GST 102: Spatial Analysis Lab Series

Lab 3: Using Attribute and Spatial Queries for Data Exploration

Document Version: 2013-07-31 (Beta)

Organization: Del Mar College
Author: Richard Smith

The development of this document is funded by the Department of Labor (DOL) Trade Adjustment Assistance Community College and Career Training (TAACCCT) Grant No. TC-22525-11-60-A-48; The National Information Security, Geospatial Technologies Consortium (NISGTC) is an entity of Collin College of Texas, Bellevue College of Washington, Bunker Hill Community College of Massachusetts, Del Mar College of Texas, Moraine Valley Community College of Illinois, Rio Salado College of Arizona, and Salt Lake Community College of Utah.

Contents

Introduction 3
Objective: Understanding Data and Using Spatial Queries 3
Lab Settings 3
1 Using the Data Dictionary and Attribute Selections 4
1.1 Data Dictionary 5
1.2 Selections 6
2 Buffering 8
2.1 Using Buffering 8
2.2 Spatial Queries 11
Conclusion 12
Discussion Questions 12

Introduction

This lab is part of a series of lab exercises designed through a grant initiative by the National Information, Security & Geospatial Technologies Consortium (NISGTC), funded by the United States Department of Labor in partnership with the Department of Education under the Trade Adjustment Assistance Community College and Career Training Grant Program (TAACCCT).

In this lab, students will explore data and decipher the data using a data dictionary or ancillary files to select the data they need. The students will need to perform some queries using ArcGIS on census data. The students will also create a buffer and learn the importance of buffering in GIS.

Your instructor may require that you provide screen captures and/or exported files. Please check with your instructor for the requirements specific to your class.

This lab includes the following tasks:

1. Using Data Dictionaries and Attribute Selections
2. Buffering

Objective: Understanding Data and Using Spatial Queries

The Objective of this exercise is to expose the importance of knowing the data you are working with as well as being able to make selections and query the data.

Lab Settings

Required Virtual Machines and Applications

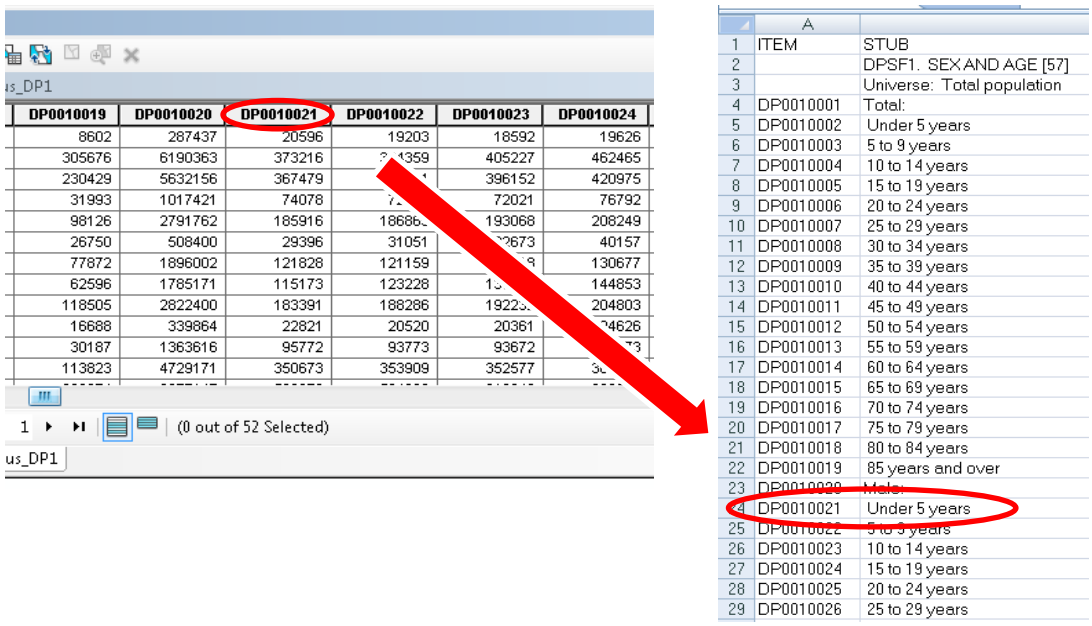
Windows Machine User Account	Train
Windows Machine User Password	Train1ng$

1 Using the Data Dictionary and Attribute Selections

Data dictionaries provide information on the fields within a dataset. They are often in electronic format and may be provided along with datasets, to allow the user to have an understanding of the data, and determine how to use it appropriately to meet their needs.

We will be using several datasets containing attributes that have been coded in a specific way to ease capture and minimize naming convention issues.

In this task, we will look at the data dictionary given to us with census data (See Figure 1). The census data captures many attributes about the population so that analysis can be done on the data and conclusions can be drawn on the trends in the population.



DP0010019	DP0010020	DP0010021	DP0010022	DP0010023	DP0010024
8602	287437	20596	19203	18592	19626
305676	6190363	373216	[illegible]	405227	462465
230429	5632156	367479	[illegible]	396152	420975
31993	1017421	74078	[illegible]	72021	76792
98126	2791762	185916	[illegible]	193068	208249
26750	508400	29396	31051	[illegible]	40157
77872	1896002	121828	121159	[illegible]	130677
62596	1785171	115173	123228	[illegible]	144853
118505	2822400	183391	188286	[illegible]	204803
16688	339864	22821	20520	20361	[illegible]
30187	1363616	95772	93773	93672	[illegible]
113823	4729171	350673	353909	352577	[illegible]

(0 out of 52 Selected)

	A	
1	ITEM	STUB
2		DPSF1. SEX AND AGE [57]
3		Universe: Total population
4	DP0010001	Total:
5	DP0010002	Under 5 years
6	DP0010003	5 to 9 years
7	DP0010004	10 to 14 years
8	DP0010005	15 to 19 years
9	DP0010006	20 to 24 years
10	DP0010007	25 to 29 years
11	DP0010008	30 to 34 years
12	DP0010009	35 to 39 years
13	DP0010010	40 to 44 years
14	DP0010011	45 to 49 years
15	DP0010012	50 to 54 years
16	DP0010013	55 to 59 years
17	DP0010014	60 to 64 years
18	DP0010015	65 to 69 years
19	DP0010016	70 to 74 years
20	DP0010017	75 to 79 years
21	DP0010018	80 to 84 years
22	DP0010019	85 years and over
23	[illegible]	Male:
24	DP0010021	Under 5 years
25	[illegible]	5 to 9 years
26	DP0010023	10 to 14 years
27	DP0010024	15 to 19 years
28	DP0010025	20 to 24 years
29	DP0010026	25 to 29 years

Figure 1. The data dictionary (shown on right) assigns a code for data that falls within an age category. These codes are then used as the field headers for the corresponding data in the attribute table (shown on left).

1.1 Data Dictionary

1. Log into the computer, using the information provided in the Lab Settings section.
2. The data for this lab is located on the lab machine. Copy the *Lab 3* folder from the shared drive into your *C:\GST 102 folder.*
3. Click **Start->All Programs->ArcGIS->ArcCatalog 10.1**. ArcCatalog will open.
4. Click **File->Connect To Folder…** navigate to, and select *C:\GST 102\Lab 3* from the folder listing and click OK (if an easily accessible connection is not already established).
5. Inside the *Lab 3* folder, there is a data folder. All the data for this lab will be stored in this folder. The **State_2010Census_DP1** contains the data that will be used for this lab.
6. There are two files in the data folder named **DP_TableDescriptions.xls** and **State_2010Census_DP1.shp**.
7. The data dictionary is located in **DP_TableDescriptions.xls**. This worksheet provides a description for each of the attributes in the shapefile.
8. Open **ArcMap 10.1** and create a blank map
9. Add the **State_2010Census_DP1.shp** to the map by dragging it from **ArcCatalog** and dropping it in the **ArcMap** display window.
10. We want to map DP0010001, however we need to know what we are mapping before we can symbolize the data.
11. In ArcCatalog expand the **DP_TableDescription.xls** and select the **Table$** file and preview the data. This is the data dictionary for this dataset. Find DP0010001 and see what this code actually means.
12. In ArcMap, select State_2010Census_DP1.shp. **Right Click->Properties** and go to the **Symbology tab**. Select **Quantities** and **Graduated Colors**. For the **Value**, select **DP0010001** and Click **OK**.
13. **Do not Exit ArcMap** as we will be using this map for attribute selections in the next task.

1.2 Selections

1. Using the map you have just created in Task 1.1, we are going to interrogate the data we have mapped. On the menu bar in ArcMap **click -> Selection -> Select By Attributes.** The Select By Attributes window will be displayed. Using this panel, we will construct the queries and select the data. See Figure 2 to get to know the Select By Attributes query builder.

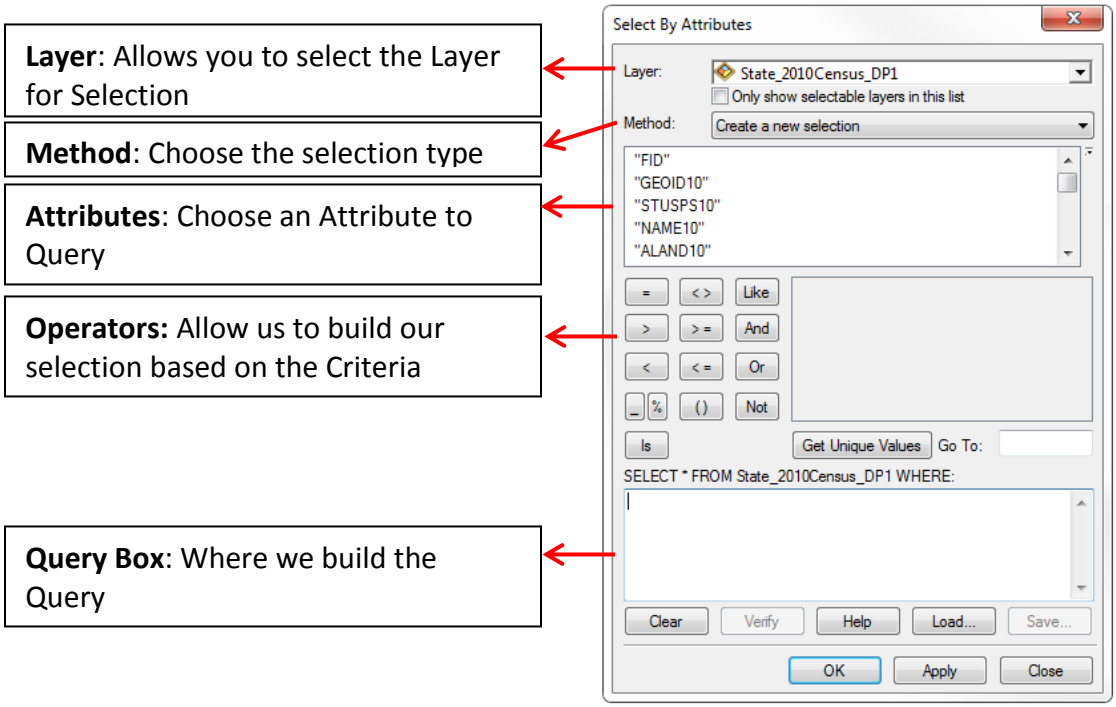


Figure 2. The Select By Attributes query builder and its features.

2. Construct a query, identifying where the total population is less than 1,000,000 people, it should look similar to Figure 3 below. We need to remember that the attribute names are coded, so we will need to make use of our data dictionary, in order to verify that the appropriate fields are used.

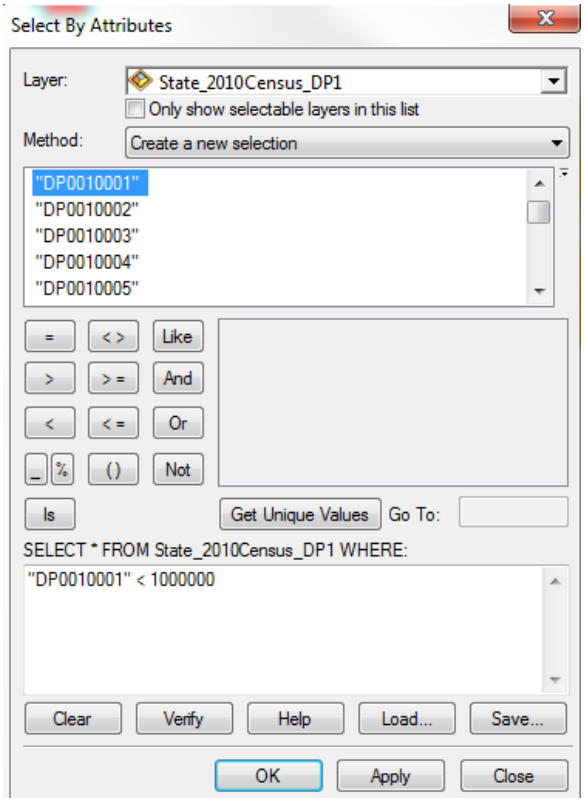


Figure 3. Query for selecting states with a total population <1,000,000

3. Before we run the query, click the **Verify** button to ensure the query will work.
4. Run the query, 8 records should be selected (verify this by looking at the attribute table).
5. Experiment with running some additional queries. Before you make another selection, clear the selected features to ensure none of them are added to the selection. We clear the selected features by clicking the clear selected features button on the toolbar.
6. Do an attribute selection to see how many states have a population of more than 10,000,000(i.e. "DP0010001" > 10000000).

2 Buffering

Buffering is a key analysis tool in GIS. It gives us the ability to create a threshold distance where we can see if features are within or outside of the threshold.

2.1 Using Buffering

1. We will now look at buffers, using the tornado dataset that is in your data file.
2. In a blank map file in ArcMap, open the **Tornado.mxd** that is inside the **tornado** folder in ArcMap. You should see a map like Figure 4 shown below.
3. On the Menu bar, click **Bookmarks->Tornado Path.** The red line on the map represents the path of a tornado that has gone through the area. The approximate area of damage was 900m around the tornado. The green polygons represent schools in the area.

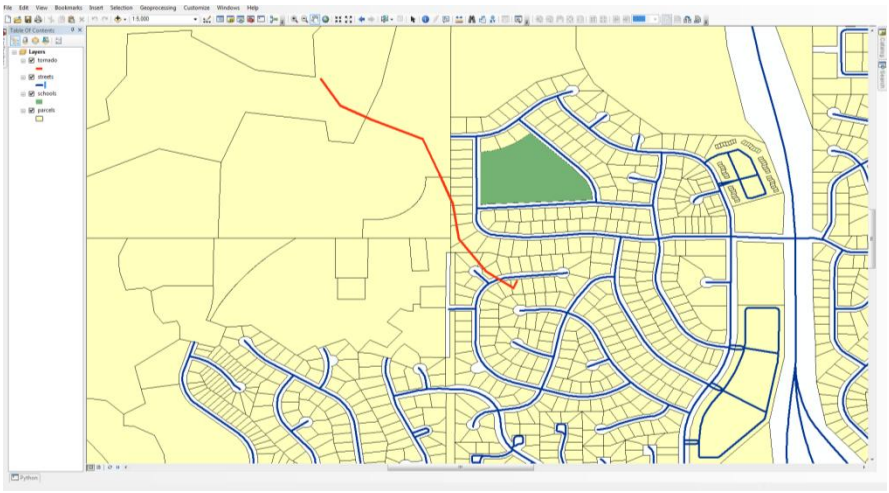


Figure 4. Tornado.mxd in ArcMap

4. To see the area affected by the tornado we are going to create a 900m buffer around the red line so we can easily see the area affected by the tornado.
   a. In ArcMap, click  to open ArcToolbox
   b. Once the toolbox is open, we must expand the Analysis Toolbox and then the Proximity toolset as shown in Figure 5.

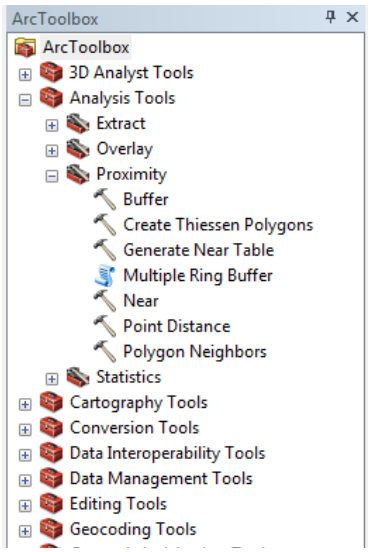


Figure 5. Finding the Buffer tool in ArcToolbox.

5. Double-click on the **Buffer** tool and enter the following parameters so they look similar to those found in Figure 6 below:
   a. Input Features : **`tornado`**
   b. Output Feature Class: Navigate to your *Lab 3* data folder and name it **`TornadoBuffer.`**
   c. Linear Unit: `900` meters
   d. Dissolve Type: ALL
   e. **Click OK** and wait for the buffer tool to build the new layer**.**

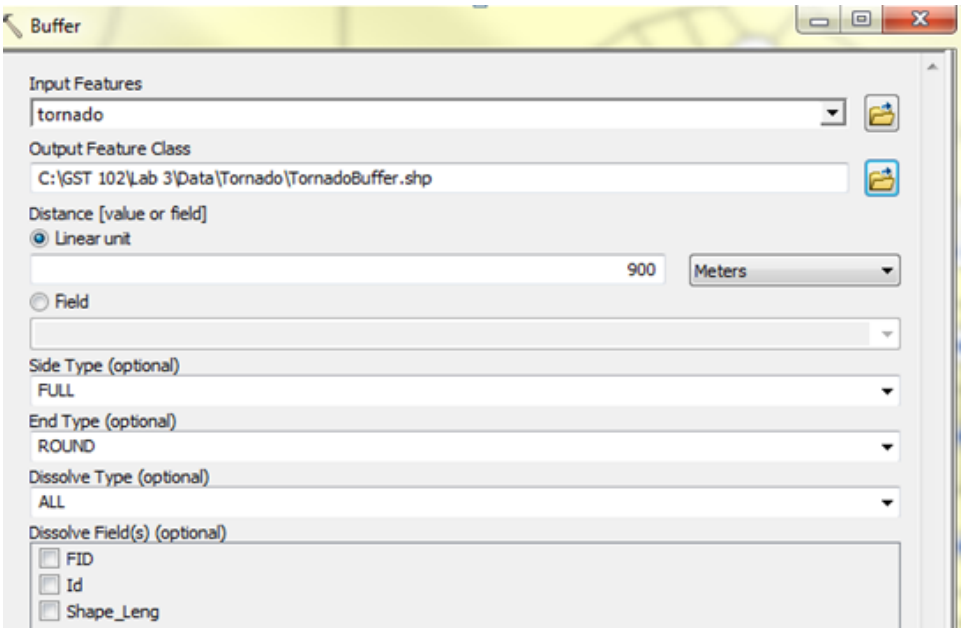


Figure 6. Settings to create a 900m buffer layer surrounding the tornado path.

6. This should create a polygon that is 900 meters away on all sides from the tornado path. In order to see what kind of locations were specifically affected by the tornado, we need to set the transparency on the buffer so we can see through our new layer. Right-click on tornado buffer layer and click **Properties**. Go to the **Display tab** and set the transparency to 50% and click **OK**.
7. Looking at the result, we can immediately see the areas affected by the tornado (see Figure 7). This is a prime example of how buffers are used. Just looking at the map, we can see the areas affected by simple inspection; however, that may not provide us with the accuracy we need for our specific analysis. In the next portion of the lab, we will be looking at more precise methods of analysis.
8. Save your map to your Lab 3 folder and keep it open for Task 2.2.

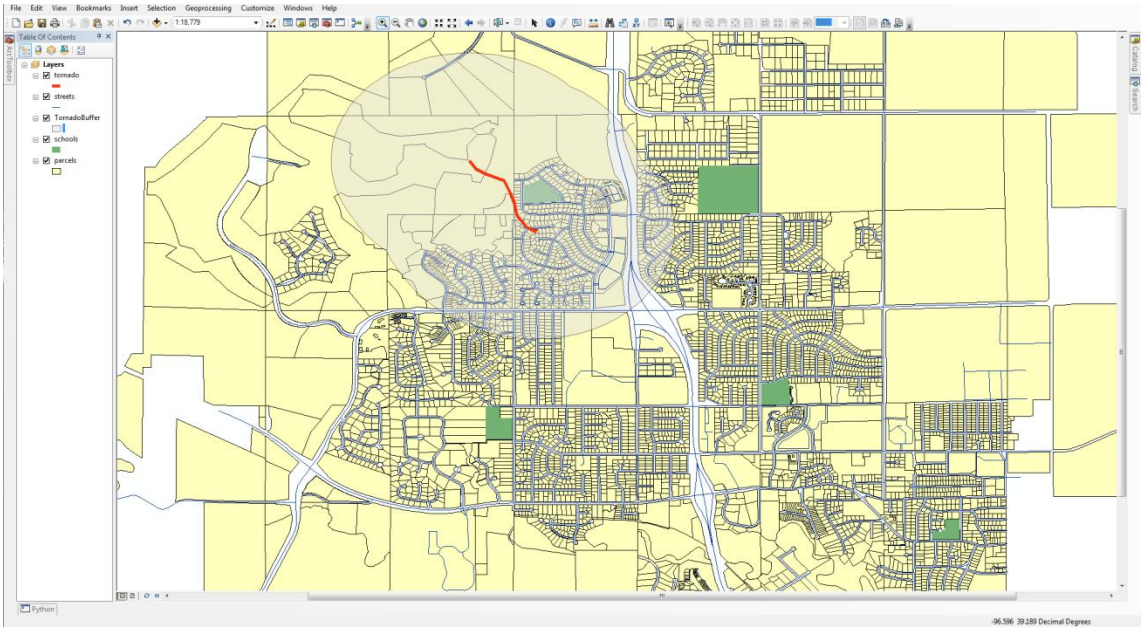


Figure 7. Tornado map showing affected areas by using the Buffer tool

2.2 Spatial Queries

1. Use the map you have just created in Task 2.1.
2. Go to Selection in the main menu and click **Select By Location**, using Select By Location allows us to make spatial queries using the layers inside the map document. Orient yourself to the Select By Location dialog box by examining Figure 8 below.

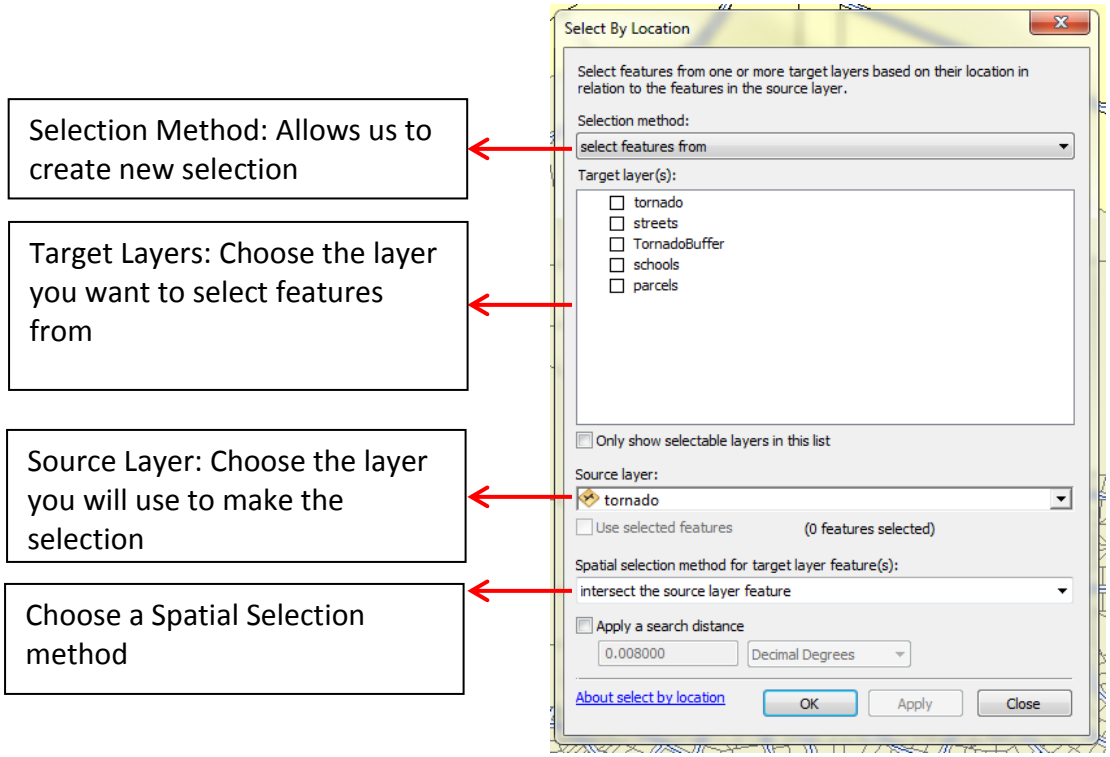


Figure 8. Select By Location dialog box explaining the different aspects of selecting by location.

3. In this selection, we will use the TornadoBuffer layer to determine the affected Parcels and Schools. Use the parameters shown in Figure 9, click OK, and view the selection.

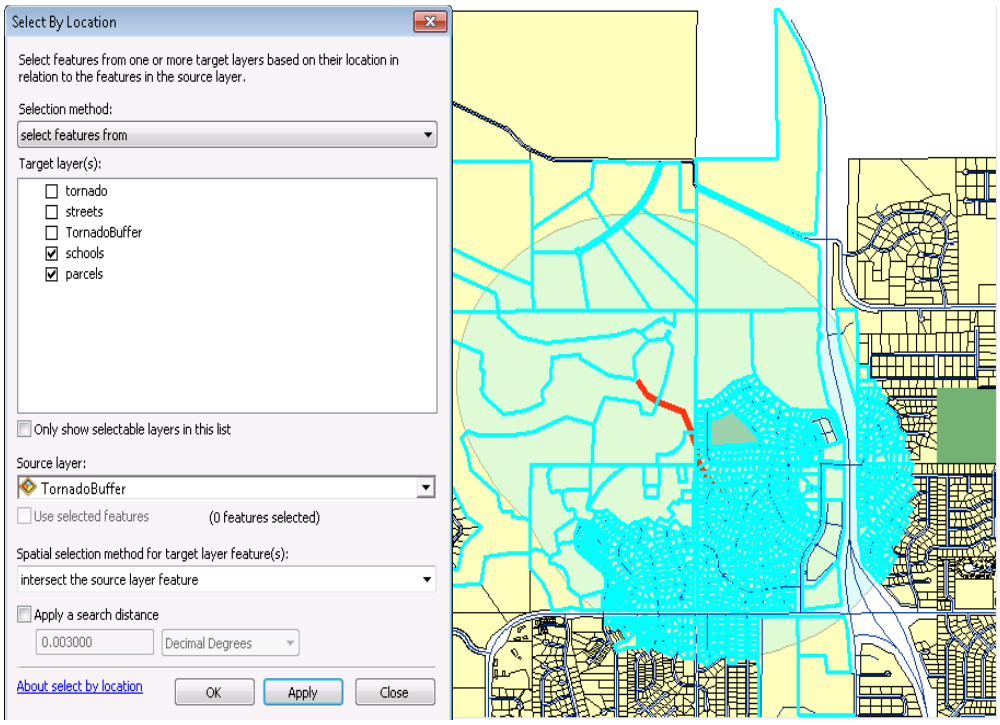


Figure 9. Select By Location box (left) with result of selection (right)

4. Looking at the result on the right in Figure 9, we can see that there are a fair number of parcels and one school that lie within the path of damage caused by this tornado. Open up the attribute tables for the **parcels** layer and the **schools** layer to view more specific details on the selections.

Using the Select By Location, we now have a more accurate selection and more specific data on the area of interest within this GIS dataset.

Conclusion

In this lab, you explored the use of a data dictionary with a dataset. We also looked at basic attribute queries.

Discussion Questions

1. Why do we need data dictionaries?
2. How are attribute selections useful in a GIS?
3. Why are buffering and spatial selections important to us?