Geographic Information Systems
Prof. A. K. Saraf
Department of Earth Science
Indian Institute of Technology – Roorkee

Lecture – 20
GIS Analysis - 01

Hello everyone! and welcome to new discussion of this geographic information systems course. As we have been discussing so far, it is all for the development of GIS spatial database and once the database is ready then the real analysis is starts and that is basically heart of the GIS that analysis part. As you know that in earlier time, when the GIS started coming in our life.

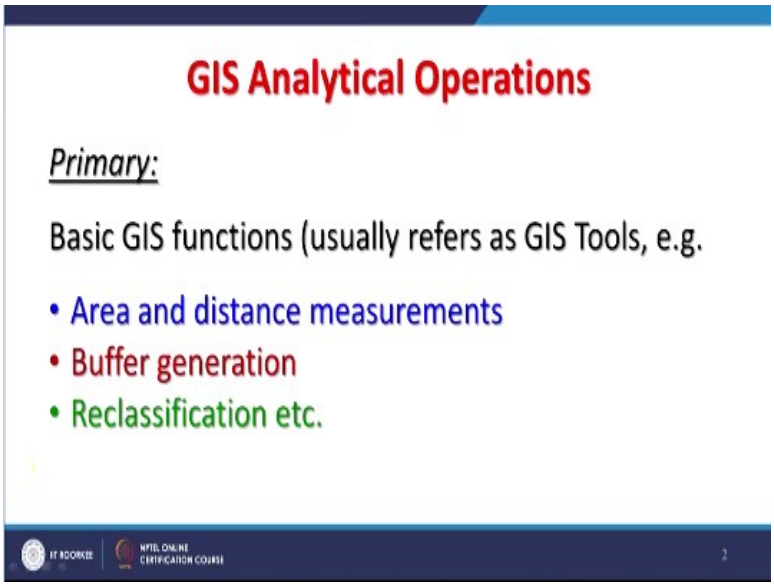
And day to day many operations, we used to spend a lot of time on development of GIS spatial database. Slowly-2 that time is getting reduced. The reason is because lot of now data is available which is compatible to GIS formats. You know, it is sort of independent of software. And that means many vector files are available in safe format which is acceptable or compatible with almost every standard GIS software.

So, the vector data can directly be used into GIS rather than we ourselves are converting format and geo referencing and then organizing in our own database. So, directly either we can get from some organizations or from portal and can start using it. If we talk about the raster data, it is almost the same that a lot of raster data is now available in TIFF or geo TIFF format. And that means they are already geo referenced.

And we do not have to do any further processing on it. Directly again, we can use in GIS. As you know that lot of remote sensing data based on different satellites of different countries is now available on net, free of cost. And that too, very well organized in geo-reference and sometimes well-organized products are also available. And those products of like MODIS's sensors products, there are various products which we can directly use in our GIS analysis.

So, what I said in the beginning that the time which we used to spend to development over our GIS database is now getting reduced. And we can spend now more time on GIS analysis that is the basically purpose of GIS to do various kinds of analysis. So in this, we can see that again, GIS analytical operations can be divided into 2 main categories.

(Refer Slide Time: 03:16)



One is the primary category which involves the basic GIS functions. Sometimes, we also call as GIS tools. For example, we do very basic measurements like area, perimeter, distance measurements say between 2 points. Area of a polygon and the length or distance along the line and likewise. So, these we put under category of primary operations. There are few more; buffer generation which we will be discussing as a separate discussion.

So, buffer generation is also, we can put in primary analytical tools or analytical operations. Buffers are very useful in many cases and if we can make more realistic buffers involving some other parameters like for example for smoke plume, so simply we can draw a circle along a point source from where the smoke is coming but if we can involve in the impedance; impedance for example, the wind direction and wind speed.

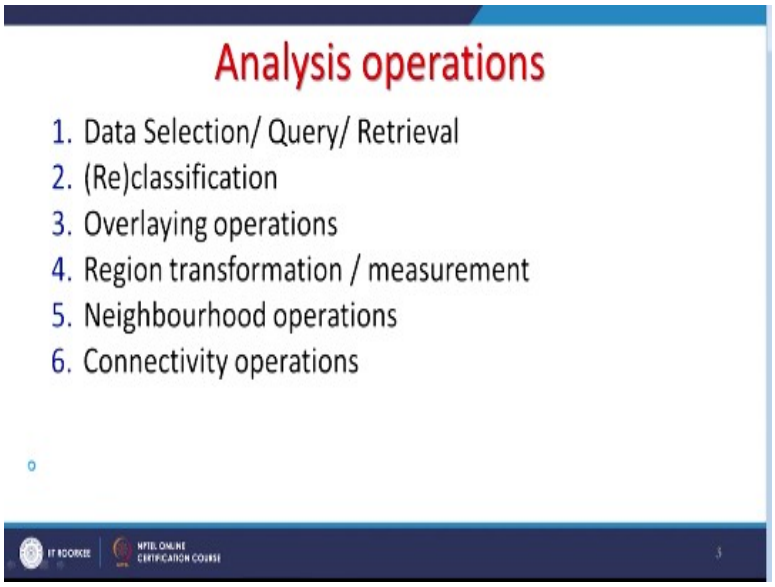
Maybe the terrain if there are undulations and maybe the vegetation or height of trees or buildings then we can really create a very genuine or very realistic buffer about the dispersion of smoke plume or pollutants. So, buffer generation, though it comes under primary operations but it is a very useful tool or a useful operation of GIS. Similarly, we can do the reclassification for example we do a lot of image classification on satellite images.

And similarly, we might be having a continuous data like raster for example a digital elevation model which is a completely continuous information. Now we want to classify into different categories and therefore from a continuous, we can make it as you know discrete or

in different classes, only for our display purpose so that the interpretation of such a continuous data can become much easier.

For example, we can reclassify digital elevation model into a relief map by giving say, 100-meter elevation above mean sea level to 200 meters; one category. 201 to 300, another category and so on so forth. So likewise, we can do the reclassification. So, that is also put under these primary simple operations basically.

(Refer Slide Time: 06:01)



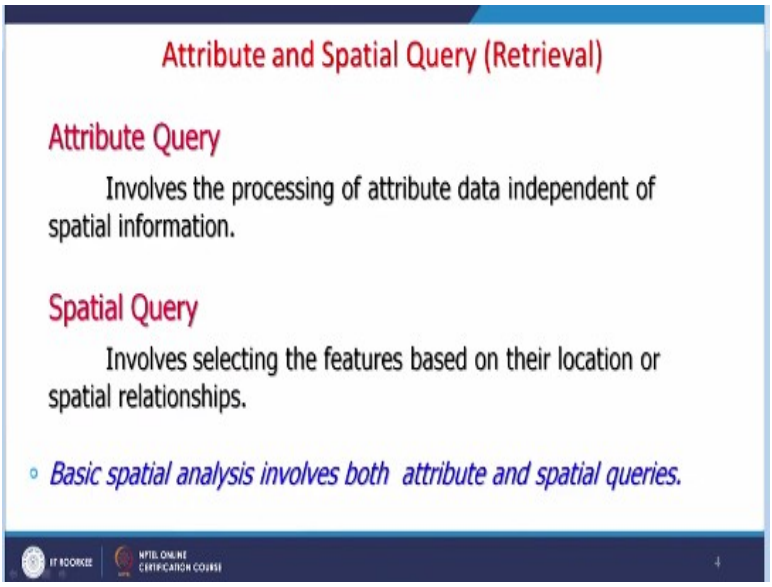
So, like for example if I go for another operation like data selection, query, retrieval; these are also very simple analytical operations. In some of the demonstrations, you have seen that how GIS database can be queried put in a very simple syntax or maybe more nested syntax, we can put and can retrieve the information from our GIS database. So, we put under this category of data selection query retrieval.

Reclassification, as I just mentioned that from continuous data to discrete data, sometimes it is required. So, we perform that too also. Then we also do overlay operations which we will see in a complete separate discussion which is based on set theory, Boolean logics and this is how the complete mathematics have been you know, sort of imported or implemented into GIS through overlaying operations that we will have also discussion.

Then region transformation, measurements; all these are also performed. And then neighborhood operations; there are many neighborhood operations because GIS is the only technology where the relationship between neighboring objects can be exploited so that also.

Then another analytical operation of GIS is the connectivity operations. So, we will also have discussion on that later on.

(Refer Slide Time: 07:49)



So, first one is the attribute or a spatial query or retrieval. Basically, the query generally done on the attribute. Suppose you are having point data and the point that corresponding attribute table might be having 100 fields. Now, I can query a single field, put a condition or I can query all the fields together putting a nested query and by which, I can retrieve the desired data from my database.

The only condition here is that whatever the output which it will generate, if I involve more than 5, 6 attributes together through a nested query then I have to interpret; I have to understand that data or that output. So, one has to be very careful. The best solution is going in steps. So, first you query using few attributes or a few fields of your database and then in the next step, again few.

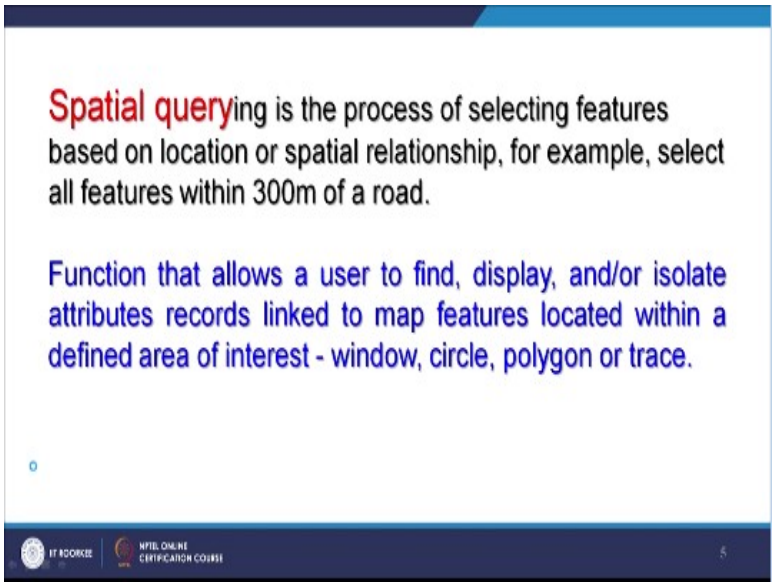
And after each operation, you can check for errors whether the results are coming alright understandable, usable or not. So, that way we can perform a very nested query in a step manner. So, this as you know that it involves the processing of attribute data independent of spatial information. And then spatial query that means performing query on the spatial objects that is also possible and this is basically done through selecting the features based on their location or a spatial relationship.

So, like we can draw a circle or we can use a polygon, that polygon maybe representing a straight boundary or a district boundary or a watershed boundary or basin boundary, something like that. And based on that we can do the selection and once the selected features are there, either we can save as a completely separate file or we can just utilize whatever the selection has been done.

So, basic spatial analysis basically involves both attribute and spatial queries. So, this is very frequent operation which is done on GIS system or on GIS database which involves basically querying in the system and retrieving the information as per user defined specification. If you recall the definition that was the last part of the definition is that, be model as per user defined specification.

So, user has to define. What are his requirements. What he is looking or taking out the data or querying them from the system.

(Refer Slide Time: 10:25)

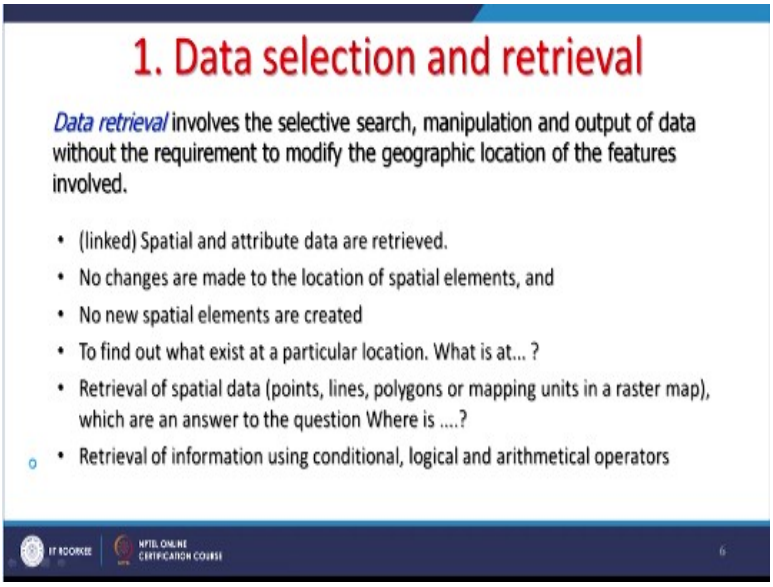


Now, spatial querying is the process of selecting features based on location or spatial relationships. For example, all features within 30 kilometer of road or 300 meter of a road that means create a kind of buffer. Using buffer, a spatial query can also be performed. And functions that allow a user to find or display or isolate attribute records linked to map features.

As you know that in our GIS database; both spatial and non-spatial that is attribute data, they are having a link or a dynamic link between these 2. So, if I do a search or query over

attribute, same times my spatial objects are also selected. And in same way vice versa is also true that when I perform a query on my spatial objects then the same time I my attributes information are also selected. This I earlier demonstrated through a demonstration class also.

(Refer Slide Time: 11:31)



Now, first is about the data selection retrieval. So, data retrieval involves the selective search. We can make different kinds of searches based on our requirements. Only thing we have to care about the syntax and brackets and other things. Sometimes we get a lot of errors if we do not put the query or the syntax properly. So, one has to be really very careful by doing these things.

And whenever the data is out or whenever a selection is done, one has to be careful to understand that one, whether right selection has been done or something wrong with my query. So, both is spatial and attributes data, we can involve. And one important point here is that when we do the spatial query or attribute query, no changes are made in our data or neither in our spatial data or vector data or raster data.

It is just querying the system, selecting whatever is available in our GIS database. And also, no new spatial elements are created. That means when I perform the query, there will be some selections that does not mean that new objects are being created or new attributes are being generated, not at all. So, this is not done. However, if I am measuring area or length or distance between 2 points then definitely a new field may be generated in this attribute database.

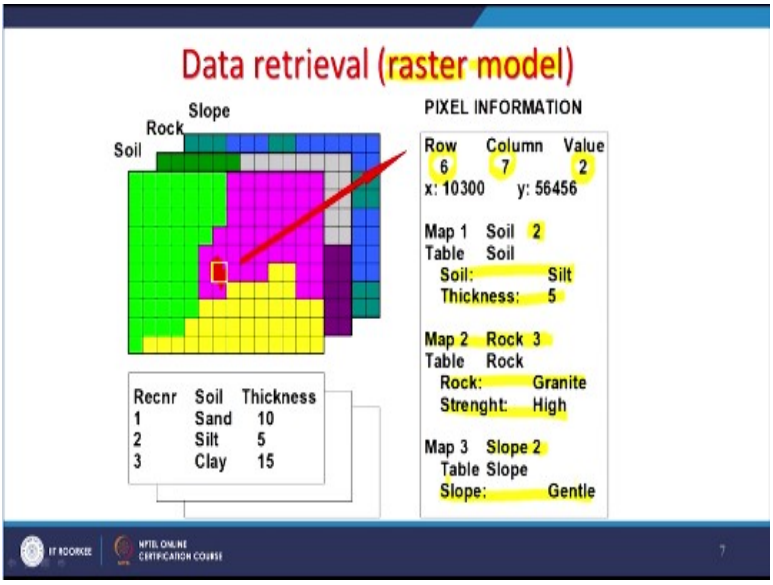
So, only when we want that a new field should be generated which will populate the area, length, perimeter or centroid then only it will be generated, otherwise in a normal query no spatial element or new attribute is generated by the system. And of course, in order to find because the purpose of GIS is to answer a few questions. So, you know, when we involve the location then we can answer a question. What is at that particular location?

So, that question can be answered very easily. And you know, all spatial object whether they are point, line, polygon or mapping units in a raster that means cell or pixels can be retrieved, can be selected and accordingly then we can utilize them. So, if we perform this kind of analysis or retrieval then we can answer a question where is? So that question can be answered very easily.

Now this retrieval; you might have seen the map calculator for example, in ArcGIS or in simple, the Query Builder in different GIS softwares. So, what do you would find that they allow us to use various operators and what are those operators? Like conditional if this is greater than this then select so these are the conditional. Logical also; you can greater than, less than or arithmetic operation, trigonometric operations.

All kinds of you know, mathematical operations which are possible, might be possible are available through these map calculator or Query Builder. So, we can perform quite easily and similarly you know, these overlaying operations are done through that also. So, this is how, these map calculators work that you are having lot of such operators available and which you can use in a single step or all operators may be in one query. But one has to understand the results also.

(Refer Slide Time: 15:25)



Now here, what I am taking an example of data retrieval and that too from the raster model. So, we are having grid on the left side and now what basically the query is that what is at? So, what it at that particular location which is shown through this arrow as you can see about this cell. So, the location is shown here, the column number is shown here and then value is shown here.

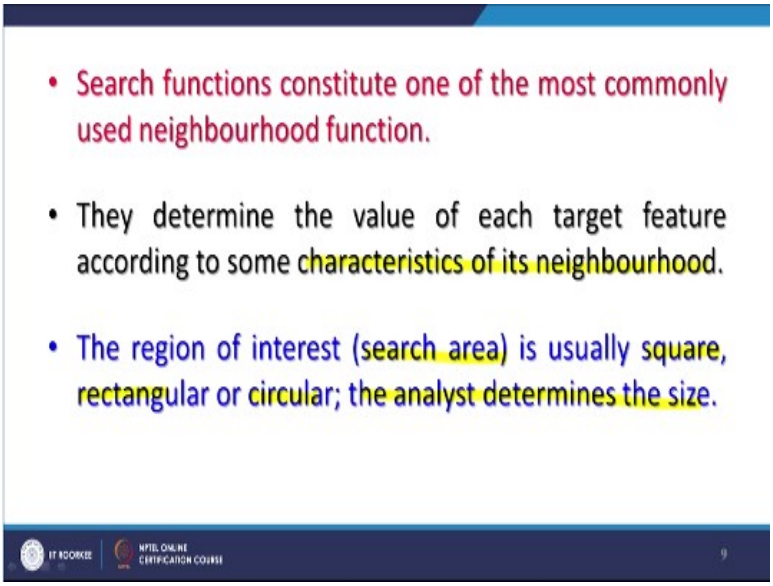
Now there are 3 layers. The top layer is the soil layer; the next layer, in the middle layer is the rock layer and the last layer or bottom layer is the slope layer. So, when I select this particular cell then I am getting the information that the soil type is silt and the thickness is 5 that means in attribute, I am having some information but though, I have mentioned that it is raster but basically, as you know that raster can have only one attribute.

So, how 2 attributes are coming? These can come through our vector also. So, we can have a vector-based grid also. So, in that way, these 2 can come. Otherwise, there can be only one attribute associated with each cell or pixel in the raster case. Same with this rock layer; we are getting the rock information also.

And similarly, from Slope map, we are getting the slope information. So likewise, just select one cell where, what is at and then I get the information and that is very simple way of data retrieval. Now as you know that these search functions constitute one of the most commonly used neighborhood operations. This is what I also described earlier that these are the very common.

You know almost every day if you are having a good GIS database available and used regularly then our search functions are used very commonly. And you try to also get information from a neighborhood also.

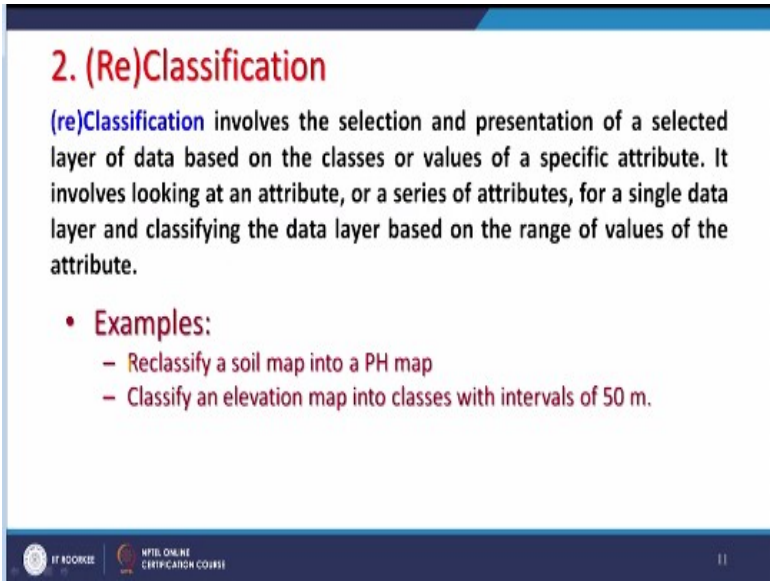
(Refer Slide Time: 18:02)



So, this determines basically through this; you determine each target feature according to some characteristics of its neighborhood and also the search area, depending on the area of interest or region of interest usually that maybe square in shape, maybe rectangular, maybe circular or analyst determine the size that means that I may have a boundary which is arbitrary boundary like a district boundary or a watershed boundary or basin boundary.

So, it is not necessary that I should have only the square, rectangular or circular. I can have any shape and still I can perform search, either on the vector data or raster data without any problem. Now after this data selection, query and retrieval, now we come to the second most common analytical operations in GIS is reclassification. As discussed earlier that the purpose of reclassification is to reduce number of classes from original map or from a continuous data to reduce to only few classes.

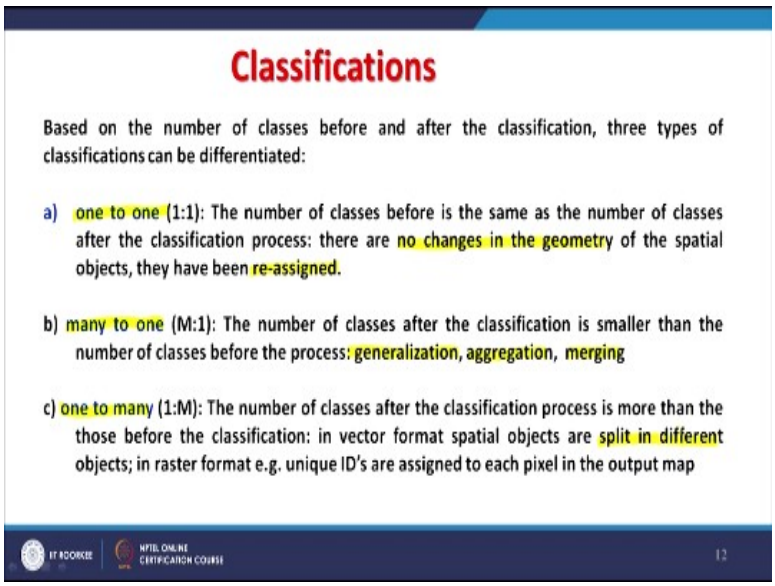
(Refer Slide Time: 19:09)



So now, we will also see some examples as well. So, if we go for definition of this that basically reclassification involves selection and presentation of a selected layer data based on the classes or values of a specific attribute. Now if I give the example of digital elevation model then there is only single attribute and that is the elevation. So, I can perform analysis on that one and can get a new map. Instead of that map, we will not call as a digital elevation model. Then we will call as a relief map and it will have different categories.

Similarly, I can do reclassification of a soil map into the pH map if I am having attribute for that. And classifying an elevation map into the classes with intervals of 50 meters, this example I have already discussed with you.

(Refer Slide Time: 20:10)



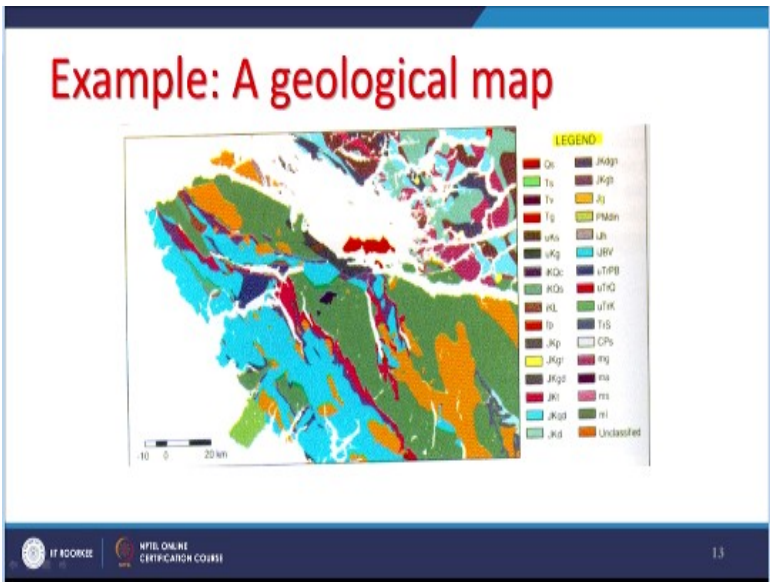
So, this classification basically or reclassification based on the number of classes; before and after a classification. So, there are 3 different types of reclassification techniques are available. The first one is one to one relationship. So that the number of classes before, is the same as the number of classes after classification process.

There are no changes in the geometry of spatial objects and they have been reassigned. Only the values are reassigned and this we put as a one to one. Now, there can be 2 other types. Next one is many to one. That means now, we are reducing number of classes. In the previous one, we are reassigning. So, there is no changes in the geometry because when we do not combine 2 classes, therefore the geometry will also not change.

So, number of classes in one-to-one example will remain same whereas many to one, then I am reducing number of classes and that involves the aggregation, merging, generalization also. So, it will go for a smaller number of classes from original one. And the third possibility is one to many. That means I am now expanding and of course, the reassignment is being done.

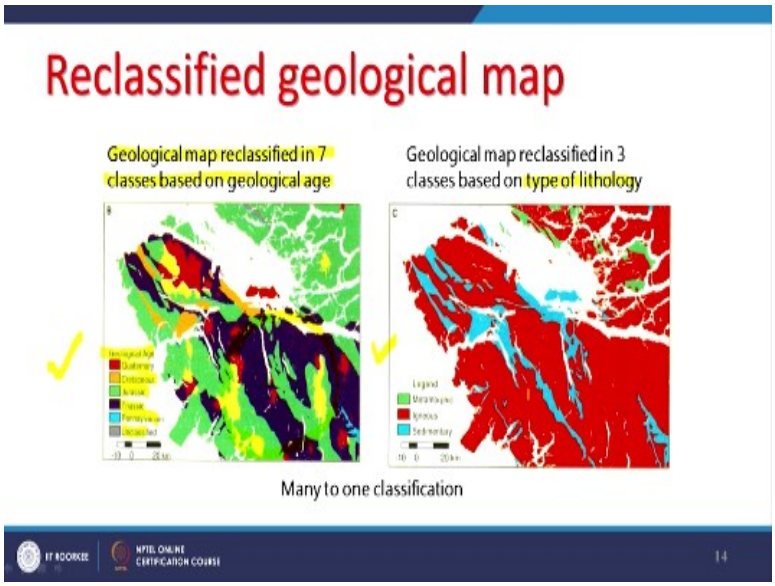
And maybe splitting my already existing objects or number of classes. So, I may be going for more number of classes. So, 3 types; one to one, many to one and one to many. All these 3 things; sometimes for some spatial purposes, we have to do it on our data so that we can then compare with other layers, other themes or other datasets.

(Refer Slide Time: 22:09)



Now I am bringing one example. Of course, this is a vector data example means it's a polygon map and there are different classes as you can see or different rocks which are present in this map which are given in abbreviated form here in the legend. And now if my attribute table is having some other information related with these units then I can reclassify this map.

(Refer Slide Time: 22:39)



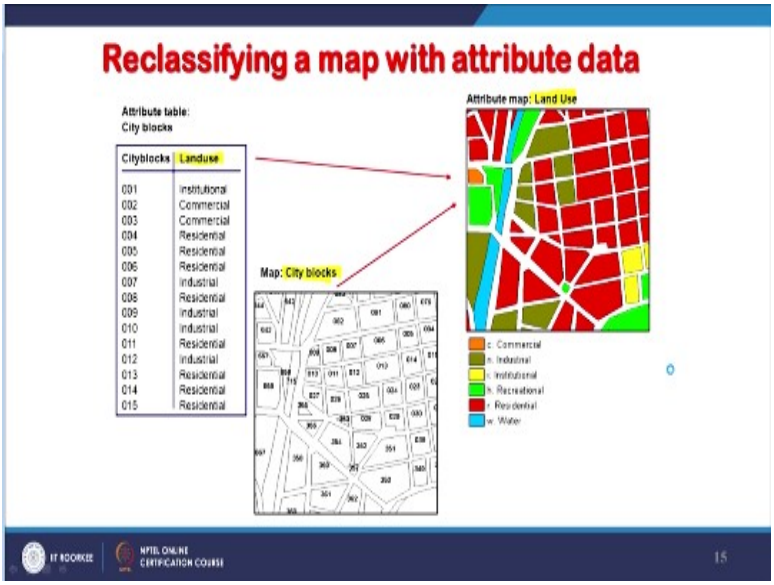
For example, I can reclassify this map into geological map; reclassified into 7 classes based on the geological age of different units. So, here you are seeing about 20 classes, those have been reduced that means many to one and those have been reduced to only based on the age. So, in my attribute, I must be having information about each polygon or against each rock type, what is their age.

So, if I am having that information, the same polygon map I can reclassify using age column or age field as he had done. So, there are now 6 classes. So, instead of 20 classes and therefore the maps become difficult to interpret, now I have reduced to just 6 classes based on their age. Earlier based on their rock types. Now also using same map, I can reclassify into just type of lithology.

Type of lithology means whether they are sedimentary rocks, igneous rocks or metamorphic rocks. And so, from 20 to 6 in this example and in the (C) example, we are having from 6 to 3. The only condition here is that our attribute database must be rich, must be having those fields which I can use to reclassify my polygon map and that way we can do it. So, a lot of information is required.

If that information is already in my attribute table, then all such operations are very easy to perform.

(Refer Slide Time: 24:28)

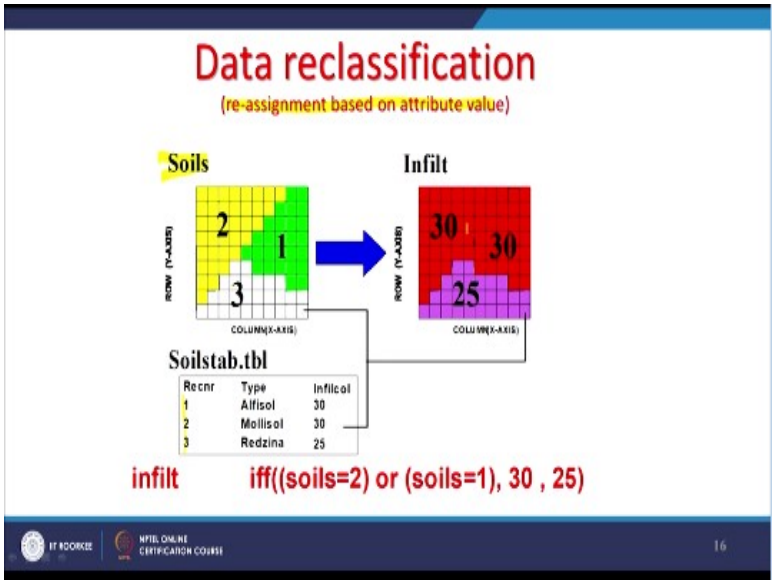


Similarly, here also little different example that reclassifying map based on attribute information. So, what is attribute I am having against each polygon is the land use. How this land or different plots are being used here? So, basically city blocks are there and this is the land use. So, the same I have reclassified into this one and this example is one to one reclassification.

That means there is no merger of 2 polygons and there is no new generation of boundary of any polygon. Whereas in previous examples, you know from 20 to 6 when we came, we have a reduced number of polygons in my display. Not really in database, only in for display. And when we went from 6 to 3, further reclassified based on the type of lithology that means sedimentary, igneous, metamorphic then only 3 types of polygons are there.

And lot of polygons which we are present here in this center part or you know, the southeast part has been merged just as a red color and that is igneous rocks. So, all kinds of possibilities are there and we can perform those analysis like that as per our requirement or as per user defined specifications.

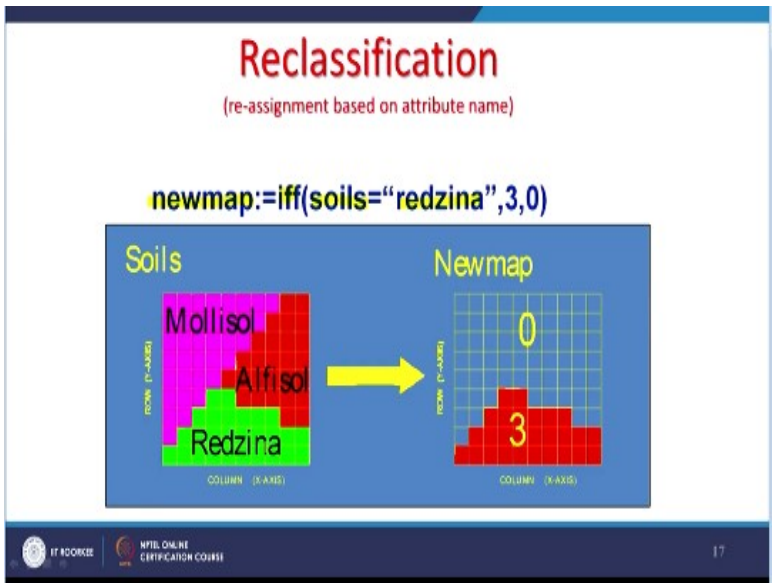
(Refer Slide Time: 25:59)



If we are having say, reassignment based on the attribute value, again example is here that I am having you know 1, 2, 3; here is the soil map. Now against each my cell; I am having another information that is infiltration. So, infiltration values if are available, I can reclassify the same map. If 2 adjacent areas are having same infiltration value like here on the screen, they are appearing as one single polygon.

But in the system, each cell or other things are stored completely differently. So likewise, you know that if this condition applies, we can do that kind of reclassification.

(Refer Slide Time: 26:48)

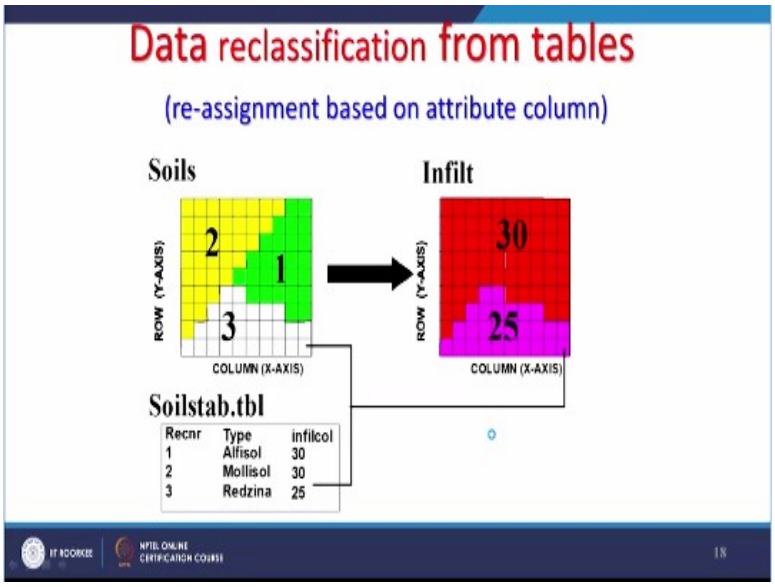


Similarly, reclassification is done that generate a new map if soil maps equal to redzina and if it is true then 3, otherwise 0. So, this is what is done. So, this green one is the redzina which is satisfying this condition of my query and rest are not. So, this is conditional operation and

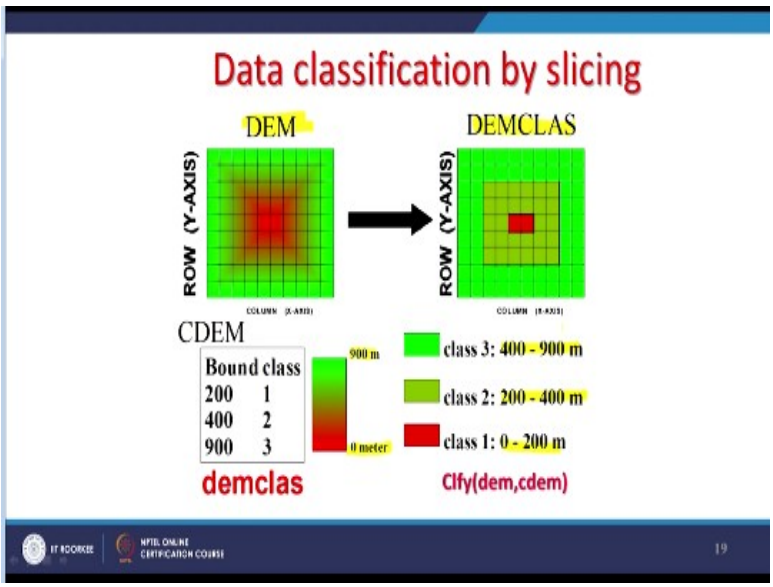
rest are not so, the zero value has been assigned for the rest area. So, as I mentioned, you can use all those operators which are available in your map calculator or Query Builder.

But the output you have to understand. So, instead of using a very complicated nested query, do it in several steps using simple-2 query.

(Refer Slide Time: 27:44)



(Refer Slide Time: 27:52)



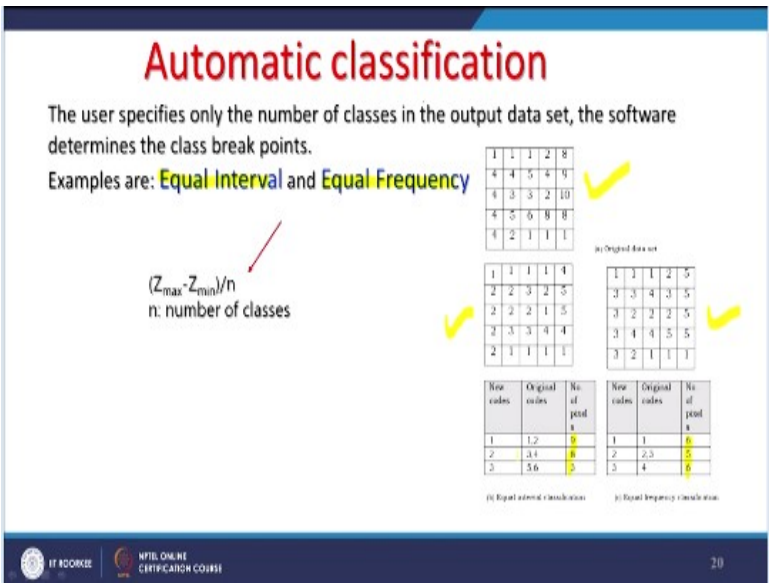
Also, one example that data classified from tables from the attribute. So, you can see here that infiltration values can be done like that. Now DEM classification; I have already given example. So, like here elevation values in my grid are varying between 0 to 90 and this is my digital elevation model. When I reclassify it or classify it then I am creating category that 0 to

200 meters should get this color and 200 to 400 meters should get this color and 400 to 900 meters should get this color.

So instead of now having a continuous digital elevation model or continuous values in my digital elevation model, now I am having just 3 classes. And I can give name that very high elevated areas or middle elevated areas and low elevated area. So likewise, usage or utilization of such reclassified maps can be very-2 useful. In from digital image processing power lines or point of view, this is also called density slicing.

So, same because the concepts are coming from different domains. So, the same term can also be used here.

(Refer Slide Time: 29:07)



GIS softwares are also capable of reclassifying based on some automatic classification methods. For example, like involving the statistics, equal interval or equal frequency classifications that is also possible. So, if this is my input map, I can classify this way or this way that means I am having either equal interval or equal frequency classification.

So, you see here, now the frequency is almost same 6, 5, 6 whereas here, the frequency was not maintained but interval between classes have been maintained. So, 1 to 2, 3 to 4, 5 to 6, something like that. So, this brings to end of this discussion and of course then we will take the next topic related with this GIS analysis operations. So, thank you very much.