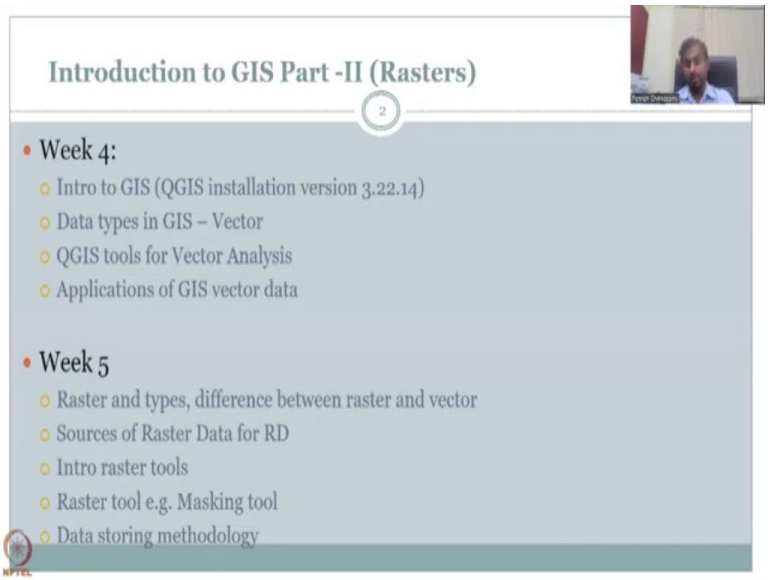
Remote Sensing and GIS for Rural Development
Professor Pennan Chinnasamy
Centre for Technology Alternatives for Rural Areas (CTARA)
Indian Institute of Technology, Bombay
Lecture – 1
Intro to Raster Data Type in GIS

Hello everyone, welcome to Remote Sensing and GIS for Rural Development NPTEL course; this is week 5: lecture 1. Let us do a recap of what we have learned in lecture 4 and how it is related to lecture, week 5.

(Refer Slide Time: 00:45)



In lecture 4, we had done an introduction to GIS which was part one vectors. In lecture week 5, we will be discussing more on rasters, which is the second type of data to be specific. We gave an introduction to GIS like what is GIS? How does GIS process the data? We looked at the flowchart of defining a problem and then understanding the issues. We also looked at multiple options for GIS software of which we will be using QGIS. We also looked into the different versions of QGIS; and we will focus mostly on the most stable version which is 3.2214 till date, in 23, early 2023. So, this might be different when you view this video in 2024, 2025 depending on the registrations; but always look for the most stable version.

Then, we went through the process of downloading the data for the software. And then we ran and install the software on my machine which was shared on the screen. Once the GIS was setup, it was also necessary to understand what are the different types of data in GIS. And we stopped

and looked at vectors and rasters. We defined that these are the two types and we went in depth to vectors in week 4. We also looked at QGIS tools for vector analysis. So, once we define the two types of data, then there are tools that work on vector and work on raster separately. So, we had looked at the specific tools in vector analysis that are installed in QGIS.

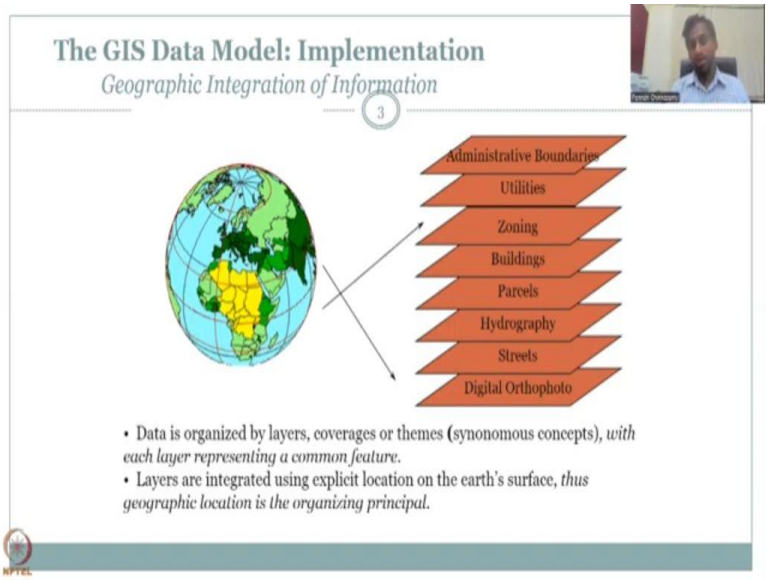
The tools may be different in whatever GIS versions or software you use. However, we cover the basic tools, so you will definitely come across the same tools, different names will be there. For example, buffer could be labeled as buffer vector, or buffer line, buffer polygon et-cetera. Buffer is always the same tool; it creates a buffer around a shape file point, polygon or line. So, the name of the tool might be different in different software's. As per GIS, the application of the tool is same. Then we look at applications of GIS vector data and analysis, so we will define the data.

We looked at some tools that can organize and play with the data, and then we jumped into the applications. Now, how is this linked to week five? We had mentioned that we have two types of data, so vector and raster. I had introduced what is a raster, but I did not get into in depth discussions; that is what we will accomplish in week 5. It is also necessary to look at the difference between raster and vector, the two types of data. Some may get confused that it is a full coverage; so is it a vector or a raster? Coverage need not always say it is a raster and vice versa. So, we should be careful about what is embedded in the data, how the data is represented; and that defines raster or vector.

Sources of raster data for rural development; we will look into some sources. And then we will jump into some intro raster tools in QGIS platform. So, there are raster tools generic, and there are raster tools that you could use throughout the GIS network and specific to certain software's. We will have an introduction of the major major raster tools. One example is the raster calculator and the masking tool. These tools are very very important and almost across the software system for GIS, it is used. Then, we will have some inputs on data storing methodologies. Throughout I will stress the fact that this is based on QGIS software, the recent stable version, and mostly for rural development.

There are multiple other applications and tools that are more specific for different applications such as atmospheric science, transportation, heritage mapping, et-cetera. But, we will be mostly focusing on rural development. Some case studies may and may come in and go depending on the basics.

(Refer Slide Time: 07:08)



So, what is the GIS data model? It implies that multi-source data is converted to digital formats, if it has a spatial location. For example, let us take this example of from the world there would be multiple layers. GIS is about layering one layer above the other and then drilling down to take, which is the analysis to take the research. So now comes the point, what are these data? All these data should have a particular input which is the data plus the spatial data. The spatial location of the data is GIS. So now, since data is organized as layers, coverages or themes, it is same concepts; but, it could be different types of data, so coverage or theme could be synonymous.

But, it has different data formats; it can have data formats. So, please understand that part that yes, it is synonymous coverage or theme, but it can be different data part of which there are two types vectors and rasters. However, each layer is representing a common theme of the problem. You are looking at rural development and I have said that crop acreage is one very very important rural development question. We cannot map on top of coverage. Something in rural India, you cannot map urban coverage for example. If industrialization has happened? Yes; but given that it is focusing on crops you have to concentrate on crops.

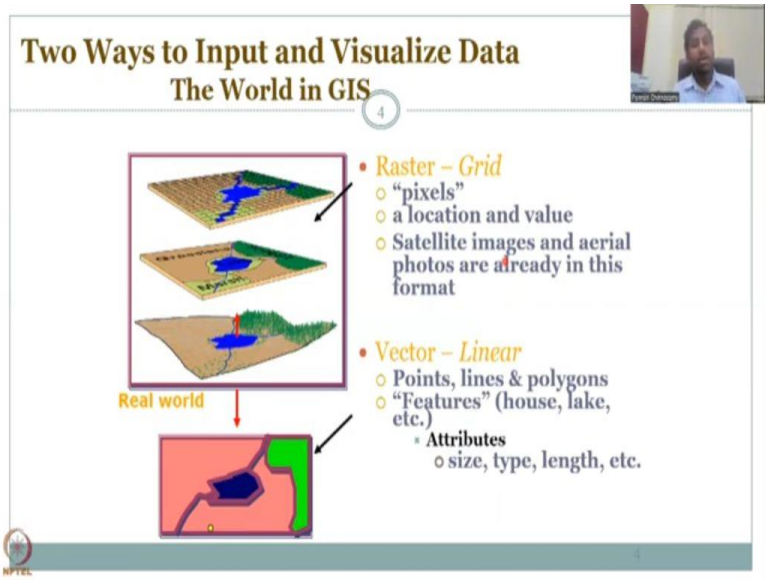
So, common theme is there. Layers are integrated using explicit location on earth's surface; because earth location is what ties all this data together. See, the data can be spread different locations. Why is it on top of each other is because the location is safe. For one location, you cannot have only one data that is the beauty of GIS you can keep on stacking; so, it is a

comparison if you would like to see. In one location you can either build one house which is called a bungalow or a separate house, or you can build an apartment which is vertically tall.

If you look at the number of levels is could have 33 levels and this is only one level; and that is what data will also give. Because here in one level, you will have only 4 to 5 people living; whereas, here every floor you will have 4 to 5 people living, provided both are the same entity, or the same people you are taking in terms of age, occupation. So, that is a common feature. So, GIS is on the left which I explained is about a vertical staggering of data and then you take an information out. So, layers are integrated using explicit location that is a single location; thus, geographic location is the organizing principle.

The location is organizing principle and now it does not matter if it is raster or a vector. Which means, one floor can be green in color, the other floor can be blue in color, and then green, and then red, brown whatever it is. So, the colors represent different layers. So, it does not matter in GIS, if how many ever different layers come up as long as your software and your computer can hold it; so, the location is what is key.

(Refer Slide Time: 11:14)



We have already looked into this in week 4, but because we are discussing raster, I will just quickly go through this again in a real world scenario which is given in the bottom. You have a lake, and a river, or a streamline that brings water to the lake and you have a grassland; around the grassland, you have some marshy lands and forest. And that has been accurately depicted as a

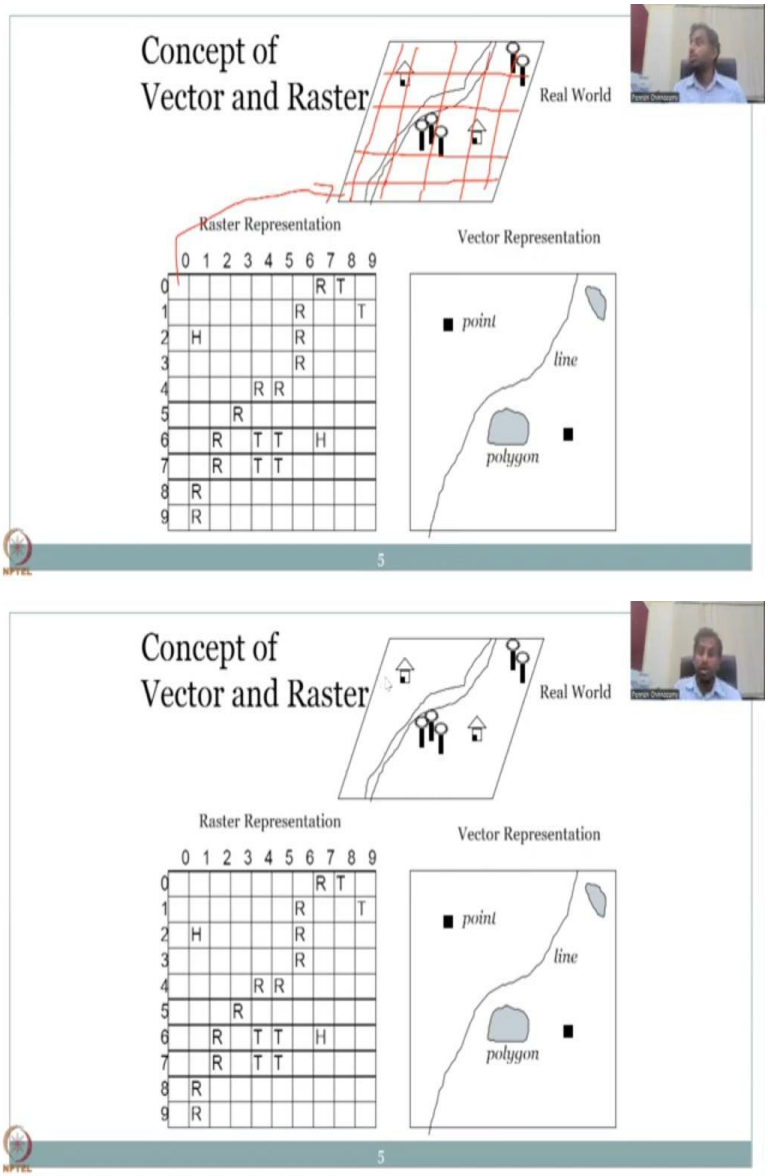
2D surface in the bottom. So, what you see in the bottom is my pointer. So, what do you see at the bottom is the real life scenario. And you do have on the top, the 3D real life world, which has been converted to a 2D surface. So rasters are in grids, which is on the top, you have divided the real world into grids.

The grids are called pixels; inside each one is called a pixel. So, you discretize it, which means gridding the surface; and then each square in the grid is called a pixel. It has a location and a value; the pixel is centered and it knows the size of the square. So, if you know the center and you know the size of the square, you can easily draw it. This size as in from the center, what is the distance to the side of the square; so, if you know that it is equidistant. From the center, you can have equidistant to the perpendiculars to the side; so that is how you could definitely create a location, a single location for a pixel.

However, the data is averaged for that location. As I clearly said, suppose this entire thing is one pixel, this entire thing may be a one pixel in a satellite image. So, what it will happen is for it is a one location but across that grid; and across the grid, what is the land use land cover? The dominant value comes in as the value, the value as given here. For example for this one, we can say the brown color is dominant; so grassland is dominant; so it will be brown color in the value. So, satellite image and aerial photos are already in this format, which is graded; and that is why you called as photographs as pixels. 4k resolution, high pixel, high definition these are all terms that say that it has been graded, and each grid has a value.

Whereas, vectors are linear, it is non-continuous; it has points, lines and polygons and they are called features with attributes. Features are the class, house, lake et-cetera. And under the class, there are data; and that data or sub-columns are called as attributes.

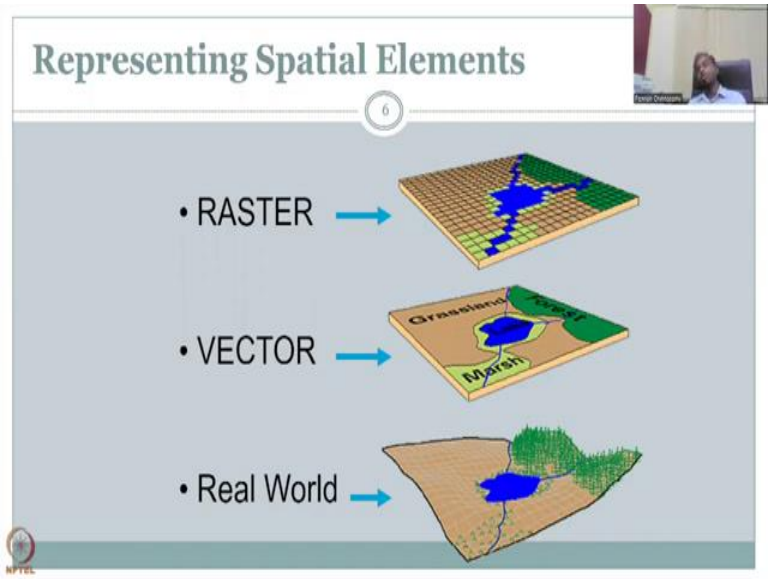
(Refer Slide Time: 14:17)

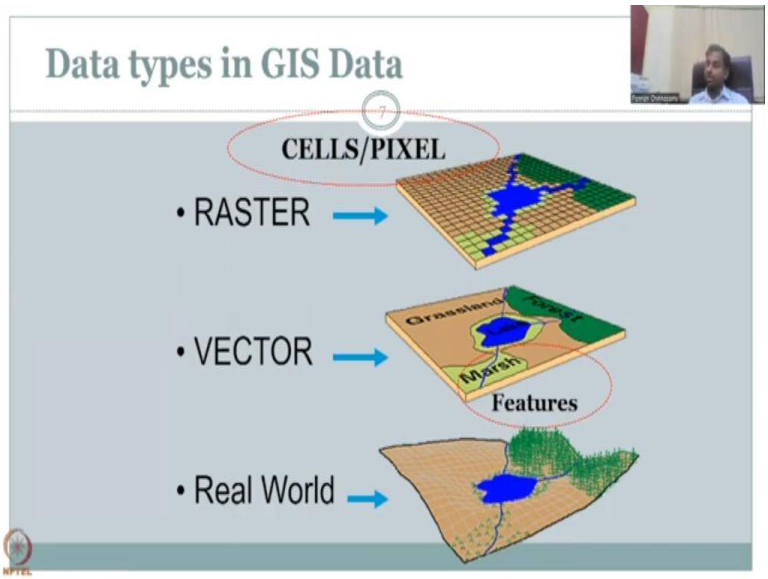


One more representation of the world we had looked at. In a raster, I will just only focus on the raster here. So, we have a real world, where you have river and trees, houses in a real world which is a 3d mesh. You have discretized it in by forming grids; so this has been divided into rows and columns. Sorry, my drawing is not as good as it should be because I am using a pointer. So, you get the idea that each pixel is what is represented here. And then you, you use the value inside the pixel as the, the dominant value inside the pixels the single value. So for here, you have rows and columns, 0 to 9, 0 to 9 it is a grid in square; it has each pixel is given here.

You have one pixel; I can color code it for just to show you the pixel. So, this is one pixel at a time and you can use this here; so that is one pixel and then you have the gridded lines. What you can do next is you could see that in the grid in the pixel, what is the dominant value it is taking? So for that location, what is the dominant value? Again the location is not a single point; but across the square, across the pixel what is the dominant? So, here is nothing-nothing which is barren land barren land; let us say land is empty, and then you have houses. There is one house in 2, one house in 6 and 7, so 6 and 7 is this one; whereas, this house is at 1 and 2. So on the column, it is 1 and then on the row it is 2. So, this is how a real life world can be represented as a raster and then vector.

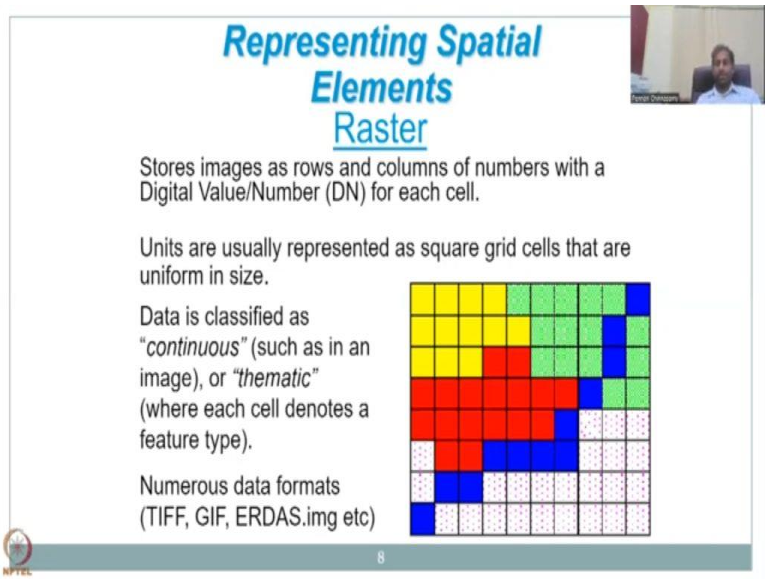
(Refer Slide Time: 16:33)





So, how are they done? Again the real world is converted to a 2D surface. Two types of data is there; one is vector and raster. In a vector, you have lines, polygons and points for each real world scenario. Whereas in raster, the entire plane has been discretized into grids and each pixel takes a value. So, when I go into the differences between these two, I will focus on why. So, you have pixels as the dominant division of the data, whereas features is the definition division of the data in the vectors.

(Refer Slide Time: 17:23)



How are spatial elements represent as rasters? Stores images as rows and columns image. So, the image here has been stored as a row and column. As I showed you the previous slide 1 and 2, we

had a house. So, as per the row and columns location, it stores the data as an information and the location. So, stores images as rows and columns of numbers with a digital value number DN for each cell; the DN for each cell is very unique. So, as I said row 1, row 2 in the previous examples, so row 2 and column 1 is unique. Only one pixel sits there, not multiple pixel takes that location.

Units are usually represented as square grid cells that are uniform in size; so, always it is square and it should be uniform. You cannot have a green bigger than red or yellow. It can be only when you do a local discretization as done in hydrological models. But for raster data, all are same size grids. And that is why your green is not bigger than the blue, or the yellow is not bigger than the green, it is the same size. With this what happens is you have uniform spread of the data with equal pixel size, and the dominant land use land cover or rainfall is taken as a value for it.

Data is classified as continuous such as in the image, you do not have an image, whatever the size is; it can be oval image, it can be a square image, rectangle image, A4 whatever size. Every image has full coverage; you do not see inside an image a white plain area; if that is a case it is called an error. So, within the frame, every single inch should be or centimeter whatever the smaller sizes should be occupied with data; in here image the data is color. So, it is continuous such as in the image given below. In the image here, you do not have any blank spaces; the pink dotted squares are also data.

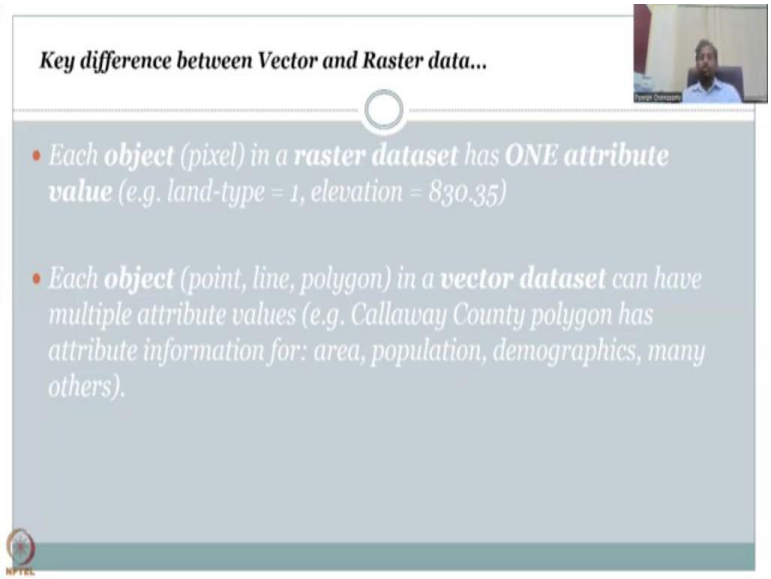
It is continuous or thematic, which as I said, rainfall land use, land cover these are the themes; thematic maps and the data is being applied across, where each cell denotes a feature type. Only one feature type, but it denotes the entire thing; so here the theme is land use land cover. The cell could be a river or a house, or a tree based on the feature which is within the, which is the data the dominant data. There are numerous data formats for grids. When we looked at shape files, we saw that it is dot SHP dot SHPX, and then other data that supports the vector. So, it is only one type of forming. But, in the one type, there are multiple multiple sub files. Whereas, in raster there are numerous grid formats, how the grid is stored and how the locations are stored; there are numerous formats.

Someone may ask why sir for raster you have numerous formats? It is one of the limitations has been discussed is that rasters have bigger size. And for the biggest size with evolution of science

and technology, with upgradation of technological interfaces, there is new formats that are being discovered or developed. Why? Because, they want to cut the image in a particular fashion to store it effectively and retrieve it so effectively. Remember that if it is too big, it is difficult to store and also difficult to retrieve. Both are important for spatial analysis, storing and retrieving. So to make that faster, there are always upgradations of the formats and that gives rise to a different or multiple formats for rasters.

People would be happier with simple shape file, so you do not see much updation; but, rasters yes, and that is the same reason why initially you could store only one two images in a pen drive. A floppy disk may have two, three, CD ROM may have hundreds of files, images; whereas, now a single pen drive can have multiple multiple GBs of data which is raters. So, now we have known what is a raster, what is the vector? Next we will see what are the key differences.

(Refer Slide Time: 22:55)

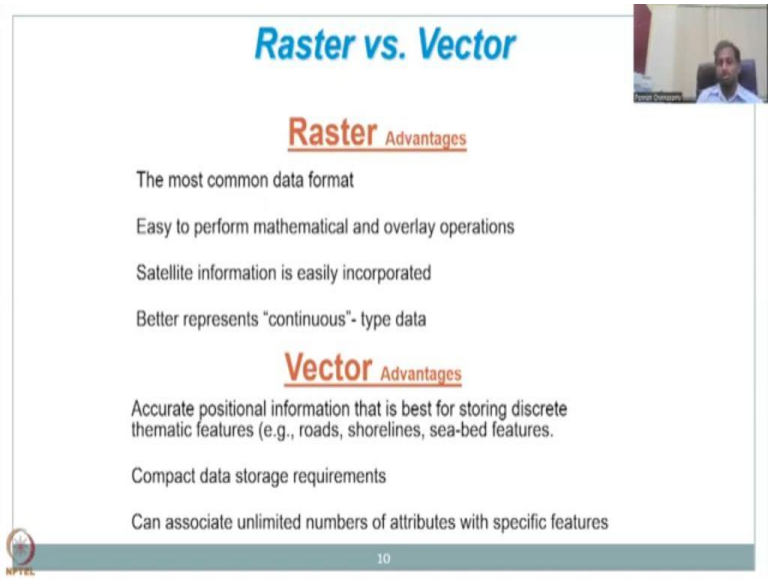


Each object is a pixel in a raster database, and has only one attribute value, one feature. So, one feature one attribute is merged here, one data; only one data per pixel. It is not like rainfall is there or not; it is just rainfall which is 1, or 0 for no rainfall. So, rainfall is a theme, rainfall is the attribute. If it is no rainfall, it is zero; but still our color is given for zero which is white. And then blue could be rainfall. So, has one attribute value; example land-type is equals to 1 is one value, elevation is how much the land is elevated from the zero level which is the sea level. And you could see that 830 meters or feet depending on the data, you have one value for the pixel.

You do not have a range for a pixel; one pixel only takes one value. However, in a vector, each object each object which is represented in a GIS framework, in a vector database can have multiple attribute values. Example, a county or a district boundary has attribute information for area, population, demographics, and many others. So, you do not stop with just the district name or the village name. There are multiple other data that are stored. As I said, there are sub-headings. The name could be the predominant column that is holding it; here the district name. And within the district, you have multiple attributes that explain the data further.

Sometimes too many attributes do not explain but spoil the data; so be careful about the size and volume relation to the quality. Do not always think that too much data always explains better; too much also has lot of errors and issues.

(Refer Slide Time: 25:10)



Some more raster versus vector definitions. So, now we have seen the raster versus vector. What are the other things that we should discusses the advantages and limitations, and how it compares between each other. In a raster, the most common data format is the raster; because most of the big data, the data that covers the entire planet comes as rasters, easy to perform mathematical and overlay operations. Why? Because for each location, you should have data if the layer is added. It is not like I do not have data, how am I going to do analytical, or mathematical operations. Since, always it will have data, you will have to use it as continuous data; and that is easier for you to do calculations.

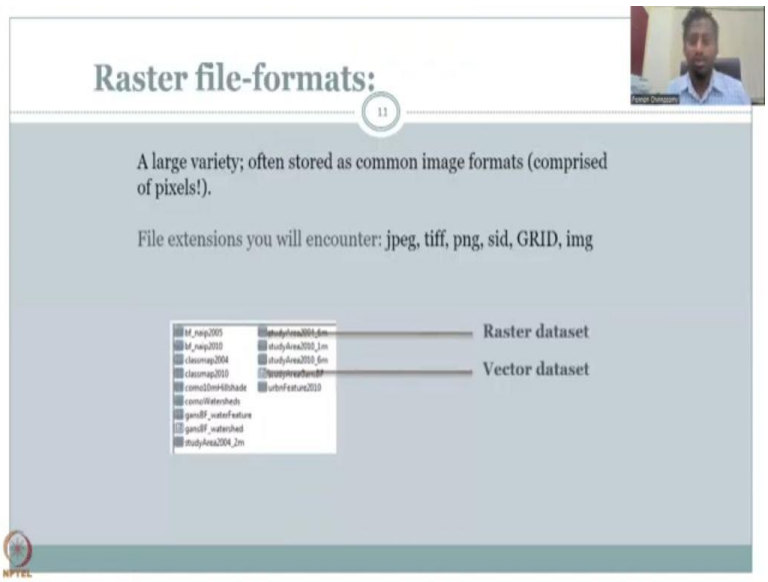
Suppose you have data gaps, you do not have data for that particular location; you are not able to do a calculation. So, that thing is removed because we use a continuous data. This data is stored in each pixel and each pixel is together mixed in this raster to as an entire data. Satellite information which is the information procured from satellites is easily incorporated in a raster; the satellite coverage is continuous. And based on that property of raster it is easily incorporated as a raster data format. It does better represent continuous form data compared to vector data, and vector advantages are accurate position is given.

So, you do not play around saying that somewhere in the pixel there is trees, either the dominant is trees or not. However, in a vector, you can add multi-dimensionality in the same pixel. So, for example, the pixel is your land holding of a farmer; the average land holding is one hectare. So, let us say that one hector is one pixel. The farmer might have some issues with growing a mono crop in the land; let us say he is not going to grow. He or she is not going to grow sugarcane. But, some tomatoes, some spinach, some banana, and maybe drumstick all these are within that small piece of land.

How will that show in vector is each plant will have a specific location or a polygon. So, it will be a point data or a polygon data. Whereas, for the raster all of this is one data; the dominant data will be uploaded as a layer for that pixel. So, that is what the advantages for vector; it is not contiguous. However, for that particular location, it is the best data available. It is the best for restoring discrete thematic feature; example roads, shorelines, seabed features, think about this. If you are going to do a map of the water bodies, ocean and rivers, the land is part. You cannot just show in a raster the lake alone without the data boundary from rasters; so, you have land and water. However, in a vector you do not need this; you can just put a boundary and say this location. This boundary is therefore water and that is where it has less storage.

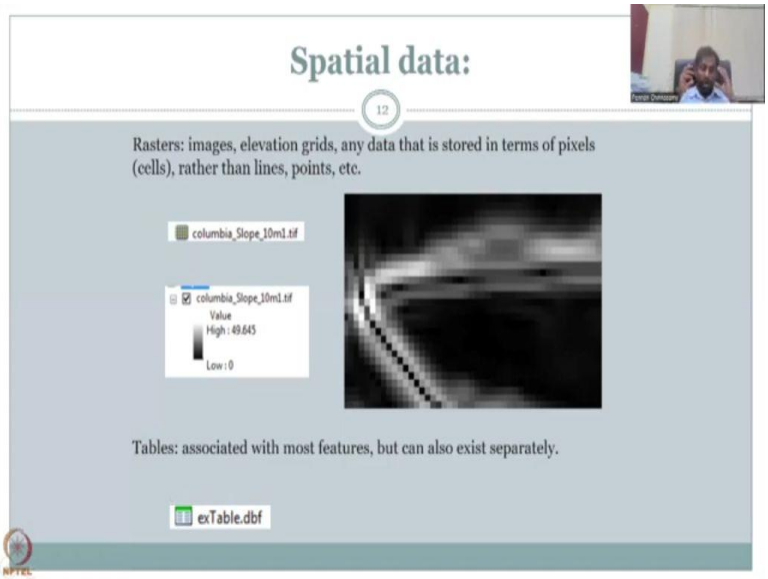
It is not continuous, so it does not have to put storage everywhere. It is compact data storage requirements is there; you do not need big supercomputers or storage facilities. Can associate unlimited number of attributes for specific features; again, I have said this in the last week 4 not unlimited, depending on your computer's speed and performance power. It will give some number of columns, the columns are the sub-attributes.

(Refer Slide Time: 29:54)



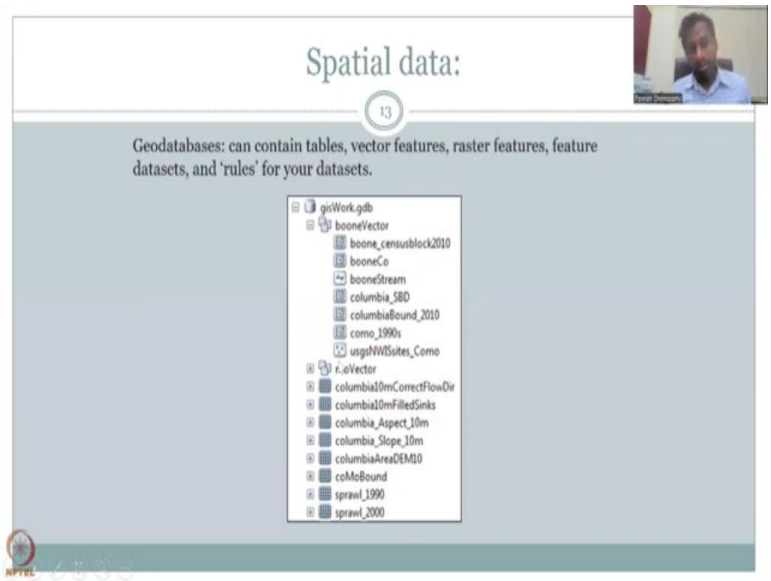
The raster file-formats, as I said they are multiple; there is not only one format. And the often ones are jpeg, tiff, png, sid, GRID, img image; and you could see that the icon looks like a gridded box. So you see, initially you would see like it is a bar of chocolate, but it is not. So, if you look at the bar of chocolate, it has lines; and then which are vertical and horizontal. Thereby, giving rows and columns, each chocolate you can break as a pixel. So that is what you could see here the icon for a raster data is given as a pixel or a gridded land. The dataset is given as different; we have seen this in the previous lecture.

(Refer Slide Time: 30:49)



Spatial data: raster images, elevation grids, any data that is stored in terms of pixels, rather than lines, points et-cetera; again, another definition. You could see here that the river line is not a line but a grid, grid which is linked to each other. This is the Columbia slope model for data. Sometimes only sometimes there are tables associated with the raster; and it depends on the data type the dot grid, dot img, the dot jpeg. So, based on the format and the size of the class, the size of the data, you will have a extension table.

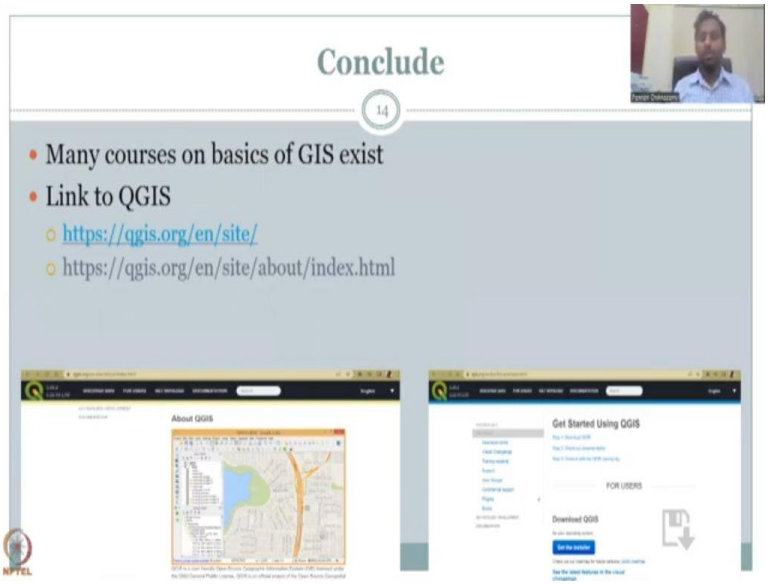
(Refer Slide Time: 31:37)



So, spatial data is stored in a database. And I would like to stress here that in a geodatabase, you can have vectors and rasters together. It is not like you have to separate your raster, separate your vectors. However, you can have them all in one working geodatabase as this is there. You have a geodatabase and you can store it. In normal circumstances, I have a folder; the folder is GIS data. And within the folder, I have subfolders for raster or rivers; and then you have subfolders for vectors administrative boundaries, river boundaries et-cetera. So, there are both limitations and advantages of using raster and vector; so we cannot say one is better than the other. It depends on your problem statement and GIS tools that you are going to use.

So, you can always have all these data in your database folder mingled together; because GIS when you open it, it will tell you what type of data it is. For example, if this is being shown, it shows that this is a polygon, a line, and then a point; and then it also shows that this is a raster.

(Refer Slide Time: 32:57)



With this I would like to conclude today's lecture. You would also go back to these forums and manuals if you have difficulties and questions, the links are given here. So, please use them as much as possible for updating or brushing up your QGIS skills; because that will you will be using much in this course. With this I conclude. Thank you.