GEOG 6304

HOMEWORK EXERCISE 2: Data Management in ArcGIS

ArcCatalog

ArcCatalog is the program used to organize and manage geospatial data in ArcGIS. ArcCatalog helps GIS users by providing an integrated and unified view of all the data files, databases, and ArcGIS documents available to ArcGIS users. Unlike other data (a photo or Word document), geographic datasets often consist of a set of files, rather than a single file. When listed in Windows Explorer, the datasets appear as a list of system folders and files, whereas ArcCatalog displays and manages the datasets as single entities. This program allows you to connect to your data source locations, browse through your workspaces, examine or explore the data, manage data, tables and metadata, and search for data and maps[1].

The ArcCatalog application provides the necessary tools, which allow you to:

- *Browse and find geographic information*
- *Record, view, and manage metadata (data about your geospatial data)*
- *Define, export, and import geodatabase data models and datasets*
- *Search for and discover GIS data on local networks and the Web*
- *Create and manage the schemas of geodatabases*

Geodatabases

Spatial data comes in many different formats, all of which can be managed in ArcCatalog, and added to ArcMap as independent layers. These formats include: coverages, shapefiles, AutoCAD files, and geodatabases. Geodatabases are the most recent addition to the repertoire of spatial data formats.

Geodatabases organize data into feature classes. A feature class (much like its predecessor the shapefile) is a group of points, lines or polygons, which represent a group of like geographic objects, such as counties, cities, rivers or roads.

Structurally; an ArcGIS geodatabase is a collection of geographic datasets held in a common file system folder, a Microsoft Access database, or a multiuser relational database (such as Oracle, Microsoft SQL Server, or IBM DB2). The geodatabase can scale from small, single-user databases built on files up to larger workgroup, department, and enterprise geodatabases accessed by many users[2].

The geodatabase contains the following dataset types:

- *Feature Classes*
- *Feature Datasets*
- *Tables*
- *Raster datasets*

[1] *ArcGIS Desktop Help – An Overview of ArcCatalog*
[2] *ArcGIS Desktop Help – An Overview of the Geodatabase.*

- *Raster Catalogs*

Feature Classes - A feature class is a collection of geographic features that are alike, that also share the same geometry type (such as point, line, or polygon), and the same attribute fields for a common area. Streets, well points, parcels, soil types, and census tracts are examples of feature classes. Related feature classes are often grouped together in a feature dataset[3].

Feature Dataset - A feature dataset is a collection of related feature classes that share a common coordinate system. Feature datasets are used to spatially or thematically integrate related feature classes. A Transportation feature dataset may contain a feature class for Highways, another for secondary roads, a third for unpaved or seasonal roads, and finally a point file for airports. Their primary purpose is for organizing related feature classes into a common dataset for building a topology, a network dataset, a terrain dataset, or a geometric network[4].

Tables - The attributes and properties of geographic objects are stored and managed in tables. Tabular information is the basis of geographic features, allowing you to visualize, query, and analyze your data. Tables in ArcGIS contain rows; all rows in a table have the same columns; each column stores a specific data type, such as a number, date, or character[5].

Raster Datasets - Raster datasets are individual rasters comprised of a single matrix of rows and columns. A mosaic of imagery would be considered a raster dataset.

Raster Catalog - A raster catalog is a collection of several raster datasets.

[3] ArcGIS Desktop Help – An Overview of creating feature classes
[4] ArcGIS Desktop Help – An Overview of working with feature datasets
[5] ArcGIS Desktop Help – An Overview of Tables

Scenario: *You've been tasked with the creation of a personal geodatabase to be used by a number of GIS analysts. These analysts are working with a NGO to analyse and map some of the data available for Malawi.*

PART 1 – SEARCHING FOR AND DOWNLOADING DATA

1. For this project you have been provide with a number of data layers. They are located in your Basedata folder (…\Geog6304…\YOURNAME\Homework02\Malawi\Basedata). The remaining datasets needed are described in this document and you will have to download these data from two (2) different online sources:
    a. DIVA- http://www.diva-gis.org/
    b. MASDAP - http://www.masdap.mw/

2. From DIVA download the following data layers for Malawi:
    a. Roads
    b. Railroads
    c. Elevation
    d. Gazetteer

3. From MASDAP download the following data layers as Zipped Shapefiles:
    a. Surface Water Bodies SRTM
    b. Contours
    c. Airports
    d. Inland Water Line (aka Rivers)
    e. Population over Enumeration Areas NSO

4. These data are in a compressed or zipped format. Before you can use these data they must be uncompressed or unzipped. Unzip each of the downloaded layers to their own folders using a utility like 7-Zip, PKZIP or some other file (de)compression utility software. It's easiest to do this in the SAL. I also recommend that you extract/unzip each dataset into its own folder. This will help keep the data organized in your personal class folder.
   HINT: See help video on Blackboard

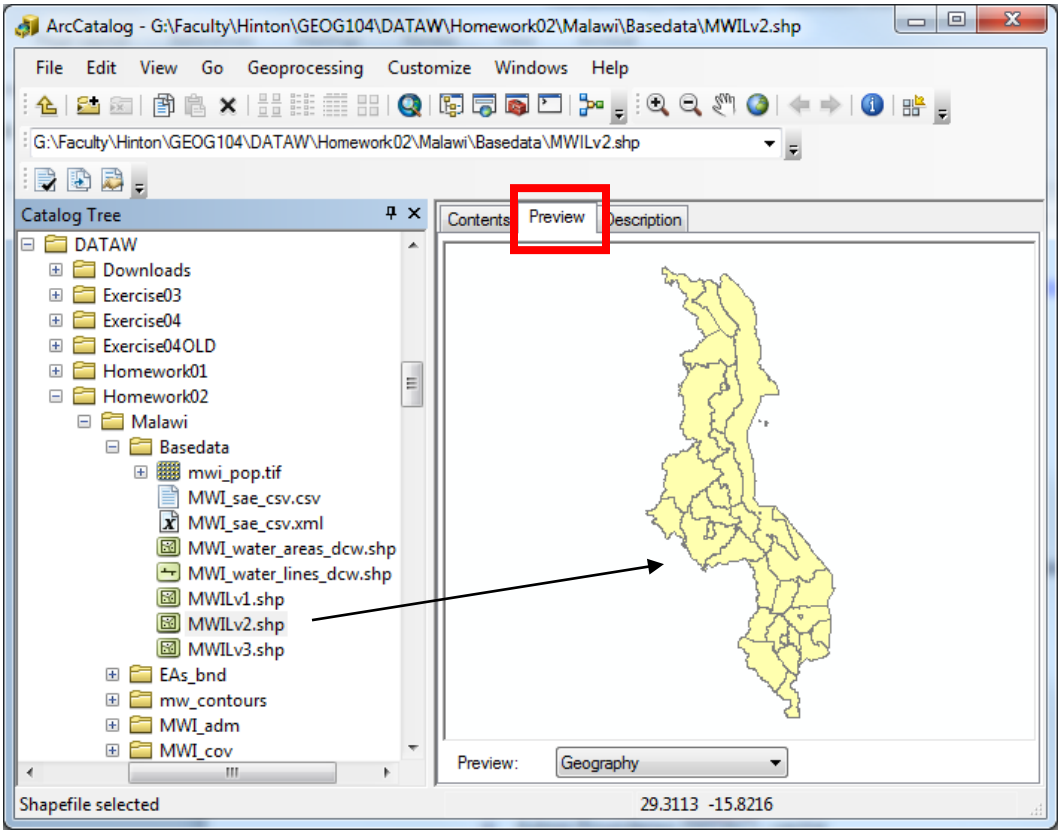
PART 2 - BUILDING A PERSONAL GEODATABASE

5. Start ArcCatalog by clicking the ArcCatalog icon.

6. If you haven't already, make a connection to your Personal folder in the class folder (…\GEOG6304…\YOURNAME). From there you should have access to the **Homework02** folder. (Use the same connect to folder button that you use within the regular ArcMap interface)

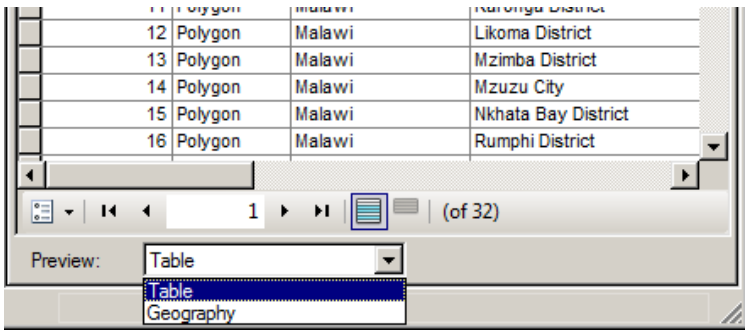
7. Once you have created the connection, expand the **Homework02** folder, and the **Malawi\Basedata** folder beneath it.

8. You will see 6 GIS files for Malawi. There are 5 vector files and 1 raster file and 1 comma delimited file, .CSV.



9. By selecting each one of these files separately, and clicking the Preview tab above the map window, you will get a snapshot of your data in the preview window to the right. Select each file in turn, and watch the preview change in the map window. Also, click on the Description tab to learn where this feature layer originally came from (NOTE: These data are missing most of their respective metadata).

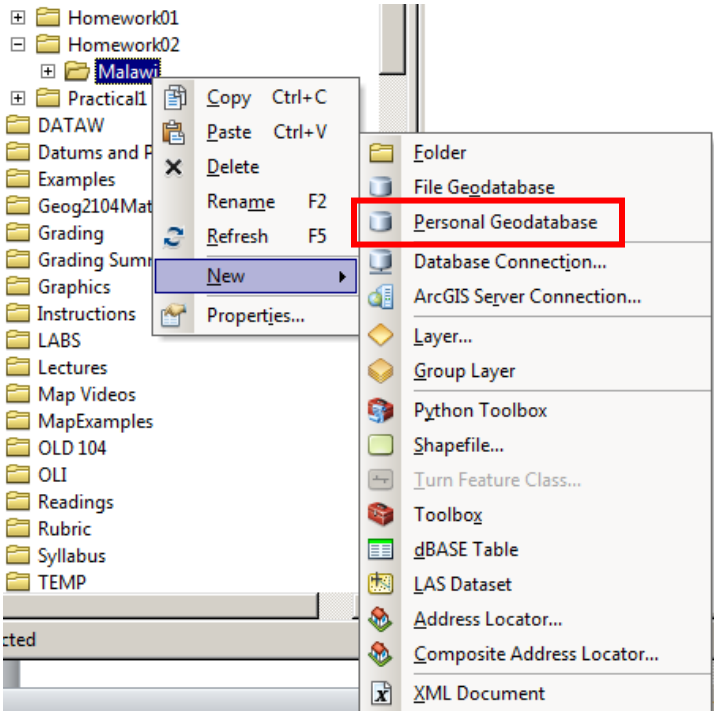
10. If you wish to preview the attributes, rather than the visual component of the file, simply go to the Preview drop down menu at the bottom left hand corner of the screen, and switch to Table.



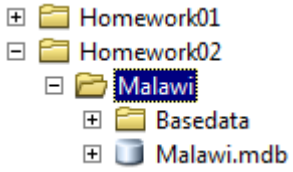
Familiarize yourself with the data, before moving onto the next step of creating a geodatabase.

CREATE A NEW PERSONAL GEODATABASE

11. Right-click on the folder **Malawi** > New > Personal Geodatabase. A new personal geodatabase is then created in the **Malawi** folder.

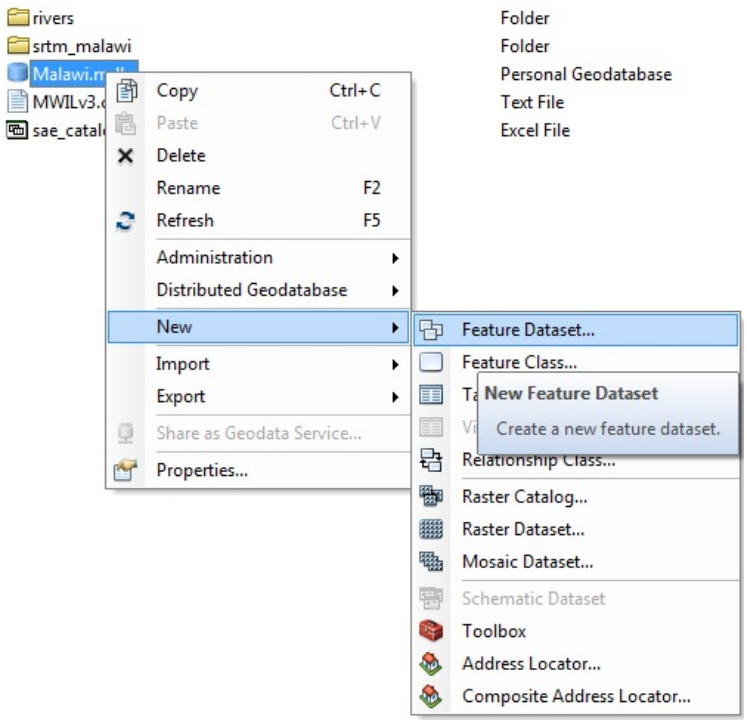


12. Call the geodatabase Malawi and press Enter on the keyboard. The geodatabase should now be named *Malawi.mdb* (make sure the file extension is visible).



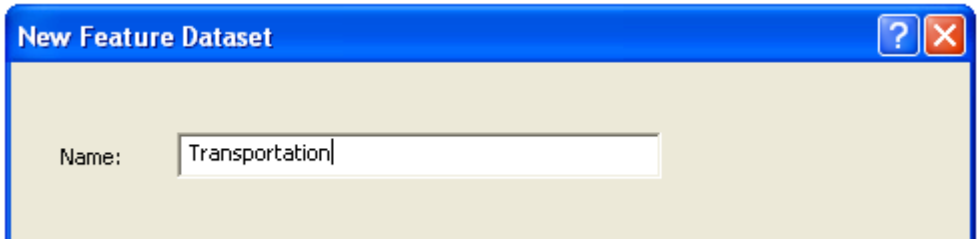
CREATE A NEW FEATURE DATASET

13. Click on the Malawi folder, so that the *Malawi.mdb* database is visible in the main ArcCatalog display window. Right-click on *Malawi.mdb*, point to New > Feature Dataset.

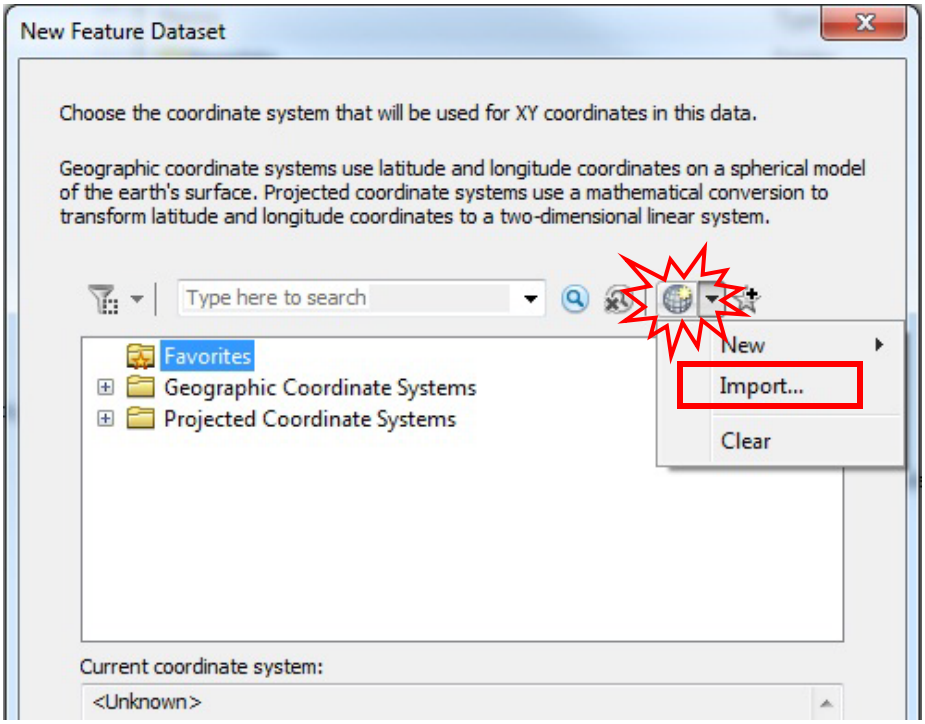


14. In the New Feature Dataset wizard, you are prompted for a name. Call the first feature dataset Transportation.

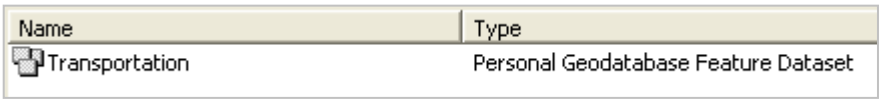
Feature datasets are a collection of feature classes sharing the same spatial reference, and usually have a similar theme.



15. Click Next.
16. In the next window you will be asked to choose a map projection for the data that will reside in the Transportation feature dataset. The easiest solution in this case is to import the map projection native to one of the files that will be stored in the Transportation dataset.
17. Click on the Add Coordinate System Button and Click on Import. (See image below)
18. From the folder where you saved the roads data (MWI_rds), click on MWI_roads.shp, and click Add.



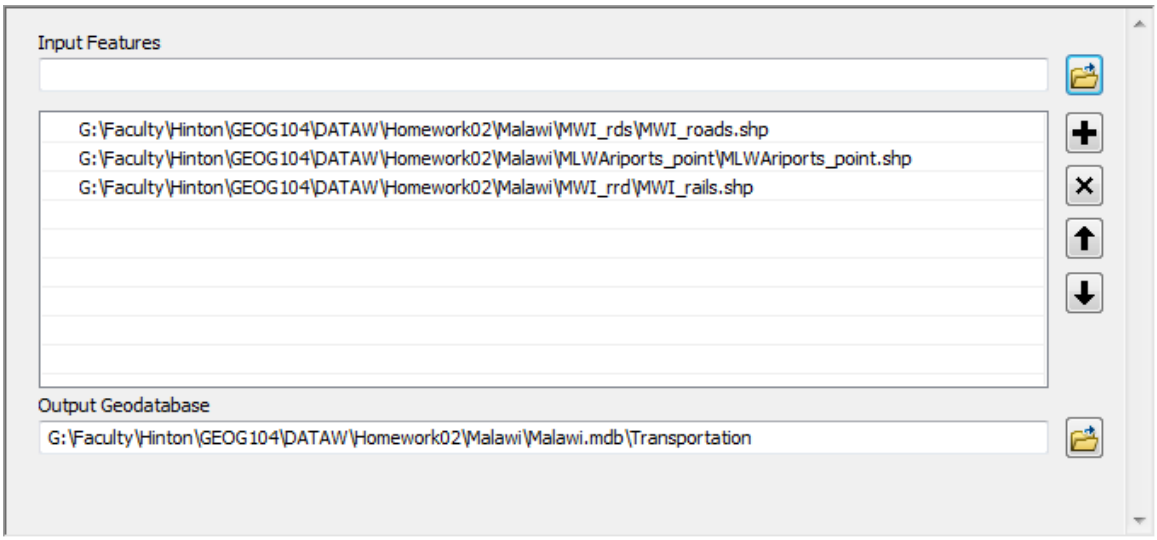
19. This assigns the map projection of GCS_WGS_1984 to the feature dataset.
20. Accept the default options for the next several screens, by clicking Next.
21. The new Feature dataset; Transportation will appear in the ArcCatalog window.



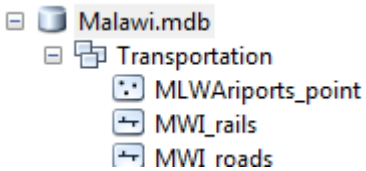
22. Right click on Transportation, navigate to Import > Feature Class (multiple)

A Feature Class is a collection of geographic features with the same geometry type (such as the vector based data types: point, line, or polygon), the same attributes, and the same spatial reference. Feature classes allow homogeneous features to be grouped into a single unit for data storage purposes. The feature class storage method that you have encountered up to now has been the shapefile (.shp)

23. Navigate to the appropriate folders and select the shapefiles: MWI_roads.shp, MLW_Airports.shp and MWI_rails.shp



24. Click OK to run the wizard, and then close the dialog box when it is completed.
25. You should now have the following file structure in your geodatabase.



26. The remaining files you downloaded can be grouped according to function, such as *Admin, Hydro, and Topography*. Therefore, you will organize these within the geodatabase as individual *Feature datasets.*

27. Right-click on Malawi.mdb, point to New > Feature Dataset. Call it *Admin*.
28. To assign a coordinate system to the new Feature Dataset, click on the Import button (as noted in Step 18), and navigate to the file MWILv1 *(You are essentially bestowing the coordinate system from MWILv1 to the new Feature Dataset, Admin)*
29. Click Next repeatedly to finish.
30. The new Feature Dataset *Admin* will appear in the ArcCatalog window.
31. To add Feature classes to this new Feature Dataset, right click on *Admin*, navigate to Import > Feature Class (multiple)
32. In the Feature Class to Geodatabase dialog, you can navigate to your **Basedata** folder, and add the following files simultaneously:
    a. *MWILv1.shp*
    b. *MWILv2.shp*
    c. *MWILv3.shp*

Alternately, you can navigate to these shapefiles in the ArcCatalog tree and simply drag them to the Import Features window.

Click OK when you have selected these files. This process can take a minute or two, so be patient.

33. Using the steps above, create *Feature Datasets* for the following categories, according to the list below (Note: Categories a. and d. have already been created). Use the Import > Feature Class (multiple) even if there is only one layer involved. The 'multiple' interface is easier to populate than the individual.

34. Once expanded in ArcCatalog, the file structure should look like the example graphic. Remember to apply a coordinate system to each new *feature dataset*, using one of the files it will eventually contain.

    a. Admin

    - *eas_bnd.shp*
    - *MWILv1.shp*
    - *MWILv2.shp*
    - *MWILv3.shp*

    b. Hydro

    - *MWI_water_areas_dcw.shp*
    - *MWI_water_lines_dcw.shp*
    - *Inland_water_line.shp*
    - *srtm_malawi.shp*

    c. Topography

    - *mw_contours.shp*

    d. Transportation

    - *MWI_roads.shp*
    - *MWI_railss.shp*
    - *mlw_airports.shp*

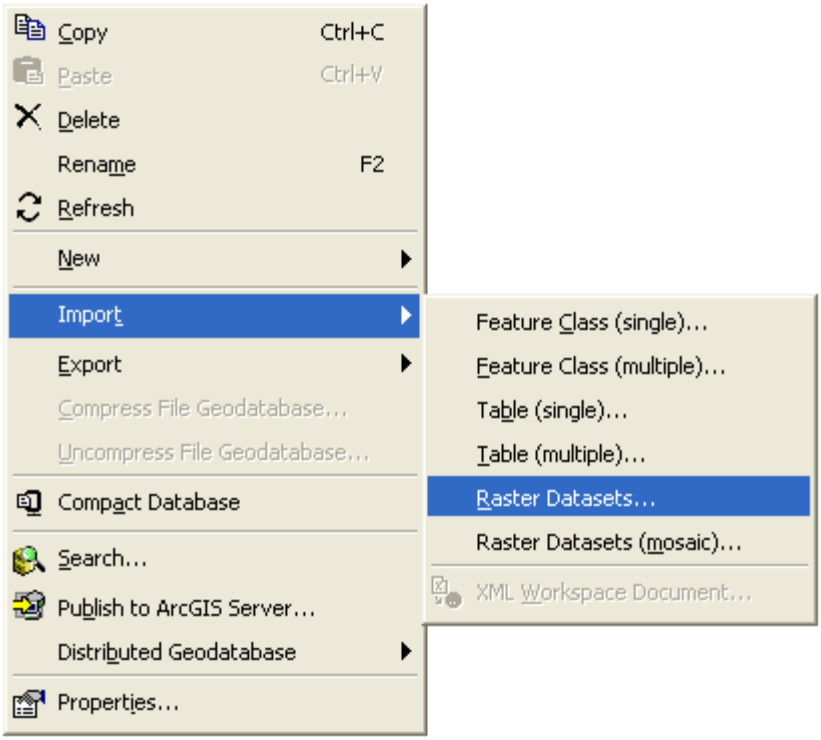
ADD RASTER DATA TO PGDB

35. Next, let's import the raster files from your folders to your geodatabase. Remember that one was provided in the **Basedata** folder.

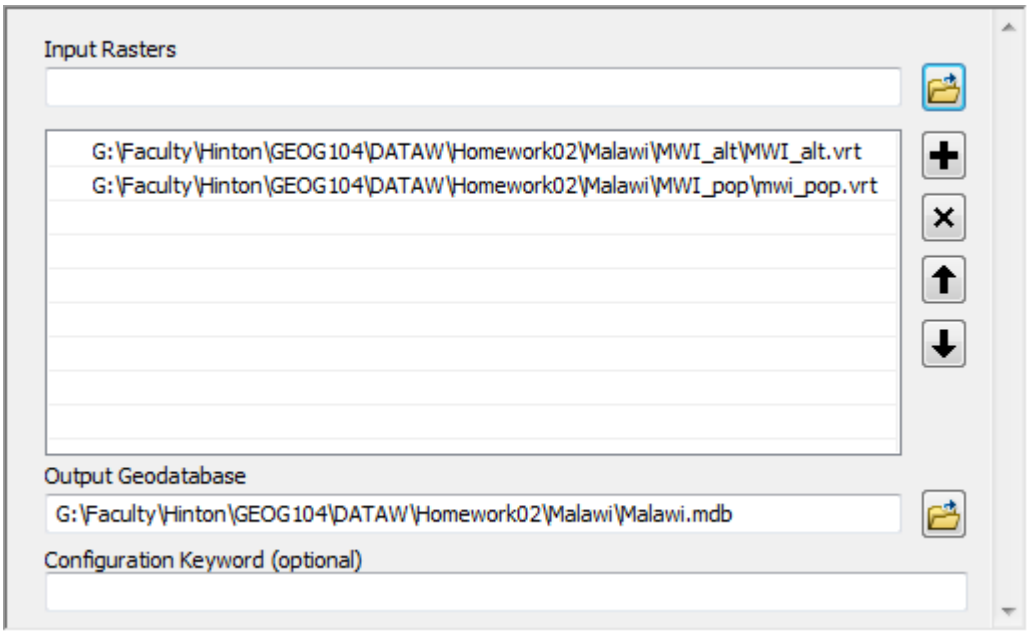
Surfaces that are in raster format can be stored in a personal geodatabase, file geodatabase, or multi-user geodatabase. There are two methods for storing rasters in a geodatabase. Raster datasets are individual rasters that are a single matrix of rows and columns. The other method for storage of rasters in a geodatabase (personal, file or SDE) is as a raster catalog. A raster

catalog is a collection of several raster datasets, that when added to ArcMap, will load all component rasters simultaneously.

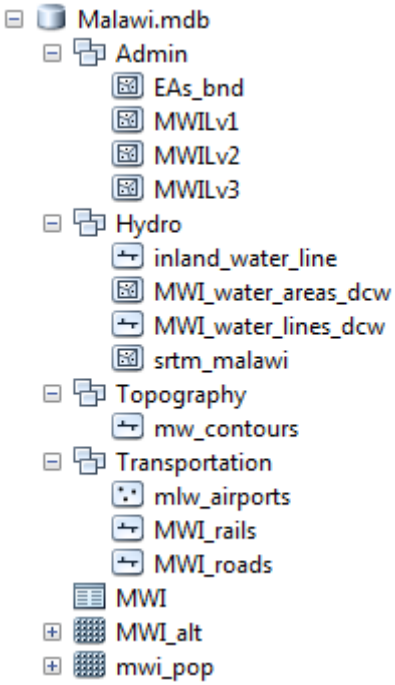
36. Right-click on Malawi.mdb, point to Import > Raster Dataset.



37. Drag the raster layers mwi_pop (population) and MWI_alt (elevation) to the Import Raster Dataset wizard.



38. The Spatial Reference Properties of the existing layers will be imported into ArcMap
39. Click OK to Import the two files.
40. Finally, you will import the gazetteer data table into the Geodatabase.
41. Once again right-click on Malawi.mdb, point to Import > Table (multiple). Drag the table: MWI to the Import table wizard, and Click OK.
42. At this stage your geodatabase, when expanded, should look like the following graphic.



43. Browse through your geodatabase and take a look at the various categories of data you have compiled and organized.

PART 3 – JOIN TABLE AND CREATE LAYER FROM X Y COORDINATES

*The admin boundary shape files in the **Basedata** folder are from SEDAC. However, they do not contain the required information on the small area estimates of poverty and inequality for Malawi. This information is found in a separate table in the **Basedata** folder, MWI_sae_csv.*

44. Launch ArcMap and add the MWILv2 feature class from your Malawi personal geodatabase and the MWI_sae_csv table from the **Basedata** folder.

45. Open the attribute table of the admin layer (MWILv2) and the MWI_sae_csv table. Look through each table and the fields contained within.

    What fields could be used to join the tables?

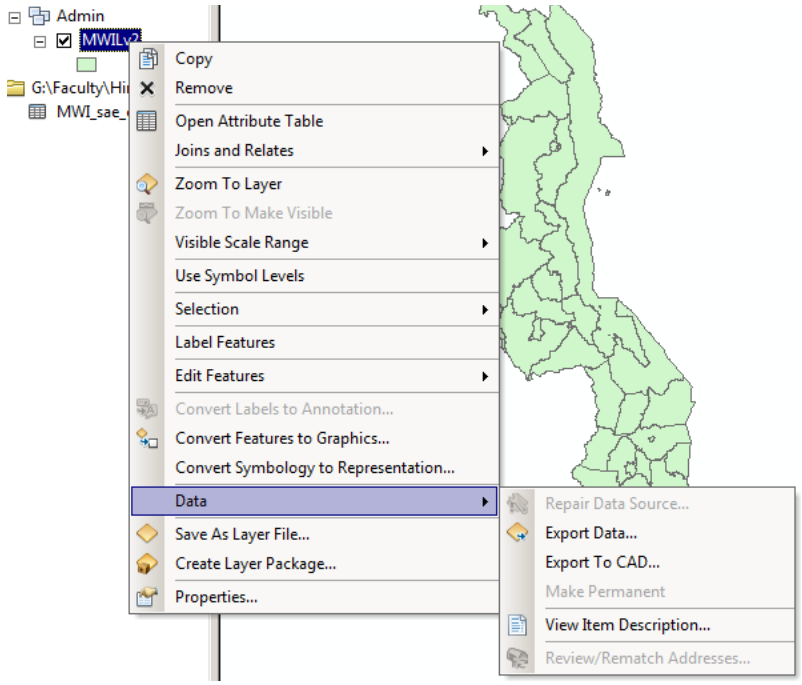
    Remember you join tables based on a primary field which contains unique values for each record. In this instance use the ADMUnitID field in both files to complete the join.

46. To start the process right click on the MWILv2 feature class (shapefile) and select 'Joins and Relates'. Remember to 'Keep all records'.

47. Once the join is complete, open the attribute table for MWILv2 again to see that the extra fields have been added with their respective values.

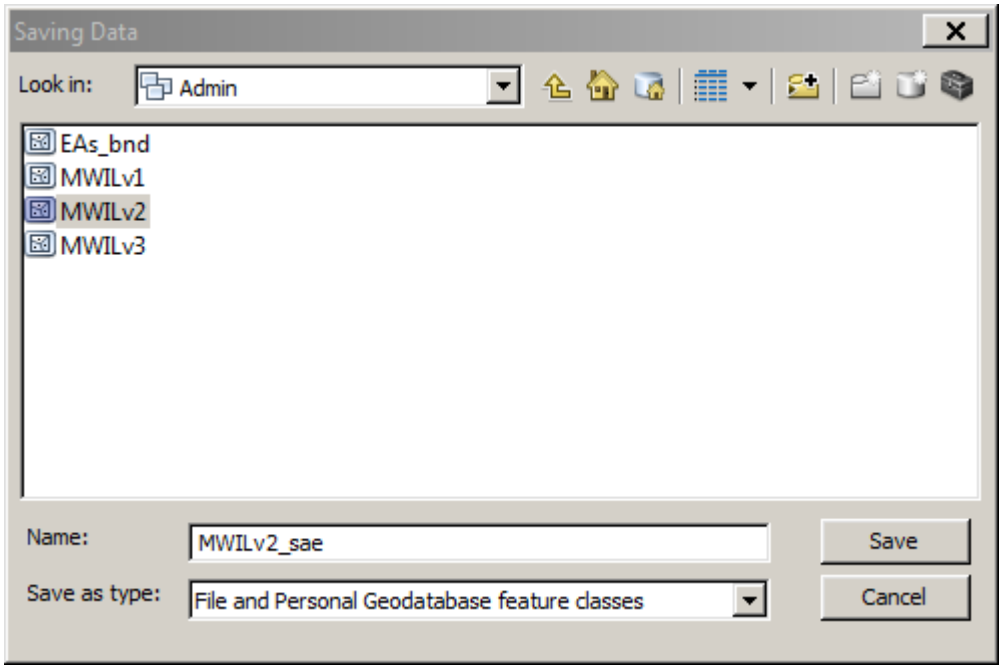
48. This table join is not permanent. To make it so we must export this file as a new feature class or shapefile. So that the analysts will have all of the data in one place let's export this file to the **Admin** Feature Dataset in our **Malawi** personal geodatabase

49. Right click on MWILv2 > Data > Export Data



50. In the Export Data window click the 'Browse' folder icon on the right side of the Output Feature class: input area.

51. Save the exported file with the following credentials:
    a. Name: MWILv2_sae
    b. Save as Type: File and Personal Geodatabase feature classes



52. Click Yes when prompted to add the exported layer to the map. The new layer is added to the top of the TOC. Open the attribute table and see that the join you created is now a permanent part of the MWILv2_sae attribute table.

CREATE LAYER FROM X Y DATA

Earlier you downloaded a gazetteer for Malawi from the DIVA website. This non-spatial table contains a list of all the cities and towns in Malawi with their locations given as latitude and longitude. Since the coordinate locations are known we can plot these points and create a new layer for our personal geodatabase.

53. Add the gazetteer table to ArcMap. File > Add Data > MWI.dbf

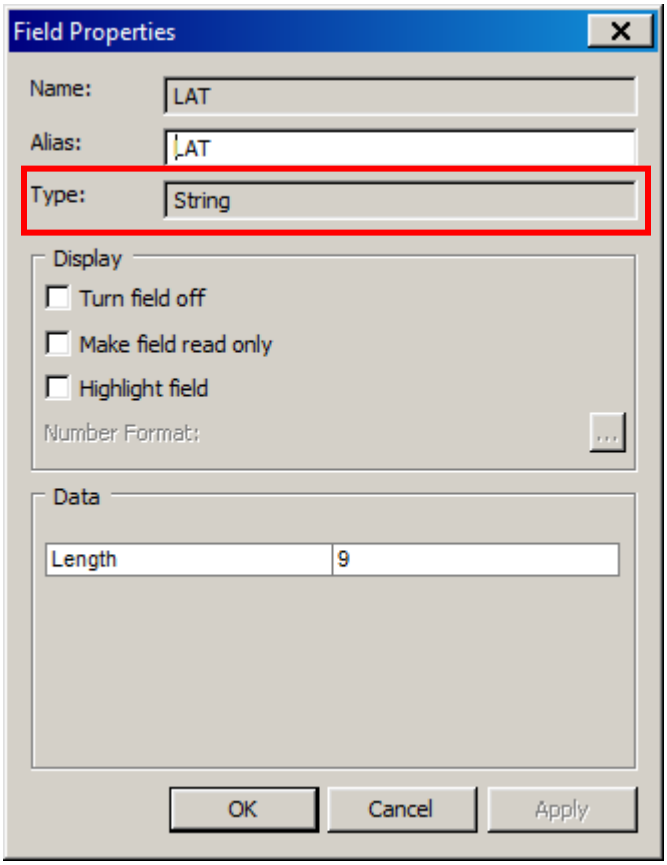
54. Open the table. Right click MWI > Open

55. Of the fields in this table you'll see a LAT and LONG field each holding the respective coordinate for each city. These look like good fields to map each city locations, so let's try it.

56. Right click on MWI.dbf > Display X Y Data

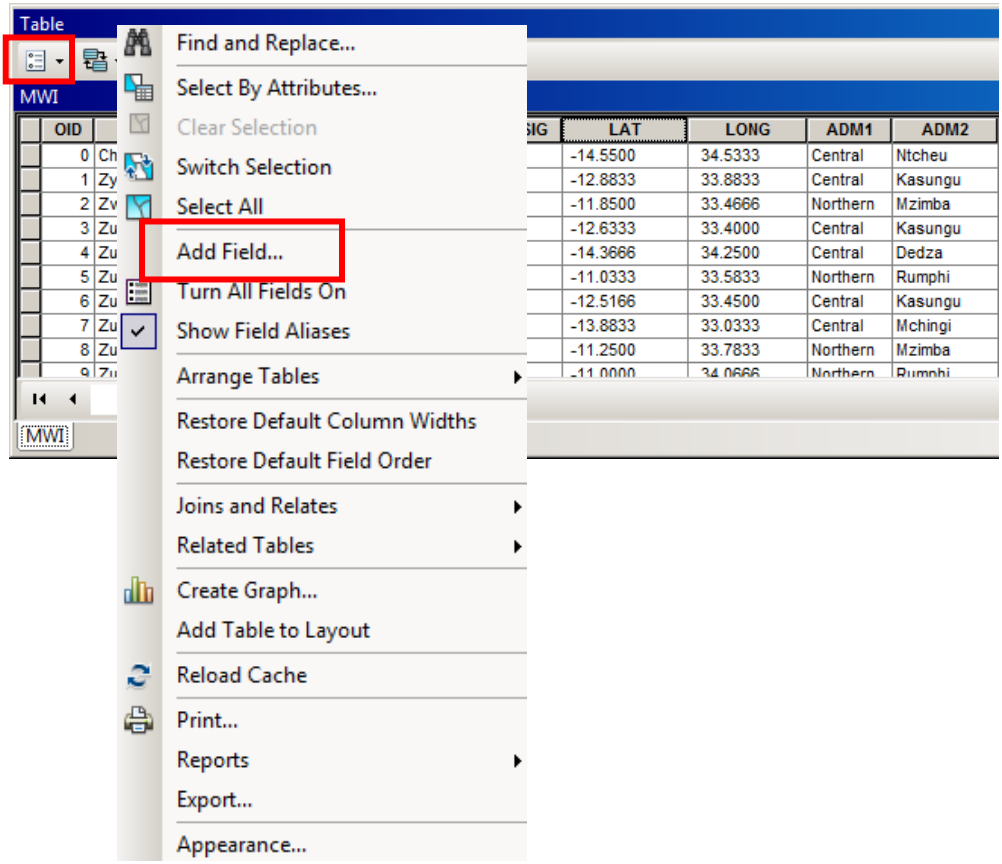
57. In the resulting window you should see options for X Field and Y Field, but there are no fields to choose from. *Why do you think this is so?* This tool is expecting input from numerical fields. In our table even though the LAT and LONG fields contain numbers, the fields are not defined as numerical fields in the table.

58. To see how these fields are defined re-open the table. Right click on LAT > Properties. You'll notice that the field type is listed as String. In GIS/database terms this means it's a text field and not numeric. Close the Fields Properties window.



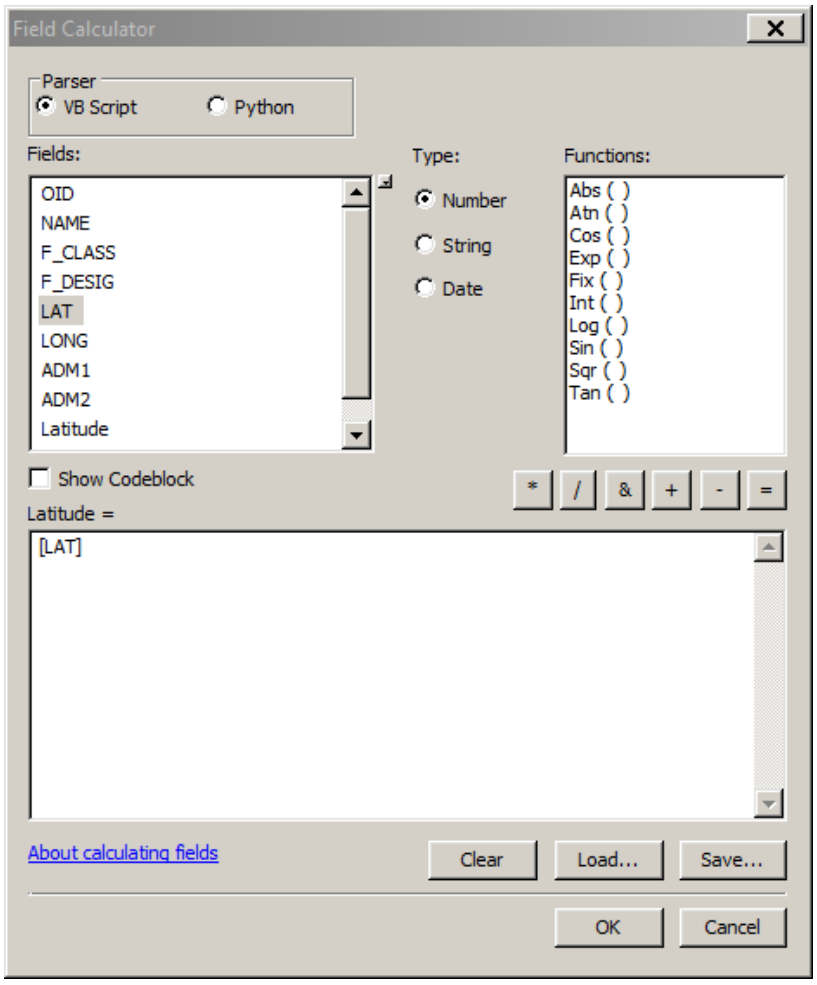
59. To create numeric fields from a text field requires two (2) steps:
    a. Create new field
    b. Calculate values into the new field

60. Add new field for LAT values.
    a. Click Table options dropdown (upper right of attribute window)
    b. Click Add Field
    c. Add field Name: **Latitude**
    d. Type: **Float**
    e. Keep default values for Precision and Scale.



61. Follow the same steps above to add a field called **Longitude**.

62. Next we need to populate the new fields based on the values obtained from the original LAT and LONG fields. Because we defined the new fields as type float, which is one of the decimal formats, ArcGIS will automatically convert the values from text to numbers.
    a. Right click on Latitude > Field Calculator
    b. Click Yes on the warning window that pops up.
    c. Double click on LAT in the list of Fields. You'll see [LAT] appears in the white space below 'Latitude = '.
    d. Click OK.
    e. In a short time, all of the records in Latitude field should be calculated.
    f. Repeat these steps for **Longitude**.



63. With both the Latitude and Longitude fields appropriately populated with their respective geographic coordinates, we can try the Display X Y tool again
    a. Right click MWI > Display X Y Data
    b. X Field: Longitude
    c. Y Field: Latitude
    d. Click OK

64. You should see a new layer **MWI Events** appear in the Table of Contents (TOC) and point symbols appear in the Map Window.

65. Much like a table join, this layer is not permanent. To make a permanent layer from these points, you must export it as a new layer.

66. To export the data, Right click on MWI Events > Data > Export Data

67. Save the file in the **Admin** feature dataset of your **Malawi** PGDB.

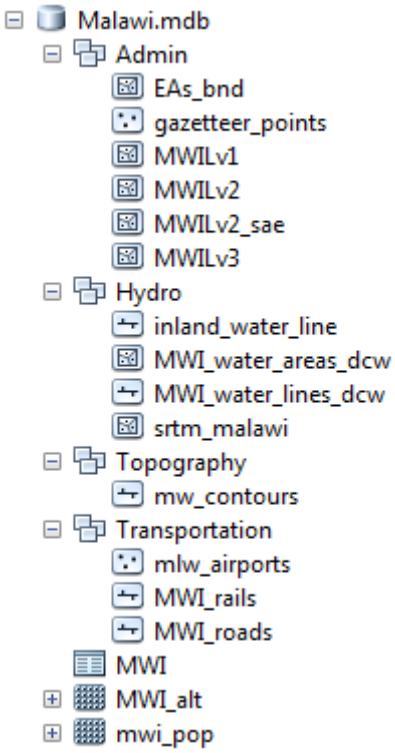
68. In the Export Data window click the 'Browse' folder icon on the right side of the Output Feature class input area.

69. Save the exported file with the following credentials:
    a. Name: gazetteer_points
    b. Save as Type: File and Personal Geodatabase feature classes

70. Click yes to add the new layer to the map.

71. Open the attribute table of gazetteer_points. You'll see that all of the attributes from the MWI.dbf file are part of the feature class attribute table.

72. Your final geodatabase should look like the following:



ArcGIS Data Types

Data type	Storable range	Size (Bytes)	Applications
Short integer	-32,768 to 32,767	2	Numeric values without fractional values within specific range; coded values
Long integer	-2,147,483,648 to 2,147,483,647	4	Numeric values without fractional values within specific range
Float (single-precision floating-point number)	approximately -3.4E38 to 1.2E38	4	Numeric values with fractional values within specific range
Double (double-precision floating-point number)	approximately -2.2E308 to 1.8E308	8	Numeric values with fractional values within specific range

ArcGIS data types table

PART 4 - DATA SELECTION AND EXPORT USING QUERIES

You have been asked to create a mini-geodatabase for Kasungu District. You have been asked to create feature classes identifying the various features in this district including transportation features, town locations and hydrography, among others.

73. In ArcCatalog, within the Malawi folder (not the Geodatabase itself, but the folder you created it in) create a new personal Geodatabase called Kasungu. (Remember how to do this? Just right click within the Malawi folder, point to New > Personal Geodatabase).

74. Open a new ArcMap session (Do not return to the ArcCatalog window)

75. Click on the Add Data button and from the newly created Malawi.mdb Geodatabase add: EAs_bnd, gazetteer_points, MWILv2_sae, MWI_water_areas_dcw, MWI_water_lines_dcw, MLW_Airports and MWI_roads.

76. You have been asked to sub select data for the District of Kasungu. First you need to isolate a boundary for the study area.

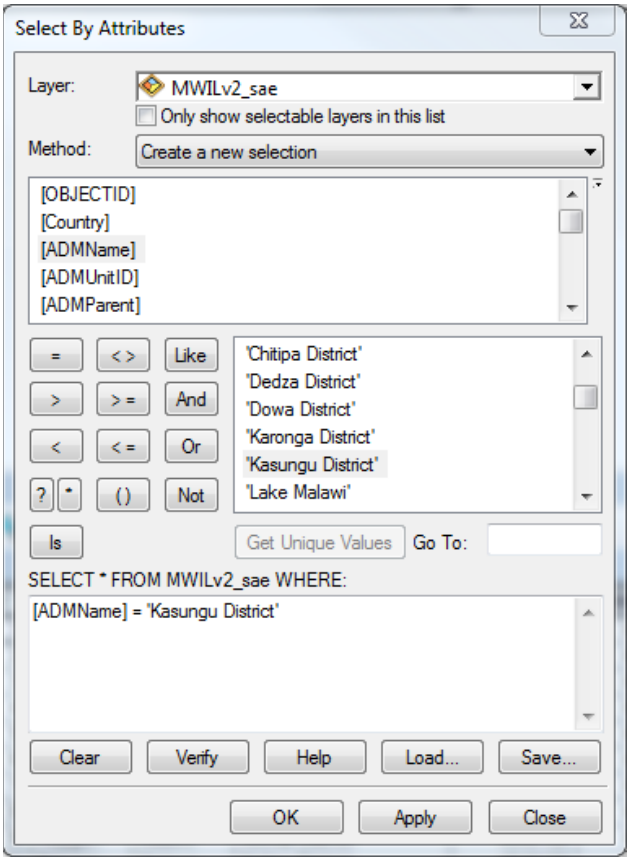
77. From the Selection drop down menu, choose the Select by Attribute option. In the Layer dropdown menu, choose the MWILv2_sae layer. By default the Method option should be Create a new selection, if not, choose that.

78. The scroll menu under the Method drop down, lists all the fields in the MWILv2_sae attribute table. The field you are interested in called ADMName. This indicates the relevant District name for each District in the layer.

79. To see a listing of all possible entries in this field, click on the Get Unique Values button. The blank window will populate with all relevant values. You will have to scroll down to see Kasungu District.

80. Using the field names, and query operators, build the following query:

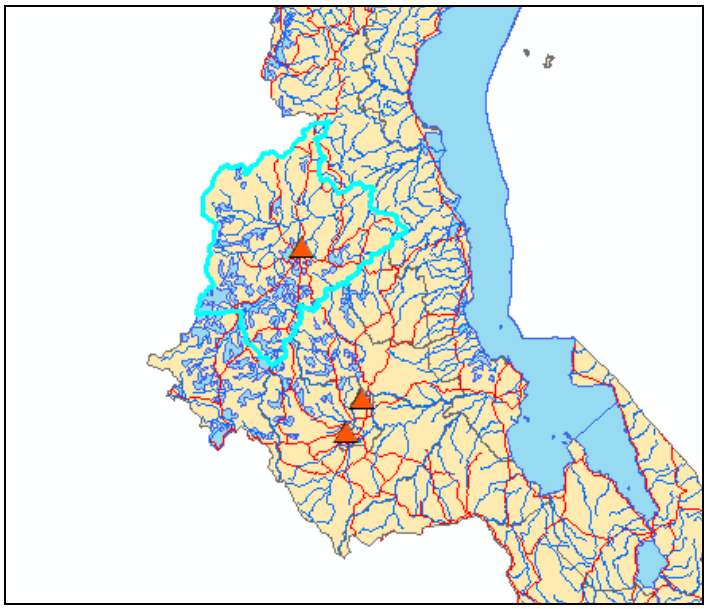
    ADMName = 'Kasungu District'



81. Your interface should look like the graphic to the right:

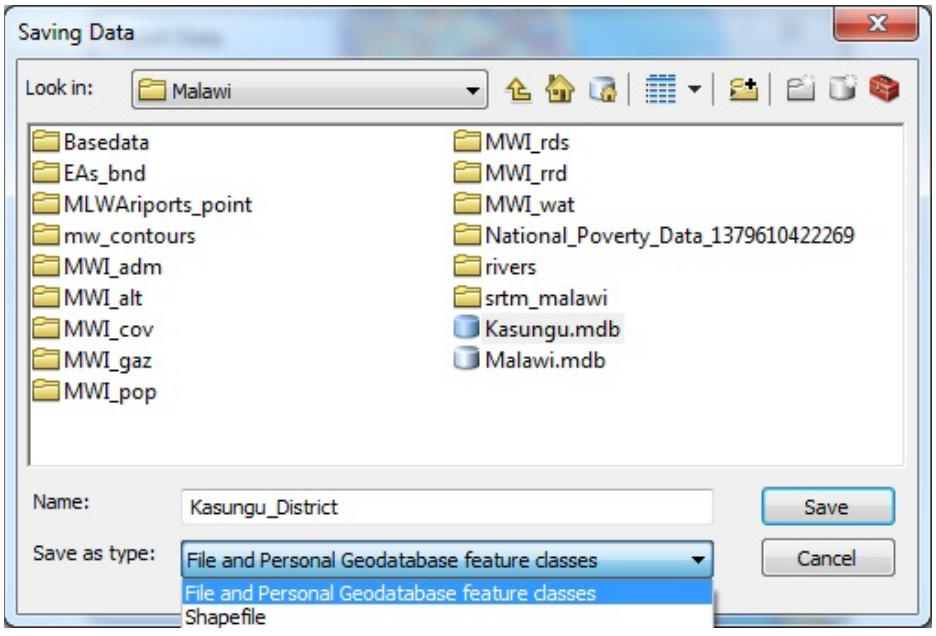
82. You may wish to check the syntax of the query before you finish, by clicking on the Verify button. Then, Click OK.

83. The Kasungu District should now be highlighted in your map



84. To export this selection as a new file, right-click on the MWILv2_sae layer, and scroll down to Data > Export Data.

85. Save the output to your new geodatabase Kasungu.mdb as Kasungu_District (You will have to change the Save as type to the File and Personal Geodatabase feature class option in order to see the Geodatabases in the Save Data wizard.



86. You should be prompted to add the new layer to your map automatically, click Yes.

87. Now that you have isolated the geographical extent of your study area, you may remove the original MWILv2_sae layer from the table of contents, by right clicking on it, and selecting Remove.

88. To zoom into your study area for a closer look, right click on the new Kasungu_District layer, and select the Zoom to Layer option

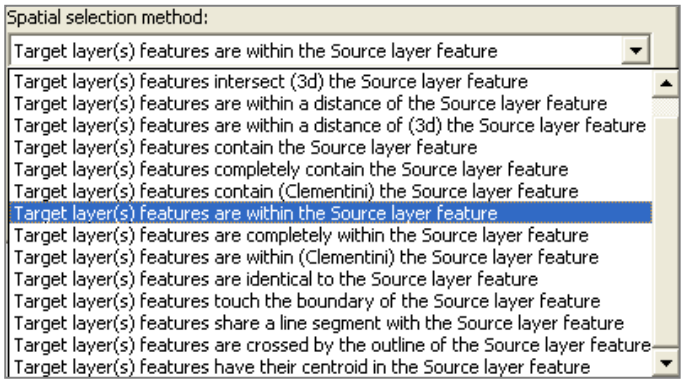
89. Next you will sub-select the Airport in this region.

90. From the main Selection drop down menu, choose the Select by Location option. In the 'Selection method:' dropdown menu, choose select features from. (This should already be the default). (See Graphic below)

91. In the next box, put a tick beside MLW_Airports. (See Graphic)

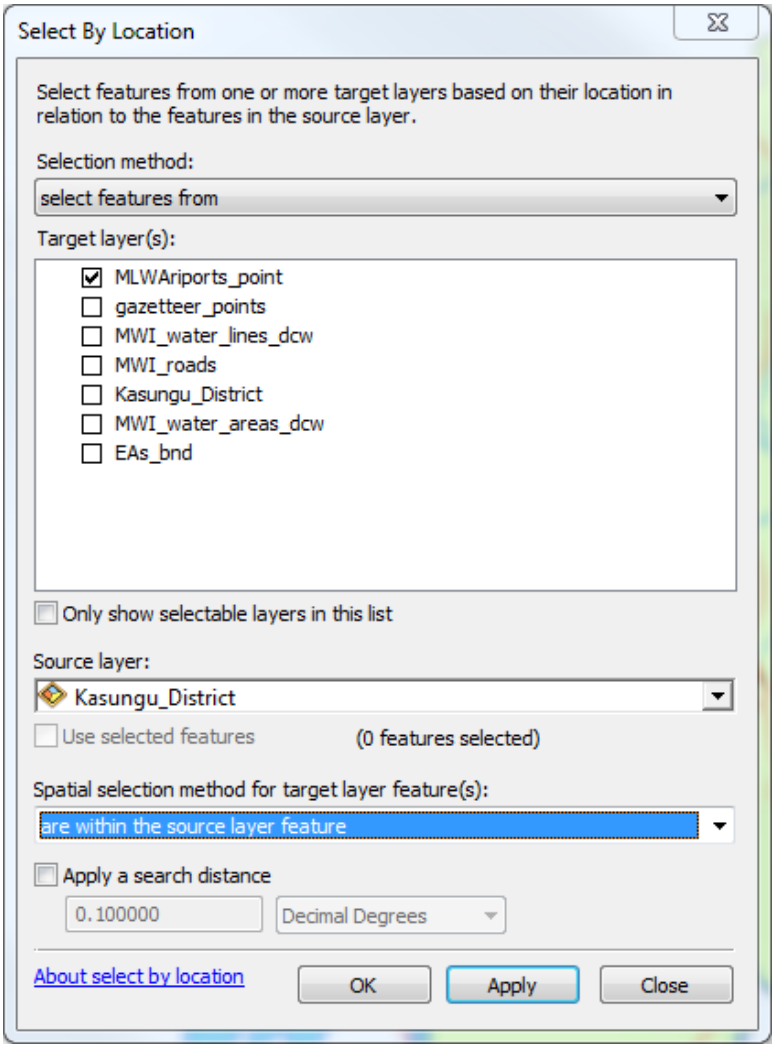
92. In the next drop down box you will identify the Source layer (or, the cookie-cutter layer as I like to call it). Choose the Kasungu_District layer. This instructs ArcMap to ***only*** select the Schools that are completely within the Kasungu_District layer.

93. The next drop down list is quite interesting. It shows all the potential spatial relationships between two layers, relationships you can use to extract data.



94. For points within polygons, several options are applicable. Such as; intersect, completely contain, are completely within, have their centroid in. Go with "are within".

95. Check to make sure your wizard matches the one below, then, click OK.



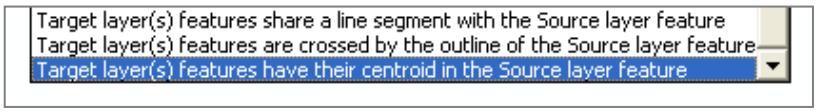
96. The airport within the Kasungu District should now be selected. (There is only 1 in this district.)

97. To export this airport right-click on the MLW_Airports layer, go to Data > Export Data, and save to your output to the new Kasungu.mdb Geodatabase as Kasungu_Airport.

98. To select the gazetteer points for the same region, repeat the previous Select By Location steps. Save the new file to Kasungu.mdb as Kasungu_Settlements

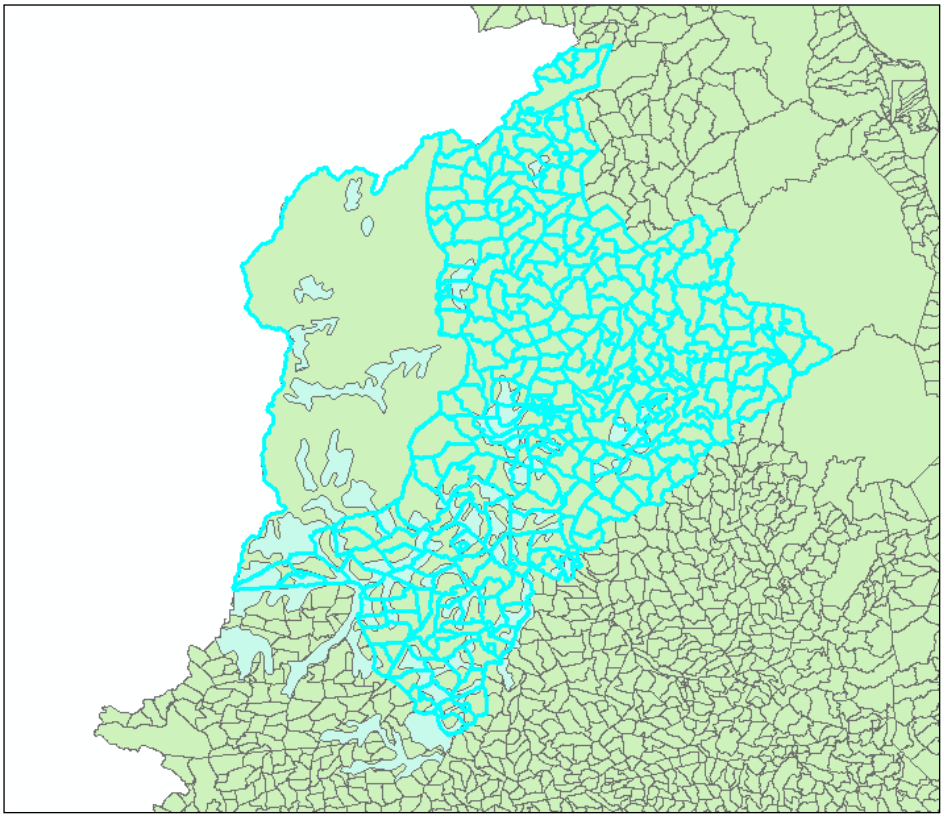
99. To select the rivers, lakes and roads within the District, repeat the previous Select By Location steps, but this time use *"intersect the source layer feature"* for the Spatial selection method. With each selection, export the selected features to the Kasungu.mbd calling the layers Kasungu_rivers, Kasungu_lakes and Kasungu_roads respectively.

100. Finally, from the main Selection dropdown listing choose Select by Location. This time you want to select features from EAs_bnd that *"have their centroid in"* Kasungu_District.



You may find it unusual to use the specification have their centroid in to select the EAs_bnd that lie within the District of Kampala. Surely Intersect or are contained should work also? Unfortunately this is not always the case. Often the boundaries for coterminous geographies do not match as neatly as they should. These discrepancies may seem microscopic to the mapmaker, but can force the query to fail, or not select all relevant elements. We use 'have their centroid in' as it will select all polygons who's middle falls within the parent polygon.

101. Your output should look like the following.



102. Finally, Export the output from the EAs_bnd layer to Kasungu.mdb as Kasungu_EAs.

IMPORTANT: The Selection options in the Select by Location Tool are not foolproof. Occasionally the geography of a specific unit might prevent it being selected using the specified options. This is especially true of geographies that consist of a collection of islands, or a mixture of island/mainland polygons. You should always check to see if the expected polygons have been selected. If a particular polygon has been omitted; hold down the Shift key, and use the manual select tool  *to add the missed polygon to the rest of the selection.*

GEOG 6304

Deliverables for HOMEWORK 2

GEODATABASES

Your deliverables from this lab will be the geodatabase:

1. Malawi.mdb
2. Kasungu.mdb

Your instructor will check the appropriate files in your Homework02 folder.

Looking for map deliverables? – No maps this week (Yeah!)

GEOG 6304

Learning Objectives

What you should know from Homework 2

There are KEY elements that you should remember from this lab:

- Create a **Personal Geodatabase**
    - Create **Feature dataset**, and specific a projections for that Feature dataset
    - **Import layers** into a Feature class
    - **Import Raster layers** into a Personal Geodatabase
- **Add Fields**
- **Calculate Fields**
- **Display XY data**
- **Select by Location**
- **Export** spatial data