INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
sustainable solutions for ending hunger and poverty
Supported by the CGIAR

ETHIOPIAN DEVELOPMENT RESEARCH INSTITUTE

Ethiopian Strategy Support Program II (ESSP II)

An Introduction to
Geographical Information Systems

Training Manual

Emily Schmidt, Helina Tilahun,
Mekamu Kedir, and Hailu Shiferaw

Ethiopia Strategy Support Program II (ESSP II)

Ethiopian Development Research Institute (EDRI)

ESSP II / EDRI Training Manual

December 2011

IFPRI-HEADQUARTERS
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
2033 K Street, NW • Washington, DC 20006-1002 USA
Tel: +1-202-862-5600 • Skype: IFPRIhomeoffice
Fax: +1-202-467-4439 • E-mail: ifpri@cgiar.org

IFPRI-Addis Ababa
http://essp.ifpri.info
IFPRI c/o ILRI
P.O. Box 5689, Addis Ababa, Ethiopia
Tel: +251 11 6 17 25 55 Fax: +251 11 6 46 23 18
E-mail: ifpri-addis@cgiar.org
Contact: Bart Minten, Senior Research Fellow and Program Leader

ETHIOPIAN DEVELOPMENT RESEACH INSTITUTE
http://www.edri.org.et/
Blue Building • Addis Ababa Stadium
P.O. Box 2479 • Addis Ababa, Ethiopia
Tel: +251 11 5 50 60 66; +251 11 5 52 53 15
Fax: +251 11 5 50 55 88
Email: exe-director@edri.org.et

Contents

Lab 11: Meteorological data presentation

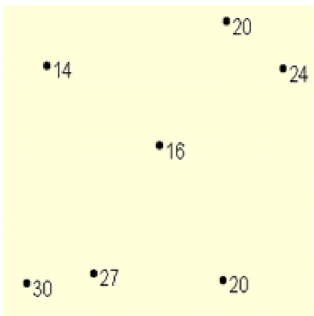
An overview of the interpolation tools

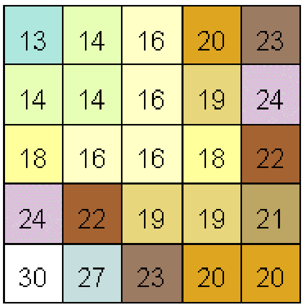
Visiting every location in a study area to measure the height, magnitude, or concentration of a phenomenon is usually difficult or expensive. Instead, you can measure the phenomenon at strategically dispersed sample locations, and predicted values can be assigned to all other locations. Input points can be either randomly or regularly spaced or based on a sampling scheme.

Surface interpolation functions create a continuous (or prediction) surface from sampled point values. The continuous surface representation of a raster dataset represents height, concentration, or magnitude—for example, elevation, pollution, or noise. Surface interpolation functions make predictions from sample measurements for all locations in a raster dataset whether or not a measurement has been taken at the location. There is a variety of ways to derive a prediction for each location; each method is referred to as a model. With each model, there are different assumptions made of the data, and certain models are more applicable for specific data—for example, one model may account for local variation better than another. Each model produces predictions using different calculations.

Understanding interpolation analysis

Interpolation predicts values for cells in a raster from a limited number of sample data points. It can be used to predict unknown values for any geographic point data, such as elevation, rainfall, chemical concentrations, noise levels, and so on. The illustrations below show the distribution and values of sample points and the raster generated from them. The left illustration shows a point dataset of known rainfall-level values. The illustration on the right shows a raster interpolated from these points. Unknown values are predicted with a mathematical formula that uses the values of nearby known points.



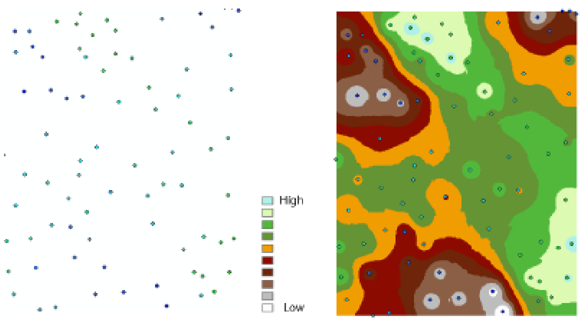


Why interpolate to raster? The assumption that makes interpolation a viable option is that spatially distributed objects are spatially correlated; in other words, things that are close together tend to have similar characteristics. For instance, if it is raining on one side of the street, you can predict with a high level of confidence that it is raining on the other side of the street. You would be less certain if it was raining across town and less confident still about the state of the weather in the next district.

Using the above analogy, it is easy to see that the values of points close to sampled points are more likely to be similar than those that are farther apart. This is the basis of interpolation. A typical use for point interpolation is to create an elevation surface from a set of sample measurements.

In the following graphic, each symbol in the point layer represents a location where the elevation has been measured. By interpolating, the values between these input points will be predicted.

Interpolation Techniques



The IDW (inverse distance weighted) and Spline interpolation tools are referred to as deterministic interpolation methods because they are directly based on the surrounding measured values or on specified mathematical formulas that determine the smoothness of the resulting surface. A second family of interpolation methods consists of geostatistical methods, such as kriging, which are based on statistical models that include autocorrelation—that is, the statistical relationships among the measured points. Because of this, geostatistical techniques not only have the capability of producing a prediction surface but also provide some measure of the certainty or accuracy of the predictions.

How Kriging works: Kriging is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z-values. Unlike other interpolation methods supported by ArcGIS Spatial Analyst, to use the Kriging tool effectively involves an interactive investigation of the spatial behavior of the phenomenon represented by the z-values before you select the best estimation method for generating the output surface.

What is kriging? Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. The Kriging tool fits a mathematical function to a specified number of points, or all points within a specified radius, to determine the output value for each location. Kriging is a multistep process; it includes exploratory statistical analysis of the data, variogram modeling, creating the surface, and (optionally) exploring a variance surface. Kriging is most appropriate when you know there is a spatially correlated distance or directional bias in the data. It is often used in soil science and geology.

The kriging formula: Kriging is similar to IDW in that it weights the surrounding measured values to derive a prediction for an unmeasured location. The general formula for both interpolators is formed as a weighted sum of the data:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

Where, $Z(s_i)$ = the measured value at the *i*th location

λ_i = an unknown weight for the measured value at the *i*th location

s_0 = the prediction location

N = the number of measured values

In IDW, the weight, λ_i, depends solely on the distance to the prediction location. However, with the kriging method, the weights are based not only on the distance between the measured points and the prediction location but also on the overall spatial arrangement of the measured points. To use the spatial arrangement in the weights, the spatial autocorrelation must be quantified. Thus, in ordinary kriging, the weight, λ_i, depends on a fitted model to the measured points, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location. The following sections discuss how the general kriging formula is used to create a map of the prediction surface and a map of the accuracy of the predictions.

Creating a prediction surface map with kriging

To make a prediction with the kriging interpolation method, two tasks are necessary:

- Uncover the dependency rules
- Make the predictions.

To realize these two tasks, kriging goes through a two-step process:

1. It creates the variograms and covariance functions to estimate the statistical dependence (called spatial autocorrelation) values that depend on the model of autocorrelation (fitting a model).
2. It predicts the unknown values (making a prediction).

It is because of these two distinct tasks that it has been said that kriging uses the data twice: the first time to estimate the spatial autocorrelation of the data and the second to make the predictions.

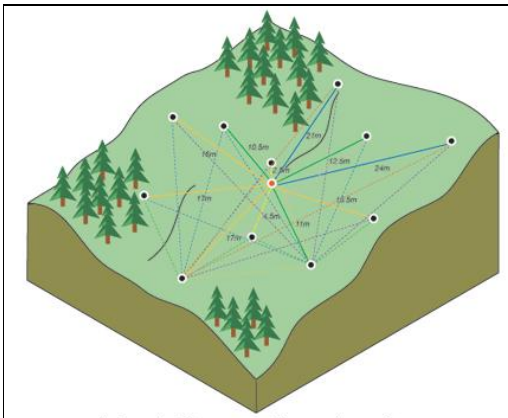
Variography

Fitting a model, or spatial modeling, is also known as structural analysis, or variography. In spatial modeling of the structure of the measured points, you begin with a graph of the empirical semivariogram, computed with the following equation for all pairs of locations separated by distance h:

Semivariogram($distance_h$) = 0.5 * average{($value_i$ – $value_j$}^{2}]

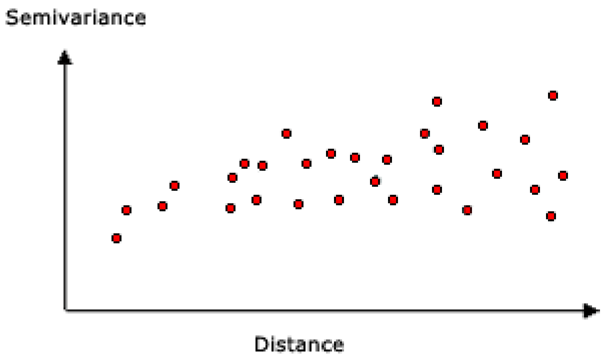
The formula involves calculating the difference squared between the values of the paired locations.

The image below shows the pairing of one point (the red point) with all other measured locations. This process continues for each measured point.



Calculating the difference squared between the paired locations

Often, each pair of locations has a unique distance, and there are often many pairs of points. To plot all pairs quickly becomes unmanageable. Instead of plotting each pair, the pairs are grouped into lag bins. For example, compute the average semivariance for all pairs of points that are greater than 40 meters apart but less than 50 meters. The empirical semivariogram is a graph of the averaged semivariogram values on the y-axis and the distance (or lag) on the x-axis (see diagram beside).



Empirical semivariogram graph example

Spatial autocorrelation quantifies a basic principle of geography: things that are closer are more alike than things farther apart. Thus, pairs of locations that are closer (far left on the x-axis of the semivariogram cloud) should have more similar values (low on the y-axis of the semivariogram cloud). As pairs of locations become farther apart (moving to the right on the x-axis of the semivariogram cloud), they should become more dissimilar and have a higher squared difference (moving up on the y-axis of the semivariogram cloud).

Fitting a model to the empirical semivariogram

The next step is to fit a model to the points forming the empirical semivariogram. Semivariogram modeling is a key step between spatial description and spatial prediction. The main application of kriging is the prediction of attribute values at unsampled locations. The empirical semivariogram provides information on the spatial autocorrelation of datasets. However, it does not provide information for all possible directions and distances. For this reason, and to ensure that kriging predictions have positive kriging variances, it is necessary to fit a model—that is, a continuous function or curve—to the empirical semivariogram. Abstractly, this is similar to regression analysis, in which a continuous line or curve is fitted to the data points.

To fit a model to the empirical semivariogram, select a function that serves as your model—for example, a spherical type that rises and levels off for larger distances beyond a certain range (see the spherical model example below). There are deviations of the points on the empirical semivariogram from the model; some points are above the model curve, and some points are below. However, if you add the distance each point is above the line and add the distance each point is below the line, the two values should be similar. There are many semivariogram models from which to choose.

Semivariogram models

ArcGIS Spatial Analyst provides the following functions from which to choose for modeling the empirical semivariogram:

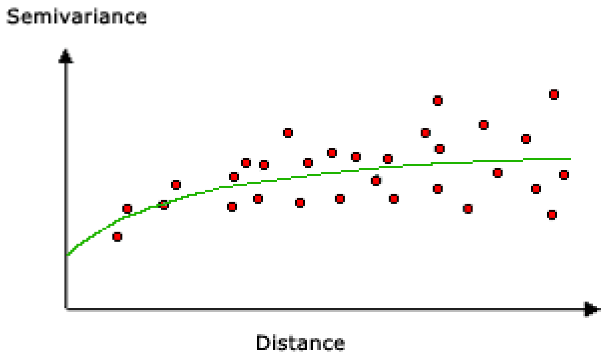
- Circular; Spherical; Exponential; Gaussian ; Linear

The selected model influences the prediction of the unknown values, particularly when the shape of the curve near the origin differs significantly. The steeper the curve near the origin, the more influence the closest neighbors will have on the prediction. As a result, the output surface will be less smooth. Each model is designed to fit different types of phenomena more accurately.

The diagrams below show two common models and identify how the functions differ:

A spherical model example

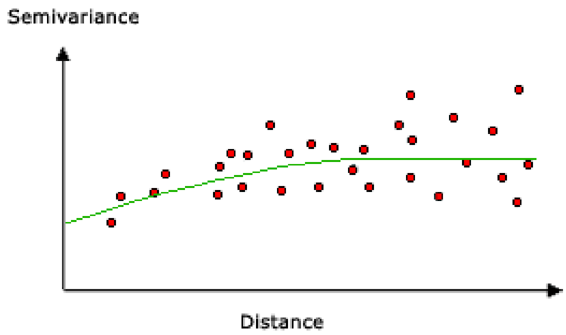
This model shows a progressive decrease of spatial autocorrelation (equivalently, an increase of semivariance) until some distance, beyond which autocorrelation is zero. The spherical model is one of the most commonly used models.



Exponential model example

An exponential model example

This model is applied when spatial autocorrelation decreases exponentially with increasing distance. Here, the autocorrelation disappears completely only at an infinite distance. The exponential model is also a commonly used model. The choice of which model to use is based on the spatial autocorrelation of the data and on prior knowledge of the phenomenon.



Spherical model example

Understanding a semivariogram—Range, sill, and nugget

As previously discussed, the semivariogram depicts the spatial autocorrelation of the measured sample points. Because of a basic principle of geography (things that are closer are more alike), measured points that are close will generally have a smaller difference squared than those farther apart. Once each pair of locations is plotted after being binned, a model is fit through them. Range, sill, and nugget are commonly used to describe these models.

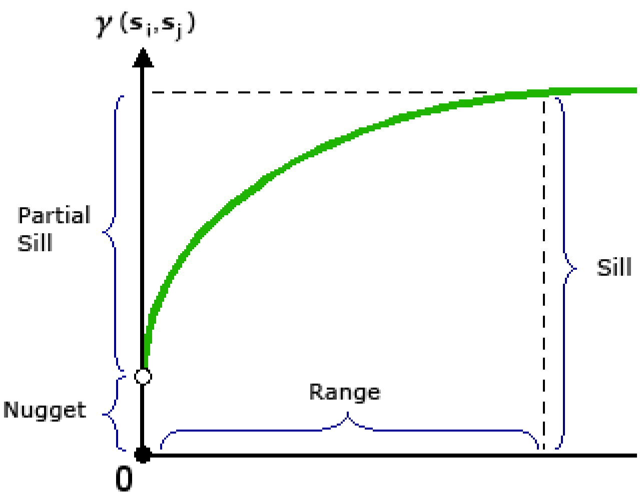


Illustration of Range, Sill, and Nugget components

Range and sill

When you look at the model of a semivariogram, you will notice that at a certain distance the model levels out. The distance where the model first flattens is known as the range. Sample locations separated by distances closer than the range are spatially autocorrelated, whereas locations farther apart than the range are not.

The value at which the semivariogram model attains the range (the value on the y-axis) is called the sill. A partial sill is the sill minus the nugget. The nugget is described in the following section.

Nugget

Theoretically, at zero separation distance (for example, lag = 0), the semivariogram value is 0. However, at an infinitely small separation distance, the semivariogram often exhibits a nugget effect, which is a value greater than 0. If the semivariogram model intercepts the y-axis at 2, then the nugget is 2.

The nugget effect can be attributed to measurement errors or spatial sources of variation at distances smaller than the sampling interval (or both). Measurement error occurs because of the error inherent in measuring devices. Natural phenomena can vary spatially over a range of scales. Variation at microscales smaller than the sampling distances will appear as part of the nugget effect. Before collecting data, it is important to gain an understanding of the scales of spatial variation in which you are interested.

Making a prediction

After you have uncovered the dependence or autocorrelation in your data (see Variography section above) and have finished with the first use of the data—using the spatial information in the data to compute distances and model the spatial autocorrelation—you can make a prediction using the fitted model. Thereafter, the empirical semivariogram is set aside.

You can now use the data to make predictions. Like IDW interpolation, kriging forms weights from surrounding measured values to predict unmeasured locations. As with IDW interpolation, the measured values closest to the unmeasured locations have the most influence. However, the kriging weights for the surrounding measured points are more sophisticated than those of IDW. IDW uses a simple algorithm based on distance, but kriging weights come from a semivariogram that was developed by looking at the spatial nature of the data. To create a continuous surface of the phenomenon, predictions are made for each location, or cell centers, in the study area based on the semivariogram and the spatial arrangement of measured values that are nearby.

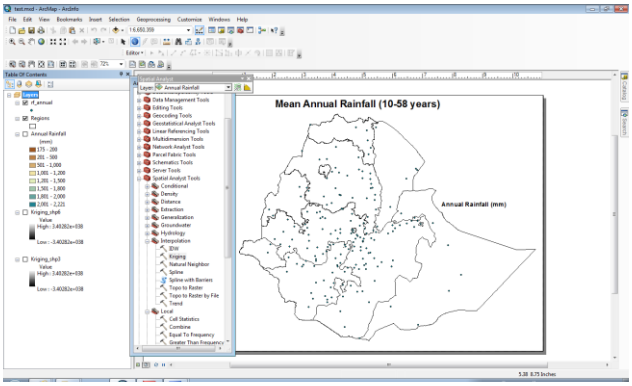
Kriging methods

There are two kriging methods: ordinary and universal. Ordinary kriging is the most general and widely used of the kriging methods and is the default. It assumes the constant mean is unknown. This is a reasonable assumption unless there is a scientific reason to reject it.

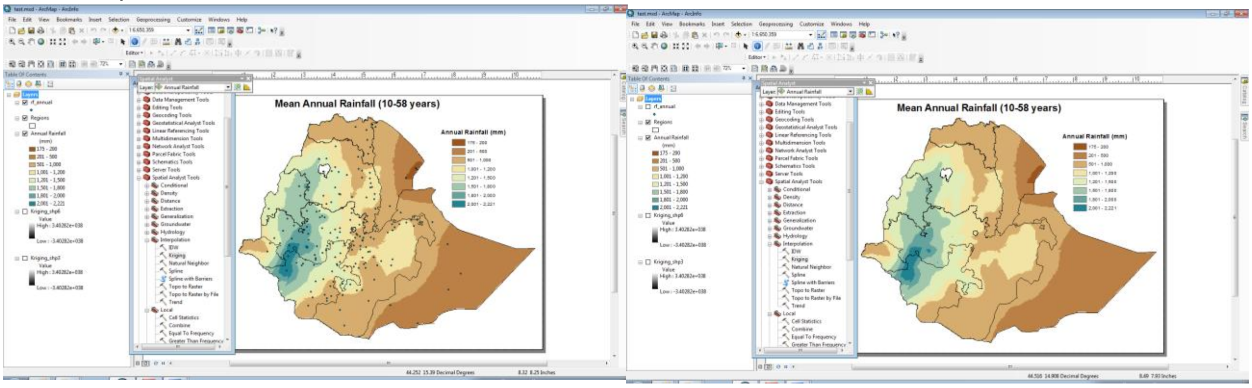
Universal kriging assumes that there is an overriding trend in the data—for example, a prevailing wind—and it can be modeled by a deterministic function, a polynomial. This polynomial is subtracted from the original measured points, and the autocorrelation is modeled from the random errors. Once the model is fit to the random errors and before making a prediction, the polynomial is added back to the predictions to give meaningful results. Universal kriging should only be used when you know there is a trend in your data and you can give a scientific justification to describe it.

Sample before and after interpolation

Before interpolation



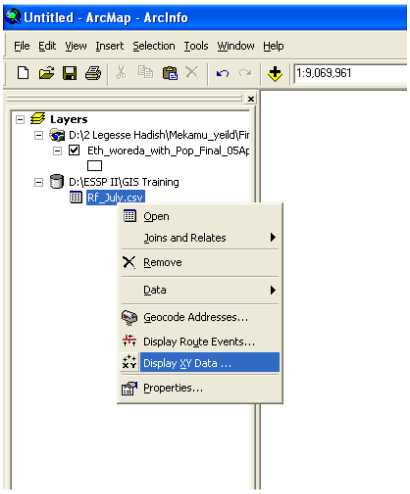
After interpolation

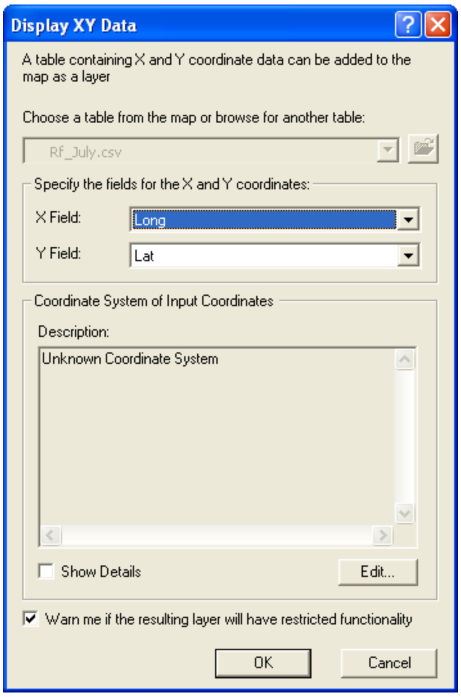


Lab 11 - Exercise Part 1: Using NMA data (Station data)

This exercise is aimed at using ArcGIS-Kriging interpolation technique to estimate and map the spatiotemporal variation of rainfall for Ethiopia. Rainfall data, for the periods between 1950 and 2000, are given in excel for 183 observed values across the country.

1. Open Excel, and go to your Lab 11. Open “RF_July.xls”. You can see that there is a latitude (“X”) and longitude (“Y”) field, along with 10-58 years observed/measured annual rainfall values for 183 locations/”stations”.
2. In order to bring an Excel worksheet into ArcGIS, you must save it as a .dbf file, or a .csv file (make sure to choose ms-dos format). This is sometimes problematic because it occasionally truncates values during the saving process.
3. Resave your file as “Rf_July.csv” into your working directory.
4. Close all worksheets in Excel. You need to have the worksheet that you intend to add to ArcGIS closed in excel in order to add correctly.
5. Open a new session in ArcGIS and add the file “Regions” from your Lab 11 and the database file: “Rf_July.csv” that you saved in your working directory.
6. Right click on the “Rf_July.csv” file and scroll to “Display XY data”. Click on “Display XY data” and a window should pop up like the window to the right.
7. Make sure that your “X Field” displays longitude for the coordinates, and your “Y Field” displays latitude for your coordinates. Press OK.





8. You may get a warning message stating that your “Table Does Not Have Object-ID Field”. This is a unique identifier that ArcGIS builds into all of its shapefiles. Press OK and ArcGIS will create this field for you.
9. Now, can you see your “Rf_July” points? Where are they? Right click on the “Rf_July.csv Events” layer, and from the menu choose “Zoom to layer”. The “Rf_July” layer should now be visible. If the “Regions” layer is not visible, please check the projection of the weather station points (in this case follow the STEPS 10 to 18, otherwise go to STEP 19.

10. Right click on the “Regions” layer and choose “Zoom to Layer”. What happens to the “Rf_July Events” layer? Magic, it disappeared…or is it a projection problem?

11. Go to the main tool bar, and select the “Zoom to full Extent” button 

12. You should now see the entire “Regions” layer incomplete overlap each other. If you use the regular zoom tool , and zoom repeatedly into these layers, you will realize that it is in fact the “Rf_July.csv Events” layer. As it is in a different projection, it is unable to locate and resize itself correctly in relation to the “Regions” Layer.

13. If we take the assumption that this information was collected by GPS, then reverting to the default coordinate system used by GPS will correct this issue.

14. So let’s try our hypothesis! First, Right click on the “Rf_July.csv Events” you select Remove.

15. Now, right click on the “Rf_July.csv” layer. Left click on the “Display XY Data”.

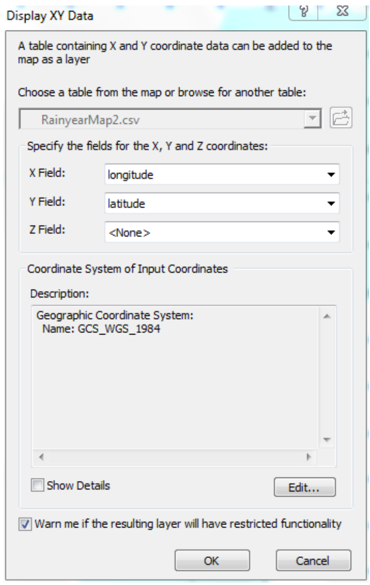
16. As you can see, the coordinate system is unknown. Click on the “Edit” button. In the next window, click the “Select” button and choose the following path:

Use either of the following projection according to your boundary projection

Geographic Coordinate Systems	Projected Coordinate Systems
World	UTM
WGS1984	WGS 1984
	WGS 1984 UTM Zone 37N.prj

17. Click Add.

18. Your “Display XY Data” window should now look like the graphic to the right. Click OK



19. Your Rainfall measurement points (weather stations points) should be geographically contiguous with your “Regions” layer.

20. Your “Rf_July.csv Events” is currently only a cosmetic layer. We know this because it has the word “Events” following the name. It is not yet a shapefile.

21. To create a permanent shapefile from this cosmetic layer, right click on the “Rf_July.csv Events”, scroll down to “Data” and select the “Export Data” option.

22. Leave all the initial options as default, but make sure to save the final file to your working directory, calling the file “Rf_July_ETh.shp”

23. A pop-up window will ask you if you would like to “Add to map”. Select OK and the new shapefile should automatically add to the data frame.

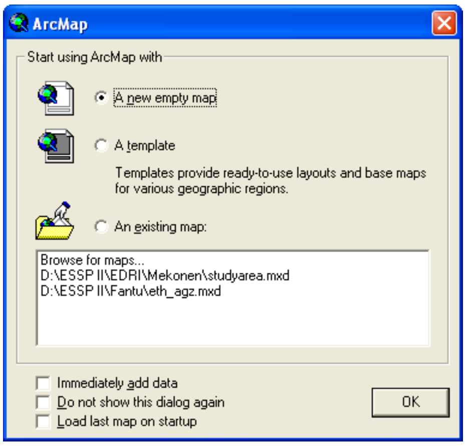
24. Look at these data, how the available rainfall observation/measurement points are representative and spatially distributed across the country and to be used to estimates other unknown locations.

25. Create a map with the rainfall measurement point's layer over the province boundary.

Creating a surface using kriging interpolation

1. Start ArcMap by double-clicking the ArcMap icon on your computer desktop. Alternately, click the "Start" menu, point to "All Programs", point to "ArcGIS" and select "ArcMap") 

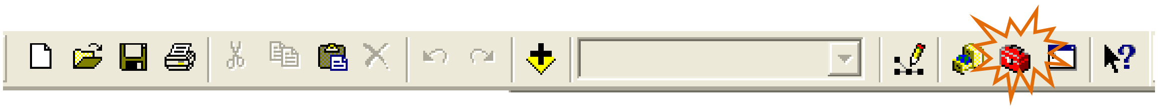
2. You may receive the following welcome screen, if so, select "a new empty map," and press OK. If you do not receive this screen, ArcMap has selected a blank map by default.

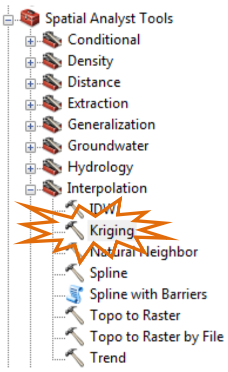


3. You are now looking at the basic ArcMap screen with its various menus and tools. To begin with, we will add rainfall data. From the "File" menu, select "Add Data".

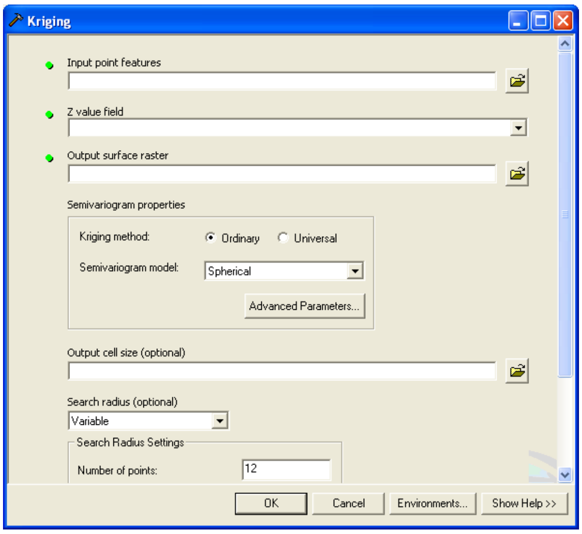
   From you working directory add "Regions" and "Rf_July_ETh" files.

4. From the standard menu click ArcToolbox window and double click the ***Spatial Analyst Tools*** from the ArcTool Box Window and you will find a number of functionality tools and when you double click the ***Interpolation tool*** you will find IDW, Kriging, Natural Neighbor, Spline,…Trend





5. Double click Kriging. Kriging window wizard opens up.



6. Click the ***Input point features*** dropdown arrow and click the rainfall dataset Rf_July_ETh point data to use.
7. Click the ***Z value field*** dropdown arrow and click the field you want to use in this case use “ANN”. The field that holds the rainfall magnitude value for each point.
8. In ***Output surface raster*** specify a name for the outputs (annual_rn) to write in your working directory, or leave the default to create a temporary dataset.
9. Click the ***kriging method*** you want to use (left click on the Ordinary radio button; please refer the Kriging methods section above).
10. Click the ***Semivariogram model*** dropdown arrow and click the model you want to use. (Read what type of semivariogram models included in ArcGIS 10 above; but for this case use spherical-default)
11. Change the default ***output cell*** size to 0.00833 degree which is equivalent to 1km (Since your dataset is in lat/long the unit is if it is in degree).
12. Click the ***Search radius*** type dropdown arrow and click Variable.

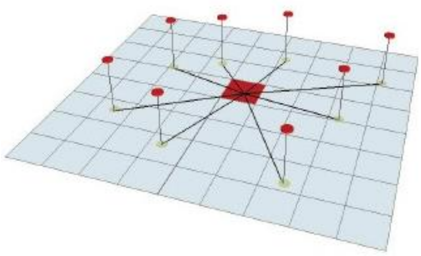
Note: Uses a variable search radius in order to find a specified number of input sample points for the interpolation.

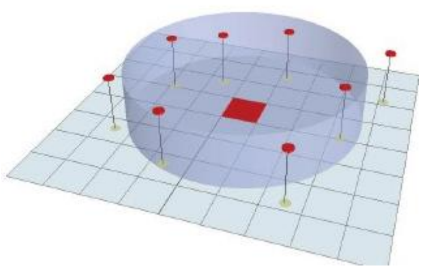
<u>Variable</u>

- ***Number of points***—an integer value specifying the number of nearest input sample points to be used to perform interpolation. The default is 12 points.
- ***Maximum distance***—Specifies the distance, in map units, by which to limit the search for the nearest input sample points. The default value is the length of the extent's diagonal.

<u>Fixed</u>

- Uses a specified fixed distance within which all input points will be used for the interpolation. Distance—Specifies the distance as a radius within which input sample points will be used to perform the interpolation.



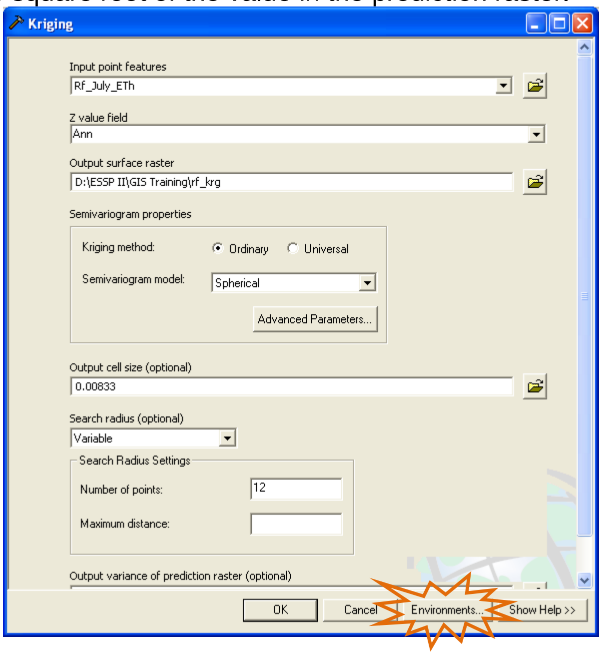


The value of the radius is expressed in map units. The default radius is five times the cell size of the output raster.

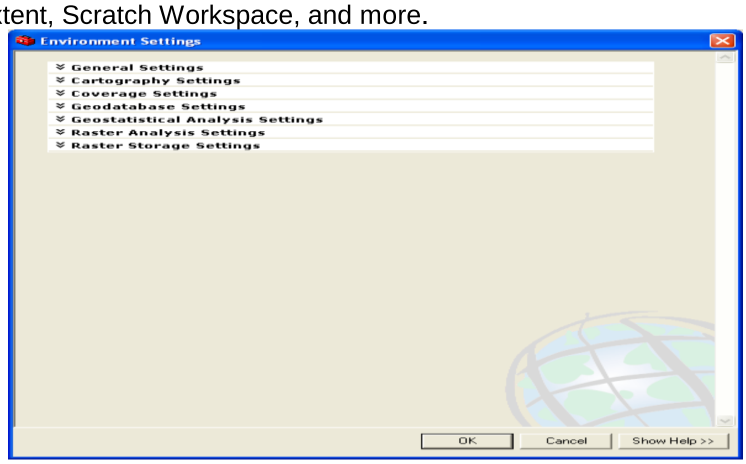
Minimum number of points—An integer defining the minimum number of points to be used for interpolation. The default value is 0.

13. Optionally, change the default number of points.
14. Optionally, specify a maximum distance.
15. Optionally, output variance of prediction raster.

    The optional output variance of prediction raster contains the kriging variance at each output raster cell. Assuming the kriging errors are normally distributed, there is a 95.5 percent probability that the actual z-value at the cell is the predicted raster value, plus or minus two times the square root of the value in the prediction raster.



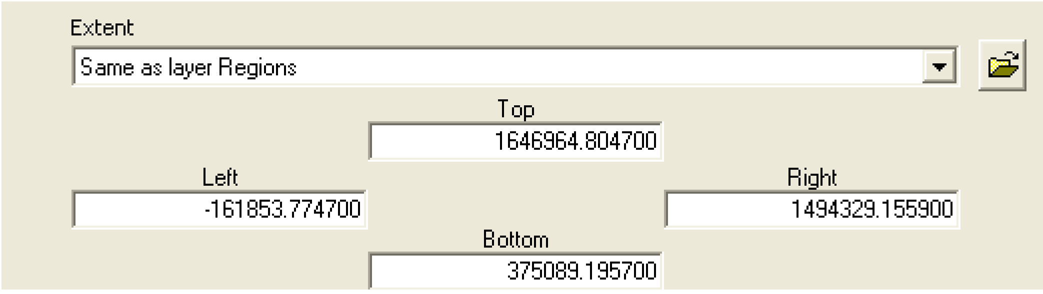
16. Click the ***Environments tab*** located at the bottom of the kriging window (as shown above figure). The Environments helps to set up Cell Size, Current Workspace, Output Coordinate System, Extent, Scratch Workspace, and more.



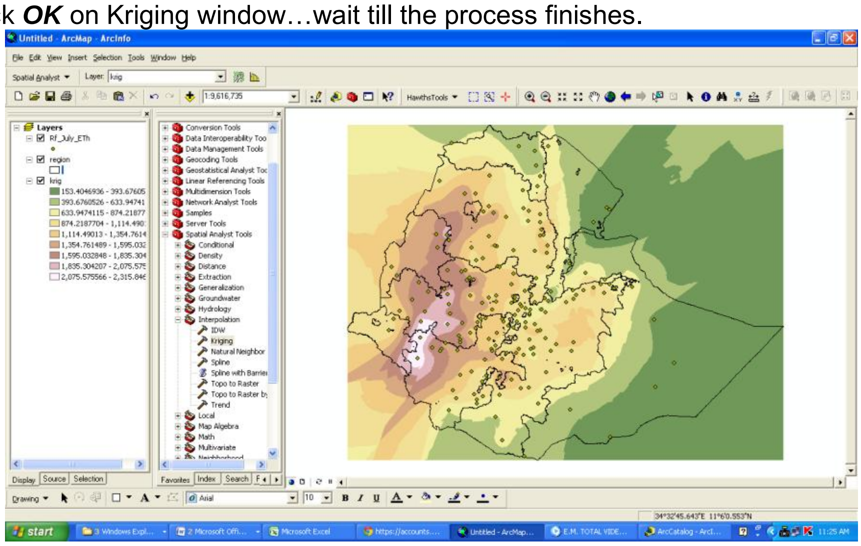
17. Click the ***General Setting*** tab and click the drop down box under Extent and you will find different extent setting options as
    - Default-No extent set
    - Union of inputs- The maximum area of all input(s)
    - Intersection of inputs-The minimum area common to all input(s)
    - As specified below (you need to specify the Left, right, bottom and top value of X and Y)
    - Same as display- the extent of the current ArcMap display wil be used
    - Same as layer <layer>-The extent will be based on the extent of the specified layer

Output Extent environment will only process features or rasters that fall within the extent specified in this setting and any feature or raster that passes through the specified extent will be processed and written to output. Note that the selected extent is used only to select features, not clip them.

18. For this particular exercise choose the ***Same as Regions*** and click Ok the Environments Setting window



19. Click ***OK*** on Kriging window…wait till the process finishes.



Following the above steps, you are expected to create rainfall grids for the remaining months (January-December), and also for temperature datasets.

Once you have 10-58 rainfall raster grids follow the following step to clip out the grid by the country boundary and create a lay out for presentation.

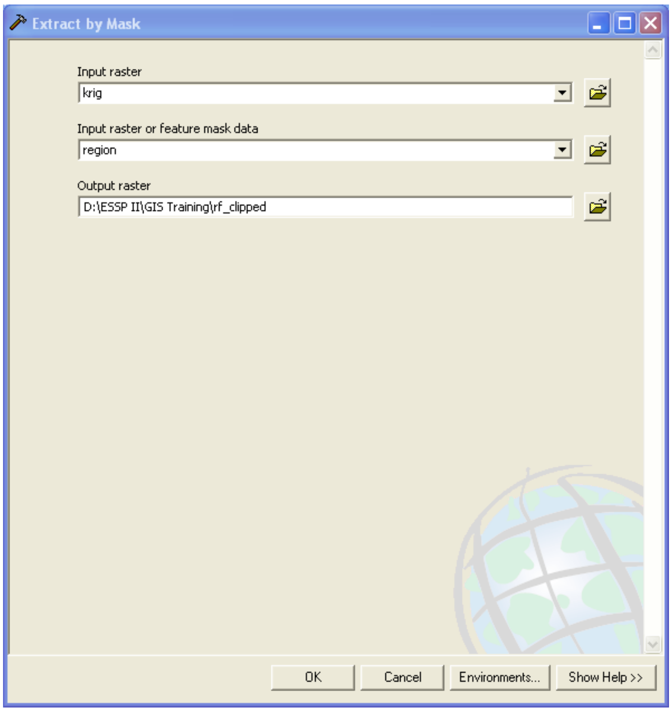
Creating layout for presentation

20. Double click the ***Spatial Analyst Tools*** from the ArcTool Box Window and you will see a number of tools
    - Double click on ***Extraction*** tool box;
    - Double click ***Extract by Mask*** tool box: This tool allows you to extract a portion of a raster dataset based on a template extent.

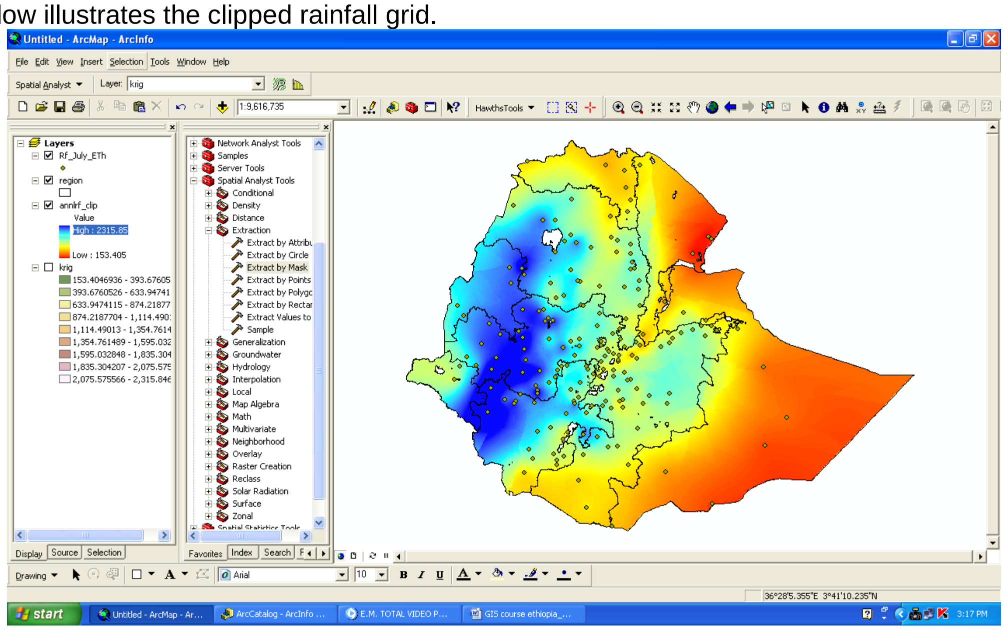
21. On the clip window under the ***Input Raster*** click the dropdown arrow and select the *annual_rn* grid as in put to be clipped.

22. In the ***Input raster*** or ***feature mask data*** option click the dropdown arrow and select the Regions boundary file as a clip feature.

23. In the ***Output raster*** write the output name in your working directory (rf_clipped)



Below illustrates the clipped rainfall grid.



Repeat the above step to clip the rest of rainfall grid files.

Prepare your map for printing

1. Now you have clipped rainfall grid and regional level administrative boundary of Ethiopia. The next step is to prepare this map for printing.

2. First, make sure that your "Region" shapefile and "RF_clipped" grids are all turned on by checking the checkmark box next to the layer name in the table of contents.

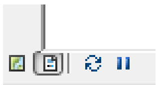
3. Reorder the layers in table of content as Region and RF_clipped. If the Region boundary filled with color it covers the underneath RF_clipped, therefore make sure that the Region boundary should be hallow.

4. Double click on the Regions symbol and Symbol selector window will display form the options, select Hollow.

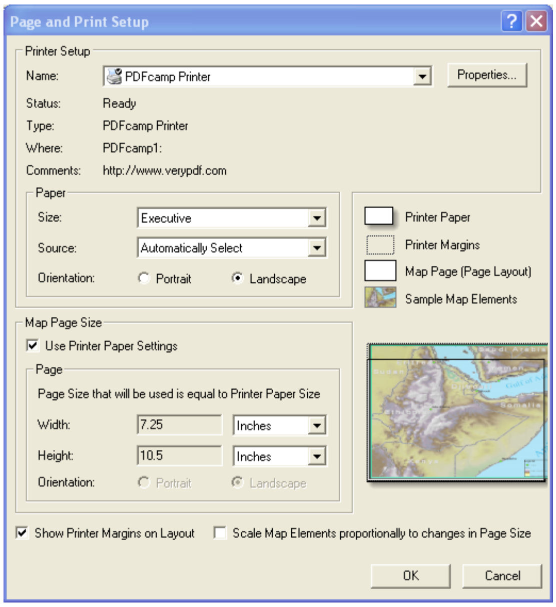
5. The map window that you have been working in up to now is called the “Data View”.

   We will now move to a different map window called the “Layout View” to prepare the map for printing. The “Data View” is where you perform the majority of your GIS editing work, while the “Layout View” is where you arrange your map(s) along with preferred graphics for printing or export.

6. Select the layout view



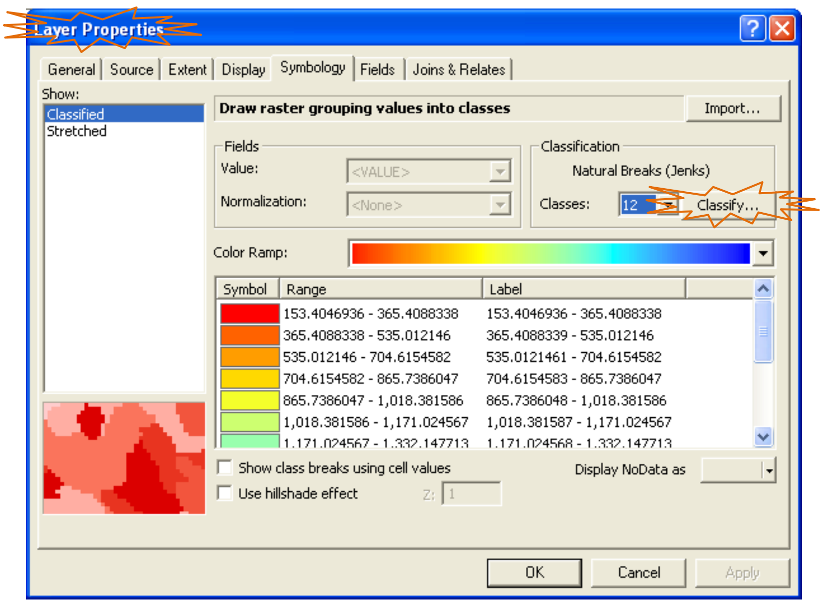
7. We want to select a paper setup that best fits the geographical nature of Ethiopia in order to use the most space on our page for our map. Click on the File tab in the upper left of your screen and scroll to “Page and Print Setup”. Make sure you select Landscape for the Orientation. Also click the checkmark to “Use Printer Paper Settings”. This will give you a dotted outline of how far you are able to expand your map. (See right for diagram)



8. Now, resize the window around your map of Ethiopia, so that it fits within the printable area.

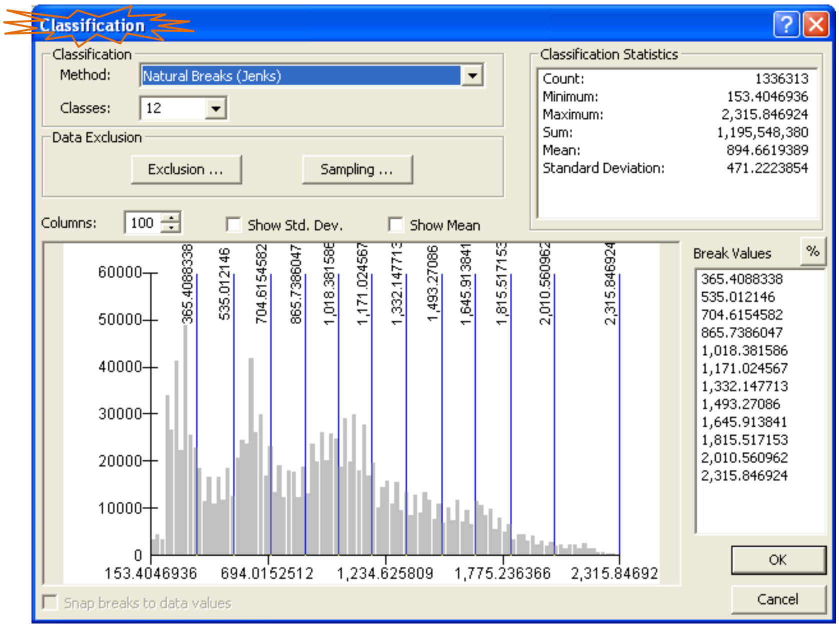
9. Double-click on the layer name “Rf_clipped”; this will take you directly to the “Layer Properties” for the layer. Select the tab “Symbology”.

10. In the “Show” box (see graphic), click on the option “Classified”, and next click on the “Classify” tab. This will take you to the “Classification” drop down menu (see graphic below).



11. Click on the arrow to the right of "Classes" and choose 12. Click on the arrow to the right of "***Method"*** and choose "Manual".

12. Choose your "Break Values" set be 174 intervals started from 500 and type the maximum value 2,233 mm of annual rainfall as there is no value more than this. (See below)



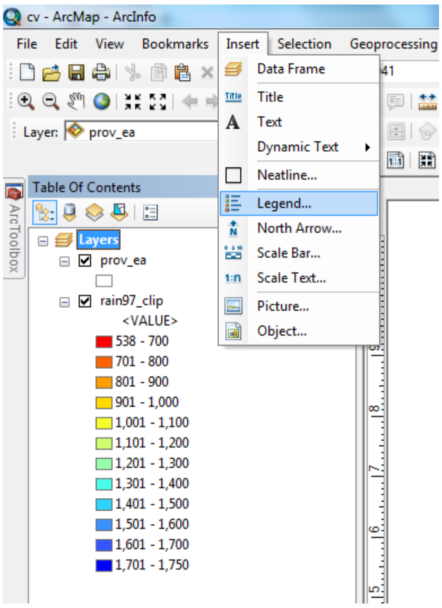
    Note: You can use your own break value and number of classifications

13. If you click ***color ramp*** you will see different color ramp options and choose your color of interest that best explains the data. For this exercise choose spectrum-full bright color ramp. To select the color ramp by name instead of graphic view right click the color ramp and uncheck the graphic view. From the drop down option find the spectrum-full bright.

    Note: You can change the color property by right on the color ramp and point to properties. Edit color ramp window lets you manipulate the color property and you can also save the style and reuse it.

14. Click ***Ok*** on *Classification* and the layer properties.

15. Now, if you want to present more than one map per page you can insert new data frame using "Insert" menu and resize it to fit page setup and map orientation.



16. The classification scheme you set for the rainfall data can be saved as a layer file that can be used for similar use.

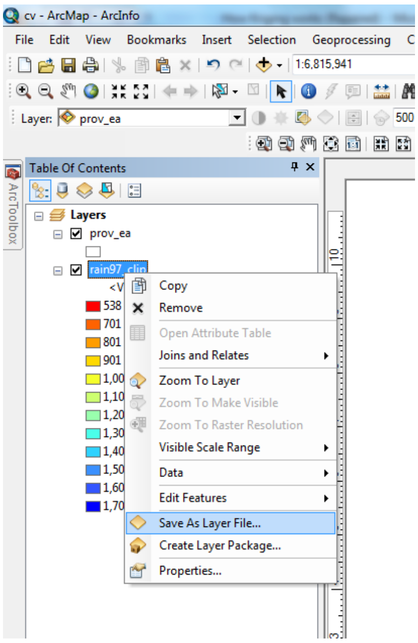
    When you save a layer to disk, you save everything about the layer, such as the symbolization and labeling. When you import/add a layer file to another map, it will draw exactly as it was saved.

    One of the main features of a layer is that it can exist outside your map as a file on disk. This makes it easy for others to access the layers you've built.

    Right click the rain98_clip layer from the table of content and click Save As Layer File… option and type appropriate name and save the layer in your working directory.

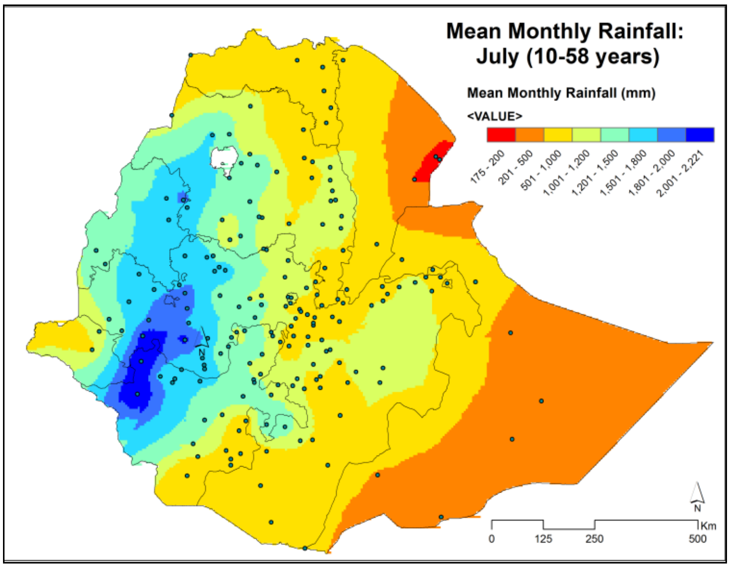
How to import symbolization in ArcMap

The symbology of the previous map will be imported into this map. This tool can be useful in workflows that require much more intricately detailed symbology that would take considerable time to recreate for each new map.

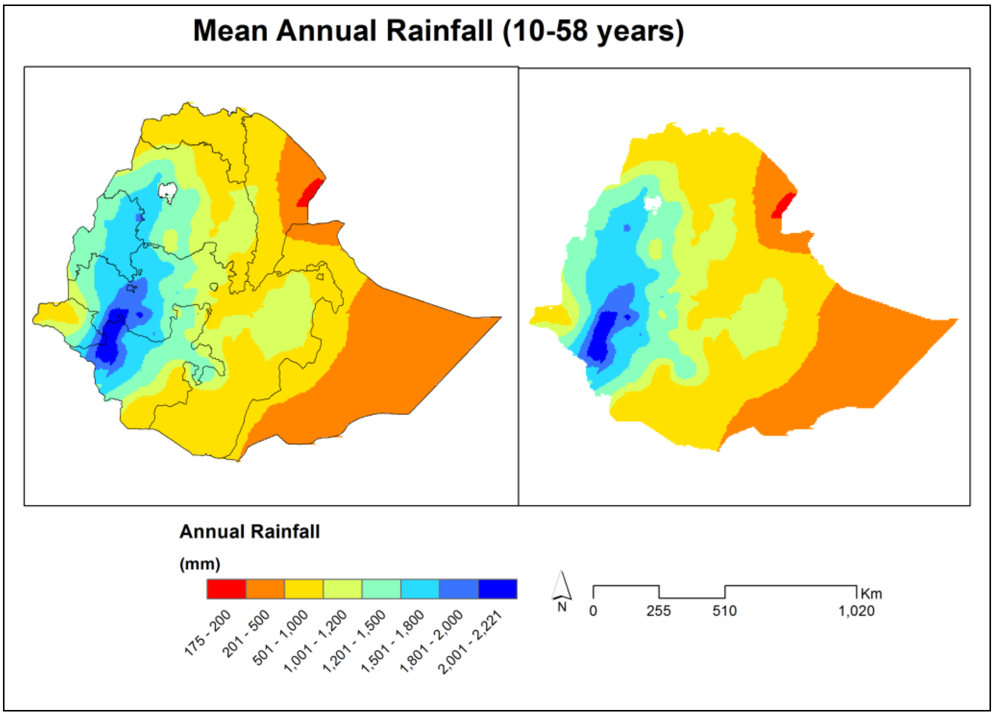


1. In the ArcMap table of contents, right-click the layer for which you want to import symbology (saved layer).
2. Click Properties.
3. In the dialog that appears, click the Symbology tab.
4. Click the Import button.
5. If importing from another layer, choose the layer from the dropdown list.
6. If importing from a layer file, click the Browse button and choose the file from which you want to import. Click OK.
7. Click OK to close the Import Symbology dialog box.
8. Using the “Insert” menu, add a Title, Scale Bar, and North Arrow to your map. Using the insert title option, name your map “Mean Annual rainfall” (or something similar).

   There are many Scale Bars, and North Arrows styles to choose from. Pick your favorite.
9. Return to the menu option “Insert” and click on “Legend”.
10. In the Legend Wizard dialog that pops up, accept all the default options by Clicking “Next” repeatedly and finally “Finish”.
11. You will notice a small legend appear somewhere in the middle of your layout window. Position it correctly.
12. If you wish to change the name of your layers in the legend, edit the corresponding text in the table of contents, and it will also change in the legend. Just click twice on the text in the table of contents to edit it (as you would to rename a file in Windows Explorer).
13. When you are happy with the appearance of your map, you can export this map as a .jpeg, .pdf, or .png, by choosing the menu option “File”, scroll to “Export”, then save with as in working directory.



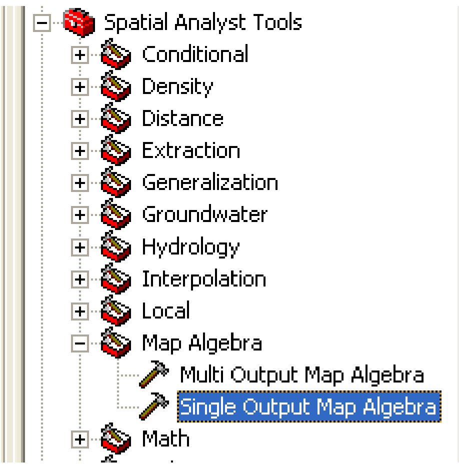
Annual rainfall two maps per page



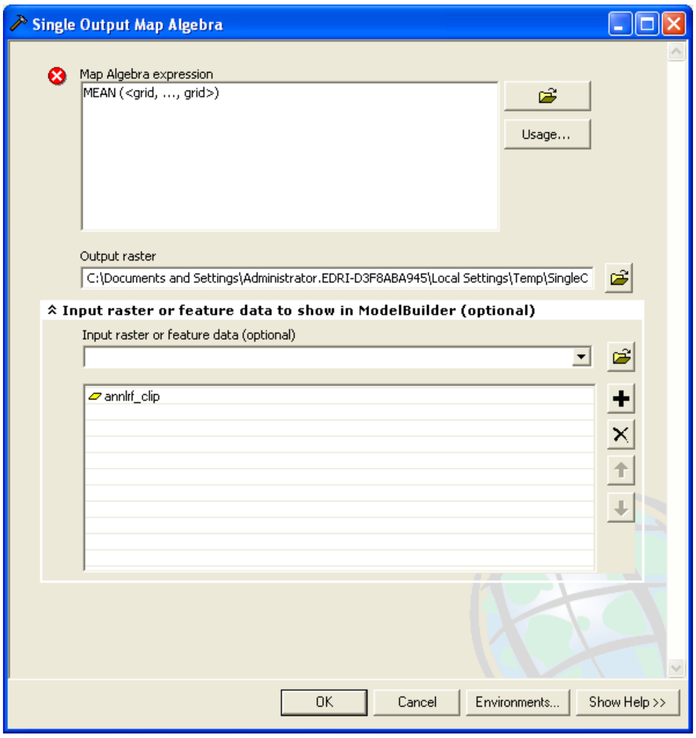
Calculating rainfall coefficient of variability

The coefficient of variation calculated as the standard deviation divided by the mean is often used to represent how rainfall is variable/ stable across the country.

1. To calculate mean: (ArcTool Box→ Spatial Analyst tools→ Map Algebra →double click on ***Single output Map Algebra***.
2. In the ***Input raster or feature data (optional)*** click the drop down arrow and select all input data for the mean calculation one by one.



3. In the ***Map Algebra Expression*** type MEAN and create an open bracket "("and from the ***Input raster or feature data (optional)*** select one layer at a time and drag and drop in the Map Algebra Expression field. Separate each layer by a comma and finally close the bracket.
4. Specify the output directory and name output as "MEAN_Rainfall" in the output raster
5. Click Ok on to calculate the mean



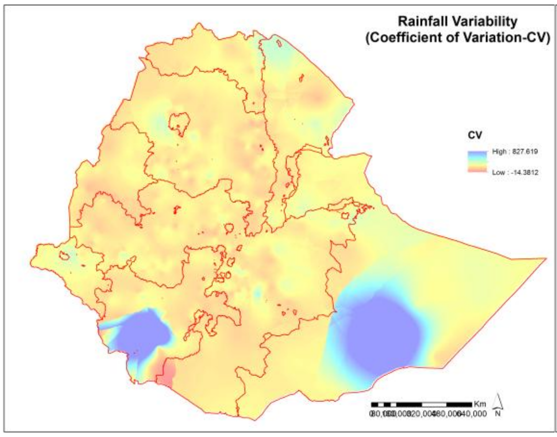
Calculating Standard deviation

1. To calculate standard deviation: (ArcTool Box→ Spatial Analyst tools→ Map Algebra →double click on ***Single output Map Algebra***.
2. In the ***Input raster or feature data (optional)*** click the drop down arrow and select all input data for the standard deviation calculation one by one.
3. In the ***Map Algebra Expression*** type STD and create an open bracket "("and from the ***Input raster or feature data (optional)*** select one layer at a time and drag and drop in the Map Algebra Expression field. Separate each layer by a comma and finally close the bracket.
4. Specify the output directory and name output as "STD_Rainfall" in the output raster
5. Click Ok on to calculate the STD

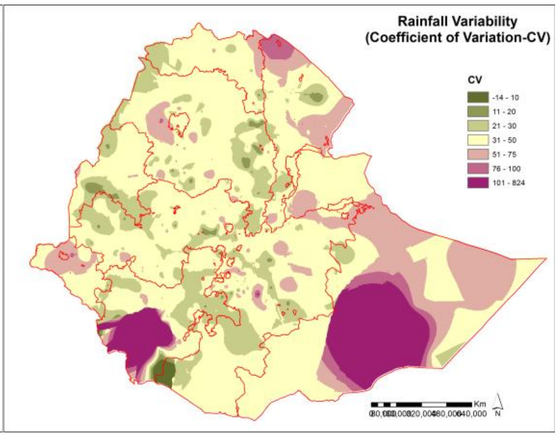
Calculating covariance or coefficient of variability

1. To calculate coefficient of variability: (ArcTool Box→ Spatial Analyst tools→ Map Algebra →double click on ***Single output Map Algebra.***
2. In the ***Input raster or feature data (optional)*** click the drop down arrow and select "MEAN_Rainfall" and "STD_Rainfall" for the CV calculation.
3. Since CV is calculated as STD divided by MEAN; therefore from the ***Input raster or feature data (optional)*** select, drag and drop the STD_rainfall and "MEAN_Rainfall" layers in ***Map Algebra Expression*** and put the division "/" between them.
4. Specify the output directory and name output as "CV_Rainfall" in the output raster
5. Click Ok on to calculate the CV on the raster calculator

6. Finally organize your map as shown below and give your answer for
    a. Where high rainfall variability is observed?
    b. What is the implication of high rainfall variability on agriculture?



Stretched values



Classified values

Use of zonal statistics function in ArcGIS

Overview

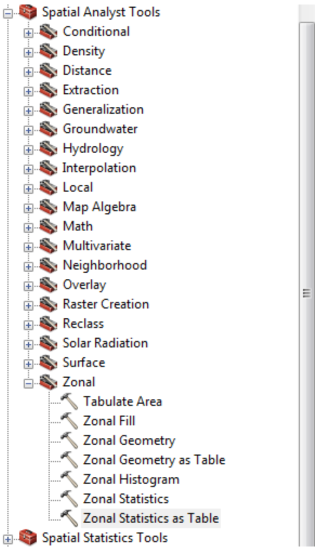
Zonal Statistics are a way of summarizing the information in a raster map layer using the boundaries of zones in a second map layer. For example, we might have a raster grid depicting the rainfall and have a second map layer depicting Woreda/Zone/Kebele boundaries. Zonal Statistics can be used to calculate the average rainfall in each of the administrative boundaries (woreda, kebele or zone). These boundaries can be stored in either raster or vector format - Zonal Statistics will work with either format for the zones.

Following parameters can be calculated (per operation, only one parameter can be calculated, to get several parameters it is necessary to save obtained table in a separate file):

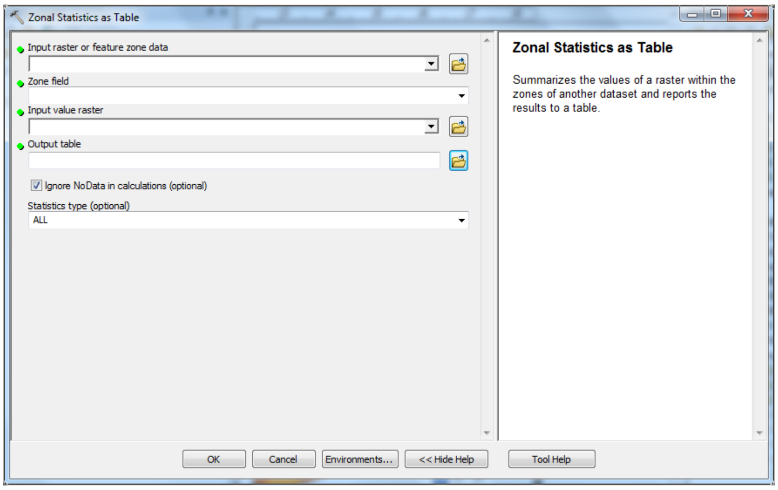
- sum - sum of all values
- min - minimum value of all values
- max - maximum value of all values
- count - number of value (number of pixels inside zone, excluding pixels with NODATA value)
- area - area covered by pixels with value different than NODATA
- range - values range
- std - standard deviation
- mean - average value (NODATA pixels, they are not used in calculation)
- median - median value (NODATA pixels, they are not used in calculation)
- majority
- minority
- variety

Using Zonal Statistics As Table

To use Zonal Statistics, make sure that you have Rf_clipped map layer open in ArcMap and another map layer with boundaries in it (woreda boundaries). You can calculate Zonal Statistics as follows:



1. In the *ArcToolBox*, choose *Spatial Analyst Tools*, then choose *Zonal* and then *Zonal Statistics as Table*. The difference between *Zonal Statistics* and *Zonal Statistics as Table* is that *Zonal Statistics* produces a map layer as output, whereas *Zonal Statistics as Table* produces a summary table. If you are working out mean, min, max, STD, rainfall for each woreda boundaries, *Zonal Statistics as Table* would produce a table showing mean, min, max, STD for each province boundaries. *Zonal Statistics* would produce the same information, but as a map display. In practice, *Zonal Statistics as Table* tends to get used a little more often.



2. Next to ***'input raster or feature zone data'***, you should enter the name of the map layer that has the zones – woreda/kebele/zone boundaries which is the shapefile.
3. The ***Zone field*** is the name of one of the fields in the table of attributes for your zone map layer. To summarize rainfall levels for different administrative boundaries, enter a W_NAME here, which is a unique name stored in it for each woreda.
4. Next to ***'input value raster'***, choose the information that you wish to summaries. Enter the name of the rainfall grid map layer here (Rf_clipped)
5. Next to ***'output table'***, you can enter in the name of a table where your results will be stored.
6. Check the Ignore NoData in calculations (optional)
7. Next '***Statistics type (optional)***' you can choose a specific type of statistics to be performed, left the default (All) as it is; but if you want only mean, you can choose MEAN.
8. If you choose OK, you should find that ArcGIS will generate a table of results for you.

9. From the table of content right click on the table and point to open (See below by regional rainfall for July)

Table

ZonalSt_shp2

OBJECTID *	R_NAME	ZONE_CODE	COUNT	AREA	MIN	MAX	RANGE	MEAN	STD	SUM	VARIETY	MAJORITY	MINORITY	MEDIAN
12	Tigray	12	803	51392000000	7	92	85	40.275219	16.384779	32341	78	29	7	37
2	Afar	2	1331	85184004000	12	214	202	86.580017	42.355373	115238	167	69	14	79
3	Amhara	3	2429	155456000000	29	270	241	155.8839	48.515057	378642	215	161	29	164
11	Special EA	11	47	3008000000	122	336	214	157.21277	60.034012	7389	22	125	123	133
10	Somali	10	5434	347775990000	84	311	227	168.59735	33.070126	916158	187	169	84	169
5	Dire Dawa	5	25	1600000000	179	215	36	195.67999	8.460355	4892	15	192	179	193
7	Harari	7	5	320000000	198	218	20	206.39999	7.088018	1032	4	208	198	208
9	SNNP	9	1674	107136000000	144	298	154	213.99821	25.883978	358233	125	222	151	215
6	Gambella	6	464	29696000000	188	281	93	227.39655	17.842283	105512	76	228	188	228
1	Addis Ababa	1	7	448000000	230	241	11	233.14285	4.12063	1632	4	231	238	231
8	Oromia	8	4662	298368010000	158	411	253	237.62869	35.415062	110782	187	233	160	236
4	Benishangul Gumuz	4	776	49664000000	185	358	173	255.22293	26.67189	198053	117	260	198	257

Lab 11 - Exercise Part 2: Using rainfall estimation (RFE) satellite data

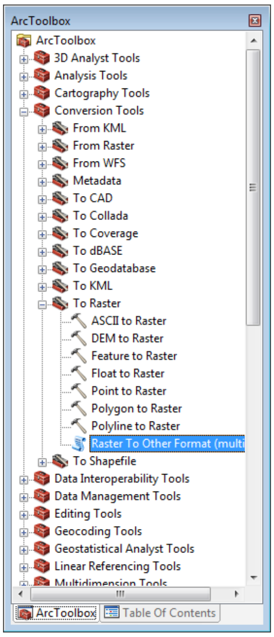
RFE is the rainfall estimate from satellite imageries, which is acquired by NOAA of NASA with 8km grid size around the world. RFE is a freely available dataset.

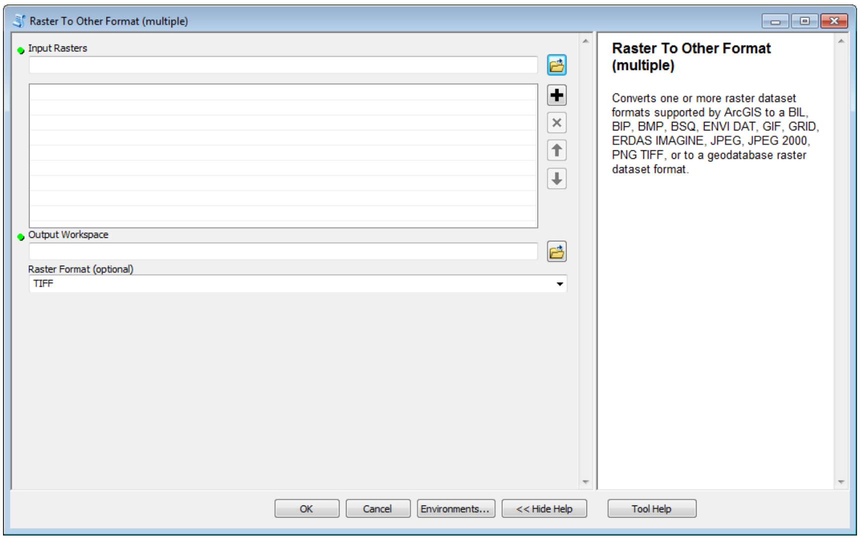
You can download the free Rainfall estimation data from the following link:

http://earlywarning.usgs.gov/fews/africa/web/datatheme.php

You will have dekadal datasets (10 days) images from NOAA and you have 3 datasets for one month. Hence, annually you can have 36 image files. The data format is .BIL (Band Interleaved by Line Format). This BIL format will be converted to ArcGIS format, and the dekadal images were summed up to develop monthly rainfall images. Finally Rainfall values can be mapped for the whole of specific areas; and also can be compared with weather station values. To do these, follow the following steps in ArcGIS.

1. After you acquired the required images as zipped files, you can open them as follows: ArcTool box→ conversion tool→ to Raster→ Raster to Other Format (multiple).

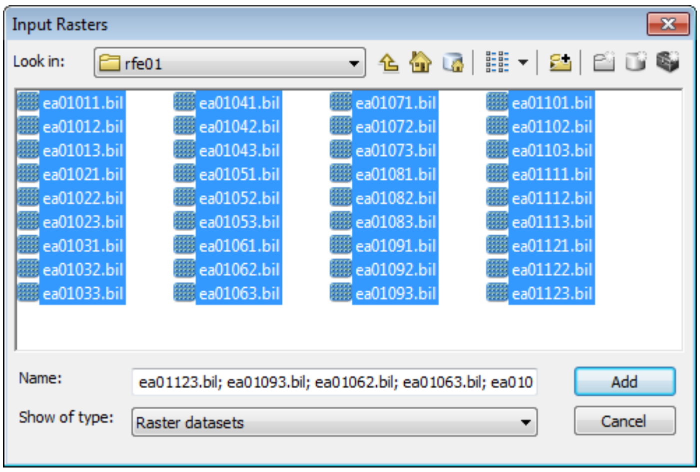




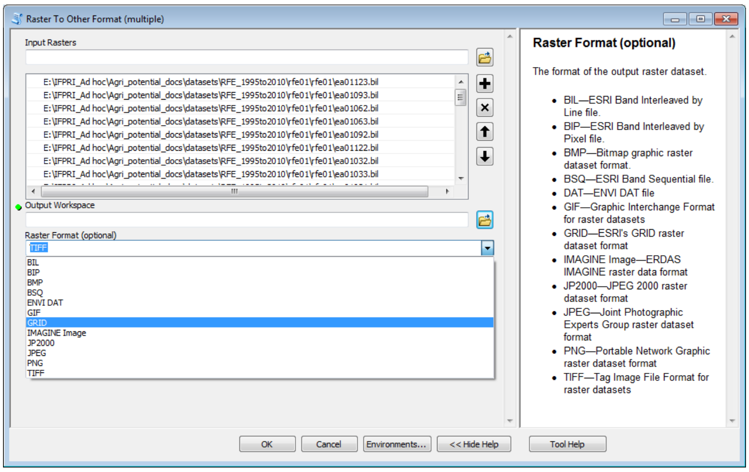
2. Unzip the file

3. Select the file of the BIL format to change to Grid format in ArcGIS (here the RFE data for 2001 in BIL format, which has 36 files for one year).

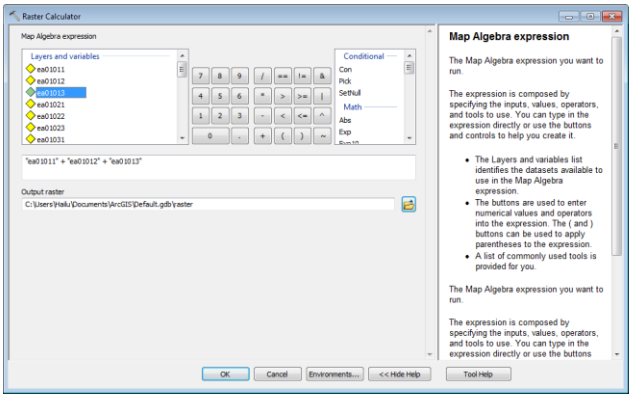
4. Click “Add”.



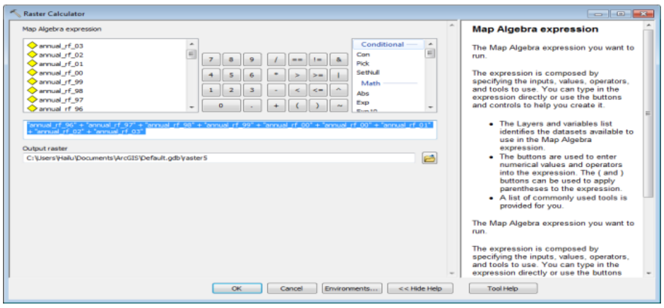
5. Specify the output directory
6. Choose the format type to “Grid”
7. Ok. Now, you have grid format of these data.



8. The data file name of ea01011 means east Africa 2001, January and first decadal (January 1-10).
9. ArcToolbox→spatial Analyst Tools → Map Algebra→ Raster calculator



10. Sum up 01011, 01012 & 01013, and give the name for output file Jan2001.
11. Do the same for other months of the same year and also other years.
12. Now, you have monthly Rainfall estimation for different years.
13. You can also sum up all month images to have annual Rainfall estimation of that year (s) as shown in the above tool.



14. Do the same for other years to have annual rainfall estimations.
15. If you need long term (>=10 years) mean annual rainfall estimation, we can add all the annual rainfall estimates and divided by the number of years. For example, if we have datasets from 2001 to 2010 for ten years, we can use raster calculator tool in ArcGIS and calculate mean rainfall estimation for the last 10 years. Using the formula:
16. **To calculate the Standard deviation follow the above steps!!!**

17. We can also do covariance (coefficient of variance) or rainfall variability across years and space. To do this, please follow the STEPS you used in the previous exercise.

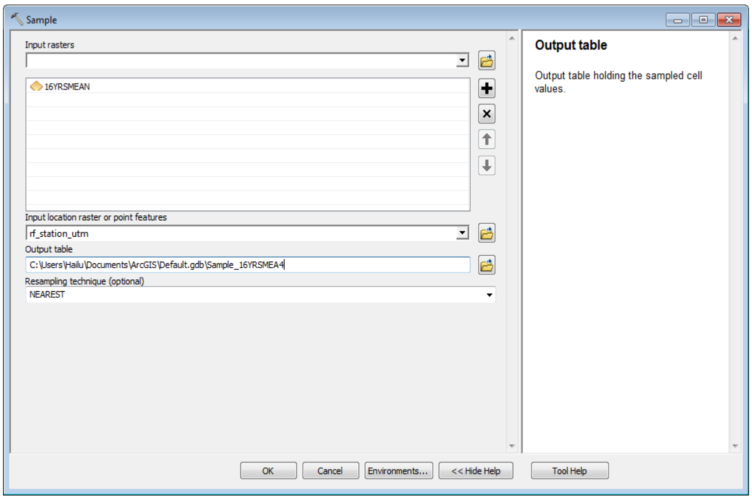
Covariance (CV) = STD / MEAN.

18. Now, you can compare the annual rainfall either for long term or for annual rainfall estimation from satellite data and weather station data. To do this, you can pick the points of RFE data by taking sample statistics from which weather stations are found.

19. Locate the mean annual rainfall from RFE in the input raster, and also locate Weather station points as input point feature, and specify the output table, and click OK.

20. Link the output table to the feature points by creating common Id (RowId for the new table, and new Id by calculating FID+1 for the point feature ID).

21. Compare the figures obtained from the RFE and weather stations; and reason out if there are differences on the same point.



Lab 12: Agriculture survey data integration

Exercise overview

You have received area by crop production information from the Central Statistics Agency of Ethiopia for the years 2007, 2008, 2009 & 2011. The data is in Microsoft Word format, and you need to integrate information on area covered by cereal into ArcMap to create a thematic map.

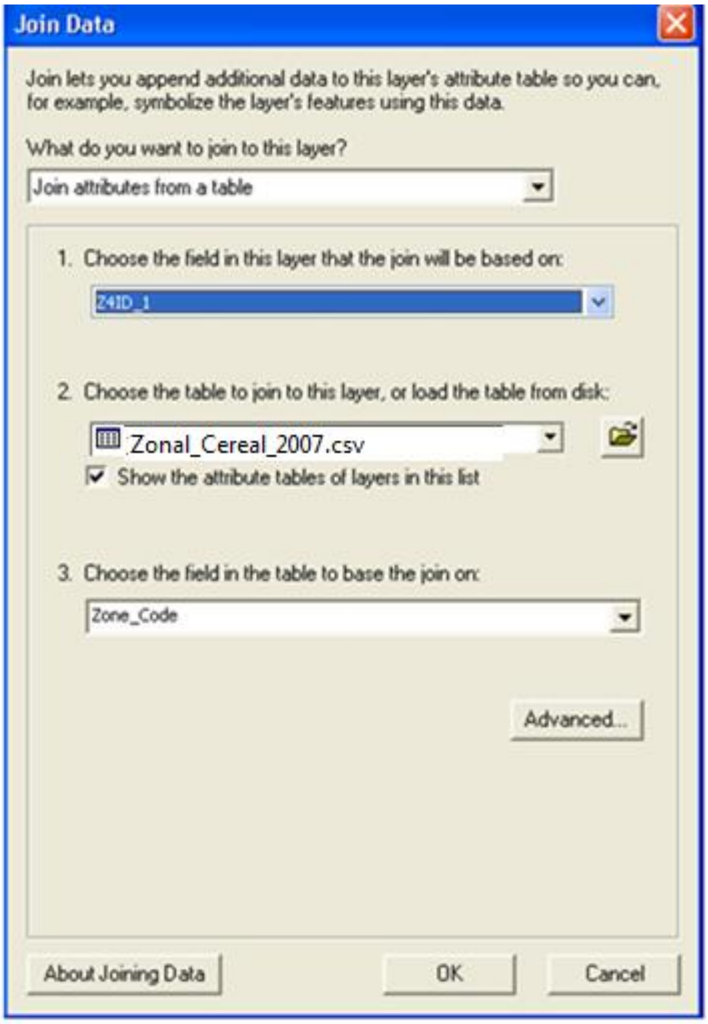
In order to do this you first need to extract the data of area coved by cereal and copy the same to Microsoft Excel. Then after you need to join this table to the attribute of a map table as discussed in Lab 07.

1. From Lab12 folder open the Word document titled "ProductionReport2006-07". This table contains data obtained from the 2006/07 Agricultural Sample Survey conducted by CSA. It lists crop type, number of holders, amount of area coved by corresponding crops, production in quintals, and yield of zones covered by the 2006/07 Agricultural Sample Survey. For this exercise your concern is only area coverage of cereals by zone.
2. Open an empty Excel sheet, write three headings "Zone_Code", "Zone_Name", "Cereal_Area" and save it as "Cereals" in Lab12 folder.
3. Starting from Table 4.1 extract and copy zone code and name (written at the top of each table). Also type the area covered by cereals which is indicated under the column heading "Area in hectare" in the "ProductionReport2006-07" to the Excel sheet you just created,.
4. Now you have an Excel sheet that lists the total area covered by cereals at zone level.

➔ **REMEMBER:** ArcMap doesn't allow space in field names, hence notice that the field names in your Excel are edited accordingly.

5. Open "Zonal_cereal_2007" excel sheet from your Lab12 folder.
6. Now, you need to join this information with the attribute of a map table.
7. To join your Excel table to the shapefile, as you practiced in Lab 07, start by saving the "Zonal_cereal_2007" table into a .csv file format so that it can be imported into ArcGIS (make sure none of your fields were truncated during the save). Name the .csv file "Zonal_cereal_2007".
8. Open a new session of ArcGIS and add the layer "Zone_Cereal_Area.shp" from your Lab12 folder.
9. Add your database file "zonal_cereal_2007.csv" from Lab12 folder.
10. Right-click on your "zonal_cereal_2007.csv" table to open it. Make sure that all field names carried through correctly, and that all data appears in working order. If it looks good, close the attribute table and move on!
11. The next step is to join the "zonal_cereal_2007.csv" table to the "Zone_Cereal_Area" layer so that you can utilize its spatial properties to visualize the area covered by cereals.

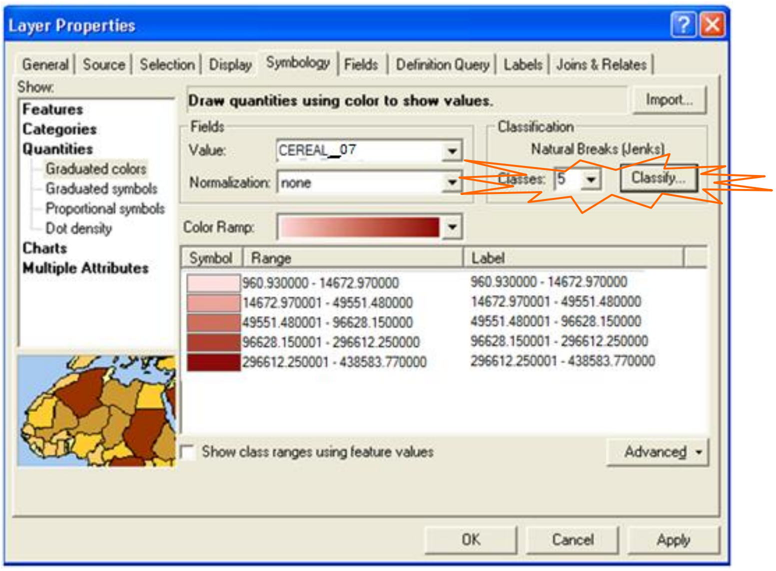
12. Right click on the "Zone_Cereal_Area" layer, Scroll to "Join & Relates" > "Join…" Make the following dropdown selections, and click OK.



13. When complete, open the attribute table of the "Zone_Cereal_Area" layer to make sure the join was successful. As you will see, some of the fields will say <null>, this is okay for this specific join because data were not collected for these specific zones.

14. Before continuing, save your map to your Lab12 folder as "Lab12_YOURNAME".

15. Double click on "Zone_Cereal_Area". The "Layer Properties" Window should now pop-up. If the tab "Symbology" is not selected, then select that tab.



16. In the box labeled “Show”, select the option “Quantities”, and click on the sub-option “Graduated Colors”.

17. In the drop down menu “Value:” scroll to the the field named “Cereal_07”: This stands for total area covered by cereals in the indicated zones in 2007.

18. Once this field has been selected, choose a color ramp that you like by clicking on the down arrow next to the color ramp. Now click, Apply, and OK.

19. Look at how the colors are distribued. The classification brackets chosen by ArcMap are based on their default statistical classification; “Natural Breaks”.

20. Reopen the “Layer Properties” dialog for the “Zone_Cereal_Area” layer. Return to the symbology tab. Under the Classification menu (top right-hand corner), Click the “Classify” button.

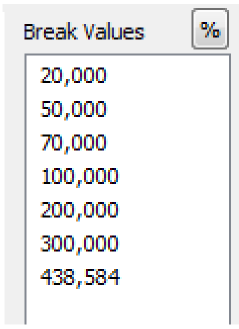
21. In the “Classificaton Wizard” you will see a histogram illustrating the data distrubution along the number line.

22. In the “Method” drop down list, you will see several classification alernatives to the “Natural Breaks” system. You also have the opportunity to change the number of classes that you use.

23. Experiment with the different classification schemes, and look at how they alter the classification breaks (blue lines) on the histogram data. By clicking OK on both wizards, you will see the effect of your class scheme changes on the map itself.

24. Return to the “Classification Wizard” screen. Choose first the “Quantiles” scheme; next change the number of classes to 7.

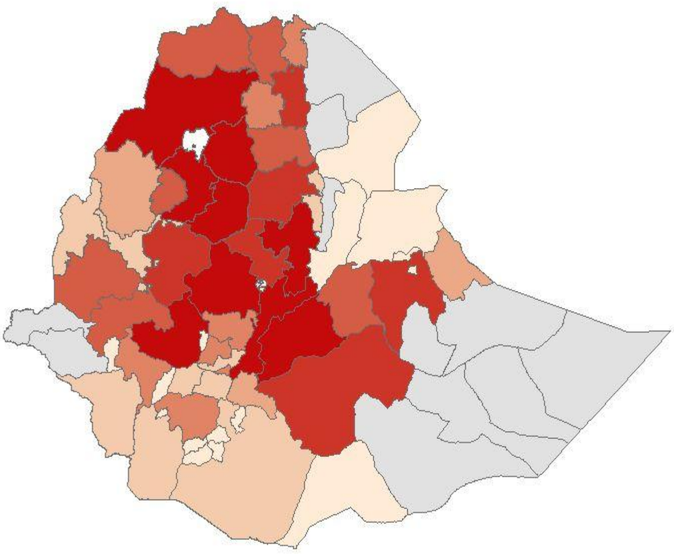
25. In the “Break Values” box to the right hand side of the wizard, set the break values to the following numbers, by simply typing over the existing values.

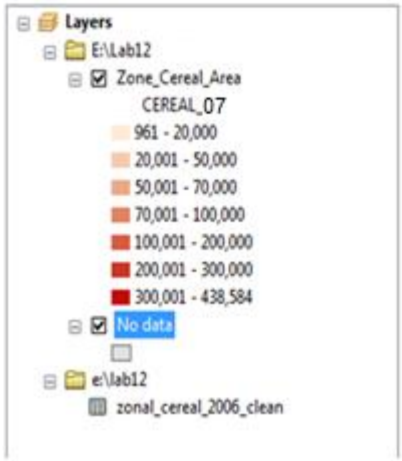


26. When done, click OK. Now you are back at the “Symbology” window. The “Label” side of the menu will reflect the changes that you make to the “Range” side, but you may also use text in your labels.

27. There are a number of Zones with “Null” values. (You should always account for these in your mapping). Given that we have only done a Table Join, and we haven’t exported our data as a new shapefile, our values reflect “Null”, we will show these by adding another layer.

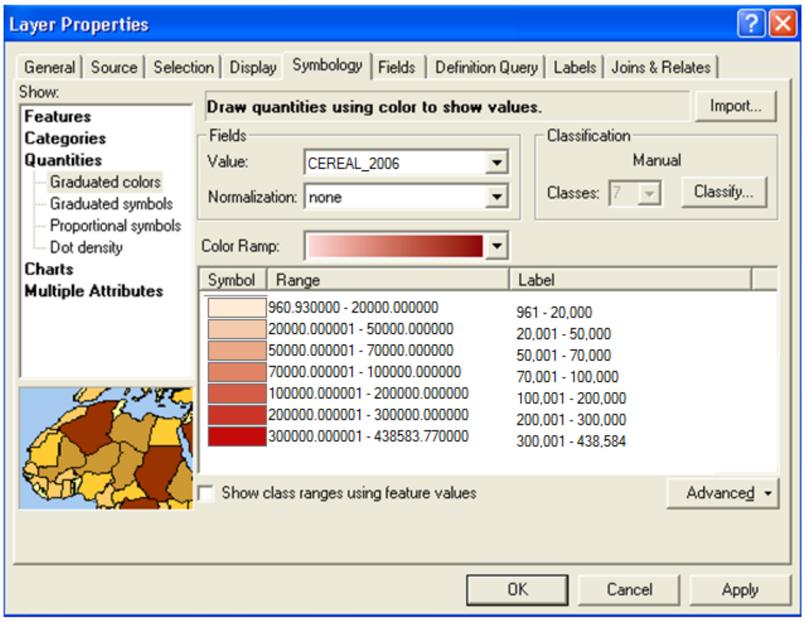
28. Copy and paste the “Zone_Cereal_Area” layer. Organize your “Table of Contents” so they look like the graphic below. Choose the underlying Zone_Cereal_Area layer to be a grey color to reflect “Null” data. Also, go to the table of contents and double click on this layer you just copied and rename it “Null”.



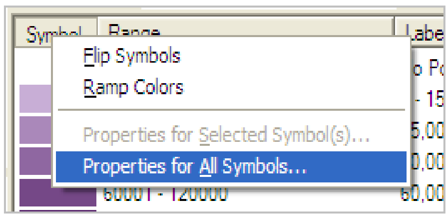


29. Now return back to your “Zone_Cereal_Area” layer with the table join. Double click on the layer and return to the Symbology tab.

30. Modify the label options with some additional text, and commas by clicking on the value under the Label column.



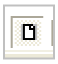
31. For a softer more subtle style, you will remove the boundaries from between the individual zones. Click on the word “Symbol” above the colored category symbols, and in the pop-up menu, choose “Properties for all symbols”. (see right)



32. In the “Symbol Selector” dialog, change the outline color to “No color”.

33. Click OK.

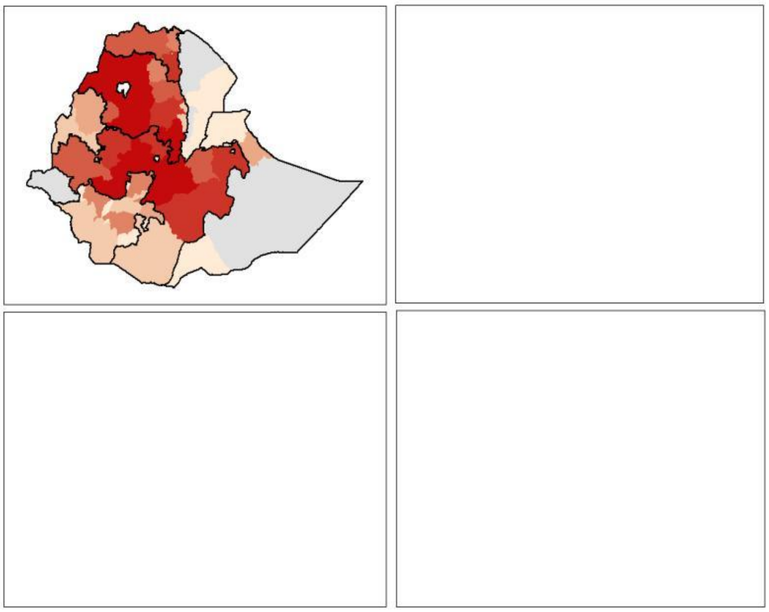
34. To distinguish the boundaries of the higher order administrative units, add the “Regions” layer from the Lab01 folder, and symbolize as hollow with an appropriate outline thickness. Remember how to do this? (hint: double click on colored box symbol below the layer name in the Table of Contents)

35. Switch to the layout view, by clicking on the layout symbol in the bottom left and corner of the map window. 

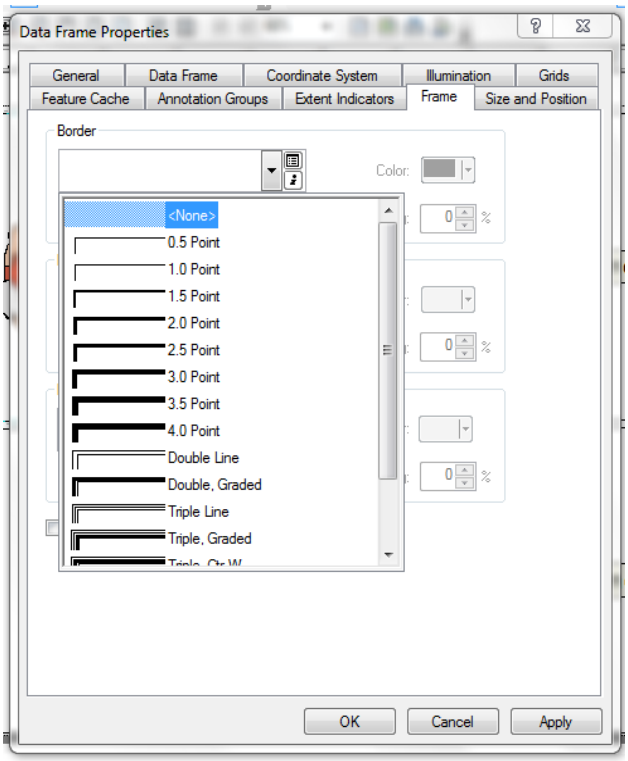
➔ **Save your map if you haven't recently!**

36. In the layout window, resize the map, so that it is roughly one-fourth the size of the page by clicking once on the map and then dragging one of the corner sizing squares to diminish the size of the map.

37. Insert three additional data frames. Resize all the frames so that your layout window should look something like this:



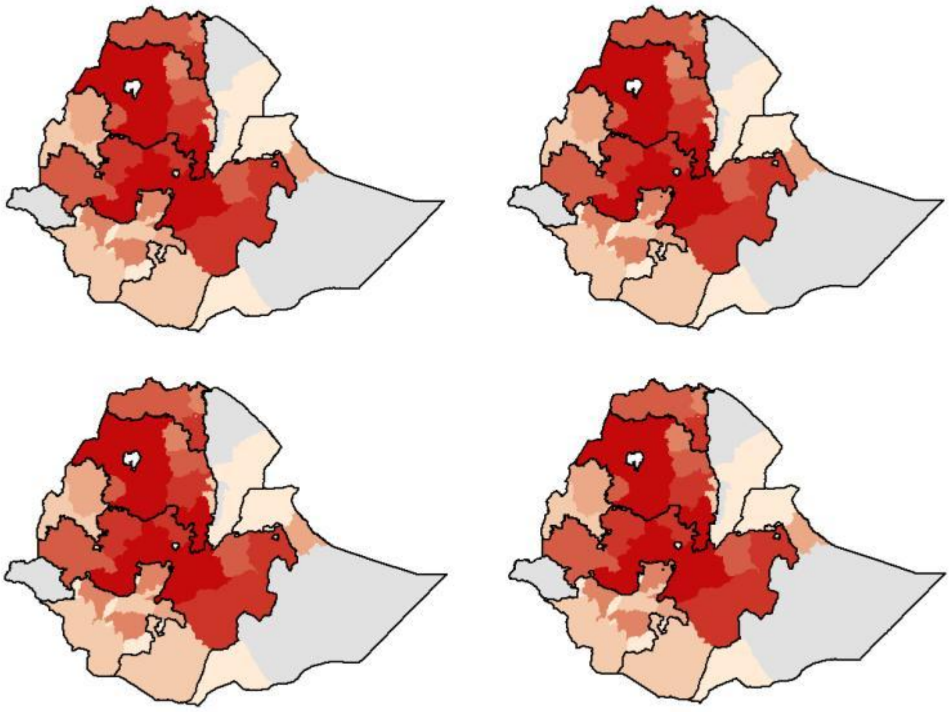
38. Right click on a data frame and go to Properties. Under the Frame tab, set Border to <None> as indicated in the figure below. Do this for all four data frames. This would remove the line around the data frame.



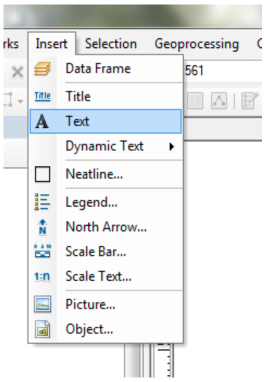
39. Copy all three layers (named “Region”, “Zone_Cereal_Area” and “No data”) under the top most data frame and paste into the remaining three data frames (namely,

- New Data Frame,
- New Data Frame 2, and
- New Data Frame 3.

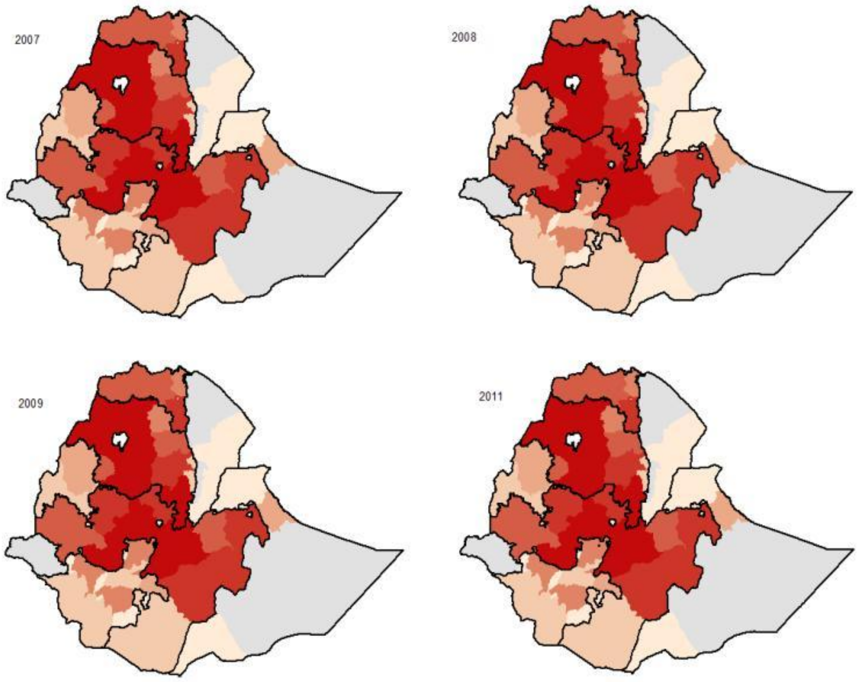
40. Now in all four data frames you should have map portraying the total area covered by cereals in 2007, shown in the figure below.



41. To show which year’s cereals coverage data is shown in which map window, go to “Insert” menu and click on “Text” as indicated beside.



42. A small box that have a “Text” inside will appear somewhere in your map. Overwrite “Text” with 2007 and place the text box in the top-left side of the first map window. Insert three additional “Text” boxes and write 2008, 2009, and 2011 inside these boxes. Place each boxes as indicated in the figure below. By looking at the years, now we can tell which map belongs to which year cereal coverage data.



43. What we want to do is to integrate and show four years (2007, 2008, 2009, & 2011) of zonal tabular data about the area covered by cereals. However, all four maps in the above figure show the total area covered by cereals in 2007. To compare change in area covered by cereals over the years, you need to join 2008, 2009, & 2011 zonal tabular data of area covered by cereals to “Zone_Cereal_Area” shapefile.

44. In the Lab12 folder you would find “Zonal_Cereal_2008”, “Zonal_Cereal_2009”, & “Zonal_Cereal_2011” Excel files. These Excels contain cereal coverage data for 2008, 2009, & 2011, respectively.

45. Following the above practiced steps you followed to join 2007 cereal area coverage data, save each of these Excels in .csv format and join each table with “Zone_Cereal_Area” in each of the three data frames.
    - “Zonal_Cereal_2008.csv” should be joined with “Zone_Cereal_Area” beside 2008;
    - “Zonal_Cereal_2009.csv” should be joined with “Zone_Cereal_Area” beside 2009;
    - “Zonal_Cereal_2011.csv” should be joined with “Zone_Cereal_Area” beside 2011.

46. Using the “Insert” menu, add a legend, and other appropriate cartographic elements to your map (North Arrow, and simple scale bar).

47. Based on your expertise, did the maps show recognizable patterns? Did the zones you know to have large cereal coverage stand out in these maps? What about zones you know to have low area covered by cereals? Are there zones for which you think the data needs re-check?