**Lab 2. Creating Spatial Data in Excel**

**Introduction**

This lab will introduce you to the importance of working in Excel when creating, editing, and manipulating data for work in GIS.

**Instructions**

Based on the assigned readings for this week, answer questions in Part I of this lab. Then proceed to the following Parts and complete the hands-on exercises. Be sure to copy all the files for the lab onto your flashdrive.

**Deliverables**

1. One MS Excel document containing the modified *USGS\_ppl* and *places2k* files. Don’t forget to save it in the following format: **LastName\_Lab#**.
2. A PDF of the questions below. Be sure to use your own words and submit based on the ‘How To” guidelines provided in the syllabus.

**PART I – Assigned Reading**

*Textbook – Bolstad (Chap. 2)*

1. Why are there different types of spatial data models?
2. Which of the following are allowable geographic coordinates?
	1. N45° 45’ 45”
	2. longitude -127.34795
	3. S96° 12’ 33”
	4. W -12° 23’ 55”
3. Characterize the following lists as nominal, ordinal or interval/ratio:
	1. worst, worse, bad, ok, good, better, best
	2. -40°F, -32°F, 0°F, 100°F
	3. corn, wheat, rye, barley, oats
	4. red, orange, yellow, green, blue, indigo, violet
4. Give examples of a phenomena that can be represented with both the vector and raster data model, and explain why.
5. What are the three basic types of vector objects?
6. When planar topology is enforced, what happens with the relationships among lines and polygons?
7. In a vector attribute table, how are unique spatial features distinguished?
8. What shape is the spatial extent of a raster?
9. Describe three advantages of using the raster data model.
10. Describe three advantages of using the vector data model.
11. When would the TIN model be more appropriate to use than the raster model?
12. Within the context of an object data model, what is a “logical model”?
13. What three files are required for a shapefile to exist?
14. Write the run length coding for this raster row: bbaacddd

**PART II – Using Excel to ‘clean up’ datasets**

Different agencies identify the latitude and longitude of any given place from different criteria; for example, where the Census Bureau may use population density of a city to determine its latitudinal and longitudinal center point, other agencies (such as the USGS) may use the geographic center of a city polygon to determine the centroid of any given place.

In this part of the lab, we will illustrate the differences in Texas populated place centroids using ESRI’s ArcMap. The lab is designed to introduce you to basic GIS functions and database management. You will manipulate data from a raw format (.csv or .txt file) using Microsoft Excel.

The lab data is in .txt format, which was retrieved from both USGS and Census Bureau websites. You will import this data into Excel for pre-processing - determining what information is needed and what information needs to be altered. After processing, it will be imported into ArcMap, mapped and analyzed.

* Download the Lab2data folder, unzip, and save to your flash drive*.*

The two .txt files provided are in two different *Text* formats. *Places2k* is fixed width, meaning that characters within the file are separated according to the place that they hold within the document (e.g.: 1, 2, 3, 4, 5, 6, etc.). USGS\_ppl has been saved as a delimited file. It is a tab-separated text file, meaning that each unique piece of information is separated by the *tab* key.

* Open a new blank worksheet in Excel
* Click on the *Data* tab to open the .txt documents



* First, let’s import *places\_2k*. On the *Data* tab, go to the drop down menu under *Get External Data*, and click on ‘*From Text’.* Navigate to *places2k* and click ok. The *Text Import Wizard* should look like the screenshot below:



* Remember that this file is *Fixed Width.* Choose the fixed width option, and click next.
* Here you are able to choose where to insert the breaks between pieces of information. Use your mouse to click to create line breaks; this will allow you to choose which data goes into a cell. Add breaks after the 2nd (state abbreviation), 4th (i.e. 01), and 9th characters (before city name begins). The remaining breaks should already be in the correct position.
* Click *Finish and then OK (to import into an existing worksheet);* places2k should now be viewable in a spreadsheet. Right click *Sheet1* and rename it *Places2k.*

This file displays more information than we need. To make it useful for our purposes, we will need to ‘clean up’ our data. Since we will only be looking at Texas, let’s start by isolating only the information that concerns Texas.

* Choose the *Filter* function under the *Data* tab. This allows you to sort data in a variety of ways by clicking on the drop down arrow.
* In Column A, sort the data so that you can only see TX (click on Select All to uncheck all states and then find TX and check).
* Once the TX rows have been isolated, select all rows (Crtl+A), copy (Ctrl+C) and paste (Ctrl+V) into a new worksheet sheet. To create a new worksheet, click the + at the bottom of the page next to where you see *Sheet 1.*
* Rename the new sheet *Places2k\_TX*.
* Delete Row 1, which still contains AL.
* Now we can remove all information that we will not be using. Delete columns E - J.
* Next, we need to add some titles. Right click Row 1 and choose *Insert*.
* In the blank cells, label the columns as follows: STATE, STATE\_FIPS, FIPS, PLACE\_NAME, LATITUDE, and LONGITUDE.

To finish cleaning up this file, we need to remove all of the extra information stored in the PLACE\_NAME column. The lower case identifiers are used by the Census to designate whether each places is a city, town, village, or CDP (Census Designated Place).

* To remove these identifiers, we will use the *Find & Select* function. This is found under the *Home* tab. Choose *replace,* and click *Options>>*:



* You will be replacing: city, village, CDP, and town. Click *Match Case*. In the *Find what* box, insert a space and then type city. Including a space before the word will remove the space between the city name and the word ‘city’. This extra space that often inhibit functions in GIS (such as a joins). Leave *Replace with* blank.



* Replace each designation one at a time to be sure that it works correctly; then choose *Replace All*.
* Repeat those same steps (Replace) for village, CDP, and town (***don’t forget to include the space in Find What!***).
* Yay! You have successfully cleaned up your *Places2k* text file! Save the *Places2k\_TX* as an excel document to your flash drive.

It is time to attack the 2nd text file from the USGS - *USGS\_ppl.*

* Add a new sheet (*Sheet3*), import the *USGS\_ppl* file using the *Get External Data* menu under *Data* tab (same as with the first file). Add the data using the *Text Import Wizard.* Make sure to click on *Delimited*, which means that each piece of information is separated with a ***tab***. After specifying that the file is delimited by tabs in the 2nd step of the wizard, click Finish.

You’ll note that this data is already has field names and looks fairly clean. Look at columns H and I. What do you notice? Does this latitude and longitude format look the same as *places2k*? Now you will need to use the *MID* function to separate out DD (degrees), MM (minutes), and SS (seconds); you will then recombine these with the *SUM* function into the Decimal Degree format.

* Begin creating new columns to the end of your Excel spreadsheet. Name these fields: LatDeg, LatMin, LatSec, LonDeg, LonMin, LonSec, Latitude, and Longitude

The **MID** function is used to extract a substring from a string of characters. For example, if I had a text string that read in cell B2: MYPROFISRAD and only wanted to extract PROF, I could use the following formula in Excel: =MID(B2,3,4). Here, *B2* is the cell where the information is found, *3* is the starting location of the substring that I want, and *4* is the number of characters to extract.

* Let’s start with the *Latitude* column:
	+ In the LatDeg column, you could use the following function to extract just the degrees: =MID(H2,1,2). This extracts only the first two characters.
	+ Once you type the formula in the first field, highlight it. This will create a box around it with a small square in the bottom right corner. Double-click on the box to copy the formula into all the fields in that column.
	+ To extract just the minutes in the LatMin column, you could use: =MID(H2,3,2).
	+ Use this logic to separate the degrees, minutes, and seconds from both Latitude and Longitude.
	+ **HINT: for longitude, you are working with three characters for degrees rather than just two (i.e. 098 are degrees).**

Once you have these six columns populated, you can recombine them into the Decimal Degree format that is most commonly used in ArcGIS.

* Convert all the original text strings into decimal degrees of longitude and latitude: Degrees + Minutes / 60 +Seconds /3600 (Minutes divided by 60 as there are 60 minutes in an hour, Seconds divided by 3600 as there are 3600 seconds in an hour).
	+ Start with the LATITUDE column and type **=VALUE((J2)+(K2/60)+(L2/3600)),** where J2 are Degrees, K2 are Minutes, and L2 are seconds. The formula divides minutes by 60 as there are 60 minutes in a degree; it also divides seconds by 3600 as there are 3600 seconds in a degree.
	+ Populate the entire Latitude column with this formula (double-click on small square)
	+ You can repeat this function for LONGITUDE. Since Texas is in the Western hemisphere, we need to designate the Longitude as negative. To do this, multiply the whole formula by -1: =VALUE((M2)+(N2/60)+(O2/3600))\*-1

Compare your *places2k\_TXsort* and your ­*USGS\_ppl* sheets. Scroll down to Austin. Do the coordinates look like they are in a similar format now? Even though they are both in Decimal Degrees, how can the location of Austin be completely different?

* Save your Excel spreadsheets in one workbook as an Excel 97-2003 Workbook. Your data has now been successfully modified for importation into ArcGIS!

**Part III: Importing Data into ArcGIS**

In the Lab2data folder, there are two shapefiles – Texas\_state and Texas\_counties. These files are GIS data files and are ready to be used in ArcMap.

* Open ArcMap and add Texas state and county shapefiles (use the  button or drag from ArcCatalog). This will place county and state outlines around your two sets of points for some context.
* Let’s change the color for the Texas state to Hollow. In the table of contents, double-click on the symbol for Texas\_state to get the *Symbol Selector* to appear.
* Change the *Fill color* to *No color* and the *Outline Width* to 2.



* Now, let’s import our prepared spreadsheets, one at a time. Go to File>Add Data> Add XY Data.



* Browse and navigate to the .xls file you created in PartII of the lab and find your Places2k\_TX worksheet. Add that to the Add XY Data wizard.
* For X choose Longitude, and for Y choose latitude.
* Click on *Edit* to change the coordinate system. Find the Geographic Coordinate System folder and go to World>WGS1984. Click OK 3 times and your points should be added to the map.
* ***THIS FILE IS ONLY TEMPORARY***. To make the file permanent, right click the Places2k\_TX$ Events and go to Data🡪 Export, and save your file as Places2k\_TX onto your flash drive. **Be sure to that it is saved as a SHAPEFILE or SHP.**
* Add the new shapefile to the map when prompted.
* Repeat those steps for the Sheet3 (USGS\_ppl file). Name this new file USGS\_Pl\_TX
* Your map should now look similar to this:



* With both files imported, change the color of your points to differentiate them (if they aren’t already different colors).
* Now you are ready to create a **join.** This is essentially a relational database. Right click on your newly created USGS\_Pl\_TX shapefile and choose **Joins and Relates**🡪 **Join**. You will be joining *Feature\_NA* from USGS to *place\_name* from Places. Check the **Keep only Matching Records** option. Validate your join. If successful, click OK.



* Export file as *TexasCombo*.
* Open the attribute table to your *TexasCombo*file – Right-click on the file names in the table of contents and click on *Attributes table*.
* Create three new fields: Difference, LatAVG, LongAVG—be sure to set the **Field Type** to **Double.** To add a new field, go to the *Table Options* drop down menu and click on *Add field.* Repeat three times – once for each field.



* Scroll down to the newly created Difference column, right-click and go to *Field Calculator*.
* Use the Field Calculator to calculate the difference between centroids: Sqr((xa-xb)^2+(ya-yb)^2), where x=latitude, and y=longitude. Once you enter the formula (using the appropriate column names), click OK.
* Sort the Difference Field by descending; Notice those outliers? Let’s eliminate them using a definition query. Double-click on *TexasCombo* in the table of contents and go to the Definition Query tab.
* Click on Query Builder and create the following query: “Difference” < 0.50
* Click OK twice to take you back to the map.
* Open the attribute table again and use the Field Calculator to calculate the average Lat & Long in the LatAVG & LongAVG fields. You can calculate the average using a similar function: (USGS\_Lat+Census\_Lat)/2
* From your current Attributes Table, export the table to your table of contents (from the *Table Option* in the top left corner go to Export. Be sure to export as DBF (database file) and name it **Centers**.
* With your new table, create an XY event to show your LatAVG and LongAVG. ***Don’t forget to edit the Coordinate system!*** Export these points and label it Centers Events. You should now have a point in between your two original place points.



* In the attributes table of **Centers Events**, add a new field. Title it Radius (make the Type a Double). Using the field calculator, find the distance between the center point and Lat/Long from USGS/Census. Use the equation: (difference)/2
* To make it easier to understand, let’s change the symbology. Double click on Centers Events in the Table of Contents and go to the Symbology tab.
* On the left, choose Quantities🡪Proportional Symbols. In the Fields section, change the Values to ‘Radius’. Change your Unit from unknown to ‘Decimal Degrees’. This should create a circle around the points from both the Census and USGS.



Yay! You finished! To wrap things up, let’s create a map of your newly created data.

* Go to View (menu bar across the top of the screen) and switch to layout view.
* Create a map showing the centroid differences in Texas. Add appropriate map elements (eg: scale, north arrow, legend, title). Most map elements can be added using the *Insert* menu bar.
* Export your map as a PDF and submit to instructor.