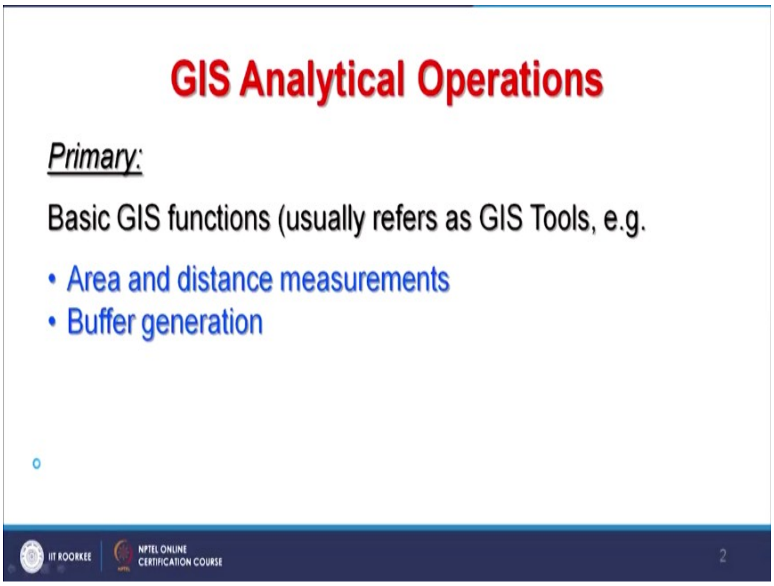
**Geographic Information Systems**
**Prof. A. K. Saraf**
**Department of Earth Sciences**
**Indian Institute of Technology – Roorkee**

**Lecture - 14**
**Pre-processing of spatial datasets - 01**

Hello everyone! and welcome to new discussion which is preprocessing of spatial datasets. And here this discussion is in 2 parts. First, we will discuss the basic types and then we will go for some advance preprocessing of spatial data. So, basically these analytical operations in GIS, there are various ways these can be divided. Some people would like to you know divide in as primary and secondary operations.

Some people will divide in some different ways. So, in the literature you may find little bit variations but basically it is in 2 parts that the basic operations which we try to do under this primary operation category. And that includes basically the basic functions usually refers that GIS tools we will be using just to simply measure area, distance. You know along line for distance, area for a polygon or perimeter. Maybe buffer generation along a line or a point or even a polygon.

**(Refer Slide Time: 01:37)**

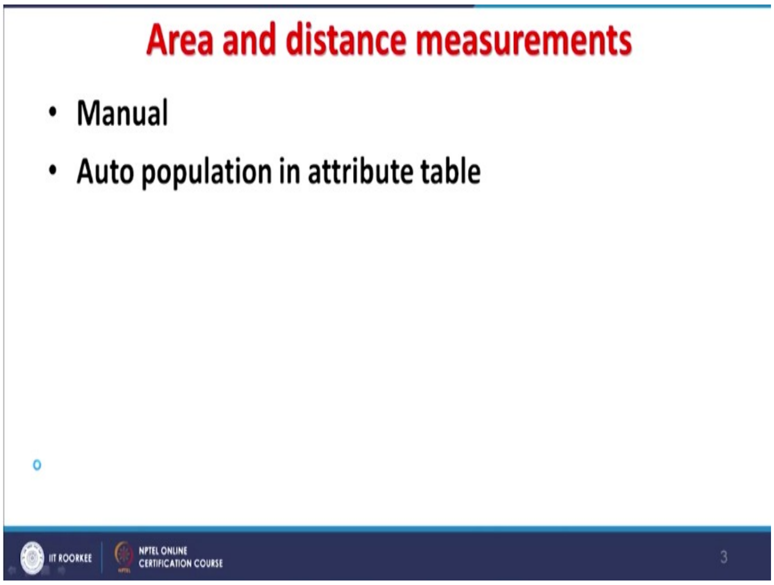



Buffer part we have not discussed yet but in future, we would be definitely covering that part. So, buffer generation is also one of the analytical operations and then reclassification etcetera. Reclassification because generally suppose if we are having a continuous data and now, we want to reclassify it into different categories.

For example, I am having a digital elevation model in which the elevation values are say varying between you know 100 to 1,000 meters. So now, I want to classify that map instead of displaying or using as a continuous, I can classify say 100 to 200, 201 to 300 and so on so forth. So, depending on my requirements I can reclassify. So, this part, we will be also discussing how to classify and reclassify.

Now, when we go for in detail about the area or distance measurements then there can be 2 methods; one is the manual method by which we can do it or auto population in attribute table because these tools are also available. So, suppose if I am having polyline theme as one of my themes in the GIS database.

**(Refer Slide Time: 03:01)**

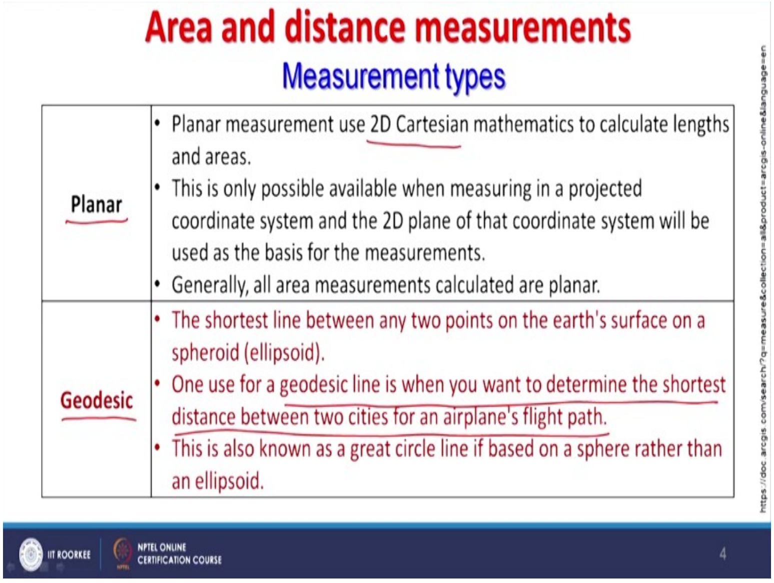



And now I want to have a field which will have the length of each segment. So that can be done automatically and this information would be residing in the non spatial database. Same I can do it for polygon like I can go for calculation of you know 3 parameters associated with polygon maybe area, perimeter and centroid. And all these things can also be done in a completely automatic manner.

Only thing which one has to remember that when you go for such auto population, please check for random. Do the random checks on certain polygons and see whether the area is coming all right or not because every time you cannot completely depend on the software. Software maybe accurate but if our coordinate system is wrong or something wrong with our units then results will also be inaccurate.

So, in order to avoid that thing, it is always better to check and recheck about the data. Now, there are different measurement types.

**(Refer Slide Time: 04:20)**



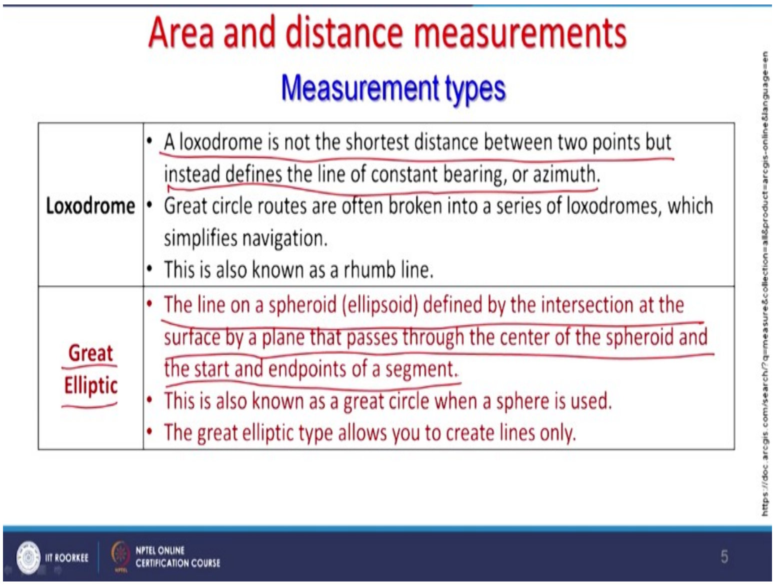

We can have one kind of measurement which are planar measurement and then another measurement geodesic measurement. Now planar measurement means basically doing the measurements in 2D. For example, like we use the survey toposheets or any other map which are in 2D Cartesian system and therefore you know, calculating area, perimeter or distances becomes much easier and accurate as well.

But when we go for the 3D then for like geodesic measurements or in 3D measurements then things become little complicated. And especially in the hilly terrain if somebody is using a digital elevation model or even a satellite image of a hilly terrain then one has to take care while doing the area measurement, planimeter you know whether that is planimeter measurement or a terrain measurement.

Sometimes we also call as terrain measurement. For example, like one use of a geodesic line is when you want to determine the shortest distance between 2 cities for an airplane flight or etcetera. So, depending on our requirements, we can choose the measurement type. And of course, depending on what type of data input is there.

Whether data has come from a 3D surface that means a part of the sphere; curved surface or data has come from a 2D map like survey toposheets or some other maps. So, depending on that we should do. And the same you know, advice here that after each and every operation in GIS, please check for errors. Now there are some other measurement types which can also be explored for you know, measurements like loxodrome.

**(Refer Slide Time: 06:21)**



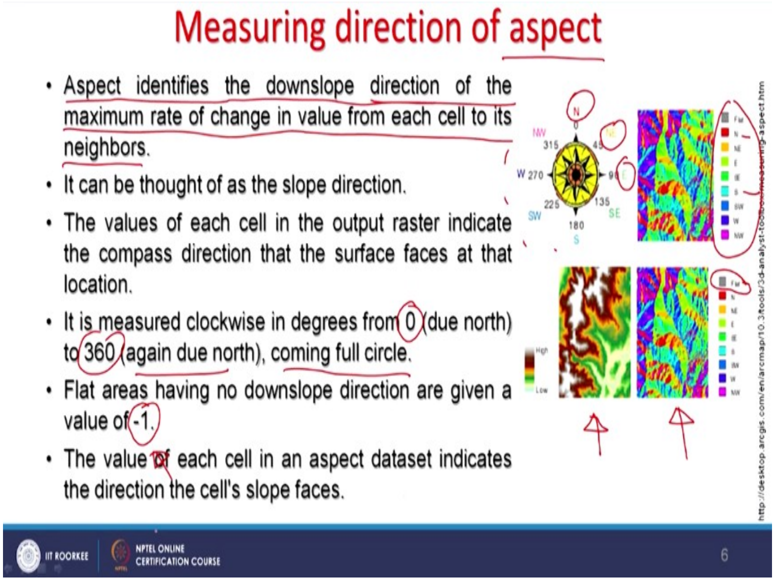

Loxodrome measurement; loxodrome is not a shortest distance between 2 points but instead defines the line of constant bearing or azimuth. So, you know instead of going through a curve if it is in a straight line, many of such distances are used mainly for aircrafts or other things or even maybe for missiles as well.

So, you know if distance is too large that means the curvature of the earth will play very important role and in what projection system, you are having your input maps that will also depend. So, like in this great elliptic that a line on spheroid or ellipsoid defined by intersection at the surface by a plane that passes through the center of the spheroid and start and endpoints of the segment that means you cannot just handle like a 2D map.

You have to take care about these things if the distances are very large. If distances are small then the curvature of the earth will not play very important role. So, if you are going for measurements even in 2D, assuming that everything in 2D may work very well. But if distances are large then all such cares are very much required. Because you know error propagates in GIS.

And therefore, we must take care about all these things in the beginning. Now there are other measurement types like for example calculating the aspect of a sloping surface and that is basically, the direction of that sloping surface with reference to the north.

**(Refer Slide Time: 08:17)**




So, this aspect for which you require input as a digital elevation model or DEM. And as you know that aspects basically identify the downslope direction of the maximum rate of the change in value from each cell to its neighbor's. So, if it is a sloping surface then what is the orientation of the sloping surface or direction with reference to north, that is what aspect is.

Here you are seeing examples also that this is input digital elevation model and this is the aspect which has been calculated for this. And aspects generally in by default, you can calculate in 8 direction starting from north, northeast and then east and likewise. And this is what you are seeing here also. And one category will always come which is for the flat. If your terrain is flat then obviously it will not have slope.

And obviously, it will not have aspect that means the orientation of the slope. So that category will always be there if that kind of features are present within your digital elevation model. Now aspect is very useful parameter which comes through digital elevation models in many kinds of modeling or analysis like somebody is working for soil erosion studies. So, they would like to see the aspects of different sloping surface.

Because maybe the rainfall may not be occurring uniformly in that terrain. Maybe generally, what happens like in Himalayan conditions rainfall occurs mainly on the southern facing slopes. Similarly, maybe the snowfall like for example if I take example of Mussoorie ridge then generally more rainfall occurs on the south facing slope which is towards the Dehradun rather than north facing slope which is towards the Uttarkashi.

So, if there is more rainfall and these are also called the sun facing slope; the south facing slopes that means there are more chances of soil erosion. And that may be also true in case of landslides also because most of the human activities in hilly regions especially in the cold areas are concentrated on the southern facing slope because it receives the sun from morning to evening.

And therefore, these slopes are more vulnerable. If you observe the satellite images, you may find that the north facing slopes are generally full of vegetation or forest but south facing slopes are devoid of forests or very little vegetation you may find because of human activities, because of the sun, because of the you know the soil erosion and because of the landslides.

So, these things; the direction of sloping surface with reference to the north is very important. And it is very useful also. People also use in you know like those who are working in hydrology or prediction of discharge in a stream which is snow fed or glacier fed then again, the aspect of a slope is very important because that receives the sun, more melting will occur on the south facing slopes.

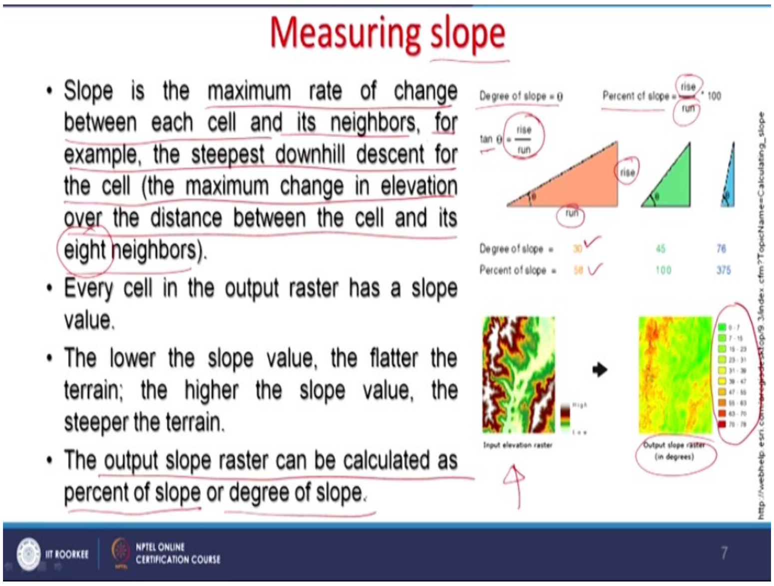
So therefore, this parameter which is derived from digital elevation model has got lot of applications in natural sciences including in civil engineering or sciences environment etcetera, even in social sciences. And as you know that it is measured in degrees from north to again go back to the north. Starting north with the 0 value and then of course, 360 degree again due north, coming full circle.

So, this is how measure but you can address you know instead of 0, it is addressed as north facing slopes and then maybe the south facing slopes, northeast, southwest and in all 8 directions, this is done. And as I have already mentioned that flat areas in output image will have a value of minus 1 that indicates that is a flat area. It does not have any sloping surface. And therefore, it does not have the aspect.

So, this is you can call as a no data value just to indicate to the system that this is a flat area. So whenever use, you must check all these things before you go for the next step. Suppose I have used the digital elevation model, I have tried aspect. And now I should check whether I am getting correct results or not. And value of each cell in aspect dataset indicates that the direction of the cells slopes faces.

Because this is after all, raster data and each cell will have its value about the aspect. So, this is how it is handled in GIS.

**(Refer Slide Time: 13:46)**




Now another very important parameter. More detailed discussion, we will have on aspect and slope when we will be discussing different derivatives of digital elevation model in the later part of this course. But for the time being and completeness from measurement point of view, we are discussing here also in very brief. So, after aspect calculations or measurements, we go for the slope calculations.

One thing that slope can be calculated in you know 2 ways; either you can calculate slopes in terms of degree and or in terms of your percentage. That means rescaling instead of 0 to 90, you rescale to 0 to 100. Most of the GIS software's will support to display the values in both ways that means in degrees as well as in percentage.

You know theoretically, what is basically slope is that rise over run that this is the rise and this is a simple you know trigonometry. That rise over run is there, that this much rise is here and this is run. So, if I want to calculate in percentage then I have to multiply rise over run multiply by 100, I get the answer in percentage. If I want to measure in degree basically, it is $\tan\theta$ and that is rise over run and by which I can get in degrees.

Now if I go for degree slopes like in this example, I am getting this is 30 or in terms of percentage, because it is rescaled so, I am getting 58. And likewise, different values I can get it and depending on how you are trying. But same digital elevation model as in aspect case I took. Now if I drive a slope map for that using the digital elevation model, this is how it is going to look.

Now it depends on me to how to display my output map. I can have classifications like 0 to 7, 7 to 15 and so on or I can have just few classes. And basically, this method is called reclassification that we will discuss little later. So, how I want the output? I can control through the classification techniques or reclassification. So basically, slope is the maximum rate of change between each cell and its neighbours.

Because it is considering what is the neighbouring cells are having values. If neighbouring cells are having say lower value, whichever will have the lowest value in among all 8 directions of a central cell of a digital elevation model then whichever is having the minimum, there slope will be there or the vector will be drawn.

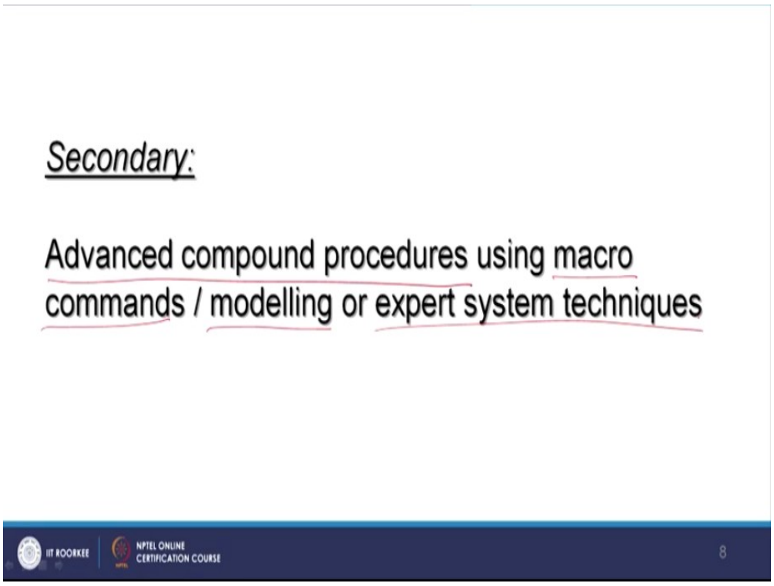
So, for example the steepest downslope descent for the cell that maximum change in elevation over the distance between the cell and its 8 neighbours. 8 neighbours are considered that means north, south, east, west, northeast, southeast, southwest and northwest. So, all 8 neighbours are always considered in case of slope measurements also. Every cell in the output raster that is your slope theme, has a basically slope value.

And it will carry the slope value. So, if you click on any of these cells then you will get the value whichever the cell is bearing. Of course, if lower the slope value, the flatter the terrain or higher the slope value, the steepest the terrain. So, these will definitely indicate. It is a very useful again like aspect, this too has got slope parameters. Slope parameter is which is derivative of a digital elevation model is, again very important parameter for various studies which I had discussed earlier with aspect.

The same way, they are also in the soil erosion, maybe in the landslides, maybe for recreation, maybe for you know many applications where slope data or slope inputs are required in the analysis. And of course, the output slope raster can be calculated as percentage or in degree of slope. These options will always be available in most of the good GIS software. So, these were basically primary analytical operations which we perform on a GIS database once the database is ready.

We may perform some secondary. And what are those secondaries? These are the advanced compound procedures.

**(Refer Slide Time: 18:53)**

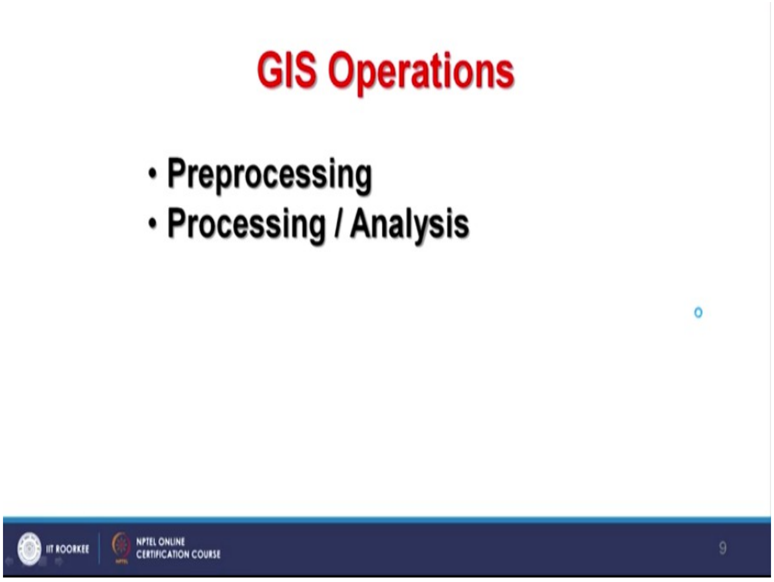



Sometimes we may use the macro commands or modelling or expert system techniques. It depends on again our requirements. Because what we are having after installation of GIS on a hardware, what we are having a system which is without any data. Database has to be generated by us for our requirements. And once that is there then we have to find how to use that data for our project; our purposes.

And for doing that purpose, we may go for very advanced compound procedures, modelling or maybe implying artificial intelligence or expert system, any other thing. So, this is one way of classifying GIS analytical operations that means in primary or secondary category. In some literature you may find people divide in these 2 like the title of this particular discussion was the preprocessing.

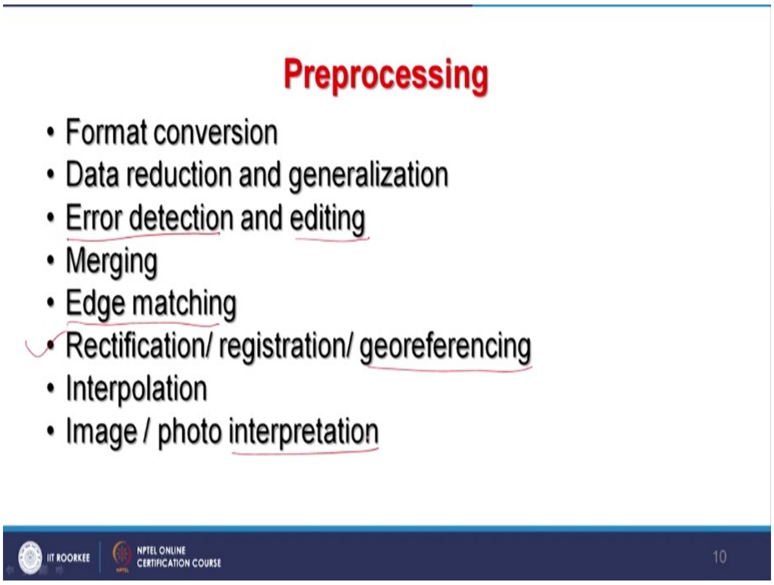
So, preprocessing that means preparing your spatial database for real analysis. So, it is basically equivalent to your primary operation. And then another one is called processing.

**(Refer Slide Time: 20:10)**




It is not post processing. So, this is processing or real analysis which involves your advance or compound analytical operations which may imply which you may go for modelling or any expert system and other things. So, what we consider in preprocessing? Some of like in primary, we have considered preprocessing.

**(Refer Slide Time: 20:33)**




Because I am covering both way of classifying our analytical operation because in literature you may find in some book or some site, you may find little differently in some way or other. So, I thought that I will cover both way of classifying. So, in this kind of preprocessing, the first is format conversion. Format conversion maybe from analog to digital. That is also considered under format conversion.

Format conversion maybe from DXF which is generated by AutoCAD or design file which is generated by MicroStation into the GIS that it means in the shapefile or a shapefile in case of ArcGIS or some other softwares. But most of the GIS softwares owners' respects or incorporate shapefiles. So, there is no issue. So, from either from DXF to shapefile or you know, your design file to shapefile.

That may be also come, this kind of operation. So, preprocessing steps may come under this format conversion. If I am having raster data, there also I may convert from format like I may get the data in TIFF, I may convert into geoTIFF means I am having now geographic coordinates or I may convert from geoTIFF to IMG which is a ERDAS IMAGINE format for raster data which is very you know compatible with like ArcGIS or many other softwares.

So, in case of vector data, shapefiles have become almost universally acceptable for vector data. And same with the raster data, IMG files or TIFF files have become universally acceptable formats for raster data. So many times, we get the data in different formats and we try to convert the data. Only thing which one has to remember while converting from one format to another, all these formats are not truly transparent conversions.

They may introduce some errors. Like for example DXF to shapefile. There may be because DXF may not carry your attribute files. There are shapefiles carries your attribute. So, you have to be very careful while converting it and must check with the results. So, that statement always true that after each and every step check for errors; check for accuracy. How accurately the conversion has been done? If it is satisfactory, go to the next step.

Now next is the data reduction and generalization. Data reduction does not mean that we will delete the data but it means that we would like to perform some redundancy checks or do some generalization. We will have more discussion in this lecture, little later. And then error detection and editing. After each and every step, check for errors. That is the error detection. And if found, do the editing, correct it and then move for next.

Like if I have digitized something. Now my topology has not been constructed properly. So, I should check for errors, construct the topology properly and then go for the next. Merging; merging means you know by if I am having 2 maps and side by side, they are like 2 toposheets having adjacent numbers then I want to join them that is also called merging. But in raster data, merging has got different meaning.

So, we will be discussing this also in detail. And then edge matching. If you are having 2 adjacent toposheets data then sometime because of projection issues. these edges that the data on the boundary may not match very perfectly automatically. So, sometimes manual edge matching is also required. Then rectification, registration and georeferencing, that we have already discussed this part under this georeferencing category in detail.

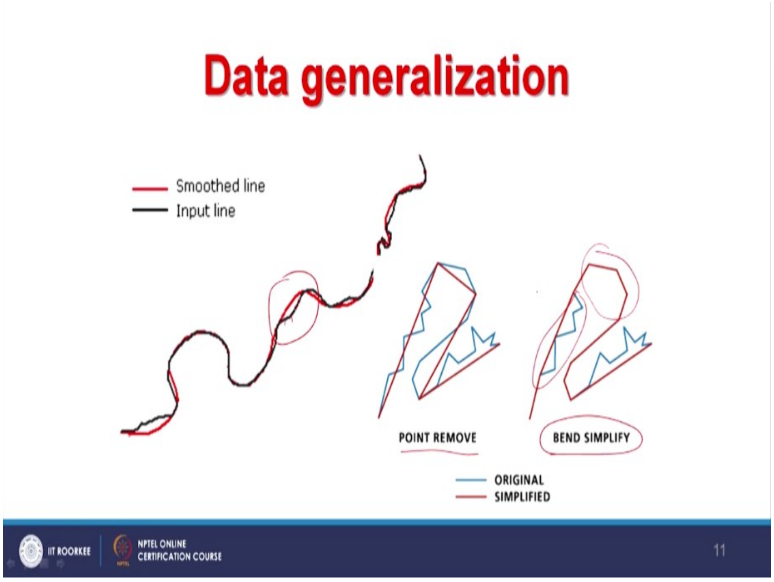
We had about an hour discussion on this. So, this part is already covered that is also considered under primary operations or preprocessing. Interpolation, we have not yet discussed. So, interpolation also comes in this preprocessing category. Basically, interpolation is that from discrete data you would like to prepare as a continuous data through interpolation techniques.

Major input in interpolation technique is your point data and rest the techniques which you choose and other parameters. This is another very important preprocessing step image or photo interpretation. Unfortunately, you know in recent years, people have started giving less emphasis on image interpretation.

Whereas in my opinion through my experience, I find that it is very-2 important that one should analyze the images; satellite data or aerial photographs and then perform image interpretations as well. Do not leave everything for the machines to do it because machines do not have intelligence, only humans are having better intelligence compared to machines. So, we should apply our interpretation keys and do the interpretation.

That will give us a very good results in our projects. So, first I will take the data generalization or data reduction.

**(Refer Slide Time: 26:26)**



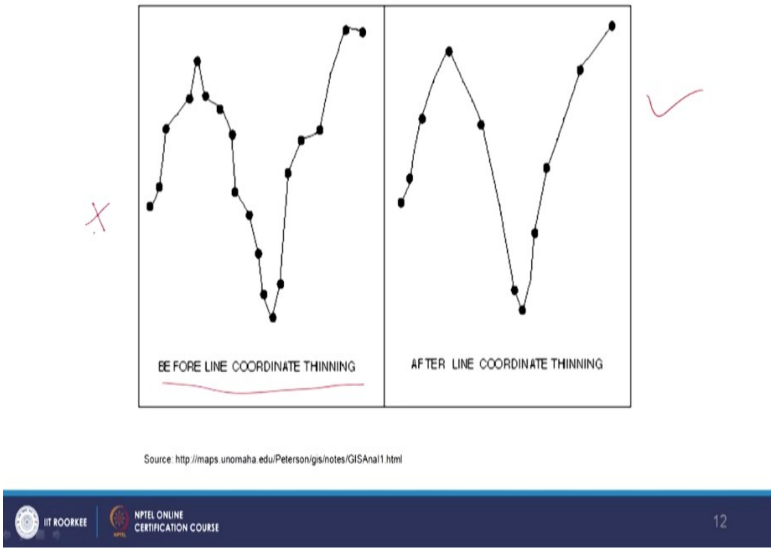

Here if I take the example this on the left side, the red lines which you are seeing here are not the original data. The black lines are the original data, red lines are the smoothed line. And when you would make a smoothed line that means you are removing you know certain errors from here and then you are removing some nodes from here or internodes from your polyline.

And as you can see here that you know this black line or polyline has gone near or has become like red line. So, if there were nodes many nodes here, they have all disappeared and only you are having now curved line. So, generalization will may smoothen your line polylines. Sometime it is very much required to perform. Therefore, these tools are supported in GIS.

Another example here like this is point removal. Point removal means node removal. So, the blue line is your original data and while digitizing, it has created very zigzag appearance. Whereas I do not want that kind of detailed data I want very simplified one, so, simplified one is the red one which you are seeing. So, the same way if I want more accurate representation of bends based on that data, it can be done.

So, bend has been preserved of original. But rest here things have been removed. And while things are getting removed, I am removing the nodes and internodes. And while doing this thing, I am reducing the data that is why data reduction and same time I am achieving generalization.

**(Refer Slide Time: 28:23)**




Another example, this is before line coordinate thinning. Some softwares in some literature, they may use word coordinate thinning. Coordinate thinning means again removing internodes from your vector data or polyline. So here you are seeing that lot of nodes are there. But here only the important nodes have been kept and small-2 nodes which are not that important from generalization point of view, are preserved or kept here. Rest have been deleted.

Again, one advice here is never delete your original vector file because vector data is expensive to create or to buy. So, vector data is you know does not require much space also as compared to raster data. So, you may perform data reduction or generalization or coordinate thinning over your vector data. But always keep the backup of your original vector. Do not delete.

You may use this right one which is generalized but you must keep this one in your database all the time.

**(Refer Slide Time: 29:39)**

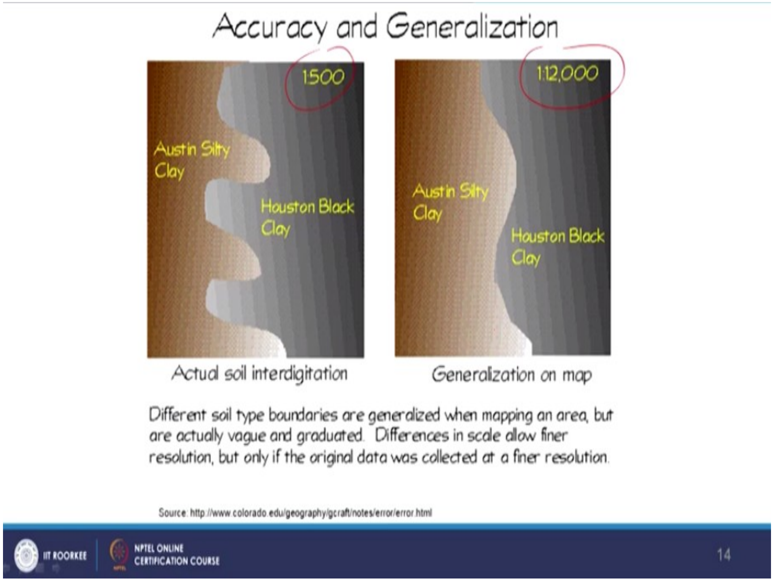

Now, more or less a real example here that like if you scan a toposheet or once it has been digitized, all layers are here, it looks cluttered. I am not interested in that detail information for my project. So, what I want just simple you know road network, rail network, some streams.

So, only 3 themes or 4 themes have been kept here. Water bodies are also here then you know, water bodies include the streams then the rail network and then roads have been here whereas there were lot of other information's which were not required for my project. So, those have been removed from my database. But again, this file must be kept and then this file should be used.

Otherwise, you know creating vector data is really very expensive. So that way, what we achieve data reduction as well as generalization. Our maps may not look cluttered because here lot of information which is not required for my project output. Many times, we find lot of details available along the coastal boundary. Now I do not want those much details in my map.

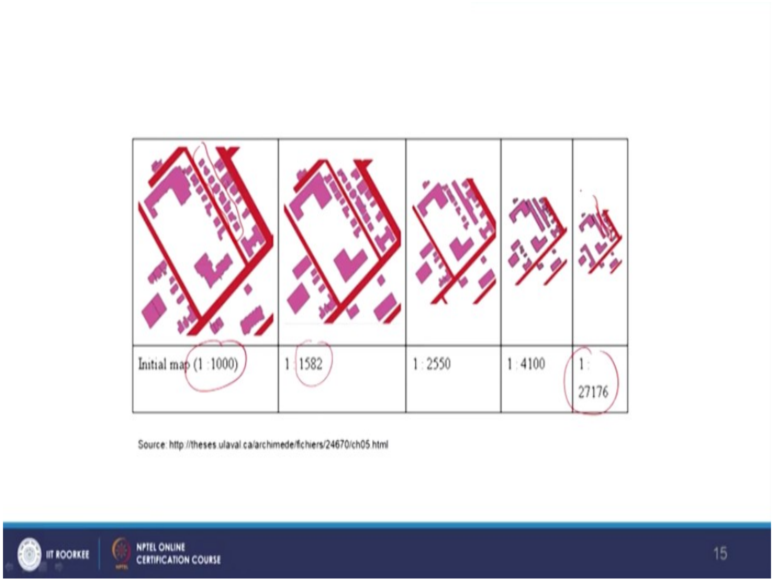
I want more smoother appearance of my map. So, I may go for generalization. Of course, when you attempt generalization, you are compromising with accuracy.

**(Refer Slide Time: 31:10)**




And that is why I am saying do not delete your original data that may be more accurate than your generalized data. So always have a backup of that one and then you can perform and use it. So, generalization depend on again scale; scale also matters. So, at this scale, all details about the coastal lines are here. But at this scale, those details I do not want. So, I will perform the generalization.

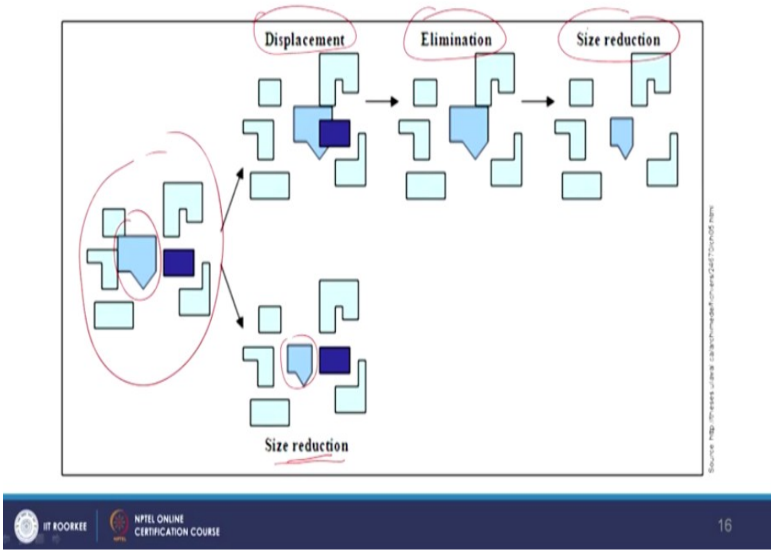
**(Refer Slide Time: 31:42)**




So many such things are there like here depending again on scale. So, this is initially scale 1:2000. Now when I go for smaller and smaller scale that means this value on the right side of the ratio increases then it becomes smaller scale. So, if I go for this, I do not want so many details. And see these buildings which were shown here, all have been merged with a single line because at that scale, those details are not required.

But if I keep then the map will look very cluttered and when it is not required, why to keep it. So, generalization or data reduction is performed depending on my output targets, output scale and based on my project requirements and analysis. So, generalization must be performed.

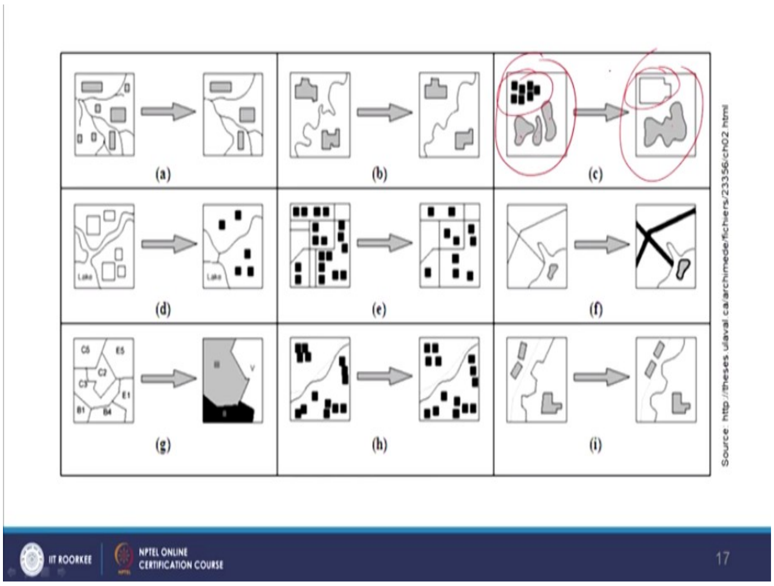
**(Refer Slide Time: 32:38)**




Another very good example of that this is my input data. Now I can do the size reduction something like this that everything has been reduced especially if you focus on this blue block like this or I can do the displacement. I can go for elimination, again the size reduction. So, there are various ways people have implemented for generalization because there cannot be one fixed base.

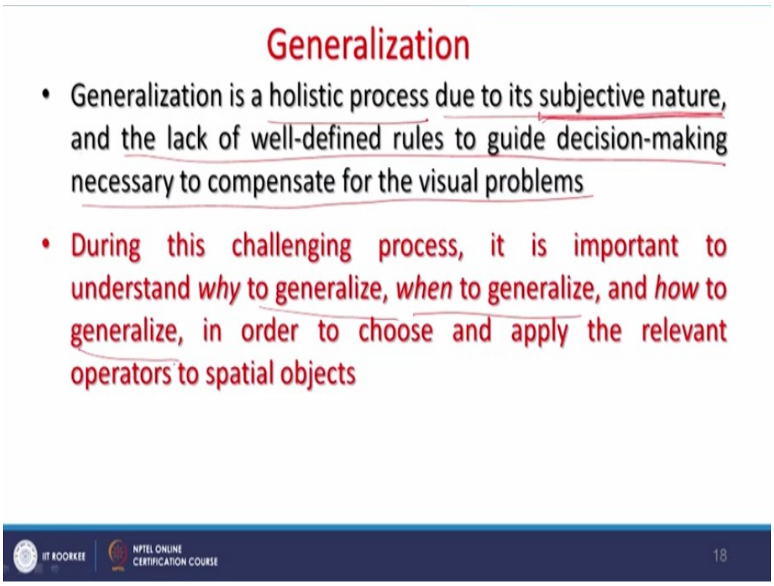
So mathematicians have you know, always venturing in these areas and they try to give us new solutions for generalization or other operations.

**(Refer Slide Time: 33:20)**




Many such examples are here. This is if I take this one then sees, it got generalized. All these 3 water bodies if I say have been merged into one at this scale and same way these houses, they have been merged as one block. So, depending on the data, depending on my requirements of output scale, I may perform generalization. Now generalization if I want to you know, define then is a holistic process due to its subjective nature.

**(Refer Slide Time: 33:56)**




And because it is the subject, how it will be merged as I have shown in previous example like this one, whether generalization I will go for displacement, elimination or size reduction or size reduction in this way. So that is why I am saying that in this definition that it is subjective nature and the lack of well defined rules to guide the decision making necessary to compensate for the visual problem.

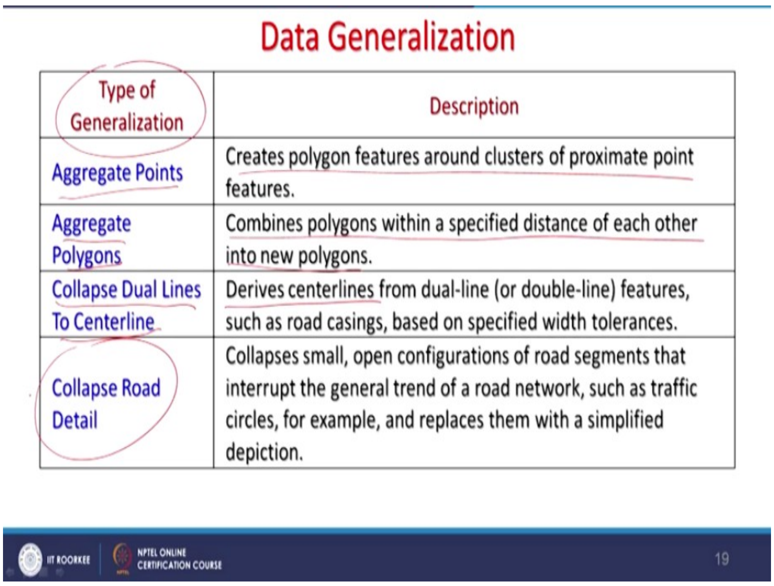
So, if you are going for generalization of a polyline then not many problems will be there. But if you go for generalization of polygons then because of the subjective nature, some problems may arise. So, one has to be careful about that. Otherwise, there is no you know well defined rules exist in GIS. Depending on how GIS developers has implemented these generalization rules.

And this is sometimes you may find a very challenging process to understand how to generalize and when to generalize and why to generalize. Why to generalize? As I have said if change in a scale is there, I am going for large scale to smaller scale. I do not want too many details; I will go for generalization. When to generalize is important. One should not go in the beginning of a project analysis or modelling but later on, one can go for that one.

And how to generalize will remain a question especially with polygon data but polyline data, that is not a big issue. And these operations differ in terms of accuracy, quickness and complexity. You know means how efficient these generalization algorithms which have been implemented and accuracy. Accuracy part, I have already discussed that must check.

Once the generalization is done, must check whether it is giving satisfactory results as per your project requirements or not. If yes, fantastic, if not then you go and try to try with some other options if they are available.

**(Refer Slide Time: 36:18)**



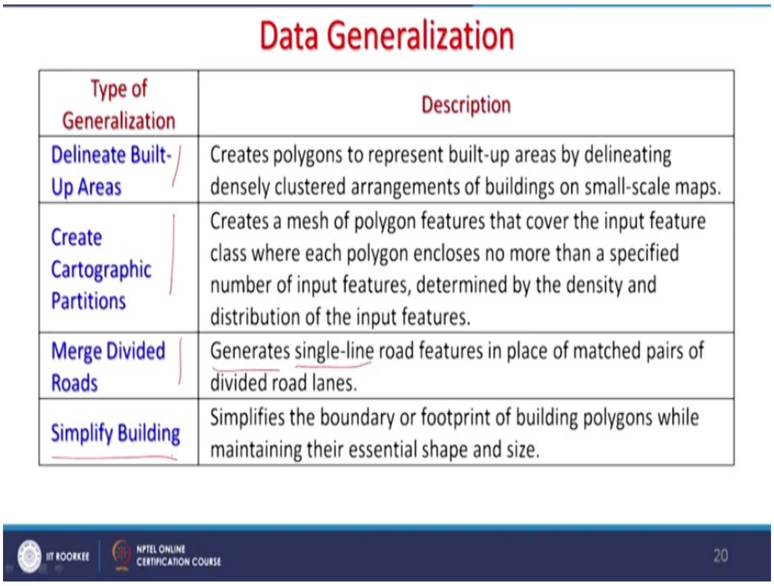
Data Generalization

| Type of Generalization | Description |
|---|---|
| Aggregate Points | Creates polygon features around clusters of proximate point features. |
| Aggregate Polygons | Combines polygons within a specified distance of each other into new polygons. |
| Collapse Dual Lines To Centerline | Derives centerlines from dual-line (or double-line) features, such as road casings, based on specified width tolerances. |
| Collapse Road Detail | Collapses small, open configurations of road segments that interrupt the general trend of a road network, such as traffic circles, for example, and replaces them with a simplified depiction. |

Now data generalization; some types of generalization which exist in theory. Many of them have also been implemented in GIS. Like you aggregate the points, create polygon features around clusters of proximate points. Aggregate polygons; we have seen some examples of, means combining polygons within a specified distance of each other into a new polygon. We have seen that example.

Now collapse dual line to the center line if it is a dual line for rail track or any other thing, you may merge into one line or drive a central line from dual line. And maybe collapse road details; too much details in a particular scale about the road are not required. So, you can get rid of them and during your generalization process.

**(Refer Slide Time: 37:11)**



Data Generalization

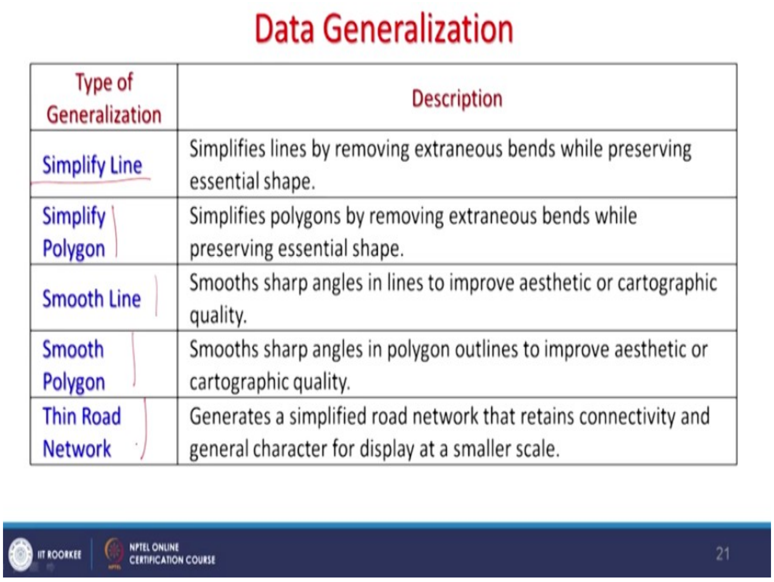
| Type of Generalization | Description |
|---|---|
| Delineate Built-Up Areas | Creates polygons to represent built-up areas by delineating densely clustered arrangements of buildings on small-scale maps. |
| Create Cartographic Partitions | Creates a mesh of polygon features that cover the input feature class where each polygon encloses no more than a specified number of input features, determined by the density and distribution of the input features. |
| Merge Divided Roads | Generates single-line road features in place of matched pairs of divided road lanes. |
| Simplify Building | Simplifies the boundary or footprint of building polygons while maintaining their essential shape and size. |

IIT ROORKEE NPTEL ONLINE CERTIFICATION COURSE 20

There are some other types of generalization like delineate built up areas, create cartographic partitions, merge divided roads and simplify. Because there may be a 2-lane road, 4-lane road, you want to merge as a single line in your output map because your requirements are just to show that there is a road. Your requirement is not to show that it's a 4-lane road or 2 lane road.

So, if it is there, you can generalize and create just single line. Or simplify buildings.; buildings will may be represented through polygons. So, you can also simplify merge into a block or a larger polygon.

**(Refer Slide Time: 37:58)**

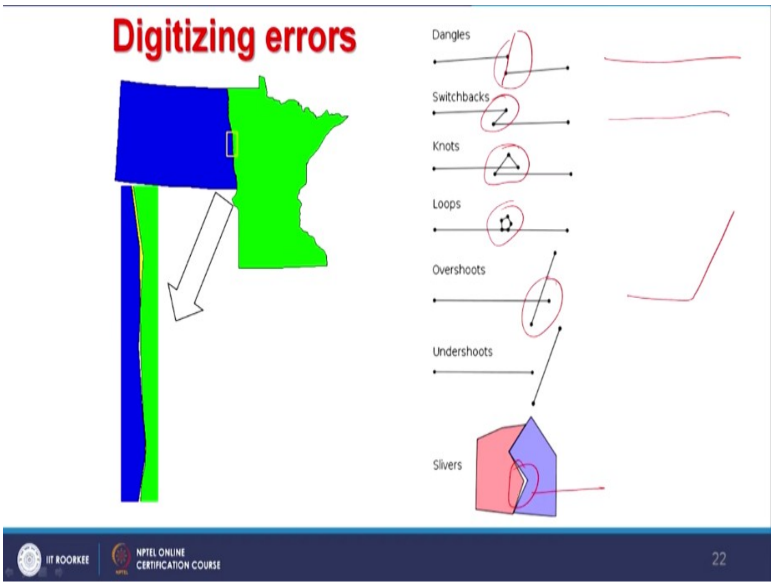


## Data Generalization

| Type of Generalization | Description |
|---|---|
| Simplify Line | Simplifies lines by removing extraneous bends while preserving essential shape. |
| Simplify Polygon | Simplifies polygons by removing extraneous bends while preserving essential shape. |
| Smooth Line | Smooths sharp angles in lines to improve aesthetic or cartographic quality. |
| Smooth Polygon | Smooths sharp angles in polygon outlines to improve aesthetic or cartographic quality. |
| Thin Road Network | Generates a simplified road network that retains connectivity and general character for display at a smaller scale. |

Now, there are other for simplifying line, for polylines, simplify polygon, smooth line. We have seen the example of a smooth line, smooth polygon if there are too many details along the polygon, we can smoothen in. Thin road network; we have already discussed this part.

**(Refer Slide Time: 38:17)**




Now one important thing when we go for digitization, how the digitization will be done? Suppose I have a scanned map say Survey of India toposheet. Now it is in the background of my screen on a GIS platform. Now I want to extract say polylines which are contours or a road network or a stream network from that.

So, while doing that digitization part means in the background, I am having a raster image which is nothing but the topographic map and, in the foreground, I am digitizing say lines or polygons. So, there may be some errors. And as I have been saying that they should be taken care about immediately at that moment. What kind of errors you may encounter while digitizing?

People have given names in literature which you will find like dangles. This line supposed to be straight line but somehow because of some extra click, there is a gap and sort of displacement which has occurred. Our geologists may call as a fault which was not intended but somehow it has been created.

Same it is the switchbacks; it is supposed to be a straight line but somehow during digitization process, error is got or maybe a knot, maybe loops, maybe overshoot like opposite to dangles, overshoots. Again, the lines should have been like this but these are now 2 separate lines and overlapping each other. If I am going for a polygon and digitization or there, I may find an overlap or a sliver that means the gap between 2 polygons.

So, this polygon was not intended. This is an artifact and this is an error also. So, this must be removed. Similarly, other errors which I have just discussed should also be removed in order to have a quality in our data because if this error continues during analysis and modelling, they may give us completely wrong results. Here I was expecting only say 2 polygons, I am now having 3 polygons.

So, the area, perimeter and centroid for the third polygon which is artifact which is due to error, will always be calculated and everything may go bad. So, one has to really take care about this. So, with this, we come to end of this discussion about preprocessing Part 1. Thank you very much.