**Adding Data:**

**Displaying XY Coordinates and Geocoding Addresses**

|  |
| --- |
| **Introduction** |

In this course, you have already been introduced to ArcGIS and learned how to access and download geospatial data in the form of shapefiles that can be added directly to your map. This week, we are learning about how to add non-spatial data to your map. These data still have some sort of information about their location (otherwise we couldn’t show them on our map at all), but we must tell ArcGIS how to use this information to place the point features on our base map.

There are two types of location information that our data might have that we can use to plot points on the map. The first (and easiest) information are XY coordinates. If our list of points already has XY coordinates, then it’s just a matter of telling ArcGIS to match up those XY coordinates with the coordinates of our base map and put point features at each location. Most lists of locations don’t come with XY coordinates, though – they come with street addresses. If we want ArcGIS to read a street address and find it on our map, we will have to build a tool called an “address locator.” ***THEN*** we use the address locator to plot the addresses on the map.

Let’s start with the easy way first.

|  |
| --- |
| **Displaying Data with XY Coordinates** |

For this part of the assignment, we will use crime data provided by the Boston Police Department (BPD). The BPD very conveniently posts their crime data on the Analyze Boston website, data.boston.gov. When I first downloaded the crime data, I was worried that we wouldn’t be able to use it for mapping – if you look in the Excel spreadsheet, the “STREET” column has *only* the street, no lot number. Uh oh! “Dyer St” isn’t enough information to find a point on a map! But then I scrolled across and – jackpot! Latitude and longitude. These are XY coordinates that we can use to plot our crime incidents on a map.



This is the easiest way to get your data on to your map. All you need to do is drag and drop your Excel file into your Table of Contents, then right-click it and choose **Display XY Data**.

A window will open and ask you to tell ArcGIS where to find the X and Y coordinates. For our purposes, it is almost always **X = Longitude** and **Y = Latitude**. There is no Z field for this exercise.

When you click OK, you will see your crime points appear on the map! If you would like to save this layer as a shapefile so that you can add it to your map more easily in the future, just right-click the new point layer, click **Data**, **Export Data**, and save the features as a shapefile (remember to save will all of your other data and to use a descriptive name!).

You can try using the Identify tool (the blue circle with the white i in the middle) to click on any crime point on your map. This will open up the Identify window to give you additional details from the data table, including the offense description, year, month, day and hour of occurrence, etc.

|  |
| --- |
| **Geocoding: Building an Address Locator** |

***For this assignment, we will use the list of registered sex offenders (RSOs) who live in Boston. You should have a “clean” Excel file of RSO addresses from our in-class exercise earlier this week.***

What if the data you want to add to your map doesn’t include XY coordinates? That is a bit more difficult. See, street addresses (like 73 Tremont St) are not spatial data. Sure, to us, they symbolize a specific place, but only because we know where certain streets are located relative to other landmarks. If someone gave you a random address, it wouldn’t help you to know where that location was unless you also knew the other streets around it. ArcGIS is the same way – street addresses mean nothing to it unless we tell it how to search our streets layer and find the right street, as well as how to find the right street number!

The tool that does this is called, sensibly, an **address locator**. The address locator tells ArcGIS that when we give it an address, it needs to check a specific layer to find that street and how far along that street it should place our point feature.

The data layer that contains this necessary information is the Address Range-Features Shapefile (ADDRFEAT) available from the [Census TIGER website](https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2018&layergroup=Relationship+Files). If you completed ArcGIS #1 successfully, you should already have this layer on your map. Otherwise, you will need to add it now – download the Address Range-Feature Shapefile (make sure you download exactly this file – many others are similarly named!) for Suffolk County, MA, and add it to your map.

ArcGIS has a built-in wizard for creating address locators. You can find it by opening the **ArcToolbox** – either click the toolbox button on your toolbar  or go to the menu **Geoprocessing** > **ArcToolbox**.

Once you open the ArcToolbox, you’ll see an option that says **Geocoding Tools**. Under there, you will find **Create Address Locator**. Double-click to start the address locator wizard.

To choose an **Address Locator Style**, click the little yellow folder next to this option. For this exercise, select **US Dual Ranges**.

Your **Reference Data** is your ADDRFEAT layer.

Make sure to set the ‘Role’ to **Primary Table**.

If you scroll down in the address locator wizard, you will see a section called “Field Map.” This is where we tell ArcGIS where to look in the ADDRFEAT layer to find all the information it needs to locate addresses. Let’s take a look back at the attribute table for ADDRFEAT to remind ourselves what’s in there (ideally you would do this before you start the Create Address Locator tool, but don’t worry, I’ve got you – see below:



In the attribute table for our ADDRFEAT layer, you can see that we have a column called **FULLNAME** that appears to have the street name in it. Then we have **LFROMHN**, **LTOHN**, **RFROMHN** and **RTOHN**. What do these columns mean? These columns tell ArcGIS which house numbers are located on this street segment and which side of the road they are on.



Okay, let’s go back to the address locator wizard. The Field Map is asking you to identify some key information from the ADDRFEAT layer. Notice that some of the fields have a \*. These are the required fields – everything else can be left blank.

Fill out the field map with the following values.

From Left: **LFROMHN**

To Left: **LTOHN**

From Right: **RFROMHN**

To Right: **RTOHN**

Street Name: **FULLNAME**

In the area labeled “Output Address Locator,” ArcGIS is asking you what you want to call your address locator and where you want to save it. Click the yellow folder to navigate to your working director, where all of your other shapefiles are saved, and title your address locator something sensible like Boston Address Locator. Once all of that is done, click OK.



Now **WAIT**. Compiling the address locator takes some time. You need to give the program time to think! Wait at least a minute before trying to move on to the next step: using your address locator to plot your list of addresses on your map.

|  |
| --- |
| **Geocoding Addresses** |

Okay, you have an **address locator**, your **addrfeat** layer, and a *clean* **list of addresses**. Now what?

First, we add our list of addresses to our table of contents. To do this, click the Add Data button  and navigate to your working directory to find your clean Excel spreadsheet of addresses. Double-click it, and it will ‘open’ to show you how many sheets are in that workbook, which in our case is just one (Excel files are technically workbooks that can have multiple sheets in them, so we COULD have multiple sheets of data in our workbook but for this exercise, we just have the one). Double-click the sheet (it should have a $ in the file name) to add it to your TOC.

You should see your table of addresses appear in your TOC, like the image to the left.

Right-click on the table and select “**Geocode addresses.**” A window will open and ask you which address locator you want to use. Your Boston locator should be in this list. If not, you can click the “Add…” button and go and find it in your working directory.



Once you find your correct locator and click “OK,” a new window will open up. This is where you tell ArcGIS where *in the list of addresses* to find the address information it needs to do the geocoding.

First it asks for where to find the street or intersection. In our Excel spreadsheet, the name of this column is **STREET**. Then it asks where to find the zip code. Remember, we did have a column called ZIP *but* all of those zip codes were incorrect (they didn’t have the leading 0). So, we created a new column called **NEWZIP**, and *that* is the one we want ArcGIS to use.

Then it wants us to tell it where to save our new layer that will be created when our addresses are all geocoded into point features. Make sure you save your layer in your working directory, where all of your other data is saved! Called it something like “Boston Level 2 and 3 Sex Offenders.” Then click OK.

A window will appear that shows you the progress of your geocoding. Eventually it will tell you that the geocoding is completed and it will give you a breakdown of the addresses matched successfully, the addresses that have multiple matches you need to sort out (“tied”), and the number of addresses that could not be matched. I got 91% matched, which is *really* good for a first try – that goes to show you the value of cleaning data before you attempt to do any geocoding! But I think we can get closer to 100%. Click the “Rematch” button.

(If you accidentally close this window without clicking Rematch, don’t worry. You can rematch later by right-clicking on your new geocoding result shapefile, clicking **Data**, then clicking **Review/Rematch Addresses…**).



|  |
| --- |
| **Review/Rematch Addresses** |

**So here is what your rematch window looks like:**



In the top right of the window, you can see your breakdown of how many addresses are matched, tied, or unmatched. This will change as you successfully rematch addresses.

In the top left of the window, we have our list of addresses. By default, ArcGIS is showing you just the unmatched addresses right now – which makes sense, because they’re the ones we need to fix.

On the bottom half of the window, you have the address information and then a large white window – this is where possible matches (candidates) will appear.

Let’s try doing some detective work.

***NOTE: Your unmatched addresses may be in a different order to mine, because your Excel file might be sorted by a different column. Don’t worry – the procedure below is still the same.***

Some of these addresses are just blank – these are probably people who do not have a mappable address.

As I use the down arrow on my keyboard to scroll through the unmatched addresses, I find one that says 1113 Harrison Ave, 02118. There are no likely candidates. The first thing I might do is delete the zip code and see if that helps – maybe I had the wrong zip code. So delete the zip code and click the ‘Search’ button.

Still nothing. Let’s see if this is even a real address – try searching for ‘1113 Harrison Ave, Boston’ in Google Maps.

This came up with an address, but it says the zip is 02119, so try entering that into our rematch window and then clicking Search again:



Bam, we got one! There’s a 100% match (see the ‘Score’ column?). Click that candidate and then click ‘match’ at the bottom of the window. Now we can move on to the next address.

My next unmatched address is 1127 Commonwealth Ave. The rematch window already has a succession for me of 1128 Commonwealth Ave. It is only a 79% match (because the house number is different), but for our purposes, being one house off won’t make a difference. Let’s match that one.

My next unmatched address is 117 Townsend, 02119. There are no suggested candidates. As always, I’ll put that in Google and see what happens. Looks like another case of incorrect zip code – Google tells me the zip should be 02121. If I fix that in my rematch window, I get a 100% match. Thank you, next!

As I move through my unmatched addresses, I can see the matches I have made so far – note that the “Status” column says U for unmatched addresses and M for matched addresses.



Keep working through the list, trying to match as many as you can. The zip code is the most common culprit. Other issues are being just one house number off or having a typo in the street name. For example, I know one of these addresses says “HUNGTINGTON” when the street name is actually Huntington. Some addresses may just be unmappable. Maybe they are so incorrect that we can’t find a reliable match at all, or maybe they are mistakenly on our list – for example, 15 Pinewood St is in Mattapan, not Boston, and so it won’t map for us because we only have Boston streets on our map. If you come across truly unmappable addresses, just skip them.

You can also try looking at Google Maps to see what else is around that might be likely. For example, 30 Cornhill St won’t map for me at all, even though it shows up on Google Maps – but I can see that right near that address, there is the New England Center and Home for Veterans. This is a likely place that someone on our list might live. The address is 17 Court St, Boston – which should sound familiar, because a *lot* of other people in our list live there. It is probably safe to assume that this person really lives at NECHV and we could change the address accordingly.

Here’s where I ended up: 98% addresses matched, only 2% unmatched. That’s amazing! It was tedious work, but I’ve made sure that my data were as clean and useful as possible, and I’ve gotten as much valid info out of them as I can. Now my map of level 2 and 3 registered sex offenders in Boston is as accurate as I can make it, excluding those with no mappable address. See the next page for my map of Boston with point features indicating where our registered sex offenders live.

