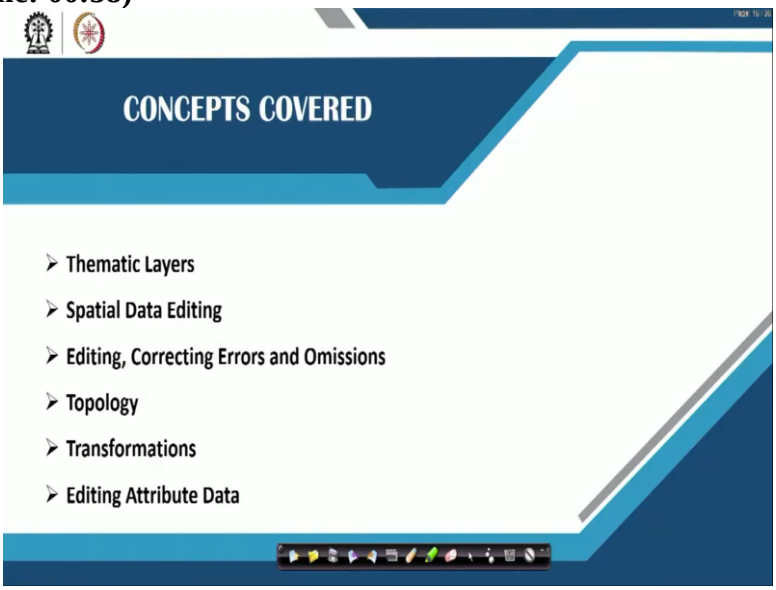
Geographic Information Systems
Prof. Bharath H Aithal
Ranbir and Chitra Gupta School of Infrastructure Design and Management
Indian Institute of Technology, Kharagpur

Lecture 24
Spatial Data Editing and Transformations

Hello, Namaste. I am back here with the next set of lectures which I promised in the previous class that I would start looking at how will graduate to look at the software part of this particular course. So, in this though it has a GIS data quality I would be looking at how do we edit a data? What are the different the transformations that we would see? And how do we actually edit a data in terms of getting a good quality data.

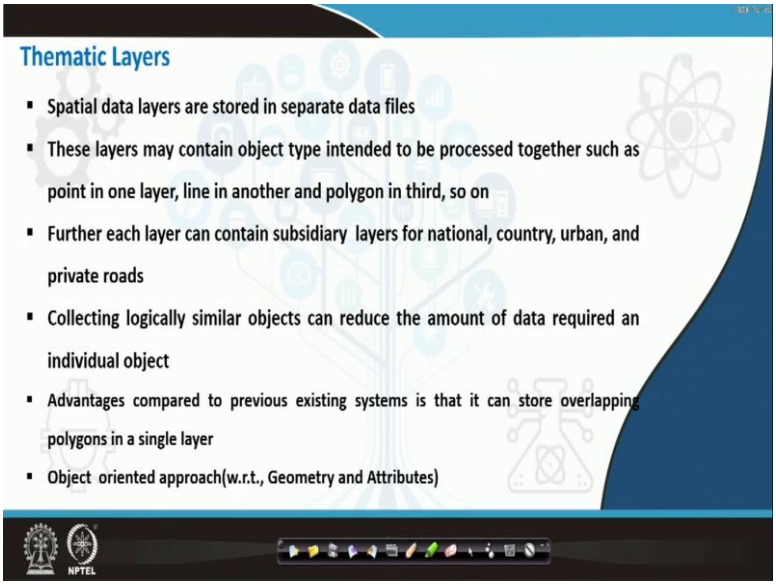
What kind of editing is necessary? What are different editing tools? That is what we will see in this particular set of slides.

(Refer Slide Time: 00:58)



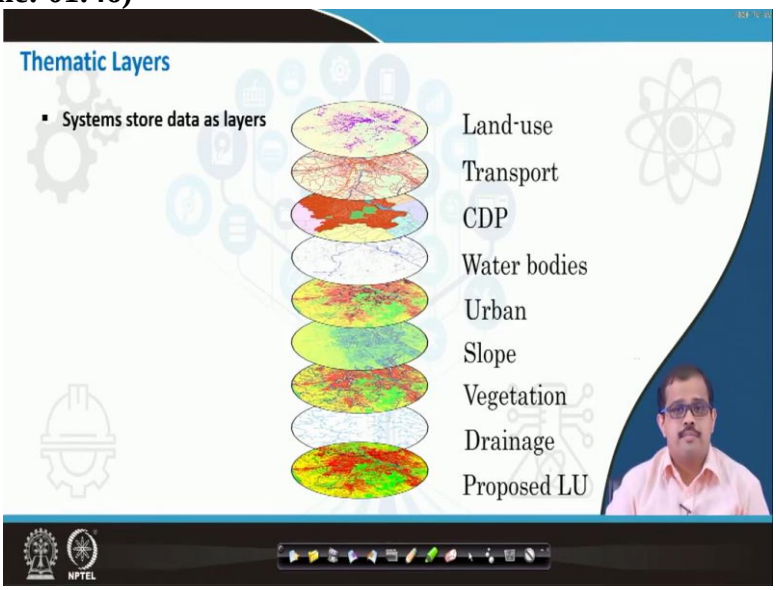
So, in this particular presentation, I are lecture, I would look at the thematic layers what do you mean by a thematic layer? Then what do you mean by spatial data editing?, so editing, correction errors and also omissions, if in case there are certain errors. Then we will look at the topology. How topologies actually created, topology is extremely important in terms of your data, data handling, then transformations that may be there in your entire data set, then editing and attribute data, how do you edit an attribute data? Why do you need an editing of an attribute data? So, all of these would look at in detail in this particular set of slides.

(Refer Slide Time: 01:41)



When we look at thematic layers probably in my previous lectures, I have spoke about this thematic layers.

(Refer Slide Time: 01:46)

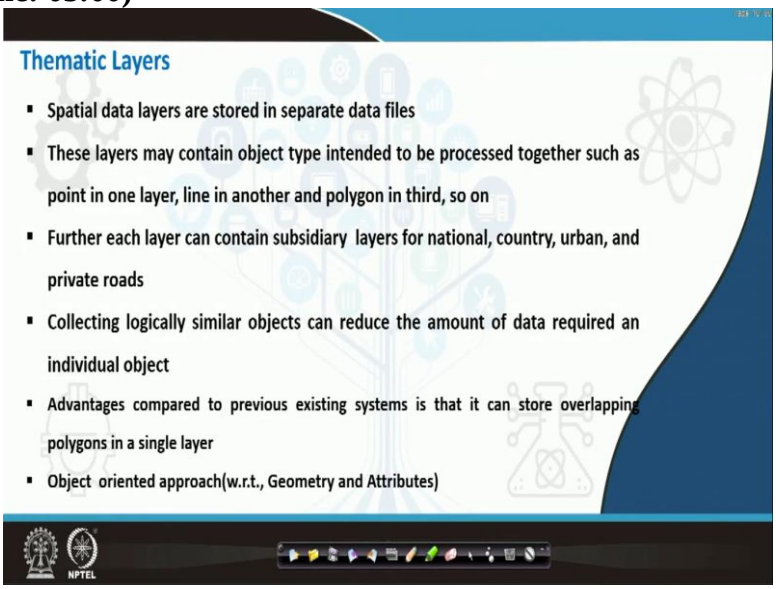


Yes, these are different thematic layers of Land use, Transport, CDP, Water bodies. Each of these are thematic layers. Why? Why are these called thematic layers? If you ask me, then for example, let us say that this is one layer of land use that have referred here. So now, I have produced a theme which means that it is representing a certain set of value representative values, which are easily visualizable. That is nothing but a theme.

Now, if I have read, in that particular data, if my in my legend red refers to urban, in my legend that green refers to vegetation, blue refers to water body. Some that is what is called a thematic layer. We are representing a theme based on that layer the data that has been

generated. So that the visualization is easy, the identification is easy and it is easy to convey message to the people or persons or to the organizations in a much easier way rather than writing it in terms of text. So, this is what is nothing but a thematic layers.

(Refer Slide Time: 03:00)



Now, when we look at any of the database, it will have several layers in terms of separate data files. Each of these layers will be stored as a separate data file. So, when we are looking at these layers, it may contain object type intended to be processed together. I mean when you have 2, 3 layers. This may have different entities which are supposed to be processed together or it may have points in one layer line in the other layer and a polygon in the third layer.

Ok, that may also happen and so on. So, when we are looking at this when you want to process. For example, you have collected only point data in a single layer, and you have collected line data in a single layer and you have collected polygon data in the other layer. So, you did not want to mix-up, when you are looking at the field data collection. So, but this has to be processed together that is where your thematic map thematic layers, that presentation becomes extremely useful.

For each layer can also if we consider that each layer has subsidiary layers for national, country, urban and private roads. There is another assumption that, for example I am making. Now collecting logically similar objects can reduce the amount of data required on individual object. So, these what, why do, I say explained previously in my lecture thematic maps is to give you an advantage to understand overlapping objects.

And to derive meaningful information without having to read the huge text that may be there underneath. So, now if you have so many layers, within a click of a button, I can understand where the new proposal of industry has to come, just play a query data set. Because, I have all the layers that is related to that particular city. I have just given a sample layers it may be under the 100, 100 and 20 layers which are generally had to be generated for a particular city modeling.

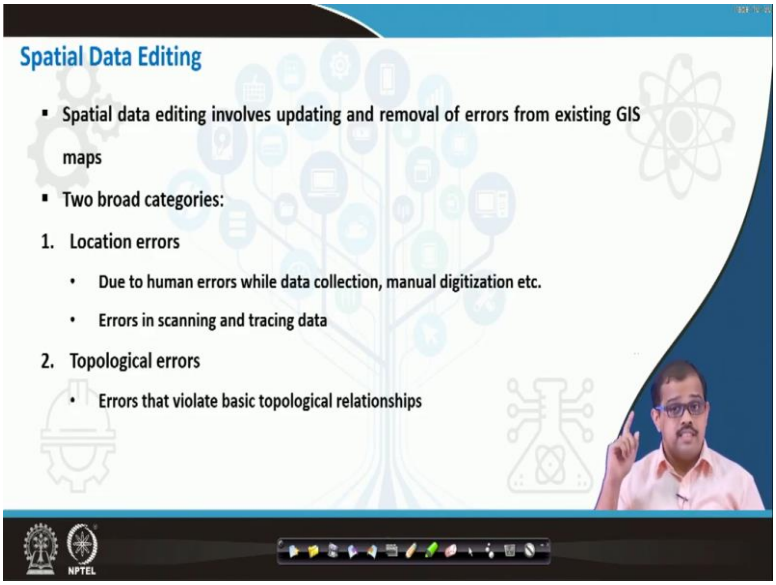
So, if you look at the entire data set of stored in a database, which is quite voluminous, now with just a query, you will be able to understand. And more importantly, this query decrease down for final. The final map is the sum or whatever the functions that use and its output as a thematic map. Now, when I get a thematic map, let us say I assume normally the theme to be RGB are representing the regions that are saturated and be representing the region.

Normally can have a greenfield or brownfield development or let us say greenfield development. So, now, I want to find out a region where actually this industry has to come. Now, what I do, I can drill down all the data find out the final map and just give it a color shades. Now, we once I give up to color shade, anyone even without a knowledge of GIS can understand by my representation that the red cannot be used.

Because I would have put the labels also, that the red are the ways places where no industry or any of the buildings has to come up in that region. For certain aspects that I would have considered I would have considered the air temperature, the air pollutants and pollutants in the atmosphere all of this stuff. But if that the same person can also understand where there is blue, it is the region where the development can happen and where there is green is extremely the region which is in between blue and red.

Where it is as sensitive as possible and the question has to be considered in terms of map in development. So, this is exclusive of those layers which are environmentally sensitive. So, I am not speaking about. So, if you have a layer of environmentally sensitive layer, so probably you can use that also to make a decision. So it is easy to find out a decision. So, always thematic layer representation and the digging down of the data is based on object oriented approach. That is based on geometry and attributes, I spoke about this.

(Refer Slide Time: 07:23)



So when that is about the thematic layer, will I, In our practical class, we will generate a thematic layer, I would suggest everyone, I have already requested or more you guys to download certain data maybe from Bhuvan and try to digitize that data if possible. If not, when we start looking at the practical class, I would also send out certain notes, as a part of this practice class.

Wherein you can look at each and every step to address every issues that are that we have already spoken in terms of this subject. Now the next concept that we are looking at is spatial data editing. The spatial data editing, when I say people think that I am editing some data to add in my, the information. It is not just that, it is about it involves updating and removal of errors from the existing GIS maps.

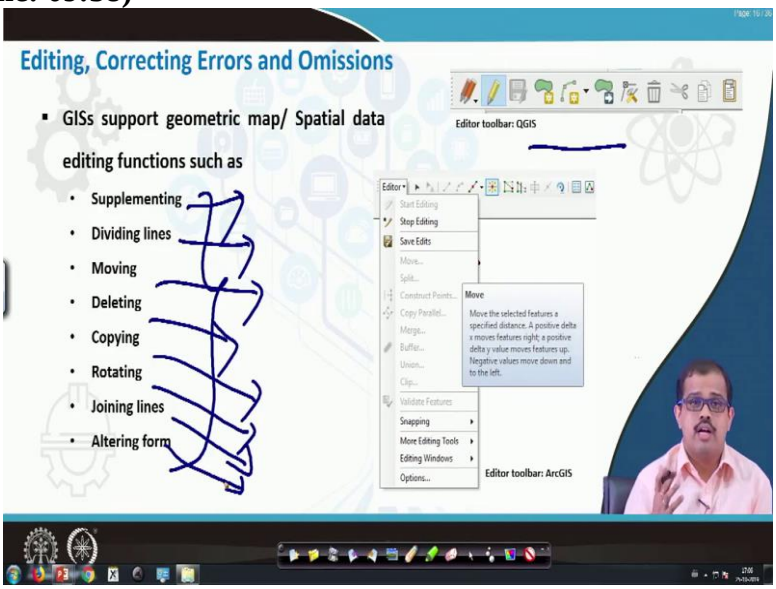
So please be careful here. I am not manipulating any data in terms of the data to be to whatever I need, it is about updating a data or removal of errors. So, that is extremely important, removal of errors as what we are more concerned about. And data updation is extremely necessary for example, a data of 5 years old has to be updated today otherwise you would not know what land parcels have changed to what regions.

So, data updation is extremely necessary. So, normally it is done in 2 broad categories, one is locational errors and a topological error. So, they have 2 kinds of errors when you are looking at spatial data editing. Especially the topological error is extremely important into in if you are looking at the topology of the space. I will come to that in the next I maybe in the later session, later the lecture, later part of the lecture.

But when you look at the locational errors, these are due to human errors, while data collection, manual digitization etc. And these are the errors which also can get in terms of data acquisition maybe through scanning and tracing of the data or digitization of the data, whereas the topological errors are those which violate the topological relationship. So, I am not speaking about the topology map here.

So, may I when I speak about topological errors, it is because it is based on the relations that topological relations that are underlying in each of the data that the it is related with. That kind of relation the correction of those errors is also extremely important.

(Refer Slide Time: 09:58)

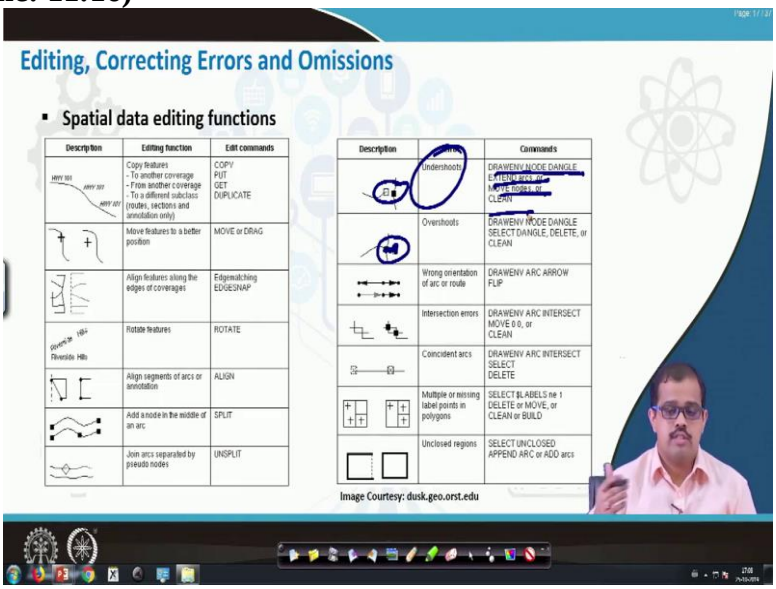


We look at how will do it. For example, when we look at how do we edit, there are a lot of GIS support packages and functions which we will be looking at it. For example, this is a screenshot in QGIS, which is giving which is showing you an editable toolbar, which here, which gives you an options of how to start editing it. So for example, if content editor in QGIS is here you can start editing it you have a lot of functions that are listed here.

For example, supplementing, you are supplementing the data that is already existing and adding in more information. Then you are dividing lines in case there are certain errors with that, then you are moving data from one table to the other table or you are deleting one data. I mean in case you have certain errors or copying certain data rotating, rotating in terms of the entire polygon set or a line set, then joining lines, then altering forms.

So, these are different ways of looking at editing this had just sample functions that we have represented. But there are huge number of functions wherein you can look at editing, correcting errors and looking at the omission part of it.

(Refer Slide Time: 11:16)



So, when we look at the sum of these when we really take up the last 2 weeks of the practical sessions, it may be last 2 or 3 weeks, where we look at the practical part of it, we will teach you all of these aspects. Now, when we look at spatial editing functions for example, you have a function which is called as copy functions, you have certain commands related to it. For example, these are certain commands which are helpful in terms of these functions to be implemented.

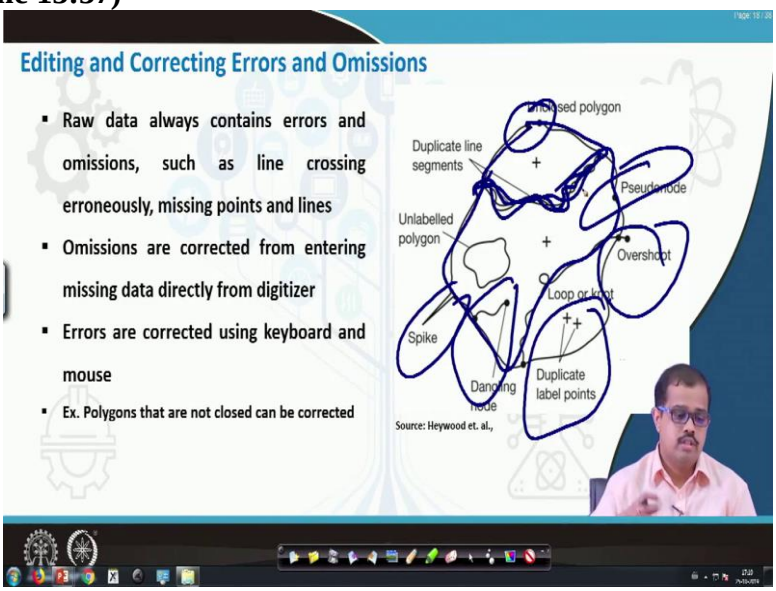
So, let us say there we discussed there are there may be certain errors, there may be huge errors in terms of a vector data creation. So, if let us say that I have an undershoot. So, what you will do is that you would you will DRAWENV NODE DANGLE EXTEND arcs or MOVE nodes, or CLEAN. So, these are the different commands which can help you that in order to connect these 2 under undershoots or this can be done even with a single click, just by a single toolbox.

There may be overshoots, that you can see here, there is an overshooted it has to be here, but this pointer has overshoots here. So, what basically is that is it first draws a NODE DANGLE, then selected, select that particular dangled, deleted, this extra note dangled will be deleted, then it will be cleaned, so that you do not have any extra line or any line segment

that is already there. So these are certain functions that are there in terms of when you are trying to edit it.

So, when you really go into this toolbox will probably understand what are the different functions and how do you do it. If you if someone is interested in coding rather than in using these toolboxes, the most of these software have a support to python, so you can have python programming and run whatever the toolbox is that just call those tool boxes to the python, python scripts and use those toolboxes the way you want to use it. So, that is also possible. So, we look at some of these toolboxes when we look at the practical part of it.

(Refer Slide Time 13:37)



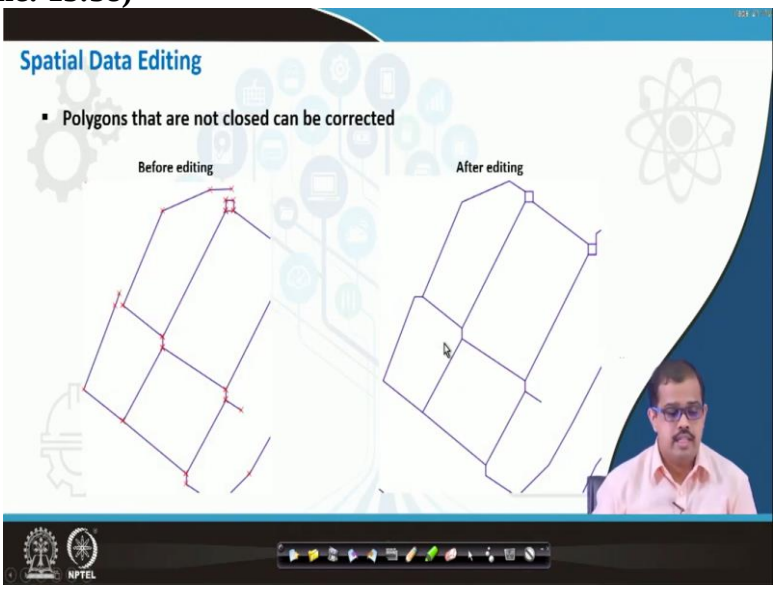
So, when we look at the raw data, it always contains errors for example, and unclosed polygons, then you have certain errors where you have this a duplicate line segment. Probably someone had distances first, but someone who again distance that they may have created it or it may be this is one boundary that was created first and this is the second boundary that was created next. So, now because it has 2 boundaries.

So, it may have a duplicate boundary or it may be having again 2 different boundaries in this particular context, which is erroneous. So, such errors has to be corrected, for example, you have a dangling note here, you have a spike here, you have duplicate level points, see the same thing is called as a loop, same thing is called as knot, you have overshoot here. So, you have a node that is not required here, which is also called as a pseudo node. So, all of these are errors that are actually there and in this particular database, so these errors have to be corrected.

For example, omissions are normally corrected by entering the missing data directly from the digitizers. If you are digitizing user using your computer, you enter, you click on it, double click on it, it opens your edit table and you start adding the omitted data. If there are errors like something is an overshoot. That also you can do just by editing it and deleting that particular overshoot and putting out the node back to its particular location. So this is what I just wanted to say about editing and correcting the errors.

Especially this image is a true representation of what different errors may occur. So this gives your picture in your mind what are the different errors that may have, that may you may have in your entire data. So, these are the error that has to be actually removed from your end from your data that you may have.

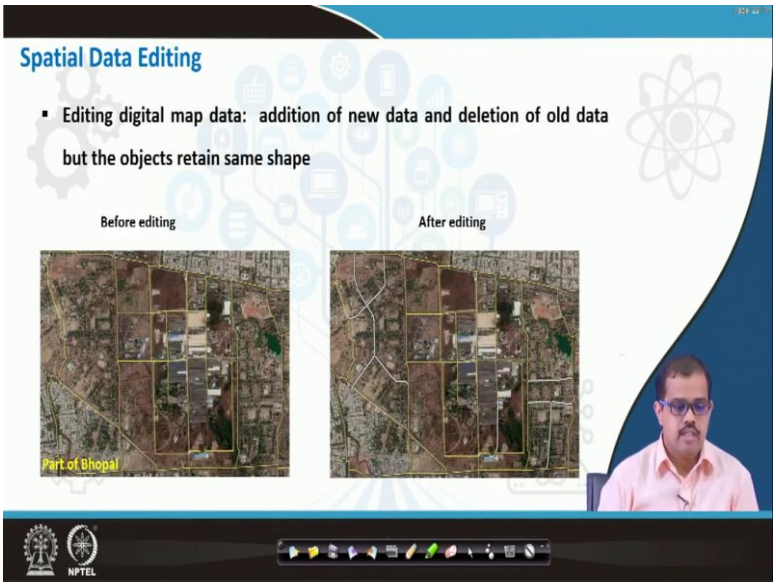
(Refer Slide Time: 15:38)



So, let us look at the next part of spatial data editing, where for example, this particular image, whatever is represented here, you can see here the polygons are not closed. So, when you are actually correcting this particular, you can just double click on it, this particular polygon computation has to be done with this particular node. So, if this node and this node, what is the error that is calculated and this node is shifted here, it becomes something like this.

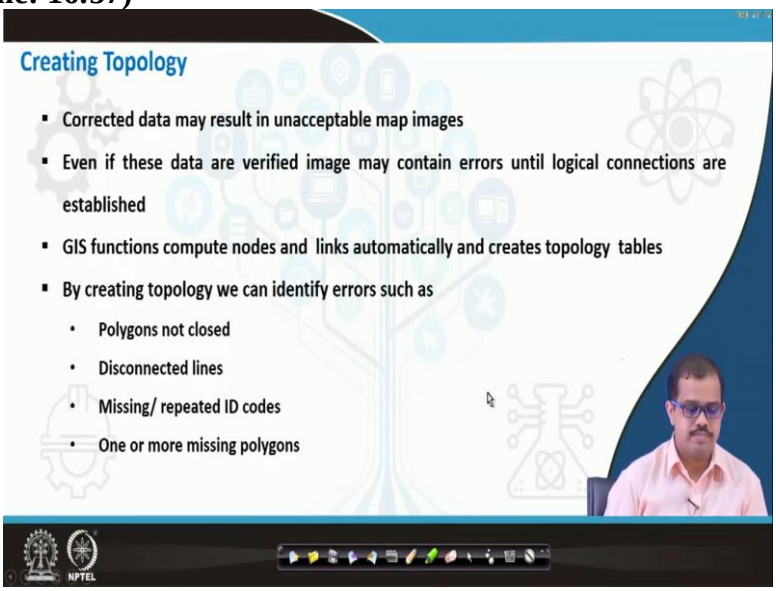
Then here also the entire thing is missing, the node is missing and that is what you do with the error correction. I am just giving an example. So that you understand what do you mean by an editing? How do you edit a particular data?

(Refer Slide Time: 16:26)



Some more examples of editing. So, before editing you can see that there are some more shapes here that are actually missing. So, these are the shapes that are added. If you see this, these are the lines, this particular road is missing here, the roads are missing in this entire region and the roads are missing this entire region. So, that is what we did with editing of the data we added those roads into that particular image that we have considered.

(Refer Slide Time: 16:57)



This is about, the first type of errors. The second type of error is topology. When you look at many of those even after corrections, these images may not be acceptable to the GIS system. Ok. Even if you have verified images, it may not be extremely acceptable to the GIS connection. Because, it lacks the logical connections, logical correctness that is that has to be

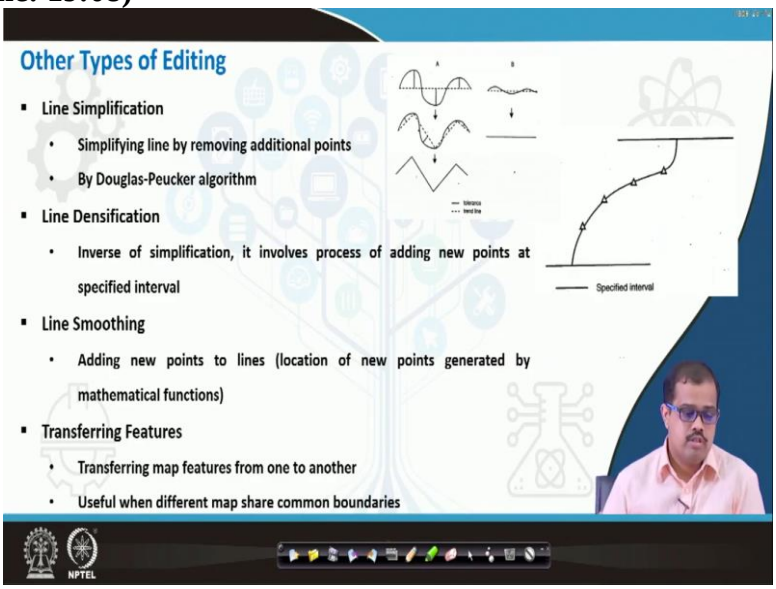
there the entire database, that is missing, that is exactly why we may have an error in such data.

So, what we will do is that we compute nodes and link automatically to create this topology tables. So, probably when you start working on the data you will the first thing that you will do is look at the topology and define the topology table. If you have look at the topology table, then, the entire if the topology table is built and if it is correct then only the GIS system will be able to accept your data in the form that is submitted.

So, by creating a topology what else you can look at. If you are looking at in terms of what may be helpful to you, the basic help that you will get us, the if you can easily look at which are the polygons that are not closed. If there are disconnected lines, connected and corrected, if there are missing our repeated ID codes. So, I gave an example in the previous, where you have 2 codes in the same polygon.

So, if there are repeated ID codes, also you can delete 1 of those codes, then you have 1 or more missing polygons, polylines or if there are missing points. So, all of this can be easily coded, if once a topology is built. So, topology gives you exact connections between the relationships and also the logical consistency in both spatial data and also in the attribute data. That is extremely important in terms of maintaining your entire data and the database.

(Refer Slide Time: 19:08)



So, then there are various other types of editing. So, for example, the lines simplification is one type of editing, where simplifying line by removing additional points. For example, or if

you look at this, their example a here that I have shown. You have this particular thing like this,. But when you look at the real scenario, this may be much useful if you have, something like this, this is a tolerance that we consider and with this be added a thread line. That becomes your line with minimal nodes.

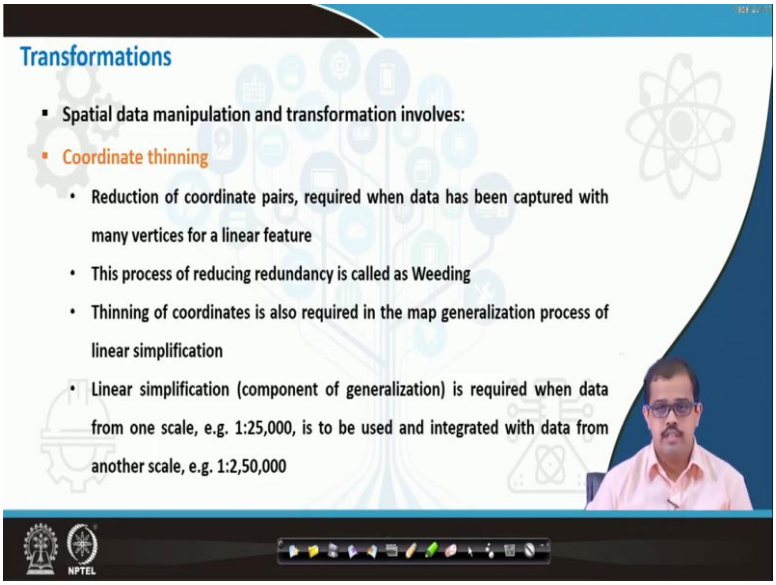
Ok. So, that may be used with removing the additional points and very well known methods of doing this, this is using a Douglas Peucker algorithm. Douglas Peucker algorithm is very well known in terms where you have to do a line simplification. There is also terms called as line densification, which means you are densifying that particular line. So, when it is the process of adding new points, so that you make the line more smoother, more better in terms of the connecting topologies.

If that also can be that also is one way of looking at the correction. So, it is dependent on user, the kind of line simplification, line densification, what kind of thing he needs into in terms of editing. So, it is dependent to the place, in dependence on the data, it is dependent on the user. Then lines smoothing, adding new points to the line, so that it becomes more smoother in terms of the mathematical function, this new most of these new points are generated through mathematical functions.

Then you have transferring features. So, transferring features from one to the other, one feature to the other features. And these are useful when you have common fear common boundaries. For example, if let us say I have a polygon, I have a polygon something like this. There is another polygon which is something like this, which is actually representing this feature.

So now, if I have, already created a data for this particularly not the same region is already threatened, instead of having a duplicate polygon here, duplicate line segment here, this feature can be easily transferred. That is what I defined here as a transferring features. So, that also can be done in terms of GIS data. So, these are some of the editing features.

(Refer Slide Time: 21:47)

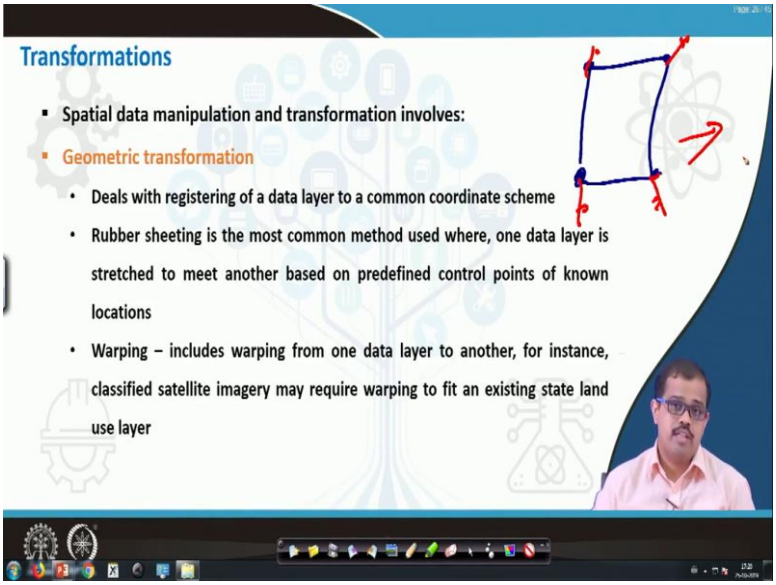


The next thing that we have to look at is transformations. So, when I am speaking of a transformation, I am not actually speaking about transforming the image. I am speaking about certain issues which we need transformations. So, spatial data manipulation and transformation actually involves the first one is coordinate thinning. The reduction of coordinate pairs required when the data has been captured with many vertices for line features.

If you have a single line feature. For example, if this is the line feature that you are trying to use. So, in line feature can be just this. So, in case you have a lot of coordinate information in between this, so, then you may have to simplify it for better representation. Then this process also reduces the redundancy in data and is normally called as reading away the data sets. Which, means that the certain information that are not required for a particular line. You read away that information or you remove that information, then it is called reading.

Then thinning of coordinates also required in terms of map generalization process of linear simplification, ok. Linear simplification is required when data from one scale that is for example, if I consider as 1:25,000 scale is to be used and integrated with the data from 1:250000 image. So, coarser image versus now high detailed image. So, if you are trying to do that then you need a linear simplification of features that has to be done. So that is one way of transformation. First is reading it out and looking at the linear simplification.

(Refer Slide Time: 23:38)

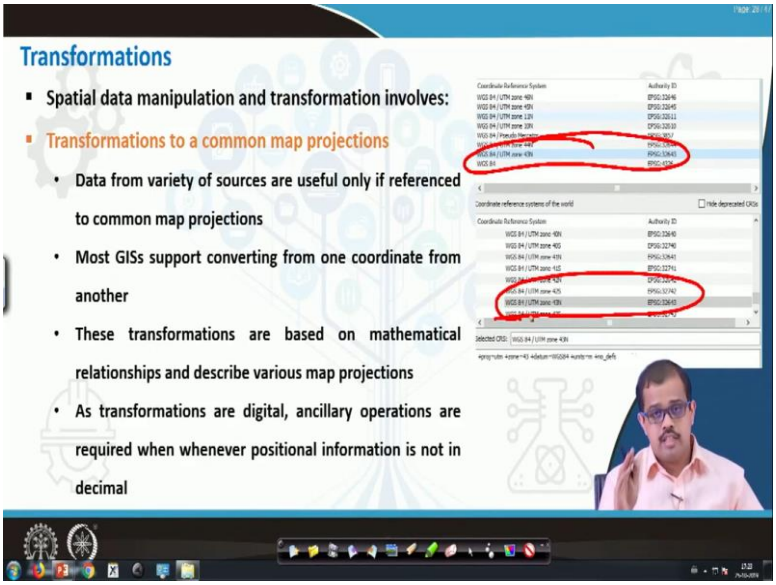


The other the second method is geometric transformations. So, geometry transformation deals with registering of a data layer to a common coordinate system. So, I have the I am repeating it again and again this is essentially extremely important. This is where most of the students make a mistake. Rubber sheeting is the most common method used where one data layer a stretch to meet another based on predefined control points.

So, for example, let us say I have one data layer is here, ok. Now I know what is a predefined, there are predefined control points, ok. These are the 4 points for this data layer. Now I will say that in my other map. I have a predefined control point which is here. So, what you do is you stretch this map or you tie these 4 points into this location and match it to the other map that is there, that is nothing but your geometric transformation.

So, normally as I say defined here this is called as a warping. Warping from one data layer to another for instance using in a classified map is very important in terms of matching 2 data and also correcting the whatever the features that are there in that particular image. This kind of method is normally used. So, there are several ways of looking at it. So, this is one of the ways of tying it out using any of the polynomial transformations or any of the other transformations and you put it into I mean corrected form. So, this is about geometric transformation.

(Refer Slide Time: 25:19)



The next thing is transformation to common map projection. Now, as I said, coordinate, common coordinate, common datum, common projection. So, these are, the things that are special data always looks at. So always keep it in a common projection system. So, for example, I have a data later say that I am building up the urban database. Now, I have certain data from google maps. I have certain data from open street maps, has certain data from one, I have certain data, that I have captured it from my own field service.

There are certain data from survey of India, there are some more data that I have collected from various gazetteers of the government sources and various other district books, census books, etc, captured. So now all of these books has to be reference to a single projection, which means this is how we see a 3d map in a 2d world. The once you project it in a proper projection then only we will be able to identify the 2d the point the entity on the 2d layer in a much accurate way.

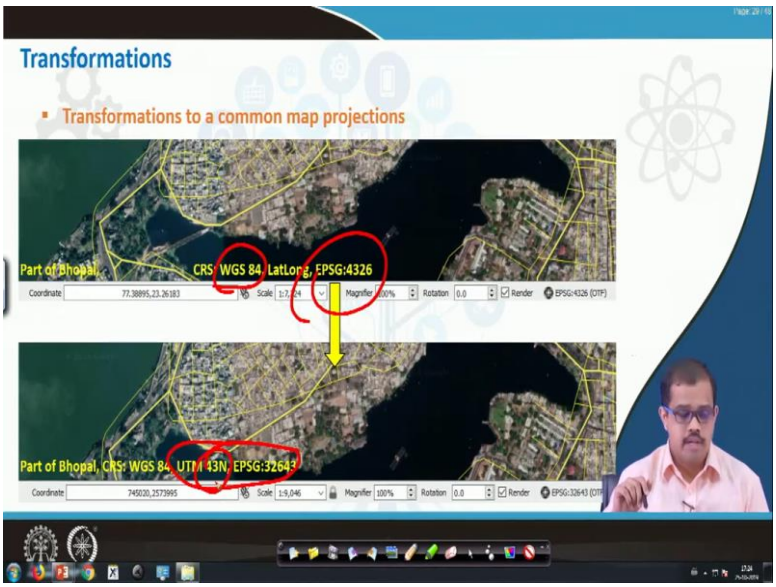
So, when you look at the GIS software, most of the GIS software supports converting one coordinate system to other, one projection to other and one datum to the other. So these transformation are based on mathematical relationship and describe various map projections and transformations are digital ancillary operations are required when wherever the positional information is not a decimal. For example, here, I am trying to this is 1 example, where I am trying to use a WGS 84 UTM zone 43 north, where EPSG code is 32643.

So, I am trying to transform that into a reference manual into our reference coordinate system. Which is WGS 84 43 north and 4386. This may be represented in a wrong way. But I

am trying to transform it into another system of projection. So, that is where you are for the other transformation is important. So, please whenever you are looking at it, please look at the geometry of objects in your database, please look at the coordinate system, please look at the datum and please look at the projections.

So all if all of these are there, then the next set is how you find out how what are the errors in your data, if you are able to understand all of this, then most of the errors, most of the issues in your entire data collection or data manipulation and data handling is solved. So, now this is about transformations to a common map projection.

(Refer Slide Time: 28:13)

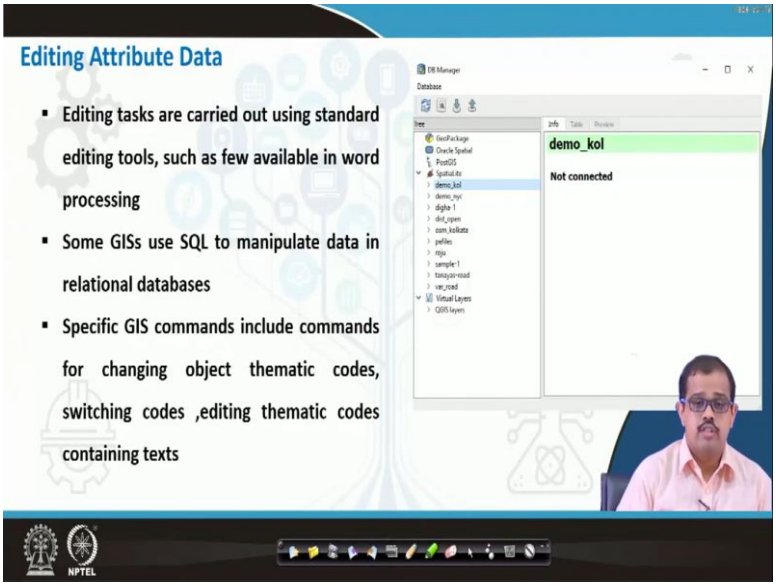


Then if you look at this is what I was referring to EPSG code here is 4326. We go back. So EPSG code here is 4326, and I am using a CRS, that is WGS 84 LatLong this is a part of Bhopal. So now I will convert into a common format that whatever the format I may need. So, what I do is I convert into EPSG code where it is UTM 43 north. I hope everyone knows that India is above the equator.

This is a common another common mistake that most of the people do, if I ask, where does India lie is above the equator, below the equator, most of them say it as below the equator. So it is and it is in the northern part of the equator. So, when you are looking at it, you have to, specially in this particular projection system you are referring it in terms of 43 north. So, WGS 84 is the datum that we have used here.

So, please remember this on how we are transforming from one location to other location. So, this is just an example to show you the transformation. So, exact transformation happens you can see this is the same data set that has been transformed and the data is intact.

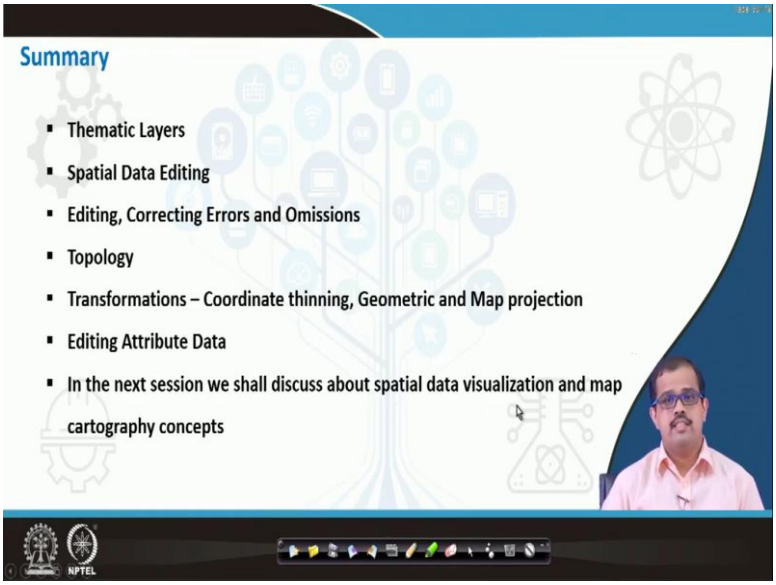
(Refer Slide Time: 29:39)



So then editing the attribute data. Also editing tasks are carried out using a standard editing tools such as a few available in even in the word processing. Most of the GIS uses the special query language. So we will also look at special query languages in our next slide. Specific GIS commands include commands for changing object thematic code, switching codes, editing thematic codes, containing different texts, etc. So these are different issues in handling the database, which we look at in the database.

When I am speaking about the database and how we built a database, how to maintain a database, how to query a database. So until then, let us think about that, there are tools for editing any attribute data in terms of creating a GIS database.

(Refer Slide Time: 30:30)



To summarize this particular class, we looked at thematic layers. Thematic layers is a visual representation in order to understand whatever the data has been here or developed and what is the output of using a certain queries in the entire data set or it may be the output of certain statistical and mathematical operations. Finally, that is given a color, color that is a respective to its representation. So, that in order to visualize. So, each color represents a different visual aspect on the surface.

Then spatial data query editing, spatial data editing is about how we edit data. So, what kind of editing is there, where there are errors and how to correct it we looked at different errors also. Then we looked at topology and topological corrections, topological errors, then transformations, we looked at coordinate training, geometric transformation, map projection, transformation, and coordinate transformation. So all of these transformations we looked. The last part, I just gave you an hint about editing attribute data.

So, in the next session, we will look at how do we visualize the data? Ok, how do we visualize the data and what are the different objects that you, you should have in your database to visualize the data, so that is about the map cartographic concepts and cartographic aspects. So we look at it and we will also look at what are the different ways of representing different objects, the same symbol cannot be used for many.

So you need to have different symbol sets, there are different representation will look at all of these different sets, when we come back in the next class. Till then, have a nice time. Thank you very much.