GIS II**: Data Management: Creation, edition and maintenance of geographic data**
Module 1: Leveraging the where of your geographic data

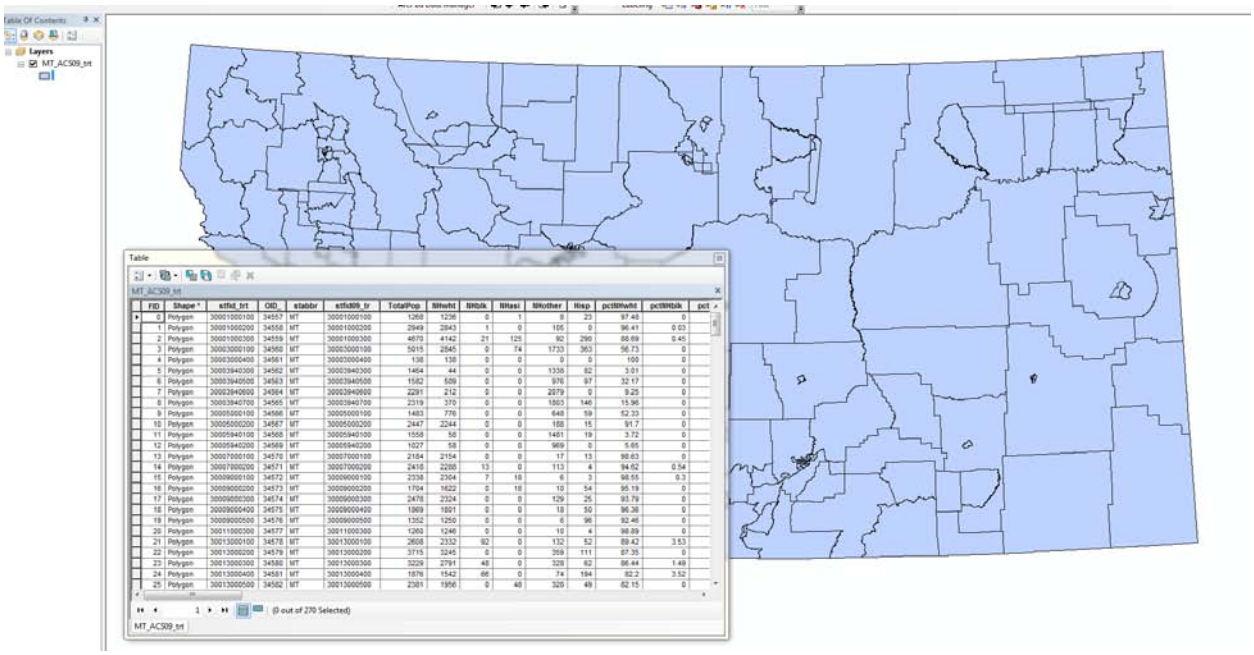
*** Files needed for exercise: *MT_ACS09_trt.shp, and McDonalds_rr.shp*

Goals: After completing this exercise, you will have experience in performing spatial joins and selections and reviewed your understanding of spatial data projections.

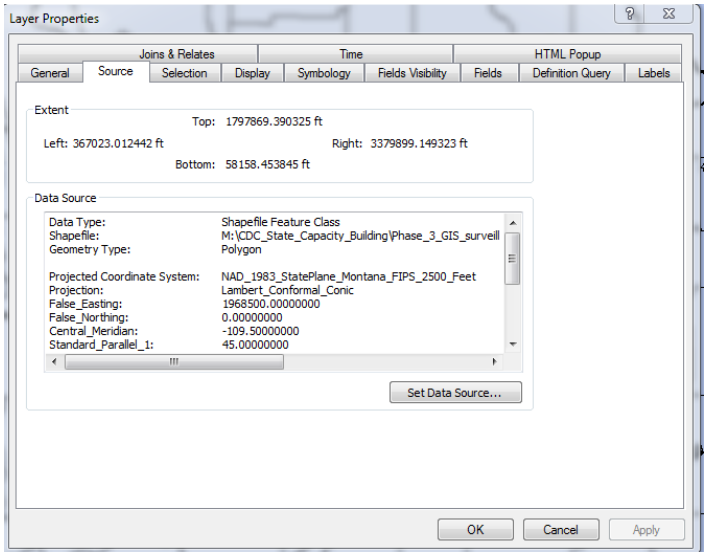
Skills: After completing this exercise, you will be able to spatially join a point dataset (businesses of interest) to a polygon dataset (American Community Survey US Census 2009 Tracts). This is a useful method for the enumeration of points within polygons: i.e. what is the count for points of interest within a geography of interest?

Creating Folder Connections:

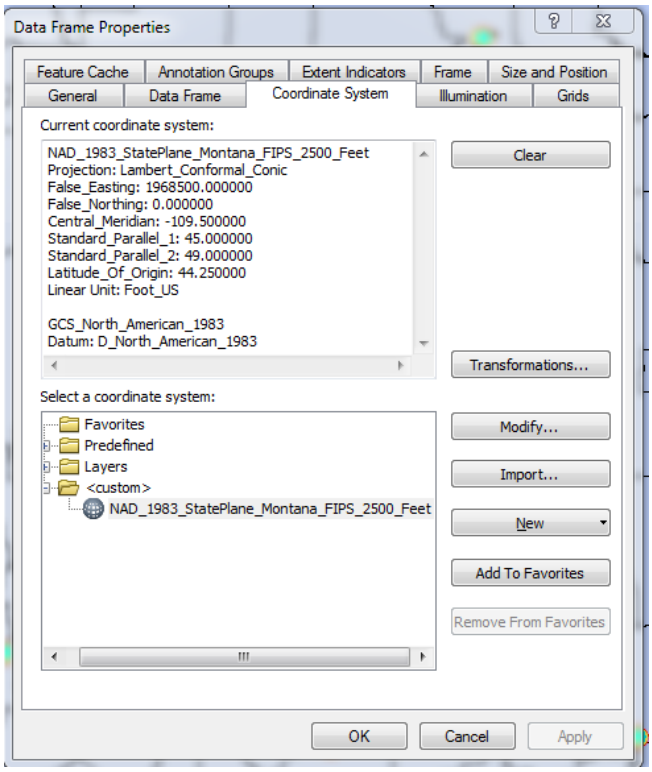
1. Open ArcMap. Choose to start a new empty map.
2. Click the Add Data button .
3. Click on the Connect to Folder button . Connect to the Exercise_01_data folder.
4. Browse to your folder and connect to it. We now have a permanent connection to that folder.
5. Double click on *MT_ACS09_trt.shp* to add it to your project. This shapefile has American Community Survey 2009 socio-economic and demographic data appended to it.
6. Open the table for the shapefile and take a look at the attribute fields.



7. This shapefile is projected properly to conform to the standard projected coordinate system for Montana. Confirm this by right clicking on the shapefile in your table of contents (TOC) and selecting the Source tab.



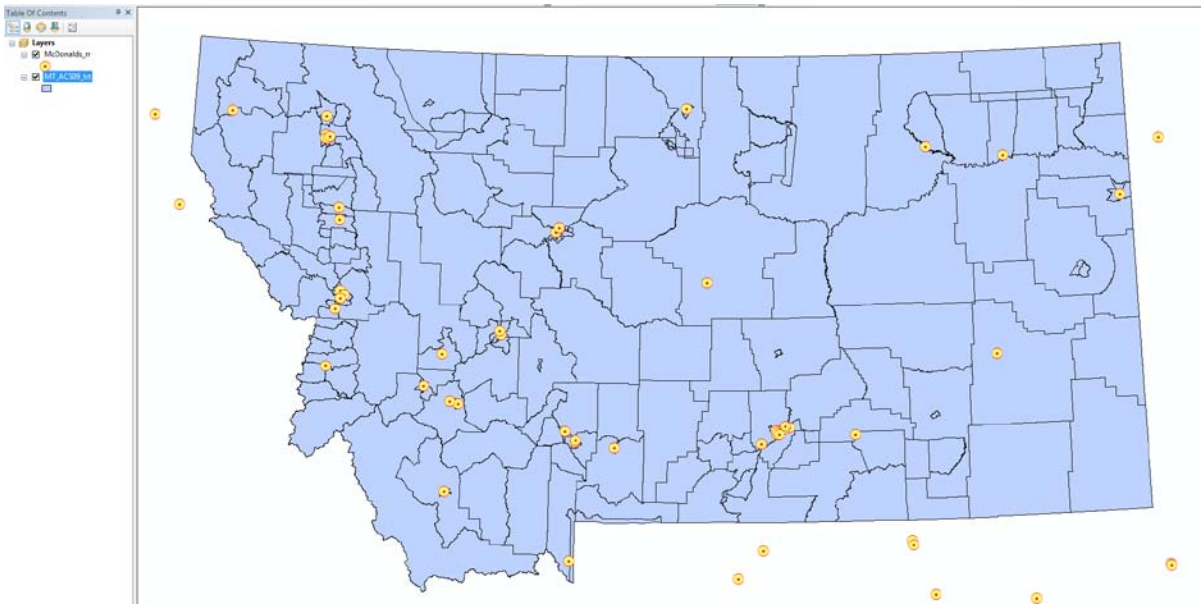
8. Since this is the first projected layer you have added to your map dataframe, the dataframe projection now matches *MT_ACS09_trt.shp.* Check this by right clicking Layers > Properties and clicking on the Coordinate System tab.



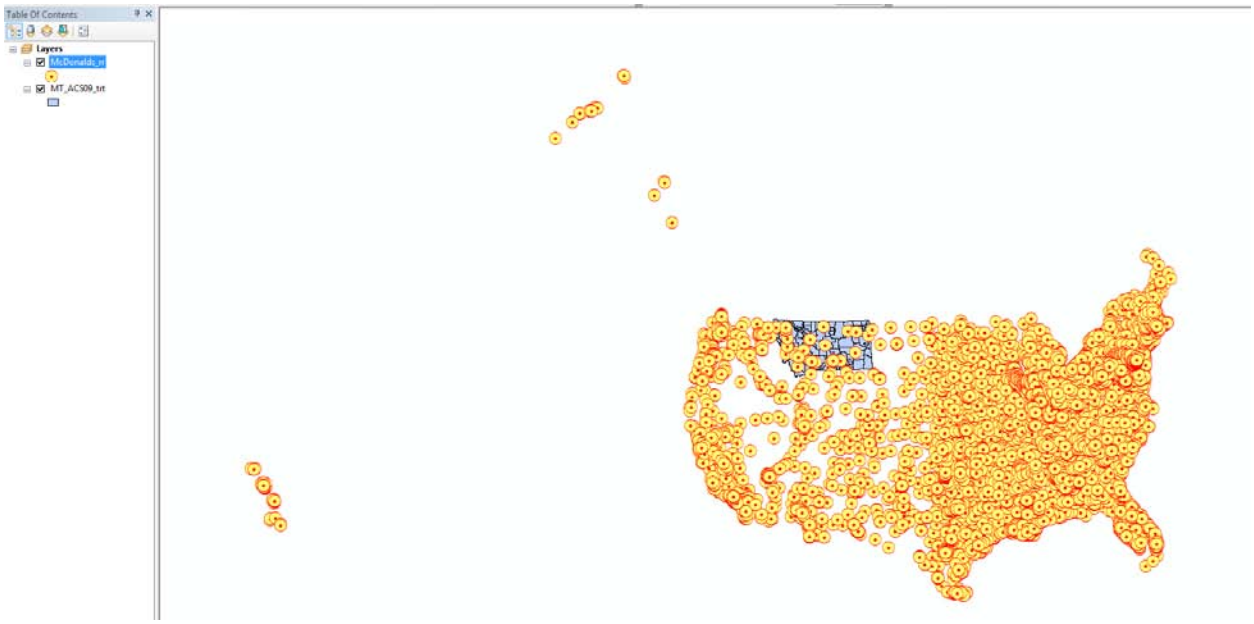
GIS II**: Data Management: Creation, edition and maintenance of geographic data**
Module 1: Leveraging the where of your geographic data

Add the *McDonalds_rr.shp* to your map:

1. Add the file *McDonalds_rr.shp* to your project. This shapefile is located in the Exercise_01_data folder. You should now see *McDonalds_rr.shp* in the TOC on the left hand side. This is a nationwide shapefile of McDonald's restaurants, and is a subset of businesses classified by 6 digit 2007 North American Industry Classification system (NAICS) as Full-Service Restaurants. Yes; apparently McDonald's is classified as a full-service restaurant. (The dataset that the McDonald's restaurants were selected from is also present in the Exercise_01_data folder *NAICS_722110.shp*). More information on the 2007 NAICS codes may be found here: http://www.census.gov/eos/www/naics/.
2. *McDonalds_rr.shp* is a nationwide dataset so it probably uses a coordinate system that differs from the Montana tract shapefile; confirm this. What is system is defined? **WGS_1984_Web_Mercator_Auxiliary_Sphere** is the de facto projected coordinate system for most of the nationwide datasets for Esri's ArcGIS 10 provided spatial data. Since the projected and geographic coordinate systems are both defined, ArcMap can project on the fly and make the data work in the same data frame as the Montana tract data.
3. The McDonald's eating establishments should now be displayed above your MT Census tracts.



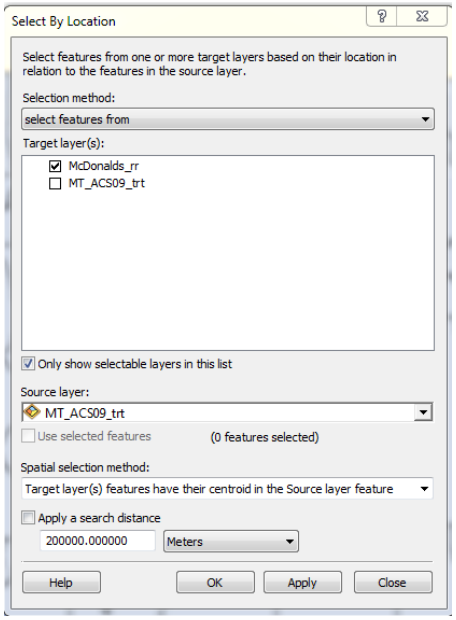
4. For kicks right click on *McDonalds_rr.shp* in your TOC and select the Zoom to layer option. You should see the many McDonald's restaurants across the USA (sorry territories are not included).
5. Open the attribute table to find out how many McDonald's are in the table.



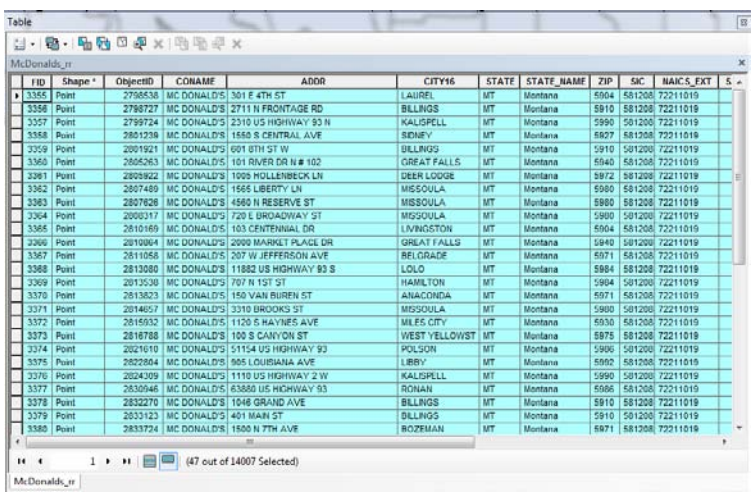
6. Once you have done this zoom back to the Montana tracts.

Perform a spatial selection to create a new shapefile of only those McDonald's in Montana:

1. Under the selection menu select choose Select by location. This will bring up the Select by location dialogue box. We want to use the location information of both the McDonald's points (Target layer) and the Montana tracts (Source layer) to determine which features in the target layer have their centroid in the source layer feature. Plainly put: which McDonald's restaurants are in Montana?

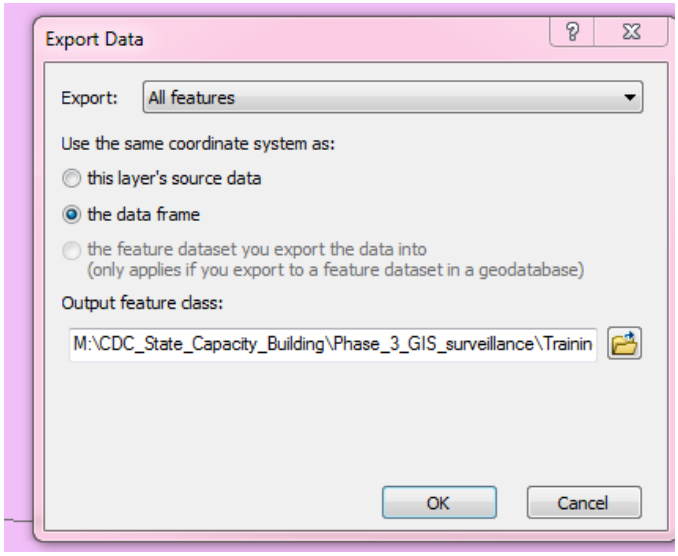


2. Apply your spatial selection and click OK.
3. Right click on *McDonalds_rr.shp* in the TOC and open the table; how many records have been selected based on your spatial query?



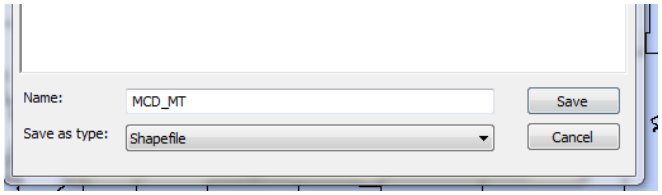
4. Once you have confirmed that **47 records** have been selected, close the table.
5. How could you have created this selection with a tabular selection?
6. Now you will export your selected McDonald's to a new shapefile. In the process you will set a new coordinate system, making sure it matches your Montana tracts. Right click on *McDonalds_rr.shp* in the TOC. Select Data > Export data.
7. Choose the radio button to indicate that you want to use the coordinate system from "the data frame". Recall that you have defined the projection for the data frame by adding

MT_ACS09_trt.shp as the first layer. With this information ArcMap can project the data to MT's projected coordinate system.



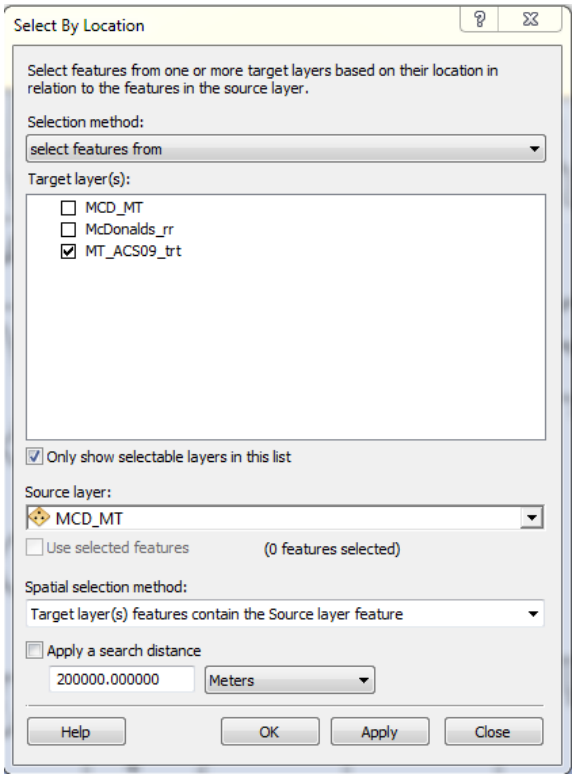
8. Name your file *MCD_MT.shp* and save it in your Exercise_01_data folder.

   Make sure you save as type: Shapefile.

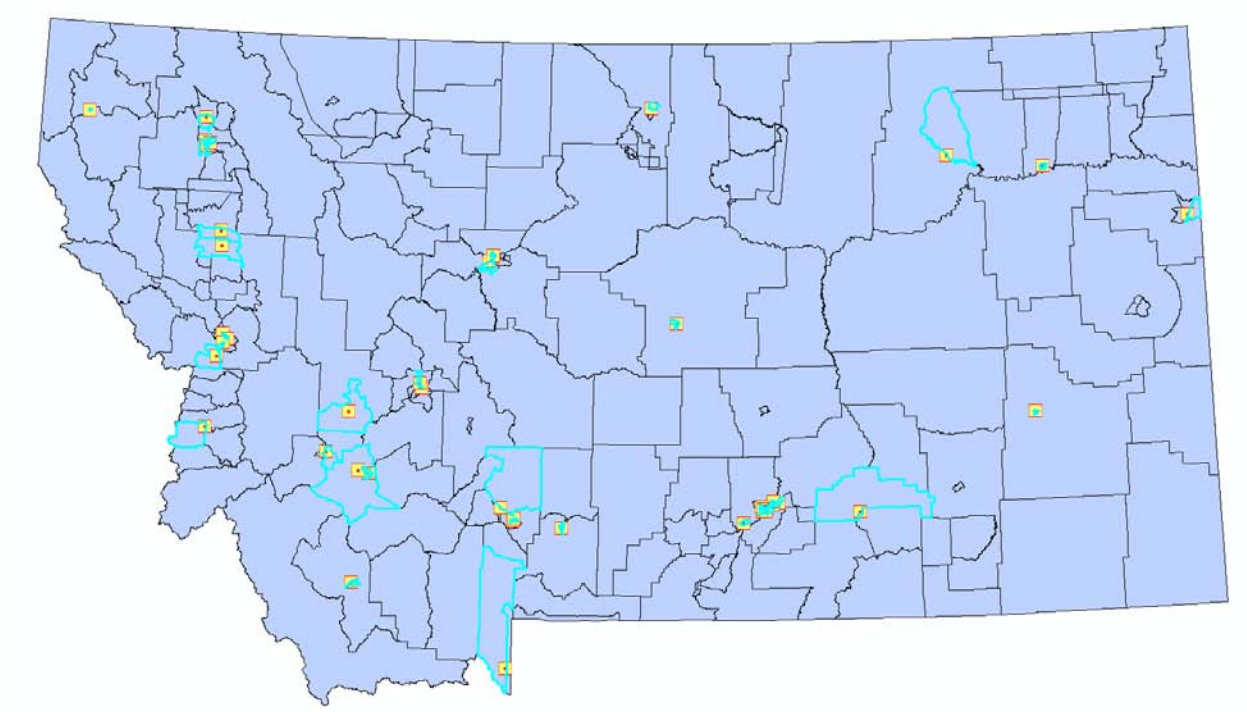


Make another spatial selection to determine which tracts contain a McDonald's:

1. You now have a shapefile that contains only McDonald's in the state of Montana: *MCD_MT.shp*. Let's say we are interested in knowing all MT Census tracts that contain a McDonald's Restaurant. You could not do this using a tabular selection and so must use a spatial selection.
2. Click on the Selection tab and choose Select by Location to answer the question: which census tracts contain a McDonald's Restaurant?
3. You will select features from *MT_ACS09_trt.shp* (Target layer) that contain features from the *MCD_MT.shp* layer (Source layer). To do this your spatial selection method will be: Target layer(s) features contain the Source layer feature, click OK once you have done this.



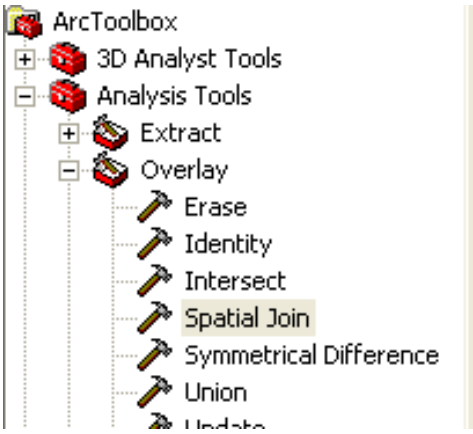
4. Take a look at your map; these are the Census tracts that contain McDonald's.



5. Now open the attribute table associated with *MT_ACS09_trt.shp* and toggle the buttons on the bottom of the table to Show only selected records. How many tracts contain McDonald's?
6. For fun, see what the median household income (MedHHinc) is for the tracts that contain McDonald's as opposed to those that do not. To do this, right click on the MedHHinc field. Select Statistics.
7. Switch the selection and check the descriptive statistics on those tracts that do not contain McDonald's.
8. This is great information and a good way to understand your data with a query, but like a tabular selection it is not permanent. What about making a more formal connection between the data? How about determining the count of the McDonald's restaurants by 2000 Census tract? We can accomplish this with a spatial join.
9. Clear your selected features.

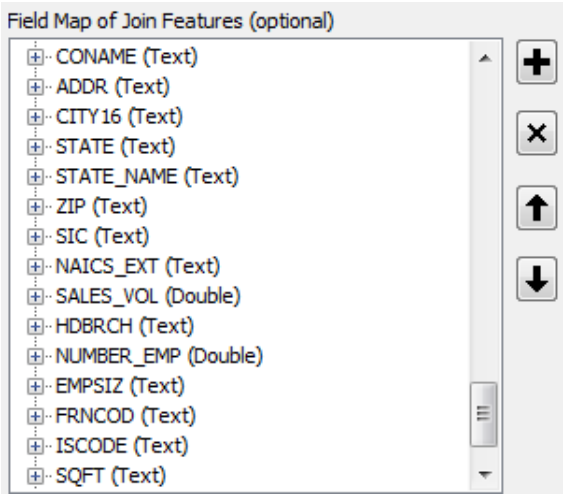
Spatial Join:

1. We are now ready to perform a spatial join between two related layers of data with the goal of answering a question that was initiated by a spatial selection: How many restaurants of interest are present with each host tract?
2. Open ArcToolbox > Geoprocessing Tab > ArcToolbox > Analysis Tools > Overlay > Spatial Join.

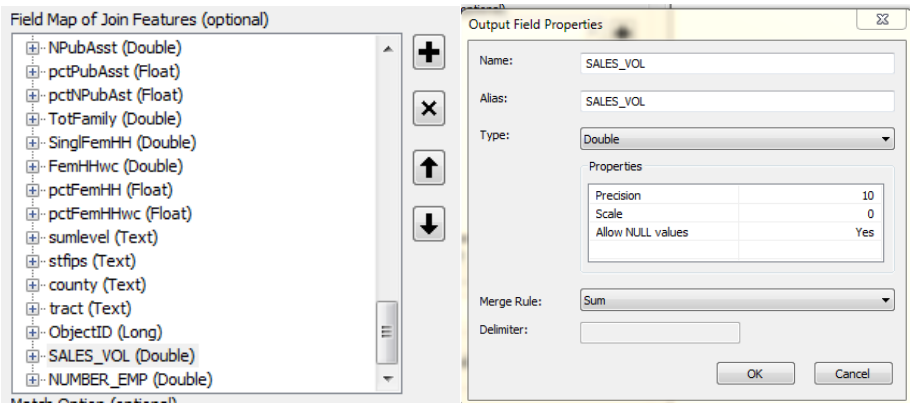


3. Our Target Features (the data to which we are appending) will be: *MT_ACS09_trt.shp*. Our Join Features (the data we are appending) will be: *MCD_MT.shp.* Join Operation: we want to JOIN_ONE_TO_ONE. Keep the default: keep all target features checked, and Match option: CONTAINS.

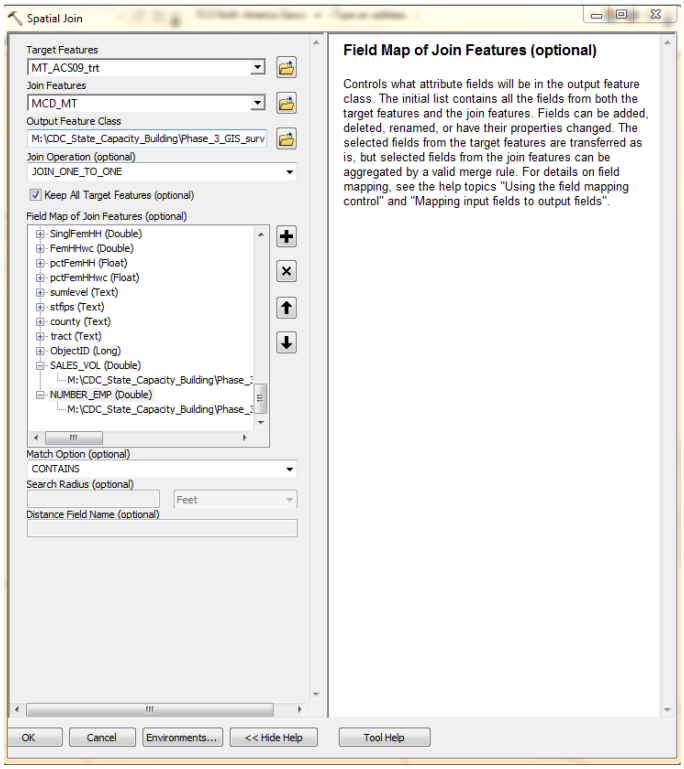
4. This is a **many** (McDonald's) to **one** (tract) join because it is possible that one tract may contain more than one McDonald's, right? So how will the attribute fields be handled for your join features (McDonald's) when multiple restaurants are present in a tract? You can adjust the merge rules for your fields in the Field Map of Join Features window. This controls what happens to fields when there is more than one record that fulfills your match option (i.e. tracts that contain more than one McDonalds).
5. Scroll down in the Field Map of Join Features; the fields from the join features (McDonald's) begin at CONAME.
6. Choose to remove all of the join feature fields with the exception of SALES_VOL and NUMBER_EMP (these are the only numeric fields in the table). This is done by clicking on the field to select it and hitting the X button to the right.



7. Right click on SALES_VOL and select Properties, this will allow you to change the merge rule for cases when more than one record that fulfills your match option (i.e. a tract contains more than one McDonalds). Select Sum as the merge rule. The default merge rule is First. Can you see why this could be an issue?



8. Repeat the process for NUMBER_EMP, adjusting the merge rule to Sum for those tracts that contain more than one McDonald's.



9. Notice that this spatial operation creates a new dataset that you will need to name and find a home for on your computer. Make sure you save it in your Exercise_01_data folder as a **shapefile** (a good name: *MT_tract_MCD_sj.shp*).
10. The new dataset will be automatically added to your table of contents. Open the attribute table and take a look the fields. For every record in your tract table (each record represents a closed polygon geometry) where a McDonalds is present; information for the restaurant has been appended to the tract.
11. You can check this by looking at the Join_count field. A > 0 count means that the tract contains a McDonalds.
12. Are there any tracts with multiple McDonald's? Sort the Join_count field in descending order by right clicking on the field and choosing Sort descending.
13. Take a look at the SALES_VOL and NUMBER_EMP fields. Since you adjusted the merge rules, these fields represent the sum of the sales volumes and number of employees for the McDonalds' in the affected tracts. Note that keeping the default merge rule (First) would display only first value detected in the spatial join. In many cases you will only be interested

in the join count, but it is important that you have an awareness of the merge rule factors in a many to one join.

Take a look at the resulting data; can you think of useful applications in your own work?

This type of examination is particularly useful in assessing the food environment for your state or community. More information on assessing the food environment may be found here:

http://www.cdc.gov/obesity/downloads/HFRassessment.pdf

Modified Retail Food Environment Index Across Census Tracts within State:

- Measures the number of healthy and unhealthy food retailers across census tracts in a state
- Lower scores indicate:
  - Less healthy food retailers
  - More convenience stores and fast food restaurants relative to healthy food retailers

Modified Retail Food Environment Index Across Impoverished Census Tracts within State:

- Measures only census tracts where 20% or more of the residents are below the poverty line

Community Food Environment

Strategies to improve the community food environment include increased access and availability to healthier food retailers. The modified Retail Food Environment Index (mRFEI) measures the number of healthy and less healthy food retailers within a census tract using the formula shown.

$$mRFEI = 100 \times \frac{\text{\# Healthy Food Retailers}}{\text{\# Healthy Food Retailers} + \text{\# Less Healthy Food Retailers}}$$

For this indicator, healthy food retailers include supermarkets, supercenters, and produce stores.† Less healthy food retailers include convenience stores, fast food restaurants, and small grocery stores with 3 or fewer employees.† Higher mRFEI scores indicate more access to healthy food retailers and fewer less healthy food retailers.

Indicator	National Data*
• Modified Retail Food Environment Index across census tracts within state	10
• Modified Retail Food Environment Index across impoverished census tracts within state	7

*The numbers shown for national data are the medians of all state scores.
† Areas without these types of healthy food retailers may still provide adequate access if smaller stores and fast food restaurants provide quality and affordable healthy foods and beverages.

More information on mRFE can be found here:

http://www.cdc.gov/obesity/downloads/NationalActionGuide.pdf

If you have time…

Try repeating the processing above for the entire 722110 NAICS code, and or try repeating the process for your state.