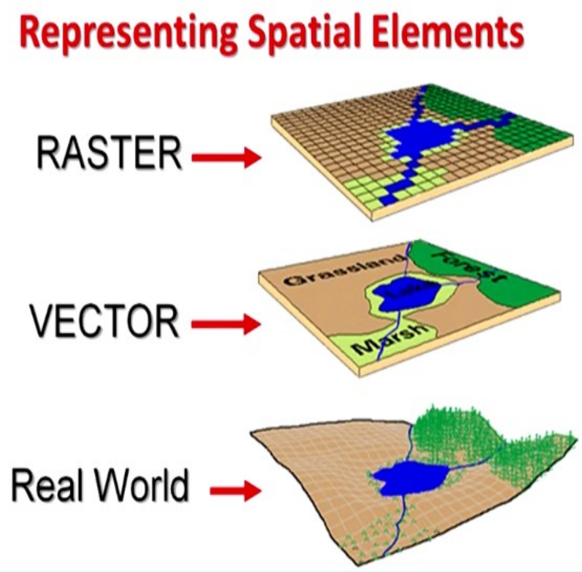
Geographic Information Systems
Prof. Arun K. Saraf
Department of Earth Sciences
Indian Institute of Technology – Roorkee

Lecture - 03
Different Types of Vector Data

Hello everyone and welcome to new discussion. In this lecture, we are going to discuss about different types of vector data. As when we have been discussing definition of GIS that time I mentioned that data in GIS comes from variety of sources and variety of formats also that is why a manipulation word is also used.

So, vector data, raster data and there are different types of vectors and raster's and also one more type is the TIN data, along with your attribute data or non spatial data. So, in this discussion, we will focus on the vector data itself. As you know that this part just to you know, connect with previous discussion is that real world which we can convert either into the vector or also in the raster like this.

(Refer Slide Time: 01:09)

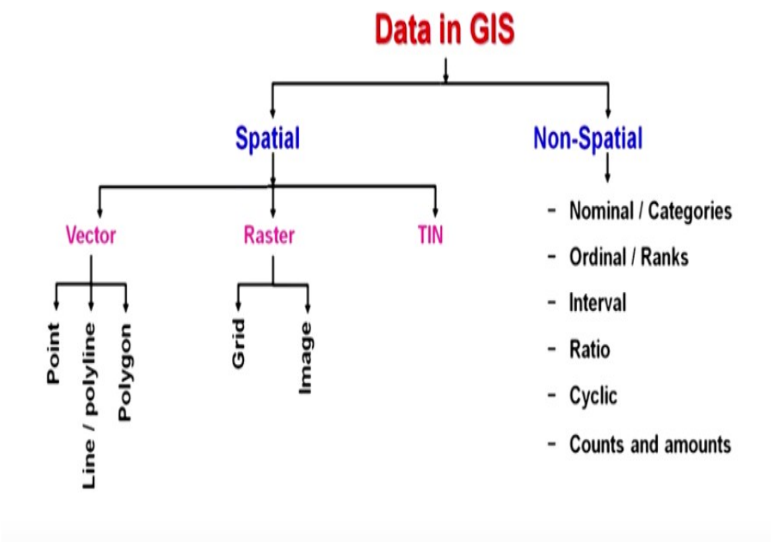


So, raster means we are having data in the grid form and raster's are 2 types whereas vectors are 3 types. So, the real world as you can see here that if I want to convert into the vector then there are 3 entities which we will also see in details that point, line/polyline and polygon. So through this, we can have a representation of the real world in the vector domain.

If you recall the history of GIS when we have been discussing so, initially the GIS were first developed to handle the vector data in the Canada that is CGIS. And later on, this raster capability also has entered into the GIS. And because of development on the Digital Elevation front which is also a one type of raster and therefore, GIS has become very powerful tool because of integration of vector, raster and one more type; the TIN data which we will discuss sometime later.

So, today our discussion is on the vector. Now, let me you know discuss everything about all types of data first in the GIS that we have also discussed this point that there are 2 types of data in GIS; major 2 types one is spatial data, another one is non spatial data. So, spatial data includes vector data, raster data and TIN data. And your non spatial data is mainly the attribute data or tabular data.

(Refer Slide Time: 03:05)



And there are different types of vectors as also mention that point, line or polyline and then polygon. So, 3 types of vectors are there. Though we have been familiar mainly with 1 type of vector that is the line vector in which we had the origin, we had the destination and this used to decide what is the direction and magnitude of vector. But here in GIS, we have point data, line data, typical vector and polygon; all 3 are considered as vector data because this is the discrete data.

This is not continuous data like contour lines. Topographic contour lines are discrete. They are representing equal heights or equal value along the line. But in between 2 contour lines, you do not have any information. So, we call a vector data as discrete data or discontinuous whereas the raster data is continuous data.

One cell of raster ends, another starts and this is continuous within that particular 2 dimensional matrix or a grid. Raster is 2 types; one is the grid type, another one is the image. There are slight differences between grid and images. So, when discussion will come on raster, we will discuss that part also. And the third type is I have already mentioned, is the TIN that is TIN stands for Triangulated Irregular Network.

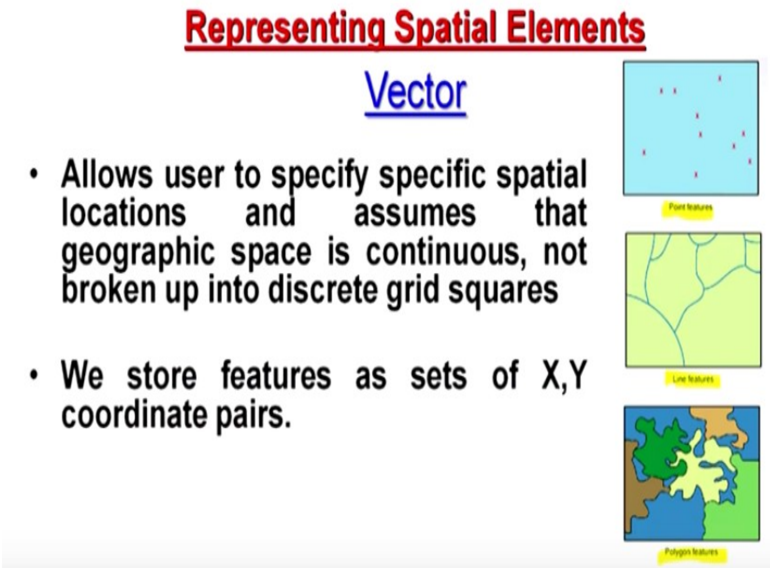
Units here in TIN are triangles whereas in raster, units are square in shape that means cells or pixels. So, if it is an image, we call as pixel. If it is a grid, we call as cell. So, the shape of the cell is always square in shape whereas in case of TIN, the shape of unit is always triangle of varying size and shapes. So, that is why it is called irregular network of triangles so, triangulated irregular network whereas in vector all are controlled by the coordinates.

The coordinates are everywhere because that is why the GIS are there to handle the geographic coordinate. Now, under this non spatial category, we are having various non spatial types of data; one is the nominal or category data, then ordinal ranks data, we will have separate discussion on this part also.

Then interval data, ratio data and cyclic data and last one is counts and amounts. So, these 6 types of non spatial data so far have been implemented into GIS. If any new development takes place in mathematical domain or computer science tomorrow, either in the representation of spatial data or non spatial data and if it is beneficial in GIS, definitely it would be incorporated in future as well.

So, just to recap the entire gamut of data in GIS that there are 2 basic types; spatial and non-spatial. And when we discuss the spatial: 3 subtypes; vector, raster and TIN whereas non spatial, they are 6 types. So, let us now focus on the vector data itself that basically what it does. It allows users to specify specific spatial location. A spatial location again here is the geographic coordinates which we have discussed.

(Refer Slide Time: 06:42)

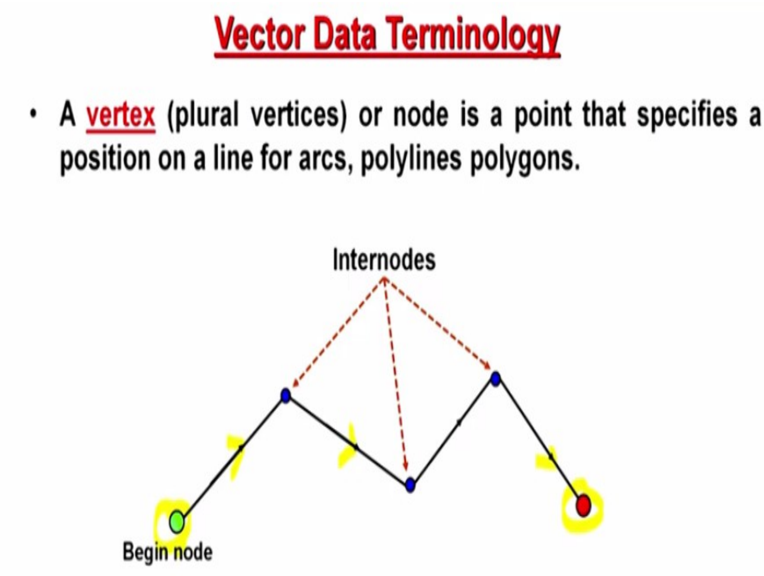


And assume that geographic space is continuous as you know, on the surface or the globe. So, it is continuous whether it is land part or water body like sea or oceans or land or any country or continent. It is all continuous throughout the globe. So, that is why the geographic space is continuous. And the same concept is also applicable in case of Moon and Mars surfaces. There also the coordinate systems have been developed.

So, like for the Earth also. So, it is specify specific spatial locations if it is a point data. It will have X and Y in geographic locations. And then you can have n number of attributes associated with that particular point data which is shown here as the point data. Then in polyline or line data, line features; we call also called as line features. There, it is a string of X, Y coordinates.

And they are there so, if we know the origin and we know the destination; we can find out the length that is that becomes magnitude as well as the direction. And whereas polyline is you know 2 dimensional vector entity where the begin points and endpoints are always same. So, it makes a complete polygon. We will see further details on this. So, in case of point data; the features or point features are stored as X, Y geographic coordinate pairs.

(Refer Slide Time: 08:06)

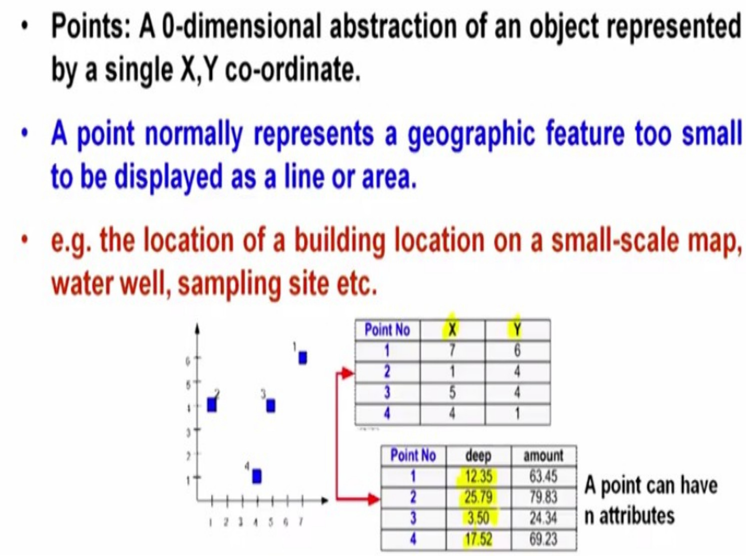


And these nodes; we also call vertex or in plural, we say vertices or node is a point that is specify a position on a line for arc, polyline polygons. So, even for point data if I take this example of a line, so, this is my begin point. This is my end point. And that is why you are seeing the arrows directions like this. So, now I know the magnitude that is the length of this line becomes my magnitude and the direction is also there.

So, these are the begin node, end node or vertex and then in between you are having internodes as well. So, we call them as internodes. So, these blue dots are shown internodes. So, when a computer or GIS platform stores a line; what it does. Basically it is storing X, Y coordinates all along the lines or of all vertices. So, it will store for begin coordinate and internodes or in between coordinates wherever the bends are there or coordinates are there and then the end one.

So, that will make this one. So this is, as I have said the begin node and the end node. Now point; we say in GIS, we consider a point as a zero dimensional vector entity. Why? Because point is nothing but just an X, Y; a pair of a single X, Y and therefore you do not have anything to measure with the point; you cannot measure. Of course, between 2 points, you can definitely measure the distance.

(Refer Slide Time: 09:35)



But I am talking about a single point. So, single point will have just 1 pair of coordinate that is X and Y. z value means that elevation value or any other attribute information will go as in non spatial data in our system. So, point is a zero dimensional entity. Therefore, we cannot calculate or estimate its area, length or anything. Even on a GIS platform if you are having a point data and if you keep zooms the point data, there will not be any size expansion of the point.

Because point does not have any dimension. So, it is only the location of a pair of coordinate. And a point normally represents a geographic feature and too small to be displayed as a line or area. Now sometimes, one may get confused about the scale. So, let me bring that discussion also here. Depending on the scale of map, suppose if I want to show the location of Roorkee within the boundary of India.

Then definitely it will come as a point because I am displaying the whole India. So, scale of my map is very small. Though Roorkee is having area; within Roorkee, IIT Roorkee also having its own area but because of smallest scale, the whole town would be shown as a just point data that means a zero dimensional entity. But you know in real sense, it does not have.

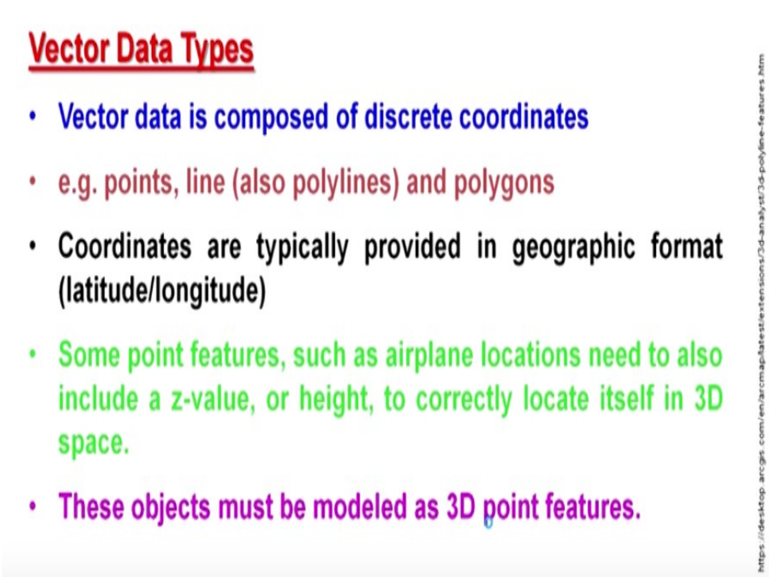
But as per its scale, it is represented. Whereas if I am displaying a map of say Haridwar district where Roorkee is part of Haridwar district and then Roorkee will show as a polygon, not as a point because now scale is large. So, Roorkee can come as an area or polyline. So, instead of zero dimensional entity then it becomes a 2 dimensional entity. So, many things about the vector data will depend on the scale.

So, one has to really understand this thing. For example, the location of a building or location on a small map, water well or sampling sites. They all will appear even a location of village or town on a small scale map will appear as point information. But we know in really, point that is not a point. It has got its own area but at a larger scale. This is one example is shown here that there are 4 points are shown in X, Y coordinate system.

And so far for simplicity, we are not showing in geographic. But in software demonstration, I would be showing in geographic coordinates to just to keep things simple. And these tables which are shown are nothing but the attribute tables. So, for like point 1, 2, 3, 4; there is 1 pair X and Y and I can have the same table and different columns can have different information or I can have a separate table. Here a separate table.

So, against point 1; what is the location of a well, what is the depth of a well or point 2, 3 and 4. So, if I am having n number of points, I will have n number of rows in my table and I can have theoretically n number of fields or columns in my table to store as much as information possible against a single point. So, here an example; four information's, we are already seeing. X, Y, deep and amount; all these things are shown here against all single points.

(Refer Slide Time: 14:04)



Now vector data types; in this discussion that vector is composed of discrete coordinates. I have already told you that vector entities in GIS; vector data are discrete, not continuous. And examples are point, line, polygons. Between 2 points if you do not have any information.

Between 2 lines, I gave the example of contours if you do not have any information. And polygons; within a polygon if the boundary is closed then you do not have any information.

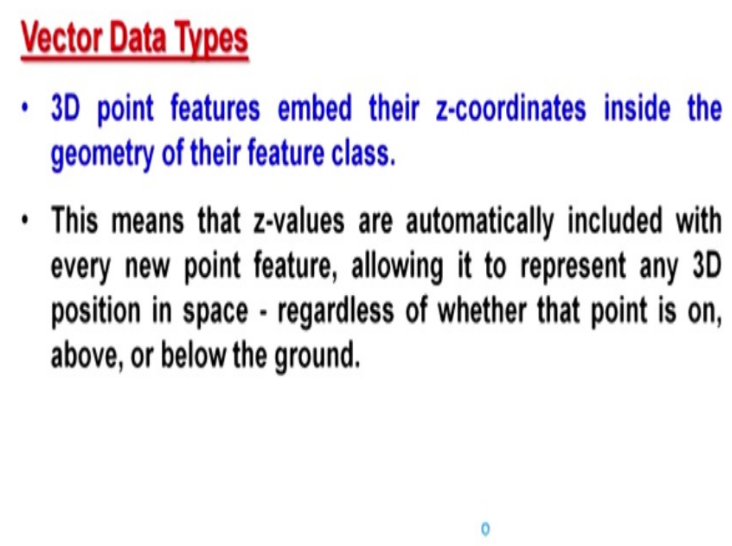
Except that you can do the measurements about the perimeter, you can do the measurement about the area. And one more information, you can have about the polygon is the centroid. So, coordinates are typically provided in geographic format. This we have been already telling you. This is true throughout our discussion in GIS that coordinates are in geographic domain or geographic coordinates.

And some point's features such as airplane locations need to also include z value or height and that will go as attribute information. And you can have n number of attributes as we have discussed but in some software like in ArcGIS, they also mention 3D points. So, whenever it is mentioned 3D point, do not get confused. Point is a zero dimensional entity. So, it does not have any dimension.

So, how this 3D point terminology has come. Whenever there is a term which is used in any software or literature, 3D points that means the point is having X, Y and z value. But that z value will always be stored as in non spatial information. So, do not get confused. Same with the 3D line also. So, that means along the line you are having height information or elevation information.

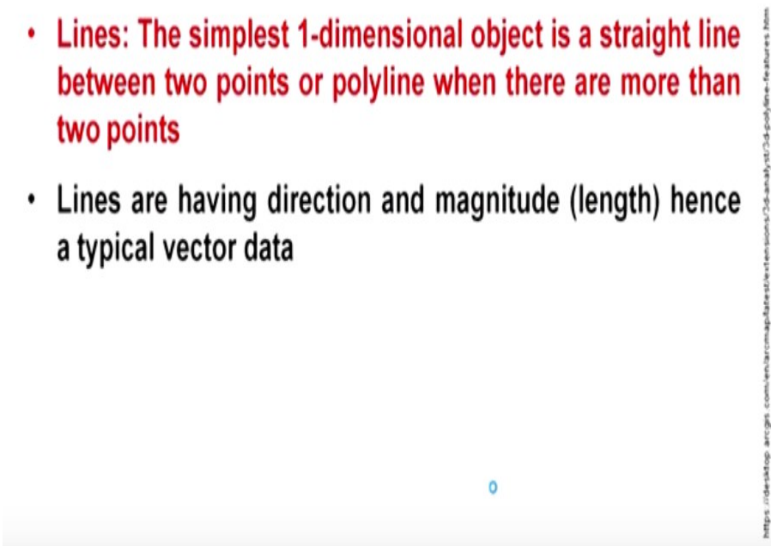
So, these objects must be modeled as 3D point features. So, if we want to have all the time z value, it has to be stored as a non spatial information in attribute table whether it is a height of a plane or even depth of a well or any other information which requires a third dimension to be stored, can be stored in that manner.

(Refer Slide Time: 16:29)



Now, 3D point features as I mentioned, they store basically Z inside the geometry of their feature class. And this means that z values are automatically included with every new point features. Now this third dimension value; z value can also come automatically if we are having in the background say digital elevation model or elevation surface. So, whenever we pick or click a point, automatically from the background layer which is an elevation surface; a Z value can be included and that can be stored against that point in attribute table.

(Refer Slide Time: 17:13)

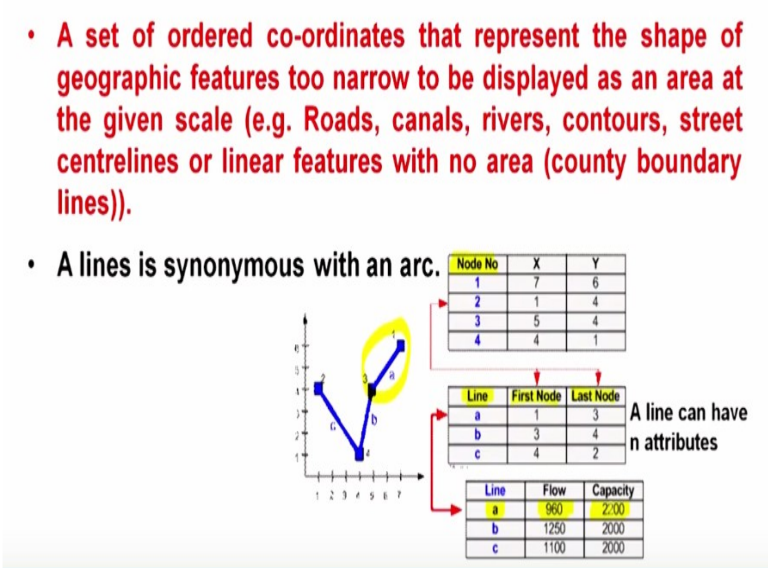


Now, the second entity among vector entities is the line or polylines. So, this is instead of zero dimensions. Now we are moving little higher in the dimensions and this is called 1 dimensional object or entity. And if it is a straight line then it is nothing but a begin point(x1, y1) and (x2, y2) or begin node, end node, that is it if it is a straight line. What if it is a polyline, I showed an example which was a polyline.

So, we had in that one begins point, one end point or node and in between, we also had 3 internodes. So, then that is called polyline. So, whenever in the GIS terminology whenever we say straight line means it will have just 2 pairs of x and y. And whenever we say polyline that means it is having more than 2 pairs of x and y. So, instead of simply calling as line we call as polyline.

And lines are having as you know a typical vector. Lines are having direction and magnitude that magnitude is equivalent to length and hence a typical vector data. Now a set of ordered coordinates that represents the line or a shape or a feature; though it will not have width but it will have only the line that is why it is called single dimensional entity.

(Refer Slide Time: 18:32)



Again this is scale dependent. For example, generally you know in a map, say 50 thousand scale or 250 thousand scales: a road, canal, river; all these features are shown as a simple line. But we know in actual or in real world, the road will have its own width. A canal will also have width as well as depth; same with the rivers. So, where that information will go? That information will be stored as attribute table.

Like in case of 3D points; the z value was stored in 3D. So, here also not only the width of the road, you can store as many as attribute information. Like when last the road was laid down, who was the contractor, who was the supervisor, who was the engineer and how much it has cost and so on so forth. So, against each route that can have information can be stored.

But on a map, it will be depending on the scale. It will be represented as a line and that is a 1 dimensional entity. That means only we can measure the length of the line, not the width of the road or canal. That information always will go there. But if I am having larger scale map like I am having a map of Roorkee, there I can show the width of the road or canal or river with the real scale.

So, many times these entities are scale dependent, just you have to remember this thing. Lines are also synonymous with an arc and that is why you know, this ArcInfo software or that word is still being continued as ArcInfo or ArcGIS. If I have to store a line like this line; there are one begin node and end node and in between we are having 2 internodes.

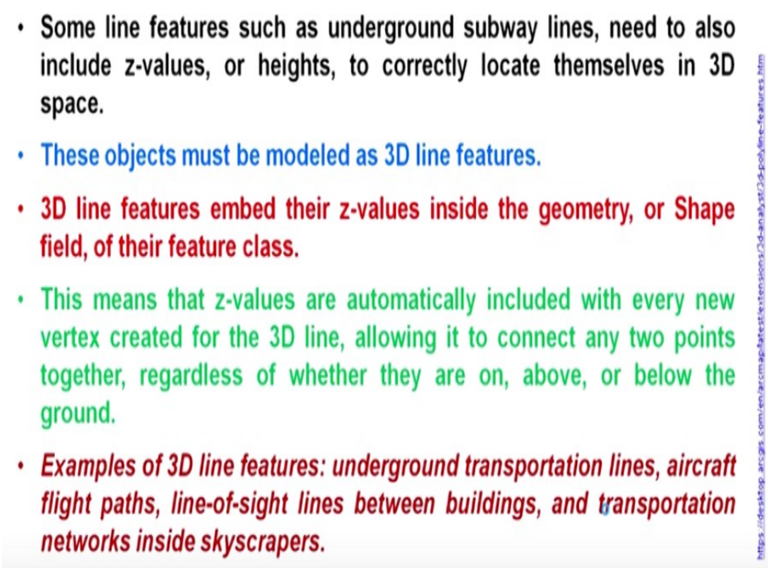
So, they are stored. Now, because it is one step higher in degree about the complicacy so, instead of zero dimensions; now we are having 1 dimension. So many more information is required to be stored. So, we will have a node table where all X and Y's are restored. We will have information about the line table where the first node and last nodes are restored.

So, if I take the segment 1 that is straight line between 1 and 3 then this first node is 1, last node is 3. Similarly, when I come to the attribute or non-spatial information then against a; you know some discharge value is written here or flow value is written here and then capacity is also written. Now, it is not necessary all this should always have a road or canal or river.

While discussing who are users of GIS? I also told that electrical engineers have also started using GIS. And there, they would like to know if there are different lines. They would like to know what kind of current is flowing and what the capacity is and so on so forth. If they want to store some information useful like height of these lines that can also be stored, the material of that line that can also be stored as attribute table.

So, as we go higher and higher in vector entities; the number of tables will increase that we will see little later in much more detail. Now when GIS is started, it is started basically for underground network and is still it is continuing of course. So, for underground features; the same information can be stored except that the depth is also stored instead of height.

(Refer Slide Time: 23:16)



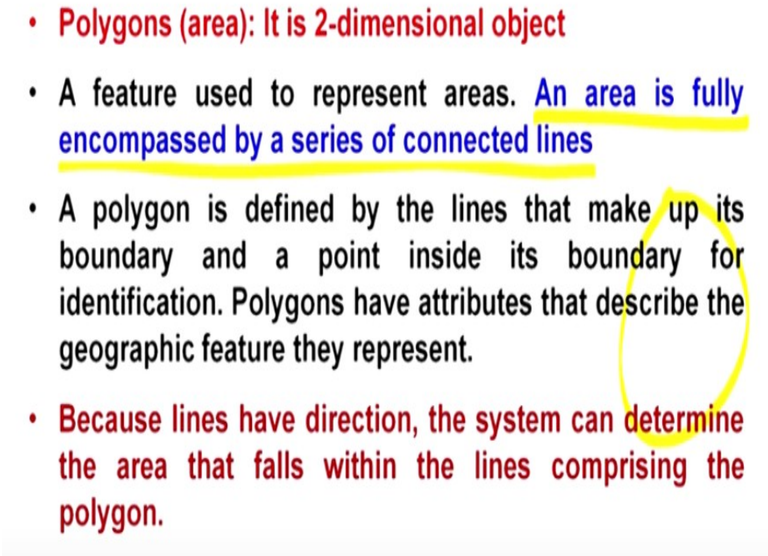
So, some line features such as underground subway lines need also to include z value or height to correctly locate them in 3D space. And these objects that mean underground lines must be modeled again in the 3D line features. So, that means the height is always attached as attribute of each line or each point. Basically line is made from continuous points which are interconnected.

So, the height value is stored against the point basically. So, 3D features embed their z values inside the geometry or a space field. And in softwares, you will find 3D line; 3D lines do not mean that you can measure the area of a line, not at all. 3D line means it is also having z value in it. And if it is not, it can be attached by bringing a digital elevation model or topographic surface below and then collecting information about the z value and putting in attribute table.

So, either you can automatically do it or later on, it can also be added in your line feature class. Line features; we have discussed few examples like underground transportation lines, lot of networks now, power or OFC or telephone line or water supply, gas supply. All these are now underground network. So, for that GIS is very appropriate tool to maintain and to know where each network is there at different depth and of course X Y location also.

Aircraft flight paths because the aircraft keep moving in altitude not only X and Y but Z also. So, these can also be stored, handled in GIS. Line of sight between buildings or in between two mountain or peaks can also be stored nicely. Line of sight analysis can also be done. So, when we will be discussing GIS analysis part, this discussion will all come; also called view set and then transportation network inside skyscrapers. Lot of networks can be handled very nicely.

(Refer Slide Time: 25:46)



Now, the third and last vector entity is the polygon. And also some literature will mention area instead of mentioning polygon, say area. Now it is 2 dimensional objects. Why it is 2 dimensional objects? Because for polygon, we can measure the area as well as the perimeter, so that is why it is called. Anything for which I can measure the area then it becomes 2 dimensional.

In case of line, I can measure only the length. So, that is why it is single dimension. In case of point, I cannot do any measurement specific to a single point. But of course, if there are 2 points, I can measure the distance between 2 points that is a different thing. So, a polygon or polygons or area is a 2 dimensional vector entity and it is used to represent many features which are present on the surface of Earth as polygons.

For example, you know land records or revenue record, revenue information. Different agricultural fields are there, different plots are there. They will be stored as polygons. Of course, here also the link of scale will come. If it is too small scale then polygons cannot be shown as polygon. You have to reduce, maybe even point. But when it is large scale then polygon or the area can be shown in form of polygon.

So, now as mentioned here that an area is fully encompassed by a series of connected lines; what does it mean that if this is my begin node and I come back after that so, my begin node and end node has to be the same. Then only it makes an area and it closes the polyline. So, it is a higher version of polyline where the condition is that begin node and end node has to be same.

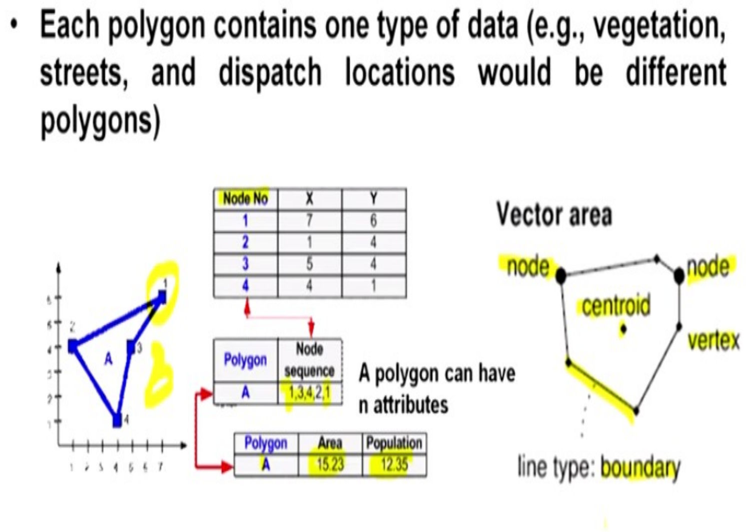
So, here like X1, Y1 and Xn, Yn has to be the same whereas in case of line or polyline, X1, Y1 and Xn, Yn need not to be the same. So, this you have to remember. Now polygon is defined by lines that make up the boundary as I have already drawn here; a point inside this boundary for identification that we also called as centroid. And polygons have attributes like point data can have attributes; line data or polyline data can have attributes, so the polygon.

Polygon can also store. Suppose if I am having revenue records or plot records so different plots are owned by different people. I want to store the area as well. I want to store the owner of the plot. Whether it is already constructed a building or not. If a building is there, how many stories are there? How many people are living and so on so forth? All that information related with a single polygon can be stored as non-spatial information.

As you know that because lines have direction; it is a typical vector and the system can determine the area that falls within the lines comprising the polygon. So once it is closed, now we know what the inside is and what the outside is. And that gives the advantages about you know, keeping the watch or information about the neighborhood. So, that is the best advantage of having vector data.

Because vector with line; so begin node and end node. Now you know that what is the direction and once it is known then you know who is outside and what is inside. Now within one polygon; you cannot have many details. All details are merged and single polygon like for a land plot. Now within land plot unless I make further divisions otherwise I do not have any further boundary in it.

(Refer Slide Time: 29:47)



So, each polygon contains only one type of data. For example, a polygon can contain an agricultural field so, one type of vegetation. Though in the field, there may be 2-3 vegetation’s. But if they are not marked or created a sub polygon, I cannot see. So, that one thing one has to remember. Now, as we have seen examples of point data; how it is stored in the system. Examples of line or polyline data; how it is stored.

The major point here; begin and end point is the same. So, is it closed polyline! Now again, you will store as a node table which is like this. So, all nodes are stored; there X and Y is also stored. Now, this A is the polygon which is the inside one within the vector. So, it is having a node sequence starting from 1, 3. And last one is also 1, first one is also 1.

Because it has to be closed then only it can be called as polygon. If it is not closed, it is not polygon. It would be called as polyline. And then polygon A; because now it is closed so we can measure the area, we can measure the perimeter or we can store as much as information which we would like to have before the polygon. Now, further we have to also store later on, the information about the outside that is the B.

If I call as the outside polygon is B, then A polygon which is the real polygon, outside one information will also be stored. So, when we will be going in further details about how system is stores vector entities, we will also see that kind of thing. Again like line, point; they can have theoretically n number of attributes, so the polygon. Why I am saying theoretically? Why I am not saying all the time?

Because, when we come for storing as many as or n number of attributes says in case of ArcGIS in a simple database management system or like in Excel, we can have maximum 512 columns. So that means I can store maximum 512 attributes against each line, point or polygon. But in some projects if I have to store more than 512 then we have to come out and resort to some other database management system where this limit is not there.

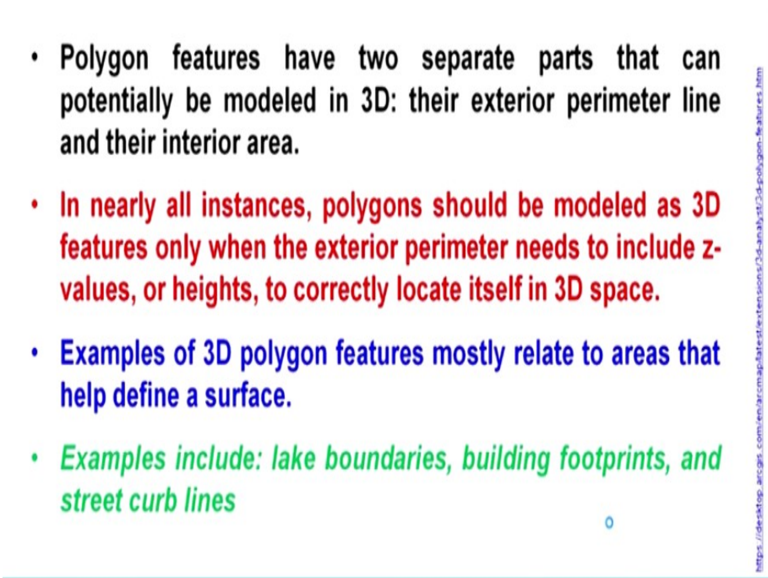
But there may be a limit of 1024 or whatever. So, somewhere theoretically no limit otherwise practically depending on the software database management system, there might be some limit. But it is still 512 fields against each point; hardly in any project have we required. Even if we require and we do not have any other way to go for another database management system, there are ways.

The best way you store in 2 tables and these 2 tables can be related any time in your processing. So, you can have multiple tables; each is having 512 fields, no issue. So, there is a way by which you can always handle more than 512 remaining in the same system. But currently this is what the support available. One more point which we sometimes for some particular studies that become very useful that is the centroid; the center point of a polygon.

That is called centroid. So, that is also stored or can be created automatically on standard GIS software. And this information becomes many times very useful because for a polygon which is a 2 dimensional entity. Now you are having a point and you can attach many things with that point and this point is unique for each polygon.

So again here begin and end nodes are there, vertices are there and then you are having the boundary. So, you can measure area; you can measure the polygon and you can determine the centroid as well.

(Refer Slide Time: 34:34)



Now I have already touched but just to recap this thing that polygon features have 2 separate parts that can potentially modeled in 3D; their exterior perimeter line and their interior area. So, because outside is open, so there we cannot have perimeter. So for inside, we can definitely measure the area but the information for outside is also stored. And nearly all examples or instances; polygons will be modeled as 3D features only when then exterior perimeter needs to be include the z value.

So, along with polygon also but if a plot or a agricultural land having undulations then against only one single z value will be stored. And you know in real sense in real world, there are undulations. So if you want to store undulations; you have to have separate polygons or there is another way of representing such thing is resort to the raster data rather than vector data or polygon data.

And 3D polygon features mostly relate to area that basically helped to define the surface. The surface and undulations can be best represented through the raster rather than vector because vector is discontinuous. You know the topographic surfaces are represented through polylines that means the contours; Iso-heights but in between contours, you do not have any information.

Though contours are representing elevation values also but it is not a perfect solution to represent a surface. Instead, there is another model which is raster representation by which we can represent the undulations, variations present within a surface. So, we should choose appropriately. But like lake boundaries; lake water is flat we can consider very well and therefore lake boundary can be represented as a polygon, no issue.

But lake can have depth and depth can be represented as a 3D polygon if that information is available. Same with the building footprints; in what area it has been constructed; constructed part. And because building might have height and therefore that information can be also stored as a 3D polygon. And there are many other such things can be stored. So, this brings to the end of discussion about the vector data.

Just to recap everything about the vector. 3 basic entities; one is point which is the simple zero dimensional entity, no measurements are possible. Only it stores X and Y coordinates. If you want to store z coordinates, it will go as attribute value. Now line or polyline is a 1 dimensional entity. You are having begin point end point or a begin node, end node. And if it is polyline then X1, Y1 and Xn, Yn and here you can definitely measure the length and of course, the direction is also known.

Because you know that what the origin is, what the destination is. And in case of polygon, the first node and last node; that mean the X1, Y1 and Xn, Yn has to be the same. And then it is nothing but a closed polyline. And closed polyline; the advantage it’s a 2 dimensional entity and therefore we can measure the area, perimeter and also determine the centroid. Thank you very much.