Marcel Rodriguez-Riccelli
2/4/2026
Satellite Image Analysis (GEOG 581)
<u>LAB 4 - UNSUPERVISED CLASSIFICATION</u>

LAB DAY: THURSDAYS
LINK TO CODE: [GOOGLE EARTH ENGINE]()

QUESTION 1. For 2013 NAIP imagery, what is the earliest and latest image date in the naip_filtered imageCollection?

The *naip_filtered* collection contains 15 images from two different dates—7 images in the collection were taken on June 19th, 2013, the earlier of the two dates, and the other 8 were taken later on July 4th, 2013.

QUESTION 2. Please interpret the number of land cover classes visible in the NAIP image. How many classes can you identify, what are they, and how do you distinguish one land cover class from the other? Are there any classes that are more difficult to distinguish or separate than others? Feel free to use example land cover types or locations in your answers.

Referring to the USGS National Land Cover database indices we can see the following land cover types in the NAIP image:

- Open water, which is easy to identify by the unique and definitive shapes and coloration of aquatic features. At the edge of the water, however, it can be hard to delineate the extent to which the water ends and bordering wetlands begin. Greenish coloration throughout from algae and other suspended particles may also be confusing.

- Low, medium and high intensity developed spaces are easy to identify from one another and other land cover types based on the regular shapes of anthropogenic structures such as roads, walkways, and buildings. Developed open spaces are somewhat more ambiguous, and difficult to distinguish from similar land cover types like grasslands, barren land, and shrub/scrub.

- Barren land, which is relatively difficult to differentiate from other brownish/greyish land cover types such as shrub/scrub, grasslands, and pasture, but still identifiable using contextual clues from shapes features on the land, such as the presence of excavation tools.

- Forest, which as a broad category is easy to differentiate from very different land cover types such as developed or water, but difficult to distinguish inter-categorically, as to whether it is deciduous, evergreen, or mixed.

- Shrub/scrub and grasslands are difficult to distinguish from each other, and also from barren land as mentioned before if they're relatively dry. Although similar, they can be differentiated from pastures and crops by using clues such as the presence of barns and nearby structures and tools associated with agriculture.

- Wetlands are easy to identify along the sides of aquatic features, where vegetation is much greener and of a different character than other types of forested areas, although it's difficult to say whether they're woody wetlands or emergent herbaceous wetlands. Other areas away from the waters edge, which may also possibly be wetlands, are harder to distinguish from grasslands or other lightly vegetated land cover types.

Fig 1. The NAIP high-resolution true-color image of our region of interest in 2013.

QUESTION 3. What Landsat bands are effective for differentiating these land cover types? Please enter these bands as visualization parameters for the landsat_visParams variable.

The common false color Near-infrared, Red, Blue combo helps to provide added information for differentiating between some of the ambiguities from the NAIP imagery, especially when trying to determine whether an area is grassland, barren land, or developed open space. *(Fig. 2)* It also helps to cleanly delineate the edges of the aquatic areas from the surrounding banks, where vegetation encroaches onto and contributes particulate matter to the water.

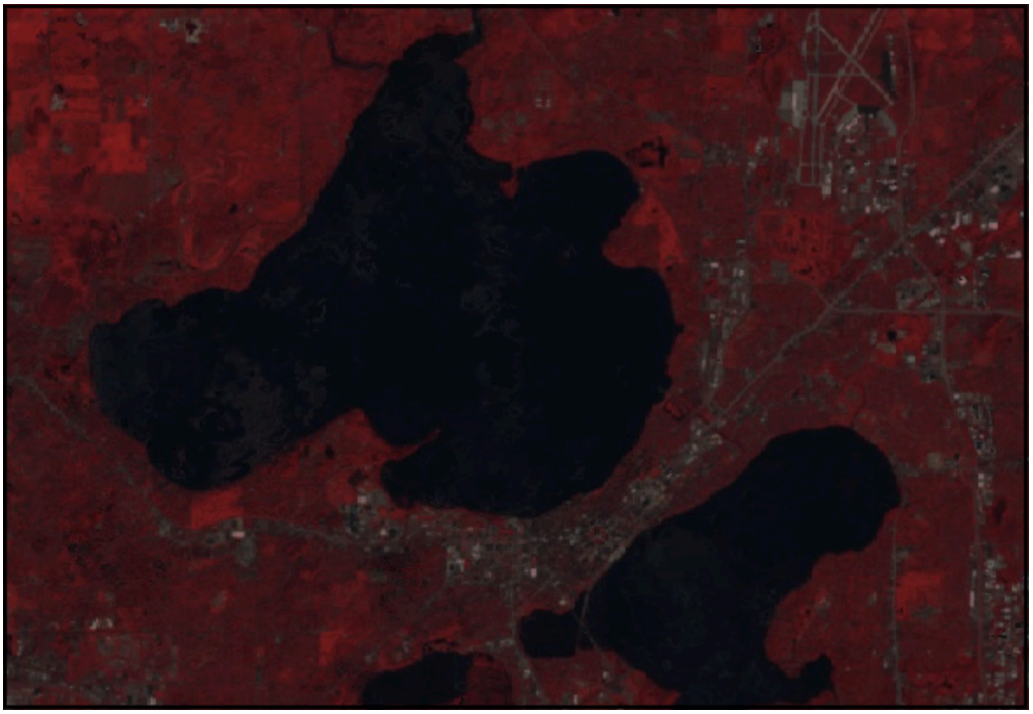

Fig. 2. False-color Landsat 7 imagery of the region of interest in 2013, with NIR encoded in the red color band, the Red band encoded in blue, and the Blue band encoded in green.

QUESTION 4. For visual comparison with the Landsat 7 composite image, why would using the NAIP imagery be preferable to referencing the default imagery used for the Earth Engine base map?

The Earth Engine base map provides toponymic context and simple classification of land cover type, both of which can be very useful, but the NAIP imagery is highly detailed, and can be used to gather strong evidence from which to deduce land cover type—for example as mentioned above, the ability to look at structures or features have smaller footprint than the square area of a Landsat 7 pixel.

QUESTION 5. For each of the six 'parameters' or 'arguments' for the image classification in the script above (including yr at the top of your script), please provide a description of the parameter's role in the classification in your own words.

Below, I'll break down each of the arguments in the classify() function argument by argument:

image

The input image to be classified.

num_clusters

The number of spectral clusters that the feature space will be divided into.

num_px

The number of pixels randomly sampled from the image to train the cluster centroids.

max_iter

The maximum number of centroid-update iterations allowed during optimization.

val_seed

The seed number controls the pseudo-random initialization of centroids and pixel sampling, to allow for reproducibility.

img_scale

The spatial resolution at which pixels are extracted from the image during sampling.

QUESTION 6. What do the pixel values in the result image refer to? How many bands are visualized for the result image, and what does each represent?

The value of a pixel in the resulting image is determined by which discrete spectral cluster it belongs to, where each cluster is represented by one of a randomly generated palette of hues. The resulting image has three bands labeled *viz-red*, *viz-green*, and *viz-blue*; these do not represent anything in and of themselves, but instead each cluster class is represented by a random combination of values of these three bands to produce a distinct hue.

QUESTION 7. Please comment on how well land cover types in the NAIP reference image that you identified in Question 2 above are captured in the 'clustered' (ie classified) result image. What land cover types appear distinguishable in the NAIP imagery but are spread across different clusters in the result image? What clusters in the result image are of the same land cover type in the NAIP imagery?

The classifier does a good job at distinguishing between the larger categories of the land cover types we had distinguished between before *(Fig. 3)*—such as highly developed areas, lightly vegetated grassland type surfaces, wetlands, and forests—and are verifiable as being of the same land cover type in the NAIP imagery. *(Fig. 1)* However, there are definite areas of confusion, especially in the aquatic surfaces, where there are discrepancies at different depths and where suspended greenish particles are visible in the NAIP imagery. At times water pixels are clustered together with land cover types that in the NAIP imagery seem to be open/low developed areas, and at other times with forested areas. The lower spatial resolution of the Landsat input imagery as opposed to the NAIP imagery also selects for the dominant land cover type where multiple are grouped together within the spatial extent of a pixel, which would be otherwise separable in the NAIP imagery.

QUESTION 8. Adjust the num_clusters variable to equal the number of land cover classes that you interpret in the reference NAIP image (from Question 2) and re-run the classification. Take a screenshot of this image, and include this image in your assignment hand-in.

I reduced the number of clusters to five, and took a screenshot of the resulting image. *(Fig. 4)*

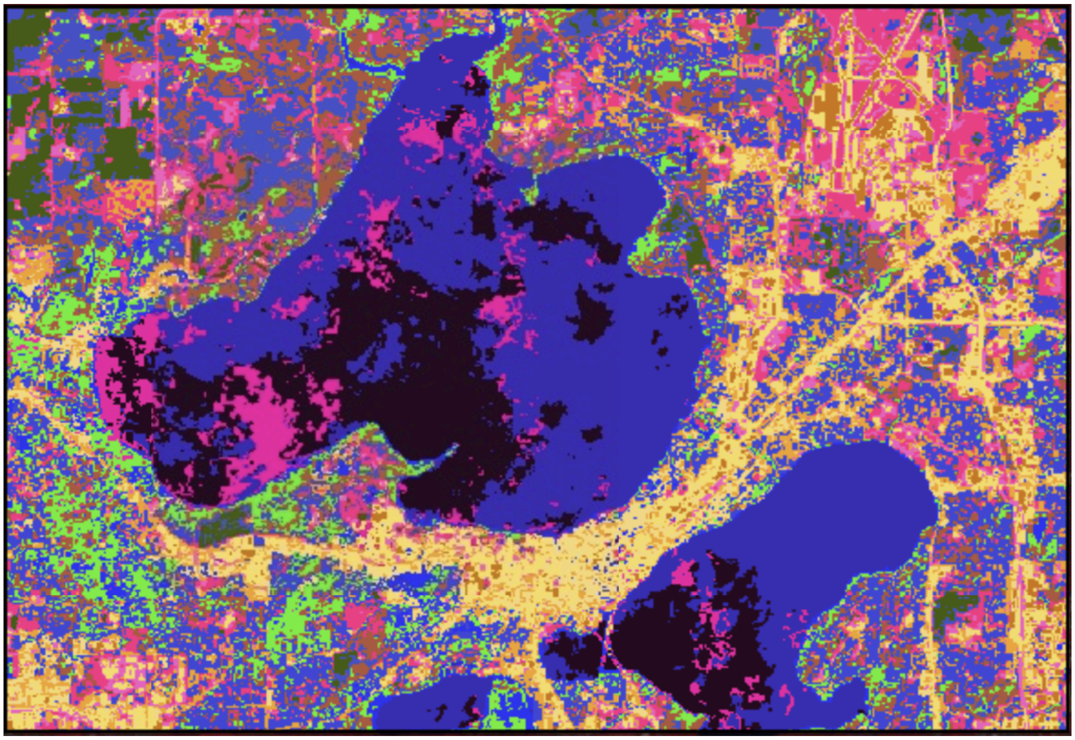

Fig. 3. The 'classified' result image from 2013, with 15 clusters, a 5000 pixel sample size, 10 iterations, and a random seed of 17.

Fig. 4. The second 'classified' result image from 2013, with 5 clusters, a 5000 pixel sample size, 10 iterations, and a random seed of 17.

QUESTION 9. Please describe how well the coverage of the clusters in the classified result image agree with the land covers visible in the reference NAIP imagery.

1. Are there disagreements between the clusters and the land covers in the reference imagery along the boundaries of land covers, or are there variations within a land cover type (i.e. building density, vegetative health, or surface water condition) that are difficult to capture within a single cluster?

There are noticeable disagreements between the clusters and the reference imagery, particularly both along land-cover boundaries and within individual land-cover types. In the 15-cluster output, aquatic areas are subdivided into three or four separate clusters, suggesting that spectral variability within water bodies—potentially due to differences in turbidity, depth, or surface condition—was captured as distinct classes rather than grouped into a single cohesive category.

In contrast, the 5-cluster output classifies water surfaces more uniformly, producing clearer and more continuous boundaries. However, this reduction in cluster number results in a loss of detail elsewhere. In the 5-cluster image, distinctions among varying levels of urban development and differences in vegetative health are less apparent. The 15-cluster result preserves greater differentiation within developed and vegetated areas, while the 5-cluster result emphasizes broader, more generalized land-cover groupings.

2. Are there any land cover types that you interpret are still not well represented by one or more clusters? If so, which land cover?

The land cover type least well represented are the moderate-to-low developed urban areas. This is likely because these areas contain a heterogeneous mix of surface materials—such as concrete, asphalt, rooftops, vegetation, and bare soil—within relatively small spatial extents. It seems that clustering more readily identifies coherent clusters in uniform land covers, while developed areas exhibit greater spectral variability that reduces classification consistency.

QUESTION 10. How do the clusters in the result image derived from a smaller sample size compare to the NAIP reference image land covers?

The result image produced using 15 clusters and a 1000-pixel sample size *(Fig. 5)* appears significantly noisier compared to the classifications generated with larger sample sizes *(Figs. 3 and 4)*. Distinct spatial patterns that were previously recognizable become fragmented, making it more difficult to delineate coherent land-cover features. Spectrally heterogeneous classes, such as urban areas, show reduced spatial consistency across the image, whereas more homogeneous surfaces—such as water bodies—remain relatively cohesive and consistently classified. This suggests that a smaller training sample size may inadequately capture the full spectral variability of complex land covers, resulting in less stable cluster assignments.

QUESTION 11. Which, if any, land covers are less effectively represented in this revised result image clusters?

Urban and suburban developed areas are less effectively represented in the revised cluster results. *(Fig. 5)* In contrast, forested regions, aquatic bodies, and grasslands that tend to be more spectrally homogeneous with a single material type dominating the pixel footprint, are still represented somewhat well. This again, shows the fragility of the classification algorithm to spectral inhomogeneity, especially when undersampled.

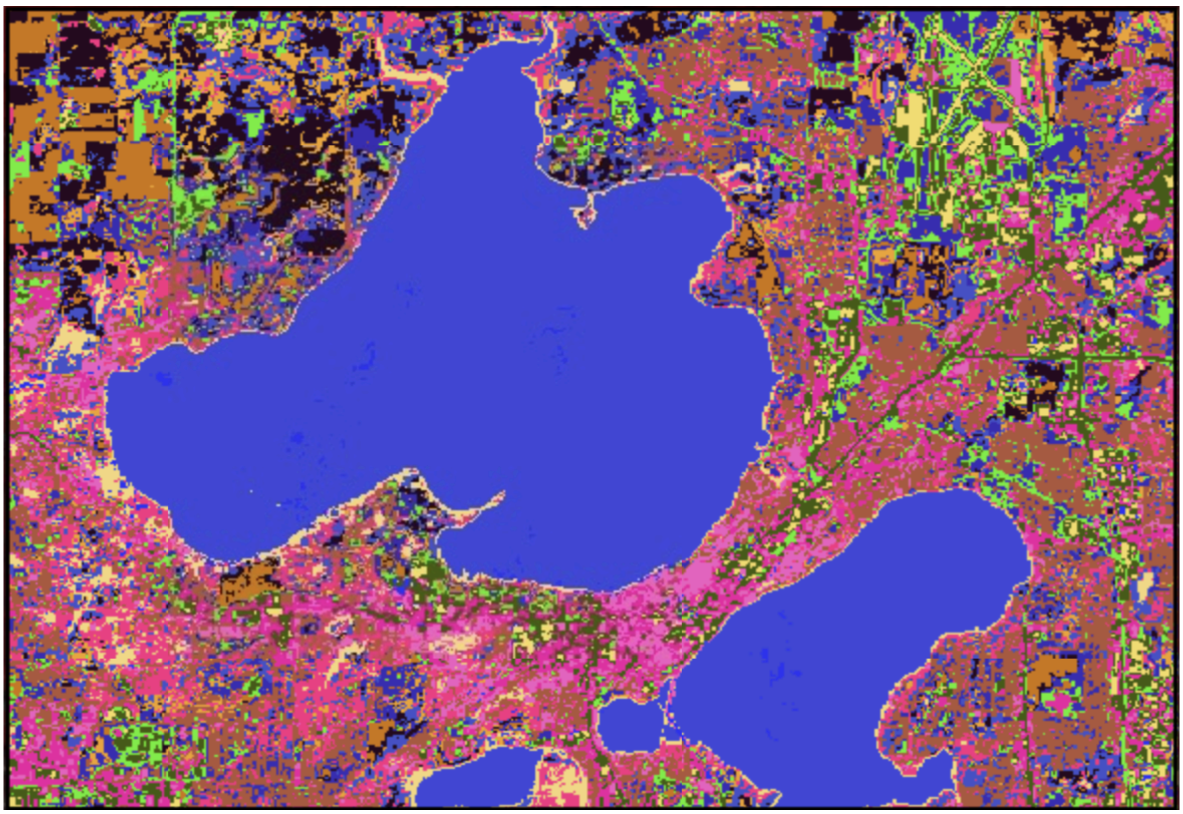

Fig. 5. The third 'classified' result image from 2013, with 15 clusters, a 1000 pixel sample size, 10 iterations, and a random seed of 17.

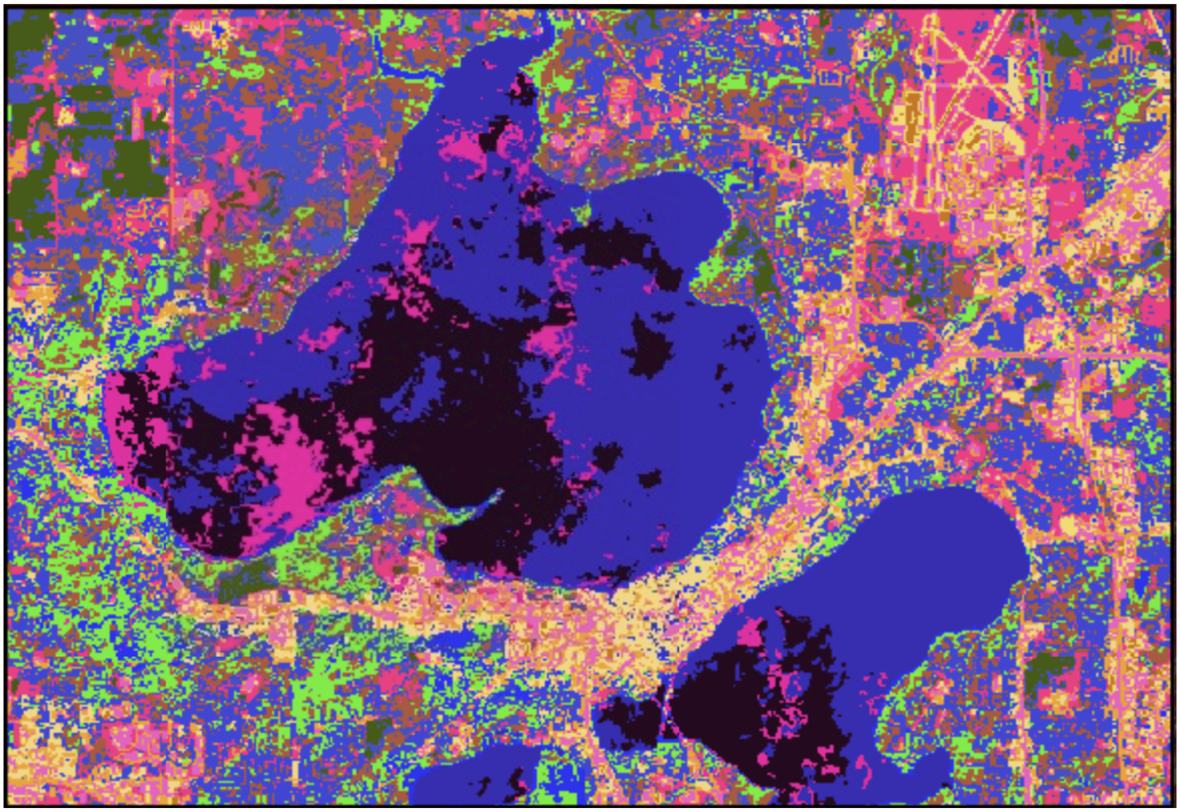

Fig. 6. The fourth 'classified' result image from 2013, with 15 clusters, a 5000 pixel sample size, 50 iterations, and a random seed of 17.

QUESTION 12. How do the clusters in the result image derived from a higher number of clustering iterations compare to the NAIP reference image?

Overall the classification with 50 iterations *(Fig. 6)* performs quite well for most surface types; especially homogenous forest and grassland areas. Again, we see the classifier struggle to uniformly classify water pixels, especially where turbidity from suspended particles and surface turbulence has affected the character of the surface in the NAIP imagery. *(Fig. 1)*

QUESTION 13. Which, if any, land covers are better represented in the result image clusters?

The urban and sub-urban moderate-to-high developed land covers are better represented in the resulting image with 50 iterations, *(Fig. 6)* and captures more detail and subtle variation overall, especially in the highly inhomogeneous inter-lake area.

QUESTION 14. How do the clusters in the result image differ when only the val_seed value is changed?

When only the *val_seed* parameter is changed, the overall spatial distribution of clusters remains largely similar, but the colors used to display each class differ.

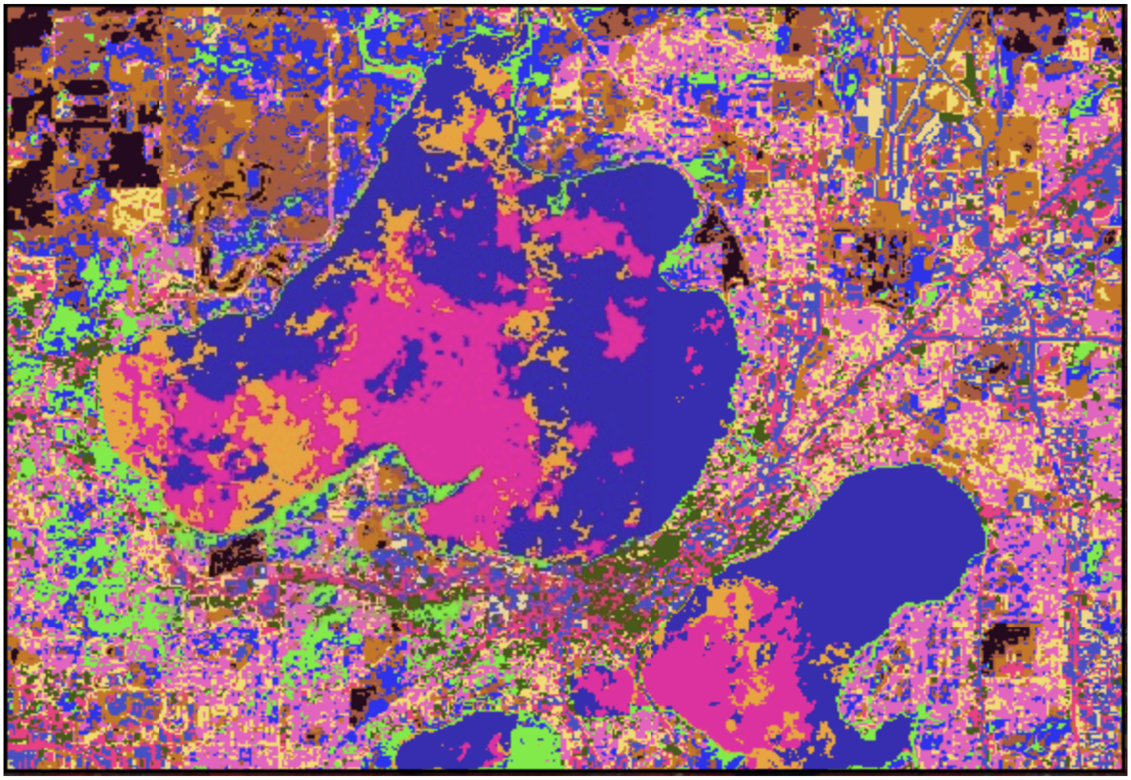

Fig. 6. The fifth 'classified' result image from 2013, with 15 clusters, a 5000 pixel sample size, 50 iterations, and a random seed of 13.

QUESTION 15. In your own words, what is the role of the seed parameter in running ee.Clusterer.wekaKMeans()? Why is it advantageous to keep the seed parameter consistent across multiple runs of wekaKMeans() even as other variable values may change?

The change in hues is due to the *randomVisualizer()* function assigning new random colors to cluster labels when seeded with a different value. Altering the seed can lead to slight differences in centroid placement and cluster boundaries, although in this case the overall classification structure appears stable. It can be advantageous to keep the seed parameter consistent for reproducibility, and direct comparison while changing other parameters.

QUESTION 16. Does any difference in colors of land covers or features in the result image between 2010 and 2013 suggest a difference in land covers or condition?

There are clear differences in the classification across the water surface, where pixels have been assigned to three different clusters, but concentrated in different areas than from the 2013 image—particularly in the tributaries and western portion of the lake. (*Fig. 7)* An interesting linear pattern has also arisen in the southern lake, which suggests some sort of reflectance-related artifact.

Wetland areas bordering the lake also appear more spatially extensive in the 2010 classification. Additionally, agricultural fields in the upper right portion of the study area are classified differently between years; large, contiguous clusters in one year appear more fragmented and heterogeneous in the other. Developed areas and adjacent land covers similarly show changes in cluster assignment.

Overall, these differences may reflect variations in vegetation health, soil moisture, phenological stage, water condition, or atmospheric and illumination differences between acquisition dates, rather than clear evidence of substantial land cover conversion.

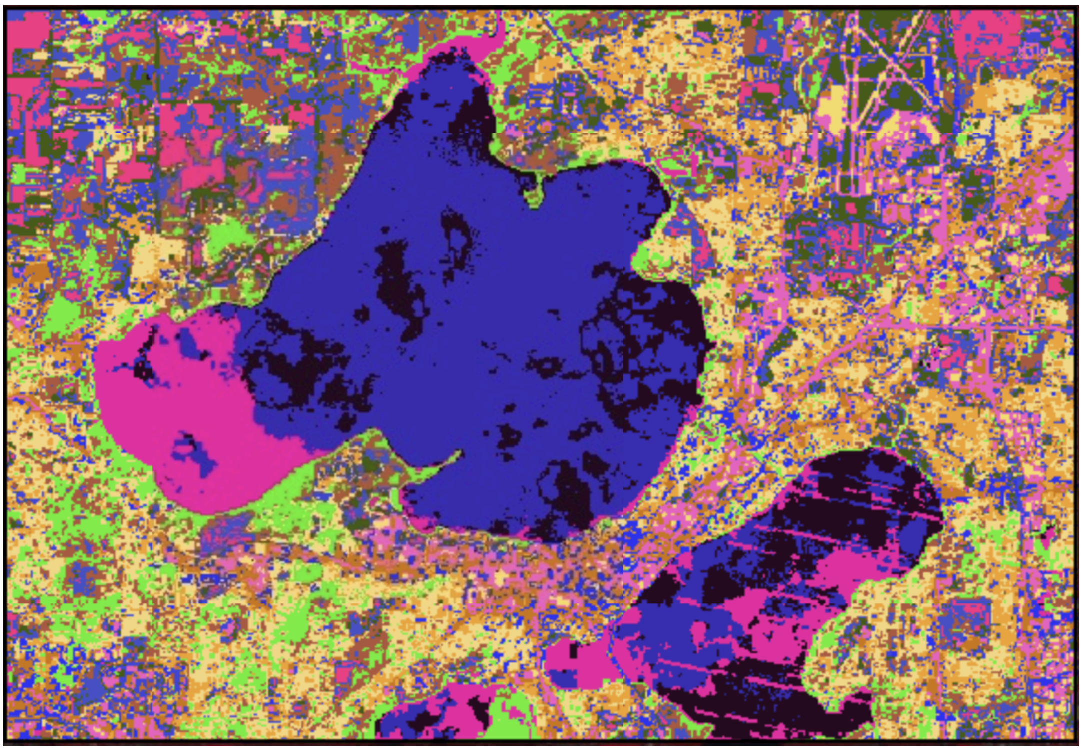

Fig. 7. The first 'classified' result image from 2010, with 15 clusters, a 5000 pixel sample size, 50 iterations, and a random seed of 17.

QUESTION 17. If you wanted to create a representative image of the 2010 and 2013 result images, could you take the mean() value of the 2010 and 2013 result images at each pixel? Why or why not?

No, taking the mean of the 2010 and 2013 result images would not produce a meaningful representative image. The pixel values in the clustered outputs are randomly assigned hues based on categorical class labels, not continuous numerical quantities with intrinsic mathematical relationships.

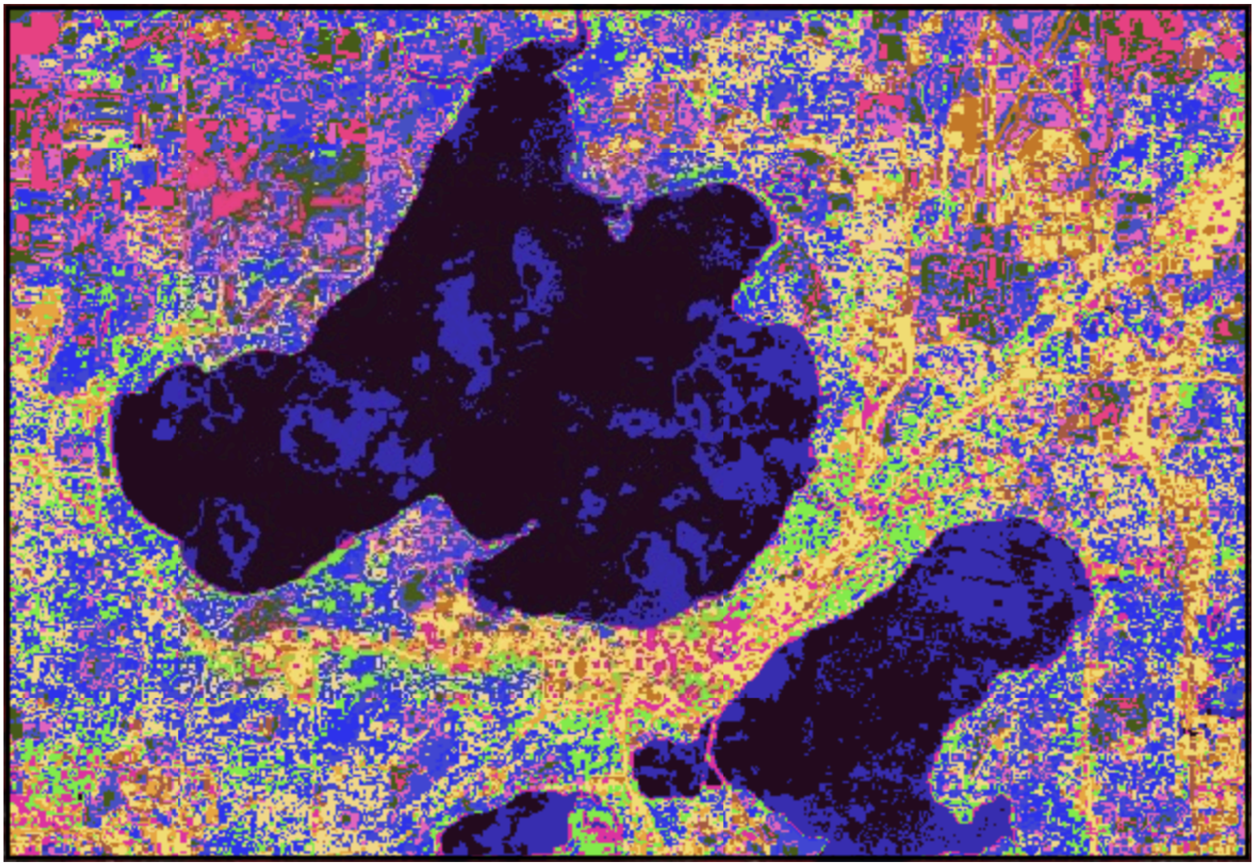

Fig. 8. The second 'classified' result image from 2010, with 15 clusters, a 5000 pixel sample size, 50 iterations, and a random seed of 17. This image only uses the blue and near infrared bands as inputs to the classifier.

QUESTION 18. Comparing the cluster results used in Question 16-17 to results based on input_nir or input_blue alone, which regions in the image show the greatest difference, and why?

The removal of the green band from consideration has reduced cluster variation seen in water surface pixels due to turbidity of a greenish hue. (*Fig 8)* Developed areas, particularly sparsely populated suburban regions in the southwest portion of the image, also show substantial differences. When considering the near-infrared band, vegetation strongly influences cluster assignment because healthy vegetation exhibits high reflectance in the near-infrared. As a result, built surfaces intermixed with lawns and trees may be grouped more closely with vegetated areas, altering the classification relative to the multi-band input.

QUESTION 19. What years of NLCD land cover data are available in the 2019 release of data?

According to the Google Earth Engine documentation for the NLCD 2019 dataset, it includes 8 epochs, including: 2001, 2004, 2006, 2008, 2011, 2013, 2016, and 2019.

QUESTION 20. How many land cover classes are represented in nlcd_2011_lc?

Using the *ui.Chart.image.histogram()* function, I generated a histogram that shows occurrences of 11 different land cover classes. *(Fig. 9)*

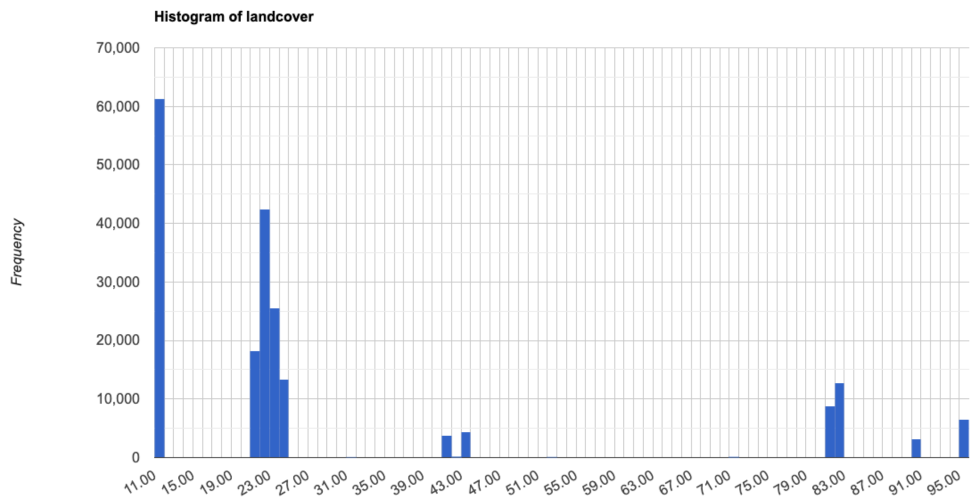

Fig. 9. Histogram generated from NLCD land cover dataset from area of interest in 2011.

QUESTION 21. What is the most common pixel value in nlcd_2011_lc? What land cover class does this value represent?

The most represented pixel value in this image is 11, which represents open water. *(Fig. 9)*

QUESTION 22. According to the histogram, is there a greater area of *Pasture/Hay* or *Cultivated Crops* in nlcd_2011_lc?

According to the histogram, there are more pixels with the value 82 than 81, meaning there is a greater area of Cultivated Crops land cover type than Pasture/Hay.

QUESTION 23. Just concerning ourselves with data within our inset_region, what values in the result image correspond to nlcd_2011_lc values representing…

A. Open Water?

Clusters 4 and 10.

B. Developed, Low Intensity?

There's a mix of different clusters for pixels which correspond to Developed, Low Intensity (Index 22) on the NLCD dataset, including Clusters 8, 5, 9 and 13 most prominently featured, amongst others.

C. Deciduous Forest?

The pixels representing Deciduous Forest (Index 41) on the NLCD dataset also have been assigned a mix of clusters in our classified output image, including Clusters 7, 1, and 41, amongst others.

QUESTION 24. What values have you included in nlcd_landcover_values?

I've included the following values in my *nlcd_landcover_values* array:

[82,81,24,23,11,41,41,81,43,23,11,21,24,43,43]

QUESTION 25. Please apply the remap() function with your completed nlcd_landcover_values list and add matched_nlcd_values to your map. Take a screenshot of this image, and include the output in your submission.

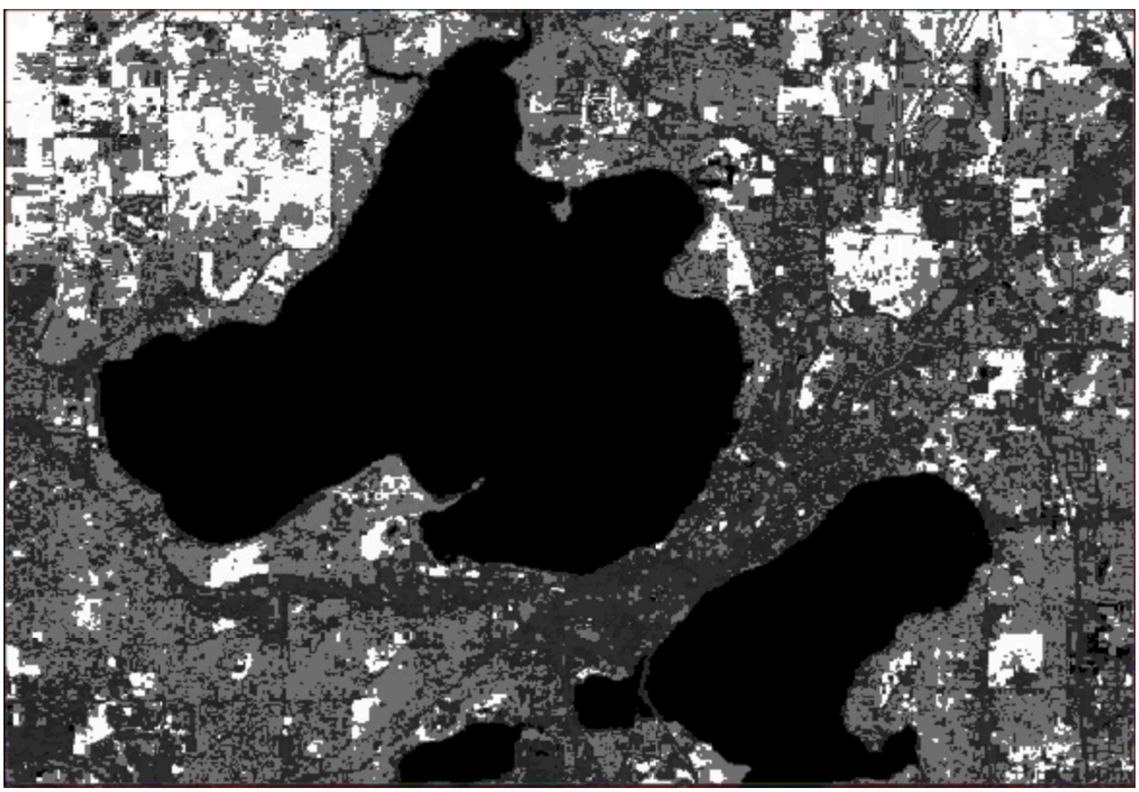

Fig. 10. Image matching NLCD dataset and classified land cover output.

QUESTION 26. Which land covers present in nlcd_2011_lc aren’t well represented in matched_nlcd_values? For NLCD land covers that were not well classified in the result image clusters, what are some reasons for why these particular land covers would be difficult to classify?

Again, we can see that moderate- to highly developed areas are the most difficult to classify, as they are highly inhomogeneous. *(Fig. 10)* These areas exhibit rapid changes in surface materials—such as concrete, trees, dirt, and grass—over short spatial distances. This high spatial variability makes it difficult for the clustering algorithm to group them into a single, cohesive class.