Image Processing Lab Exercise #7:

Dr Bagher Bayat, RS/GIS Group, Faculty of Earth Sciences, SBU

Classification II: Training Samples and Getting Ready for Supervised Classification

Objectives:

- **Determine the ideal classes for a supervised classification (of Gainesville case)**
- **Collect data in the field to be used as training samples**
- **Create reference data from Google Earth to be used as training samples**
- **Convert Google Earth Pro KML/KMZ files to shapefiles**
- **Read XY data from CSV files and make shapefiles**

DATA Files:

Lab7-data-collection-CIPEC-training-sample-form.docx = this is the data form you might use when planning for field training sample (TS) collection in your future studies.

Lab7-CIPEC_TS_Protocols_Explanation.pdf = this file explains all the categories to be completed in the TS file you will use. This file is very detailed!

2020_Training_and_Error_Analysis_Samples.xlsx = spreadsheets of past TSs have been collected for the Gainesville region, for use in Part III of this lab, where you will load and display the training samples by land cover class. You will then use these data files in Lab 8, to undertake a supervised classification analysis of the Gainesville region. Look through this file to see what LC classes we used in the past. There is a column of the detailed class, and the 5-class equivalent in the first tab: All_TS_All_Time (More classes upto 10 mentioned in the file. However, the five classes Forest, Urban (urban, commercial and residential), water, grassland/pasture,and wetlands are important classes).

Satellite image *gnv-subset-etm-11-feb-2003-17-39* for use in Part III of the lab

Software programs needed:

Google Earth Pro
Arc GIS Pro

Part I: Training Samples (TSs) from CIPEC forms – Where you have a project!

This part of the lab will not be completed today. Its purpose is simply to introduce you to the procedures you will need when collecting your own field samples in the future, for example for your research project or any other case study. Therefore, this will show you how you can collect your own training samples for anywhere in the world. To do this you will complete the CIPEC Training Sample forms for any DIFFERENT Land Cover types. You might want to take points of a nearby park, urban areas, neighborhoods, golf course, lake etc.

How do you select an area for a training sample?
We usually select representative areas of the land cover of interest (one of our land cover classes), and then find an area of at least 90 m by 90 m (so 9 pixels for Landsat, assuming you will use Landsat images), although a larger area is better (easier to find on the image). You must have the following:

1. an ability to record GPS coordinates (an App on the phone, your own GPS etc.)
2. an ability to take photos of the site – again a cell phone, digital camera etc. all work
3. access to a printer to print out desired numbers of the worksheets to complete [Lab7-data-collection-CIPEC-training-sample-form.docx] or the ability to fill them out on a device. It is easiest to fill them out on paper and not in your device.
4. You may need access to a scanner to upload the completed forms – if you do not have this, simply use one of the many scanning apps for your phone, compile all the images into 1 document and save them as training sample files.

SAFETY: Keep safety as the #1 concern when undertaking Training Samples. The purpose of Part I is for you to learn how to go about completing training samples in research or for data collection. However, do not put yourself in harm's way to do this – take the TSs in public areas, with lots of people around, in safe areas, etc. Be careful crossing streets. Do not put your safety at risk to get a TS point.

To help guide you when completing the training sample forms, use the document which describes (in way too much detail!) the meaning of every category on the form (*Lab7-CIPEC_TS_Protocols_Explanation.pdf)*. Many are obvious and are simply confirmed with this form. Use this form to ensure you complete all the necessary sections of the form. Note that the form can be modified to suit your own research question or project objectives.

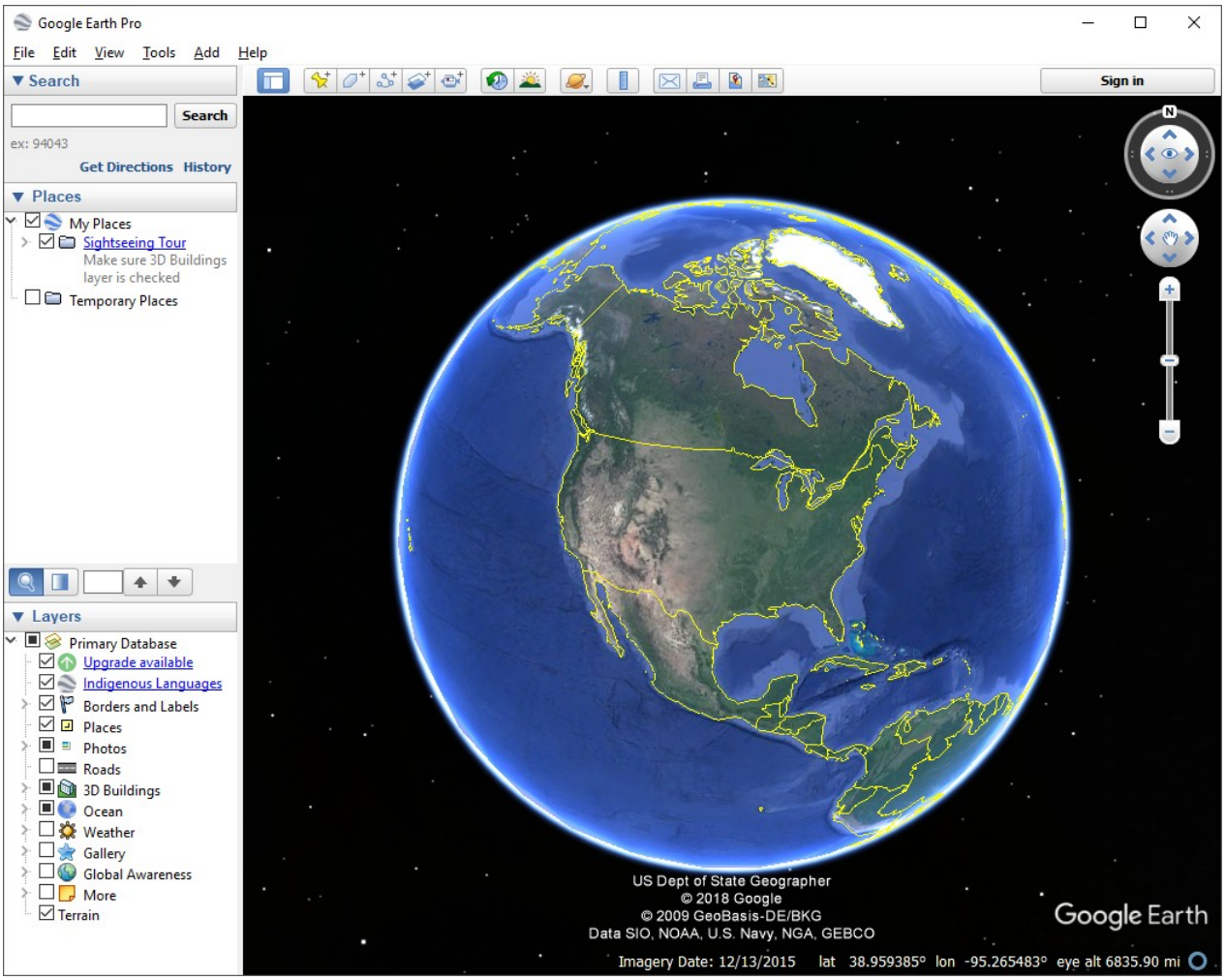
Part II: Google Earth Training Samples for Gainesville, FL

Another way to collect TSs, and one used more and more increasingly today, is to use Google Earth (GE) to collect them. Part II of this lab will deal with the collection of TSs via Google Earth. We will walk you through how to collect TS's in Google Earth.

With almost all remote sensing data analysis there is a need for field data to validate classifications or train the software to carry out the analysis, whether it is a classification or any other kind of study. Traditionally this has led to extensive field seasons in which hundreds of detailed training data were collected including tree diameter at breast height (DBH), % canopy cover, and the most difficult factor, land-use history, among many other things. The field work may be great fun for the participants, but not only are such field seasons expensive but also the observations are a snapshot in time. The lack of historic land-use and land-cover data has led many remote sensing scientists to turn to time series imagery from Google Earth to develop training samples that are either (1) stable land cover over time, or (2) document landscape change. At first there was push-back in the remote sensing world for using such training data, but more and more journals, government organizations, and NGO's are accepting such "field" data.
Below is a tutorial on how training samples can be collected in Google Earth. This tutorial will introduce the basics on how to use GE to view time series imagery. Remember the training data collected are always dependent on the research question for the study. You are familiar with the CIPEC training sample sheet by now. Note that while it is recommended to use this data collection sheet, sometimes not ALL of these data are required or this sheet may be missing something vital

to your study. So, it is supplied to you in .docx format so that you can modify it to suit your project or research questions later.

1. This is the basic Google Earth Layout:



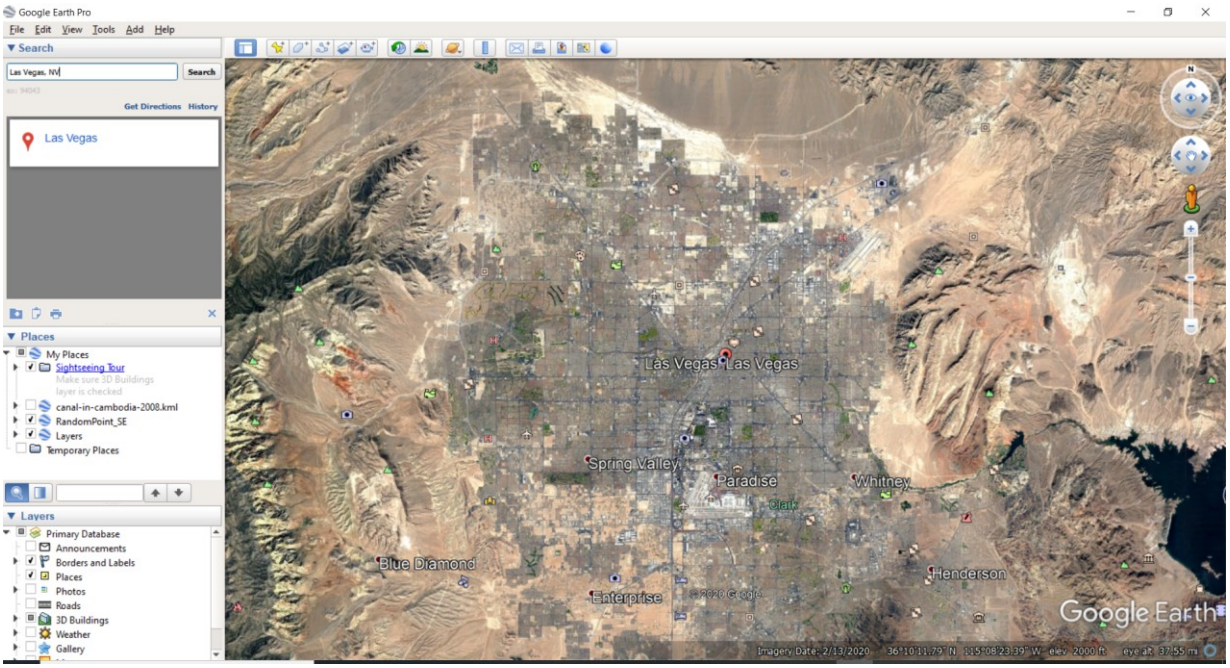
If you are not familiar with GE there are a couple of different panels to understand. On the right is the zoom in and out slider and tools to change the orientation of the earth. Across the top is the file, edit, view, tools, add, and help buttons. File has a **Save** function that is commonly used to save training sample points. Under the **Tools** menu there is also an **Options** function which is commonly used to switch between coordinate descriptions (Lat-Long in decimal degrees, DMS, UTM, etc.) and measurement units (metric or English). Below the menus are a row of buttons, we will get familiar with those in the next step. In the left hand portion of the window is the search (try entering a city name!) and places panel. When taking samples in GE you will use the places panel a lot.

Lastly, you can grab and scroll to any place on Earth. Additionally, you can double click anyplace on Earth for an automatic zoom to that location.

2. Now we will get familiar with the Button Bar across the top of GE.

There is enormous functionality in these buttons. The first on the left simply hides the panels on the left and make the imagery on the earth larger. The second button is very important. It is how we mark places on earth that are good training samples. The other critical button on this bar is the clock. This is what we will use to scan through the time series of imagery.

3. In the search bar enter Las Vegas, NV (not NM) and click on the Search button.

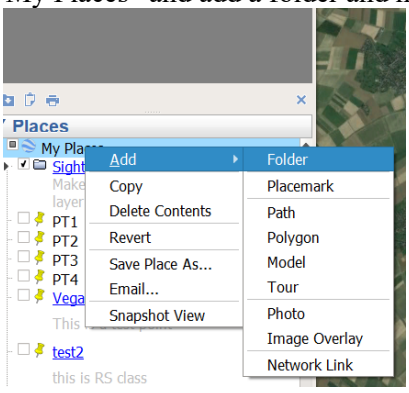


The image that comes up is a clean crisp high resolution image from 2024 from Digital Globe. Google earth has a huge repository of historical imagery which can give you a short (30 or so years but in some, especially urban areas, cases much longer) historical view of the area.

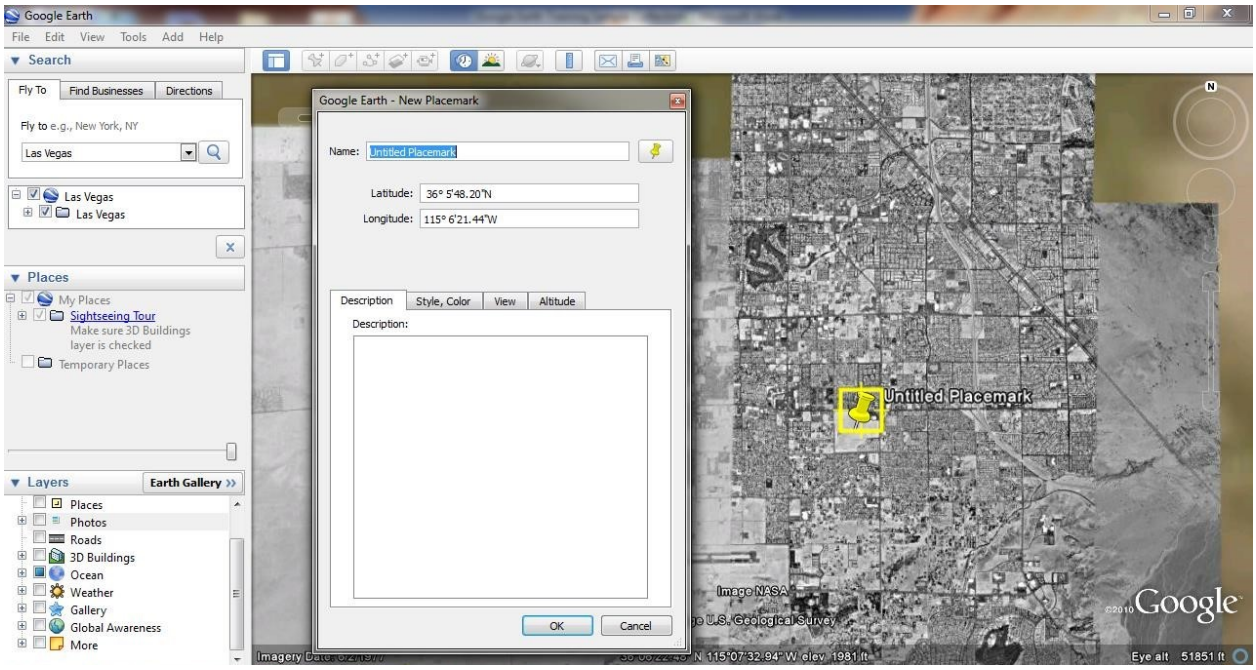
4. Zoom to the Vegas airport (south of the Las Vegas: see below).

5. Hit the Clock button  and move the slider back and forth with either the slider button itself or the two arrows on the right and left of the slider if you want to move one year at a time. Note that the data of the imagery is shown in the slider to the month, and at the bottom of the image to the day. You will see that there is much historical imagery for the Vegas area, stretching as far back as 1965.

6. Grab the time series bar and scroll to the first image. You will notice that in this image there is little of what we know Vegas to be today. There is an airport and a couple of small roads. It was a town built in the middle of the desert and built very quickly. Think of how this information would be useful in a classification. Not only can we get an idea of what the original landscape may have been like but we now know that our large features, the airport and runways, have been long standing.
7. Scroll through the time series of Las Vegas imagery and see the birth and growth of the city. Think of how we could collect numerous samples, by dropping pins, documenting when land changed and to what.
8. Flip back and forth between the 1977 and 1990 images. In the interim years Vegas underwent huge change.
9. Now let's collect a sample.
10. First right-click on the "My Places" and add a folder and name it as you wish:



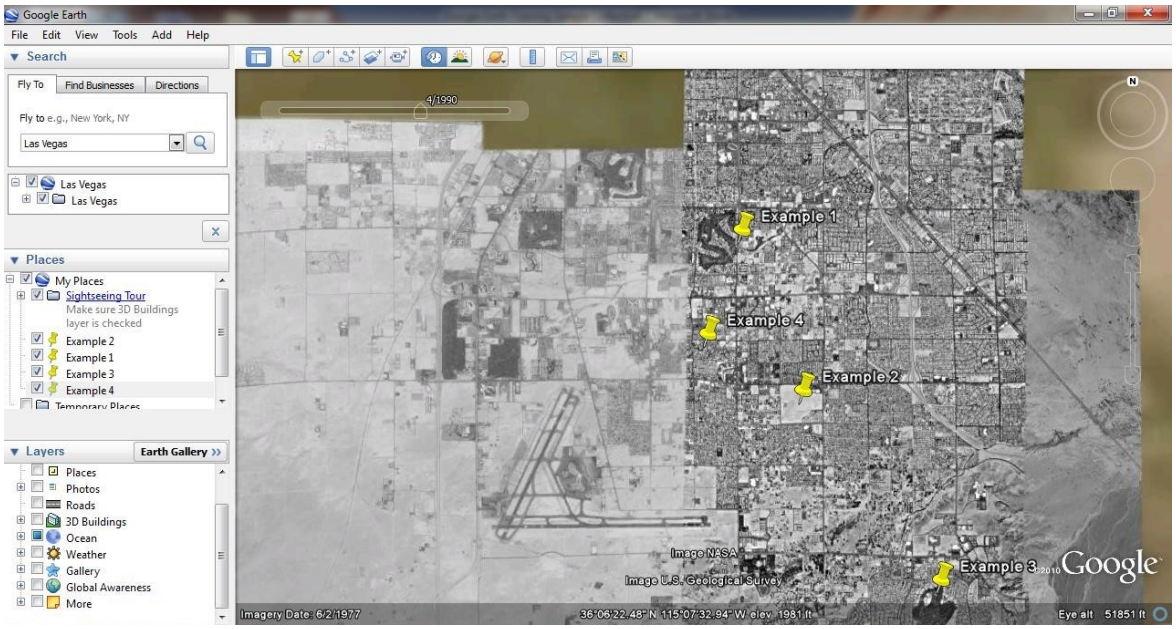
Then click on the yellow pin  in the Button Bar across the top of GE:



When you drop a pin, using the pin button  , a new placemark window appears in addition to the pin. You can move the pin to the exact location you want for the training sample, the latitude and longitude (or UTM coordinates if you have set this coordinate system in the Tools:Options window) of the new placemark GUI will automatically update. Once in the correct location you can name the training sample (Use your best estimate of the land cover or land use type and add a number, e.g. Desert 01, or blacktop runway 01 – remember the 30 classes in the unsupervised classification; your first pass through a supervised classification training sample collection will resemble those classes) and add a description of the sample. Things that could be included in this description: dominant vegetation or land cover type, land use if you are confident of inferring it, approximate size of this cover type (you can use the measurement button to aid in this), and even number of trees or shrubs (with the high resolution imagery its simple, just count!).

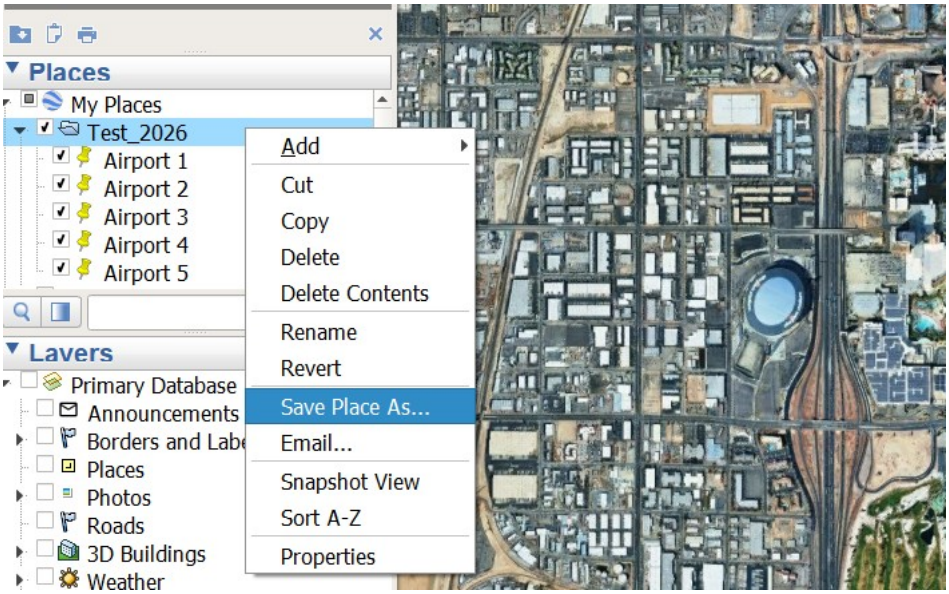
Make a mistake in the point location or the text in the description? Simply right click the yellow pin and hit properties. This will reopen the new Placemarker GUI.

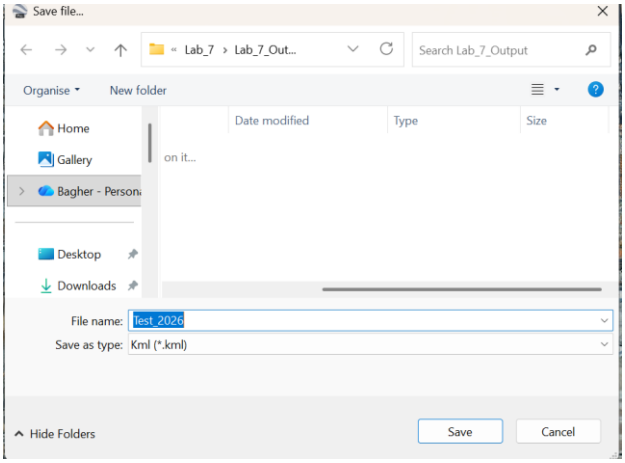
Once you have entered all of the information just click OK and the point is complete. You will notice that the new point appears in the places tab.



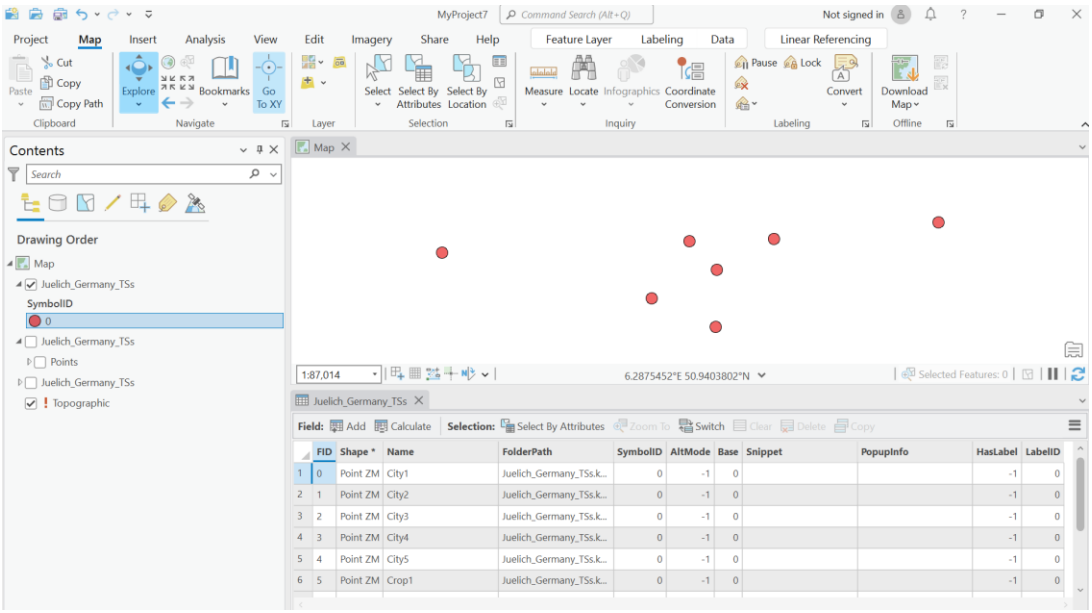
11. Once complete you can export your samples to a kml or kmz. These are file types native to Google Earth, but we can convert them for use in ENVI and ArcGIS.

12. In the Places tab right click on the folder you created under my places and select Save Place As... You can then name your training sample file and save it to your working directory either as a .kml or .kmz file.

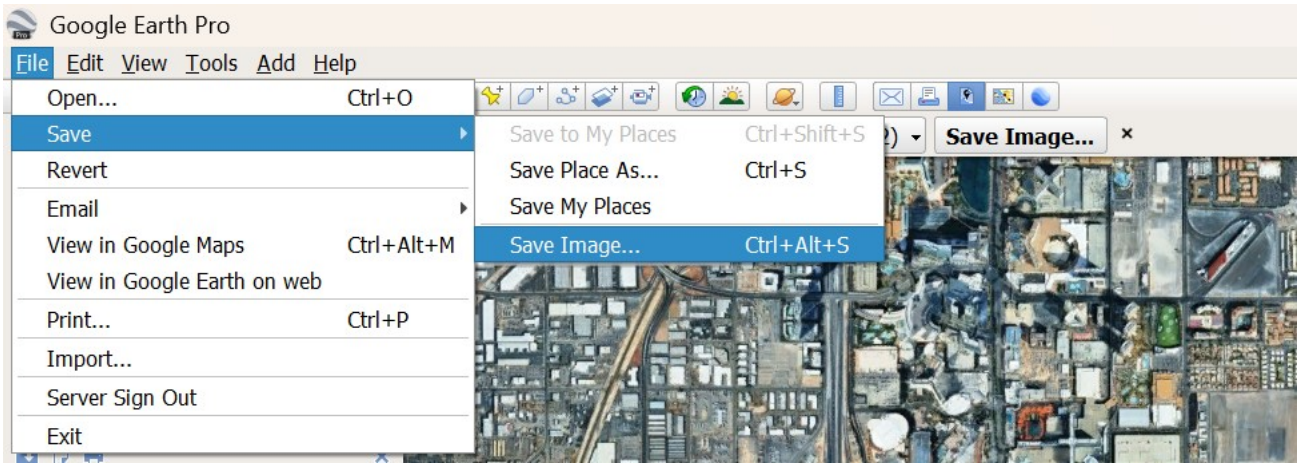




The saved KML/KMZ files containing your TSs can be added into ArcGIS (pro) and then converted to a shapefile using the 'KML to Layer' tool. Once the layer is created, right-click on the resulting feature (point) layer and select 'Data>Export Features' to convert it to a shapefile.



In addition, you can save an image file (jpg or png) with ALL your TSs entered (this is under File>save>save image) and export your TSs as a picture (that could be used as a map showing the locations of your TSs).



Part III: Using available TS's for Gainesville area, to start the process of Supervised Classification Preparation (Supervised Classification using this data will be undertaken in Lab 8).

The image we will use in this lab, *gnv-subset-etm-11-feb-2003-17-39,* was seen in a previous lab. It can be found in the ***Lab 7 input data folder***. The image is a subset of a larger Landsat 7 ETM scene. See the figure below for more general information about the scene.

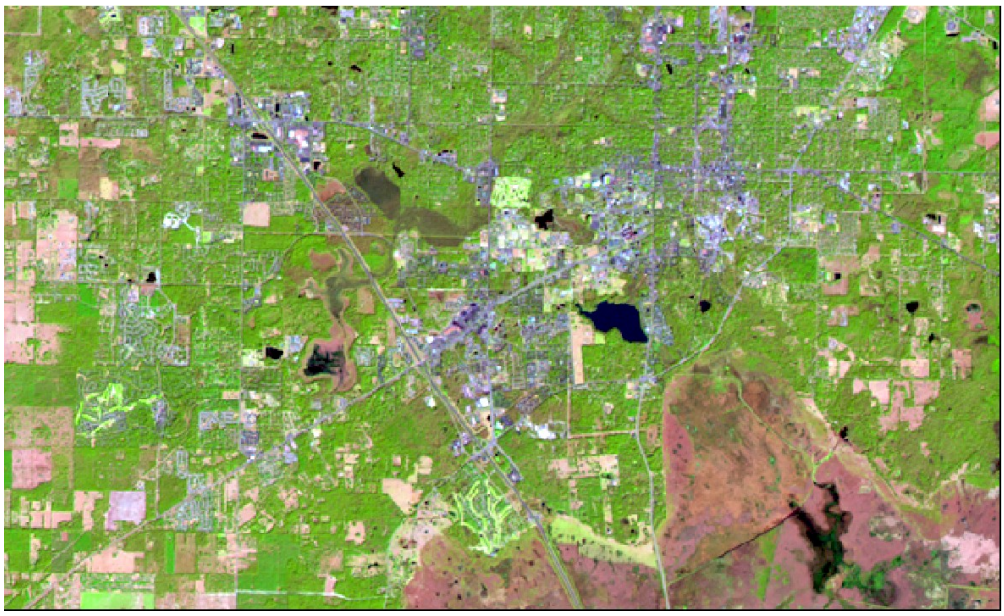


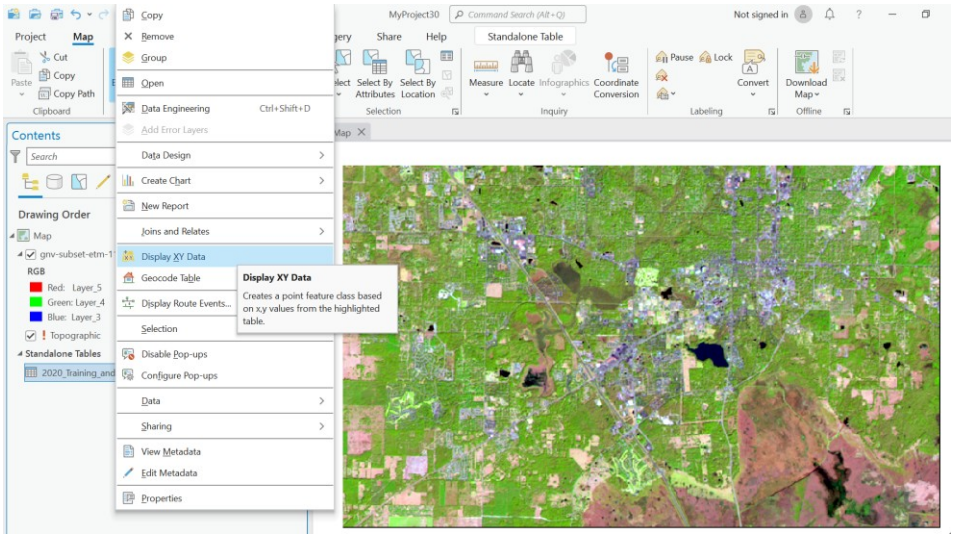
Figure 1. General information for Gainesville image with false color composite (RGB=543)

To run the supervised classification we need our training sample data to be in ENVI vector format, but that will wait until next supervised classification lab. For now we will use ArcGIS to make a shapefile of our training samples, then a map of the data. In order to do this the data must be in .csv

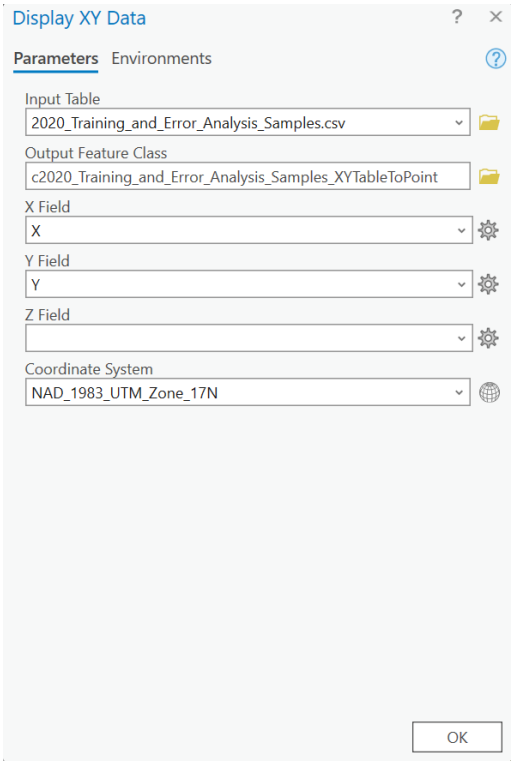
format. We will use the *2020_Training_and_Error_Analysis_Samples.csv* file in the ***Lab_7*** data folder to complete this section of the lab.

Displaying XY Data

1. Open ArcGIS and add the Gainesville image (*gnv-subset-etm-11-feb-2003-17-39.img)* to the screen (change the band color composite to 5-4-3).
2. Add "2020_Training_and_Error_Analysis_Samples.csv" file.
3. You will notice that a table of your training samples opens up in the table of contents of ArcMap.
4. Right-click the file name in the table of contents and select the **Display XY Data** (see image below).



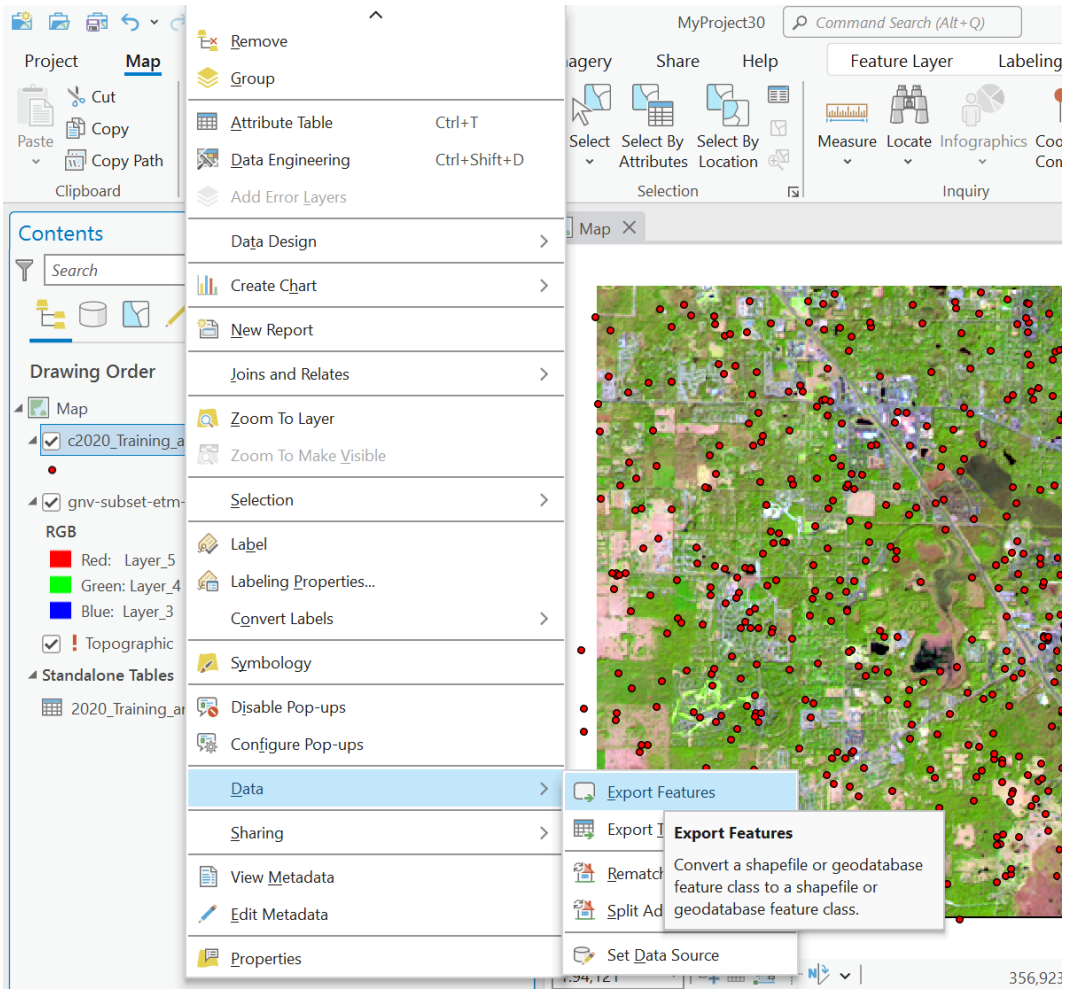
5. In the Display XY Data window, select X for the X field and the Y for the Y field (if not selected automatically) and for the coordinate system option choose "gnv-subset-etm-11-feb-2003-17-39.img" which eventually will show "NAD_1983_UTM_Zone_17N":



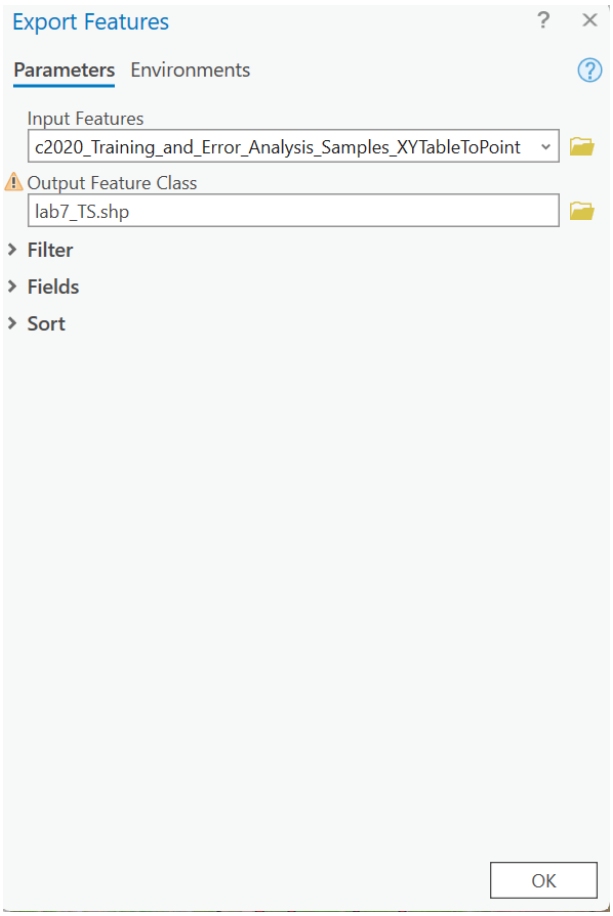
6. Then click **OK** again to display the XY data. A warning message window "⚠ WARNING 040401: NULL geometry ignored. ObjectID = 1379, 1396" will pop up showing the table does not have an object-id! This is no problem, just close the window.

You will notice that the points now display on the screen, over your image of Gainesville. However, technically it is still not a shapefile (and it is called the Events). The next step will rectify this.

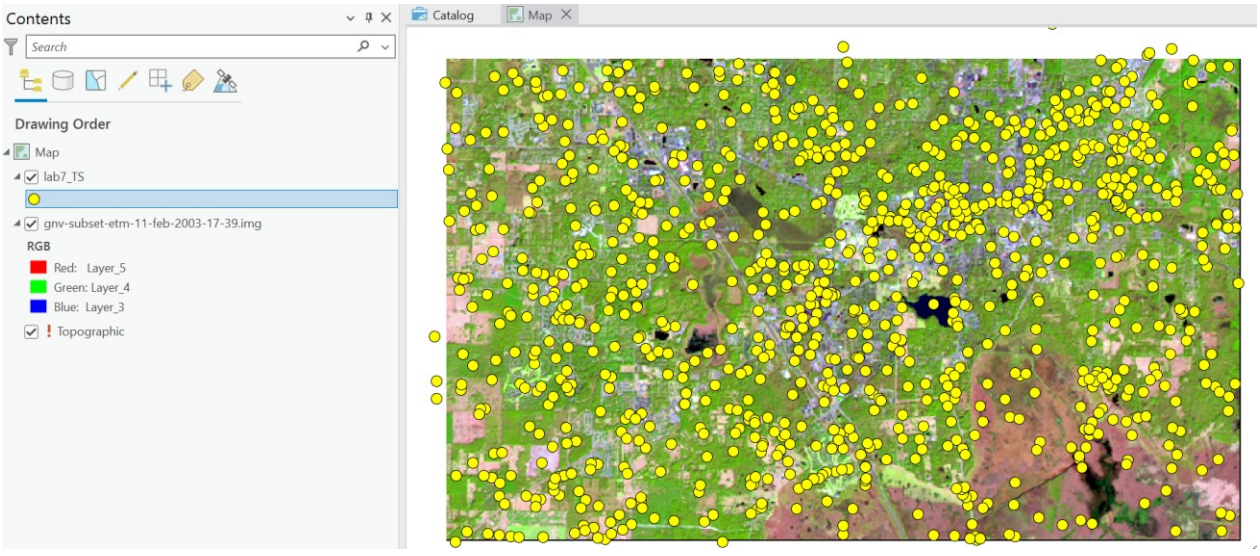
7. Right click the point event, what you just made, select Data > Export Data. The Export data dialog will open (see image below)



8. Use this export data dialog to save the shapefile to your lab 7 folder. BE SURE TO SAVE IT TO THE CORRECT FOLDER! Click the folder on the export data dialog and navigate to your lab 7 output folder. Name the file lab7_TS. Make sure you select Shapefile in the select as file type portion of the screen. Click **Save**. Then click **OK**. Lastly, after the data are converted you will see a message asking if you want to display the data select **Yes.**



9. Right click the csv and the event and remove them. You may want to change the symbols that indicate each point to something that is yellow and 8-points in size. The window should look like this if you do:



Ok, you are all set! Congrats!