Basic Queries Exercise - Haiti

Written by Barbara Parmenter, updated by Marcia Moreno-Baez for ArcGIS Pro – January 22, 2022

Background

An earthquake of 7.2 magnitude struck Haiti on August 14th 2021, overwhelming hospitals, and heavily damaging buildings and trapping people under rubble in at least two cities in the western part of the country's southern peninsula: Les Cayes and Jeremie. Phone lines were down in *Petit Trou de Nippes*, about 80 miles west of Port-au-Prince, the epicenter of the quake. Seismologists said it had a depth of seven miles and was felt as far away as 200 miles in Jamaica.

In this exercise, we will use spatial information to explore the geographic area of Haiti and ask some questions that might help us understand how many people were affected by the earthquake and visually explore the area.

Some of the question we will ask in this tutorial are:

- Where do the earthquake happened?
- How many hospitals are and where are they located?
- How many hospitals are in the most affected area?
- How many people live by a main road?
- How many healthcare providers are in each of Haiti's Departments?
- How many square kilometers of each type of *Livelihood Zones* are there?
- How many people were affected the most by the earthquake?

Introduction and Objectives

In this tutorial, we will ask some questions and use basic analysis tools to answer them. We will use data available for Haiti to bring context to the exercise.
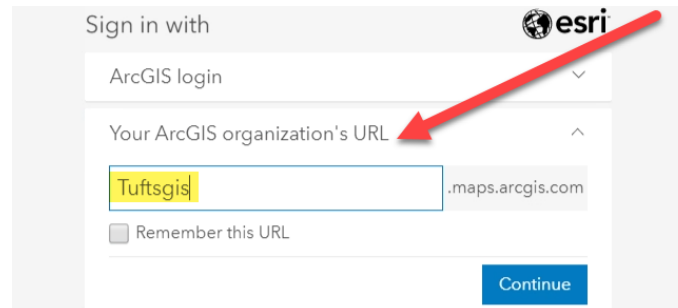
The learning objectives for this exercise are to:

- Describe data sources and data quality by reviewing file formats and data structures
- Explore querying geospatial data with selection tools (Select by Attribute and Select by Location)
- Calculate basic statistics for selected features
- Manage attribute tables (e.g., add a field, calculate geometry, and summarize by an attribute category)
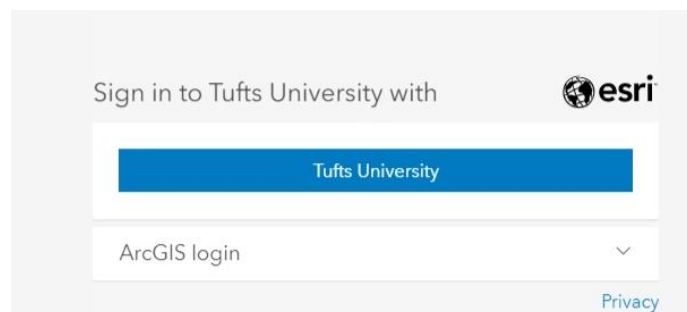- Perform spatial analysis such an overlay analysis (i.e., spatial join) and zonal statistics

Starting an ArcGIS Pro Project File

Important: When starting a new ArcGIS Project, you need to determine where you will be saving all GIS data and project files before you start, since Pro makes you choose a home folder when opening the software. You have several options including **Tufts Box, your H Drive** (for Tufts students), **your Desktop or Documents folders** (if it's a personal computer) or an **external drive**. If you plan to use Tufts Box, you MUST be logged into **Box Drive** before starting the tutorial.

1. Open ArcGIS Pro and Sign In link in the top right to sign in with Tufts Organization's license.

2. Click **Sign in with your ArcGIS Organization's URL.** Type in **Tuftsgis.** If this is your personal computer, click **"Remember this URL".**


3. Click Sign into Tufts University with **Tufts University.** Enter your **Tufts Username** and **Password** and go through Duo Authentication.

4. Once you are signed-in, click on the option to start with a new blank **Map.**

5. A new dialogue box should open with the option to name your project file and where to save it. Name your project "*ArcGISPro_BasicQueries_Haiti*". Under *Location*,   click on the **folder icon** and navigate to where you want to save your project.

   Note: this tutorial will use Tufts Box; alternatively, you might save it elsewhere such as the H drive or a USB.

   Make sure **Create a new folder for this project** is checked so that it creates a sub-folder for this activity.

   Press OK.
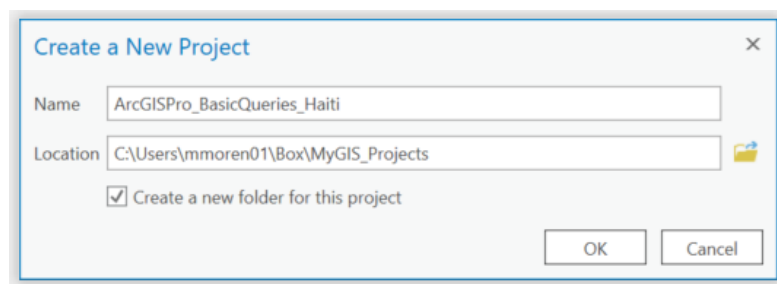
6. A new blank map will open. On the left, you will see your **Contents** pane. That shows all layers currently in the map (in this case, nothing other than a base map). On the right, you will see the **Catalog** pane. This is where you can access everything associated with this project, including GIS data, saved maps, toolboxes, etc. Double click on **Folders** and you will see your connection to the **ArcGISPro_BasicQueries_Haiti** Folder in Tufts Box.

Download Exercise Data

1. If you have not already done so, you will need to download and unzip the tutorial data into a folder on your Box drive. Do not include any spaces, hyphens or characters other than letters in the folder pathway names.
2. The data is saved here: https://tufts.box.com/v/ArcGISPro-Basic-HaitiData
3. Download the file named ArcGISPro_BasicQueries_HaitiData.zip. Depending on the browser you are using, it might ask if you want to save, put it in the downloads bar at the bottom or save it directly into your Downloads folder.
4. In the Windows File Manager (AKA Windows Explorer), navigate to your Downloads folder. Copy this zip folder over to your Box → GIS_Tutorials→ ArcGISPro_BasicQueries_Haiti folder and paste it there. There will be several other files in the windows folder including the ***.aprx** (ArcGIS pro project file) and several other folders that were created when you started this ArcGIS Pro project.
5. Right click on the ArcGISPro_BasicQueries_HaitiData.zip file → 7 Zip → Extract All. Alternatively, you might not have 7-zip installed so just look for Extract here or Extract All or something similar. This unzips all the components that are contained within this GIS folder so you can use them in ArcGIS Pro. You should now be able to see an unzipped file called "ArcGISPro_BasicQueries_HaitiData". The zipped file will also still be there, which is useful to keep as a backup in case the data somehow becomes corrupted.

6. Now that you've unzipped the data go back to the ArcGIS Pro project. On the right side in **Catalog**, right click on the folder "ArcGISPro_BasicQueries_HaitiData" and press **Refresh.** This will allow us to see any new files or folders added to this folder.

   Note: Whenever you add new files to a folder and the project is ALREADY open in ArcGIS Pro, you must **refresh** the folder in order to see what has been added.
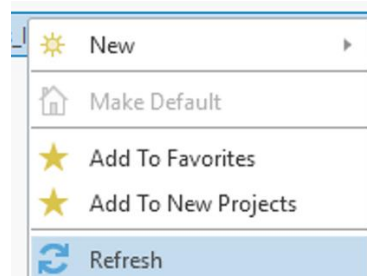
Data Sources

What you will see in the ArcGISPro_BasicQueries_HaitiData folder is:

A geodatabase (*Haiti_Geodb*):

- *gpw_v4_popdens_adj2015* **-** UN WPP-Adjusted Population Count, v4.11 (2020) Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets. Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H45Q4T5F. Accessed 8/20/2021. Spatial resolution 30 Sec (~I km^2) arc-minute in GeoTiff format, for 2020 population density. The file was extracted using a mask for the island only for purposes of this exercise.

- *healthcare_haiti_osm_20210820* – Healthcare (hospitals, pharmacies, doctor, dentist, etc.). Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org. Retrieved from https://planet.openstreetmap.org on 8/20/2021

- *hti_boundaries_communes_adm2_cnigs* – Communes

Three shapefiles:

- *hti_rdsl2_minustah.shp* - Roads, administrative geographies for level 1 and 2 (departments and communes) - USAID Data Repository of the Geographic Information Support Team (GIST) - https://gist.itos.uga.edu/ (data no longer available)

- *Haiti_ADM3_stats.shp* – Demographics (3rd-level sections) - downloaded from Harvard University, Haiti Earthquake Data Portal - https://dataverse.harvard.edu/dataverse/earthquake_data

- *Haiti_Livelihoods_Zones.shp* **-** Livelihood Zones - *Famine Early Warning System Network* (data no longer available). https://fews.net/am%C3%A9rique-centrale-et-les-cara%C3%AFbes/haiti/livelihood-baseline/march-2015

Add data to your project and change the base map

1. Find your Folders connection in the Catalog pane and navigate to where the downloaded material is. On the right side in Catalog, right click on the folder "ArcGISPro_BasicQueries_HaitiData" and press **Refresh.** This will allow us to see any new geodatabases or files or folders added to our project.
2. Since this exercise is about asking questions (querying a database), it's critical that you are familiar with the features, attributes, and attribute values in that database.
3. Take a few minutes to add the files to your project and explore the spatial data, including the attributes that are available within each layer
4. Last, let's change the base map to the OpenStreetMap
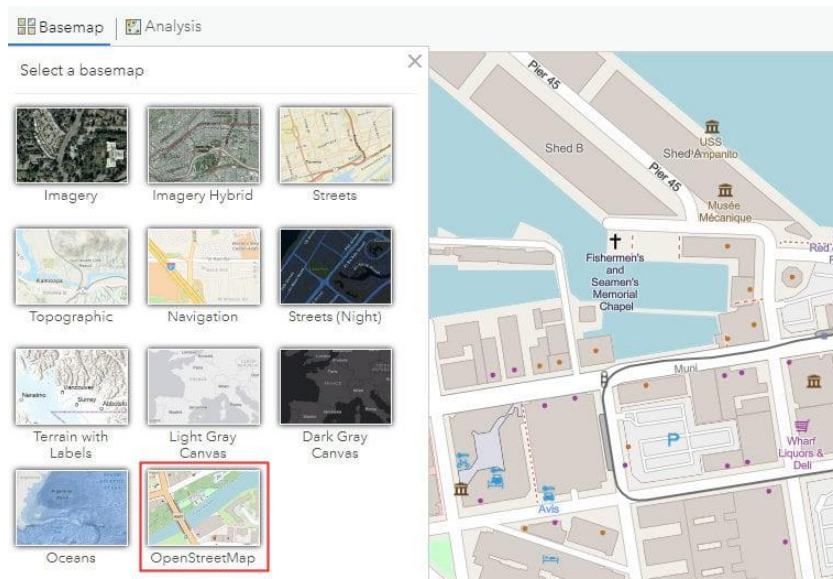
This will allow us to see more detail throughout the process if we need to.

About data and layers:
- What is the difference between the layers we see through ArcCatalog and the files we can see using Windows Explorer?
- How many layers do we have in total?
- How many features do the different layers have?
- Can you distinguish between the raster and vector layers? What differences do you see?

Adding data from the web

If we want to know the exact location where the earthquake happened, we have several options. One, is to ask for the location of the three most affected towns located. You can use the Inquiry ribbon and the tool called **Locate**, which is under the menu Map.

1. Click on the search space and type *Petit Trou de Nippes*.
2. Explore the different options provided and define which one you think is the one we are looking for.
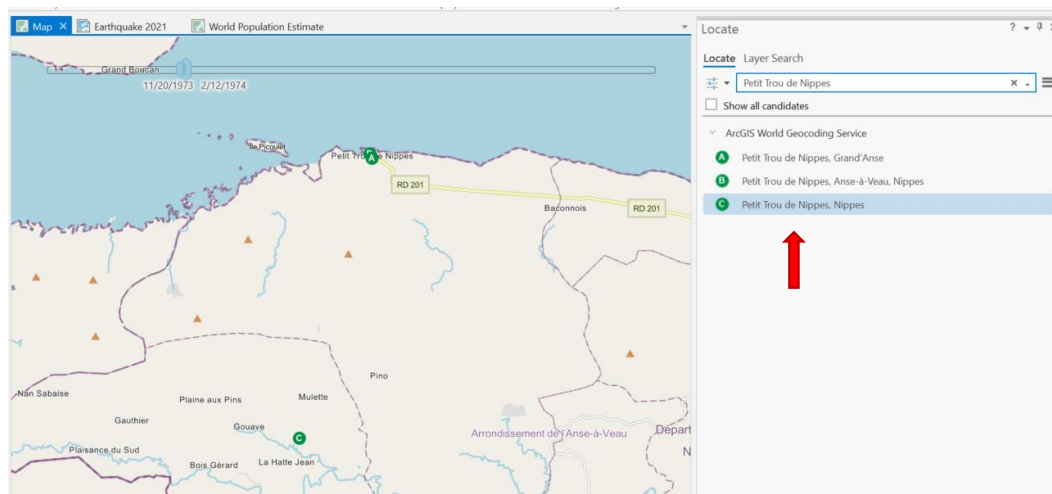
3. Repeat the same process to locate Jeremie (make sure you type *Jeremie, Haiti*) and Les Cayes (type <u>Les Cayes, Haiti</u>).

Exploring the Living Atlas

Another – easy to access – source of information to explore is The Living Atlas, which is a collection of geographic information from around the globe. This source includes maps, apps, and data layers that can be access through ArcGIS Pro and other ESRI platforms.

Now, let's try to answer our first question:

Where do the earthquake happened?

For us to get some current information about earthquakes, we will explore The Living Atlas through ArcGIS Pro:

1. Open the Catalog pane.

Note: If you closed it, you can re-open it. Click on the menu <u>View > Catalog Pane</u>

2. There are three tabs: Project, Portal, Favorites. Click on <u>Portal</u>.
3. There are several options to choose from, click on the last option <u>Living Atlas.</u>
4. Type in *Earthquake*

Results will appear (you might see options than the image below – the spatial data made available by ESRI changes over time and/or is updated in a regular basis).

Select the layers *Recent Earthquakes* and *Global Earthquake Archive*
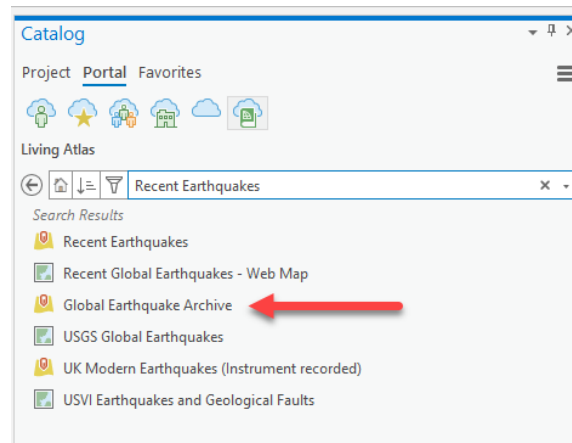
We will explore both layers.

Note: You will notice that the layers available might seem differently. Some of them are what we called **Feature layer** – more about it here and others are Web maps – more about it here

5. Add the Feature Layers (red arrow) to your project by right click > <u>Add to current map</u>

Select by Attributes

One of the main questions a decision-maker needs to ask after an earthquake is where can people go to get help? We have a layer with information about hospitals, pharmacies, etc.; hence, a first question to ask could be:

How many healthcare facilities are and where are they located?

There are a couple different ways to answer this question. Given that one of the first things you do when exploring new datasets is to look at the attributes, let's fist open the attribute table and look at the different fields; this will give you an idea if that kind of information is available and, if we can find out how many hospitals are in the area. There are different ways to access Select By Attributes:

1. Open the attribute table. You can see the option of **Select By Attributes** at the top of the table.

2. Click on **Select By Attributes** and a window tool will open. In the Input Rows the table selected is the layer *healthcare_haiti_osm_20210820*. The Selection type is New Selection. In Expression, click on New expression. Now select the attribute "*healthcare*" = "hospital" (see the following graphic):
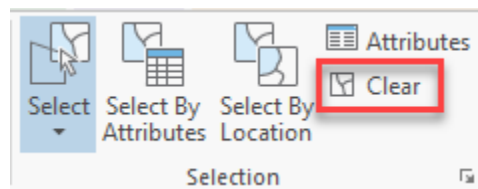

3. Click Apply and then, OK

> - How many of Haiti's hospitals did you find? (Look at the bottom of the attribute table)

4. At the bottom of the attribute table, you will see the number but also, you can select Show Selected Records (as seen the in the figure below) to just look at those records selected.
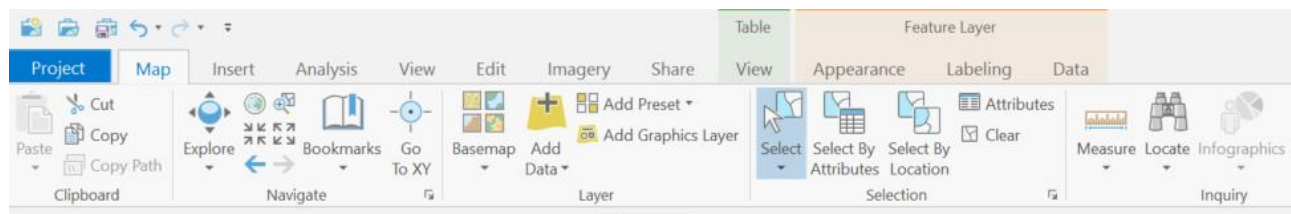
5. Clear the selection by clicking on Clear

Take a screenshot of your attribute table for Lab Exercise 2.

In order to get to the same, **Select By Attributes** tool, you can look at the ribbon Map and under the section Selection, the same option is available to access.

6. Perform a different search by your own. Let's change the question to:

- how many clinics are registered in the layer?

7. Close the attribute table.

Changing Symbology

You can identify the location of all hospitals that are in the area by displaying the data such that you can distinguish between hospital, dentist, doctor, etc. using the layer's symbology. The attribute table will be used to do the classification using symbology.

There are different ways to access the symbology pane for a layer. Let's start with the following:

1. Right click on the *healthcare_haiti_osm_20210820* layer in the **Table of Contents** > **Symbology**.

2. The **Symbology** pane will open on your left. It is time to select what Primary symbology we will use. In this case, we will select **Unique Values;** given that the symbology is based on the layer's attributes for its classification, we need to select the field that has the values we want to use for its classification or representation, in this case it is: ***healthcare***

By default, we will have 13 different symbol classes but if you want to easily distinguish between each symbol, you can play with the colors (Color scheme) or with the symbols.

3. Let's change just one (**healthcare**) so the change can be applied only on that particular class:
    a. Click on top of the symbol
    b. The Gallery under the same Symbology tool will appear. Select a **Cross 3** class and return to the previous window by clicking on the arrow pointing left.
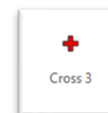
As a note: You can also group classes manually within the symbology pane if you wish. In order to do that, you need to select those classes and right click to it, followed by selecting **Group Values.**

Now, you can see where hospitals are located.

4. Get close to the three towns (i.e., Les Cayes, Jeremie and in Petit Trou de Nippes) to see the location of hospitals in those areas.


Select By Location

Sometimes we want to ask questions that are based on a particular location (area). In this case, we want to know all the hospitals located in the three regions where the earthquake affected the most: Les Cayes, Jeremie and Petit Trou de Nippes. Again, there are different ways to get there but we will use a two-step process to do it.

How many hospitals are in the most affected area?

First, we must select one of the communes among the most affected. Let's start with **Les Cayes**. The process will be more or less the following: 1) Select Les Cayes, and 2) select all the hospitals inside Les Cayes. If you think about it, what they need for this particular selection is a common location. Les Cayes is a "commune" (municipality) and its boundaries can be found in the ***hti_boundaries_communes_adm2_cnigs*** layer.

1. To select Les Cayes we can do **any** of the following:
    - Open the attribute table and sort the attributes (Sort Ascending or Sort Descending); and find the attribute (Les Cayes) we are looking for. You can manually highlight them when you find it by clicking on the left side of the attribute table. It will be highlighter in blue.
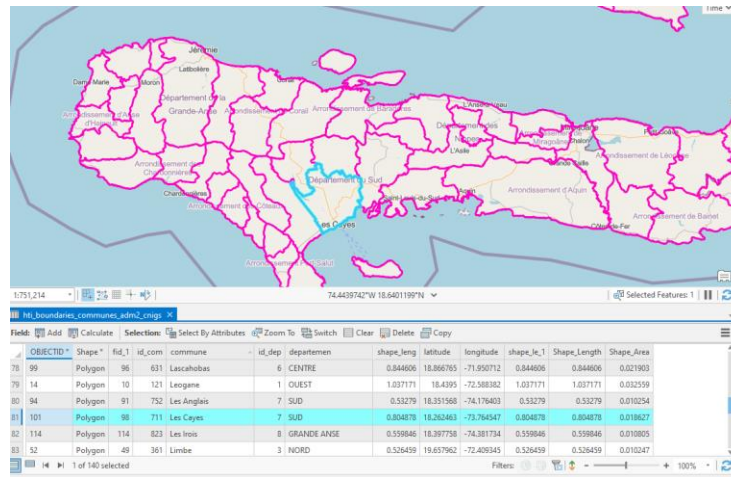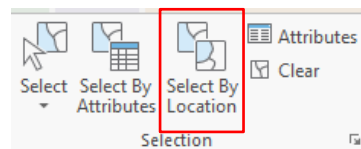
- Another way to select it is by zooming (Zoom) in on the area where Les Cayes is, and select it using the **Select Features** tool
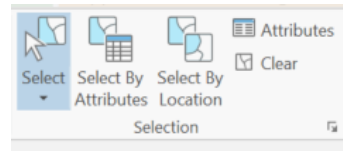
- Also, you can **Select by attribute** - from the *Commune* field where "commune" = 'Les Cayes'
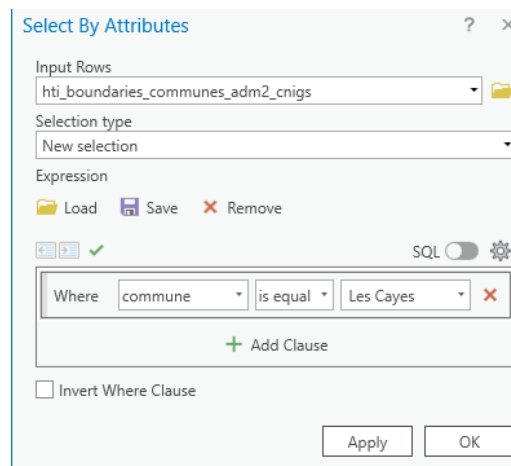
2. Once you have Les Cayes selected, exit out of the attribute table and click on the **Selection** menu at the top and choose **Select by location.**


Make sure Les Cayes stays selected because our goal is to find all the hospitals within Les Cayes.

3. A window will open.

   - The Input Features will be the layer you want to apply the selection: *healthcare_haiti_osm_20210820*
   - The Relationship will be *Intersect*

- Selecting Features is the layer you will use to perform the selection; in this case, the *hti_boundaries_communes_adm2_cnigs*
- Leave <u>Search Distance blank</u> and go to Selection type; leave it as a <u>New selection</u>
- Click <u>OK</u> and check the results

---
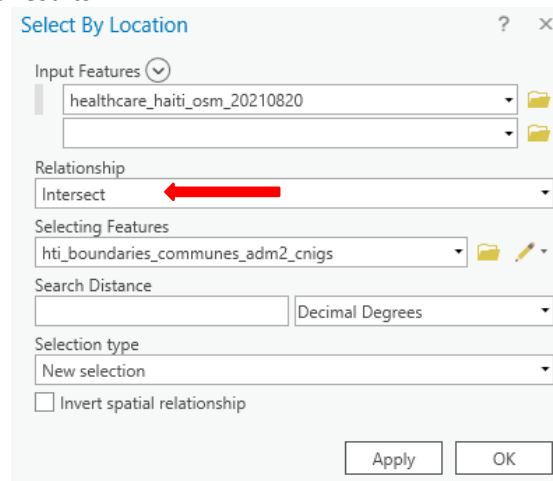
- *How many features were selected within Les Cayes?*

There are a couple ways to look for the results:

1. To simply look at the table and find the number of features selected
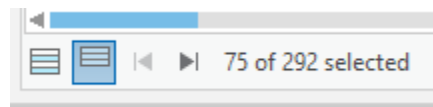
2. It is also possible to look under the table of contents. There are different tabs that allow us to see our datasets based on its characteristics. Click on **List by Selection** and you will see the list of the different layers. You will notice a number on the right of the layer, indicating the number of features selected. You will see both layers with selected features (*healthcare_haiti_osm_20210820* and *hti_boundaries_communes_adm2_cnigs*).
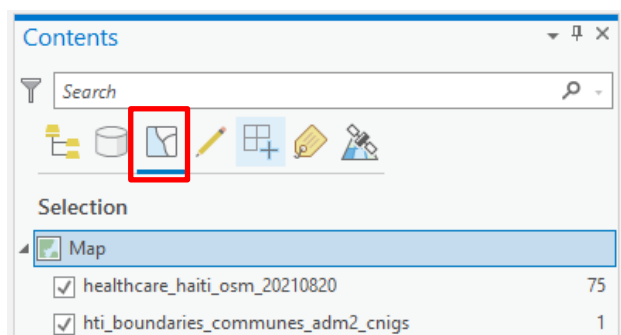

- *From our selection, are they all only hospitals?*

When we have selections, we can continue querying our results that is, we can add questions to our results. For example, if we want to know which of the **Les Cayes** features are only **hospitals**, we can add a query on top of the one already done.

We will use *Select by Attributes* with the *Method* set to **Select Subset from the current selection** (since all Les Cayes healthcare locations are already selected). We will also add more than one clause to build on our selection. See below:

3.   Click on **Select By Attributes**

4.   Under Input Rows, the layer *healthcare_haiti_osm_20210820* should be selected

5.   Under Selection type, pick the *Select subset from the current selection*

6.   Click New Expression

7.   Add first

- Where healthcare > is equal > hospitals

- Add Clause

- Or healthcare > is equal > clinic
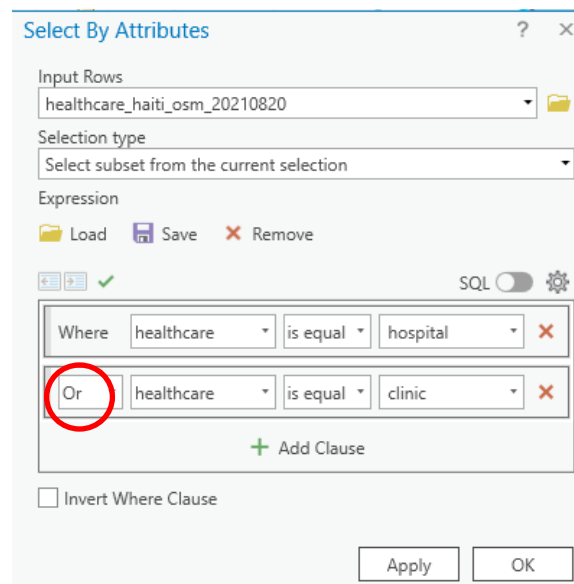
- *How many hospitals/clinics does Les Cayes has based on this dataset?*

Now, apply the same process to find the number of hospitals/clinics in the other two communes, Jeremie and Petit Trou de Nippes.

- *How many hospitals/clinics were potentially available in the other two communes?*

4. **Clear** all the selected features again using the *Clear Selected Features* icon ⬚ or the *Selection* menu option.

Explore Statistics

This section focuses on a question that might help on understanding accessibility and might help decision-makers plan ahead to send help.

How Many People Live by a Main Road?

In order to answer the question, we will need the layers of roads (*hti_rdsl2_minustah)* and the layer of population density (*Hait_ADM3_Stats)*. If we are thinking about main roads, look at the attributes of the layer roads and think about what field provides the information we are looking for.

> - *What is the name of the field?*

1. First, look at the different classes we have and determine what you would consider as a "main road".
2. S**elect** the features that belong to main roads from the *hti_rdsl2_minustah* layer by using the attribute of **Class =" Principal".** Make sure your selection method is set to *"Create a new selection".*

Note: follow the previous instructions on how to make New selections

3. Next, we will use **Select By Location** to select all *Haiti_ADM3_stats* polygons that are within 1 mile of those selected roads (see the following graphic).
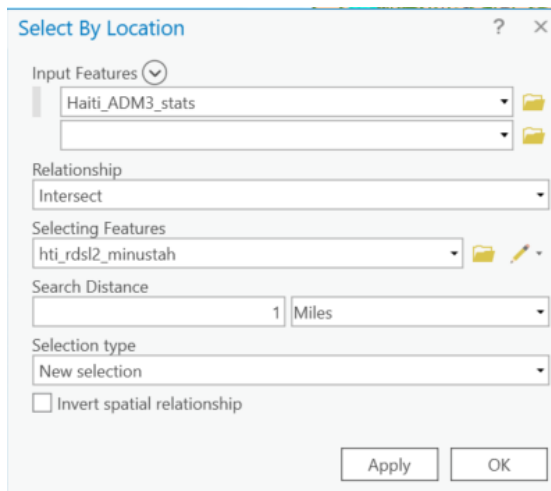

Take a screenshot of your attribute table for Lab Exercise 2.

4. Open the *Haiti_ADM3_Stats* attribute table to see the selected features that are within 1 mile of a main road.
5. Right click on the field *POPULATION* and select **Statistics**
6. A new window will open with the distribution of population.

Check the results – On your right, you will have a window with the Distribution of POPULATION providing some basic statistics on the data selected. For example, we have a total (Sum) of 5,152,393 people living within these

220 Administrative zones. On average (Mean), there are 23,419 people PER Administrative zone. Also, there's a distribution chart that shows the data distributed on a graph.
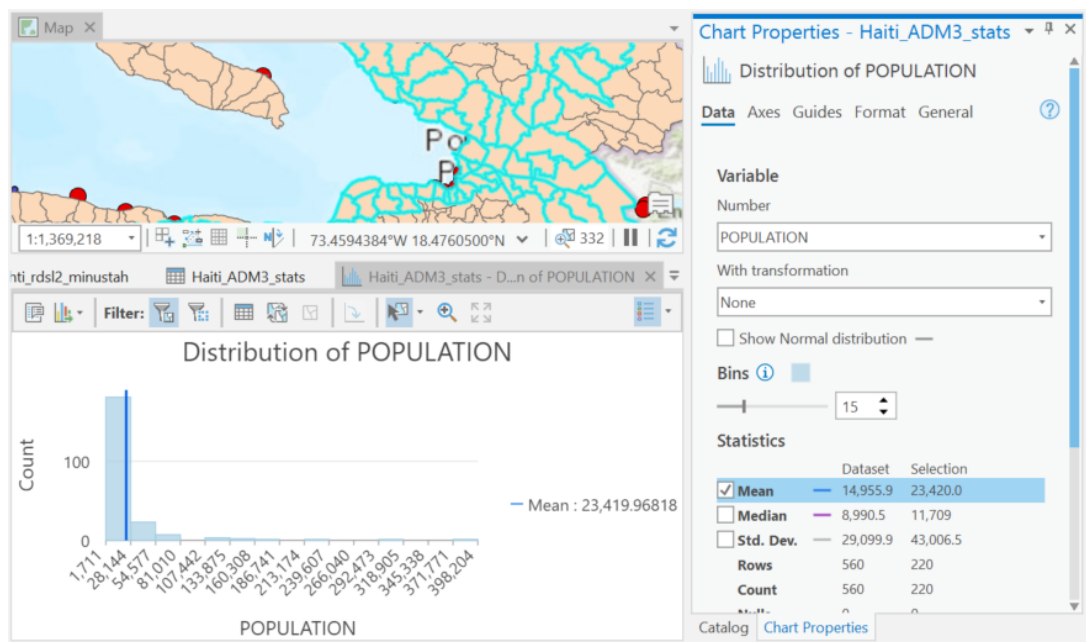
- *Why do you think we want to use the Sum field to understand the total population of those selected districts?*
- *What does the mean field represent?*

Note: To see the total population of ALL the communes (and thus the total population in Haiti), unselect the communes and run the *Statistics* function again.

Take away point: The *Statistics* function gives you basic statistics describing numeric values for EITHER a selected set of features OR for all features if no features are selected.

Clear the selections before going to the next section.
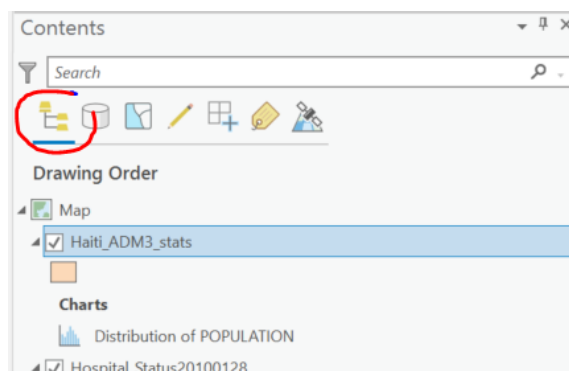
Spatial Join

A Spatial join is a GIS operation that affixes data from one feature layer's attribute table to another from a spatial perspective (more information at: Spatial Join).

A question to contextualize Haiti's healthcare system in terms of access to hospitals or other healthcare entities can be:

How many healthcare providers are in each of Haiti's Departments?

In order to answer this question, is important to think spatially. This about the elements you have in the map to represent departments (administrative level of representation), and healthcare entities.

There are different possible methods to approach the problem. The first one is to repeat a similar approach to what we have previously applied in this exercise. However, this would take us repeating the process over and over for each one of the departments: 1) select a department; 2) apply a spatial selection to hospitals based on the department selected; and 3) repeat the process for each one of the departments. However, there's a tool called Spatial Join, which joins attributes from one feature to another based on the spatial relationship that is, if they are located in the same geographic space.

First, we need to make sure we are back to the display Drawing Order in our Contents window.

1. Right-click on *Haiti_ADM3_stats* (i.e., Department Boundary) and choose **Joins and Relates → Spatial Join.**

A Geoprocessing tool window will open on your right.

2. Fill out the dialog box as in the following graphic – make sure you check on the following link Spatial Join Analysis > Syntax for more information about each one of the Parameters.

Within the Field Map section, you can determine how you want to see the fields that will be associated to your layer, summarized. For example, if I only want to know the number of healthcare facilities in each one of the communes, I can request to get the "count" (see figure on your right). The attribute table resulted from this analysis, will provide that number for each commune.

Note: In ArcGIS Pro, the file will always be saved in the geodatabase created for the project. If you want to save it as a shapefile, you will have to specify so by changing the location to an actual Folder, outside the geodatabase.
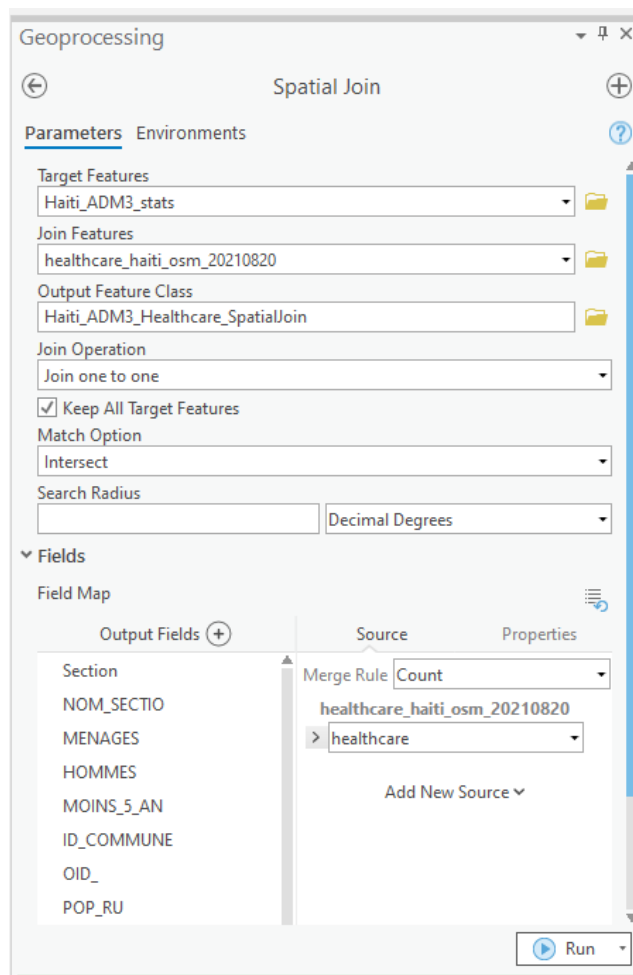
3. Click Run.

4. The output is a new layer file that is added to your ArcGIS Pro session – open the attribute table to see what happened – you'll see a count of healthcare facilities in each district at the end of the table, under the field called healthcare.

Make sure to take a screenshot of your attribute table for Lab Exercise 2.

> - *What information do you need to know if you would like to know the person per healthcare facility ratio in a region?*
>
> *This would be a challenge, and you don't need to do this today. But why would you say it might – or not - be difficult to calculate?*

Add Fields

Sometimes we receive or get datasets that can be enriched with additional information. In our particular case, we have a dataset (*Haiti_Livelihoods_Zones*) that can help us to provide more context to our exploration of Haiti. This layer provides information about the ways in which households support themselves from year to year and survive during hard times. There is a growing interest in the use of livelihoods analysis as a first step in gaining a better understanding of many development-related issues, ranging from emergency response and disaster mitigation to long-term development. Hence, one of the questions we can ask is:

How many square kilometers of each type of Livelihood Zone are there?

This might provide some context additionally to simply visualizing the size of livelihood zones and the territory it cover; however, one of the missing data is the area (check on the attribute table and look for area).

It is possible to add a field in your attribute tables and calculate the area. There is a tool called Calculate Geometry and we can generate summaries of our data to generate a report. We will start by adding a field to hold the square kilometer value (area):

1. Open up the *Haiti_Livelihoods_Zones* attribute table.

Note: you can always close other attribute tables if you are not using them.

2. Click on Add 🏷 Add - and the *Fields view will open. A list of attributes will appear together with their characteristics including name of the field, alias, data type, etc.

3. At the bottom of the list, you can add more fields.

4. Click on Field and add a new field with the name **Sq_Km**

5. Under Data Type, you can include Double.

6. For the name, remember to use short names without spaces (use hyphens instead) or special characters. The name should be no longer than 8 characters.

Note: Choosing the correct data type allows you to correctly store the data and will facilitate your analysis, data management needs. The available types include a variety of number types, text types, date types, binary large objects (BLOBs), or globally unique identifiers (GUIDs). For more information, click the following link ArcGIS field data types or copy paste the following address: https://pro.arcgis.com/en/pro-app/2.7/help/data/geodatabases/overview/arcgis-field-data-types.htm

7. Once you create a field, you will need to Save in order for it to be available. Under Feature Layer at the top (ribbon), you will have that option.
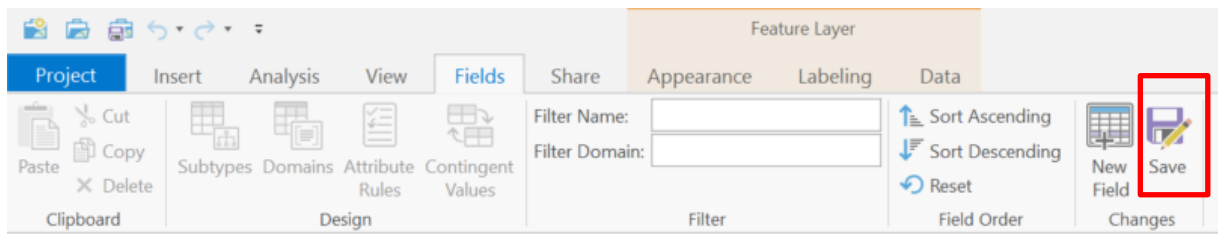
Note: New fields are always at the end of your attribute table – scroll to the right and find the new field.

8. Go back to the attribute tab and right click on the new field. Choose **Calculate Geometry** and a window will appear.

9. Fill out the form making sure that it looks like the image below. To select the Coordinate System, we need to select one that allow us to make the right calculation. Please select UTM by clicking on the little globe you will see on your right

10. A new window will appear. On the Coordinate Systems tab, on to the <u>search bar</u>, type UTM Zone 18N

11. Navigate to UTM > WGS 1984 > Northern Hemisphere and select <u>WGS 1984 UTM Zone 18N</u>

12. Click OK
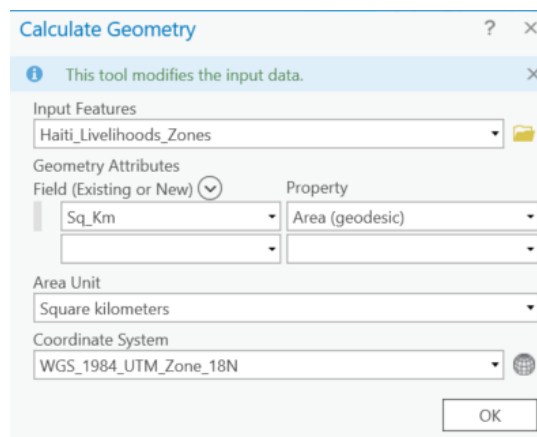
## A note on Projections/Coordinate Systems

Before we can <u>calculate area</u>, we need to choose the right projection. The reason why is because every map deforms the earth. It is impossible to represent a curved surface (the earth) on a flat one (a map) without stretching, skewing, and tearing it. Hence, to calculate distance and areas, we need to make sure we have the right map projection.
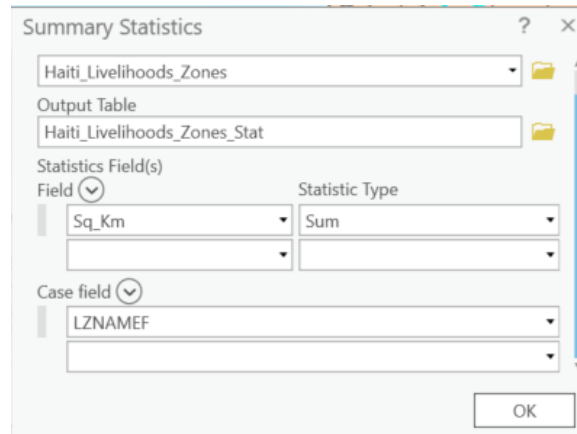
Universal Transverse Mercator (UTM) is a system that is commonly used for <u>large-scale maps</u>. In our particular case, our map is narrower than 6° of longitude, hence we can map it with a UTM projected coordinate system.

Haiti coordinate system in UTM falls in Zone 18 N, and we will use WGS 1984.
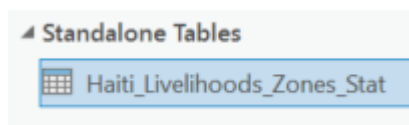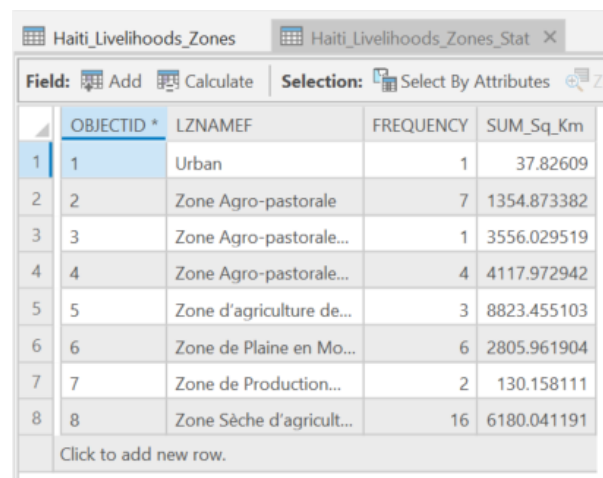
Summarizing Categorical Data

Next, we can focus on answering the question of area per zone given that we have calculated the area in square kilometers of each polygon. By selecting manually each type and use the *Statistics* function to tell us, the process might take us a long time. However, there's a way to summarize the information by using the tool Summarize.

1. Make sure to clear any selection ⌧ Clear

2. Right-click on the *LZNAMEE* field type and choose **Summarize.**

3. Fill out the form as follows – be sure to give the new table a name or in this particular case, you can leave the default name however, make sure you pay attention to it such that the data don't get lost.

4. Click OK and look for the table added to the **Contents** pane.

Important: *Summarize* can ONLY be used on **Category** (Nominal) type values (like livelihood zone or land cover or clinic type) and is used to aggregate numerical values (like square kilometers). The function always gives a count, but you pick how else to aggregate the values (e.g., max, min, average, variance, standard deviation). It will work on ALL features if nothing is selected, **OR** on the selected set.

Summing up what you've learned

The tools you learned in this exercise are ones you will use repeatedly in GIS. They form the foundation of basic GIS analysis. There are many more advanced analyses tools, but you should become very familiar with the ones we used here so that they eventually become second nature to you.

- You can make **queries (Select By Attributes)** on individual layers to select out features based on certain attribute values in their tables.

- You can chain selections together in various ways to select out subsets of features or add to or remove from selected sets.

- You can **Add a Field** and **Calculate** area (**Geometry**) for polygon features, and length for lines.

- You can **Summarize** information based on categories.

- Most importantly, and this is the true value of GIS over maps, you have learned that you can look at **Spatial Relationships** *between* layers – some of the ways you can look at these relationships include the following:

  - Selecting features based on their spatial relationship (**Select by Location**) to other features (e.g., inside another feature or within a specified distance other features)

  - Passing information from one layer to another layer (using **Spatial Join** for vector data or **Zonal Statistics** for underlying raster data)

References

Clarke, Keith (1997) Getting Started with Geographic Information Systems, Upper Saddle River, NJ: Prentice Hall.