



## Background Information

**Geocoding** is the process of assigning a geographic location to point data based on a description. The description usually comes in the form of street addresses, postal codes, or cities. The geocoding service then converts this descriptive information into a point feature on a map with precise location coordinates. In order to create the point feature, a **reference layer** such as a street file with addresses is required.

An **Address Locator** defines the process for converting these descriptions to points on a map by setting the parameters of the transformation. It is possible to rerun the geocoding service in order to match unmatched points interactively and thereby increase the percentage of matched points.

It is important to note that geocoding is not an exact science. Each point that is inputted into the geocoding service is compared with potential candidates in a **Reference Table**. The points are then assigned a score based on their sameness to points in the reference table. Scores that exceed a user defined percentage are automatically matched. Points that fall below the designated grade are not matched but can be rematched interactively.

Once the points have been geocoded a new output table containing 4 new auto generated columns will appear in your map view.

- The **status** field simply indicates whether the record was matched or not. The value “**M**” for matched addresses and “**U**” designates unmatched addresses.
- The **score** field lists the score of the candidate to which the address was matched as a percentage.
- The **side** field lists the side of the street that the address was matched on if that information is available. “**L**” means that the address is located on the left side of the street and “**R**” stands for right side of the street.
- **Arc\_Street** contains the address information that was geocoded.

Enumeration Area (EA) is a small area composed of one or more neighbouring blocks, used by Statistics Canada for distributing questionnaires to households and dwellings (census collection). All of Canada is divided into enumeration areas.

**Note:** Keep in mind if you are using ArcView 9.3/9.3.1 the steps to complete this tutorial will be different.

## Overview

### In this tutorial, you will:

- Create a geocoding service in ArcCatalog
- Add the geocoding service in ArcMap
- Geocode addresses in ArcMap
- Interactively rematch addresses
- Conduct a series of queries on the newly geocoded data

**Note:** To complete this exercise, you will need to download the dataset designed to be used in this tutorial. Download the data files by clicking either the **WinZip** or the **Self Extracting** link located just below this tutorial. The **Self Extracting** link will automatically extract the files to a folder called **geocode** on your **c: drive**.



Data for this tutorial has been provided courtesy of DMTI Spatial. If you are interested in similar data for your school area, ESRI Canada and DMTI Spatial have partnered to offer Canadian K-12 schools a Local Dataset. The Local Dataset is a complete, custom designed set of local data surrounding a school and includes a wealth of census data at the enumeration area level, a comprehensive street network file with address ranges, points of interest, land use and building footprints (where applicable), topographic features, and much more. Please contact [education@esricanada.com](mailto:education@esricanada.com) for more information.




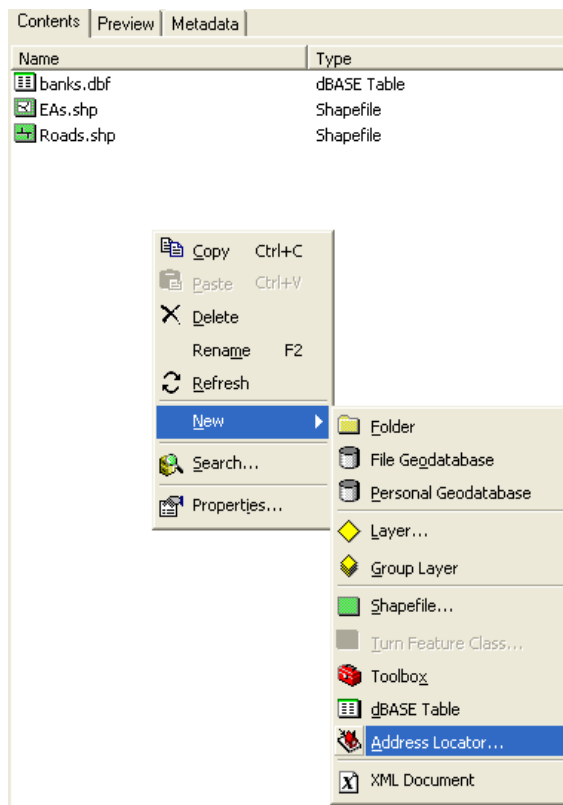
## Scenario

In this tutorial you are provided with a table of data that consists of 26 banks in the Oakville, Ontario area, a shapefile with enumeration boundaries data and a roads shapefile. You are going to geocode the banks using the roads file as a reference. Once the banks have been geocoded, then you will query the data to conduct some spatial analysis.

## Creating an Address Locator in ArcCatalog

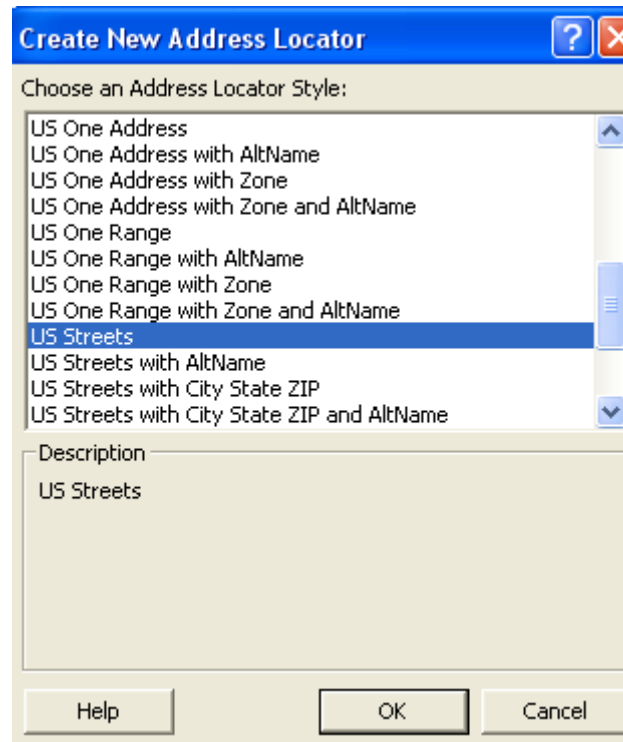
In order to geocode files in **ArcMap**, you must first create a **New Address Locator** in **ArcCatalog**.

1. Start Menu > Program Files > ArcGIS > ArcCatalog or if you are in ArcMap, double click on the ArcCatalog icon .
2. Navigate to the *geocode* folder (C:\geocode). Ask your teacher if you are not sure.
3. Inside the geocode folder, right click anywhere in the white space and then scroll down to **New** and then **Address Locator**.





- In the resulting dialog box, scroll down, select **US Streets** and click **OK**.



- In the **New US Streets Address Locator** dialog box, type in **Oakville** as the name of the service. Under the *Primary table* tab, click the **Browse** button  and browse to your *Reference data* (**geocode** folder). Select **Roads.shp**, and click **Add**.
- Decrease the **Spelling sensitivity** to **60**, accept all the other defaults and click **OK**.

**Note:** Reducing the spelling sensitivity and the minimum match score is important. Data in your table will sometimes have mistakes and data in the reference table could have mistakes or could be recorded in a different manner. If you set both scores equal to 100, the geocoding service would only match addresses that were identical in both tables. Reducing the scores allows the geocoding service to pick up on subtle differences in spellings and the way addresses are recorded.



**New US Streets Address Locator**

Name:

Description:

Primary table

Reference data:

Store relative path names

Fields

House From Left:

House To Left:

House From Right:

House To Right:

Prefix Direction:

Prefix Type:

Street Name:

Street Type:

Suffix Direction:

Input Address Fields

The field containing:	is recognized if it is named:
Street	Address
	Addr
	Street

Buttons: Add..., Delete, ↑, ↓

Matching Options

Place Name Alias Table...

Spelling sensitivity:

Minimum candidate score:

Minimum match score:

Intersections

Connectors:  Separate connectors by a space, e.g. "& @ , /"

Output Options

Side offset:  in

End offset:  %

Match if candidates tie

Output Fields

X and Y coordinates  Standardized address

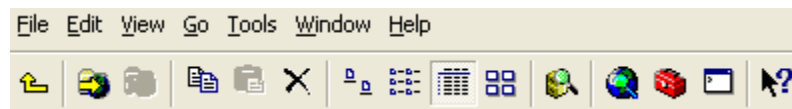
Reference data ID  Percent along

Buttons: Help, Advanced..., OK, Cancel

This will create the **Geocoding Service** that you can reference in **ArcMap** to geocode points. It is possible to change the **Matching Option** settings such as spelling sensitivity while you are geocoding in **ArcMap**, however it is important to note that those changes are not permanent. Those temporary changes are most often used when rematching unmatched records. In order to permanently change the geocoding service you must use **ArcCatalog**.

## Geocoding in ArcMap

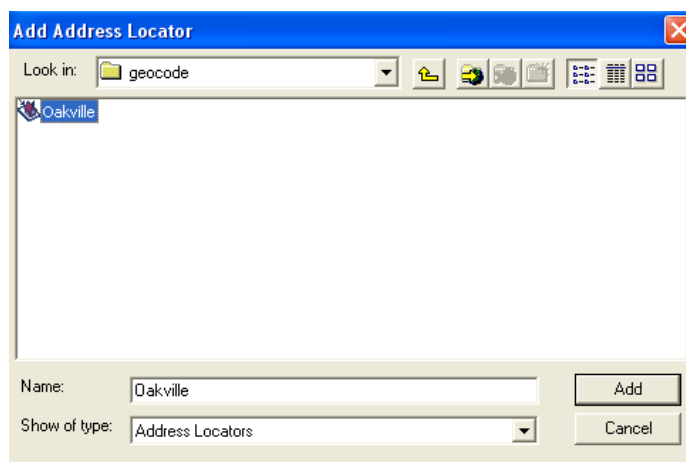
1. Start ArcMap by clicking the ArcMap  button at the top of the ArcCatalog screen.



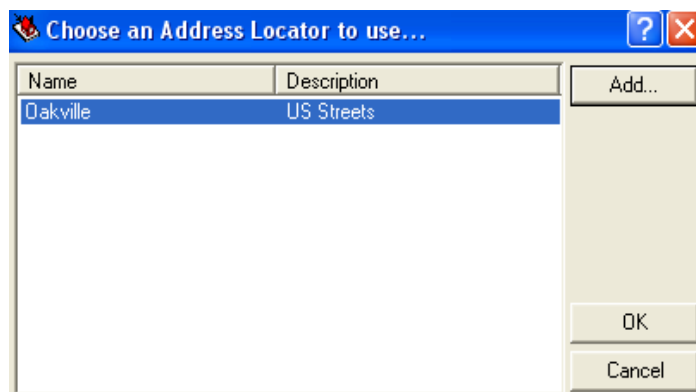
2. In the welcome window select "Start using ArcMap with:" **A new empty map**. Browse to the **:\geocode** or appropriate folder on your hard drive and select **banks.dbf**, **EAs.shp**, and **Roads.shp**. To select all the layers at once, hold down the **Shift** key. Click **Add**.




3. Right click on **banks** in the table of contents and select **Open** in order to examine the table.
4. Notice that the table consists of 26 records of various banks in Oakville. The column that you will be using to geocode the points is the **Address** field. **Close** the attributes of banks table.
5. Click on the **Tools** menu at the top of the screen and select **Geocoding** → **Geocode Addresses**.
6. You will need to add the **Address Locator** that you created in **ArcCatalog**. Click the **Add** button and browse to the **Geocoding services** folder.
7. Select the **Oakville** service/locator that you created and click **Add**.

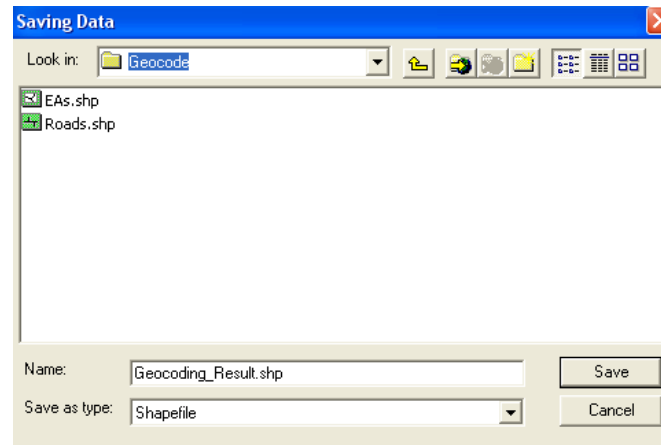


8. Click **OK** in the following dialog box.

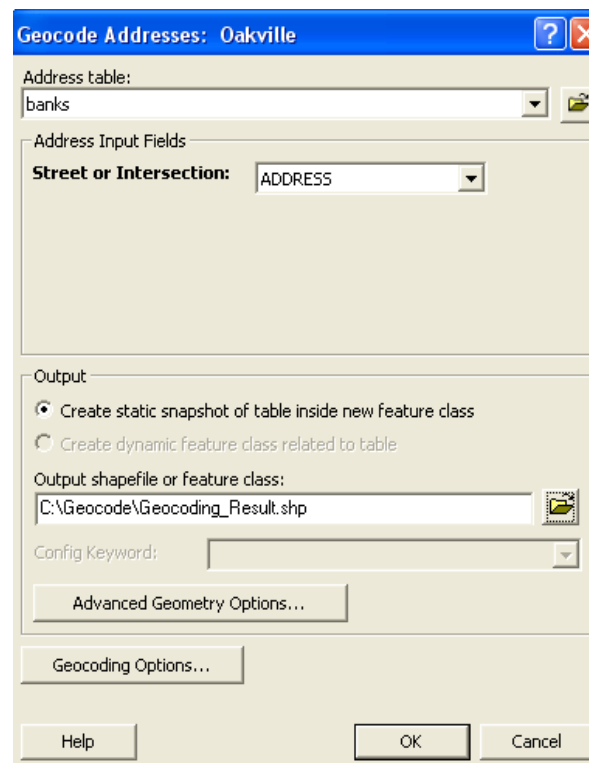




9. In the **Geocode Addresses** dialog box make sure that banks is the **Address table** and that **ADDRESS** is the Address Input Field. In the **Output** box, click the Browse button  and browse to your geocode folder. Leave the default as **Geocoding\_Result** and click **Save**.



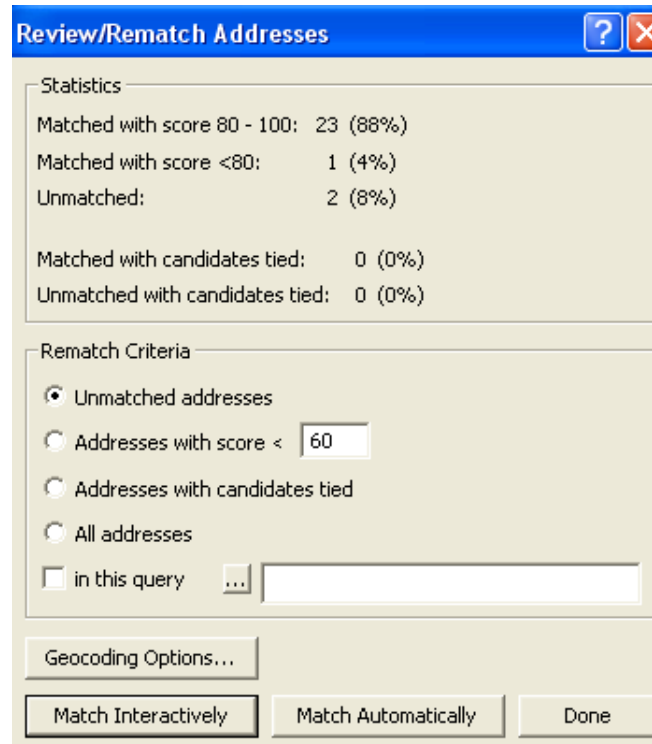
10. Click **OK** to execute the geocode.




**Note:** This tutorial was developed using ArcView 9.2, results may vary if you are using a different version of the software.



Once the geocode has been processed you will see a **Review/Rematch Addresses** window appear with the results of your geocoding. It should say that 23 addresses (or 88%) have a match score of 80% or greater and 2 addresses (or 8%) are unmatched. This is a very reasonable score however, considering that there are only two outstanding records it is worth trying to rematch them interactively



- Click on the **Match Interactively/Rematch** button in order to step through each address in an attempt to match them individually. Maximize  the *Interactive review* window.

**Note:** In ArcView 9.3/9.3.1 under the show results drop down menu, change to unmatched addresses.

- Highlight the first record which should be 277 EAST LAKESHORE RD. The *Candidate* window at the bottom provides a list of potential match for this address, however the current score for it is only 52% which is quite low. Examine the candidate carefully though. What are the differences between it and the record you're attempting to geocode?



**Interactive Review**

FID	Shape	Status	Score	Side	ARC_Street	
11	Point	U	0		277 EAST LAKESHORE RD	POI3225867
15	Point	U	0		282 AKESHORE RD E	POI3227742

Record: 1 Show: All Selected Records (of 2)

Street or Intersection:  
277 EAST LAKESHORE RD

Standardized address:  
Modify... 277 | E | LAKESHORE | RD |

1 Candidate

Score	Side	LeftFrom	LeftTo	RightFrom	RightTo	PreDir	PreType	StreetName	StreetType	SufDir
52	L	245	279	250	276			LAKESHORE	RD	E

Geocoding Options... Zoom to: Candidates Original Extent Search Match Unmatch Close

Notice that the directional specification *EAST* is listed before *LAKESHORE* in the **Banks** table address spelling at the top of the interactive review window. Click on the **Modify** button (just above the Candidate window). This will show you how the address has been broken down by the geocoding service and allow you to conduct any necessary edits. The geocoding service has placed the letter “E” in the PreDir (PreDirectional) window because the directional word *EAST* came at the beginning of the address. However, in the candidate window *EAST* is listed as a **SufDir**.

**Edit Standardization**

HouseNum: 277

PreDir: E

PreType:

StreetName: LAKESHORE

SufType: RD


SufDir:

- With a little research, you discover the address order is incorrect because the “East” should come after *Lakeshore*. (Hint: Compare this Lakeshore record with other records in the table.) In order to fix this mistake in the address, delete the letter “E” in the **PreDir** textbox and enter it into the **SufDir** textbox. Press the **Enter** key.

This should prompt some more potential *candidates* to appear in the lower window. Notice the **score** in the column on the left. The *candidates* are automatically sorted by **score** with the highest appearing at the top. The initial candidate should now have a score of 100. Notice that this is the same candidate as before but the score has been improved since you corrected the data.

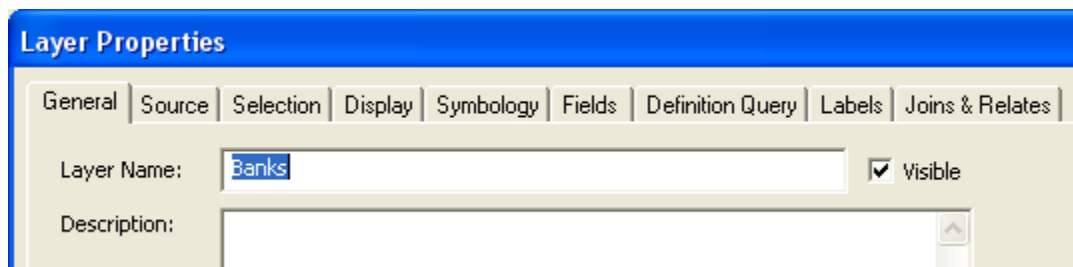


57 Candidates											
Score	Side	LeftFrom	LeftTo	RightFrom	RightTo	PreDir	PreType	StreetName	StreetType	SufDir	Match_addr
100	L	245	279	250	276			LAKESHORE	RD	E	277 LAKESHORE RD E
53	L	283	315	282	312			LAKESHORE	RD	E	283 LAKESHORE RD E
52	L	207	239	216	240			LAKESHORE	RD	E	239 LAKESHORE RD E
52	L	321	351	326	350			LAKESHORE	RD	E	321 LAKESHORE RD E
51	L	165	197	166	200			LAKESHORE	RD	E	197 LAKESHORE RD E
50	L	389	407	390	410			LAKESHORE	RD	E	389 LAKESHORE RD E
50	L	125	159	126	162			LAKESHORE	RD	E	159 LAKESHORE RD E
49	L	417	435	422	438			LAKESHORE	RD	E	417 LAKESHORE RD E
47	L	489	515	492	512			LAKESHORE	RD	E	489 LAKESHORE RD E
46	L	521	535	528	544			LAKESHORE	RD	E	521 LAKESHORE RD E
38	R	0	0	474	474			LAKESHORE	RD	E	474 LAKESHORE RD E
38	R	0	0	468	468			LAKESHORE	RD	E	468 LAKESHORE RD E

- It is clear the first candidate is the better match. Make sure it is highlighted and then click on the **Match** button (bottom right). This will add this record and its score to the table of geocoded/matched addresses and the status will change from **U** to **M**.
- Close  the **Edit Standardization** window and highlight the remaining unmatched address (282 LAKESHORE RD E). What is the problem with this last record?
- Hopefully you noticed that the spelling of this address is incorrect. The street should in fact be spelled like the previous record, *LAKESHORE RD E*. Click on the **Modify** button and change the *StreetName* to **LAKESHORE** in the *Edit Standardization* window and press the **Enter** key.
- This will prompt 57 potential candidates to appear in the candidate window. The candidate with the high score of 100 is the most appropriate match.

57 Candidates											
Score	Side	LeftFrom	LeftTo	RightFrom	RightTo	PreDir	PreType	StreetName	StreetType	SufDir	Match_addr
100	R	283	315	282	312			LAKESHORE	RD	E	282 LAKESHORE RD E
53	R	245	279	250	276			LAKESHORE	RD	E	276 LAKESHORE RD E
52	R	207	239	216	240			LAKESHORE	RD	E	240 LAKESHORE RD E
52	R	321	351	326	350			LAKESHORE	RD	E	326 LAKESHORE RD E
51	R	0	0	360	370			LAKESHORE	RD	E	360 LAKESHORE RD E
51	R	165	197	166	200			LAKESHORE	RD	E	200 LAKESHORE RD E
50	R	389	407	390	410			LAKESHORE	RD	E	390 LAKESHORE RD E
49	R	125	159	126	162			LAKESHORE	RD	E	162 LAKESHORE RD E

- Select that record and click **Match**.
- All 26 points should now be matched. Close the **Edit Standardization** window and the **Interactive Review** window and click **Done** on the **Review/Rematch Addresses**.
- The new geocoded layer will be added to your table of contents. Click on its symbol to bring up the **Symbol Selector**. Click on the **More Symbols** button to show the other symbol themes and select **Business**. This will add the business symbols to your selection list. Choose one of the **Bank** symbols and click **OK**.
- Double click on the new layer name to bring up the **Layer Properties**. Click on the **General** tab and change the layer name to **Banks**.

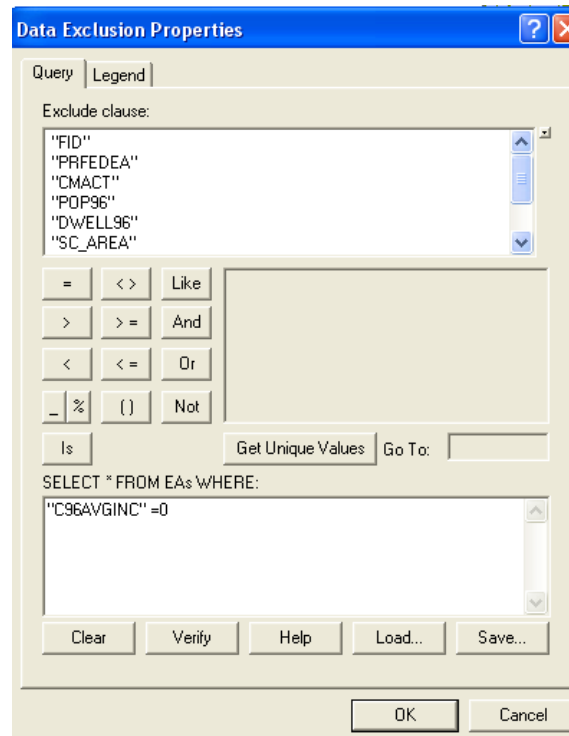


Now you can conduct some analysis about the location of the banks. What areas might you expect the banks to be located in?

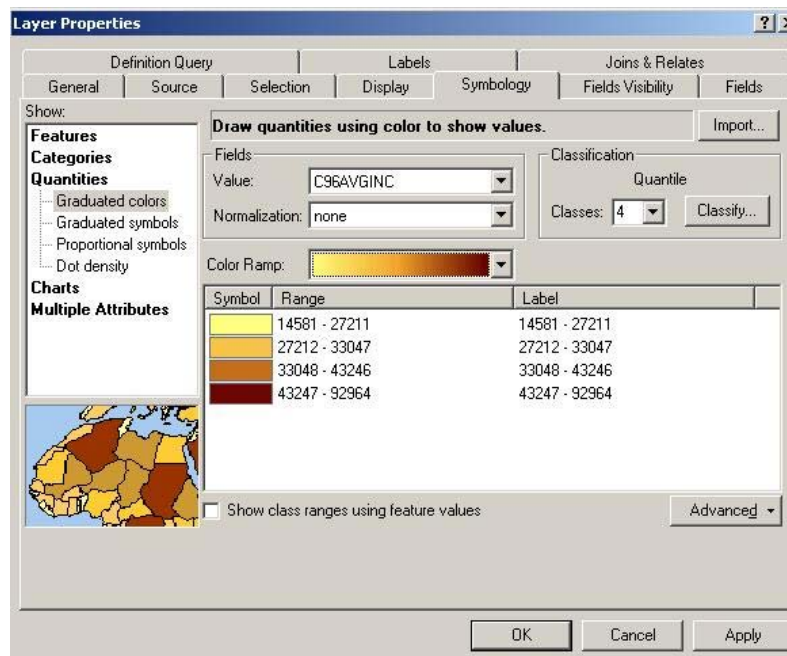
### Thematic Map

*In this section you are going to make a graduated color map of average income to visually analyze whether there is a correlation between the location of banks and the wealthiest areas in Oakville.*

1. Double click on the **EAs** theme in your table of contents to bring up the **Layer Properties**.
2. Click on the **Symbology** tab and click on **Quantities** → **Graduated Color**.
3. Click on the **Value** drop down menu and choose **C96AVGINC**.
4. Click on the **Classify** button. Choose **Quantile** for the **Method** and **4** for the number of **Classes**. This way you are roughly dividing up Oakville's population into quarters based on wealth.
5. You will notice from the graph that there are a few **0** or **Null** values that are skewing the intervals. Click on the **Exclusion** button. Double click on the **C96AVGINC** in the **Field** window. Click on the **=** sign and click on or type **0**. Click **OK twice** to go back to the *Layer Properties* window.



6. Click **OK** to apply the changes to your map.



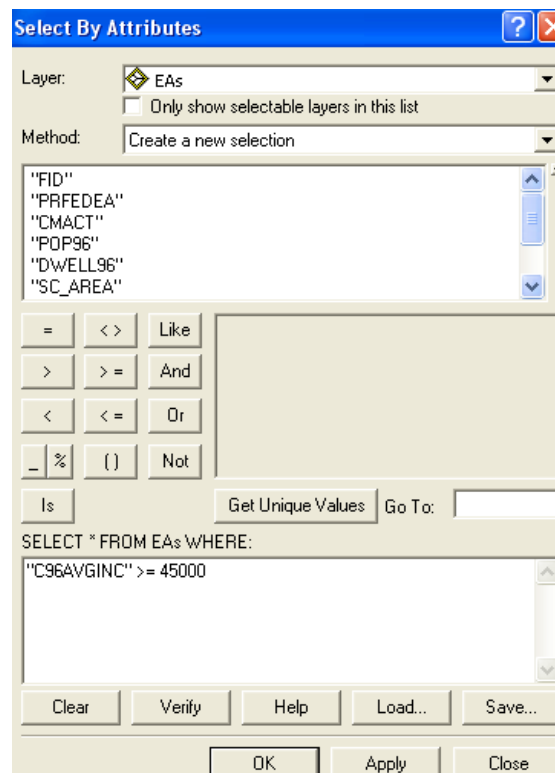
You will now have a graduated color map of average income in Oakville. Where is most of Oakville's wealthiest population located?



### Querying based on Attributes

You are now going to select the wealthiest areas in Oakville; you want to select all EAs with an average income of \$45,000 or more.

1. Click the **Selection** drop down menu and choose **Select by Attributes**.
2. Choose **EAs** as the layer and **Create new selection** as the **Method**.
3. Double click on **C96AVGINC** in the **fields** window, click on the “>=” sign, and type “**45000**”. Click **OK**.



You should now see a number of enumeration areas highlighted in blue. Right click on the **EAs** theme in your table of contents and select **Open attribute table**. You should see **12** of **71** records selected. Close the attribute table for **EAs**.

### Querying by Location

You are now going to see how many **Banks** are contained by or within 500 metres of these 12 enumeration areas.

1. Click on the **Selection** menu and choose **Select by Location**.
2. In the resulting *Select By Location* dialog box, choose **select features from** the following layer(s): **Banks** that: **are within a distance of** the features in this layer: **EAs**.



- Make sure the **Use selected features** box is checked, so that only the 12 selected features will be incorporated into the query. **Apply a buffer to the features in EAs of 500 Meters.** Click **Apply** and **Close**.

- Right click on the **Banks** theme in the table of contents and select **Open Attribute Table**. Notice that **15 out of 26 Banks** are located within 500 metres of one of the wealthiest census tracts.

*More than half the banks are located within close proximity to the wealthiest quarter of the population. What other conclusions can you draw about the location of banks? Conduct some analysis on your own! Pretend you are the owner of a chain of banks, where would be an ideal location to build a new one?*