

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

**Lecture - 46**  
**ArcGIS Session 5**

Welcome to the last session of ArcGIS, in this series of lectures on Spatial Statistics and Spatial Econometrics. In this session, we are going to look at raster data, how to visualize such raster data, how to look at its property, and what to be careful about, what can we do to manipulate these data. And then finally, our quest of taking this stuff to an Excel sheet, to statistical software, and maybe on R and analyzing this data from here.

After this session, you will go into learning hands-on exercises on R. So, it is kind of a natural progression, different sophisticated techniques like building a variogram model, estimating a variogram model, and spatial regressions, they will all be taught on R because there are very nice routines, packages which can be used.

ArcGIS on the other hand is a very nice visualizer, manipulator, and data management tool. All of those things can be done here as well by the way, you could estimate a variogram model in ArcGIS, but R happens to be much more powerful, natural software where you can do many other things.

So, what we are doing here, is we are trying to learn how to visualize, you get a hands-on feeling of these data to sort of zoom in and look at what is happening and do some cross-validation and things like that on ArcGIS, while mostly analysis. So, data management, and manipulation maybe happens here, you want to create pictures, and visual maps for your research, papers, and articles or you could do it all on here.

But when it comes to analysis perhaps it is best to take the data to a statistical software. So, that is the journey that we are completing here. So, in that spirit let us go on to look at the raster data.

So, an example of raster data as we have looked at in your classroom is the Bhuvan ISRO data. So, how do we get there? If we simply Google Bhuvan land use land cover data something that is out there. So, the first thing that comes out is these data are these data.

So, we have seen these data like you can select whatever theme you want and then there are these statistics you can look at, there is an entire 15 to 20-minute module in your lecture series dedicated to the Bhuvan portal. But ultimately, if you want to get these, you can download these data, there is a form you have to fill out and all that.

So, like here, I am just showing one type of land use product, you know under web services you can figure out how to bring this data to your computer. I have done that for you here, but you will have to then download these data to be able to work with them. So, say we have downloaded this data for an area of interest of our choice, right?

So, I am going to show you such a data set that I acquired from ISRO. So, in my folder again, you know these data are to be found here. So, I have lulc, land use land cover 250k, that is the map, the scale, 2056 and then it is a tif file. So, instead of shape now, I have tif files. I also have such data for 2014-15 and I have some other files.

So, I am going to drag this to my map and see if I get something, it seems like it does not get dragged; well, it does fantastic, it got dragged.

So, here now I have this data for a separately requested from the Bhuvan portal and acquired from there, which does not exactly look like the UP data, it is a data set that looks more like Indo-Gangetic plains. This is the Gangetic plains, here is the river and this is these are the Gangetic planes. So, the first thing that I could do is I would say I am only interested in analyzing UP data. So, I am going to now clip the data according to that.

So, I am going to go to my Geoprocessing and I am going to say Clip Raster, I have just opened it for you.

So, I have just searched Clip Raster, and it says there in data management tools, I am going to say Open, Input Raster, there is only one raster on the window, on the left-right. By the way, one thing that we should always do is we should always look at the properties of our data.

So, I should look at the spatial extent, and spatial reference, well it is WGS 1984 that is fantastic. So, I can overlay the vector files on it and the locations I see are a good representation of where certain land use types are.

The data can also be seen here, I mean this is in the physical folder. In the other folder, you only see one file `lulc_0506.tif`, you only see this file, and you see this file. Just like shape files these data also come as packets. So, they will have projections, they will have attributes.

So, they are not just a single file they are multiple files. So, if you want to copy and paste these data, and move these data around, it is not a good idea to do them from physical folders. Rather we should go back to our catalog and we should say I want to copy these data and I want to take it to a different location, let us say here and I am going to say I want to paste data ok; sorry about that.

So, I want to copy these data, I want to go to this folder. Let us say I go to the folder right here, India Admin Data, and I say I want to paste these data, so that is done. So, I have this data now in India Admin as well let me go to the physical folder and figure out whether that happened. So, I am going to say India Admin Data, I am going to look for here we go.

So, `lulc250k`, I hope you can see `lulc250k` has a total, I think 6 files that came with it. They all moved together when I moved them with a single file from the ARC catalog So, again that is the best practice and you should always follow this. So I have been able to open these data, but what are these different colors right? So, I can on these data if you see, I have an arrow which I click and I have different colors that represent different classes, right?

I have class 0 which is black color, and class 1 which is red color. If I pay attention here the NCR region seems like one represents urban or built-up areas. Then there is 2, 3, 4, 5, 6 all the way to 18, 19 till all the way down, there is a singular color and I do not see those things. So, I guess there are no classes here. How do I know what these classes are? Well, when I download these data, I request these data from ISRO, they also provide me with a lot of metadata right?

As part of those metadata, first of all, there is this readme file, the readme file has the license; how do you cite these data? What are the purposes you can use these data for and so on, right? So that is one.

Second is the `lulc` classes, well one is built up that is what we guessed, 2 is kharif crop, 3 is rabi crop, zaid crop and 5 is double and triple crop. So, let us pay attention to the first 5. So, built up kharif, rabi, zaid, and double-triple. So, let me go back, the first is called, can I rename this class?

Let us see if I can rename 1, 2. Let us see display source, metadata, and general, no I thought I could just anyway. So, I know that one is built up, I can write on a piece of paper if I am not able to edit here, I think one can edit these things. So, you should look up on your own time, how to edit these classes.

So, 1 should be called built-up, and 2 is kharif which is yellow; where do I see this yellow on my map? Well, I see a lot of it here, with this place well let us look at the districts and let us see what this district is.

I am going to open the district's file, I am going to label the districts and I know that a lot of kharif cropping going on in Bijnor, just trying to bring these things up here. Lot of kharif cropping going on in Bijnor.

Rabi is the winter cropping non-monsoon so on around late November or early December and then harvested around March, early March mid-March, that is rabi right? So now, a lot of rabi south of the river right? So, here is Ganges south of the river, you have a lot of rabi going on right? Look at Jalaun, Jhansi, and Hamirpur a lot of rabi cropping in 2005-6, zaid is the summer crop, it happens in April, May, and June, so that is it. So, March, April, May, June.

Where is zaid? Let us try to find out, can I find one single pixel of zaid, not quite right, I mean there is not much of a zaid going on, well there is some here it seems, no that is not zaid.

So, we need a brown color, if I have zaid, it is going to be brown color. So, we could probably look at the east, there we go, there is some zaid in Bihar it seems right, this is all. So, look at the water bodies nearby. So, where the summer crop is getting its sustenance, that is how we do visual analysis, then 5 is double and triple.

So, this is this green, there is a there is quite a bit of this green by the way like I mean so there are this double triple cropping means, they are the land is being cropped three times in a year, at least twice or 3 times all three seasons are cropped, that is interesting. So, there is a lot of double triple cropping in Uttar Pradesh, which makes sense, I mean these are good soils very rich in natural resources and so on.

So, I have this data, and here are the colors, this is also downloaded from Bhuvan, I have taken this from Bhuvan. So, you can tell what is what here. So, you can compare this file with

this file here and you can name your classes. Now, I want to work with only UP data. So, I am going to clip the raster, I am going to go back to my Geoprocessing, I am going to clip the raster.

And say find clip raster, Input Raster is only 1, output extent I want to work with UP data. So, I am going to say UP state from the DIVA-GIS, it automatically fills the extent and it says clip. So, I am going to now take my raster data to the raster data folder, I am going to say lulc0506 and I am going to say UP clipped. And I am going to make it a tif file, I could also use dot img by the way, img is a good extension for the raster data, save and no data is 0. So, no data means 0, right?

So, I am going to say that is ok, to maintain clipping extent well that is interesting. What does it really mean? Well, it really means that if it is checked the number of columns and rows will be adjusted on the pixel that will be resampled exactly to match the clipping extent. What does that mean?

Let us first see what it means. If you go to the boundary of Uttar Pradesh, right? I am just going to go to symbology and make it a slightly thinner boundary.

What happens is that at the boundary here you see different pixels and if I pay attention to a particular pixel, the boundary is crossing through the pixel. And the question that this checking and unchecking is about is that, whether I want to count the pixel inside the boundary, right that is overlapping slightly outside the boundary, or if should I be strict and say no.

If you are anyway, anywhere outside the boundary I am not going to include it in my clip. So, I am going to leave it unchecked, I am going to be a bit liberal about it, I am going to be ok, that is fine, just because the pixel is always a square and the administrative boundary is crossing through some of these pixels. I do not want to throw them out, I want to take them in, that is a decision as part of an analysis. You may decide to throw them out, throw them out you should try anyway.

So, I am going to say run, it's going to run it 100 percent done, right? It says building pyramids, clip clip raster is completed perfectly. So, if I uncheck, I am going to have a clipped raster. So, the clipping now has happened according to the rectangular extent it did not exactly happen according to the state. Let me now redo this with the maintain clipping

extent, you know command and say clipped 2 and we will say run. It is running, it is working right here very well.

It is done. Still, I have the same thing, but now the pixels that are going to be counted are going to be a bit different.

So, I have now a smaller set of data that is not as large, which is approximately according to the boundary of the UP file. The next step that I might want to do, just like for groundwater data and railroads data is that I might want to figure out the land use profile inside the districts. So, let me just change the color of this thing, because it's harder to view that color now.

So, I might want to know which district has what percentage of which data, and which land use type. For that we have this tool, under Geoprocessing we have this tool called tabulate Area. So, the tool it says is not licensed. So, we have to license the tool. So, let us figure out a way to do that.

So, now that we have our clipped file for Uttar Pradesh, it is a separate file that we have created. What we can do is, we can now navigate through different districts by name which I know by name, for example, Hardoi, I can visualize the land use makeup in Hardoi in 2005 and 6, right?

I can see which districts have large urban bodies, which have a higher proportion of agricultural land versus built-up land, and so on. If you go in, you will find these small speckles of red. A question arises what are these, right? I mean are these really built-ups or are these errors? Well for that one could you know let us say which district are we in, we are in Hardoi.

So, let us go and open Google Maps, and maps dot google dot com, I am going to open Google Maps, I am going to now look at the satellite layer and I am going to say Hardoi, show me Hardoi, alright. Now, let us look at the district of Hardoi. So, if we go on to, if we zoom out a little bit, this is a satellite picture and you can see these little speckles of white in green right?

These little speckles of white, what these speckles are small townships or villages which have built up areas, right? So, this sort of also gives a sense of population pressure.

So, if you zoom in on any speckle, what you are going to see is a built-up area. So, there is a little built-up area with different names, there are these settlements inside the rural areas, these may be different villages town centers, and so on, right? Bigger town centers will have bigger footprints. So, if I go back to my map, that is what little speckles are showing me right?

Little villages or town centers we are looking at the locations at that fine detail as actual data. What I could also do is instead of districts, say if I had an interest in the groundwater data and land use data.

So, I can go back to my symbology and I could work with say something like proportional symbols, and you know, I did this earlier, I will just want these from 1 to 40. there we go fantastic and I have 5 classes. Now, if I wanted, I could just say I do not want any fill, no color Apply. Do I get it? No, I do not get it.

Wait, So I am getting no outline color, I want, let us say blue, and I will say Apply, So, Apply.

So, I want an outline color, but I do not want Apply, I want no color, Apply done, fantastic. Now, the bigger the circles we saw earlier, is the depletion, a characteristic of a depletion problem. Now, the bigger colors are near this green right or they are near the urban built-up.

Smaller colors, let us figure out where the smaller circles are in oranges. What are the oranges? Well, they are, you know let us go back and look at the data, I know this is built up kharif and rabi. So, rabi lands have a proportionately lower amount of depletion than when lands are double or triple crops, this is a very interesting thing that I have learned from the data.

So, by simple visualization, you can start to make these connections between land use types and groundwater extraction or groundwater depletion. So, of course, the next step would be to formally connect them and run regressions based on groundwater levels and how are they dependent on land use types and so on.

So, for that, we will as I said that type of analysis is more convenient in statistical software like R, R is also really powerful because of its open-source nature. So, anybody can download it, it is free of cost and like ArcGIS. So, we will cover those types of analyses in R.

So, with that, we are about done with our session on ArcGIS. Again, we have learned how to open, how set up the software, how to open vector data, raster data, manipulate them, clip them, and how to sort of walk between an Excel file and a vector file of the attribute table. So, we have learned quite a bit about tools, I hope they will be helpful for you, but I want to end by saying that this is just a ramp-up, there is an entire toolbox in front of you.

We saw under geoprocessing, all these tools and they can all do very interesting things. So, it depends on how far you go in terms of experimenting and spending time, I highly encourage you that you do that. So, thank you very much for your attention, we are done with ArcGIS hands-on sessions, and next up our sessions on hands-on exercises with R.

Thank you.