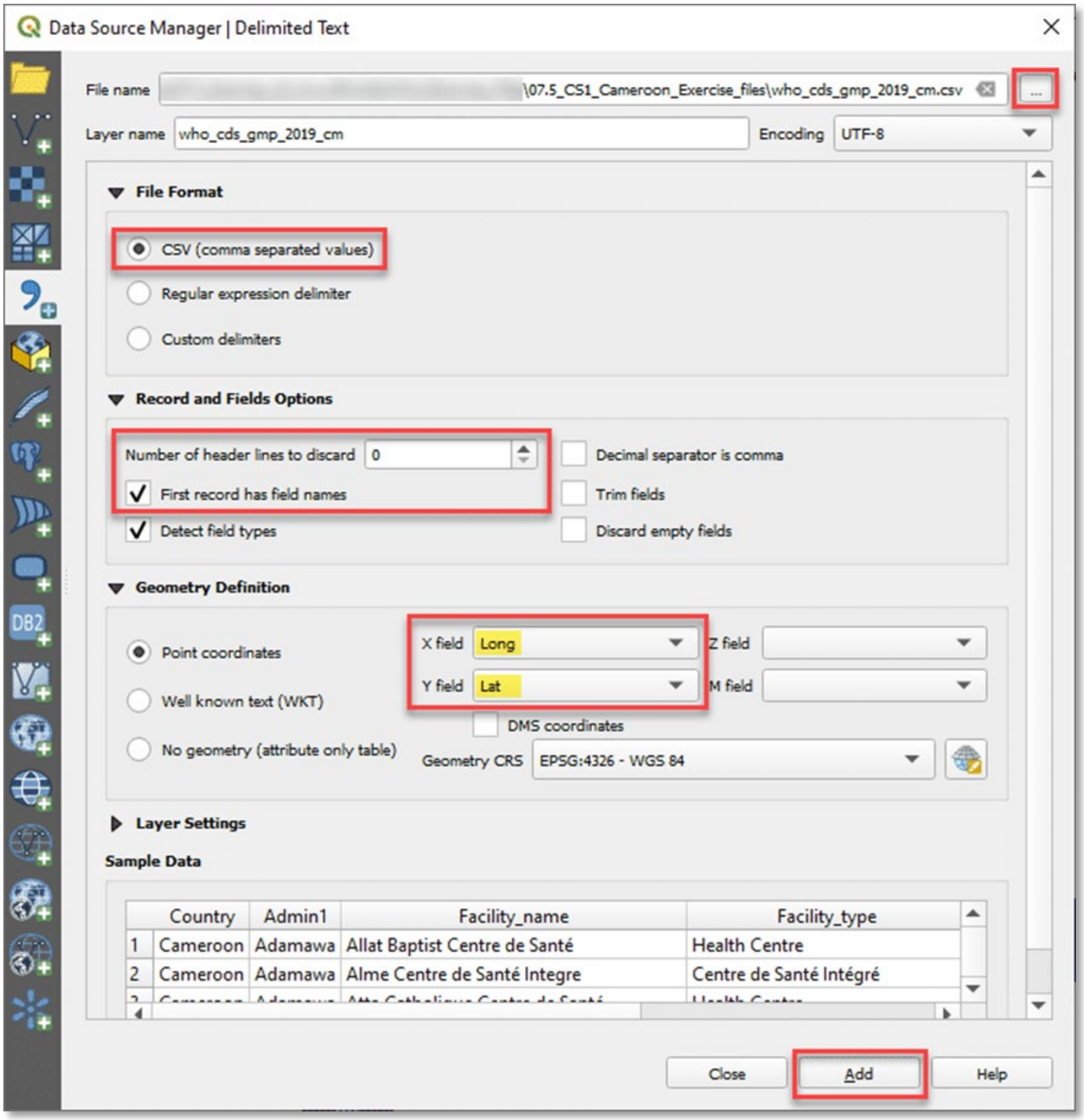
In other GIS software, spatial join tools may give you the option to count the number of points within a polygon. In QGIS, there is a separate tool used to accomplish this. In this activity, you will be counting the number of health facilities within the regions of Cameroon.

This section corresponds to Activity 7.5.1 of the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\7.2.3_CountPointsInPolygons\ folder. For this exercise, the health facility data for Cameroon is from WHO's Global Malaria Programme (GMP) and the administrative boundaries come from The DHS Program's Spatial Data Repository.

1. Open the QGIS project, Counting_points.qgz.

2. With the QGIS project open, you will see that there is currently one layer in your map. Your case study's administrative boundaries will appear in the map's center. For this exercise, we want to count the number of health facilities in the country's administrative units. So, we need to add the health facilities to the map.

3. The health facility locations are stored in the CSV document. To add them to your map, click on the Layer menu → Add Layer → Add Delimited Text Layer.



4. In the Delimited Text tab of the Data Source Manager, configure the options as they are displayed in the image. Be sure that the following conditions are set. The File format should be CSV. The Record and Fields Options should have the Number of header lines to discard set to 0 and the box next to First record has field names should be checked. Lastly, 'Long' should be set as the X field and 'Lat' should be selected as the Y field. Once these settings are selected, click Add.

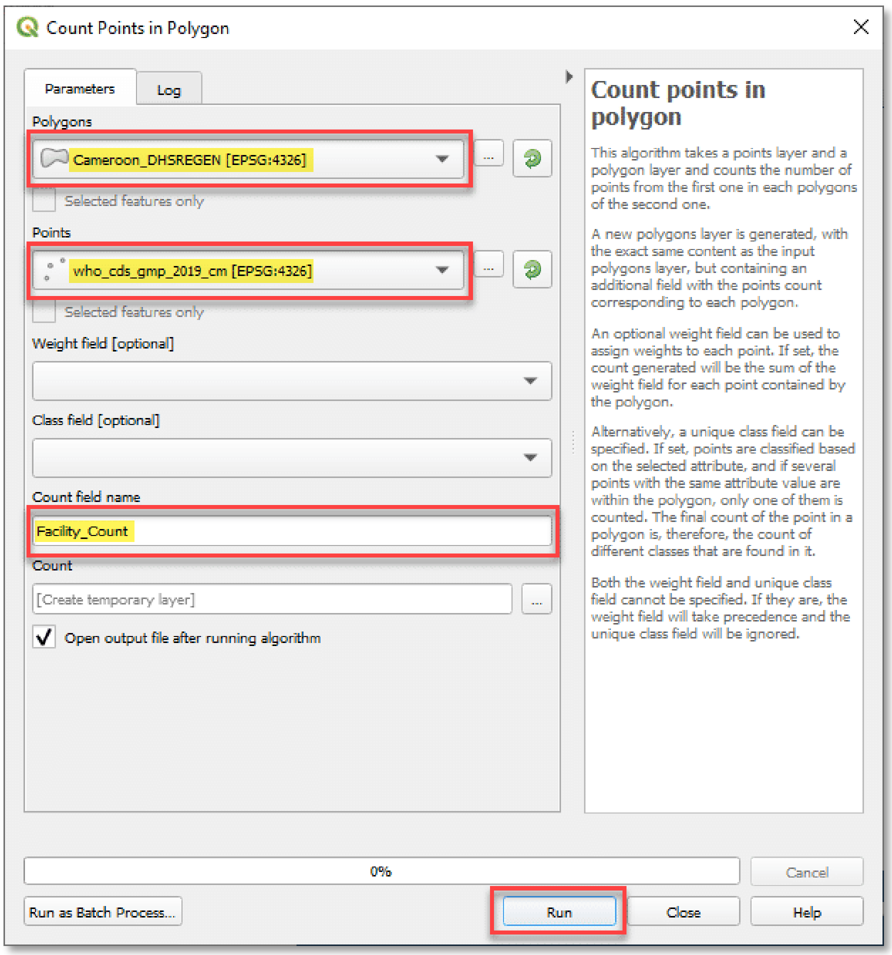
5. Close the Data Source Manager window, and you should see the health facility layer added to your map and the Layers Panel.

6. With our CSV data displayed in our QGIS project, we can now use the Count Points in Polygon function to count the number of health facilities in each administrative unit.

7. Click on the Vector menu → Analysis Tools → Count Points in Polygon.

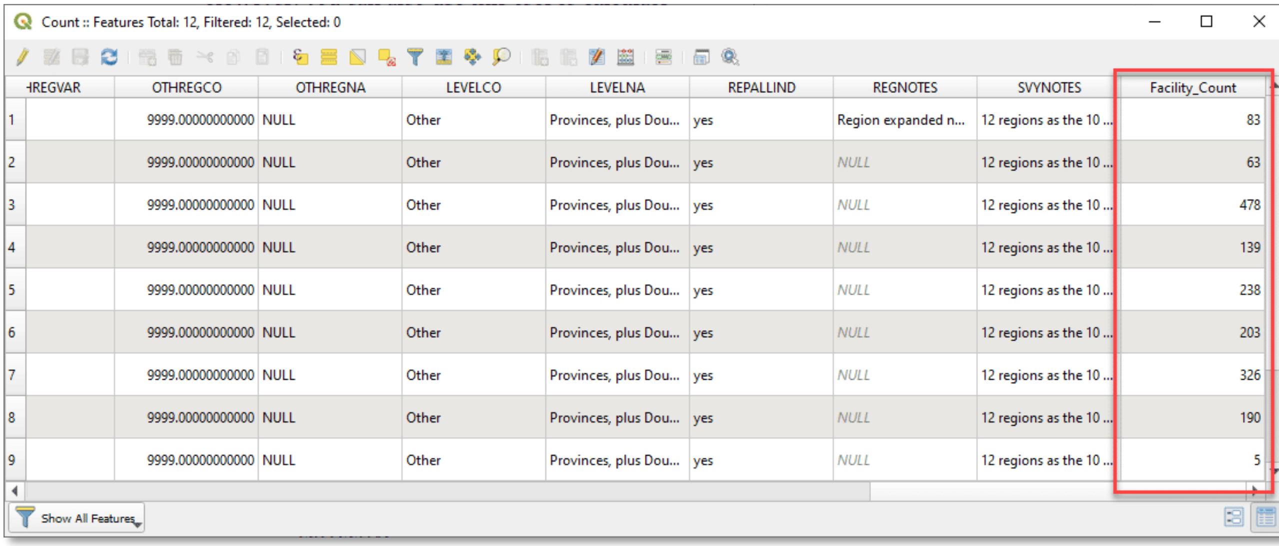
8. In the Count Points in Polygon window, select the administrative boundary layer from the Polygons dropdown and the health facility data from the Points dropdown. Change the Count field name to something clearer than 'NUMPOINTS'. This is the attribute that will be added to the polygon file that contains information on the count of points. For this exercise, change this name to 'Facility_Count'. With all of these settings selected, you can now click Run to execute the function.

9. The QGIS Count Points in Polygon function is a powerful tool. In this exercise, we are simply calculating the count of points in a polygon. However, you can also use this tool to calculate the sum of the specific attributes of points. For example, if we had a point layer that contained population data, we could use the Count Points in Polygon to add up the total population of all points that fall within each polygon. To do that, we would select the population data from the Weight field dropdown menu.



10. After running the function, you will see a second polygon layer has been added to your map and Layers Panel. This temporary layer called 'Count' should contain the count of all facilities within each administrative unit. Before we export and save this layer, we should check its attributes to make sure the Count Points in Polygon function executed correctly.

11. To do this, right-click on 'Count' in the Layers Panel and select Open Attribute Table. Scroll to the right until you reach the end of the table. You should see a new attribute added to the table called 'Facility_Count'.



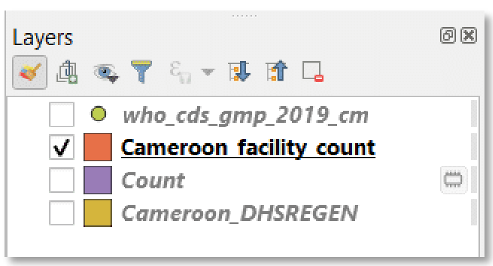
Count :: Features Total: 12, Filtered: 12, Selected: 0

	‑REGVAR	OTHREGCO	OTHREGNA	LEVELCO	LEVELNA	REPALLIND	REGNOTES	SVYNOTES	Facility_Count
1		9999.00000000000	NULL	Other	Provinces, plus Dou...	yes	Region expanded n...	12 regions as the 10 ...	83
2		9999.00000000000	NULL	Other	Provinces, plus Dou...	yes	NULL	12 regions as the 10 ...	63
3		9999.00000000000	NULL	Other	Provinces, plus Dou...	yes	NULL	12 regions as the 10 ...	478
4		9999.00000000000	NULL	Other	Provinces, plus Dou...	yes	NULL	12 regions as the 10 ...	139
5		9999.00000000000	NULL	Other	Provinces, plus Dou...	yes	NULL	12 regions as the 10 ...	238
6		9999.00000000000	NULL	Other	Provinces, plus Dou...	yes	NULL	12 regions as the 10 ...	203
7		9999.00000000000	NULL	Other	Provinces, plus Dou...	yes	NULL	12 regions as the 10 ...	326
8		9999.00000000000	NULL	Other	Provinces, plus Dou...	yes	NULL	12 regions as the 10 ...	190
9		9999.00000000000	NULL	Other	Provinces, plus Dou...	yes	NULL	12 regions as the 10 ...	5

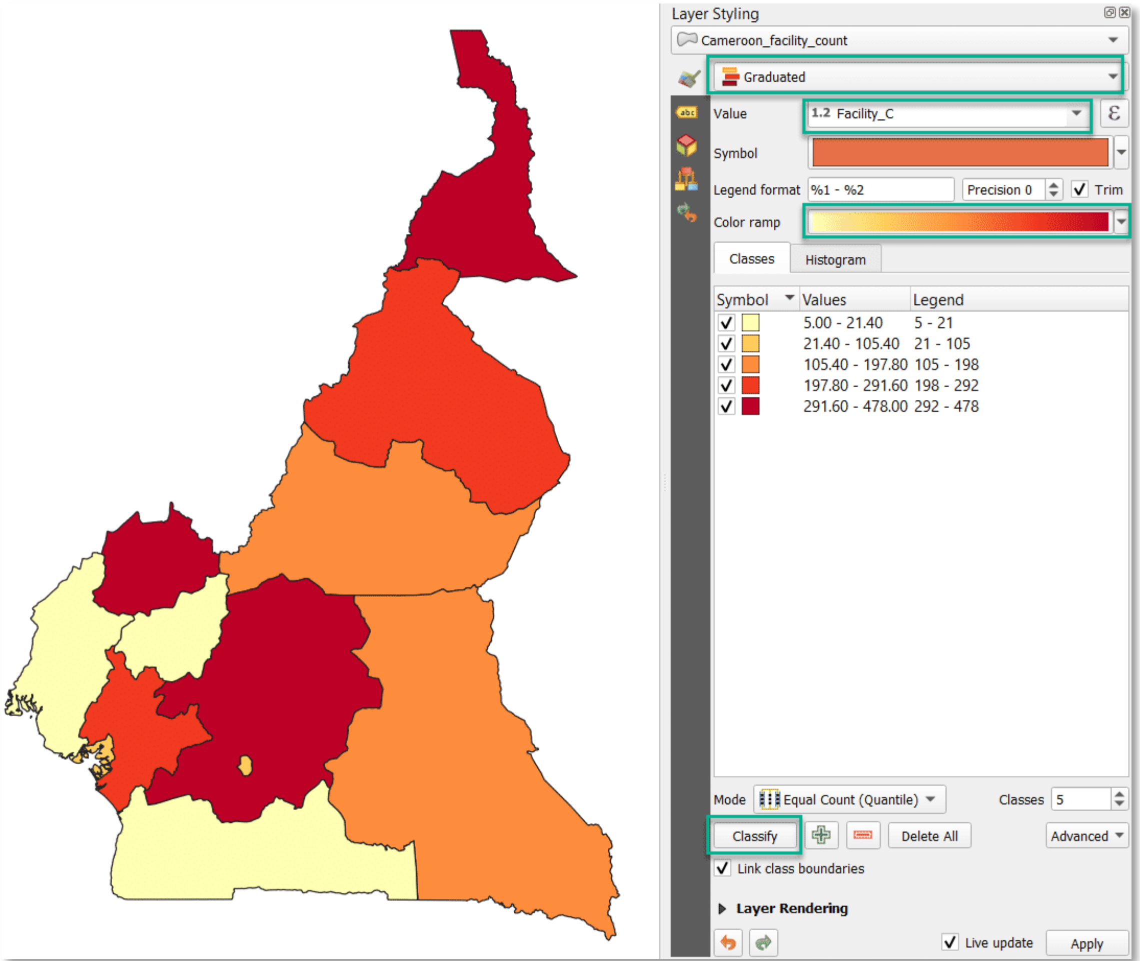
Show All Features

12. Now that you have verified the function executed correctly, let's save the 'Count' layer as a shapefile by exporting it from QGIS. Right-click on the layer in the Layers Panel and select Export → Save Features As.

13. Now that the new facility count polygon layer has been added to your map and Layers Panel, we can begin formatting the map we will export from QGIS. Before we change the symbology of our new layer, let's turn off the other layers in our Layers Panel. Uncheck the boxes next to the CSV, 'Count', and original administrative boundaries layers.



14. With only our new layer visible, let's change its symbology. To do so, open the Layer Styling Panel. (This can be opened by right-clicking on the QGIS menu and checking the box next to the panel in the dropdown menu. Or, you can click the first icon along the top of the Layers Panel.)



15. In the Layer Styling Panel, first make sure that the proper layer has been selected. The first dropdown in the menu should contain the name of the layer you want to style. In the second dropdown menu, choose 'Graduated' instead of 'Single symbol'.

16. Next to the Value option, select the 'Facility_Count' attribute from the dropdown menu. (Please note, the name of this attribute may have changes to 'Facility_C' during the export of the layer.)

17. To create graduated symbology for your layer, click the Classify button near the bottom of the Layer Styling Panel.

18. Before we leave the panel, let's change the colors. Next to the Color ramp option, click the dropdown arrow and select a different color ramp. Choose one that makes the classes look distinct. You can invert the colors by clicking the Invert Color Ramp option in the dropdown menu.

19. Our layer is now colored by the number of facilities in each administrative unit. The last thing to do before creating a print layout will be to center the polygon layer in our map. Right-click on the layer in the Layers Panel and select Zoom to Layer so that it is centered in your screen.

20. In the QGIS menu, click on Project → New Print Layout from the dropdown. In the popup box, enter a title for your Print Layout. Once you enter a title and click OK, your new Print Layout will appear in a separate window.

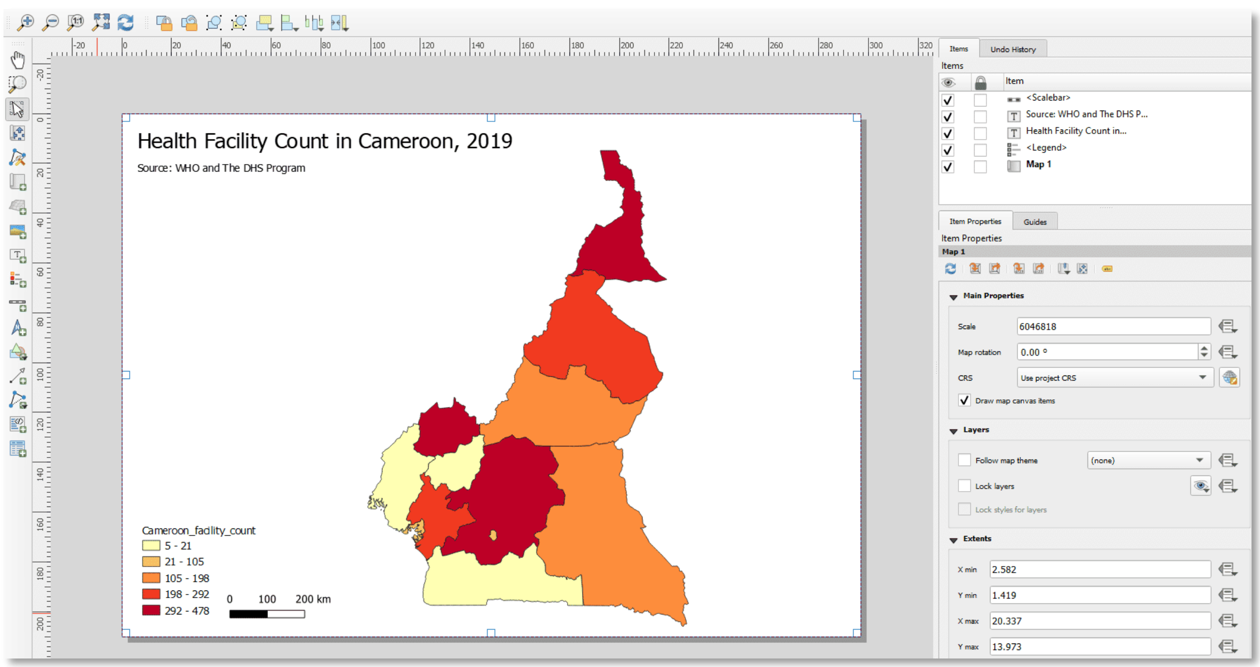
21. The Print Layout window will display a blank canvas on which you can add different map elements. To add map elements, you will select the elements from the list of icons along the left-hand side of the screen, click-hold-and-drag your mouse on the canvas, and release your mouse. Be sure your map includes all five of the elements listed below. For this exercise, be sure your map includes the following elements.

    a. Title

    b. Data source (Please note, for every case study the source should be written as 'Source: WHO GMP and The DHS Program'.)

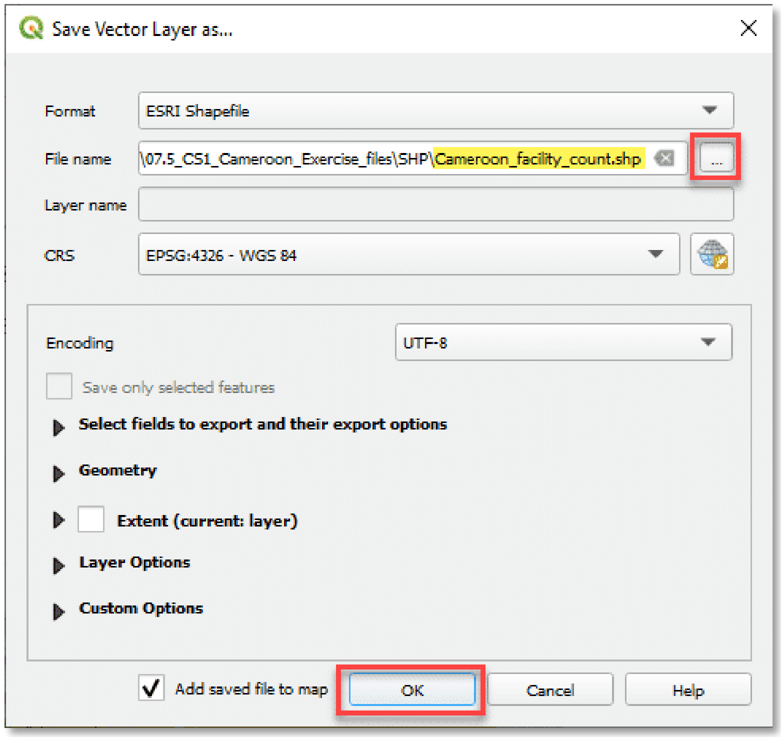
    c. Legend

    d. Scale bar



22. With these three elements added to your map and organized to your liking, you are now ready to export your map. Before we export the map, let's save our Print Layout. To do that, click on the Layout menu → Save Project.

23. To export the map, click on the Layout menu in your Print Layout window and select Export as PDF, depending on desired format.



24. In the Save Vector Layer As popup window, make sure the Format is set to 'ESRI Shapefile' and the CRS is set to 'EPSG:4326 - WGS 84'. Be sure to save the shapefile in the 'SHP folder where your current administrative boundaries are stored. Please save this new shapefile as 'Cameroon_facility_count'.

8. LABELING MAP FEATURES

8.1 CONCEPTS

Labels are essential for helping readers understand the maps you make by providing context for the different features you include. Labeling features can be tedious and complex in GIS. While the label generators built into GIS software are useful, you will often need to manually manipulate the labels after you generate them. By the end of this section, you will be able to:

1. Express the function and importance of labels in a map,
2. Create labels for features in a QGIS map,
3. Change the style of map labels in QGIS,
4. Manipulate the positioning of labels in QGIS, and
5. Create a label key in QGIS.

8.1.1 Definitions

The following terms are used in this section.

- **Callouts/leader lines**: lines indicating which labels refer to which features in a map
- **Field calculator**: a tool in QGIS used to populate the values of a field in the attribute table
- **Font style**: properties of the font that include italics, bold, and underlines
- **Label key**: a map element that matches numbers in the map that represent labels with the label content itself
- **Title/proper case**: font case that capitalizes the first letter in a word (with the exception of prepositions and articles)
- **Typeface**: the style of the characters used in text

8.1.2 Overview of Map Labels

This lesson corresponds to Activity 8.1 of the Health Data Mapping Workshop. Labeling is an essential part of map making. It can mean the difference between maps that are busy and difficult to read and maps with clear labels that direct your eyes. Using the maps below as an example, the map of Uganda on the left is busy and difficult to read while the map on the right is well-balanced.

When designing a map, there are three key questions to keep in mind: What, Where, and Who.

- What kind of data are you mapping and the purpose of the map?
- Where are you mapping?
- Who will be using your map?

While some of these questions can be addressed by other elements of a map, such as its title and cartography, labels can also help answer them. Labels can help tell the viewer what they're looking at. They can help provide context for where the content of the map appears in the real world.

Labeling is a time-consuming but very important part of the map design process.

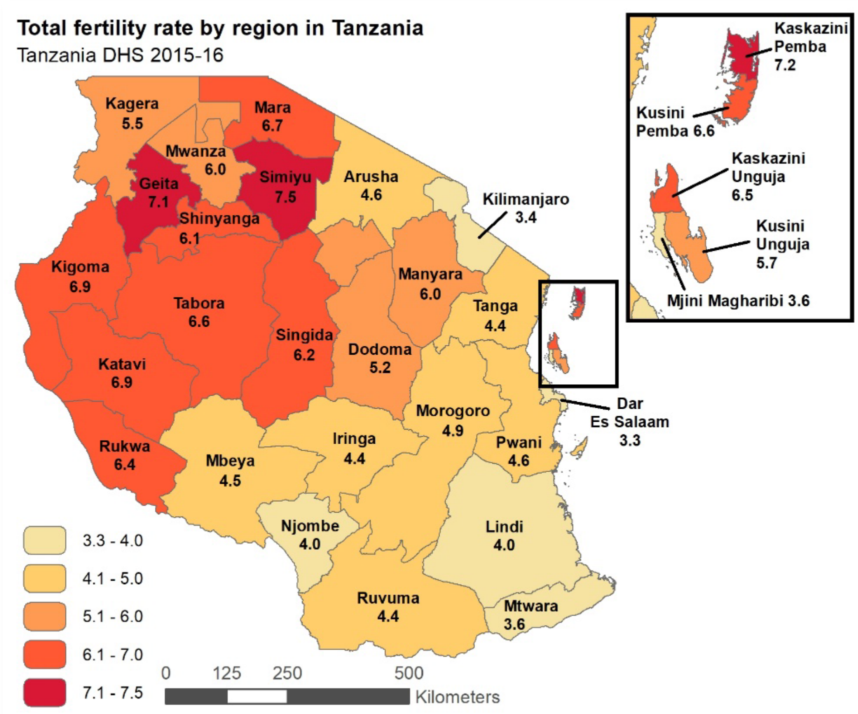
When we label, we need to consider five main features: size, color, typeface, font style (e.g. bold, italics, underlined), and position. These features all communicate aspects of map data, help create a visual hierarchy, and give a map style.

For example, part of the reason National Geographic maps are recognizable is because they consistently use a particular typeface, font style, and color scheme in their maps. Maps from The DHS Program and the United States Geological Surveys also fall into this category.

Position and Orientation

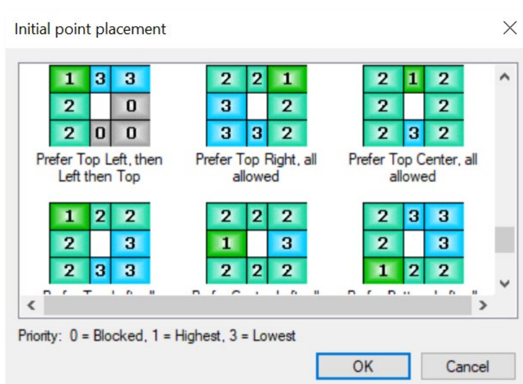
Labels should be clearly associated with the features they refer to. They should be placed in a clear position and not overlap other labels or other map content. Labels should help the reader understand the distribution of map features as well as their size. How you orient your label in relation to features depends on the type of feature you are mapping: polygons, points, or lines.

Labels for polygon features – such as regions, deserts, and forests – should be centered inside the polygon. Sometimes long names may need to be shortened, or, if the polygon is too small, need to be written outside of the polygon. In cases where the name is written outside the polygon, leader lines should be added to connect the label to the feature. As you can see, in this map of Tanzania, seven of the regions have labels outside of their borders. Leader lines identify to which boundaries these labels belong.



Labels for **points**, such as cities, are usually arranged so they are parallel with other text on the page. In GIS software, the cartographer can often choose where the labels will be placed in relation to points.

The primary location for labels is often in the top-right corner of the point. If there is no room there, because of the placement of other labels, the GIS software will place the label either in the top-left corner or bottom-right corner, which, as can be seen, both have a weight of 2. As a last resort, if there is no room around the point in other directions, the GIS software will place a label in the bottom-left corner.



In this example, we can see that the labels for different cities in Nigeria are displayed at different place in relationship to the points.

Labels for **line** features, such as rivers or roads, are typically rotated and curved to follow the shape of the line. Even in cases where the river or road runs North to South, the label is rotated to follow the path of the line.

Size

You may want to choose different font sizes for different features. For example, you could use a larger font size for cities with populations over 200,000 and a smaller font size for cities with populations of less than 200,000 people. A rule of thumb is to use a difference of at least 2 points for different classes. Labels should not be smaller than around 6-7 points for print maps and 9-10 points for maps displayed on screen. In this map, the font size for city labels reflects the maximum population estimate of that city. For example, the label for Lagos is much larger than that for Nsukka.

Color

Color variation in labels is used to denote different types of features. Labels for rivers and lakes are usually in blue, and black is often used for cities. The difference in color helps map readers differentiate the different kinds of features on your map. In this map of Tanzania, labels for water features are blue and cities are black. The country labels are light gray, making the city labels stand out more. Since the country names only provide context for the map, their colors are slightly muted.

Typeface and Font Styles

The map below highlights the use of different typefaces and font styles. There are two major types of typeface: serif and sans-serif. Serifs, like Times New Roman, have finishing strokes on letters. Sans-serifs, like Arial, on the other hand, do not have these finishing strokes.

Times New Roman (Serif font) **|** Arial (Sans-serif font)

Sans-serif has a clean and modern look whereas serif has a more classic appearance. The typeface that you chose can make a map feel formal, informal, historical, or modern. You can combine serif and sans-serif typefaces in a map. In this map, for example, the labels for natural features, specifically mountains and bodies of water, are written in a serif typeface, while all other labels are written in a sans-serif typeface. This difference in typeface makes the serif labels, the border labels, stand out.

© Sémhur / Wikimedia Commons / CC-BY-SA-3.0

Not only are the labels in the map above written in different typefaces, they also implement different font styles. Some labels are written simply in proper case. Others are written in uppercase. We also see the use of italicized text for water labels and bold text for capital cities.

As you can see, the combination of text size color, typeface, font style, position and orientation all contribute to a visual hierarchy that makes the map easy to read and interpret.

8.2 QGIS INSTRUCTIONS

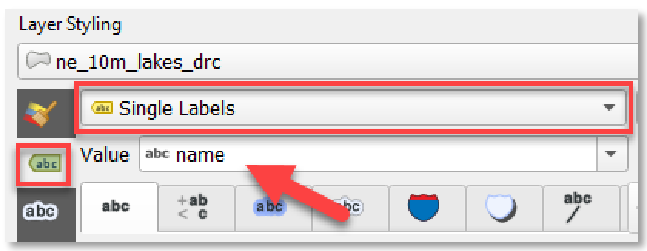
8.2.1 Labeling Features in QGIS

In this activity, we want to create a simple map of the Democratic Republic of the Congo (DRC) that shows the primary rivers and lakes in the country. We want these rivers and lakes to be named. As a reference, we will also include cities in the DRC. The cities use a graduated symbology by size. We will want the cities to be labeled as well.

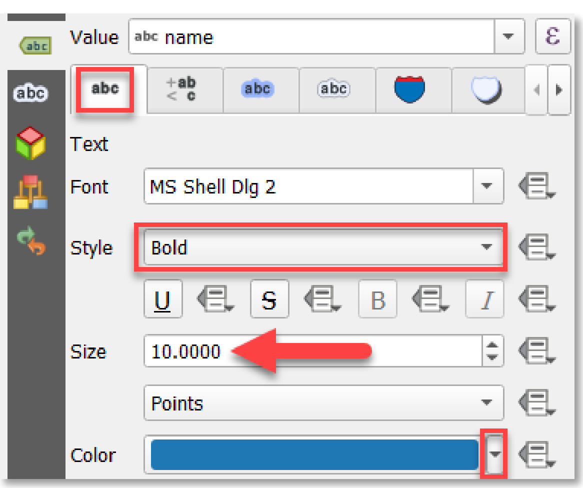
This section corresponds to Activity 8.2 of the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\8.2.1_LabelsInQGIS\ folder.

1. Open the DRC_Rivers.qgz project. There are three layers in the Layers Panel and map frame. We have DRC's rivers, lakes, major cities, and national borders.

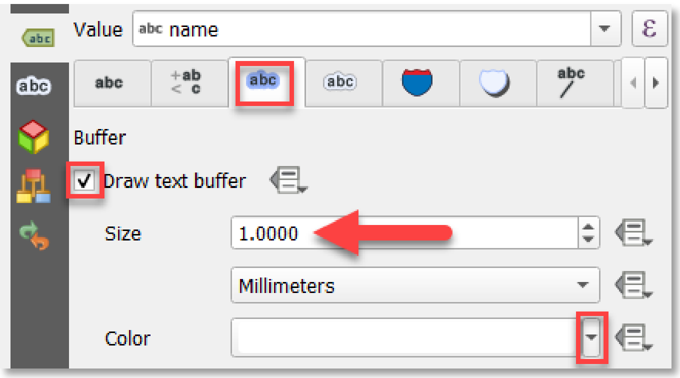
2. Let's begin by adding labels to the lakes. To do this, open the Layer Styling Panel and click on the second tab called the Labels Tab. Change the label type from 'No Labels' to 'Single Labels'. Be sure the Value is set to the variable, 'name'. This is demonstrated in the screenshot below.



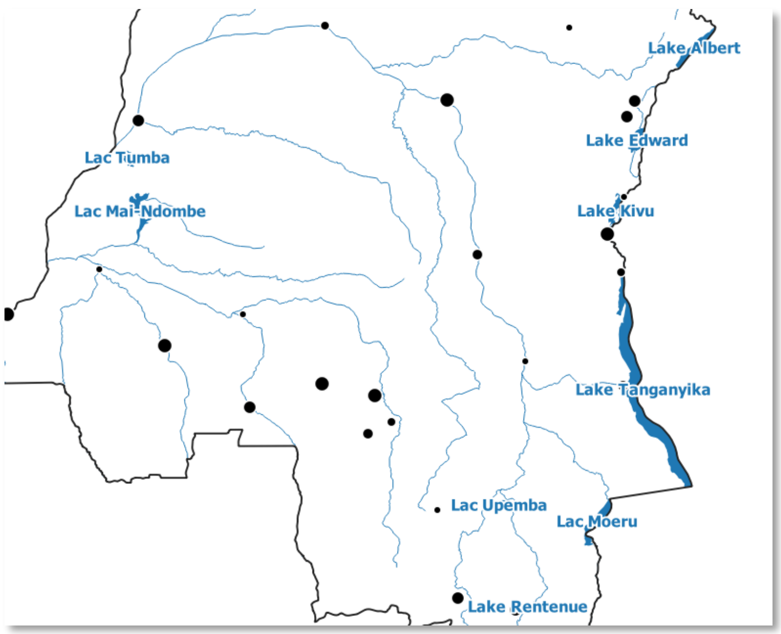
3. In the Text Subtab (the first subtab in the Labels Tab), set the Style to 'Bold', the Size to 10 points, and the Color to blue. These settings are shown below.



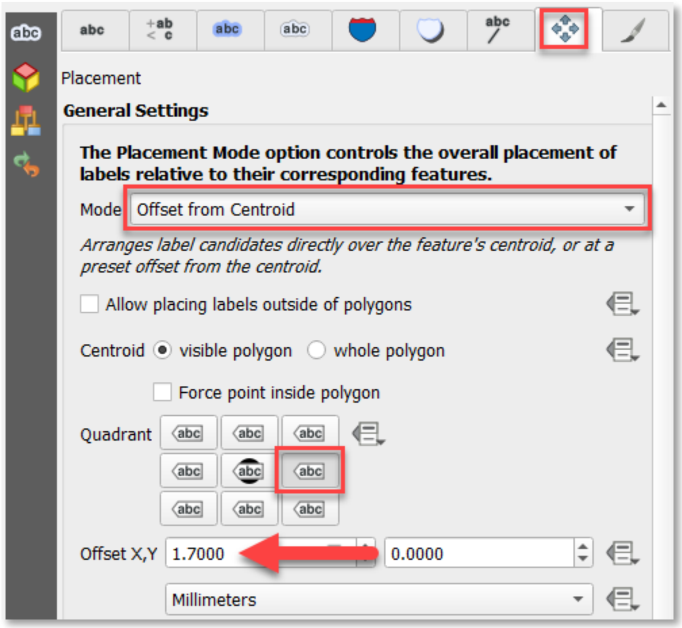
4. Next, we're going to add a buffer (also called a halo) to our labels. To do this, click on the Buffer Subtab (the third subtab) and check the box next to 'Draw text buffer'. Set the size to 1 mm and the Color to white. See the screenshot below.



The labels should now appear similar to those in the screenshot below. Note how the lake labels overlap the lakes themselves. Let's change them so that they appear off-center, to the right of the labels.

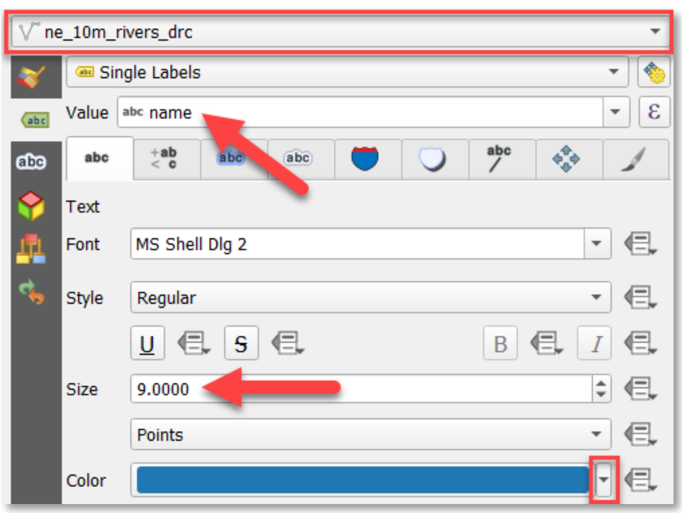


5. Open the Placement Subtab (identified below). Change the Mode from 'Around Centroid' to 'Offset from Centroid'. Set the Quadrant to the center-right quadrant and the Offset X,Y to 1.7 mm. These configurations are demonstrated in the screenshot below.

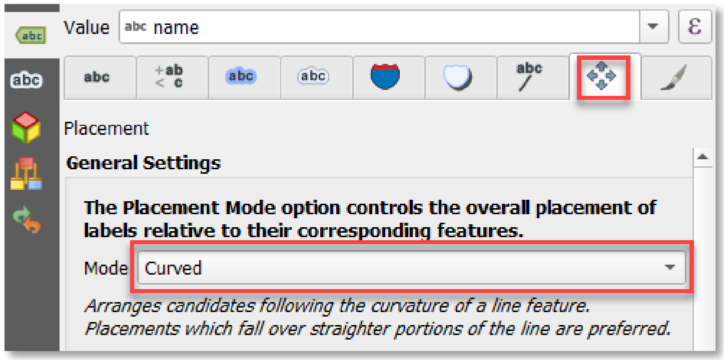


Now, let's add labels for our rivers. For the river labels, we'll configure them so that they follow the shape of the river.

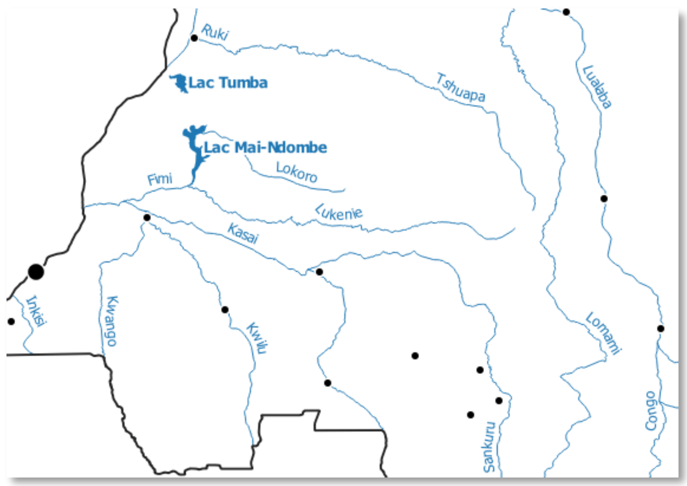
6. Open the rivers layer in the Layer Styling Panel's Label Tab. In the Text Subtab, set the Value to 'name', the Style to 'Regular', the Size to 9 points, and the Color to blue (as shown below).



7. Then, open the Placement Subtab (identified below) and set the Mode to 'Curved'.

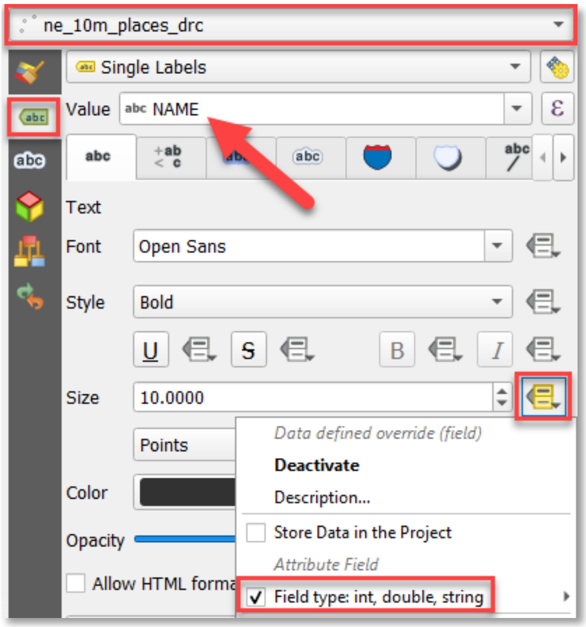


After changing the Mode to 'Curved' the labels should appear similar to the ones in the screenshot below.

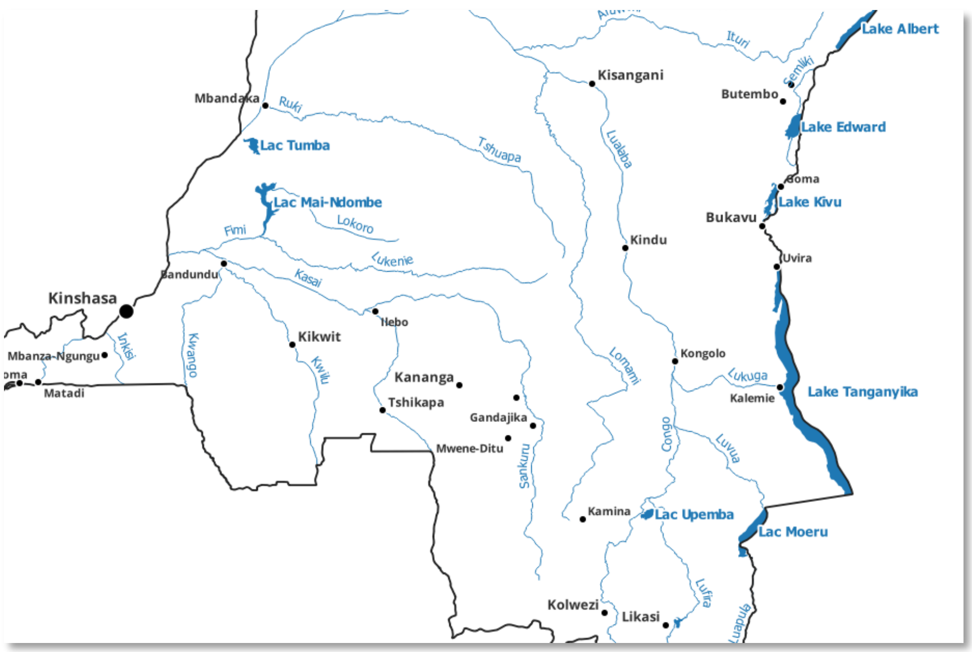


Now, we'll add labels to the cities. Rather than use a single label size for our cities, we will use a graduated size based upon an attribute in our cities layer (ne_10m_places_drc). The attribute called Label_size contains a number that corresponds to the size each label should be. In the next step, we're going to link the label's size property to the Label_size field in its attribute table.

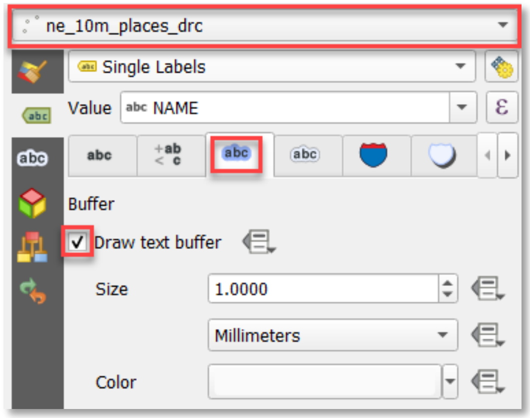
8. In the Label Tab of the Layer Styling Panel, open the properties for the cities layer (ne_10m_places_drc). Change the labeling style from 'No Labels' → 'Single Labels'. Set the Value to the field called 'NAME' and the Style to 'Bold'. Lastly, click the Data Defined Override button (identified below), hover over the Field Type setting, and select the 'Label_size' field from the list of attributes.



Your labels should now appear like those in the screenshot below.



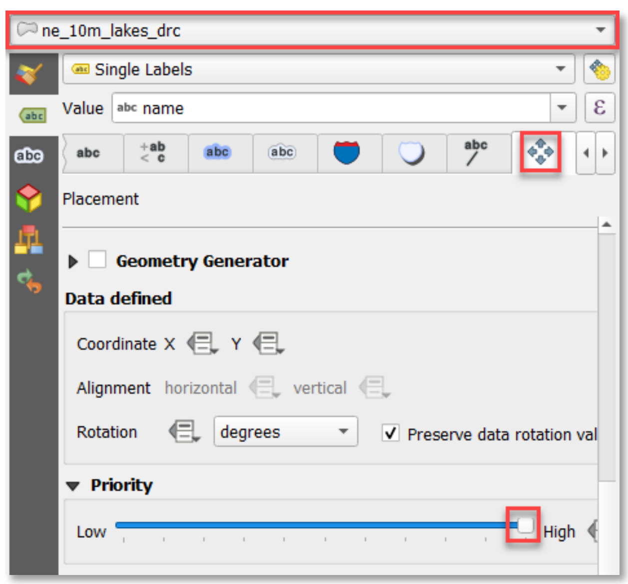
9. Now, let's add a buffer (halo) to our city labels. To do so, open the Buffer Subtab (the third subtab) and check the box next to 'Draw text buffer'. Use the default buffer settings: the Size should be set to 1 mm and the Color should be set to white.



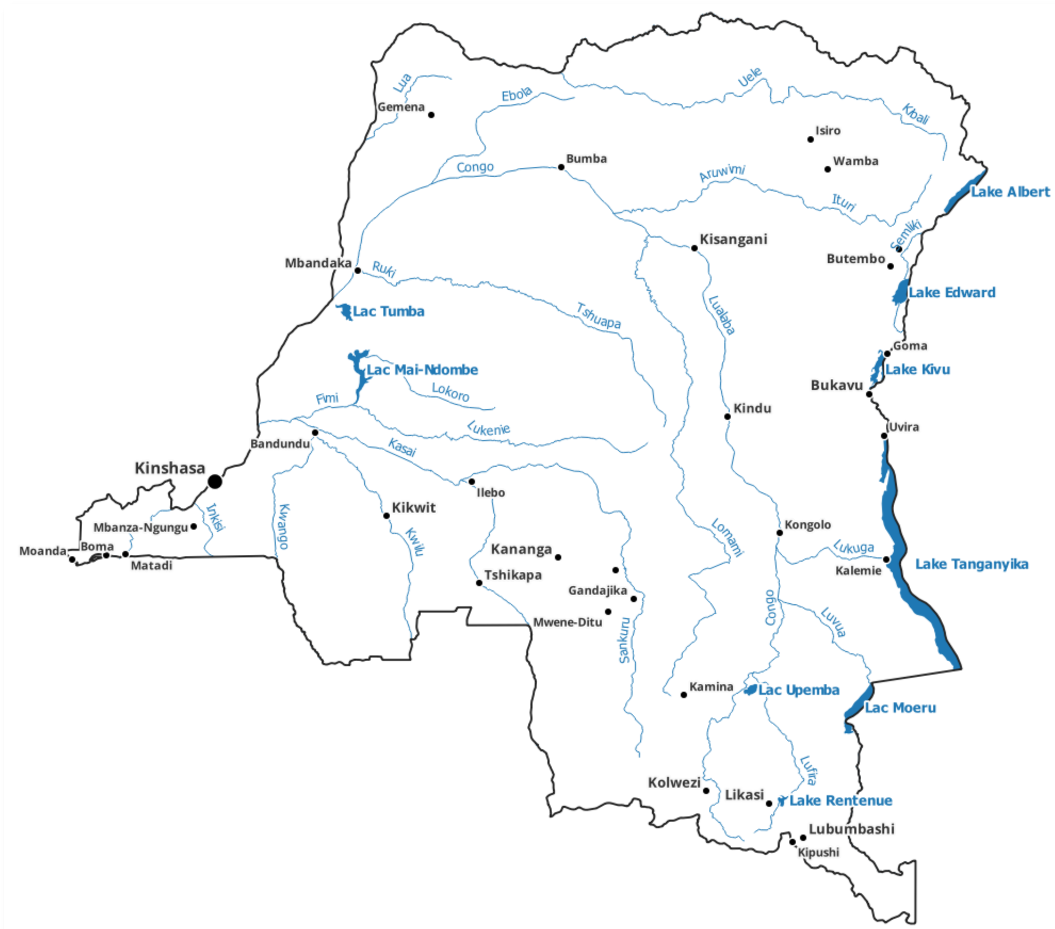
There's one small labeling detail that we need to fix. It may be difficult to notice, but one small lake in the southeast of the DRC no longer has a label. This is because there are too many labels around it, so it doesn't have the space to appear. QGIS by default does not allow labels to overlap. We only override this setting if we want to manually move the maps (a technique which will be introduced in Section 8.2.2).

To ensure that our lake labels appear, we can change their priority setting. Whereas all labels have a default priority of medium, we can change the lake labels so that their priority is relatively higher.

10. Open the lakes layer (ne_10m_lakes_drc) in the Layer Styling Panel and open the Placement Subtab (identified below). Move the indicator on the Priority scale to the right, the High side of the scale. You can set the lakes to have the highest priority labels.



You should now see Lake Rentenue's label in the southeast corner of the map. The result should look like the map below.

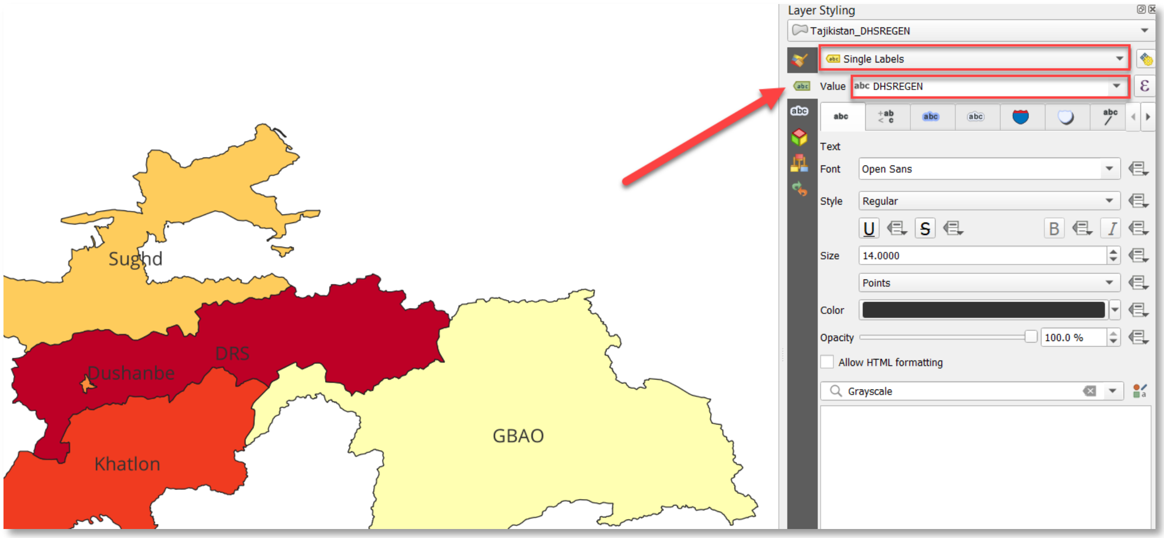


8.2.2 Creating Custom Labels in QGIS

In this activity, you will be customizing map's labels by moving them around on the map. To manually relocate a label, you need to set up attributes to store its location and configure its settings.

This section corresponds to Activity 8.3.1 of the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\8.2.2_CustomLabels\ folder.

1. To begin, open the 'Labels.qgz' project in the exercise folder. After opening the project, explore the map's content and the layer's attribute table. The country represented here is Tajikistan, and the indicator is the percentage of currently married or in union women with an unmet need for family planning (FPNADMWUNT). The indicator has already been displayed in the map.



2. Open the Layer Styling Panel if it is not displayed. (To do this, click View → Panels → Layer Styling.) With that panel open, select the second tab, the Labels tab.

In the dropdown menu, select 'Single labels'. Labels will appear on the map, but we need to change the attribute being displayed. Next to the Value text, select 'DHSREGEN' from the dropdown menu. You should now see the region names displayed on the map.

3. For this exercise, we want to display both the country names and indicator values as labels. So, we need to add both attributes to the Value option. To do this, we need to click the Expression button (Σ), to the right of the Value dropdown menu. This will open the Expression Dialog window.

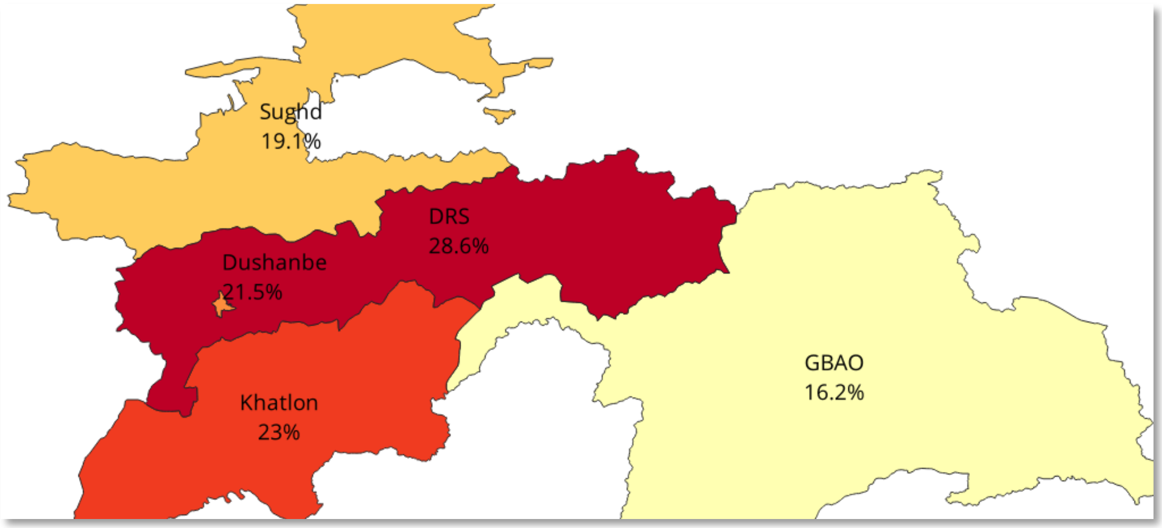
4. In the Expression Dialog window, we need to combine the region name and indicator attributes into one expression. To do this, we're going to use the `concat()` function. Within the function, we want to include the region name attribute, **DHSREGEN**, along with a line break and the indicator name, all separated by commas. The notation for a line break is `'\n'`.

5. To add the indicator attribute, begin typing the attribute (**FPNADMWUNT**) in the formula. As you type a dialog box will appear in which you can select your attribute name.

6. Then, add a percent symbol between single quotes, `‘%’`, before the end of the parentheses. Your label formula should be written as it is printed below.

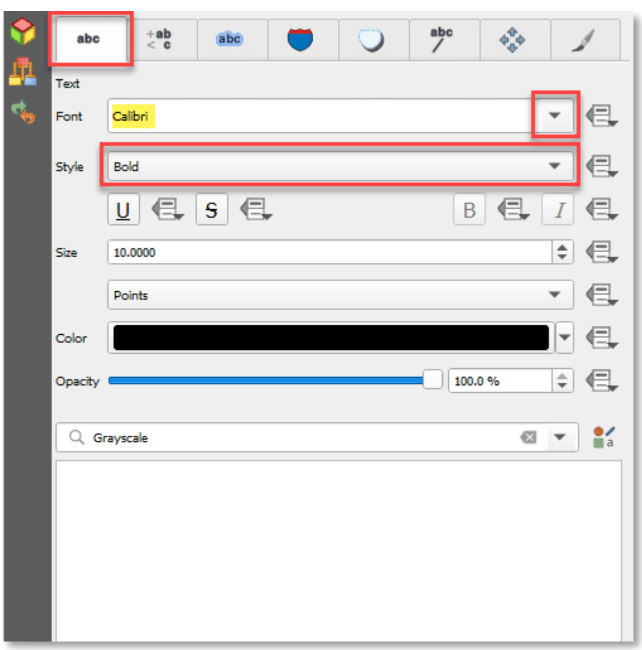
```
concat(“DHSREGEN”, ‘\n’, “FPNADMWUNT”, ‘%’)
```

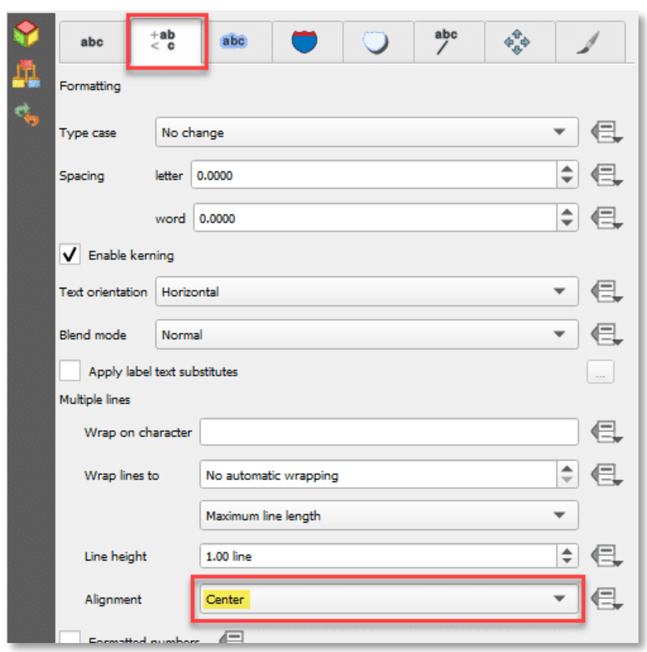
Your map should now resemble the example below.

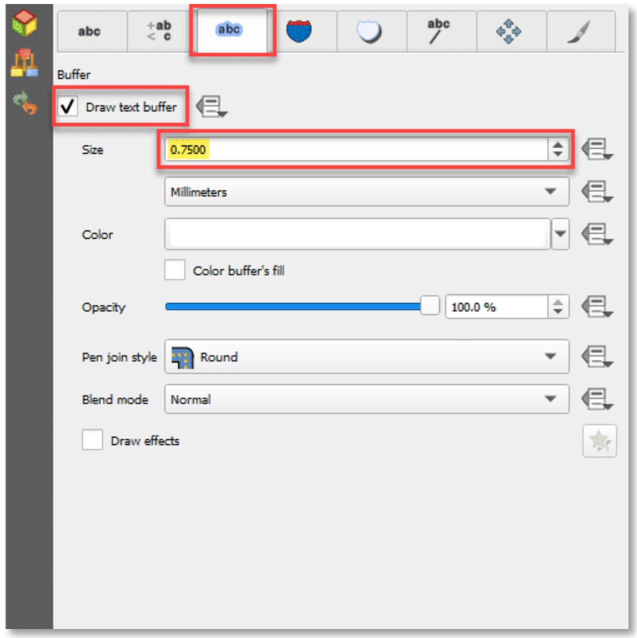


7. With the labels displayed correctly on the map, we can now change their style. The three images below show how to change the following label style settings.

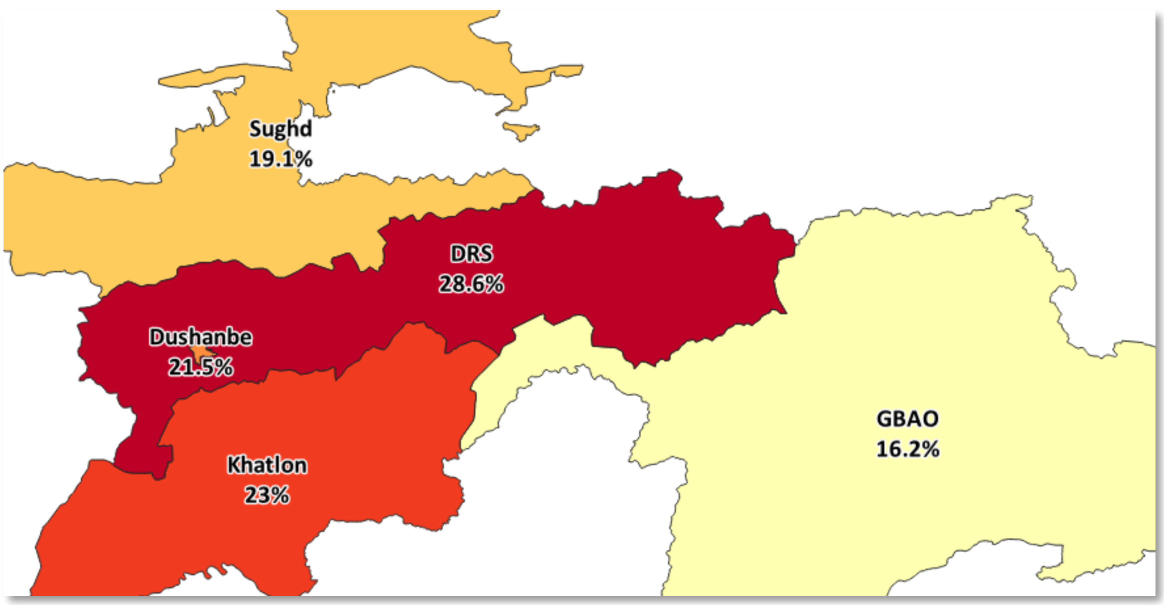
   - Font and font style
   - Alignment
   - Buffers/halos around the text







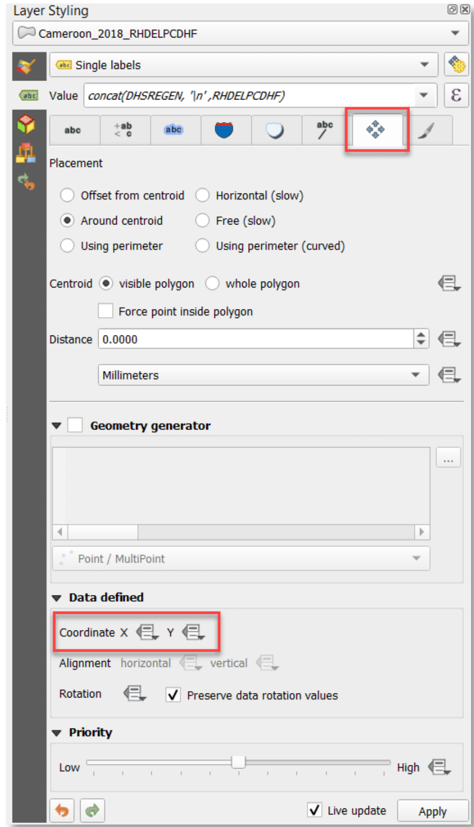
After making the changes depicted above, your labels should look similar to those in the map below.



8. In QGIS, you cannot simply drag around labels without configuring certain settings. If you navigate to the Placement Section (as shown in the image below), you can configure the settings necessary for moving the layer's labels.

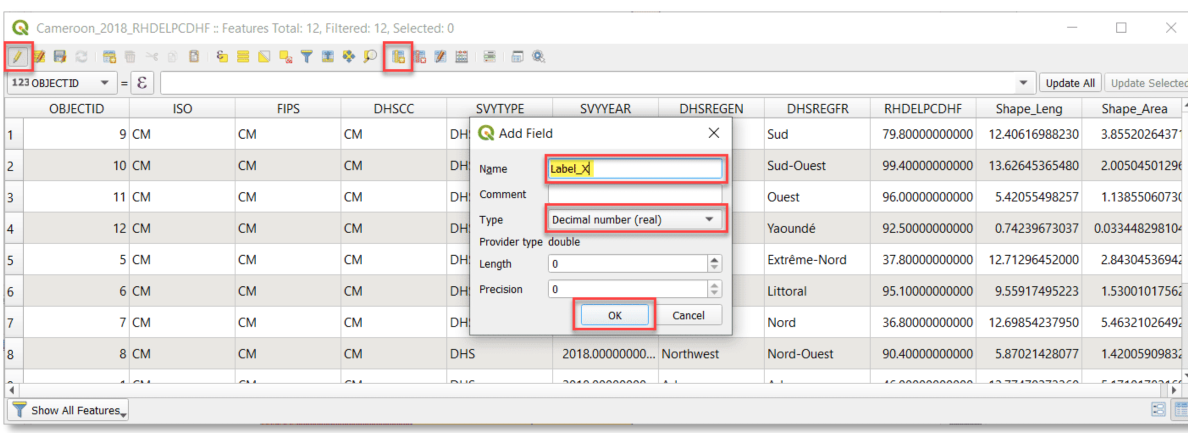
Under the Data Defined heading, you will see a Coordinate X and Y option. This option allows you to choose two attributes in your layer's attribute table within which you can save the location of your labels. In other words, here we can select two attributes in our layer's attribute table that will store the X, Y coordinates of each label. That way, QGIS can read our attribute table and know where to display the labels.

Before we can do this, however, we need to create two new attributes in our layer's attribute table.



9. To do this, right-click on the layer in the Layers Panel and select Open Attribute Table.

10. With the attribute table open, we can now turn on editing and add our new attributes. The buttons you need to click are identified in the image below. In the Add Field dialog box, create a new field called 'Label_X'. Make sure you select the Type of 'Decimal number (real)'.



11. After creating the 'Label_X' attribute, repeat **Step 3** to create a second attribute called 'Label_Y'. Be sure that this attribute is also a 'Decimal number (real)'.

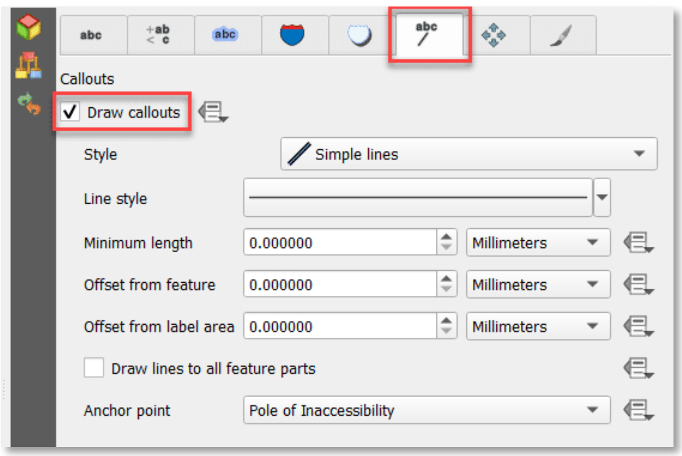
12. With the new attributes created, let's return to the Placement section of the Layer Styling Panel's Labels tab.

Below the Data defined heading click on X → Field type → 'Label_X'. Once you have defined this variable, the icon next to the letter X will turn yellow. Now, click on Y → Field type → 'Label_Y'.

The icons next to both X and Y, now defined, should appear yellow.

13. With our labels formatted and our attributes set, we can now move our labels. Before we do so, we need to adjust one more setting in the Layer Styling Panel. For small regions that might be completely covered by a label, we may want to move a label outside the boundary of the administrative boundary. For any labels outside the polygon they are labeling, it is best practice to use a leader line (i.e., callout) to connect the label to the location it describes.

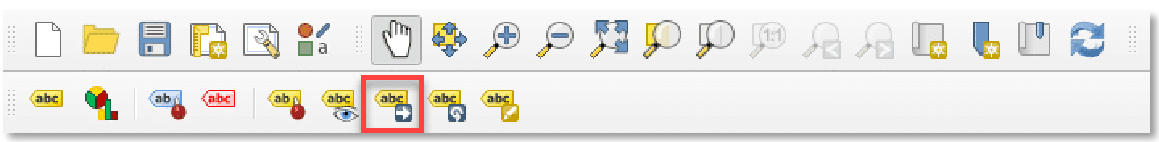
To add leader lines in QGIS, click on the Callouts section of the Labels Tab (as shown in the image below) and check the box next to the text, Draw callouts. This will automatically draw callouts for labels that are moved outside the administrative boundaries of your layer.



There are additional settings below this checkbox to modify the leader lines. For this exercise, we are going to leave these settings as they are.

14. We're now ready to move the labels around our map. To access the tools necessary to move the labels, we need to add the Label Toolbar to QGIS. Right-click on the QGIS menu bar and check the box next to Label Toolbar in the dropdown menu.

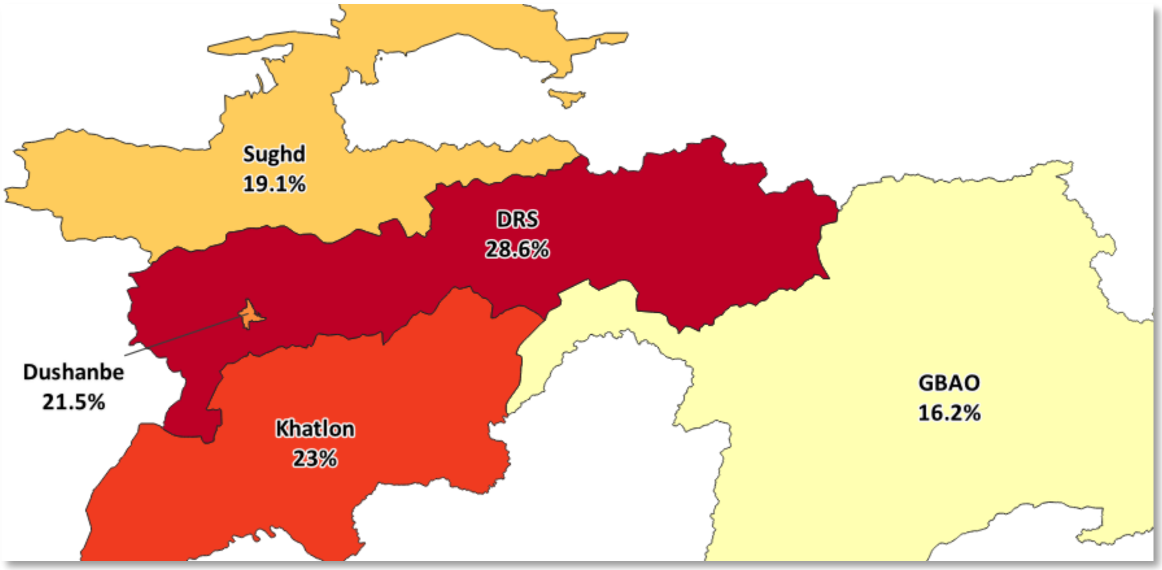
15. With the Label Toolbar added to QGIS, select the Move a Label tool (identified below).



When you click the tool, it will automatically turn editing on for your layer. Now, when you click and move a label in your map, the values in the 'Label_X' and 'Label_Y' attributes will update to reflect the coordinates of the label.

16. Using the Move a Label tool (identified above), click and move the labels around your map so that the labels are centered in our administrative boundaries. For labels that are larger than the administrative boundaries, move them outside the boundary so that the leader lines connect them to the shape which they describe. Specifically, move the Dushanbe label outside of Tajikistan's boundary and center the Sughd label within its borders.

See the image below for an example of how your labels may appear.



In the example above, note how the leader lines connect the Dushanbe label to the capital.

17. Before we create a print layout and export our map, we need to save the changes we made to our layer. Remember, when we moved the labels, the attribute table was edited to reflect the coordinates of the labels. To save these changes, right-click on the layer in your Layers Panel → Save Layer Edits. Now, turn off the layer editing by right-clicking on the layer and choosing Toggle Edits.

18. In the QGIS menu, click on Project → New Print Layout from the dropdown. In the popup box, enter a title for your Print Layout. Once you enter a title and click OK, your new Print Layout will appear in a separate window.

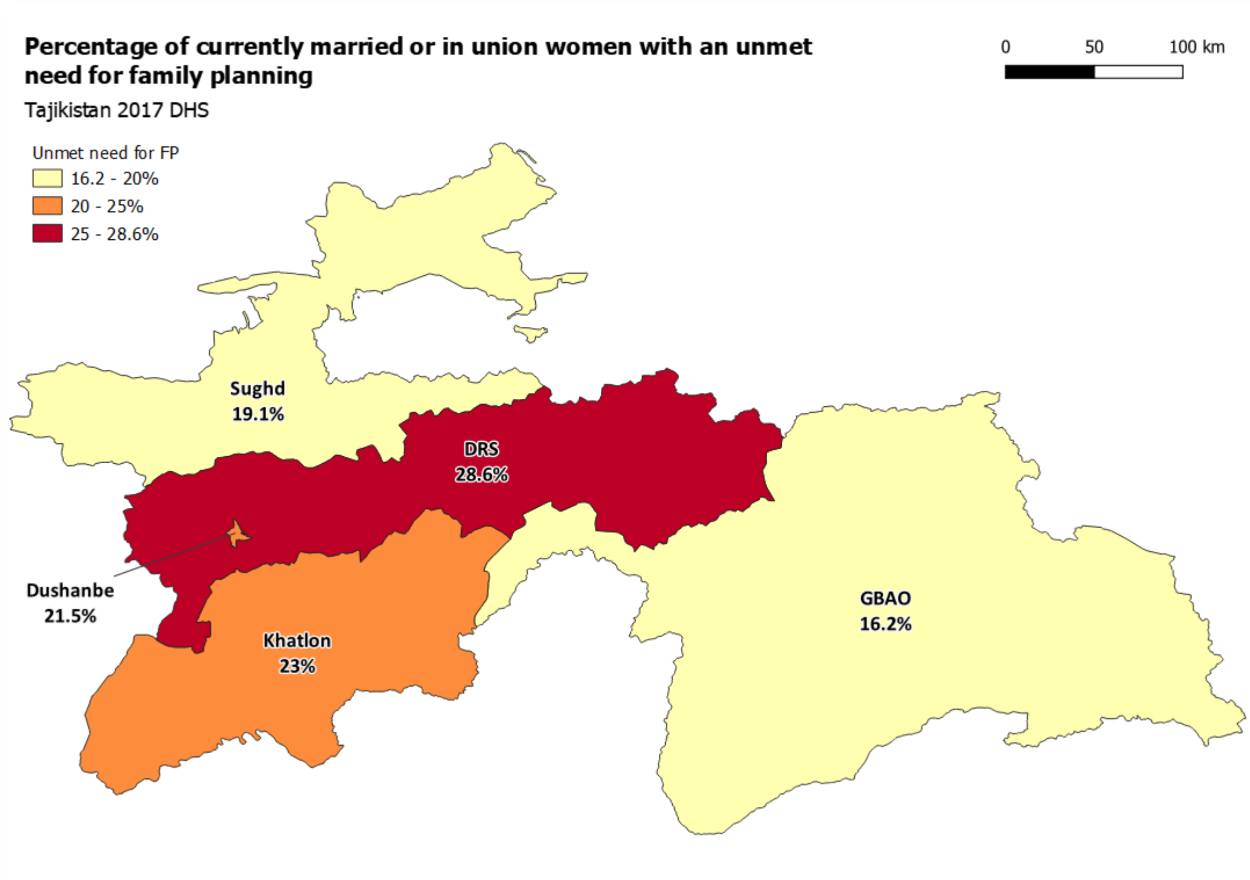
19. The Print Layout window will display a blank canvas on which you can add different map elements. To add map elements, you will select the elements from the list of icons along the left-hand side of the screen, click-hold-and-drag your mouse on the canvas, and release your mouse. Be sure your map includes all five of the elements listed below. For this exercise, be sure your map includes the following elements.

    - Title
    - Data source
    - Legend
    - Scale bar

20. For the title, please write the indicator's name. The data source should be the survey from which the data derives. Below the indicator and survey for the case study is written.

    - Tajikistan 2017 DHS
    - Percentage of currently married or in union women with an unmet need for family planning

After adding the elements to your print layout, it should look similar to the one displayed below.



21. With all the elements added to your map and organized to your liking, you are now ready to export your map. Before we export the map, let's save our Print Layout. To do that, click on the Layout menu → Save Project.

22. To export the map, click on the Layout menu in your Print Layout window and select Export as PDF, depending on desired format.

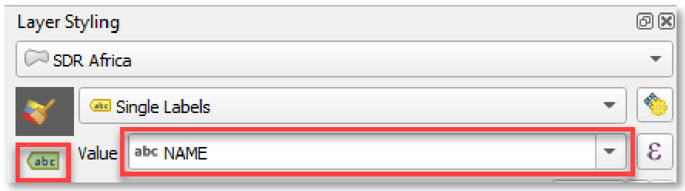
8.2.3 Creating Label Keys in QGIS

In this activity, you will be creating a map that identifies all African countries for which there is data on the Spatial Data Repository. All identified countries should be labeled. Given the varying sizes of these countries, we will be using a label key to identify countries what are too small to be labeled directly in the map frame.

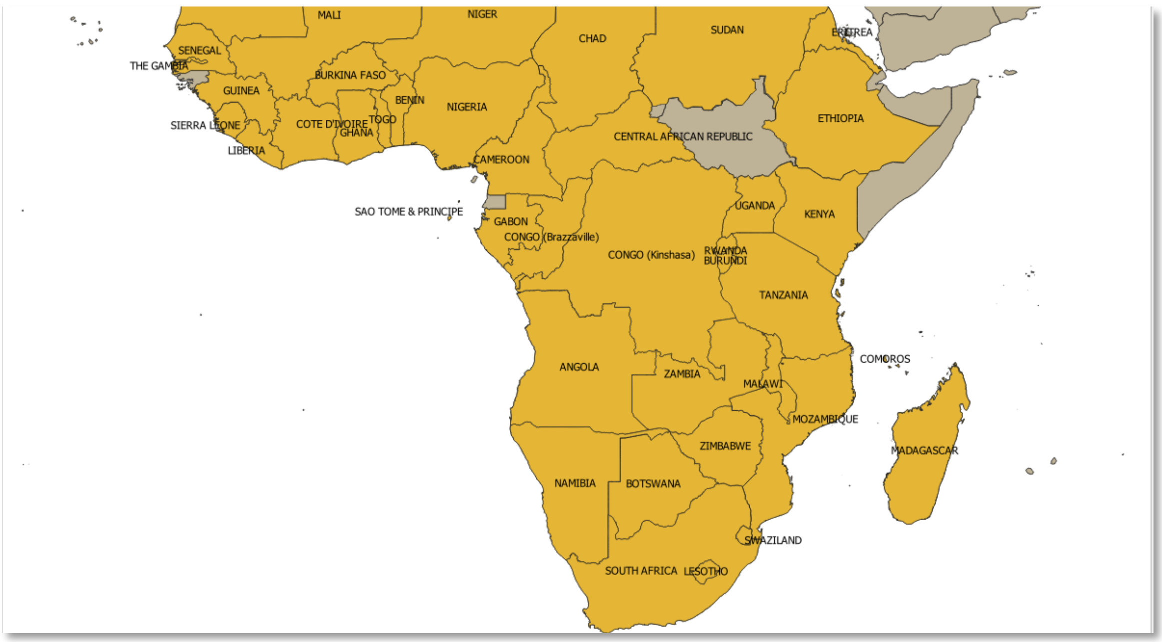
This section corresponds to Activity 8.4 of the Health Data Mapping Workshop. The exercise files for this section can be found in the \DHS_QGIS_2022\Exercise_Files\8.2.3_LabelKeys\ folder.

1. Open the Africa.qgz project in the exercise folder mentioned above. You'll find two shapefiles in the Layers Panel: SDR Africa and NE World. The first is data for countries in Africa from the Spatial Data Repository. The latter is a shapefile of all countries from Natural Earth Data.

2. Let’s add labels to our SDR Africa layer. To do so, open the Layer Styling Panel and click on the second tab, the Labels tab. Change the labeling style from ‘No Labels’ to ‘Single Labels’ and set the Value to the NAME field (as shown below).



The result should look similar to the map below.

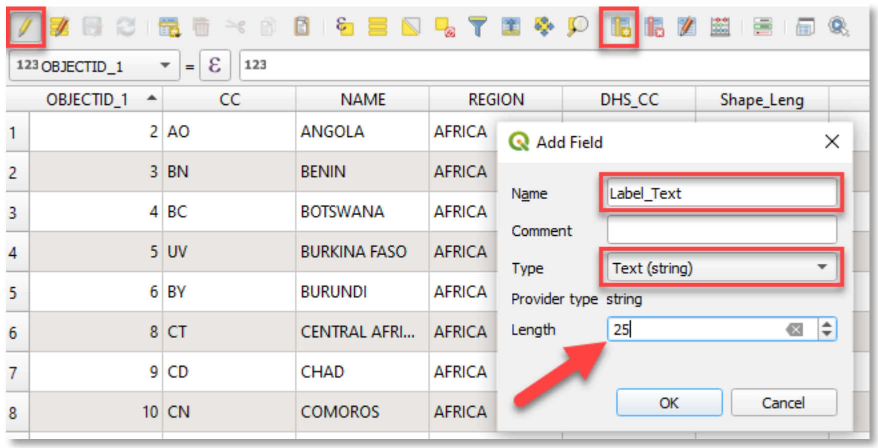


3. Note which labels will be best to relocate and identify with callouts (leader lines) and which labels should be replaced with numbers. While there’s no ‘correct’ answer (as this is a subjective question), for this exercise let’s replace the following overlapping labels with numbers and a label key: Benin, Burkina Faso, Burundi, Central African Republic, Congo (Brazzaville), Eritrea, Eswatini (called Swaziland in this dataset), Lesotho, Malawi, Rwanda, and Togo. The following labels will be relocated and identified with callouts: Cote d’Ivoire, The Gambia, Liberia, Mozambique, Sierra Leone, and Tunisia.

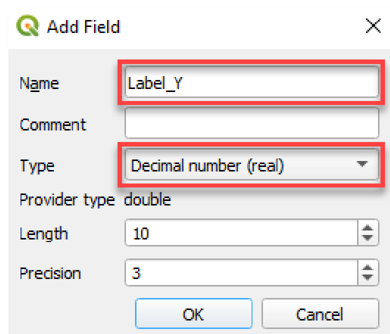
Before creating the label key, we need to create the labels themselves. To correctly format the labels, we need to add three new fields to the SDR Africa attribute table: one field for custom label text and two fields for the label coordinates.

4. Open the attribute table for the SDR Africa layer by right-clicking on it in the Layers Panel and selecting Open Attribute Table.

5. Click the pencil button on the far left of the attribute table’s toolbar to toggle editing on and click the New Field button (both identified below). Create a new field called Label_Text as Text (String) field with a length of 25 characters.



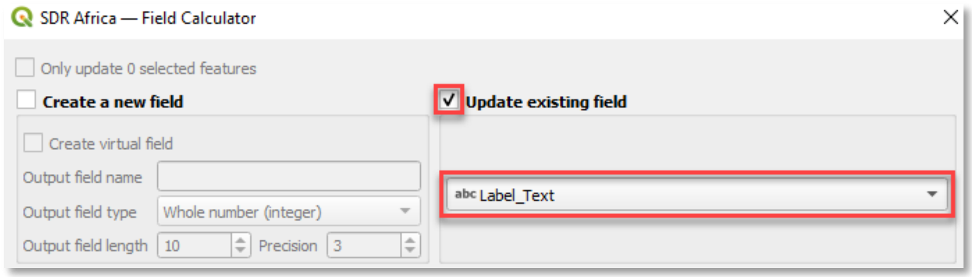
6. Then, create two additional fields called Label_Y and Label_X to store the coordinates that will determine where the labels are placed. These fields should both be set to Decimal Numbers.



Now that we've created the three new fields in the attribute table, let's populate the Label_Text field with the values we want displayed in our labels. For most countries, we want the names in title case (proper case). For the labels identified as those to be replaced with numbers (see **Step 3** above) we need to populate the Label_Text with those numerals. First, we'll populate all the values in the Label_Text field with the country names in title case, then we'll replace some with numbers.

7. Click on the Open Field Calculator button (identified below).

8. In the Field Calculator, check the box next to 'Update Existing Field' and set the field to Label_Text.



9. In the expression box type (or create) the following expression. The `title()` formula converts text into text in title case (proper case) where the first letter of each word is capitalized. Click OK to run the expression.

`title("NAME")`

This formula works for most countries in the shapefile, but there are three that were improperly altered. The two countries with Congo in their name should have the capital city (the name between parentheses capitalized) and Cote d'Ivoire should be written as such. The *d* should be lower case and the *I* should be capitalized.

10. Change the following names in the Label_Text field: Congo (kinshasa) → Congo (Kinshasa) and Cote D'ivoire → Cote d'Ivoire.

Now, we'll replace some label names with numbers. These numbers will be used to identify labels in our label key. Rather than alphabetical order, the labels will be in order from west-to-east. So, we will replace the labels with numbers in the following order: Burkina Faso, Togo, Benin, Congo (Brazzaville), Central African Republic, Burundi, Rwanda, Lesotho, Eswatini (called Swaziland in this dataset), Malawi, and Eritrea.

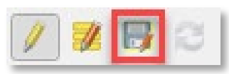
11. Change the following names in the Label_Text field.

Burkina Faso → 1	Rwanda → 7
Togo → 2	Lesotho → 8
Benin → 3	Swaziland → 9
Congo (Brazzaville) → 4	Malawi → 10
Central African Republic → 5	Eritrea → 11
Burundi → 6	

The result should look similar to the screenshot below.

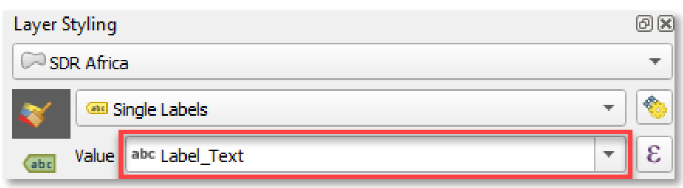
EDLITRWLIT	HCELECHELC	Label_Text	Label_Y	Label_X
58.10000000000	41.60000000000	Angola	NULL	NULL
34.10000000000	35.60000000000	3	NULL	NULL
9999.00000000000	9999.00000000000	Botswana	NULL	NULL
22.50000000000	13.10000000000	1	NULL	NULL
67.90000000000	8.70000000000	6	NULL	NULL
9999.00000000000	3.00000000000	5	NULL	NULL
22.10000000000	7.70000000000	Chad	NULL	NULL
63.30000000000	69.30000000000	Comoros	NULL	NULL
82.20000000000	41.60000000000	4	NULL	NULL
58.90000000000	15.20000000000	Congo (Kinshasa)	NULL	NULL
37.70000000000	55.80000000000	Cote d'Ivoire	NULL	NULL
9999.00000000000	22.90000000000	11	NULL	NULL

12. Save the edits you made to the attribute by clicking the Save Edits button (identified below).



Now that we've updated the label text. Let's change our labeling properties so that the new text field is shown on the map.

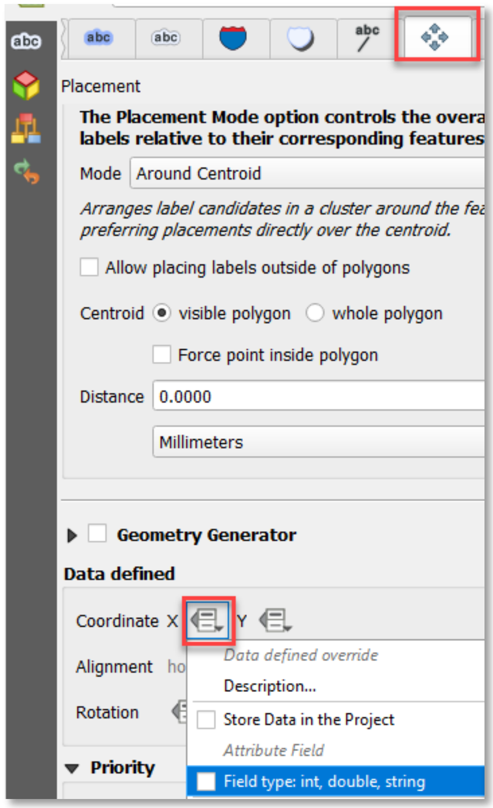
13. In the Label tab of the Layer Styling Panel, change the Value from 'NAME' to 'Label_Text' (shown below).



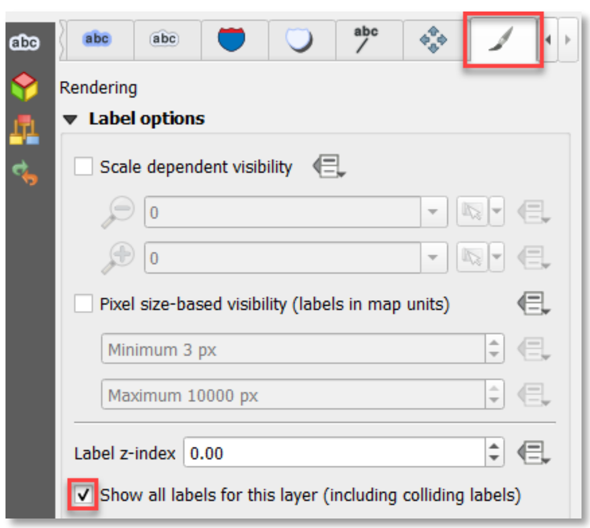
Your map should now show several of the country names replaced by labels.

The next step will be to adjust the label placement for the labels that need to be moved. To do this, we need to assign the Label_Y and Label_X fields to the labeling properties that determine where the labels should be placed.

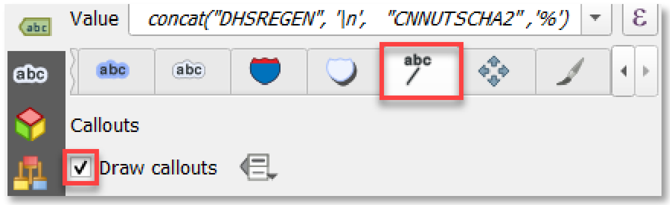
14. In the Label Tab of the Layers Styling Panel, navigate to the Placement Options. Under the Geometry Generator section, set the Coordinate X to the Label_X field by clicking on the Coordinate X button → Field Type → Label_X. This is shown in the screenshot below. Then, assign Coordinate Y to the Label_Y field.



15. Open the Rendering Options (the last tab in the Label Tab of the Layer styling Panel). In the first section, check the box next to 'Show all labels for this layer'.



16. Click on the Callouts Option of the Label Tab (in the Layer Styling Panel) and check the box next to 'Draw callouts' (shown below).



17. Before we can adjust our label placement, we will need to make sure the Label Toolbar is activated. Click on View → Toolbars → Label Toolbar to turn it on.

18. Now, use the Move a Label Tool on the Label Toolbar (shown below) to move the labels around. To move a label, activate the Move a Label Tool, click on the label, and then click where you want it on the map. Move all the labels so that they don't overlap with regional borders. Since the Dar es Salaam and Kilimanjaro labels will not fit in their respective administrative areas, move them outside of Tanzania's borders. We'll use callouts to connect them with their admin areas.

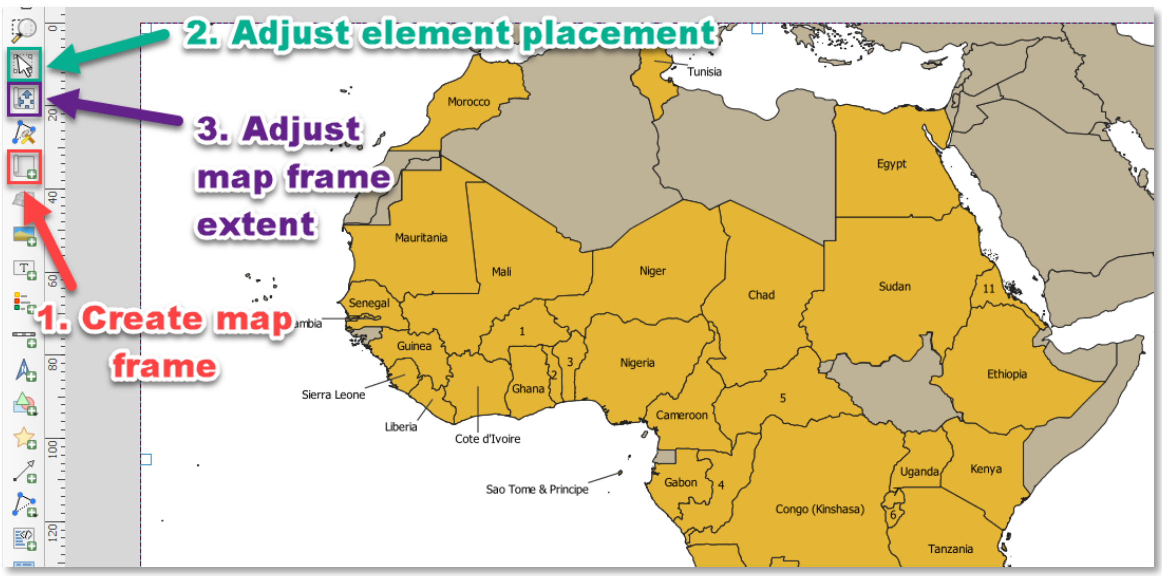
Now, let's save the edits to our SDR Africa layer. It's important to save the edits to our layer because we have generated coordinates in the Label_Y and Label_X fields by moving labels around the map. The formatted labels should look similar to those in the screenshot below.

19. To save the edits we made to SDR Africa, right-click on the layer in the Layers Panel and select Save Layer Edits. We can also toggle off editing, since we will not be making additional edits until later in this section. So, click Toggle Editing to deactivate layer editing.

With the labels formatted, we can now create the map layout.

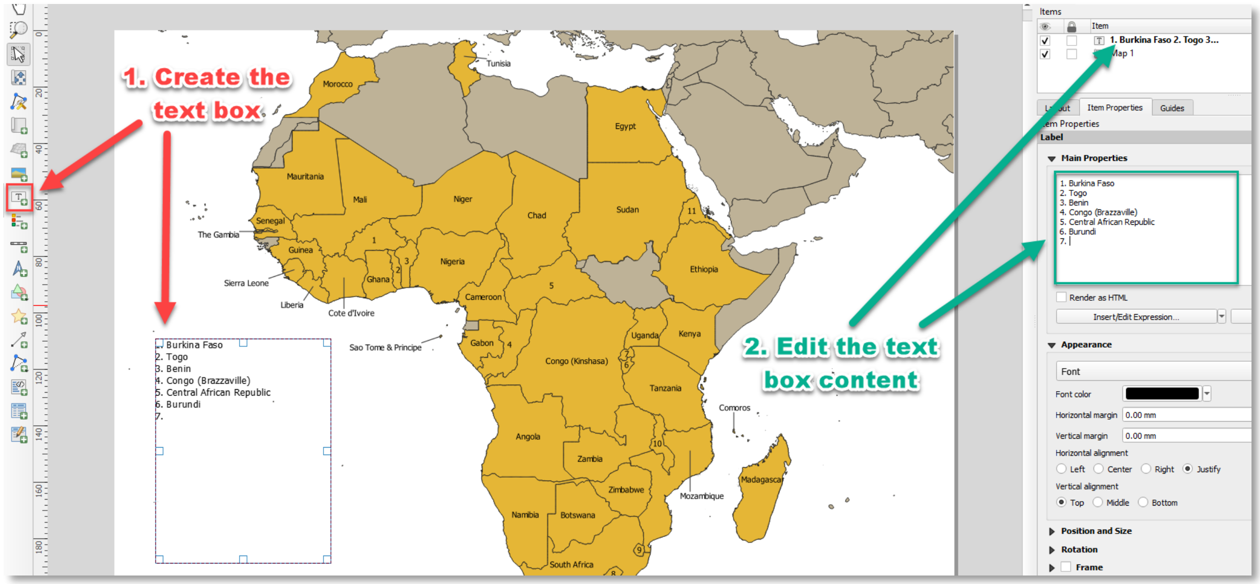
20. Create a new print layout by clicking Project → New Print Layout. You can call this layout "Map with Label Key".

21. In the new print layout, begin by adding the map frame. For a reminder about how to add elements to a print layout, see **Subsection 4.2.2** of this document.



One simple way to create a label key is to add a text box to our map that will contain the numbers and names of our different countries.

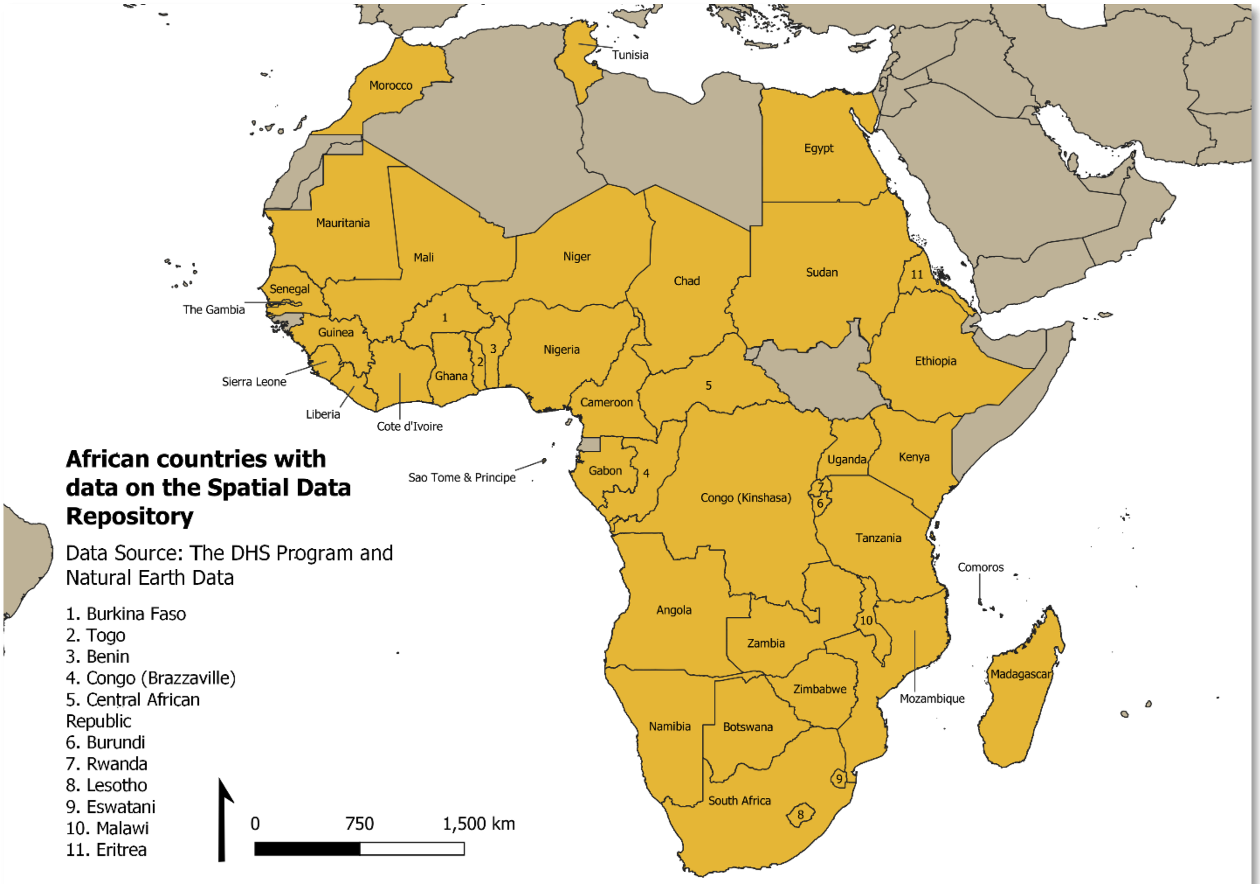
22. Add a textbox to your map layout by using the Add Label button identified (in red) below.



23. In the Main Properties of the textbox you created, add the label numbers and text with each label on a new line. This is demonstrated (in green) above. See **Step 11** of this activity for a list of which numbers correspond to which country labels.

24. After you have formatted your label key, add the following elements to your map layout: detailed title ("African countries with data on the Spatial Data Repository"), data source (The DHS Program and Natural Earth Data), scalebar, and north arrow.

The result should look similar to the map below.


African countries with data on the Spatial Data Repository
Data Source: The DHS Program and Natural Earth Data
1. Burkina Faso
2. Togo
3. Benin
4. Congo (Brazzaville)
5. Central African Republic
6. Burundi
7. Rwanda
8. Lesotho
9. Eswatani
10. Malawi
11. Eritrea
0 750 1,500 km
Tunisia
Morocco
Egypt
Mauritania
Mali
Niger
Chad
Sudan
Senegal
The Gambia
Guinea
Sierra Leone
Liberia
Cote d'Ivoire
Ghana
Nigeria
Cameroon
Ethiopia
Sao Tome & Principe
Gabon
Congo (Kinshasa)
Uganda
Kenya
Tanzania
Comoros
Angola
Zambia
Mozambique
Madagascar
Zimbabwe
Namibia
Botswana
South Africa

9. ADVANCED MAP LAYOUTS

9.1 CONCEPTS

In some cases, maps will require complex layouts that include multiple map frames. This section will explore how to create map layouts with extent indicators and insets using QGIS. By the end of the section, you will be able to:

1. Create map layouts with extent indicators and
2. Create map layouts that include inset maps.

9.1.1 Definitions

This section uses the following three terms.

- **Extent indicator**: in a map layout with two maps, the extent indicator is a smaller map that shows where the larger map is located in reference to surrounding geographic features
- **Inset map**: in a map layout with two maps, the inset map is a smaller map that shows a portion of the primary map at a larger scale (i.e., more zoomed in)
- **Overview box**: a rectangle that indicates the extent of an inset map within the primary map

9.1.2 Advanced Map Layouts

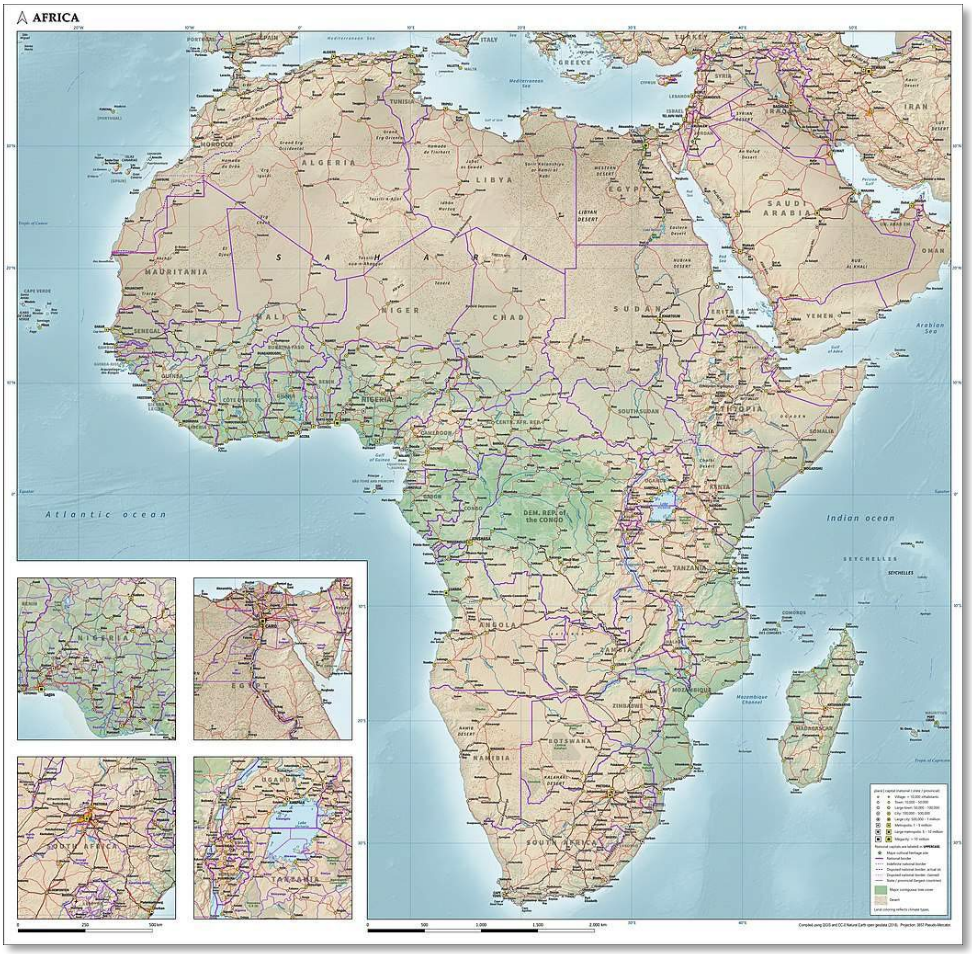
In the previous section, **Subsection 8.2.2**, we explored how to create a label key in QGIS. Label keys are an important element in map layouts where labels would otherwise be crowded. This advanced layout may be particularly useful in the country or region where you work, especially if administrative units or geographic features vary greatly in size.

While label keys are useful, they are but one advanced map layout we can use to add further context to our maps. Both extent indicators and inset maps can be used in QGIS Print Layouts to provide more information as well.

Extent indicators (also called locator maps) can provide additional context for our map readers. The map below of Tanzania has an extent indicator in the bottom-left corner.

In this example, the extent indicator tells the map reader where Tanzania is located in Africa. A cartographer can use extent indicators on larger scale maps (maps zoomed-in more) as well. On city maps, it's common to include an extent indicator of where the city is located regionally.

In the bottom-left corner of the map below, the cartographer added four inset maps. Inset maps and label keys are used similarly. Both seek to provide detail about features in a map that are too small to see.



In the case of label keys, features are labeled with shorter names or numbers so that the user can identify them using a key.

Inset maps are small maps featured in the layout of a larger map. Often these maps are displayed at a larger scale (i.e., they are zoomed-in more than the main map). In this example, the inset maps provide more detail on southern Nigeria, the Nile River, the area around Lake Victoria, and Johannesburg.

9.2 QGIS INSTRUCTIONS

9.2.1 Extent Indicators in QGIS

For this activity, we will be creating a map of different national income levels for countries in sub-Saharan Africa. To orient map readers who may be unfamiliar with this region, we will include an extent indicator to show where these countries are located in relation to Europe and western Asia.

This section corresponds to the first part of Activity 9.1 of the Health Data Mapping Workshop. This activity uses the shapefile in the \DHS_QGIS_2022\Exercise_Files\9.2.1_ExtentIndicators\ folder.

1. Open a new QGIS project and save it in the exercise folder mentioned above.

2. We’re now going to add two copies of the World_ne.shp to our project, but let’s begin by adding one copy. In the Browser Panel, expand the Project Home folder and add the World_ne.shp file from the \Shapefiles\ folder into your QGIS project by dragging it into the map frame.

3. So that we don’t get the two duplicate layers confused, let’s rename this first copy of World_ne so that it reads “SSA” for sub-Saharan Africa. To do so, right-click on the layer in the Layers Panel and click Rename.

4. Drag a second copy of the World_ne.shp from the \Shapefiles\ folder in your Project Home folder into your project. You should now see a second copy of World_ne layer in your Layers Panel.

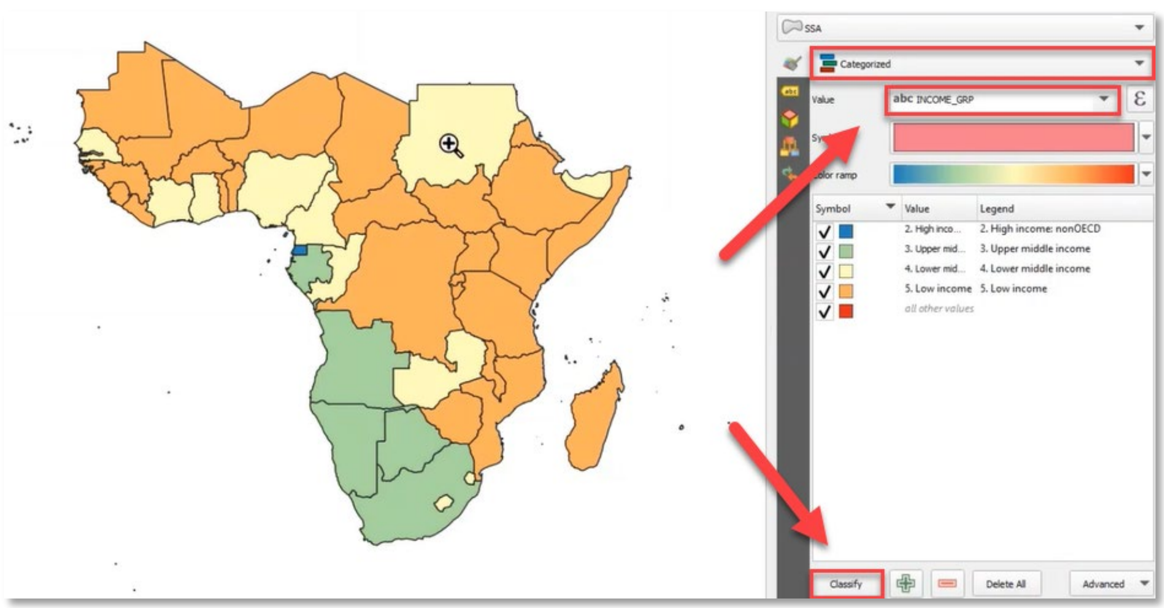
5. Uncheck the box next to the World_ne layer in the Layers Panel to turn it off.

6. As mentioned in the activity description above, we only want to display data for countries in sub-Saharan Africa. We’re now going to filter the SSA layer. Right-click on the SSA layer in the Layers Panel and select Filter.

7. Create the following query and click OK.

```
"REGION_WB" = 'Sub-Saharan Africa'
```

8. We now want to add categorical symbology to the layer. Open the Layer Styling Panel. Be sure that the layer is set to ‘SSA’, the symbology type is set to ‘Categorized’, and the value is set to the variable ‘INCOME_GRP’. Click classify to add this symbology classification to your layer (as shown in the screenshot below).



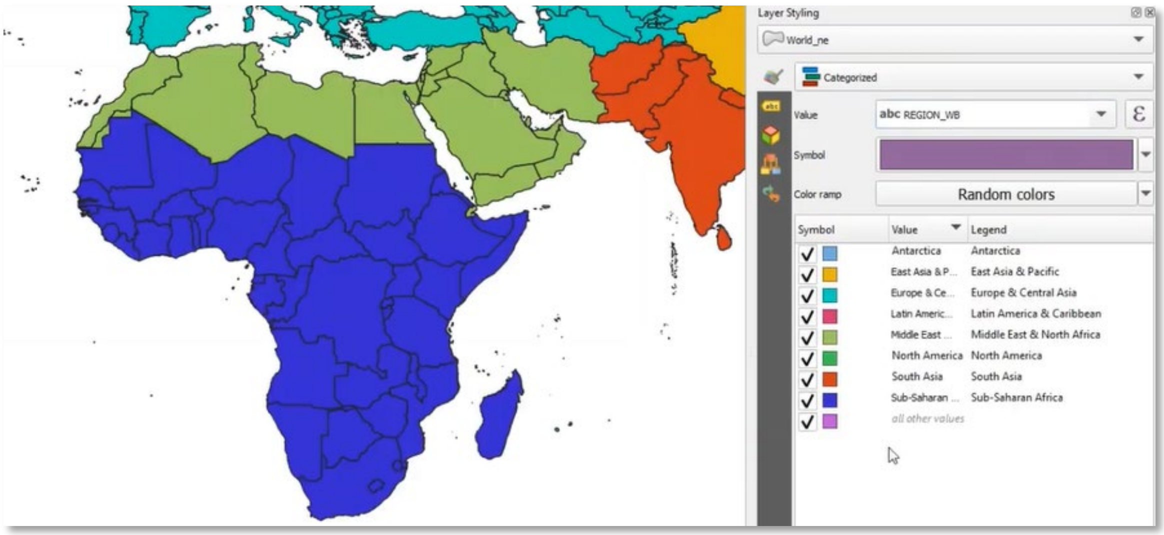
9. Change the color ramp and category colors to colors that you think are appropriate for this indicator.

Now that we have the symbology set for our primary layer (the SSA layer), we can add symbology to the layer being used in our extent indicator (the World_ne layer).

10. In the Layers Panel, turn off the SSA layer and turn back on the World_ne layer.

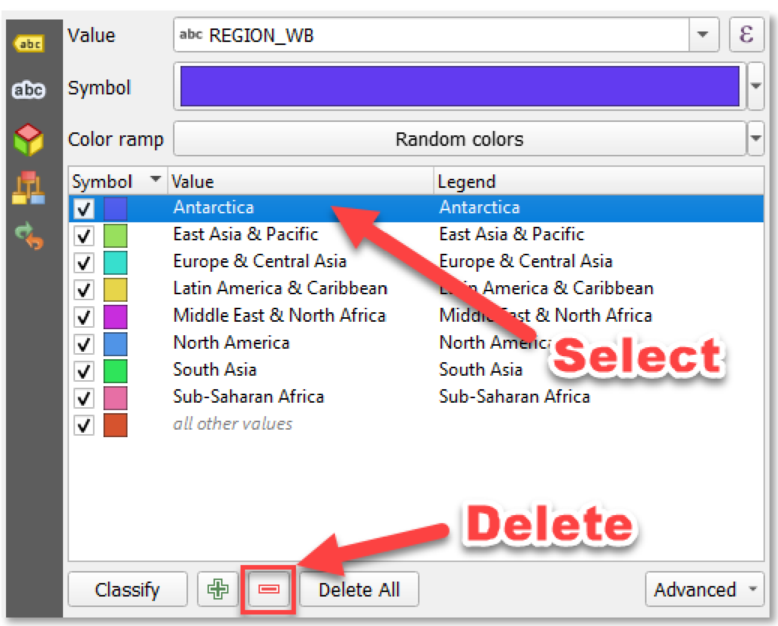
11. In the Layer Styling Panel, change the target layer from ‘SSA’ to ‘World_ne’. Then, select ‘Categorized’ symbology and set the value to ‘REGION_WB’ (the World Bank regions). Click Classify to add the symbology to the layer.

After you do **Step 11**, your map should look like the map below (though the colors will probably be different). Notice that all of the World Bank regions are displayed in different colors.

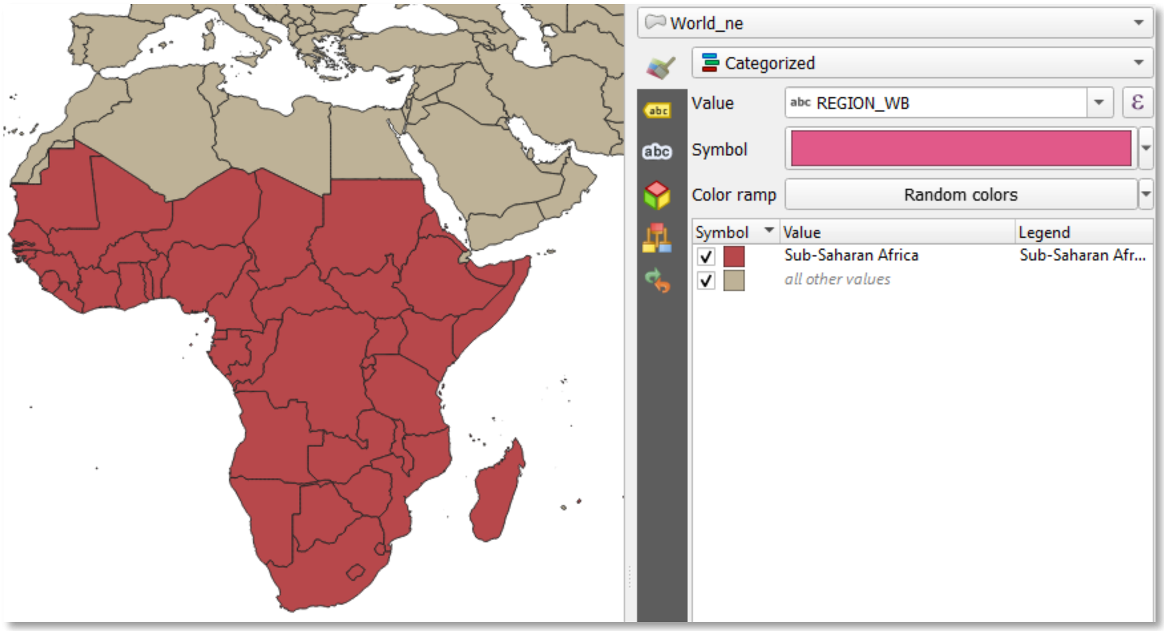


We want our map to highlight sub-Saharan Africa, so we want to remove the classifications for the other World Bank regions.

12. To remove the other World Bank region classifications, click on a category to highlight it in blue and click the red minus (-) button. This is depicted in the screenshot below. Remove all classifications except two: ‘Sub-Saharan Africa’ and ‘all other values’.



13. Now you should only see two categories for the World_ne layer. Change the 'Sub-Saharan Africa' color to a color that stands out. Make 'all other values' to a muted color (like a light grey) so that the countries in sub-Saharan Africa stand out. The result should look like the screenshot below.



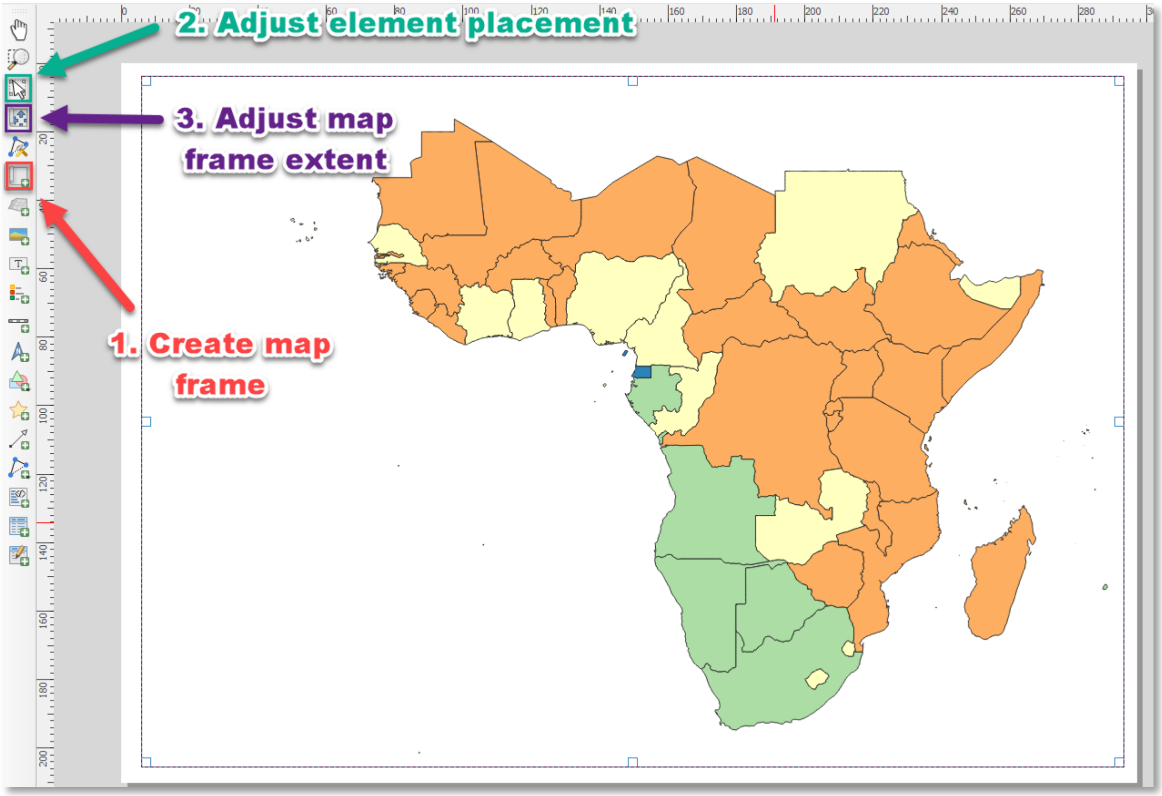
Now that both layers have appropriate symbology, we can now begin creating our print layout.

14. Before creating the print layout, be sure that the SSA layer is turned on and the World_ne layer is turned off in the Layers Panel.

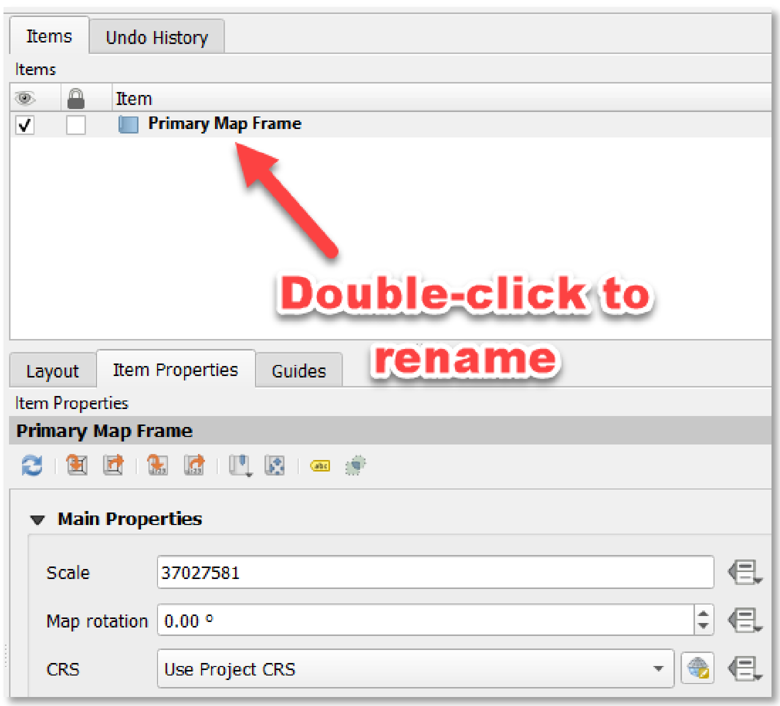
15. Create a new print layout by clicking Project → New Print Layout. You can call this layout "Map with Extent Indicator".

In our print layout, we will have two distinct map frames. The first map frame will be the Primary Map Frame and will show the SSA layer and its symbology. The second map frame will be our Extent Indicator.

16. In the new print layout, begin by adding the first map frame. Resize the map frame so it looks similar to that in the screenshot below. For a reminder about how to add elements to a print layout, see **Subsection 4.2.2** of this document.

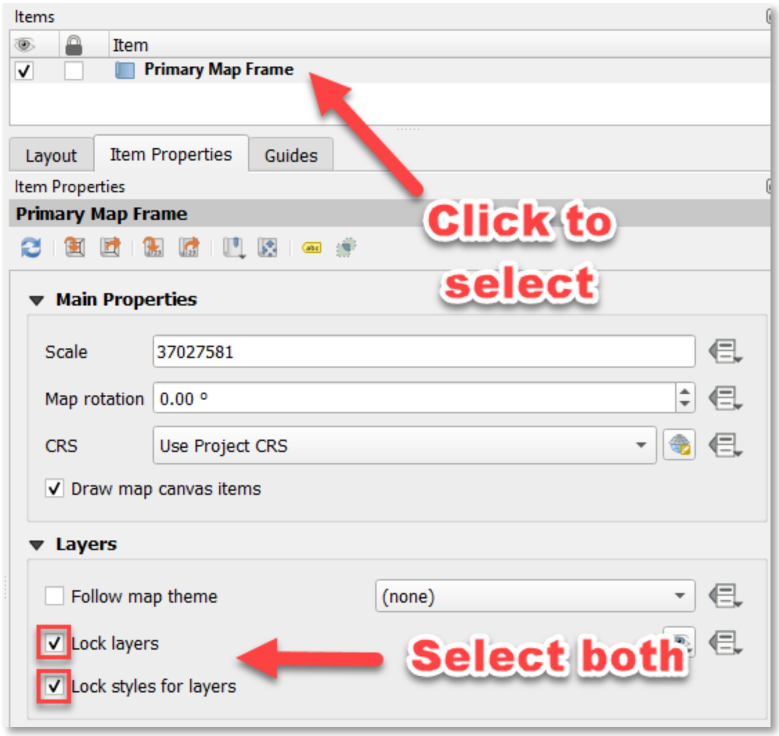


17. Now that we added our first map frame to the print layout, let's rename it "Primary Map Frame" by double-clicking on Map 1 in the Items Box (see screenshot below).



After formatting the Primary Map Frame in your layout, we need to lock the layers. By locking the layers and their symbology in this map frame, we ensure the Primary Map Frame will not change when we set up our second map frame, the Extent Indicator.

18. To lock the layers in our Primary Map Frame, select the Primary Map Frame in the Items Box and check the boxes next to ‘Lock layers’ and ‘Lock styles for layers’ under the Layers settings (see the screenshot below).



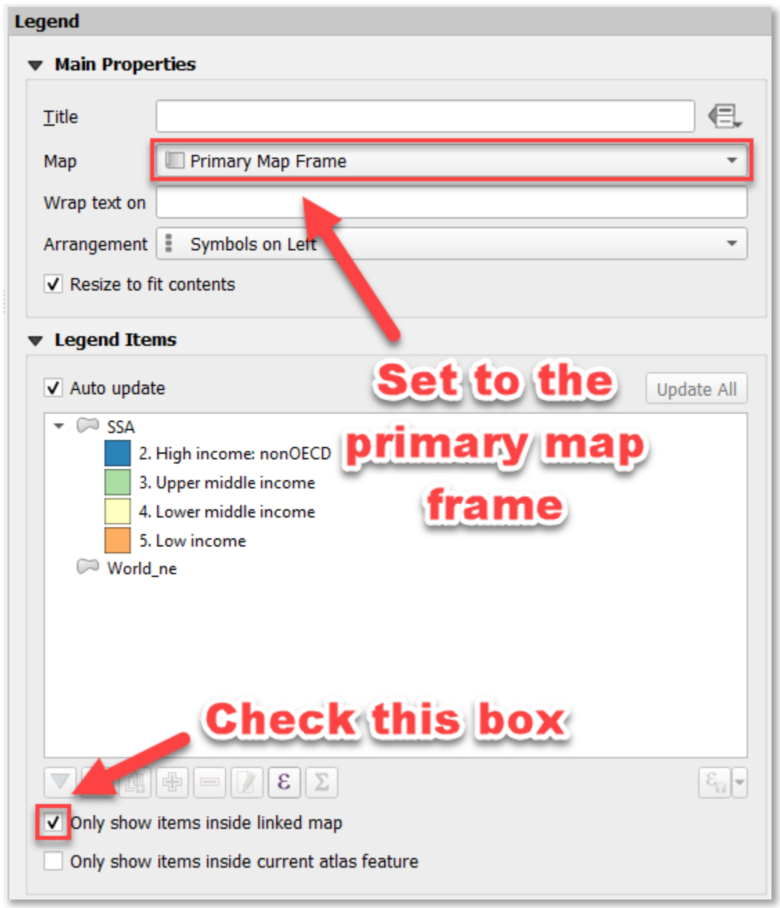
19. Save your print layout and return to your QGIS project window. Back in the Layers Panel of your QGIS project, uncheck the box next to the SSA layer and check the box next to World_ne. Save your QGIS project and return to your Map with Extent Indicator print layout.

20. Now, let’s add the extent indicator to the print layout. Add the second map frame to your project using the Add Map tool. Since you changed which layer was activated in your QGIS project. Place the second map frame in the bottom left corner of the print layout and change the extent to show a bit more of the world.

21. Change the name of this second map frame to “Extent Indicator” by double-clicking on Map 2 in the Items Box.

22. We also want to put a black border around the Extent Indicator to show it as a separate element. To do so, select Extent Indicator in the Items Box and check the box next to Frame in its settings. See the screenshot below.

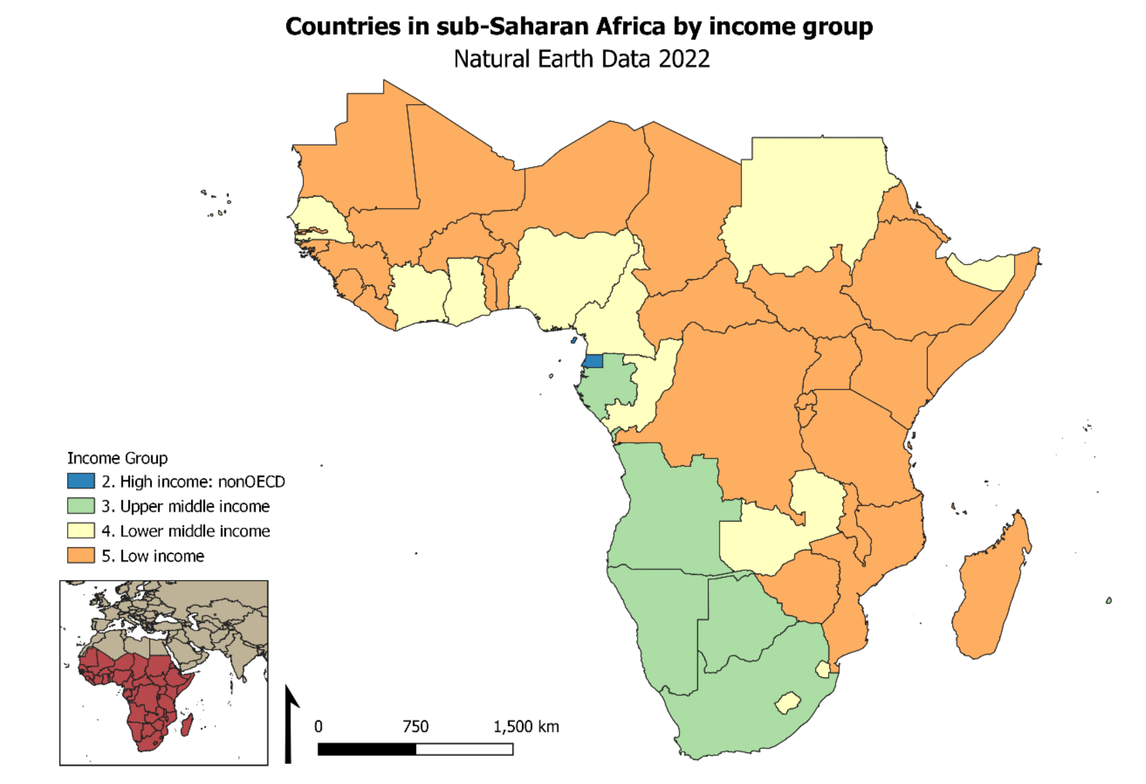
23. Next, we'll add a legend to our print layout.



24. The legend should now only show the symbology for the SSA layer. Let's make the title of this layer a bit more descriptive. Save your changes to the print layout and go back to the QGIS project. There, right-click on the SSA layer in the Layers Panel and select Rename Layer. Change the name to "Income Group". Save the changes and return to your print layout.

25. Now that the legend is set up, add the rest of your map elements: descriptive title, data source, scale bar, and north arrow. The data source for this data is Natural Earth Data 2022.

The final map should look similar to the map below.



9.2.2 Inset Maps in QGIS

For this map, we're going to explore stunting in children under age 5 according to the 2015 Tanzania DHS. Off the coast of Tanzania is an archipelago called Zanzibar. The islands of Zanzibar are small compared to the regions of Tanzania, so we'll be using an inset map to show the islands at a larger scale.

This section corresponds to the second part of Activity 9.1 of the Health Data Mapping Workshop. This activity uses the shapefile in the \DHS_QGIS_2022\Exercise_Files\9.2.2_InsetMaps\ folder.

1. Open a new QGIS project and save it in the exercise folder mentioned above.

2. Expand the Project Home folder (the green folder) in the Browser Panel and drag the Tanzania_dhs.shp into the map frame to add it to the project. You should now see the Tanzania_dhs layer in the Layers Panel.

The shapefile contains an attribute for the percentage of children age under age 5 stunted from the 2015 Tanzania DHS. The variable is called CNNUTSCHA2 (sometimes written as CN_NUTS_C_HA2 in STATcompiler). This is the variable we want displayed in the map.

3. Let's format the symbology for this indicator.

    a. Open the Layers Styling Panel and change the symbology for the Tanzania_dhs layer from Single Symbol to Graduated.

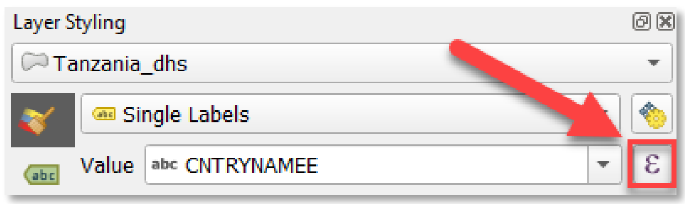
    b. Set the Value to the CNNUTSCHA2 attribute.

c. Change the color ramp to the yellow-orange-red ramp by clicking the down arrow to the right of the Color Ramp parameter and selecting All Color Ramps → YlOrRd.

d. Change the Mode from 'Equal Count (Quantile)' to 'Pretty Breaks'. And click Classify.

4. Now, let's add the labels to this map. Open the Labels tab (the second tab) in the Layers Styling Panel and change the setting from 'No Labels' to 'Single Labels'.

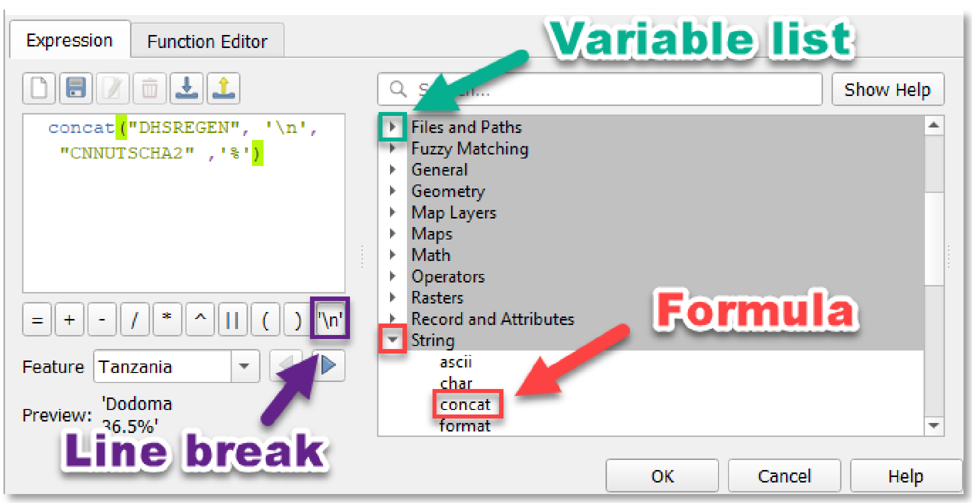
5. The default Value is set to the country name (CNTRYNAMEE). We're going to change this to a two-line label with the region name above the stunting indicator value. To do so, click on the formula button to the right of the Value parameter (shown below).



6. In the Expression Dialog window, use the concatenate formula to combine the region name (**"DHSREGEN"**), a line break (**'\n'**), the indicator value (**"CNNUTSCHA2"**), and a percent symbol (**'%'**). Type the expression below to create the labels and click OK.

```
concat("DHSREGEN", '\n', "CNNUTSCHA2", '%')
```

Instead of typing the expression above, you can also create it with buttons shown in the screenshot displayed below the expression.



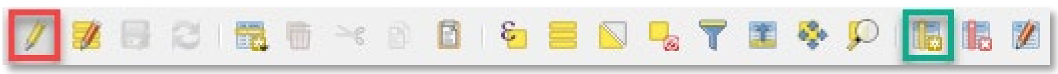
Note that the attributes are enclosed between two double quotes while the text (the line break and percent symbol) are enclosed between two single quotes.

7. Let's change the symbol used for our labels. Under the Text tab (the first tab) of the Label settings, change the Style from 'Regular' to 'Bold'. Lastly, change the Color from the default grey to black. Click Apply at the bottom of the Layer Styling Panel.

8. The labels are aligned to the left, but we want them centered. To change this, open the Formatting tab (the second tab) and change the Alignment from ‘Follow Label Placement’ to ‘Center’.

We don’t want to show all of our labels on the map. The islands of Zanzibar off the coast of Tanzania are small and their labels are clustered too closely together. Our inset map will be zoomed into Zanzibar. We want the labels for the Zanzibar regions to only be displayed in the inset map and not in the primary map frame. We will also want to move our labels, so we need two variables to store the coordinates for our label placement. In the next step, we’ll add all the attributes we need for labeling our map.

9. To add fields to our attribute table, right-click on the Children_stunted layer in the Layers Panel and select Open Attribute Table. Then, click the pencil button to toggle-on editing (identified in red below). You will then use the New Field button to create three new fields (identified in green below).



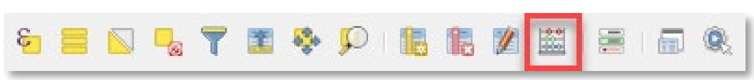
    - Create Label_Cat as a text (string) field. We will use this variable to identify that label as part of Zanzibar or part of the mainland. We can use this field to toggle on and off labels for these two categories.

    - Create Label_Y and Label_X as decimal number (real) fields. These fields will store the latitude and longitude coordinates for custom label placement.

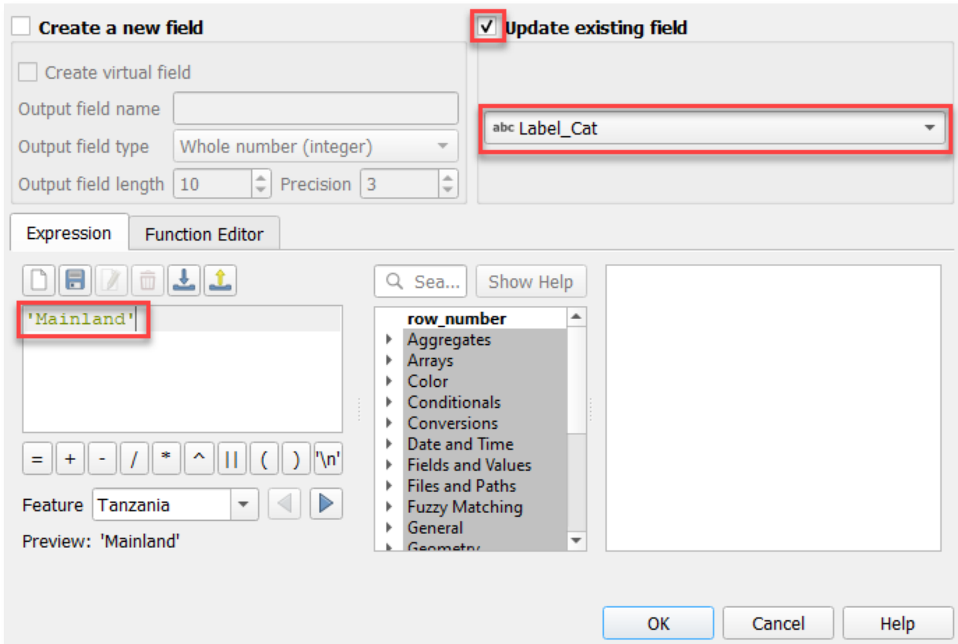
10. After you create the new fields, save the edits to the layer by clicking the Save Edits button (the third button on the menu bar). Now, let’s populate the Label_Cat field. To do this, we’ll need the Selection Toolbar. So, be sure to activate that toolbar by clicking View → Toolbars → Selection Toolbar.

11. Open your attribute table again and toggle editing back on by clicking the pencil icon (the first icon on the attribute table’s toolbar).

12. Now open the Field Calculator by clicking on the icon identified below.



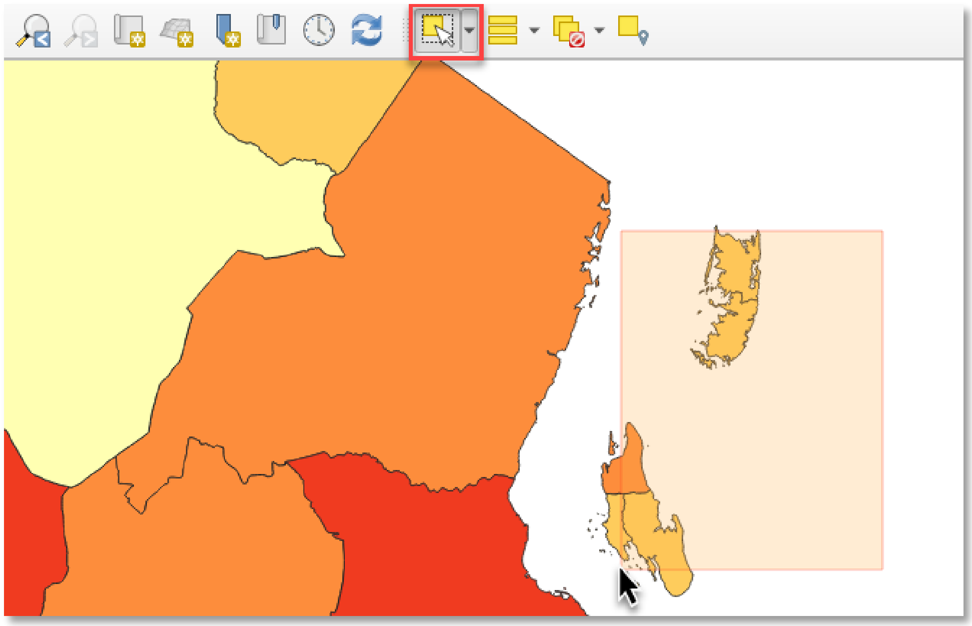
13. In the Field Calculator, check the box next to the Update Existing Field option and select Label_Cat from the dropdown menu. Then type the text `‘Mainland’` (with the single quotes around it) in the Expression Box. Then, click OK. This is depicted in the screenshot below.



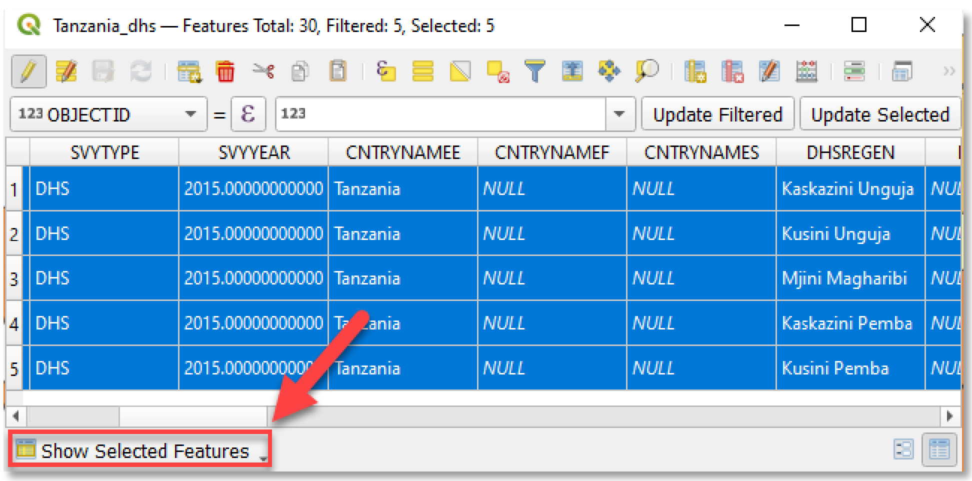
14. Save the edits to the Tanzania_dhs layer by clicking the save button on the attribute table's toolbar.

You have now updated all the values in the Label_Cat variable to contain the text, "Mainland". Now, let's change that variable for the 5 regions in Zanzibar. We want the Label_Cat variable to contain the text, "Zanzibar", for these regions.

15. Let's begin by selecting the islands of Zanzibar. We could do this by selecting the administrative areas by their names or we could manually select these islands. We'll do that latter. Close the attribute table and zoom into Zanzibar. Use the Select Features by Area or Single Click tool (shown below). Either click, hold, and drag your cursor to select the islands with an area or individually select the five polygons by clicking on each while holding down the Ctrl key.



16. Once you have the Zanzibar islands selected, open the attribute table again. In the bottom lefthand corner, change the view setting from ‘Show All Features’ to ‘Show Selected Features’. This will now only show the regions of Zanzibar in the attribute table (shown below).



17. Now, we’ll use the Field Calculator again, but this time we’ll change the value of Label_Cat for the selected features. Open the Field Calculator as illustrated above in **Step 12**.

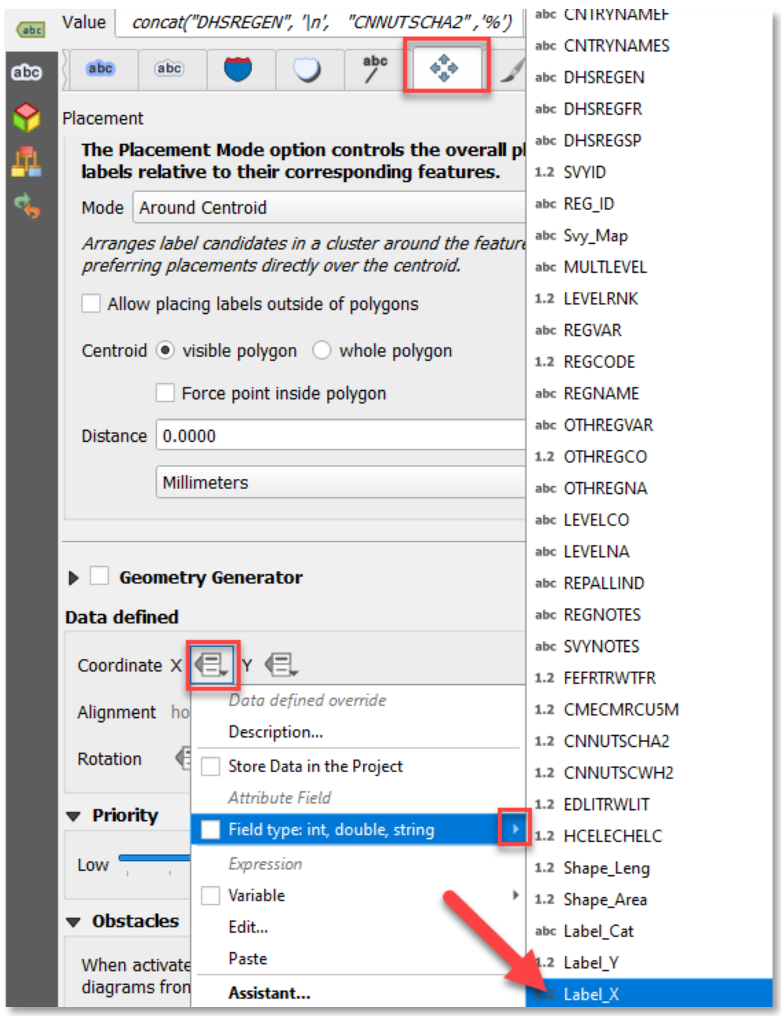
18. In the Field Calculator, check the box next to the Update Existing Field option and select Label_Cat from the dropdown menu. Then type the text `'Zanzibar'` (with the single quotes around it) in the Expression Box. Then, click OK. The tool configuration is depicted in the screenshot below **Step 13**.

19. Save your edits to the Tanzania_dhs layer and close the attribute table.

20. To deselect the regions of Zanzibar click the Deselect button (identified below) on the Selection Toolbar.

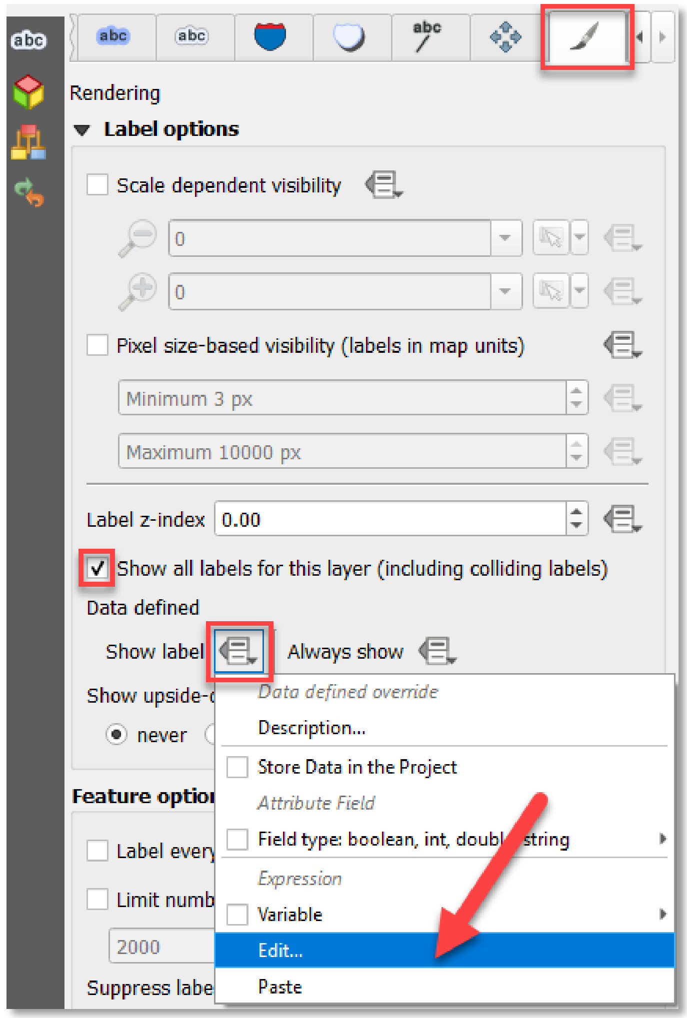
Now that we have our label attributes configures, we're going to connect these attributes to the labeling properties in the Layer Styling Panel.

21. The first label attributes we'll link to our label properties will be the two attributes that will contain our label placement coordinates: Label_Y and Label_X. In the Label Tab of the Layers Styling Panel, navigate to the Placement Options. Under the Geometry Generator section, set the Coordinate X to the Label_X field by clicking on the Coordinate X button → Field Type → Label_X. This is shown in the screenshot below. Then, assign Coordinate Y to the Label_Y field.



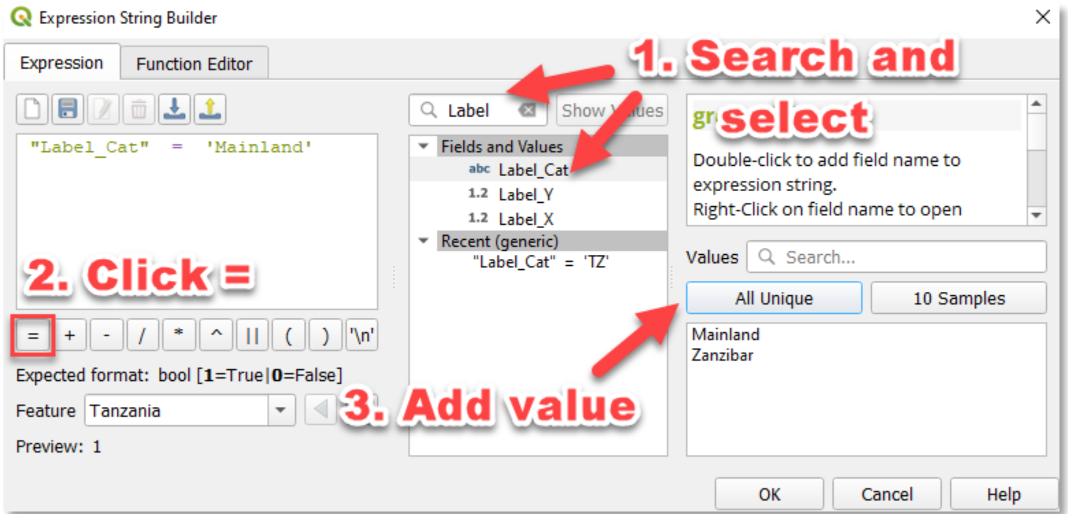
Now that we've set the label coordinate attributes, we can move labels around our map. Before we do so, let's filter out the Zanzibar labels. We'll adjust the mainland labels first and then set up the Zanzibar labels.

22. Open the Rendering Options (the last tab in the Label Tab of the Layer styling Panel). In the first section, check the box next to 'Show all labels for this layer'. Then, click on the Show Label option under Data Defined and select Edit. This is depicted in the screenshot below.



23. In the Expression Builder window, either type the following expression or click to add elements (demonstrated in the screenshot below the expression). Once the formula is written, click OK.

```
"Label_Cat" = 'Mainland'
```



After clicking OK, the labels for Zanzibar should disappear. We'll now adjust the placement for our mainland labels.

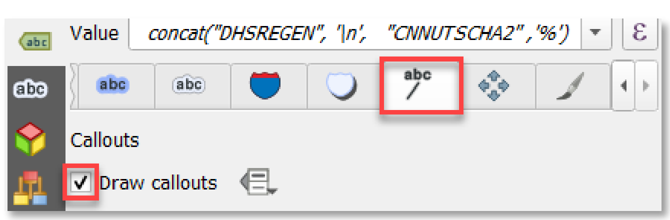
24. Before we can adjust our label placement, we will need to make sure the Label Toolbar is activated. Click on View → Toolbars → Label Toolbar to turn it on. We also need to start editing the Tanzania_dhs layer. To do this, right-click on the layer in the Layers Panel and select Toggle Editing.

25. Now, use the Move a Label Tool on the Label Toolbar (shown below) to move the labels around. To move a label, activate the Move a Label Tool, click on the label, and then click where you want it on the map. Move all the labels so that they don't overlap with regional borders. Since the Dar es Salaam and Kilimanjaro labels will not fit in their respective administrative areas, move them outside of Tanzania's borders. We'll use callouts to connect them with their admin areas.

Now, let's save the edits to our Tanzania_dhs layer. It's important to save the edits to our layer, because we have generated coordinates in the Label_Y and Label_X fields by moving labels around the map.

26. To save the edits we made to Tanzania_dhs, right-click on the layer in the Layers Panel and select Save Layer Edits. We can also toggle off editing, since we will not be making additional edits until later in this section. So, click Toggle Editing to deactivate layer editing.

27. To add callouts for the Dar es Salaam and Kilimanjaro labels, click on the Callouts Option of the Label Tab (in the Layer Styling Panel) and check the box next to 'Draw callouts' (shown below).

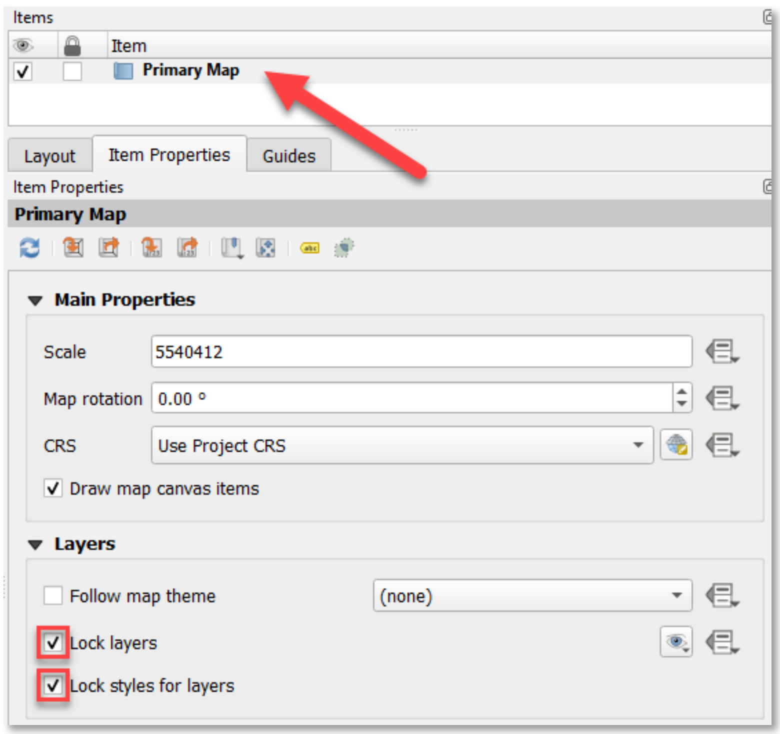


Now that we have configured the labels for our primary map, let's set up our map layout before configuring our inset map.

28. Click on Project → New Print Layout. Name the new layout "Map with Inset".

29. Create a map frame in your layout and rename it from "Map 1" to "Primary Map" in the Items Box. For a reminder of how to set up and rename your primary map frame, see Steps 16 and 17 in [Section 9.2.1](#).

30. With the primary map frame added and renamed, we need to lock its layers and symbology so that we configure the inset map. To do so, select the Primary Map from the Items Box and open the Item Properties Tab. Under the Layers heading, check the boxes next to 'Lock layers' and 'Lock styles for layers'. This is shown in the screenshot below.



31. Now, save your print layout and return to the QGIS project.

Now, we're going to remove the mainland labels and add back in the labels for Zanzibar. We'll then move the Zanzibar labels so that they are easy to read in the inset map we create.

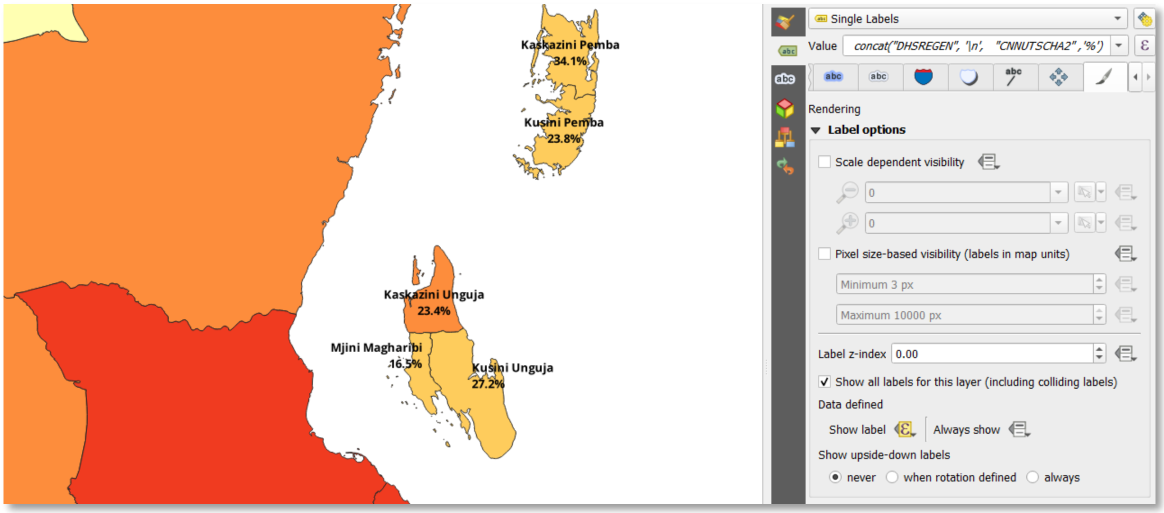
32. Zoom into the Zanzibar islands.

33. In the Label Styling Panel, open up the Rendering Options in the Labels Tab. Click on the Show Label option under Data Defined and select Edit. This is depicted in the screenshot below **Step 22** of this activity.

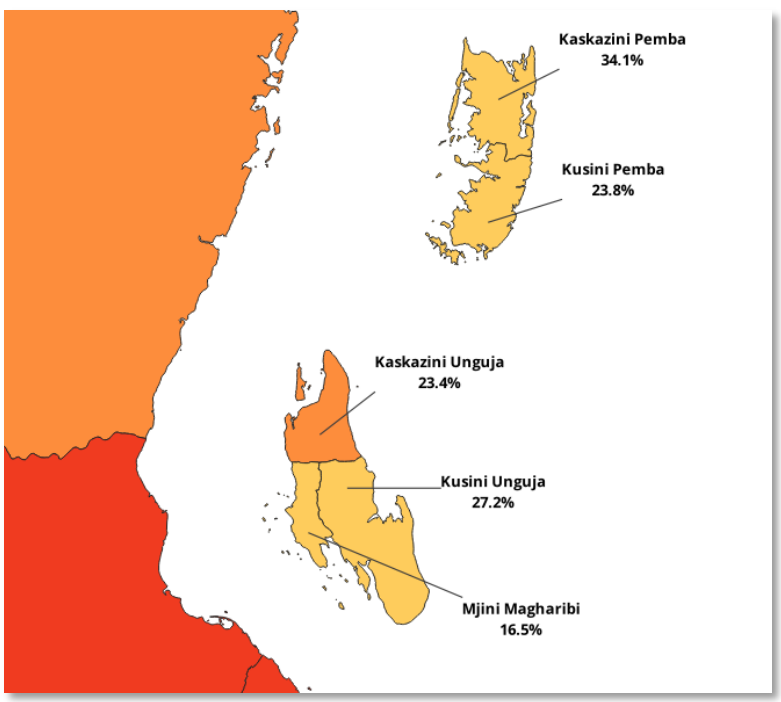
34. In the Expression Builder window, change the text, `'Mainland'`, to `'Zanzibar'`. The expression box should contain the text below. Click OK once it does.

```
"Label_Cat" = 'Zanzibar'
```

You should now see the Zanzibar labels appear and the mainland labels disappear as depicted below. The next step will be to adjust the placement of these labels.



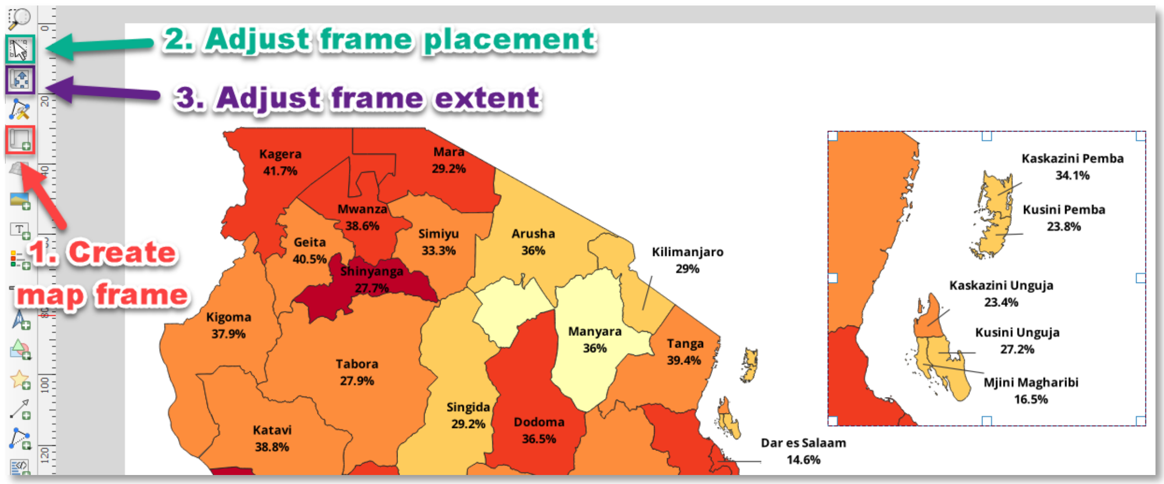
35. Before we can move the labels, we need to toggle editing back on. Right-click on Tanzania_dhs in the Layers Panel and select Toggle Editing. Then use the Move a Label Tool to move the labels for the Zanzibar regions. For a reminder of how to do so, see **Step 25** of this activity. Move the labels so that they do not overlap with any administrative borders. The result should look similar to the screenshot below.



36. After you have adjusted the labels to your liking, save the edits. Right-click on the Tanzania_dhs layer in the Layers Panel and select Save Layer Edits. Then, click Toggle Editing to turn editing off.

37. Save the QGIS project and return to your print layout.

38. Now, add a second map frame to your print layout. Rename this map from "Map 2" to "Inset Map" in the Items Box. Use the Select/Move Item and Move Item Content tools (labeled 2 and 3 in the screenshot below) to adjust the placement and extent of the inset map. It should look similar to that in the screenshot below.

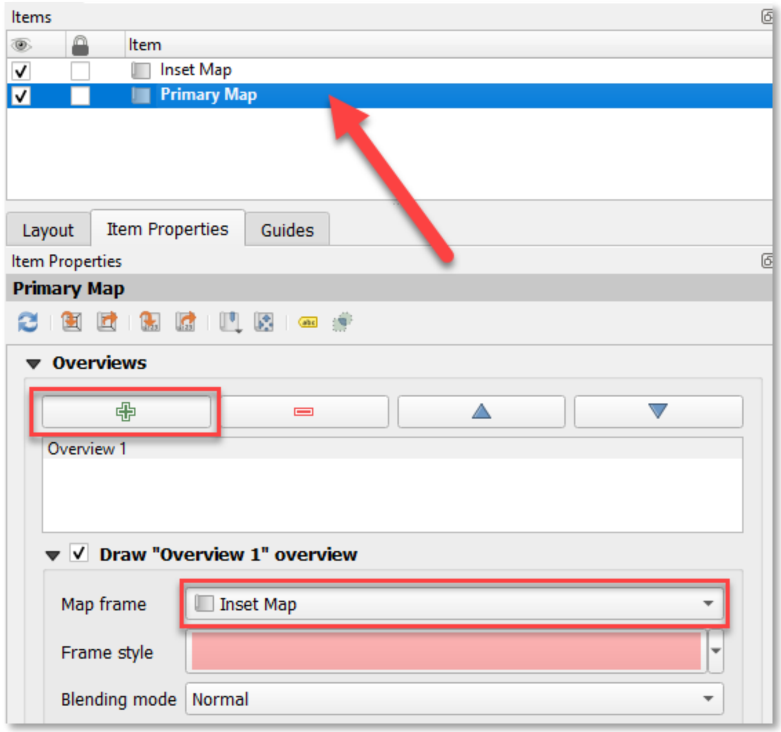


You may need to go back and forth between your map layout and QGIS project to make slight adjustments to the label styles and placement.

39. Once the inset map is placed, add a frame around it by checking the box next to 'Frame' in the Inset Map's Item Properties.

Next, we will add an overview box to our primary map. The overview box is a box that indicates the extent of our inset map on top of the primary map.

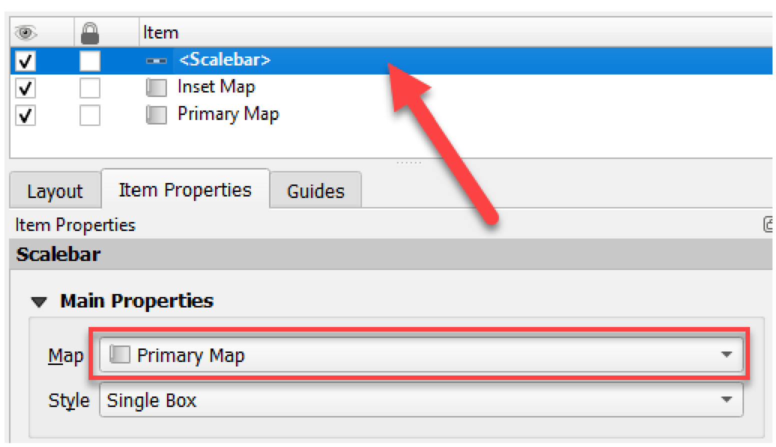
40. To add an overview box to our primary map, select the Primary Map from the Items Box and open the Item Properties Tab. Expand the Overview section and click the green + button. Under the Draw "Overview 1" Overview section, set the Map Frame to our Inset Map. This is depicted in the screenshot below.



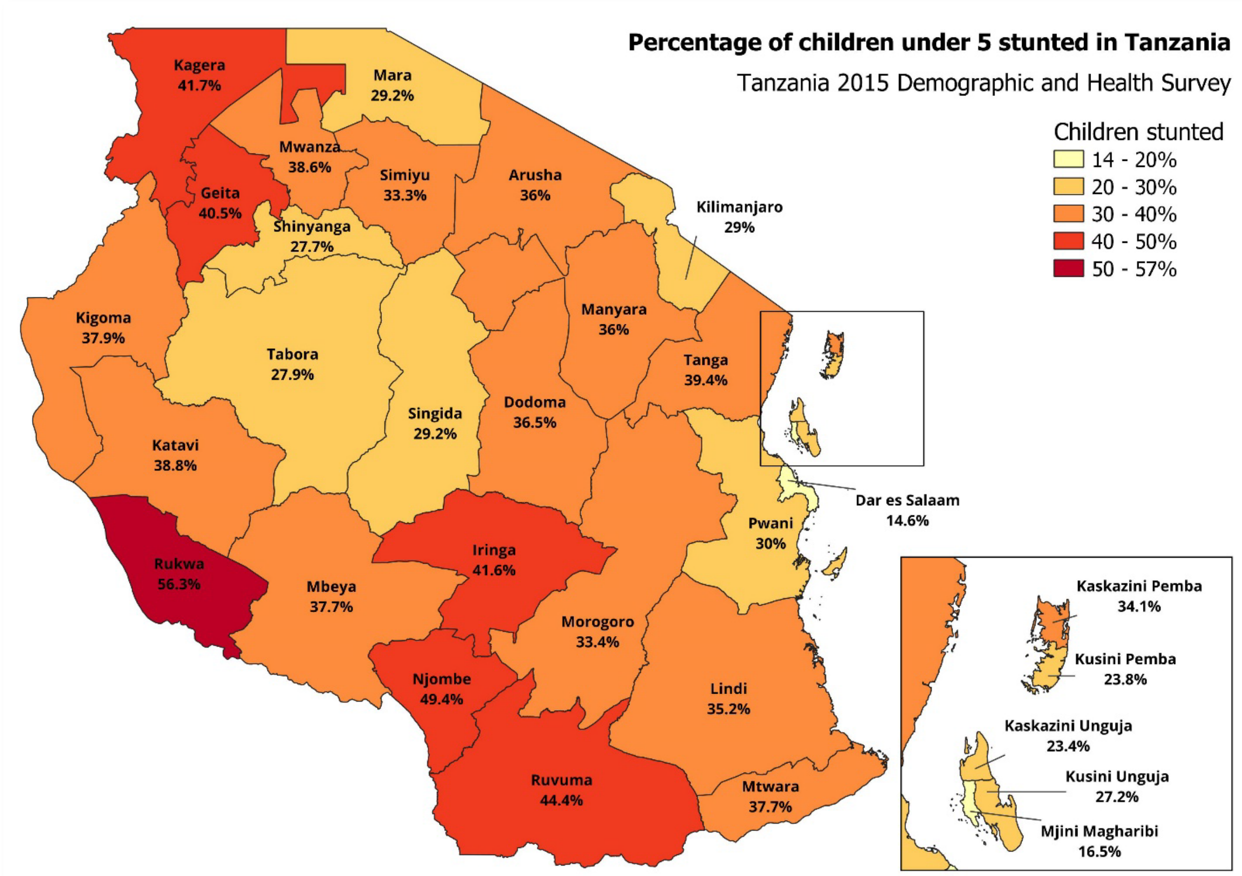
A slightly transparent red box will now appear over the Zanzibar islands in your primary map frame. It's best practice to have the overview match the border of our inset map. So, we're going to change the style of our overview to match it.

41. To change the overview's style, click the dropdown arrow next to the Frame Style option and select Configure Symbol. Click on the Simple Fill and change the configuration so that the Fill Style is set to 'No Brush' and the Stroke Style is set to 'Solid Line'. This is depicted below. Click on the back button until you return to the Item Properties Tab.

42. Now that you have configured your primary and inset map frames, add the remaining map elements: a descriptive title, the data source, a formatted legend, and a scale bar. For the scale bar, be sure that the Map is set to the Primary Map (as shown in the screenshot below).



After you add the remaining map elements to your layout, the final map should look similar to the map below.



10. ADDITIONAL INFORMATION

10.1 GEOSPATIAL DATA RESOURCES

Several geospatial data sources are mentioned in throughout this guide. The four data sources below are good places to start when looking for geospatial data, including shapefiles.

- **GADM**: This free set of national and subnational boundaries available at a high resolution and with rich attributes. You can access this data at the following link: gadm.org/data.html.

- **Humanitarian Data Exchange (HDX)**: The HDX has more data than can be listed here. The dataset come from a wide variety of sources, so you can expect to find all sorts of geospatial datasets. However, many of the datasets on this site are not geospatial data, so be sure to use the site's filters to view only the types of data you want. You can access this data at the following link: data.humdata.org.

- **Natural Earth Data**: This is great data for creating a basemap or reference map. This source has everything from rivers and lakes to populated places and ports. You can access this data at the following link: naturalearthdata.com.

- **Spatial Data Repository (SDR)**: The DHS Program's SDR has rich geospatial data on demographic and health indicators. You can access this data at the following link: spatialdata.dhsprogram.com.

10.2 FREQUENTLY ASKED QUESTIONS

What version of QGIS should I use?

We recommend using the latest long-term release version of QGIS. The long-term release version is available on the QGIS download webpage, qgis.org/en/site/forusers/download.html, under the 'Standalone installers' heading and 'Long term release (most stable)' subheading.

If I installed a newer version of QGIS, can I still use the exercise files distributed with this guide?

Yes, QGIS software is backwards compatible. You can open older *.qgz and *.qgs projects in newer editions of QGIS without any problem. However, it should be noted that the interface and menus in newer versions of the software may be different.

If I have trouble installing or using QGIS, where can I get assistance?

The support page, qgis.org/en/site/forusers/support.html, is an excellent place to start. The page provides links to various resources that will help you answer your QGIS-related questions.

Aside from QGIS, what other GIS software packages can I use?

Esri ArcGIS is the most common software used in the field of GIS, but it can be cost prohibitive. Aside from QGIS, comparable free alternatives include: gvSIG Desktop (gvsig.com/en/products/gvsig-desktop), WhiteboxTools Open Core (whiteboxgeo.com), and SAGA GIS (saga-gis.sourceforge.io). Google My Maps (google.com/maps/about/mymaps) is a helpful tool for creating simple web maps, but it does not contain the geoprocessing tools available in the other software mentioned here.

Where can I expand upon the GIS skills I learned in this course?

The interactive QGIS Training Manual (docs.qgis.org/3.22/en/docs/training_manual/index.html) is a great place to start. The *QGIS for Digital Cartography in Censuses and Surveys* created by UNFPA is another great resource. This document is available in several languages here: unfpa.org/resources/qgis-digital-cartography-censuses-and-surveys.

How can I collect GPS data using a smartphone or tablet?

The applications listed in **Subsection 6.1.5** provides a list of mobile applications that can be used to leverage the GPS receivers built into smartphones and tablets. Specifically, QField (qfield.org), ODK (getodk.org), Input (inputapp.io), and Geopaparazzi (osgeo.org/projects/geopaparazzi) are mentioned in that section. At The DHS Program, we use a software called CSPro (census.gov/data/software/cspro.html), developed by the United States Census Bureau. We use this software to build the electronic questionnaires used to collect DHS data and collect GPS coordinates.

While the GPS receivers in smartphones and tablets work without an Internet connection (due to the technology described in **Subsection 6.1.3**), many of the GPS data collection applications include basemaps that show the user where they are in the world, which do not work without an Internet connection. However, in some cases, these basemaps can be downloaded for offline use. This is an important factor to consider when selecting an application for your project.

How can I access GPS data from The DHS Program?

To access GPS data for DHS Program surveys, you need to create an account on our website (dhsprogram.com) and submit a formal request. This data is not available publicly on the Spatial Data Repository given its sensitivity. When submitting a request, you will need to provide a rationale for requiring the GPS data. These requests will need to be approved by The DHS Program, so be sure to factor that into your project schedule. For more information on downloading GPS data from our website, please visit our webpage on GPS datasets: dhsprogram.com/Methodology/GPS-Data.cfm.

Where can I download subnational indicator data from The DHS Program?

Indicator data can be downloaded from the Spatial Data Repository (spatialdata.dhsprogram.com) and from STATcompiler (statcompiler.com). To download subnational indicator data from SDR, click on the Indicator Data tab (spatialdata.dhsprogram.com/data) and select whether you want to download data from a Single Country or Many Countries. If you choose the Single Country option, you can select which surveys for which you want data. The Many Countries option, on the other hand, will only give you data from the most recent survey in each selected country. After you select the surveys/countries for which you want data, you will be prompted to select the indicators. Once the indicators have been selected, you will be presented with several download options. Be sure to check the box next to 'Subnational' under the Geography Level. You can then download the subnational data as a shapefile to be used in GIS software. To watch a video of this process, go to the following link: dhsdata.com/2Cq6vqA.

To download subnational indicator data from STATcompiler, you must be viewing subnational data in the application. If you are viewing a table of subnational data on STATcompiler, you can download that table by clicking the Export menu in the top right corner and selecting Table Format. Shapefiles can also be downloaded from this application. If you are viewing subnational data in STATcompiler and click the Export button, you will see an option to Export Map Data. This will export the highest level of subnational data as a shapefile for use in GIS software. You can access the first part of our tutorial on STATcompiler at the following link: dhsdata.com/2xVK3ow.

10.3 GLOSSARY

Altitude: the height of a point or location above sea level; to collect altitude data with a GPS receiver, you need at least four satellites

Attribute join: merging two datasets based on a common variable called a join key

Attribute table: a tool used in QGIS to explore a layer's attributes

Attributes (of a GIS layer): the variables and values for each feature in a GIS layer

Balance (map design): the positioning of elements on a map layout so that they are spread evenly across the page and white space is limited

Binary data: data consisting of two possible values, often 1 and 0 or TRUE and FALSE

Callouts/leader lines: lines indicating which labels refer to which features in a map

Cartography: the science and art of drawing maps

Categorical data: qualitative or quantitative data divided into groups (e.g., countries broken into income groups or communities categorized as urban or rural)

Cells (Microsoft Excel): the boxes in a spreadsheet that contain values

Choropleth map: a thematic map that uses colors or shading within areas to indicate the value of an indicator in those areas (e.g., a choropleth map of stunting in a country may use increasingly darker shades of red to indicate a higher prevalence of stunting in some administrative areas compare to others)

Clean data (for QGIS): a cleaned dataset that (1) has a single header row with variable names that do not use special characters, (2) has no merged cells, (3) is formatted so that every row is a single observation (e.g., each row is a household cluster or health facility), and (4) uses the same format for any geographic information in a column

Continuous data: quantitative data with a theoretically infinite number of possible values between a given high and low value (e.g., elevation data or precipitation data where each data point is equal to any value between a given maximum and minimum value)

Coordinate systems: reference frameworks with a set of rules used to determine where places are located on or near Earth's surface

Crosswalk tables: a table comprised of values pulled from multiple datasets; use the =VLOOKUP() function to create a crosswalk table in Excel

Decimal degrees: coordinates in DDD.DDDDD format (e.g., 34.73459), this is the format preferred by many GIS software including QGIS

Degrees, minutes, seconds: coordinates written in DD° MM' SS.S" format (e.g., 34° 44' 04.5")

Equator: an imaginary line on Earth's surface that is equal distance from the North and South poles and divides the planet into the Northern and Southern hemispheres

Extent indicator: in a map layout with two maps, the extent indicator is a smaller map that shows where the larger map is located in reference to surrounding geographic features

Feature: individual data points in a GIS layer are called features (e.g., in a file of GPS locations, each location is considered an individual feature; in GIS layer that contains administrative boundary, each individual boundary polygon is considered the feature)

Field calculator: a tool in QGIS used to populate the values of a field in the attribute table

Figure-ground organization (map design): the use of contrasting colors to differentiate map elements in the foreground from those in the background (i.e., the use of contrasting colors to focus the attention of the map user)

Font style: properties of the font that include italics, bold, and underlines

Geographic information systems (GIS): software used to organize, display, and analyze different geographic data

Geoid: the spherical shape of Earth

Geographic/geospatial data: any data that contains information about a location on Earth's surface (examples of location information include X/Y coordinates or geographic names)

Geography (of a GIS layer): the graphic component of a GIS layer that depicts the features' physical location on Earth's surface

GPS: the Global Positioning System, developed and maintained by the United States government, uses a satellite constellation for navigation on Earth's surface

GPS receivers: a processor that uses radio signals from satellites to calculate the user's location (e.g., these receivers are found in handheld GPS devices and cellphones)

Inset map: in a map layout with two maps, the inset map is a smaller map that shows a portion of the primary map at a larger scale (i.e., more zoomed in)

Interference: when the radio signal from the satellite gets distorted by an object (e.g., a building, hard ground, mountains, heavy tree cover, etc.) before it reaches the GPS receiver

Join data: when joining two datasets, this is the dataset that will be added to the target data

Join key: a common variable used to merge two datasets, often a unique identifier

Label key: a map element that matches numbers in the map that represent labels with the label content itself

Latitude: measurement north or south of Earth's equator

Layer: dataset displayed in a GIS are called layers, because the user visually stacks these datasets on top of each other to create maps and conduct analyses

Legend: a visual guide of the symbology used in a map (e.g., if a capital city is represented with a star and all other cities are represented with a circle, the legend will show a star and circle labeled according to what they represent)

Longitude: measurement east or west of the prime meridian

Map elements: different components that together comprise the map layout (e.g., the map frame, legend, title, scalebar, etc.)

Map layout: what is depicted on a virtual or printed map page

Overview box: a rectangle that indicates the extent of an inset map within the primary map

Panels (QGIS): the QGIS user interface includes small windows called 'panels' that allow the user to change content in the project and conduct analyses; by default, panels appear on the left- or right-hand side of the QGIS map area; to open panels in QGIS, go to View → Panels (e.g., the Layers Panel is where you see your map layers listed and the Layers Styling Panel is where you change the symbology of those layers)

Pivot tables: a table that aggregates values from a larger dataset (e.g., if you have a dataset of every health facility in a country with information about the administrative areas, you can create a pivot table of the number of health facilities in each administrative area)

Prime meridian: the imaginary line chosen to represent 0° longitude that passes through Gao in Mali and Lake Volta in Ghana

Print layout (QGIS): the user interface in QGIS used to create map layouts

Projections: the geometric transformations used to flatten the spherical Earth so that it can be depicted in a two-dimensional map or GIS

QGIS (formerly Quantum GIS): the open source/free GIS software used in this guide

Quantitative data: numerical data

Qualitative data: non-numerical data

Query: the process of filtering features out of a dataset by certain criteria

Raster data: geographic data in the form of a collection of pixels where each pixel represents a square piece of the Earth's surface or atmosphere

Reference maps: map designed to show where places or other geographic features are located in relation to each other (e.g., a city map used by tourists or a topographic map used by geologists)

Scale: the relationship between the dimensions of our map and the actual world

Scalebar: a graphical representation of a map's scale

Selection: a subset that is highlighted in a GIS dataset based on its attributes or location

Shapefile: vector datasets that can be used in GIS software; they are comprised of up to 11 separate files, but only require three: *.shp, *.shx, and *.dbf

Spatial join: merging two dataset based upon the geographic location of their features

Subset: a smaller dataset created by querying out, selecting, or removing features based on given criteria

Symbology: the use of symbols, shapes, and colors to represent physical places and geographic phenomena in a map

Target data: when joining two datasets, this is the dataset to which the join data will be added

Thematic maps: maps designed to convey information about a single topic or theme (e.g., percent of households with electricity or population density)

Title/proper case: font case that capitalizes the first letter in a word (with the exception of prepositions and articles)

Toolbars (QGIS): QGIS tools are organized into various toolbars, each with icons representing the different tools; by default, the toolbars are docked above the QGIS map area; to turn on toolbars, go to View → Toolbars (e.g., the Map Navigation Toolbar contains the tools used to zoom and pan around the map)

Triangulation: the process of using three intersecting circles to determine location

Typeface: the style of the characters used in text

Unique identifier: a variable in which each observation is unique (e.g., a cluster or facility ID)

Vector data: geographic data in the form of points, lines, or polygons that represent places or phenomena on Earth's surface

Visual hierarchy (map design): the prominent placement of items in the map layout to emphasize their importance

WGS 1984: the coordinate system used by the Global Positioning System (GPS) and commonly used in cartography and GIS application

Workbook (Microsoft Excel): a Microsoft Excel file, often ending in the *.xlsx or *.xls file extension

Worksheet (Microsoft Excel): an individual page or sheet in a Microsoft Excel workbook