Introduction to GIS: Lab 8

Geostatistics[[1]](#footnote-1)

This lab will explore the geostatistical analyst toolbox in ArcGIS, as well as a few other tools related to understanding **surfaces** and **spatial distributions**. Previously, you have worked with vector data analysis and raster data analysis. This lab will look at a few more tools that **combine** vector and raster data analysis.

**Before you begin, make sure the Spatial Analyst and Geostatistical Analyst extensions are active (Project -> licensing).[[2]](#footnote-2)**

The data that you’ll be working with for the lab portion are related to the distribution of an invasive grass called cheatgrass, which has invaded over 100,000 square kilometers of the western U.S.[[3]](#footnote-3) Here’s what an invaded landscape (left) looks like compared to a native landscape (right):

 vs. 

Throughout this lab, you’ll use this real-world scenario (and a few others) to contextualize your analytical approaches. **Each section of this lab introduces a new and useful tool to analyze some combination of raster and vector data.** These will all be necessary for the lab production, which **uniquely requires no map, but instead will be all statistics based**.

While the lab production activity is what you will turn in to Moodle for a grade, please note that practice questions throughout to test your own understanding of the concepts are usefully noted in **bold**. These questions are very useful for testing your capacities and will be helpful to complete the lab production as well![[4]](#footnote-4)

At the beginning of each lab part, the section is summarized as well as connected to the Geospatial Technology Competency Model (GTCM) so that you can see the links to broader practice of the topic at hand. In addition, a hyperlink to the walkthrough for that part is also present for ease of reference.

This lab is divided into eight parts, which you should complete in order:

**Part A:** Zonal Statistics

**Objective:** Use overlays between vector and raster data to create summaries and perform analysis on layer characteristics.

**Part B:** Extract Values to Points

**Objective:** Connect information from a raster layer to a shapefile point layer to enable analysis and visualization.

**Part C:** Create Random Points

**Objective:** Use ArcGIS to create a random sample of points to form the basis of comparison for other data samples.

**Part D:** Spatial Join

**Objective:** See how to join based only on location – no attribute tables necessary!

**Part E:** Spatial Clustering I

**Objective:** Explore how to measure spatial distributions and encounter an entirely new way to read output within ArcGIS Pro.

**Part F:** Spatial Clustering II

**Objective:** Take another shot at pattern analysis with the classic Moran’s I.

**Part G:** Interpolation

**Objective:** Use different methods to apply statistical techniques to ‘fill in’ areas of a dataset without observed values.

**Part H:** Lab Production: Suitability Analysis of White Top (*Lepidium draba)* in Idaho

**Objective:** Combine your spatial toolsets to explore some invasive species distribution data in Idaho

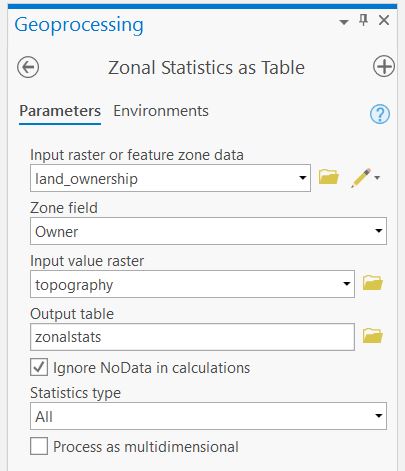
# Part A: Zonal Statistics

*Summary: This part explores the idea of using area information or classifications stored in a vector or a raster format to understand the distribution of other data stored in a raster format. With this capacity, we can understand how data is distributed by some attribute or characteristic, rather than just as a whole raster.*

[*Geospatial Technology Competency Model*](https://www.careeronestop.org/competencymodel/competency-models/geospatial-technology.aspx)*: 5.2.19.4, 5.2.8, 4.1.13.2, 4.1.11.2, 4.1.9.1, 4.1.6.7, 2.4.3.1, 2.1*

[*Link to Video Walkthrough*](https://youtu.be/50PRo34uENs)

1. As we’ve experienced, vector and raster data operate in different ways and allow us to understand different components of our problems through their comparative advantages. Zonal Statistics works as a way for us to understand information stored in a raster based on discrete zones (stored as polygons) or classifications (stored as discrete values in a raster)[[5]](#footnote-5)
2. This tool works with a polygon shapefile and some other information stored in a raster. This raster represents the information we want to know something about, aggregated by the zones of the polygon. By applying zonal statistics, you will calculate the statistics of the raster grid for each class specified within the polygon shapefile.
3. For this example, our goal is to compile some statistics about topography (from a DEM (digital elevation model – a raster, remember?) within different land classes (in a polygon shapefile).[[6]](#footnote-6)
4. Add the layers *topography* and *land\_ownership* to your map.
5. Open the attribute table of *land\_ownership* and notice that the column ‘owner’ lists different landowners. This study site is in Nevada, so a lot of the land is public.
6. Change the symbology of the land\_ownership layer to classify each owner in a different color. NOTE: When you examine the attribute table, you’ll there are a lot of rows in this feature layer; before you start a complicated procedure on a layer with lots of observations, you should save your project just in case Arc crashes during the operation.
   1. **Practice: What agency owns/manages the largest land parcel in this study area?[[7]](#footnote-7) How big is the parcel?**
7. With all this information available to us, we could use some computational assistance in making sense of it. In Toolboxes, go to Spatial Analyst Tools -> Zonal -> Zonal Statistics as Table.[[8]](#footnote-8)
8. The first line, ‘Input raster or feature zone data’ is the layer that defines your zones.[[9]](#footnote-9) Set it to land\_ownership, since we are interested in understanding our statistics through the lens of who owns the land.
9. The ‘Zone field’ line defines which values will be used to compute the zonal statistics.[[10]](#footnote-10) Set the zone field to owner, since we are interested in the owner component of the land\_ownership layer.
10. The ‘Input value raster’ should be the layer that has the data we want to compute: set this one to topography.[[11]](#footnote-11) This command will calculate basic statistics about topography (elevation) for each type of land ownership.
11. Remember to choose the appropriate folder for the output raster and name the file "zonalstats.”



1. The nice thing about this command is that you can have multiple polygons of a particular class (e.g., BLM owned) and zonal statistics will consider them all, i.e. aggregate them, when calculating statistics. This way, you don’t have to calculate statistics for individual polygons.[[12]](#footnote-12)
2. Open up the new file and read the results.[[13]](#footnote-13)
   1. **Practice: The mean elevation of BLM lands within our study site should be ~1605 m. What is the mean elevation of USFS (Forest Service) lands within our study site?**
3. The output stats file is a .dbf, which you can export to Excel, or another spreadsheet program to create charts. You can also view the information as a table to compare the elevation ranges of the different land ownership classes.

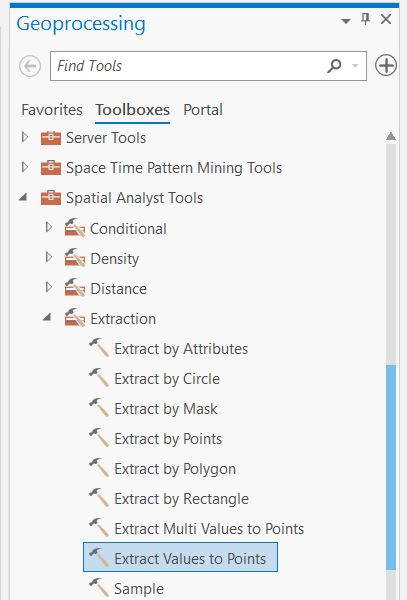
# Part B: Extract Values to Points

*Summary: This part furthers connections between vector and raster data by using a point shapefile as a lens to gather information from a raster layer. This method allows for quick linking of raster values into a shapefile attribute table.*

[*GTCM*](https://www.careeronestop.org/competencymodel/competency-models/geospatial-technology.aspx)*: 5.2.19.4, 5.2.8, 4.1.13.2, 4.1.11.2, 4.1.9.1, 4.1.6.7, 2.4.3.1, 2.3.6, 2.1*

[Link to Video Walkthrough](https://youtu.be/jpbaIPzpXM8)

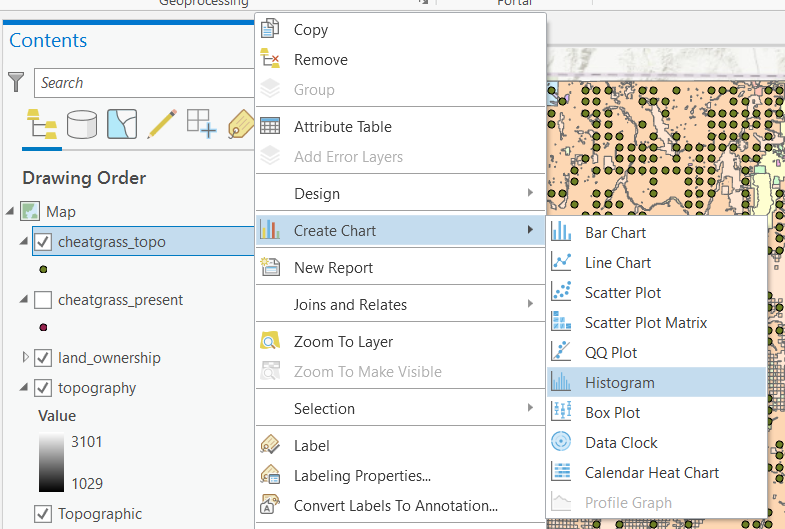
1. Now, let’s say we have a point distribution of a species that we’re interested in. In this case, we have point distribution data for cheatgrass within the study area.
2. Add the shapefile cheatgrass\_present to the map. We want to know whether cheatgrass is more likely within particular ranges of elevation. But, the elevation (topography) data are a raster grid. We can extract the values of the raster grid into our point shapefile using the ‘Extract Values to Points’ command, which you can find under Spatial Analyst Tools -> Extraction.[[14]](#footnote-14)



1. Select cheatgrass\_present as your Input point feature, and topography as your input raster.[[15]](#footnote-15) ArcGIS will output a new shapefile for this process, so specify a name and location for that shapefile.[[16]](#footnote-16)
2. Open the attribute table of the new shapefile you just created and notice that it has a new column called ‘Rastervalu’, which in this case is the elevation of each point location.
   1. **Practice: What unit is the elevation in? Why? How could you confirm if you didn’t know for sure?**



1. We have a few options for analyzing this information: You could open the .dbf file in excel to figure out basic statistics of these 872 points. Or, you can calculate mean and standard deviation by right clicking the column heading RASTERVALU and then statistics.
2. In ArcGIS Pro, you can also explore your data by creating a histogram or other kinds of graphs. Right click on the layer you just created and choose Create Chart > Histogram.



1. In the Chart Properties pane, set Number to ‘Rastervalu,’ the elevation values that you just extracted. At the bottom of the pane, you can see the summary statistics for these elevations. The mean elevation is 1448.7 m and the standard deviation is 154 m.
2. In Chart Properties, you can change the number of bins in your histogram, show or hide a normal distribution curve, display mean, median, or standard deviation lines, and perform a transformation on your data if they were skewed (you can see the value of data skewness in the statistics section – the larger the absolute value, the more skewed your data). In this case, the distribution is close to normal so there’s no need for a transformation, but you can check out what happens if you transform the histogram anyway.[[17]](#footnote-17)
3. Click to select the bins at the low elevation end of the histogram and notice that the points corresponding to these elevations are also selected in the data view. This way, you can quickly identify and highlight any outliers from your distribution data.[[18]](#footnote-18)
4. If you want to put this histogram into a map, you can add it to your layout and include it in a presentation of your data later on.[[19]](#footnote-19)
5. We’ve figured out that our cheatgrass presence points have a mean elevation of 1448.7 m. But, is that elevation significantly different from the available land area? To calculate summary statistics for the study area, we can use zonal statistics.[[20]](#footnote-20) But first we need a polygon shapefile of the study area.
6. Create a new polygon shapefile outlining the study area (approximate extents are okay). You have a few options, including tracing out a new polygon by hand[[21]](#footnote-21) , or [creating a minimum bounding geometry](https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/minimum-bounding-geometry.htm)[[22]](#footnote-22) (for example a convex hull).
   1. Make sure to set the projection to be the same as topography when creating the layer.
   2. Your study area polygon should roughly match the extent of the topography layer.[[23]](#footnote-23)
      1. To be very clear: You need to create a shapefile (a polygon) that is the same extent as the topography layer.[[24]](#footnote-24)
      2. Since the topography layer is a big square, the shapefile (polygon) you create (just like you traced the campus pond, for example, in Lab 5), should also look like a big square.
      3. How can I say this in another clear way? You need a shapefile that shows your study area. Your study area is the whole topography layer. So your study area shapefile should be a big square – like the topography raster![[25]](#footnote-25)
      4. We’ll use this file later on! So, save it in a good place and name it something better than ‘New\_Shapefile.’[[26]](#footnote-26)
   3. Use the Zonal Statistics as Table tool to calculate the mean elevation of the study area. The number should be around 1580 with a standard deviation of 300.
7. Based on these numbers, we can begin to infer that cheatgrass might prefer to invade a slightly lower elevation than the mean of the study area.[[27]](#footnote-27)

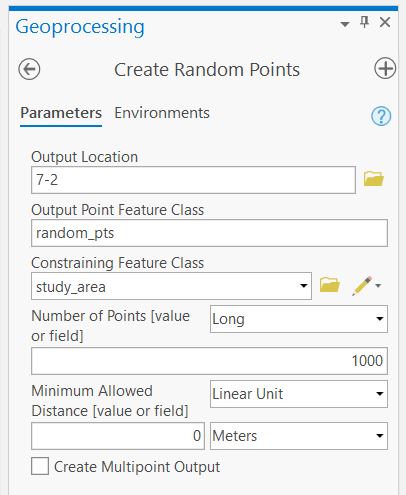
# Part C: Create Random Points[[28]](#footnote-28)

*Summary: To support further investigation of the nature of the distribution of these cheatgrass points, this section outlines how to create a set of random points for comparison.*

[*GTCM*](https://www.careeronestop.org/competencymodel/competency-models/geospatial-technology.aspx)*: 5.2.19.4, 5.2.8, 5.2, 4.1.13.2, 4.1.11.2, 4.1.9.1, 4.1.6.7, 2.4.3.1, 2.1*

[Link to Walkthrough Video](https://youtu.be/CNV2Ik5iPLs)

1. If you want more information than just the mean topography value for comparison to the values for cheatgrass distribution (e.g., you want to compare two histograms), then you might want to create a set of random points to compare to the species distribution.
2. Create Random Points is hiding out in Toolboxes under Data Management Tools -> Sampling -> Create Random Points.[[29]](#footnote-29)
3. In the Geoprocessing pane, choose your output folder, and name your file. You’ll also want to specify output extents for your points in the ‘constraining extent’ box.[[30]](#footnote-30) Set this to the study area polygon that you created earlier[[31]](#footnote-31). Also specify an output of 1000[[32]](#footnote-32) points (the more points you make, the longer it will take to process).[[33]](#footnote-33) The remainder of the optional inputs we will leave blank.



1. Click Run to create your random point distribution within the study area extents.
2. Now, like you just did with the cheatgrass points shapefile, use ‘Extract Values to Points’ to pull the topography values from the grid into the random points shapefile.
3. Create a histogram for this new shapefile. Rename your two histograms in the Contents pane so that you can tell them apart (having both named RASTERVALU Distribution is not very helpful).
4. Add the histograms from the cheatgrass points and the random points to a layout so that you can compare them side by side. To add a chart frame to a layout, you first have to add a map frame to your layout with the layers containing the charts you want to use. Then you can insert the charts from the Map Surrounds group.
5. The random points histogram tells you about topography in the whole study area, while the cheatgrass points histogram tells you about topography where cheatgrass is present.
   1. **Practice: What differences do you see between the histograms?**
   2. (In an ideal world, for analytical purposes we’d prefer to compare the environmental characteristics of locations where a species is *present* to locations where a species is *absent*. But, if we don’t have absence data for our target species, we can instead compare presence to total available land area.)
   3. What we’ve done here is a form of exploratory spatial analysis – looking for patterns visually without performing any formal analytical steps. Useful for identifying patterns as a prelude to such analysis.[[34]](#footnote-34)

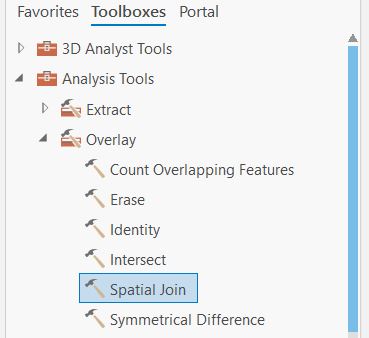
# Part D: Spatial Join

*Summary: Merging the concepts of a database join and spatial overlay, this part explores how to connect vector attributes together when no database key exists.*

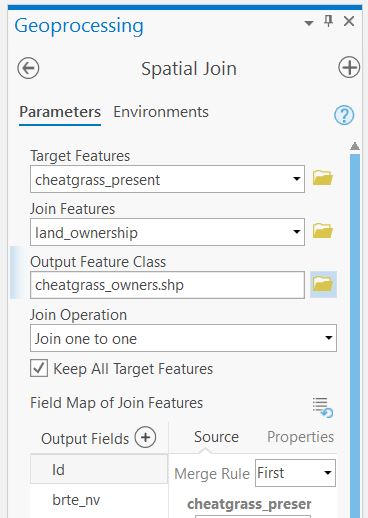
[*GTCM*](https://www.careeronestop.org/competencymodel/competency-models/geospatial-technology.aspx)*: 5.2.19.4, 5.2.8, 5.2, 4.1.13.2, 4.1.11.2, 4.1.11.1, 4.1.9.1, 4.1.6.7, 2.8.6, 2.4.3.1, 2.1*

[Link to Walkthrough Video](https://youtu.be/-Dy5_Nen4gY)

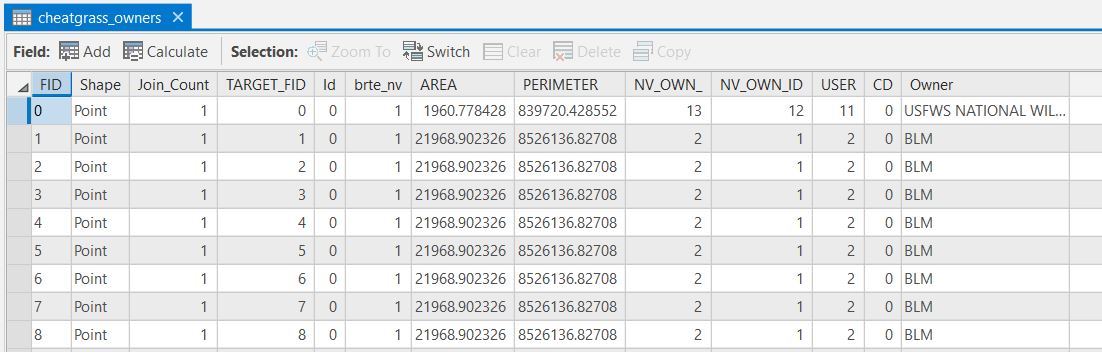
1. You’ve calculated the distribution of cheatgrass present relative to topography. Now, let’s look at its distribution relative to land ownership. If you removed it previously, add land\_ownership back into your map.
2. In this case, we can’t use the extract values to points command because land ownership is a polygon shapefile, not a raster grid. And since we want to preserve the detailed shapes of the polygons, we don’t want to convert that file to a raster grid. Instead, we need to use what’s called a spatial join, which combines the attribute tables of two shapefiles based on their location.[[35]](#footnote-35)
3. Find spatial join in the Analysis Tools -> Overlay -> Spatial Join.[[36]](#footnote-36)



1. The attributes of the Target Features and the Join Features will be combined into the Output Feature based on their spatial location. So, let’s make our target features (or, the features who we want the geometry from) the cheatgrass\_present file, to which we will join land\_ownership as the join feature (or, the features we want to append to our other feature’s geometry. Choose a logical output shapefile name and location and click Run.[[37]](#footnote-37)



1. Your new output should contain the attributes from land ownership, including ‘owner.’

`

* 1. **Practice: Summarize this chart from the owner column to calculate a count of points in each land ownership category. What percentage of cheatgrass points are on BLM lands?**
  2. **Practice: Now, follow the same procedure using your random points. What percentage of random points are on BLM lands? Does that seem very different from the cheatgrass points? Do you think that cheatgrass is preferentially invading BLM lands?**

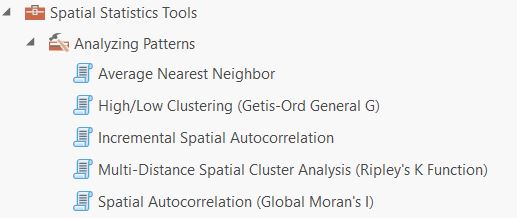
# Part E: Spatial Clustering I

*Summary: Diving into more formal measures of spatial relationships, this part begins to explore spatial statistics with more specific tools.*

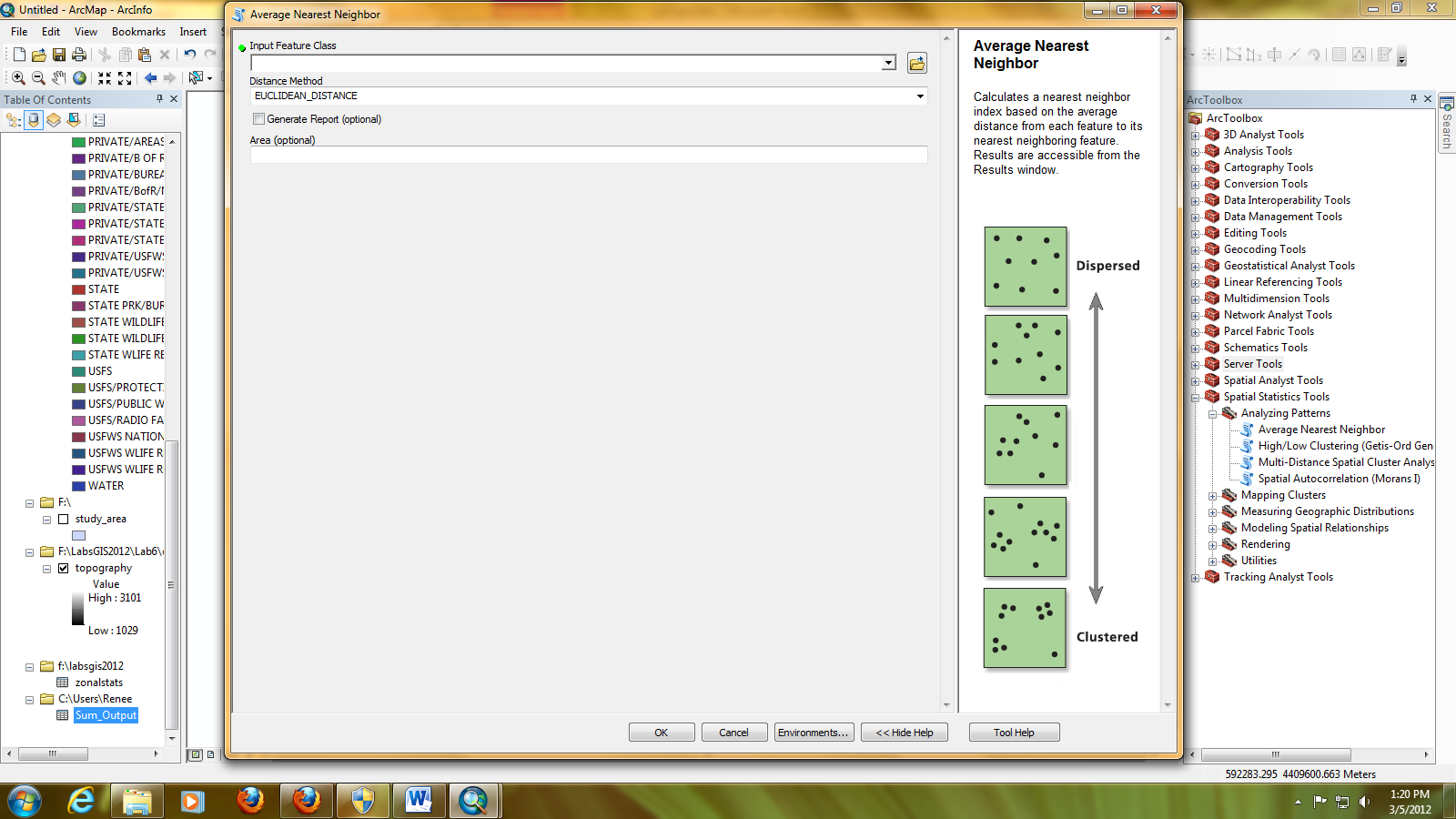
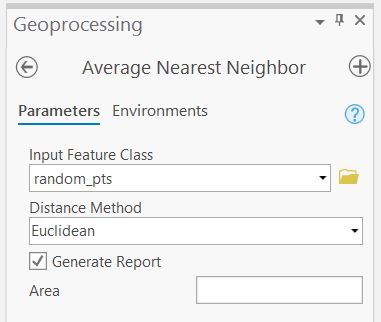
[*GTCM*](https://www.careeronestop.org/competencymodel/competency-models/geospatial-technology.aspx)*: 5.2.19.4, 5.2.8, 5.2, 4.1.13.2, 4.1.11.2, 4.1.9.1, 4.1.6.7, 2.4.3.1, 2.1*

[Link to Video Walkthrough](https://youtu.be/tKqicOLdb1g)

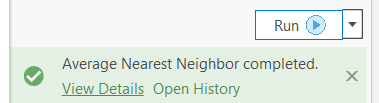
1. Clustering is relevant for point, polygon, or line distributions. A cluster analysis could help you determine whether points that you collected are spatially biased, or whether species presence points are clustered vs. randomly distributed.[[38]](#footnote-38)
2. If presence points are clustered, that suggests that something is affecting their distribution (e.g., a species only grows on south facing slopes, or on sandy soils, or near roads). If something is affecting a species distribution, then chances are good that we can try to predict what that is and better understand why that species is where it is (the essence of biogeography)[[39]](#footnote-39).
3. The ArcGIS tools used to measure clustering can be found in Toolboxes under Spatial Statistics Tools > Analyzing Patterns

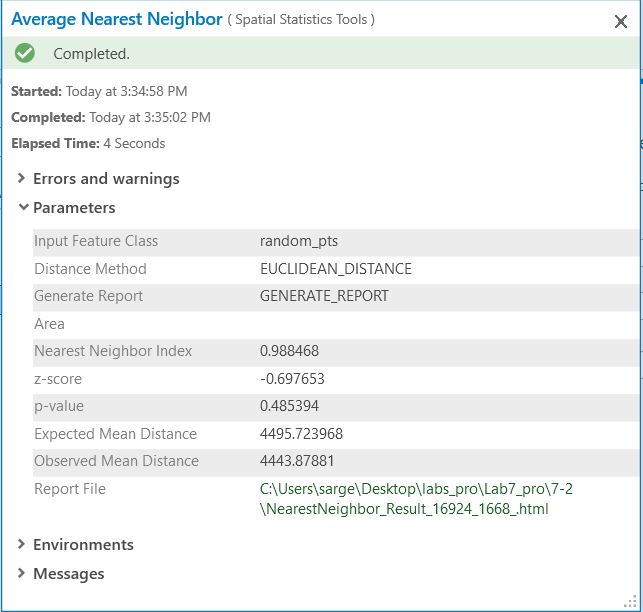


1. Open ‘Average Nearest Neighbor’, which compares the locations of points to determine whether they are clustered, random, or evenly distributed. Select the random points file you created earlier. Make sure you click ‘generate report’ as an option so that ArcGIS produces an html output.[[40]](#footnote-40) Click Run.

1. **Once this process finishes, it will look like nothing has happened**. That’s because ArcGIS ‘conveniently’ stores the .html report in the folder where your default geodatabase is located. To view the report results, click on ‘View Details’ in the green progress and messages box that appears at the bottom of the Geoprocessing pane. The output values are displayed in the Parameters section of this pop-up window. You can view the full .html report by navigating to it through Windows explorer, or by clicking on the Report File link in the pop-up.

****



1. Note – the details and messages box is a helpful feature to know about because when processing doesn’t work properly, the View Details option will include some description of what went wrong, under Errors and Warnings (not always, but occasionally useful).
2. Click on the Report File link. The output .html document should tell you that your point distribution is random…which it is! So, that’s good.[[41]](#footnote-41)



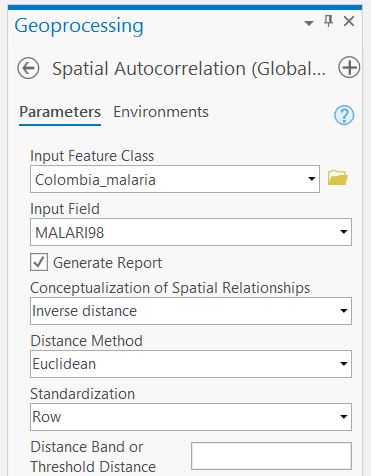
# Part F: Spatial Clustering II

*Summary: Moran’s I is the focus of this part, a key spatial pattern analysis tool for use with polygon shapefiles.*

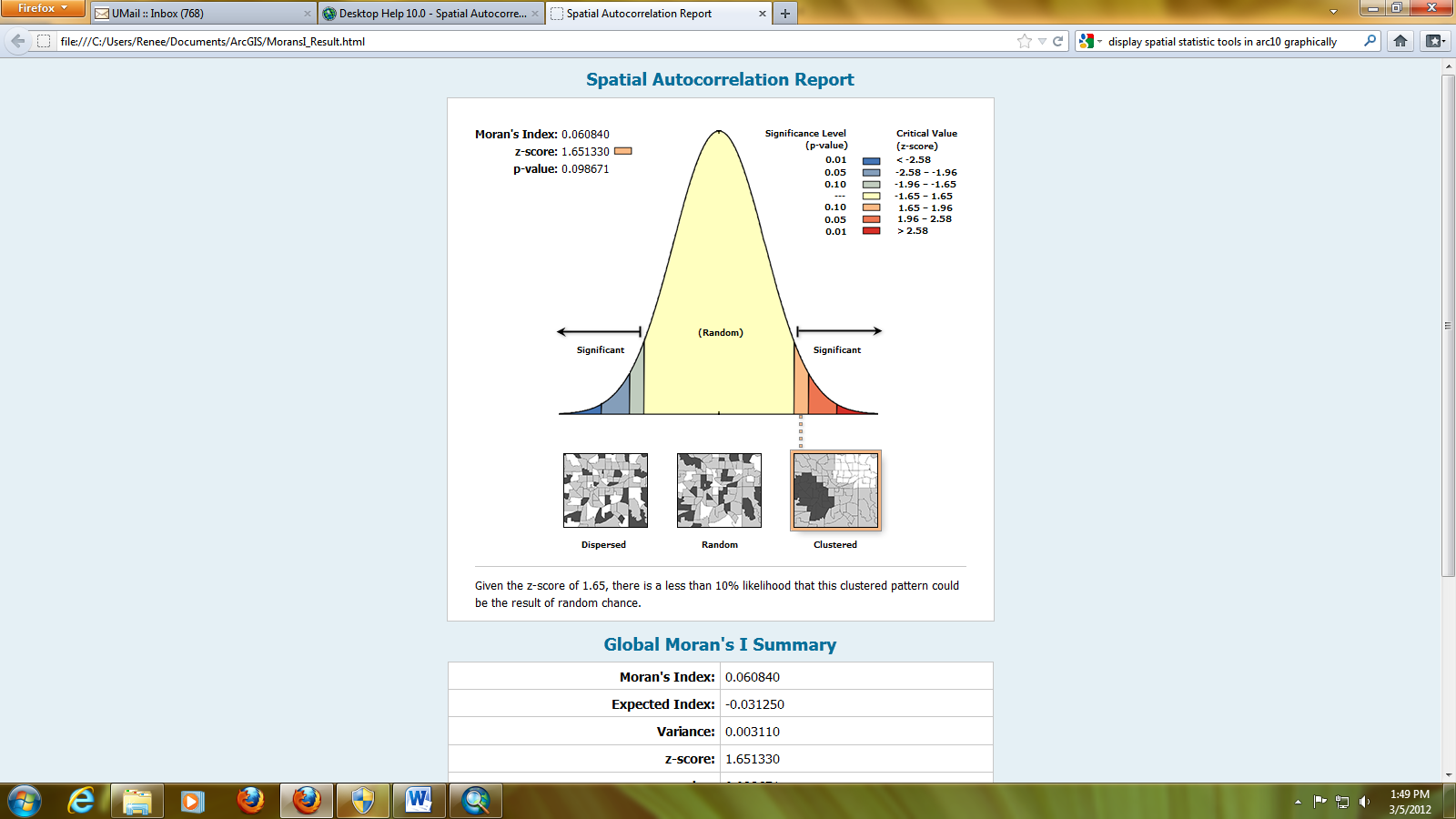
[*GTCM*](https://www.careeronestop.org/competencymodel/competency-models/geospatial-technology.aspx)*: 5.2.19.4, 5.2.8, 5.2, 4.1.13.2, 4.1.11.2, 4.1.9.1, 4.1.6.7, 2.4.3.1, 2.1*

[Link to Video Walkthrough](https://youtu.be/_BEL5JPeBYU)

1. Another pattern analysis tool is Moran’s I, which analyzes how spatially autocorrelated (clustered) the data in an attribute table are relative to the distribution of the shapefile.
2. Add a new Map to your Project and add Colombia\_malaria.[[42]](#footnote-42) Open the attribute table and notice that, along with lots of other information, there is a column of malaria incidents in 1998. You should set the display to symbolize Colombian provinces by malaria incidents in 1998.
3. Do you notice any obvious patterns? Do provinces with high or low malaria incidence seem to be clustered, or interspersed?
4. From the Spatial Statistics Toolbox > Analyzing Patterns, select ‘Spatial Autocorrelation (Global Moran’s I).’
5. Use Colombia\_malaria as the input shapefile, and Malari98 as the input field. Make sure to select Generate Report.



1. Click Run. ArcGIS will now calculate whether there is a spatial pattern of number of incidents of Malaria within the Colombian states. To view the results, find the .html document in your default geodatabase folder or click on View Details in the Geoprocessing pane and click the Report File link.



1. The Moran’s I output shows that the data are clustered rather than randomly distributed. The Moran’s I score (0.06) also gives you a confidence range. 0.06 means that if we created 100 random distributions, only 6 of them (0.06 \* 100) fswould be as clustered or more so than this one. The other 94 would be less clustered. So, that gives us some confidence that the clustering we’re measuring here is real and not just a random occurrence.
   1. **Practice: Change the symbology of Colombia\_malaria to display by graduated color, and make ten classes based on quantile. Does the map display also suggest some clustering? Why do you think malaria outbreaks would be spatially clustered in Colombia?**
2. Why would you bother with any of this stuff? Only if you have data *or attributes of data* that you think might be clustered. The previous two analyses provide spatial statistics (nearest neighbor and Moran’s I) to support any assertion on your part that your data are indeed clustered (or, that your data are most likely randomly distributed).

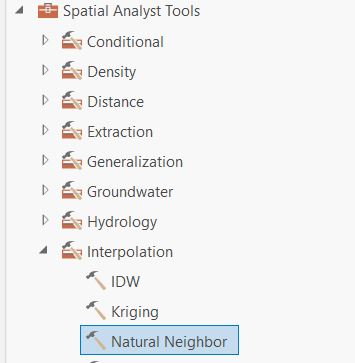
# Part G: Interpolation

*Summary: Using spatial statistics to infer the values between measurements is a powerful tool. This part explores how to create surfaces using some familiar data from a previous part of the lab.*

[*GTCM*](https://www.careeronestop.org/competencymodel/competency-models/geospatial-technology.aspx)*: 5.2.19.4, 5.2.8, 5.2, 4.1.13.2, 4.1.11.2, 4.1.9.1, 4.1.6.7, 2.4.3.1, 2.1*

[Link to Video Walkthrough](https://youtu.be/ry-9fQJexEg)

1. Let’s say you have a bunch of points, and you want to estimate values in between those points. Here’s a case where you need interpolation to create a raster surface from your point data.[[43]](#footnote-43)
2. Add a new Map to your Project and add the point shapefile you extracted the values to in Part B.
3. You might imagine a situation where, many iterations of this data being shared down the line, this is the only record of the elevation you have left!
   1. **Practice: Change the colors of the points to a graduated color scheme showing low to high elevations. Generally, where are the high elevations and the low elevations?**
4. We’ll start with a Natural Neighbor interpolation. This method assumes that every value is weighted equally, and it does not interpolate beyond the boundaries of the point distribution. You can find it in Spatial Analyst Tools -> Interpolation.



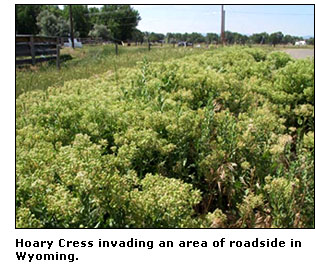
1. Make sure to select RASTERVALU as your Z value field, and for ease of comparison to the original topography, enter 30 as the cell size.[[44]](#footnote-44) Click Run to create the interpolated grid. Pick a color ramp you think helps visualize the pattern.
2. Run the IDW and Spline interpolations from the same toolbox, using the same Z value and cell size as before. Leave any other settings as default.
3. Change the color ramps of the three interpolations to be the same and compare the resulting interpolations. IDW weights closer points much more heavily than far away points (weighting default is 1/distance2)[[45]](#footnote-45), which can create a ‘bullseye’ effect. Splines are a mathematical best fit that work well along regular gradients but can over and/or under predict interpolated values relative to your data values.
   1. **Practice: What is the lowest pixel value for the spline interpolation? Does that seem reasonable to you? Can you explain why that happened?[[46]](#footnote-46)**
4. If you want to get fancy with interpolations, you would also want to check out kriging, which is a much more mathematically complex method that also has numerous options you could play with. We won’t do that here, but if your GIS project requires some interpolation, kriging might be a method you’ll want to look into further.[[47]](#footnote-47)
5. If you want to have bunches of fun, add the original topography layer into this map. What does the interpolation capture? What does it miss?

# Part H, Lab Production:

**Suitability Analysis of White Top (*Lepidium draba)* in Idaho**

[Link to Video Walkthrough](https://youtu.be/xIWiSd6RgFM)

**Problem Statement:** White top (*Lepidium draba*) is an invasive plant that has become more and more problematic in western states.



White Top invasion in Wyoming

Concerned about this invasion, the Idaho state department of agriculture (ISDA) has collected point locations of White Top invasion within the state using GPS. They have asked you to figure out if the data that have been collected so far are representative of the state as a whole.

ISDA has provided you with a list of lat/lon coordinates where White Top has been found in the file *white\_top\_idaho.csv*. Their GPS receivers used the NAD27[[48]](#footnote-48) datum. Convert this table to a shapefile.

**Data**:

*Idaho\_major\_roads* (a line shapefile of major roads in the state)

*Idaho* (a polygon outline of the state)

*Idaho\_land\_ownership* (a polygon shapefile of land ownership in the state)

*Id\_ann\_ppt* (a raster grid of interpolated annual precipitation in mm)

First, since we’re only interested in Idaho, remove[[49]](#footnote-49) all the points that are outside the state boundaries.

Second, set the geoprocessing environments processing extent to be the same as the Idaho layer.[[50]](#footnote-50)

**Your lab production will consist of the answers to the questions below along with the two histograms of white top and random points relative to annual rainfall in Idaho (#10).[[51]](#footnote-51)**

**Questions to be Submitted:**

Distribution relative to roads[[52]](#footnote-52)

1. For the collected White Top points, what is their average distance to a major road?
2. Within the entire state of Idaho, what is the average distance to a major road?
3. Explain how you calculated 1&2.
4. Are the White Top points skewed relative to distance to road? Explain how this analysis helps you know this. Give one hypothesis why the point location data might (or might not) be skewed relative to distance to major roads.[[53]](#footnote-53)

Distribution relative to land ownership

1. Calculate the area (in km2) of each land ownership class in Idaho. Explain how you did this.
2. Only BLM, Private & State lands are available for future sampling. These lands represent a total land area of 126,123 km2. Relative to the total land area available for future sampling, what percentage of land area is owned by:
   1. BLM
   2. Private
   3. State
3. Calculate the percentage of White Top points on land owned by:
   1. BLM
   2. Private
   3. State
4. Of the three land ownership classes, which have been oversampled? Which have been under sampled?

Distribution relative to annual rainfall

1. Are the White Top points found in areas of lower or higher annual rainfall than the state as a whole?
2. Create a histogram of the White Top points relative to annual rainfall. Create a histogram to compare this to for a random set of points in Idaho.[[54]](#footnote-54)

1. University of Massachusetts – Amherst, ArcGIS Pro Edition

   Written by Bethany Bradley, Forrest J. Bowlick, Sophie Argetsinger, Steven Bittner, Brit Laginhas, Chloe Thompson, Connor Hughes, and many others [↑](#footnote-ref-1)
2. Just like last week, but this time with another extension to add on! [↑](#footnote-ref-2)
3. Massachusetts is about 20,000 square kilometers. [↑](#footnote-ref-3)
4. Practice answers are also provided. See the link on the GitHub site in the Labs section [↑](#footnote-ref-4)
5. Very useful, and a once-upon-a-time GIS pipe dream. [↑](#footnote-ref-5)
6. Perhaps answering a hypothetical question of ‘what are the comparative topographies of these different land cover types?’ [↑](#footnote-ref-6)
7. Wait, what unit is that ‘Area’ field in? No, I don’t know either. You should re-calculate it. [↑](#footnote-ref-7)
8. You could search for Zonal Statistics as Table in the search tab too. [↑](#footnote-ref-8)
9. ‘Input zone data’ would be the easier way to read this – but those zones could be rasters or vectors. Fun! [↑](#footnote-ref-9)
10. Intuitive for once. [↑](#footnote-ref-10)
11. Where are the numbers? The numbers are here. [↑](#footnote-ref-11)
12. But you could if you wanted to by using the FID of each polygon. [↑](#footnote-ref-12)
13. The output is a table, hence the tool name. No map layer to see here. Move along. [↑](#footnote-ref-13)
14. Screenshot for that on the next page. [↑](#footnote-ref-14)
15. Which file has your points? Which file is a raster? [↑](#footnote-ref-15)
16. This is important to remember: you get a new shapefile out of this process, so don’t worry about the old one! [↑](#footnote-ref-16)
17. Remember back to all of your stats courses on how distributions function. [↑](#footnote-ref-17)
18. ooooooooooooooooooh [↑](#footnote-ref-18)
19. aaaaaaaaaaaaaaaaaaaah [↑](#footnote-ref-19)
20. We just did this in Part A! [↑](#footnote-ref-20)
21. We just did this…in Lab 5! [↑](#footnote-ref-21)
22. Check out this handy help entry from ESRI [here at this link.](https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/minimum-bounding-geometry.htm) [↑](#footnote-ref-22)
23. That is, draw a big square. [↑](#footnote-ref-23)
24. Oh no, you did not just name your new shapefile ‘New Shapefile’. [↑](#footnote-ref-24)
25. I’m not kidding when I say just draw a big square. [↑](#footnote-ref-25)
26. Oh snap, did he just write something in the main text that is also a footnote? [↑](#footnote-ref-26)
27. But it would be really nice if we could quantify this better. [↑](#footnote-ref-27)
28. Aka, quantifying it better [↑](#footnote-ref-28)
29. A search would be easier, as always [↑](#footnote-ref-29)
30. Don’t make Arc calculate random points for the entirety of existence. [↑](#footnote-ref-30)
31. Remember the big square [↑](#footnote-ref-31)
32. A nice round number [↑](#footnote-ref-32)
33. Sometimes you might have specific random point needs, but in this case, we’re going for a roundish, high number. [↑](#footnote-ref-33)
34. And the primary point of GTCM 5.2 – Analysis and Modeling Don’t know what you have until it’s ~~gone~~ explored! [↑](#footnote-ref-34)
35. Many of you have wanted to do this all semester – here we go! [↑](#footnote-ref-35)
36. No, the lab didn’t end. Keep going. [↑](#footnote-ref-36)
37. Since we have some space down here, another way of thinking how spatial join works is to think of the order of file input as a little sentence. First file, geometry. Second file, appending. And doing it wrong does nothing bad except give you another file to delete. Now, where’d that screenshot go? [↑](#footnote-ref-37)
38. Those advanced stats we’ve talked about. [↑](#footnote-ref-38)
39. Pretty much any field that deals with spatial data is interested in clustering too [↑](#footnote-ref-39)
40. No, really, click this. If you don’t, Arc won’t generate a report, and you’ll never know the answer! [↑](#footnote-ref-40)
41. If it tells you it isn’t random, congratulations! You’ve randomly created a non-random distribution. The wonders of statistics. [↑](#footnote-ref-41)
42. Don’t ask [↑](#footnote-ref-42)
43. Or at least, one where we are pretending to need such a thing [↑](#footnote-ref-43)
44. Yes, you can do that by typing! [↑](#footnote-ref-44)
45. That’s a squared, not a footnote. [↑](#footnote-ref-45)
46. Use the tool help to read up on how Spline is calculated. [↑](#footnote-ref-46)
47. Kriging it also the most fun interpolation method to say. [↑](#footnote-ref-47)
48. Geographic Coordinate Systems -> USA and territories -> NAD 1927 [↑](#footnote-ref-48)
49. What tool could you use to subset the data just within Idaho? What would be the inputs? [↑](#footnote-ref-49)
50. Due to Idaho’s funky shape, this is an important step to save you some analysis headaches later! [↑](#footnote-ref-50)
51. That’s right, no map this time! Sometimes your GIS output is not a map. Fun! [↑](#footnote-ref-51)
52. You do remember how to calculate a raster layer that show distance from a vector feature from Lab 6, riiiiiiiight? [↑](#footnote-ref-52)
53. Wow, look at all these footnotes! [↑](#footnote-ref-53)
54. If you wanted to make the histograms look nicer and have identical bins, you could create them in excel, based on the extracted point values. But, for our purposes, the ArcGIS output is fine. [↑](#footnote-ref-54)