Clouds, shadows & cloud masks in R

Learning Objectives

After completing this tutorial, you will be able to:

- Describe the impacts that thick cloud cover can have on analysis of remote sensing data.
- Use a cloud mask to remove portions of an spectral dataset (image) that is covered by clouds / shadows.
- Define cloud mask / describe how a cloud mask can be useful when working with remote sensing data.

What you need

You will need a computer with internet access to complete this lesson and the data that we already downloaded for week 6 of the course.

{% include/data_subsets/course_earth_analytics/_data-week6-7.md %}

About Landsat scenes

Landsat satellites orbit the earth continuously collecting images of the Earth's surface. These images, are divided into smaller regions - known as scenes.

> Landsat images are usually divided into scenes for easy downloading. Each Landsat scene is about 115 miles long and 115 miles wide (or 100 nautical miles long and 100 nautical miles wide, or 185 kilometers long and 185 kilometers wide). -*wikipedia*

In the previous lessons, we learned how to import a set of geotiffs that made up the bands of a landsat raster. Each geotiff file was a part of a Landsat scene, that had been downloaded for this class by your instructor. The scene was further cropped to reduce the file size for the class.

We ran into some challenges when we began to work with the data. The biggest problem was a large cloud and associated shadow that covered our study area of interest - the Cold springs fire burn scar.

Dealing with clouds & shadows in remote sensing data

Clouds and atmospheric conditions can present a significant challenge when working with spectral remote sensing data. Extreme cloud cover and shadows can render the data in those areas, un-usable given reflectance values are either washed out (too bright - as the clouds scatter all light back to the sensor) or are too dark (shadows which represent blocked or absorbed light).

In this lesson we will learn how to deal with clouds in our remote sensing data. There is no perfect solution of course. We will just discuss some options to dealing with the uncertainty surrounding clouds and shadows in spectral data.

Let's begin by loading our spatial libraries.

```r
# import spatial packages
library(raster)
library(rgdal)
library(rgeos)
# turn off factors
options(stringsAsFactors = F)
```

Next, we will load the landsat bands that we loaded previously in our homework.

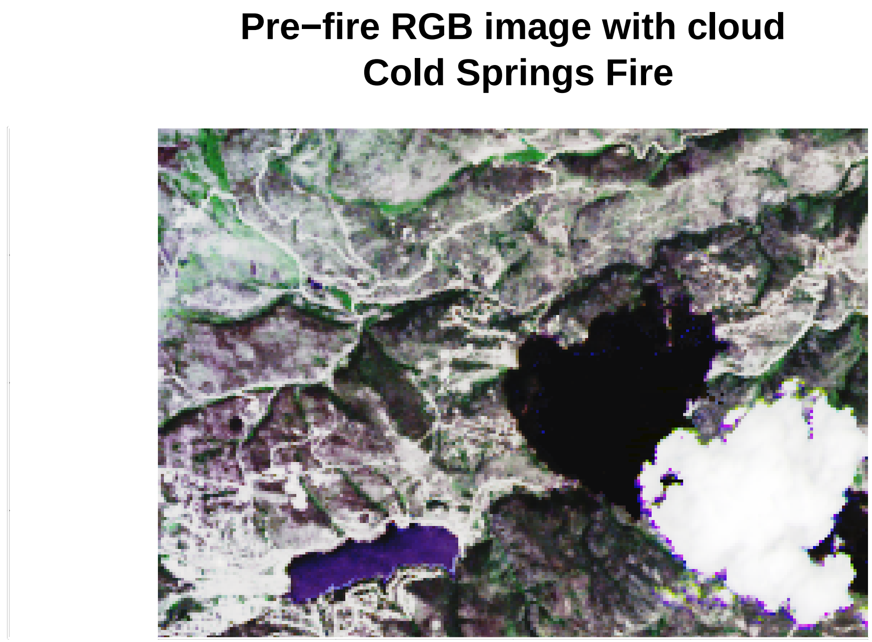

Figure 1: RGB image of our landsat data.

```
# create a list of all landsat files that have the extension .tif and contain the word band.
all_landsat_bands <- list.files("data/week6/Landsat/LC80340322016189-SC20170128091153/crop",
           pattern=glob2rx("*band*.tif$"),
           full.names = T) # use the dollar sign at the end to get all files that END WITH
# create spatial raster stack from the list of file names
all_landsat_bands_st <- stack(all_landsat_bands)
```

When we plotted the pre-fire image, we noticed a large cloud in our scene. Notice as i'm plotting below, i'm adding a few *par*ameters to force `R` to add a title to my plot.

Data Tip: Check out the additional "How to" lessons for this week to learn more about creating nicer plots in `R`. {: .notice }

```
# turn the axis color to white and turn off ticks
par(col.axis="white", col.lab="white", tck=0)
# plot the data - be sure to turn AXES to T (we just color them white)
plotRGB(all_landsat_bands_st,
        r=4, g=3, b=2,
        stretch="hist",
        main="Pre-fire RGB image with cloud\n Cold Springs Fire",
        axes=T)
# turn the box to white so there is no border on our plot
box(col="white")
```

Raster masks

Often (but not always) remote sensing data come with mask layers. These layers identify pixels that are likely representative of a cloud or shadow that have been generated by whomever processed the data. When

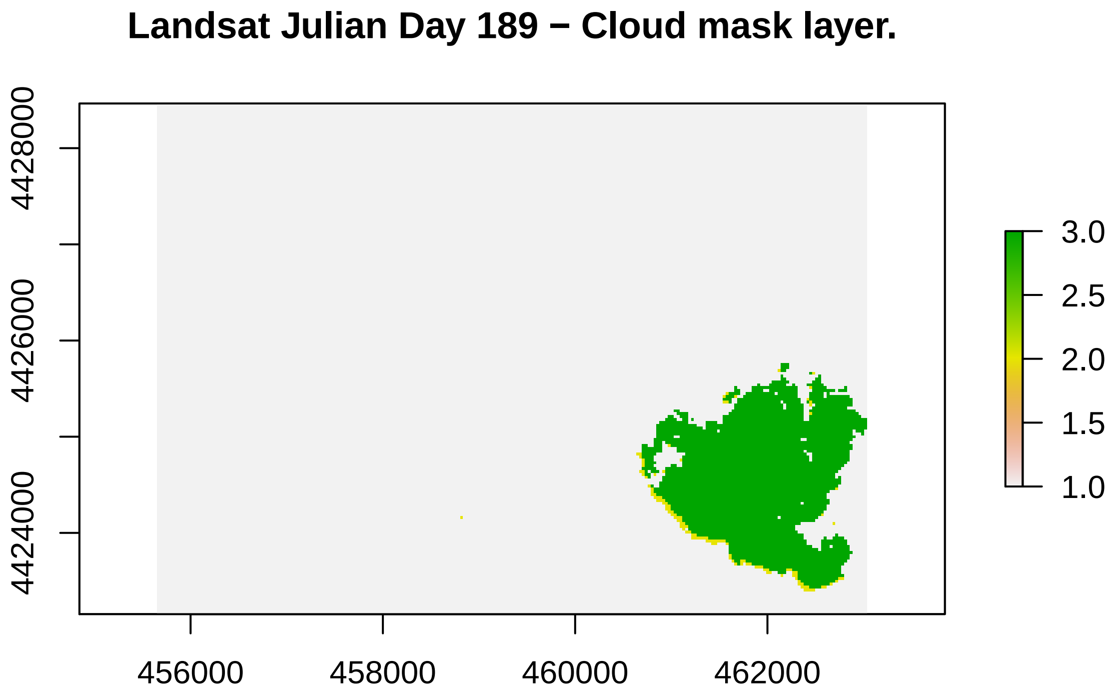

Figure 2: cloud mask - no shadows.

we download Landsat 8 data from Earth Explorer, the data came with 2 processed cloud mask raster layers.

1. `LC80340322016189-SC20170128091153/crop/LC80340322016189LGN00_cfmask_crop.tif`
2. `LC80340322016189-SC20170128091153/crop/LC80340322016189LGN00_cfmask_conf_crop.tif`

Let's have a look at these layers next.

```
# open cloud mask layer
cloud_mask_189_conf <- raster("data/week6/Landsat/LC80340322016189-SC20170128091153/crop/LC8034032201618
plot(cloud_mask_189_conf,
  main="Landsat Julian Day 189 - Cloud mask layer.")
```

Next, we can plot the second mask layer. Do you notice any difference between the two?

```
# apply shadow mask
cloud_mask_189 <- raster("data/week6/Landsat/LC80340322016189-SC20170128091153/crop/LC80340322016189LGN
plot(cloud_mask_189,
  main="Landsat Julian Day 189 - Cloud mask layer with shadows.")
```

What do the metadata tell us?

We just explored two layers that potentially have information about cloud cover. However what do the values stored in those rasters mean? We can refer to the metadata provided when we downloaded the Landsat data to learn more about how each layer in our landsat dataset are both stored and calculated.

Let's open the metadata file: `data/week6/landsat/LC80340322016189-SC20170128091153/LC80340322016189LGN00.xml` What does it tell us?

```
<a href="~/Documents/Github/earthlab.github.io/images/course-materials/earth-analytics/week-7/cloud-mas
<img src="~/Documents/Github/earthlab.github.io/images/course-materials/earth-analytics/week-7/cloud-ma
</a>
```

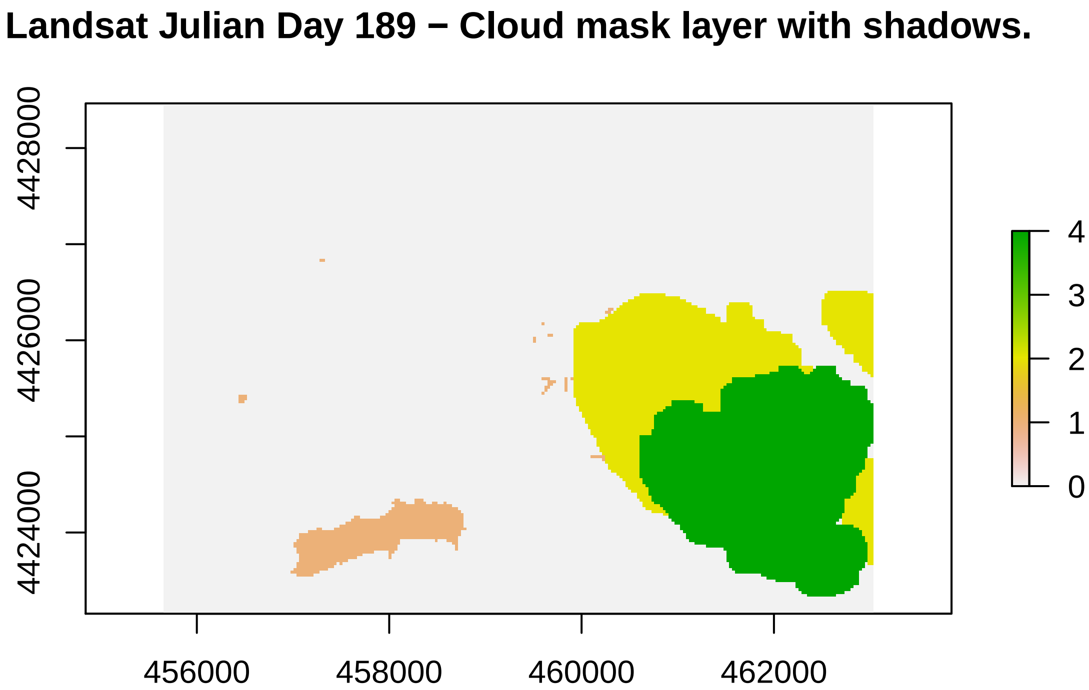

Figure 3: cloud mask with shadows

```
<figcaption>A snippet of metadata for Landsat 8 downloaded from USGS Earth Explorer website. It's impor
</figcaption>
```

When we download remote sensing data, often (but not always), we will find layers that tell us more about the error and uncertainty in the data. Often whomever created the data will do some of the work for us to detect where clouds and shadows are - given they are common challenges that we need to work around when using remote sensing data.

In this case, if we study the metadata we can see that our `cfmask.tif` file contains several classes. 1, 2, and 4 represent clouds and shadows. These might be values of pixels that we want to mask from our analysis. More on this later.

Cloud masks in R

We can use the cloud mask layer to identify pixels that are likely to be clouds or shadows. We can then set those pixel values to `NA` so they are not included in our quantitative analysis in R.

When we say "mask", we are talking about a layer that "turns off" or sets to NA, the values of pixels in a raster that we don't want to include in an analysis. It's very similar to setting data points that equal -9999 to NA in a time series data set. We are just doing it with spatial raster data instead.

```
<a href="~/Documents/Github/earthlab.github.io/images/course-materials/earth-analytics/week-7/raster_mas
<img src="~/Documents/Github/earthlab.github.io/images/course-materials/earth-analytics/week-7/raster_ma
</a>
<figcaption>When we use a raster mask, we are defining what pixels we want to exclude from a quantitati
</figcaption>
```

To create the mask this we do the following:

1. We make sure we use a raster layer that is the SAME EXTENT and the same pixel resolution as our landsat scene. In this case we have a mask layer that is already the same spatial resolution and extent

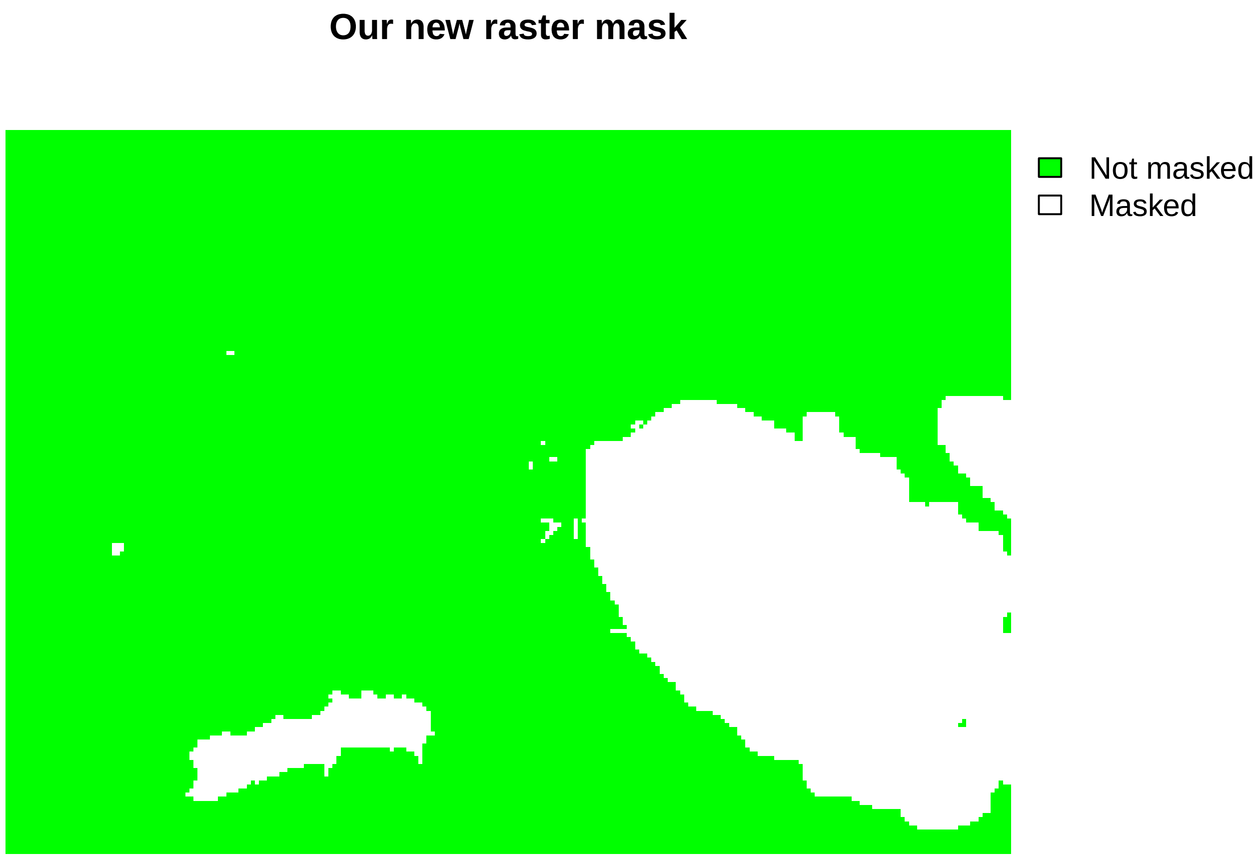

Figure 4: raster mask. green values are not masked.

as our landsat scene.
2. We then set all of the values in that layer that are clouds and / or shadows to `NA`
3. Finally we use the `mask()` function to set all pixel locations that were flagged as clouds or shadows in our mask to `NA` in our `raster` or in this case `rasterstack`.

In this case, we want to set all values greater than 0 in the raster mask to `NA`.

```
par(xpd=F, mar=c(0,0,1,5))
# create cloud & cloud shadow mask
cloud_mask_189[cloud_mask_189 > 0] <- NA
plot(cloud_mask_189,
     main="Our new raster mask",
     col=c("green"),
     legend=F,
     axes=F,
     box=F)
# add legend to map
par(xpd=T) # force legend to plot outside of the plot extent
legend(x = cloud_mask_189@extent@xmax, cloud_mask_189@extent@ymax,
       c("Not masked", "Masked"),
       fill=c("green", "white"),
       bty="n")
```

Notice in the image above, all pixels that are green represent pixels that are OK or not masked. This means they weren't flagged as potential clouds or shadows. All pixels that are WHITE are masked - these are areas of clouds and shadows.

Apply a mask

We can apply a mask to all of the bands in our raster stack which is convenient! Let's use the `mask()` function to mask our data.

```
# mask the stack
all_landsat_bands_mask <- mask(all_landsat_bands_st, mask = cloud_mask_189)
# plot RGB image
# first turn all axes to the color white and turn off ticks
par(col.axis="white", col.lab="white", tck=0)
# then plot the data
plotRGB(all_landsat_bands_mask,
        r=4, g=3, b=2,
        main="RGB image - are all of the clouds gone from our image?",
        axes=T)
box(col="white")
```

Notice above that I didn't have to use the stretch function to force the data to plot in R. This is because the extremely bright pixels which represented clouds, are now removed from our data.

```
# plot RGB image
# first turn all axes to the color white and turn off ticks
par(col.axis="white", col.lab="white", tck=0)
# then plot the data
plotRGB(all_landsat_bands_mask,
        r=4, g=3, b=2,
        stretch="lin",
        main="RGB image - are all of the clouds gone from our image? \n linear stretch",
        axes=T)
box(col="white")
```

Next, we can calculate a vegetation indices.

Optional challenge

- Overlay the fire boundary on top of the landsat pre-fire image.
- If you were asked to QUANTIFY the pre vs post fire burn area extent, what are some problems that you can anticipate running into with the cloud cover - even with using the mask?

A cloud's covering our study area - what's next?

Now that we have discovered a problem with our data that will impact quantitative analysis of the data, what do we do?

Well, there are several options, most of which we won't discuss in this class. However, one option is that we could go find a better image. We happen to know that the conditions before the fire were rather stable in 2016. So what if we could find an image from say - June that doesn't have clouds?

In the next lesson, we will talk about using the EarthExplorer website to download remote sensing data.

Figure 5: apply raster mask to stack and plot.

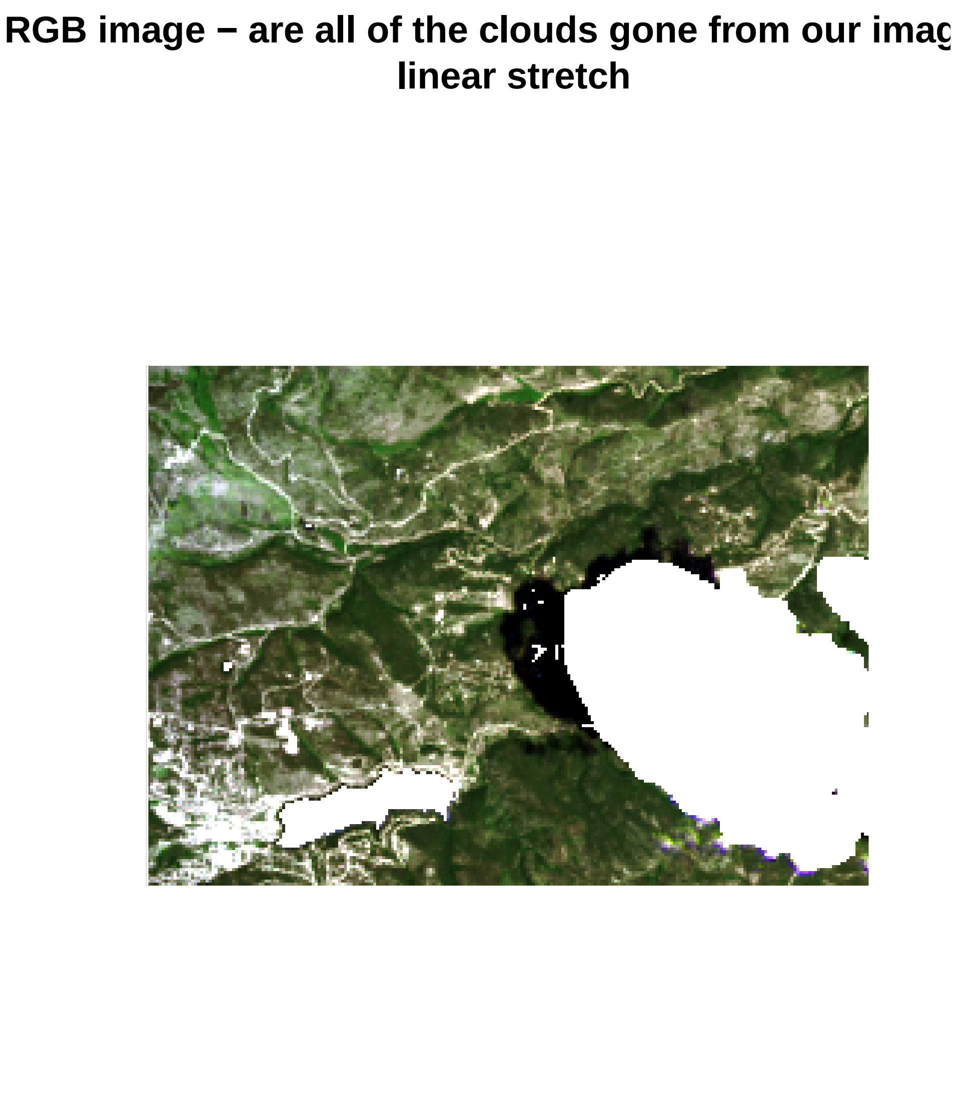

Figure 6: apply raster mask to stack and plot.

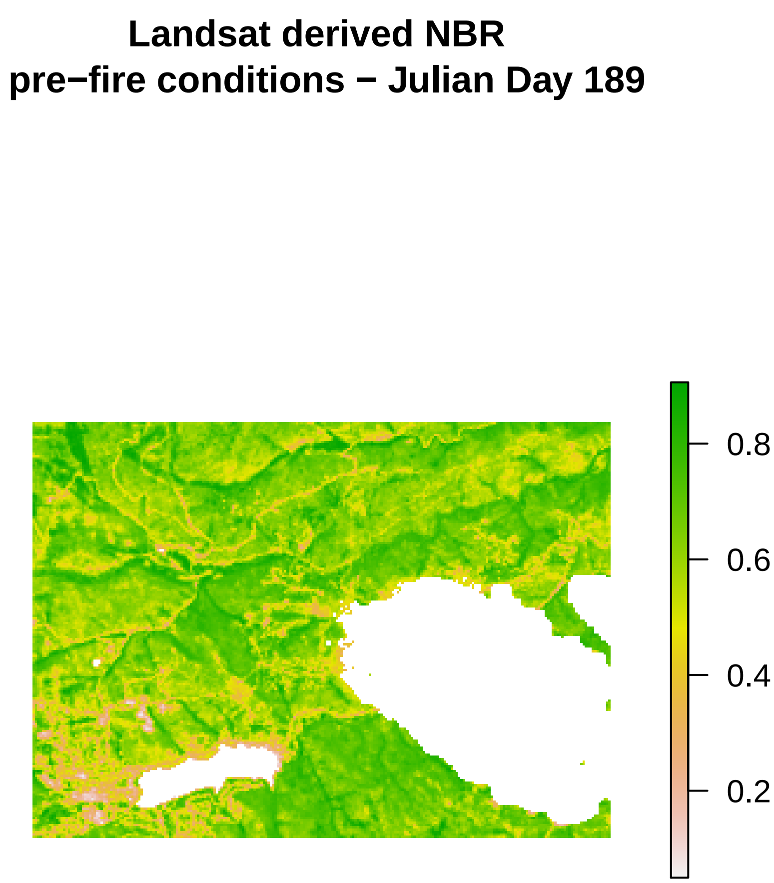

Figure 7: landsat NBR plot