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

Lecture - 43
ArcGIS Session 2

Welcome back to the 2nd session of learning or working with ArcGIS. So, here in this session, we will work, we will study or will learn to work with vector data. What are vector data? Well, we know that there are two major data structures under which information is stored, the spatial information is stored – one is the vector data, and the second is the raster data.

Raster data are a bunch of pixels continuously distributed in a given datum or a given domain of interest, whereas, vector data are stored as either points, lines, or polygons. Vector data are a discrete form of data wherein you may have on a particular point in space, you may have a measurable value, but then in its neighborhood, one delta ball around may have no value, right?

And, then a little further you may have another area with some value and without value, and so on. So, no data and yes data are sort of discretely distributed in space as polygons, lines, and points. That is it, right?

So, I am going to on your screen what you see here is a source of vector data which we have introduced in our lectures as well called the DIVA-GIS. So, I have just Googled DIVA-GIS and it gives me this first link as www.DIVAS-GIS.org

When I click on this when, I click on DIVA-GIS free, simple, and effective, this particular page opens up. So, now, it has a lot of information about the data. You should always read the information of the data that you are using, very very important.

So, I am going to go into this without sort of spending time on other aspects which you should definitely look at whenever you are writing papers or publishing reports for academic journals or professional you know business-related or any other professional report which is going to be in the public domain or is going to be reviewed by some of your peers, you must document the details about spatial data. It is incredibly important because the data come with its metadata, right?

So, its definition is embedded in the metadata. So, if you do not read this documentation you are going to misrepresent it. You are very likely to misrepresent your analysis as well, right?

So, I am just going to click on free spatial data. Now, it gives me data on a country level, global level, global climate data, species occurrence data, and crop genebank collection data. Near global 90-meter resolution elevation data: now, that world hill shade that we saw on the map on the ArcGIS portal in the previous session is an elevation map. So, you can have one downloaded from alternate sources.

So, I am going to go to a very good list of data sources from the Eden Project. So, just look up the Eden project, these are very interesting starting points honestly. One of the basic queries of people who want to work with spatial data is where will I get these data from? Well, it turns out that there are now incredible amounts of resources as far as the spatial data are concerned.

Such queries are now lesser and lesser, but when I teach students for first anxiety will I be able to access this data or will I get these data? Well, yes, you will be able to access these data. A simple Google search would lead you to incredible amounts of spatial data of both vector and raster kinds, right? It may not be stored in your traditional dot CSV formats that you are very used to, but they are very accessible in their own right.

So, let us look at the country-level data. So, I am going to click on download country-level data for any country in the world. It will give me administrative boundaries, roads, railroads, altitude, land cover, and population density fantastic. So, I am going to go for country-level data.

I am interested in it. So, now, again there is a lot of information. All this information must be, you should make a note of it. Now, look at the administrative boundaries, the format is a vector; inland water – vector, line, and area; roads – vector, it is a line vector. So, you know. So, it is a line format elevation is a grid format, and so on.

Now, I am going to go to download the Indian data. So, I am going to go to India. Now, I can download many types of data. I can download Administrative areas, Inland water, Roads, Railroads, Elevation, and so on.

First, I will just say Download Administrative Areas, and it says Download and I say OK. So, here now I download a zip file. It says IND_adm.zip. I can go Back.

And, I can say ok I also want Roads data. So, OK and I say Download. So, it says ok.

Now, I am downloading the IND underscore roads dot zip file. So, you will get these files as zip files and then you will have to unzip them. So, let us do it for one of them. So, I am going to say Show in a folder.

And, I am going to say, Copy X, I am going to take it to where I am used to.

I am going to my practice sessions and I am going to say alright India adm right click Extract All and ok, you can say Extract, alright.

So, I have all these files that are now available in India adm.

This is the extracted version, this was the zipped version on your screen – IND adm is the zipped version and there is this extracted version.

So, the extracted version now has many different files.

It also has a license which is a text document you must open, and you must read it, because it will tell you that, these data can be used for what purpose, it is not allowed to redistribute these data, and so on. So, I am not going to provide you with this data anywhere you have to download it yourself. I am not going to distribute it.

And, it has been extracted from this source called the GADM database. So, I am not making these data ok. These data are publicly available on the GADM database. And, this is not the only source for India's administrative data. You can get administrative data from Indian government sources as well, right? So, it is up to you how you search the data where you download them, it is all up to the analyst.

Here I am only providing you an example, a popular source of data for different countries let us say you are doing inter different country-level analyses. So, you need these you might use these data, and so on. So, far as India's data is concerned India has its own, India would have its database.

If it is some other country like the United States or China, or Afghanistan, they will have their own sources and it depends on how accessible those data are, whether are they freely downloadable, you know what kind of request process you have to go through to get to those data – all of those things are subject matters very contextual to countries and all that, their data distribution, privacy properties and so on.

But, the point is that it is not allowed to redistribute or use them for commercial purposes without prior consent. So, this is something to keep in mind while you are using this data. So, now, let us look at the data by itself. So, we have IND underscore adm0s. So, I have dot cpg dot csv dot dbf and dot prj and dot shp and dot shx.

So, there are these six files that come under India adm0. And, then similarly there will be six other files that will come under ad India adm1, 2, and 3. So, these are all vector data. If you have heard of vector data earlier you would have heard of this name called the shape file.

So, one of the extensions is dot shp whether it is India adm0, India adm1, 2, or 3 there is always a dot shp file this is a shape file. Any vector format is always shared as or all or is mostly usually stored as a shape file. When we talk of shape files we do not just talk about one file. We talk about a bundle of files that come together.

So, shape files or vector data always come as bundles. What do these bundles comprise? Well, you know for starters apart from shp, I have a file called dot prj. Now, you can take guess what dot prj would constitute. It will constitute the geographic projection system, the geographic coordinate system. What is a projection system? Is it GCS 19, WGS 1984 is it Alber's equal you know? So, there are various projection systems from which this data might be projected.

This file is absolutely mandatory for me to project it onto ArcGIS. Can we construct these files on our own? Well, all of these are matters of training, right? But, the point is that these data do not come just as a dot in one file. shp is; obviously, the file that is going to be used to visualize the data onto ArcGIS, we will see that in a minute.

Then we also have these files called dbf and dot csv; dot csv is something we are used to, dot csv looks like a comma-separated version. So, it is something like an Excel sheet where the information is stored in a gridded format. Now, how does that get integrated into a shape file which is a visual vector file that we will see in some time when we visualize these data?

So, we have India adm0, 1, 2, 3. How are they different? Well, the best way is to look at them in the software itself.

So, you will see that under this I have already downloaded and stored these data under India adm. So, I have India adm roads, railroads, and inland waterways. So, I have downloaded four types just for education. These data are not constructed by me. They are coming from a source. I have shown you the source, right?

So, if they have any kind of a mismatch with the official Indian government data that is not my doing. It is just a source it is just an easily accessible source that I am using for purely instructional purposes here.

So, here let us go and look at the data on ArcGIS. So, I have now opened the SSSE ArcGIS project that I had saved in the previous module, right? So, I saved this in the previous session in session 1. So, in session 2, I am starting where I had left. So, I have my Contents pane, I have my Map pane, and I have my Catalog pane.

Under Catalog and Folders, I have this Gaurav ArcGIS and under practice sessions now I have my project. I have also my India admin data, but where did the other data go? So, you know let me go back to the Windows Explorer yeah, ok. So, under Practice Sessions I have now you know one second I have India admin data, but I also have these data these two zip files India adm, India rds and India adm which is the extracted version.

The question is why do not they show up here? Well, they do not show up here because I did not Refresh this.

So, let me just left-click on this one under Practice Sessions and I am saying to say Refresh.

So, it shows me India adm now. It does not show me the zip files. So, the arc catalog does not catch or does not read the zip files, but it reads the extracted files. So, now, I have three types of one is the SSSE ArcGIS project, India adm, and India admin data. Now, I can also by the way delete these data right from here.

So, I can say Delete. Do you want to delete this item? I am going to say, Yes.

If I do that and if I go to my Explorer and I do refresh, the India adm file is gone from here. I can again extract India roads. I can say Extract All, Extract.

And, if I do again I go to Practice Sessions I say Refresh, it brings it here. So, I am navigating folders from within ArcGIS, I am just trying to show you that.

Now, let us go to India adm again.

So, here are my files India adm, here are my files. So, I see a lot of these you know India shp, shx, and so on. There are some of these desktop files which you can ignore, but you know now the packet is much larger, but the idea is that we have all these files still.

Now, under this project, if I go to India adm I see only four files India adm0 shp, India adm1 shp, India adm2 shp, and India adm3 shp. So, what happens is within the arc catalog, it only is reading the shp file, but in the original folder, I have a packet of files not just one file, right? So, it is very important to note this.

Now, if I move to India adm0 let us say, I say Copy and I say can you take it to India adm roads? Ok, I can say Paste. Now, here, what it does? it is still working, but it actually copies the India adm0 for Ind adm folder to Ind underscore roads folder. Now, let us look at what happens to the physical folder. Now, this is the ArcGIS catalog, let us look at the physical folder.

I am going to go to Practice Sessions and India Roads and it actually brought all the files related to India adm0. So, when I move files from within the arc catalog it is moving the bundle automatically. This is very very good for me. Remember, when I am working when each layer when each image layer is going to have a packet of 6 to 7 files and I am moving them around, then it may be very problematic for me if I were to do them physically. I can make a mistake while doing them manually.

So, it is always advised to move files from within the arc catalog because you cannot make a mistake when you do that ok, alright. So, now, we have learned that the utility of arc catalog is that we can move files as full packets from one location to another. In physical folders, I will have to manually select, and make sure nothing is left out. If anything is left out the whole packet becomes useless because it may not then project these things, right.

If you do not bring the projection, it does not know how to visualize where to put the coordinates; if you do not bring the csv, it will not have the data that is embedded, right? So,

the best practice is to always move spatial data from the arc catalog and not physically right. So, that is knowledge stack number 1.

So, now, that I have done that let me work with India adm data. So, I am going to now start to work with India adm data. So, let us say, I have India adm0. So, now, what is the difference between 0, 1, 2, and 3? Well, I have to actually look at the data and then make sense of it. So, I am going to simply click and drag the data onto the Map.

And, here we go. So, it is a data set that has a bit of a polygon sitting inside of it, right? So, it should not have been like that, but you know this is the data set that has come from DIVA-GIS. So, I can write, I want to rename it and I am going to say rename it to DIVA-GIS. So, it has come from a source. So, it will help me keep track of let us say if I am getting my data from different sources, I can say ok I am going to say, alright.

So, rename, I am just for everyone, I am just going to always keep adding sources just as best practice, a recommended best practice is to always add sources to the data in their names itself, it is very helpful when you are writing papers that are based on multiple sources of data and so on. We can move from 1 to 0 to 1.

What is the difference between 0 and 1? So, now, it seems that the 1 is also providing me with states. So, the first polygon was an India-level polygon; the second polygon is a state-level polygon. So, now, you have India adm, and let us say you have adm1, then you have India adm2.

Now, India adm2 has even smaller polygons. What are those, right? How do I know what these polygons are? Well, let me go and look at India adm3, and what that has, and then we will come back to this question. Let me project India adm3.

Wow, it has even further small polygons. And, I guess if you are from India, you know you will start to make even if you are from any other country in the world, you will start to make sense of what these things, what these smaller entities are, they are basically smaller administrative units. For example, in India we have states as administrative units, within states, we have districts as administrative units, within districts, we have taluks or tehsils as administrative units.

So, going from 0 national level map to the second layer state level map to district level map and tehsil level map. How do I know that? Well, on the left-hand side in the contents pane what has happened in the process of looking at these things was that I added layers in addition to World Topographic Map and World Hillshade.

If you were not working with either of them, that is no problem. if you are working with an older version that did not have World Hillshade or World Topographic it is not a problem you will see the map like this on your screen.

You will see it like this on your screen where you will not have a base map that is fine. It does not change anything for me.

What is important is that each layer is somehow sitting on top of the other, right? Each layer is sitting on top of the other.

And, if I right-click, let us say, I right-click on India adm1, let us work with the state-level layer. Now, if I right-click on the state-level layer, I can see various things, I can see the attribute table which is going to be very important for me and I see properties.

So, I will always start with properties I like to first look at my data, look at its properties. It directly throws me to this tab called source under properties and it tells me that this is the shape file feature class. So, it is a shape file, we know it is polygon data. So, it is a shape file and it is a feature class. So, shape files are often referred to as features.

So, polygons, lines, and points are often referred to as feature vector data. Instead of saying vector explicitly if it says feature data, features are often referring to vector data in ArcGIS terminology, geometry type is polygon we know that the shape file shows me the exact location of it.

So, if you see some data open and you do not know where it is located you can get to the location from here from the properties.

Then you have your spatial reference WGS 1994, very good. So, I can actually know from my previous session or I can just go back and look at World Topographic data, and I can just say Properties, under Spatial Reference, I have WGS 1994. So, that is why I can overlay the state-level file onto the World Topographic map and start comparing regions, ok. Let me try to do that.

So, let me do this now the trouble is because of the color scheme which is called a symbology. So, another thing here when I write click on to India adm1, you know a very interesting thing that I see is Symbology.

If I go to Symbology, I can actually work and get a different color let us say I get a black outline. In this black outline, I do not see the solid color blue, I only see the boundaries, right? So, what I have here is let us say data on Rajasthan; here is a state in Rajasthan called Jodhpur and Jaipur. It makes sense, you know these are two cities in Rajasthan, then we have New Delhi it is there where it is supposed to be.

We have Mumbai, Pune, Nasik, you know they are where they are supposed to be, and we have Uttar Pradesh, and so on. So, now, what tells me that this is Uttar Pradesh, this is Rajasthan, this is Gujarat, this is Maharashtra? Well, because of my existing knowledge, but if I am working for different countries of data I may not be aware of the states of different countries.

So, where do I make sure that I am on the right track that this is indeed going to be West Bengal and this is indeed going to be Jharkhand and this is indeed going to be Bihar and here I am looking at Bhopal and Indore which is in Madhya Pradesh?

So, how do I really make sure? So, for that when we looked at the bundle packet of the shape file, it had this file called dot csv, that csv file stores such information in what is called the attribute table. So, I am going to right-click and then select the attribute table open table, and what it does is it opens attribute tables.

Now, here I have an FID which is a field id it says shape which is a polygon and it has an ID number, ISO name, and it has NAME 0, ID 1, NAME 1 and there is other information here.

It has latitude longitude which is the center of mass, but the point is I have a data set in terms of India adm1 where every entity, sub-entity is a polygon. There are 37 such polygons. On the lower bottom bar, it says 0 to 37 selected which tells me there are 37 data points.

So, there are going to be 37 unique polygons here.

And, as soon as I click on these polygons, I get what is also stored in the attribute table. So, if I click on let us say Madhya Pradesh it says FID 18. So, if let me go to FID 18, and then India

NAME 1 is Madhya Pradesh, NAME 0 is India. I am now going to my attribute table and going to FID 18. FID 18 here we go.

It is a polygon perfect it is now selected. So, if I click on the row it selects the polygon for me I have again ISO name India, NAME 0 is India, and NAME 1 is Madhya Pradesh. So, I have a cross-validation of what my attribute table is showing. So, similarly, I can go and I can select Uttar Pradesh. So, on the map, if I go up, I look at Uttar Pradesh it is going to be FID 34. So, let me go to FID 34.

Here we go FID 34 and it should select Uttar Pradesh.

If I want to select two different states at once, I can press the control key, I want to just look at 18 which is Madhya Pradesh and I can select these two and now you have both of them selected at once ok.

So, this way you can navigate through the data, because under Map because Explore tab is selected right now. You know it is selected. Click it gives me an information. It gives me a popup. It gives me all the information extracted from this attribute table itself. So, the dot csv file is equally very very important.

Now, if I want to remove selections, I can on this, it is a very interesting one. I can just click on clear, right? I can also Delete them but do not do that because it is permanently deleted. So, you will have to then again download the data. So, you have to be very careful with these functions.

So, here I can actually go on to this top left corner, and in this white box if I click it de-selects everything. So, it does the function, in the older versions you will not have these clear switches, and all these buttons, but you can use you can just explore and conduct all those functions even without explicit tabs for them.

So, I can also close the attribute table, right? So, I have understood that I have the States here, let me go to the next shape. I go to the next shape. Now, I am in this shape, I do not know it seems to me that these will be probably districts. Because of the symbology again, I cannot visualize anything. I am going to double-click on the box here.

Earlier I right-clicked and I went to symbology. This time I am not going to do that. I am going to take an alternate route and I am going to click on this box with the color with the

pinkish color that you know India adm2 has. I am going to click it is going to directly take me to symbology, fantastic, right?

So, I am going to say ok, I can do many things here, right? I can vary symbology by attribute, I can allow symbology property connections, and so on. I can go to properties and I can play with a lot of things here.

So, under properties, I am going to now say, I need no color. I do not need a fill. With the outline I am going to change the color I am going to make it, I am going to make that colorful, and I am going to say Apply and that is what we get.

So, now, I have my states and I have what is inside them?

Let us go back to India adm2 and open the Attribute Table and see what we have. We have districts. So, after my states, I have India, I have States, right?

And, now, apart from that I also have districts. So, I now know that I have another file which is districts. So, I have India adm1, adm0 is just national. I am going to rename it and make sure that. So, instead of adm0, I am going to call it national; now adm1, I am going to right-click and Rename.

And, I am going to call it states. adm2 now I have just learned that this is districts. I am going to just go and call it districts.

And, now what remains is adm3, so, to understand what adm3 is, I am going to go back and open adm3.

So, once I do that I can see that I have, you know I can go in and I can say, I want to what if I want to zoom in if I just scroll my mouse, I can zoom in, I can zoom in, zoom in, zoom in.

So, let us say I am focusing on this district. If I click on it I will get some information about it.

I click on it and select India adm2 because that is by list order which is the topmost layer that has been checked. So, it is clicking India adm2 if it's giving me FID 594, the NAME 0 which is the nation is India that is the country name, then the state name is Uttar Pradesh, and then finally, NAME 2 which is the district is Hardoi.

So, for now, I am going to focus on Hardoi, right? I am going to focus on Hardoi keep your eye on Hardoi while let me just select Hardoi.

So, let me go to India adm2. I will just open my Attribute Table. I will search for FID 594; FID 594 because that is where I know Hardoi is 594 here you are in India Uttar Pradesh district.

I must make sure that I have. So, let me try there we go I see Hardoi here NAME 2, right? So here we go and it selects Hardoi. So, I am going to uncheck this.

Now, when I check India adm3, it gives me even smaller boundaries within Hardoi. So, what is this higher-resolution administrative unit? I am going to just right-click.

I am going to go to its Attribute Table.

And, what I am going to learn is that I now look at taluk. Now, I am looking at taluks within Hardoi.

So, if I click here, I am going to get India, Uttar Pradesh, Hardoi, and Hardoi.

Click here I am going to get India, Uttar Pradesh, Hardoi, and Bilgram. So, now, I have this district; this subdistrict is Bilgram and this subdistrict is Hardoi.

What about this one here on the eastern corner? It is called Sandila, right? So, now I know that I am working with taluks and they have different names. How do I make sure that I can visualize them one over the other? Well, I will have to go to Symbology. I am going to just close the Map for a minute. I have closed the Map. Now, I am again going to Symbology double-clicking. Click to modify symbology, alright.

Let us do that. It takes me directly to properties. Under galleries, I can select some pre-defined symbologies. Under properties, I can create my own. So, I do not need a fill. So, I am going to say No color. So, instead of the boundary or the outline I am going to use a color that resembles the fill. So, let me do that.

I am just going to try and try my best to replicate it, but I am not going to try too much. So, I am going to just make it a little thicker, right?

And, I am going to say Apply. I am going to say ok. When I apply it is still applying. It is working, working, working ok. it is worked, alright.

So, now what happens is that I only see taluks, what happened to my districts and my states? Well, let us go back to the Symbology for the districts and make them a bit thicker. I am going to make it let us say thicker up to 3 points and then say Apply.

And, now I can see my districts somewhat right. Let me change the color to a darker color. Apply and Close.

So, now, very interestingly I need to deselect everything. I am going to go back to the Attribute Table of India adm3.

And, I am going to say it re-select or clear selection, I do not want this selection. Now, it is district-level data. So, Attribute Table, please clear the selection, thank you, alright.

So, I cannot see states. I am going to go to the state symbology again and I am going to say hey I want to also view states. So, I am going to make its thickness the most.

So, here we go, alright. So, now, in black whenever I see these black hashes I know that those are state boundaries.

Wherever I see the red hashes I can say that those are my district boundaries and the blues obviously, reflect the Taluk boundaries. So, I need to change the name. So, I need to go to my, I cannot see my catalog pane. I have lost my catalog pane. I am going to now recreate it for a minute just a second. I am going to say Insert or I am going to go to View, I am going to click on Catalog pane.

And, now here I am going to rename India adm3 to say taluks. So, now, I have renamed the data in my folder. I am going to try and see whether that has happened. I am going to say Refresh, now I am going to the physical folder. So, let us see where are you, India adm data. Practice Sessions, India adm data, India adm, fantastic.

Now, look at this, wonderful. We have India districts. So, India adm2 has changed to India districts. India adm0 has changed to India national, all of them in the packet at once. So, that is how powerful the arc catalog is. So, arc catalog is a database management software within the ArcGIS package, and it is very efficient that you can see in front of your eyes.

If you were to change this data on your own, you would have to go and change it to every physical file separately, make sure the spelling is the same, you do not use capital letters somewhere, small letters somewhere. So, it is very complex if you are going to physically manage data. It is not possible and remember we are just working with five files; on a particular project, one is working with at least 20 files, 30 files at once, right?

So, you cannot possibly be thinking of working with these data using physical sort of you know processes, alright? Let us go back. So, now, again let us say this is India. So, I know all the state names and all that and you know, but let us say if this was not India how would I know? So, for that, I have this Label feature, right? So, I have this Label feature. I am going to go to India adm2 adm one where you have States and I am going to say Label Properties.

When I click on this, the right-hand side pane gives me Label Class.

Symbol, Position, Appearance, and so on.

So, I am going to go to Class.

And, I am going to say, I want NAME 1 feature double click NAME 1. So, let me just delete everything on this expression and I am going to double click on NAME 1 which is what I know the state name is and I am going to say Apply.

And, under Symbol I can then go and check the Font name, I want Times New Roman. I am a fan of Times New Roman. I want 10 points is too small for me; I want 14 points let us say, I am going say apply. I can also check and figure out the color of the font and so on, right?

So, after I am done, I can close this pane I can go back to India adm1 and click on Label. As soon as I do that it wonderfully puts the names of the shape onto each state and because of the way I have structured my symbology. So, what I have done is I have checked the state file the district file, and the taluk file I have changed their symbology to make sure that I can visualize all three at once, right?

And, then what I am doing is that I can improve the symbology a little bit just give me a second here. So, I am going to try and Apply. So, this is too big Apply.

And, maybe I will make it a little lighter. So, I know that maybe a color like yellow might just, ok.

So, now, it is too much ok, apply ok.

So, you can play around with Symbology, but you know ultimately the idea is to sort of figure out a way to visualize all the different details. So, interestingly I have different States, within Uttar Pradesh, I can you know which is demarcated by this black boundary. Within that the red or the crimson boundaries are the district boundaries; within that, the green boundaries, which can be a little darker.

Let us try a little darker. Now, these are my taluk boundaries.

So, here is a symbology example a labeling feature example, where I can visualize different parts of India and I can create a very nice map which is sort of you can take this map, clip this map, download this map directly from ArcGIS and take it to publication.

It identifies different features within ArcGIS, we are playing with symbology, we are playing with labeling features and we are trying to create a visual picture of multiple layers put together one over the other, where polygons of different sizes are visualized at once their location, their shapes, their sizes their respective locations, and so on. One last tool I want to just talk about before we move on is this tool called a measure.

So, let us say if I want if I go on to this area where I am at the border between Uttar Pradesh and Madhya Pradesh. I can go to this, you know use this tool called measure, and I can go and look at the distance from a particular district let us say the neighborhood district.

Here of your taluk of Madhya Pradesh its lowermost point the uppermost point of this taluk of you know Uttar Pradesh. It is 107 kilometers. So, any two points in space that I am interested in or any path that I want to traverse, I can create a measure of them, right?

So, the measure tool is another thing that I am just mentioning, I am just showing you a closing sort of tool here that you can use to quickly get measures of distances between units on a map this will become useful more and more useful in future exercises.

So, thank you, this is it for session 2. In the next session, we will start manipulating vector data. Right now we have just visualized these things. So, see you in the next session for ArcGIS in learning vector data.

Thank you.