

# Introduction to Raster Data

## ⌚ Overview

**Teaching:** 15 min

**Exercises:** 10 min

### Questions

- What format should I use to represent my data?
- What are the main data types used for representing geospatial data?
- What are the main attributes of raster data?

### Objectives

- Describe the difference between raster and vector data.
- Describe the strengths and weaknesses of storing data in raster format.
- Distinguish between continuous and categorical raster data and identify types of datasets that would be stored in each format.

This episode introduces the two primary types of geospatial data: rasters and vectors. After briefly introducing these data types, this episode focuses on raster data, describing some major features and types of raster data.

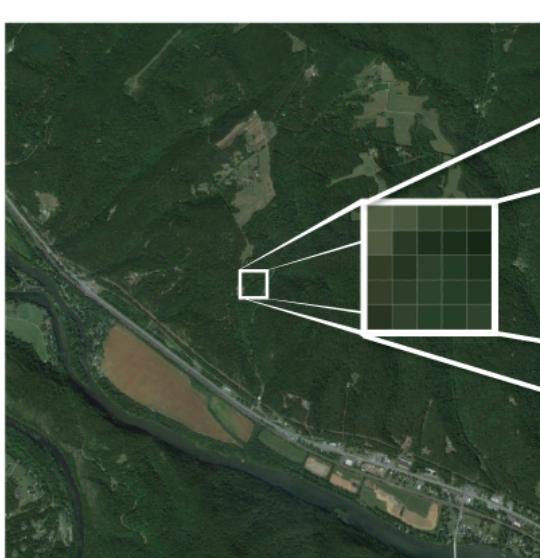
## Data Structures: Raster and Vector

The two primary types of geospatial data are raster and vector data. Raster data is stored as a grid of values which are rendered on a map as pixels. Each pixel value represents an area on the Earth's surface. Vector data structures represent specific features on the Earth's surface, and assign attributes to those features. Vector data structures will be discussed in more detail in the next episode ([/02-intro-vector-data/](#)).

The R for Raster and Vector Data (<https://datacarpentry.org/r-raster-vector-geospatial/>) lesson will focus on how to work with both raster and vector data sets, therefore it is essential that we understand the basic structures of these types of data and the types of data that they can be used to represent.

## About Raster Data

Raster data is any pixelated (or gridded) data where each pixel is associated with a specific geographical location. The value of a pixel can be continuous (e.g. elevation) or categorical (e.g. land use). If this sounds familiar, it is because this data structure is very common: it's how we represent any digital image. A geospatial raster is only different from a digital photo in that it is accompanied by spatial information that connects the data to a particular location. This includes the raster's extent and cell size, the number of rows and columns, and its coordinate reference system (or CRS).



0,0 5,0

1	3	9	7	7
2	8	7	7	8
6	7	3	5	7
7	6	5	5	6
8	6	5	6	4

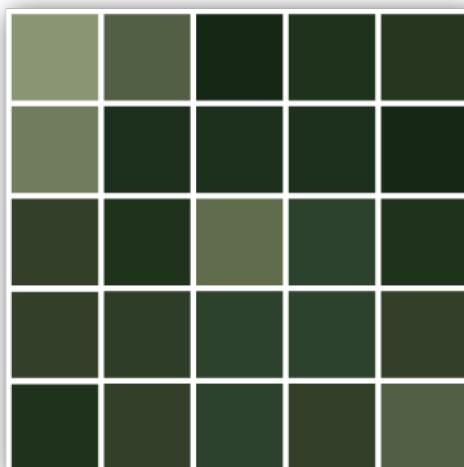
0,5 5,5

Legend



0,0 5,0

Plot Raster



0,5 5,5

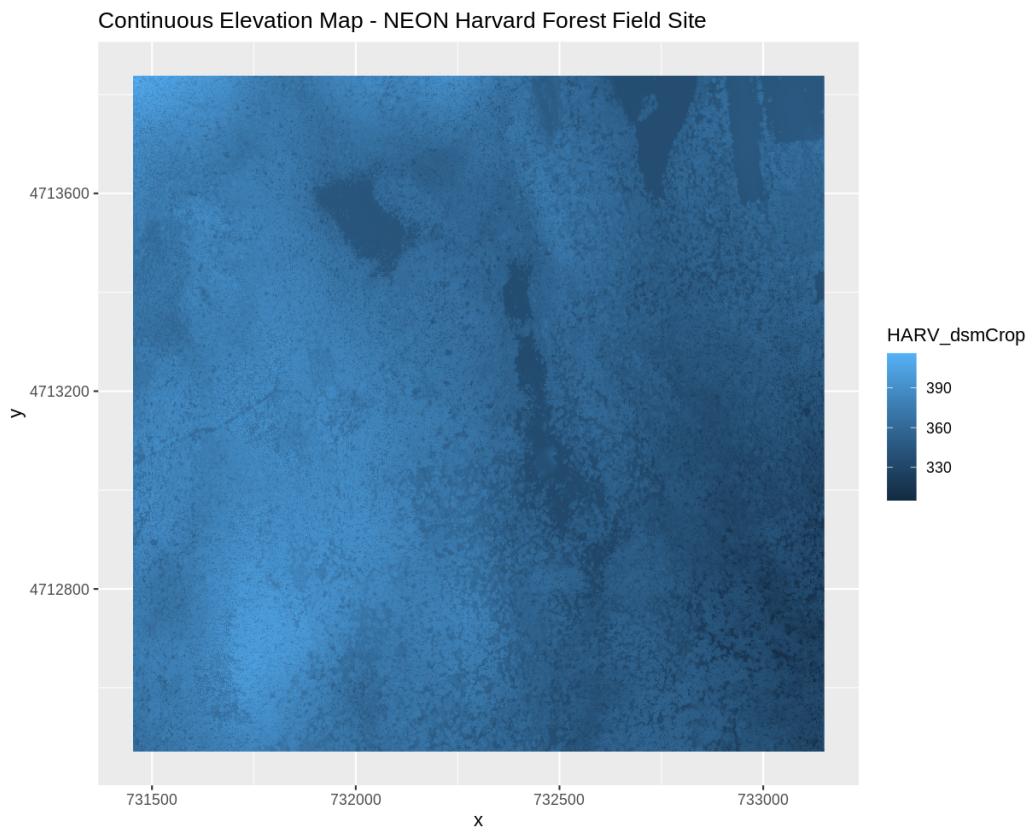
neon®

Source: National Ecological Observatory Network (NEON)

Some examples of continuous rasters include:

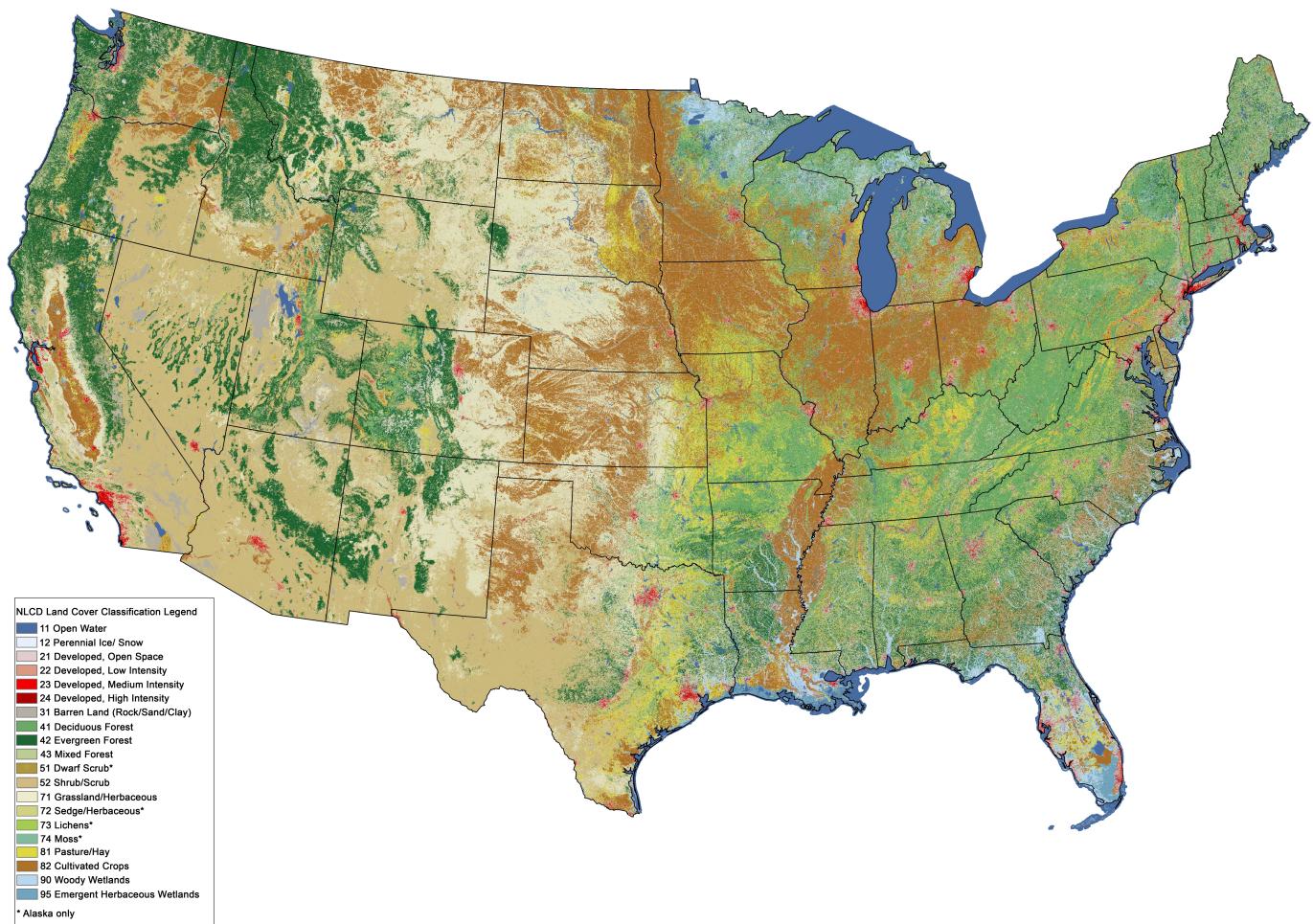
1. Precipitation maps.
2. Maps of tree height derived from LiDAR data.
3. Elevation values for a region.

A map of elevation for Harvard Forest derived from the NEON AOP LiDAR sensor (<http://www.neonscience.org/data-collection/airborne-remote-sensing>) is below. Elevation is represented as continuous numeric variable in this map. The legend shows the continuous range of values in the data from around 300 to 420 meters.

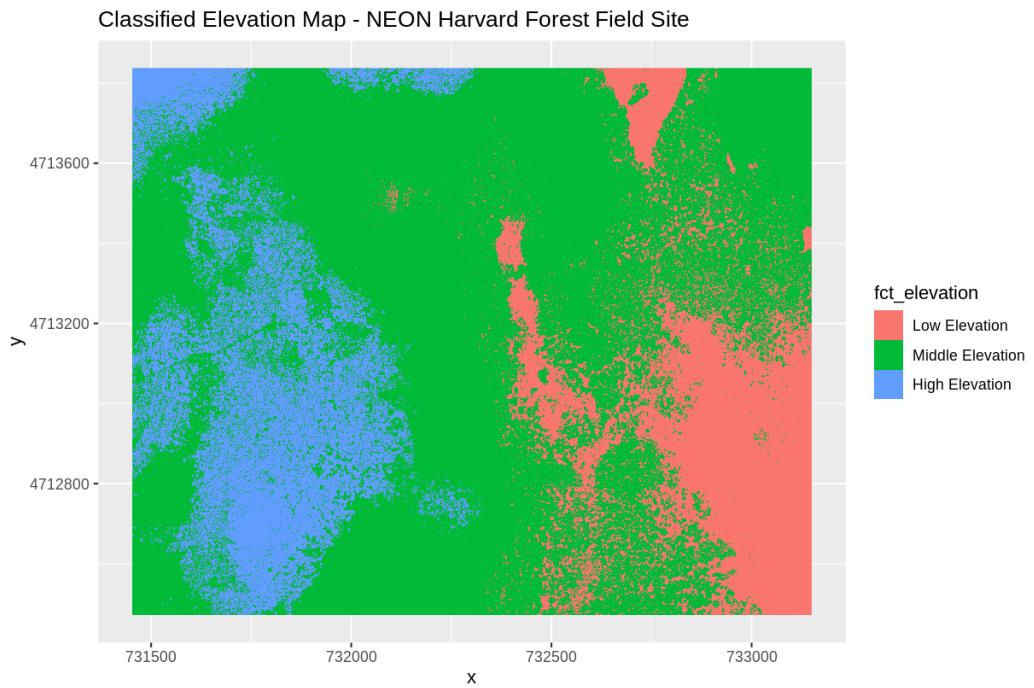


Some rasters contain categorical data where each pixel represents a discrete class such as a landcover type (e.g., “forest” or “grassland”) rather than a continuous value such as elevation or temperature. Some examples of classified maps include:

1. Landcover / land-use maps.
2. Tree height maps classified as short, medium, and tall trees.
3. Elevation maps classified as low, medium, and high elevation.



The map above shows the contiguous United States with landcover as categorical data. Each color is a different landcover category. (Source: Homer, C.G., et al., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing, v. 81, no. 5, p. 345-354)



The map above shows elevation data for the NEON Harvard Forest field site. We will be working with data from this site later in the workshop. In this map, the elevation data (a continuous variable) has been divided up into categories to yield a categorical raster.

## Advantages and Disadvantages

With your neighbor, brainstorm potential advantages and disadvantages of storing data in raster format. Add your ideas to the Etherpad. The Instructor will discuss and add any points that weren't brought up in the small group discussions.

## Solution

Raster data has some important advantages:

- representation of continuous surfaces
- potentially very high levels of detail
- data is ‘unweighted’ across its extent - the geometry doesn’t implicitly highlight features
- cell-by-cell calculations can be very fast and efficient

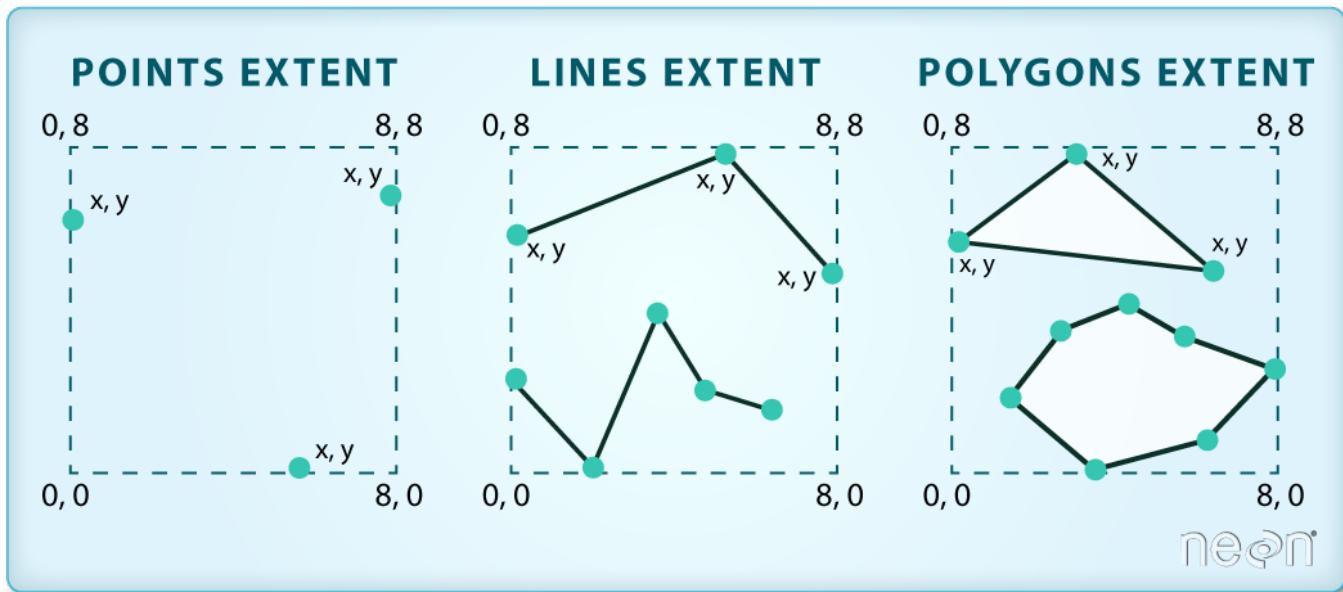
The downsides of raster data are:

- very large file sizes as cell size gets smaller
- currently popular formats don’t embed metadata well (more on this later!)
- can be difficult to represent complex information

# Important Attributes of Raster Data

## Extent

The spatial extent is the geographic area that the raster data covers. The spatial extent of an R spatial object represents the geographic edge or location that is the furthest north, south, east and west. In other words, extent represents the overall geographic coverage of the spatial object.



(Image Source: National Ecological Observatory Network (NEON))

### ✍ Extent Challenge

In the image above, the dashed boxes around each set of objects seems to imply that the three objects have the same extent. Is this accurate? If not, which object(s) have a different extent?

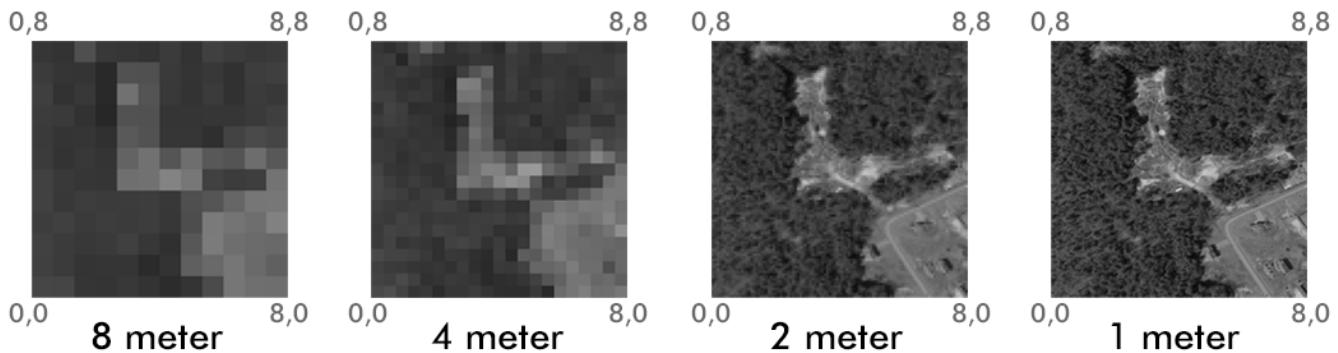
### 👁 Solution

The lines and polygon objects have the same extent. The extent for the points object is smaller in the vertical direction than the other two because there are no points on the line at  $y = 8$ .

## Resolution

A resolution of a raster represents the area on the ground that each pixel of the raster covers. The image below illustrates the effect of changes in resolution.

## Raster over the same extent, at 4 different resolutions



(Source: National Ecological Observatory Network (NEON))

## Raster Data Format for this Workshop

Raster data can come in many different formats. For this workshop, we will use the GeoTIFF format which has the extension `.tif`. A `.tif` file stores metadata or attributes about the file as embedded `tif` tags. For instance, your camera might store a tag that describes the make and model of the camera or the date the photo was taken when it saves a `.tif`. A GeoTIFF is a standard `.tif` image format with additional spatial (georeferencing) information embedded in the file as tags. These tags should include the following raster metadata:

1. Extent
2. Resolution
3. Coordinate Reference System (CRS) - we will introduce this concept in a later episode (/03-crs/)
4. Values that represent missing data ( `NoDataValue` ) - we will introduce this concept in a later lesson (<https://datacarpentry.org/r-raster-vector-geospatial/01-raster-structure/index.html>).

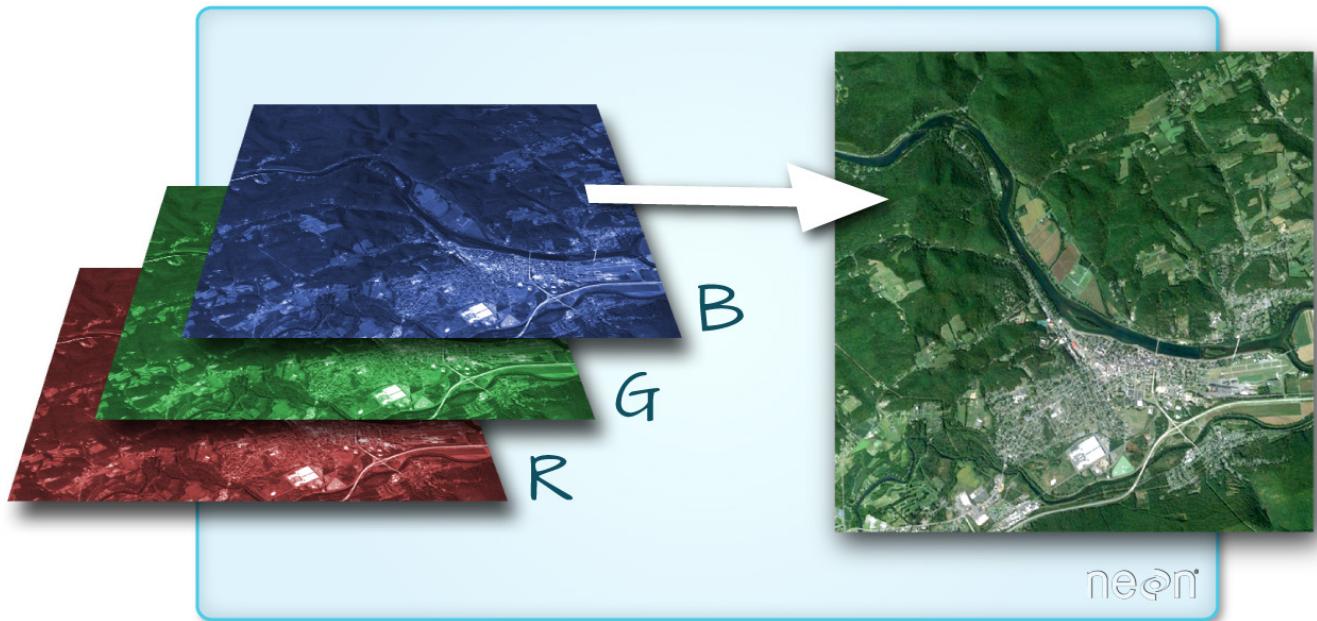
We will discuss these attributes in more detail in a later lesson (<https://datacarpentry.org/r-raster-vector-geospatial/01-raster-structure/index.html>). In that lesson, we will also learn how to use R to extract raster attributes from a GeoTIFF file.

### ↗ More Resources on the `.tif` format

- GeoTIFF on Wikipedia (<https://en.wikipedia.org/wiki/GeoTIFF>)
- OSGEO TIFF documentation (<https://trac.osgeo.org/geotiff/>)

## Multi-band Raster Data

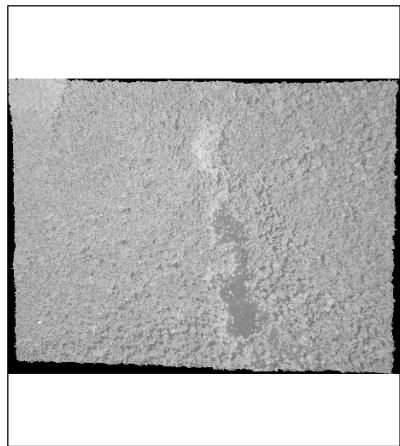
A raster can contain one or more bands. One type of multi-band raster dataset that is familiar to many of us is a color image. A basic color image consists of three bands: red, green, and blue. Each band represents light reflected from the red, green or blue portions of the electromagnetic spectrum. The pixel brightness for each band, when composited creates the colors that we see in an image.



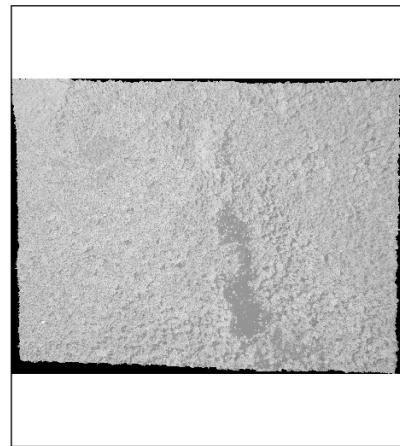
(Source: National Ecological Observatory Network (NEON).)

We can plot each band of a multi-band image individually.

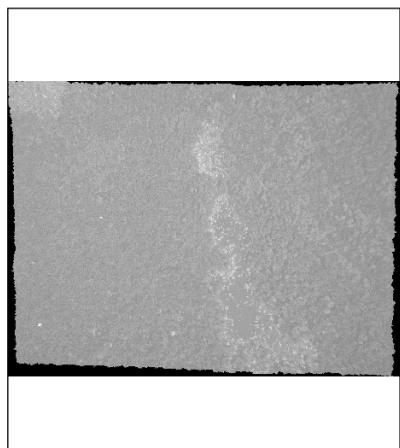
**Red.Band**



**Green.Band**



**Blue.Band**



Or we can composite all three bands together to make a color image.

### 3 Band Color Composite Image NEON Harvard Forest Field Site



In a multi-band dataset, the rasters will always have the same extent, resolution, and CRS.

## ➤ Other Types of Multi-band Raster Data

Multi-band raster data might also contain:

1. **Time series:** the same variable, over the same area, over time. We will be working with time series data in the Raster Time Series Data in R (<https://datacarpentry.org/r-raster-vector-geospatial/12-time-series-raster/index.html>) episode.
2. **Multi or hyperspectral imagery:** image rasters that have 4 or more (multi-spectral) or more than 10-15 (hyperspectral) bands. We won't be working with this type of data in this workshop, but you can check out the NEON Data Skills Imaging Spectroscopy HDF5 in R (<https://www.neonscience.org/hsi-hdf5-r>) tutorial if you're interested in working with hyperspectral data cubes.

## ❶ Key Points

- Raster data is pixelated data where each pixel is associated with a specific location.
- Raster data always has an extent and a resolution.
- The extent is the geographical area covered by a raster.
- The resolution is the area covered by each pixel of a raster.