Basic Queries Exercise - Haiti

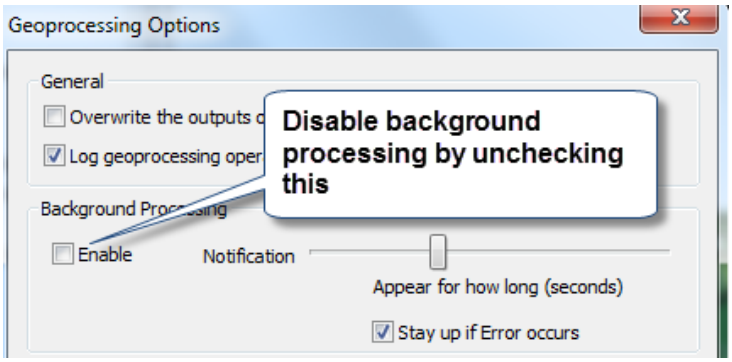
Written by Barbara Parmenter, revised by Carolyn Talmadge on September 18, 2016

In this exercise you will learn how to use the basic analysis GIS tools. We will use data for Haiti that was put together in the days after the 2010 earthquake. Functions to be covered:

Select by attribute
Select by location
Statistics for selected features
Add a field to a table
Calculate geometry
Summarize by an attribute category
Spatial join
Zonal statistics

Setting Up

1. On the S: drive, navigate to S:\classes\DHP_P207\Basic_Query_Practice
2. Copy the **Basic_Query_practice.mxd** to your H: drive.
3. Double – click on it in your H: drive to open our Haiti map file.
4. Once ArcMap opens with the Haiti map, click on the **Geoprocessing** menu, and choose **Geoprocessing Options**, then disable *Background Processing* as shown below:



5. Click OK.

The data in this exercise comes from:

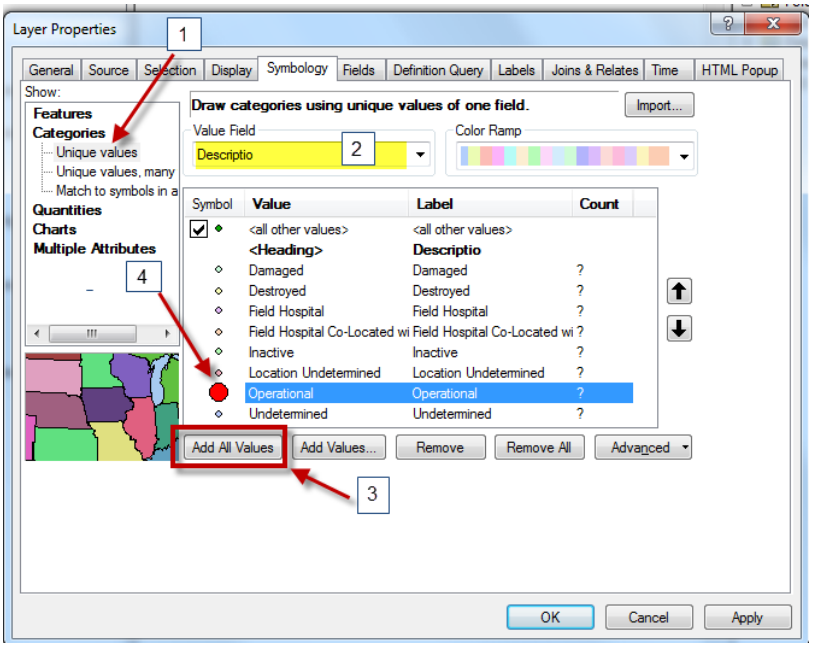
- Livelihood Zones - *Famine Early Warning System Network* (FEWSNet) - http://www.fews.net/central-america-and-caribbean/haiti/livelihood-zone-map/thu-2009-11-19 (data no longer available)
- **Roads, administrative geographies for level 1 and 2 (departments and communes)** - US AID Data Repository of the Geographic Information Support Team (GIST) - https://gist.itos.uga.edu/
- **Haiti_ADM3_stats** – Demographics (3rd-level sections) - downloaded from Harvard University, Haiti Earthquake Data Portal - http://cegrp.cga.harvard.edu/haiti/?q=resources_data
- **Hospital damage status (Hospital_Status20100128**) – US Department of Health and Human Services
- **glp10ag** – Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT). 2005. *Gridded Population of the World Version* 3 (GPWv3): Population Grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw. (3/10/2011). We have downloaded the 2.5' resolution version in ESRI Grid format, for *2010 population estimates*.

When querying a database, it's critical that you are familiar with the features, attributes, and attribute values in that database. Take a few minutes to explore the data sets and what attributes are available in the data set's attribute table

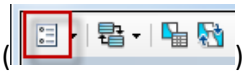
Where are the hospitals that are still operating? (Select by attribute…)

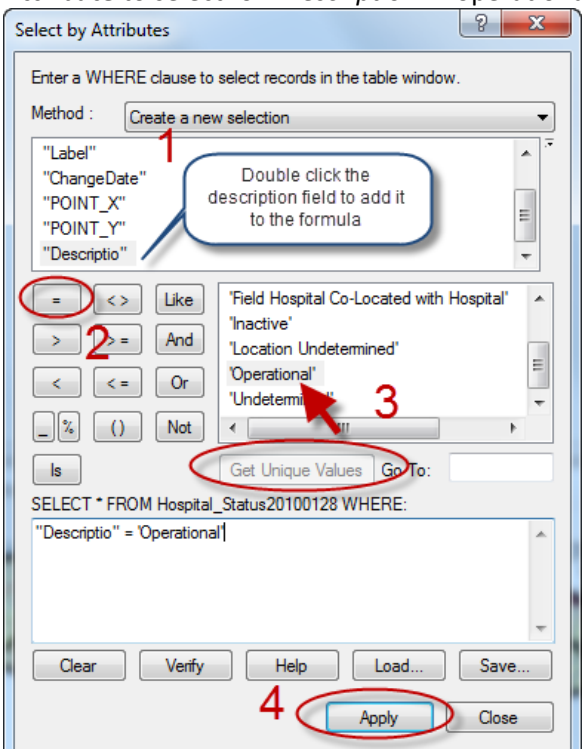
There are a couple different ways to do this:

1. You can visualize this information by adjusting the Hospital layer's symbology properties.
    a. In the **hospitals** Layer Properties box, go to *Symbology* and choose Categories-unique values (1);
    b. Use the *Descriptio* in the *Value Field* (2)
    c. Click on *Add All Values* tab (3) to assign symbols to different status types.
    d. Click over the dot of "Operational" so that you can change the size and then see it on the map(4)

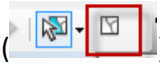


2. Or you could select operational hospitals by using **the Select by attribute function**:

    a. In the **hospitals** attribute table, click the **Table Options** Icon () and use **Select by Attribute** to select for "*Descriptio"* = "operational" (see the following graphic):



How many of Haiti's hospitals are operational (see the bottom of the attribute table)?

3. Look at the map and attribute table to see the results.
4. Clear the selected features with the *Clear Selected Features* icon ()

Where are the hospitals in Port-au-Prince? (Select by attribute, select by location…)

To answer this question using the *Selection* tools is a two-step process – you must first select Port-au-Prince, and then select all the hospitals inside Port-au-Prince. Port au Prince is a "commune" (municipality) and its boundaries can be found in the *Commune Boundary* (Admin2) layer.

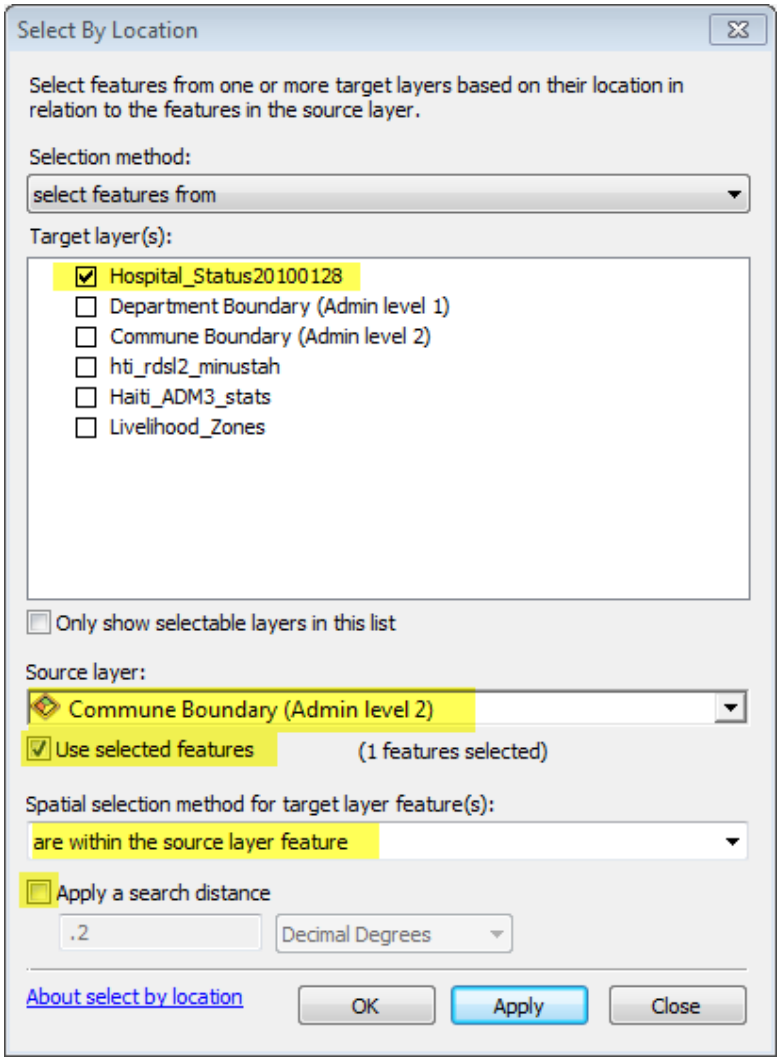
1. To select Port-au-Prince you can do **any** of the following:

    a. Zoom in on it if you know where it is and click it with the **Select Features** tool 

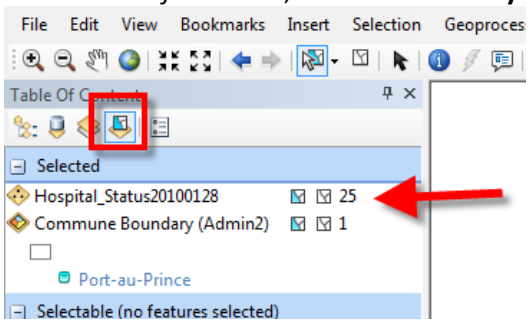
    b. **Select by attribute** - from the *Commune Boundary* layer select for where "A2_Name" = 'Port-au-Prince'

    c. Find it in the attribute table and highlight that row

2. Once you have Port au Prince selected, click on the **Selection** menu and choose **Select by location** to find all the hospitals within the selected feature of the *Commune Boundary* layer (Port-au-Prince), as shown in the following graphic:

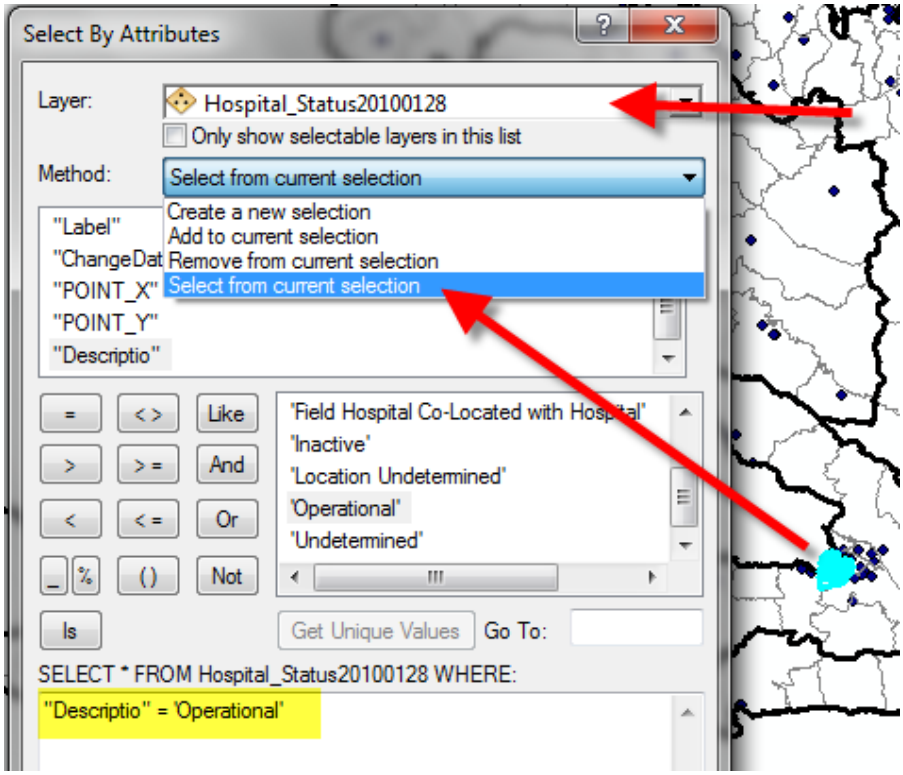


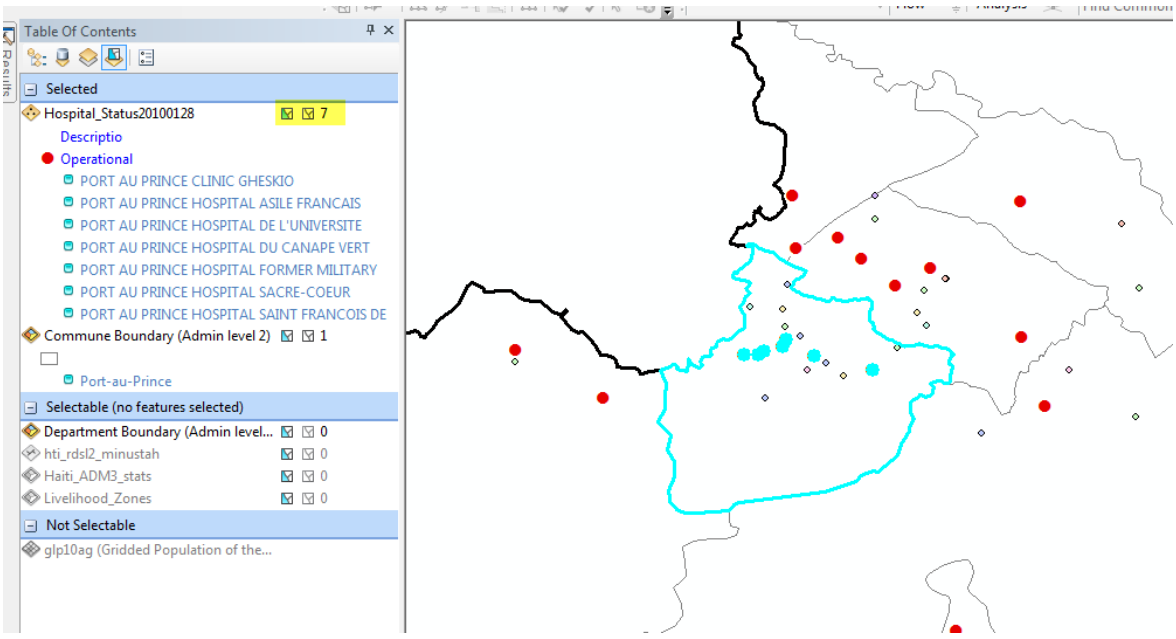
3. Check your results. How many hospitals are selected? There are a couple ways to find out:
    a. In the *Table of Contents*, click on the **List by Selection Tab** to see selection results:



    b. Open the attribute table – at the bottom it tells you how many are selected, and you can choose to see only the selected features.

4. Finally, what if we wanted to know which of **the Port au Prince** hospitals were **operational**? Again, there are several ways, but one is to use *Select by Attributes* with the *Method* set to **Select from the Current Selection** (since all Port Au Prince hospitals are already selected). See below.





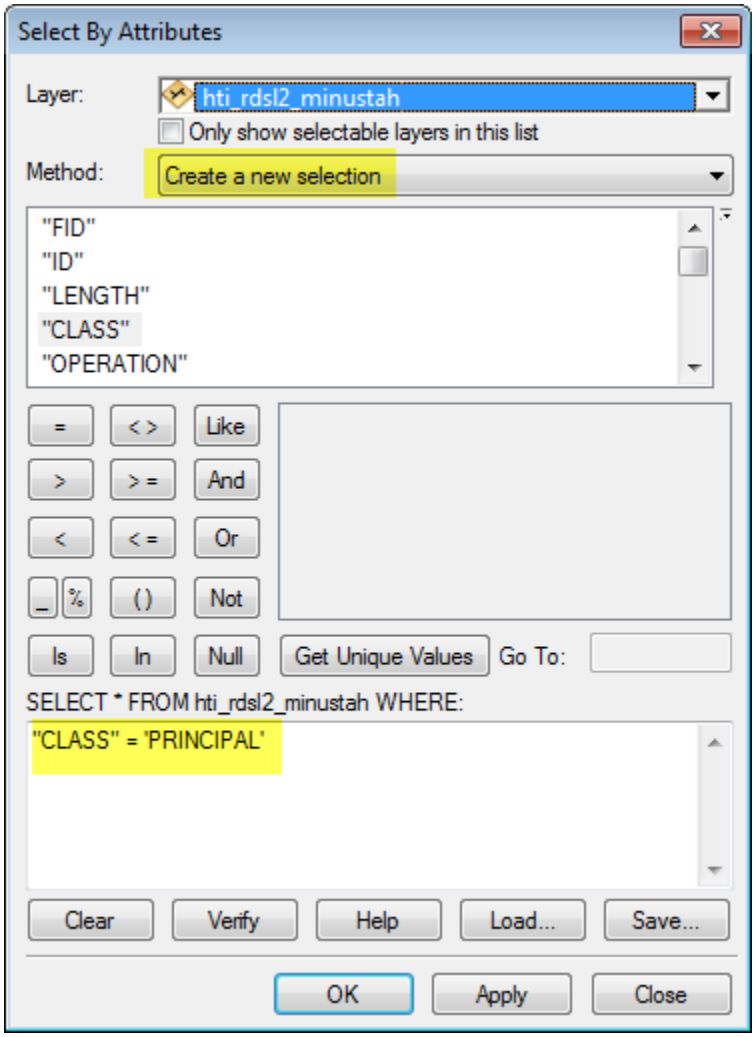
Now we have all the operational hospitals within Port-au-Prince selected. How many hospitals were operational in the days after the quake? Why might your answer be wrong?

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

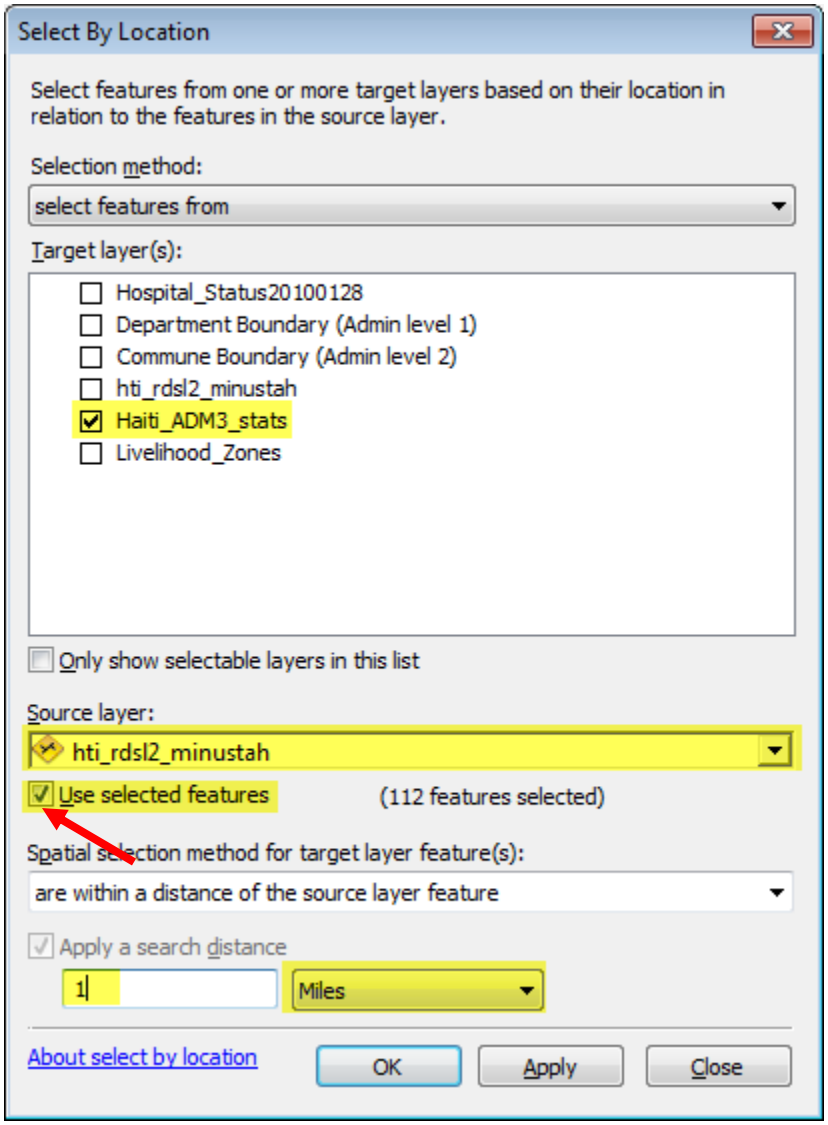
How Many People Live by a Main Road? (Select by Location, Statistics)

Select by location and then view Statistics

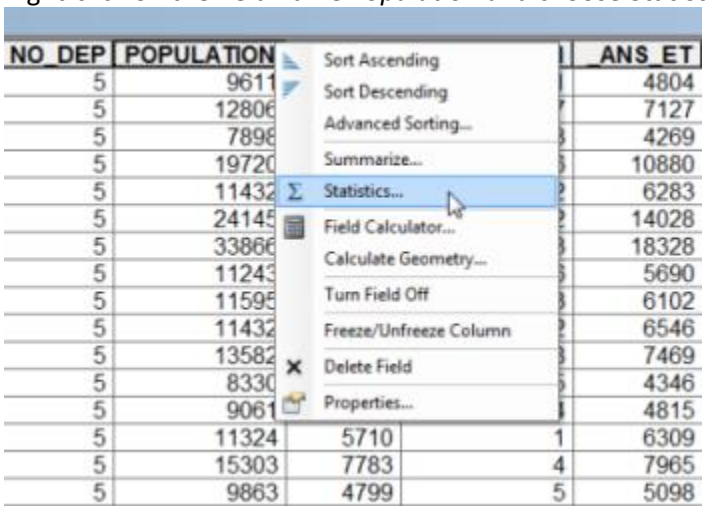
1. Turn on the *roads* (*hti_rdsl2_minustah*) and Hait_ADM3_*Stats* layer which contains population data.
2. First select the main roads from the **hti_rdsl2_minustah** layer, using the attribute of *Class="Principal"*. Make sure your selection method is set to ***"Create a new Selection"***.



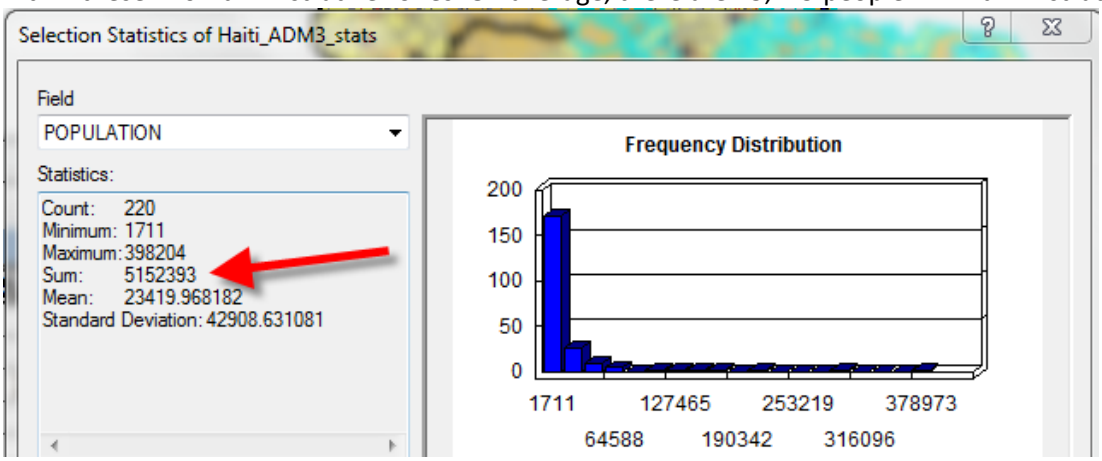
3. Next, select all *Haiti_ADM3_stats* polygons that are within **1 mile** of those ***selected*** **roads** (see the following graphic).



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 name *Population* and choose **Statistics…**



6. Check the results – **Sum** aggregates all the ***selected*** population values. This means that 51,52,393 total people live within these 220 Administrative zones. On average, there are 23,419 people PER Administrative zone.



7. 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.

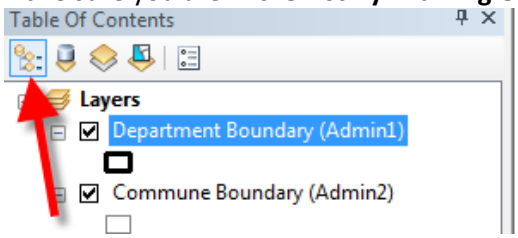
Clear the selected features again.

How many hospitals are in each of Haiti's Departments (Administrative level 1)? (Spatial Join)

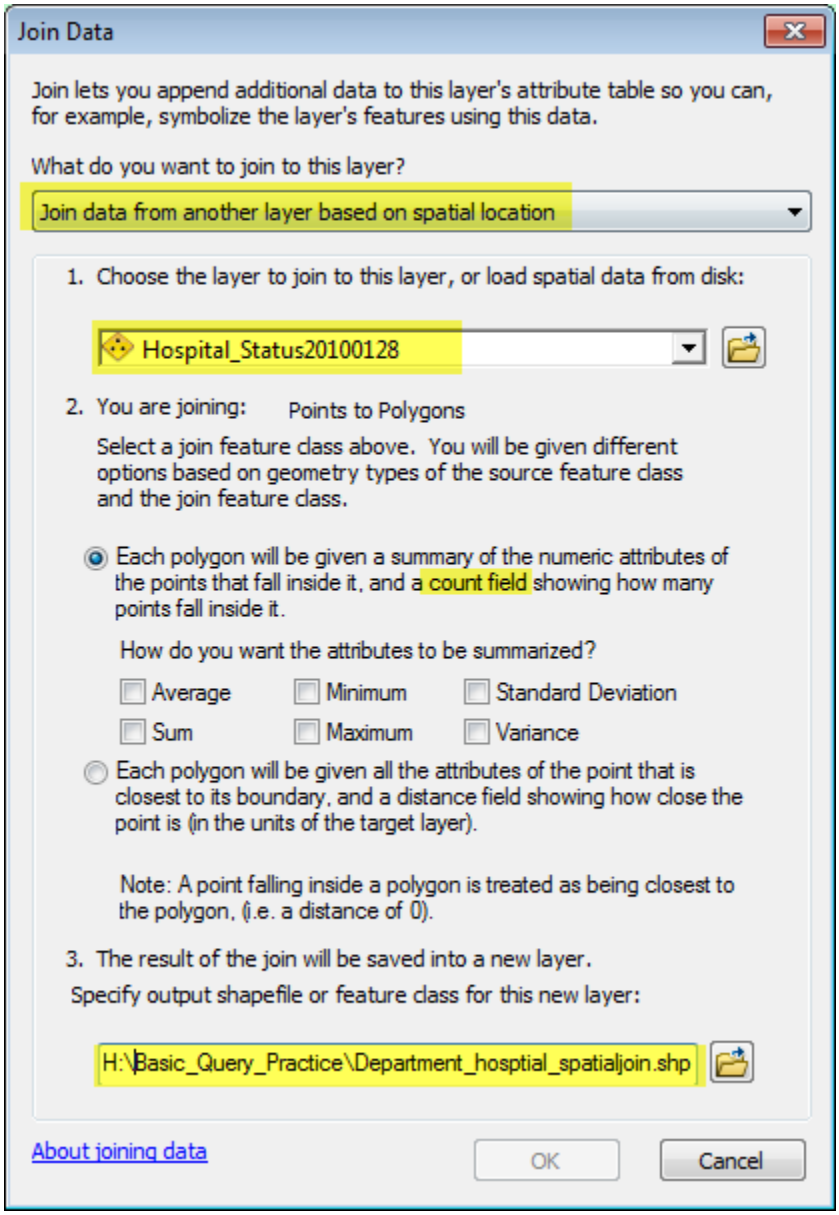
Possible Methods:

- Select each department, then select the hospitals inside that department – do that over and over and over…..not a good use of time!
- Use a spatial join – much more efficient!

1. Make sure you are in **the List By Drawing Order** view on the *Table of Contents*:



2. Right-click on *Department Boundary* and choose **Joins and Relates →Join.**
3. Fill out the dialog box as in the following graphic – make sure you select the join based on spatial location, and you give the new output file a name and location you'll remember. **The default format is not usually a shapefile. Be sure to save it as a shapefile.**



4. Click OK.
5. The output is a new shape file that is added to your ArcMap session (at the top) – open the attribute table to see what happened – you'll see a count of hospitals in each district at the end of the table.

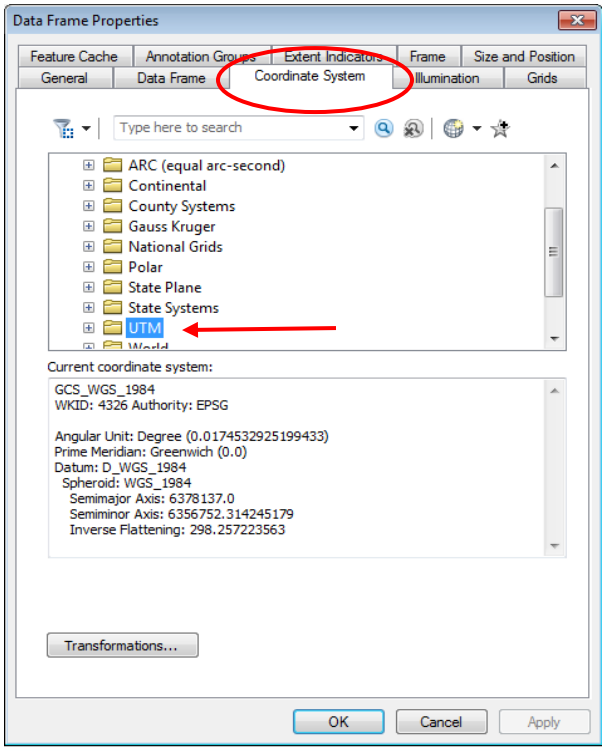
What if we wanted to know the person per hospital ratio in a region? (This would be a challenge, and you don't need to do this today. But why is it difficult?)

How many square kilometers of each type of Livelihood Zone are there? (Add Field, Calculate Geometry, Summarize)

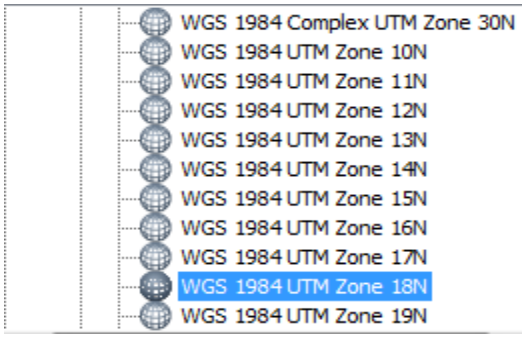
Right now we can visualize the size of livelihood zones, but we don't know how big each zone is. There is no area attribute field in the attribute table. So we have to add one!

Before you can calculate area, your data frame MUST be projected! The **UTM coordinate system** is an appropriate choice in this case. Recall from the *Haiti Mapping* exercise that Haiti is in UTM Zone 18N (WGS 1984 datum). You can always use Google Earth to find the UTM zone.

1. To set the projected coordinate system to UTM Zone 18N, WGS 1984:
    a. Click on **View → Data Frame Properties.**
    b. Go to the *Coordinate System Tab* and in the top selection box, find *Projected Coordinate Systems – UTM – WGS 1984 – Northern Hemisphere:*



    c. Then scroll to find **WGS 1984 UTM Zone 18N**:

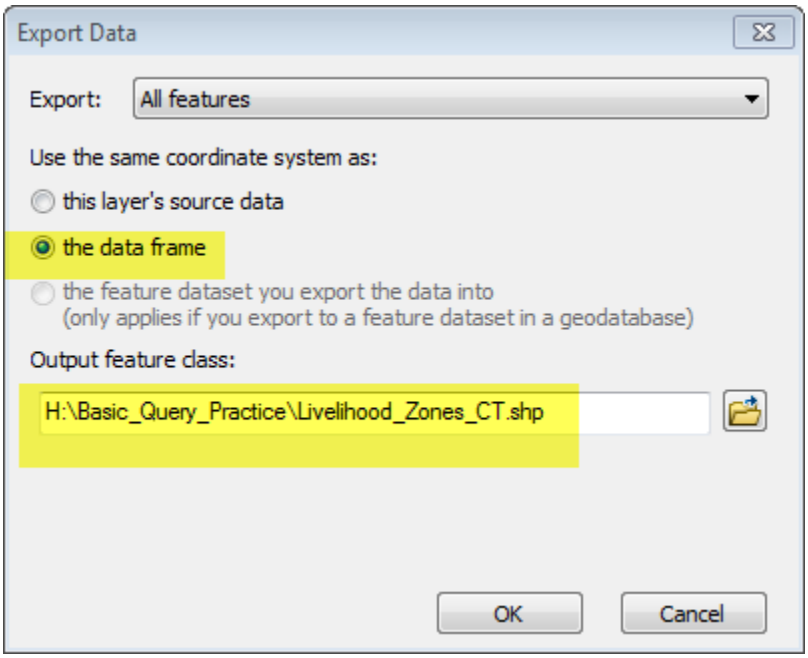


    d. Click **OK.** Did you see the map shift? It was changing the projection!
    e. Click on **File → Save as** to save your map file to your H: drive.

Now you need to make a copy of the **Livelihood Zone** data set and place it in your own personal folder (H: drive) so that you can make changes to it. To do this:

2. Make sure no features are selected using the *Clear Selection* icon ()
3. Right-click on **Livelihood Zone** data layer and choose **Data → Export Data.**

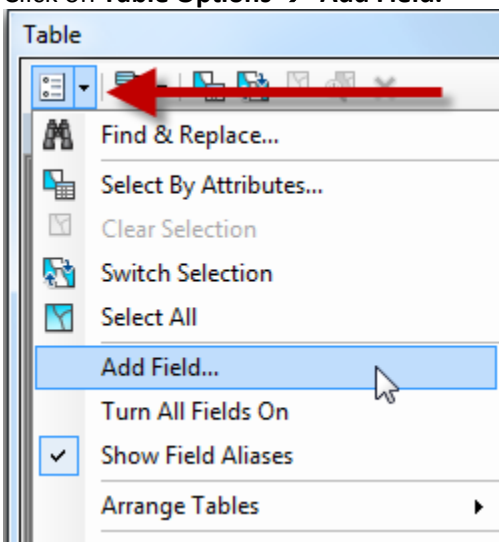
4. Fill out the dialog box as you see here, and substituting with your initials in the new file name. Be sure you **save it as a shapefile.**



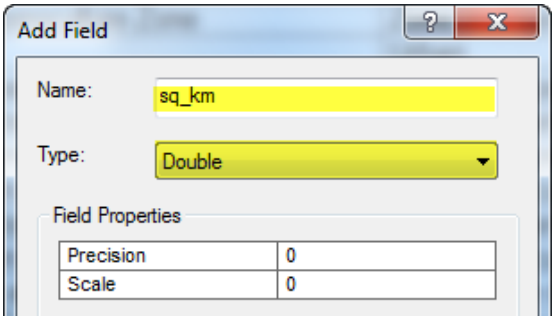
5. Click yes when asked whether to add it to the map.

Now you'll add an attribute field to hold the square kilometer value:

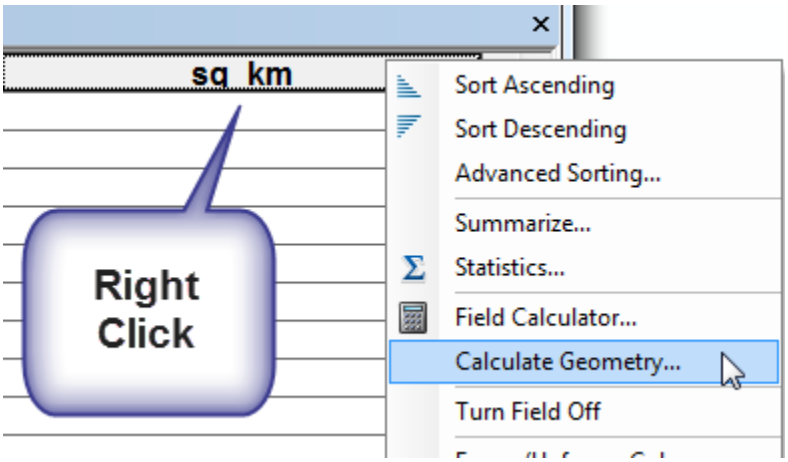
6. Open up the *Livelihood_Zone_yourinitials* attribute table.
7. Click on **Table Options → Add Field.**



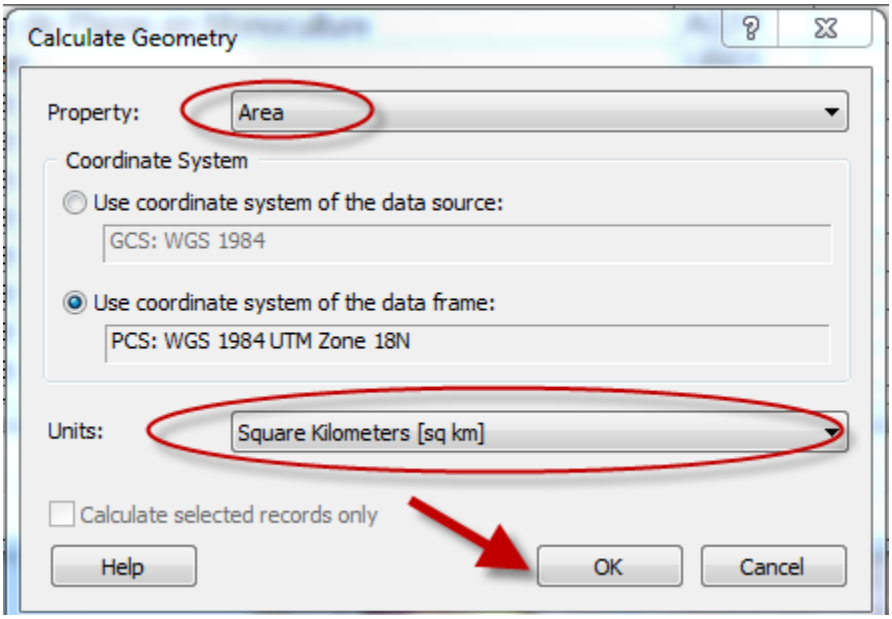
8. Create a new field for **Sq_Km** that is of the type ***double*** - note you cannot have spaces or hyphens or special characters in your field name and they must be 8 characters or less:



9. Press OK. New fields are always at the end of your attribute table – scroll to the right and find the new field.
10. Right-click on the *Sq_km* field name and choose **Calculate Geometry**…



11. Ignore the warning (hit yes).
12. Fill out the form:

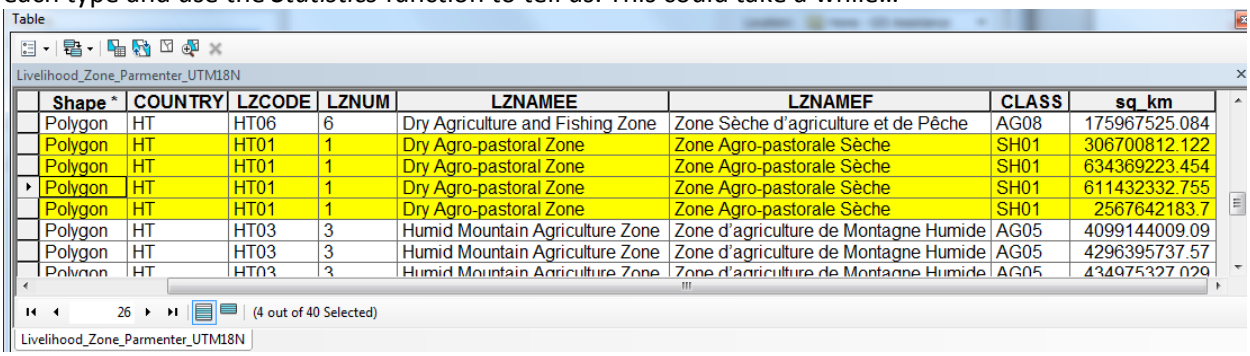


Important! Before you can calculate area or distance, you **MUST** be sure that your data frame is in a **projected** coordinate system, **NOT** GCS WGS 1984 or GCS NAD 1983! The latter are not projected coordinate systems! You will see the *Area* and *Distance* calculations disabled if your data frame is not projected.

Summarizing Categorical Data

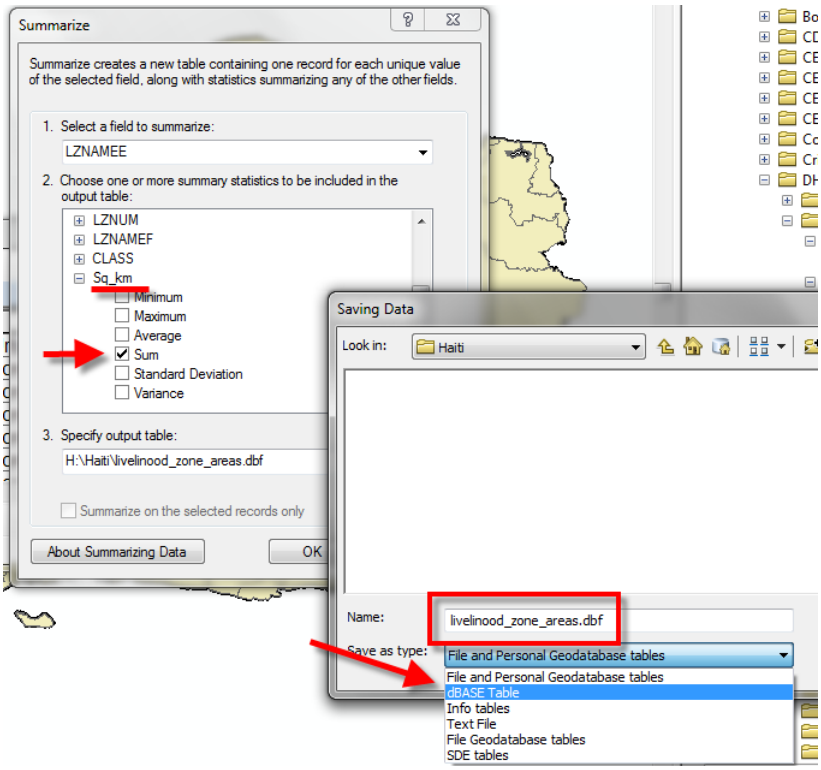
So, our question was how many square kilometers of each type of *Livelihood Zone* are there?

We can answer this now because we know the size in square kilometers of each polygon, so we could manually select each type and use the *Statistics* function to tell us. This could take a while…

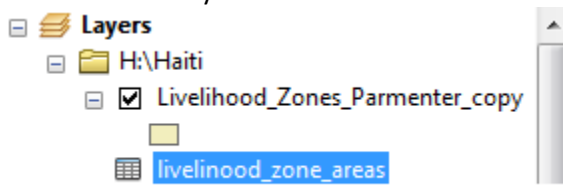


But there's an easier way – the **Summarize** tool!

1. Make sure to clear any selection (**Field Options → Clear Selection**).
2. Right-click on the *LZNAMEE* field type and choose **Summarize.**
3. Fill out the form as follows – be sure to give the new .dbf table file that will be created a clear name, and save it as a dBASE Table format, then click OK.



4. Click yes when asked if you want to add it to ArcMap.
5. Find this table in your table of contents. Go to **list by source view**, right-click on the layer and **Open.**



OID	LZNAMEE	Count_LZNAMEE	Sum_sq_km
0	Agro-pastoral Zone	7	1354.70691
1	Dry Agriculture and Fishing Zone	16	6183.94028
2	Dry Agro-pastoral Zone	4	4120.144552
3	Humid Mountain Agriculture Zone	3	8830.515074
4	Plains under Monoculture Zone	6	2808.843014
5	Plateau Agro-pastoral Zone	1	3561.797106
6	Sea Salt Production Zone	2	130.237343
7	Urban	1	37.870043

Count of Polygons

Total Area of Each Zone

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 or enrollment). 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.

NEVER try to *Summarize* on a numeric field that represents continuous values (e.g., population numbers). Why not? Try it on the *Admin3_Stats* population field. What happens?

How can we estimate the population of each livelihood region? (Zonal Statistics)

To answer this involves using an overlay operation. We will be overlaying our **Livelihood Zones** data layer over each of the following data sets to create two estimates of population per zone.

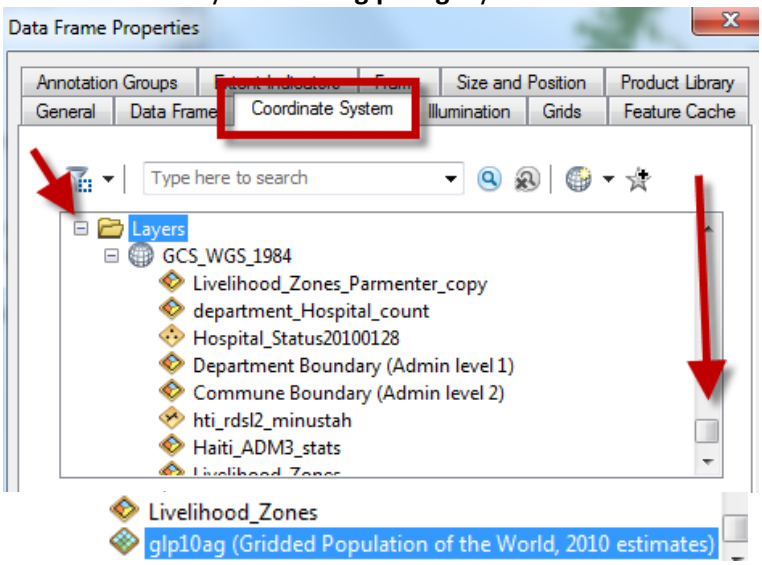
We have two data layers that attempt to estimate the population of Haiti:

- **glp10ag**
- **Haiti_ADM3_stats**

We'll first show you a way to do this using the **Gridded Population Data** set (raster) which is available for the entire world, and thus is very useful.

1. Right click on the **glp10ag (Gridded Population of the World)** and choose **Zoom to Layer**. Note the world is very distorted because the data frame is in the UTM coordinate system. Let's set the coordinate system to the same one as the Gridded Population data set.
2. Double click on **Layers** in the *Table of Contents.*

3. Go to the Coordinate System tab.
4. Scroll all the way down until you see **Layers** and click on the **+** next to it.
5. Click on the **+** next to **GCS_WGS_1984.**
6. Scroll down until you see **the glp10ag** layer and click on it.



7. Click OK to close the Data Frame Properties dialog box. This just changed the data frame properties to match the gridded population layer's coordinate system. Notice how it changed again.
8. In the *Table of Contents*, right click on the **GLP10AG layer** again and choose **Zoom to Layer**. Explore the data set by zooming to different places in the world. Each raster cell is an estimate of the population with that cell.
9. Zoom back to Haiti when you are finished and look at the gridded population for Haiti. Notice how big the raster cells are!

Overlay using Zonal Statistics

The **Gridded Population Data** (*glp10ag* in our exercise) is a raster data set. The value of each raster cell is the estimated population in that cell.

The *Zonal Statistics* function allows us to overlay a zone layer (livelihood zones in our case) over a raster data layer and aggregate the underlying raster cell values up to the zone layer.

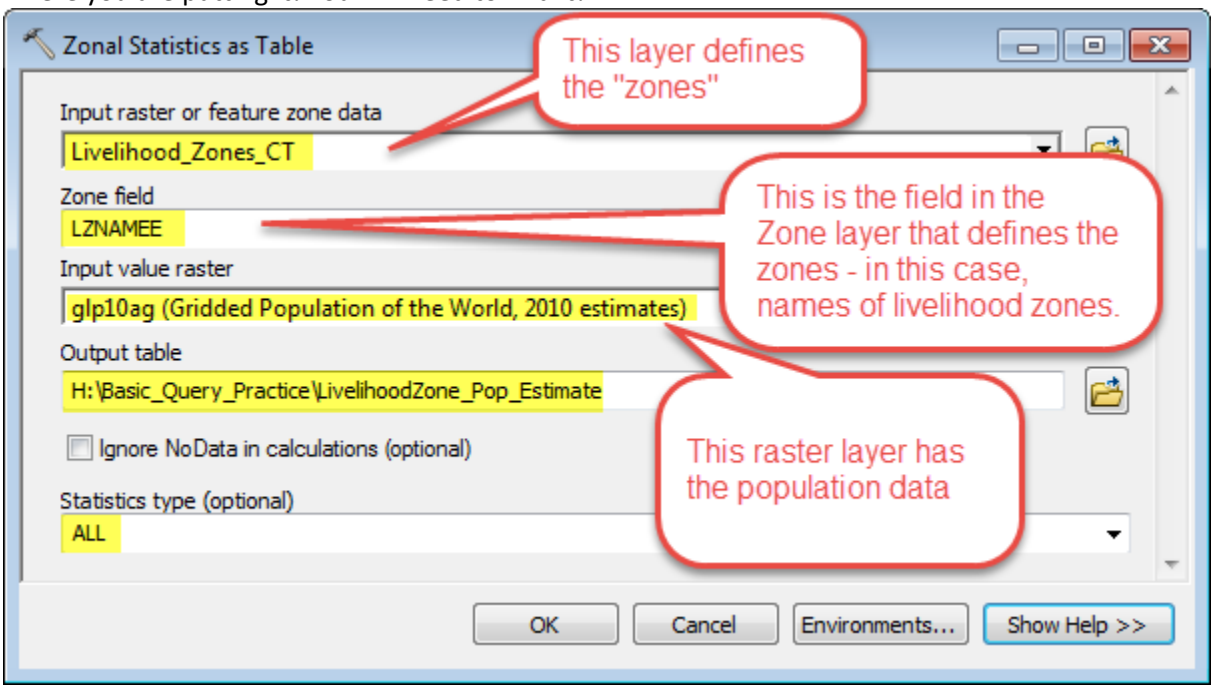
The *Zonal Statistics* function is part of an extension called Spatial Analyst – this extension allows us to use raster data in many different kinds of analysis. To use the *Spatial Analyst* functions, you first have to enable it. To do this:

1. Click on **Customize → Extensions** in the menu bar.
2. Check mark *Spatial Analyst* if not already checked.
3. To run the **Zonal Statistics** function, open *ArcToolbox*

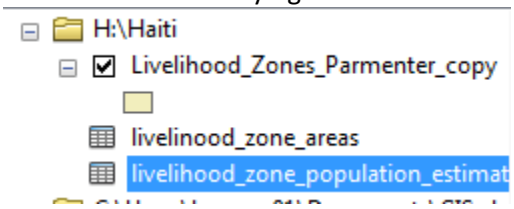
4. Go to **Spatial Analyst Tools → Zonal → Zonal Statistics as Table.**

   Note: If you receive an error message saying "Unable to Execute the Selected Tool-There is no Spatial Analyst License Currently Available or Enabled," go to **Customize → Extensions** on the menu and check Spatial Analyst!

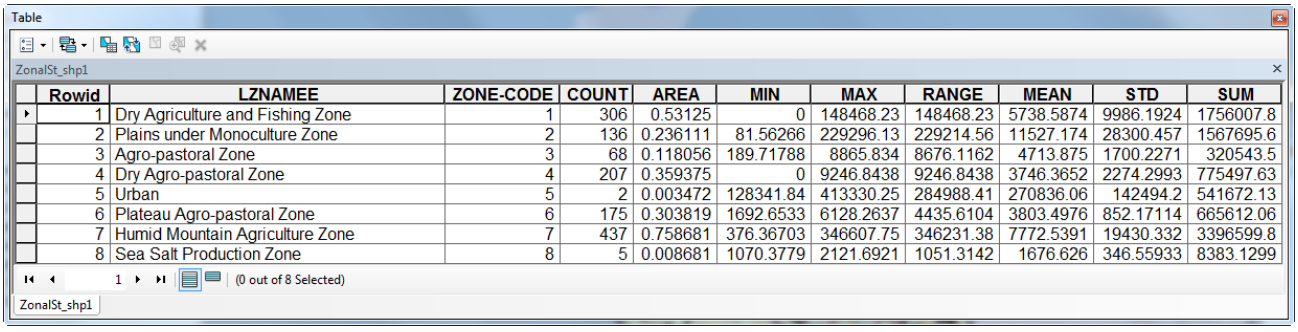
5. Fill out the dialog box as you see here, making sure that you *remember* what you are calling the output table and where you are putting it! You will need to find it!



6. Click OK – you have created a table in .dbf (dBase) format that has population estimated for each livelihood zone from the underlying GPW raster. It will be added to your *Table of Contents.*



7. Open it up to see the table. You'll see several statistics, but the SUM will tell you the estimated population of each zone based on the underlying raster population data set.



Table

ZonalSt_shp1

Rowid	LZNAMEE	ZONE-CODE	COUNT	AREA	MIN	MAX	RANGE	MEAN	STD	SUM
1	Dry Agriculture and Fishing Zone	1	306	0.53125	0	148468.23	148468.23	5738.5874	9986.1924	1756007.8
2	Plains under Monoculture Zone	2	136	0.236111	81.56266	229296.13	229214.56	11527.174	28300.457	1567695.6
3	Agro-pastoral Zone	3	68	0.118056	189.71788	8865.834	8676.1162	4713.875	1700.2271	320543.5
4	Dry Agro-pastoral Zone	4	207	0.359375	0	9246.8438	9246.8438	3746.3652	2274.2993	775497.63
5	Urban	5	2	0.003472	128341.84	413330.25	284988.41	270836.06	142494.2	541672.13
6	Plateau Agro-pastoral Zone	6	175	0.303819	1692.6533	6128.2637	4435.6104	3803.4976	852.17114	665612.06
7	Humid Mountain Agriculture Zone	7	437	0.758681	376.36703	346607.75	346231.38	7772.5391	19430.332	3396599.8
8	Sea Salt Production Zone	8	5	0.008681	1070.3779	2121.6921	1051.3142	1676.626	346.55933	8383.1299

(0 out of 8 Selected)

ZonalSt_shp1

8. Use the **Statistics** function on the *Sum* column to see what the total population adds up to.

How does the total population compare with what you calculated earlier for Haiti? Can you find the current population of Haiti on the web?

Are there other ways you might estimate the population for each livelihood zone using the Admin3 population layer? (We'll talk later in the semester about some other options.)

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 analysis 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** 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 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 relationships *between* layers – some of the ways you can look at these relationships include the following – we will be learning many others over the semester:
    - Selecting features based on their spatial relationship to other features (e.g., inside another feature or within a specified distance other features – **Select by location**)
    - Passing information from one layer to another layer (using ***Spatial Join*** for vector data or ***Zonal Statistics*** for underlying raster data)