Spatial Statistics and Spatial Econometrics
Prof. Gaurav Arora
Department of Social Sciences and Humanities
Indraprastha Institute of Information Technology, Delhi

Lecture - 45
ArcGIS Session 4

Welcome everyone to another session of working with ArcGIS. In this session, we are going to pick up from where we left off. So, we were working with vector data.

We introduced ourselves to polygon-type vector data and we looked at districts, states, and taluks for India and we sort of focused down to Uttar Pradesh. We looked at the districts level data. On your screen, you can have a bit of a visual.

Let me just zoom in to layer for your convenience.

I will also remove labels, things that we have done till now.

And what we did previously, we use this intersect tool in the toolbox, the geoprocessing toolbox. Under view in geoprocessing, we used this tool under analysis overlay and intersect to provide an intersection of a line data that is the railroad's data for India, downloaded from Diva GIS and the districts level data for UP.

And consequently, we created this table which is called UP districts railroads intersect table 2 using this tool. On your screen, if we go to the catalog, you can see the location of these data, right? We talked about managing data using the arc catalog and on the arc map, we have a visual of the data and in the table, we have this intersection of the rail lines as they pass through each district. We calculated the length of these entities.

We actively added a field called length and we calculated the geometry and calculated the length of rail lines passing through each of them.

And we figured that if we focused on Saharanpur, we said ok, select by attributes and I want to select by NAME 2 and I want Saharanpur, right? So, we did all that in the previous towards the end of the previous.

And I have these 11 of 437 units selected and if we focus on, where these selections are; well, they will happen to be in the Saharanpur district.

So, this is going to be a Saharanpur district. If I open the attribute table, I say select by attributes and I say apply.

You know all the rail lines that are passing through Saharanpur are then selected as individual entities, right? So, the FID Uttar Pradesh which is 62 which refers to NAME 2, the district name being Saharanpur intersected with the FID of the railroads which provides us with individual unique railroad types, and then, from these, we calculated the length. So, if I were to sum the length of all the selected portions, the 11 selected rows.

Then, I will have the length of railroad lines in the Saharanpur district in UP, and it's pretty powerful, right? We understand now if we do data analysis, we can get these data in a comma-separated format which is the dot csv or dot Excel format.

Then, we can do a lot of things on our traditional software like Python, on Stata; if you are an economist or you have experience working with Stata or with SPSS or with R, you can take these data there and then, start manipulating them, right? So, if I could sum the length by unique districts, then I will get the exact length of railway lines in each of these you districts.

So, my next query or my next quest is really to be able to then export these data into a dot csv format which I can then take to my statistical software and start manipulating there, merging them with different data files which I may have different district level data for let us say nutrition, for some health outcomes, for some agricultural outcomes, from urban development outcomes and I might just want to merge the railroad length by the district.

So, I could construct the railroad length using the visual maps and then, export them as csv, and then, merge them separately in statistical software, right?

So, something of interest to us right now is to convert these data.

So, what you could do under geo-processing, you have this under favorites, you have this option of fine tools, right?

You may access this option from anywhere.

And in fine tools, I am just going to say export table, right? So, it is very convenient. The arc is really convenient. So, now I have this conversion tool called export table, it is going to be

under conversion tools briefcase that signs that we have if we go back this conversion tools; this toolbox, if I open, I will have a conversion export table, excel to table, table to excel.

So, I have many different functions that I can use. So, I am going to just use the find function because I guess it is very convenient. We can do it for different other things as well as we will see as we go. So, we do not need to exactly know the location of each tool, we could simply find them from the geo-processing tab ok.

So, it says ok, what does this do, let us go to the question mark. It says to export the rows of a table or table view to a table ok. So, it says table inside which is an object of ArcGIS or arc map which is typically a dot dbf file and we want to convert it to a dot csv or dot xlx extension. So, we say ok, I want to what is my input table? Well, it's the UP district railroads intersect table and my output table, I am going to now save it at a known location. So, I am going to say ok; I want to go to folders.

I want to go to my practice sessions and I want to save these data here. So, let me do it here.

So, I am going to say UP districts railroads intersect dot c s v ok. So, I am going to just provide this extension on my own, I am going to say save ok and I will then say run.

So, it runs it; it says ok export table is completed.

And it gives me a dot csv on standalone tables. But I do not want the table to be inside ArcGIS, rather I want to be able to open it on Excel, as an Excel file.

So, I will go to my windows explorer.

I will go to the desktop.

I will go to my location of the data.

And it is going to be under ok; there we go. So, UP districts railroads enter comes out as an Excel file right away.

So, if I double-click it, there we go it is an Excel file now. I can save it in any location; I can save it in different formats, you could save it as a text file, you could save it as a dot csv file, you could take it to software like SAS to R to Stata, what not, to python or whatever you are comfortable working with. So, the entire data set has now been transported into this particular

format, right? So, I have my FID Uttar Pradesh which is the original ID for Saharanpur, it is 62. So, wherever I see 62, it is all Saharanpur and here, we go the name is Saharanpur here.

And then, I have my railroad ID right here and then, I have length and we figured in the last class by cross-validating using the measure tool that these are in hundreds of kilometers.

So, I am just making sure that I have the unit right; I am going to still save as.

I am going to say, I will say convert arc converted UP district level railroads ok. So, I have my lengths. Now, I can take it and I can do ok. So, I want to just save this file differently, I can just import it on any system, and I can take it across systems. So, now, I have a way to take a data set from an arc table to an Excel table; something I am very used to.

But what about the other way around? So, what if I were to begin with an Excel file and I were to then convert it into a visual map, and then, analyze the data from there? So, let us go on to doing that as a next step. So, I will go back to my SSSE ArcGIS Project.

I am going to close this table, I am done with it. So, I am going to remove it.

And I am going to save my project. I am always going to save my project. So, I can return to it whenever I like as best practice.

Then, I will do zoom to layer. So that, things look not so bizarre and now, I am going to show you a points data set. So, the vector data occur as polygons and lines, and points. So, we have seen polygons, we have seen lines, and we have seen how they can both interact with each other. Now, we will look at points.

For that, I want you to pay attention to this India water resource information system database of the government of India. So, you can simply Google India WRIS and it will give you the first thing, it will give you is India WRIS portal.

On this portal, you can navigate to groundwater data. So, you can go to, let us say figure out for a given season, any season in any year starting from 1993 till as recent as 2022. Let us say, we say 1998 and we say ok; so, there are 4 seasons. January to March, April to June, July to September, and October to December, all these seasons are quarters in a year. But they also map your monsoon seasons and the crop calendar of India, right?

So, groundwater monitoring in India logically happens according to the monsoons and non-monsoon seasons. So, of course, July to September is the monsoon season; April to June is pre-monsoon; October to December is post-monsoon right, and January to March is something like a post-monsoon winter season. So, you can choose the season and you can download this data. You can submit a request, you can download the data as an Excel file. So, I have done that for you beforehand. So, I am going to close this window.

And I am going to show you a data set that you can get from there. So, this is groundwater levels data ok; cancel, cancel ok, something wrong?

Ok, there we go. So, here is the groundwater level data, this data is for the state of UP. Separately downloaded; ID is a what is this ID, we will see in a minute. Well, this ID belongs to a well code. So, India's groundwater management agencies have or own certain wells in different locations in the country, from where they go in and they monitor groundwater levels, right?

So, they measure the level beneath the ground surface; how far down the groundwater is; where is the water head; if the water head falls below that means, depletion and that means, alarm, and if it so alarms, if it depletes quite a bit let us say within a season.

If there is a dry season, if there is less rainfall, well you are going to have some problems right? These are the data that I was showing you through the lecture series as well right? So, you can now use these data go back, and revisit all the different things that we talked about these data, all of those functionalities that we talked we can do all of those on ArcGIS, right?

We would not have the time to do that through this particular module. But once you are able to use these data, play with these data, analyze, and manipulate these data on ArcGIS, you can do all those functions on your own. So, what do we have? We have these well codes, against each well, I have a latitude and longitude.

So, I have my x y coordinate exact location on a map, where that well is located and the data is in the public domain. Then, I have the year which is 1998. So, this is the data for 1998, and it is post-monsoon kharif; kharif is the season of cropping which happens during monsoons.

So, it is going to be the July, August, September, and post-monsoon kharif, then would probably mean either those three months or October, November, or December. So, it is one of

those months which is after monsoon probably, just after the kharif season, and what you see as values here are meters below ground level.

So, at this particular well in the Baghpat district at this location in 1998, the water head was 8.8 meters below the ground surface. So, this is now powerful data. I am used to working with these data. You know it is a cross-sectional data set, for all, it is worth it as applied statisticians or applied econometricians or econometricians, we know that this is a cross-sectional data set.

Now, how do we think of this as a spatial data set? Well, there is a spatial delineation. I have the lat long, the question is how do I now exploit this spatiality of the data, that is where ArcGIS will come into the picture. So, we are going to go back to our software. We are going to the soft folders. So, I have my folder that I had added previously.

And in that folder, I have this point vector data and I have my groundwater xlx. So, I am going to just drag it, I am going to just double click it and hope that this will open in my table; it seems it will not. So, I am going to do one thing; I am going to go into the map.

I am going to click on add data, I am going to go to my folders ok.

So, vector, then I have my xlx file ok.

So, this says groundwater level 1988; it is a typo, it should be 998 and this is a dollar sign. So, this should be the sheet name. So, let us go back to our Excel file and see for sure if we have that name. So, we have groundwater level. So, it does not say that. So, let us see ok one second; I have a little snag here. So, the groundwater level ok and the groundwater level ok.

So, I will go back and I will just try and import the sheet.

Here we go. Let me open this data set; groundwater level 1988.

It should be 1998. So, by the way, I can simply correct the typo ok; I am done. Now, I can open it so that everything remains correct; perfect. So, I have my ID; I have my state, I have my district, my well code, and my lat long and year 1998 post monsoon kharif 8.8.

So, here is my Excel sheet and I have added this data using this tool called add data ok. I was not able to drag this data directly from here; just like I was able to do previously remember like if you recall from previous sessions when we had a shape file, we could simply drag it from the catalog to the map window and it would simply just be added to my contents pane and I could visualize the data. Here that did not work.

So, I had to manually add data from here; I could click here.

You know it will take me to that folder, where I have kept the data.

And then, I can from there, I can simply import this data; I can import the xlx file that is the Excel format or the csv file, it does not matter.

So, the quest that I have is I want to now you know I will remove all the maps and the quest is that I want to now visualize these data spatially.

So, for that, we go to the data set, we click, we say display x y data; we click on display x y data.

So, it gives me a new window, it says the input table is groundwater level 1998; perfect. Output feature class, is asking me where to put a feature class.

I am going to give it a location, I am going to say ok give it to me right here. So, I want groundwater level 1998. So, I am saying Excel to shape or feature right dot shp because that is the format I like.

And then, x field and y field; so, x is longitude and y is latitude. You can imagine; I mean latitudes are vertical lines on the Earth's sphere. So, if the latitudes are changing, we are going along the y-axis, and for longitudes when they change, we are going along the x-axis. So, y is latitude; z, we do not have a z field, why? because we do not have a height, we do not. Our data is 2D, it is not 3D, right?

So, with regards to the location spatial domain, our data is 2D, we have two coordinates right; we do not have three coordinates ok. Very very important, coordinate system is W G S 1984 which is very critical because remember all my vector files are W G S 1984.

So, if I am to overlay them on a map and make sure that if I see a well location at a given sort of with some reference of a district, it is indeed in that district; I need the same projection

system. So, that is great. It has already provided me with the projection system that I like. So, I am going to say ok.

There we go. So, here on this screen, on your screen what you see is a groundwater level data set for 1998 as a shape file, right?

I can now right-click and open the attribute table. I will the same attribute table as earlier opens, right? There are 1309 units in this attribute table; that means, there are 1309 unique wells distributed across the Uttar Pradesh state, right?

And not only the shape resembles UP, but if I now, check UP state vector file, see what happens is that I am able to find the envelope of these wells. Now, this is phenomenal. I can go from an attribute table in the arc to an Excel file and I can start from an Excel file and come to the arc and do all the manipulations I like.

So, what else can I do ok? So, one thing is I can probably look at the properties of these data, well let us look at the spatial reference. Well, WGS 1984, I specified the gcs, right? So, it is the same that I like the angular unit in degrees and so on, right? So, there are all these details. Now, I can also view which district has which data.

So, I can go and check the district's file.

Here, we go. We can see that some districts have a much higher density of monitoring relative to some northern districts; maybe these are hilly regions or remote areas. You know it is not possible to put wells or monitor them, it may be very costly and hence, you may have a lower density of wells in those regions. It could be a research question to ask why some districts are sampled more aggressively than others; I mean you might want to figure out that is a good place to start in terms of analyzing the data.

Very interestingly, the western districts which are all sort of that are counted as NCR. So, if I were to label these data right.

These districts of Ghaziabad the district called Gautam Budh Nagar is Noida, are part of the national capital territory, Baghpat, and often we come across popular press about the depletion problem in these districts, and yet what we see is that they are not so densely monitored, there are let us say three wells in Gautam Budh Nagar and three wells in Ghaziabad right.

By contrast, Meerut has many more wells, Muzaffarpur Nagar has many more wells and so, does Bijnor and Harpur and Bulandshahr, and so on, and for that matter, Mathura, right? So, these are some interesting things, these are starting points in a research journey, right?

Now, I might want to figure out, if can I actually visualize the water levels across districts. Well, let us try. Let us remove the labels because we do not want to clutter. We go back to our point data and we say symbology.

Now, under symbology, I will say what I want to look at in proportional symbols.

I want to look at proportional symbols and I want to look at the field which is post-monsoon kharif. Well, I do not see it here. So, ArcGIS is not reading it. So, for that, I will go and I will now open my attribute table.

I will add a new field, I will call the field groundwater 1998 meters or let us say met or just m and I like to have these things ok.

So, I am done with that I would like to have the data type double and I will just say I am done. So, it will say do you want to save this change, I will say yes, I want to save this change, perfect.

So, I have my new field groundwater 1998 meters, I am going to right-click, I am going to say calculate field; not calculate geometry, calculate field ok.

I go there I say groundwater 1998 meter, I want this to be equal to post-monsoon.

And I will say apply, it will give me, it will do, it has done calculations right. So, I can see the values are similar; but it has given me a lot of warnings and if you pay attention the warnings are where the post-monsoon level was NA and it does not know what NA is. NA is a text string. So, it converts it to zeros ok. So, this is a caveat. So, you have to be very careful about this caveat right ok. So, apply ok.

So, we have a new stream now. Let us again go back to symbology, there is.

And catching the field groundwater 1998 meter, I am going to use it. Wonderfully, it gives me a visual understanding of groundwater status in 1998 across districts.

Well, it does it in four classes or four or let us say five classes which go from 1 meter to 10 meters and above right? So, these are just less than or equal to 1 meter, 1 to 2.5, 2.5 to 5, 5 to 7.5 greater than 5, 10. So, a larger circle resembles a worsening groundwater depletion problem. So, larger circles are a reflection of the greater problem.

So, I can play with the template. So, you know the minimum size, I do not want the sizes to be so large. So, I am going to say the minimum size is 1, and the maximum I want the size is 40 ok.

Now, what that does is makes my graph look much clearer, and my visualization becomes clearer. So, I am done with symbology; but something I learned from this symbology is that I have some depletion problems in 1998 in the bordering districts of NCR, not so much; maybe on the Southern side and the depletion problem was certainly higher in the Southern UP belt that is South of the Ganges river than in the North. So, I see no real depletion problem happening in the North; whereas, the South was already stressed so far as groundwater resource was concerned and this is 1998 right?

So, I am talking about 25 years ago. So, I have visualized, I have done some symbology, and I visualized this data now, I may have a query that I want to learn about the average groundwater level inside a particular district or across all districts. So, what do I do? Well, I transport my knowledge from how I intersected the line feature with the polygon feature, I am going to do the same; I am going to go to geoprocessing.

I will say show me geoprocessing, go to an analysis overlay, and intersect. What are my input features?

Well, my first input feature is UP districts; my second input feature is the groundwater levels. I am going to rank 1 and 2, just like I did previously; I am going to give the address of my output feature class and I am going to say UP districts groundwater 1998 post monsoon; so, pm intersect dot shape.

I want to join all attributes, I want the output to be the same as the input right? I actually wanted to be a polygon. So, let us see if it does it. It would not do it because it is only giving me two options here. But if I were to change ranks, I am going to make it 2, make it 1, and let us see if it allows me to save these things as a polygon; it does not. So, I am just keeping my

ranks, my districts are my primary query, I want to understand the groundwater situation within a district.

So, I want to take an average or I might want to know the standard deviation of groundwater levels in a district; I might want to know the median level in a district. So, I basically want to sort of do some kind of a spatial joint between the district and groundwater file.

I know that wells are also identified by the district. So, I could directly do it in the Excel sheet; but I could also do it if I did not have that identifier in the groundwater data, I could still use the lat long information and then, join it to the districts. You could do it to Taluks, right?

So, the original groundwater file does not identify which well falls in which taluk. Well, by intersecting the data, you could join the taluks to the groundwater wells, individual groundwater wells, you should do that as homework, right?

So, I am just going to say run; it is going to run it.

And it has run it; fantastic. So, I am going to open the attribute table and it is going to now show me for Sonbhadra, right?

I going to say select by attributes, select a field I want NAME 2 and now, I will look at, let me look at something like a Gautam Budh Nagar ok.

So, I have these three wells in Gautam Budh Nagar; fantastic, I can see that they are now selected.

So, I am going to now only work with these and I am going to see where are they selected on my data. There you go. Alright, alright, alright, alright; Gorakhpur, Amethi, Raebareli, Amethi now Azamgarh, Ballia ok, Ghazipur; yes ok.

For some reason, I am not able to ok. Wait for a second, let me see if I can ok. So, for some reason, I am not able to find the selections in this.

I am just thinking about whether they may be in another attribute table. So, whatever, but the thing is I could just take an average of the levels at these three wells; I could calculate the

standard deviation from these three well. Of course, it is a small sample matric, but it is what it is. So, now, just like earlier, I could export these data from this table to Excel. So, for that, you can fall back to what we did with the railroads data just 5 minutes ago, right?

We could take these files and then, merge them with some other files, auxiliary Excel files. We could do these things at the state level, we could do these things at taluk levels and if you have a village shape file, you could do them at village shape levels. So, you understand how we can sort of manipulate the data accordingly.

So, that is about it. We have seen vector data of three types; we have seen how to manipulate them, how to store those data I mean and how to visualize those data, use symbologies to create informative maps and publication-ready maps.

And finally, now is the time to look at some raster data, right? So, something another data type that we have worked with and see if we can now sort of find ways to not only visualize raster data, manipulate them, but also sort of read them into polygons like districts or something around an outer ring of a well, I could just draw an outer ring of a well and read those data in the districts. And then, export tables into Excel and do some manipulation, some analysis.

So, thank you very much for your attention. That is about it for session 4. We will do a short session on raster data shortly later. See you then.

Thank you.