Fundamentals of Raster Analysis in QGIS 3.4

Ryan Lennon

May 2019

Abstract

This is an introductory raster tutorial which provides instructions on how to perform basic raster functions in QGIS 3.4 (Madeira), which is the most recent stable long-term release to date of QGIS. This tutorial is written as an extension of the introductory workshop and manual "Introduction to GIS Using Open Source Software (with QGIS)".

Contents

1 Introduction 2

2 Loading in the Data 2

3 Retrieving Basic Information about your Raster Datasets 5
   3.1 Managing Raster Projections 6
   3.2 Filling NoData Values 7
   3.3 Produce Raster Unique Values Report 8

4 Using the Raster Calculator 10
   4.1 Converting Temperature from Celsius to Fahrenheit 10
   4.2 Subsetting Raster Pixels 12

5 Clipping Rasters 13
   5.1 Clip Operations 13

6 Extracting Raster Statistics 15
   6.1 View Raster Layer Histogram 17
   6.2 Sampling Raster Values 17

7 Converting Raster Layers into Vector Layers 18
   7.1 Performing an Attribute Query 20
   7.2 Perform an Intersect Between Raster Cells and Vector Polygons 21
   7.3 Raster to Point 23
   7.4 Calculate Distance to Water from Each Raster Cell 24

8 Overlaying Raster Datasets with NYC Orthoimagery 24

9 Raster Styles 27
   9.1 Loading a Print Composition 28
1 Introduction

This document should give users an introductory overview of how to carry out basic analytical operations on raster datasets on QGIS. This is an updated version of the raster tutorial written for QGIS 3.4 that explores different raster functions. This tutorial should expand the user's ability to extract information from general raster datasets, which could include a wide variety of information.

Anyone is welcome to use this document under a Creative Commons Attribution Noncommercial No Derivative Works 4.0 License (CC BY-NC-ND 4.0): http://creativecommons.org/licenses/by-nc-nd/4.0/ for personal or classroom use.

The question of what users go to raster data for is best answered by looking at its advantages over vector data for spatial analysis. One of the main advantages of raster data over vector data is that you can represent large areas with continually changing values in a more intuitive way.

In this tutorial we will analyze temperature as it relates to discrete features in the landscape. Temperature is one example of a measurement that is best represented in raster format, because it is a continuous feature in the landscape. Other features which might influence or be influenced by temperature, such as parking lots, buildings, or trees are better represented as discrete objects which is why we typically store them as vector shapefiles.

You will see that QGIS includes a number of tools to extract information from raster datasets by using the discrete geometries of vector objects as a spatial reference. These tools can help us to answer research questions, to learn about our local environment or just to create useful map products.

For the purposes of this tutorial we will be using temperature datasets derived from the Landsat 8 OLI tile which includes New York City. The Landsat satellite collects infrared thermal radiation which is used to create a raster image representing land surface temperature. This temperature data can be used to study variations of land surface temperature throughout New York City, and anywhere in the world.

2 Loading in the Data

The first step is to download the working data from http://www.baruch.cuny.edu/geoportal/practicum/raster/. Unzip the file and you will see that it contains several raster temperature datasets and several shapefiles (.shp file extension) for features that we will be referencing the raster data to.

First, we will load in the surface temperature raster which is derived from the Landsat 8 satellite.

1. Launch QGIS Desktop 3.4 (Madeira).

2. In the upper-left hand corner, click New Project.

3. Click Open Data Source Manager in top menu bar and click on the Add Raster tab.
4. In the Source section, next to the Raster dataset(s) box, click the ellipsis \( \ldots \). Navigate to the working folder C:\Downloads\Raster Tutorial Data in the browse dialog and select the NYCLandsatST20170831.tif image.

5. Click Add.

QGIS should render the image onto the map canvas. If you would like to center the map canvas on the image, you can right-click on the image in the layers list in the bottom-left and click Zoom to Layer.

Load in the New York Borough Boundaries shapefile. This is a vector shapefile that represents the boundaries of each of New York’s five boroughs and is clipped to the shoreline.

1. Click Open Data Source Manager in top menu bar and click on the Add Vector tab.

2. In the Source section, next to the Vector dataset(s) box, click the ellipsis \( \ldots \). In the browse dialog and select the nybb.shp shapefile.

3. Click Add.

4. Once it is added, click on the nybb.shp in the layers list in the lower-left hand side of the map canvas and drag it to the top of the drawing order.
In the layers list on the left, right-click on the `nybb.shp` layer and click Properties. Click on the Symbology tab. Set the opacity slider to 50 percent and click OK. This should make it so that you can see the temperature layer along with the NYC borough boundaries, so you have some spatial reference to what you are looking at.
Notice that now you can see the other layers underneath the New York borough boundaries outline. We will talk more about setting raster symbology later on in the tutorial, but for now you can play with these to arrange your map so that you can see what you are doing.

3 Retrieving Basic Information about your Raster Datasets

Before we can use many tools in QGIS we have to enable the processing toolbox. In order to do this we have to go through the plugins menu. Click on the plugins dropdown in the toolbar and click Manage and Install Plugins.

In the search bar type Processing, and the plugin for Processing should come up in the results. Click the box next to Processing and click Install plugin. You should now see the Processing button in the toolbar. Click this button to open the Processing Toolbox.
3.1 Managing Raster Projections

It is important that when we add raster files to a project we know which projection they are in so that we can accurately compare them to other datasets spatially. When you add your first spatial dataset to a QGIS project, the map canvas assumes the projection of that dataset. The map canvas will remain in this projection unless you change it. This is important to know for when we overlay datasets that have different projections.

1. In the toolbar at the top of the screen, click on the Raster dropdown.

2. Select Miscellaneous, then select the Raster Information tool.

3. The Raster Information tool dialogue should appear.

4. Click on the Input layer dropdown and select the NYCLandsatST20170831.tif raster that was just added to the project.

5. Because we want to save the output statistics to a file, click the ellipsis next to the Layer [save to temporary file] box and select Save to File. Then select the working directory C:\Downloads\Raster Tutorial Data, and let's name the new output file LandsatST0831.html to stay consistent with the naming convention for the original file.

6. Click Run.

Once the process is complete, a link to the output should appear in the lower left-hand corner of your screen. When you click on the link it should open in your computer's default handler for HTML files, which may be a web browser like Firefox. The results should look like this:

Notice that the projection and datum for the raster is listed as USER:100000 - * Generated CRS . . .

We can also access this information by right-clicking on the raster dataset in the layers list, clicking on Properties, and clicking on the Information tab .

R. Lennon, Baruch CUNY, 2019
We can also see that the NoData value is -3.39999995214436425e+38. This NoData value is assigned to pixels for which there was some error during processing, usually due to cloud interference with the sensor. QGIS gives you the ability to set the NoData value and the projection for rasters that you produce using raster functions.

Click on the Zoom to Full Extent button to zoom out so that we can see the temperature dataset overlayed on top of the National Land Cover Dataset.

Run the Raster Information tool on the National Land Cover Dataset (NLCD) and you will see that the projection for this raster is EPSG:5070 - NAD83 / Conus Albers - Projected.

Because the temperature raster was added before the nlcd2011_ny.tif raster, the map canvas is currently in the projection of the temperature raster. If we wanted to change this, we could right-click on the nlcd2011_ny.tif dataset in the layers list, click CRS, and click Set Project CRS from layer.

### 3.2 Filling NoData Values

When working with satellite imagery we often have to deal with cells where data was removed due to cloud cover or other obstructions which prevented the sensor from being able to get an accurate measurement for that pixel. Because the number that we use to represent NoData pixels is arbitrary (in this case -3.4e+38), when we run certain operations on a raster dataset, the algorithms might not run correctly if there are NoData values present in the dataset, or we just might not get an output that is accurate.

To resolve this problem, we can fill the NoData values so that the algorithms will run smoothly. We can choose to fill the NoData values a number of ways.

1. Click on the Raster dropdown menu on the toolbar.
2. Click on the Analysis menu.
3. Click on the Fill NoData Values tool.
4. Select the NYCLandsatST20170831.tif layer as the Input layer.

5. Next to Filled, click the ellipsis and click Save to File. Name the output raster TempNYC20170831.tif.

6. Click Run. The output should be a raster where NoData values are replaced with an approximation of what we think the values for those cells should be based on the values of the cells surrounding them. In the layers list the output raster will be named Filled. Right-click on this layer in the layers list, click Rename layer and rename it to TempNYC20170831.tif. Uncheck the box next to the existing (unfilled) NYCLandsatST20170831.tif layer in the layers list so that it is removed from the map canvas (but still accessible in the project).

3.3 Produce Raster Unique Values Report

Some raster datasets, like the temperature dataset that we’ve been working with, contain continuous values, which could fall anywhere within the range for that raster dataset. Other raster datasets, like land cover classifications, contain categorical values, or discrete integers that are used as a code for some feature type being represented by each cell in the raster dataset. Because categorical rasters only have a limited number of possibilities for what value each cell could be, they often take up less disk space for the amount of area that they cover.

For data analysis purposes, it is important that we know how many values we are working with. In spatial datasets, it is particularly important that we know what areas we have data for, and what the reliability of our data is. We can find this information by generating a Raster Unique Values Report. First, we will add another raster to the project: the 2011 National Land Cover Dataset for New York State.

1. Click on the Data Source Manager.

2. Click on the Add Raster tab.

3. Click the ellipsis, navigate to the Tutorial Data folder and select the nlcd2011_ny.tif layer as the input raster.
4. Click Add.

Now that the layer is added, we can generate a Raster Unique Values Report.

1. Click on the Raster Analysis dropdown in the Processing Toolbox.
2. Click on the Raster Unique Values Report tool.
3. Next to Input Raster, click the ellipsis and select the nlcd2011_ny.tif layer as the input raster.
4. Click Run, and then click on the link to the report once the process is finished.

You will see that the Raster Unique Values Report is output as an HTML file, which will open in your browser. This report is also useful because it tells us how many cells have each unique value in the dataset, which gives us the most precise calculation of how much space is taken up by each land cover type.

In the NLCD layer each number corresponds to a different land cover type which is specified in the metadata. You can see that the most common value (the value that takes up the most space in the NLCD layer) is 41, which accounts for 19,448,414,583 square meters (about 4.8 million acres). 41 is the classification number assigned for deciduous forest.

This metadata file is also available in your data folder as an image named NLCD_Colour_Classification_Update.jpeg. If you open it you will see that there are many designations for different land cover types. The second most common land cover type, 11, corresponds to open water. The values 21-23 all correspond to different intensities of developed (or urban) land cover.
4 Using the Raster Calculator

The compact data structure of raster data is one of the format’s strengths relative to vector data. This structure allows you to perform calculations over a large area without needing too much computer memory, which decreases the likelihood of the process to crash.

4.1 Converting Temperature from Celsius to Fahrenheit

For example, if we want to convert our temperature data from degrees Celsius into degrees Fahrenheit, we can use the raster calculator.

1. In the toolbar at the top of the screen, click on the Raster dropdown.
2. Select Miscellaneous, then select the Raster Calculator tool.
3. The Raster Calculator tool dialogue should appear.
4. Double-click on the TempNYC20170831.tif layer to add it to the expression window.
5. Enter the formula to calculate degrees Fahrenheit based on the existing raster, which is in degrees Celsius. The formula is tempFahrenheit = tempCelsius * 1.8 + 32.

   ![Raster Calculator Expression]
   
   "NYCLandset3120170831.tif" * 1.8 + 32

In the area where it says Result Layer, click the ellipsis and choose to save the file in the tutorial data folder, named Temp_f_20170831_NYC. Under file type choose GeoTIFF.
6. Click OK.

After the raster calculator runs you should get an output raster that shows temperature in degrees Fahrenheit. Click and drag the Temp_f_20170831_NYC raster to the top of the layers list. With the layer highlighted, click on the identify tool in the toolbar above the map canvas. Click on one of the darker areas that represent water. Notice that the identify results window gives us the value and the coordinates for the pixel that we clicked on.

Click on an area for land and notice that the temperature value is higher. In the Identify Results window, click the Mode dropdown menu and change the mode to Top down. Click on another point and you will see the values returned for each layer in the layers list. Change the View to Graph and you will be able to see the values for each layer in a graph next to each other.

Add the other two temperature rasters in the tutorial data folder, Temp_f_20170909_NYC and Temp_f_20171027_NYC, to the project. All three of these temperature rasters are in degrees Fahrenheit, and were taken from the same sensor at different dates. If we want to see how temperature changes throughout the year, this would be one way that we could do it. Use the identify tool to look at how temperature varies from August to September to October. Click on a pixel and switch the Identify results view to Graph.
You can see that in some areas the temperature change from August to September is not quite as
significant as the temperature change from September to October.

4.2 Subsetting Raster Pixels

Another application of the raster calculator is to subset raster pixels. If we wanted to create another
raster based on our land cover dataset that is just water, we could select all of the cells in the NLCD
2011 raster that correspond to open water, and then create a new raster containing only those cells
(discarding all other information in the dataset).

Remember from the metadata file for NLCD 2011 that the designation for open water is 11.
1. In the toolbar at the top of the screen, click on the Raster dropdown.

2. Click Raster Calculator.

3. The Raster Calculator tool dialogue should appear.

4. In the layers list in the top left, double-click the NLCD 2011 layer to add it to the expression
   box. Next to the layer name, type =11.

5. Click on the ellipsis next to output layer and name the output layer WaterNYC.tif.

6. Click OK.

The output raster should be a binary file having only two values: 1 for water and 0 for not water.
Right-click on the TempNYC20170831.tif layer and click Remove layer to remove the original layer from the project. Do the same for NYCLandsatST20170831.tif. Click the Save button , and save the project in the tutorial data folder as NYC Temperature.qgz.

5 Clipping Rasters

Clipping a raster dataset is a fundamental process for raster analysis. We can use the clip tool to subset our data - to create a new dataset that just covers a geographic area we are interested in - and then compare that area to the larger extent.

Clipping rasters is also useful to minimize processing requirements. Some processes that are complex and take up a lot of computer memory will run more efficiently if we break them up into smaller chunks. For this reason, we can break up our dataset into a few key regions of interest and run our algorithms on those areas first before looking at areas that aren't necessarily as important for whatever hypothesis we are testing.

5.1 Clip Operations

Clip Raster by Mask can be used to clip a raster to the outline, or mask, of the study area. To do this you will need an input polygon shapefile representing the outline of the area you want to clip the raster to. The process may not work if the shapefile is a multipolygon, so it is best to dissolve the polygon (using the Dissolve tool in the processing toolbox) into one feature if you want to use it as a mask.

Clip Raster by Extent can be used to clip the raster to a bounding box, or extent, of a study site. A shapefile can be used, or you can manually put in the coordinates for each corner of the extent you want to clip the raster to.
Right now the temperature raster Temp_f_20170831_NYC is already clipped to the extent (minimum bounding box) of New York City, but the land cover raster nlcd2011_ny.tif covers all of New York State, which is more data than we need. We will clip the nlcd2011_ny.tif to the area around New York City.

1. In the Menu bar, click on the Raster dropdown menu.

2. Click on Extraction, and then click on the Clip by Extent tool.

3. In the dialog that pops up, select the nlcd2011_ny.tif raster.

4. Next to the field for Clipping extent, click the ellipsis, click Use layer extent, and select the nybb.shp shapefile. Notice that the field populates with the extent (coordinates of all four corners) of the nybb.shp layer.

5. Scroll down to the field that says Clipped (extent) and click the ellipsis. Click Save to File... and save the output in the tutorial data folder as nlcd_NYC.tif.

6. Click Run.

The output should look like this:

Notice that the extent of the raster now matches the minimum bounding box of the NYC borough boundaries shapefile. In the layers list, the output is called Clipped (Extent). Right-click on this layer and click Rename layer. Rename the layer to NLCD NYC 2011.
6 Extracting Raster Statistics

Now that all of the rasters are clipped to the extent of the study area, we can calculate the mean temperature for all of New York City. To do this, we just need to calculate and export some statistics on the raster.

In the toolbox, click in the search bar at the top and type in "Raster Statistics". The first tool that comes up should be the Raster layer statistics tool. Another way to get to the tool is to navigate through the Processing toolbox:

1. In the Processing Toolbox, click on the arrow next to Raster Analysis to expand the toolbox.
3. In the dialog that pops up, click the dropdown arrow next to Input layer and select the temperature raster NYCLandsatST20170831.tif as the input layer.
4. The dialogue box should look like this:

![Raster Layer Statistics Dialogue Box]

5. Click Run.

The Raster Layer Statistics tool will output an HTML file with summary statistics for the raster layer. It calculates the mean for all raster cells in the layer, as well as a few other summary statistics including standard deviation. You can access this summary statistics report either by navigating to the Tutorial Data folder where we chose to save it, or you can click the link in the bottom left corner that pops up.

![Summary Statistics Report]

Notice that this mean (around 24 degrees Celsius) is lower than what we would expect the average temperature to be for New York in August. The reason for this is that a large proportion of the
The raster layer that we used as the input layer is water, which has a lower surface temperature than land in NYC.

Since we are interested in what the temperature is over land, we can clip the raster to the borough boundaries shapefile using it as a mask, rather than clipping it by extent. Since the borough boundaries shapefile is already clipped to the shoreline, our output will be a raster whose geometry is also clipped to the shoreline, thus representing land.

1. In the toolbar, click on the Raster dropdown menu.

2. Click on Extraction.

3. Click on Clip Raster by Mask Layer...

4. In the field that says Input layer, click the dropdown arrow and select the original temperature raster NYCLandsatST20170831.tif as the input.

5. In the field for Clipping Extent, click the ellipsis, and choose the nybb.shp shapefile as the mask.

6. Uncheck the box next to Match the extent of mask layer.

7. Next to the field Clipped (mask), click on the ellipsis and click Save to File... Name the output (clipped) raster NYCtemp_land_20170831.tif.

8. Click Run.

Notice that the output raster is clipped to the exact polygons for the New York City borough boundaries. In the layers list, it will show up as Clipped. Rename the clipped raster by right-clicking on it in the layers list and clicking Rename Layer. Name the layer NYC Temp Land.

Navigate to the Raster Layer Statistics tool again and calculate the layer statistics for the raster that is clipped to the mask outline of the New York borough boundaries (clipped to shoreline/land). Name the output file tempNYChist_clip.html.

Open the resulting raster layer statistics file and notice that the mean is about 25 degrees Celsius, which is higher than the mean for the raster that was clipped to the layer extent.
We can also do this by right-clicking on the raster layer and clicking Properties, then clicking on the Metadata tab of the Layer Properties menu that pops up.

### 6.1 View Raster Layer Histogram

One other way to visualize summary statistics of a raster file is to use the Raster Layer Histogram tool in the Processing Toolbox. This tool should tell us if the temperature values in our raster dataset are normally distributed.

1. In the Processing Toolbox, expand the Graphics section.
2. Double-click on the Raster Layer Histogram tool.
3. Next to Input layer, click the dropdown menu and select the NYC Temp Land layer.
4. In the field that says Histogram, click on the ellipsis and click Save to File...

![Save to File]

Name the output Temp_hist.HTML.

5. Click Run.

In the results window, click on the line of text that is the output file directory. This should open the HTML file containing the histogram. Notice that most of the values fall between 25-30 degrees Celsius and the shape of the distribution is similar to a bell curve. This means that the values are more or less normally distributed.

You can also view a raster layer's histogram by right-clicking on the layer, clicking on Properties, and clicking on the Histogram tab.

### 6.2 Sampling Raster Values

To extract the raster values for a set of point features, we can use the sample raster values tool. The sample raster values tool extracts the values of raster cells that intersect a point vector layer. The point vector layer could be any other feature that we have spatial reference for and want to compare to a raster dataset. Let’s say that we want to get the average value for temperature in parking lots in NYC. In this case, we will sample the temperature values at the point locations of parking lots in New York City. Add the ParkingLots.shp layer to the map canvas by clicking on the Data Source Manager, clicking the Vector tab, clicking on the ellipsis, and selecting the ParkingLots.shp file in the dialog. Click Add.

Next we will run the Sample Raster Values tool.

1. In the Processing Toolbox, expand the Raster Analysis section.
2. Click on the Sample Raster Values tool.
3. Next to Input Point Layer, click the dropdown arrow and select the ParkingLots.shp layer.
4. Select NYC Temp Land as the Raster Layer to sample.

5. Click the arrow next to Advanced parameters to expand it. In the field that says Output column prefix, type Temp.

6. Click Run.

The output should be a point vector layer named Sampled points. The layer should have the same geometry as the parking lots layer used as an input. Right-click on the layer and open the attribute table and you should see a column inserted for Temperature on the right. This is the value extracted from the raster pixel that intersects each parking lot in the ParkingLots.shp dataset.

If you want to see what the mean temperature was for parking lots in New York City, you can click on the Show Statistical Summary button at the top, select the Sampled Points layer in the dropdown menu under Statistics, and select the Temp column. You should see summary statistics for that field (including the mean) populate in the table below. The mean is 24.9 degrees Celsius, which is slightly higher than the mean for all of New York City but not quite as high as what we might expect.

Save the current project.

7 Converting Raster Layers into Vector Layers

In the Landsat metadata file you can see that Landsat images have a resolution of 30m by 30m. This means that each pixel in a Landsat image is 30m wide and 30m long. When we import a raster into QGIS that originates from the Landsat sensor, the geometry of each cell is equal to this resolution, so each cell is 30m wide and 30m tall.

If we were to convert this geometry into vector format, we could create a polygon layer composed of many adjacent polygons with the same geometry as the Landsat pixels, which is perfect squares measuring 30m x 30m. Then we could perform other operations on the data that we would normally perform on vector files, such as intersects and joins.

The data structures for raster datasets are designed to contain a large amount of information using relatively little disk space. Once we convert a raster dataset into a vector dataset, the resulting data will take up considerably more disk space since each vector feature in a dataset has its own unique geometry that is independent of the other features in the dataset whereas raster cells always have the same geometry relative to other raster cells in the same image.

To conserve disk space, before we convert our raster datasets into vector datasets, we will clip both of them to the outline of a much smaller study area, Floyd Bennett Field in South Brooklyn.

1. In the upper-left hand corner, click New Project.

2. Click Open Data Source Manager in top menu bar and click on the Add Raster tab.

3. In the Source section, next to the Raster dataset(s) box, click the ellipsis. Navigate to the working folder C:\Downloads\Raster Tutorial Data in the browse dialog and select the temperature raster that was clipped to the shoreline and converted into degrees Fahrenheit, NYCtemp_land_20170831.tif.
4. Click Add.  

Add the Floyd Bennett Field shapefile FloydBennettField.shp to the QGIS project and clip the NYC temperature raster to the mask of the shapefile using the Clip Raster by Mask tool. Name the output raster FBFtemp_20170831.tif.

Now that we have a much smaller raster dataset (less than 10,000 pixels). We are now ready to convert our raster data into a vector polygon dataset.

1. In the Processing Toolbox, click the dropdown arrow next to Vector Creation.

2. Double-click on the Raster Pixels to Polygons tool.

3. Select the clipped FBFtemp_20170831.tif raster as the input.

4. Under where it says Vector polygons, click on the ellipsis and click Save to File...

Name the output layer FBF_Temp_Polygons.shp.

5. In the box under Field name type Temperature.

6. Click Run.

In the layers list you should see a new layer called Vector polygons, which is the output from the Raster Pixels to Polygons tool. As you can see, the raster has now been converted to a polygon vector layer where each raster cell is an individual polygon with the same geometry as the original raster cell (30m by 30m in this case). Right-click on the layer click Open Attribute Table and you will see that the temperature attribute was preserved. Right-click on the Vector polygons layer in the layers list to the left and click Rename layer. Name it FBF Temp Polygons.

Right-click on the FBF Temp Polygons layer in the layers list and click Properties. Click on the Symbology tab and change it to Graduated. Next to Column, select Temperature. Click the Classify button in the the bottom left. Click OK.
The temperature vector polygons layer is now displayed using the appropriate range of values, where cells with a high temperature are red and cells with a low temperature are white.

Convert the NLCD raster into vector polygons the same way that we converted the temperature dataset into vector polygons. Name the output layer NLCDpolygons.shp.

Click Save project, and save the project as FloydBennettField.qgz.

7.1 Performing an Attribute Query

Now that we have a vector polygon representation of land cover types, we can do a query to select only the cells that represent a particular land cover type. Now that we know what the average temperature is for parking lots in New York City, let’s do an attribute query to see how much of Floyd Bennett Field is paved, or urban land cover type.

1. Right-click on the NLCDpolygons.shp layer in the layers list.

2. Click on Open Attribute Table button.

3. Click the Select by Expression button.

4. Click the Fields dropdown menu and double-click the DN field to add it to the expression box.

5. Next to 'DN', type "= 21", which is the number designation for Developed, Open Space.
6. Click Select Features and close the expression box (do not close the attribute table).

Notice that now, in the map window, all of the polygons representing developed open space are selected. At the top of the attribute table you should see the number of polygons that were selected.

Features Total: 6184, Filtered: 6184, Selected: 475

Now that we have selected all of the cells representing developed open space, we can export them as a shapefile representing that type of land cover. Close the attribute table. Right-click on the NLCDpolygons.shp layer in the layers list, click Save Selected Features As..., and save it as a shapefile Urban_LC_FBF.shp. Make sure that in the export window the box is checked that says "Save Only Selected Features".

Next, we can clip the temperature dataset FBFtemp_20170831.tif to areas representing urban land cover and calculate raster statistics, which will give us some indication of how temperature over developed space compares to temperature over other land cover types.

Zoom out and save. Hit the zoom to full extent button to zoom out to the full extent of your layers. Then hit the save button.

7.2 Perform an Intersect Between Raster Cells and Vector Polygons

Next, we can use the Intersection tool to create a subset of raster pixels that are close to or touching water. We can either use the shapefile for the borough boundaries that is clipped to shoreline as a vector representation of water, or we could take the subsetted raster dataset for water that we created in the Raster Calculator section (WaterNLCD.tif) and convert it into vector polygons to produce a vector representation of water.

1. In the toolbar, click on the Vector dropdown menu.
2. Click on Geoprocessing tools, and then click on Intersect.
3. Next to Input layer, select the FBF_Temp_Polygons.shp layer.
4. Next to Overlay layer, select the WaterNLCD.shp layer.
5. Name the output Temp_shoreline.shp.
6. Click OK.

We now have a vector layer representing all of the raster pixels that intersect the water dataset that we created from the NLCD raster. Looking at the average temperature values for these pixels, you can see that the mean temperature is lower than pixels representing urban land cover, but not quite as cold as pixels over water.
Right click on the NLCDpolygons.shp layer in the layers list to the left and click Properties. Click on the Symbology tab. Click the dropdown menu at the top that says Single Symbol and change it to Categorized, and click the Classify button at the bottom left. You will see all of the land cover classes appear in the symbology list. Expand the section for Layer Rendering. Use the opacity slider to set the transparency of the layer to 50 percent. Click OK.

Drag the NLCDpolygons.shp layer to the top of the layers list. You should be able to see the land cover classes overlayed on top of the temperature dataset. Displaying both of these variables at the same time is called bivariate mapping, and it helps us to visualize relationships between land cover and temperature. You can see that each combination of temperature and land cover has a very specific color attributed to it.
If we wanted to visualize variations of temperature over urban land cover, for example, we could change the color for urban to green in the NLCD, and then urban pixels would show up as different shades of blue depending on how they vary in temperature.

### 7.3 Raster to Point

Since we know that water is playing a role in temperature, we should do an analysis to see if areas further from water have a higher average temperature than areas closer to water. In this section we will convert our temperature layer into a point layer where each point is located at the centroid of each raster cell. Then we will calculate the distance of each point from water, adding that distance to the point layer as an attribute.

1. In the Processing Toolbox, click the arrow next to Vector Creation.
2. Double-click on the Raster Pixel to Point tool.
3. Select the temperature dataset clipped to Floyd Bennett Field, FBFtemp_20170831.tif.
4. Name the output layer temp20170831points.shp.
5. Under Field Name, type Temperature.
6. Click Run.

   Notice that now we have a vector point layer that has the same geometry as the centroids for each raster cell. If you use the Select Features tool, you can click on one of the points and view its attributes. Use the Identify tool to click on the raster cell underneath the point and notice that the values are the same as the points.
7.4 Calculate Distance to Water from Each Raster Cell

Once the point layer has been created, we can use the Distance to Nearest Hub tool to calculate the distance between each point in the point layer and the water layer.

1. In the toolbar, click the Raster dropdown menu.

2. Click Analysis.

3. Select the Distance to Nearest Hub tool.

4. Select the points temperature dataset temp20170831points.shp dataset as the Source points layer.

5. In the field that says Destination hubs layer, select the New York borough boundaries shapefile nybb.shp.

6. Click Run.

The output should show up in the layers list named Hub distance. It is a point layer with the same geometry as the input point dataset and it includes fields for the temperature at each point, the nearest polygon feature in the "Hub" layer, and the distance to the "Hub" layer.

Now right-click on the temp20170831points.shp dataset in the layers list and click Open attribute table. Notice in the attribute table there is a column populated with the distance from each point to water, HubDist. Double-click the HubDist column header to sort the dataset by distance from water. Notice that raster pixels that were further from water generally had a higher temperature, whereas pixels closer to water have a lower temperature value.

This could be used to substantiate the idea of building a pavilion further inland, on pavement so that park visitors have a place to sit in the shade.

8 Overlaying Raster Datasets with NYC Orthoimagery

For this tutorial we have gotten to look at how temperature varies in the landscape in New York City. Let’s say, for the purpose of this tutorial, that we are tasked with the construction of a pavilion to provide shade for park visitors at Floyd Bennett Field in South Brooklyn. As a GIS Analyst for the...
National Park Service, it is your job to find out what areas of the park get the hottest in the Summer, and what areas would benefit the most from shade.

One easy way to manually generate a spatial reference for something is to compare it to panchromatic aerial orthoimagery. For this section we will use aerial imagery obtained from NYC Open Data. First we will add the tile designation layer.

1. Click on the Add Vector button.

2. Select the NYC orthoimagery tile index layer 2016 Ortho Index.shp. This a layer showing the tile number for each tile in the NYC orthoimagery dataset. Because the full orthoimagery dataset is rather big, it is good to download it one or a few tiles at a time.

3. Click Add.

Navigate the map to the area of interest, Floyd Bennett Field in South Brooklyn. Right-click on the Floyd Bennett Field layer in the layers list and click Zoom to Layer. Notice that this is a rather large park. Only part of it is pavement/parking lot, and a large chunk of it is forest. What we will do now is we will try to select raster cells in this park that we think are parking lot, and compare the temperature in those cells to the temperature in the parts that are forested.

1. Click on the orthoimagery tile designation layer and drag it to the top of the layers list.

2. Right click on the orthoimagery tile designation layer and click on Properties.

3. Click on the Symbology tab and set the opacity to 50 percent.

Click OK.

Notice that now the tile designation layer is transparent so we can see which tiles overlap the park. Right-click on the orthoimagery tile index and go to Properties. Click on the Labels tab, and choose Single Labels. Next to Label with, click the dropdown and select the IMAGE column to label the tiles with the image number column.

Now we should be able to see the filenames for each orthoimagery tile that corresponds to Floyd Bennett Field.
Write down the tile numbers for the tiles that overlap the park and go back to the Data Source Manager. Click on Add Raster, look in the Orthos folder and you will notice that the filenames for the image tiles (each file with a .tif file extension) in that folder match the tiles that overlap the park. Select all of them (click on the top one, hold down the Shift key, then click the bottom one) and add them to the project.

Click on the vector polygons layer and drag it so that it is above the orthoimagery in the layers list. We sometimes call the layers list the drawing order, because it is the order in which layers are visible when they are drawn on the map canvas. Right now we want to be able to select pixels in the
temperature dataset that overlap pavement in Floyd Bennett Field so that we can compare temperature over pavement to the temperature in the rest of the park. Right-click on the Vector polygons layer, click Properties, and go to the Symbology tab. Using the opacity slider, set the transparency of the layer to 50 percent.

With the Vector Polygons temperature layer highlighted in the layers list, click the Select Features button in the toolbar and select the raster cells that appear to be in the parking lot by clicking on them. To select all of the cells that appear to be parking lot, hold down the Ctrl key while you are clicking (be careful, because the first time you click without holding down the Ctrl key you will lose your selection!) Once you have all of the cells selected, right click on the layer and click Export Selected Features. Name the layer Temp_urban_FBF.shp. Keep the box checked that says Add saved file to map.

You now have a temperature layer just representing temperature over pavement in Floyd Bennett Field. Run the raster layer statistics tool on this layer and you will notice that the mean for this layer is higher than the layer for the rest of Floyd Bennett Field.

*Zoom out and save.* Hit the zoom to full extent button to zoom out to the full extent of your layers. Then hit the save button.

### 9 Raster Styles

Applying styles to rasters helps us to visualize spatial patterns in the data without needing to extract any statistics. For this project we will look at how temperature varies in Floyd Bennett Field from August to October.

Try right clicking on the vector layer for Floyd Bennett Field in the layer manager, and click Zoom to Layer. Now, right click on the temperature raster and click Stretch Using Current Extent.

QGIS uses the raster values within the current map view to stretch the raster style so that we can visualize the difference between cell values. Since we zoomed in to Floyd Bennett Field, we are getting the best representation of how temperature varies spatially within Floyd Bennett Field.

In QGIS, we can create a number of rasters in a project, save that project, and then apply the same style to each raster so that they can be visualized side-by-side in an intuitive way.
Note that changing the symbology does not change the values present in the dataset. Stretching the symbology to match the current view simply changes the ranges of values associated with each color you see in the map view.

1. In the Processing Toolbox, click on the Raster Tools toolbox.
2. Click on the Set Style for Raster Layer tool.
3. Set the clipped Floyd Bennett Field temperature raster as the input raster `FBF_temp_20170831.tif`.
4. Choose the `temp.qml` style sheet as the style file.
5. Click Run.

Setting the raster style to something consistent helps so that we don’t create maps that are misleading. Just to make sure that all three rasters are displayed using the same symbology, right-click on the `FBF_temp_20170831.tif` in the layers list and click Copy style. Now right-click on the other two temperature rasters clipped to Floyd Bennett Field and click Paste style. Each one of your rasters should now be stretched to the same minimum and maximum values.

### 9.1 Loading a Print Composition

Now that each raster is displayed using the same symbology, we can load a print composition displaying these rasters to show a time-series. For this section a customized print composition was created for this tutorial, but you can create and save your own print compositions whenever you want. QGIS gives you the ability to change and save print compositions for later use.

1. Click on the Layout Manager.
2. Double-click on the Raster Time Series print composition to open it.
3. The print composition window opens with the images rendered according to the chosen layout.
4. Use the Add Label button to add a label underneath each temperature raster showing the date for the image (i.e. August 31, September 9, etc.).
5. Use the pan tool to center each image on Floyd Bennett Field.
6. When you are finished, click the Export as Image button to export your finished map.

Note that if you want to export the view within the current map canvas at any time in QGIS, you can click on Project in the top-left, click Import/Export, and click Export Map as Image.

You can see from the exported map layout that in the temperature raster from earlier in the year there was less variability in heat throughout the city and that the average temperature was much lower.

Right-click anywhere on the toolbar and enable the Raster Toolbar.

The Raster Toolbar enables us to easily change the raster style.
Zoom in on an area and click the Local Histogram Stretch button. Notice that the values are now stretched based on their distribution, rather than evenly based on the range within the current map extent. Stretching the values based on the histogram is a more intuitive way of displaying a raster image. This is similar to what would happen if you opened the layer properties for the raster, clicked on the symbology tab, and chose to classify the values based on natural breaks. QGIS gives you the ability to display rasters in a number of different ways.

Let's say that we want to see which areas in the city drop below freezing in October. We can change the classification to show areas below 0 degrees Celsius in red, and all other areas in green by modifying the classification scheme.

Zoom out and save. Hit the zoom to full extent button to zoom out to the full extent of your layers. Then hit the save button.

10 Conclusion

As you can see, QGIS gives us a wide variety of tools to extract information from spatial datasets in raster format. Analyzing raster datasets takes careful planning and conceptualizing what we want to learn from the data and what structure the data needs to be in for us to answer the questions we need to answer.

10.1 More Information on Raster Data Sources

The temperature data that we used for this tutorial was derived from the Landsat 8 satellite, which you can read more about here: https://landsat.usgs.gov/landsat-surface-temperature

The National Land Cover Dataset (NLCD) is derived from 30-m Landsat images and is available for the entire continental US at this link: https://www.mrlc.gov/data. Soon, the 2016 NLCD will be released: https://www.mrlc.gov/national-land-cover-database-nlcd-2016

There are a number of other sources for raster imagery available, including:

- NYC has aerial orthoimagery uploaded twice a year: https://data.cityofnewyork.us/City-Government/2016-Orthoimagery-Manhattan/xuek-2su9
- USGS Earth Explorer: https://earthexplorer.usgs.gov/
- Google Earth Engine - requires some knowledge of JavaScript: https://earthengine.google.com/
- The NYC Borough Boundaries shapefile dataset was obtained from https://www1.nyc.gov/site/planning/data-maps/open-data/districts-download-metadata.page
- Boundaries of National Parks can be downloaded as a shapefile from: https://public-nps.opendata.arcgis.com/datasets/national-park-service-park-unit-boundaries
- Parking Lots were derived from a Parking Lots polygon dataset on NYC Open Data: https://data.cityofnewyork.us/City-Government/Parking-Lots/e2f7-cs7i