Josh Virene
NR323-001
6 May 2022
Remote Sensing and Image Interpretation
Lab Seven: Final Project

Lab One: Basics and Introduction to Google Earth Engine

Purpose:

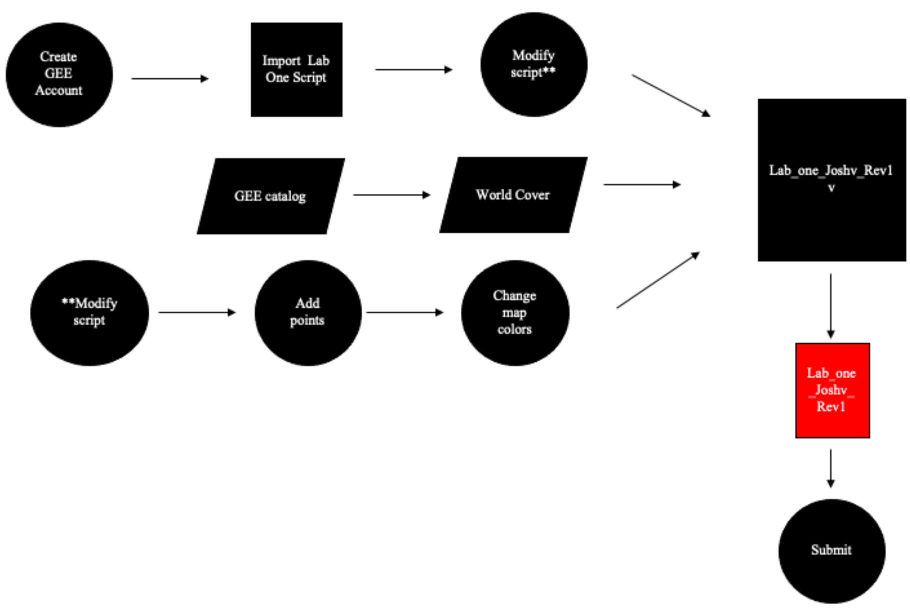
What was the overall goal of this lab?

- The goal of this lab is to explore the Google Earth Engine interface, and develop familiarity with the JavaScript coding language, as well as their applications within remote sensing.

What were individual objectives pursued in this lab to "answer" the overall goal?

- Acquire a Google Earth Engine account and become familiar with its interface- the Code Editor
- Work with and navigate land cover data using the Code Editor, share results with the "permalink" URL
- Modify scripts in the code editor and save them to the repository

Overview:



Results:

The band name for this image is called "Map" and the properties associated with the image include the date, a description of the map, keywords, the provider and a link to their webpage, and the image type. Within the inspector tab, there are additional properties associated with the image including class names, class palette (the colors associated with the map), class values, and others. This dataset represents imagery obtained in 2020. General metadata for this dataset are in the table:

Image Name	Description	Date / Scale	Citation	Url	Date Accessed
ESA WorldCover	European Space Agency	2020 10m	ESA WorldCover 10m v100	https://esa-worldcover.org/	24 January, 2022

The land cover classes provided in the dataset are included in the following table:

Class Values	Class Names	Class Palette (Color)
10	Trees	
20	Shrubland	
30	Grassland	
40	Cropland	
50	Built-up	
60	Barren / sparse vegetation	
70	Snow and ice	
80	Open water	
90	Herbaceous wetland	
95	Mangroves	
100	Moss and lichen	

In evaluating the accuracy of this map, I found that it does a good job of approximating general land cover data; it conveys the land cover theme and users of this map can get a sense for what sort of landcover is in a given area. When looking closer (i.e., at individual raster cells), there are classification errors where the map incorrectly classifies certain cells. Three instances of such classification error are reported in the following table:

	Location (Fort Collins, CO):	WorldCover Classification:	What it should be:
Cell one:	College & Edwards intersection	Grassland	Built-up
Cell two:	CSU Lagoon	Built-up	Open water
Cell three:	Road surrounding the Oval	Trees	Built-up

Discussion:

This WorldCover map is valuable for many reasons. Urban developers may use the map to determine the suitability of land for expanding developments in Fort Collins. Natural resource managers may use the land cover data to evaluate animal habitats or determine where they should allow / prohibit certain recreation activities. Farmers may use the data to determine the suitability of an area for the crops that they are growing.

The biggest limitation of this dataset is that there are many cells on the map that are incorrectly classified (i.e., cells are labeled "built up" when they should be "open water"). This limitation is important to consider when using this map because it can impact planning decisions. For example, if urban developers or farmers allocate resources according to a certain percentage of land cover that they believe will be suitable for their expansion, they will have misallocated their resources due to the incorrectly classified cells. Another limitation is classifications within certain cover types. For example, the classification "trees" could have sub classifications such as deciduous and coniferous.

Because of this first limitation, it is important that users of this map understand that there is uncertainty in the land cover types and that additional research and planning should take place before making large financial decisions.

Appropriate uses:

This map is appropriate for getting a general idea for land cover types across the world. For example, users of this map can explore regions of the United States and distinguish grassland, cropland, shrubland, and the other classes. Farmers could approximate what percentage of land in their state is cropland, game managers could determine what percentage is forest.

Limitations and caveats:
As mentioned in the discussion, these are approximations, and it is important to recognize that there are inaccurately classified cells on the map. Because of this, the map is not appropriate for getting highly precise estimates of landcover and in a practical setting, allocating funds to expansion projects based on those estimates.

Next steps:
A helpful step to improve this map would be to increase detail and implement more precise technology that more accurately distinguishes land cover types.

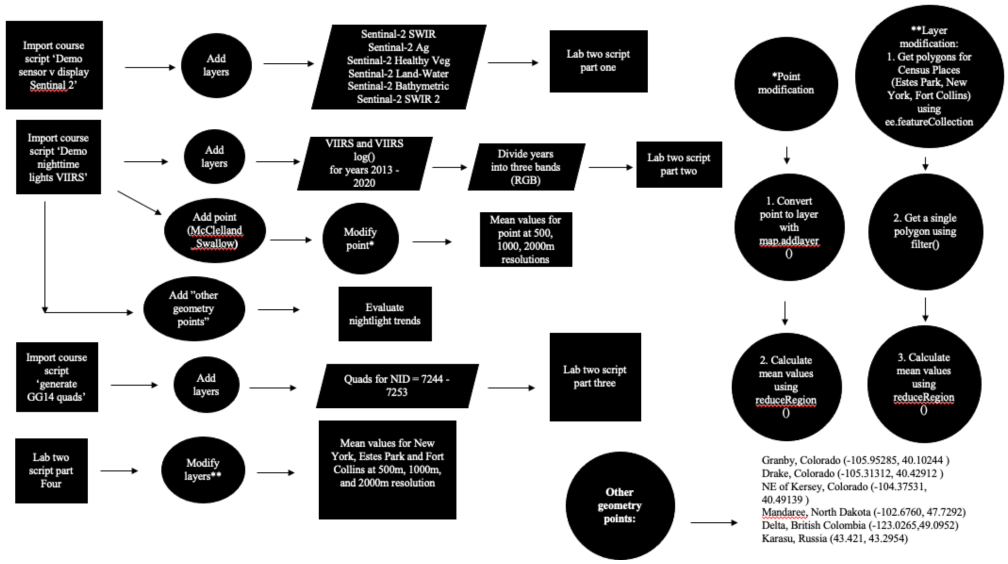
a. **Permalink:**
https://code.earthengine.google.com/4c3a09a8513dbec16484f67c3586b9e6
b. **Script Path:**
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabone%2FLab_one_joshv
c. **Helpful resources:**
https://gis.stackexchange.com/questions/395679/reclassify-values-of-the-land-use-land-cover-class-of-copernicus-global-land-cov
Example scripts (users/DavidTheobald8/NR323)
Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, Linlin, Tsendbazar, N.E., Ramonio, F., Arino, O., 2021. ESA WorldCover 10 m 2020 v100. https://doi.org/10.5281/zendono.5571936
Virene, J.W. 2022. Visualize World Cover v100 data. Google Earth Engine script, URL: **https://code.earthengine.google.com/4c3a09a8513dbec16484f67c3586b9e6**

Lab Two: Visualizing, querying, and summarizing imagery

Purpose:
- The purpose of this lab is to interact with many different types of remote sensing data including true / false color imagery, VIIRS satellite imagery, and to practice interpreting satellite imagery.

The individual objectives for achieving this goal include:
- Visualizing imagery in true and false color
- Data analysis with VIIRS satellite data at different resolutions
- Interpreting imagery under global grid plots (NID)

Overview:



Results:

Part I: Visualize:

a. The display combination most useful for understanding wildlife habitat will be the layer: R=B8, G=B11, B=B2, Name= Sentinel-2 Healthy Veg. This layer brings out the forested landscapes and vegetation very well.
b. The display combination that is most useful for understanding large rivers is the layer: R=B8, G=B11, B=B4, Name= Sentinel-2 Land-Water. This layer contrasts bodies of water, including rivers with the surrounding landscape very well.
c. The visual patterns marked by the "mysteryFeature" geometry points are defined in the following table:

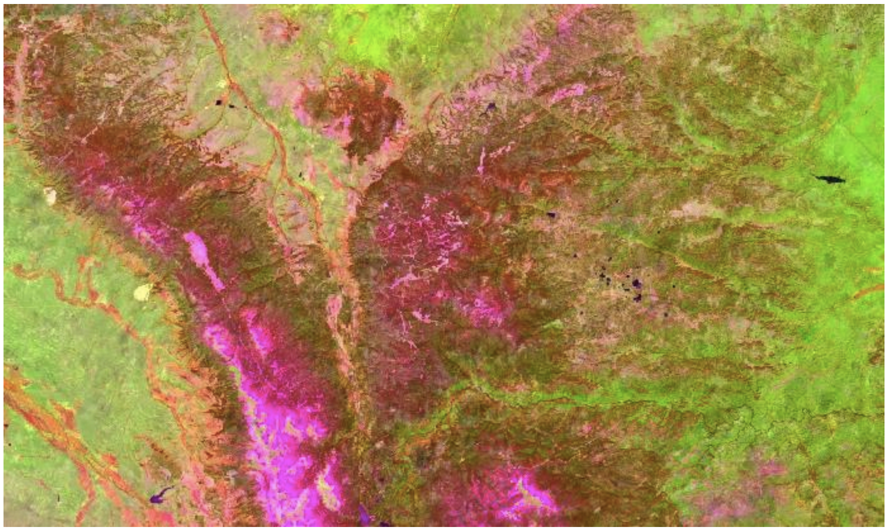
	Color	Description
Feature one	Red	Shadow cast by a cloud
Feature two	Green	Cloud
Feature three	Purple	Null data from cloud-removal algorithm **in the 'inspector' window, the values for this area are masked

These visual patterns are most likely a result of clouds and the shadows that clouds cast.

d. An abrupt change that has occurred is the Cameron Peak Wildfire, this fire is the largest wildfire on record in Colorado's history. This fire took place from August to December 2020. Below are pictures showing the damage that the fire had in the region:

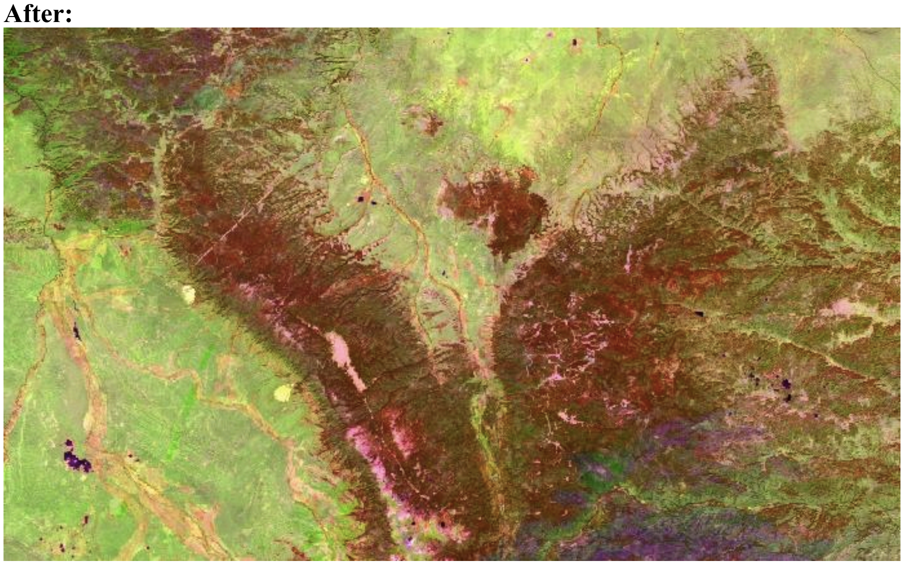
I analyzed both of these images using the 'Sentinal-2 Healthy Veg layer. In the first image which was prior to this fire, this layer highlights the vegetation and dense forest that spans throughout the region with the pink and bright green layers. In the second image which spans the duration of the fire, the vegetation in the region is vastly reduced and the burn scar becomes very apparent. To support that the color being displayed is forest, band eight in the sentinel-2 imagery is near infrared, which has high values for 'red' in the inspector tab.

Before:



Start / End dates: 1/1/2020 – 8/1/2020
Layer: R=B8, G=B11, B=B2, Name= Sentinel-2 Healthy Veg.

After:



Start / End dates: 8/13/2020 – 12/2/2020
Layer: R=B8, G=B11, B=B2, Name= Sentinel-2 Healthy Veg.

Part II: Query

a. Describe the general patterns you see in Northern Colorado

Looking across the timeframe from 2013 to 2020, the first key trend that the VIIRS datasets show is an expansion of the light into regions that were previously dark. Another trend that these datasets show are an increase in brightness in the larger cities of Fort Collins, Greeley, and Loveland. Red lights in the Northern Colorado region are becoming more spread out, which indicates that there is development and expansion in the region. The city centers appear white in all three cities, which indicates equal brightness across the 2013-2020 period.

b. Identify the trend for the locations provided in the table below:

Location	Coordinates	Trend
Drake, Colorado	-105.31312, 40.42912	Peaked in the middle
Grandby, Colorado	-105.95285, 40.10244	Increasing
Northeast of Kersey, Colorado	-104.37531, 40.49139	Decreasing

c. Identify the mean night light values for the point: -105.079036, 40.54588 at different resolutions

Resolution	Mean values
500m	60.5
1000m	44
2000m	58.5

d. The information for the three locations, their coordinates, and trends is reported in the table below:

Location description	Coordinates	Trend
Mandaree, North Dakota	-102.6760, 47.7292	Increasing
Delta, British Colombia	-123.0265, 49.0952	Decreasing
Karasu, Russia	43.421,43.2954	Decreasing

Part III: Interpret

NID	Classification	Intensity of human modification (I)	F	Degree of modification (H)
7244	Desert	0%	0	0
7245	Croplands and pasture	50%	20%	0.1
7246	Grassland / undeveloped	10%	1%	=<0.01
7247	Industrial + cropland agriculture	90%	60%	0.5
7248	Desert	0	0	0
7249	Marshland / undeveloped	0	0	0
7250	Cropland agriculture + pasture	40%	60%	0.2
7251	Cropland agriculture	50%	90%	0.5
7252	Ocean + dock / road	60%	10%	0.1
7253	Wilderness / recreation + cropland agriculture	50%	30%	0.2

Part IV: Summarize

	Mean Values – Fort Collins, CO	Mean Values - Estes Park, CO
500m	16.479	5.771
1000m	17.583	4.129
2000m	15.486	4.695

- Based on this data, Fort Collins is clearly more of a densely developed urban area than Estes Park. This is because the mean average nightlight brightness values for Fort Collins are higher than those for Estes Park at all resolutions.

Discussion:

The first section of this lab demonstrates the power of displaying satellite data in different bands and how doing so can bring out particular features. This is important and has many practical applications. An example is using the Sentinal-2 Healthy Veg layer to monitor regions of forest for wildfires or beetle kill. Another example might be the USDA using Sentinal-2 Ag to monitor the growing patterns and crop health in the midwestern US. These applications and many others show the power of false color imagery. These alterations to produce false color images change the spectral resolution of the image.
Another key consideration with these images is their temporal resolution. Based on the time frame of the imagery, users of these images can determine changes in the landscape. I looked at the impacts of the Cameron Peak Wildfire in 2020. The two images showed a significant reduction in healthy vegetation across the region due to the devastating impacts of the fire on the forest.

In the second section of this lab, we look at nightlights across the world using VIIRS imagery. This imagery is useful for looking at light pollution and the growth and expansion of cities across the world.

A key concept here is that having a higher spatial resolution will improve the accuracy of measurements. When we looked at the mean value for nightlights in various locations, the readings were more precise as the resolution in meters decreased (i.e., 500m produced a more precise mean than the 2000m resolution). The reason for this is that Google Earth Engine is deriving the mean value over a smaller range, so, when using the inspector and clicking on a certain point, it is highly likely that the mean value given by the 500m resolution will be closer to that points value than the 2000m resolution. There is a tradeoff with this; running a coarse analysis is quicker whereas a finer analysis is more precise. As with the Sentinal-2 images, the temporal resolution is important in determining the timeframe across which the night light data is monitored.

Another important component considers the temporal resolution of the VIIRS imagery. When creating the RGB layers for this map, different years correspond to different bands: 2013-15 – blue, 2016-18 – green, 2019-20 – red. We interpret reflectance in a given color as those years having the highest nightlight value, which is how this imagery can be used to monitor development in a region.

Looking at the 'image interpretation' component of this lab, the GG14 plots specified using randomly generated values for NID returned rectangular areas that needed to be appropriately identified and classified. I classified these images based on their land use and the human activities in the area. Classifying these images requires looking at the surrounding area, as well as distinguishing features in the area such as roads, buildings, and patterns on the land itself which may indicate grazing, farming, or timber harvesting. The impact of human activities on a region is determined by the intensity of the land use (I), as well as the percentage of the area where the activity takes place (F). Intensity is the degree of modification that human activities have on the land. Low impact activities (0-30%) include undeveloped, natural areas with passive use. Moderate impact activities (40-70%) are agriculture and residential development. High impact activities (70-100%) include resource extraction, mixed development, and industrial. The other component is how much of the area on which the specified activity takes place.
Based on these two measures, we can determine the impact of human activities on these plots. This classification requires user discretion and, in some contexts, will differ based on the users' interpretation.

Appropriate uses:

Sentinal-2 imagery has a wide range of applications, especially because there are many types of false color images that specialize in highlighting and bringing out different features based on the bands that are put into (RGB).

There are many additional uses for Sentinel- 2 imagery. SWIR imagery aids scientists in estimating how much water is in plants and soil. More generally, Sentinal-2 imagery can be used to monitor changes in a landscape over time. VIIRS imagery can be used for tracking the growth and development of cities across the world, and the light pollution produced by these cities.

The GG14 plots are useful for determining the land use and intensity of human activities in a specified region. The applications and uses of satellite imagery apply to section three of the lab. Developers and urban planners can use this imagery to determine the type of land cover in a region. As an environmental economist, I could see satellite data and human modification information being used to determine the pressure on ecosystems and consider the impacts of these activities accordingly.

Limitations and caveats:

A key limitation that applies to the imagery for every part of this lab is the classification; computing mean values is contingent on the spatial resolution, these decline in accuracy as the spatial resolution decreases. When computing a mean value for Fort Collins, the 2000m resolution mean value for night lights will not be as accurate as the value provided by the 500m resolution. Lower resolution sensors may not correctly distinguish features in dense / urban environments, which will reduce the accuracy of the map. *Resolution is a factor that impacts the certainty of the data we are analyzing*. There is also a tradeoff between speed and image quality.

Another limitation is within the temporal component of the imagery, some satellites capture imagery more frequently than others, which is important to consider when using these images and conducting analysis because landscapes will change over time. Users of sentinel-2 data should ensure that the timeframe of the imagery satisfies their purpose otherwise their data may not reflect the true nature of the area.

A limitation within the interpretation of imagery as conducted in part three of the lab is that it can be subjective; if somehow one of my classmates and I got the same set of NID values, it is highly likely that we classified their degree of modification differently. This limitation is important to consider in a professional setting to ensure that collaborators can draw consensus on classification and report effectively.

Next steps:

Another incredibly useful tool to examine is how well the VIIRS datasets and the resulting mean values of nightlight in a region predict the level of development and the intensity of human activities in that region. Map users can examine the nightlight trends for a city or any sort of land area and compare these against the degree of human modification that is occurring in that area.

These two measures complement each other because they provide two different metrics for human development in a region.

In the field of environmental economics, an interesting study that researchers could conduct using these metrics would be the impact of development on the economic welfare of the region. The mean values of nightlight and degree of human modification would be the independent variables that predict economic welfare (i.e., household income, wealth distribution etc.).

Another, more environmentally focused way to use this data would be a study on the proximity of increasing nightlights to national parks and wilderness areas. Researchers could determine if development and increased light pollution is decreasing the utility of visitors to these parks.

Sources, citations, and resources:

- **https://custom-scripts.sentinel-hub.com**
- **https://www.sciencedirect.com/topics/earth-and-planetary-sciences/spatial-resolution**
- Salafsky N, Salzer D, Stattersfield AJ, Hilton-Taylor C, Neu- garten R, Butchart SHM, Collen B, Cox N, Master LL, O'Connor S, Wilkie D. 2008. A standard lexicon for biodiversity conservation: unified classifications of threats and actions. Conserv Biol 22(4):897–911.
- Users/DavidTheobald8 'demo scripts' + demo videos
- Virene, J.W. 2022. Google Earth Engine script, URL (permalink): https://code.earthengine.google.com/9c5edef19322db63e906c2121fb68201

Permalink:

https://code.earthengine.google.com/9c5edef19322db63e906c2121fb68201

Script Path:

https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabtwo%2Flab_two_submission

Extended analysis:

I computed the mean nightlight values for New York City. These brightness values indicate that the city is brighter and more developed than Fort Collins and Estes Park. Looking at the imagery itself, New York city shows as white on the RGB layer. This means that the brightness values are reflecting in all colors; the interpretation for this is there hasn't been much change in nightlights over these years.

	Mean Values – New York City
500m	34.012
1000m	34.976
2000m	35.210

Given that New York City is a dense urban area and there is likely inaccuracy resulting from this resolution as was previously discussed, I computed the nightlight value at even higher resolutions as reported in the table. These values will give a more precise reading for the nightlight values in New York City.

	Mean Values – New York City
250m	34.637
100m	34.462

When I tried to compute this at the 10m resolution, the console returns an error message stating that there are too many pixels in the region. A limitation within the code of this analysis is that Google Earth Engine cannot compute mean nightlight values at this high of a resolution. A potential solution would be to write code that increases the maximum number of pixels, and then compute values at even higher resolutions.

<u>Lab Three: Image Interpretation</u>

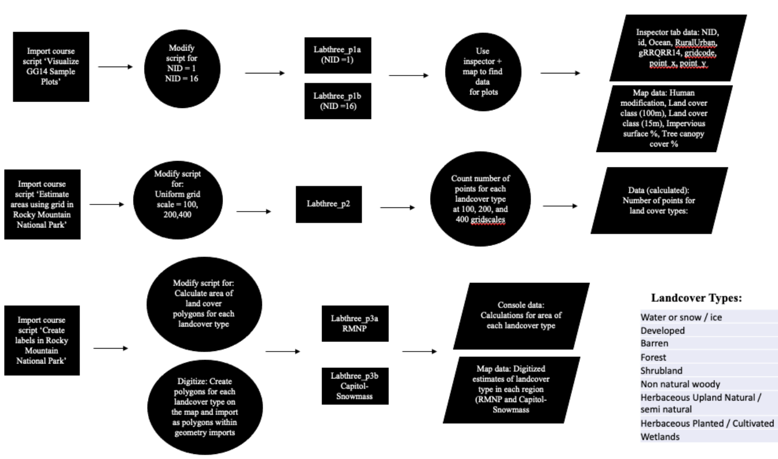
Purpose:

- The purpose of this lab is to further explore image interpretation and understand how different scales impact the precision of estimates. Additionally, this lab introduces two methods of classification for land cover- validation versus training, and under the training approach, digitizing polygons.

Individual objectives for achieving these goals include:

- Interpreting imagery and calculating human modification under global grid plots
- Estimating land cover types with grid overlays
- Digitizing land cover types and classifying them accordingly based on satellite imagery

Overview:



Results:

Part I: Quantify land use / cover from imagery

This section of the lab analyzes data from 10 samples on two plots, NID = 1, NID = 16. Clues from the satellite imagery such as texture, pattern, and color help to determine the human activities that take place within the samples on these plots. The accuracy of the footprint was increased compared to last lab due the ‘uniform grid’ layer. This layer places evenly spaced dots across the entire grid and there are roughly 25 per plot. The footprint of activities can then be calculated by computing the proportion of dots over which the activity takes place, divided by the total number within a sample (25). The two tables below give the data for NID = 1 and NID = 16. The mean values row gives the average human modification over the ten samples within their respective plots.

Tables for part one:

NID	1
id (plot)	000045f12aa7e0c14b06
Human modification (mean value)	2.22
Ocean	0 – terrestrial
RuralUrban	3 – urban edge
gRRQRR14 / gridcode	254

NID	16
id (plot)	000062b8546c02405664
Human modification (mean value)	0.94
Ocean	0 – terrestrial
RuralUrban	1
gRRQRR14 / gridcode	3631

Part II: Estimate the areas of land cover types within RMNP using a grid overlay

This attachment provides the tables for number of points, area, and proportion of the plot for each of the landcover types. Additionally, the fourth table shows the area and proportion of the plot as calculated in section three of the lab. It is useful to compare these calculations to see how well they match each other.

Dot count table:	**Uniform grid scale**		
Land cover classification	100m	200m	400m
Water or snow / ice	3	1	0
Developed	22	7	2
Barren	0	0	0
Forest	397	111	24
Shrubland	177	39	15
Non natural woody	0	0	0
Herbaceous Upland Natural / semi natural	236	48	11
Herbaceous Planted / Cultivated	27	8	2
Wetlands	164	52	16
Total dots	1026	266	70

Area table: (square meters)	**Uniform grid scale**		
Land cover classification	100m	200m	400m
Water or snow / ice	23247.50585	29889.6504	0
Developed	170481.7096	209227.553	227161.343
Barren	0	0	0
Forest	3076419.941	3317751.19	2725936.11
Shrubland	1371602.845	1165696.36	1703710.07
Non natural woody	0	0	0
Herbaceous Upland Natural / semi natural	1828803.793	1434703.22	1249387.39
Herbaceous Planted / Cultivated	209227.5526	239117.203	227161.343
Wetlands	1270863.653	1554261.82	1817290.74
Total land	7950647	7950647	7950647

As a proportion of total land	**Uniform grid scale**		
Land cover classification	100m	200m	400m
Water or snow / ice	0.003	0.004	0.000
Developed	0.021	0.026	0.029
Barren	0.000	0.000	0.000
Forest	0.387	0.417	0.343
Shrubland	0.173	0.147	0.214
Non natural woody	0.000	0.000	0.000
Herbaceous Upland Natural / semi natural	0.230	0.180	0.157
Herbaceous Planted / Cultivated	0.026	0.030	0.029
Wetlands	0.160	0.195	0.229

Area estimates from part three (square meters)		
Land Cover Type	Area Estimate	As proportion of total
Water	28210.086	0.0035
Builtup	147716.22	0.0186
Barren	0	0.0000
Forest	3194333.867	0.4018
Shrubland	1472459.94	0.1852
Grassland	1996464.038	0.2511
Agricultural	175848.057	0.0221
Wetland	955891	0.1202
Total area:	7970923.208	

The grid spacing choice is very important when estimating these land cover types and there are tradeoffs with each that should be considered. Tighter grid spacing (i.e., uniformGridScale = 100) means there are more

points on the plot. This will yield a higher level of precision because the landscape is heterogenous and the points we are using to compute land cover are over a smaller area. The disadvantage with using a smaller grid scale is that the computation is more intensive; computers or in this case, the users need to count more points for landcover types. The 100m grid scale has 1026 points. Larger grid scales (i.e., uniformGridScale = 400) have the advantage of being less complex. The 400m scale only has 70 dots, so there is less of a burden in computing the number of dots and their corresponding proportions. The disadvantage with this scale is that these points are being generalized over a larger extent, so their values may not be as precise with respect to land cover classification.
Another challenge with using these points is when they lie on the threshold between one classification and another; it can be difficult to determine which land cover type it should be classified as.

Part III: Draw land cover polygons to generate labels and estimate area within RMNP

A table giving the area estimates and their proportion of total landcover is provided in part II.
Questions:

1. If we drew polygons representing features using the level II classes, there would be more polygons and it would take longer to digitize. Using the level II classes, there are more features in the imagery to distinguish such as evergreen versus deciduous forest. This alone will add more time and additionally, the person creating this land cover with the level II classes will need to make more polygons to digitize the land cover within these classes.
2. The area estimates from the polygons are proportionally, very similar to those using the grid overlay method. If I needed to make a map, I would use the area estimates from the polygons instead of the ones from the grid. The polygon method likely gives the most precise area estimate for each class.
3. The minimum size polygon that could be digitized under a zoom scale where the image is displayed at 5m/px and digitizing in increments of 5 pixels would be 625m^2. The way to calculate this is to create a square, each vertex will be five pixels apart, so the side of each length is 25m. The area of this square is 25^2 or 1250m^2.
   For a triangle, with side lengths a = 25 b = 25 c = $\sqrt{1250}$
   = **625 m^2**

Discussion:

Section I of this lab is similar to section III of lab two. In both of these labs, we are analyzing the impact of human activities in the plot. Human modification (H) is determined by the intensity of land use (I), and the footprint-the extent on the sample over which the activity takes place (F). The first important difference is that in this lab, there is a uniform grid layer that improves the accuracy of our estimates for the footprint of the activity. Rather than just "eyeballing" the extent over which an activity takes place, we can estimate by using the proportion of dots for the activity divided by the total number of dots in the sample. The other key difference is that in this lab, we are determining human modification in the plot based on 10 samples within the plot whereas in the previous lab, we were just computing the value over the whole plot. Introducing these samples increases the precision of the estimates of human modification because the samples are dispersed throughout the plot, so they are more representative of the entire area. The previous estimates did not account for the heterogeneity of the plots and human modification in this manner, and they were less accurate.

Section II estimates the areas for each land cover type using the uniform grid scale and satellite imagery. The process for determining the area of each land cover type entails counting the number of dots for that land cover type and using total landcover area, and total number of dots to algebraically calculating landcover for each type. As discussed in the results section, there are tradeoffs associated with using tighter dot spacing (100m); these estimates will be more accurate, however they are more computationally demanding. The reverse is true for further spaced dots (400m), they are quicker to calculate but are less accurate. As the value for uniformGridScale decreases, the precision and computational demand increase.

Section III estimates the areas for each land cover type with the user digitizing polygons for each landcover type and calculating their area. Digitizing can be a long and tedious process depending on the zoom scale at which the user does this process. At the 200m scale (5m/px), it can be done quickly however, as the zoom increases, there is more detail per pixel. This will increase the accuracy of the polygons capturing a given landcover type, but again there is a tradeoff - between time spent digitizing and the accuracy of the landcover map.

Appropriate uses

The satellite imagery in the first section of the lab, coupled with the uniform grid can be used to get accurate estimates of human modification in a region. Another use of this imagery is determining the land cover and area within a region.

There are two approaches to estimating area of satellite imagery that have other applications as well: the uniform grid approach (validation), and the digitizing polygons approach (training).

Other uses for this approach may be:

a. Military- estimating the size / extent of enemy forces in a region.
b. Forestry- estimating the extent of a burn scar, beetle kill
c. City planning- projecting growth of urban development or calculating current rate of growth

Again, these are estimates, more precise readings may be needed for these applications.

Limitations and caveats:

A key limitation with the first section of this lab- using satellite imagery to determine human modification is that interpretation can be subjective; users of this imagery may interpret human modification differently in terms of both intensity and extent. This can result in different estimates for human modification, which should be acknowledged when reporting calculations. An example of this is that I may interpret landcover in a region of Asia as pasture, whereas a local who is much more experienced with the agriculture in a region, knows that it is cropland.

When using the uniform grid to estimate landcover, there are two sources that impact precision and potentially introduce inaccuracy into calculations. First, it can arise when linking dots to the correct landcover type as discussed above. Second, there is a component of human error when counting the number of dots- there may not be much if the uniformGridScale is set to 400m and there are only 70 dots however, at 100m, there are 1026 dots, and it becomes much more difficult to keep track of which dots are counted in their respective landcover type.

When using polygons and digitizing to estimate landcover, inaccuracy arises if the map zoom is not large enough, it is reduced as zoom increases, though the tradeoff here is that this takes more time and effort. In the extended analysis, I further compare this tradeoff.

For both of these approaches for calculating error, given the sources of inaccuracy identified, it is important that those using this approach acknowledge the sources of inaccuracy and treat their calculations as estimates. Steps can be taken to make the estimates more precise such as zooming in when digitizing.

Next steps:

An important consideration when using these two approaches is that users of these can double check their work by comparing the values for area and proportion of land cover across methods.

Both of these methods are user intensive; they involve someone manually doing all of these steps and getting the calculations. A powerful advance in remote sensing is automating this process and allowing the computer to make these estimates. This is known as unsupervised classification; the computer decides classes based on statistical analysis of the spectral characteristics of the image. (Source: ArcGIS Pro- see below)

Scripts, citations, and resources

- https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/overview-of-image-classification.htm#:~:text=Unsupervised%20classification%20is%20where%20you,class%20categories%20within%20your%20schema)
- Users/DavidTheobald8 'demo scripts' + demo videos
- Virene, J.W. 2022. Google Earth Engine script

Permalinks:

Part one (a): https://code.earthengine.google.com/1410a50c9026f14a991b266e06910cb6
Part one (b): https://code.earthengine.google.com/ae5f0a7ca244639eca86aa184670b0f7
Part two: https://code.earthengine.google.com/d8fbeed938cef1f403ce297ae731fae6

Part three(a): https://code.earthengine.google.com/c252c5cd56ce9f183c5eee5b878ba167
Part three(b): https://code.earthengine.google.com/f76ee44a6782e1e9fd2352c2553a4778

Script Path: (Parts 1-3, and extended analysis)
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabthree%2Flabthree_p1a
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabthree%2Flabthree_p1b
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabthree%2FLabthree_p2
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabthree%2Flabthree_p3a
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabthree%2Flabthree_p3b

Extended analysis:

I created another script and digitized landcover data for a different region in Colorado. The region that I digitized, which I will call Capitol-Snowmass due to the two fourteeners it contains is located in the Elk Mountain range, just outside of Aspen, Colorado. I obtained the coordinates for this region in Google Earth Pro. Also, this is one of my favorite recreation areas for backpacking and hiking.

Land Cover Type	Area Estimate (Sq m)	As proportion of total
Water	419556.715	0.004955988
Builtup	0	0
Barren	54799756.73	0.64731874
Forest	20822440.14	0.245963788
Shrubland	5795245.92	0.068455984
Grassland	2819527.998	0.0333055
Agricultural	0	0
Wetland	0	0
Total area:	84656527.51	1

One consideration with this is that the area that I digitized was significantly larger than the area in Rocky Mountain National Park. As a result, I digitized at a scale of 1000m instead of 200m because for the extent of this grid, digitizing at 200m would take hours.

As mentioned in the limitations section, there is a tradeoff between accuracy and time spent digitizing; as zoom increases (i.e., from 1000m to 200m), the map becomes more accurate, though digitizing landcover will take more time. Comparing these two landcover scripts, the accuracy of the RMNP data is more precise than the Capitol-Snowmass landcover however, the Capitol-Snowmass script took less time. Furthermore, the Capitol-Snowmass region is largely barren landscape, so depending on how we are using the map, it may be less important to identify small details.

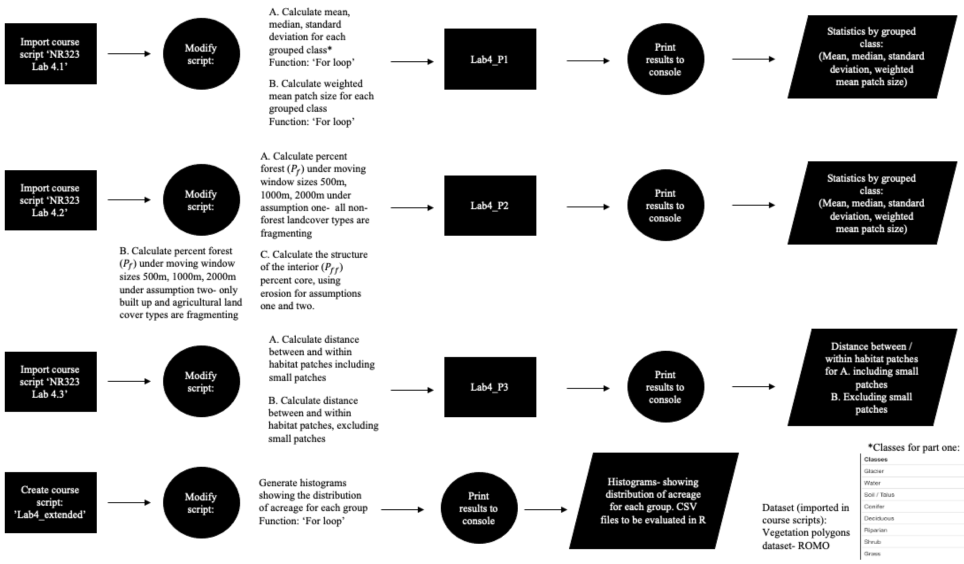
<u>Lab Four- Measuring Landscape Patterns</u>
Purpose:

- The purpose of this lab is to use landscape metrics to quantify geometric patterns that emerge in remote sensing data. We use these metrics in spatial analysis to quantify ecological / human aspects of interest.

Individual objectives for achieving this goal include:

- Using common metrics of landscape patterns, particularly forested landscapes
- Evaluating the mean, median, standard deviation of patches to determine their distribution and to understand how scale can impact these metrics

Overview:



Results:

Part I: Composition

Figure one: mean, median, standard deviation of patch area for each for each of the group vegetation classes:

Classes	Mean	Median	StandardDeviation	Skewness
Glacier	3.85	1.61	5.83	1.15
Water	7.59	1.28	65.57	0.29
Soil / Talus	9.19	1.88	72.43	0.30
Conifer	14.69	2.58	46.46	0.78
Deciduous	1.88	0.86	3.82	0.80
Riparian	8.26	1.38	26.12	0.79
Shrub	6.34	0.91	19.41	0.83
Grass	10.49	1.76	46.50	0.56

The assumption that polygons represent a homogenous unit of land under a comprehensive classification system does not entirely hold true. Factors that would contribute to this assumption being faulty would generally include human error in interpreting and classifying each land cover type and the scale at which the classification takes place. In the case of this lab, a large source of error is merging classes into more generic classes or splitting them into more specific ones. Any of these sources of error can lead to an area being classified as one landcover type, meaning the model assumes all of the area within the patch is under one specific class when in reality, the area represented by the polygon may have other landcover classes.

As an example, for clicking on one polygon, in the inspector tab Google Earth Engine may report that the landcover type is deciduous, but looking closer at the map, there is a small section of landcover within that polygon that is shrubland. This is a violation of the assumption that the land within the patch is homogenous. These sources of error can be reduced by digitizing at a smaller scale and paying close attention to detail when making classification judgement. Also, landcover type estimates can be more precise with a larger number of classes because the model is not lumping what are likely very different landcover types into a small number of classes, and if done correctly, more classes means that the classification is more representative of the actual landcover in the region.

Distribution evaluation:
In evaluating the distribution of each group vegetation classes, I used Pearson's Coefficient of skewness to determine the values that quantify the skewness of the data:
Formula: $skewness = \frac{3(mean - median)}{standard\ deviation}$
Given the context of this study and surrounding literature (cited below) on the interpretation of the values for Pearson's Coefficient, the following interpretation is defined:

$coefficient > |0.75|, strong\ positive\ (negative)\ skew$
$coefficient < |0.75|, weak\ positive\ (negative)\ skew$
** If value is positive, data is right skewed, if negative, data is left skewed.
$coefficient = 0,\ no\ skew$

Based on this, most of these group cover classes exhibit strong right skew in their distribution (see figure one), as their values for skewness are greater than 0.75. The classes that fall under this include glacier, conifer, deciduous, riparian, and shrub land covers.

There are a few classes with a weak positive skew, so they are more closely approximated by a normal distribution, though still not a perfect normal distribution because their values are different from zero. The classes that fall under this are water, soil / talus, and grass.

None of these classes have a perfect normal distribution.

Source: https://www.statology.org/pearson-coefficient-of-skewness-excel/

Within the context of this study, a normal distribution for the landcover types means that the frequency of values (acres) for patch size increase moving closer to the mean, there are few (if any outliers) and the data will not be skewed in any way. Right skew would suggest that there are some polygons in the study area that are larger than the mean size of the estimated landcover type. In this study, it makes sense that the landcover types with strong right skew have this distribution if there are many smaller patches, and then a small number of large patches. For the groups with weaker right skew, their patches have smaller outliers. Visually, this would mean there are a large number of polygons with similar areas- a high frequency of patches near the mean value and few polygons that are different from the mean in either direction. Because they still have a weak skew as defined above, there are likely still outliers present that are influencing their distribution and letting it take on a shape that is not perfectly normal.

Figure Two: Mean patch size, and weighted mean patch size

Mean vegetation patch size (MPS)	Weighted-mean patch size (WMPS)
11.82	122.85

The weighted mean patch size is significantly larger than the mean patch size.

Part II: Structure

In this figure, the percentage of forest cover is calculated with moving window sizes of 500m, 1000m, and 2000m. The second column 'Assumption One' assumes that all non-forested cover classes are fragmenting, and the third column 'Assumption Two' assumes that only urban and agricultural land cover classes are fragmenting.

Figure three: Proportion of forest at different moving window sizes.

Moving window size	Assumption One	Assumption Two
500m	0.602717625548402	0.9982551717400944
1000m	0.5973228192086085	0.9989920579020231
2000m	0.6502955228649019	1

Based on the differences in percentage of forest cover between the first and second assumptions, it is clear that using the second assumption yields a significantly higher percentage in forest cover compared to the first assumption. The reason for this is that the second assumption is much less strict in what fragments the forest, there are only two landcover types that fragment forest habitat whereas in the first assumption, any non-forest landcover type fragments.

This consideration is important in the design of ecological or human impact studies because the parameter of what fragments forest can significantly impact the value for percentage of forest landcover. Furthermore, depending on the study it may be important to fragment for different landcover types and make the definition of forest either more or less strict.

The mean P_f values also change based on the size of the moving window. From 500m to 1000m, the mean P_f value decreased, then moving to 2000m, it increased relative to 500m and 1000m. The neighborhood size impacts these values because it is based on the nearby values and uses these as context for the location of interest.

<u>Figure Four:</u>

This calculates the structure of the interior of forested areas, which accounts for the fact that other land uses, and the usage and processes taking place in those areas can impact the edge of forest habitats, so the model adjusts to account for this.

Assumption One:	**Assumption Two:**
0.43663397489335026	0.9815359131410921

Accounting for erosion at the edge of forest habitats will reduce the percentage of forest land cover for both assumptions one and two. This is because the edges of forest habitats are removed to account for their location being on the fringe of forest landcover and some other landcover type.

Another noticeable point is that the percentage of forest landcover is reduced much more significantly in assumption one compared to assumption two because the forest is on the edge with many landcover types, whereas in assumption two, the calculation is only accounting for the edge between urban and agricultural landcover types.

Part III: Process / Function

<u>Figure Five: Mean GIS fragmentation distances for between and within, **including small patches**</u>

Within	1675.153836904637
Between	1411.6115223493928

<u>Figure Six: Mean GIS fragmentation distances for between and within, **excluding small patches**</u>

Within	338.2290822737853
Between	588.0143921074633

Comparing these two figures, changing the definition of forested pixel patch by removing small patches will cause the landscape to become more connected. The interpretation of mean GIS fragmentation states that as the value increases, it indicates lower connectivity and therefore higher fragmentation. When small patches (those less than 10 hectares) were removed, the values for between- distance away from the forest edge, and within- distance towards the patch core both decreased significantly.

The interpretation of this result is that in GIS studies, it is important to consider the definition of what qualifies as habitat and what does not. This can depend on a number of factors such as the species in question, the ecology of the region, or others. As a practical example, small habitats may be included as suitable habitat in this study area if the species in the study is falcons or mice, but would not be included for wolves, because a patch size under 10 hectares is likely too small.

Defining a habitat removing small patches in this example decreased the habitat fragmentation; demonstrating that small patches are in this case, more susceptible to fragmentation.

Discussion

Section I of this lab generates statistics about patch size of the patches for each landcover type that include the mean, median, standard deviation, and weighted mean patch size. These statistics are useful in determining if there is skew in the frequency distribution of patches. Pearson's coefficient of skewness was used to calculate a coefficient to quantify the degree of skew for each group's frequency distribution. These coefficients showed that all observations had some level of right skew, meaning that the mean exceeded the median. The statistical interpretation of the observed right skew in these groups is that there are outliers whose patch size exceeds the mean patch size by some degree depending how heavy the skew is (a higher coefficient means more extreme skew). Using the landcover class conifer as an example, at Rocky Mountain National Park, this would mean that there are a high number of conifer classes taking on values near the mean, and a few very large patches of coniferous forest above the mean that skew the data.

Section II looks at forest land cover and how different assumptions and techniques can impact the proportion of forest in Rocky Mountain National Park. In the script, percentage of forest was calculated under two assumptions. The first set of calculations follows the assumption that all non-forest land cover classes are fragmenting and returned values for forest integrity (P_f) with different neighborhood sizes for the moving window. The second set of calculations is the exact same, except it alters this assumption so that only agricultural and built-up land cover classes are fragmenting. Changing from assumption one to assumption two made the criteria for forest landcover less strict and resulted in a larger proportion of forest (P_f). Changing the radius variable in the script changes the neighborhood size of the moving window, which affects the values for forest integrity. One more assumption that was assessed in this script draws on the concept that there is erosion at the fringes of forest landcover that is adjacent to other landcover types; this is because the forest land in these areas is likely impacted by the other uses. This returns values interpreted as the percent that is "core" or not fragmented (P_{ff}). These values will be lower than (P_f).

Section III uses measures of distance to determine the structural connectivity from patches- the degree of fragmentation (GISFrag). 'Distance between' is the distance from the edge of forest patches into non-forest areas, 'distance within' measures distance into the forest patch from the edge. The interpretation of these is that as the value for GISFrag increases, there is higher fragmentation and lower connectivity. In the script, the values of GISFrag between and within were calculated first, including small patch sizes < 10 hectares, then were calculated again removing these small patches, which caused the landscape to become more connected as shown through a decrease in the values for GISFrag. The importance of patch size in this context is that it can affect the degree of habitat fragmentation. As mentioned in the results section, careful study design that is sensitive to the ecology and species in question will ensure that measurements are accurate and representative for the needs of the study.

Appropriate uses

Part I: Determining the extent of landcover by class is an important tool for many reasons. Practical applications for doing so may include:

- (Wildlife biology): Determining the extent of habitat for different species
- (Dendrology): Calculating the size of forests for different tree species
- (Firefighting): Assessing wildfire risk and potential severity

The distribution of patch size within different classes can be used in all of these applications to determine the overall heterogeneity of a region with respect to each class type.

Part II: The applications above are likely to be extended to determine the percentage each landcover type, and different applications and study designs can define assumptions about what fragments habitats, which will change this percentage. In the script from this lab, the percentage of forest changed depending on whether habitat was fragmented by all other landcover types, versus only by agricultural and built-up land. Ecologists may define different landcover types that fragment habitats for a species they are studying.

Part III: Looking further into habitat fragmentation, the GISFrag metric is useful in understanding species extinction risk; a study found that one of a components of comparative extinction risk modelling is habitat pressures on species geographic ranges (Ramirez-Delgado et al. 2022).

Limitations and Caveats

One key limitation is the concept of lumping / splitting classes as done in section one. Initially, there were 42 different landcover classes which were then grouped into 8 classes. When grouping classes like this, it is important to acknowledge that the statistics will be less precise because these 8 classes are not representative of all of the different landcover types in the heterogenous landscape of Rocky Mountain National Park. The model of this script is generalizing these landcovers into eight classes when in reality, there are many more. Another limitation is that human error in interpreting landcover types can also lead to inaccuracy of classification. In both of these cases, the result of this inaccuracy is an incorrect classification of landcover. As an example, shrubland may be classified as wetland and the model overestimates the area for shrubland and underestimates the area for wetland.

For section two of the lab, moving window sizes impacted the value for proportion of forest. It is important to choose appropriate values for the moving window in order to get accurate calculations. This is supported by study in *Ecologist Processes* which found that window size does have an effect on prediction accuracy when calculating forest structural variables including stem number, basal area, volume, and mean height for different tree types. Source: (Ozkan and Demirel, 2021).

In section three, the cutoff for small patches changes the fragmentation of a habitat; this cutoff must be defined correctly based on the needs of the study, otherwise the habitat fragmentation will be incorrect. Another consideration is that this is another type of study that depends on the correct classification of landcover types; it is important to ensure that these landcover types are correctly classified, otherwise habitats that should be fragmenting will be incorrectly classified and the degree of fragmentation will be underestimated. An example of this may be a residential area in an otherwise heavily forested area; if this is not identified as built up, it will not be factored into fragmentation of the forest land cover.

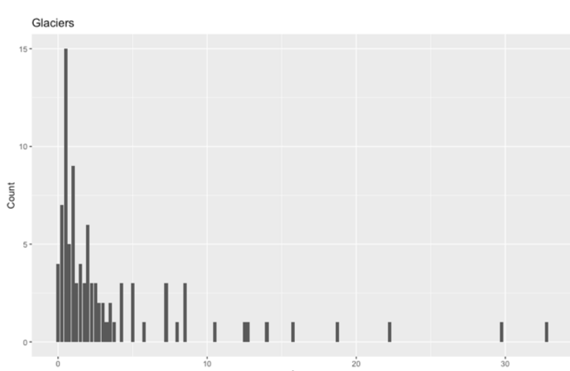
Next Steps

The scripts that we have explored are incredibly important in understanding landcover classes and the factors that contribute to their composition, extent, and fragmentation. This data can be used in many applications.
One important next step onward from the analysis conducted so far would be to use other software packages such as RStudio to validate the results (especially the normality assessments) shown in Google Earth Engine and even conduct other analyses.

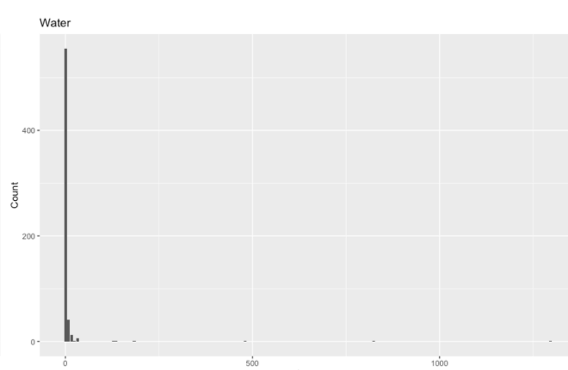
Extended Analysis:

Using Pearson's coefficient to evaluate skewness of the acreage for each class is only one approach, there are many others that can be used to demonstrate skew. Another way to do this is to generate histograms to visualize the skew in each group.
The following figures show the frequency distribution of acreage by class.

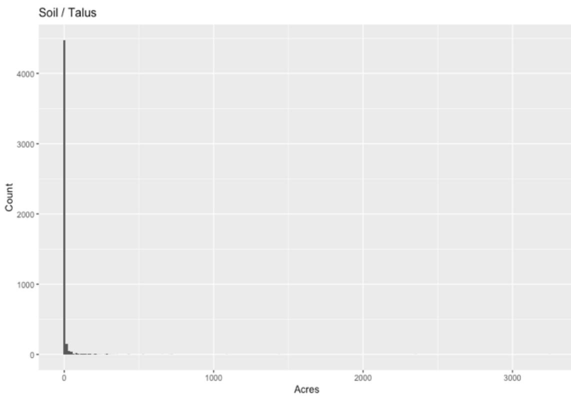
Group 0: Glaciers



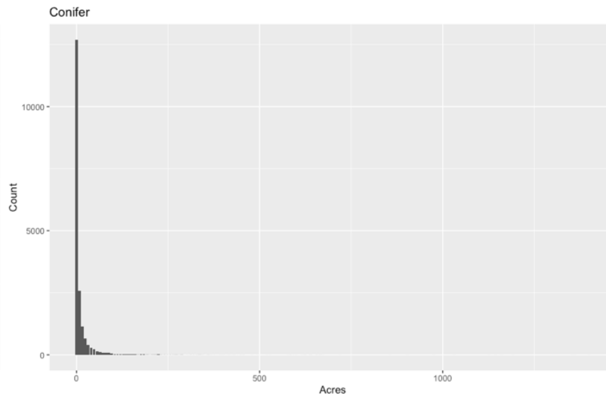
Group 1: Water



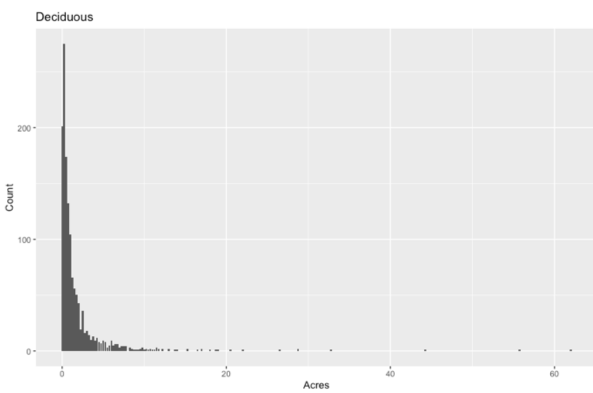
Group 2: Soil / Talus



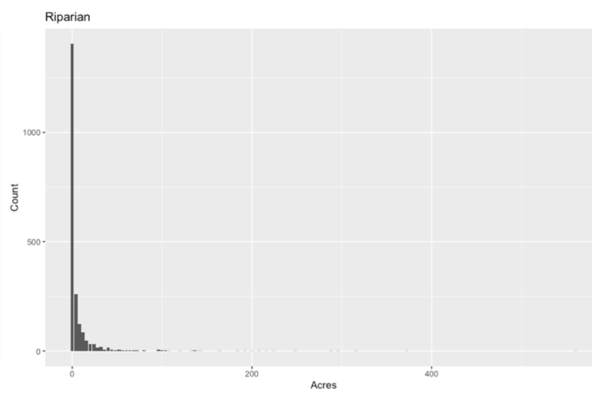
Group 3: Conifer



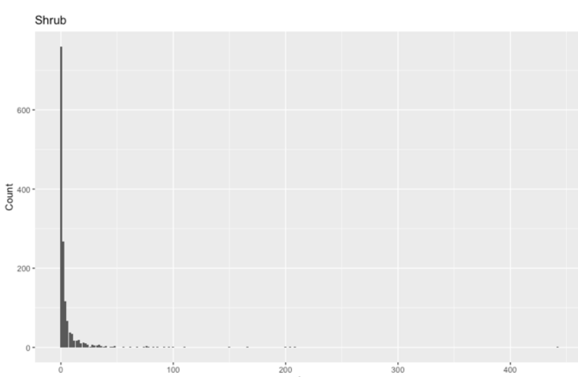
Group 4: Deciduous



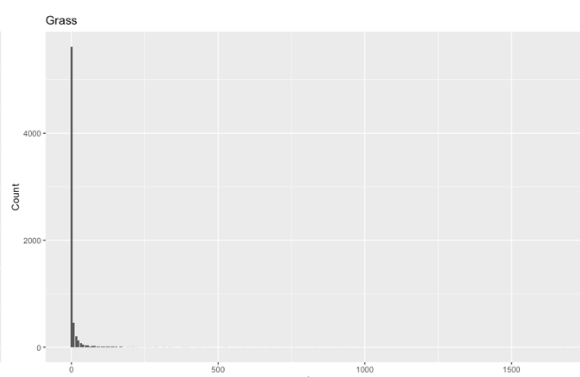
Group 5: Riparian



Group six: Shrub



Group 7: Grass



Next, using each group's values for acreage weighted by frequency, the skewness can be evaluated in R, using the "e1071" package, and the values compared against those from Pearson's coefficient of skewness.

values	finalskew	classes
0	1.011247	Glacier
1	1.314738	Water
2	1.181338	Soil/Talus
3	2.869612	Conifer
4	1.037504	Deciduous
5	1.431155	Riparian
6	1.138364	Shrub
7	1.627157	Grass

These graphs, and this table support the estimates for Pearson's coefficient of skewness because each class exhibits right skew.

Scripts, citations, and resources
Permalinks: (Parts 1-3, and extended analysis)
https://code.earthengine.google.com/89c0ffad2190a0b4d64284acfaf496f6
https://code.earthengine.google.com/65a05bb97912fb5a51a102ee39241b20
https://code.earthengine.google.com/5781c3334dd1bc5c00444d609ded7a7c
https://code.earthengine.google.com/1bee886fd6da5067532d993f1741ee80

Other resources:
https://www.statology.org/pearson-coefficient-of-skewness-excel/
https://www.nature.com/articles/s41467-022-28270-3.pdf
https://ecologicalprocesses.springeropen.com/track/pdf/10.1186/s13717-021-00330-4.pdf
RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Users/DavidTheobald8 'demo scripts' + demo videos
Virene, J.W. 2022. Google Earth Engine script

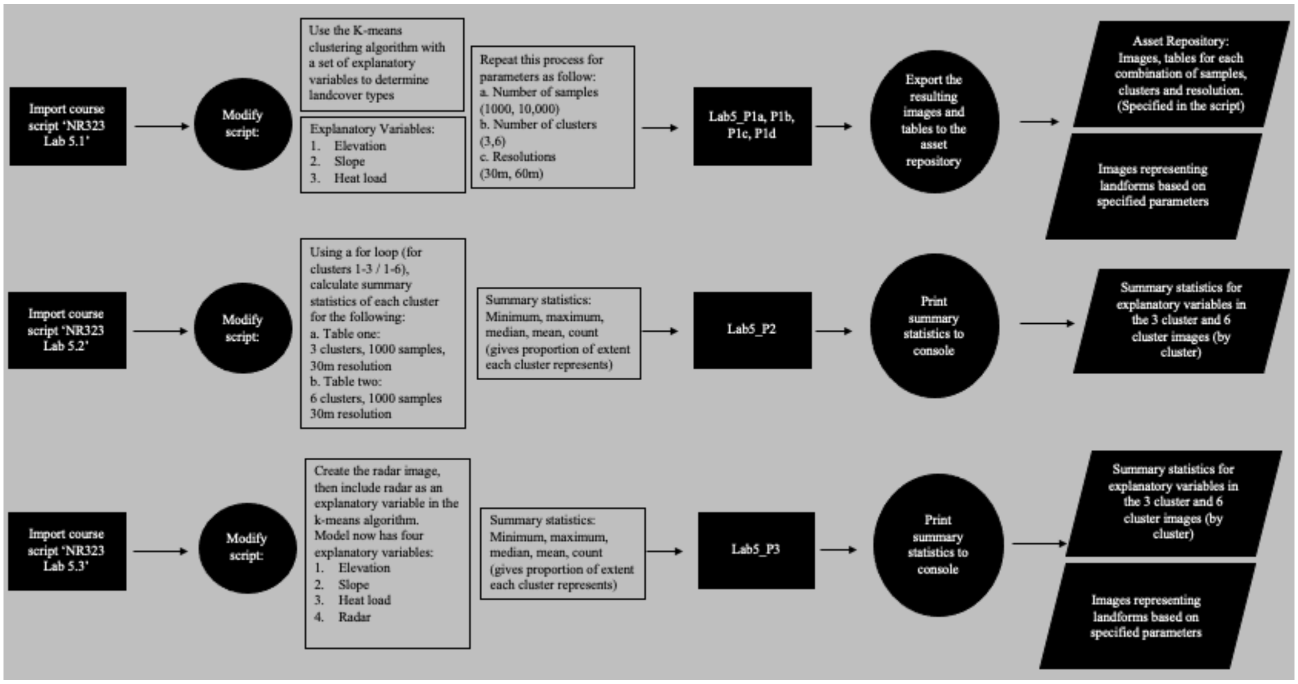
Script Paths: (Parts 1-3, and extended analysis)
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3ALabfour%2Flab4_p1
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3ALabfour%2Flab4_p2
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3ALabfour%2Flab4_p3
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3ALabfour%2Flab4_extended

Lab Five- Unsupervised Classification of Landforms by Clustering

Purpose:
The purpose of this lab is to learn how to use Google Earth Engine for unsupervised classification. In the process of unsupervised classification, the platform is trained based on samples for the variables of interest and it extrapolates the data from these samples to the entire defined region.

Individual objectives for achieving this goal include:
- Understanding the workflow of unsupervised classification
- Using the k-means clustering algorithm
- Summarizing the classes which were clustered and mapping landforms

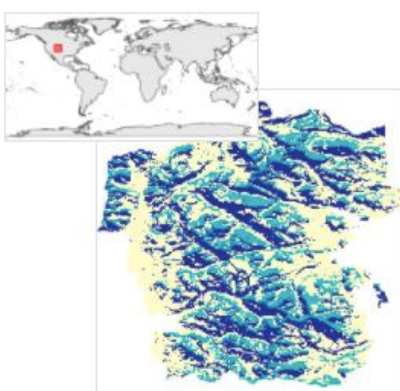
Overview:



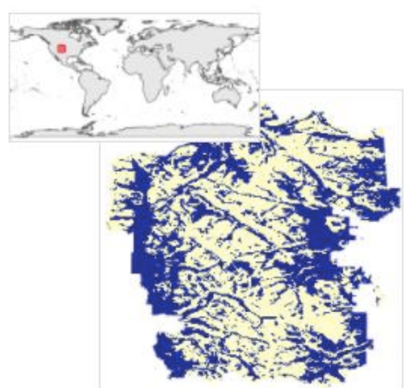
Results:

Part I: Cluster terrain variables into landform types and compare visually

<u>Image one:</u> 3 Clusters, 1000 samples, 30m resolution

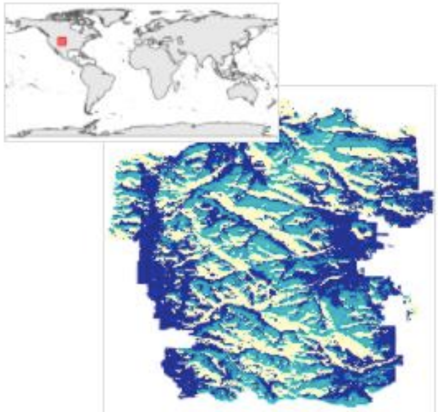


<u>Image two:</u> 3 Clusters, 1000 samples, 90m resolution



<u>Image three:</u> 3 Clusters, 5000 samples, 30m resolution

<u>Image four:</u> 3 clusters, 1000 samples, 30m resolution, heat load removed



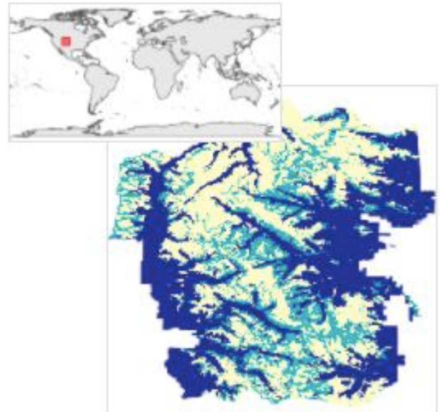
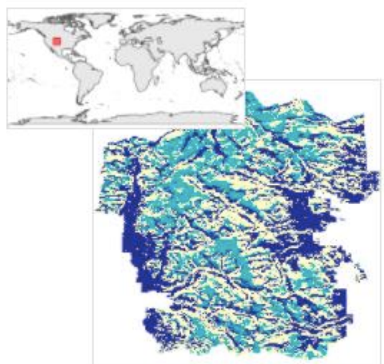


Image five: 6 Clusters, 1000 samples, 30m resolution



Interpretation / Questions:

1. Changing the resolution from 30m to 90m created a coarser image. The 30m resolution means that the model is clustering over samples whose cell size is $30m^2$, whereas the 90m resolution clusters over samples whose cell size is $90m^2$. Because the latter model is generating an image given cells that are large in value, the image will be less detailed and less precise.
2. Increasing the number of samples from 1000 to 5000 appears to have inverted the colors between the two images. This effect is likely due to more accuracy in the image based on a larger sample size. The model is generating the image and data based on the data for explanatory variables under the defined number of samples. When the number of samples increases, there is more data. Unsupervised classification models are extrapolating the data from the number of samples to the population; as the number of samples increases, it becomes more representative of the population and therefore more accurate in defining landforms.
3. Removing the heat-load variable has also changed the classification of this map. Again, the model uses the explanatory variables for the defined number of samples to generate these maps, when the heat-load variable is removed, this is no longer factored into the model so it will change how the map is classified.
4. When the number of clusters was increased from three clusters to six clusters, this has the effect of making the map more accurate in defining landforms. The reason for this is that the samples are now grouped across six clusters instead of three. One particular change that is very important is that the edges of the map in image one (3 clusters, 1000 samples, 30m resolution) are tan whereas, in image five (6 clusters, 1000 samples, 30m resolution) are blue. What changed is that samples are now being clustered / grouped

differently and the edges highlight how increasing the number of clusters makes the samples more representative of the landforms in those areas.

Part II: Compare Quantitatively
<u>Table one:</u> 3 Clusters, 1000 samples, 30m resolution

	Heat load	Elevation	Slope	Proportion (% of total extent)
Median, Cluster One	160	3083.77	11.94	46.3
Median, Cluster Two	101.5	3326.11	27.42	24.8
Median, Cluster Three	221	3357.59	25.66	28.9

<u>Table two:</u> 6 clusters, 1000 samples, 30m resolution

	Heat load	Elevation	Slope	Proportion (% of total extent)
Median, Cluster One	104	3180.56	25.86	20.2
Median, Cluster Two	86.5	3552.90	37.97	5.8
Median, Cluster Three	223	3153.93	24.89	14.3
Median, Cluster Four	162	3417.96	13.19	25.6
Median, Cluster Five	231	3468.39	31.61	10.7
Median, Cluster Six	163	2859.032	10.05	23.4

Question three:
Ranking the explanatory variables in order from most to least important, the order is: heat load, slope, elevation.

- **<u>Heat load</u>** is the most explanatory variable because the differences in heat load across terrain types are very apparent on the map. This is because there are many factors across the heterogenous landscape that impact heat load and draw out these differences. An example of this is comparing heat load on different aspects of a mountain. For example, looking at the differences between a north facing and south facing slope make it very clear that the feature in question is a mountain. This applies to other terrain types as well, there are heat load differences between roads and rivers, and differences between high elevation mountain tops and lower elevation valleys. All of these differences are captured very well by heat load.
- **<u>Slope</u>** is the second most important explanatory variable because it becomes apparent on the map for any change in elevation across a landscape. This variable can highlight mountains and features including gullies / couloirs, cliff bands, and ditches. There are many features for which slope can illustrate on the map.
- **<u>Elevation</u>** is the least important on this list of explanatory variables. This does not mean that elevation is not an important variable, though of these three it has the least power in assessing features on the landscape. Elevation can illustrate large changes such as the elevation difference between a valley and a mountain top, but there are many features in a landscape that elevation cannot capture due to there being only a small change in elevation across the feature.

Part III: Using Radar Imagery
Image:

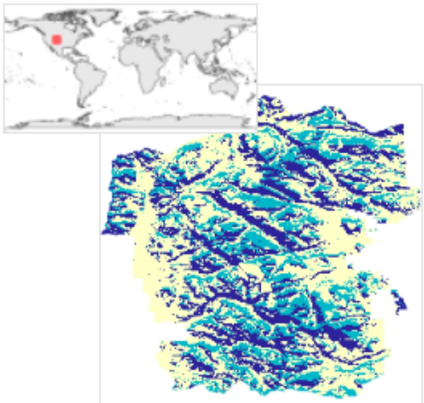


Table 3: 3 Clusters, 1000 samples, 30m resolution
Using the Copernicus satellite (radar) imagery and subsequent dataset

	Heat load	Elevation	Slope	Radar	Proportion (% of total extent)
Median, Cluster One	162	3138.75	12.32	12.32	49.7
Median, Cluster Two	99.5	3305.17	27.42	27.42	24.4
Median, Cluster Three	224	3326.12	27.66	27.67	25.9

Table Four: Standard Deviation of variables by cluster

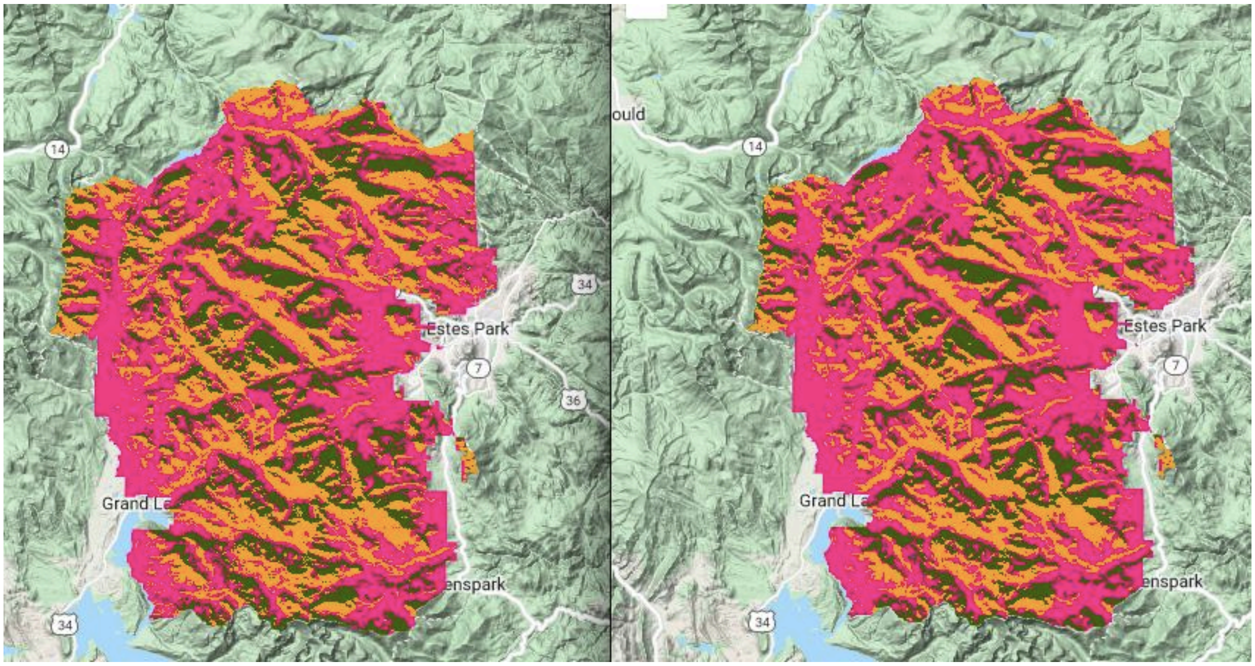
	Heat load	Elevation	Slope	Radar
Standard Deviation, Cluster One	26.46	329.03	6.10	6.10
Standard Deviation, Cluster Two	30.05	304.86	10.50	10.50
Standard Deviation, Cluster Three	24.73	307.22	9.58	9.58

Table Five: Minimum and maximum values of variables by cluster

	Heat load	Elevation	Slope	Radar
(Min, Max) Cluster One	(98,220)	(2352,3084)	(0,31.94)	(0,31.94)
(Min, Max) Cluster Two	(4,166)	(2529,4276.91)	(9.93,64.74)	(9.93,64.74)
(Min, Max) Cluster Three	(154,255)	(2492,4007)	(9.05,64.41)	(9.05,64.41)

Question two:
Adding the radar imagery changes the landform classes because the model is now including the radar imagery and using it as a factor in determining the landforms within the park boundary.

Image six: Visual comparison of the model without radar (left) to the model with radar (right)



Question two (continued):

Based on the visual inspection, it is tempting to draw the conclusion that these models are the same and state that radar does not have an impact however, this is not the case. There are small differences in the map that are brought out by including radar as a covariate. The radar image makes the map on the right more precise in estimating landscape features because it introduces variables outside of those derived from the DEM (elevation, slope, and heat load).

This is supported quantitatively by the change in the percentage of total map extent attributed to each cluster. In the radar image, a larger percentage of map extent is accounted for in cluster three. Comparing these two maps with the satellite image, and then against each other, it is clear that the radar image better matches the satellite image. With this said however, the difference between these two maps is not very large.

Discussion:

Section I of this lab uses the k-means clustering algorithm to classify terrain based on a set of defined explanatory variables. The variables used in this section of the lab were slope, elevation, and heat load (derived from the DEM). The stack of explanatory variables was then used in the clustering process to classify the terrain. From here, the script generates layer, images, and tables to show terrain classification visually, and quantitatively. Another important element of this section is that the images and tables were calculated for different combinations wherein sample size, number of clusters, and resolution changed. These parameters changed the resulting image output, as well as the summary statistics derived from the tables in the asset repository. Increasing the number of samples increases the precision of the model because it becomes more representative of the true nature of the landforms in the region. Increasing the number of clusters means grouping samples across a smaller range which to a certain extent, can also increase precision. Last, reducing the resolution means the cell size over which the samples are drawn becomes smaller which can also increase precision. These factors all impact how representative the model is of the actual landscape in the region, so they are important to consider when creating these images.

Section II of this lab reports the summary statistics for each explanatory variable by the number of clusters over which the image was classified. These summary statistics were calculated for the images and tables with 1000 samples, 30m resolution, for 3 clusters and 6 clusters. These statistics provide a quantitative means of evaluating the results from section I. The importance of each explanatory variable in the model can be determined based on both the images and the summary statistics (i.e., min, max, med, mean, standard deviation, and count- % of total extent). By evaluating the summary statistics, as well as comparing the image output to the satellite imagery, the ranking in order of importance (most to least respectively) of the explanatory variables was heat load, slope, and elevation. Despite this order, all these variables have their value in conveying information about landforms in the region, so elevation being last on the list does not mean it is not important.

Section III introduces a new explanatory variable to the model- 'Radar' which is a radar image from Sentinel GRD. This variable is included as an additional covariate into the k-means clustering algorithm, so the model now evaluates and classifies terrain over the four covariates by number of clusters instead of three. Similar to previous sections, the script generates layers, as well as creates a table and image to be placed into the asset repository. Next, summary statistics (min, max, med, mean, standard deviation, and count- % of total extent) were calculated for each of these covariates by the number of clusters (3, and 6). Though visual inspection and statistical analysis, the layer for the model with 3 explanatory variables in the original model can be compared to the model that introduces radar as an additional variable in order to determine which model best explains the terrain in the region. Although the radar model is slightly more accurate, overall, the difference in these models is not very large. Adding the radar variable is important because this model incorporates data from the radar image, in addition to the variables derived from the DEM. This additional variable means there are two data sources rather than just one, which will make the model more robust.

Appropriate Uses:

Clustering and unsupervised classification are incredibly useful tools with a broad variety of applications. In lab three, landcover was classified by digitizing. Digitizing, although a useful tool in GIS, can be tedious and time consuming. Unsupervised classification is a means of reducing the effort in classification by using algorithms and statistical models to automate the process of classification.

This lab uses the k-means algorithm to evaluate terrain in the Rocky Mountain National Park region, but this could be applied to other situations including different regions and evaluating + classifying for other study purposes. Other studies might use this algorithm to classify landcover, habitat suitability, or even wildfire damage over a study area. An example of another use of the k-means clustering algorithm is in evaluating geyser eruption segmentation in Yellowstone National Park, WY. This study grouped geysers into clusters based on eruption time, and time between eruptions (Dabbura, 2020).

Limitations and Caveats:

The first key limitation of unsupervised classification is that it can become **computationally expensive**. When inputting data into the model and specifying parameters such as resolution, clusters, and number of samples, this can impact the time it takes to generate the image, layer, and tables needed to analyze the model. In this lab, the parameters did not take very long to run however, in professional settings, the datasets will be larger, and the models will be more complex. An example of such a study may be monitoring the climate impacts from a large number of emission sources across a specified region or the entire globe.

Another limitation of machine learning is that the model and results can only be as good as the **data inputs** for that model; if the model uses data that are not accurate or representative of the region / study purpose, the results will be inaccurate. Furthermore, **poor model design** can further worsen the issue and contribute to inaccuracy within the model. These considerations are important when using unsupervised classification; if these sources of error are present, the model will be poor so they should be accounted for and limited (Stewart, 2020).

Next steps:

These models give predictions for the landscapes of the study area in Rocky Mountain National Park. A good next step from here would be to validate these results by going out into the field and comparing the terrain predictions given by the model to what is actually there. Another model to compare this output to would be a digitized model of landscapes in the study area.

Sources, Scripts, Citations

Part one: (a-d)

https://code.earthengine.google.com/343afa046ce95dc95e7cb469661394b4
https://code.earthengine.google.com/e9275c4f8fb906b194265753dca3a30e
https://code.earthengine.google.com/b2017b08bd98f9b7f08398b42eea3e20
https://code.earthengine.google.com/62b1cb3f3067a5c136a651283f6e5e8a
Part two:
https://code.earthengine.google.com/4bad8dea27a944dbfd3c04b3116c330a
Part three:
https://code.earthengine.google.com/df83e87a57aa5342d8404b30769ec65c

Dabbura, I. (2020, August 10). *K-means clustering: Algorithm, applications, evaluation methods, and drawbacks*. Medium. Retrieved April 15, 2022, from https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

Matthew Stewart, P. D. R. (2020, July 29). *The limitations of machine learning*. Medium. Retrieved April 15, 2022, from https://towardsdatascience.com/the-limitations-of-machine-learning-a00e0c3040c6

Script Paths (Parts 1 (a-d), 2, and 3)
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabfive%2Flab5_p1a
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabfive%2Flab5_p1b
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabfive%2Flab5_p1c
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabfive%2Flab5_p1d%20(ext.)
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabfive%2Flab5_p2a
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabfive%2Flab5_p2b%20(ext.)
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3Alabfive%2Flab5_p3

Extended Analysis:
This section evaluates how increasing the sample size can change the map data- the images / layers generated, in addition to the summary statistics within the model:

Image seven: Layer

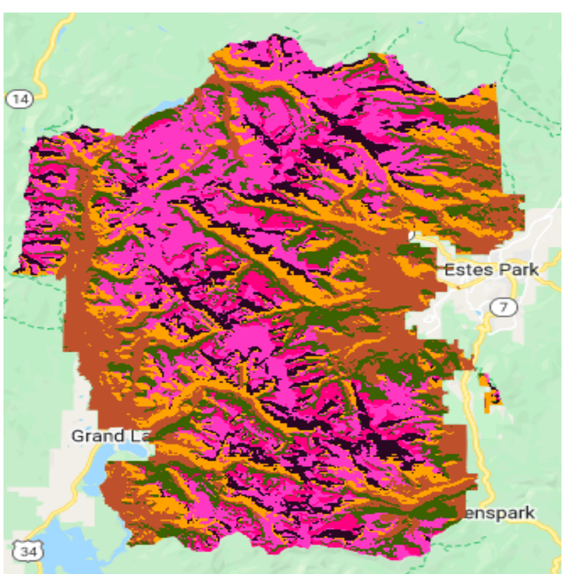
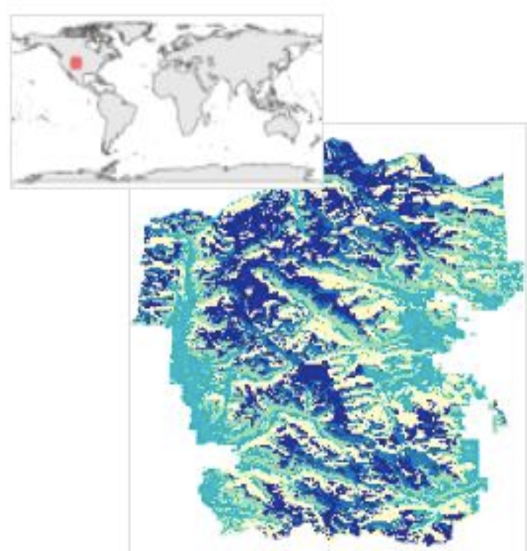


Image eight: Image asset



Median values:

	Heat load	Elevation	Slope	Proportion (% of total extent)
Median, Cluster One	80	3525.17	34.62	8.93
Median, Cluster Two	108.5	3091.88	22.37	17.36
Median, Cluster Three	222	3084.25	24.61	16.84
Median, Cluster Four	165	2761.77	8.93	19.96
Median, Cluster Five	233	3540.09	31.13	10.50
Median, Cluster Six	162	3411.78	12.89	26.20

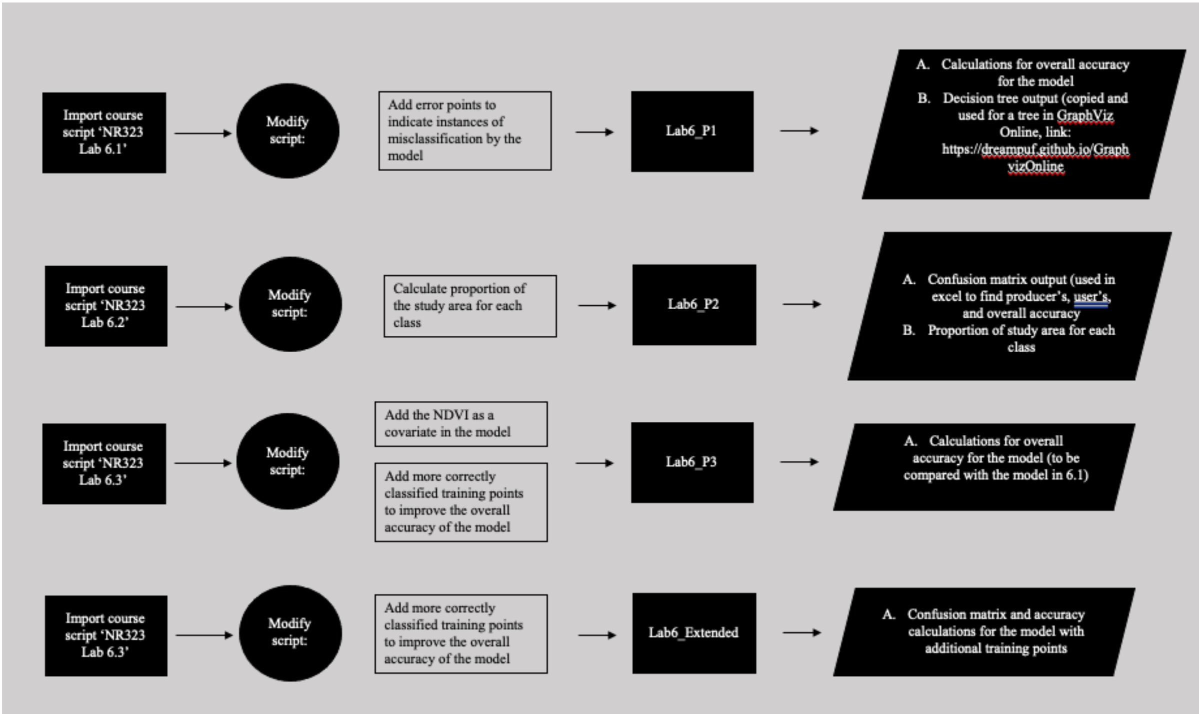
Changing the number of samples makes the model more representative of the landscape of the region. This is supported by comparing the layer image with satellite imagery, in addition to evaluation with the summary statistics, including the median as reported above.
When running this analysis, generating the image and table assets took longer (around 5 minutes compared to 1-2 minutes in previous models), which demonstrates how increasing the number of samples makes the analysis more computationally taxing.

<u>Lab Six- Supervised Classification of Land Cover</u>

Purpose:
The purpose of this lab is to use Google Earth Engine for unsupervised classification. The process of unsupervised classification entails creating training points for the different groups into which the map will be classified; then providing these to the supervised classification algorithm and creating a map that classifies the region of interest.

The individual objectives for achieving this include:
- Understanding supervised classification workflow
- Using and displaying categorical and regression trees
- Evaluating the accuracy of the classification using the confusion matrix
- Mapping canopy cover and impervious surface

Overview:

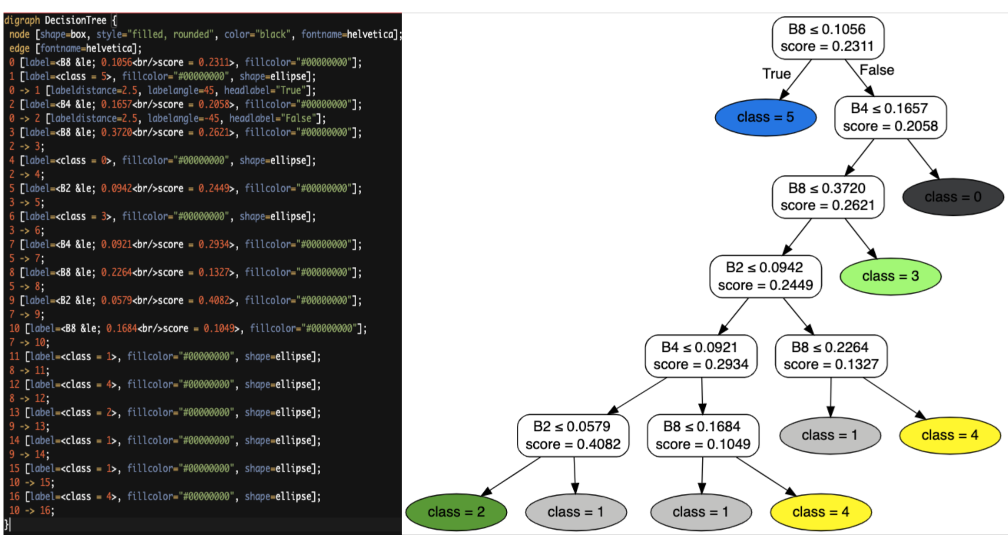


Results:

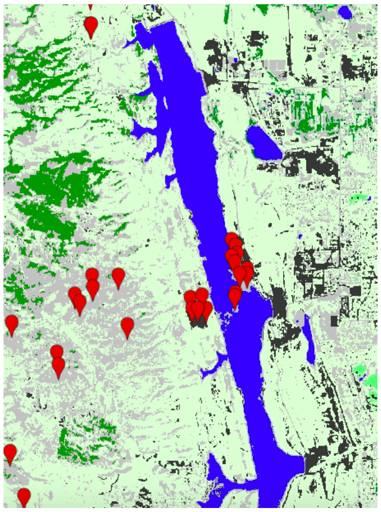
Part I: Training supervised classification using interactive image interpretation

Questions:

1. The primary variables / covariates in this model that are the most important are in order from most important to least important B8, B2, B4, B3
   This is given by both the importance value in Earth Engine, as well as the order these bands appear (or don't in the case of B3) on the regression tree.
   Classification Tree:



2. There are many instances of classification error within the study area:



- On the west end of Horsetooth reservoir, the land is likely either pervious tree or pervious nonirrigated, though it is classified by the model as impervious nonbuilding. There are houses in this area, though not as densely clustered as the model would indicate.
- In Horsetooth reservoir, what should be classified as water is classified by the model as pervious nonirrigated
- There is a road on the north end of Horsetooth reservoir; some of the cells are correctly classified as impervious nonbuilding, while other cells on this road are classified as pervious nonirrigated.
- In one of the bays of Horsetooth reservoir, there is an area that the model classifies as impervious building though it should be classified as impervious nonirrigated. Looking at the Sentinel II imagery, there is a cloud above the area, which is a potential explanation for this incorrect classification.

Visually, these errors are highlighted in the map

In terms of classes and their overall accuracy on the map, **there is lots of error between classification of impervious non-building, and pervious non-irrigated**. Due to this being a large source of error in the model, a means of improving its accuracy would be to create more training points for these landcover types. This will let the machine learning algorithm get a better sense for the band reflectance of each of these landcover types and lead to less error and more accurate classification.

3. The accuracy of the results generated by the training data is 100% because the accuracy of the figure is assessed with respect to the data that is already included in the model.

Part II: Accuracy Assessment

1. The overall accuracy of the map is 54%, this is too low to say that this is a good map of impervious / pervious surfaces and improvements within the training data can and should be made in order for the map a to provide more accurate representation. A level of accuracy between 70%-80% as in the NLCD maps would be sufficiently accurate to state that the map provides an accurate representation of impervious / pervious surfaces.
2. Confusion matrix: Looking at the validation confusion matrix, there is a large amount of error to be attributed to the classes impervious nonbuilding, and pervious nonirrigated. This supports the findings from visual interpretation above, because these two classes had many instances of misclassification.
   *Included producers' and users' accuracy on these matrices as extended analysis.

Confusion matrix										
Matrix one: Based on training data										
	Impervious Building	Impervious nonbuilding	Pervious Tree	Pervious Irrigated	Pervious Nonirrigated	Water	**Totals**	User's Accuracy		Total accuracy:
Impervious Building	13	0	0	0	0	0	13	100.00%		100.00%
Impervious Nonbuilding	0	16	0	0	0	0	16	100.00%		
Pervious Tree	0	0	5	0	0	0	5	100.00%		
Pervious Irrigated	0	0	0	13	0	0	13	100.00%		
Pervious Nonirrigated	0	0	0	0	18	0	18	100.00%		
Water	0	0	0	0	0	22	22	100.00%		
Totals	13	16	5	13	18	22	87			
Producer's Accuracy	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%				
Matrix two: Validation										
	Impervious Building	Impervious nonbuilding	Pervious Tree	Pervious Irrigated	Pervious Nonirrigated	Water	**Totals**	User's Accuracy		Total accuracy:
Impervious Building	9	8	0	0	7	0	24	37.50%		54.21%
Impervious Nonbuilding	11	5	0	0	9	0	25	20.00%		
Pervious Tree	0	3	6	0	1	0	10	60.00%		
Pervious Irrigated	0	1	2	12	0	0	15	80.00%		
Pervious Nonirrigated	0	4	1	0	10	0	15	66.67%		
Water	0	2	0	0	0	16	18	88.89%		
Totals	20	23	9	12	27	16	107			
Producer's Accuracy	45.00%	21.74%	66.67%	100.00%	37.04%	100.00%				

3. Table: Proportion of study area for each class

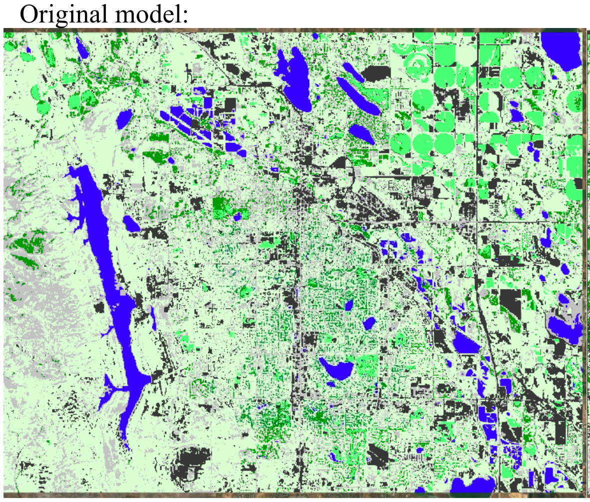
Class	Impervious Building	Impervious nonbuilding	Pervious Tree	Pervious Irrigated	Pervious Nonirrigated	Water
Proportion of study area:	12%	20%	5%	6%	50%	6%

Part III: Improving the Classification Model

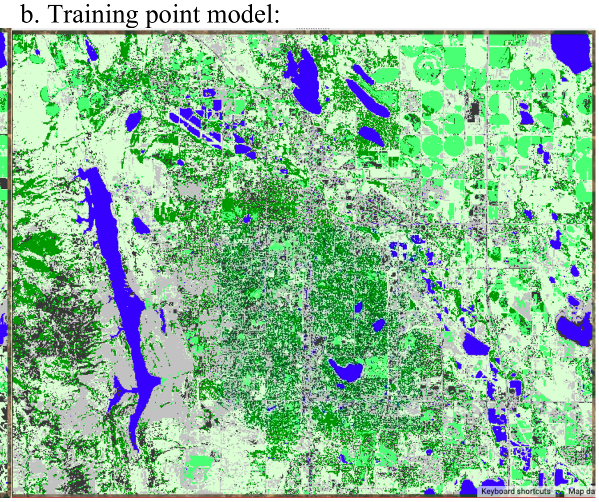
1. Adding, moving, or deleting points will result in a better classified map provided that the points being changed overall make improvements to the training data. This means points are placed into the correct classification category. This is shown by changes in the overall accuracy between the original model and the new model with additional training points.
   Original model overall accuracy = 55%
   New model accuracy = 59%

Visually:

a. Original model:



b. Training point model:



One visual note here, is that there is still a large amount of error for the impervious nonbuilding on the south end of Horsetooth reservoir, this can be improved by making more training points for impervious nonbuilding, as well as the true landcover type of that area, pervious non irrigated. This will be explored in the extended analysis.

2. Including NDVI in this model does slightly improve the accuracy. In the map without NDVI in parts I and II, the overall accuracy without any alterations to the model was 0.54, and in the map that includes NDVI has an overall accuracy of 0.55, which shows that including NDVI will yield a model with higher overall accuracy.
   Changing the order in which the covariates in this model are added will not impact the confusion matrix output, and the overall accuracy does not change either. The order of the covariates does not matter in this model.

Discussion:

Section I of this lab created a model that classified the land in the study area into the following classes: water, impervious (building or nonbuilding), and pervious (tree, irrigated, and non-irrigated). The covariates in the model that were used to create this classification are reflectance bands (B8, B2, B4, B3). The decision tree in the results section illustrates how these classes were determined from the bands. This model classifies using the covariates and looks at the samples for each classification, and it is important to recognize that the model depends on this 'training data' being accurate and there being enough data to correctly classify. Based on prior knowledge of the study area, it is clear that there are classification errors in this model which are highlighted above in the results section.

Section II expands on this consideration of accuracy by evaluating the accuracy of the model using the confusion matrix. Using the calculation for total accuracy, the model accuracy is around 54%, which is not excellent in terms of illustrating these classes. This supports the findings from the first section of the lab, where errors in classification were identified on the map layer, because the confusion matrix quantifies the degree of error in that map. Based on both the errors that were identified visually in the map (section I), as well as the total accuracy of the model provided in this section (section II), it is clear that this model does not accurately classify impervious / pervious surfaces and a better model will be needed to illustrate these classes.

Based on error in this map, section III of this lab seeks to create a more accurate map of pervious / impervious surfaces by making modifications to the training points. A higher degree of accuracy (evaluated through both visual inspection and calculating total accuracy in the confusion matrix) can be obtained by adding, moving, or deleting the training points. It is important when modifying these points to ensure that they are done accurately (i.e., points represent the correct class), or else the results will not yield a more accurate model. By adding roughly 10-15 more training points per class, the model accuracy increased from 55% to 59%, which illustrates how models can be improved through better and more accurate data. In the extended analysis, the highest degree of accuracy is calculated.

Appropriate Uses:

Similar to unsupervised classification, supervised classification is an advancement in classifying different types of landcover in a region using machine learning instead of carrying out the process manually by digitizing. This tool is beneficial because digitizing can be tedious and time consuming. By creating models and in the case of supervised classification, training points, the process of classification is automated.

Supervised classification can be used for a wide variety of scenarios within remote sensing; it can classify landcover into NLCD classes, pervious / impervious, vegetation type, habitat suitability, soil type, and many other categorical classifications that are of interest to the study.

Limitations and Caveats:

A key disadvantage of supervised classification is that it **requires prior knowledge of the study area**, which represents additional complexity and nuance compared to unsupervised classification. To illustrate this, errors in the classification around Horsetooth reservoir were easily identified due to prior knowledge of the study area; had this study been conducted in a different region such as a national park in Africa, it would be difficult to identify these areas of misclassification without prior knowledge or some other means of validating classification from the model.

Extending on the previous point, the training process where researchers label the points also draws on this prior knowledge. Inaccurate or insufficient knowledge can reduce the accuracy of the model because **the model can only be as good as the data that researchers provide it with**. This was illustrated in the model where accuracy was 55%, this low level of overall accuracy shows that the training data given to the model was not highly accurate, leading to error in the model visualized in the map layer, as well as calculated from the confusion matrix. The benefit of supervised classification over unsupervised is that there is validated data to compare the model output against (Daburra, 2020).

Last, supervised classification models **require lots of data and with more data; the models become more computationally expensive.** This will be illustrated in the extended analysis, where lots of training data attempts to create a more accurate model (Dhiraj, 2020).

Next steps:

The models in this lab illustrate the power of supervised classification in determining pervious and impervious landcover types. A good next step from this analysis would be to validate the results by using models including unsupervised classification and digitizing, then comparing the results from each of these approaches to see how well they match up.

Sources, Scripts, Citations (Lab six, parts 1-3 + extended analysis, in order)
https://code.earthengine.google.com/d88391f0af27b083fee5c8d391eda61d
https://code.earthengine.google.com/6a10a92add9ff5ce060dbbb92d03ae67
https://code.earthengine.google.com/5348134090f19e1562e46ffd43591ead
https://code.earthengine.google.com/e376eeb4a0e538028675af1197ed5d2d

Agarwal, Dhiraj. "Supervised vs Unsupervised Learning: Algorithms, Example, Difference." *Blog For Data-Driven Business*, 24 May 2020, https://www.intellspot.com/unsupervised-vs-supervised-learning/.

Dabbura, I. (2020, August 10). *K-means clustering: Algorithm, applications, evaluation methods, and drawbacks*. Medium. Retrieved April 15, 2022, from https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

Script Paths:

https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3ALabsix%2Flab6_p1

https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3ALabsix%2Flab6_p2

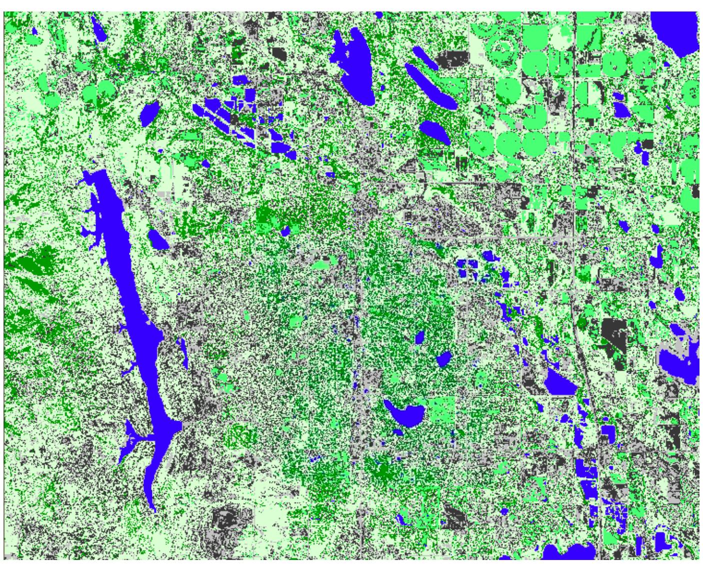
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3ALabsix%2Flab6_p3

https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3ALabsix%2Flab6_extendedanalysis

Extended Analysis:
This section seeks to test the limits of the model in terms of how accurate (within reason) this model can become by revising the training points (adding, moving, deleting). Furthermore, it illustrates how adding more data can make the model more computationally taxing which can serve as a limitation within the framework of supervised classification. The results are discussed below:

For the training data, there are roughly 50 data points per class for a total of over 300 points (I found that this process of creating training points became a bit tedious, similar to digitizing in lab three).

Image:



Permalink: https://code.earthengine.google.com/5578eca48c4a9d81d7e92db5a5ad9957

Adding these additional training points to the model did increase the time it took to generate this map and calculate its accuracy; however, it was only about 1 minute longer, so not enough to say that the additional data imposed a computational burden.

The overall accuracy of this map was only around 61%, not much of an improvement from previous models with accuracies of 55% and 59%. Although the accuracy of this map was slightly higher, it is still far from the desired level of around 70-80%.

In order to reach a higher level of accuracy, adding additional covariates (i.e., bands) to this model may be a viable approach.

Resources: (Other useful resources for labs 1-6, permalinks and script paths listed above)

Lab one:

https://gis.stackexchange.com/questions/395679/reclassify-values-of-the-land-use-land-cover-class-of-copernicus-global-land-cov

Example scripts (users/DavidTheobald8/NR323)

Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, Linlin, Tsendbazar, N.E., Ramonio, F., Arino, O., 2021. ESA WorldCover 10 m 2020 v100. https://doi.org/10.5281/zendono.5571936

Users/DavidTheobald8 ‘demo scripts’ + demo videos

Lab two:

https://custom-scripts.sentinel-hub.com

https://www.sciencedirect.com/topics/earth-and-planetary-sciences/spatial-resolution

Salafsky N, Salzer D, Stattersfield AJ, Hilton-Taylor C, Neu- garten R, Butchart SHM, Collen B, Cox N, Master LL, O’Connor S, Wilkie D. 2008. A standard lexicon for biodiversity conservation: unified classifications of threats and actions. Conserv Biol 22(4):897–911.

Users/DavidTheobald8 ‘demo scripts’ + demo videos

Lab three:

https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/overview-of-image-classification.htm#:~:text=Unsupervised%20classification%20is%20where%20you,class%20categories%20within%20your%20schema)

Users/DavidTheobald8 ‘demo scripts’ + demo videos

Virene, J.W. 2022. Google Earth Engine script

Lab four:

https://www.statology.org/pearson-coefficient-of-skewness-excel/

https://www.nature.com/articles/s41467-022-28270-3.pdf

https://ecologicalprocesses.springeropen.com/track/pdf/10.1186/s13717-021-00330-4.pdf

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Users/DavidTheobald8 ‘demo scripts’ + demo videos

Lab five:

Dabbura, I. (2020, August 10). *K-means clustering: Algorithm, applications, evaluation methods, and drawbacks*. Medium. Retrieved April 15, 2022, from https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

Matthew Stewart, P. D. R. (2020, July 29). *The limitations of machine learning*. Medium. Retrieved April 15, 2022, from https://towardsdatascience.com/the-limitations-of-machine-learning-a00e0c3040c6

Lab six:

Agarwal, Dhiraj. “Supervised vs Unsupervised Learning: Algorithms, Example, Difference.” *Blog For Data-Driven Business*, 24 May 2020, https://www.intellspot.com/unsupervised-vs-supervised-learning/.

Lab seven:

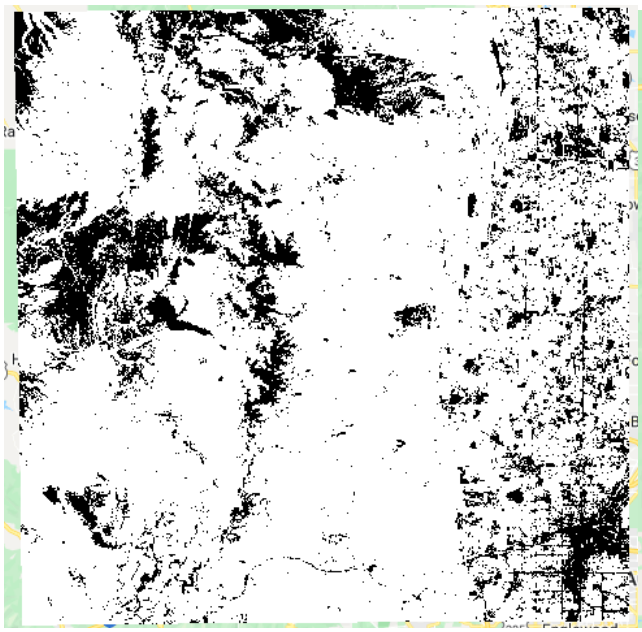
https://stackoverflow.com/questions/57060903/reclassify-ndvi-raster-in-intervals-on-google-earht-engine

https://gis.stackexchange.com/questions/392578/calculating-ndvi-for-small-study-area-using-google-earth-engine

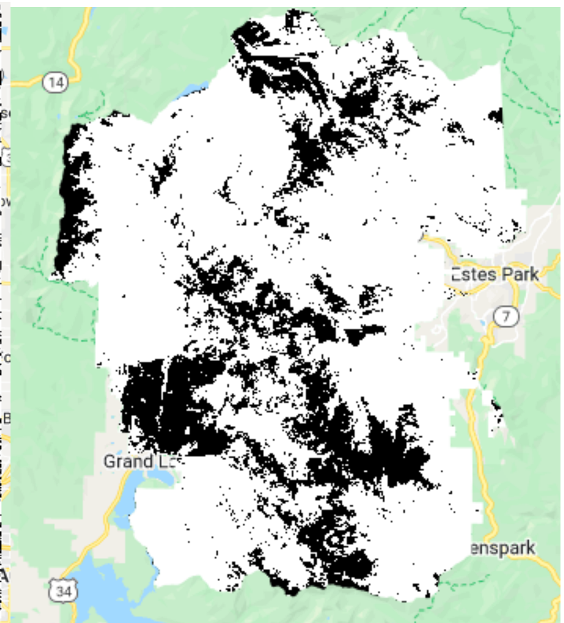
Comprehensive Lab Report (A new script with an overview of the bulk of the material covered throughout the semester).

Question: How many square kilometers of forested area are there in Rocky Mountain National Park?
Forest area = 205,000 acres = 831.4 square kilometers

Overall area for the NDVI



ROMO Boundary



The following proportions for each landcover type are estimated:

Forest landcover (white)	Nonforest Landcover (black)
0.77	0.23

Based on these, and given a total acreage for the ROMO boundary of 266,817 acres, the square kilometers for each land cover type are estimated:

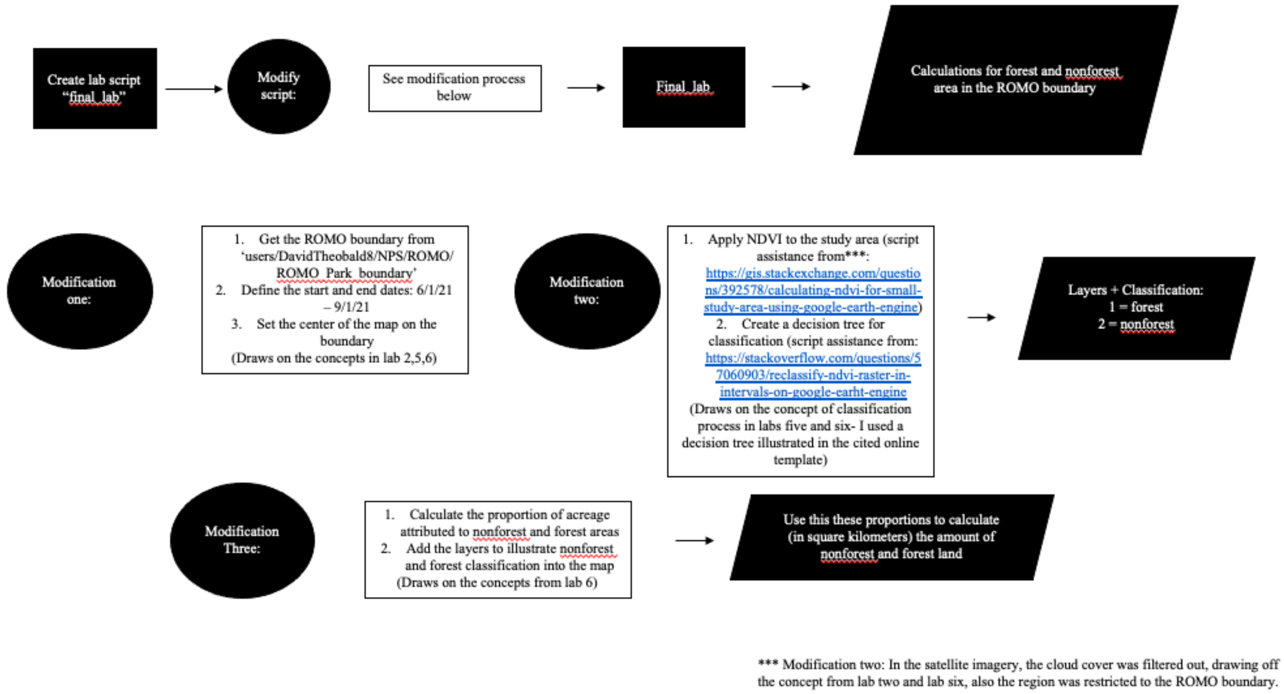
Forest landcover (white)	Nonforest Landcover (black)
831.42 square kilometers	248.34 square kilometers

Sources, Scripts, and Citations:

https://stackoverflow.com/questions/57060903/reclassify-ndvi-raster-in-intervals-on-google-earht-engine

https://gis.stackexchange.com/questions/392578/calculating-ndvi-for-small-study-area-using-google-earth-engine
https://code.earthengine.google.com/f64b358d2945010ea4ecc3eca8d6eeea (permalink)
https://code.earthengine.google.com/?scriptPath=users%2Fjoshvirene%2Flabs%3AFinal_lab (script path)

Overview Diagram:



Other tables, synthesis, scripts, that are useful:

Pros and Cons of Google Earth Engine:

Pros: This software is incredibly versatile, and it can accomplish many different analyses across the world for many different purposes including assessing human modification, landcover classification, and machine learning through supervised and unsupervised classification to achieve classification for other purposes (i.e., pervious vs impervious surface). Another advantage of the software is that it is widely used and as a result, there are many resources for users to consult when they encounter issues or need assistance with certain objective

Cons: The API on Google Earth Engine is limited, and conducting analyses requires skill in JavaScript in order to accomplish the above applications (though, there are many resources that can aid users in achieving their objectives. The only other major weakness of the software is its inability to accomplish studies within hydrological analyses.

Comparison to other GIS software programs:

Comparing Google Earth Engine to ArcGIS, there are many similarities. Both programs can accomplish an extensive range of objectives and be used across many studies. One advantage that ArcGIS has over Earth Engine is that it can be used for cartography and the maps can be directly placed into research papers because it integrates all the elements required for making maps that convey information to map readers in a clear manner. Another advantage of ArcGIS is that it can conduct hydrological analyses, and ESRI maintains many different analytical tools all of which, are on a very user-friendly interface that can be done with or without programming.
Despite these advantages, Google Earth engine is excellent for analyzing raster data, and likely better than ArcGIS for this type of data. Furthermore, JavaScript is a very user-friendly form of programming. Furthermore, as mentioned, there are many resources by google, stack exchange, and other resources to aid users.

Another application I've used is RStudio, this program is used extensively in economics and data analysis, though it can also be used to analyze geospatial data and validate the results from Google Earth Engine including histograms, confusion matrices, and many other things.

**
[redacted], thanks for a great semester and have a nice summer!