Data Analysis Instructions

Part I: Field-Methods Data Analysis

Last week, we collected data on the spatial dispersion of southern live oaks using two different methods (density and T-square). Remember, that the main difference between these two methods is that the plot-based technique is scale-*dependent,* and the T-square method is scale-*independent*. In theory, regardless of what spatial scale we consider, the southern live oak distribution should remain the same using the T-square method.

During Part I of this lab, we will use the data obtained from the plot-based and T-square techniques to examine several questions: (1) what is the spatial dispersion of southern live oaks given these differing collection methods? (2) How do these dispersion patterns vary with spatial scale? (3) What abiotic and biotic factors might be influencing the spatial dispersion of southern live oaks?

Exercise 1: Plot-based sampling

Mapping the location of individuals within your study area can be a simple way to infer spatial dispersion. Accurate maps of large areas may be difficult to obtain due to logistical constraints and observer error, but map generation may be the only method available to determine spatial dispersion in some cases. Here, we will use the plot-based data to create a map of the southern live oak at the largest spatial scale we observed (i.e., whole-plot, 50x50 m). We will also use the known relationship of tree size (DBH) to age to examine whether these patterns of dispersion relate to the age of these live oaks.

1. Open the class data that has been uploaded to Canvas and begin on the **"Plot-Based"** tab.
2. The locations of individuals tree (x and y coordinates) have been entered into excel for you. Highlight the x and y coordinate columns and insert an x-y scatter plot. This will create a map of the 50x50 m plot and the location of each tree in that plot.
3. Resize each data point to equal the DBH of each tree. To do this, double-click an individual data point. Make sure only a single point is selected and not the entire data series.
4. Right click on the data point and select "Format Data Point." (**<u>NOTE:</u>** If you see "Format Data Series" it means every data point is selected and you need to click on the single point again).
5. Select the paint bucket → "Marker" → "Marker Options" → "Built-in" and resize the point to equal the DBH (Figure 1). Repeat these steps until every tree has a DBH entered.
6. To finish your map, be sure to add in appropriate axes labels and a caption. Include this in your post-lab report.

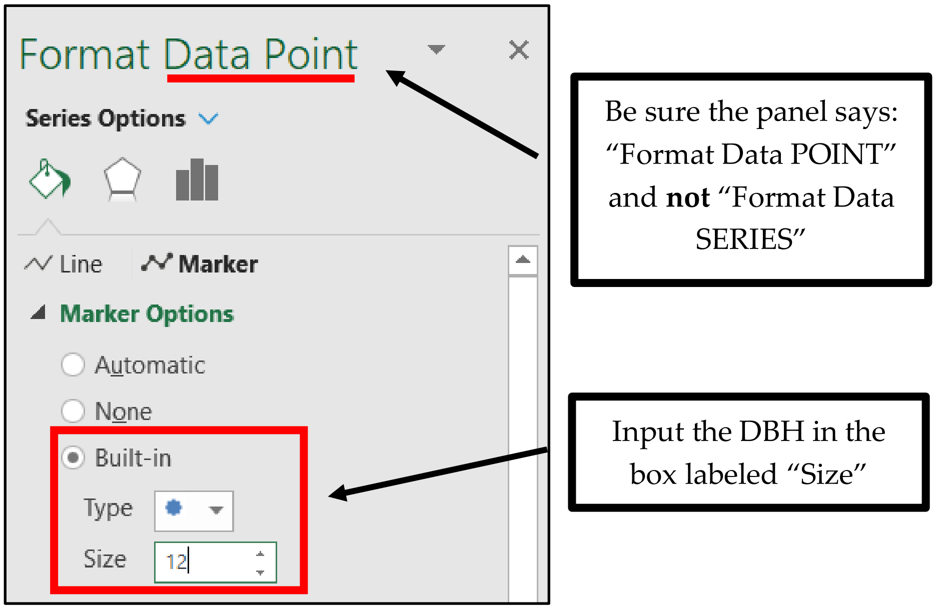


Figure 1: Graphic illustrating how to change each datapoint so its size represents the DBH of each tree.

Exercise 2: T-square sampling

Next, let's analyze the data from the T-square method and contrast the spatial dispersion result inferred from this plotless technique. To do this, we will calculate an **Index of Spatial Dispersion** (*C*) and then use a statistical test (e.g., ***Z*-test**) to determine which spatial pattern best fits the data.

1. Switch to the **"T-Square"** tab.
2. Calculate the Index of Spatial Dispersion (*C*), which is a numerical outcome that can be used to identify a regular, random, or clumped dispersion pattern.
    a. Type the formula "=B2^2" in cell D2 to calculate the first quantity needed to calculate *C* (d_1^2). Drag this formula down for all replicates.

    b. Next, type the formula "=C2^2" in cell E2 to calculate the second quantity needed to calculate *C* (d_2^2). Drag this formula down for all replicates.

    c. Third, type the formula "=E2*0.5" in cell F2 to calculate the final quantity needed in the numerator [$0.5 * (d_2)^2$]. Drag this formula down for all replicates.

    d. Now, we can determine the numerator in Column G by entering the formula "=D2/(E2+F2)" in cell G2. Drag this formula down for all replicates.

    e. To finish calculating *C*, sum (Σ) all values in Column G and divide that value by the number of replicates (*n*) you have in the sample. This will produce your Index of Spatial Dispersion (*C*):

$$C = \frac{\sum\left[\frac{(d_1)^2}{(d_1)^2 + 0.5(d_2)^2}\right]}{n}$$

- $C < 0.50$ for regular dispersion
- $C \approx 0.50$ for random dispersion
- $C > 0.50$ for clumped dispersion

Now that we have the Index of Spatial Dispersion (C), let's use a statistical test to determine whether or not the live oak individuals are dispersed randomly in space ($C \approx 0.50$). The "random" spatial dispersion pattern will serve as our **null hypothesis**, or the simplest explanation for the data collected. Our **alternative hypotheses** are that individuals are either uniform ($C < 0.50$) or clumped ($C > 0.50$).

Specifically, we will use a *Z*-test to estimate the probability that our *C* value is significantly different from random ($C = 0.5$). To do this, we will need to calculate a **test statistic**. Remember, that a test statistic (in this case *z*), is a numerical value that summarizes the data you collected in the context of the null hypothesis. We have previously only calculated *t*-statistics for our *t*-tests. Finally, we can then convert that test statistic to a **probability value** (*p*), which will tell us whether to reject or FAIL to reject our null hypothesis that the spatial dispersion is random. If $p > 0.05$, we fail to reject H_0, but if $p < 0.05$, we reject H_0 in favor of one of the alternative hypotheses (regular or clumped, depending upon the value of *C*)

1. Use your knowledge of Excel and the process above to calculate the z-score with the equation below. Remember to use the appropriate order of operations. For the square root function, use "=SQRT( )" to calculate the denominator quantity.

$$z = \frac{C - 0.5}{\sqrt{1/(12n)}}$$

2. Once you have your *z*-score, use this website to calculate your p-value. Input your z-score that you calculated, leave the significance level at 0.05, and check off a **TWO-TAILED** test. Next press "Calculate"!
3. Create a table with the Index of Spatial Dispersion and p-value you just calculated. Don't forget a proper caption.

Part II: ArcMAP Data Analysis

While scientists often use these field-based sampling methods to calculate spatial dispersion for small numbers of organisms, it's not always feasible to use these methods for large numbers of samples due to limited financial or labor resources. Luckily, more advanced spatial analysis software exists to fill this gap. During this part of your lab, we will explore the program ArcMAP to complete the same analyses as above, but this time with a new dataset to demonstrate how spatial dispersion can be determined with a larger sample. In this case, we'll examine the spatial dispersion patterns of wetlands within the Tampa Bay region. Consider during this analysis, what abiotic and biotic factors are most likely influencing their spatial dispersion pattern and how they differ from the live oak trees.

1. Understanding GIS Layers

ArcGIS maps are created from different shapefiles that project as layers on your map. These layers can be turned on and off in your display based on your needs. All the shapefiles that you will need for this assignment can be downloaded as a "zipfile" on Canvas labeled "Spatial Analysis Dataset". In addition to the shapefiles, this folder also contains a map that you can open with the shapefiles preloaded. You should not have to add any shapefiles to this map. To use this dataset, you will first have to extract the data from the folder.

1) Download the folder "Spatial Analysis GIS Dataset" on Canvas to your desktop/downloads folder.
2) Once the folder completes downloading, right click the folder and select "**Extract All**."
3) The extraction location should also be in your desktop.
4) Now you should be able to open the folder and find one labeled **PCB-3043L_SpatialAnalysis**. That is what you should open.
5) If the Map has **!** next to each of your layers, it's because the map can't find that layer's data files. You'll thus need to tell the map where your shapefiles are located. Select the **!** and then find the shapefile you selected in the folder you just downloaded. Fixing one **!** should fix them all (Figure 2).

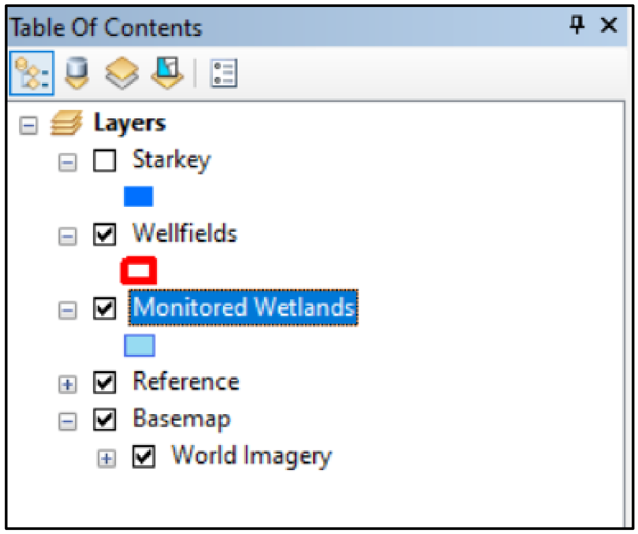


Figure 2: Table of contents with preloaded GIS layers. If you see !, this is an error message signifying you need to select where the data is located in your files.

Once the map is open, you will see the following shapefiles as layers:

a. **Monitored Wetlands:** This polygon layer shows the locations of all the wetlands in Tampa Bay monitored by the Southwest Florida Water Management District. These wetlands are primarily within parks where groundwater is extracted for drinking water to observe any negative changes to wetland health.
b. **Wellfields:** This polygon layer indicates areas where groundwater extraction is occurring and demonstrate the different spatial scales at which we can examine these wetlands – all of Tampa Bay versus within a particular wellfield.
c. **Starkey:** This polygon layer shows the locations of all the wetlands within the Starkey wellfield ONLY. This will be important when we consider different spatial scales.

2. **Explore Map Document**

To begin the analysis, we first need to know what we're looking at. In ArcMAP, data is stored in the "Attribute Table" and displayed using the "Symbology" tab under "Properties." To access either of those, all you have to do is right click on any layer and select one of those options (Figure 3).

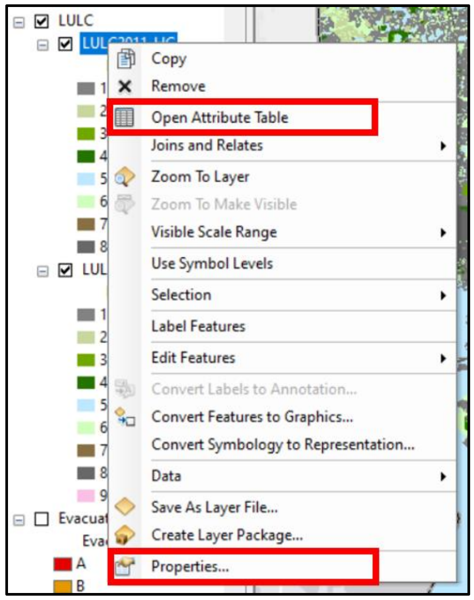
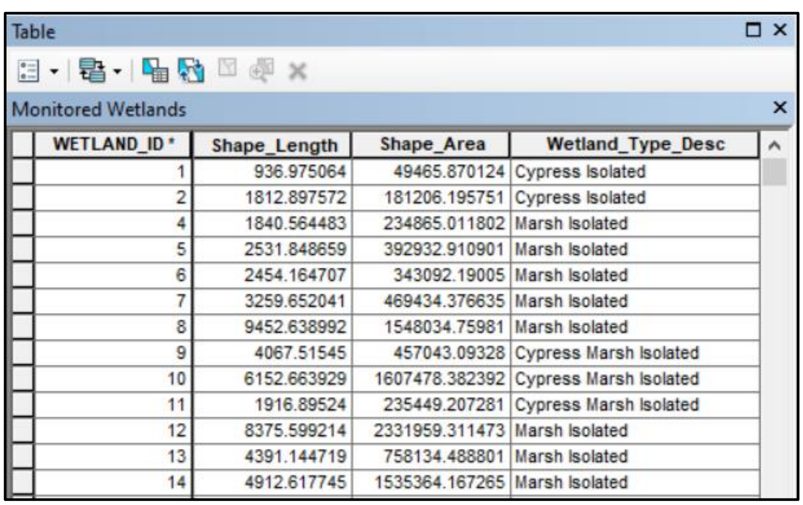


Figure 3: View different data properties for each layer.

The attribute table includes identifier and spatial data. Displayed below is the attribute table for the *Monitored Wetlands* dataset. There, you can see several columns: some which are necessary for spatially projecting the dataset and others that are descriptions you might collect when sampling. The "WETLAND_ID" column numbers the wetlands and the "Wetland_Type_Desc" indicates which type of wetland each ID corresponds to. The "Shape Length" and "Shape Area" columns refer to the size of each wetland polygon and helps spatially project them (Figure 4).

Any one of these attributes (e.g., WETLAND_ID, Wetland_Type_Desc) can be visualized using the "symbology" tab (Figure 5). You can edit the symbology (color ramp, size of points, etc.) and alter the map display by adjusting the value field. Take some time to play around with this. Hitting "Apply" will change it on the map and hitting Apply → Okay will close the window.



Table

Monitored Wetlands

WETLAND_ID *	Shape_Length	Shape_Area	Wetland_Type_Desc
1	936.975064	49465.870124	Cypress Isolated
2	1812.897572	181206.195751	Cypress Isolated
4	1840.564483	234865.011802	Marsh Isolated
5	2531.848659	392932.910901	Marsh Isolated
6	2454.164707	343092.19005	Marsh Isolated
7	3259.652041	469434.376635	Marsh Isolated
8	9452.638992	1548034.75981	Marsh Isolated
9	4067.51545	457043.09328	Cypress Marsh Isolated
10	6152.663929	1607478.382392	Cypress Marsh Isolated
11	1916.89524	235449.207281	Cypress Marsh Isolated
12	8375.599214	2331959.311473	Marsh Isolated
13	4391.144719	758134.488801	Marsh Isolated
14	4912.617745	1535364.167265	Marsh Isolated

Figure 4: Attribute table for Monitored Wetlands dataset.

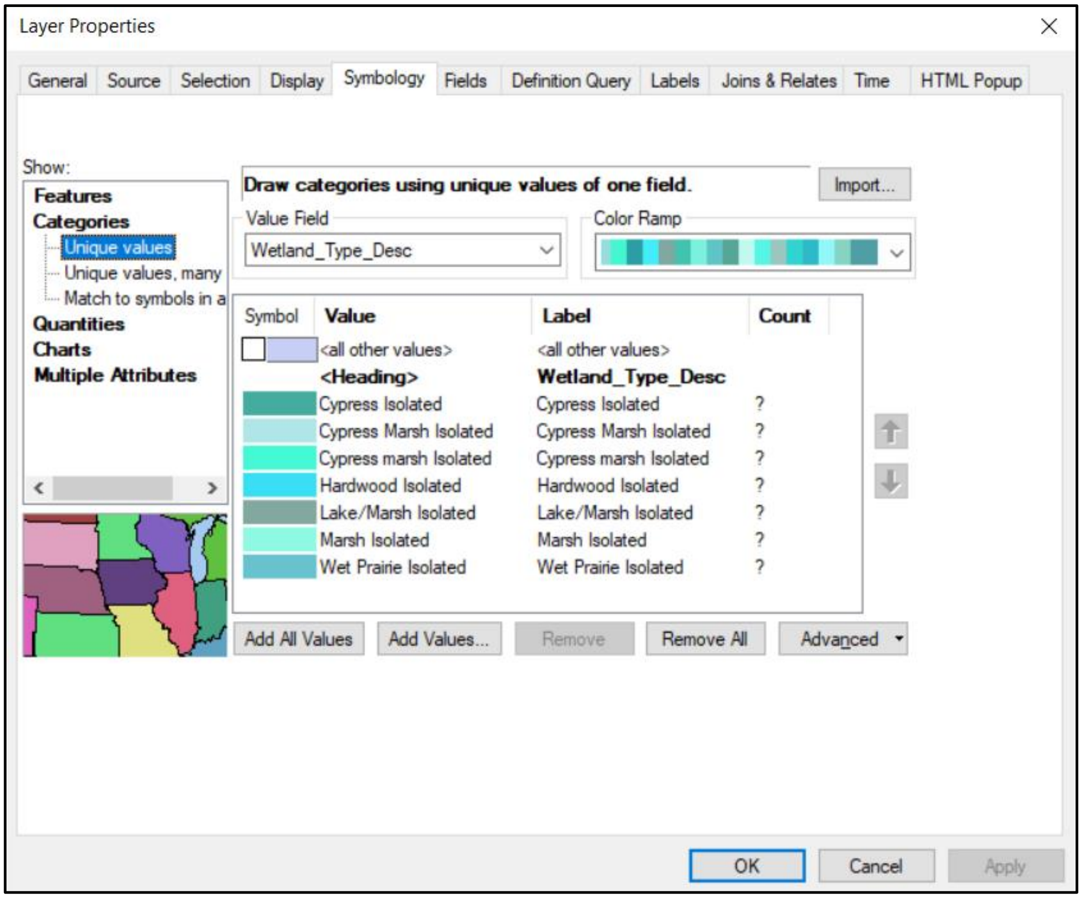


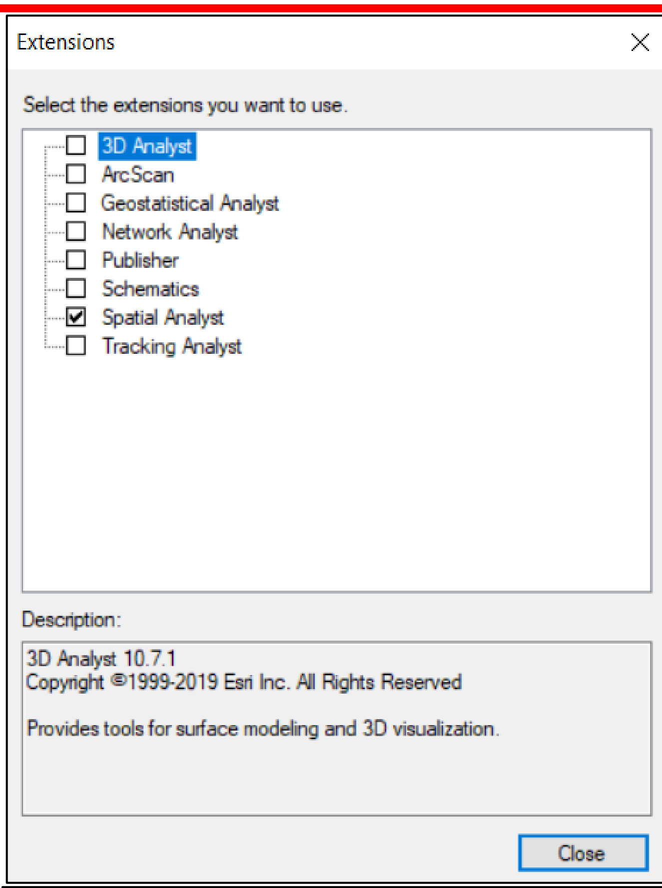
Figure 5: Symbology for Monitored Wetlands

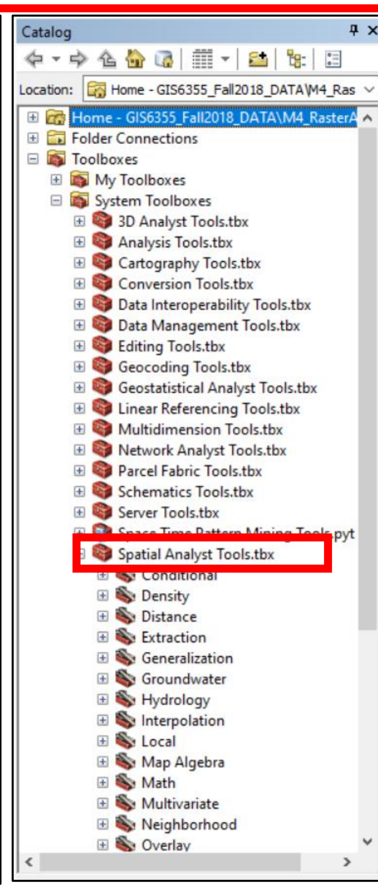
3. Use spatial analyst tools to determine spatial distribution patterns at different scales

You can do more in ArcGIS than just visualize spatial data. You can also manipulate this data same as you can in Excel to calculate new values, layers, and statistics. In this assignment, you will use some of the ***Spatial Statistics Tools*** to answer specific questions.

Before we start:

- **Turn on the license for Spatial Analyst from the main menu (similar to adding the Data Analysis toolpack in Excel): Go to Customize → Extensions (Figure 4).**





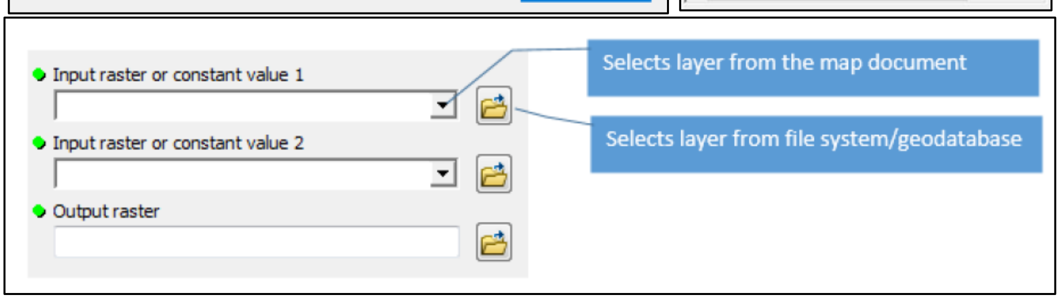


Figure 6 (top left): Window illustrating how to enable the Spatial Analyst tools we'll be using throughout this lab; Figure 7 (bottom): Illustration of the input options for many of the tools we'll be using. Refer to this, when tools have error options; Figure 8 (top right): ArcCatalog displaying where the Spatial Analyst tools we'll be using are located (circled in red).

1. Let's calculate the Index of Spatial Dispersion (*C*) for the entire monitored wetlands layer first. To do this, go to the Toolboxes → System Toolboxes → Spatial Statistics Tools → Analyzing Patterns → Average Nearest Neighbor
2. Select the "Monitored Wetlands" layer as the Input Feature Class. Leave the distance method as EUCLIDEAN_DISTANCE and the area as blank. Check the "Generate Report (optional)" box. Select okay (Figure 9).

- To see the results, select GEOPROSSESSING → RESULTS. You'll then see the NNRatio value which is similar to the Index of Spatial Dispersion value calculated before and can be interpreted the same way, where $C < 0.50$ for regular dispersion; $C \approx 0.50$ for random dispersion; and $C > 0.50$ for clumped dispersion (Figure 10).
    a. The p-value indicates whether the spatial pattern is significantly different from random.
    b. If you double click the "Report File" a report appears interpreting the results for you.

3. Let's now do the same analysis but for the Starkey layer to see how scale can influence spatial dispersion patterns. Follow the same steps above but select "Starkey" as your input feature class this time.
4. Add the NNRatio and p-values for both spatial scales (Tampa Bay and Starkey) to the table you created above. Be sure to interpret their results in your table caption.

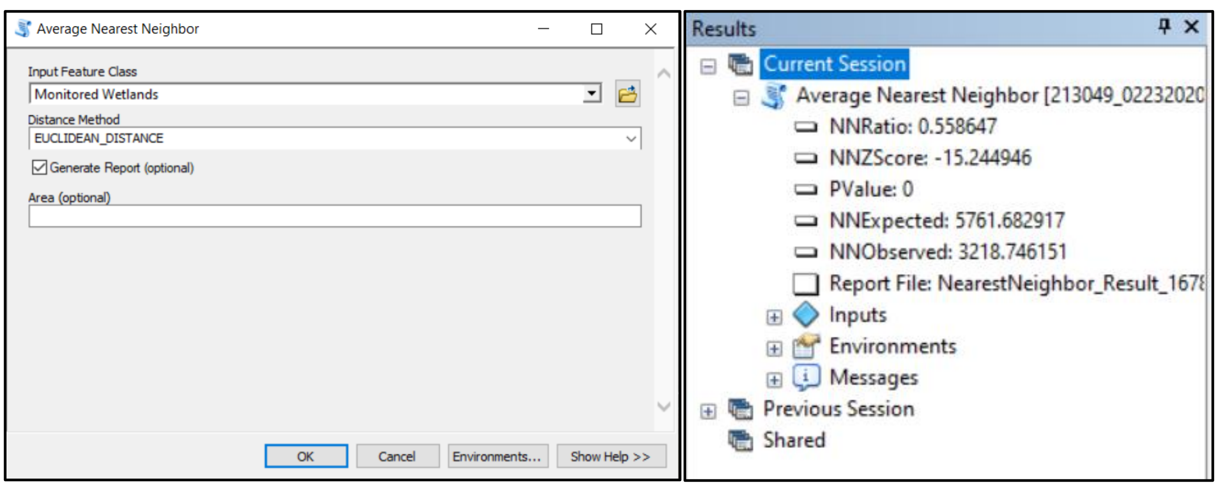


Figure 9 (left): Average Nearest Neighbor analysis options; Figure 10 (right): Results window.

Summary Questions (max 2-3 sentences each):

1. What is the major difference between the two field-sampling methods you used to determine the spatial dispersion of southern live oaks?
2. What type of spatial dispersion pattern does *Quercus virginiana* possess based on your 50m x 50m plot? Be sure to reference your figure in your response.
3. If you examine the field map, you should see little variation in the diameter at breast height (i.e., dot size). Provide an explanation as why southern live oaks show very little variation in DBH. Keep in mind that DBH can be a proxy for tree age.
4. What type of spatial dispersion pattern do the live oaks possess based on the Index of Spatial Dispersion (C) you generated from the T-square method? Explain how you used the results from the Z-test to come to this conclusion and be sure to reference your table in your response.
5. Provide one abiotic and biotic explanation for the spatial dispersion observed from the T-square method.
6. What type of spatial dispersion pattern did the monitored wetlands have at the full county scale compared to the Starkey wellfield based on the NNRatio values? Were the patterns significantly different from random? Explain why or why not and be sure to reference your table in your response.

Post-Lab Report Submissions

Your results (PDF preferred), along with your edited Excel workbook, should be uploaded to Canvas for your Spatial Dispersion & Analysis Post-Lab Report. Results include: full sentence responses to the thought and summary questions as well as any tables and figures you created and reference. Title each file "LastName_Spatial_Dispersion" and with the appropriate extension (.doc, .xlsx, .PDF). You can review the rubric for your submission at any time on Canvas under the assignment to ensure you have all the components.