

Laboratoire n°5 : Télédétection et SIG raster

Objectif : L'objectif de ce laboratoire est de vous présenter un processus de classification de base par télédétection et d'évaluer la détection des changements dans les régions forestières à l'aide d'un SIG de type raster.

Résultats de l'apprentissage : À la fin de ce laboratoire, vous serez capable de :

- ❖ Effectuer une amélioration spectrale en utilisant des combinaisons de bandes et de ratios.
- ❖ Entraîner un classificateur d'images.
- ❖ Expliquer un fichier de signatures.
- ❖ Réaliser une classification supervisée en utilisant des images télédétectées à haute et moyenne résolution.
- ❖ Nettoyer et généraliser une couche raster thématique.
- ❖ Évaluer la précision d'un résultat classifié.

Procédure:

1. Lisez et travaillez sur les exemples.
2. Répondez aux questions à la fin du laboratoire.

Données/matériels:

- ❖ Toutes les données sont fournies dans **lab5.zip**.

Travail:

- ❖ Répondez aux questions (#1-8) dans les espaces fournis.
- ❖ Le total des points disponibles pour ce travail est de 17 points, soit 15% de la note totale du cours.
- ❖ Tous les travaux pratiques doivent être réalisés et remis individuellement.

Format:

- ❖ Les étudiants remettront les questions avec les réponses complètes.

Les tableaux et les figures doivent être produits selon les normes du Département de géographie, environnement et géomatique de l'Université d'Ottawa.

Pour des directives sur le formatage et la présentation appropriés des tableaux et des figures, veuillez consulter la **section d'introduction de ce manuel de laboratoire.**

Nom de l'étudiant: _____ # d'étudiant: _____

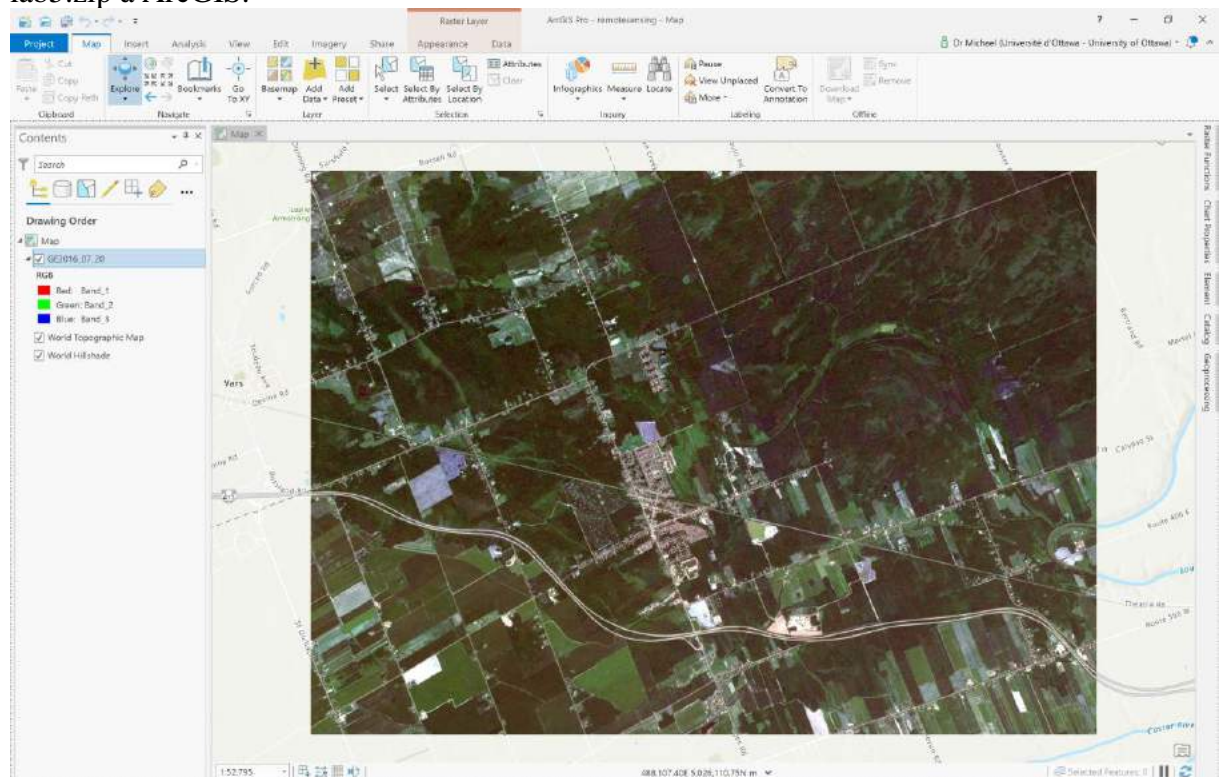
La disparition des forêts est un problème qui ne cesse de s'aggraver dans le monde d'aujourd'hui (voir <https://earthenginepartners.appspot.com/science-2013-global-forest>). Dans les questions de l'exemple, vous apprendrez à classer les forêts et les zones non forestières sur deux images satellites datant respectivement de 2001 et de 2016, puis à déterminer la quantité de forêt perdue ou gagnée entre-temps dans une petite région du sud-est d'Ottawa. Pour obtenir une estimation de la perte de forêt, dans ce laboratoire, vous ferez de la détection de changement post-classification en utilisant une image Landsat 7 ETM+ de 2001 à une résolution de 30 m, et une image RapidEye de 2016 à une résolution de 5 m. Une détection de changement post-classification vous permet d'éviter de nombreuses étapes de prétraitement nécessaires pour d'autres méthodes de détection de changement, et vous permet également d'utiliser directement les nombres numériques dans les images sans, par exemple, entreprendre des corrections radiométriques absolues ou des corrections de la réflectance du sol pour les différentes images utilisées.

Cette section présente des exemples de choses que vous devrez savoir pour répondre aux questions que vous remettrez pour obtenir vos notes. Votre tâche est de revoir et de répéter les exemples de cette section, et les couches dérivées que vous produisez dans cette section sont utilisées directement dans les questions à la fin de l'exercice. D'autres questions sont pertinentes pour vous aider à répondre aux questions à la fin de l'exercice.

Q1: Comment puis-je visualiser différentes combinaisons de bandes d'une image télédéteectée?

Diverses combinaisons de bandes peuvent vous aider à identifier différentes caractéristiques dans les images de télédétection.

1. Ajoutez l'image appelée "RE2016_07_20" du fichier IMAGERY.GDB dans le fichier lab5.zip à ArcGIS.



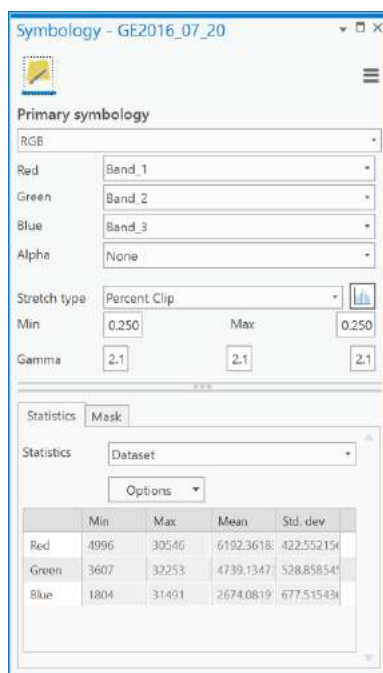
Nom de l'étudiant: _____ # d'étudiant: _____

2. Il s'agit d'une section d'une image multispectrale RapidEye acquise le 20 juillet 2016. Cette constellation de satellites est utilisée pour la surveillance et la recherche agricole et phénologique en raison du temps de revisite quotidien (hors du nadir, 5,5 jours au nadir) dû à une constellation de 5 satellites. Le capteur est donc utile pour surveiller le stress des cultures ou l'apparition de maladies dans de vastes régions. Le capteur RapidEye est multispectral, avec les bandes suivantes:

440 - 510 nm (bleu)
520 - 590 nm (vert)
630 - 685 nm (rouge)
690 - 730 nm (bord rouge)
760 - 850 nm (proche IR)

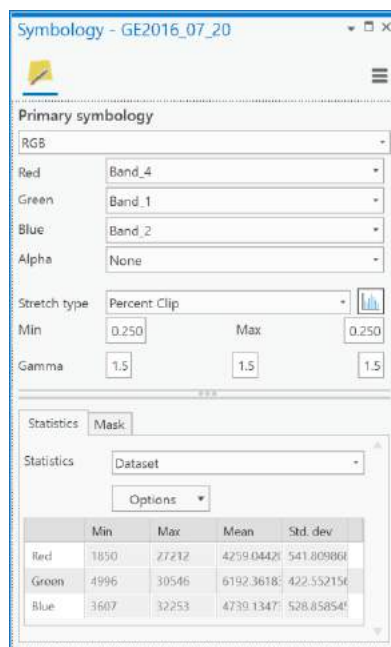
Cette image n'est pas corrigée atmosphériquement. Dans la plupart des cas, l'utilisation de cette image pour évaluer le stress des plantes ou la phénologie nécessiterait une correction atmosphérique, comme vous pourrez l'apprendre dans un autre cours, mais nous ne le ferons pas ici. De plus, l'image que vous utilisez ici est orthorectifiée, mais uniquement à partir d'une élévation constante, ce qui n'est pas très précis.

3. Pour visualiser d'autres combinaisons de bandes de l'image, cliquez avec le bouton droit de la souris sur la couche "RE2016_07_20" dans le volet de contenu (« Contents »), puis sélectionnez symbologie (« Symbology »):

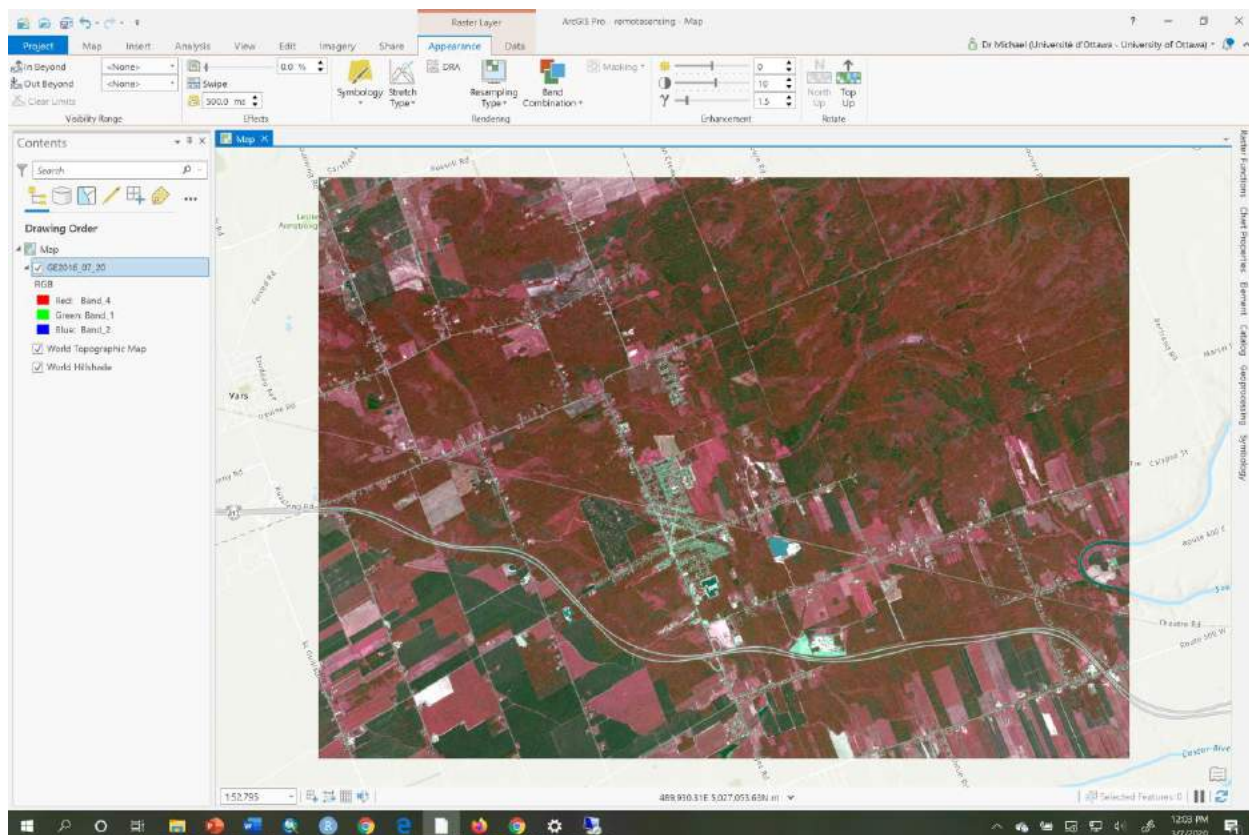


4. Par défaut, ArcGIS affiche la combinaison de bandes de couleurs visibles 3, 2, 1 pour le rouge, le vert et le bleu. Vous pouvez changer cette combinaison en Fausses couleurs infrarouges (« False Color Infrared »), par exemple, en choisissant 4,1,2 (Band_4 pour la bande rouge, Band_3 pour la bande verte et Band_2 pour la bande bleue) dans la colonne Band de l'onglet Symbology:

Nom de l'étudiant: _____ # d'étudiant: _____



5. La carte résultante est un "composite fausses couleurs" et indique en rouge les endroits où la végétation est présente.



6. Vous pouvez essayer d'autres combinaisons de bandes pour créer différents composites de fausses couleurs, dont certains peuvent visuellement mettre en valeur différents éléments du paysage imagé.

Nom de l'étudiant: _____ # d'étudiant: _____

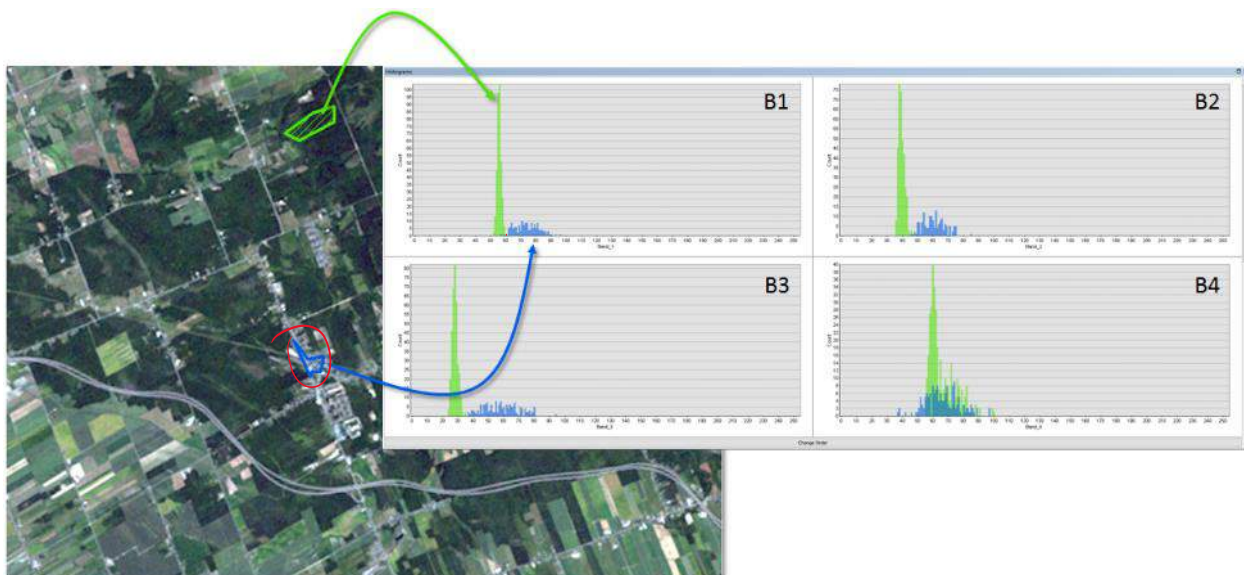
Q2: Comment classifier une image télédétectée?

Dans ce laboratoire, nous allons effectuer ce que l'on appelle une "classification supervisée", dans laquelle l'analyste d'images (vous) fournit à l'ordinateur des exemples de pixels contenant différentes caractéristiques du paysage à traiter en tant que classes d'information. Une classe d'information est une catégorie nominale que vous souhaitez extraire d'une image télédétectée. Par exemple, si vous souhaitez extraire les zones forestières d'une image, vous pouvez créer une classe d'information appelée **Forêt** et une autre appelée **Non-Forêt**. Vous pouvez ensuite trouver des exemples de pixels appartenant à la classe d'information définie par l'utilisateur appelée **Forêt** et un autre ensemble de pixels pour la classe appelée **Non-Forêt**. Vous pouvez avoir de nombreuses autres classes d'information en fonction du type d'information que vous souhaitez extraire d'une image. Par exemple, les classes d'information peuvent inclure "eau", "terres cultivées", "résidentiel", "commercial", "forêt dense", "forêt ouverte", etc. En fonction du type d'image obtenue par télédétection et de sa résolution spatiale et spectrale, il est possible d'extraire de nombreux types d'informations.

Pour effectuer une classification supervisée, vous avez besoin de deux choses:

1. Données d'apprentissage : Les données d'apprentissage sont constituées de pixels de l'image appartenant à différentes classes d'informations que vous souhaitez extraire de l'image pour créer une couche raster thématique. Par exemple, les pixels appartenant à la catégorie **Forêt** et les pixels appartenant à la catégorie **Non-Forêt** sont deux exemples de classes d'information. Examinez la figure ci-dessous. Le polygone vert est un polygone d'entraînement qui englobe un ensemble de pixels (toutes bandes confondues) que l'analyste a identifiés comme appartenant à la classe d'information appelée Forêt dans l'image. La fréquence absolue (le nombre de pixels) des valeurs de pixels "**Forêt**" avec différents valeurs numériques (VN), à l'intérieur du polygone vert, est représentée sous forme d'histogrammes verts pour chaque bande. De même, le polygone bleu contient des pixels non forestiers et constitue un site d'entraînement contenant des pixels non forestiers dans toutes les bandes de l'image. La fréquence absolue des VN dans la catégorie d'information **Non-Forest** pour chaque bande est représentée sous forme d'histogrammes bleus. Pour que la classification soit précise, les pixels d'entraînement de chaque classe d'information doivent avoir séparé au moins un des histogrammes, c'est-à-dire que les histogrammes de chaque classe d'entraînement doivent se chevaucher le moins possible. Ce point est abordé en détail plus loin. Dans ce cas, la séparabilité est observée dans les bandes B1, B2, B3, mais pas dans la bande B4.

Nom de l'étudiant: _____ # d'étudiant: _____



2. Un fichier de signature : Un fichier de signature porte l'extension ".gsg" et contient un résumé statistique des différentes valeurs de pixels dans chacune des classes d'information spécifiées dans l'ensemble de données d'apprentissage :

```

# I7_2001_08_25tsgsg - 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
# 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
# -----

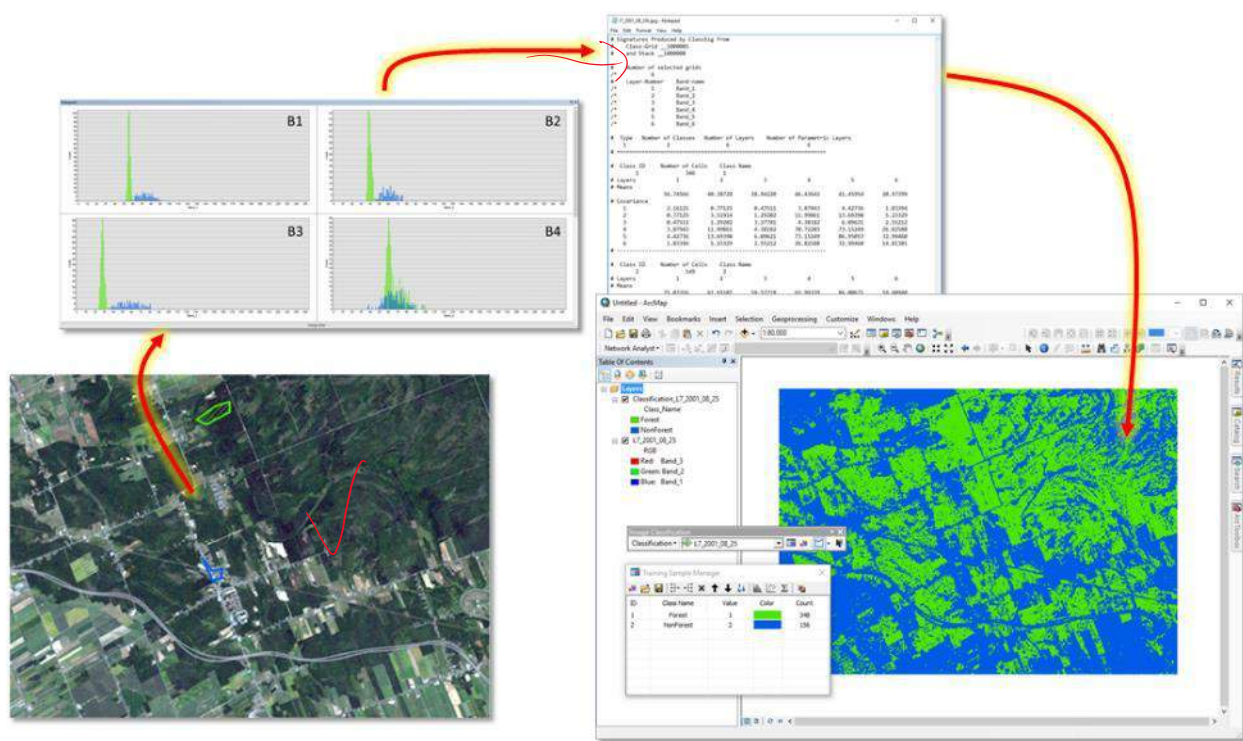
# Class ID Number of Cells Class Name
# 2 149 2
# Layers 1 2 3 4 5 6
# Means
# Covariance
1 56.16779 50.66044 78.50077 -58.31058 7.27680 58.92536
2 50.66044 51.12062 77.10820 -46.60371 17.95506 61.78573
3 78.50077 77.10820 128.92137 -81.39475 32.30015 104.02562
4 -58.31058 -46.60371 -81.39475 119.56077 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

```

Nom de l'étudiant: _____ # d'étudiant: _____

Le fichier de signature ci-dessus spécifie les valeurs moyennes des histogrammes et leur variabilité (covariance) au sein de chaque bande d'image. Le fichier de signature est un résumé statistique des histogrammes de chaque bande, pour chaque classe d'information. Par exemple, le fichier contient la moyenne et la covariance des VN pour les pixels appartenant à la classe Forêt, et de même pour Non-Forêt, pour chaque bande. Ces informations permettent à un classificateur automatique de placer chaque pixel de l'image dans sa classe respective.

À l'aide du fichier de signatures créé à partir des histogrammes d'entraînement, une classification supervisée attribue tous les pixels de l'image aux classes d'information spécifiées dans le fichier de signatures. Le classificateur examinera votre fichier de signatures et attribuera chaque pixel de votre image à l'une des catégories d'informations que vous avez définies dans l'ensemble de données d'apprentissage. Le résultat du classificateur est une couche raster thématique qui peut ensuite être utilisée dans l'algèbre de cartes ou dans d'autres processus de modélisation raster:

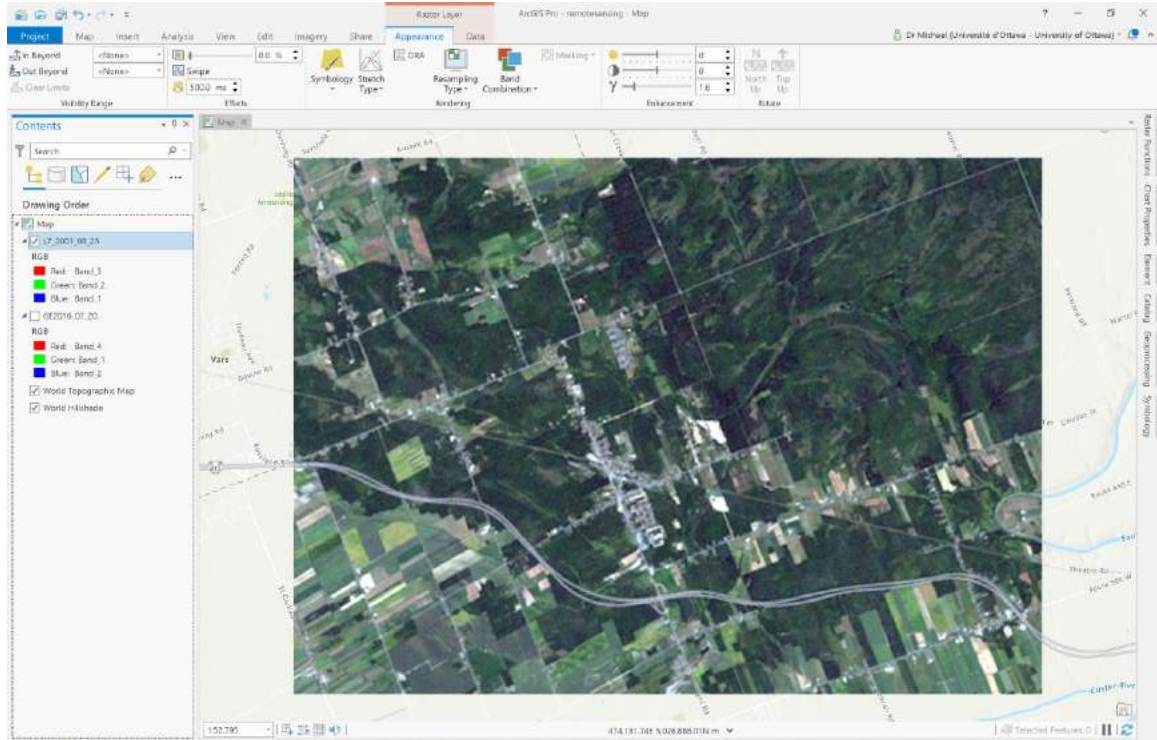


Les classificateurs sont des algorithmes d'apprentissage automatique qui apprennent, à partir de vos données d'apprentissage (dans le fichier de signature), à classer n'importe quel pixel de l'image, y compris ceux qui ne font pas partie de votre ensemble d'apprentissage. Il existe de nombreux types de classificateurs qui présentent des avantages et des désavantages différents ; ils portent des noms tels que Maximum de vraisemblance, Réseaux neuronaux, Machines à vecteurs de support, K-means, Forêt aléatoire et Iso-Cluster. Quel que soit le classificateur utilisé, la qualité des données d'apprentissage est primordiale et a le plus grand effet sur la précision de la couche cartographique résultante.

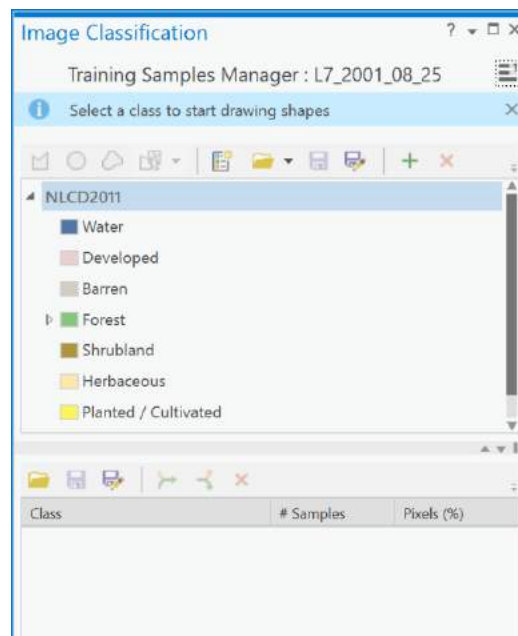
Ok, commençons. Nous allons d'abord travailler avec une image Landsat 7 du 25 août 2001 appelée "L7_2001_08_25" dans le fichier IMAGERY.GDB:

Nom de l'étudiant: _____ # d'étudiant: _____

3. Ajoutez le fichier "L7_2001_08_25" à ArcGIS et veillez à le sélectionner dans le volet *Contents*:














4. Cliquez sur l'onglet *Imagerie* (« *Imagery* »), puis sur le bouton Outils de classification (« *Classification Tools* ») et choisissez le gestionnaire d'échantillons d'entraînement (« *Training Samples Manager* »). Il se peut qu'il y ait un ensemble de classes d'information par défaut, ne vous inquiétez pas, nous nous en occuperons dans les étapes suivantes.



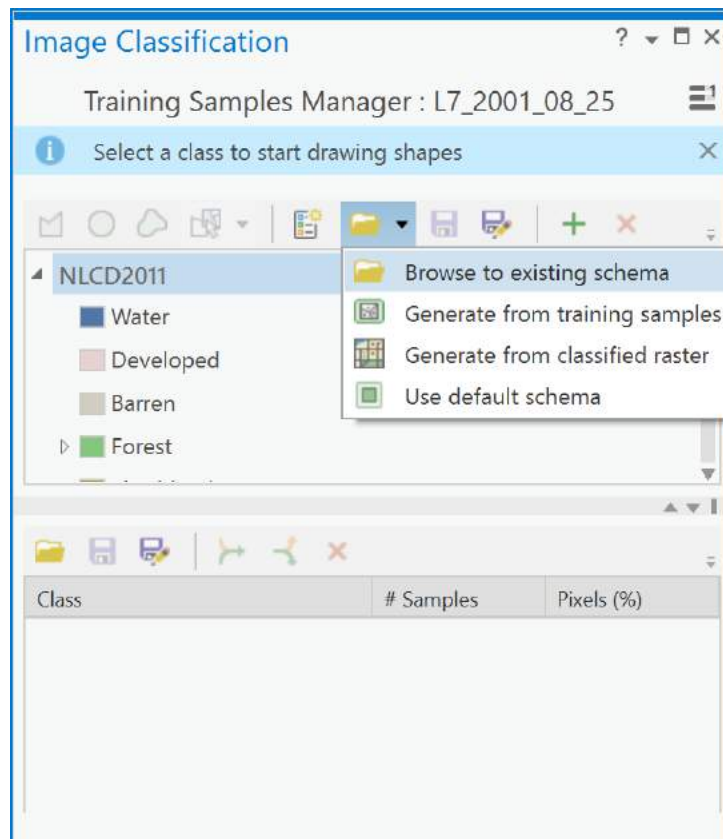
Nom de l'étudiant: _____ # d'étudiant: _____

5. Voici la description du gestionnaire d'échantillons d'entraînement (« *Training Samples Manager* »).

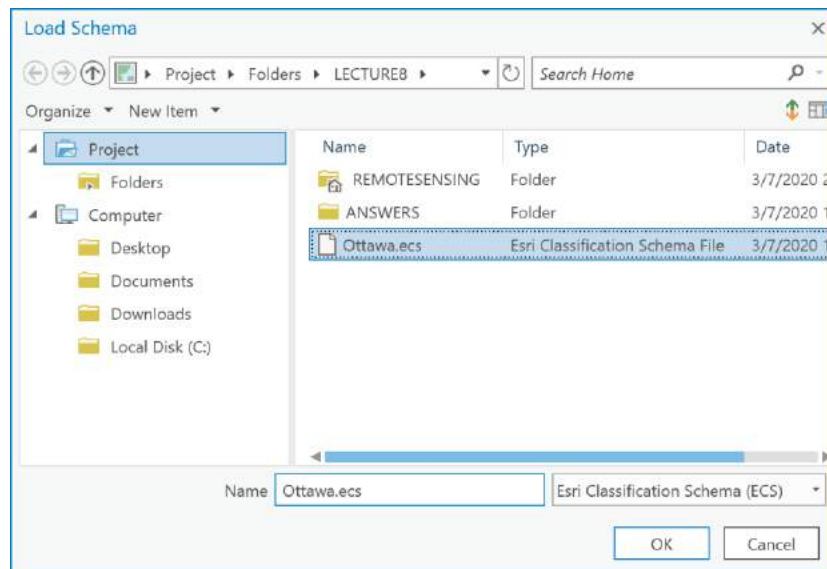
Outil	Fonction
	Créer un échantillon d'entraînement en dessinant un point sur les pixels ou les objets dans le raster.
	Créer un échantillon d'entraînement en traçant un polygone autour des pixels ou objets dans le raster.
	Créer un échantillon d'entraînement en traçant un cercle autour des pixels ou objets dans le raster.
	Créer un échantillon d'entraînement en traçant une forme à main levée autour des pixels ou objets dans le raster.
	Créer un échantillon d'entraînement en sélectionnant un segment à partir d'une couche segmentée. Cette option est seulement disponible s'il existe une couche segmentée dans la fenêtre Contents (Contenu) . Activer Segment Picker (Sélecteur de segment) en mettant en surbrillance la couche segmentée dans la fenêtre Contents (Contenu) , puis sélectionner la couche dans la liste déroulante Segment Picker (Sélecteur de segment) .
	Créer une structure de classification. Cliquer avec le bouton droit sur le titre New Schema (Nouvelle structure) et cliquez sur Add New Class (Ajouter une nouvelle classe) .
	Sélectionner une option relative à la structure de classification. <ul style="list-style-type: none"> • Accéder à une structure existante. • Générer une structure à partir d'une classe d'entités d'échantillons d'entraînement existante. • Générer une structure à partir d'un raster classé existant. • Utiliser la structure par défaut, 2011 National Land Cover Database.
	Enregistrer les modifications apportées à la structure.
	Enregistrer une copie de la structure.
	Ajouter une catégorie de classe à la structure. Sélectionner le nom de la structure pour créer une classe parent au niveau le plus élevé. Sélectionner le nom d'une classe existante pour créer une sous-classe.
	Supprimer la classe sélectionnée ou la catégorie de sous-classe de la structure.

6. En utilisant le gestionnaire d'échantillons d'entraînement, vous ouvrirez un jeu d'entraînement préexistant de polygones dans une classe d'entités appelée "L7_2001_08_25_Training" dans la géodatabase IMAGERY.GDB. Cliquez d'abord sur le bouton Schéma de classification (« *Classification schema* ») et choisissez Parcourir vers un schéma existant (« *Browse to an existing schema* »):

Nom de l'étudiant: _____ # d'étudiant: _____

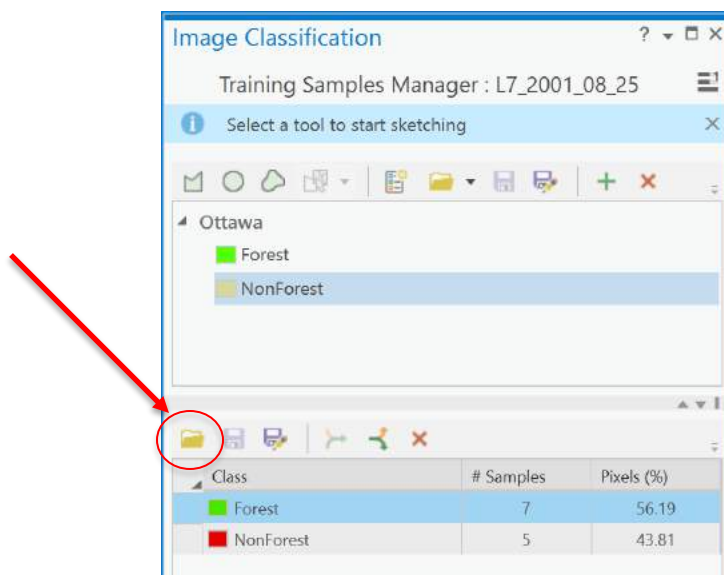


7. Naviguez jusqu'à l'endroit où vous avez dézippé le fichier lab5.zip, choisissez le fichier "Ottawa.ecs" et cliquez sur OK:



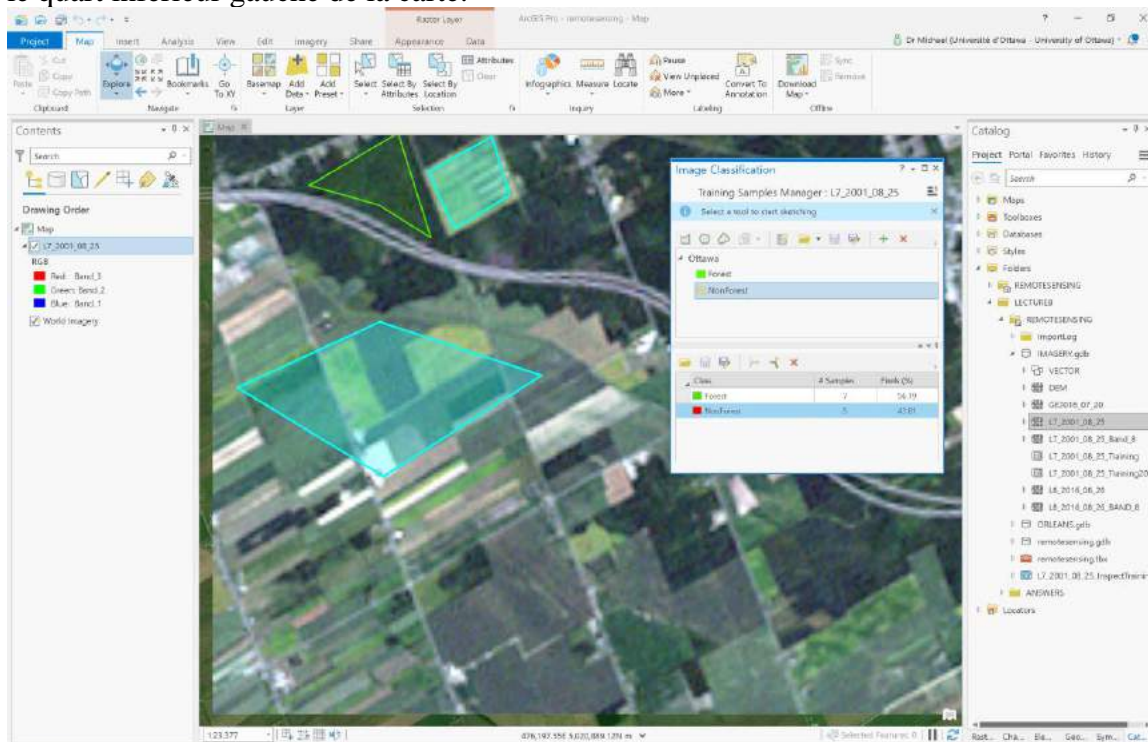
8. Cliquez maintenant sur le bouton Charger des échantillons d'entraînement (« Load training samples ») dans le demi-volet du bas, naviguez jusqu'à la classe d'objets "L7_2001_08_25_Training" et choisissez-la dans IMAGERY.GBD. Votre volet Gestionnaire d'échantillons d'entraînement devrait maintenant ressembler à ceci:

Nom de l'étudiant: _____ # d'étudiant: _____



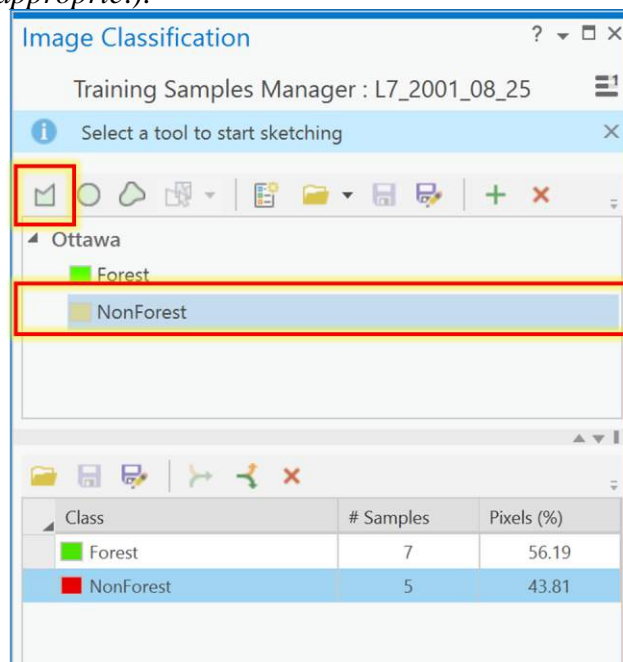
Vous pouvez voir que la classe **Forêt** correspond à la couleur verte et que les sites d'entraînement sont représentés par des polygones verts sur la carte. De même, pour la classe d'information **Non-Forest**, les polygones sont représentés en rouge. Ce qui est important ici, c'est que les deux classes d'information, **Forêt** et **Non-Forêt**, aient approximativement le même pourcentage de pixels dans chaque classe. À l'heure actuelle, la classe **Forêt** contient 56 % des pixels et la classe **Non-Forêt** 44 %. Ce n'est pas si mal, mais nous allons ajouter un nouveau polygone à la classe **Non-Forêt** pour augmenter son pourcentage.

9. Pour ajouter un autre exemple de Non-Forêt aux échantillons d'entraînement, zoomez sur le quart inférieur gauche de la carte:

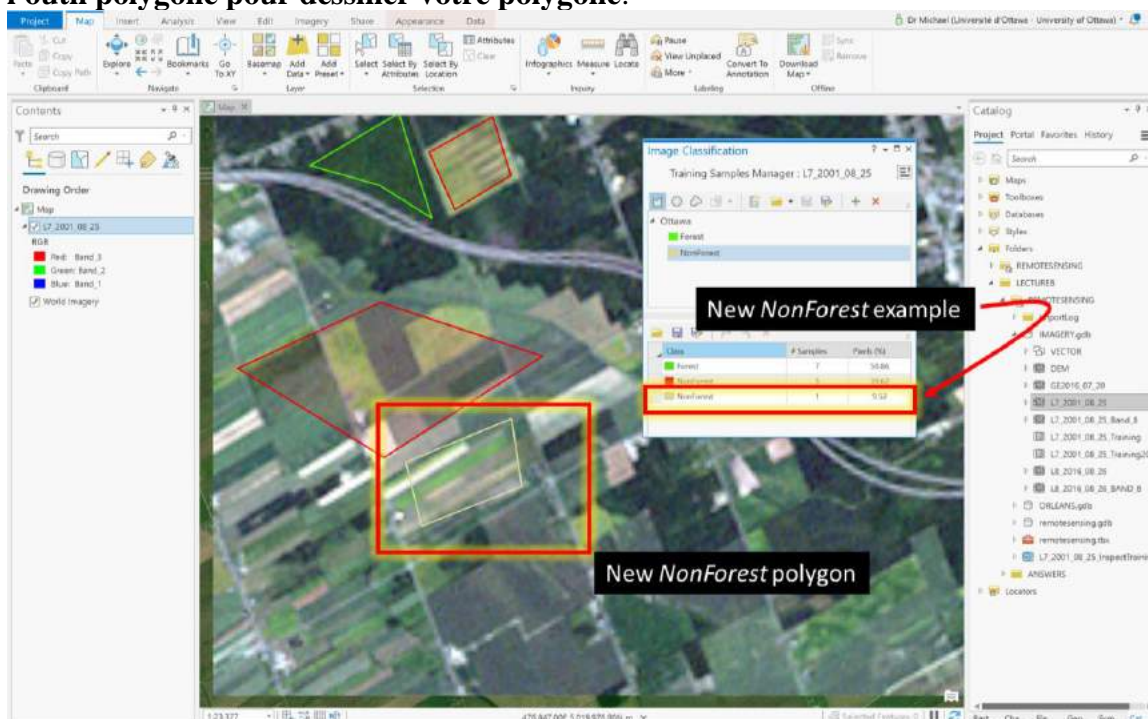


Nom de l'étudiant: _____ # d'étudiant: _____

10. Ensuite, dans le Gestionnaire d'échantillons d'entraînement, dans le demi-volet du haut, sous le schéma appelé Ottawa, cliquez sur la classe d'information **NonForest** (non forestier). Vous verrez que l'outil polygone de la barre d'outils devient actif (il s'agit de l'outil le plus à gauche; *des versions plus récentes d'ArcGIS peuvent avoir un bouton Rectangle en premier, ce qui fait que le bouton Polygone est le deuxième. Veuillez à sélectionner l'outil approprié.*).

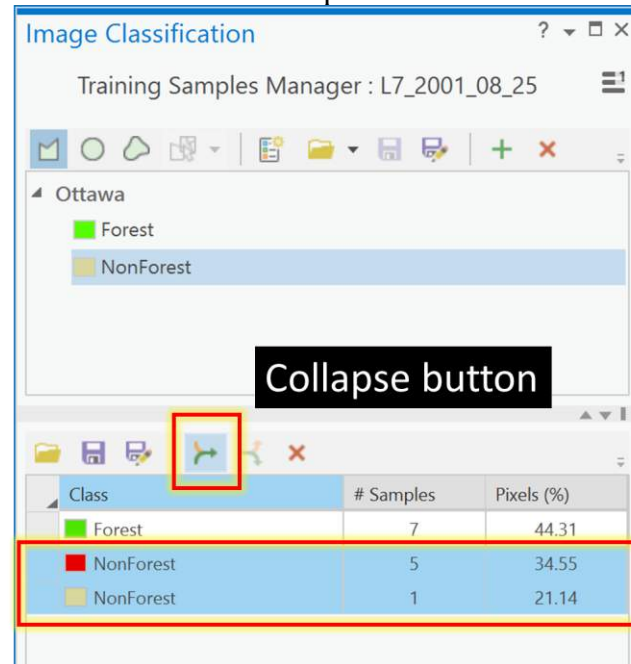


11. Cliquez maintenant sur l'outil polygone et dessinez un polygone dans n'importe quelle zone non boisée de l'image satellite. **Ne choisissez pas d'autre outil de dessin que l'outil polygone pour dessiner votre polygone:**

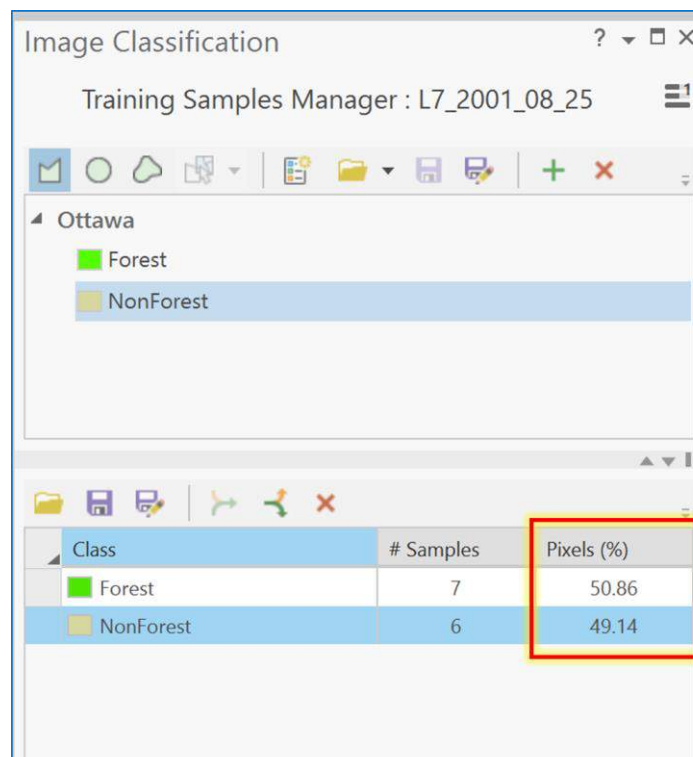


Nom de l'étudiant: _____ # d'étudiant: _____

12. Examinez le volet Gestionnaire d'échantillons d'entraînement, le demi-volet du bas, et vous verrez maintenant les polygones de l'exemple non forestier d'origine en rouge et le nouvel exemple non forestier que vous venez de créer. Appuyez sur la touche Ctrl et cliquez sur les deux exemples non forestiers, puis cliquez sur le bouton Réduire (« *Collapse* ») pour fusionner les deux exemples non forestiers en un seul ensemble:



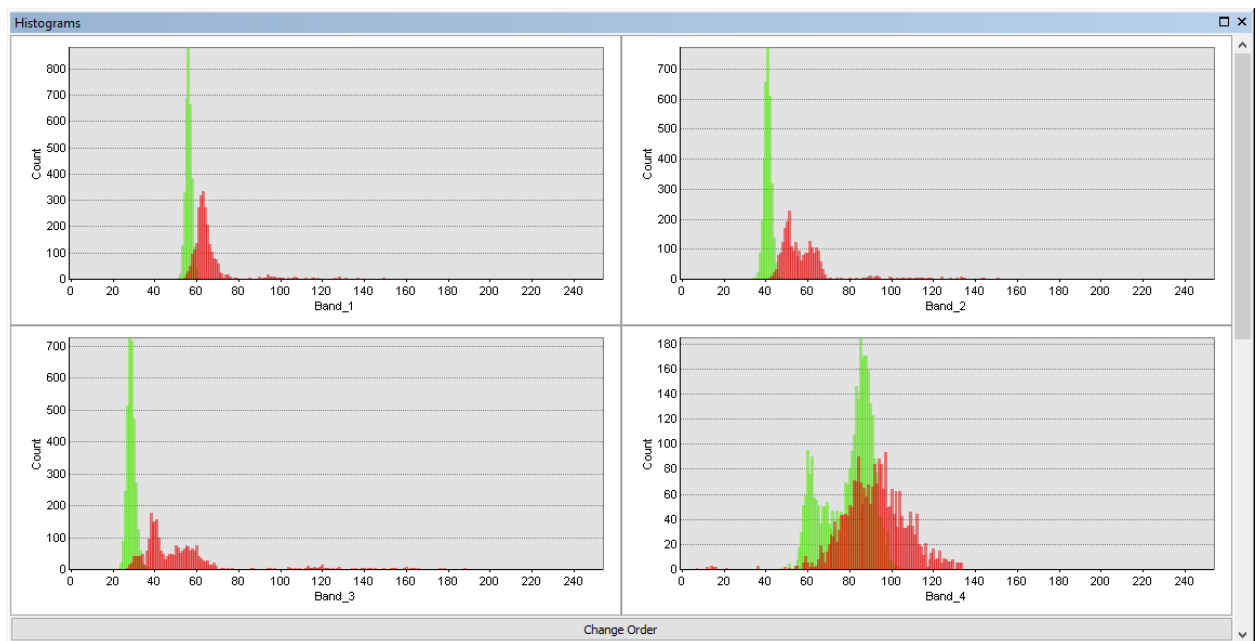
13. Le volet Gestionnaire d'échantillons d'entraînement devrait ressembler maintenant à ceci:



14. Vérifiez que les deux classes d'information ont maintenant environ 50 % de pixels chacune. Si le polygone que vous avez ajouté était trop grand, la classe **Non- Forêt** aura maintenant beaucoup plus que 50 % des pixels, ce qui signifie que vous n'avez pas réussi à équilibrer la taille des échantillons d'entraînement des deux classes. Dans ce cas, vous pouvez soit ajouter un polygone à la classe **Forêt** (et ainsi de suite, jusqu'à ce que vous parveniez à un résultat correct), ou vous pouvez faire ceci:
- Sélectionnez la classe **NonForest** dans la moitié du bas du gestionnaire d'échantillons d'entraînement et cliquez sur le bouton "*Expand*" (à côté du bouton "*Collapse*").
 - Vous verrez maintenant plusieurs lignes étiquetées **NonForest**, chacune correspondant à un polygone individuel.
 - Supprimez la dernière - celle que vous venez de créer.
 - Ajoutez-en un nouveau (plus petit).
 - Sélectionnez tous les polygones **NonForest** et réduisez-les ("*Collapse*").
 - Vérifiez que la répartition entre les deux classes est plus ou moins égale.
15. Examinez attentivement le volet Gestionnaire d'échantillons d'entraînement, qui fournit de nombreuses informations importantes sur les échantillons d'entraînement pour chaque classe d'information. Une chose qui n'est pas montrée est que **Forest** (la première classe listée) reçoit une valeur de pixel de 1 et **NonForest** (la classe suivante listée) reçoit une valeur de pixel de 2 (et ainsi de suite, si vous avez plus de classes). Ces valeurs deviendront les valeurs des cellules pour toute classification de l'image. La colonne "Pixels %" indique combien de pixels se trouvent dans chaque classe d'apprentissage. Dans le cas ci-dessus, 50,86 % des pixels d'entraînement se trouvent dans la classe d'information **Forêt** et 49,14 % dans l'ensemble d'entraînement **Non-Forêt**. Nous avons donc obtenu une proportion presque égale, ce qui est acceptable et meilleur que notre point de départ.
16. Ensuite, nous devons enregistrer nos nouveaux polygones d'entraînement dans une classe d'objet. Pour cela, cliquez sur le bouton "Save" dans le volet du bas du Gestionnaire d'échantillons d'entraînement, naviguez jusqu'au fichier IMAGERY.GDB et enregistrez le nouveau fichier d'entraînement sous le nom "L7_2001_08_25_Training2023". Nous n'avons pas besoin de sauvegarder le schéma "Ottawa" dans le demi-volet du haut parce qu'il n'a pas changé. Si nous avions ajouté une nouvelle classe d'information, nous aurions également dû enregistrer le nouveau schéma.
17. L'étape suivante est fondamentale car elle nous permettra d'explorer le degré de "pureté" de nos classes d'information sur la base des pixels que nous avons choisis pour chaque classe. Nous voulons savoir quelles sont les bandes qui présentent la plus grande *séparabilité* entre les forêts et les zones non forestières. La *séparabilité spectrale* fait référence à la distance spectrale entre les valeurs des pixels qui composent chaque classe d'information dans un ensemble de données d'échantillons d'entraînement. En fin de compte, nous voulons savoir quelles bandes de notre image de télédétection seraient les meilleures pour classer notre image en **Forêt** et **Non-Forêt**. Les bandes les plus performantes permettront de distinguer clairement nos deux classes d'information et il n'y aura que peu ou pas de chevauchement entre les valeurs des pixels correspondant à la catégorie **Forêt** et **Non-Forêt** dans ces bandes.

Nom de l'étudiant: _____ # d'étudiant: _____

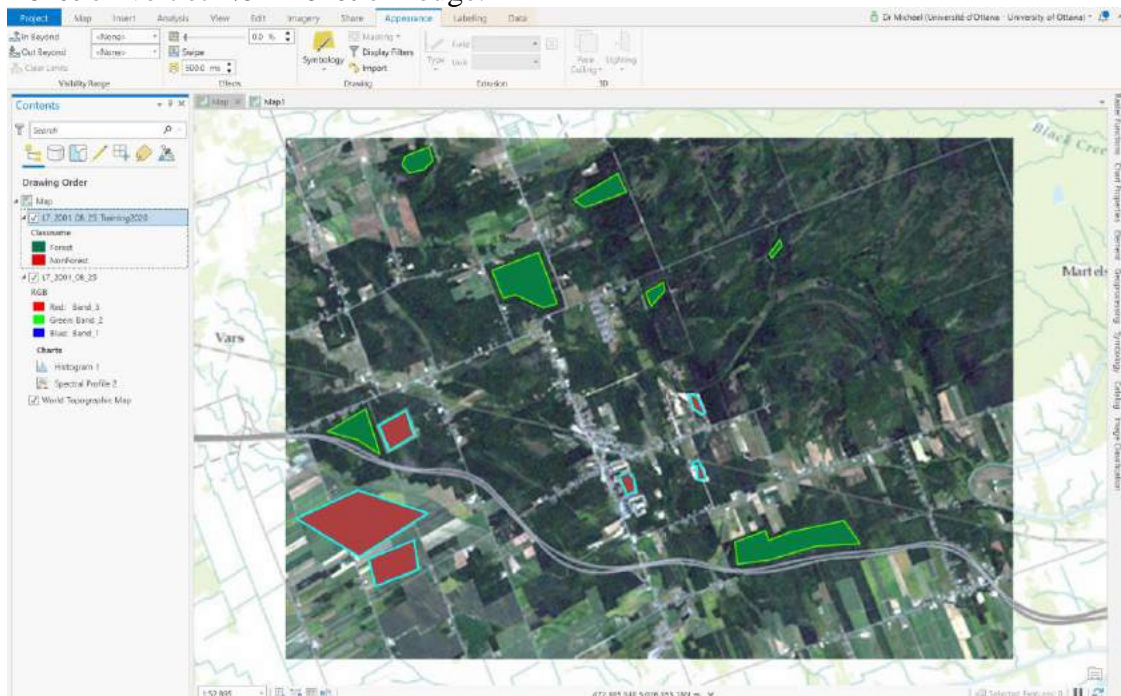
Pour comprendre la séparabilité spectrale, regardez la figure ci-dessous pour voir les histogrammes pour quatre bandes et deux classes d'information différenciées par les couleurs rouge et verte. Chaque histogramme montre la fréquence des valeurs numériques pour chaque classe, classe rouge ou classe verte, sur l'axe des ordonnées (y) et l'étendue des valeurs numériques associées à la classe rouge ou verte sur l'axe des abscisses (x). Par exemple, dans la Bande_1, nous voyons que la classe verte se situe entre 52 et 60 et la classe rouge entre 55 et 150. Ainsi, bien que les deux histogrammes soient clairement différents, il y a également un certain chevauchement autour des valeurs de 55-60, indiquées en rouge foncé. Toutefois, les pics des histogrammes rouge et vert ne se chevauchent pas, ce qui est une bonne chose. En général, une "bonne" bande pour la classification est une bande dont les histogrammes rouge et vert ne se chevauchent pas du tout. S'il n'y a pas de chevauchement entre les classes d'information dans une bande, les classes rouge et verte sont distinctes et complètement séparées. Par conséquent, une telle bande permettrait de classer l'image dans les classes d'information rouge et verte. Ici, aucune des bandes de l'image ne présente une séparabilité complète. La bande présentant le plus faible chevauchement est la bande_2. Pourquoi ? Parce qu'il y a peu de chevauchement entre les histogrammes rouge et vert. Inversement, la bande la plus défavorable est la bande_4, où le chevauchement est considérable et où la bande_4 n'aiderait donc pas à classer l'image correspondante dans les bandes rouge et verte.



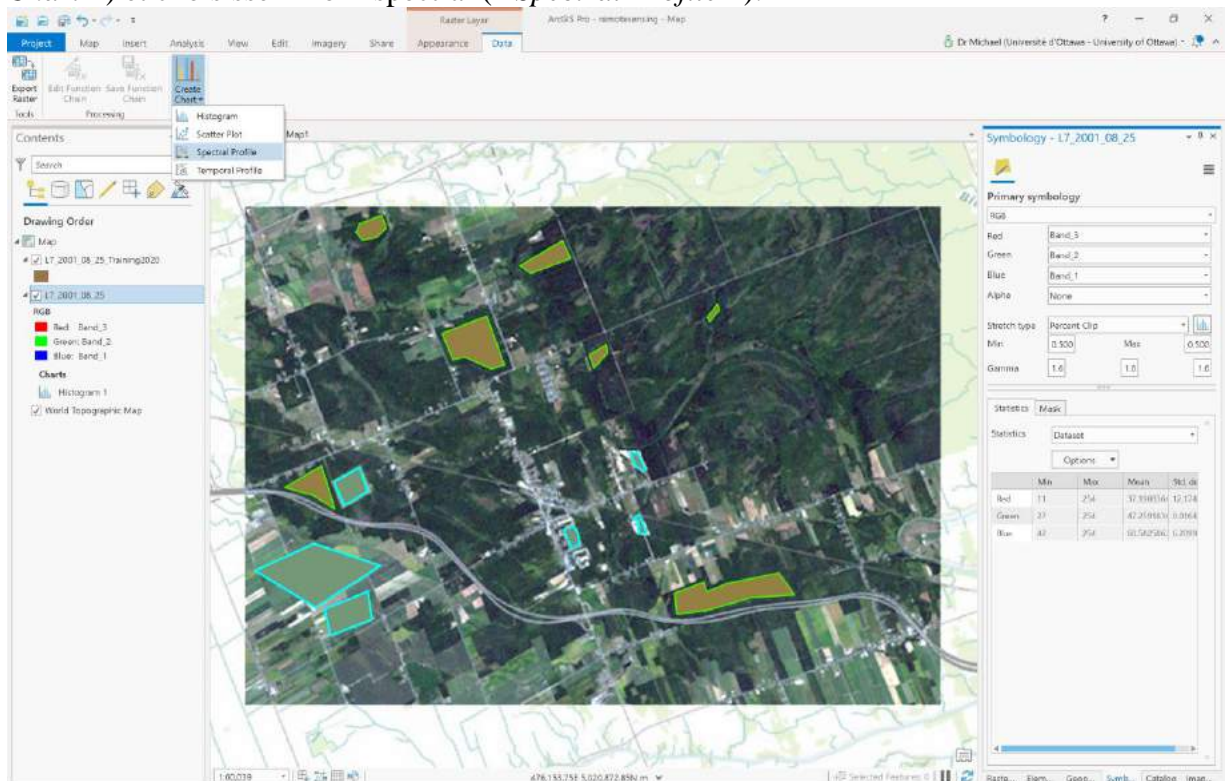
Il n'est pas nécessaire de générer des histogrammes pour chaque bande afin d'examiner la séparabilité. La façon la plus courante d'examiner la séparabilité spectrale entre les différentes bandes est de générer un graphique des statistiques de chaque histogramme pour chaque bande dans un profil spectral.

Nom de l'étudiant: _____ # d'étudiant: _____

18. Pour créer un graphique spectral, assurez-vous d'abord que vous avez ajouté votre couche "L7_2001_08_25_Training2023" au panneau de contenu, allez dans la symbologie, et symbolisez-la en utilisant des valeurs uniques avec le champ « *ClassName* », et mettez **Forêt** en vert et **Non-Forêt** en rouge.

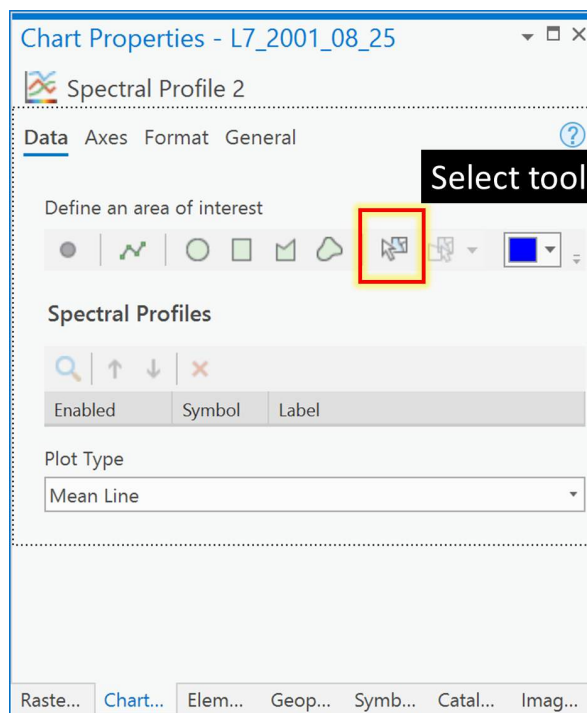


19. Ensuite, choisissez la couche "L7_2001_08_25" dans le volet de contenu. Cliquez ensuite sur l'onglet Données (« *Data* »). Cliquez maintenant sur Créer un graphique (« *Create Chart* ») et choisissez Profil spectral (« *Spectral Profile* »):

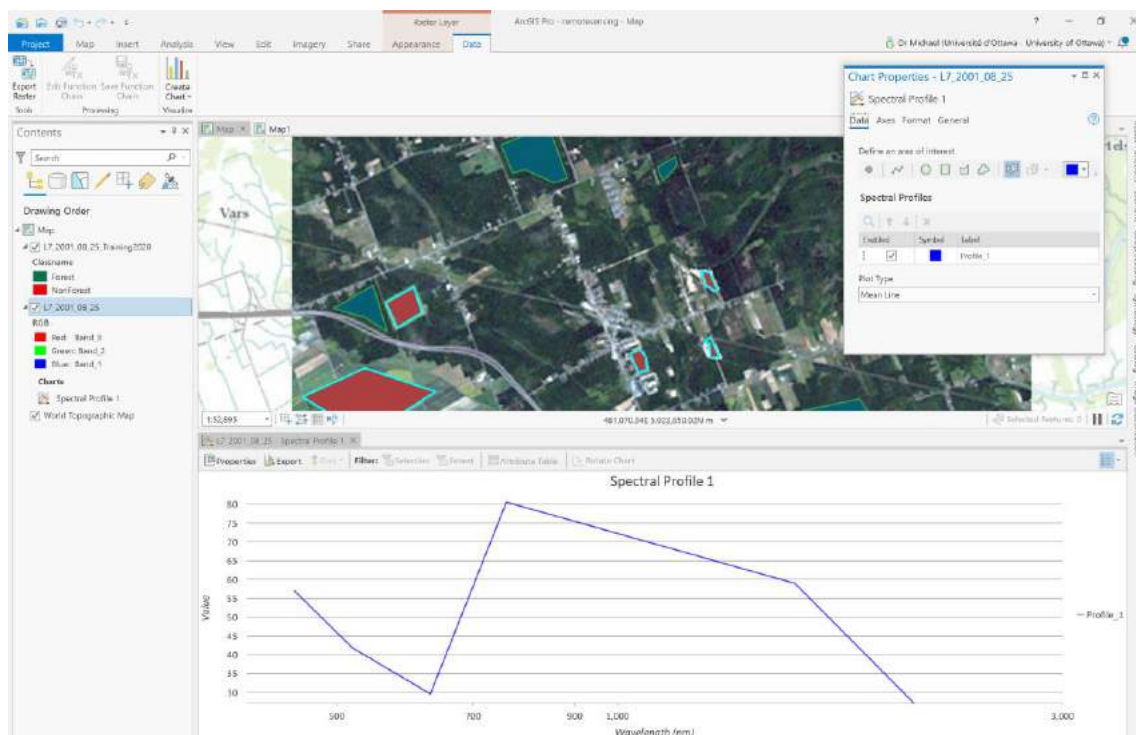


Nom de l'étudiant: _____ # d'étudiant: _____

20. Ensuite, dans le volet Propriétés du graphique (« *Chart Properties* »), cliquez sur le bouton Sélecteur d'entités (« *Feature selector* »):

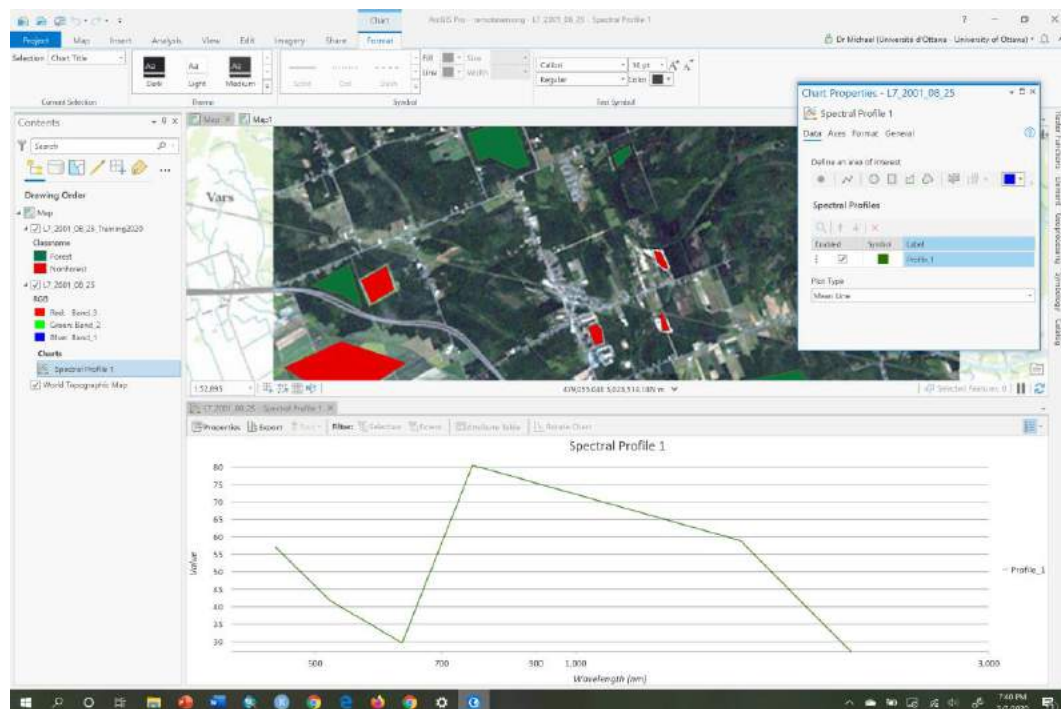


21. Cliquez maintenant sur l'un des polygones de **forêt** dans le panneau de la carte. Vous verrez un graphique créé avec une seule ligne appelée *Profile_1* qui représente la moyenne des VN sur toutes les bandes pour cette classe d'information appelée **Forêt**.

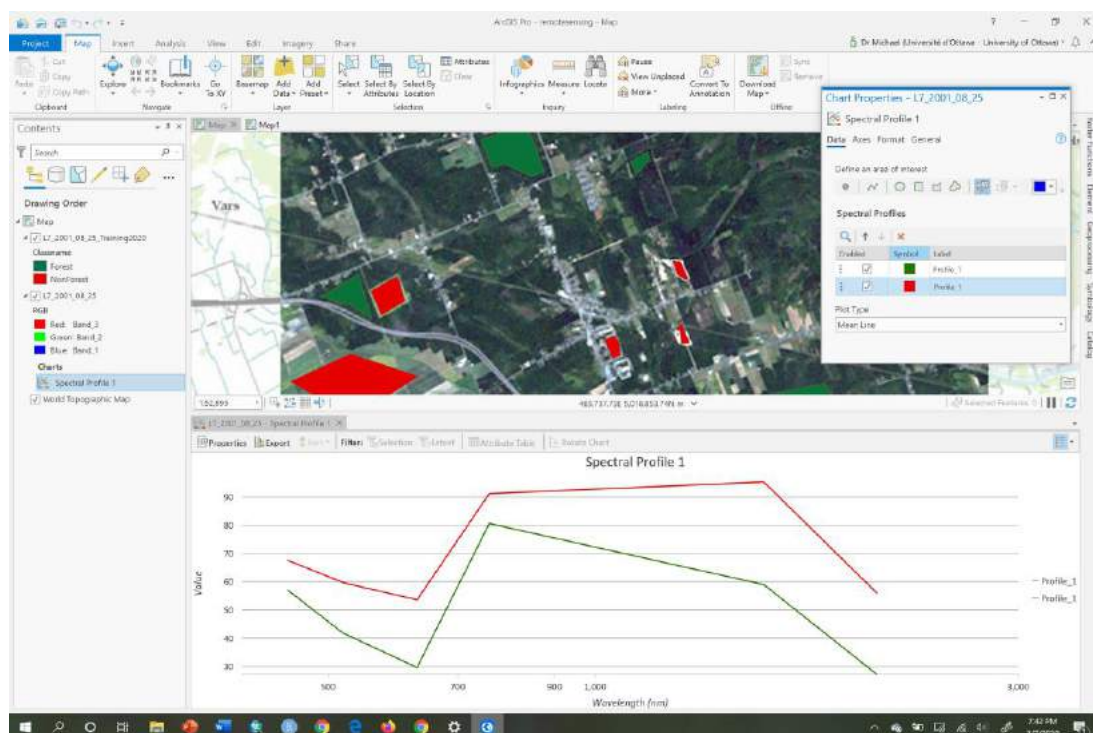


Nom de l'étudiant: _____ # d'étudiant: _____

22. Maintenant, dans le volet Propriétés du graphique (« *Chart Properties* »), changez le label de cette ligne de moyenne spectrale de "Profil_1" à **Forêt** (si cela fonctionne!) et changez la couleur en vert et vous verrez votre ligne devenir verte:



23. Répétez maintenant les étapes 21-22 mais pour les polygones d'entraînement **NonForest** et rendez la nouvelle ligne rouge pour correspondre à **Non-Forêt** et renommez le label si vous le pouvez dans votre version d'ArcGIS Pro:

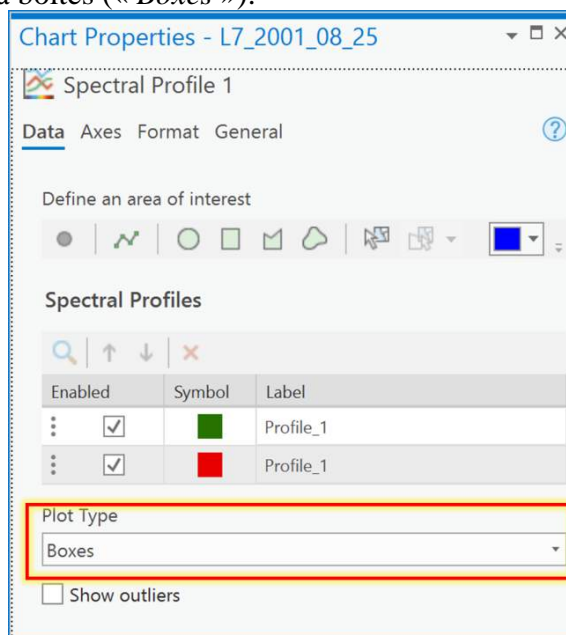


24. Examinez maintenant le graphique *Spectral Profile 1*:



Ce graphique montre la longueur d'onde sur l'axe des x plutôt que les noms des bandes parce que les métadonnées de l'image contiennent des informations sur la longueur d'onde de chaque bande. Vous pouvez déduire les bandes en observant où les lignes ont leurs "cassures". Par exemple, Band_1 se trouve à l'extrême gauche, Band_2 juste après 500 nm et Band_3 juste avant 700 nm, etc. Nous constatons que, dans chaque bande, il existe une différence entre les valeurs moyennes de chaque classe d'information. C'est peut-être dans la bande 4, juste après 700 nm, que la séparabilité est la plus faible, car c'est là que les lignes sont les plus proches. Nous pouvons vérifier cela à l'aide d'un diagramme statistique plus détaillé appelé diagramme en « boîtes et moustaches ».

25. Pour créer le graphique en boîte et en moustache de la séparabilité spectrale pour chaque bande, il suffit de modifier le type de graphique dans le volet Propriétés du graphique (« *Chart Properties* »). Dans le volet, changez le type de tracé (« *Plot Type* ») de moyenne (« *Mean* ») à boîtes (« *Boxes* »):

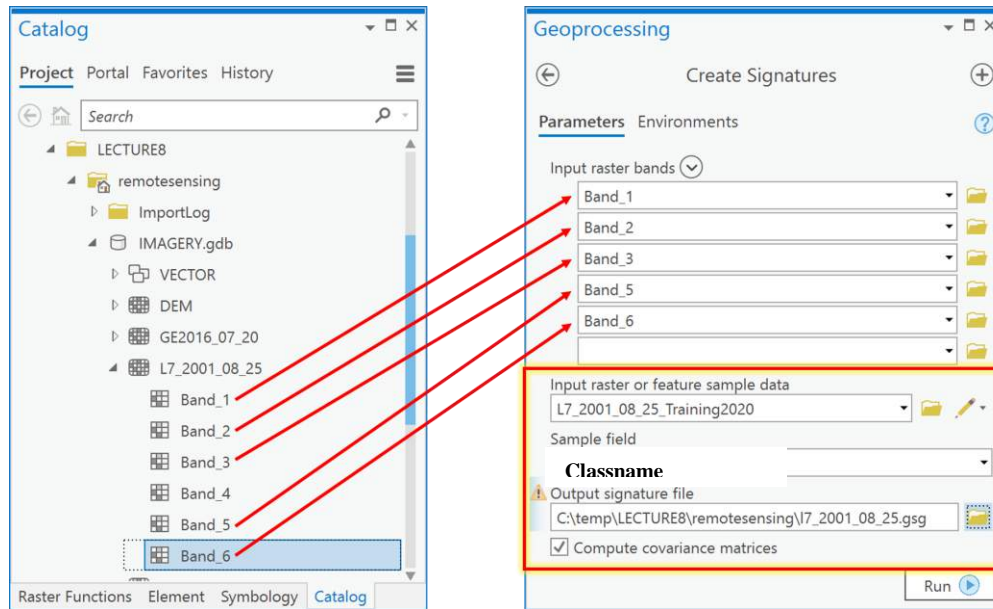


26. Examinez maintenant le diagramme en « boîtes et moustaches ».

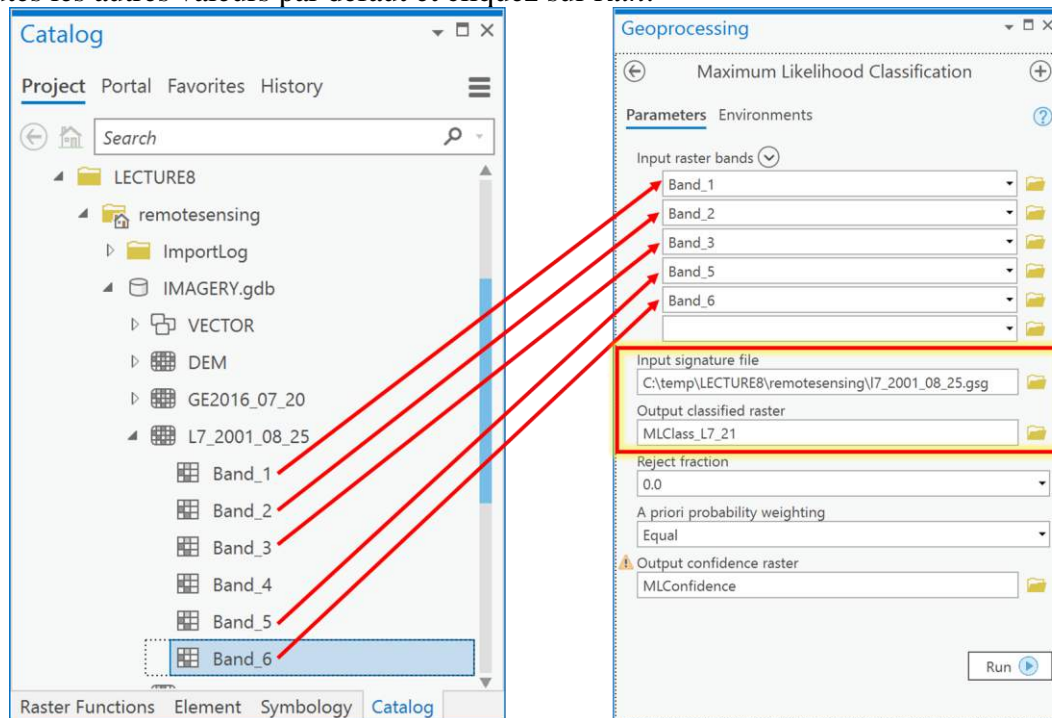


Dans ce diagramme, il y a en fait un diagramme en boîte et moustache pour chaque bande, chacun montrant les VN **forestiers** (vert) et **non forestiers** (rouge). Le haut et le bas des boîtes colorées sont les quartiles supérieurs et inférieurs de l'histogramme des VN pour la classe d'information. Les moustaches sont les deux lignes à l'extérieur de la boîte qui s'étendent jusqu'aux VN minimum et maximum pour la classe d'information. Il s'agit donc d'un résumé visuel des histogrammes pour chaque classe d'information, dans lequel nous pouvons maintenant examiner quelles bandes présentent la meilleure séparabilité. Il y a beaucoup plus d'informations ici que dans le graphique du profil spectral. Ce qui importe le plus ici, c'est que les boîtes ne se chevauchent pas, et ce pour toutes les bandes à l'exception de la Bande_4, qui est celle qui était également la plus proche dans le graphique du profil spectral. Nous pouvons donc envisager d'exclure la bande_4 de toute classification des **forêts** et des zones **non forestières**, mais nous devrions probablement conserver toutes les autres bandes.

27. Créez un fichier de signatures à partir des sites d'entraînement que nous pourrions utiliser pour la classification. Pour créer le fichier, allez dans l'onglet Analyse (« *Analysis* »), cliquez sur Outils (« *Tools* ») et recherchez l'outil Créer des signatures (« *Create Signatures* ») qui se trouve dans la boîte à outils *Spatial Analysis*. Désactivez l'outil *Create Signatures* afin de pouvoir le voir en même temps que votre volet Catalogue. Ensuite, dans le volet *Catalog*, naviguez jusqu'à *IMAGERY.GDB* et agrandissez l'image "L7_2001_08_25" de manière à voir toutes les bandes individuelles (Band_1, Band_2,...,Band_6). Faites glisser toutes les bandes à l'exception de la Bande_4 (puisque nous savons qu'elle a une mauvaise séparabilité) dans "Input raster bands", sous "Input raster or feature sample data" choisissez la classe objet "L7_2001_08_25_Training2023", sous "Sample field" choisissez **Classname**, et sauvegardez le fichier dans le dossier où vous avez vos données, et appelez-le "l7_2001_08_25.gsg". Cliquez sur "Run" (Exécuter).

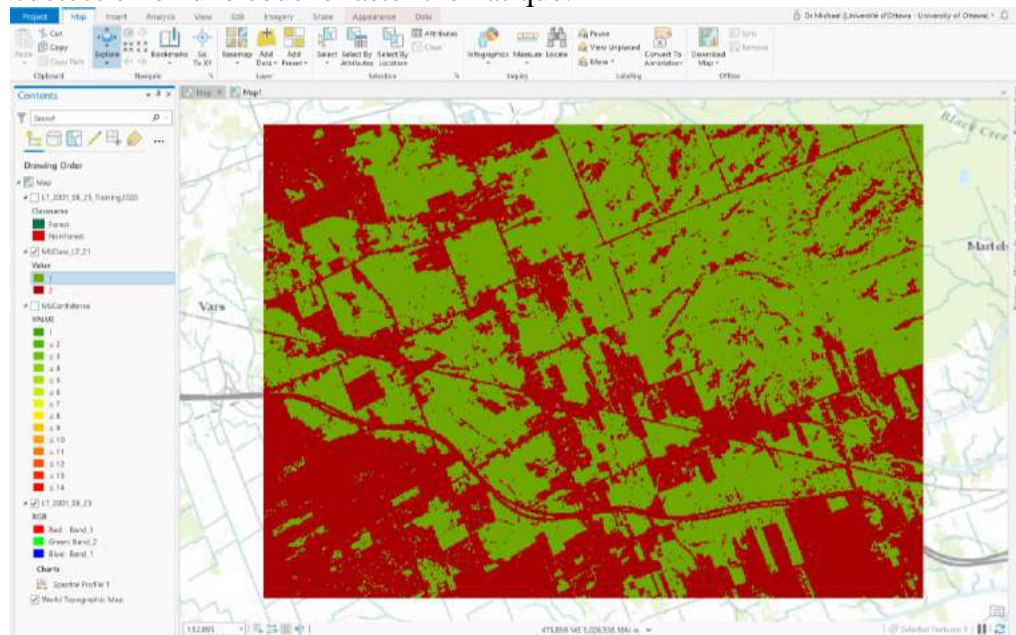


28. Utilisez maintenant le fichier de signatures pour classifier l'image dans les deux classes d'information Forêt et Non-Forêt. Pour cela, recherchez maintenant l'outil d'analyse spatiale "Maximum Likelihood Classification" (Classification par maximum de vraisemblance). Désactivez encore l'outil afin de pouvoir le voir en même temps que votre volet Catalogue. Ensuite, dans le volet Catalogue, naviguez jusqu'à IMAGERY.GDB et agrandissez l'image "L7_2001_08_25" de manière à voir toutes les bandes individuelles (Bande_1, Bande_2, ..., Bande_6). Faites glisser toutes les bandes à l'exception de la Bande_4 dans "Input raster bands", sous "Input signature file", choisissez le fichier de signature créé à l'étape précédente appelé "l7_2001_08_25.gsg", sous "Output classified raster" appelez-le "MLCLASS_L7_21", et enregistrez-le dans IMAGERY.GDB. Laissez toutes les autres valeurs par défaut et cliquez sur *Run*.



Nom de l'étudiant: _____ # d'étudiant: _____

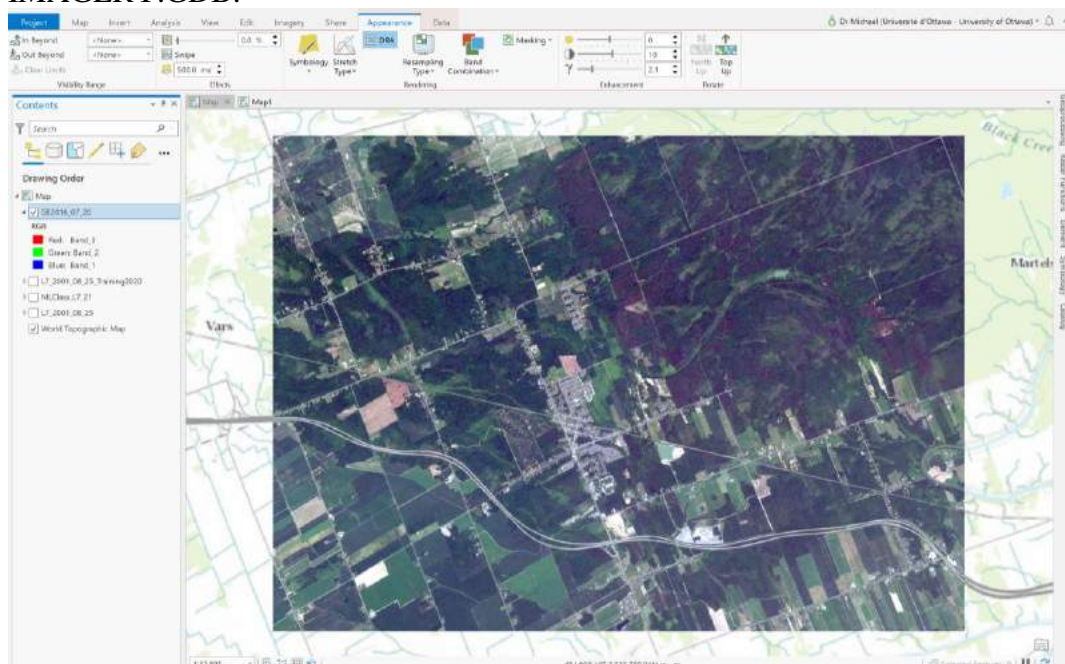
29. Examinez le raster classifié résultant (notez que vous pouvez changer la valeur 1 - Forêt en vert et la valeur 2 - Non-Forêt en rouge). Nous avons maintenant converti une image de télédétection en une couche raster thématique.



Q3: Comment créer un ensemble de données d'entraînement correspondant à mes classes d'information?

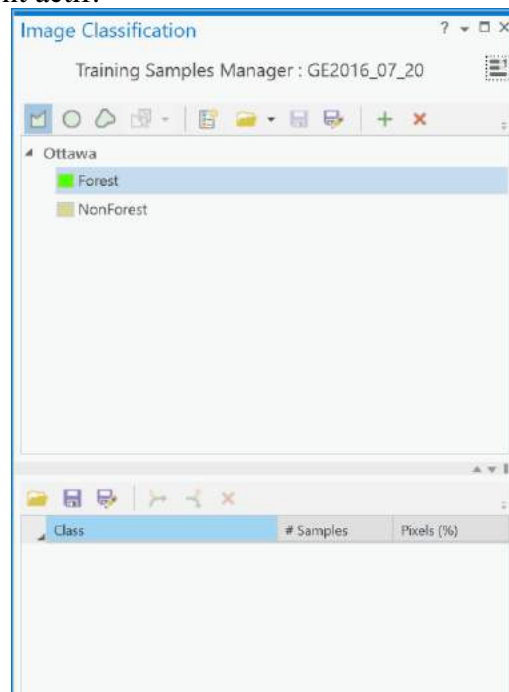
Les données d'entraînement sont constituées de pixels appartenant à différentes classes d'information. Par exemple, certains pixels appartiennent à la classe appelée **Forêt**, et un autre ensemble de pixels à la classe appelée **Non-Forêt**.

1. Si elle n'est pas déjà présente, ajoutez l'image "RE2016_07_20" à ArcGIS à partir de **IMAGERY.GDB**.



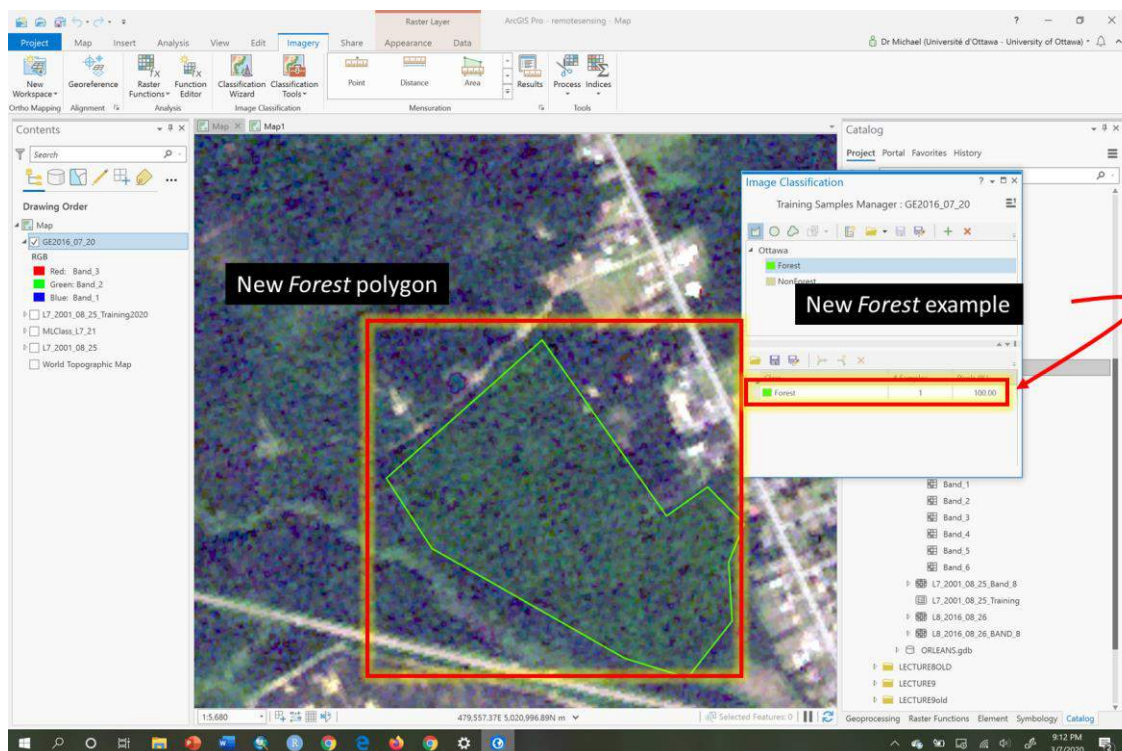
Nom de l'étudiant: _____ # d'étudiant: _____

2. Cliquez sur l'onglet Imagerie (« *Imagery* »), puis sur le bouton Outils de classification (« *Classification Tools* ») et choisissez le gestionnaire d'échantillons d'entraînement (« *Training Samples Manager* »).
3. En utilisant le gestionnaire d'échantillons d'entraînement, ouvrez le schéma Ottawa.ecs, comme vous l'avez fait précédemment.
4. Nous allons maintenant créer un nouvel ensemble de polygones d'entraînement pour les classes d'information Forêt et Non-Forêt, à utiliser avec l'image RapidEye. Notez que nous ne pouvons pas simplement réutiliser les polygones que nous avons utilisés pour l'image Landsat, car les zones à l'intérieur de ces polygones peuvent avoir changé de Forêt à Non-Forêt, ou vice versa, dans la période entre 2001 (lorsque l'image Landsat a été acquise) et 2016 (lorsque l'image RapidEye a été acquise). À cette fin, dans le Gestionnaire d'échantillons d'entraînement, dans le demi-volet du haut, sous le schéma appelé Ottawa, cliquez sur la classe d'information Forêt. Vous verrez que l'outil polygone de la barre d'outils devient actif.

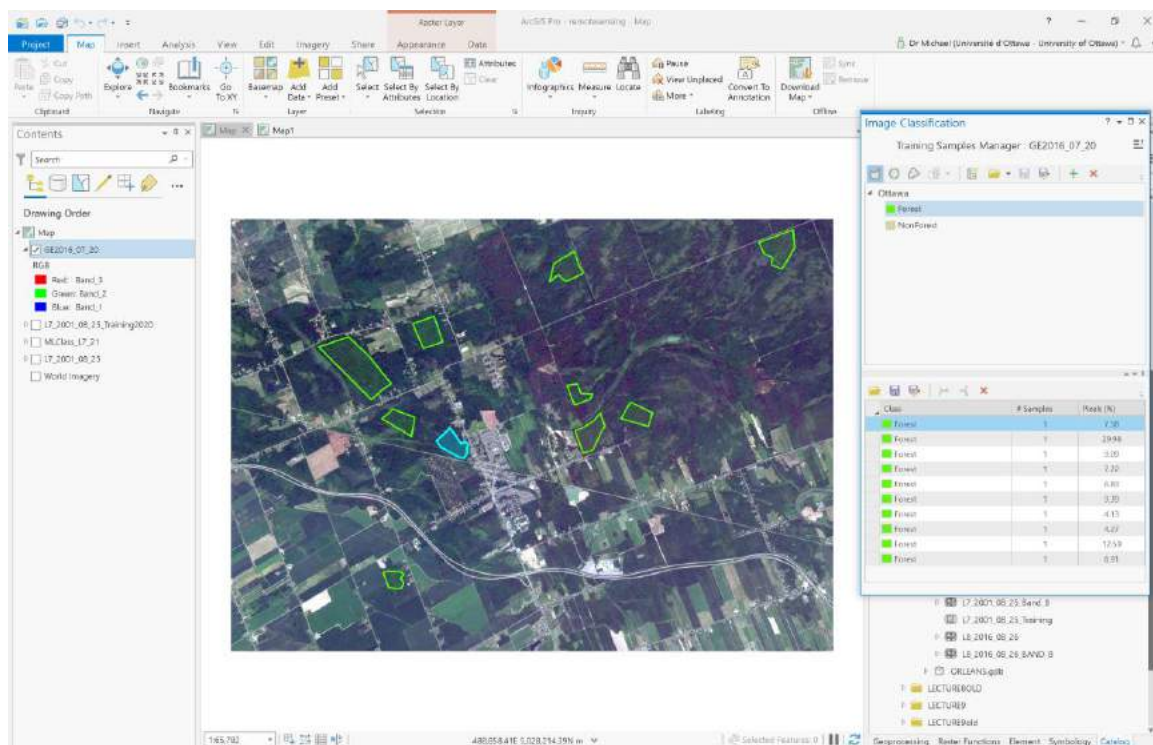


5. Cliquez maintenant sur l'outil polygone et dessinez un polygone dans n'importe quelle zone boisée de l'image satellite.

Nom de l'étudiant: _____ # d'étudiant: _____

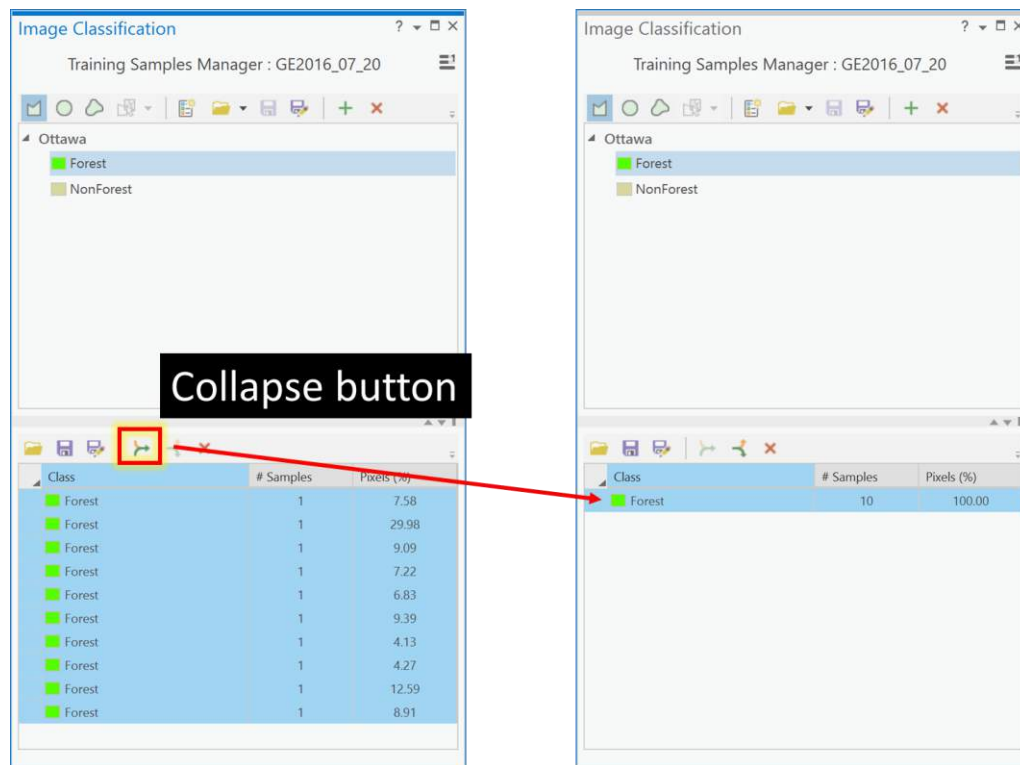


6. Répétez l'étape précédente neuf fois dans différentes zones forestières. Vous obtiendrez ainsi quelque chose qui ressemble à ceci:



Nom de l'étudiant: _____ # d'étudiant: _____

7. Examinez le volet Gestionnaire d'échantillons d'entraînement, le demi-volet du bas, et vous verrez maintenant les 10+ polygones de **forêt** énumérés. Tenez la touche *Shift* (Maj) enfoncée et cliquez sur le premier et le dernier exemple de polygone forestier pour sélectionner tous les exemples de **forêt**. Cliquez ensuite sur le bouton Réduire (« *Collapse* ») pour fusionner les exemples de **forêt** en un seul groupe:



8. Ensuite, dans le panneau Gestionnaire d'échantillons d'entraînement, cliquez sur **Non-Forest** et répétez les étapes 5 à 7 pour la classe d'information **Non-Forest**. Au final, vous devriez obtenir un panneau Gestionnaire d'échantillons d'entraînement qui ressemble à l'illustration ci-dessous, avec un pourcentage presque égal de pixels dans chaque classe d'information. Pour vous assurer que vous créez des classes égales, le nombre de polygones n'est pas ce que vous surveillez au fur et à mesure que vous ajoutez des exemples supplémentaires de **Non-Forest**, mais vous observez le % de pixels de la classe **Forest** et essayez de l'amener à environ 50 %. Fusionnez ensuite tous les polygones non forestiers à ce stade:

Nom de l'étudiant: _____ # d'étudiant: _____

Image Classification

Training Samples Manager : RE2016_07_20

Ottawa

- Forest
- NonForest

Class	# Samples	Pixels (%)
Forest	14	50.86
NonForest	1	2.00
NonForest	1	1.13
NonForest	1	1.34
NonForest	1	1.00
NonForest	1	2.13
NonForest	1	1.53
NonForest	1	1.14
NonForest	1	2.84
NonForest	1	5.53
NonForest	1	2.20
NonForest	1	2.05
NonForest	1	9.68
NonForest	1	3.29
NonForest	1	5.02
NonForest	1	1.67
NonForest	1	1.72
NonForest	1	1.03
NonForest	1	3.84

9. Le volet Gestionnaire d'échantillons d'entraînement ressemble maintenant à ceci:

Image Classification

Training Samples Manager : RE2016_07_20

Ottawa

- Forest
- NonForest

Class	# Samples	Pixels (%)
Forest	14	50.86
NonForest	18	49.14

10. Ensuite, nous devons enregistrer nos nouveaux polygones d'entraînement dans une classe d'objet. Pour cela, cliquez sur le bouton "Enregistrer" dans le volet du bas du gestionnaire d'échantillons d'entraînement, naviguez jusqu'à IMAGERY.GDB et enregistrez le nouveau fichier d'entraînement sous le nom "RE2016_07_20_training".

Nom de l'étudiant: _____ # d'étudiant: _____

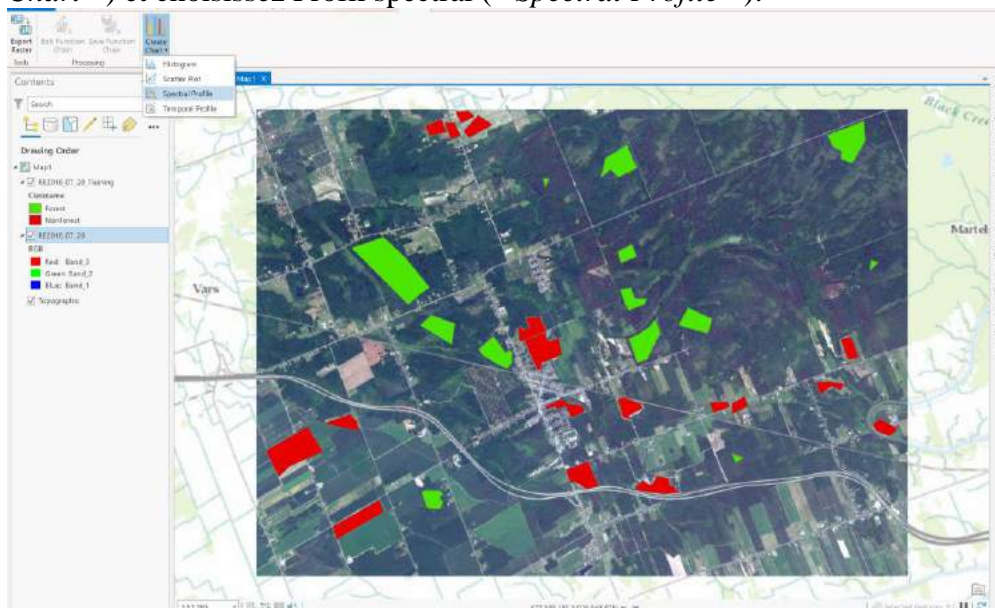
Q4: Comment créer un fichier de signature à partir de mes polygones d'entraînement?

Pour créer un fichier de signature, vous avez besoin d'un ensemble de données de polygones d'entraînement et d'une image multibande que vous souhaitez utiliser pour extraire les informations des pixels de chacune des classes d'informations d'entraînement. Vous devez ensuite évaluer à nouveau la séparabilité avant de créer le fichier de signature, car vous utilisez une nouvelle image.

1. Ajoutez la classe d'éléments "RE2016_07_20_training" de la Q3 ci-dessus au volet Contenu et symbolisez-la à l'aide de valeurs uniques et du champ **Classname**, et donnez à la classe "**Forest**" la couleur verte et à la classe "**NonForest**" la couleur rouge.

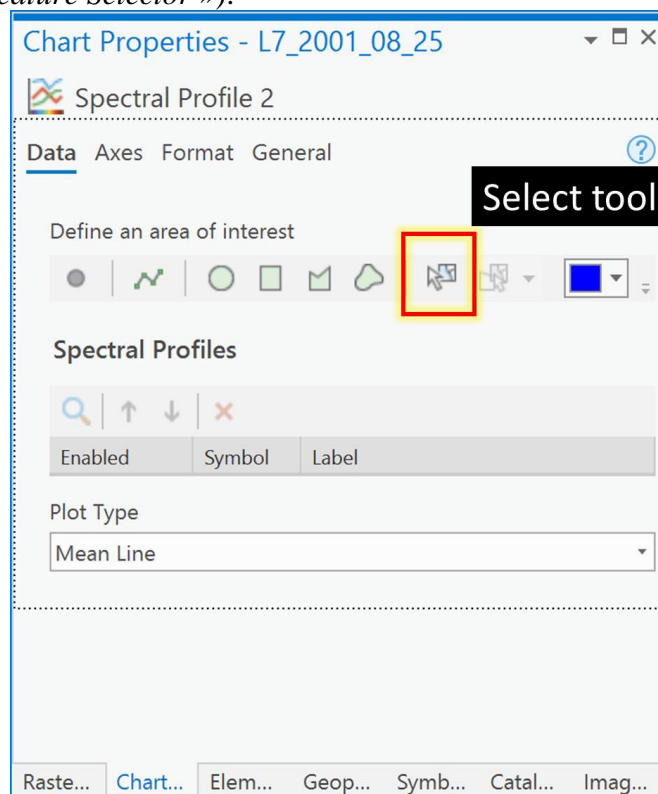


2. Ensuite, choisissez la couche "RE2016_07_20" dans le volet de contenu. Cliquez ensuite sur l'onglet Données (« Data »). Cliquez maintenant sur Créer un graphique (« Create Chart ») et choisissez Profil spectral (« Spectral Profile »):

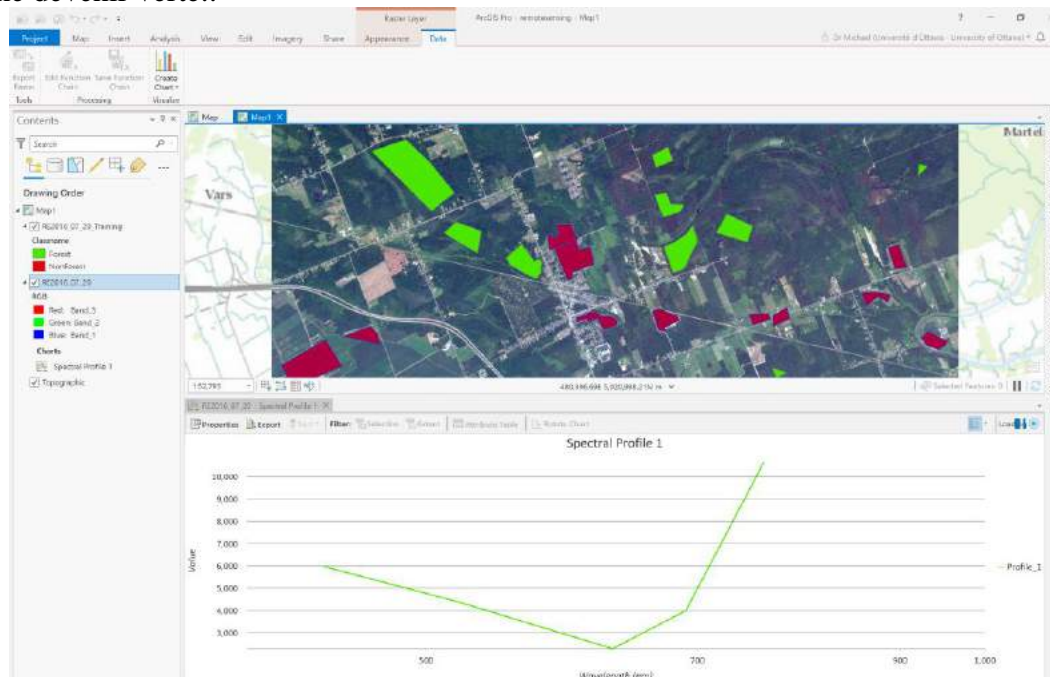


Nom de l'étudiant: _____ # d'étudiant: _____

3. Ensuite, dans le volet Propriétés du graphique (« *Chart Properties* »), cliquez sur le bouton Sélecteur d'objet (« *Feature Selector* »):

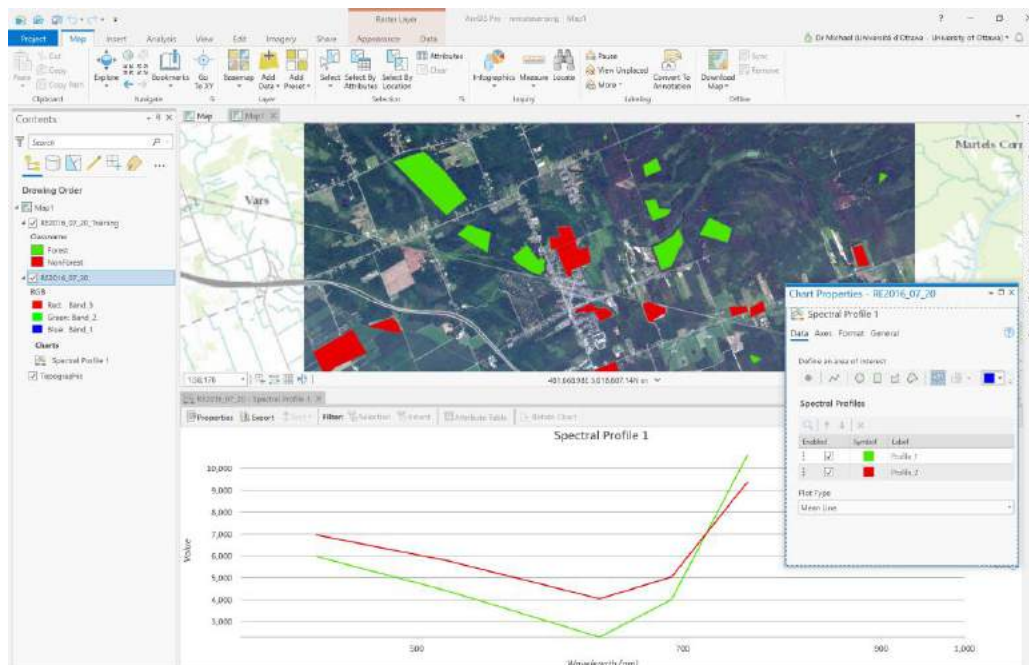


4. Cliquez maintenant sur l'un des polygones de forêt dans le volet de la carte. Vous verrez un graphique créé avec une seule ligne appelée "Profile_1" qui représente la moyenne des VN sur toutes les bandes pour cette classe d'information appelée **Forest**. Dans le volet Propriétés du graphique, changez l'étiquette de cette ligne de moyenne spectrale de "Profil_1" à **Forêt** (si cela fonctionne !) et changez la couleur en vert et vous verrez votre ligne devenir verte.:

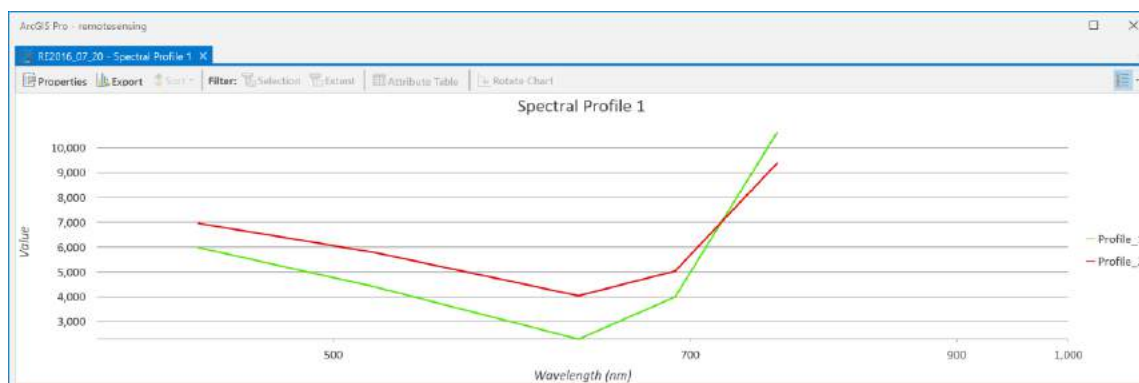


Nom de l'étudiant: _____ # d'étudiant: _____

- Répétez les étapes 3-4 pour les polygones d'entraînement **NonForest**, et faites en sorte que la nouvelle ligne soit rouge pour correspondre à **NonForest**. Renommez le label si vous le pouvez dans votre version d'ArcGIS Pro:



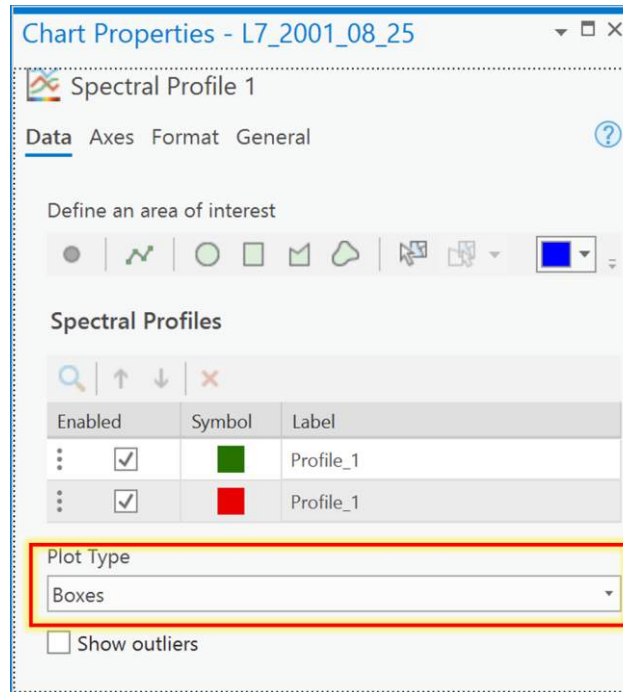
- Examinez maintenant le graphique du profil spectral:



Tout comme dans Q2, ce graphique montre la longueur d'onde sur l'axe des x plutôt que les noms des bandes, parce que les métadonnées de l'image contiennent des informations sur la longueur d'onde de chaque bande. Les VN moyens montrent des différences constantes, mais il est difficile de juger de la séparabilité à partir des seules moyennes. Nous devons donc examiner les diagrammes en « boîtes et moustaches » pour mieux choisir les bandes dont les boîtes ne se chevauchent pas et qui présentent donc une bonne séparabilité.

Nom de l'étudiant: _____ # d'étudiant: _____

7. Pour créer le graphique en boîte et en moustache de la séparabilité spectrale pour chaque bande, il suffit de modifier le type de graphique dans le volet Propriétés du graphique (« *Chart Properties* »). Dans le volet, changez le type de tracé (« *Plot Type* ») de moyenne (« *Mean* ») à boîtes (« *Boxes* »):



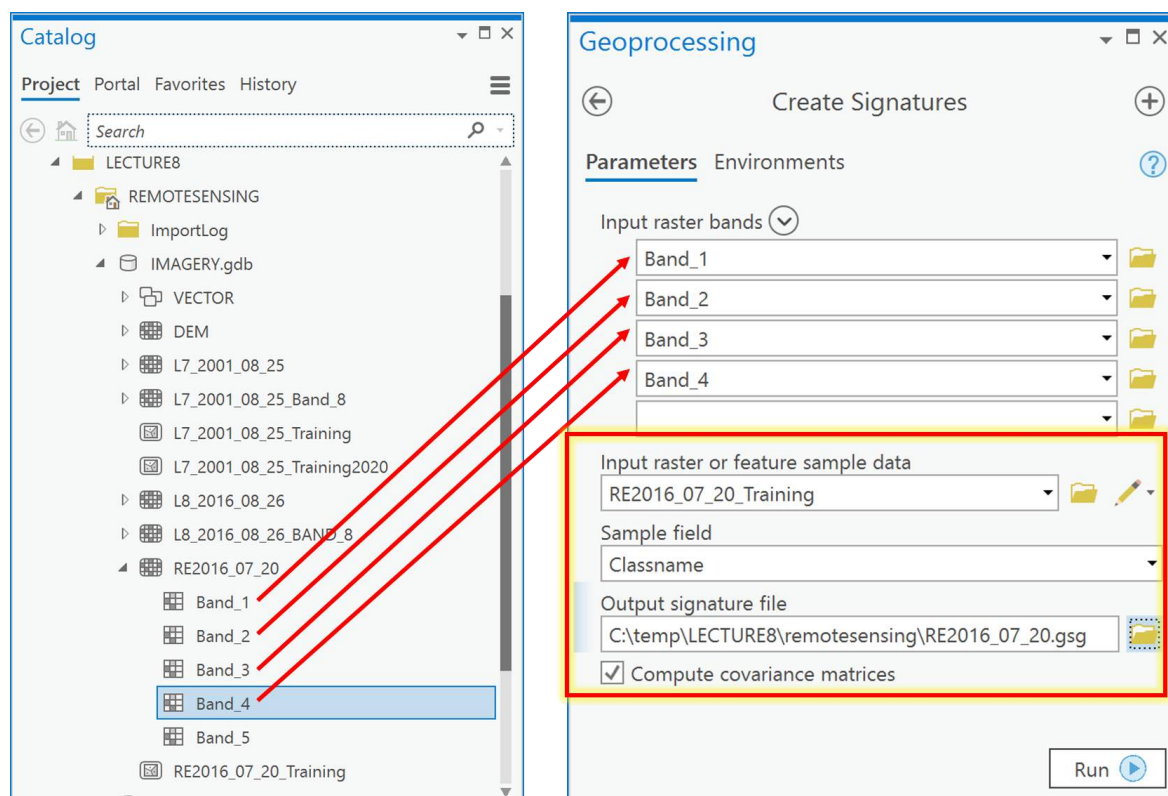
8. Examinez maintenant le diagramme en « boîtes et moustaches ».



Nom de l'étudiant: _____ # d'étudiant: _____

La plupart des bandes présentent une bonne séparabilité, à l'exception de la Bande_5 où les cases se chevauchent. La largeur de bande spectrale de la Bande_5 sur ce capteur RapidEye est de 760-850 nm, ce qui correspond à la partie proche de l'infrarouge du spectre. C'est dans cette même partie du spectre que l'image "L7_2001_08_25" a montré une faible séparabilité entre forêt et non forêt (dans la bande 4, qui couvre 750-900 nm pour le capteur Landsat). Ainsi, pour notre image actuelle, la bande 5 sera exclue de toute classification des **forêts** et des zones **non forestières**.

9. Créez un fichier de signatures à partir des sites d'entraînement que nous pourrions utiliser pour la classification. Pour créer le fichier, allez dans l'onglet Analyse (« *Analysis* »), cliquez sur Outils (« *Tools* ») et recherchez l'outil Créer des signatures (« *Create Signatures* ») qui se trouve dans la boîte à outils *Spatial Analysis*. Désactivez l'outil *Create Signatures* afin de pouvoir le voir en même temps que votre volet Catalogue. Ensuite, dans le volet *Catalog*, naviguez jusqu'à IMAGERY.GDB et agrandissez l'image "RE2016_07_20" de manière à voir toutes les bandes individuelles (Band_1, Band_2,...,Band_5). Faites glisser toutes les bandes à l'exception de la Bande_5 (puisque nous savons qu'elle a une mauvaise séparabilité) dans "Input raster bands", sous "Input raster or feature sample data" choisissez la classe objet "RE2016_07_20_training", sous "Sample field" choisissez **Classname**, et sauvegardez le fichier dans le dossier où vous avez vos données, et appelez-le "RE2016_07_20.gsg". Cliquez sur "Run" (Exécuter).



10. Vous avez maintenant un fichier de signatures que vous pouvez utiliser pour la classification des images.

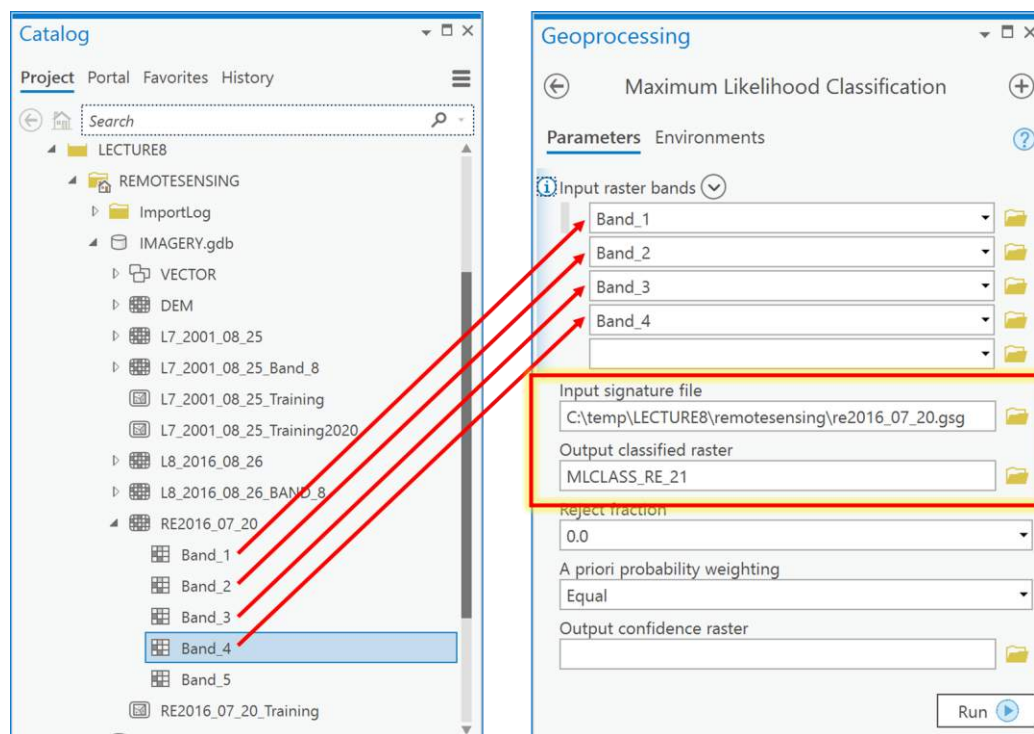
Nom de l'étudiant: _____ # d'étudiant: _____

Q5: Comment puis-je classer mon image en utilisant la classification d'image supervisée?

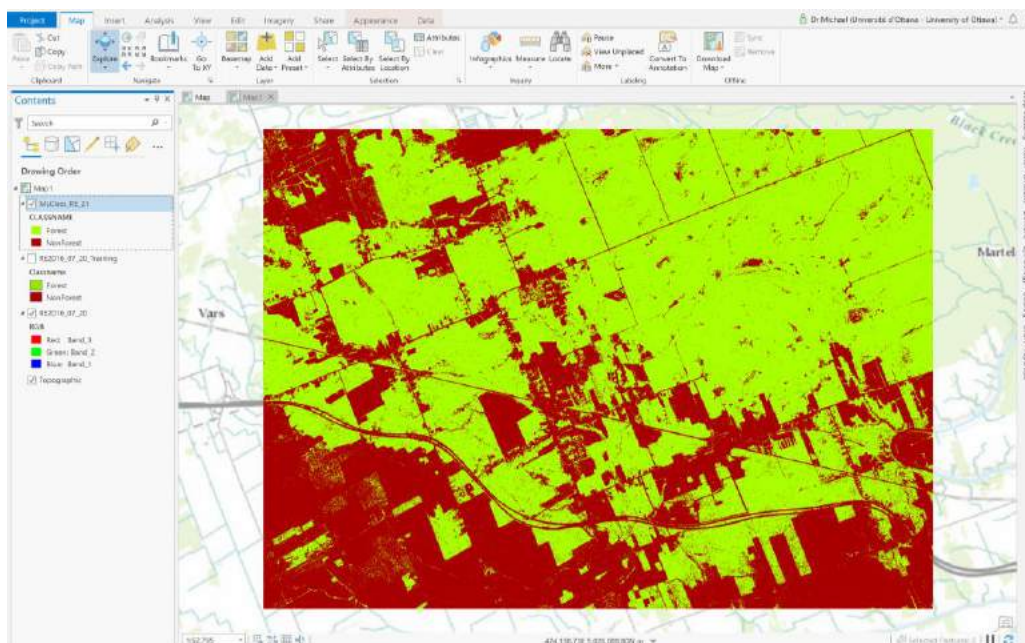
Pour classer une image, vous devez avoir répondu aux questions ci-dessus car vous avez besoin du fichier de signature "RE2016_07_20.gsg" pour l'image RapidEye "RE2016_07_20" afin de procéder à une classification supervisée.

Maintenant que vous avez obtenu un fichier de signatures à partir de l'exemple précédent, vous pouvez procéder à une classification supervisée:

1. Utilisez maintenant le fichier de signatures pour classer l'image dans les deux classes d'information **Forêt** et **Non-Forêt**. Pour cela, recherchez maintenant l'outil d'analyse spatiale "Maximum Likelihood Classification" (Classification par maximum de vraisemblance). Désancrez encore l'outil afin de pouvoir le voir en même temps que votre volet Catalogue.
2. Ensuite, dans le volet Catalogue, naviguez jusqu'à IMAGERY.GDB et agrandissez l'image "RE2016_07_20" de manière à voir toutes les bandes individuelles (Bande_1, Bande_2, ..., Bande_5). Faites glisser toutes les bandes à l'**exception** de la Bande_5 dans "Input raster bands", sous "Input signature file", choisissez le fichier de signature créé durant l'exemple précédente appelé "RE2016_07_20.gsg", sous "Output classified raster" appelez-le "MLCLASS_RE_21", et enregistrez-le dans IMAGERY.GDB. Laissez toutes les autres valeurs par défaut et cliquez sur *Run* (Exécuter):

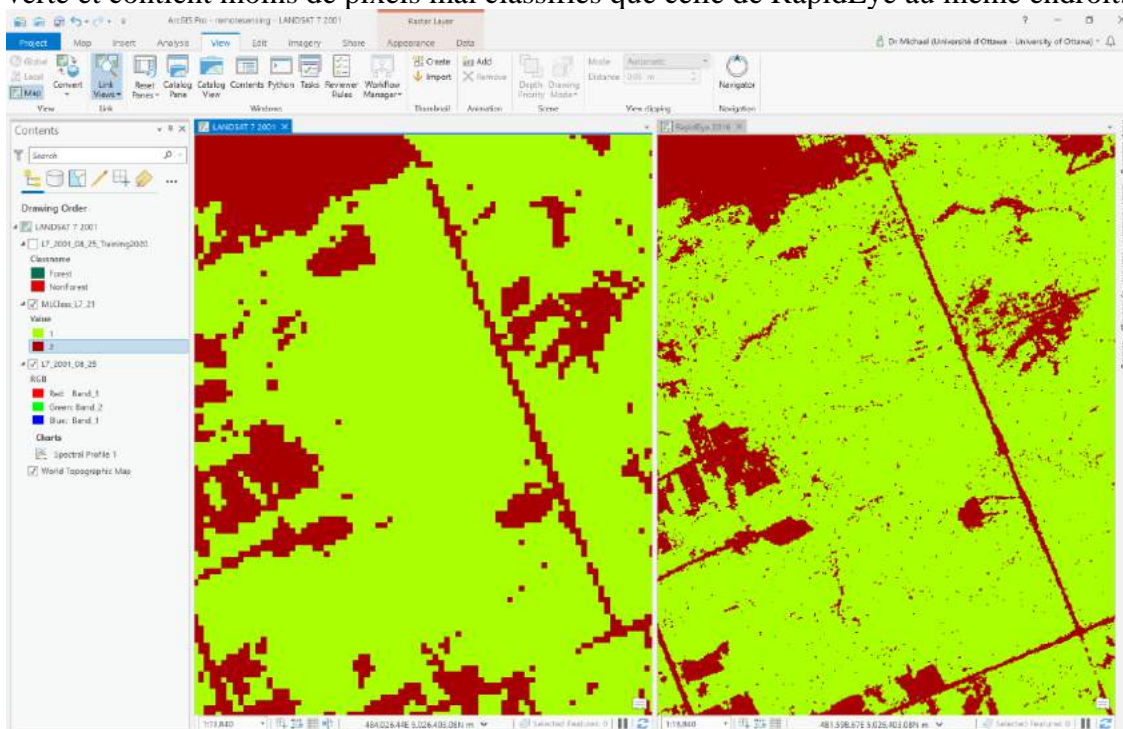


3. Examinez la couche raster classée résultante (notez que vous pouvez changer la valeur 1 - **Forest** en vert et la valeur 2 - **NonForest** en rouge). Nous avons maintenant converti l'image en une couche raster thématique.



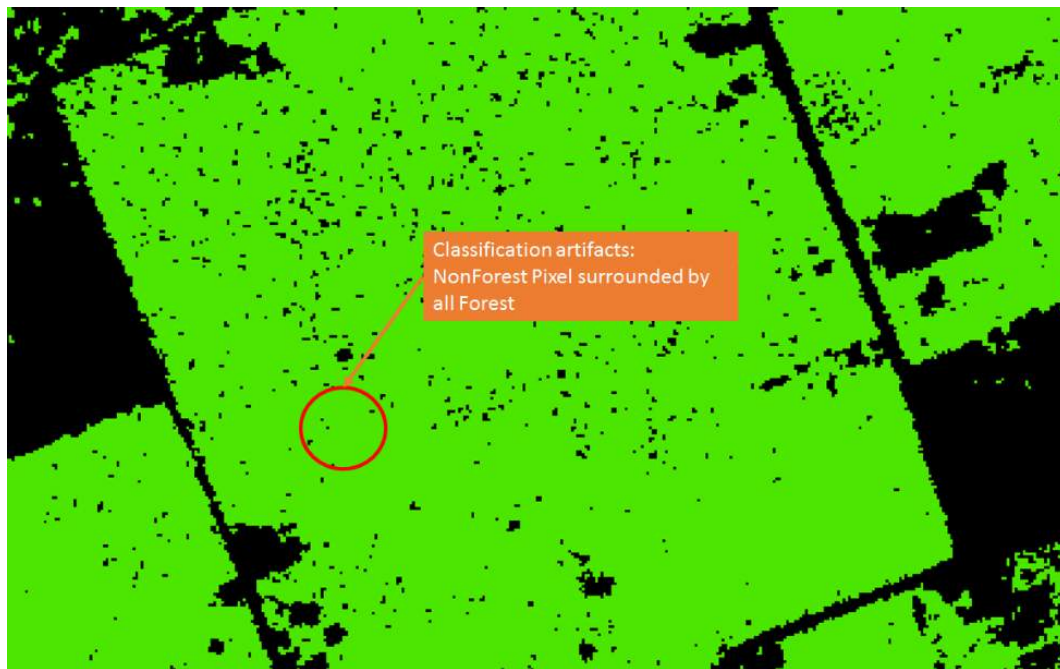
Q6: Comment puis-je nettoyer ma couche raster classifiée?

Dans le résultat classifié, vous remarquerez de nombreux artefacts du processus de classification; ceux-ci sont particulièrement évidents lorsque des pixels de **forêt** sont entourés de pixels **non forestiers** ou vice-versa. En règle générale, plus la résolution de l'image est élevée (plus les pixels sont petits), plus il y a de pixels mal classifiés. Étant donné que la couche Landsat 7 avait une résolution de 30 m et que la couche RapidEye avait une résolution de 5 m, il n'est pas surprenant qu'il y ait plus de pixels mal classifiés dans la carte réalisée à partir de l'image RapidEye. Par exemple, en regardant les résultats de classification de Landsat et de RapidEye côte à côte (comme indiqué ci-dessous), nous pouvons voir que la **forêt** de Landsat 7 est plus verte et contient moins de pixels mal classifiés que celle de RapidEye au même endroit.



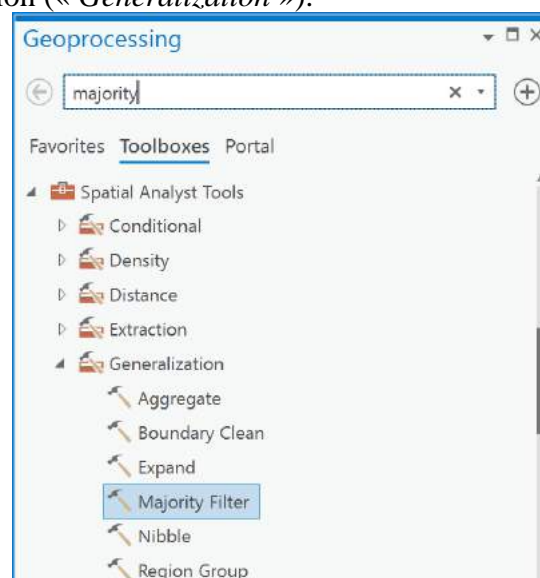
Nom de l'étudiant: _____ # d'étudiant: _____

Il est peu probable que des pixels non forestiers entourés de forêts représentent une perte de forêt sur ce pixel dans la majorité des cas. Cela dépend en partie de la définition de ce qu'est une "forêt" !!? Bien qu'il ne s'agisse pas d'une définition formelle, beaucoup considèrent qu'une forêt est une zone dominée par des arbres - mais cela ne signifie pas qu'il doit y avoir des arbres partout ! Les forêts peuvent avoir des clairières sans arbres, mais ces clairières font toujours partie de la forêt. Avec l'imagerie RapidEye, de telles clairières, ou des zones qui ne contiennent qu'un petit arbre, ou un arbre malsain, peuvent avoir été mal classifiées comme non forestières (voir les exemples dans la figure ci-dessous).



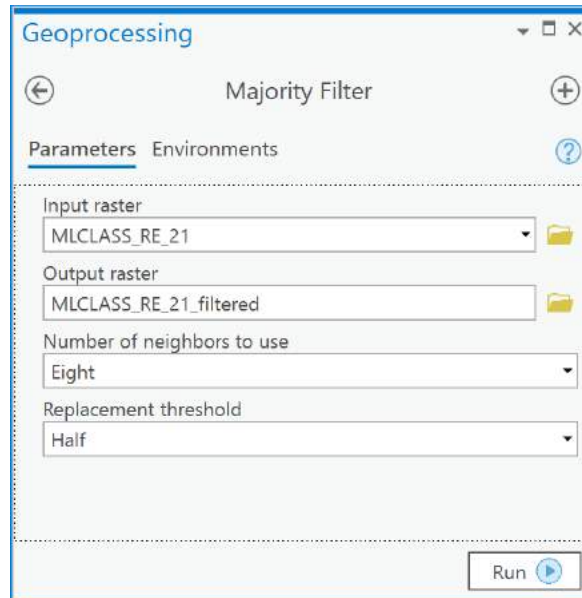
Nous pouvons filtrer certains des pixels erronés à l'aide d'un filtre majoritaire:

1. Ouvrez l'outil Filtre de majorité (« *Majority Filter* ») sous Outils (« *Tools* ») d'analyse spatiale -> Généralisation (« *Generalization* »):



Nom de l'étudiant: _____ # d'étudiant: _____

2. Dans le menu déroulant « Input raster », sélectionnez votre couche raster classifiée "MLCLASS_RE_21". Appelez la sortie "MLCLASS_RE_21_filtered" et enregistrez-la dans IMAGERY.gdb. Choisissez le nombre de voisins à utiliser (optionnel) comme huit voisins, et cliquez sur *Run*.



3. Le résultat est une couche classifiée plus propre avec moins de pixels mal classifiés/abhorrés:

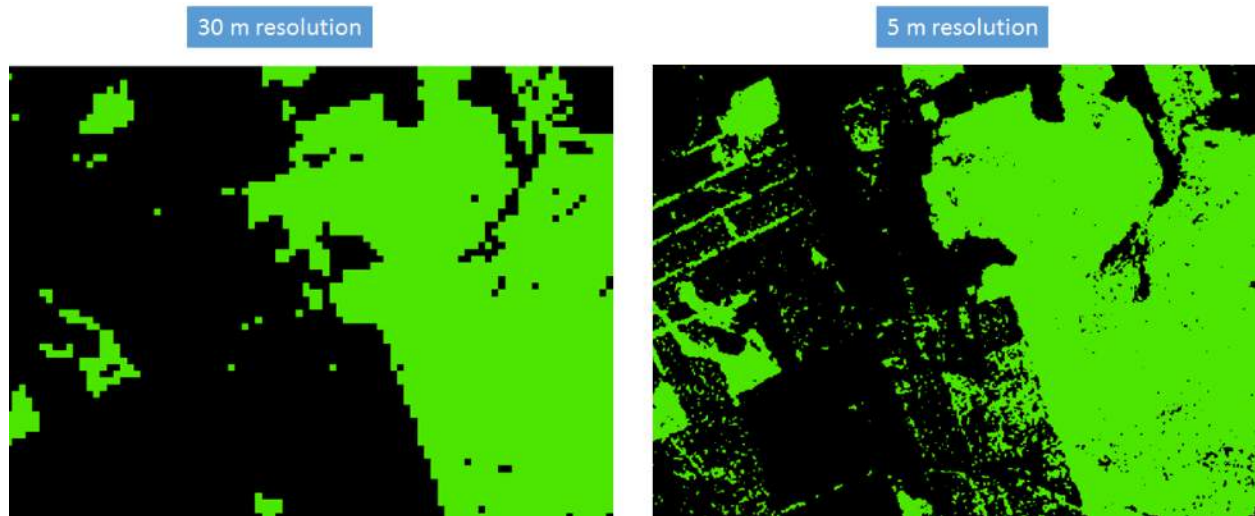


Il est parfois nécessaire d'exécuter plusieurs fois le filtre de la majorité (ou des outils similaires) pour obtenir le résultat au niveau de généralisation requis pour un objectif particulier.

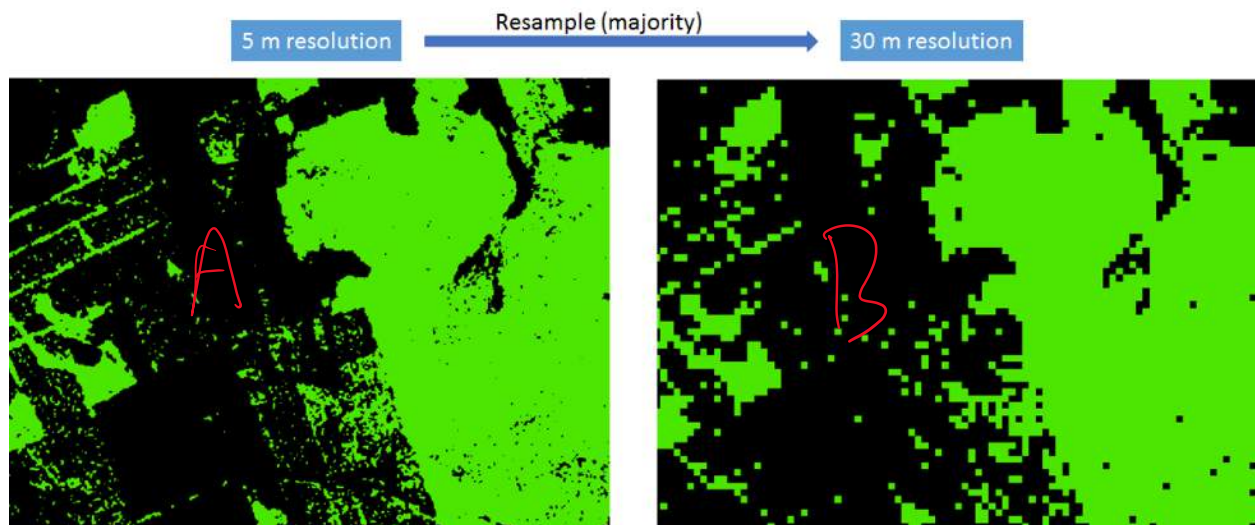
Q7: Comment généraliser mon raster classifié "MLCLASS_RE_21_filtered" pour que la taille des cellules soit la même que celle de la couche "MLCLASS_L7_21" de Q2?

La scène Landsat 7 qui a été classifiée dans Q2 est dans la couche appelée "MLCLASS_L7_21" et a une résolution spatiale de 30 mètres (c'est-à-dire que chaque cellule est de 30 x 30 mètres). Votre couche "MLCLASS_RE_21_filtered" a la même résolution que l'image RapidEye qui est de 5 m (voir comparaison ci-dessous).

Nom de l'étudiant: _____ # d'étudiant: _____

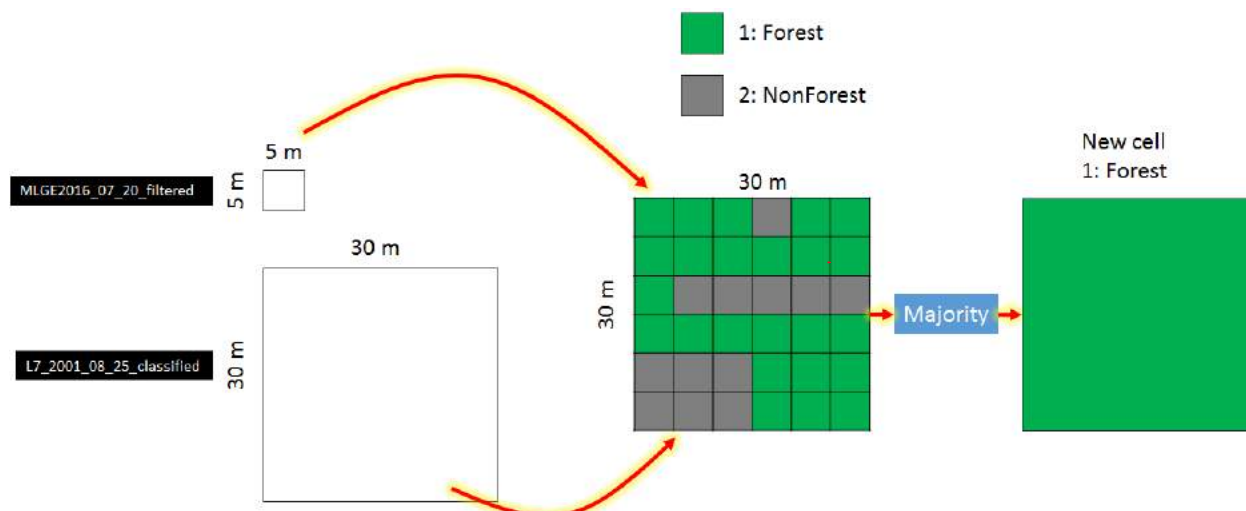


Comme il y a 36 cellules dans la couche "MLCLASS_RE_21_filtered" pour chaque cellule dans "MLCLASS_L7_21", vous augmenterez la taille des cellules de "MLCLASS_RE_21_filtered" pour qu'elle corresponde à la taille des cellules de "MLCLASS_L7_21".



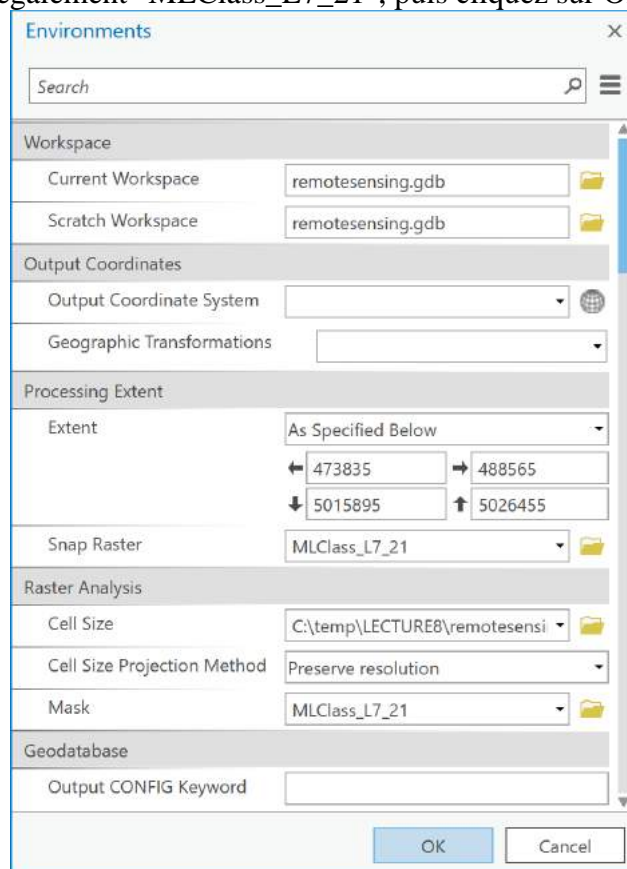
Vous devez donc indiquer à ArcGIS comment résumer les cellules "MLCLASS_RE_21_filtered" lorsque vous augmentez la taille des cellules de cette couche. En d'autres mots, dans chaque nouvelle cellule de 30 x 30 m, ArcGIS doit regarder les 36 cellules plus petites qui la composent et doit décider comment calculer la valeur (Forêt ou Non-Forêt) à l'intérieur de la nouvelle taille de cellule. Comme il s'agit de données nominales (Forêt et Non-Forêt), ArcGIS prend par défaut la valeur de la cellule de 5 x 5 m la plus proche du centre de la nouvelle cellule de 30 x 30 m. Cependant, avec des données nominales, une fonction d'agrégation plus appropriée serait que la nouvelle cellule de 30 x 30 m prenne une valeur correspondant à la majorité des valeurs qu'elle contient. Par exemple, s'il y a 24 cellules forestières et 12 cellules non forestières dans la nouvelle cellule de 30 m, la nouvelle cellule à 30 m doit être une cellule forestière, comme dans l'exemple ci-dessous.

Nom de l'étudiant: _____ # d'étudiant: _____



Lorsque vous généralisez votre couche raster à la même taille de cellule que le résultat de la classification Landsat 7, vous devez également vous assurer que les cellules de la couche généralisée s'alignent correctement sur les cellules de la couche Landsat 7. Pour ce faire, vous devez définir le paramètre « *Snap Raster* » dans les paramètres de l'environnement global ainsi que le paramètre « *Processing Extent* »:

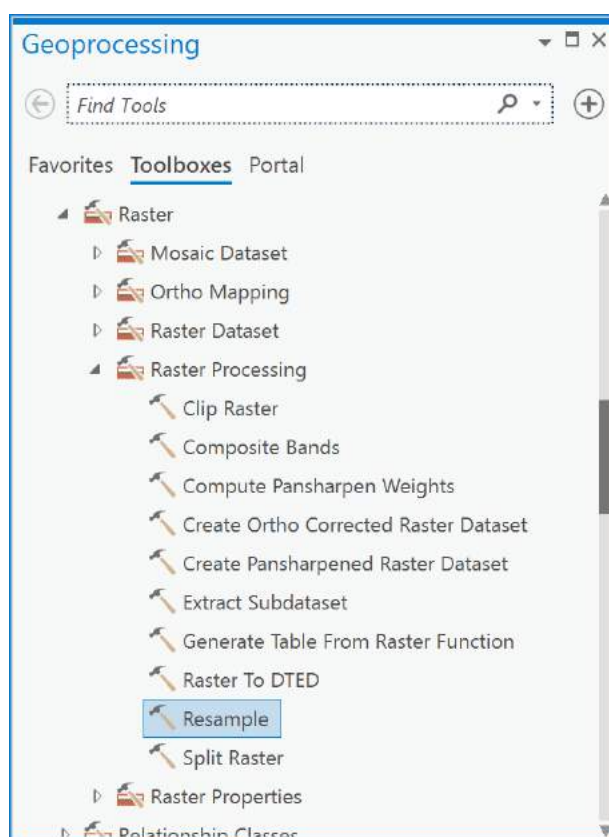
1. Dans la boîte de dialogue « *Environment Settings* », sous « *Processing Extent* », choisissez "MLClass_L7_21" et sous « *Snap Raster* », choisissez également "MLClass_L7_21". Ensuite, sous « *Raster Analysis* », « *Cell Size* », choisissez "MLClass_L7_21" et sous « *Mask* », choisissez également "MLClass_L7_21", puis cliquez sur OK:



Nom de l'étudiant: _____ # d'étudiant: _____

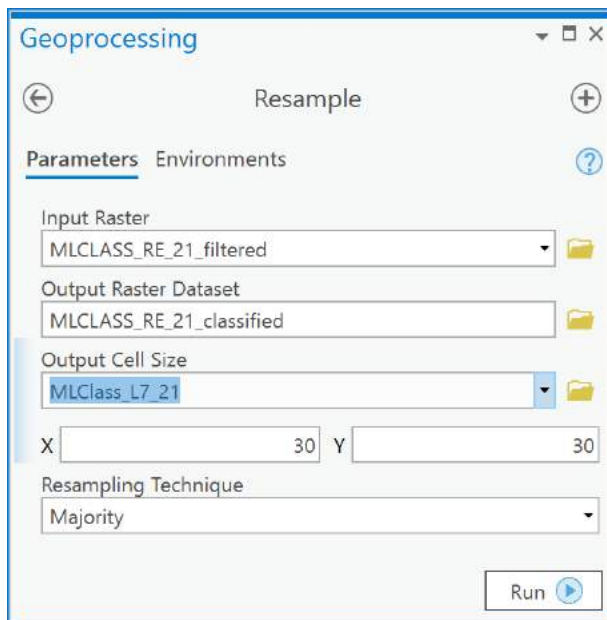
En complétant l'étape 2, vous vous êtes assuré que tout processus de traitement dans l'analyse raster sera conforme à la taille des cellules et à l'étendue de la couche "MLCLASS_L7_21". N'oubliez pas que ces paramètres seront conservés jusqu'à ce que vous fermiez et rouvriez ArcGIS ou que vous les modifiiez à nouveau. Ainsi, si vous décidez d'effectuer d'autres analyses, avec d'autres couches, dont l'étendue est différente de "MLCLASS_L7_21", les résultats ne seront produits qu'à l'intérieur de l'étendue spécifiée. Il se peut donc que vous deviez effacer la propriété d'étendue ou la ramener à sa valeur par défaut lorsque vous aurez terminé les analyses raster.

2. Pour convertir le produit RapidEye à une résolution de 30 m, vous utilisez l'outil « *Resample* » sous « *Data Management* » -> « *Raster* » -> « *Raster Processing* ». Ouvrez cet outil:

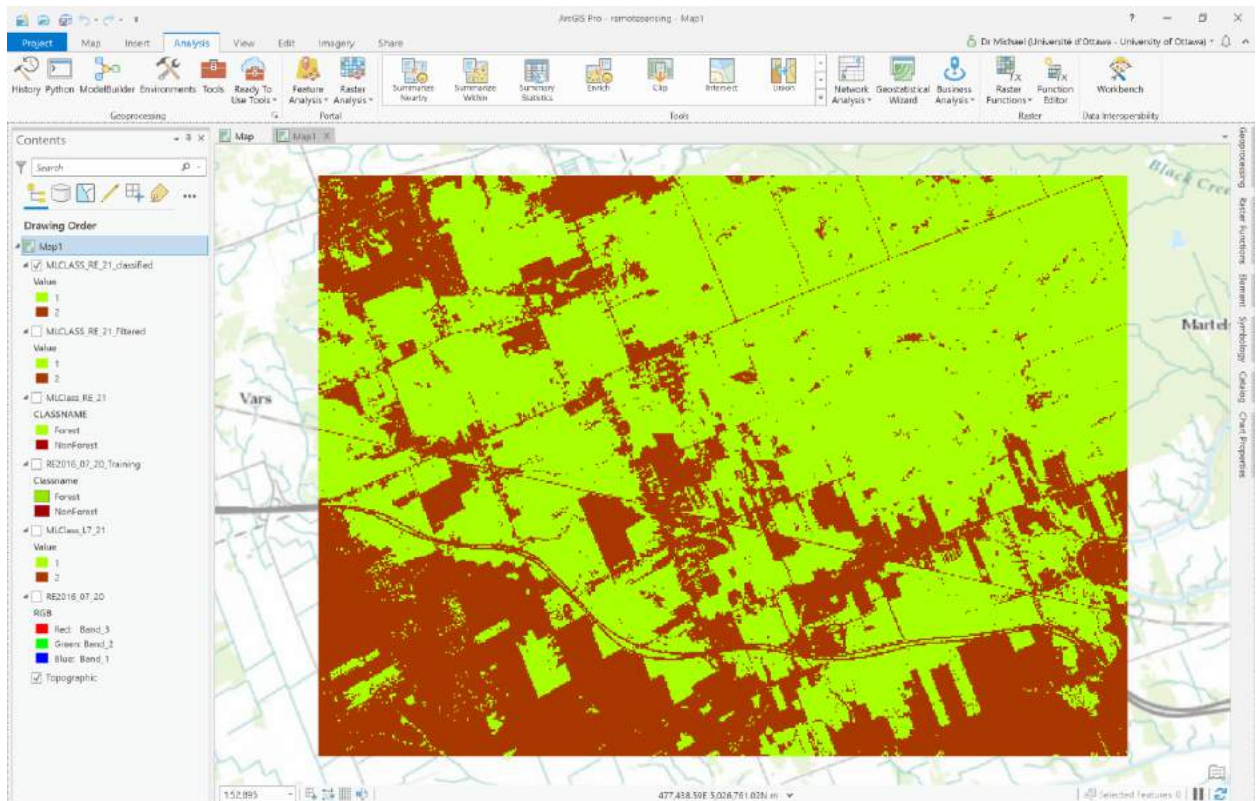


3. Dans la boîte de dialogue « *Resample* », votre raster d'entrée (« *Input Raster* ») doit être spécifié comme "MLCLASS_RE_21_filtered", le raster de sortie (« *Output Raster Dataset* ») doit être appelé "MLCLASS_RE_21_classified" et la taille de cellule de sortie (optionnelle) doit être spécifiée "MLClass_L7_21". **Plus important encore**, la technique de rééchantillonnage facultative (« *Resampling Technique* ») doit être réglée sur MAJORITÉ (« *Majority* »). Cliquez ensuite sur OK.

Nom de l'étudiant: _____ # d'étudiant: _____



4. Votre résultat agrégé ressemblera à ceci:

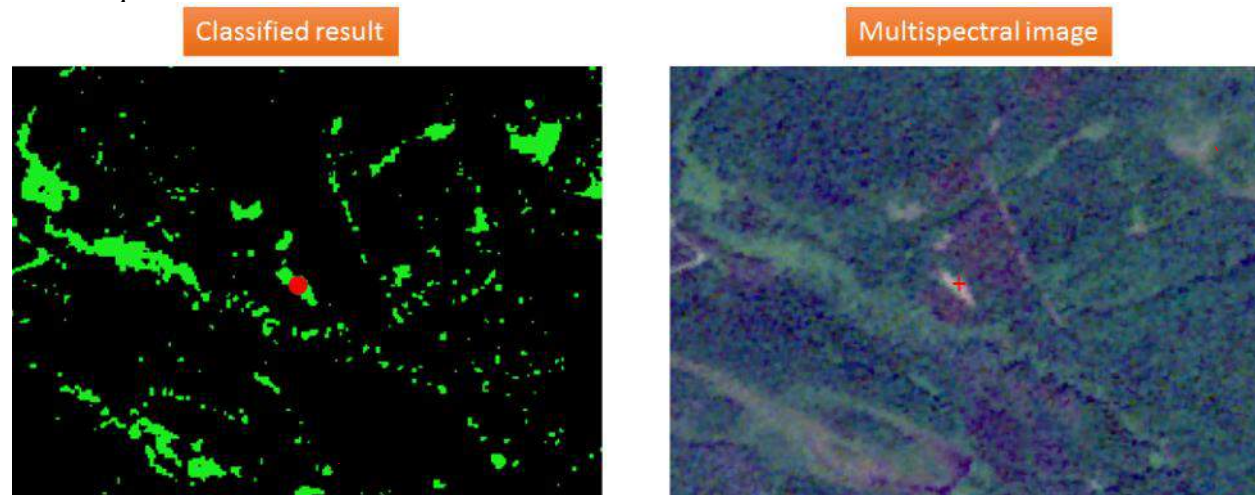


Nom de l'étudiant: _____ # d'étudiant: _____

Q8: Comment puis-je vérifier l'exactitude d'un résultat classifié?

La précision d'une classification est testée en comparant les pixels classifiés à ce que l'on appelle les données de référence. Les données de vérité terrain (VT) sont des lieux qui n'ont pas été utilisés dans la classification mais dont on sait qu'ils appartiennent à l'une des classes d'information que vous avez définies, c'est-à-dire qu'ils appartiennent à la catégorie Forêt ou Non-Forêt. En d'autres termes, les données VT constituent un "ensemble de données de test" et sont indépendantes des données utilisées pour créer le fichier de signatures qui a entraîné le classificateur. Les données VT peuvent être acquises en allant sur le terrain avec une unité GPS et en observant plusieurs endroits pour détecter la présence d'une classe d'information particulière. Toutefois, sauf nécessité absolue, d'autres méthodes de collecte de données VT indépendantes sont généralement utilisées, car il est souvent difficile, lent et coûteux d'acquérir des données VT sur place. Ainsi, vous pouvez généralement collecter des exemples de données VT en observant directement les caractéristiques de l'image pré-classifiée et en les étiquetant comme l'une des classes d'information, dans notre cas comme Forêt ou Non-Forêt. Pour évaluer la précision d'une classification, vous devez collecter ces données en plusieurs points choisis au hasard et calculer une matrice de confusion qui vous permettra de quantifier le nombre de fois où la classification est correcte et le nombre de fois où elle ne l'est pas.

Note : Le terme "vérité de terrain" est mal vu par certains, car il implique que ces données sont intrinsèquement "vraies". Ce n'est évidemment pas le cas, car ces données sont produites par des humains, et les humains font des erreurs. Néanmoins, nous utilisons ce terme ici parce qu'il est courant et que vous le rencontrerez à l'avenir si vous travaillez davantage avec l'imagerie satellitaire. D'autres termes utilisés pour la même chose sont "données de validation" ou "données de test", mais ces termes eux-mêmes peuvent avoir des significations différentes pour différentes personnes, et ne sont donc pas idéaux non plus. Nous nous en tiendrons à "vérité terrain" pour cet exercice !



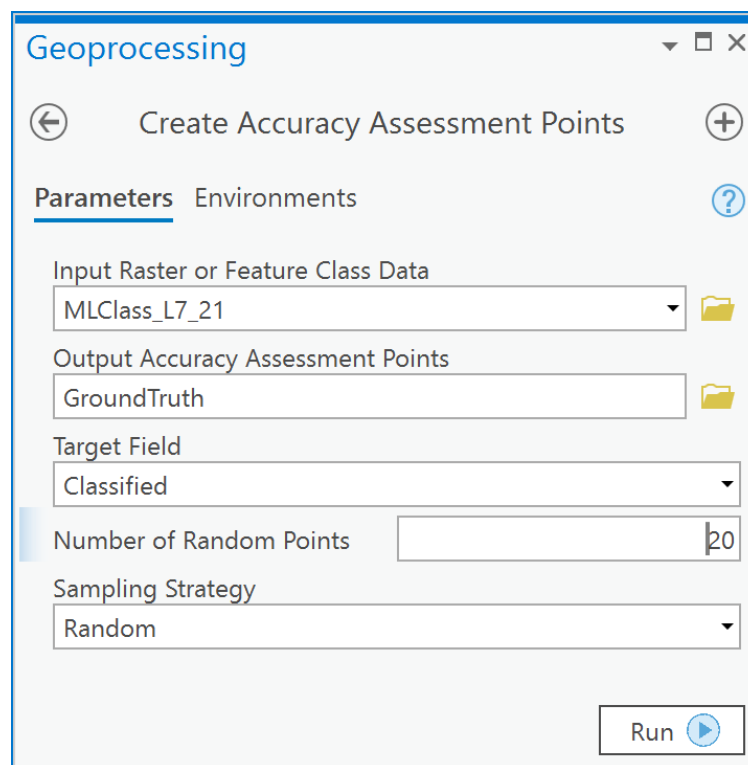
Exemple de classification erronée. Le pixel classé comme forêt (à gauche) n'est manifestement pas une forêt (à droite), comme le montre un point de vérité terrain (croix rouge, à droite) identifié manuellement à partir d'une image multispectrale.

Vous calculerez la précision du résultat "MLCLASS_L7_21", puisque tout le monde aura le même résultat à partir de la question 2 ci-dessus.

Nom de l'étudiant: _____ # d'étudiant: _____

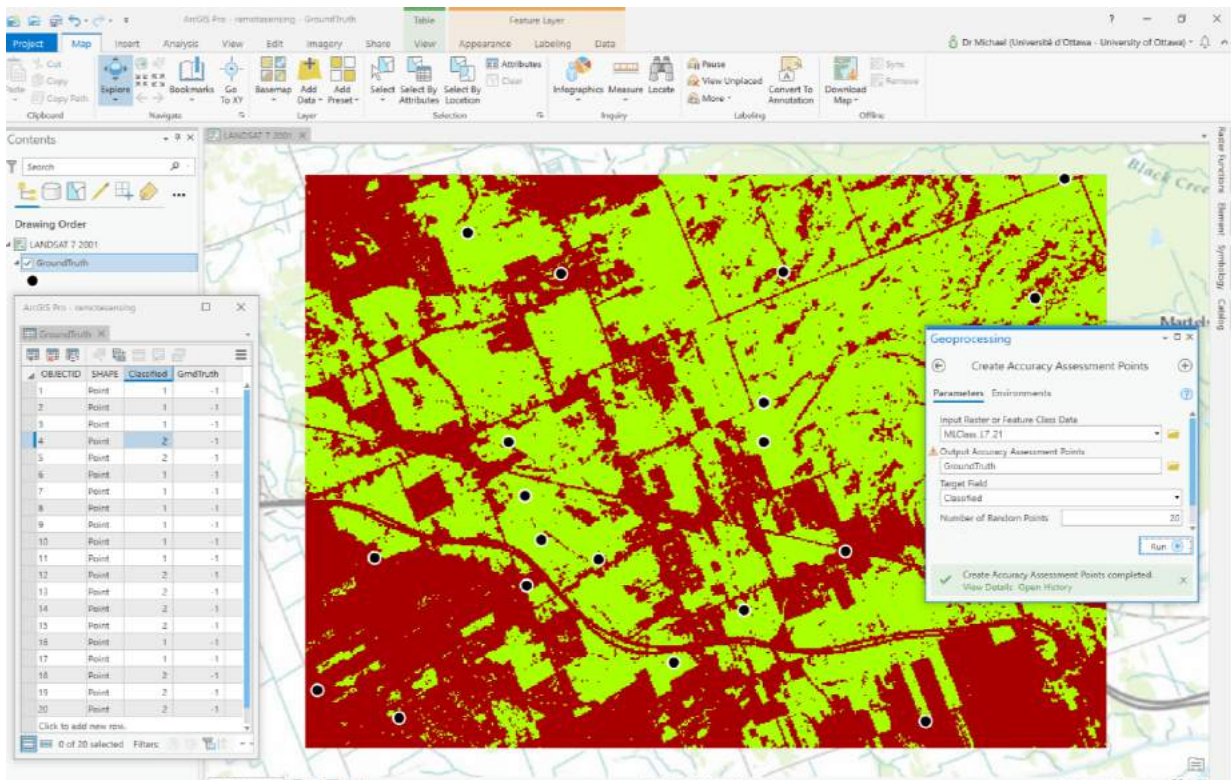
1. Allez dans la boîte à outils, trouvez « *Spatial Analyst* » -> « *Segmentation and Classification* » et choisissez l'outil « **Create Accuracy Assessment Points** » (Créer des points d'évaluation de précision).
2. Dans l'outil, sous « *Input raster or feature class data* », choisissez la couche "MLCLASS_L7_21" et sous « *Output Accuracy Assessment Points* », appelez le résultat "GroundTruth" et assurez-vous de l'enregistrer dans la géodatabase IMAGERY.gdb. Sous « *Target Field* », choisissez « *Classified* ». Sous « *Number of Random Points* » (nombre de points aléatoires), choisissez 20, et enfin sous « *Sampling Strategy* » (stratégie d'échantillonnage), choisissez « *Random* » (aléatoire).

Cela créera 20 points aléatoires (où les coordonnées x et y de chaque point sont choisies, dans l'étendue de l'image, à partir de la fonction de densité de probabilité uniforme, indépendamment et avec une probabilité égale) dans chacune des classes "Forêt" et "Non-Forêt". Pour chaque point aléatoire, la valeur du pixel de la couche "MLCLASS_L7_21" sera extraite à l'emplacement du point dans la couche "MLCLASS_L7_21". Ces valeurs seront stockées dans la table « *GroundTruth* » dans une colonne appelée « *Classified* » contenant les valeurs de pixel classées pour chacun des points.



3. Le résultat sera 20 points situés de manière aléatoire dans l'image:

Nom de l'étudiant: _____ # d'étudiant: _____



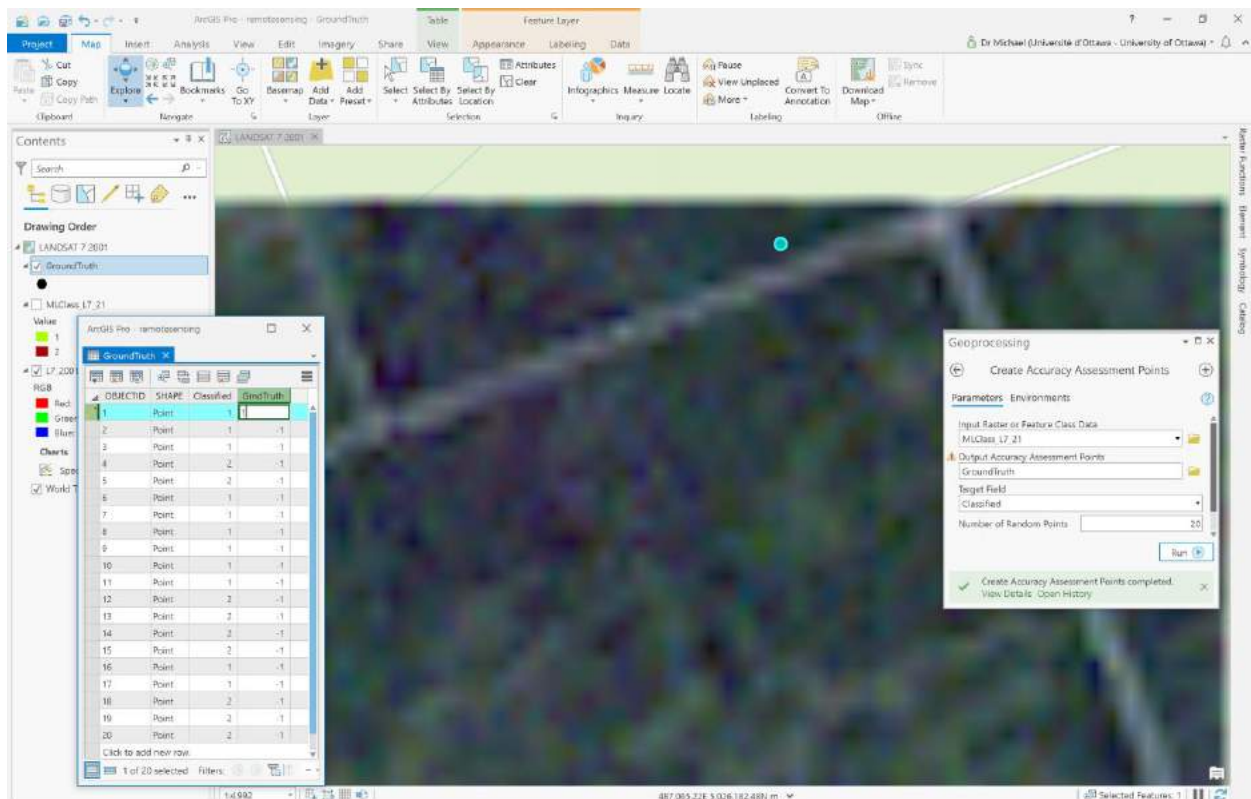
4. Ouvrez maintenant le tableau appelé « *GroundTruth* ». Il devrait ressembler à ceci:

The screenshot shows the 'GroundTruth' table in ArcGIS Pro. The table has columns for OBJECTID, SHAPE, Classified, and GrndTruth. The data is as follows:

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

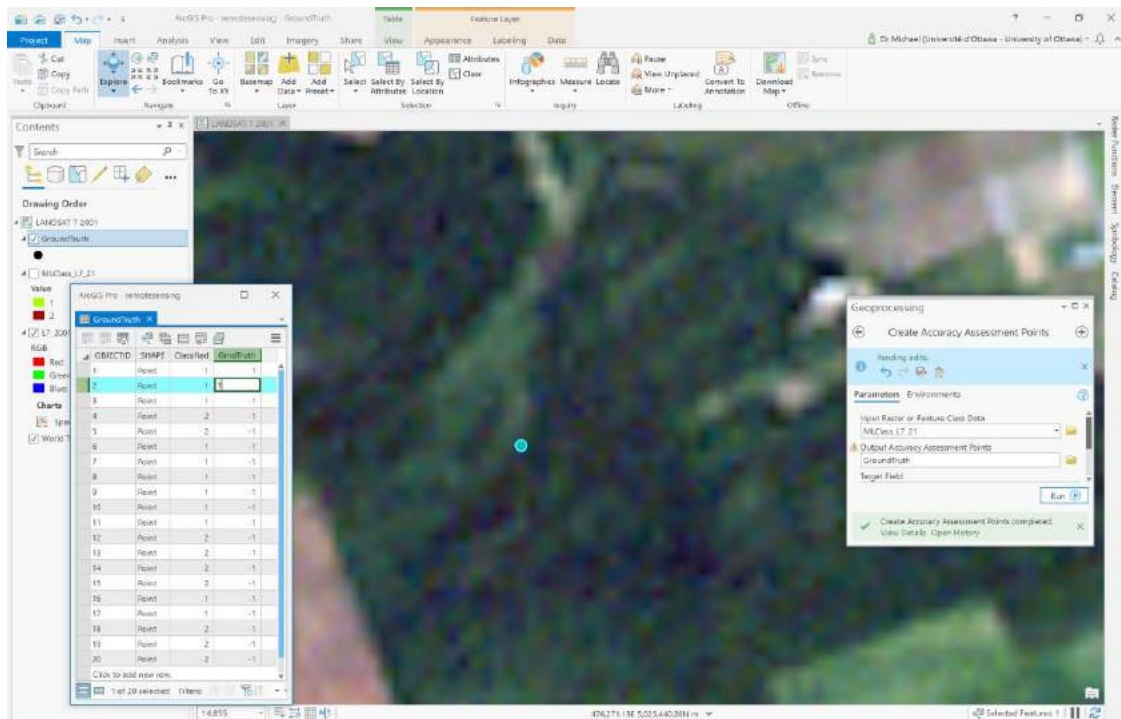
Nom de l'étudiant: _____ # d'étudiant: _____

- Chaque point aléatoire est associé à la valeur de la classe d'information Forêt = 1 ou Non-Forêt = 2, en fonction de la classe dans laquelle se trouve le point. Ces données figureront dans la colonne « Classified » du tableau d'attributs « GroundTruth » (voir le tableau ci-dessus).
- Nous devons maintenant remplir la colonne **GrndTruth** du tableau, dont les valeurs sont actuellement de -1. Les valeurs de -1 nous indiquent que la "vraie" classe pour ce point n'a pas été observée et entrée dans le tableau. Nous devons donc remplir ces "vraies" valeurs de classe, en nous rappelant que la valeur 1 signifie Forêt et la valeur 2 signifie Non-Forêt. Pour cela, assurez-vous que les seules couches visibles sur la carte sont "L7_2001_08_25" et « GroundTruth ». Sélectionnez maintenant la première ligne du tableau "GroundTruth", afin d'identifier le premier point. Pour la ligne actuellement sélectionnée, cliquez avec le bouton droit de la souris et choisissez « Zoom To ». Zoomez ensuite un peu plus près jusqu'à ce que vous puissiez voir clairement dans quelle classe d'information se trouve le point sélectionné. Saisissez ensuite un 1 ou un 2 dans la colonne **GrndTruth**, en fonction de votre interprétation de l'image à ce point. Dans l'exemple ci-dessous, vous pouvez voir que la forêt se trouve sous le point sélectionné, vous devez donc saisir une valeur de 1 dans la colonne GrndTruth pour cette ligne double-cliquant sur la cellule où vous souhaitez inscrire le 1.



- Sélectionnez maintenant la deuxième rangée, "Zoom To", et examinez s'il y a de la forêt ou non à cet endroit et entrez une valeur de 1 ou 2, respectivement, dans la colonne "GrndTruth" du tableau "GroundTruth". Dans l'exemple ci-dessous, il s'agit à nouveau d'une forêt, vous devez donc entrer une valeur de 1:

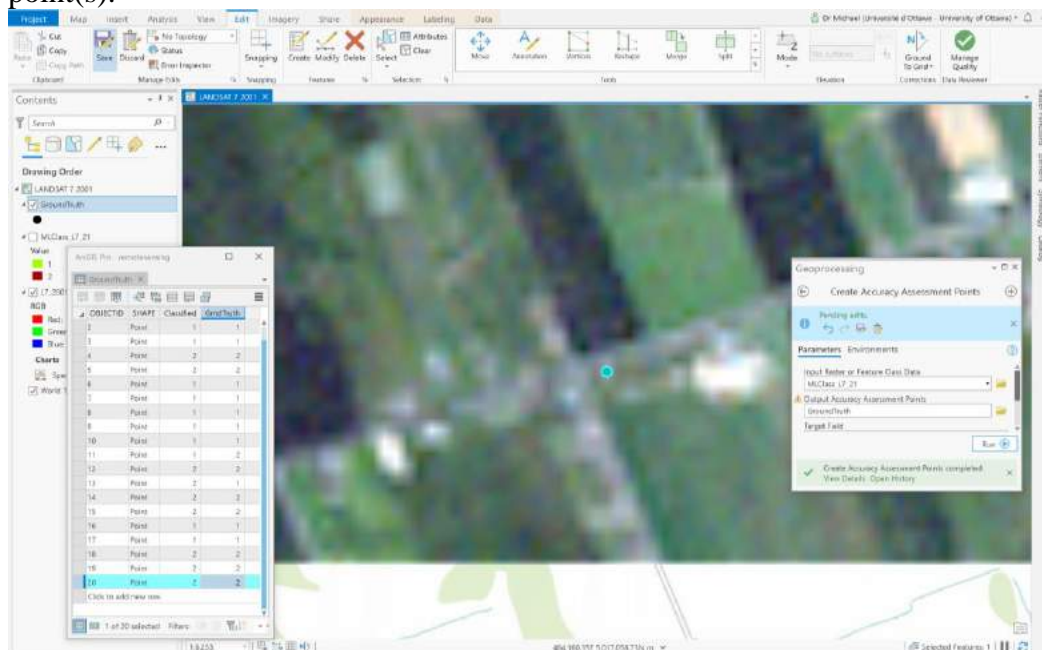
Nom de l'étudiant: _____ # d'étudiant: _____



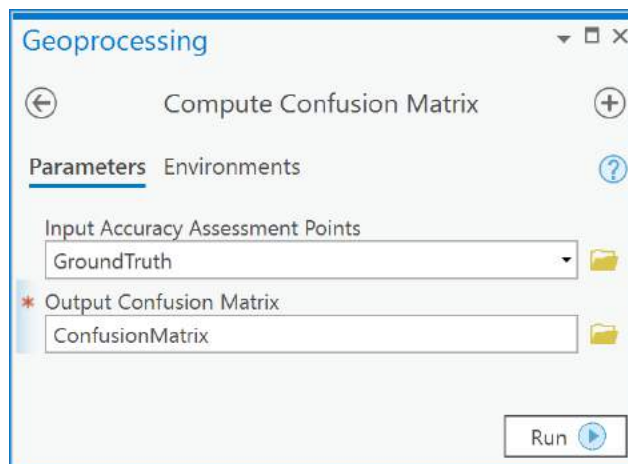
8. Répétez ces étapes pour tous les points du tableau "GroundTruth", afin de créer un tableau complet et prêt pour l'évaluation de la précision, par exemple:

GroundTruth			
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

9. Allez maintenant dans l'onglet « *Edit* » et cliquez sur « *Save* » pour sauvegarder les modifications apportées au tableau "GroundTruth". Cliquez ensuite sur le bouton « *Clear* » dans le groupe « *Selection* » du même onglet pour désélectionner le/les point(s):



10. Votre tableau "GroundTruth" contient maintenant une colonne appelée **Classified** qui contient le résultat du classificateur à chaque point et une colonne appelée **GrndTruth** qui contient la classe d'information observée à chaque point. Maintenant, vous voulez calculer combien de fois **Forest** (valeur 1) correspond à **NonForest** (valeur 2), combien de fois **Forest** (valeur 1) correspond à **Forest** (valeur 1), combien de fois **NonForest** (valeur 2) correspond à **NonForest** (valeur 2) et enfin combien de fois **NonForest** (valeur 2) correspond à **Forest** (valeur 1). Comme nous ne traitons que deux classes dans cet exemple, vous pourriez facilement le faire manuellement, mais ArcGIS dispose d'un outil qui le fait pour nous, appelé « *Compute Confusion Matrix* », qui se trouve sous « *Spatial Analyst* » -> « *Segmentation and Classification* » dans ArcToolbox. Ouvrez cet outil et, sous « *Input Accuracy Assessment Points* », choisissez votre table "GroundTruth". Appelez le résultat ConfusionMatrix et sauvegardez-le dans votre fichier IMAGERY.gdb :



11. Ouvrez le tableau ConfusionMatrix:

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

Dans le tableau, C_1 signifie la valeur de la classe 1 (Forêt), et C_2 signifie la valeur de la classe 2 (Non-Forêt). Les valeurs indiquées dans les cellules où une colonne et une rangée du même nom se croisent indiquent le nombre de fois où les deux classes correspondantes ont été correctement classées. Par exemple, dans le tableau illustré ci-dessus, il est arrivé 10 fois qu'un C_1 corresponde à un C_1 (où la colonne C_1 croise la rangée C_1) et 1 fois qu'un C_1 corresponde à un résultat classifié C_2. Ainsi, 10 points de Forest ont été correctement classifiés et 1 a été incorrectement classifié. Similairement, il y a eu 8 fois où un C_2 a été correctement classifié comme un C_2, et une fois où il a été classifié comme un C_1. Ainsi, 8 fois sur 9, NonForest a été correctement classifié. La valeur de Kappa est une mesure, allant de -1 à 1, qui indique dans quelle mesure les performances du classificateur sont supérieures à celles d'un classificateur aléatoire. Ici, elle est de 0,797, ce qui est considéré comme un résultat relativement bon. Vous pouvez consulter le fichier d'aide de l'outil « *Compute Confusion Matrix* » pour plus d'informations.

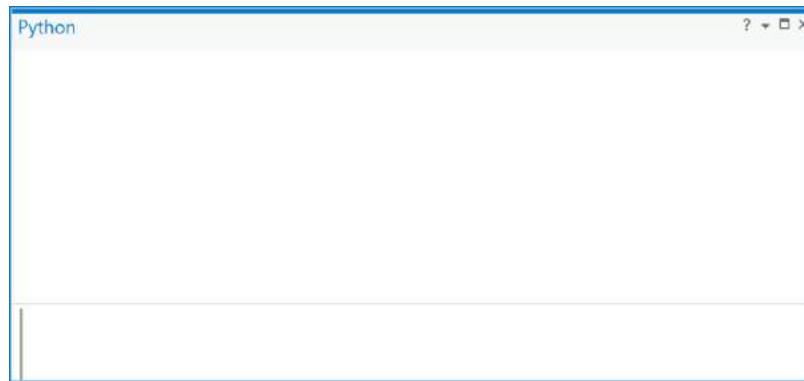
12. Maintenant, répétez toutes les étapes pour cette question avec votre couche "MLCLASS_RE_21_classified" en utilisant l'image RE2016_07_20 comme couche de vérité terrain.

Q9: Comment puis-je créer un raster contenant uniquement les forêts à partir de chacun de mes résultats classifiés, et déterminer où les forêts ont diminué?

Puisque vous avez maintenant deux rasters, tous deux contenant des classes d'information appelées **Forest** et **NonForest**, vous pouvez utiliser une simple expression d'algèbre de carte pour créer une couche raster booléenne où la valeur de 1 signifie **Forest** et 0 signifie **NonForest**. Vous pouvez le faire de plusieurs façons, ici nous allons essayer d'utiliser le langage de programmation Python qui est intégré dans ArcGIS :

1. Ouvrez le volet Python:

Nom de l'étudiant: _____ # d'étudiant: _____



2. Tapez les trois lignes suivantes et appuyez sur la touche Entrée (notez que vous n'avez pas besoin de taper les lignes en vert qui sont précédées du signe #, puisqu'il ne s'agit que de commentaires pour vous indiquer ce que fait chaque ligne):

```
# accéder aux fonctions de Spatial Analyst
from arcpy.sa import *

# Créer une variable raster à partir des résultats de la classification
filtrée
x = Raster("MLCLASS_RE_21_classified")

# Créer un raster booléen en utilisant l'opérateur relationnel pour les
égalités logiques
forest_2016 = x == 1
```

La première ligne indique à ArcGIS de vous permettre d'accéder aux fonctions de Spatial Analyst dans le volet python. La deuxième ligne utilise la fonction Raster() pour créer un objet raster qui peut être utilisé dans une expression d'algèbre de carte Python. La troisième ligne est une expression d'algèbre de carte relationnelle utilisant des égalités logiques (==). Cette expression examine chaque cellule de la couche raster appelée x, et si la cellule a une valeur de 1 (**Forêt**), alors la nouvelle valeur de la cellule raster aura également une valeur de 1. Si la valeur de la cellule est différente de 1, la valeur du raster de la cellule en question sera égale à zéro.

3. Le résultat sera une couche matricielle appelée "forest_2016" dans le volet Contenu (« Contents »), mais avec des valeurs de pixel de 1 pour les cellules qui sont des forêts et de 0 sinon.
4. Modifiez maintenant ces lignes pour créer un raster binaire pour l'image Landsat classifiée:

```
# accéder aux fonctions de Spatial Analyst
from arcpy.sa import *

# Créer une variable raster à partir des résultats de la classification filtrée
y = Raster("MLClass_L7_21")

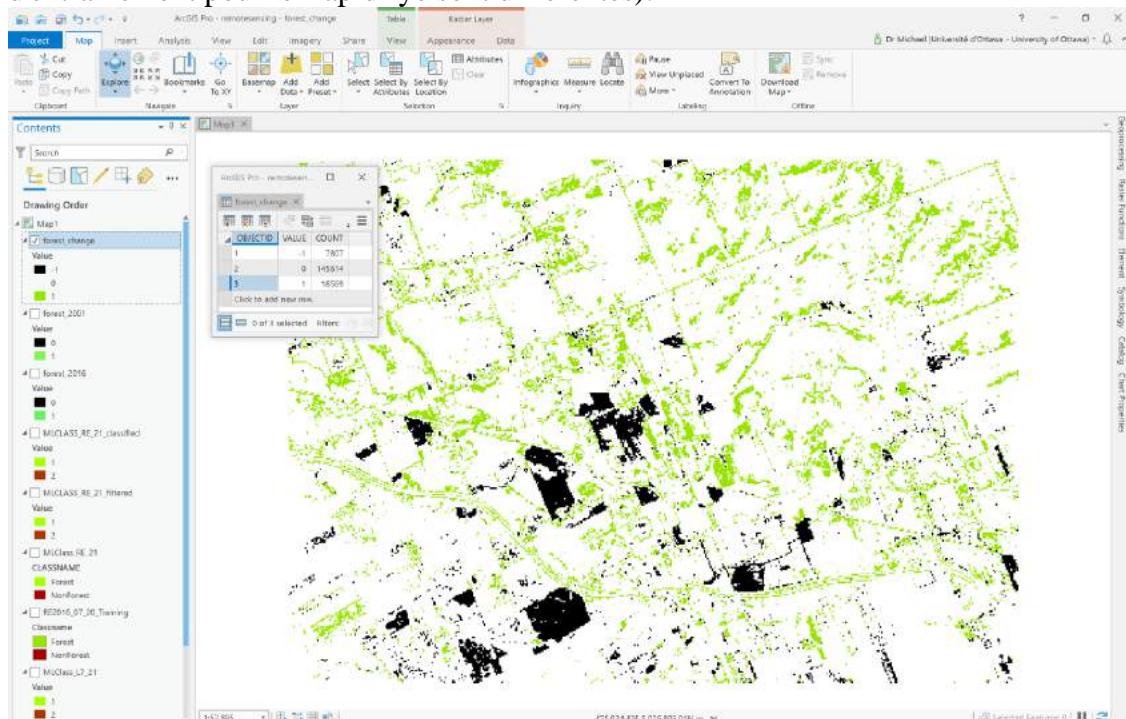
# Créer un calque booléen en utilisant l'opérateur relationnel pour les égalités
logiques
forest_2001 = y == 1
```

Nom de l'étudiant: _____ # d'étudiant: _____

5. Maintenant, pour déterminer la quantité de forêt qui a été perdue ou gagnée, nous pouvons soustraire la forêt_2001 de la forêt_2016:

Soustraire 2016 de 2001 pour voir où la forêt a gagné et/ou perdu du terrain
 $\text{forest_change} = \text{forest_2016} - \text{forest_2001}$

6. La carte obtenue ressemblera à ceci (elle ne sera pas exactement la même car les données d'entraînement pour le RapidEye sont différentes):



Ici, la table d'attributs nous indique dans le champ « *Count* » le nombre total de pixels pour chaque valeur.

Nom de l'étudiant: _____ # d'étudiant: _____

Répondez aux questions suivantes.

1. [1 point] Quelle est l'incertitude positionnelle due à la résolution des cellules dans la couche "MLCLASS_L7_21" ?
2. [3 points] Dans Q2, vous avez créé une couche raster thématique appelée "MLCLASS_L7_21", et dans Q3-Q7, vous avez créé une couche appelée "MLCLASS_RE_21_classified". Dans Q8, vous avez ensuite créé des matrices de confusion pour les deux couches. Présentez une carte pour chaque résultat et chaque matrice de confusion. Expliquez quel résultat est le plus précis, comment vous le savez et quelle en est, selon vous, la raison.
3. [3 points] Dans le raster que vous avez créé à la fin de la Q9, vous aviez des valeurs de cellule de -1, 0 et 1. Décrivez ce que ces valeurs représentent dans le paysage, par exemple, que signifie une valeur de cellule de -1, que signifie une valeur de cellule de 0 et que signifie une valeur de cellule de +1 ? Présentez la carte avec votre explication. Quelle est votre conclusion générale sur la perte de forêts dans la région?
4. [3 points] Dans la couche différenciée que vous avez créée à la fin de la Q9, vous pourriez être intéressé par les grandes régions contiguës de cellules avec des valeurs qui indiquent une perte de forêt (conversion de forêt à non-forêt). Où se trouvent ces grandes régions contiguës ? Utilisez l'outil « *Region Group* » sous « *Spatial Analyst* » -> « *Generalization* » avec votre résultat de Q9 comme couche d'entrée. Lisez le fichier d'aide de l'outil « *Region Group* » en cliquant sur le bouton Aide de l'outil. Expliquez ce que fait l'outil « *Region Group* ». Maintenant, ouvrez la table d'attributs du raster résultant du « *Region Group* », cliquez avec le bouton droit de la souris sur la colonne "Count" et triez par ordre décroissant (« *Sort Descending* »). Examinez le champ "Link". Comment les valeurs du champ "Link" se relient-elles aux valeurs des cellules dans le résultat de la Q9 ? Fournir un tableau des surfaces en unités cartographiques pour les 10 plus grands groupes contigus de cellules qui sont passées de **Forêt** à **Non-Forêt** entre 2001 et 2016. Présentez également une carte montrant ces régions.
5. [2 points] Quelle est la taille moyenne des régions (taille des parcelles) en mètres carrés pour les régions qui sont passées de la catégorie Forêt à la catégorie Non-Forêt entre 2001 et 2016 ? Pour répondre à cette question, vous devrez effectuer une requête sur la table d'attributs de la couche des régions groupées (de #4) et faire un clic droit sur la colonne "Count" et choisir "Statistics" pour obtenir la moyenne, pour répondre à cette question.
6. [2 points] Sélectionnez uniquement les régions qui sont passées de la catégorie **Forêt** à la catégorie **Non-Forêt** entre 2001 et 2016 et dont la superficie est supérieure ou égale à 3,000 mètres carrés. Quelle est la pente moyenne de chaque région ? Pour répondre à cette question, vous devrez utiliser l'outil « *Spatial Analyst* » -> « *Surface* » -> « *Slope* » avec les données DEM (il s'agit d'un Modèle Numérique de Terrain) trouvées dans IMAGERY.gdb, ainsi qu'une requête d'attribut sur la couche des régions groupées (de #4) et « *Spatial Analyst* » -> « *Zonal Statistics as Table* », pour répondre à cette question. Présentez un tableau montrant la pente moyenne pour chaque région sélectionnée.

Nom de l'étudiant: _____ # d'étudiant: _____

7. [1 point] Examinez la couche appelée L8_2016_08_26, qui est une image Landsat 8 OLI. Utilisez votre outil d'identification pour cliquer sur quelques endroits différents et notez la VN à chaque fois. Comparez ces VN avec les VN pour des caractéristiques similaires dans la couche "L7_2001_08_25". Pourquoi les VN de l'image Landsat 8 sont-ils plus élevés que ceux de l'image Landsat 7 pour des éléments similaires ?
8. [2 points] Créez et présentez une carte NDVI (Indice de Végétation par Différence Normalisée) en utilisant l'image L8_2016_08_26. Consultez vos notes de cours ou l'Internet pour déterminer les bandes appropriées à utiliser pour calculer le NDVI avec les données Landsat 8, car elles ne sont pas les mêmes que pour Landsat 7. Qu'est-ce que cette carte vous apprend ? Pourquoi ?