Student Name: ______________________________ Student Number: _____________________________

Laboratory #5: Remote Sensing and Raster GIS

Purpose: The purpose of this lab is to provide you with an introduction to a basic remote sensing classification workflow and to assess change detection in forested regions using raster GIS.

Learning Outcomes: By the end of this lab, you will be able to:

- Undertake spectral enhancement using band combinations and ratios.
- Train an image classifier.
- Explain a signature file.
- Undertake a supervised classification using high and medium resolution remotely sensed imagery.
- Clean and generalize a thematic raster layer.
- Assess a classified result for accuracy.

Procedure:

1. Read and work through the examples.
2. Answer the questions at the end of the lab.

Data/Materials:

- All data are provided in lab5.zip.

Assignment:

- Answer the questions (#1-8) in the spaces provided.
- The total marks available for this assignment = 17 marks for 15% total course grade.
- All lab assignments must be completed and handed in individually.

Format:

- Students will hand in questions with completed answers.

Tables and figures should be produced to the standards of the Department of Geography, Environment and Geomatics, University of Ottawa.
For guidelines on proper formatting and presentation of tables and figures, please see the introductory section of this lab manual.

Student Name: ____________________________   Student Number: ___________________________

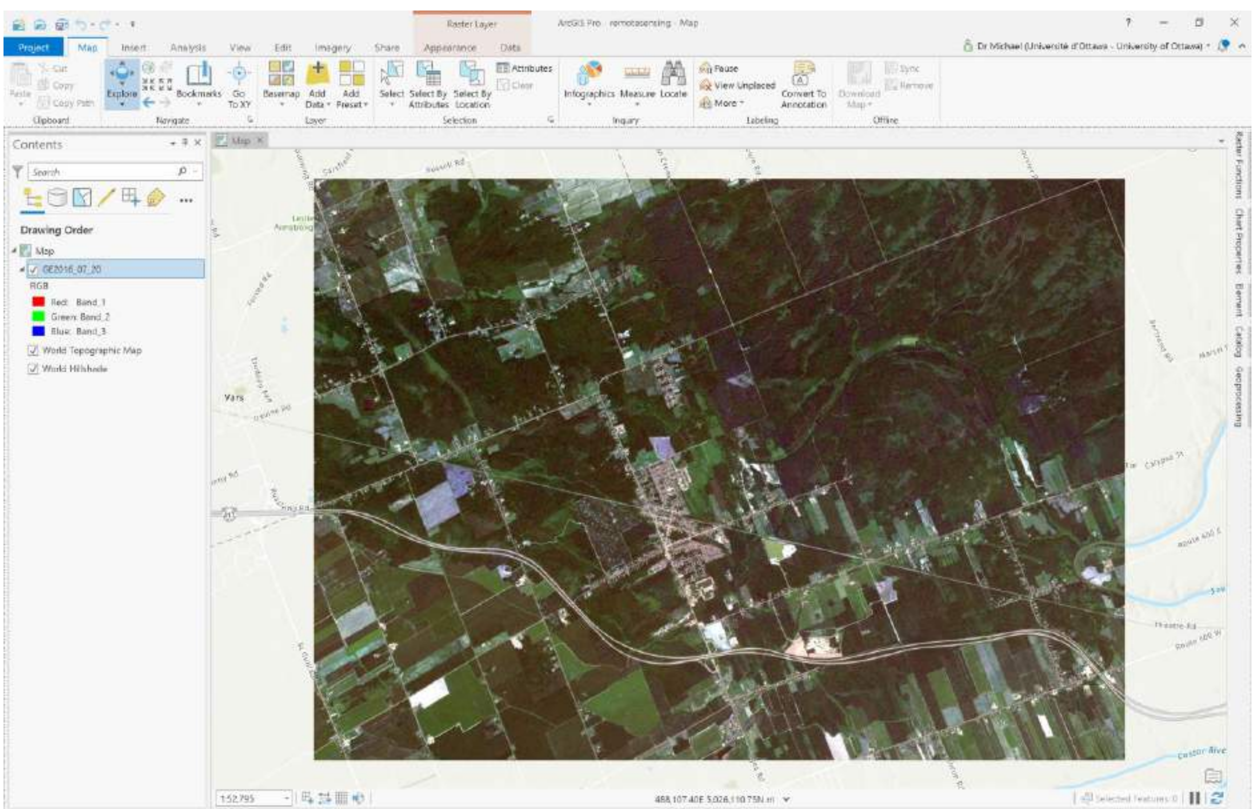
Forest loss is an ever-increasing problem in the world today (see https://earthenginepartners.appspot.com/science-2013-global-forest). In the example questions you will learn how to classify forest and non-forest in two satellite images from 2001 and 2016 respectively, and then determine how much forest has been lost or gained in the meantime in a small region in the south-east of Ottawa. To achieve an estimate of forest loss, in this laboratory, you will be undertaking post-classification change detection using a 2001 Landsat 7 ETM+ image at 30 m resolution, and a RapidEye image at 5 m resolution from 2016. A post-classification change detection allows you to avoid many pre-processing steps necessary for other change detection methods, and it also allows you to directly use the digital numbers within the images without, for example, undertaking absolute radiometric corrections or corrections to ground reflectance for the different images used.

This section presents examples of things you will need to know to address the questions you will be handing in for marks. Your task is to review and repeat the examples in this section, and the derived layers you produce in this section are used directly in the questions at the end of the exercise. Other questions are relevant to help you address questions at the end of the exercise.

Q1: How do can I view different band combinations of a remotely sensed image?

Various band combinations can help you identify different features in remotely sensed images.

1. Add the image called “RE2016_07_20” from the IMAGERY.GDB in the lab5.zip file to ArcGIS.

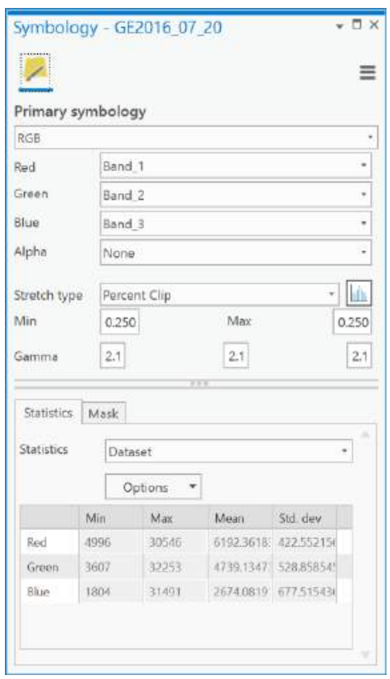


2. This is a section of a RapidEye multispectral image acquired on July 20th, 2016. This satellite constellation is used in agricultural and phenological monitoring and research because of the daily revisit time (off-nadir, 5.5 day at nadir) due to a 5-satellite constellation. This makes the sensor useful for monitoring crop stress or disease outbreaks in large regions. The RapidEye sensor is multispectral, with the following bands:

   440 – 510 nm (Blue)
   520 – 590 nm (Green)
   630 – 685 nm (Red)
   690 – 730 nm (Red Edge)
   760 – 850 nm (Near IR)

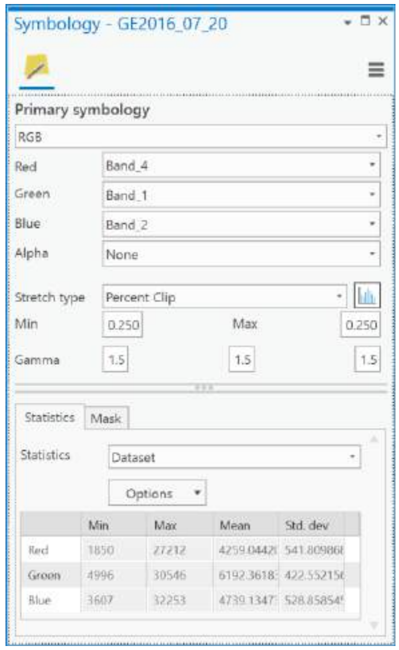
   This image is not atmospherically corrected. In most cases, using this image for assessing crop stress or phenology would require atmospheric correction, as you can learn about in another course, but we will not do that here. Furthermore, the image you are using here is orthorectified, but only using a constant elevation, which is not super-accurate.

3. To view other band combinations of the image, right-click the “RE2016_07_20” layer in the contents pane and then select Symbology:

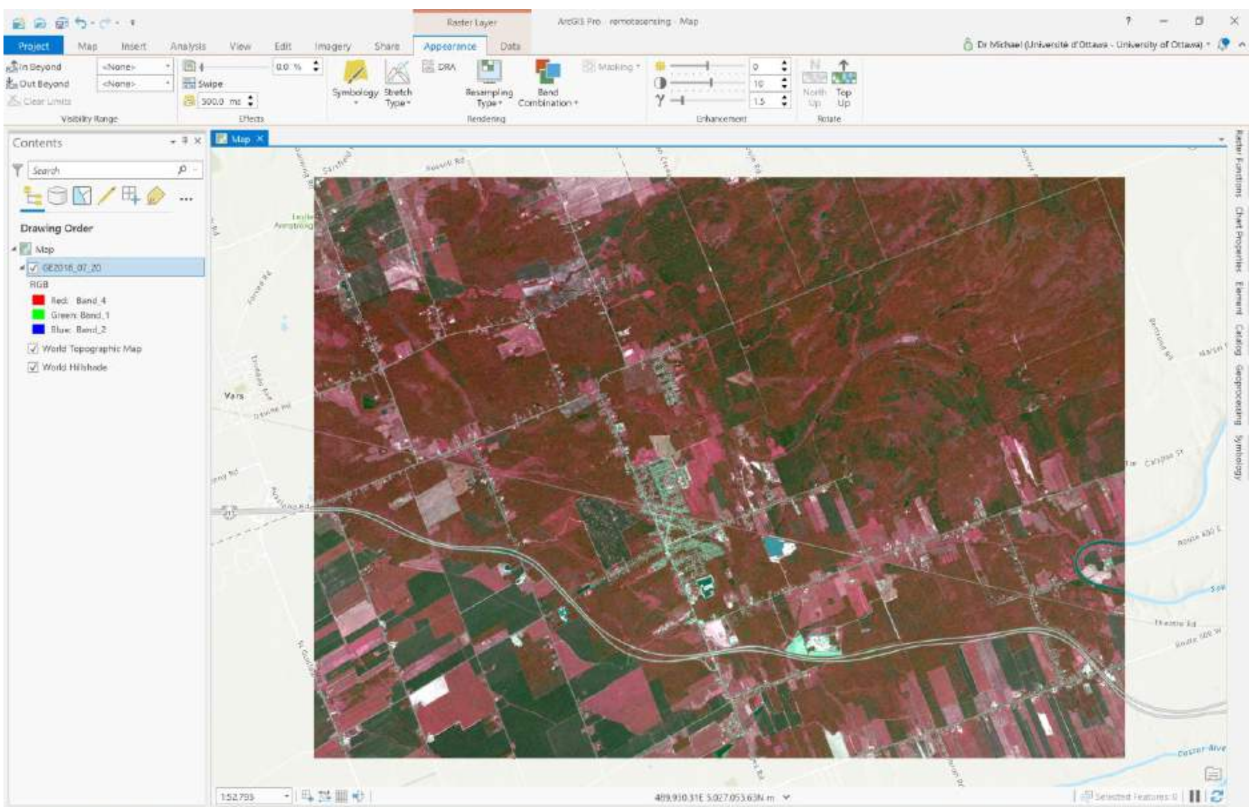


4. By default, ArcGIS will display the visible color band combination 3, 2, 1 for Red, Green and Blue. You can change this to False Color Infrared, for example, by choosing 4,1,2 (Band_4 as the Red band, Band_3 as the Green band, and Band_2 as the Blue band) in the Band column of the Symbology tab:

Student Name: ____________________________ Student Number: ___________________________



5. The resulting map is a “False color composite” and will show red where vegetation is present.



6. You can try other band combinations to create different false color composites, some of which may visually highlight different elements of the imaged landscape.

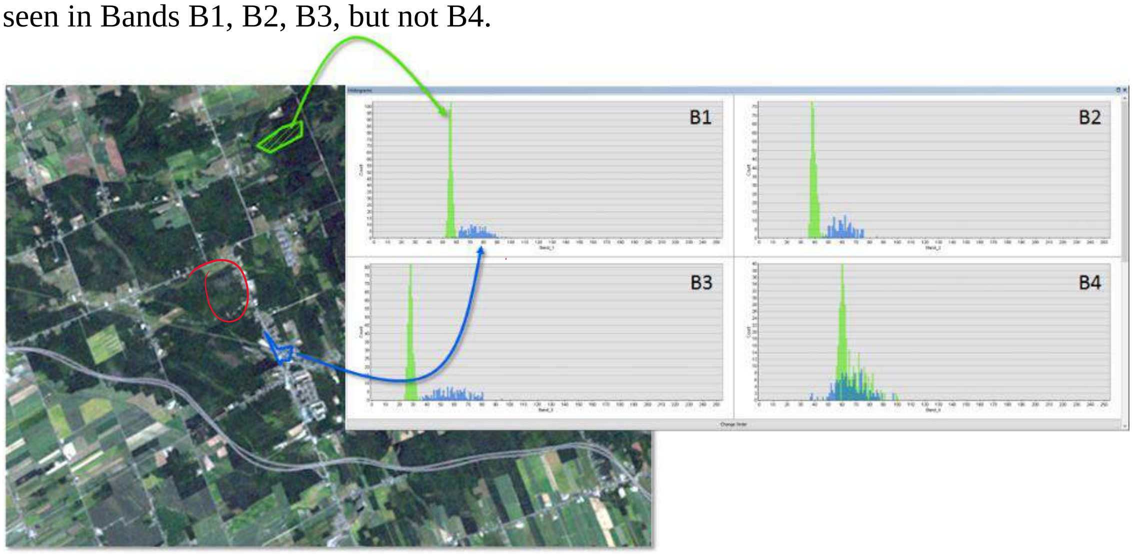
Student Name: ______________________________ Student Number: ______________________________

Q2: How do I classify a remotely sensed image?

In this lab, we will conduct what's called "supervised classification", in which the image analyst (you) provides the computer with examples of pixels containing different landscape features to be treated as information classes. An information class is a nominal category that you want to extract from a remotely sensed image. For example, if you want to extract forested areas from an image then you could create an information class called **Forest** and one called **NonForest**. You could then find examples of pixels that belong to the user-defined information class called **Forest** and another set of pixels for the class called **NonForest**. You could have many other information classes depending on the type of information you want to extract from an image. For example, information classes could include 'water', 'cropland', 'residential', 'commercial', 'dense forest', 'open forest', etc. Depending on the type of remotely sensed image and its spatial and spectral resolution there are many kinds of information that are possible to extract.

To conduct a supervised classification, you need two things:

1. <u>Training data</u>: The training data consists of pixels within the image that belong to different information classes that you want to extract from the image to create a thematic raster layer. For example, pixels that belong to **Forest** and pixels that belong to **NonForest** are two examples of information classes. Examine the figure below. The green polygon is a training polygon that encloses a set of pixels (across all bands) that the analyst has identified as belonging to the information class called **Forest** in the image. The absolute frequency (the count of pixels) of 'Forest' pixel values with different digital numbers (DN), within the green polygon, are shown as green histograms for each band. Likewise, the blue polygon contains **NonForest** pixels and is a training site containing non-forest pixels across all bands in the image. The absolute frequency of DNs within the **NonForest** information class for each band are shown as blue histograms. For the classification to be accurate, the training pixels for each information class should have separation at least one of the histograms, i.e., the histograms for each training class should have minimal overlap. This is discussed in detail later. In this case, separability is seen in Bands B1, B2, B3, but not B4.



2. <u>A signature file</u>: A signature file has a ".gsg" file extension and contains a statistical summary of the different pixel values within each of the information classes specified in the training dataset:

I7_2001_08_25t.gsg - Notepad

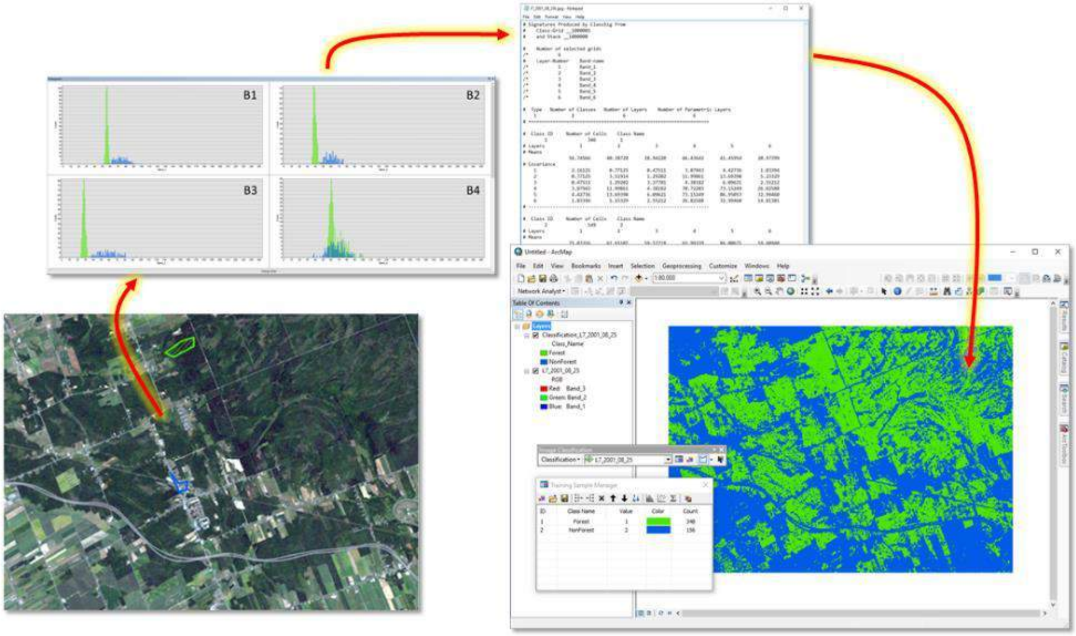
File Edit Format View Help

```
# Signatures Produced by ClassSig from
#    Class-Grid __1000001
#    and Stack __1000000

#    Number of selected grids
/*           6
#    Layer-Number    Band-name
/*           1       Band_1
/*           2       Band_2
/*           3       Band_3
/*           4       Band_4
/*           5       Band_5
/*           6       Band_6

#  Type   Number of Classes   Number of Layers    Number of Parametric Layers
    1            2                  6                          6
# ==================================================================

#  Class ID     Number of Cells    Class Name
       1              346          1
# Layers             1           2           3           4           5           6
# Means
                56.74566    40.38728    28.94220    66.43642    41.45954    20.97399
# Covariance
    1            2.16121     0.77125     0.47511     3.87943     4.42736     1.83394
    2            0.77125     3.51914     1.29202    11.99861    13.69398     5.23329
    3            0.47511     1.29202     3.27781     4.38182     6.09621     2.55212
    4            3.87943    11.99861     4.38182    70.72203    73.15249    26.82588
    5            4.42736    13.69398     6.09621    73.15249    86.95053    32.99460
    6            1.83394     5.23329     2.55212    26.82588    32.99460    14.81381
# ------------------------------------------------------------------

#  Class ID     Number of Cells    Class Name
       2              149          2
# Layers             1           2           3           4           5           6
# Means
                75.03356    61.65101    59.57718    65.99329    86.00671    59.40940
# Covariance
    1           56.16779    50.66044    78.50077   -58.31058     7.27680    58.92536
    2           50.66044    51.12062    77.10820   -46.60371    17.95506    61.78573
    3           78.50077    77.10820   128.92137   -81.39475    32.30015   104.02562
    4          -58.31058   -46.60371   -81.39475   119.56077    46.36491   -45.51075
    5            7.27680    17.95506    32.30015    46.36491   147.54725    84.61886
    6           58.92536    61.78573   104.02562   -45.51075    84.61886   122.89207
```

The signature file above specifies the mean values of the histograms and their variability (covariance) within each image band. The signature file is a statistical summary of the histograms for each band, for each information class. For example, the file contains the mean and covariance of DNs for pixels belonging to the class **Forest**, and likewise for **NonForest**, for each band. This information allows a machine classifier to place each pixel in the image into its respective class.

Using the signature file created from the training histograms, a supervised classification is one that assigns all pixels in the image to the information classes specified in the signature file. The classifier will look at your signature file and then assign each pixel in your image to one of the information categories that you defined in the training dataset. The outcome of the classifier is a thematic raster layer that could then be used in map algebra or other raster modelling processes:

Student Name: ______________________________ Student Number: _____________________________



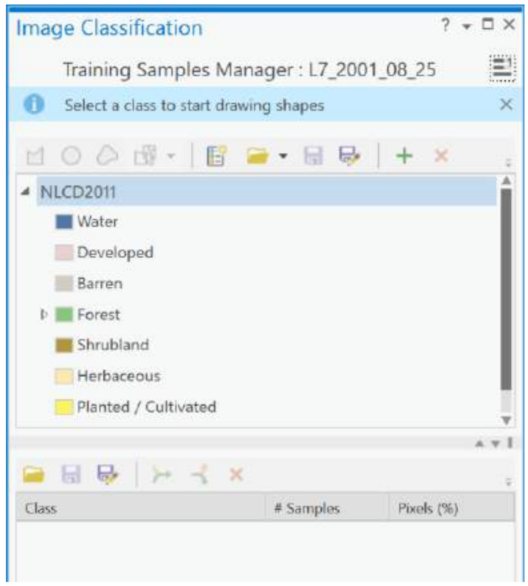
Classifiers are AI/machine learning algorithms that learn from your training data (in the signature file) how to classify any pixel in the image, including those pixels that are not part of your training set. There are many types of classifiers that have various benefits and drawbacks; these go by names such as Maximum Likelihood, Neural Networks, Support Vector Machines, K-means, Random Forest, and Iso-Cluster. Regardless of which classifier you work with, the quality of the training data is paramount and has the greatest effect on the accuracy of the resulting map layer.

Ok, let’s get started. We’ll first work with a Landsat 7 image from 2001, August 25th called “L7_2001_08_25” in the IMAGERY.GDB:

3. Add the “L7_2001_08_25” to ArcGIS and ensure that you choose it in the Contents pane:

Student Name: ______________________ Student Number: ______________________

4. Click on the Imagery tab and then click on the button 'Classification Tools' and choose the Training Samples Manager. There may be a default set of information classes, don't worry, we will deal with that in the next steps.

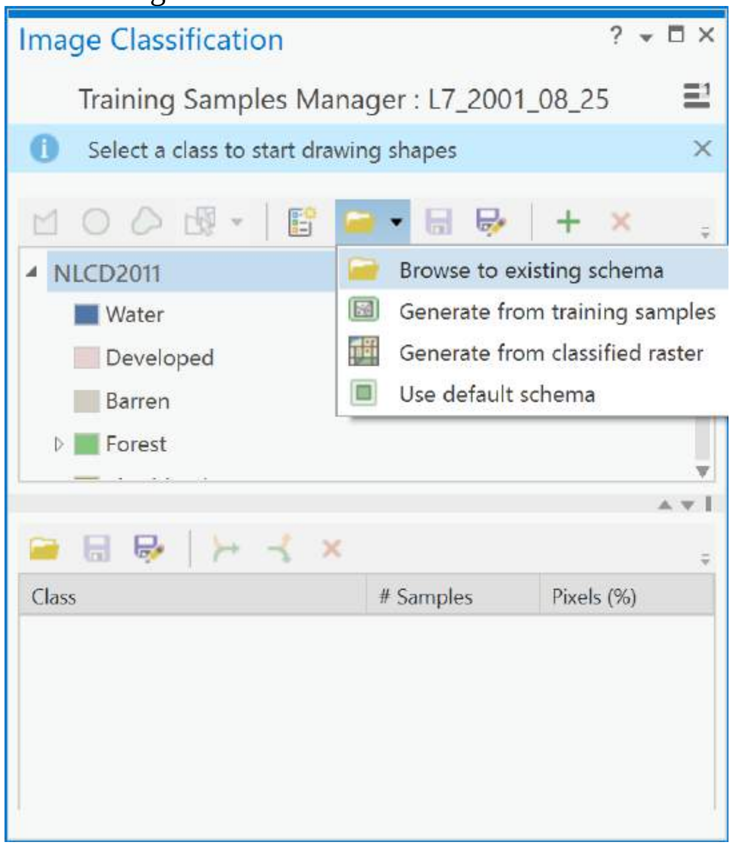


5. Below is the description of the ***Training Samples Manager***.

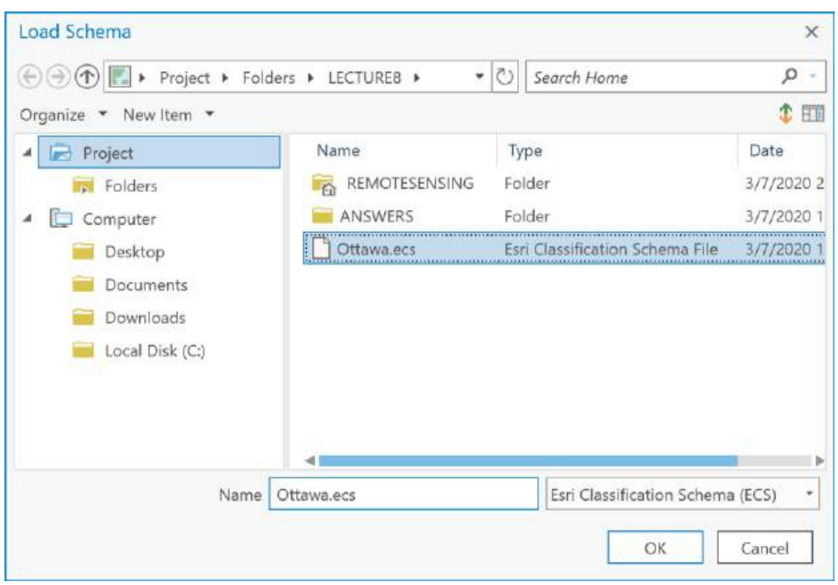
Tool	Function
	Create a training sample by drawing a polygon around pixels or objects in the raster.
	Create a training sample by drawing a circle around pixels or objects in the raster.
	Create a training sample by drawing a freehand shape around pixels or objects in the raster.
	Create a training sample by selecting a segment from a segmented layer. This option is only available if there is a segmented layer in the **Contents** pane. Activate the **Segment Picker** by highlighting the segmented layer in the **Contents** pane, then select the layer from the **Segment Picker** drop-down list.
	Create a new classification schema. Right-click the **New Schema** title and click **Add New Class** to begin creating class categories.
	Select a classification schema option. • Browse to an existing schema. • Generate a new schema from an existing training sample feature class. • Generate a new schema from an existing classified raster. • Use the default 2011 National Land Cover Database schema.
	Save changes to the schema.
	Save a new copy of the schema.
	Add a class category to the schema. Select the name of the schema first to create a new parent class at the highest level. Select the name of an existing class to create a subclass.
	Remove the selected class or subclass category from the schema.

Student Name: ___________________________ Student Number: ___________________________

6. Using the Training Samples Manager, you will open a pre-existing training set of polygons within a feature class called "L7_2001_08_25_Training" within the IMAGERY.GDB geodatabase. First, click on the 'Classification schema' button and choose Browse to an existing schema:

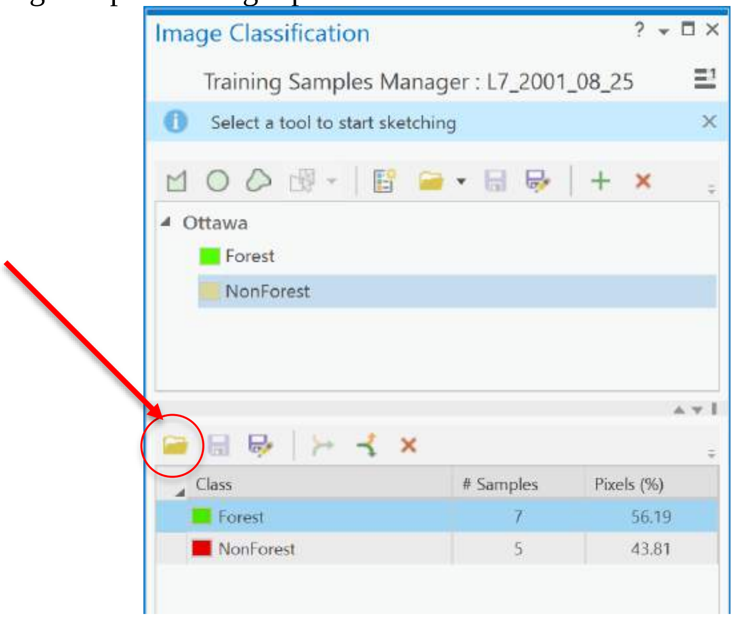


7. Navigate to where you unzipped the lab5.zip file, choose the "Ottawa.ecs" file, and click OK:



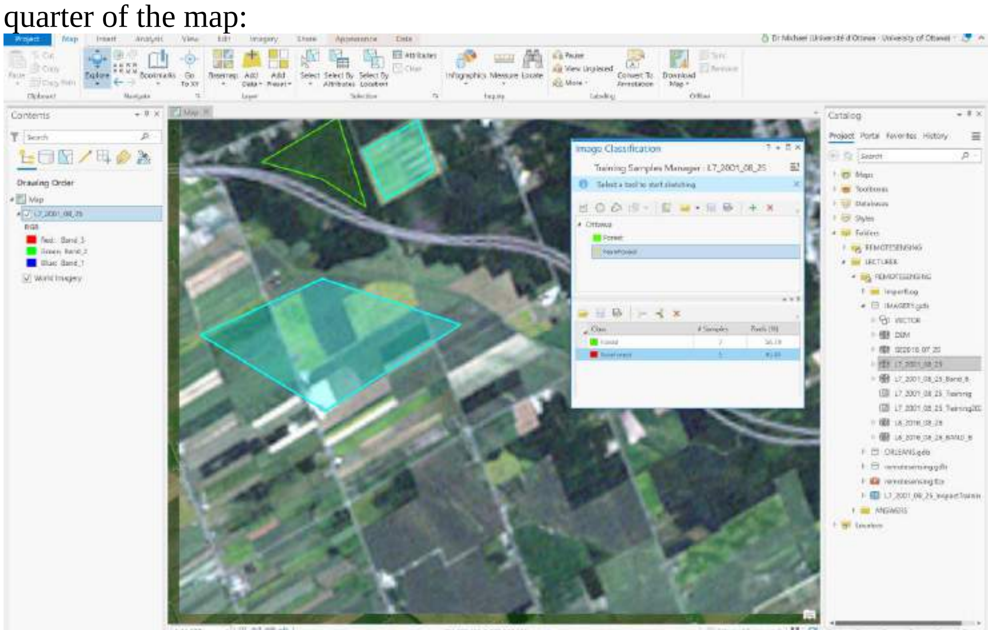
Student Name: ______________________________ Student Number: ____________________________

8. Now, click on the ‘Load training samples’ button in the lower half-pane and navigate to the and choose the “L7_2001_08_25_Training” feature class from IMAGERY.GBD. Your Training Samples Manager pane should now look like this:



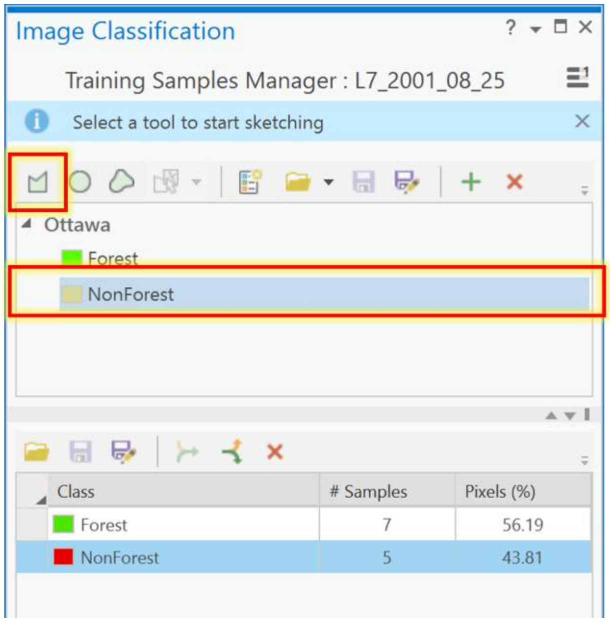
You can see that the **Forest** class corresponds to the color green and the training sites are shown as green polygons in the map. Likewise, for the **NonForest** information class which has polygons shown as red. What is important here is that the two information classes, **Forest** and **NonForest** have approximately the same percentage of pixels in each class. Right now, the **Forest** class has 56% and the **NonForest** has 44%. That’s not too bad, but we will still add a new polygon to the **NonForest** class to increase its percentage.

9. To add another example of **NonForest** to the training samples, zoom into the bottom left quarter of the map:

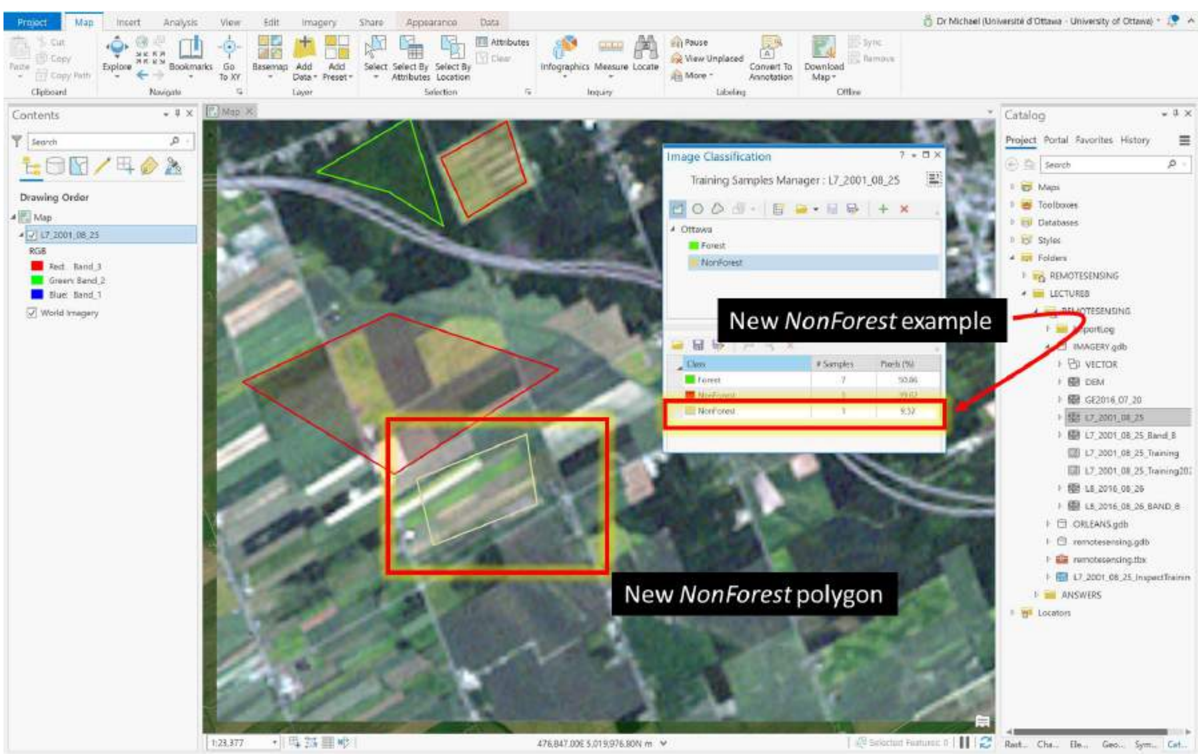


Student Name: ______________________________ Student Number: _____________________________

10. Next, in the Training Samples Manager, in the upper half-pane, under the schema called Ottawa, click on the **NonForest** information class. You will see that the polygon tool in the toolbar becomes active (this is the left-most tool; *newer versions of ArcGIS may have a rectangle button first, making the Polygon button second. Make sure to select the proper tool*).

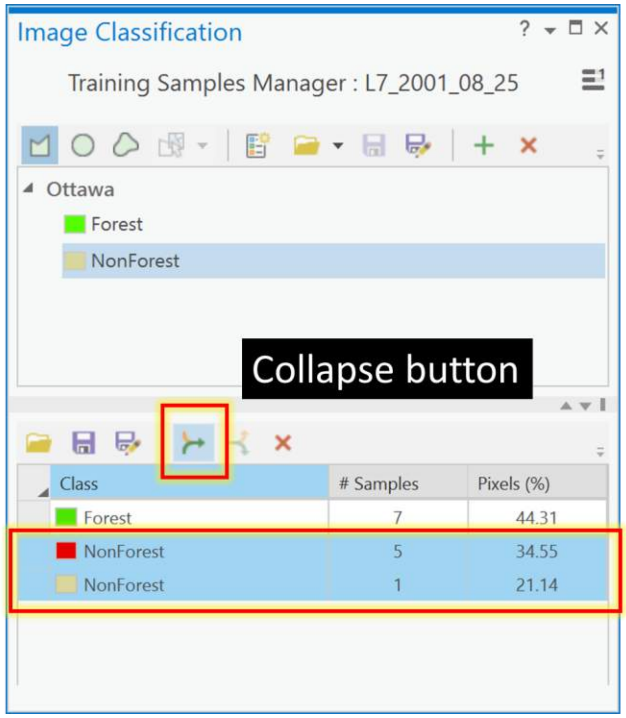


11. Now, click on the polygon tool and draw a polygon in any non-forested area on the satellite image. **Do not choose any other drawing tool than the Polygon tool to draw your polygon**:

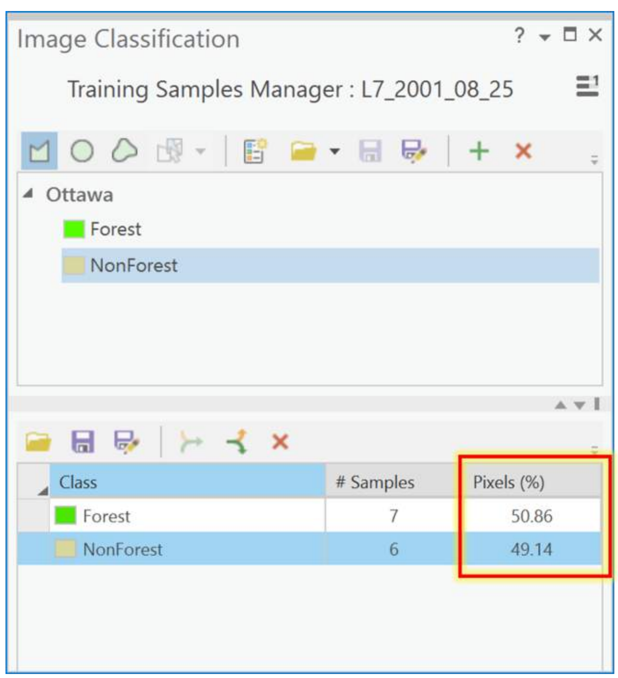


Student Name: ____________________________ Student Number: ___________________________

12. Examine the **Training Samples Manager** pane, the lower half-pane, and you will now see the original **NonForest** example polygons in red and the new **NonForest** example that you just made. Hold down your Ctrl key and click on both **NonForest** examples and click on the 'Collapse' button to merge the two **NonForest** examples into a single set:



13. Now your Training Samples Manager pane will look like this:



14. Check that the two information classes now have approximately 50% pixels each. If the polygon you added was too large then the NonForest class will now have much more than 50% of the pixels, in which case you did not succeed in balancing the size of the training samples from the two classes. If that happened you can either add a polygon to the Forest class (and so on, until you get it right), or you can do the following:

    - Select the NonForest class in the bottom half of the Training Samples Manager, and click the “Expand” button next to the “Collapse” button.
    - You will now see several rows labeled NonForest, each corresponding to an individual polygon.
    - Delete the last one – the one you just created.
    - Add a new (smaller) one.
    - Select all the NonForest polygons, and collapse them.
    - Check that you ended up more or less at a 50/50 distribution between the two classes.

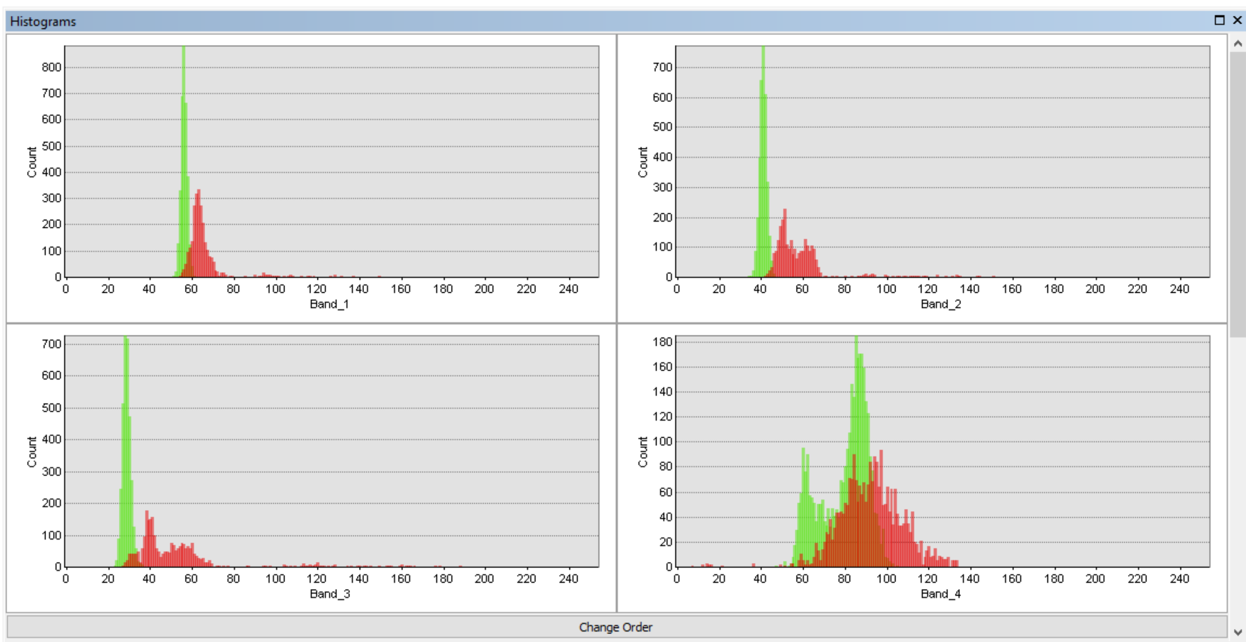
15. Examine the **Training Samples Manager** pane carefully, it provides a lot of important information about the training samples for each information class. One thing not shown is that **Forest** (the first-listed class) is given a pixel value of 1 and **NonForest** (the next-listed class) is given a pixel value of 2 (and so on, if you had more classes). These values will become the cell values for any classification of the image. The ‘Pixels %’ column tells you how many pixels are in each information training class. In the above case, 50.86% of the training pixels are in the **Forest** information class and 49.14% are in the **NonForest** training set. So here we have achieved an almost equal proportion and that is acceptable and better than what we started with.

16. Next, we need to save our new training polygons to a feature class. To do so, click on the ‘Save’ button in the lower-pane of the **Training Samples Manager** and navigate to the IMAGERY.GDB and save the new training file with the name “L7_2001_08_25_Training2023”. We don’t need to save the ‘Ottawa’ schema in the top half-pane because it hasn’t changed. Had we added a new information class then we would also have had to save the new schema.

17. This next step is fundamental because we will explore how ‘pure’ our information classes are based on the pixels that we chose for each class. We want to find out which bands have the most *separability* between **Forest** and **NonForest**. ***Spectral separability*** refers to the spectral distance between the pixel values that compose each information class in a training samples dataset. Ultimately, we want to know which bands of our remotely sensed image would be best for classifying our image into **Forest** and **NonForest**. The bands that are best will clearly distinguish between our two information classes and there will be little to no overlap between the pixel values for **Forest** and **NonForest** in those bands.

Student Name: ____________________________ Student Number: ___________________________

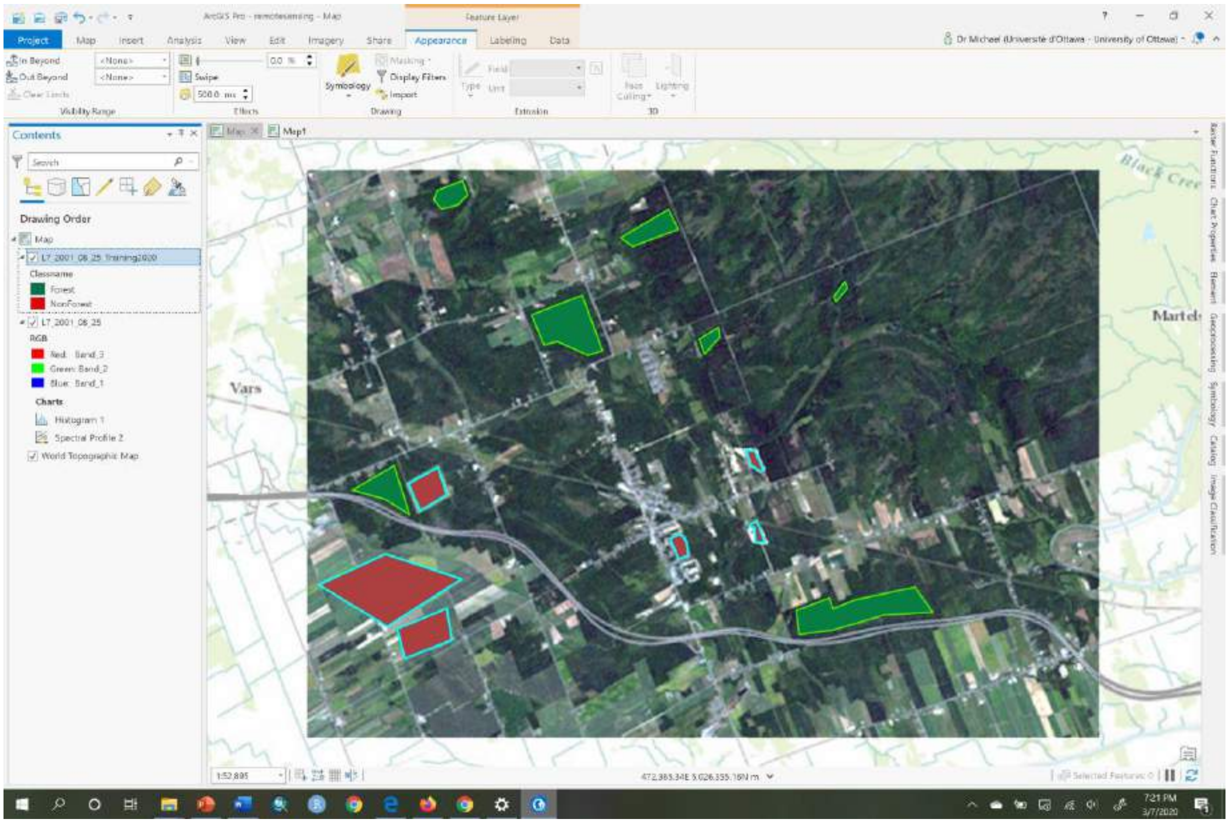
To understand spectral separability, look at the figure below to see histograms for four bands and two information classes differentiated as red and green colors. Each histogram shows the frequency of digital numbers for each class, red class or green class, on the y-axis and the range of digital numbers associated with the red or green class on the x-axis. For example, in Band_1 we see that the green class ranges from about 52-60 and the red class from around 55-150. So, while the two histograms are clearly different, there is also some overlap around values of 55-60, shown in dark red. However, there is no overlap in the peaks of the red and green histograms which is a good thing. In general, a 'good' band for classification purposes is one that would have no overlap at all between the red and green histograms. If there is no overlap between information classes in a band then the red and green classes are distinct and completely separated. As such, such a band would be good for classifying the image into red and green information classes. Here, none of the image bands show complete separability. The band with the smallest overlap is Band_2. Why? Because there is little overlap between the red and green histograms. Conversely, the worst band is Band_4, where there is considerable overlap and so Band_4 would not help in classifying the corresponding image into red and green bands.



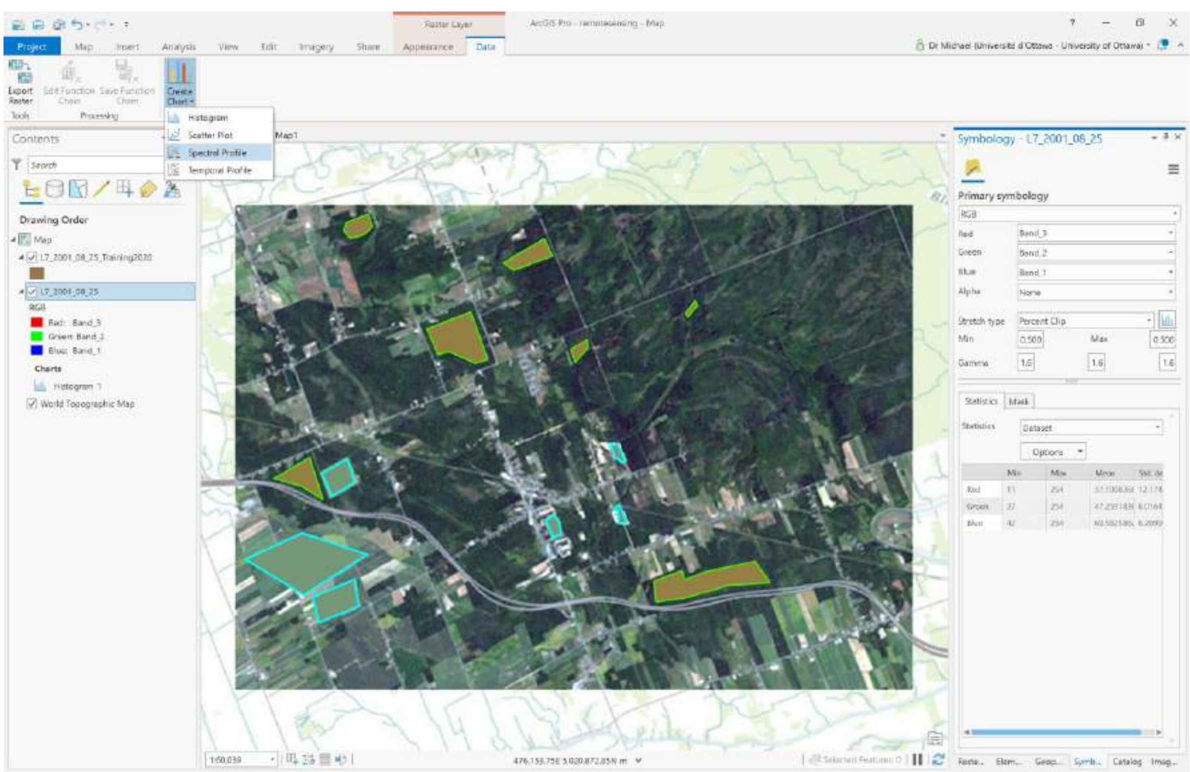
We don't need to generate histograms for each band to examine separability. The most common way to examine spectral separability among the different bands is to generate a plot of the statistics of each histogram for each band in a spectral profile.

18. To create a spectral plot, first, ensure you have added your "L7_2001_08_25_Training2020" layer to the contents pane and symbolize it using unique values and the ClassName field, and make **Forest** as green and **NonForest** red.

Student Name: ___________________________ Student Number: __________________________

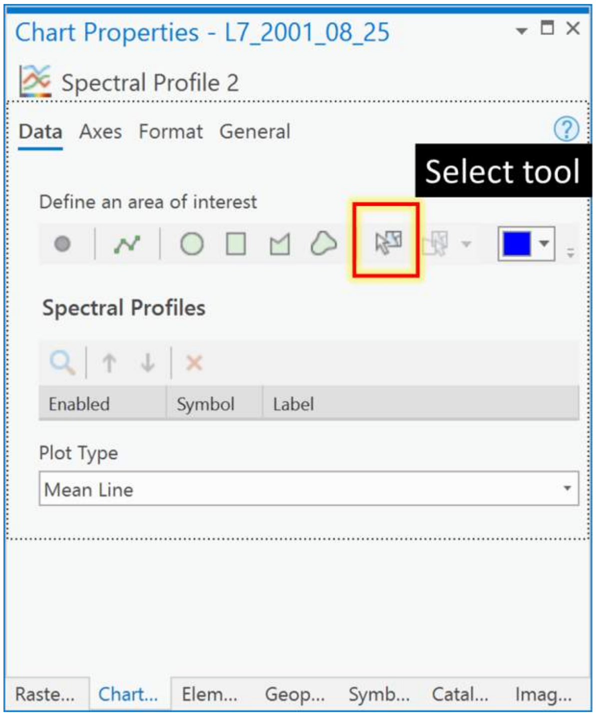


19. Next, choose the “L7_2001_08_25” layer in the contents pane. Then, click on the Data tab. Now click on the Create Chart and choose ‘Spectral Profile’:

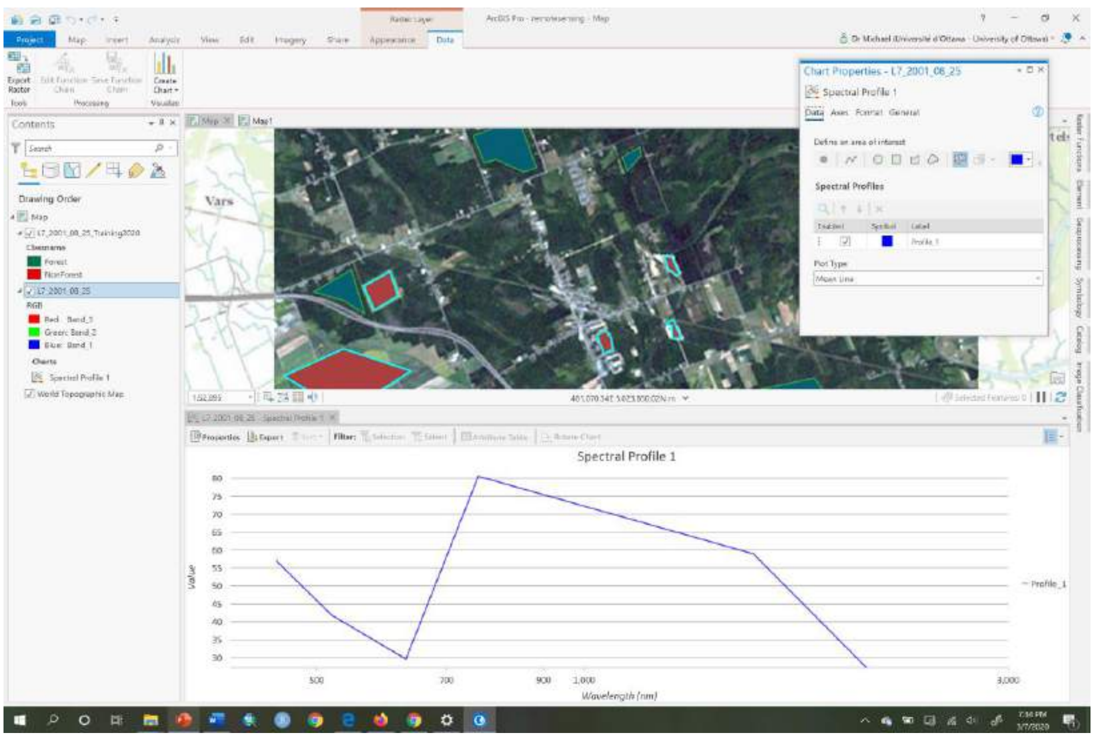


Student Name: ______________________________ Student Number: ______________________________

20. Next, in the Chart Properties pane, click on the Feature selector button:

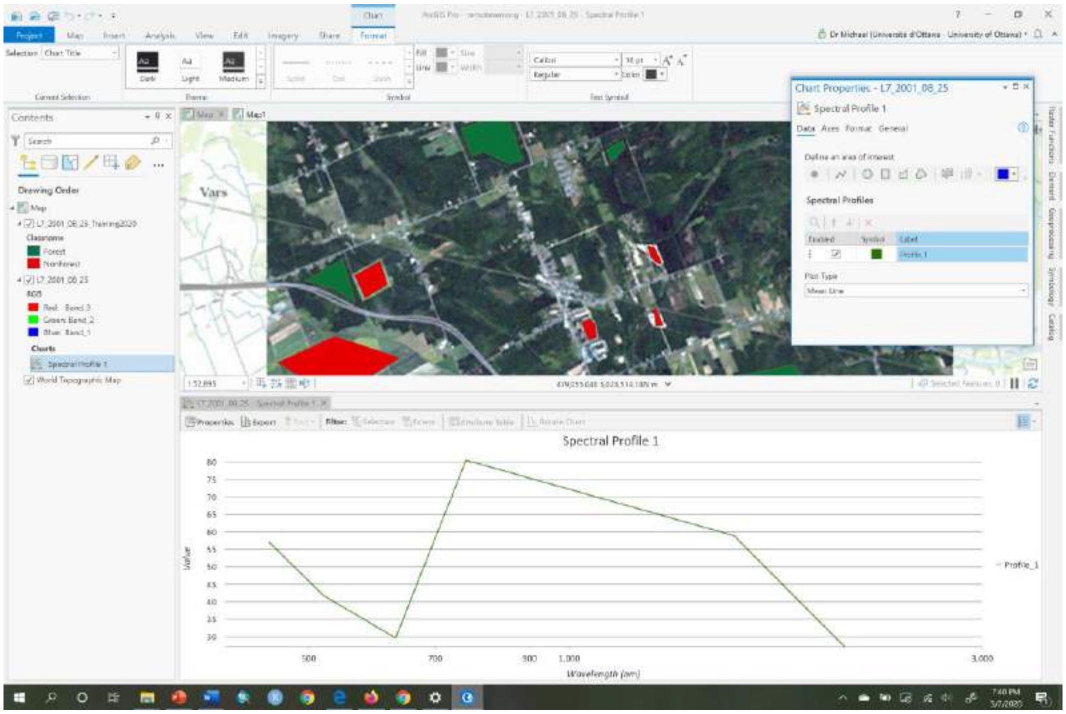


21. Now, click on any of the **Forest** polygons in the map pane. You will see a chart created with a single line called "Profile_1" that represents the average or mean of the DNs across all bands for that information class called **Forest**.

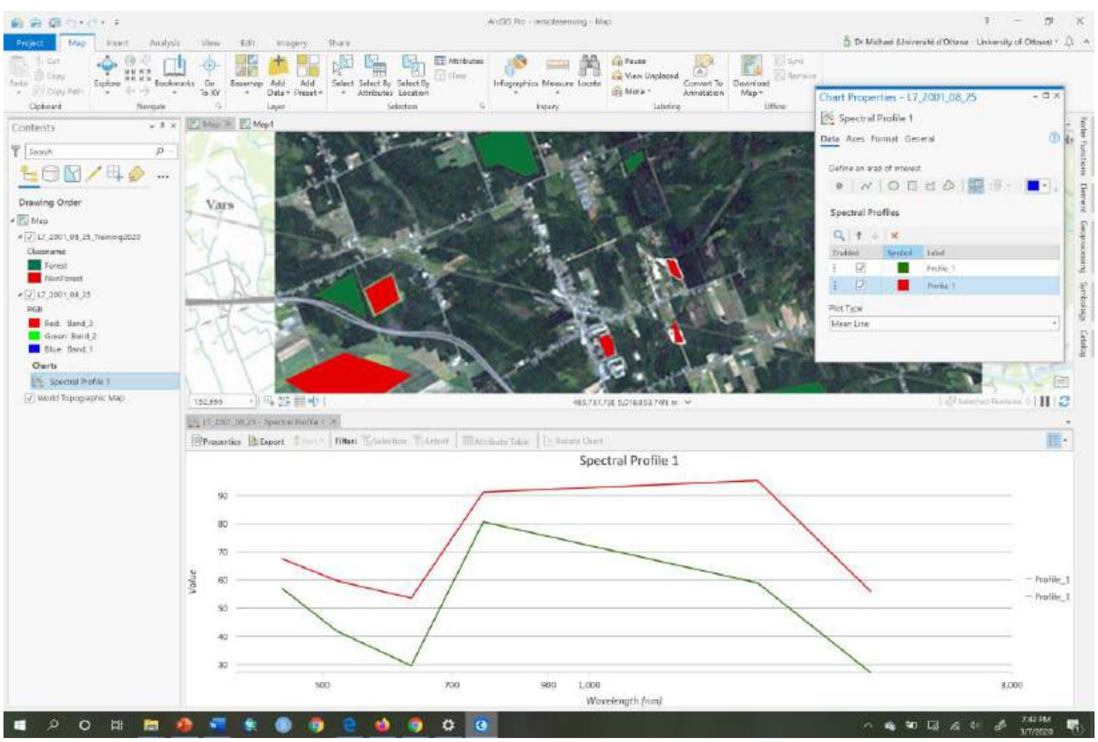


Student Name: ______________________________ Student Number: ______________________________

22. Now, in the Chart Properties pane, change the label for this spectral mean line from “Profile_1” to **Forest** (if it works!) and change the color to green and you will see your line become green:

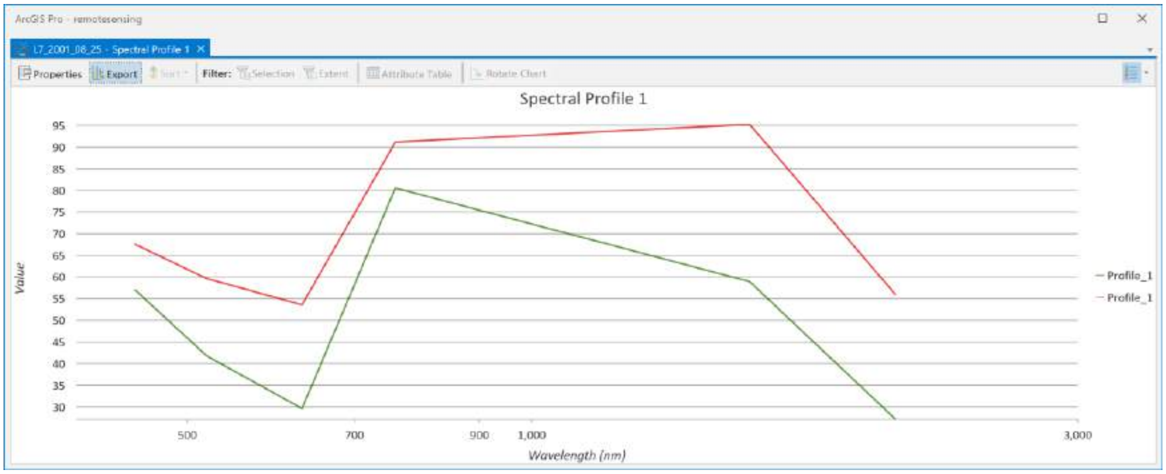


23. Now repeat 21-22 but for the **NonForest** training polygons and make the new line red to correspond to **NonForest** and rename the label if you can in your version of ArcGIS Pro:



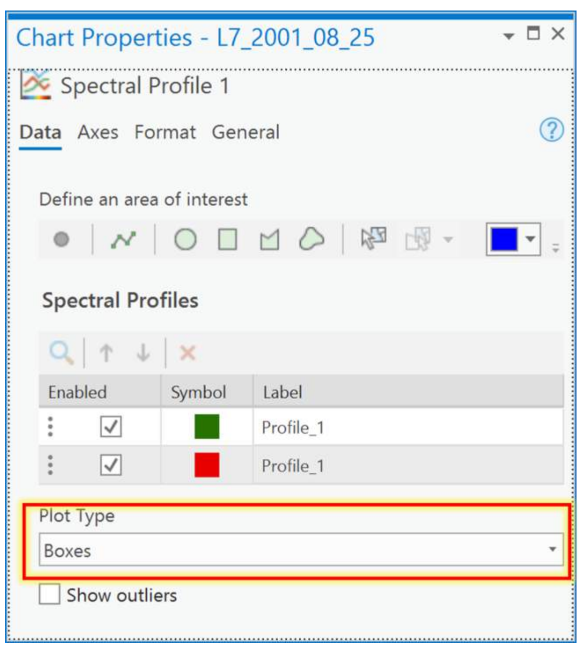
Student Name: ____________________________ Student Number: ___________________________

24. Now examine the Spectral Profile 1 graph:



This graph shows wavelength on the x-axis rather than band labels because the metadata of the image contains information on the wavelength of each band. You can infer the bands by seeing where the lines have their 'kinks'. For example, Band_1 is at the far left, Band_2 is just after 500 nm and Band_3 is just before 700 nm, etc. What we see is that in each band there is a difference between the mean values for each information class. Perhaps the least separability might be in Band 4, just after 700 nm, since the lines are closest there. We can check with a more detailed statistical plot called a box-and-whisker diagram.

25. To create the box-and-whisker plot of the spectral separability for each band we only need to change the plot type in the Chart Properties pane. In the pane change the Plot Type from 'mean' to 'boxes':

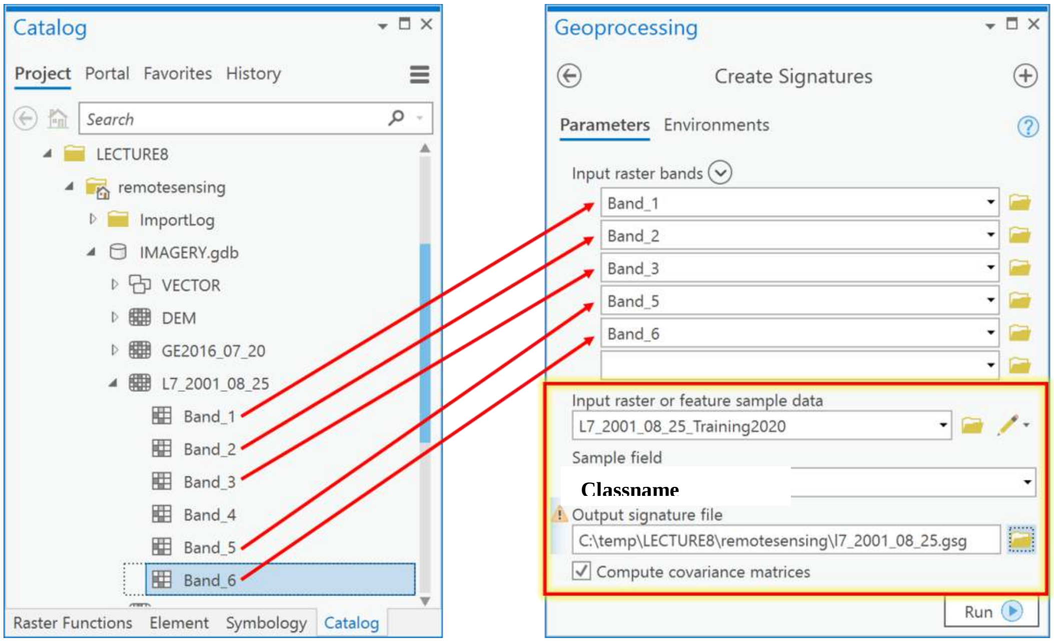


26. Now examine the box and whisker plot.

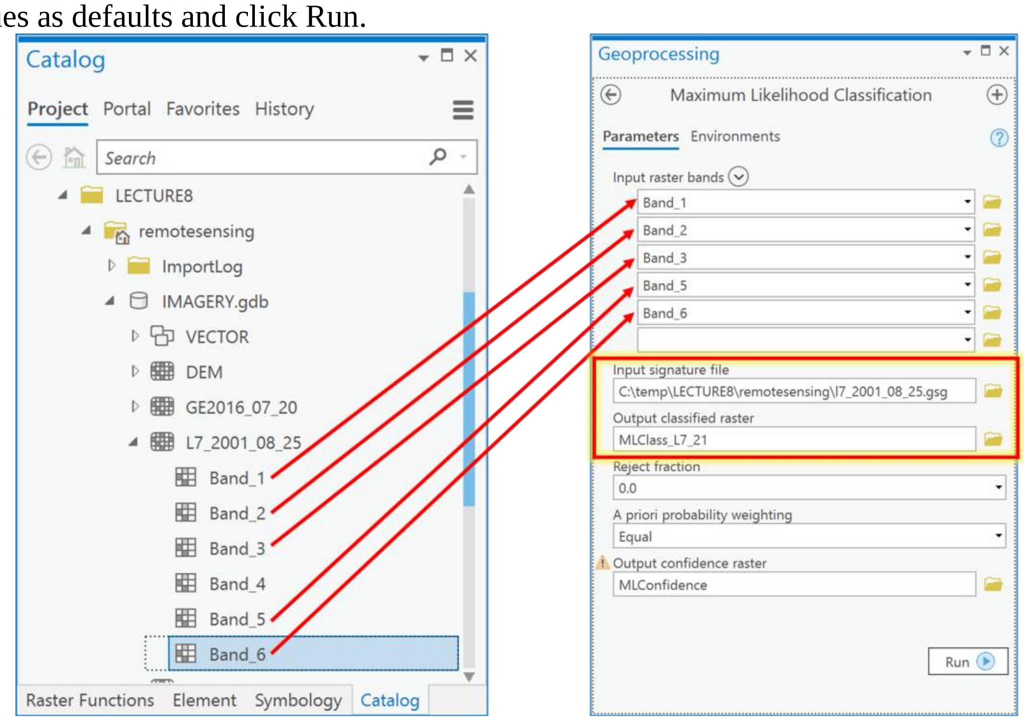
In this plot, there is actually a box-and-whisker plot for each band, each one showing the **Forest** (green) and **NonForest** (red) DNs. The top and bottom of the solid-colored boxes are the upper and lower quartiles of the DN histogram for the information class. The whiskers are the two lines outside the box that extend to the minimum and maximum DNs for the information class. So basically, these are a visual summary of the histograms for each information class, in which we can now examine which bands have the best separability. There is much more information here than in the Spectral Profile graph. What really care about seeing here is that the boxes don't overlap, and they don't for all bands except for Band_4, which is the one that was also closest together in the Spectral Profile plot. So, we can consider excluding Band_4 from any classification of **Forest** and **NonForest**, but we should probably keep all the others.

27. Create a signature file from the training sites that we can use in classification. To create the file, go to the Analysis tab, click on Tools, and search for the Create Signatures tool from the Spatial Analyst toolbox. Undock the Create Signatures tool so that you can see that and your Catalog pane at the same time. Then, in the Catalog pane, navigate to IMAGERY.GDB and expand the "L7_2001_08_25" image so that you can see all the individual bands (Band_1, Band_2,…,Band_6). Drag all bands **except Band_4** (since we know it has bad separability) into the 'Input raster bands', under 'Input raster or feature sample data' choose the "L7_2001_08_25_Training2020" feature class, under 'Sample field' choose Classname, and save the file to the folder where you have your data, and call it "l7_2001_08_25.gsg". Click Run.

Student Name: ______________________ Student Number: ______________________

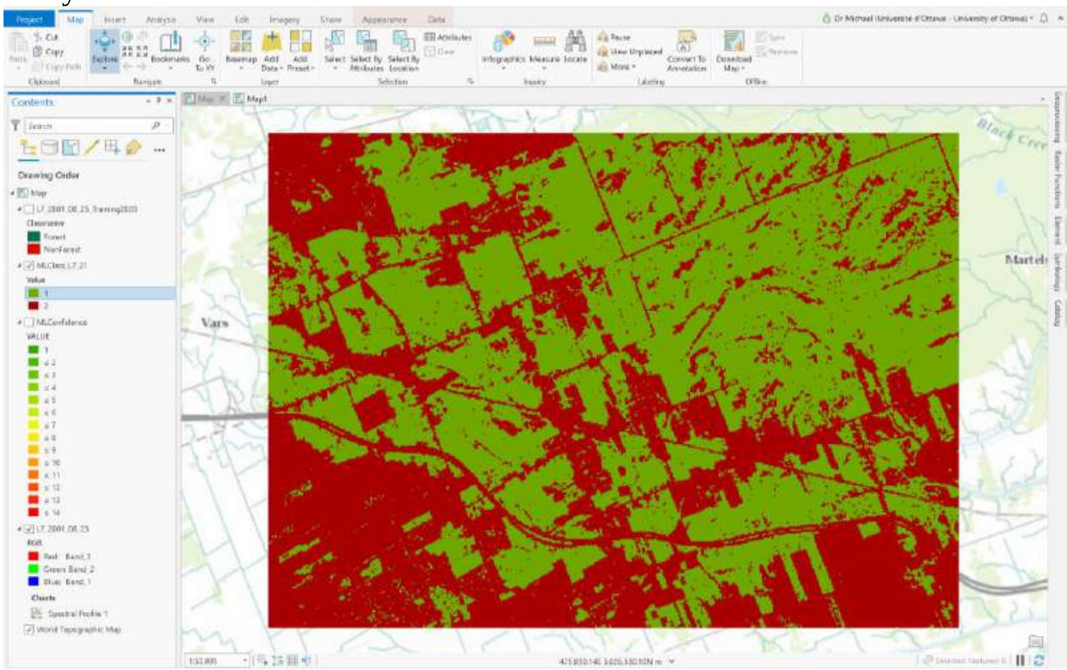


28. Now, use the signature file to classify the image into the two information classes of **Forest** and **NonForest**. To do so, search for the Spatial Analyst tool "Maximum Likelihood Classification". Undock the Maximum Likelihood Classification tool so that you can see it with your Catalog pane at the same time. Then, in the Catalog pane, navigate to IMAGERY.GDB and expand the “L7_2001_08_25” image so that you can see all the individual bands (Band_1, Band_2, …, Band_6). Drag all bands **except** Band_4 into the ‘Input raster bands’, under ‘Input signature file’, choose the signature file created in the previous step called “l7_2001_08_25.gsg”, under ‘Output classified raster’ call it “MLCLASS_L7_21”, and save it in IMAGERY.GDB. Leave all other values as defaults and click Run.



Student Name: ____________________________ Student Number: ___________________________

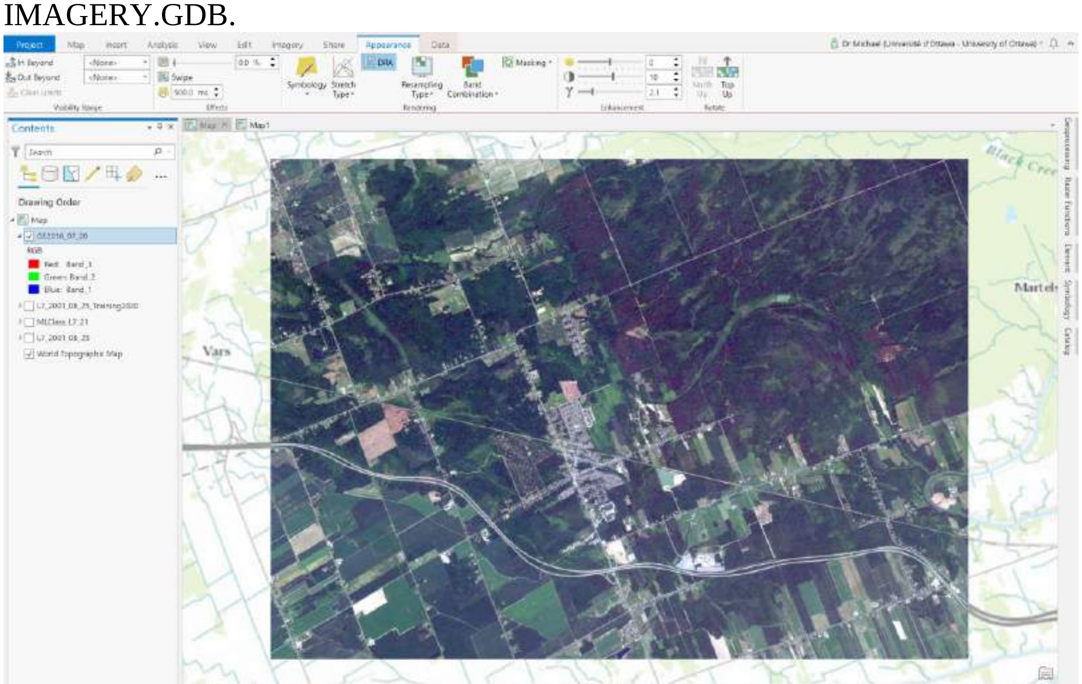
29. Examine the resulting classified raster (note that you may want to change value 1 to green and value 2 to red). We now have converted a remotely sensed image to a thematic raster layer.



Q3: How do I create a training dataset corresponding to my information classes?

Training data consists of pixels that belong to different information classes. For example, all pixels that belong to the class called **Forest**, and another set of pixels for the class called **NonForest**.

1. If it's not there already, add the "RE2016_07_20" image to ArcGIS from IMAGERY.GDB.

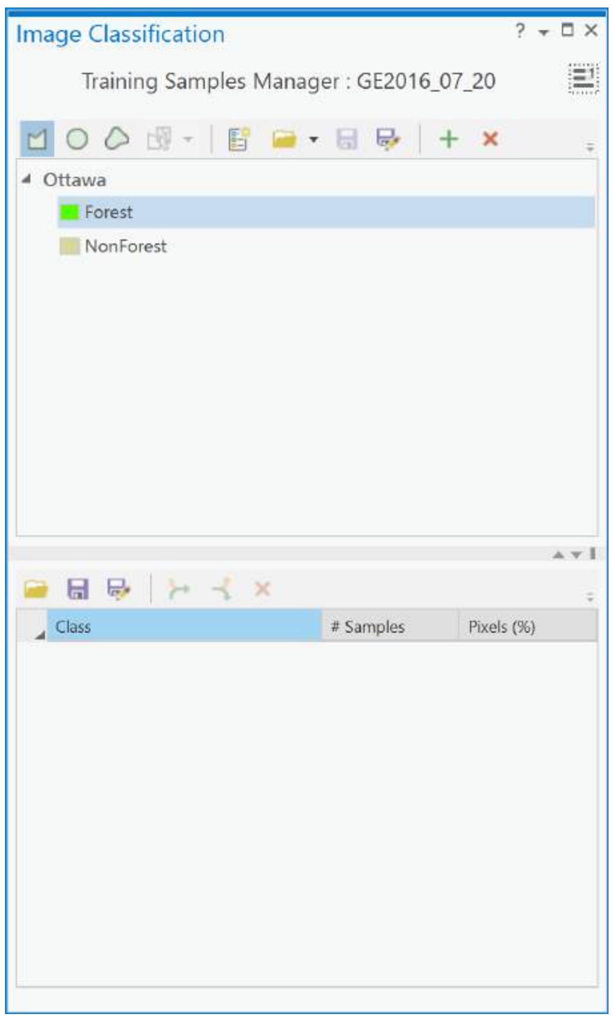


Student Name: ______________________ Student Number: ______________________

2. Click on the Imagery tab and then click on the button 'Classification Tools' and choose the Training Samples Manager.

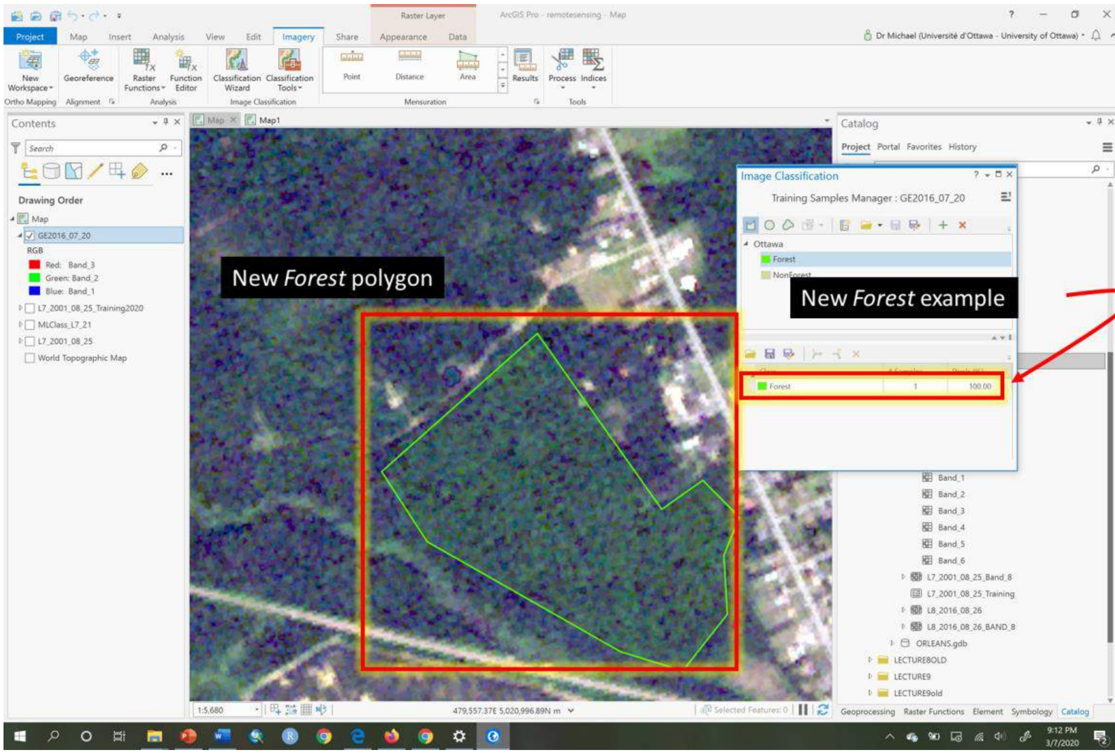
3. Using the Training Samples Manager, open the Ottawa.ecs schema, as you also did earlier.

4. Now we will create a new set of training polygons for the **Forest** and **NonForest** information classes, to be used with the RapidEye image. Note that we cannot simply reuse the polygons we used for the Landsat image, because areas within those polygons may have changed from **Forest** to **NonForest**, or vice versa, in the period between 2001 (when the Landsat image was acquired) and 2016 (when the RapidEye image was acquired). To do this, in the Training Samples Manager, in the upper half-pane, under the schema called Ottawa, click on the **Forest** information class. You will see that the polygon tool in the toolbar becomes active.

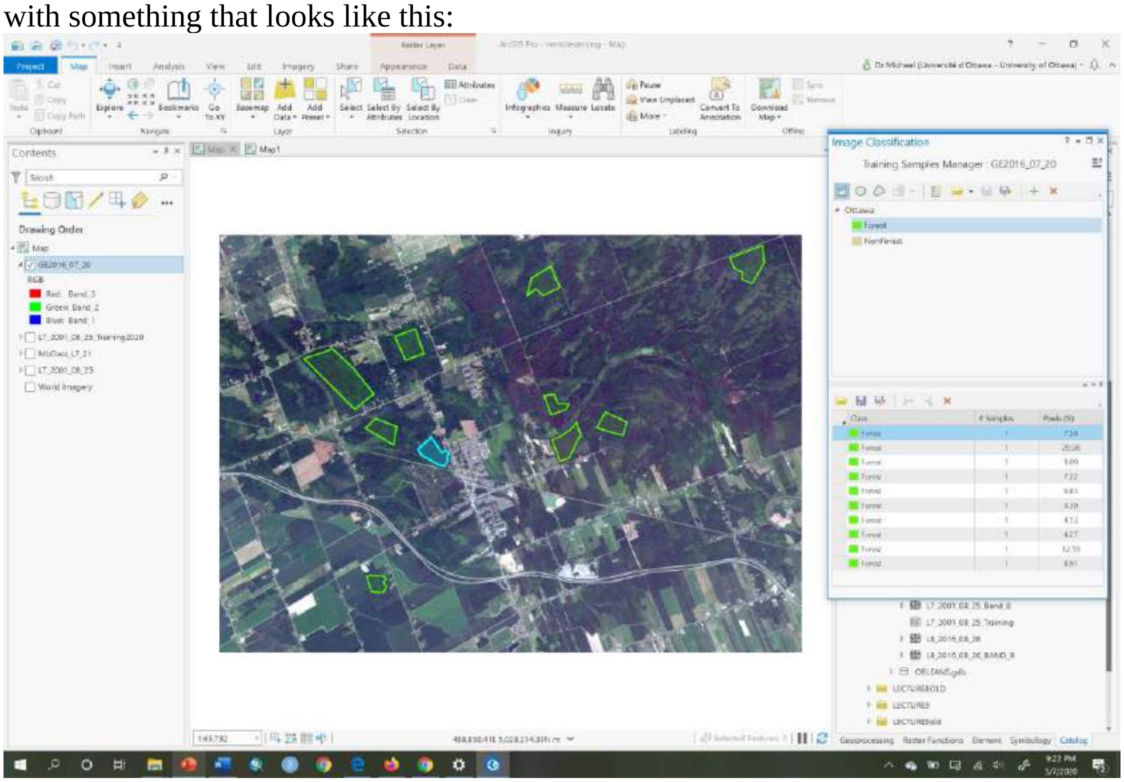


Student Name: ___________________________ Student Number: ___________________________

5. Now, click on the polygon tool and draw a polygon in any forested area on the satellite image.

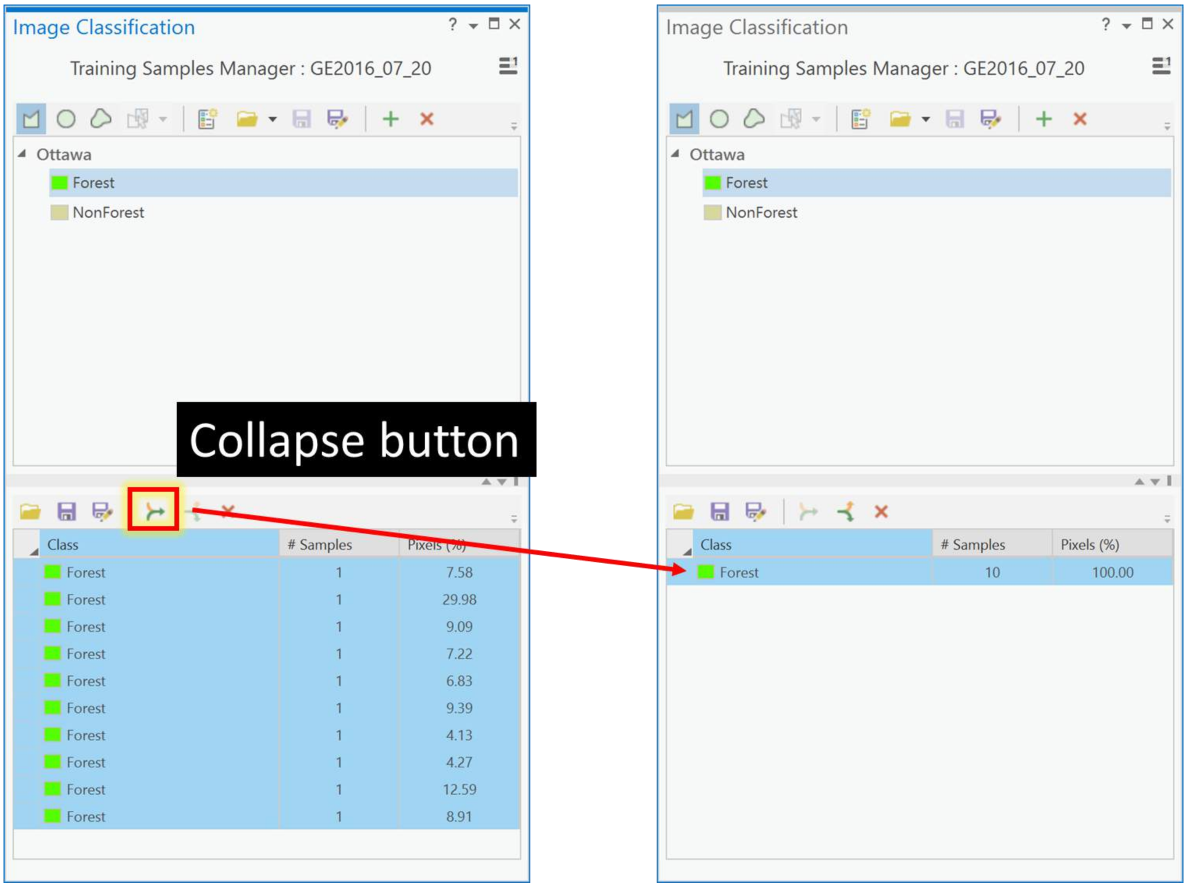


6. Now repeat the previous step nine more times in different forested areas. You will end up with something that looks like this:



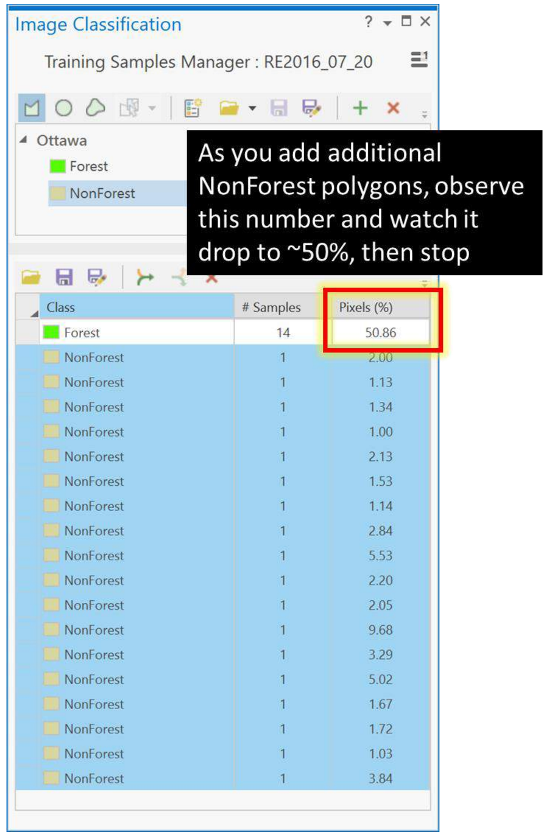
Student Name: ______________________ Student Number: ______________________

7. Examine the Training Samples Manager pane, the lower half-pane, and you will now see the 10+ **Forest** polygons listed. Hold down your Shift key and click on the first and last **Forest** polygon example to select all the **Forest** examples. Then click the ‘Collapse’ button to merge the **Forest** examples into a single group:

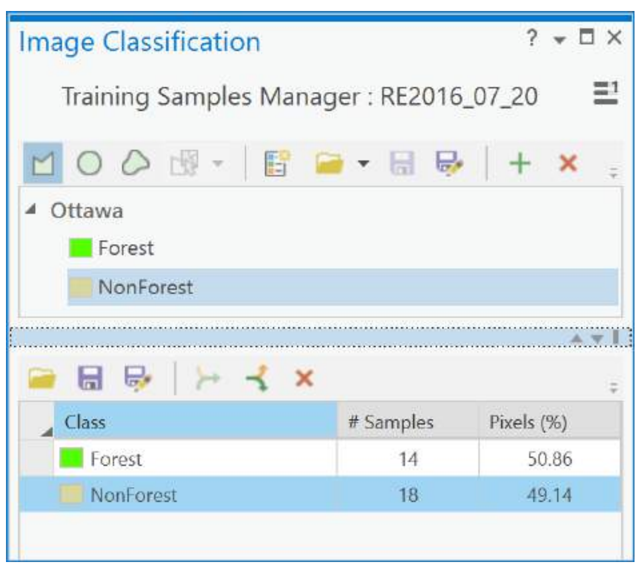


8. Next, in the Training Samples Manager pane, click on the **NonForest** and repeat steps 5-7 for the **NonForest** information class. In the end you should have a Training Samples Manager pane that looks like the one below, with near equal % of pixels in each information class. To ensure you create equal classes, the number of polygons is not what you monitor as you add additional examples of **NonForest**, instead you observe the Pixel % of the **Forest** class and try to get it to about 50%. Then merge all the **NonForest** polygons at that point:

Student Name: ______________________________ Student Number: _____________________________



9. Now your Training Samples Manager pane will look something like this:



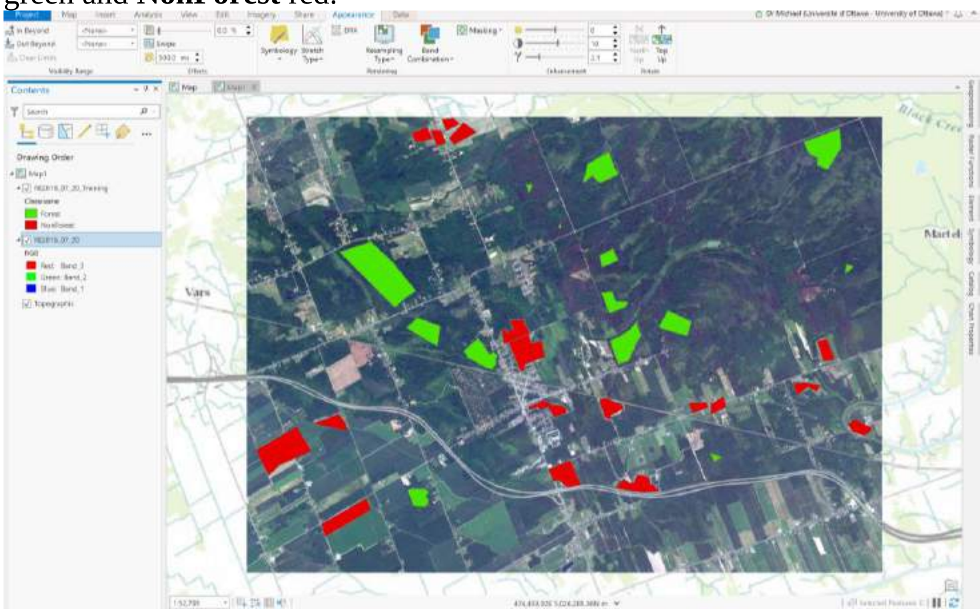
10. Next, we need to save our new training polygons to a feature class. To do so, click on the ‘Save’ button in the lower-pane of the Training Samples Manager and navigate to IMAGERY.GDB and save the new training file with the name “RE2016_07_20_training”.

Student Name: ___________________________ Student Number: __________________________

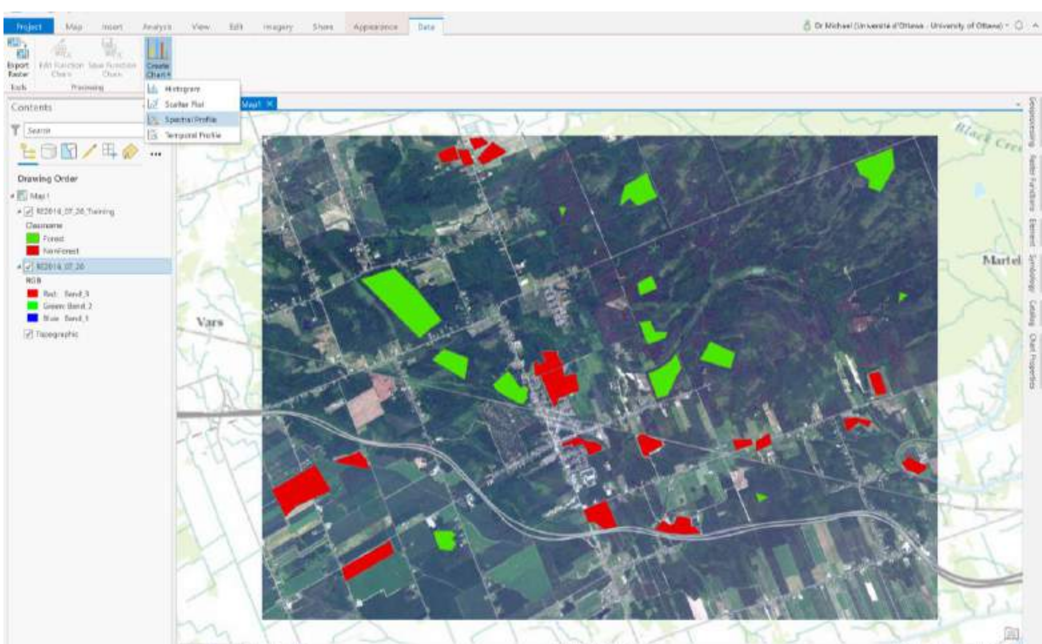
Q4: How do I create a signature file from my training polygons?

To create a signature file you need a training polygon dataset and the multiband image that you want to use to extract the pixel information from each of the training information classes. Then you need to assess separability again, before creating the signature file, because you are using a new image.

1. Add the "RE2016_07_20_training" feature class from Q3 above to the Contents pane pane and symbolize it using unique values and the Classname field, and make **Forest** as green and **NonForest** red.

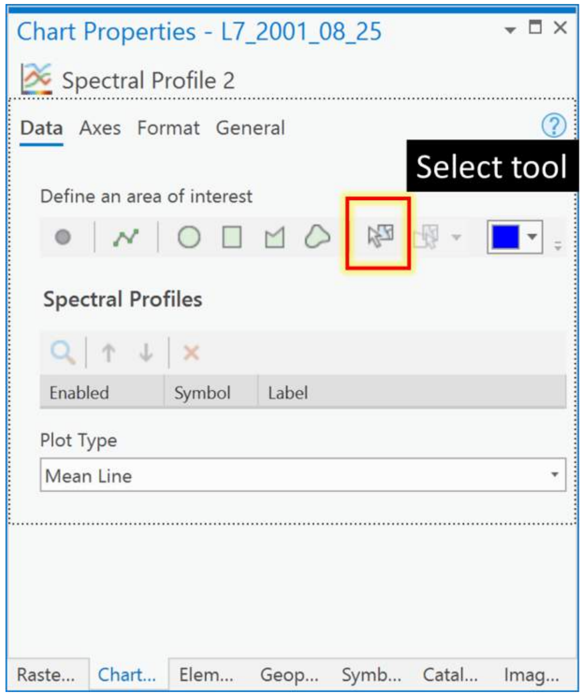


2. Next, choose the 'RE2016_07_20' layer in the contents pane. Then, click on the Data tab. Now click on the Create Chart and choose 'Spectral Profile':

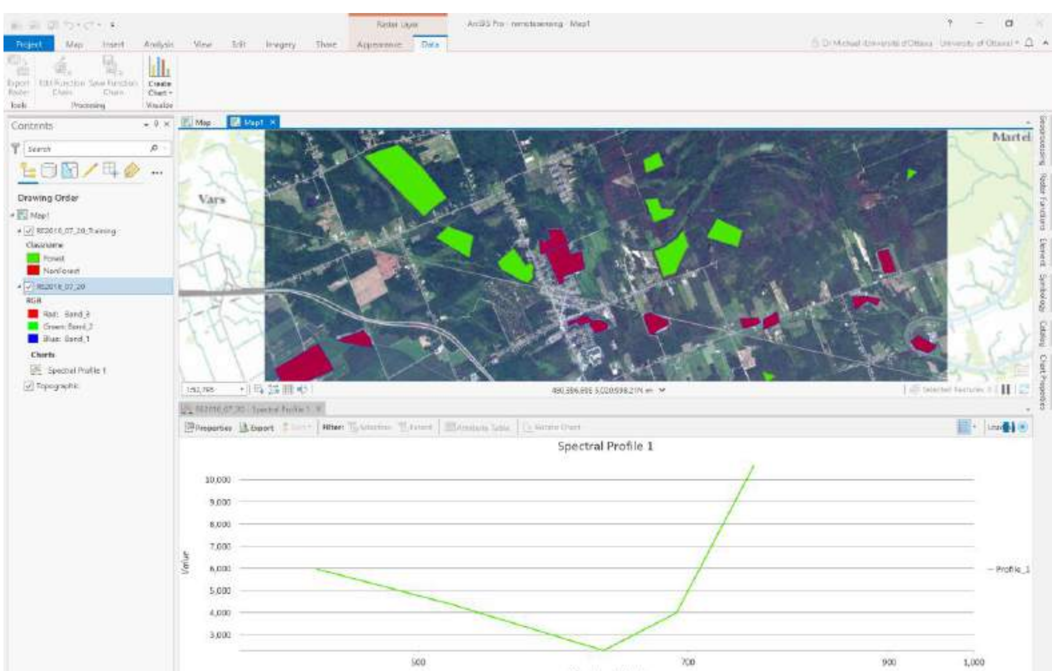


Student Name: ______________________________ Student Number: ______________________________

3. Next, in the Chart Properties pane, click on the Feature selector button:

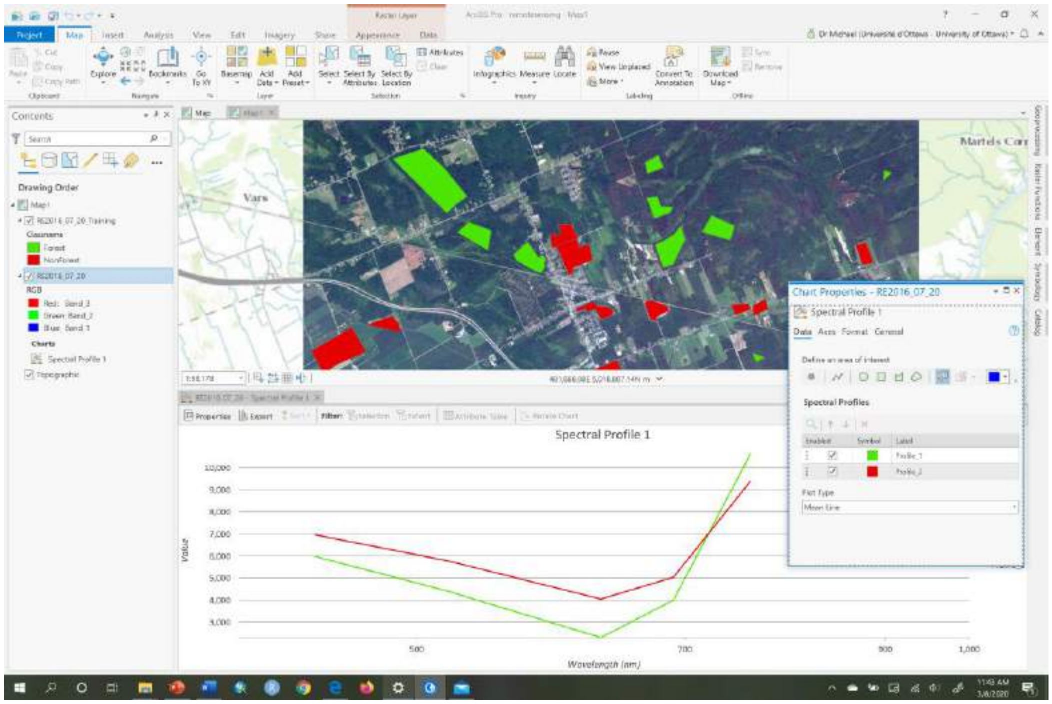


4. Now, click on any of the **Forest** polygons in the map pane. You will see a chart created with a single line called "Profile_1" that represents the average or mean of the DNs across all bands for that information class called **Forest**. In the Chart Properties pane, change the label for this spectral mean line from "Profile_1" to **Forest** (if it works!) and change the color to green and you will see your line become green:

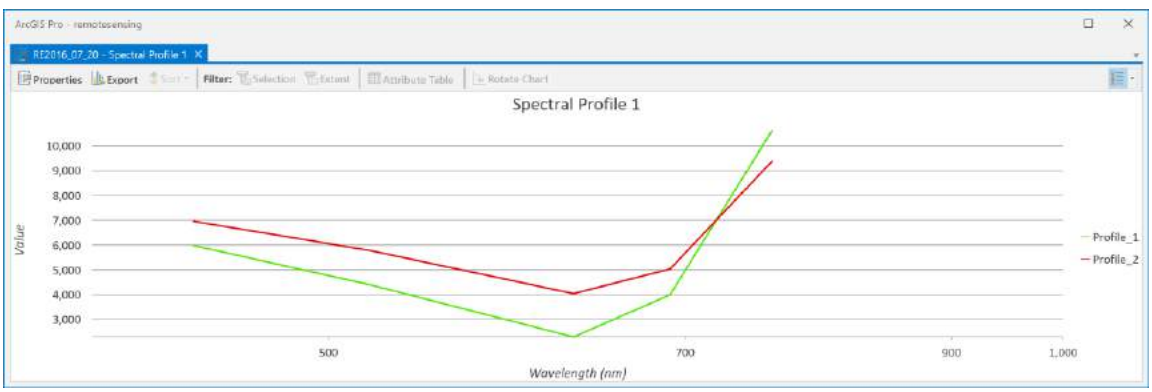


Student Name: ___________________________ Student Number: ___________________________

5. Now repeat steps 3-4 for the **NonForest** training polygons, and make the new line red to correspond to **NonForest.** Rename the label if you can in your version of ArcGIS Pro:



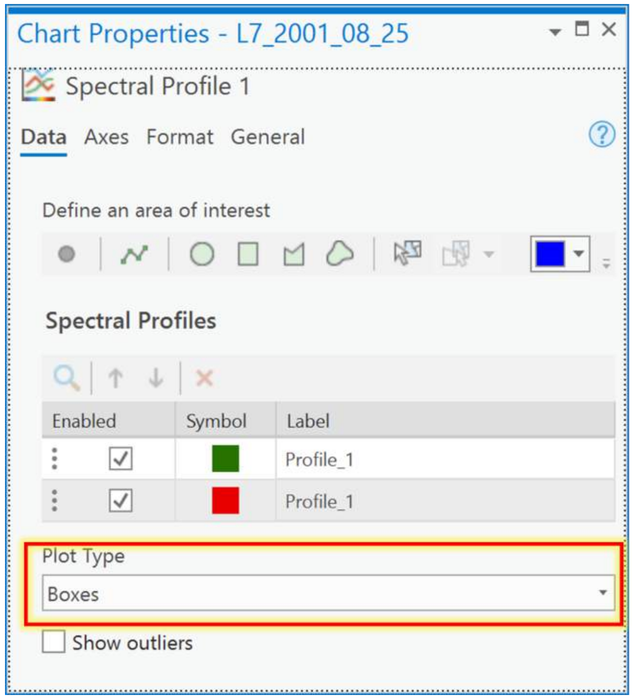
6. Now examine the Spectral Profile graph:



Just as in Q2, this graph shows wavelength on the x-axis rather than band labels, because the metadata of the image contains information on the wavelength of each band. The mean DNs show consistent differences, but it's hard to judge separability from the means alone so we need to examine the box-and-whisker plots to better choose those bands whose boxes do not overlap and so have good separability.

7. To create the box and whisker plot of the spectral separability for each band we only need to change the plot type in the Chart Properties pane. In the pane change the Plot Type from 'mean' to 'boxes':

Student Name: ______________________________ Student Number: _____________________________

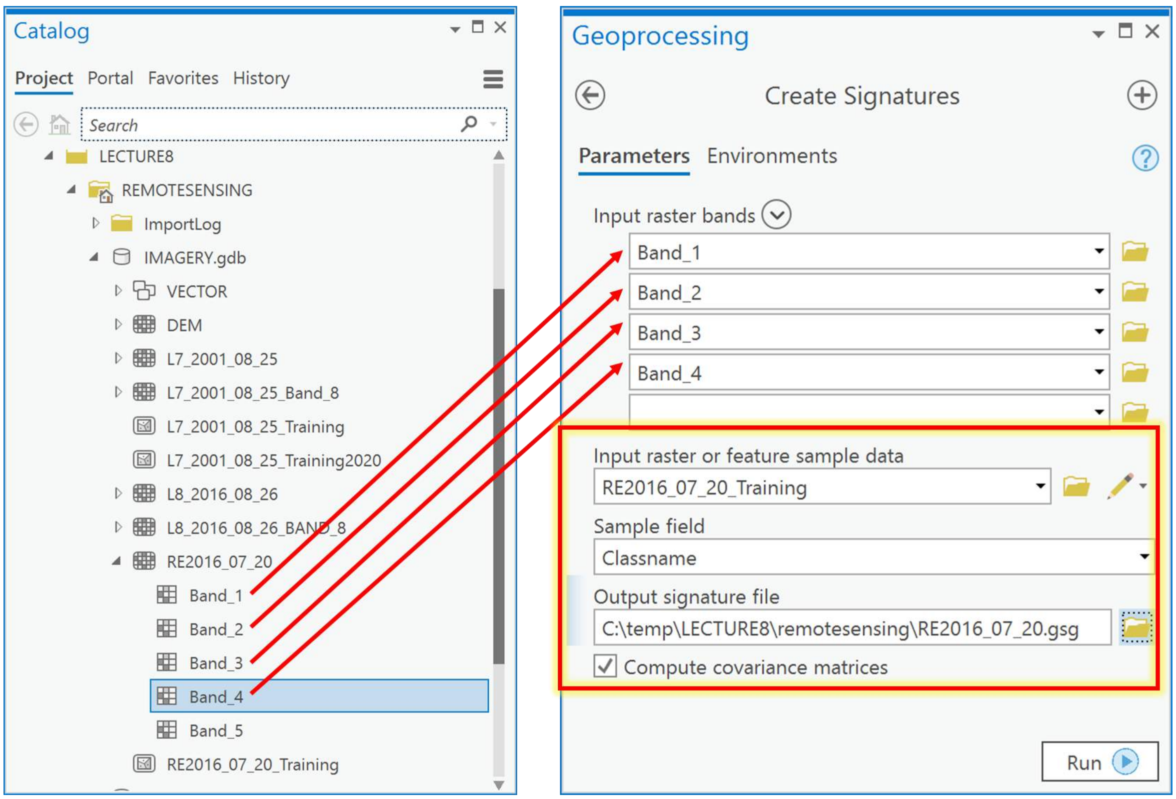


8. Now examine the box and whisker plot.

Most bands show decent separability, except for Band_5 where the boxes overlap. The spectral bandwidth of Band_5 on this RapidEye sensor is 760-850 nm, which is in the near-infrared part of the spectrum. This is the same part of the spectrum in which the "L7_2001_08_25" image showed poor separability for **Forest** and **NonForest** (in Band 4, which covers 750-900 nm for the Landsat sensor). So, for our present image, Band_5 will be excluded from any classification of **Forest** and **NonForest**.

Student Name: ___________________________ Student Number: ___________________________

9. Create a signature file from the training sites that we can use in classification. To create the file, go to Spatial Analyst -> Multivariate and choose the Create Signatures tool. Undock the Create Signatures tool so that you can see that and your Catalog pane at the same time. Then, in the Catalog pane, navigate to IMAGERY.GDB and expand the “RE2016_07_20” image so that you can see all the individual bands (Band_1, Band_2,…,Band_5). Drag all bands **except** Band_5 (since we know it has bad separability) into the ‘Input raster bands’, under ‘Input raster or feature sample data’ choose the “RE2016_07_20_training” feature class, under ‘Sample field’ choose Classname. Save the file to the folder where your data are and call it “RE2016_07_20.gsg”. Click Run.



10. You now have a signature file that you can use in image classification.

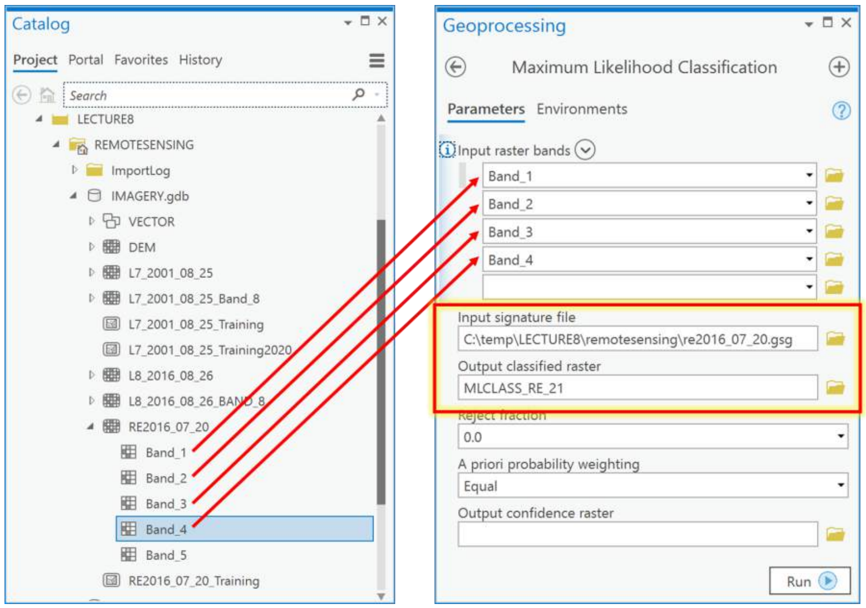
Q5: How do I classify my image using supervised image classification?

To classify an image, you need to have completed the above example questions because you require the signature file “RE2016_07_20.gsg” for the ‘RE2016_07_20’ RapidEye image in order to undertake supervised classification.

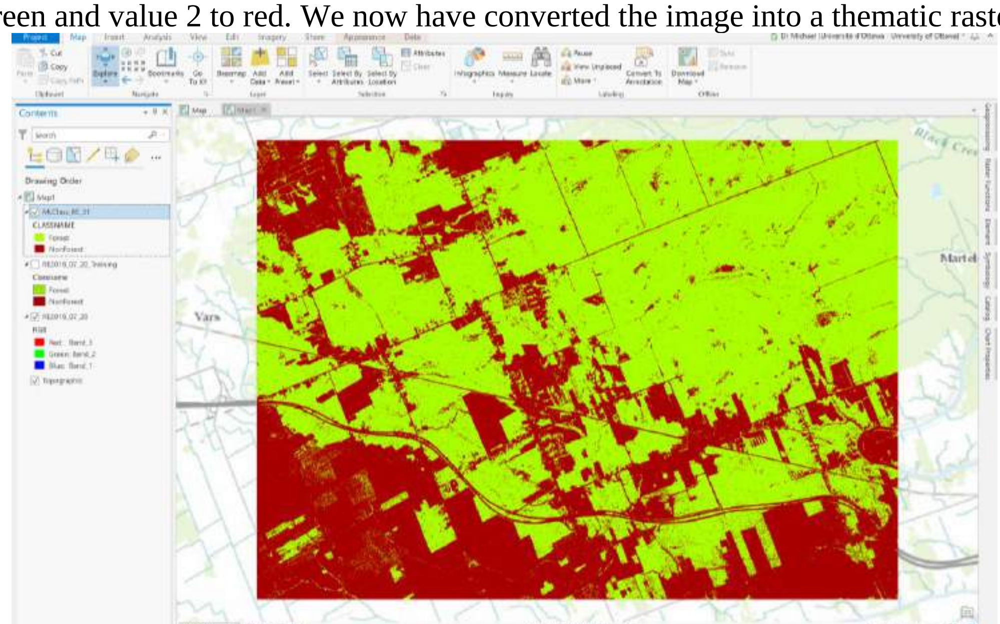
Now that you have a signature file from the previous example question, you can go ahead and undertake a supervised classification:

Student Name: ______________________________ Student Number: ____________________________

1. Use the signature file to classify the RapidEye image into the two information classes of **Forest** and **NonForest**. To do so, go to Spatial Analyst -> Multivariate and choose Maximum Likelihood Classification. Undock the Maximum Likelihood Classification tool so that you can see it with your Catalog pane at the same time.
2. In the Catalog pane, navigate to IMAGERY.GDB and expand the RE2016_07_20 image so that you can see all the individual bands (Band_1, Band_2,…,Band_5). Drag all bands **except** Band_5 into the 'Input raster bands', under 'Input signature file', choose the signature file created in Q4 called "RE2016_07_20.gsg", under 'Output classified raster' call it "MLCLASS_RE_21", and save it in IMAGERY.GDB. Leave all other values as defaults and click Run:



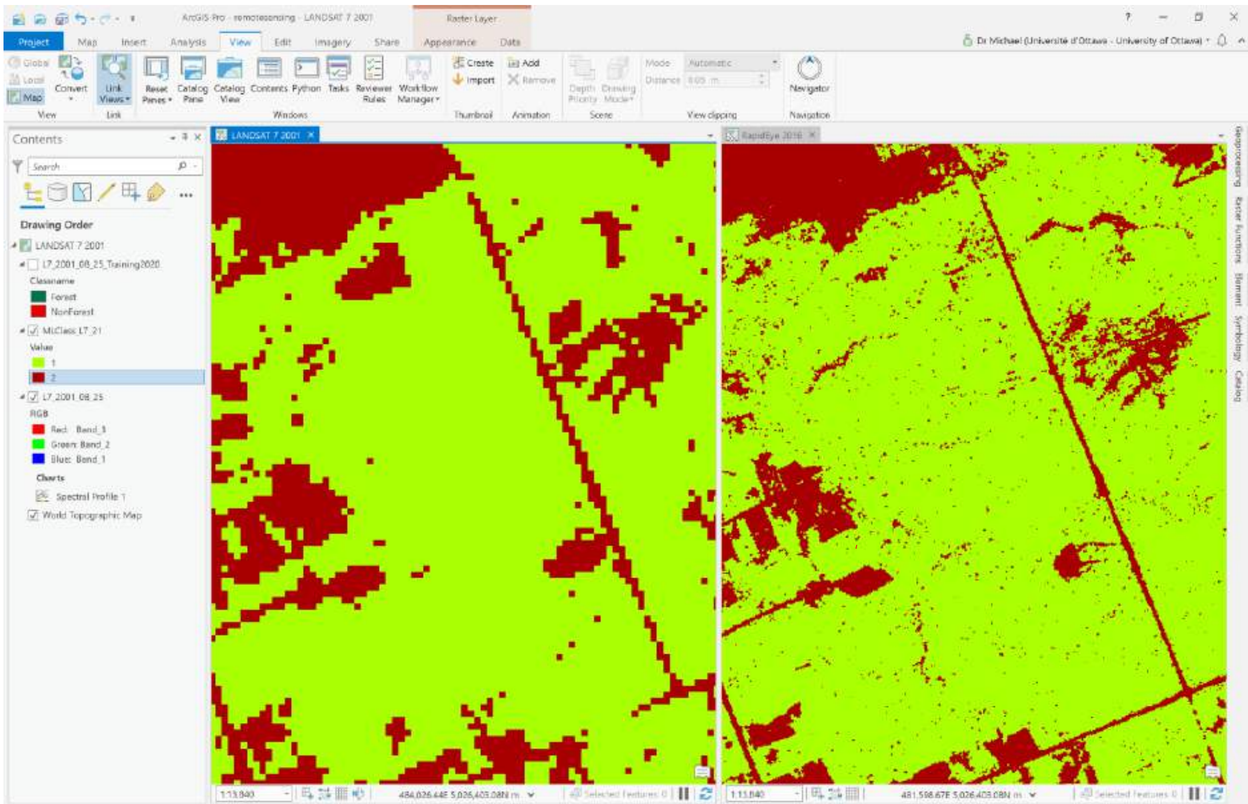
3. Examine the resulting classified raster (note that you may want to change value 1 to green and value 2 to red. We now have converted the image into a thematic raster layer.



Student Name: ______________________________ Student Number: ____________________________

Q6: How can I clean up my classified raster layer?

In the classified result, you will notice many artifacts of the classification process; these are particularly evident as **Forest** pixels surrounded by **NonForest** pixels or vice-versa. As a general rule, the higher the image resolution (the smaller the pixels), the more misclassified pixels there are likely to be. Given that the Landsat 7 layer was at a 30 m resolution, and the RapidEye was at 5 m resolution, it is not surprising that there are more misclassified pixels in the map made from the RapidEye image. For example, looking at the Landsat and RapidEye classification results side-by-side (as shown below), we can see that the **Forest** from Landsat 7 is more solid green with fewer misclassified pixels than the RapidEye in the same location.

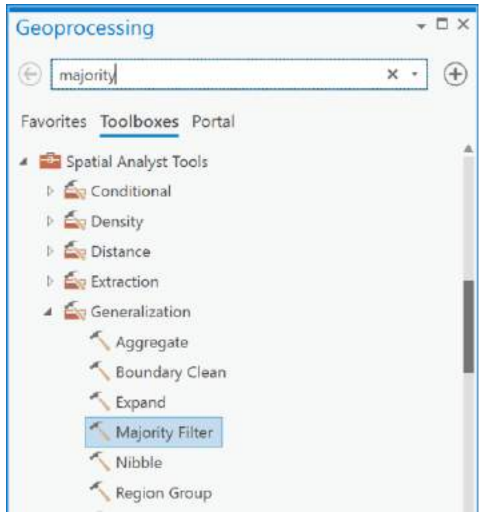


Single **NonForest** pixels surrounded by **Forest** are unlikely to represent forest loss at that pixel in the majority of cases. It partly comes down to the definition of what a "Forest" actually is?!? While this is not a formal definition, many would consider a forest an areas that is dominated by trees – but that doesn't mean there has to be trees everywhere! Forests can have clearings without trees, but those clearings are still part of the forest. With the RapidEye imagery, such learings, or areas that only contain a small tree, or an unhealthy tree, may have been misclassified as **NonForest** (see examples in the figure below).

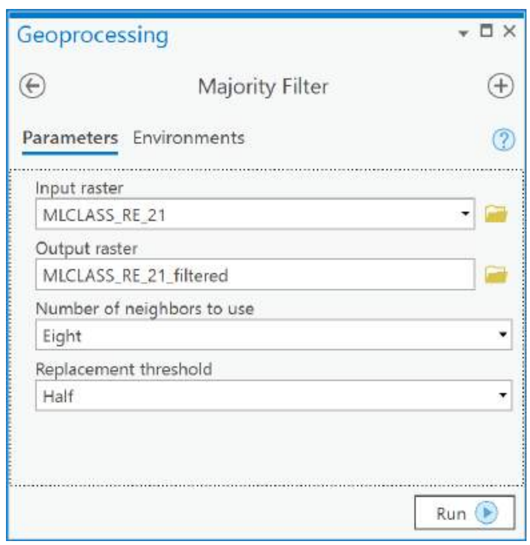
We can filter out some of the individual erroneous pixels using a majority filter:

1. Open the Majority Filter tool under Spatial Analyst Tools -> Generalization:

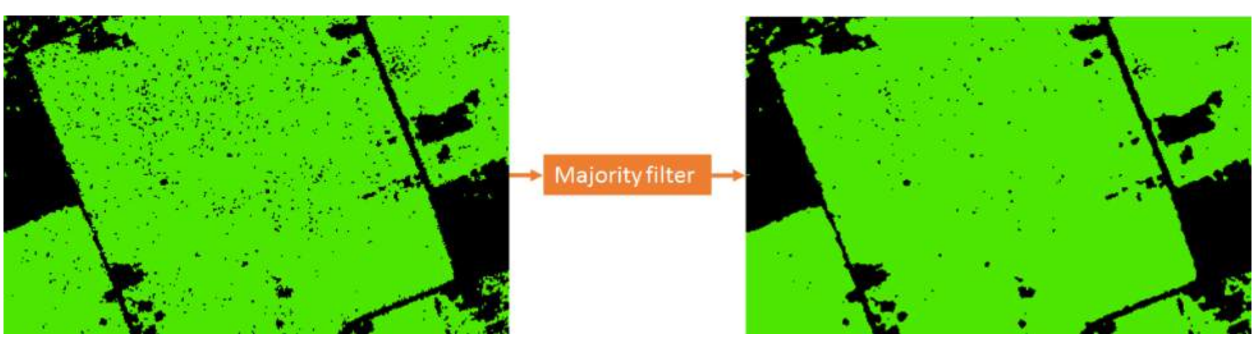
Student Name: ______________________ Student Number: ______________________



2. In the Input raster drop-down, select your "MLCLASS_RE_21" classified raster layer. Call the output "MLCLASS_RE_21_filtered" and save it to IMAGERY.gdb. Choose the Number of neighbors to use (optional) as eight neighbors, and Run it.



3. The result is a cleaner classified layer with less misclassified/abhorrent pixels:

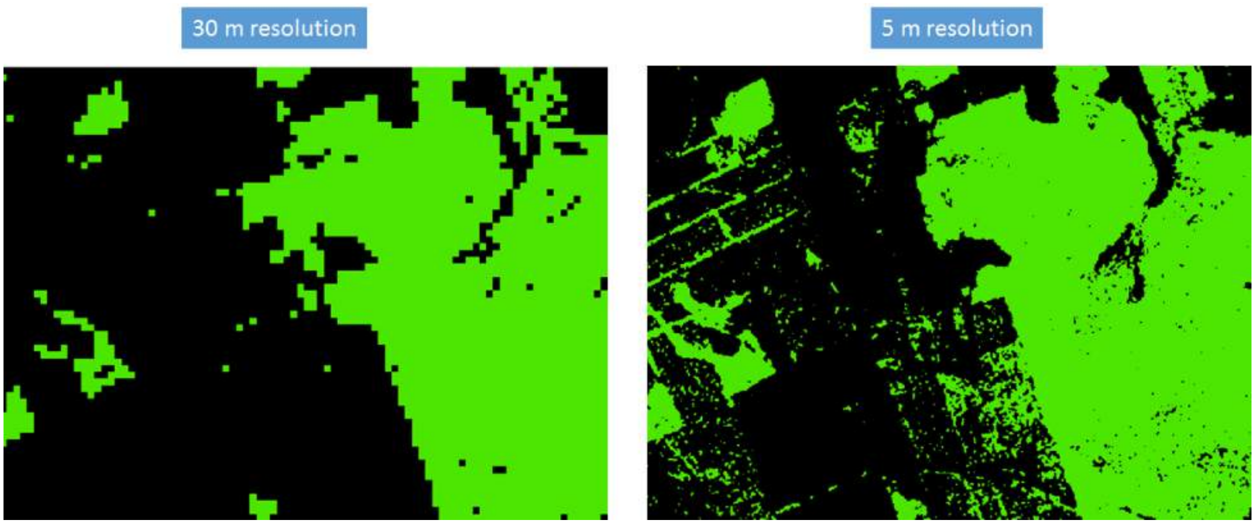


Sometimes you may run the majority filter (or similar tools) a few times to get the result at the level of generalization that is required for a particular purpose.

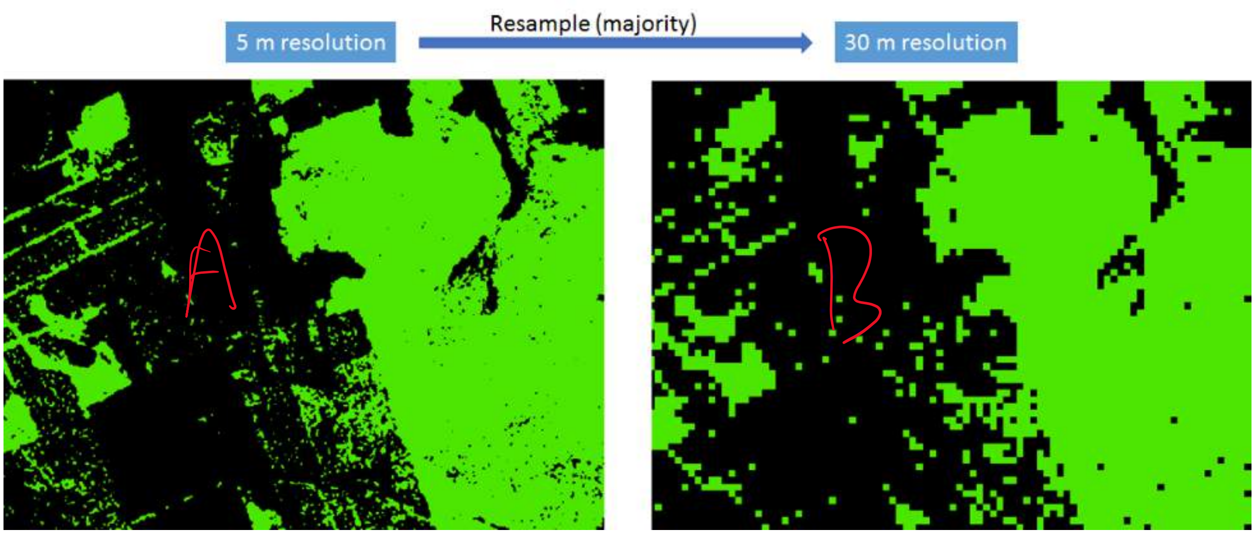
Student Name: ______________________ Student Number: ______________________

Q7: How do I generalize my "MLCLASS_RE_21_filtered" classified raster so that the cell size is the same as the layer "MLCLASS_L7_21" from Q2?

The Landsat 7 scene that was classified in Q2 is in the layer called "MLCLASS_L7_21" and has a spatial resolution of 30 meters (i.e., each cell is 30 x 30 meters). Your "MLCLASS_RE_21_filtered" layer has the same resolution as the RapidEye image which is 5 m (see comparison below).

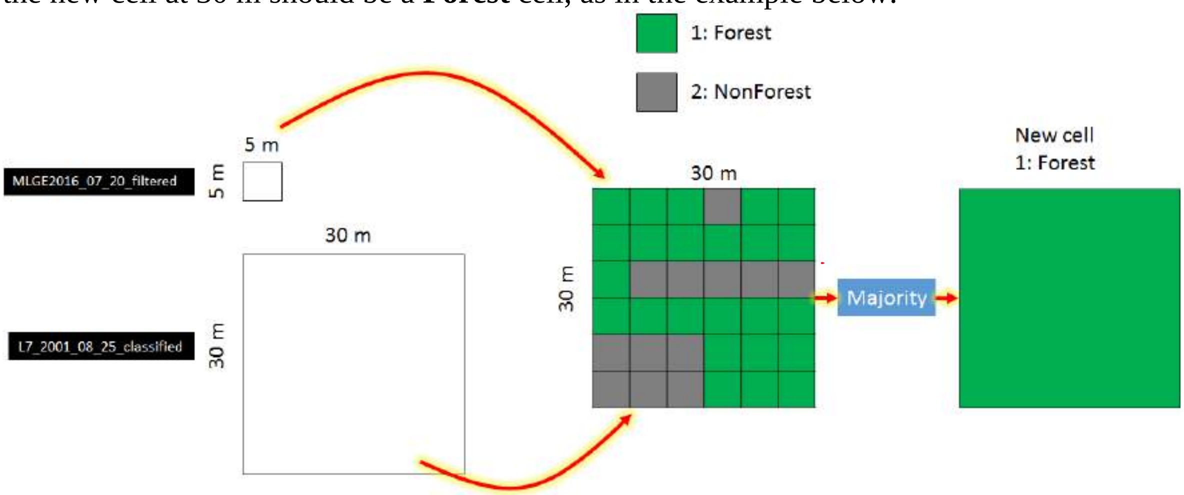


Because there are 36 cells in "MLCLASS_RE_21_filtered" layer for every cell in "MLCLASS_L7_21", you will be increasing the cell size for "MLCLASS_RE_21_filtered" to match the cell size of "MLCLASS_L7_21".



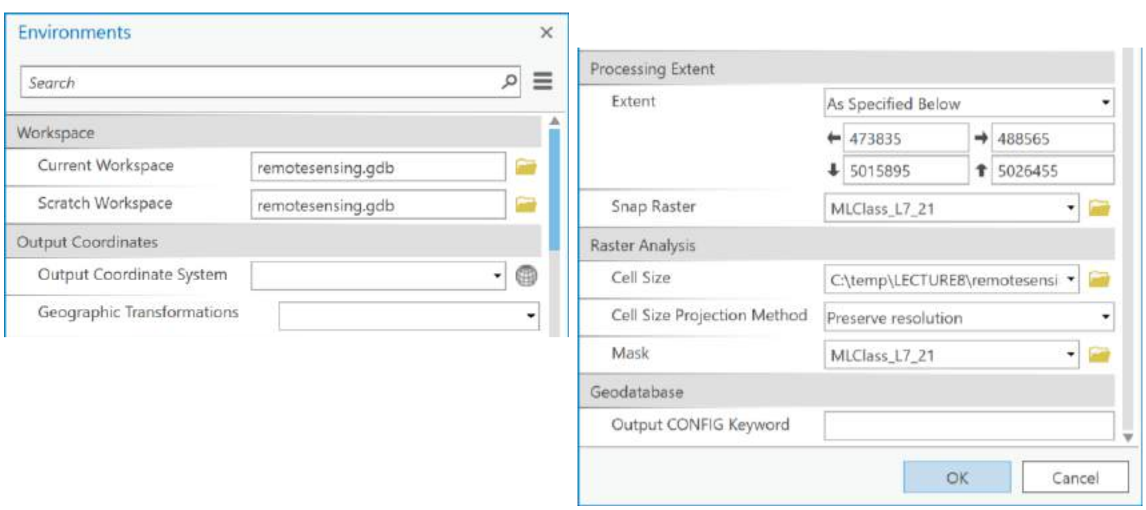
Student Name: ___________________________ Student Number: __________________________

So, you must tell ArcGIS how to summarize the "MLCLASS_RE_21_filtered" cells when increasing the cell size of that layer. In other words, within each new 30 x 30 m cell, ArcGIS must look at the 36 smaller cells that it is made of and decide how to calculate the value (**Forest** or **NonForest**) within the new cell size. Because we're dealing with nominal data (**Forest** and **NonForest**), ArcGIS will by default simply take the 5 x 5 m cell value closest to the cell center of the new 30 x 30 m cell. However, with nominal data, a more appropriate aggregation function would be for the new 30 x 30 m cell to take on a value of the majority of values within it. For example, if there are 24 **Forest** cells and 12 **NonForest** cells and within the new 30 m cell, then the new cell at 30 m should be a **Forest** cell, as in the example below.



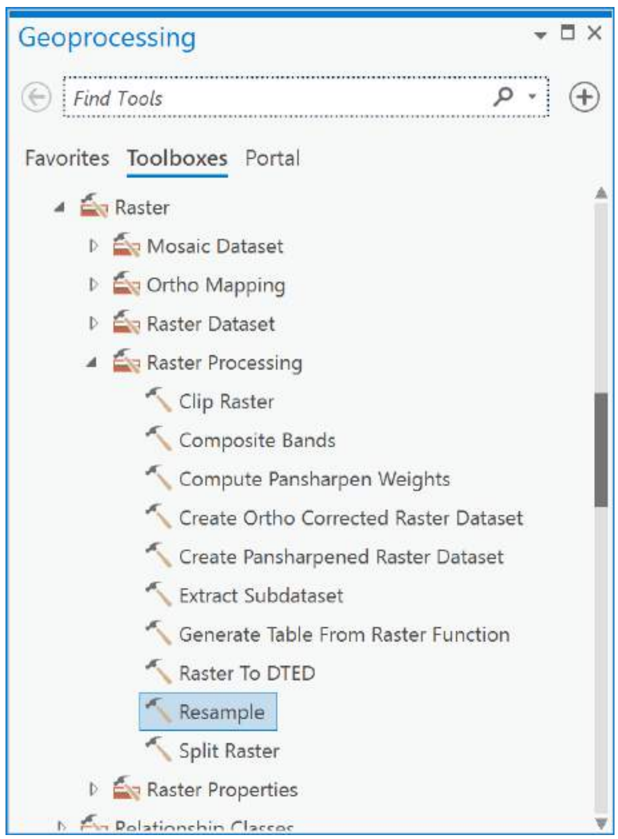
When you generalize your raster layer to the same cell size as the Landsat 7 classification result, you will also want to ensure that the cells in the generalized layer line up correctly with the cells in the Landsat 7 layer. This is done by setting the "Snap Raster" in the global environment settings as well as the Processing Extent:

1. Click on the Analysis tab and then click on the Environments button.
2. In the Environment Settings dialog, under Processing Extent, under Extent choose "MLClass_L7_21" and under Snap Raster, also choose "MLClass_L7_21". Then under Raster Analysis, Cell Size choose "MLClass_L7_21" and under Mask also choose "MLClass_L7_21" and then click OK:



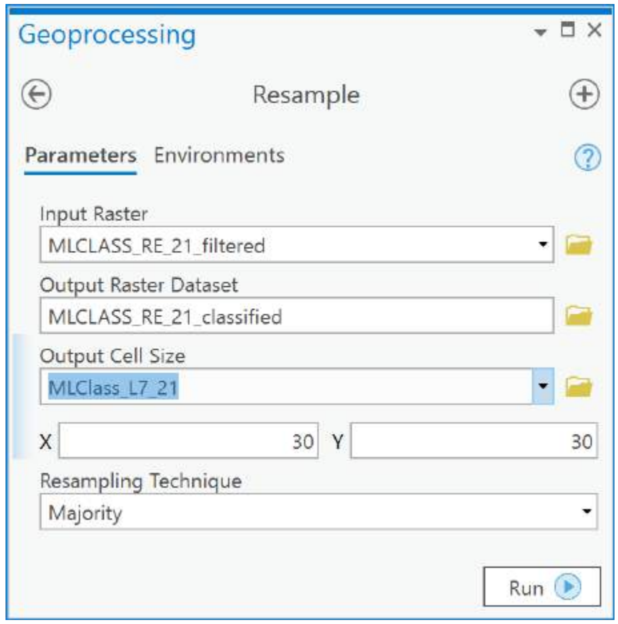
By completing Step 2, you have ensured that any processing output in raster analysis will conform to the cell size and extent of the "MLCLASS_L7_21" layer. Keep in mind that these settings will be maintained until you close and reopen ArcGIS or change them again. So, if you decide to do other analyses, with other layers, that have a different extent than "MLCLASS_L7_21", results will only be produced within the specified extent. Thus, you may need to clear the extent property or set it back to its default value when you are done with the raster analyses.

3. To actually convert the RapidEye product to 30-m resolution, you use the Resample tool under Data Management -> Raster -> Raster Processing. Open that tool:

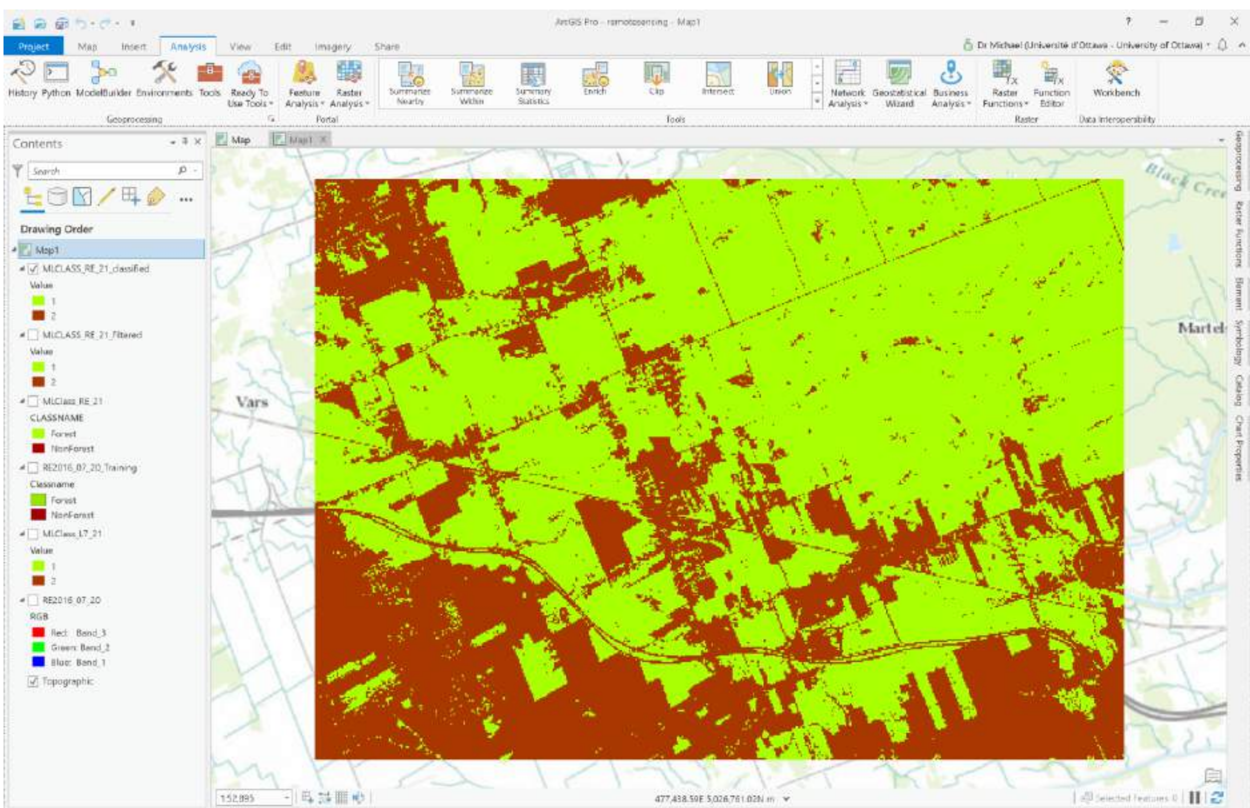


4. In the Resample dialog, your input raster should be specified as "MLCLASS_RE_21_filtered", the Output Raster should be called "MLCLASS_RE_21_classified" and the Output Cell Size (optional) should be specified "MLClass_L7_21". Most importantly, the Resampling Technique (optional) should be set to MAJORITY. Then click OK:

Student Name: ______________________________ Student Number: ____________________________



5. Your aggregated result will look like this:

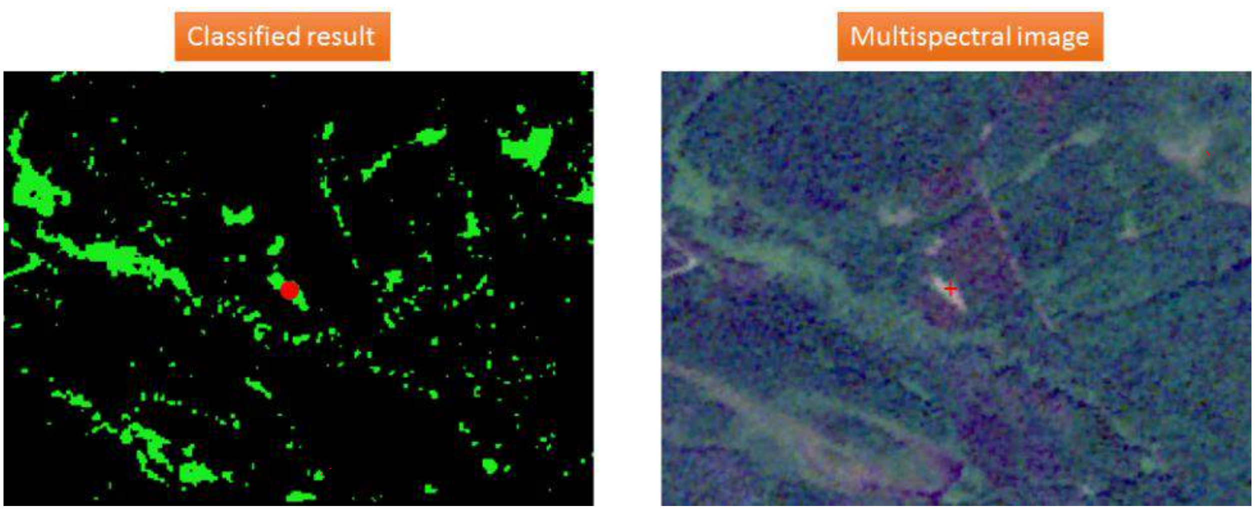


Student Name: ____________________________ Student Number: ___________________________

Q8: How do I test the accuracy of a classified result?

The accuracy of a classification is tested by comparing classified pixels to what are called 'ground-truth' data. The ground-truth (GT) data are locations that have not been used in the classification but are known to belong to one of the information classes that you have defined, either belonging to **Forest** or **NonForest**. In other words, the GT data is a "test dataset" and is independent of the data used to create the signature file that trained the classifier. GT data can be acquired by either going into the field with a GPS unit and observing multiple locations for the presence of any particular information class. However, unless absolutely necessary, other methods of collecting independent GT data are typically used because field-based GT data is often difficult, slow, and expensive to acquire. Thus, you can usually collect examples of GT data by directly observing features within the pre-classified image and labelling those as one of the information classes, in our case as either **Forest** or **NonForest**. To assess the accuracy of a classification you would collect such data at several randomly selected points and compute a confusion matrix that would allow you do quantify how often the classification is correct, and how often it isn't.

Note: The term 'ground-truth' is frowned upon by some, because it implies that these data are inherently "true". This obviously isn't correct, because they are made by humans, and humans make mistakes. Nonetheless we use the term here because it is common, and you will encounter it in the future if you work more with satellite imagery. Alternative terms used for the same thing is "validation data" or "test data", but those terms themselves can mean different things to different people, so they are not ideal either. We'll stick with 'ground-truth' for this exercise!



Example of a misclassification. Pixel classified as Forest (left) is clearly not a forest (right) as can be seen in a ground truth point (red cross, right) identified manually from a multispectral image.

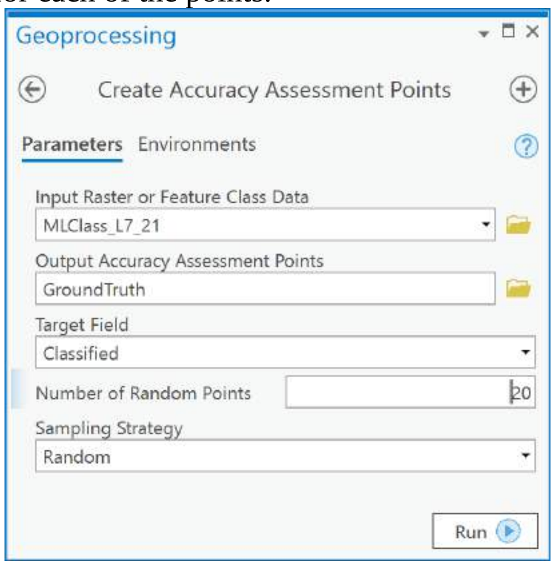
You will calculate the accuracy of the "MLCLASS_L7_21" result, since everyone will have the same result from Q2 above.

1. Go to Spatial Analyst -> Segmentation and Classification and choose the tool Create Accuracy Assessment Points.

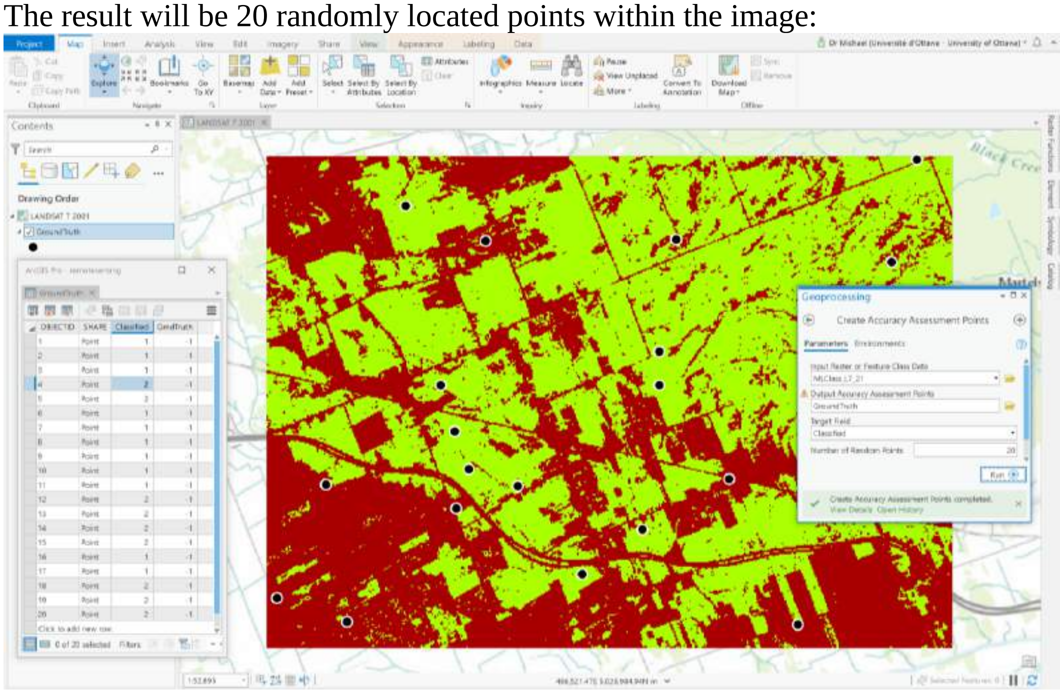
Student Name: ___________________________   Student Number: ___________________________

2. In the tool, under Input raster or feature class data, choose the "MLCLASS_L7_21" layer and under Output Accuracy Assessment Points, call the output "GroundTruth" and ensure that you save it in the IMAGERY.gdb geodatabase. Under Target Field choose Classified. Under Number of Random Points choose 20, and finally under Sampling Strategy choose Random.

   This will create 20 points randomly (where the x and y coordinates for each point are chosen, within the extent of the image, from the uniform probability density function independently and with equal probability) within each of the classes **Forest** and **NonForest**. For each random point, the pixel value from the "MLCLASS_L7_21" layer will be extracted at the point location within the "MLCLASS_L7_21" layer. These values will be stored in the "GroundTruth" table in a column called "Classified" containing the classified pixel values for each of the points.

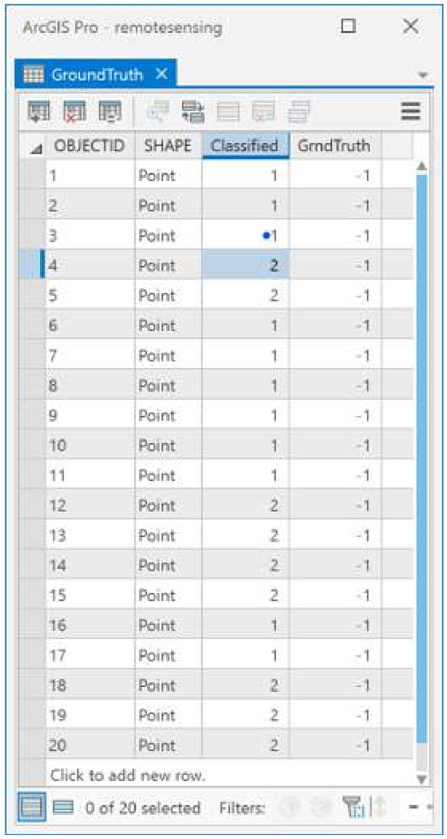


3. The result will be 20 randomly located points within the image:



Student Name: ______________________________ Student Number: ______________________________

4. Now open the table called “GroundTruth”. It should look something like this:

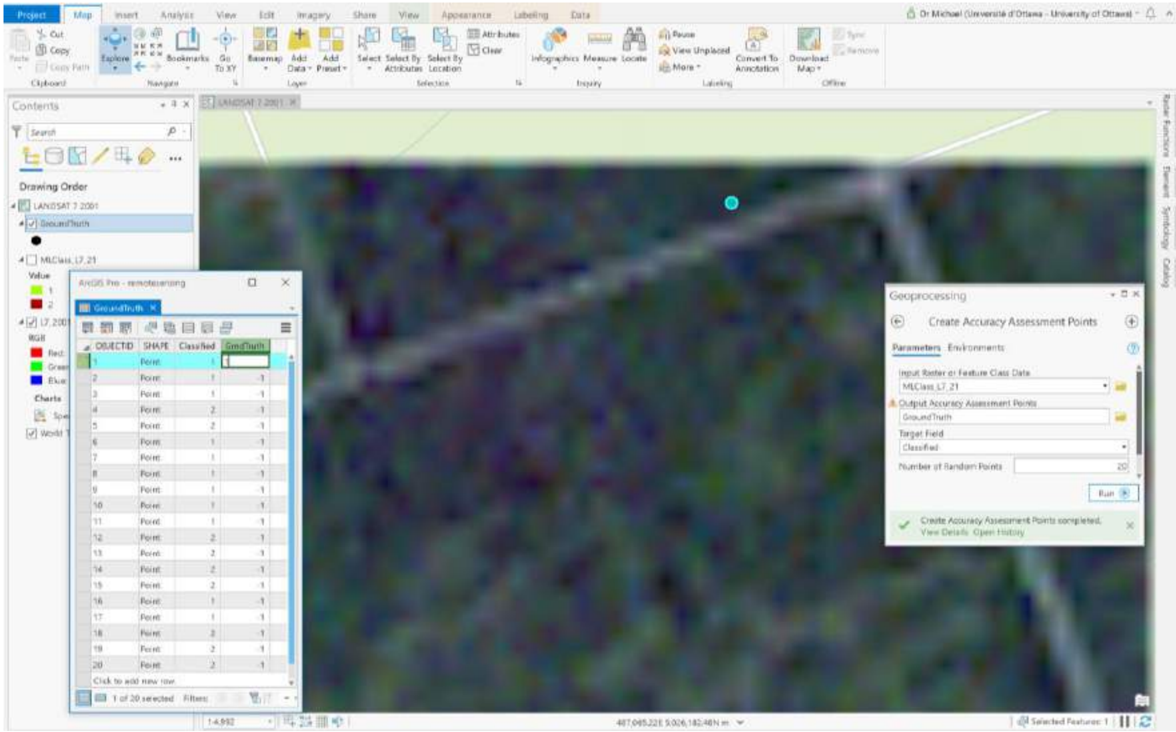


OBJECTID	SHAPE	Classified	GrndTruth
1	Point	1	-1
2	Point	1	-1
3	Point	1	-1
4	Point	2	-1
5	Point	2	-1
6	Point	1	-1
7	Point	1	-1
8	Point	1	-1
9	Point	1	-1
10	Point	1	-1
11	Point	1	-1
12	Point	2	-1
13	Point	2	-1
14	Point	2	-1
15	Point	2	-1
16	Point	1	-1
17	Point	1	-1
18	Point	2	-1
19	Point	2	-1
20	Point	2	-1

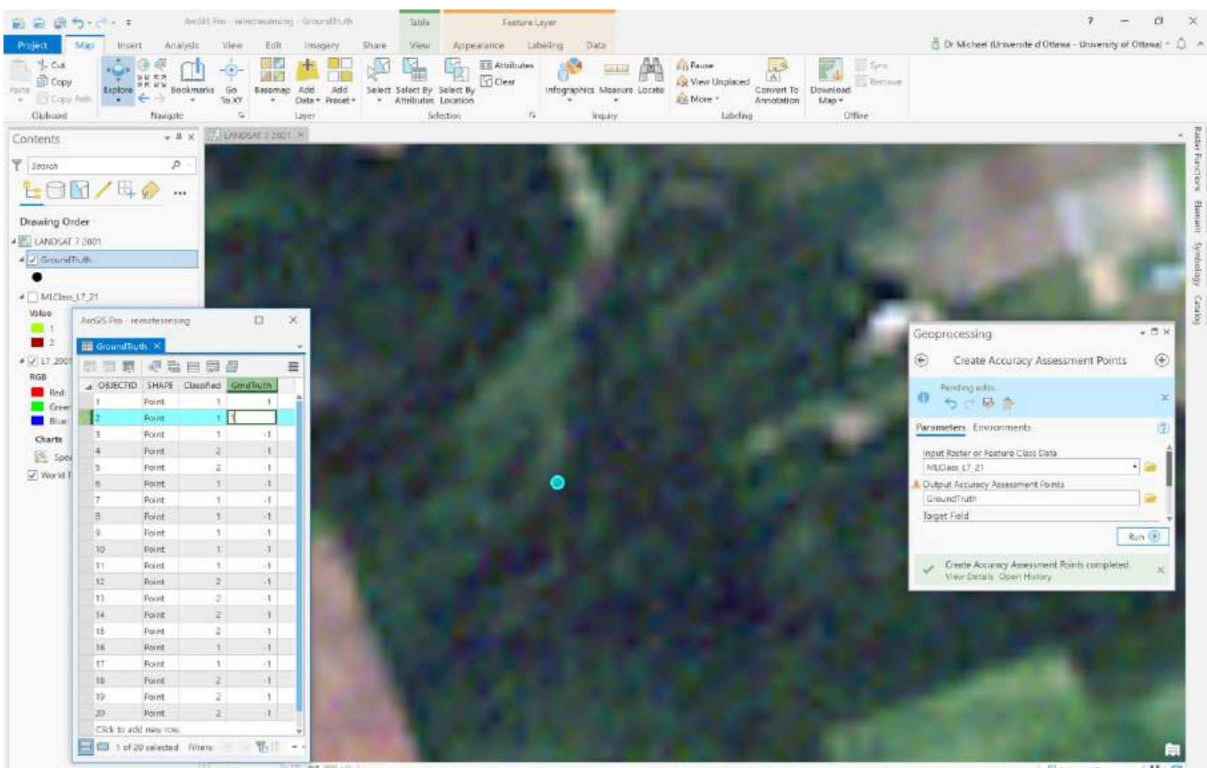
5. Each random point has assigned to it the information class value for either **Forest** = 1 or **NonForest** = 2 depending on which class contained the point. This data will be in the “Classified” column of the “GroundTruth” attribute table (see the above table).

6. Now we need to fill in the GrndTruth column of the table which has values of -1 right now. The -1 values tell us that the ‘true’ class for that point has not been observed and entered into the table. So we need to fill these ‘true’ class values, remembering that the value 1 means **Forest** and 2 means **NonForest**. To do so, ensure the only layers visible on the map are “L7_2001_08_25” and “GroundTruth”. Now select the first row in the “GroundTruth” table, to identify the first point. For the currently selected row, right-click and choose Zoom To. Then zoom a little closer until you can clearly see what information class the currently selected point is in. Now, enter either a 1 or 2 in the GrndTruth column, depending on your interpretation of the image at that point. In the below example, you can see that **Forest** is under the selected point, so you would enter a value of 1 in the GrndTruth column for that row by double-clicking the cell where you want the 1.

Student Name: ____________________________ Student Number: ___________________________

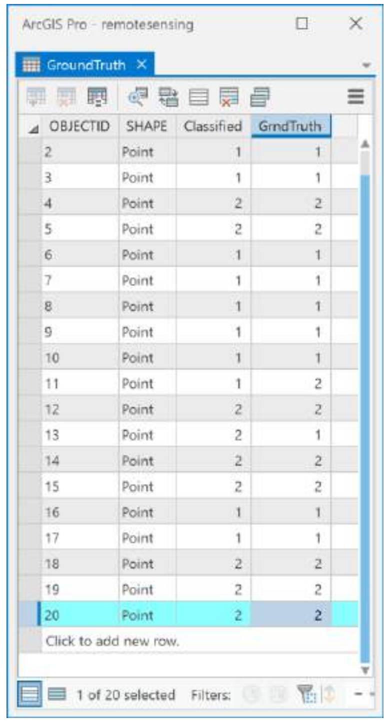


7. Now select the second row, Zoom To, and examine whether or not there is **Forest** or **NonForest** at that location and enter either a value of 1 or 2, respectively in the GrndTruth column of the "GroundTruth" table. In this example, again, it is **Forest** so you would enter a value of 1:



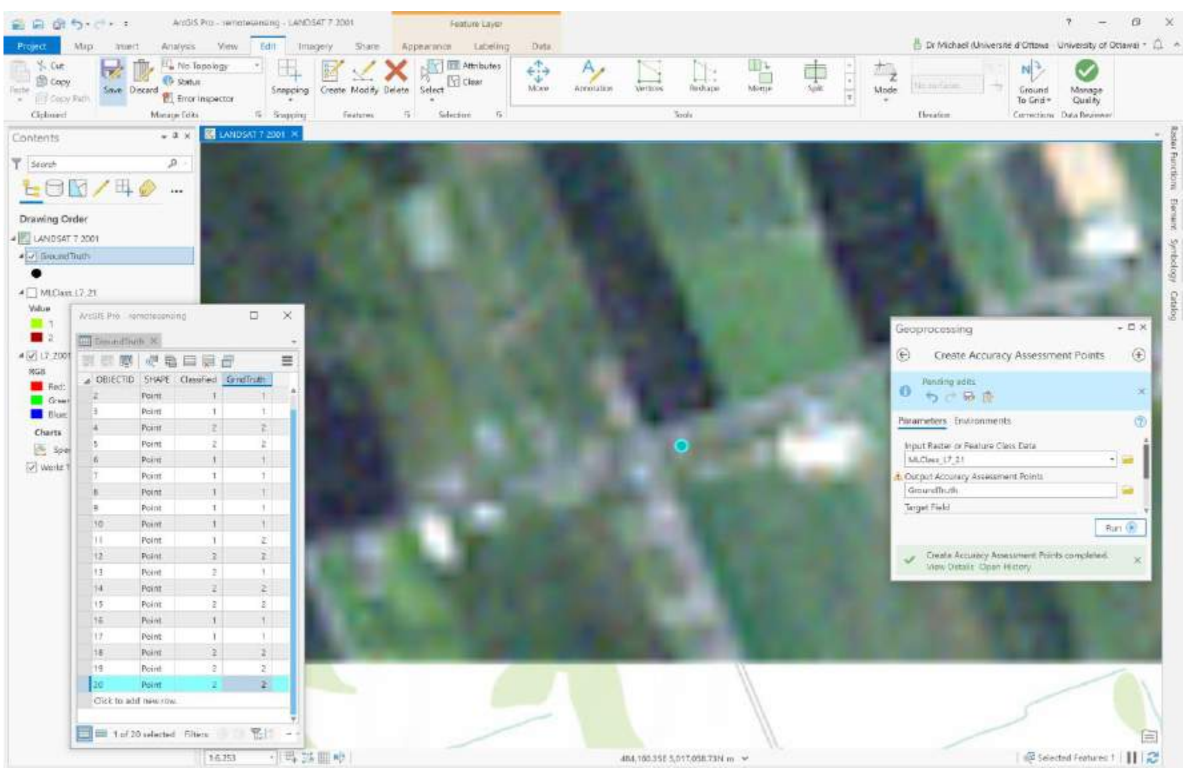
Student Name: ____________________________ Student Number: ___________________________

8. Repeat these steps for all the points in the "GroundTruth" table, to create a table that is complete and ready for accuracy assessment, e.g.,:



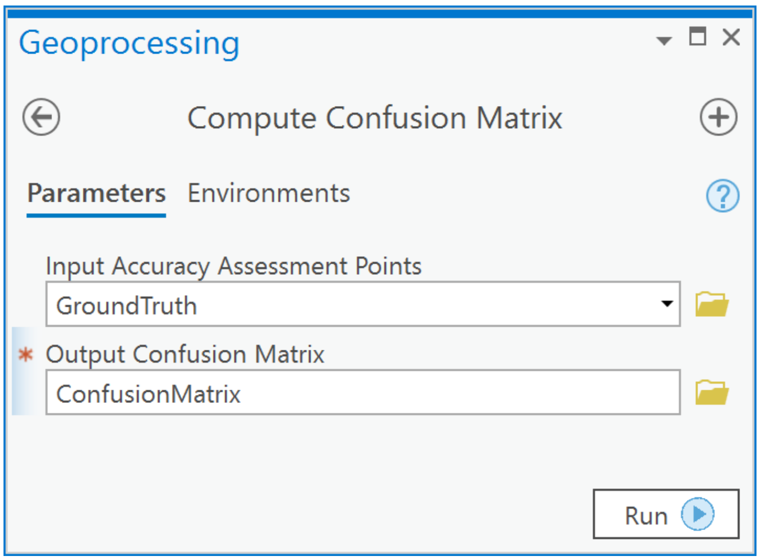
OBJECTID	SHAPE	Classified	GrndTruth
2	Point	1	1
3	Point	1	1
4	Point	2	2
5	Point	2	2
6	Point	1	1
7	Point	1	1
8	Point	1	1
9	Point	1	1
10	Point	1	1
11	Point	1	2
12	Point	2	2
13	Point	2	1
14	Point	2	2
15	Point	2	2
16	Point	1	1
17	Point	1	1
18	Point	2	2
19	Point	2	2
20	Point	2	2
Click to add new row.			

9. Now, go to the Edit tab and click on Save to save the edits to the "GroundTruth" table. Then click on the Clear button in the Selection group on the same tab:

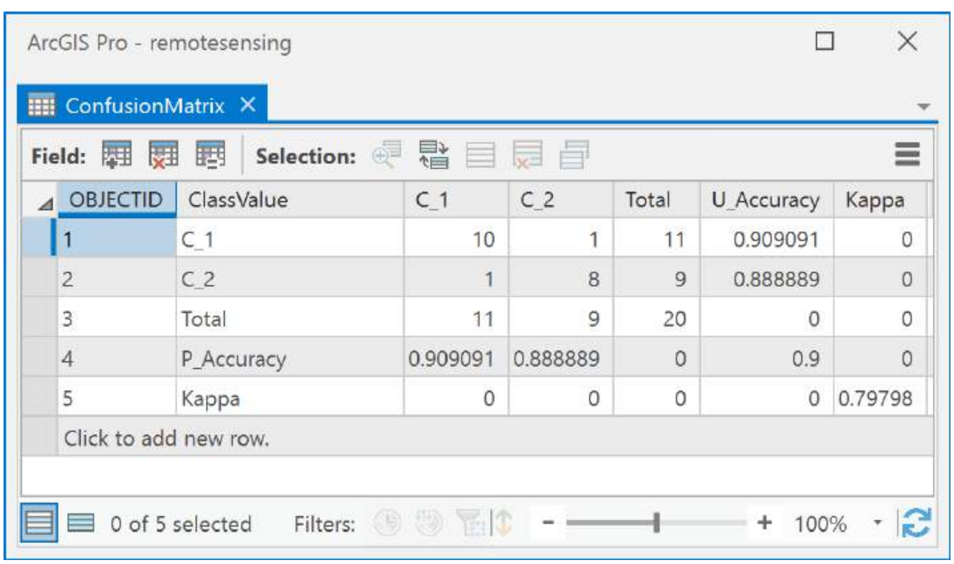


Student Name: ______________________ Student Number: ______________________

10. Your “GroundTruth” table now contains a column called Classified that contains the output of the classifer at each point and a column called GrndTruth that contains the observed information class at each point. Now you want to tabulate how many times **Forest** (value of 1) corresponds to **NonForest** (value 2), how many times **Forest** (value of 1) corresponds to **Forest** (value of 1), how many times **NonForest** (value 2) corresponds to **NonForest** (value 2) and finally how many times **NonForest** (value 2) corresponds to **Forest** (value of 1). Because we’re only dealing with two classes in this example you could easily do this manually, but ArcGIS has a tool that does it for us, called the Compute Confusion Matrix which is under Spatial Analyst -> Segementation and Classification in ArcToolbox. Open that tool and under Input Accuracy Assessment Points choose your “GroundTruth” table. Call the output ConfusionMatrix and save it to your IMAGERY.gdb:



11. Open the ConfusionMatrix table:



OBJECTID	ClassValue	C_1	C_2	Total	U_Accuracy	Kappa
1	C_1	10	1	11	0.909091	0
2	C_2	1	8	9	0.888889	0
3	Total	11	9	20	0	0
4	P_Accuracy	0.909091	0.888889	0	0.9	0
5	Kappa	0	0	0	0	0.79798

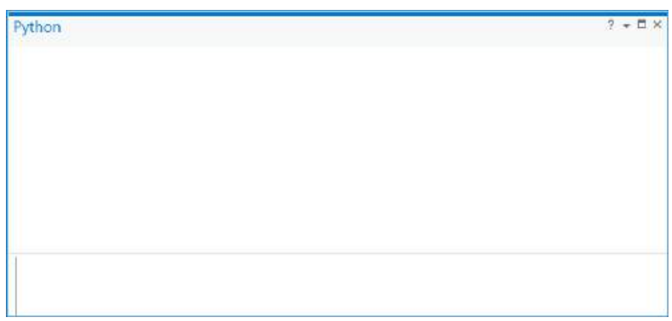
In the table, C_1 means class value 1 (**Forest)**, and C_2 means class value 2 (**NonForest)**. The numbers in the cells where a column and row of the same name intersect indicate the number of times the two corresponding classes were correctly classified. For example, in the table illustrated above, there were 10 times where a C_1 corresponded to a C_1 (where C_1 column intersects with C_1 row) and 1 time a C_1 corresponded with a C_2 classified result. Thus, 10 **Forest** points were correctly classified and 1 was incorrectly classified. Similarly, there were 8 times where a C_2 was correctly classified as a C_2, and one time where it was classified as a C_1. Thus, 8 out of 9 times **NonForest** was correctly classified. The Kappa value is a measure, ranging from -1 to 1, of how much better than random the classifier performed. Here it is 0.797, which is considered a modestly good result. You can read the help file for the Compute Confusion Matrix tool for more information.

12. Now repeat all the steps for this question with your "MLCLASS_RE_21_classified" layer using the RE2016_07_20 image as the ground truth layer.

Q9: How can I create a raster with only Forest from each of my classified results, and determine where forest has been lost?

Since you now have two rasters, both containing information classes called **Forest** and **NonForest**, you can use a simple map algebra expression to create a Boolean raster layer where the value of 1 means **Forest** and 0 means **NonForest**. You can do this in several ways, here we'll try to use the Python programming language that's integrated as part of ArcGIS:

1. Open the Python pane:



2. Type the following three lines and hit the Enter key (note that you don't need to enter the lines in green that are preceded by the # sign, since these are only comments to tell you what each line does):

```
# access Spatial Analyst functions
from arcpy.sa import *

# Create a raster variable from the filtered classification results
x = Raster("MLCLASS_RE_21_classified")

# Create a boolean layer using the relational operator for logical equals
forest_2016 = x == 1
```

The first line tells ArcGIS to allow you to access Spatial Analyst functions within the python pane. The second line uses the Raster() function to create a raster object that can be used in a Python map algebra expression. The third line is a relational map algebra expression using logical equals (==). That expression looks at each cell in the raster layer called x, and if the cell has a value of 1 (**Forest**), then the new raster cell's output will also have a value of 1. If the cell's value is anything other than 1 then the output raster will have a value of zero for the cell in question.

3. The result will be a raster layer called "forest_2016" in the table of contents but with pixel values of 1 for cells that are forest and 0 otherwise.

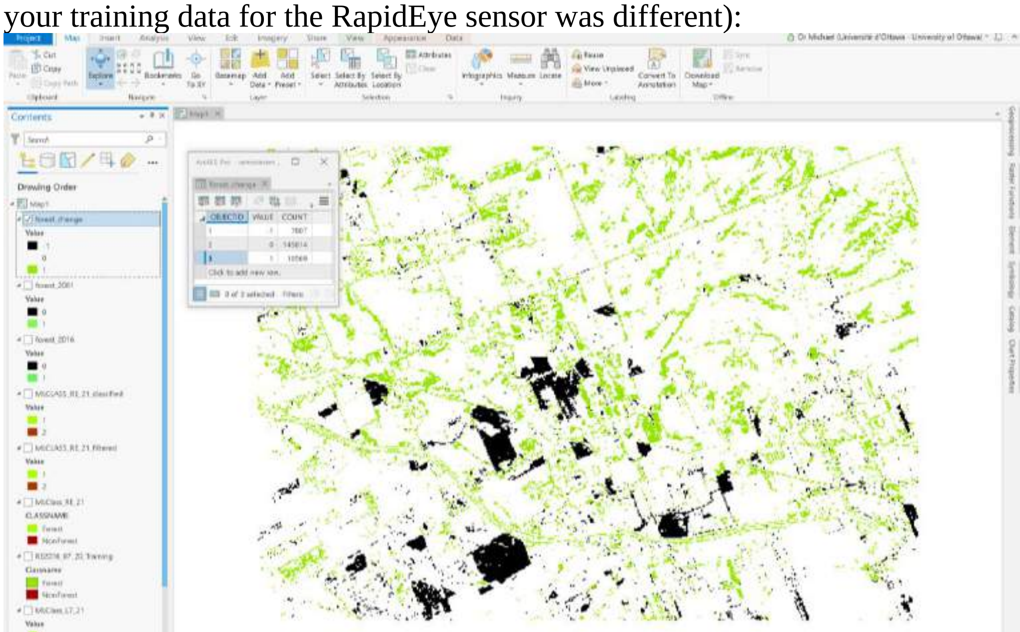
4. Now modify these lines, to create a binary raster for the classified Landsat image:

```
# access Spatial Analyst functions
from arcpy.sa import *

# Create a raster variable from the filtered classification results
y = Raster("MLClass_L7_21")

# Create a boolean layer using the relational operator for logical equals
forest_2001 = y == 1
```

5. Now, to examine the amount of forest that has been lost vs. gained we can subtract the forest_2001 from the forest_2016:

```
# Subtract 2016 from 2001 to see where forest has been
# gained and/or lost.
forest_change = forest_2016 – forest_2001
```

6. Your resulting map will look something like this (it won't be exactly the same because your training data for the RapidEye sensor was different):



Here the attribute table tells us in the Count field how many total pixels are for each value.

Student Name: ______________________________ Student Number: ____________________________

Answer the following questions.

1. [1 point] What is the positional uncertainty due to cell resolution within the "MLCLASS_L7_21" layer?

2. [3 points] In Q2 you created a thematic raster layer called "MLCLASS_L7_21", and in Q3-Q7 you created a layer called "MLCLASS_RE_21_classified". In Q8 you then created confusion matrices for both layers. Present a map for each result and confusion matrix. Explain which result is more accurate, how you know, and what you think the reason is.

3. [3 points] In the raster you created at the end of Q9, you had cell values of -1, 0, and 1. Describe what these values represent in the landscape, e.g., what does a -1 cell value mean, what does a 0 cell value mean and what does a +1 cell value mean? Present the map with your explanation. What is your overall conclusion about forest loss in the region?

4. [3 points] In the differenced layer you created at the end of Q9, you may be interested in the large contiguous *regions* of cells with values that indicate forest loss (conversion from forest to non-forest). Where are these large contiguous regions? Use the Region Group tool under Spatial Analyst -> Generalization with your result from Q9 as the input layer. Read the help file for the Region Group tool by choosing the Tool Help button. Explain what the region group tool does. Now, open the attribute table of the region group result raster, right-click on the "Count" column and sort in descending order. Examine the "Link" field. How do the values in the "Link" field relate to the cell values in the result from Q9? Provide a table of area in map units for the 10 largest contiguous groups of cells that have changed from **Forest** to **NonForest** from 2001 to 2016. Also present a map showing these regions.

5. [2 points] What is the average region size (patch size) in square meters of regions that have changed from **Forest** to **NonForest** from 2001 to 2016? You will need to undertake an attribute query on the attribute table from the region grouped layer from (4) and right click on the "Count" column and choose Statistics to get the mean, in order to answer this question.

6. [2 points] Select only those regions that changed from **Forest** to **NonForest** between 2001 to 2016 and are greater than or equal to 3000 square meters in size. What is the average slope for each region? You will need to use Spatial Analyst -> Surface -> Slope with the DEM data found in IMAGERY.gdb, as well as an attribute query on the region grouped layer from (4) and Spatial Analyst -> Zonal Statistics as Table, to address this question. Present a table showing the average slope for each selected region.

Student Name: ______________________________ Student Number: ______________________________

7. [1 point] Examine the layer called L8_2016_08_26, which is a Landsat 8 OLI image. Use your identify tool to click on a few different locations and note the DN each time. Compare those values with the DN values for similar features in the "L7_2001_08_25" layer. Why are the DNs in the Landsat 8 image greater than those in the Landsat 7 image for similar features?

8. [2 points] Create and present an NDVI map using the L8_2016_08_26 image. Consult your lecture notes, or the Internet, to determine the appropriate bands to use for calculating NDVI with Landsat 8 data, as they are not the same as for Landsat 7. What does this map tell you? Why?