

Utilizing GIS Data Workshop Series

Distribution Analysis with Population Data

GIS Resources:

GIS Listserv: <https://listserv.miami.edu/scripts/wa.exe?SUBED1=gis&A=1>

UM Library GIS Guide: <http://sp.library.miami.edu/subjects/gis/>

## Scenario:

World population has surpassed 7 billion people. Where do they all live? You are a researcher looking for current population distribution trends in the world at the country and lower levels. You want to find concentrations of urban and rural populations as these are two common types of population distributions. Finally, you want to look more closely at a specific part of the world to see how population dynamics affect public health.

## Outcomes:

This workshop will show you how to find current world population data formatted as a table from an Internet source and join it to a vector country boundary file so distribution of the population data can be mapped. The workshop will also cover working with raster (gridded cells of data) population data to investigate population distribution in more detail.

## Skills Covered:

Internet: Data searching, tabular and raster data download/extraction, metadata examination.

Excel: Tabular data manipulation in Excel, formatting table for use in ArcGIS.

ArcGIS: Table joining with key field, table editing, metadata examination, data dissolve, create new fields with field calculator, symbology (stretch, chart, single) , float to integer raster conversion, zonal statistics with majority and variety, extract by mask, subset, select by location, cost distance, near analysis, raster calculator, raster to polygon, change coordinate systems, and cartography.

## Download Tutorial Data for this Workshop:

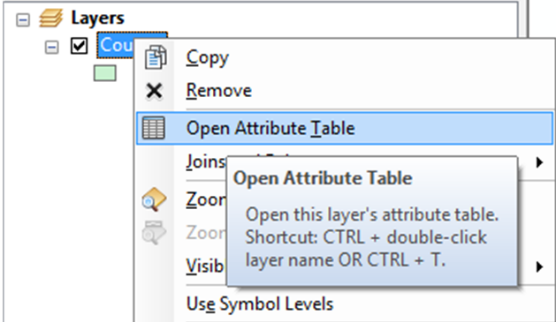
1. Go to Subjects Plus guide: <http://sp.library.miami.edu/subjects/gis>
2. Download the Dataset for the **Population Distribution** Workshop (Right-click> Save Target As> Save to C:\temp)

## Extract the Data to the c:\Temp folder

1. Browse to the c:\Temp folder, where you saved the Data
2. Right-click on the Population.zip and select Extract All…
3. Accept all defaults to extract the data file to C:\Temp

**Distribution of World Urban and Rural Population**

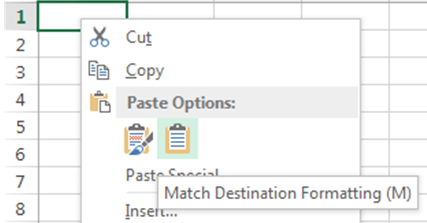
**Open Countries file in ArcMap and examine table**

* ****Open a blank map document. Add the Countries layer and open the attribute table. A limited amount of attribute data limits the information you can communicate through symbology or use to calculate other data or reveal other information. None of the attributes contain urban and rural population; we will need to find this data in an external table and join it to this table using a common field (country in unique identifier).
* However, there appear to be two separate country fields. Examine CountryAff field- what does that mean?
* Open ArcCatalog  and examine metadata. Right click Countries>Item Description. Only brief metadata displayed. Open ArcCatalog program from Start menu, go to Customize>ArcCatalog Options>metadata tab, then select North American ISO profile, refresh. Now should have more detailed metadata.
* Open ArcCatalog back in ArcMap and examine metadata (right click Countries>Item Description. Scroll down to Fields section and find CountryAff. It explains that “*The country name if there is no affiliated sovereign country or the name of the affiliated sovereign country*.” Keep that in mind when we do the table join later- we will eventually use the country field because dissolving on CountryAff will give us less unique countries.

**Acquire and manipulate data in Excel**

* Search Google for “world population” and select the first result from WorldOMeters.
* On Worldometers website, click Population, then select population by country.

<http://www.worldometers.info/world-population/population-by-country/>. The table only contains urban population attribute, but it’s a start.

* Highlight the entire table and right-click and click Copy.
* Open a blank Excel Workbook. In the top left cell, right-click and hit Match Destination Formatting under the Paste Options.
* Fix headers. ArcGIS will only accept one header row with field names with no spaces and only 10 characters long. They can’t start with a number or contain special characters either. Can add aliases later.
* Make the following header changes:

Population (2014) to Pop\_2014

1 Year change to One\_Yr\_Chg

Population Change to Pop\_Chg

Migrants (net) to Mgrnts\_Net

Median Age to Med\_Age

Aged 60 + to Aged\_60\_Ov

Fertility Rate to Fert\_Rate

Area (Km2) to Area\_KM2

Density (P/Km2) to Dens\_PKm2

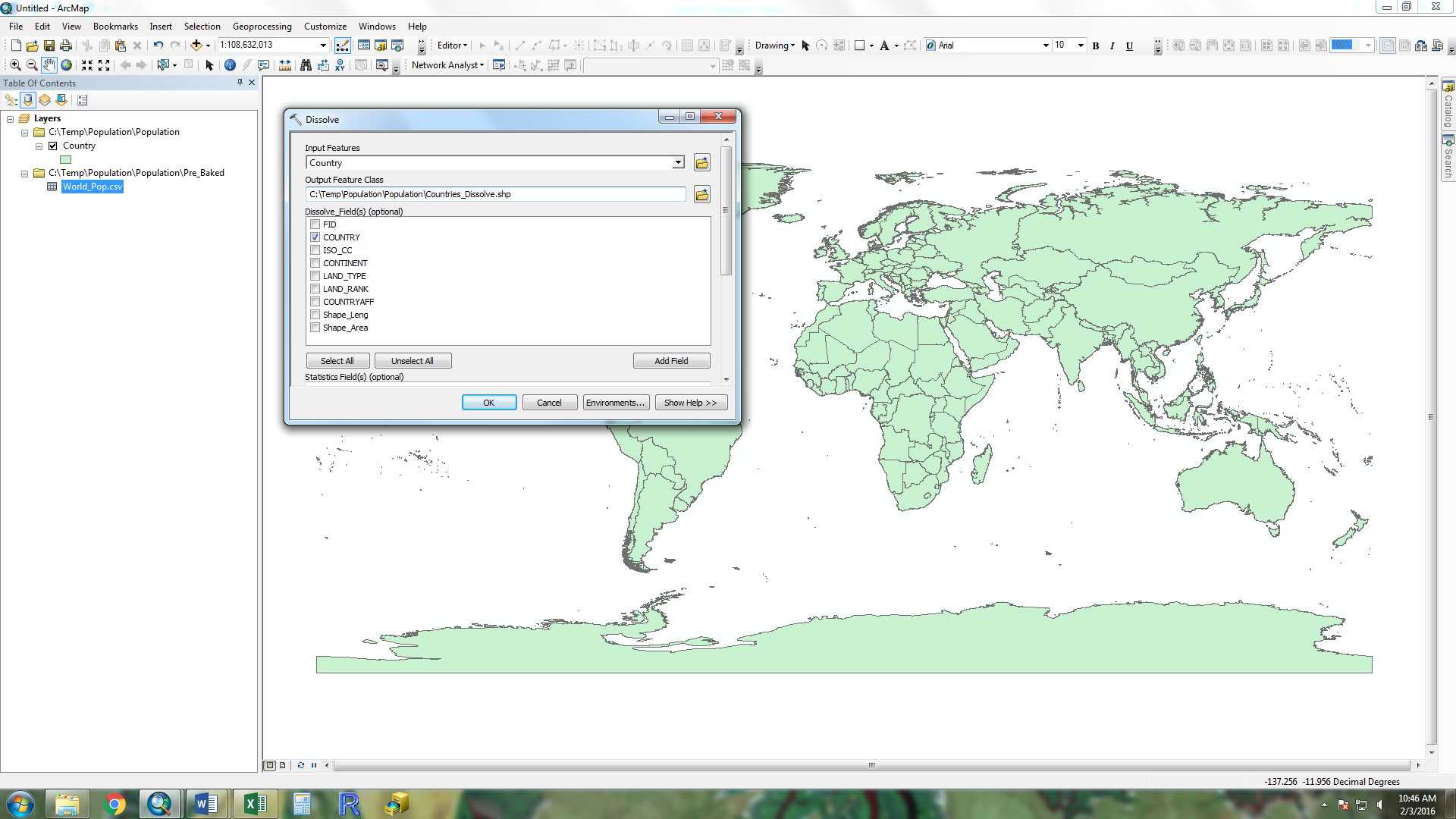
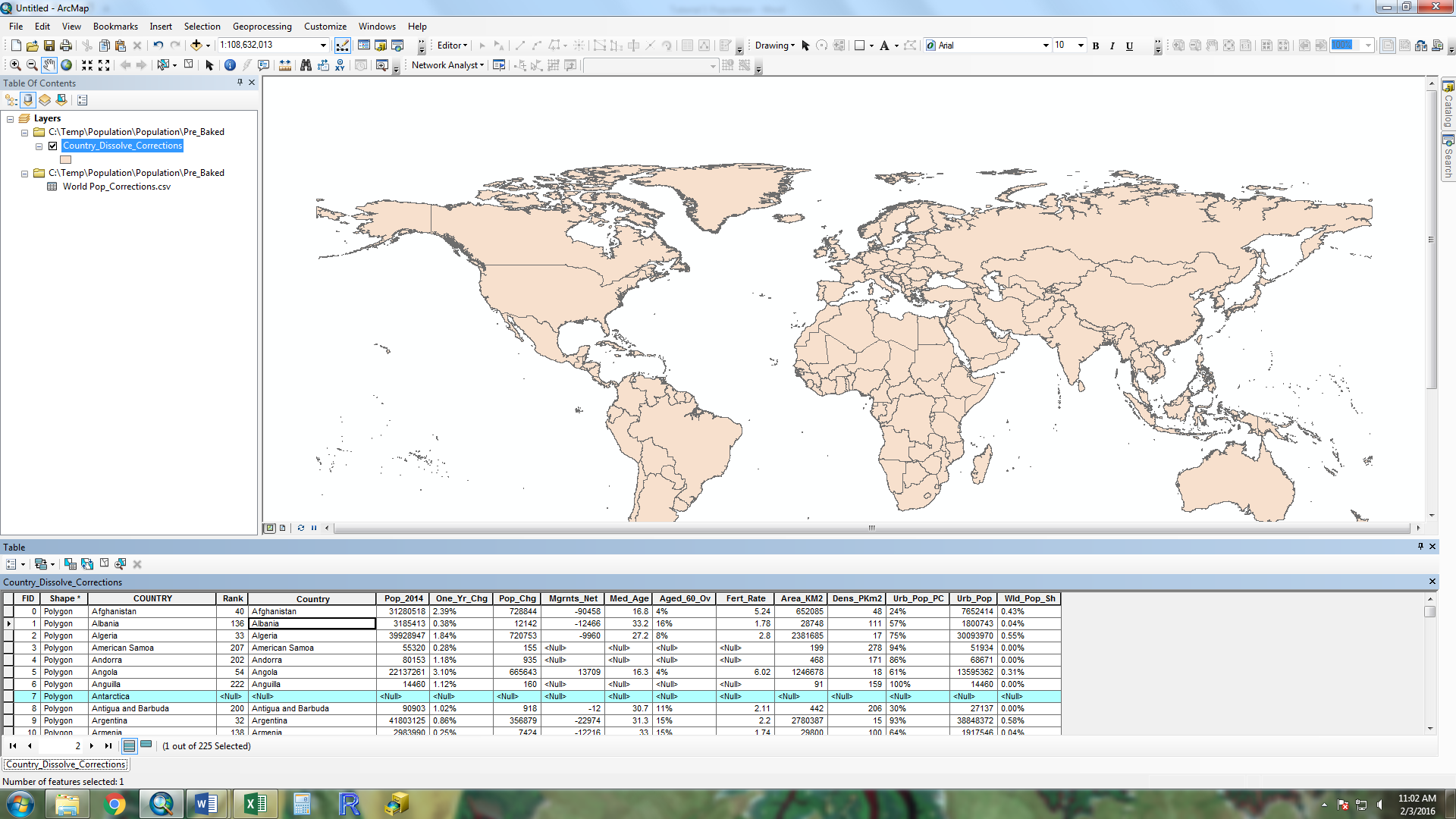
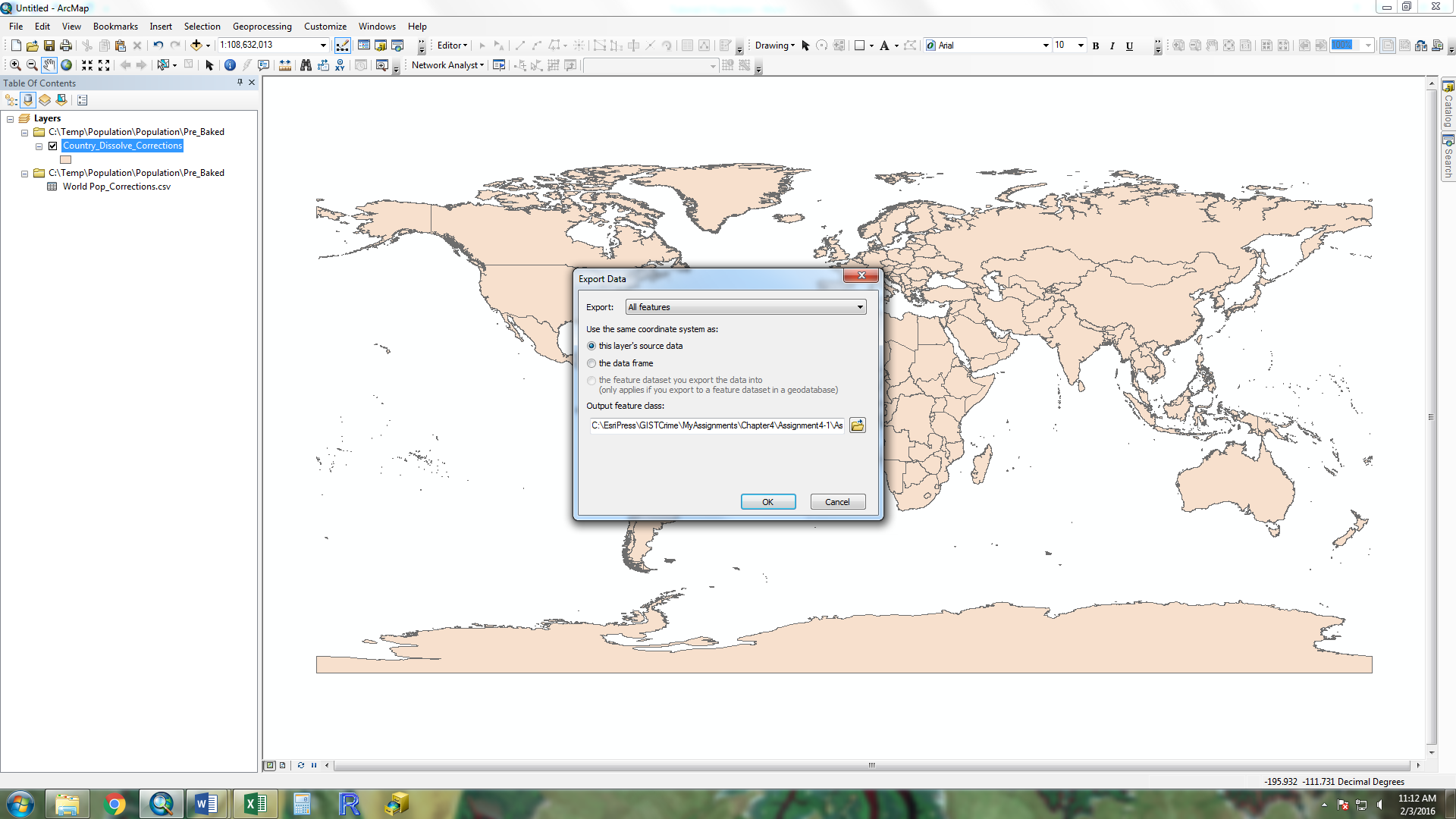
Urban Pop % to Urb\_Pop\_PC

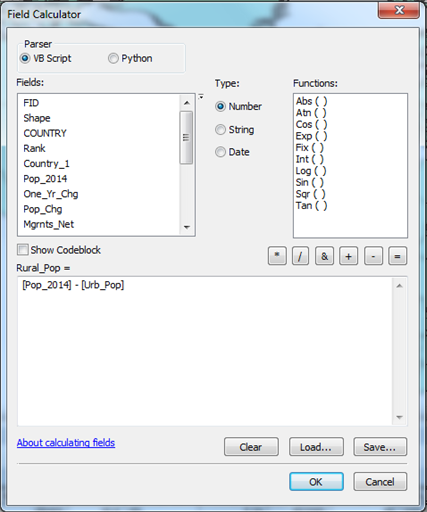
Urban Population to Urb\_Pop

Share of World Pop to Wld\_Pop\_Sh

* Delete extraneous second header row. Save file first as Excel workbook, then as a Comma Separated Value file (.csv) in the World\_Pop folder. ArcMap will read Excel, but csv is the most reliable format. If you were not able to keep up, there is a pre-baked Excel and CSV in the Pre-Baked sub folder of the World\_Pop directory.

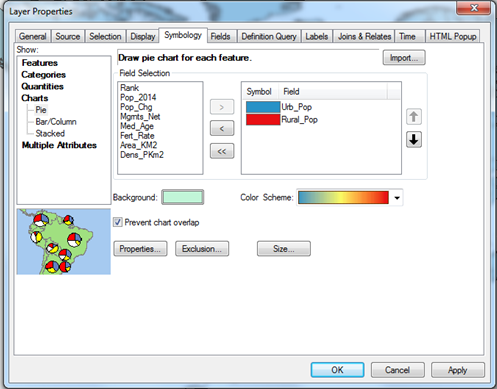
**Add World data to Map and Join to Countries Layer**

* Add World\_Pop.csv to ArcMap.
* Open the attribute table for the country layer again and examine the number of records representing the countries in the Country layer- there are 669 records- why? Because there are some countries represented by more than one polygons- like for islands.
* Open the table for the World\_Pop.csv table. There are 232 rows representing countries in this table. We need to fix the Country file so we get one country per row in the table.
* Search for and open the Dissolve (Data Management) tool. Select Country as input feature, call the output Countries\_Dissolve (save in Part 2 directory) and select Country as the dissolve field. Once the dialog box looks like the box on the right, click OK. Add Countries\_Dissolve to the map.
* Do a table join of the World\_Pop.csv file to the Country\_Dissolve layer by right-clicking on the Countries layer and going to Joins and Relates > Joins. Input Country for Fields 1 and 3 and click Validate Join. There is a discrepancy of about 40 countries that did not join. We have to find these errors and correct them in either the World\_Pop.csv file or the Country\_dissolve layer table (or both). As this will take too long for this workshop, let’s add the Pre\_Baked files I have already corrected: Add World\_Pop\_Corrections.csv and Country\_Dissolve\_Corrections layer from the World\_Pop>Pre\_Baked directory. You can remove the other layers.
* Perform the table join again on the Country\_Dissolve\_Corrections layer with the World\_Pop\_Corrections.csv file using country as the join field. Restore default column widths in the table menu if necessary by clicking Table Options > Restore Default Column Widths. Query second country field for null value- should only get Antarctica as the only country that didn’t match.
* Deselect Antarctica ( ) and re-save layer as Country\_Joined (join is not preserved in layer- just in the map document). This is done by right-clicking on the layer in the Table of Contents and clicking Data > Export Data. Click on the folder symbol  , name the file and change the Save as Type to Shapefile. File naming convention important when you are creating may different iterations of a layer.

**Calculate new data fields in Table**

* Next we will use the field calculator to calculate the rural population for each country.
* Open table for the Country\_Joined layer.
* Click Add Field in the Table Options menu. Call the field “Rural\_Pop” and make it a long integer.
* Right click the Rural\_Pop field, select “field calculator” and enter the following: [Pop\_2014] - [Urb\_Pop].
* Re-Save the layer as Country\_Final.

**Symbolize Urban and Rural Population**

* Right click on the Country\_Final layer and open the layer properties for Country\_Final and select the symbology tab.
* Choose Charts>Pie for the type of symbology and select Urb\_Pop and Rural\_Pop as the fields (make them two distinguishable colors). Click apply- the pie charts take up too much space on the map. Click Size and change the pie size to 10. Make two different maps for each hemisphere to see data clearly.

**Raster Analysis of Population Data**

This use of vector data form the WorldOMeters web site joined to existing polygons representing countries is one way to quantify and visualize population data. Now we will explore a very different type of population data called raster that reveals a little more detail in the distribution of population beyond the country level.

**Obtain Raster Data from CIESIN**

* In a web browser, navigate to the Socioeconomic Data and Applications Center (SEDAC) web site hosted by CEISEN at Columbia University (<http://sedac.ciesin.columbia.edu> ).
* Navigate to DATA>Data Sets and search for “population count grid future estimates” and select the download link to the “Population Count Grid Future Estimates, v3 (2005, 2010, 2015)” data set.
* Make sure the following parameters are set and download the data set (gl\_gpwfe\_pcount\_15\_wrk\_25.zip).

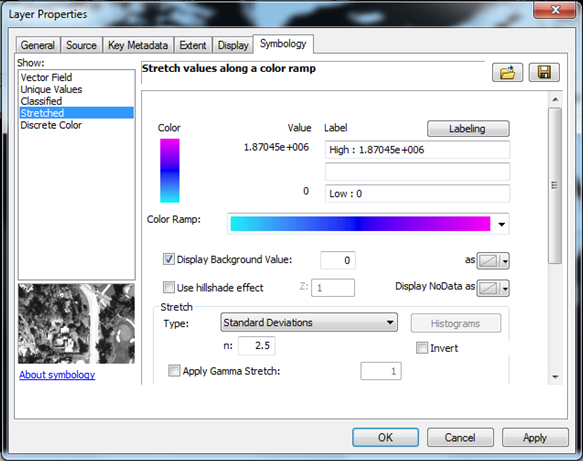
**Geography:** Region>Global

**Data Set:** Population Count Grid Future

**Data Attributes:** Format: Grid, Resolution: 2.5’, Year: 2015

* Unzip the gl\_gpwfe\_pcount\_15\_wrk\_25.zip file and extract it to your C:/Temp/Population folder. This should produce a folder titled “glfecount15”. Examine the contents to see how a grid file is organized.

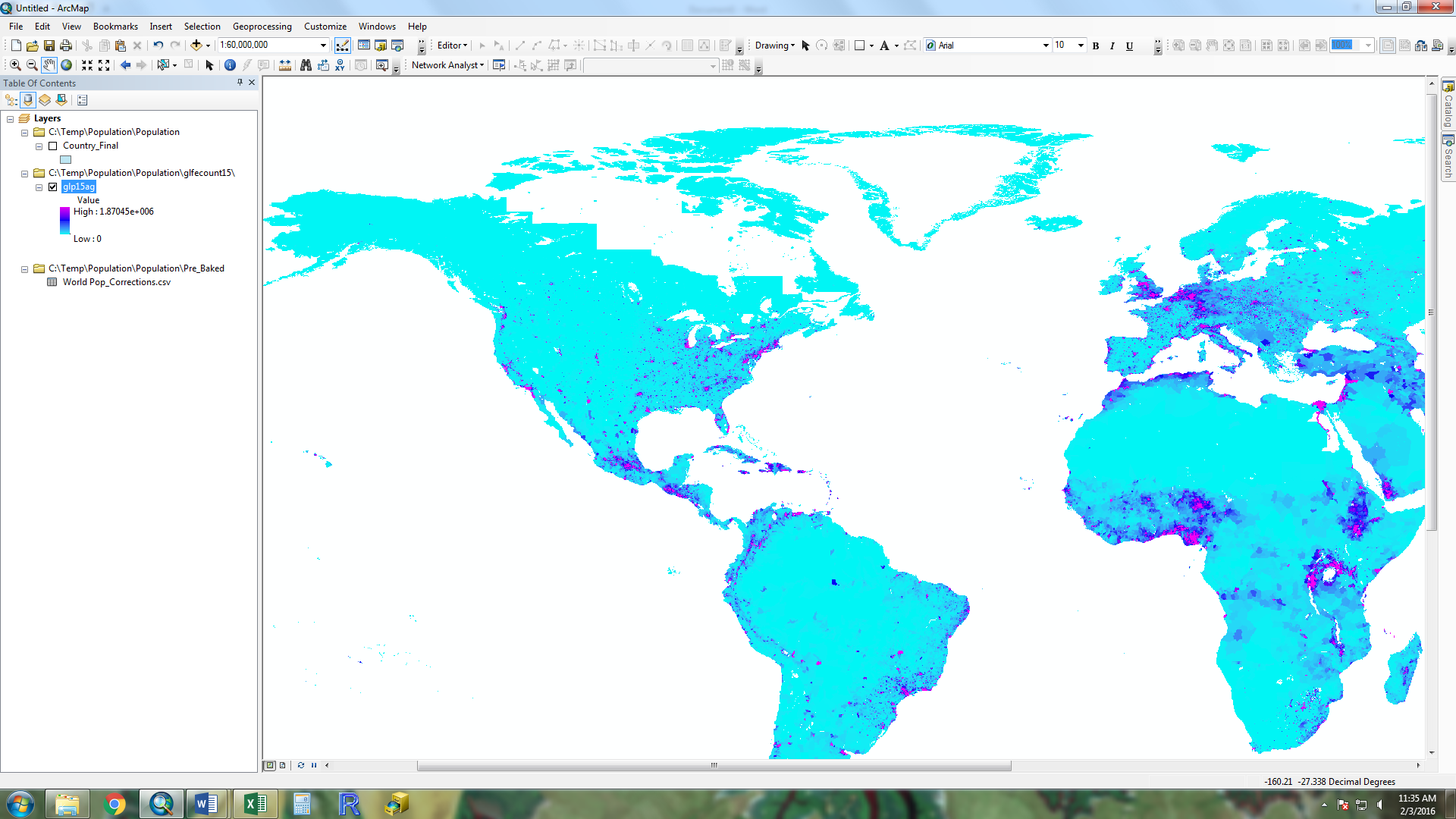
**Symbolize the Grid Population Data**

* Add the grid file to the Population.mxd project in ArcMap software (File>Add Data>glp15ag). Click yes if it asks to build pyramids. Remove the fill from the Country layer (just outline) and make sure it is above the grid file in the Table of Contents.
* Right click the glp15ag grid file in the Table of Contents and select the symbology tab.
* Keep the symbology on stretched (this is the default), but change the color ramp to light blue to dark blue to violet. Check the Display Background Value box, set to zero and select no color. Click OK. Turn off the country\_Final layer.

**Examine the Population Distribution in the Grid**

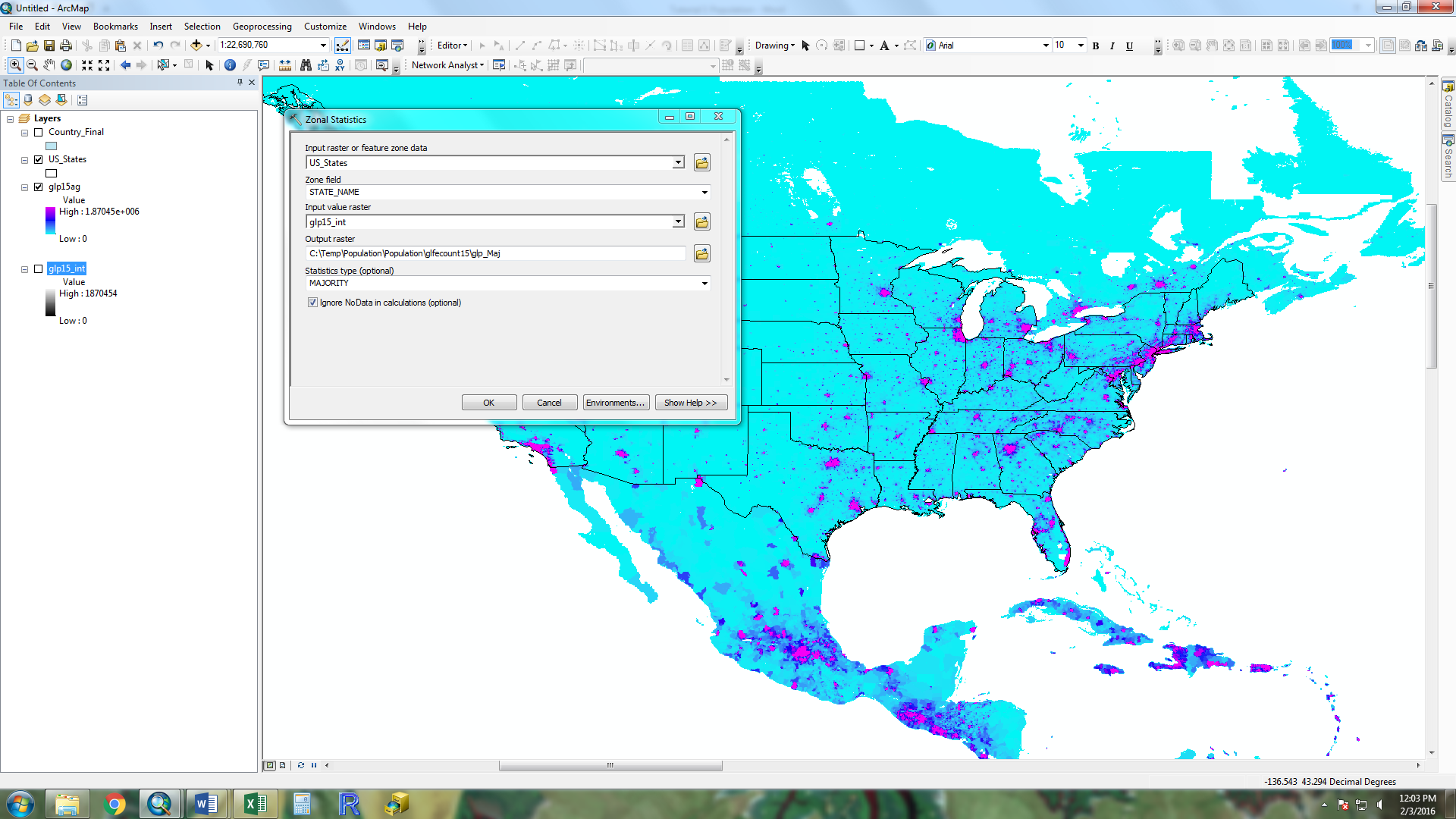
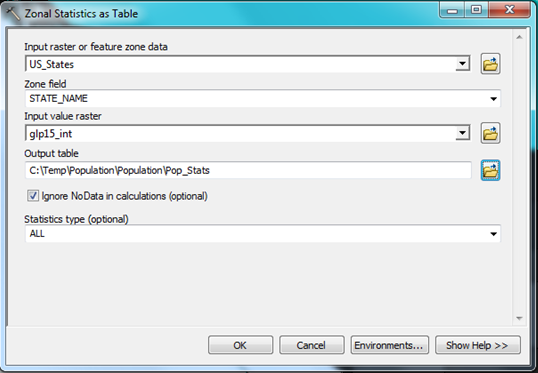
* Zoom in to about 1: 40,000,000 scale and examine the distribution of violet cells (representing the most populated areas) versus the light blue (least populous areas). This data set gives much more detail on how world population is distributed compared to the vector data derived from WorldOMeters, which was summarized at the country level, whereas the grid is divided up into geographic units that are 2.5 minutes.
* What can the cell distribution tell you about the population in Egypt, India, Yemen, the Caribbean island of Hispaniola that you were not able to determine from the WorldOMeters data?

**Examine the Grid Resolution**

* Zoom into about 1:100,000 scale to an area of the grid near the equator with a variety of different shades (near Lake Victoria in Africa is a good place). Use zoom tool or just type in 100,000 in the scale box.
* Select the measure tool  and set distance to miles and measure a grid cell height and width. Both should be approximately 2.9 miles, giving a resolution of about 8 square miles per grid cell.
* Now measure a grid cell near one of the poles (the tip of Argentina near the South Pole is a good place). It also has a measurement of about 2.9 miles for the height of the cell, but the width of the cell is about 1.7 miles, giving a resolution of about 5 square miles.
* Examine the metadata from the CIESIN web site where the grid was downloaded. It indicates and average input resolution of about 18 square kilometers, which is equal to about 7 square miles.
* Add the world30 shapefile to the data frame from the C:/Temp/Population directory in ArcCatalog. Change the projection from Geographic to Mollweide. Examine how the lines of latitude are equidistant and the lines of longitude converge at the poles. This coordinate system arrangement is the reason for different

**Compare USA States Population Distribution Through Zonal Statistics**

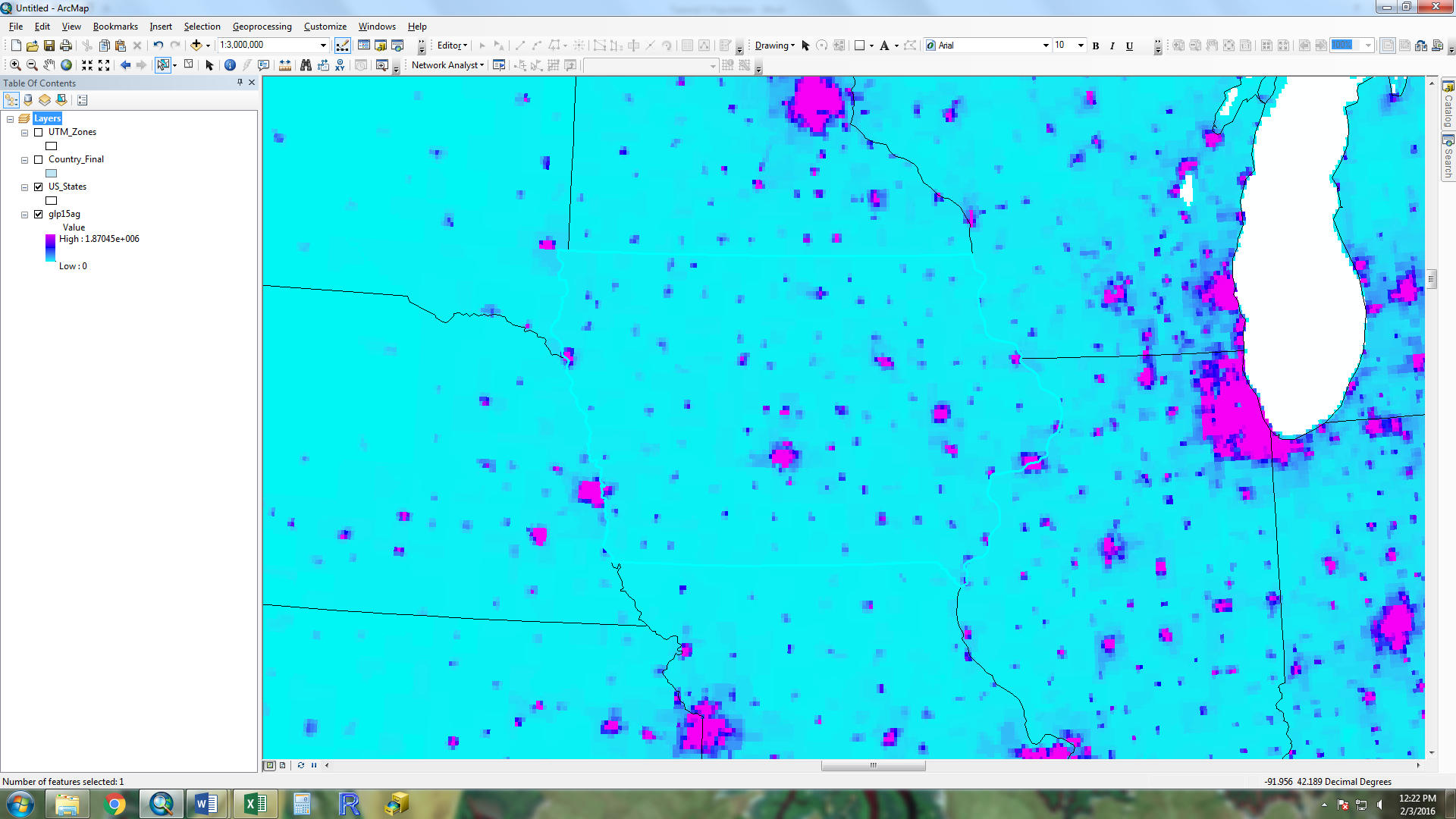
Another way to examine grid population data other than through a simple stretched symbology is to perform statistical calculations and rankings based on other geographies such as States. Our goal at this point is to move from the world scale and look at the population distribution of a particular country- the USA. Our goal is to examine how each state compares to all the other in its population distribution.

* Add the US\_States shapefile from ArcCatalog (drag and drop into data frame from Catalog directory). Symbolize it with a hollow fill by clicking on the colored box under the layer.
* In order to use the zonal statistics majority and variety, the grid values need to be integers rather than floating point values, so we will first convert them with Int tool. In the Search tab, type Int and open the Int(spatial analyst) tool.
* In the Int dialog box, select glp15ag as the Input raster or constant value and save the output as C:/Temp/Population/glp15\_Int and click OK.
* Now we can use this integer grid in the zonal statistic tool. In the Search tab, search for “zonal statistics” and open this tool (zonal statistics spatial analyst).
* In the Zonal Statistics dialog box, enter US\_States as the Input raster or feature zone data, State\_Name as the Zone field, glp15\_Int as the Input value raster, Majority as the statistics type, keep NoData checked, and save as C:/Temp/Population/glp15\_Maj. This identifies the value that occurs the most often of all cells within a particular zone (state). What can that tell you about population distribution among the different states? Which population value and states cover the most area? Which states have the highest homogenous population distribution (use table with sort function)?
* Run the Zonal Statistics tool again with every parameter the same, except change the statistics type to Variety and call the output file “glp15\_var.” This identifies the number of unique values for cells with in a zone (state). What can this tell you about the population distribution among the different states? Which ones are the most heterogeneous/diverse in their population distribution? Does Pennsylvania have more varieties in its population distribution than it does in its Heinz ketchup?
* Go back to the search tab and run the Zonal Statistics as Table (spatial analyst) tool. Choose all types of statistics, save the output table as C:/Temp/Population/Pop\_Stats, join it to the US\_States shapefile, and resave as US\_States\_Stats.

**Population Distribution Effects on Public Health in Iowa**

Next we are going to examine population distribution specifically in the state of Iowa to see if we can determine if the population distribution can explain its unique distribution of hospitals compared to other states and determine the furthest distance anyone would have to travel to the nearest hospital.

**Re-Project Map to Coordinate System measured in Meters**

* In order to do some calculations with raster data, the data needs to be in a coordinate system with a linear unit of measure (such as meters) rather than angles, as the current un-projected WGS 84 (examine Spatial Reference in Raster Dataset Properties in ArcCatalog). We are going to change the coordinate system to UTM, which is measured in meters. In order to determine the UTM zone, add the UTM\_Zones shapefile to the Table of Contents from ArcCatalog. Symbolize the layer with a hollow fill like the US\_States layer and label with the Zone field by going to the Label tab under the Layer Properties. You will see that Iowa falls within zone 15 North.
* Right click inside the Data Frame and change the coordinate system in the Data Frame Properties to UTM Zone 15 (Projected Coordinate Systems>UTM>North America>NAD 83 (2011) UTM Zone 15N). Click yes in the next box. You can turn off the UTM Zones layer now.
* Next we are going to re-save the glp15ag grid with the UTM projection clipped to the State of Iowa. Use the Select Features button  , select the state, then right click US\_States in the Table of Contents > Data > Export Data. Save the subsetted Iowa as Iowa.shp. Turn off US\_States and zoom to the Iowa shapefile. Make the Iowa layer hollow fill. Turn on the glp15ag file.
* Search for and open the Extract by Mask tool. In the Extract by Mask dialog box, select glp15ag and the Input raster, Iowa as the Input raster or feature mask data, and save as glp15\_IA. Change the Output Coordinates in the Environmental Settings to UTM Zone 15 N. Click OK.

**Euclidean Distance to Hospitals**

* Add the US\_Hospitals shapefile to the data frame. Turn on US\_States and examine the distribution of hospitals across the US. How is the distribution pattern of hospitals different in Iowa compared to other states? Iowa has a large rural population as you can see from the population distribution in the grid file. About a third of the state lives in rural areas and it is the 12th largest rural population compared with other states.
* We want to determine what the furthest distance any person in the state of Iowa would have to travel to the nearest hospital. We will use Euclidean Distance to determine this.
* On the Menu bar go to Selection > Select by Location and put US\_Hospitals as the Target Layer and Iowa as the Source Layer. Change the Spatial Selection Method to “are within the source layer feature” and click OK. Now use the Export Data technique learned above to save the selection as Iowa\_Hospitals.
* Search for and open the Euclidean Distance (spatial analyst) tool in ArcMap.
* Enter US\_Hospitals and the Input raster or feature source data, keep the defaults, and change the output to C:/Temp/Population/Hosp\_Dist. In Environmental Settings, change the output coordinates to UTM zone 15. In Raster Analysis, select Iowa as the mask and cell size the same as glp15\_IA.
* With the Hosp\_Dist layer, you can visualize the areas of Iowa that are remote in relation to hospitals and can see from the symbolized values that the furthest distance to a hospital for anyone in Iowa is about 26 kilometers or 16 miles.

**Near Distance to Hospitals**

* Suppose you wanted to determine the nearest hospital from a particular address in Iowa. The Near tool can be used for this.
* Add the Melrose\_IA shapefile to ArcMap. This represents a sample address.
* Search for and open the Near tool. Select Melrose\_IA for the Input features, Iowa\_Hospitals for the near features, check location, planar for method and click ok. Open the table for Melrose\_IA and there will be new fields with the Near\_FID, Distance, and coordinates for the nearest hospital. Perform a join of the Iowa\_Hospitals table to the Melrose\_IA table using the Near\_FID and FID key fields to discover the name of the hospital.

**Further exploration**

Make maps with the data you have obtained in this workshop or use the tools you have learned to calculate new data.