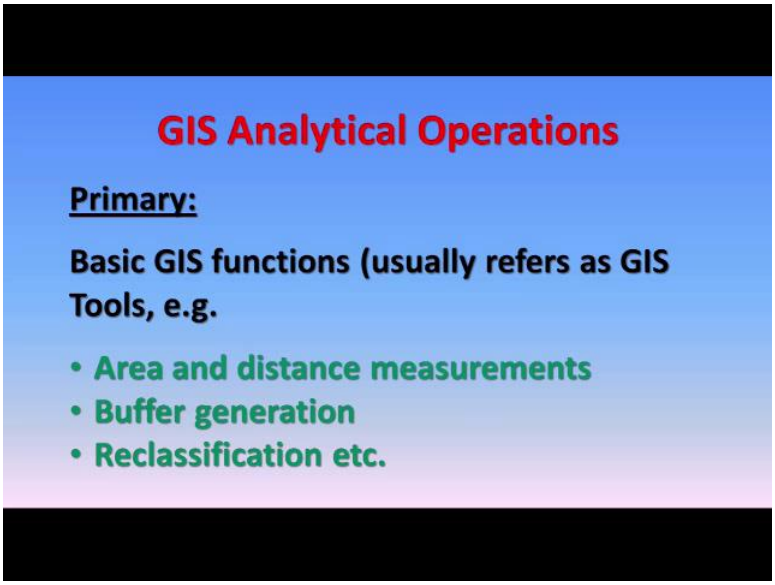
Introduction to Geographic Information Systems
Dr. Arun K Saraf
Department of Earth Sciences
Indian Institute of Technology, Roorkee

Lecture – 15
GIS Analysis – 01

Hello everyone and today we are going to start the very important part of GIS that is analysis. As you know that there we started with the first what is GIS definition of GIS and then different components of GIS, different types of data which GIS can handle, how we bring a data into GIS, then data base system which we have also discussed and finally, now we have reach to the real operation of GIS and this is in three parts and the GIS analysis.

So, in first we will discuss the simple primary operations in GIS or in little detail and as you know that we have discussed about primary operations and secondary operations.

(Refer Slide Time: 01:17)

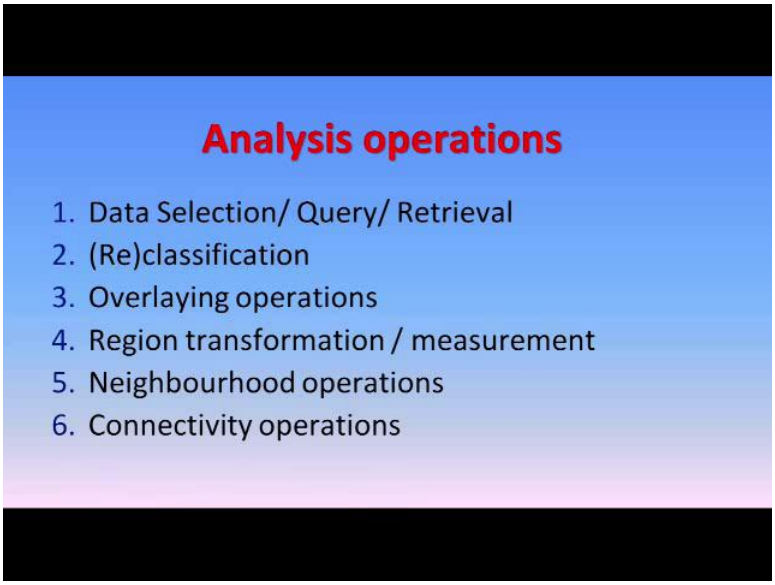


So, in primary operations or basic functions of GIS usually we call them as tools and this as we have discussed about area and distance measurements buffer generation which we will discuss in later lectures and reclassification which we will discuss here now. So, in

data of analysis operation the first thing is starts with the data selection and query, or data retrieval because if you recall the definition of GIS say efficient retrieval of the data that is one of the purposes of GIS.

So, this say retrieval has to be very efficient and there are different tools and techniques are available which we will see.

(Refer Slide Time: 02:00)

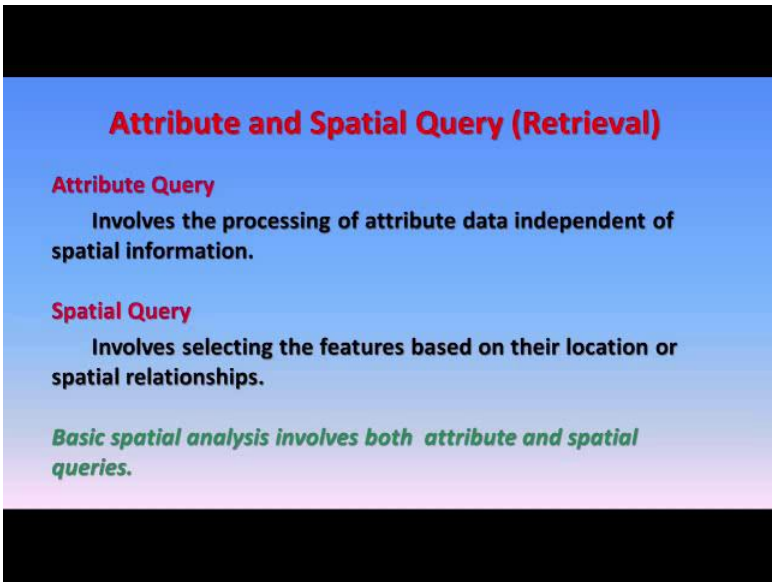


And we will also discuss are classification in this particular lecture - overlaying operations and other neighboring, neighborhood operations connectivity operations all these we will discuss in later analysis - one analysis two and the three and lectures.

So, first is about attribute and spatial query retrieval. As you know that there are two sets of data and GIS we will put - one is a spatial data, another one is a attribute data or we have also called as none spatial data. So, if we take the example of say point data point data is a spatial data which is having just x by co-ordinates no dimensions, but it can have n number of attributes and this there is a dynamic linkage between attribute data and a spatial data. So, whenever we query a spatial data within the same the objects which are satisfying this query on the attributes then they are also selected and in reverse if we do certain selections or a query over a spatial data, that the same time our attribute

data are selected and then we can retrieval; that means, we can save them as a separate file.

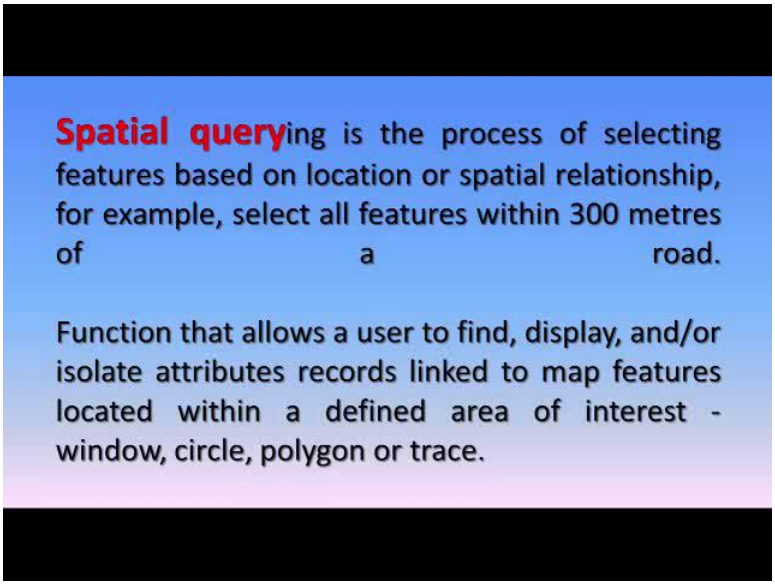
(Refer Slide Time: 03:31)



So, if subset of the data after the selection can also be created very easily in all most all standard GIS software. So, just say attribute query you can perform query either on a spatial data or on none spatial or attribute data. So, this query involves the processing of attribute data independent of a spatial information and in the reverse way in the spatial query it involves the selecting of features based on their location or a spatial relationship.

As I have mentioned there is a dynamic linkage or a hot linkage between a spatial and none spatial data. So, any operations you perform on one set of data automatically the similar operations are also performed on another set of data. And the spatial analysis functions which are in GIS these tools are available which involves both attribute and spatial data queries.

(Refer Slide Time: 04:35)



Now, especially query is basically the process of selecting features based on certain conditions and these conditions the GIS user will put based on may be on locations based on may be the spatial relationships for example, selecting all features within three hundred meter of a road this. So, you are putting these two three conditions that there is a feature which is a road feature or either road network and then you are saying that whatever the features which are falling within 300 meter on both sides of roads should be selected.

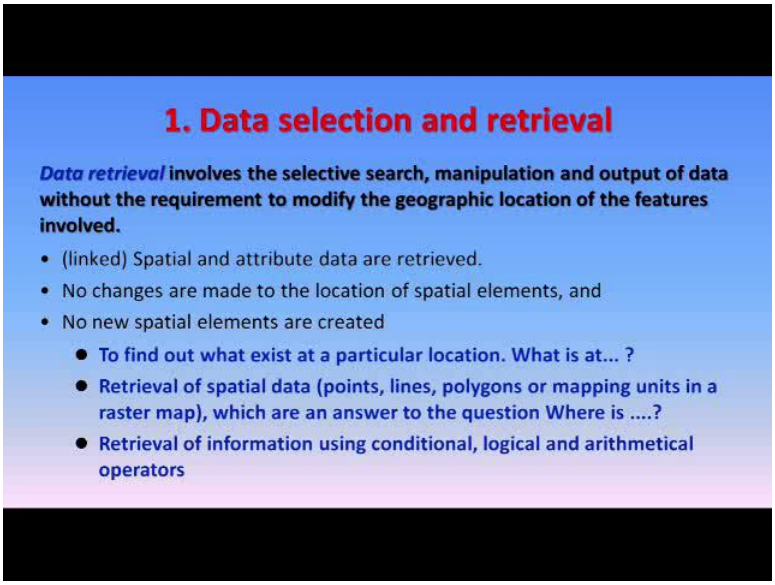
Once they have selected you can say there is a subset of the data and function that allows a user to find display or isolate attributes records linked to map features; that means, there is a dynamic link between both things and this can be a only spatial data you can select you can make selections based on windows and these windows can be of any shape and sizes you can have a circles polygons trace whatever the graphics which you can draw only spatial data.

Now, that base on that basis the selection can be made if you are a already have been certain boundary for example, if you are having a you know blazes of India plotted in a GIS and which are been displayed now you want to only select a blazes of say Uttarakhand. So, you need to have the boundary polygon boundary of Uttarakhand and

once you instruct or make a spatial query that whatever the blazes which are following within the boundary of Uttarakhand these would be selected.

Once they are selected you can save as separate files. So, it is a sub set of blazes of India I will go as a subset of Uttarakhand villages. So, this way that is why it is called also retrievals selection and retrieval.

(Refer Slide Time: 06:20)



And a data retrieval as you know (Refer Time: 06:19) the first you have to search manipulation; that means, a only selected features you are doing and then output of data without requirement to modify geographic locations of the features. When you made the subset of the data you are not modifying the data, neither are you changing their spatial location.

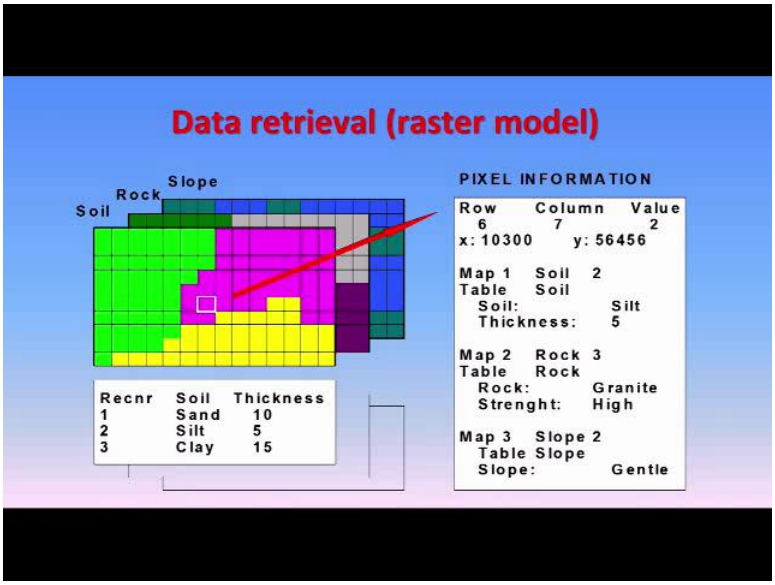
What you are doing you are just extracting the data making a subset and saving as a separate file in your system and as it has been mentioned already that is spatial and a none spatial or attribute data are having linked and there are no changes are made to the location of a spatial elements this is important and whenever you will perform a query original data will remain all the ways in depth unless you go for editing and change the data. Otherwise query will never modify data and neither you retrieval and no new

spatial elements attributed, once as since you are not modifying you are not editing and therefore, new addition will not be made only the subset of the data will be made unless you instruct the system and to find out what exists as a particular location.

So, you start getting answer in during selection and query processes that what is at that location, retrieval of data you can have a point data a line polygon or mapping units of a raster. So, on any type of data, any type of vector any type of raster you can perform a spatial queries, but remember one thing a vector data can have n number of attributes whereas you know that raster data will have only one attribute. Based on that particular attribute you can make the selection and it can retrieve the data which we will see it little later. And retrieval of information using conditional logical or arithmetic operations these tools are also available under a category which called a spatial analysis and sometimes depending on the software.

If you are using say a RGIS software in which there is a spatial analysis a module or a extension is there when you use switch on for this extension all these tools are available and also these tools like conditional that if as a map is having say elevation more than 1000 meter then all though cells which are having and satisfying this condition should be selected. So, you can make these conditional queries you can have a logical queries greater than a smaller than and so on so forth, you can have a arithmetic operations you know plus minus multiplication and subtraction all kind say and in combination also you can have logical operations, based on Boolean logics can also be performed like a you know a union b intersection b all those thing, but we can keep them as a separate category which we called as overlaying operations which we will discuss and later in other lectures of GIS analysis.

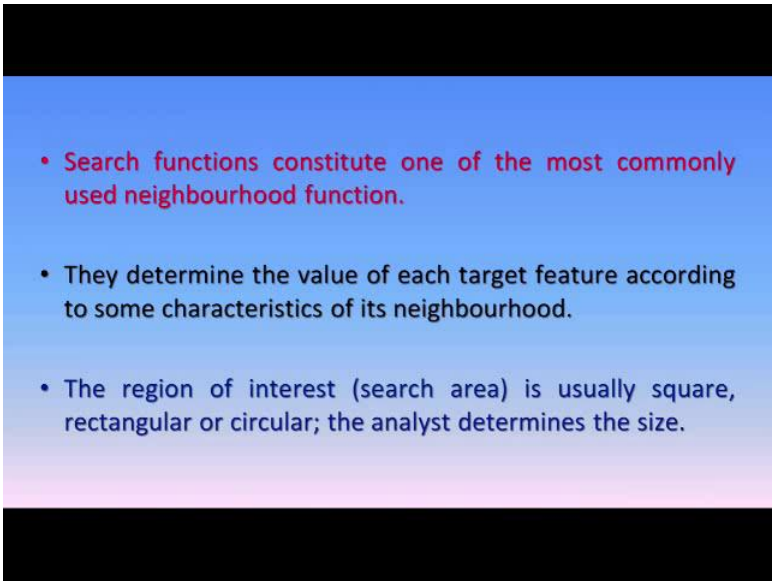
(Refer Slide Time: 09:27)



So, let us look first the data retrieval using a raster model, raster model will only have one a attribute as you can see here there are three raster layers are there one is soil, another one is a rock, third one is slope and it depends on if you just a select a cell over raster all corresponding cells in all remaining layers will also be selected and you get information that here the thickness of the sand in case of soil is this much and all those details can be retrieved. So, map one is having a soil say two thickness a that soil type is silt taken as a soil, in map two which is the rock map is having rock a granite and is a strength is high if that attribute information you have a kept there or map three here you can have a slope map So, it is telling the (Refer Time: 10:28) slope.

So, all this is information it can be retrieved and there is no limit basically, theoretically there is no limit you can have a retrieval of data in case of raster or vector of tens of layers together. So, that is very good advantage having working in GIS.

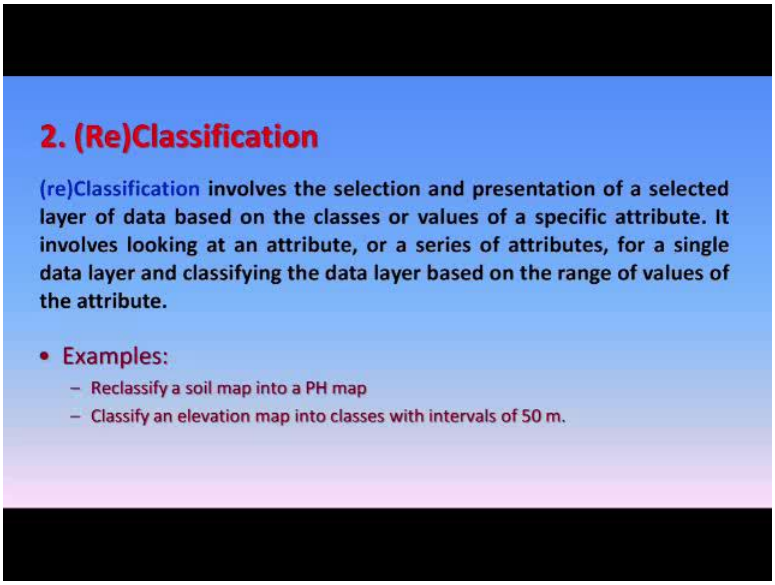
(Refer Slide Time: 10:49)



Now such functions it can also be used which constitute one of the most commonly used neighbourhood function. So, search functions will do the searching yes per your given conditions in the surroundings in the neighbourhood and they will determine the value of each target feature according to some characteristics of its neighbourhood and these characteristics if they are a stored in the system then you can definitely retrieve. Region of interest or search area is usually can be square or rectangular circular or it can be of any arbitrary size, but these have to be polygon in shape and then you can retrieve also using poly line data can also be used and the all these will be a determined by the analyst the size and shape and other things it can only be fed by the analyst himself the GIS expert.

Now, the second one after the data selection query retrieval and now we come to the reclassification. Let us look by reclassification because sometimes you are having a map, say a polygon map which is having geological information. So, I take that as example here.

(Refer Slide Time: 12:07)



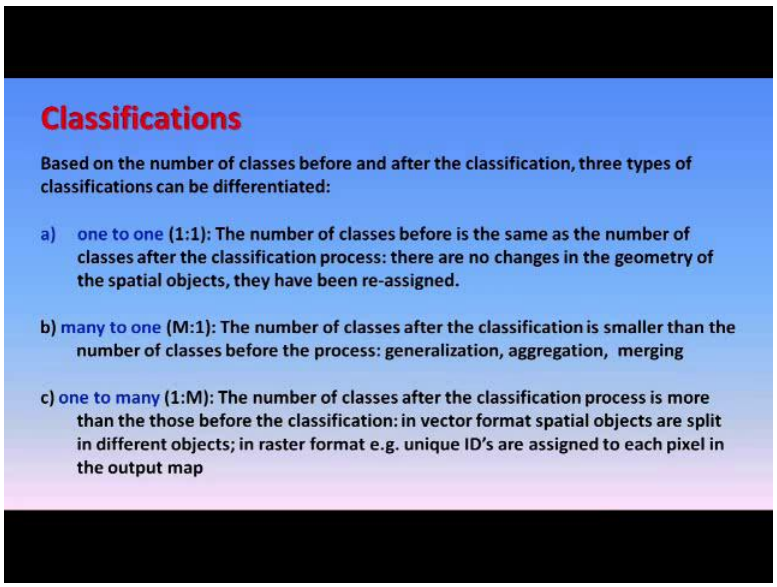
And there they here basically we want to classify that map instead of simple a as a geological map do we want to classify the rocks which are present based on their edges. So, there is what the reclassification is. Now same thing a same map if we are having information about strength of different rocks, we can also classify the same map based on their strength and this type of these types of operations we call them as reclassification.

So, reclassification involves the selection and presentation of a selected layer of data based on classes or values of specific attribute. That attribute has to be there and based on that attribute you can reclassify the data. So, that the presentation style becomes different and that map becomes more appropriate then a very general map and it involves looking at an attribute, that attribute if it is already having the edge of the rocks for example, then you can use that edge attribute and reclassify. Or a series or attributes you can also have of a single data layer and generally the reclassification is done on a single layer single say polygon layer and reclassifying the data layer based on the range of values of the attribute.

You can also if you are having a continuous map like a digital elevation model in which elevation are changing in continuous passion. So, supposing you are having a range of

elevation the minimum elevation is zero maximum elevation is thousand now you can re classify this map instead of a continuous you can segmented and you say that I want to have ten classes starting from 0 to 100, 1 to 200 and so on so forth. So, the range of values can also be used of attributes to reclassify a map and a then reclassification of a soil map if you are having already a ph value against a individual polygons of soil map you can reclassify, reclassifying an elevation as I have already explained about this part that in to different classes and the range user can define or if you do not want to define you want to use from a statistical methods then though two tools are also available which we will discuss in later classes of GIS analysis.

(Refer Slide Time: 14:33)



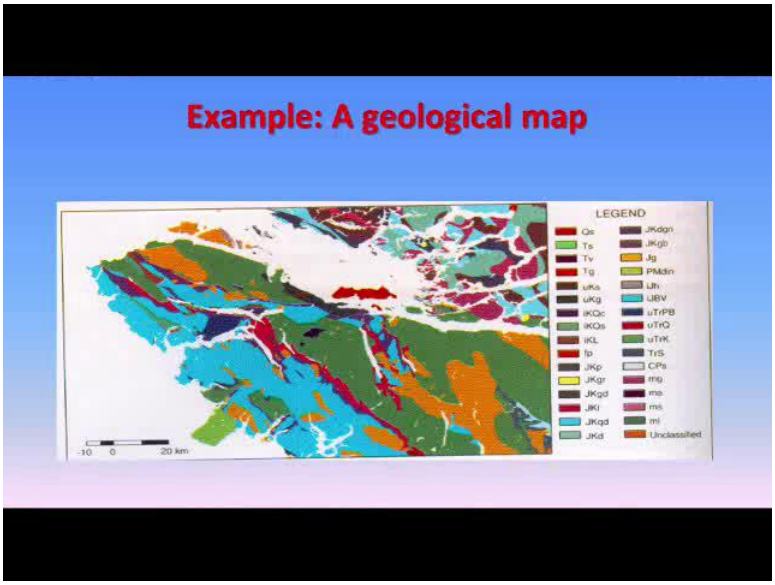
So, the classification basically based on the number of classes before and after classification and three types of classification can be a used in GIS - first one is one to one. So, these are looking basically the relationships and second one is many to many and third one is a one too many. So, one to one in both the number of classes before is the same as number of classes after the classification and there are no changes in geometry of the spatial objects they have been reassigned.

So, whatever say a polygon map bass there you have to reclassified the number of polygon maps which were earlier being displayed same numbers will be there only their

attributes will change and many to one is the number of classes after classification is a smaller than number of classes before the process and this brings the generalization or aggregation or merging also. These generalization and merging and aggregation these things we have discussed in pre other lectures and there is last one is one too many that initially you head few polygons. Now, you may reclassified map you are having many polygons.

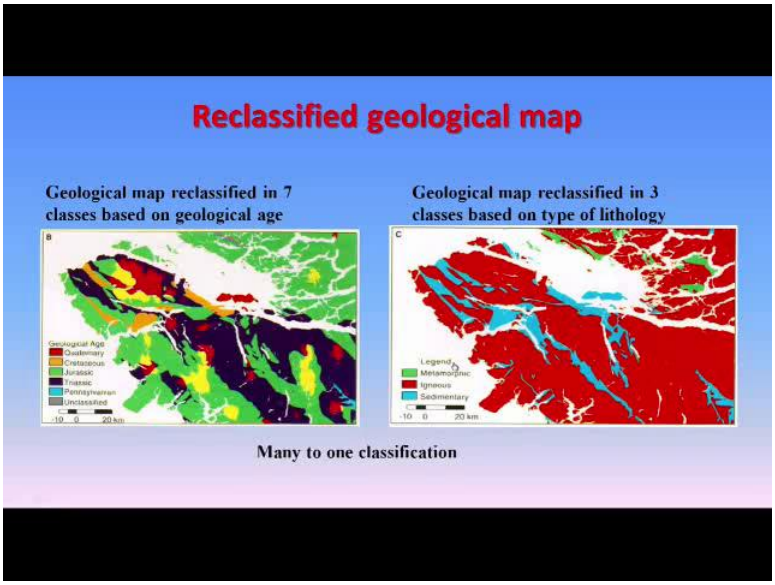
So, the number of classes after the classification is more than the before the classification in vector format is a spatial objects are a split in different objects in raster format for example, unique id's are assigned to each pixel in the output map.

(Refer Slide Time: 16:11)



Let us look the example, example is a geological map which is being displayed here different types of lithologist are presented here now we can if we are having in our attribute table see this is the classified.

(Refer Slide Time: 16:19)



So, we have now number of classes here we have more than 20 say, now we have reclassified and we have used the age as a one of the criteria and reduced the many to one.

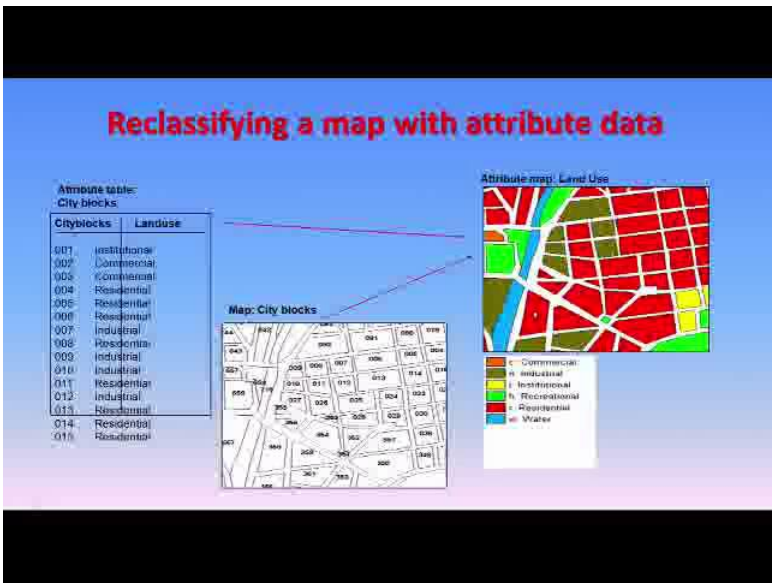
So, we have reduced from many more than 2 classes to 7 classes. So, what happened that many polygons which where we are there now I have reduced and a this whole area we see there were many types of lithologist were represent, but when we brought the ages of these rocks and these a ages are quite a broad ages and therefore, lot of polygons have disappeared and only large you know this classes are coming here. So, number of classes has reduced and therefore, the interpretation b is start interpreting this map it is rather difficult because classes are too many, but here number of classes have reduced say in this example 7 and therefore, interpretation of just maps or presentation of these maps becomes much easier and useful also. Because, now we can understand different rocks based on their ages and that is say in our science study graphic we always try to present rocks in their ages.

So, that is the advantage of reclassifying, but this information has to be there in your attribute data if this information is not there then it cannot be done; however, if we by

some source if you can add that information into a attribute then definitely you can achieve such type of classification.

Now, the same map which is been displayed here if in attribute data the different lithology information's are there we can likely we can merge them here the lithology are presented, but we have merge into three main types of igneous sedimentary metamorphic and therefore, number of polygons and number of items in the legend has reduce just to three. So, these this is really an example of many to one classification, and map has become further simpler and a very you know easily understandable even that where are the rocks, where are sedimentary rocks, where are the igneous rocks and metamorphic rocks. So, this is the advantage of reclassification.

(Refer Slide Time: 18:48)

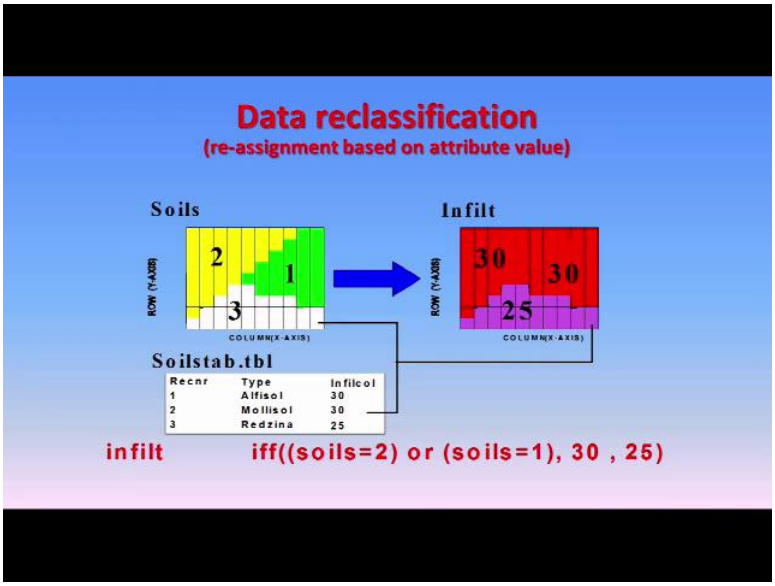


Similarly, a reclassification of a map here is a given here that attribute here we are using land use and there are street maps and there are city blocks or the where the houses are built.

So, you can use a table from here and then change to say land use. So, land use this map as we change to land use now land use is showing say red color is showing residential. So, all residential plots have been given a color which is a red color, all recreational area

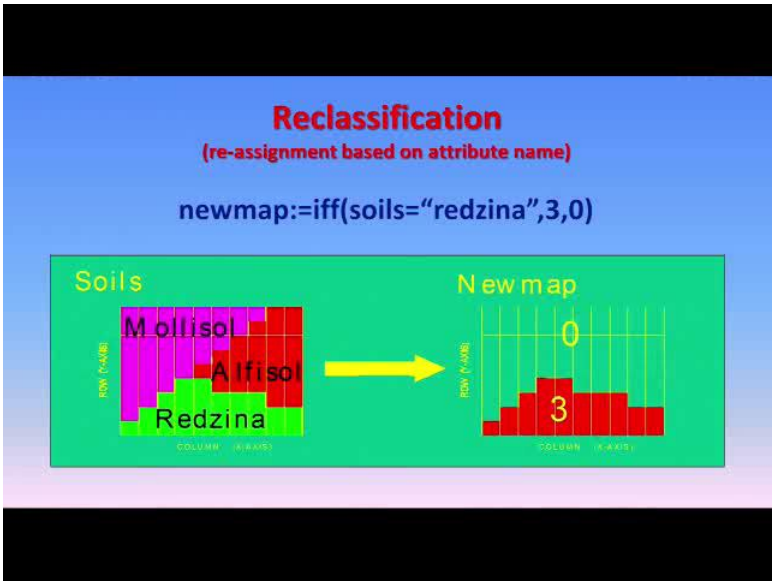
say garden or some play grounds have been given green color, so on so forth. So, the this map initially which had attributes like this if we use this land use attribute and reclassify this map then the this maps becomes much easier to understand then the map which having the and different numbers and you cannot identify easily that which are the residential areas unless you see their attributes. But here using that attribute we can re classify and map become easier. So, that is the main purpose of a reclassification is to either merges aggregate or generalized it that is the main purpose. So, this is the more or less example of generalization and merging together.

(Refer Slide Time: 20:16)



Now, in case of roster again reassigned a based on the attribute value. So, now, one to one these no changes in number of polygons, initial map have the three polygons and three different soils also had in attribute information about the infiltration information. So, soil type 1 2 3 have been reclassified based on their infiltration value. Now class 1 of soil and 2 are having the same infiltration values. So, there about the boundary has disappeared and they are all getting one color and then class 3 is having infiltration value 25. So, now, it has reduced here number of polygons, but as you can see that since both are having same color therefore, they are appearing as one, otherwise there are three polygons and in this particular example. Similarly here assigned based on the attribute name we can also use you know that nominal attribute and reassign this thing.

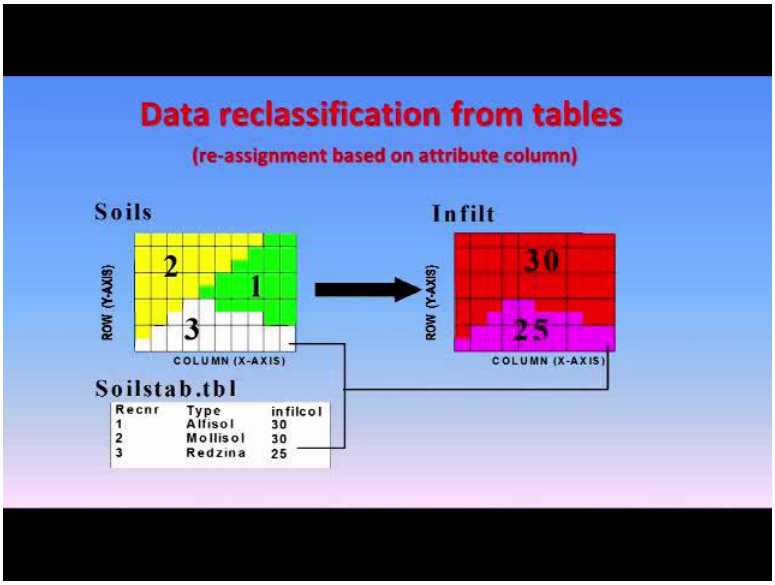
(Refer Slide Time: 21:22)



So, if they are having certain names which are present we say these are the equal one, this has been assigned to zero value and then next one is assigned three. So, again you have reduced number of polygons here.

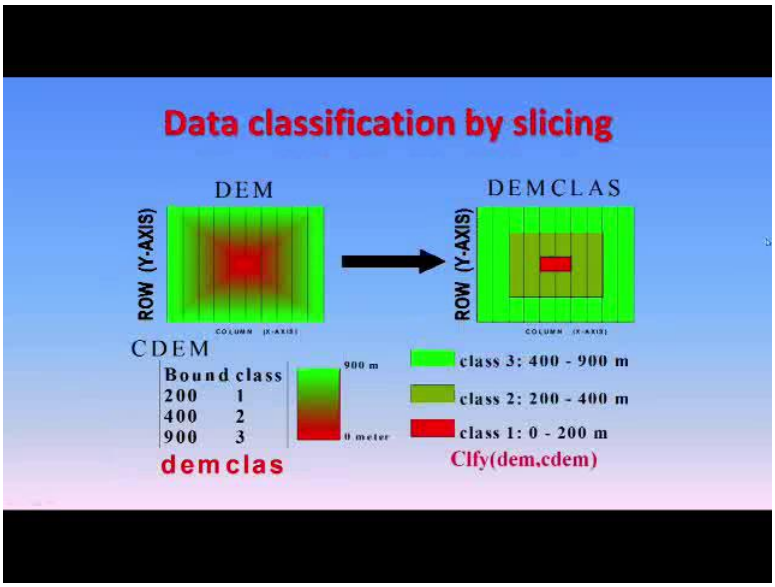
Because the condition which was put is if the soil is equal to Redzina then the value should become, if it is then a values would be three otherwise it should be zero. So, this kind of syntax you have to put into every spatial query. Now another example of reclassification is a using a table again here that they are initial they are three units same example and then it has gone to this one.

(Refer Slide Time: 22:01)



Now, data classification by slicing this is a very common technique in digital image processing of remote sensing data as well where we use the term which we call as density slicing.

(Refer Slide Time: 22:10)

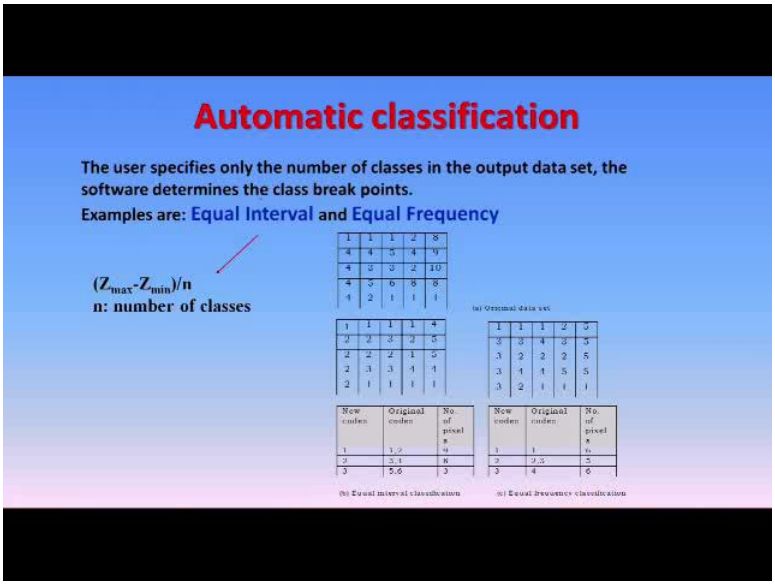


So, on a single layer you can regroup the number of values pixel values into say you are having a 8 bit image and which will have a value between 0 to 55; that means, initially they will be having values of 256 classes. Now you these 256 classes you can reduce to 8 classes by giving a range of values and assigning a color and once you do it we say is slicing or density slicing.

Similarly here, that a digital elevation model is there which is the continuous; in this example it is representing continuous data. Now we can reclassify and give the range of a elevations that a the range of elevations are between say 0 to 100 and then 201 to 400 and remaining 900 and when we assigned different colors then here we get this kind of map which is now classified map and we also sometimes call these kind of map; if the continuous data I have especially elevation data been reclassified into certain classes we call them as relief map as example is shown. It is a very simplified example here, but just for understanding that instead of having continuous data you describe and discretize the data by classifying. So, legend can also be seen that (Refer Time: 23:55) it is a continuously legend 0 to 900 meter where is the legend here you have a three distinct colors and their range are also here.

Now, you can also do automatic classifications because once you start involving your statistics here or the values and their ranges and their variations then you can bring a statistic directly.

(Refer Slide Time: 24:28)



For example, you can use equal number equal interval or equal frequency and then you can classify. So, there examples are shown and this topic of a classification techniques based on a statistical methods this we will be discussing in much detail in later lectures.

So, here the example are given that this is the original limit has been classified based on the interval, it is classified based on the frequency, number of occurrence of values. So, this brings to the end of this presentation on the first one on GIS analysis 1. And in next lecture we will be going much deeper into analysis 2 and then next analysis 3.

For time being thank you very much.