Creating & Adding Geospatial Data:
Finding & Adding Geospatial Data to ArcGIS

*****NOTE:** This is a standalone lab that is not associated with any chapter in the textbook.

There are several different ways to acquire data to incorporate into ArcGIS. This includes online data resources, field collection of data, and the digitizing of data.

The most common way to incorporate data into ArcGIS is to find data produced by someone else. In addition to becoming familiar with finding the data, this lab will also explore how to download, evaluate, and incorporate the data into ArcGIS.

For this tutorial you will be exploring the city of Youngstown, OH.

1. Since you will be collecting data from different sources online, you should first create a folder for all of the data you acquire online. I create a folder (in the folder for the lab (ie. Lab04)) labeled "*Downloads*" to save all of your downloaded data into.

The *National Map* is the public interface for the National Geospatial Program. This program was established to provide geospatial data to the public. The data comes from a variety of different sources and is of a high-quality.

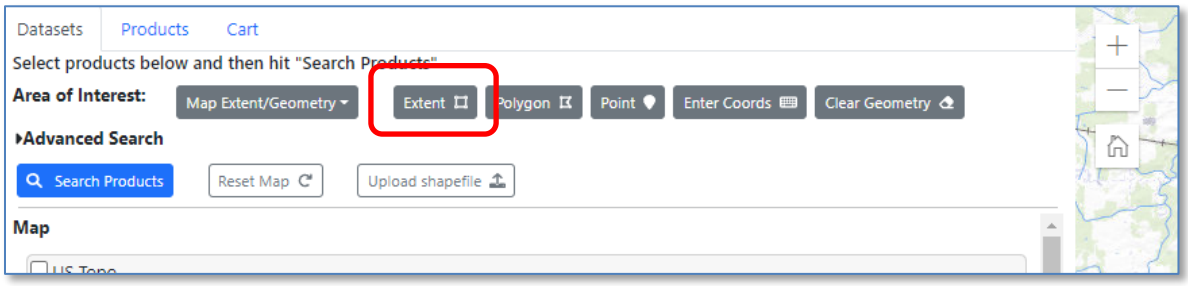
2. Open the *National Map* website: https://www.usgs.gov/programs/national-geospatial-program/national-map

You will see lots of information on the home page including Training Courses, Data, etc. All freely available to the public as a digital download.

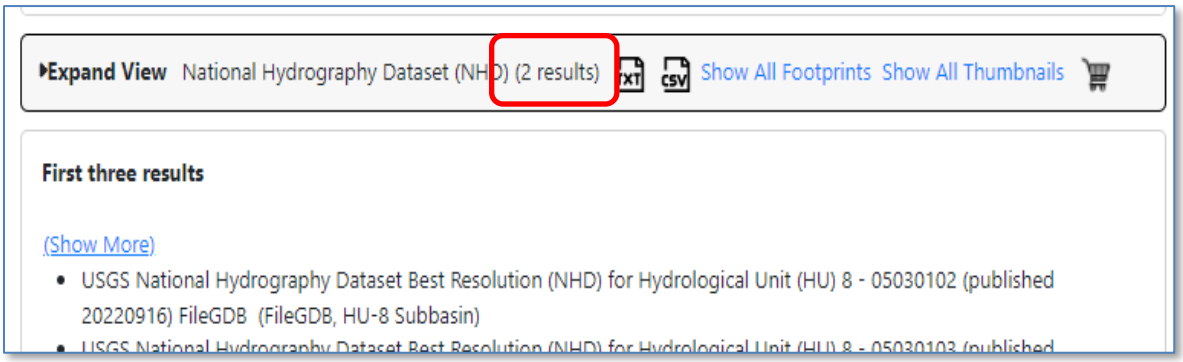
What we want is the *National Map Viewer* to identify what information is available for a specific location. In this case, for Youngstown, OH

- Select **The National Map Viewer** option.
    - This will open a zoomable map you can use to search for a location.
- Click **Data Download** on the upper left side of the screen (under USGS logo).
    - This will open a searchable map.
- In the search box type *Youngstown* and select *Youngstown, OH, USA* from the dropdown list.
    - The map should zoom in to the greater Youngstown area

- On the left side of the screen, select the Extent button and draw a rectangle that includes Youngstown (make sure NOT to include Pennsylvania in the box). This bounding box will limit your results to just those datasets that include data inside the extent.



- On the left side of the screen will be a list of all of the different datasets that are available through the national map. Select the following data layers:
    - Boundaries – National Boundaries Dataset
    - Hydrography – (NHDPlus HR, NHD, WBD)
    - Structures – National Structures Dataset
    - Transportation
- When you select a data layer it will expand and show several options under it. Make sure that you select **State** for the *Data Extent* and select **FileGDB** for the *File Format* for each dataset.
    - For *Hydrography*, make sure that you select **National Hydrography Dataset (NHD)** for *Subcategories*, and **HU-8 Subbasin** for *Data Extent.*
- Click **Search Products**
- A new tab should open with a list of the layers you selected. The results will be grouped according to the datasets you selected.
    - The **Structures**, **Transportation** and **Boundaries** datasets should only contain 1 result since the files are for the entire state.
    - You may find more than 1 result for the **Hydrography** dataset depending on how you drew your bounding rectangle.



    - **Note:** if more than one dataset shows up for a given layer, it is because both datasets cover your search area – Youngstown.
    - To determine which layer would be best to use click on the "**(Show More)**" button (see above). This will expand the list of results for that dataset.
    - Click on the **Thumbnail** button next to each layer. It will display the area covered on the map. Use this to select the dataset most appropriate for your needs.

    - For this case you will want to use the dataset for **Hydrological Unit (HU) 8 – 05030103**, because this dataset covers all of the city of Youngstown.
- Click on the **(Show More)** link for each dataset.
- Click on the **Download Link (Zip)** button for each result to download the file.
    - **NOTE:** Be sure to save the downloaded file in the *Downloads* folder you created for the lab.
    - **NOTE:** Notice that there is also a *Vendor Metadata* button. Click this to see the metadata of the files you selected. Often, metadata for downloadable files will be on the website and not with the file itself.

All of the National Map data you downloaded will be in a compressed format. So, the first thing you will have to do is unzip the data. The zip files are self-extracting, so they are easy to unzip.

3. Unzip the files in Windows Explorer
- Navigate to your Downloads folder. Create a new folder for each download and label appropriately (e.g., transportation, structures, hydrography & boundaries). The default folder names can be confusing.
- Unzip each file.
- **NOTE:** I would recommend that you keep the original downloaded zipped files. IN case your data becomes corrupted accidentally you will not have to go back to the national map to redownload the data. If you are worried about space on your hard drive, you can always delete the data once this lab is completed.

Before you do anything else, you should verify that all of the data you downloaded is in the correct location and readable by ArcGIS.

4. Open ArcGIS Pro
- Create a new Map.
- Using the Catalog pane, navigate to the location where you saved the unzipped files.
- There should be 3 files in each folder. A **.jpg**, a **.gdb**, and **.xml**.
- Explore the Metadata for each file (right-click on the file in the Catalog pane and select *View Metadata*).

Question 1: What do you notice about the metadata for the three files? What is wrong here in terms of what you have learned in class? Why do you think this is the case? (**HINT**: What is the purpose of the .xml file?)

If you expand the geodatabases for each of the datasets you downloaded, you will see a large number of feature classes contained within them (You will also find Feature Dataset that will need to be expanded to see the feature classes within them). We do not need all of them for the tutorial. In fact, some of them may be empty because the features they contain may not be present in the current location (Youngstown, OH).

5. Add the following data to the map:
- *GU_IncorporatedPlace* – Boundaries for the area.
- *Struct_Point* – Different structures in the area.

- *Trans_RailFeature* – Railroads in the area.
- *Trans_RoadSegment* – Roads in the area.
- *NHDLine* – Non-stream water features (ie. Levees) in the area.
- *NHDFlowline* – Streams in the area.
- *NHDWaterbody* – Bodies of water (ie. Lakes) in the area.

For reference, you may want to add the state boundary for Ohio. You can use the **states** feature class from the **Usdata** geodatabase.

Question 2: What datum, coordinate system, and units of measurement are the National Map data in? (Keep in mind that all of the downloaded layers have the same spatial reference.)

When you add the data in you will notice that it is for the entire state of Ohio (except for hydrography), and it may take a long time to load some of the layers. This is a situation where you have a very large data file, and it is too cumbersome to work with the entire layer when you are only concerned with one area of it. Therefore, you need to reduce the size of the data layers you are working with. In this case, it is the roads layer.

NOTE: This is very common when downloading GIS data. The data extent provided is often at a larger scale than you will need for a given analysis.

NOTE: To reduce the size of a file you will want to reduce the area that is included in the file. There are two ways to do this. One is through a query of the attribute data (ie. *State Name = Ohio*), and then creating a new layer from the selection. The other method is by cropping the data layer to the extent of a different layer. This is known as *Clipping*.

- **Select by Attribute**: This tool allows you to identify a feature (or several features) based upon a given field in the attribute table (for example, the states feature class contains a field labeled *State_Name*. This field can be used to select the state of Ohio by selecting the feature where *State_Name = Ohio*. The selected feature(s) can then be saved as a new feature class.
- **Clip Tool**: This tool allows you to "cut" the features in one layer based upon the features in a second layer. (For example, if you have a layer of rivers that extends across the state border between Ohio and Pennsylvania, and you clip the rivers layer using the Ohio boundary layer, then the output would be a feature class that contains only the river segments present within the state of Ohio (think of it like using a cookie cutter to cut out the shape that you want to keep). This output can then be saved as a new feature class.

Since we are creating a whole new set of data files for just the Youngstown area, we want to clip down the data to the extent of Youngstown.

NOTE: Don't forget when clipping data, it is a good opportunity to convert all of the data to an appropriate projected coordinate system for the location of interest.

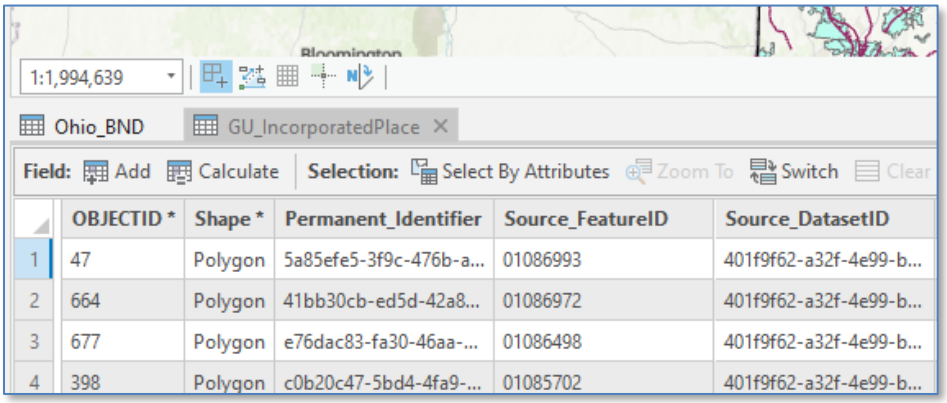
Question 3: What would be an appropriate coordinate system for data related to Youngstown, OH to be projected in?

6. First, set the map to an appropriate coordinate system for Youngstown.
    a. In this case use *State Plane, Ohio North (NAD 1983 – US Feet)*
7. Next, look at the *GU_IncorporatedPlace* layer on the map.

a. You can see that it is comprised of a bunch of polygons and if you click on a polygon with will contain the name of a city, village, etc.
b. This layer contains the boundaries for all incorporated political entities within the state of Ohio. Youngstown is one of them.

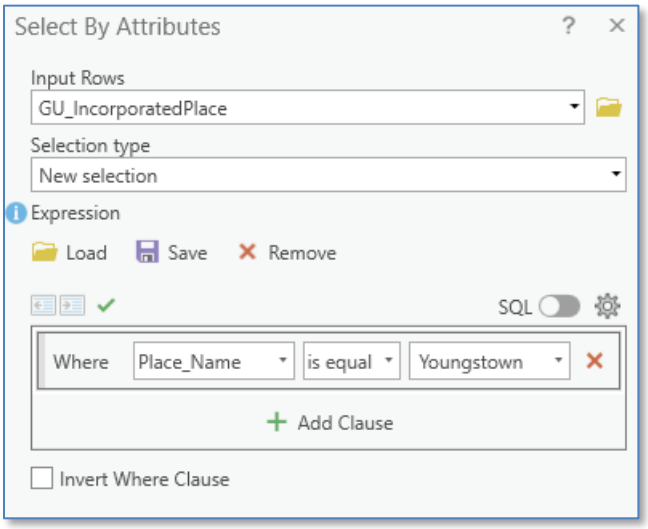
8. Next, we need to create a layer that contains the boundary for the city of Youngstown. This we can do with a query of attribute table to find the city boundary.

a. Open the attribute table for the *GU_IncorporatedPlace* layer
b. Look at all of the field headings and also the field entries to find a field that may be useful for identifying the city of Youngstown. (*Place_Name* or *GNIS_Name*).
c. Click on the Select by Attributes button at the top of the attribute table.



	OBJECTID *	Shape *	Permanent_Identifier	Source_FeatureID	Source_DatasetID
1	47	Polygon	5a85efe5-3f9c-476b-a...	01086993	401f9f62-a32f-4e99-b...
2	664	Polygon	41bb30cb-ed5d-42a8...	01086972	401f9f62-a32f-4e99-b...
3	677	Polygon	e76dac83-fa30-46aa-...	01086498	401f9f62-a32f-4e99-b...
4	398	Polygon	c0b20c47-5bd4-4fa9-...	01085702	401f9f62-a32f-4e99-b...

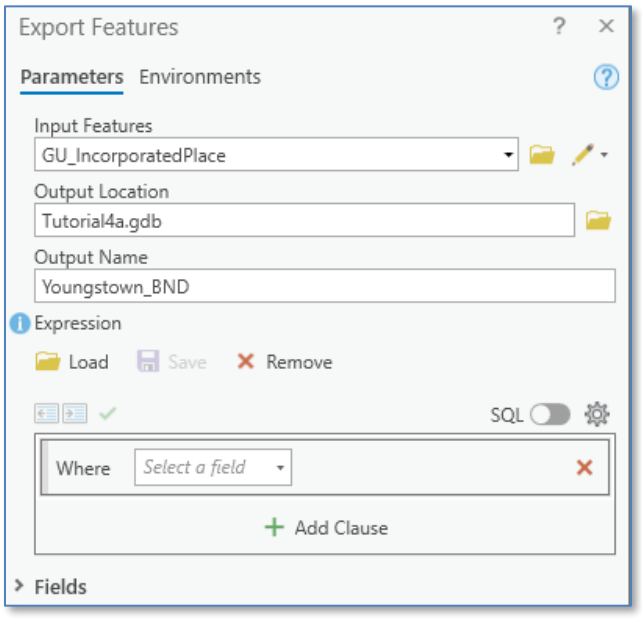
d. The Select by Attribute tool window opens.



e. Enter the expression where "*Place_Name*" "*is equal to*" "*Youngstown*". Click Apply.
f. The boundary polygon for Youngstown should be selected now (At the bottom of the attribute table it should state "*1 of 926 selected*").
g. To verify it right-click on *GU_IncorporatedPlace* in the Contents pane and select *Selection > Zoom to Selection.*
h. Now, we need to create a new feature class that just contains the Youngstown boundary.
    i. Right-click on *GU_IncorporatedPlace* in the Contents pane and select *Data > Export Features*. The Export Feature tool window opens.

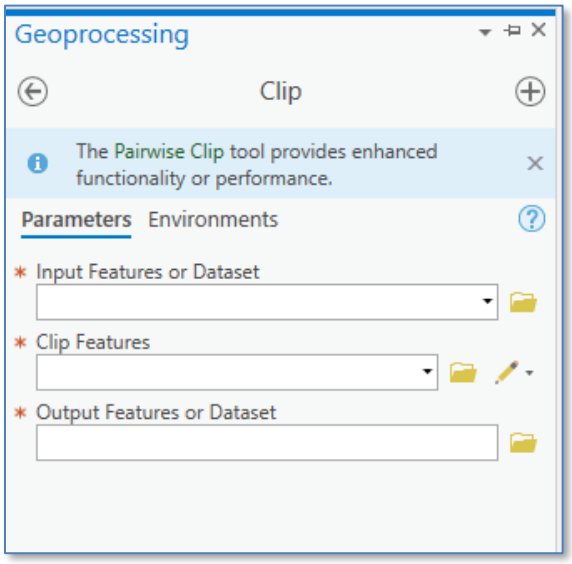
NOTE: When any features are selected, the export function will only save the selected features, not the entire feature class.

ii. Make sure the output location is the geodatabase for the lab.
iii. Give your output file a name (ie. Youngstown_BND).
iv. Click OK.
v. The new feature class should be automatically added to the map. (Turn off the *GU_IncorporatedPlace* layer).

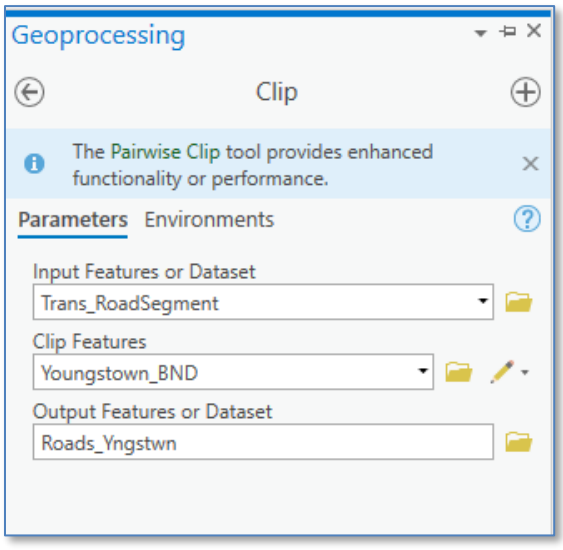


9. Next, use the **Clip** tool to crop the transportation, structures, and water features to the extent of Youngstown (ie. so the layers only include those features located within the city limits of Youngstown).
   a. Open the Geoprocessing pane.

b. In the search bar type Clip. Select Clip (Analysis Tools). This will open the Clip tool window.



c. Enter the following variables into the respective fields:
   i. **Input Features or Dataset** = *Trans_RoadSegment* (use dropdown box to ensure everything typed correctly).
   ii. **Clip Features** = *Youngstown_BND* (or whatever you labeled your Youngstown boundary layer)
   iii. **Output Features or Dataset** = *Roads_Yngstwn* (or any name you will recognize).
   iv. Click Run.

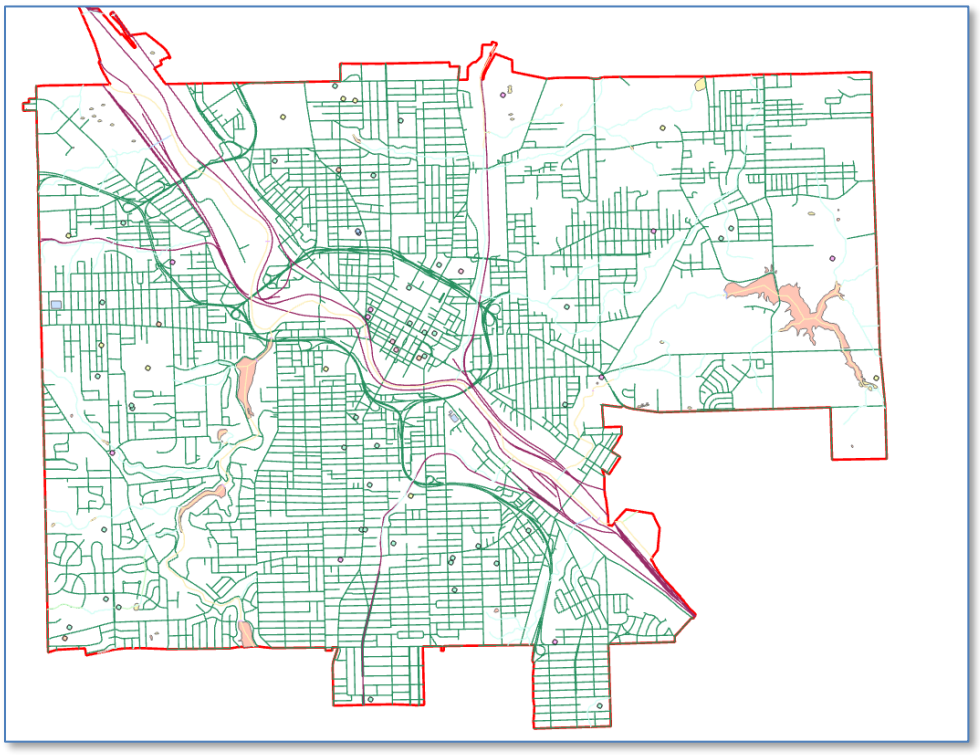


Note: It may take a minute or two to process all of the road features depending on how big the input file is. The output file should be automatically added to the map.

d. Repeat the process for all of the other layers. Use the Youngstown boundary for the clip feature.

i. *Trans_RailFeature*.
ii. *Struct_Point*
iii. *NHDFlowline*
iv. *NHDLine*
v. *NHDWaterbody*

e. Turn off all of the original files so only the clipped lavers are displayed on the map. All of the features should stop at the boundary of Youngstown.



Now you can explore your data a bit more easily.

An easy way to find features, is to use the Locate tool (You can also use Select by Attributes, but if you are not sure what layer a feature may be found in the Locate tool is easier).

NOTE: When using the Locate tool, the features are case sensitive, so if you cannot find a feature, check the spelling and capitalization.

10. Open the **Locate** tool

- Select **Layer Search**
- In the **Search** bar, type what you are looking for
- Click on a result and the location will be highlighted on the map.
- Rick-click on the result and select "*Zoom to*". This can help you more easily find the feature and explore the features around it.

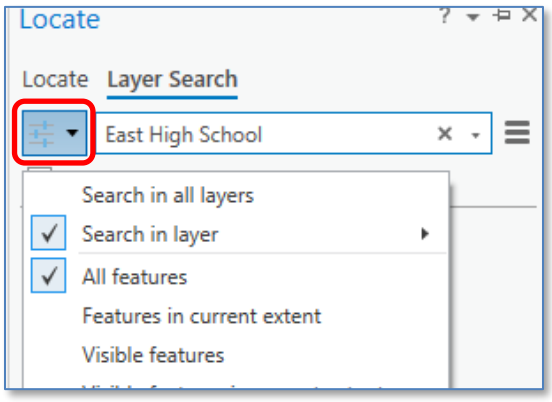
Question 4: Use the Locate tool to find and identify the following features:

a) *Youngstown Fire Department Station 6* is at the corner of which two Youngstown roads?
b) What water feature flows between *Lake Cohasset* and *Newport Lake*?
c) *Vittorio Ave* is adjacent to what **type** of water body?
d) *Coitsville Ditch* connects to what body of water?

Now, let's check out some not-so rare issues that you may encounter with GIS data.

One thing to keep in mind with any data set that you acquire is the accuracy, reliability, and usability of the dataset. This is partly what the metadata is for. It allows you to assess the data by identifying who created it, how old it is, etc. Nevertheless, data that seems very accurate may not always be. The world is constantly changing, and it often takes time for GIS data to catch up with reality. (This is an issue we encountered a couple years ago in class).

11. Turn off the structure layer so it does not display.
12. Go to **Canvas** and download the file labeled *Struct_Ohio2017* from the Tutorial 4a section.
    a. Unzip the file
    b. Add the *Struct_Point2017* layer to your map. Rename the layer if necessary to distinguish it from the other structure layer.
13. Using the **Locate** tool and the new structure layer, search for *East High School.*
    a. To limit the search to a single layer, click on the *Options* button left of the search bar.



    b. Select Search in *Layer* > *Struc_Point2017* > *All Fields*

Question 5: When you search for *East High School* in the new structure layer, what did you get as a result? Why? (**HINT:** Check the attribute table for the layer. Look in the *FType* field.) Look at the Table of Contents. Should East High School be listed? (East High School is an actual school in Youngstown).

This is the Structure file we downloaded from the National Map a couple of years ago. Since the data we were interested in (schools) was missing for this data set, we had to find the data elsewhere.

(The structure data set has since been corrected. If you search for *East High School* in the structure dataset you downloaded from the National Map you will find the school).

In this case we needed to find a data layer that contains schools for the Youngstown area. We can get this from the Mahoning County GIS website.

14. Navigate to the Mahoning County GIS Portal
- Search for *Mahoning County GIS*
- Select the first hit (**GIS/Tax Map | Mahoning County**)
- Select *Data Downloads* on the left-hand side of the screen.

- Click the "Proceed to Site" button
- Select the directory labeled *Shape_Files* from the list
- Scroll down until you find the link for *Schools (zip)* and download the file.
- Uncompress the file
- Add the shapefile into ArcGIS Pro

For the above case it seemed that the data on the National Map had been edited, and many features were removed from it. In fact, the description of the structures data (the metadata) from the National Map at the time stated that the structures file just contained information relevant to emergency response. And if you look at the attribute table you will find Emergency Response, Hospital and some landmarks listed, and that is it. They changed the data but did not change the symbology. The Table of Contents lists many different data structure categories. The Structures datafile has since been updated and corrected to include all of the features listed in the Table of Contents.

Remove the 2017 structure layer from your map.

Now let's explore another aspect of data reliability. We will explore the railroads of the Youngstown region. There are a lot of them. Be sure that the railroads layer is displayed on the map.

Question 6: What is the name (owner) of the major rail lines coming into Youngstown (city limits) from the North? The South? The Northwest? The Southeast? (**HINT:** look for the *RailOwner* field)

In this case it seems that several fields in the attribute table have not been filled in (<*null*>). The topology is present, but the attribute information is lacking. So, it looks like we need to go elsewhere to find the data to answer our questions. Let's try another rail dataset. This time from ArcGIS online.

15. Open the *Portal > ArcGIS Online* tab in the Catalog pane.
   a. Search for *Ohio Rail*
   b. Sort the results by *Relevance.*
   c. Add the *rail_line* dataset (created by *odot88*)
   d. Change the symbology for both the *rail_line* layer and the *Trans_RailFeature* layer so that you can tell them apart from each other.

Question 7:

a) How do the data layers compare? Do they overlap? Is information contained in one but not the other?

b) How do the rail lines in downtown Youngstown compare? (**HINT:** Find Federal Plz on the road layer). Which dataset contains this rail line? Why do you think the rail line is only present in one of those layers? (**HINT:** you may want to add the *Imagery Hybrid* basemap to help understand the situation)

Let's compare the status of lines for both layers.

16. Change the symbology for the *rail_line* layer
   a. Select *Primary Symbology > Unique Values*
   b. Use *RR_Status* as the field to display
   c. Adjust the symbology so that they are easy to distinguish

Question 8: What are the different status types (*RR_status*) for Railroads listed in the *rail_line* layer?

Question 9: Which railroads (name of rail line/ company) have *Active* lines coming into Youngstown from the North, South, Northwest & Southeast?

Question 10: Which railroads (name of rail line/ company) have *Abandoned* lines coming toward Youngstown from the southeast? Northwest?

Question 11: Which rail dataset is more accurate? How can you tell? (**HINT:** Add the *Imagery Hybrid* basemap)

From comparing these two maps you can see that even though one is more accurate in terms of current rail lines; it is also more incomplete in terms of updated attribute fields. One came from the National Map (a reliable source) and the other came from ODOT (a reliable source). So even though the source is reliable, there can still be problems with the data.

Now let's verify some of the data in the *Trans_RoadSegment* layer for Youngstown. Add the *OpenStreetMap* basemap. Find Youngstown State University (just north of downtown Youngstown [you can also search the schools layer to find the university]). Compare the *Trans_RoadSegment* to the *OpenStreetMap* basemap for YSU.

Question 12: What roads in the Youngstown State University campus area (present in the National Map *Trans_RoadSegment* layer) don't match up with the *OpenStreetMap* basemap (list street names)? How do they not match? (**HINT:** Look in the area near the green oval around a rectangle – northwest corner of campus).

Question 13: What is occurring in the locations identified in Question 12? (**HINT:** If you cannot identify the features, try adding the *Imagery Hybrid* basemap).

Here again is reliable data from the National Map, but it is not reflecting recent changes on the ground. This often occurs with GIS data sets, due to the time between when changes on the ground occur and when the datasets are updated.

Question 14:

a) What is the currency of the *Trans_RoadSegment* layer (ie. What is the time period of the data?)

b) When was the sports complex created? (**HINT:** based upon the *Trans_RoadSegment* layer what can you say about the time frame when those features were created