# Geography 281 Map Making with GIS
# Project Four: Comparing Classification Methods

Thematic maps commonly deal with either of two kinds of data:

- **Qualitative Data** showing differences in kind or **type** (e.g., type of land use or type of vegetation).
- **Quantitative Data** showing differences in **amount** (e.g., level of population density or average amount of precipitation).

The first kind of map involves assigning a unique symbol to each named type or a symbol to a group of similar types. The second involves the grouping of data into numeric classes (0-10, 10-20, 20-30) and the assignment of a set of graduated symbols or different hue values to the classes. This activity deals with the second kind of map. It introduces three of the most widely used data classification methods:

- Natural Breaks (Jenks Method) classification - *Class boundaries conform to gaps in data.*
- Equal Interval classification – *Class boundaries placed at regular intervals.*
- Quantile classification - *Equal number of observations in each class.*

It also shows how you can modify a classification to achieve a more legible and reader-friendly classification system.

In this activity you will also learn two important new ArcMap techniques:

- Joining an external data table to a shapefile.
- Setting layer properties to eliminate a no-data feature from the map.

The data needed for proj4 is located in the \\Geogsrv\data\geog281\proj4\data directory.

| **Project 4 files:** | **Description:** | **Feature Type:** |
|---|---|---|
| africa shapefile | Base map of Africa | polygon |
| africa_demog.dbf | Database file containing demographic information | |

The first part of the activity guides you though the steps and decisions you need to take to create maps of population density in Africa. The second part lets you apply what you've learned to additional variables.

---

After copying the data from the server to your local working directory, start ArcMap. Project 4 does **NOT** start with a map document file; you will need to create one.

**Create a map document file for Project 4**

- Name the file proj4.mxd

**Add the following files to the Data View**

- africa.shp
- africa_demog.dbf

A map showing the countries of Africa with a single default color should appear in the map window. In the table of contents, you should see references to both the shapefile (africa) and the database file (africa_demog).

**Joining an External Data Table to a Layer**

Explore the data that are included with the Africa layer:

- In the table of contents, **right-click** on **africa** (the shapefile) and choose **Open Attribute Table.**

Note the attribute field: CNTRY_NAME. This file contains information on country name but no information on population. Note also the Shape field. This confirms that you are working with the attribute table of a Polygon shapefile.

- Close the **africa** attribute table.

Next, look at the separate database file:

- In the table of contents, **right-click** on **africa_demog** and choose **Open.**

This file contains information on country name and a variety of demographic measures including population density (POPDEN00). But unlike the previous file, it has no Shape field. This confirms that it is not an attribute file attached to a shapefile but rather a simple database file. You can learn about the population of Africa countries from this table, but you can't map the data in its present form. Since you want to map population density, you need to be able to work with data from both tables. Before you can do this, you must join together the two files.

ArcMap allows you to join an external data file to a shapefile's attribute table as long as both the external data and the attribute table each contain a field containing identical feature identification information. In this case, the information in the CNTRY_NAME field of the attribute table and the COUNTRY field of the database file contain identical references to the names of the countries of Africa (including identical capitalization and abbreviation).

- Close the **africa_demog** table.

To join the tables using this common information:

- In the table of contents, **right-click** on **africa,** choose **Joins and Relates,** then choose **Join.**
- Under "*What do you want to join to this layer*?" choose **Join attributes from a table.**
- Under "*Choose the field in this layer that the join will be based on*:" press the arrow and highlight **CNTRY_NAME.**
- Under "*Choose the table to join to this layer, or load the table from disk:"* press the arrow and highlight **africa_demog** (it may already be highlighted since it is the only available table in your project).
- Under "*Choose the field in the table to base the join on:"* press the arrow and highlight **COUNTRY** (it may already by highlighted since COUNTRY is the only text field in this table).
- Press **OK.**
- If you are asked if you want to index the join field, you can press **Yes.**

Now open the attribute table of the shapefile to view the changes that have been made:

- In the table of contents, **right-click** on **africa** and choose **Open Attribute Table.**

Scroll across and note that the demographic information from africa_demog has been appended to the area information from the africa shapefile. The field names tell you which source file provided the data. For example, africa.SQKM is part of the original shapefile attribute table while africa_demog.POPDEN00 is

from the separate africa_demog data table. This also reminds you that the data still reside in two separate tables that have been joined within the context of your project, but not physically joined on disk.

**Filtering Out a Missing Data Feature**

If you scroll down the attribute table, you'll see that data exist for all countries except one. Western Sahara has values of -99 which indicate missing data. One way to deal with this is to redefine the Africa layer to exclude Western Sahara:

- Close the attribute table.
- In the table of contents, right-click on **africa** and choose **Properties.**
- Click on the **Definition Query** tab.
- Click on **Query Builder.**
- In the **Fields** list (the top portion of the form containing the attribute fields), double-click on **africa.CNTRY_NAME.**
- Click once on the **Not Equal To** (<>) button.
- Press the **Get Unique Values** button.
- Scroll down and double-click on **Western Sahara.**
- The query including all punctuation should now read: "africa.CNTRY_NAME" <> 'Western Sahara'.
- If the query syntax does not match the above, highlight it, delete it, and try again.
- Once the query is correct, press **OK** twice.

Note that the Western Sahara has disappeared from your map. It's still there on disk, but your project will now treat it (along with its -99 missing data values) as if it were not part of the Africa shapefile.

**Exploring the Distribution of Population Density Values**

Now you are ready to begin the process of classifying your data. Before you decide which classification system to apply, it's usually a good idea to explore your data.

Start by arranging your data from lowest to highest:

- Right-click on **africa** and choose **Open Attribute Table.**
- Right-click on the field name **africa_demog.POPDEN00** and choose **Sort Ascending.**
- Scroll down the list of population density values. Note, how many low values there are.

Next view a frequency distribution histogram of your data:

- Right-click on the field name **africa_demog.POPDEN00** and choose **Statistics.**

The graph shows that population densities are concentrated toward the low end of the range with a few high end values.

**Classification Method Option #1- Equal Interval -** *Class boundaries placed at regular intervals.*
With the Equal Interval classification method, each class has the same range. In other words, the class boundaries are set an equal distance apart. Here are the steps for setting class boundaries using the Equal Interval method:

1. Class interval = range / # of classes
   Range = difference between highest and lowest value
   Data range for Population Density: 362- 2.32 = **359.68**
   Class Interval = 359.68 / 5 classes = **71.94**
2. Set Upper Limits
   Upper Limit of class 1= lowest value + class interval = **74.26**
   Upper Limit of class 2= upper limit of class 1 + class interval = **146.20**
   Upper Limit of class 3= upper limit of class 2 + class interval = **218.14**
   Upper Limit of class 4= upper limit of class 3 + class interval = **290.08**
   Upper Limit of class 5= upper limit of class 4 + class interval = **362**
3. Set Lower Limits
   Lower Limit of class 1= lowest value
   Lower Limit of class 2= upper limit of class 1 + .01 = **74.27**
   Lower Limit of class 3= upper limit of class 2 + .01 = **146.21**
   Lower Limit of class 4= upper limit of class 3 + .01 = **218.15**
   Lower Limit of class 5= upper limit of class 4 + .01 = **290.09**

.01 was added to create the lower ranges because the data precision was 2 decimal places for the population density variable.

You do not need to manually calculate the upper and lower limits because ArcMap will calculate them for you.

What do you think a map of population density will look like, if we use the Equal Interval method of classification? Hint: think about the distribution of the population density data.

- Close the Statistics window and the attribute table window.
- In the table of contents, right-click on **africa** and choose Properties.
- Click on the Symbology tab.
- Under **Show**, click on **Quantities- Graduated Colors.**
- Under Fields, set the **Value** to **africa_demog.POPDEN00.**
- Press the **Classify** button and set the **Classification Method** to **Equal Interval.**
- Press **OK** twice.

Most of Africa falls in the lowest class (2.32 to 74.26) with only a small number of countries falling in the higher classes. Why? As you may recall from looking at the data arranged from low to high, the majority of the data was concentrated below 74.26. As a result most of the countries will fall into the lowest category.

This Equal Interval view of African population densities accurately shows that most African countries have low population densities compared with a small number with much larger densities. But it also hides most of the spatial variation among the vast number of countries in the lowest category. The Equal Interval method is better applied to evenly distributed data.

**Classification Method Option #2- Quantile -** *Equal number of observations in each class.*
With the Quantile classification method, each class will have the same number of observations. If the number of observations cannot be equally divided among the classes, each class will contain a similar number of observations. Here are the steps for assigning observations to each class using the Quantile method:

1. Determine number of observations in each class:
   Total Observations / number of classes
   48 observations/ 5 classes = 9.6 observations in each class
   Since we can not divide an observation between two classes, some classes will contain 10 observations and some will contain 9.

2. Arrange data from lowest to highest.

3. Set class break after each $9^{th}$ or $10^{th}$ observation.
   Class 1- contains 10 observations
   Class 2- contains 10 observations
   Class 3 – contains 10 observations
   Class 4 – contains 9 observations
   Class 5 – contains 9 observations

You do not need to manually assign the observations, ArcMap will assign observations for you.

Quantile classification is useful for "spreading out" the variation that is clustered toward the low end of the scale and "collapsing" the variation at the high end. Let's see what effect this has:

- Retrace your steps and choose **Quantile** as the classification method.
- Press **OK** twice to view the map.

Note how the Quantile method spreads the countries that previously fell into one large category over four smaller ones. You can now see important distinctions within the previous low class. For example, you can see differences between countries of the Sahel (very low density) and many of the countries of Sub-Saharan Africa (somewhat more densely populated). The ability to see important distinctions within the low class comes at the expense of being able to distinguish among the denser countries that are now collapsed into one group ranging from 83.9 to 362.0. Depending on the purpose of your map, this may be acceptable.

**Classification Method Option #3- Natural Breaks -** *Class boundaries conform to gaps in data.*
The Jenks optimization is a statistical method used by ArcMap to find existing groups of values and put them together, thus exploiting the natural gaps in the data. This is the default classification method of ArcMap. It is also a good choice for unevenly distributed data such as population density- where there are many low and medium density countries, but a few high densities countries. To set class breaks for the Natural Breaks classification method:

1. Class breaks are placed where natural gaps occur in the data. This can be done manually by viewing a histogram of the data. For a statistical approach select Natural Breaks (Jenks method) as the classification method in ArcMap. ArcMap uses the Jenks method to minimize differences within the classes, while maximizing differences between the classes.

Let's see what effect this has:

- Retrace your steps and choose **Natural Breaks** as the classification method.
- Press **OK** twice to view the map.

The Natural Breaks classification shows the difference between countries with very low densities (less than 25) and those countries with midrange densities (under 50 and under 100). Higher end densities (over 300) are placed in the highest class; this differs from the Quantile approach which grouped all data over 83.9 together.

**Classification Method Option #4- User Defined Classifications**
One advantage of the Equal Interval, Quantile, and Natural Breaks (using Jenks method) classification methods is that they are based on widely accepted statistical measures. They also produce range breaks such as 2.32 - 11.74 or 11.75 - 34.85 that can be very confusing to someone trying to read the map because they may suggest an unneeded precision. Another common approach is to create categories that respect the general shape of the distribution but that employ round numbers as range breaks. For example, elevation usually has a concentration of low values. Most elevation maps will use round number breaks with an interval that is narrower at the low end and broader at the high end. i.e.: 0 - 1000 feet, 1000 - 2000 feet, 2000 - 5000 feet, 5000 - 10000 feet. Here is another example using the Africa population data:

- Return to the **Symbology** properties window for the **africa** layer.
- Double-click on the first range (2.32 – 11.74), type in **10** (the new upper value) and press **enter.**
- Double-click on the subsequent range values and enter **25**, **50**, **100** as the new upper values for ranges two, three and four.
- The value or the fifth range is automatically calculated – you only have to press **OK** to accept the changes.

Note the new values in the Africa legend in the table of contents. To make your ranges even more legible, you can edit the descriptive labels:

- Reopen the Symbology properties window.
- Under Label, double-click on the value for the lower range and enter **10 or less.**
- Double-click on the value for the highest range and enter **Above 100.**
- Press **OK.**

The resulting map will be different from one generated with a statistical method, but in some circumstances the improvement in readability is more important. If you decide that a "statistically" based approach might confuse your audience, it is OK to apply a user based system based on round numbers and a clear pattern of increasing or decreasing class intervals. It is not appropriate, however, to select a system simply to "force" particular areas into the classes you want them to occupy.

**Background Layer for Missing Data**
For legibility, you should fill in the space vacated by Western Sahara.

- Press the **Add Data** button and bring in a second copy of **africa.shp.**
- **Press and drag** the new layer below the Africa layer showing population density, so that it will draw below the population density layer (you may need to click on the Display button at the bottom of the Table of Contents before ArcMap will let you change the drawing order).
- Open the properties page for the new layer, press the **Symbology** tab, click on the symbol, and change the **Fill Color** to Gray 10%.
- Open the properties page for the new layer, press the **General** tab, and change the **Layer Name** to **Missing Data.**
- If you wish, you can also set the **Definition Query** for the new layer so that the country name will be equal to Western Sahara. This step isn't really necessary in terms of the accuracy of your map.

**On Your Own**
Use the histogram function to explore the data distribution for other variables contained in the Africa_demog data file. Find a variable where the Equal Interval approach produces useful information and create a choropleth map of this data. Switch to the Layout View and create a finished map for that variable.