

This week and next we will use ArcGIS to develop a hedonic house price model for the single-family housing market in Rutherford County. Our first step is to *geocode* our housing data: we will use street address to place each house in the correct location on the map.

I. Geocode a data table with street address information

Census provides a good set of instructions on geocoding with ArcGIS and TIGER shapefiles, though some of the web links in the instructions are now dead:

Document 1: S:\TEFF\450\2020\arcview6\Geocoding.pdf [[link](#)]

Start by acquiring the TIGER/Line shapefile you'll need for geocoding addresses in Rutherford County. Go to:

<https://www.census.gov/cgi-bin/geo/shapefiles/index.php>

Download, under *Relationship Files*, the *Address Range-Feature Shapefile* for Rutherford County.

Follow the directions in Document 1 above to *Create Address Locator* from the shapefile you downloaded.

In the directory S:\TEFF\450\2020\arcview6\ there are other files we will use tonight. Copy the entire S:\TEFF\450\2020\arcview6\ directory to your *My Documents* folder. Load all of the shapefiles into ArcMap. Also add the data table *ruth2000.dbf*; this table contains the house price data we will use. Follow the instructions in Document 1 above to geocode *ruth2000.dbf*. You will need to pick the field *mailaddr* for *Street* and *mailzip* for *Zone*.

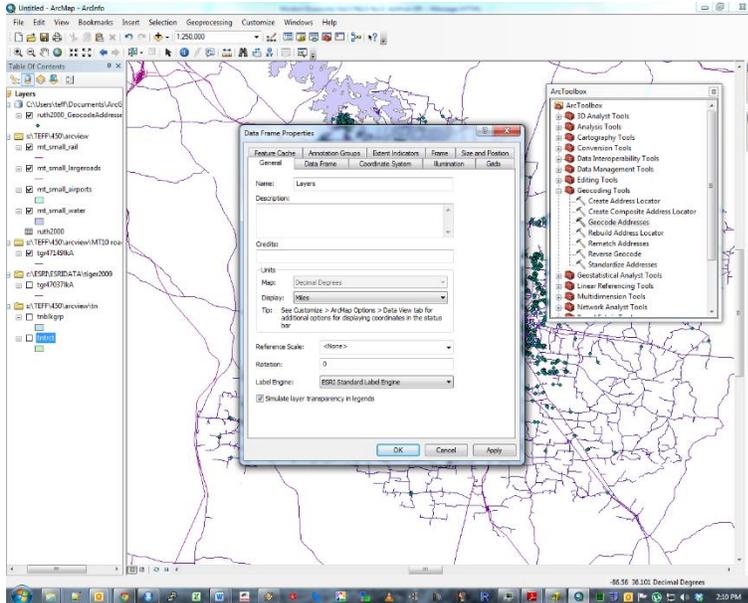
Table 1: Field names in *ruth2000.dbf*

variable	description
CITYNAME	city name
YR	year of sale
COUNTY	county
PRICE	sale price
LNDVAL	appraised land value
IMPVAL	appraised improvements value
AGE	age of home at time of sale
YRBLT	year actually built
EFYRB	effective year built
SQFT	square feet
ACRES	acres (not accurate)
MONTH	month of sale
SOUND	dummy: sound condition==1
VACANT	dummy: vacant at time of sale==1
DRYWALL	dummy: interior walls are drywall==1
BRICK	dummy: exterior walls are brick==1
SEWER	dummy: connected to sewer==1
LEVEL	dummy: built on level lot==1
PRACC	dummy: poor road access==1
MAILADDR	mailing address
MAILZIP	zip code

II. Variables measuring proximity to a feature

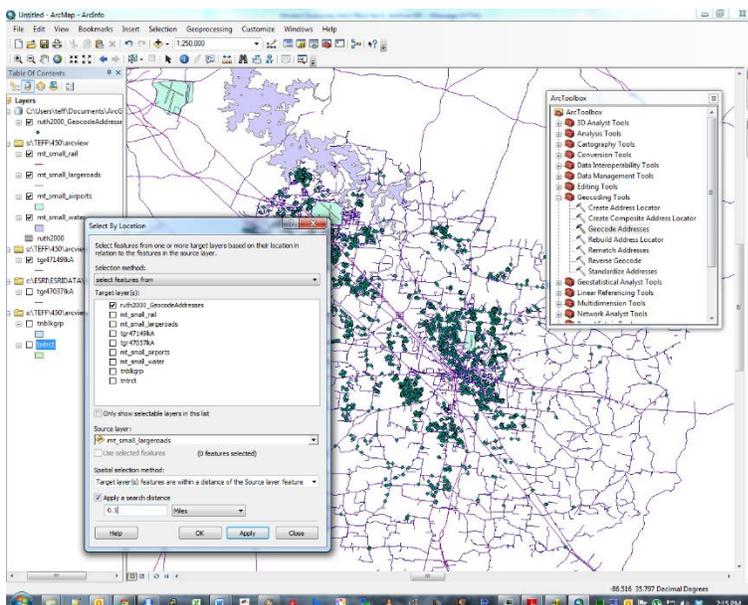
a. Creating a dummy variable

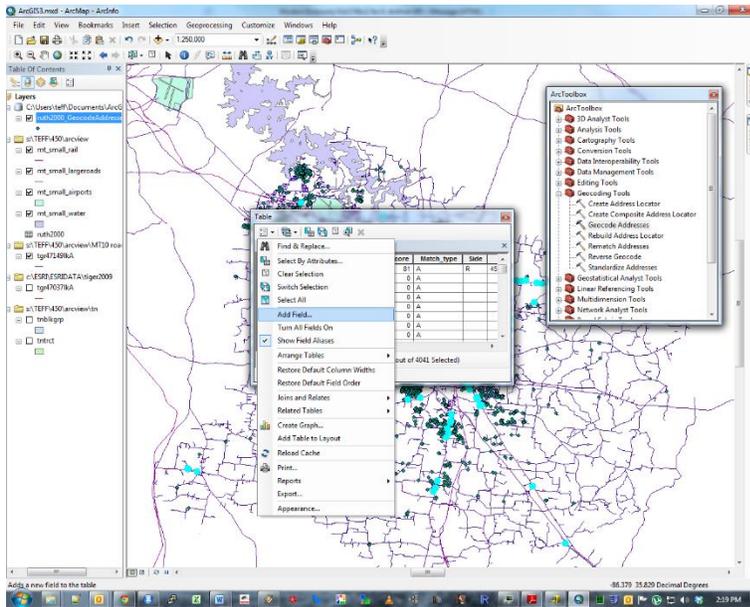
Click on *View* on the top menu, then click on *Data Frame Properties* near the bottom of the drop-down box. A dialog box appears: under the tab labeled *General*, enter *Miles* in the box for *Display*. Click *OK*.



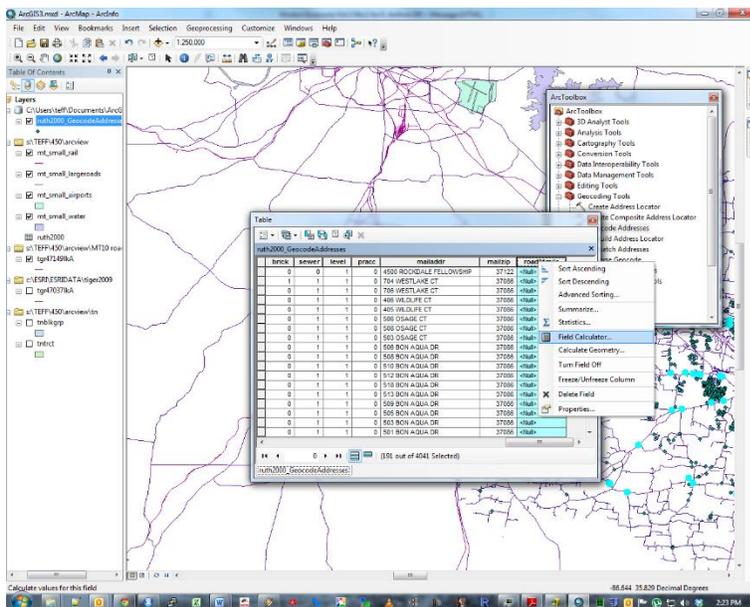
Next right click on the geocoded *ruth2000* layer and select *Open Attribute Table*. Click on the icon at the left on the menu bar of the attribute table, and then select *Add Field* in the drop-down menu. Call your new field *rail25mile*, and give it the type *Double*. Right-click the field name and then select *Field Calculator* in the drop-down menu. Enter the numeral “0” in the formula box, and then click *OK*.

Now go to *Selection* on the top menu bar, and click on *Select by Location*. A dialog box appears. In the first box, enter “Select features from” then click the small box next to the name of your geocoded home data. In the next box pick the layer *TNrails*. Then, for “spatial selection method”, select “are within a distance of the source layer feature.” For the distance, pick 0.25 miles. Click *OK*.

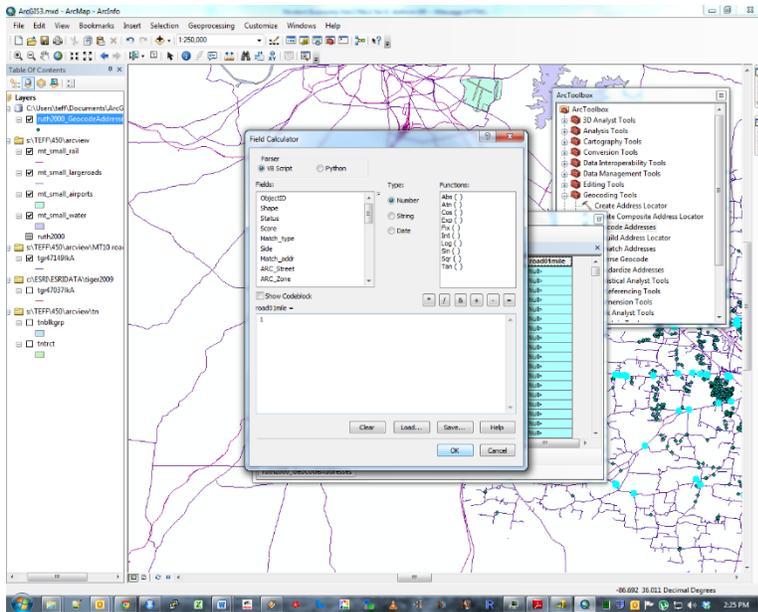




Return to the attribute table, right-click the field name and then select *Field Calculator* in the drop-down menu.



Simply enter the numeral “1” in the formula box, and then click *OK*.



Close the attribute table and de-select the homes by going to the top menu bar, clicking *Selection* and then *Clear Selected Features*.

We can repeat this procedure to get new fields indicating whether a home is within a distance of airports, interstate highways, water, etc. The point landmarks shapefile (*tl_2014_47_pointlm*) has a field in the attribute table called *MTFCC*, which gives a code for the kind of landmark. The code descriptions are given here:

<https://www.census.gov/library/reference/code-lists/mt-feature-class-codes.html>

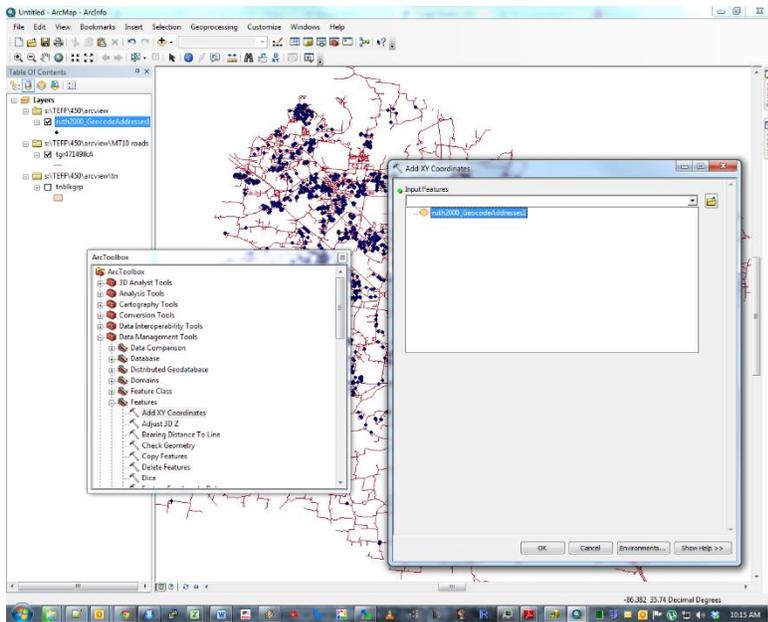
So, for example, you believe that proximity to a cemetery would affect house price, and you wish to create a dummy variable for all homes within 1/8th of a mile of a cemetery. Open the layer properties window for *tl_2014_47_pointlm*, and click the *Definition Query* tab. Using the Query Builder, restrict the displayed points to cemeteries ("*MTFCC*" = 'K2582'). Now, using the steps we did previously for railroads, select the homes within 0.125 miles, add a field to the attribute table, and use the field calculator to create a dummy variable.

b. Calculating distance to nearest point on feature.

In ArcToolbox, click *Analysis Tools*, then *Proximity*, then *Near*. In the pop-up box, use your geocoded house point shapefile for the *Input Feature*, and use *tl_2014_47_areawater* for the *Near Feature*. Click OK. You will now have two new fields added to your attribute table: *NEAR_FID* and *NEAR_DIST*. The first of these tells which of the polygons in the areawater shapefile is nearest, and the second tells the distance.

III. Add XY coordinates to attribute table

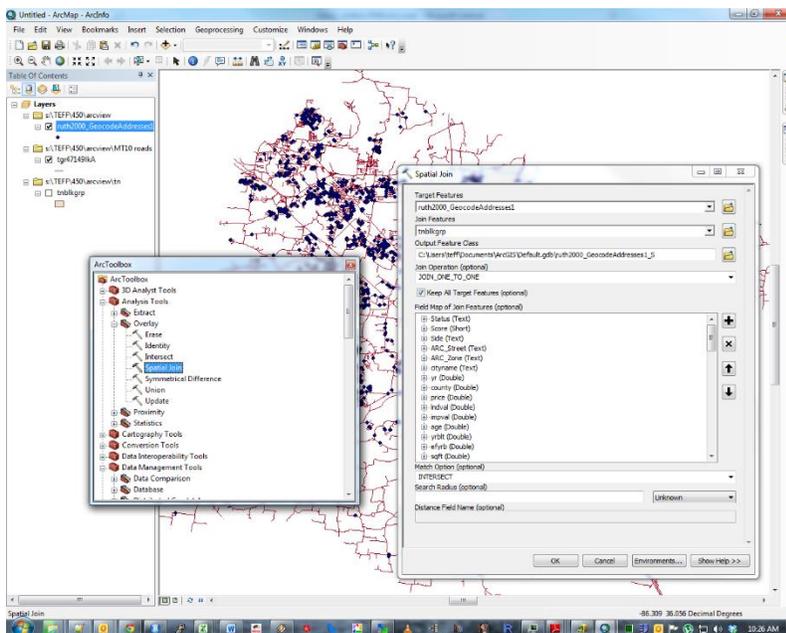
Open ArcToolbox (the red box icon on the top menu bar). In the toolbox table of contents, click the box for *Data Management Tools* and then click the box for *Features*. The top hammer icon says *Add XY Coordinates*—double-click this. In the resulting dialog box, activate the drop-down menu and pick your geocoded homes file, then click *OK*. ArcMap will then add two new fields to your dbf file: *point_y* (latitude) and *point_x* (longitude).



IV. Spatial Join to polygon data

It's often useful to include Census data in a home price model. This can be done through a *Spatial Join*. The shapefile *ruthBG* contains a few variables, at the block group level, from ACS 2014 5-year sample. The fields are described in *S:\TEFF\450\2020\arcview6\ruthBGfields.csv*; you might find it worthwhile to collect other fields, relevant to housing, for this exercise.

Click the box in the toolbox table of contents called *Analysis Tools*, then the sub-box called *Overlay*. Double-click the hammer icon for *Spatial Join*, and in the resulting dialog box, pick your geocoded homes file as the *Target Feature*, then *ruthBG* as the *Join Features*. For *Match Options* pick *INTERSECT*. Then click *OK*. Look at your attribute data to see what you now have for each home.



Not all of your homes were successfully geocoded, so that you have values equal to zero for the census fields and values equal to one for latitude and longitude. The following steps will remove these. First, double-click on the name of your file to bring up the *Layer Properties* dialog box. Click the tab for *Definition Query* and then the button for *Query Builder*. Double-click the field “status” and then click the button for *Get Unique Values*. Click the symbol for “not equal to” (\neq) and then click “U” (*unmatched*). Click *OK*. Click *OK* again. Now extract your data to a new shapefile (right-click on file name, then *Data*, then *Export Data*), and bring this new shapefile in the view, removing the old one.

Once you have created sufficient location-related variables in the attribute table, export the attribute table as a dBase format file. Save it on your computer in a location where you have write permission. We’ll use the data next week in a hedonic house price model.