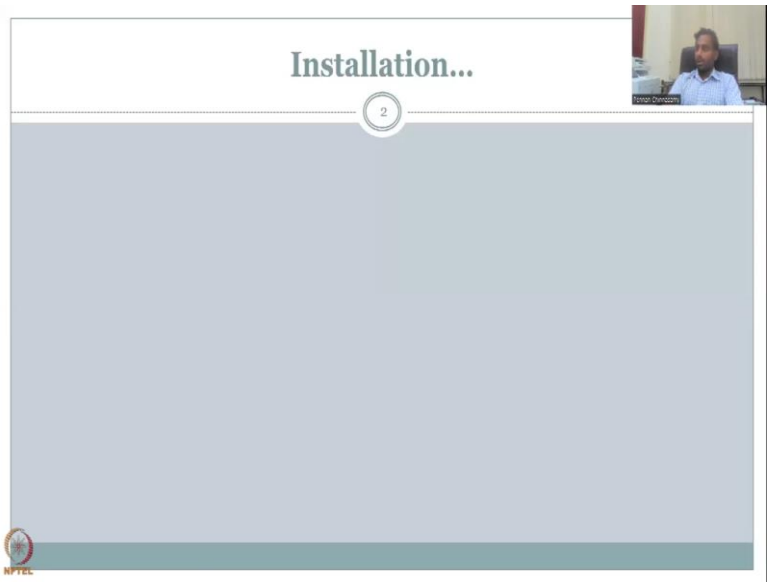
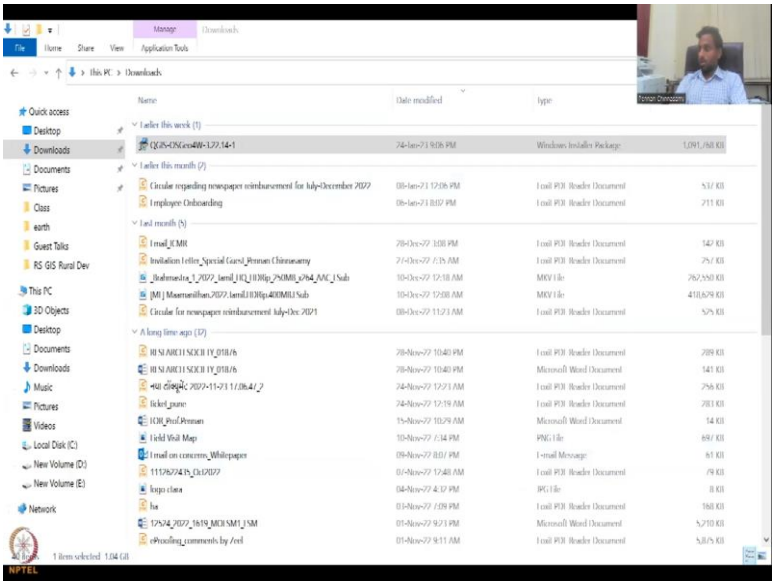
Remote Sensing and GIS for Rural Development
Professor. Pennan Chinnasamy
Centre for Technology Alternatives for Rural Areas (CTARA)
Indian Institute of Technology, Bombay
Week – 4
Lecture – 2
Introduction to GIS Data Types and Download

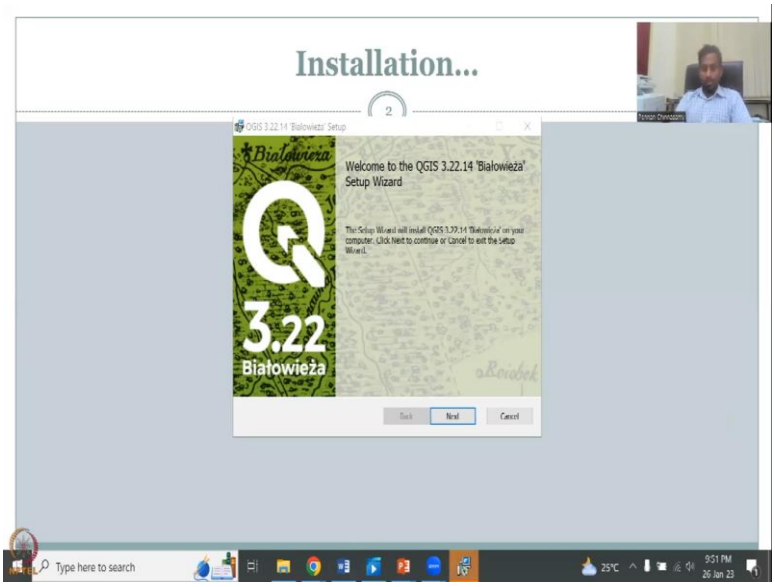
(Refer Slide Time: 00:15)

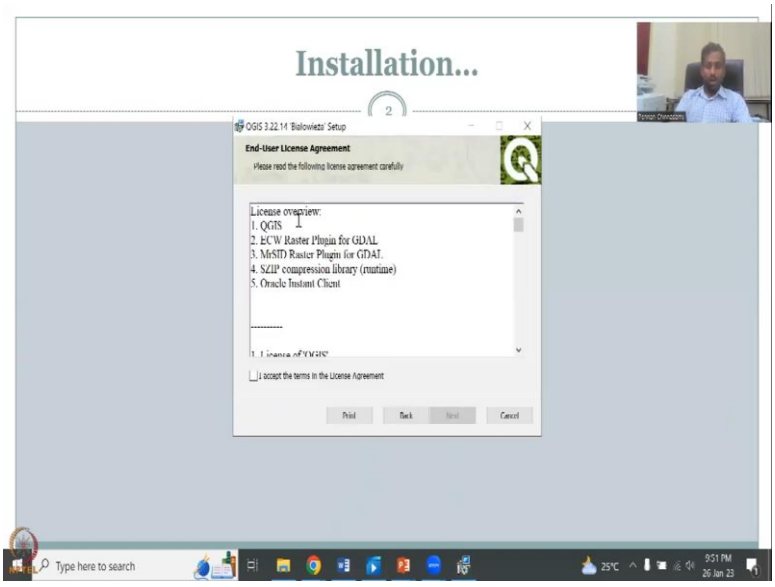
Hello everyone. Welcome to Remote Sensing and GIS for Rural Development NPTEL course. This is week 4, lecture 2. In this week's lecture, we have been looking at using QGIS and introducing GIS concepts so that we can analyze the remote sensing data.

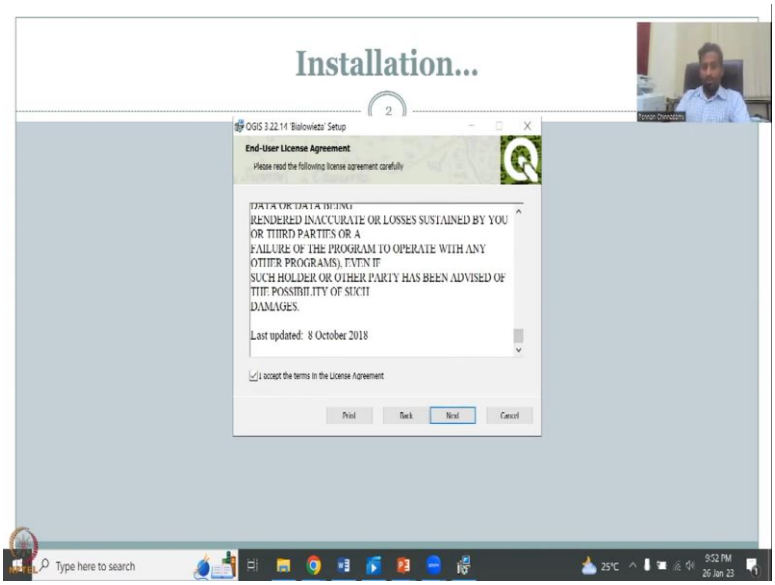
(Refer Slide Time: 00:40)

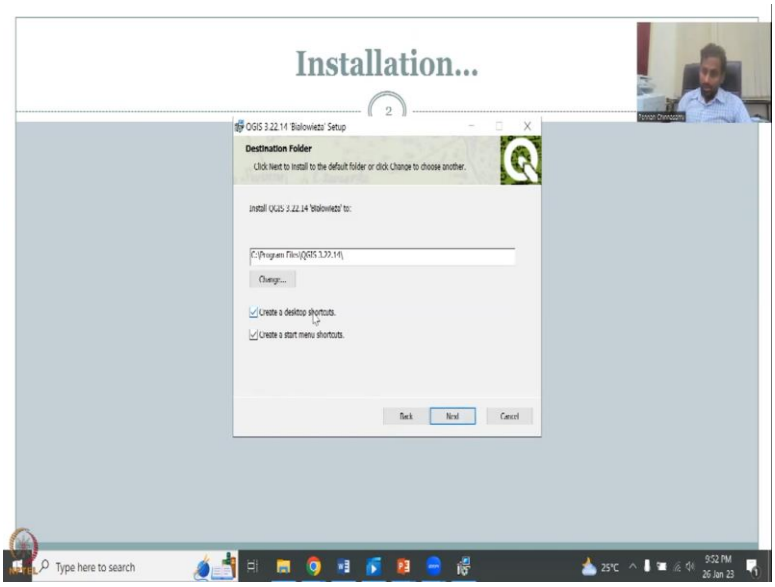



Manage
Downloads
File
Home
Share
View
Application Tools
This PC > Downloads
Quick access
Desktop
Downloads
Documents
Pictures
Class
earth
Guest Talks
RS GIS Rural Dev
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Local Disk (C)
New Volume (D:)
New Volume (E:)
Network
Name
Date modified
Type
Earlier this week (1)
Earlier this month (2)
Last month (5)
A long time ago (17)
1 item selected 1.04 GB
NPTEL


Installation...
2
QGIS 3.22.14 'Białowieża' Setup
Welcome to the QGIS 3.22.14 'Białowieża' Setup Wizard
3.22
Białowieża
Back
Next
Cancel
Type here to search
25°C
9:51 PM
26 Jan 23


Installation...
2
QGIS 3.22.14 'Białowieża' Setup
End-User License Agreement
Please read the following license agreement carefully
License overview:
1. QGIS
2. ECW Raster Plugin for GDAL
3. MrSID Raster Plugin for GDAL
4. SZIP compression library (routine)
5. Oracle Instant Client
I accept the terms in the License Agreement
Print
Back
Next
Cancel
Type here to search
25°C
9:51 PM
26 Jan 23





I hope all of you have downloaded the QGIS software which was shown in the last week and last class. So, today we will install it, just to show that how the acceleration process goes on and to make sure that you do not get stuck in the way of installation. So, let us go ahead with the installation.

I am going to share my folder. So, in your folder, you would have seen this part where you have installation. You have double click this part where you could go for installation which I downloaded in the previous time. And you will come across the insulation window like this. I hope it is visible. I will just make it visible again for those I have to stop share and share. So, now, you will be able to see the installation web page and I am going to showcase the installation now.

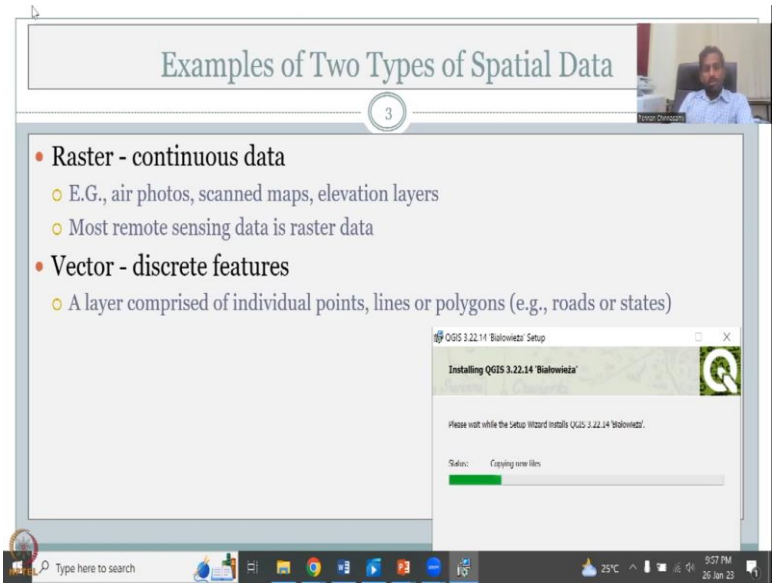
So, in the installation, you will be seeing this part. So, now, you could see that the insulation process has started. It does take a little bit time and you will click the bottom screen. So, you just say next. In the next window, you will have overview of the license, the raster plugin which is a type of data, we will cover it today and then MrS IT, raster plugins from GDAL SZIP Oracle Instant Cclient etcetera, etcetera. So, you just go through all the license agreement. We cannot read it in the class time but I recommend you to fully read it.

I have read it and so I will accept the terms, the long page. So, make sure you read it, it is a generic statement, it is 4 years 5 years old. So, then you go next and then pick a folder in the C drive where you want the software to be installed. Normally automatically, it takes a C drive. C drive is more stable, do not put it on your desktop, do not put it on your other drives where if any error comes it will just get lost.

So, C drive is normally used for installation, create a desktop shortcut if you want and create a start menu shortcut. I click both but if you do not want you can unclick and then click next, click install to begin. So, it will start running but before that in the bottom of the screen you can see the pointer.

It will be asking can you change the software, is the software allowed to change the setting on your system. You should say accept because it is going to put software on your desktop or lapto, you will have to accept it. So, in the meantime while it is installing, I think we should be able to see slowly the status improving, it takes time and also I am recording the lecture.

(Refer Slide Time: 05:25)



So, it will take time but let it go through. In the meantime, we will look at the types of spatial data. So, this is a very important topic that we need to cover for the lecture. GIS data is of two types. So, you download the data from remote sensing and you collect data from observation data. There are two types of data. So, this is kind of the basics of GIS. I will not get fully into the basics but just some part of the basic we will cover.

So, of the two types, the first type is raster which is very important for this course. It is a remote sensing and GIS. So, most of remote sensing, if not all are raster data. So, the data that you are going to use in GIS for spatial data is two types, one is raster one is vector, and the first one raster is also a continuous data wherein throughout the map, throughout the canvas, you will have data not in only some pockets. So, that is important to understand because you will be looking at the status. I will just check the status now as well.

So, here now it is asking me, my software is asking me should I allow the software to run. I say yes. Once I say yes, it starts computing, validating and installing. So, let it install in the background while we discuss the course topic of the data. I will keep it here so that you can also see how it runs.

So, as I said there are two types of data, the raster data is more continuous. Examples are air photographs taken from the conson survey, planes that take data from the ground up and also scan the maps, topographic maps, any image which has fully covered with data. For example, if you have a sheet of paper, so you have all this data in it, all the colors and so which means the data is full, all of the pixel, every pixel is covered with data.

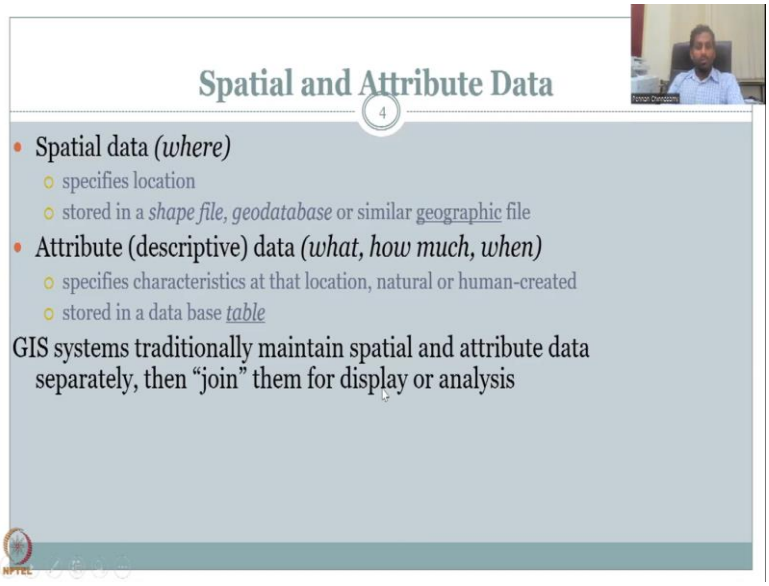
So, like that you will need to look at data as across the canvas or only in some points. If it is across the canvas, it is called raster data. Most remote sensing data is raster data. The vector data is of discrete types where vector is discrete features a layer comprised of individual points, either points, lines or polygons.

Example, statues are points, just point on each location, rows are lines and state boundaries are polygons. We will cover each of this separately in this week's lecture. This week we will look at vector, in particular, in the remaining lectures raster will come next week. I will devote one week of lectures for raster because raster is more important for remote sensing data.

You could see here as I mentioned, most data is, remote sensing data is an raster data, whereas your vector data is discrete features. You can convert between these two data

formats but it is advised that some features are lost. So, to keep these two separate and only change if it is needed. So, let us look at this data concept in GIS, in particular.

(Refer Slide Time: 09:27)



We have spatial data where is the location. So, spatial and attribute data is imported in GIS. Where is the data taken from is called the spatial data. It gives you the specific location of the data. In GIS, each data has a data and information which is attributes but also the location where the data is collected and that is what makes GIS different from other softwares.

So, we should be able to capture that part very very carefully because we have space location data inbuilt in GIS data. So, it is stored in the shape file, geodatabase or similar geographic file, whereas attribute data is descriptive data about what, how much and when. So, this is the actual data.

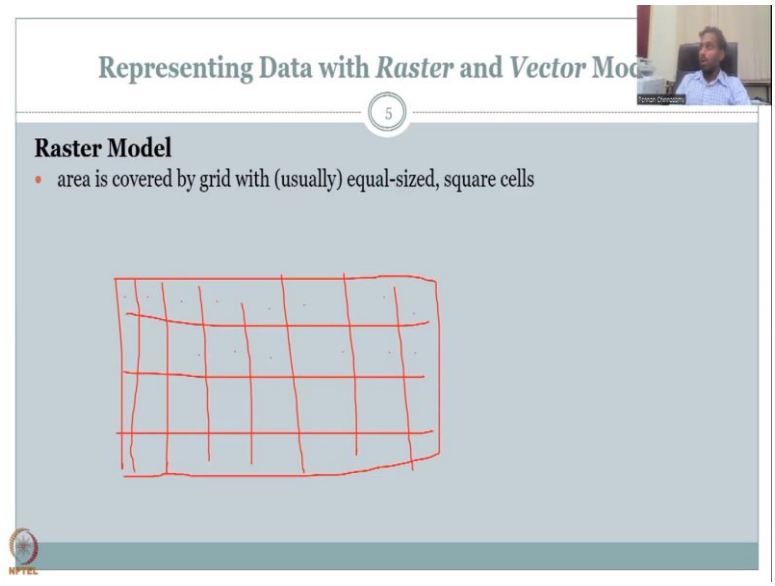
Let us say census data. So, census data is population data. It is data about how many people live, how many males, females, let us say. If I am just giving data itself, it is attributes. You have males, females, total population attributes. But we are also having a location attached to it.

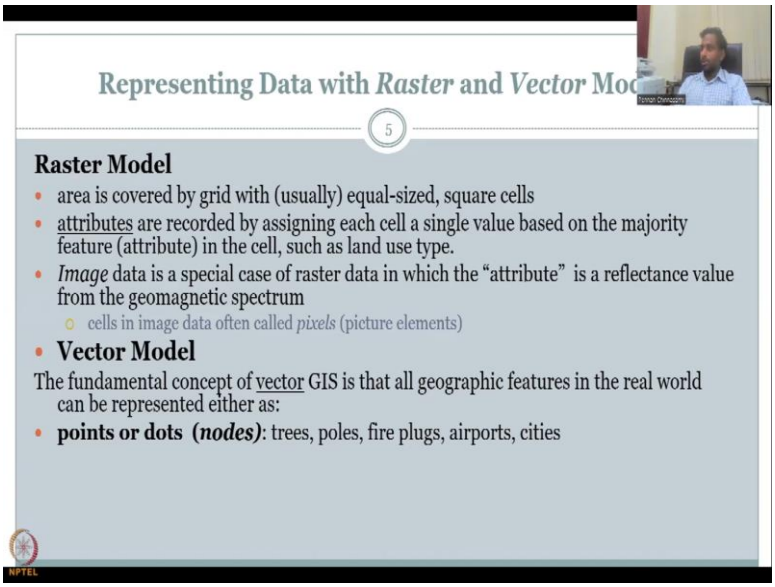
Let us say Tamil Nadu's population. So, Tamil Nadu is the spatial location which is the first bullet, then the second bullet is what data comes into the system. So, these are specific characteristics at that location, natural or human created. These are stored in database table in a table format. So, we have both spatial location now and table data associated with that location in GIS.

So, in GIS normally, they are maintained separately. In a GIS environment, you will see that the data being managed separately. Still copying the files, I will just check if we have done with the files. Still is going on. Let me go on. And what we will do is the, so, we have GIS system separately traditionally storing them in different files but, and they join them and display for analysis.

So, when I talk about the structure of data, you will see that one GIS file will have seven to eight associated files. And in one of the associated files, the location of the data is stored, whereas the data as table is stored in another file. When you open this file, this shape file in GIS software, it combines this and projects the data as one data. So, that is what the last bullet means.

(Refer Slide Time: 12:49)





And then we have representing data with raster and vector models. So, we have here different models of data. So, we have data that represents both raster and vector. We will see in a real-life scenario, how it is done. Raster model, it is area covered by grids. As I said, if you have the map throughout, let me draw it for you. So, you have a map and in the map or the map area, every single area is having the data. This is in raster format. So, how is this accomplished?

It is divided into pixels of equal size. So, these are grids. So, I am not doing straight straight, but in a normal situation, each data is going to be graded. I will show you an example but I am just showing you how, and each point each grid will have data in it. So, that is why it is called continuous equal sized and square cells. Let me clear this page.

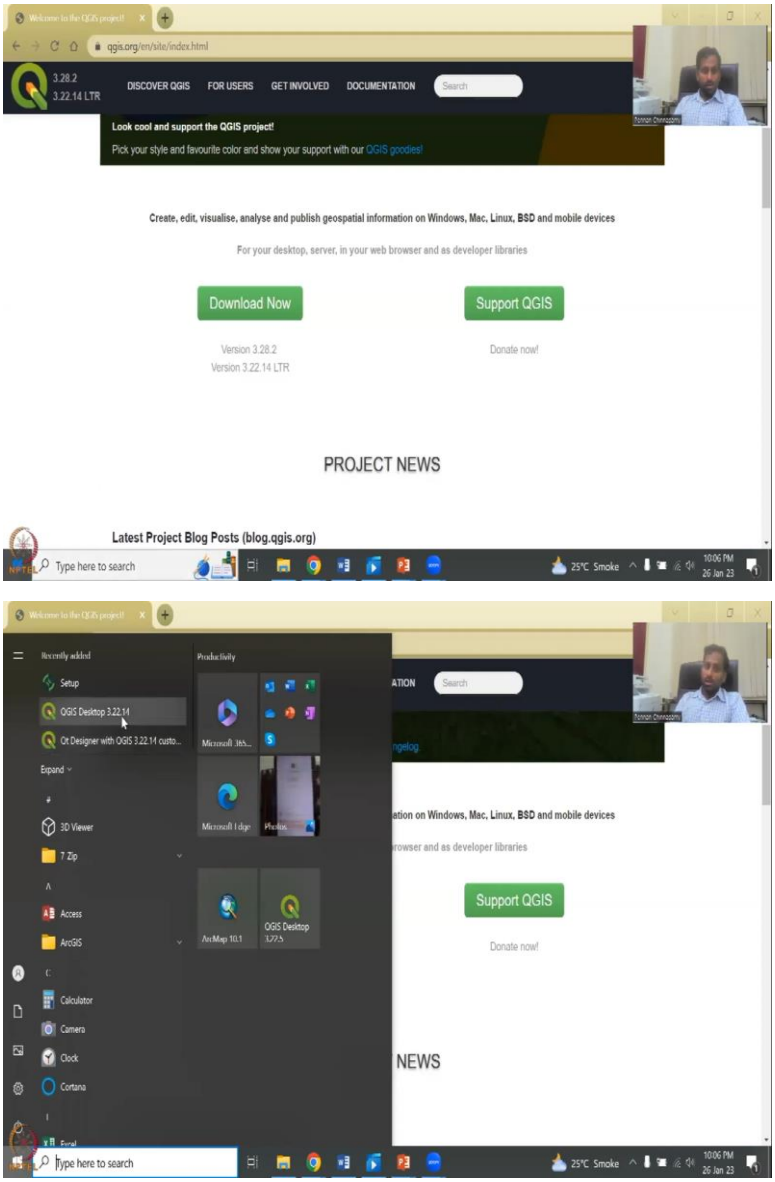
And so, what happens is you have equal size cells, attributes are recorded at each single cell where the value is going to be placed and you will have a majority of the features in the cell, for example, land use. If in a pixel you have 70 percent as forest 30 percent as urban, the pixel will be colored green which represents forest, yes, 30 is urban but because the pixel can take only one color, one value, it will take the majority color.

Image data is a special case of raster in which the attribute is a reflectance value from the geomagnetic spectrum. Again as I said, a satellite or a plane is taking an image, remote sensing object is taking an image of the planet of the land and it is, when it takes an image is the reflectance, how much light is reflected and in what colors. So, when the reflectance happens, the reflectance is monitored and captured in the camera and converted into digital image.

In the digital image, coloring is based on how many colors the camera can capture. So, cells and image data often called pixels, called picture elements. In a vector model, the fundamental concept of vector GIS is that all geographic features in the real world can be represented either as points or data or dots, nodes, trees, poles, fire plugs, airports, cities like, they are as points. So, imagine, we have, I will just check if the software has been updated. It is still running. So, let it run.

You can see it here, yes. So, what we will do is we will have more, so, we have more points or dots representing trees, poles, figures in the vector model. In the meantime, I have noted that the software has downloaded successfully. Let us quickly look at the software for which I will open. So, now, if you could see my other screen, I will just open the other screen so that you can see.

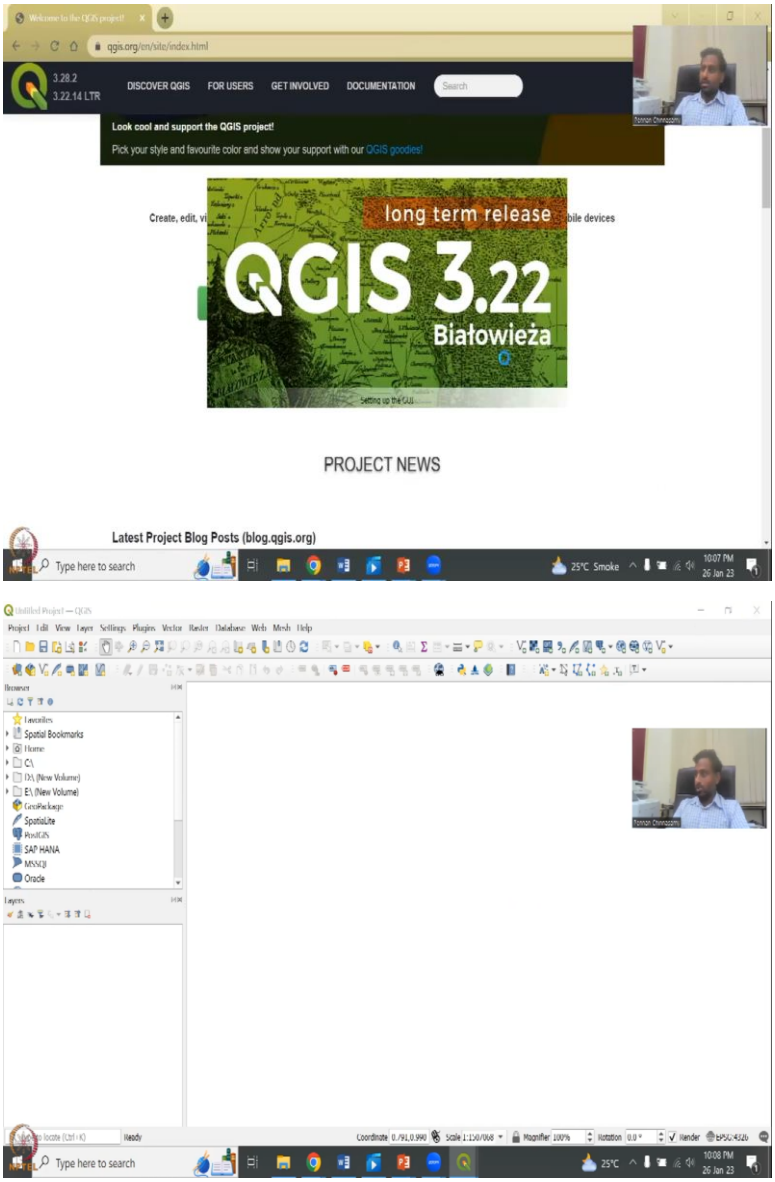
(Refer Slide Time: 17:08)



So, if you could see the screen, this is where we downloaded the software and if you click your start button on the top, you could see that your GIS has been set up. Basically, your GIS 3.22.14 and what location it has come has been there. So, I am going to open the 3.22 14 which is the long stable version and then your 3.28 is your newest version with multiple features.

The feature which one is there. Remember that you should remove your previous ones. I have my older version also but that is the beauty of QGIS. It lets you run multiple versions because some people like the older version, they have the older version and the newer version also in the same system.

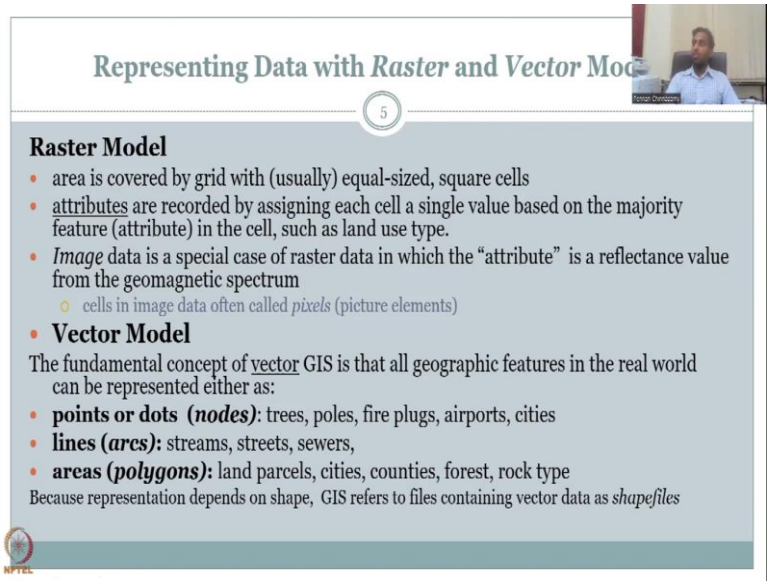
(Refer Slide Time: 18:04)



So, once you click your QGIS, this thing opens up which says that welcome to the long-term release, the version is given and the name. Each a model is given a name Biatowieza or and how to pronounce the last dot is that a. So, these names are taken and given by QGIS and it starts.

So, always you open a new product and then it says some plugin has crashed, you can just ignore the warning but now your QGIS has been set up. So, in the next course lectures I think week 5, 6, we will have hands-on on how to use this. So, I will use this video for that but I just wanted to show you how to install it into your system and then take it up. So, let us get back to the lecture notes that we were presenting. Yes.

(Refer Slide Time: 19:07)



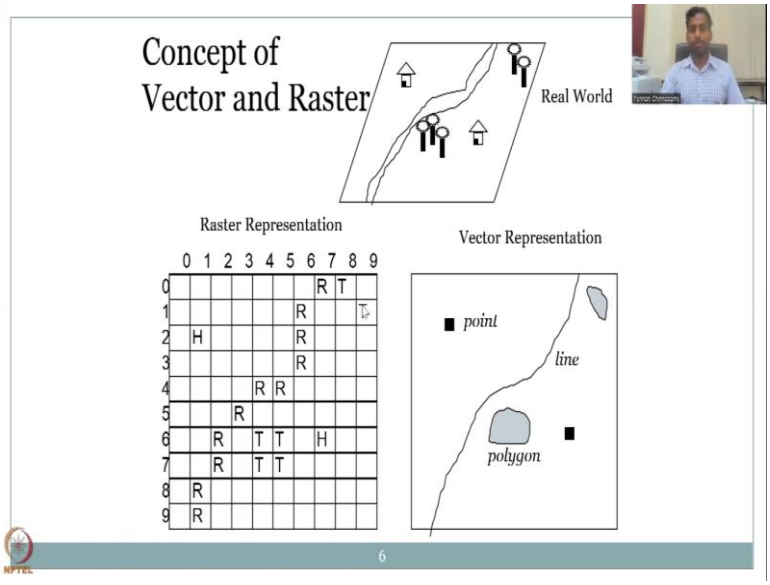
So, we stopped here where we said points or dots, nodes, trees, fireplugs, airports and cities where you have a vector model, a location is taken for these as a point and for the point data is being stored. The data can be anything but examples are trees, poles, fire plugs, airports and cities.

And that is one type of vector data, the other type of vector data is lines where you have a combination of points joined together as a line and the location of the line is also split in different points. We will not see that part but we will definitely see the line in your image and a perfectly placed line in the geolocation. It is not like randomly placed a line, it has a location and it will be placed likewise.

And then we have areas. So, a combination of lines become polylines and then polylines become polygons, if they are connected and closed. So, areas are polygons where it is again starting with points, it converts to lines lines convert to polygons and polygon examples are land parcels, cities, counties, forest, rock type. So, these boundaries can also be put on a GIS map and with the location and with the data.

So, the last part we would like to see here is, my video has gone up let me let me bring it down. The last part we would like to see here is, because representation depends on shape, GIS refers to files containing vector data as shape files. Since it has a shape, point is a shape, line is the shape and a polygon is a shape, we call this as point files.

(Refer Slide Time: 21:08)



Let us look at a real world so that you can understand how the system works. In a real world of concept, the top part is a real world where you have a river flowing. So, the line here is a river and then you have houses, forestries and then houses on both sides of the river. So, now I am going to convert it into the two types of data we have in GIS, raster representation and vector representation.

Let us do the raster first because we have discussed that. As I said the image is first discretized into grids, equal size square grids. The size depends on the spatial resolution of the camera. So, here you could see that. It it has been divided into grids, each grid gets a unique number. This is 0 0 1 1 1 2 1 3 1 4 etcetera.

So, we do not care about the labeling scheme but for now, rows, columns are given nine rows, nine columns and then the grids are being established. So, each grid is of equal size and inside the grid there is only one value for data. You cannot have houses and rivers in the same box. It is either house, river or a blank or a tree.

So, there is only four data types, there is blank, this is land or let us say land and then there is a house, there is a tree or a river, only four types of data that can be represented in this image. So, here in the raster representation, you see the blank which represents the land, you see the H which is the house and in the river going as connected grid as R and then you have T for trees.

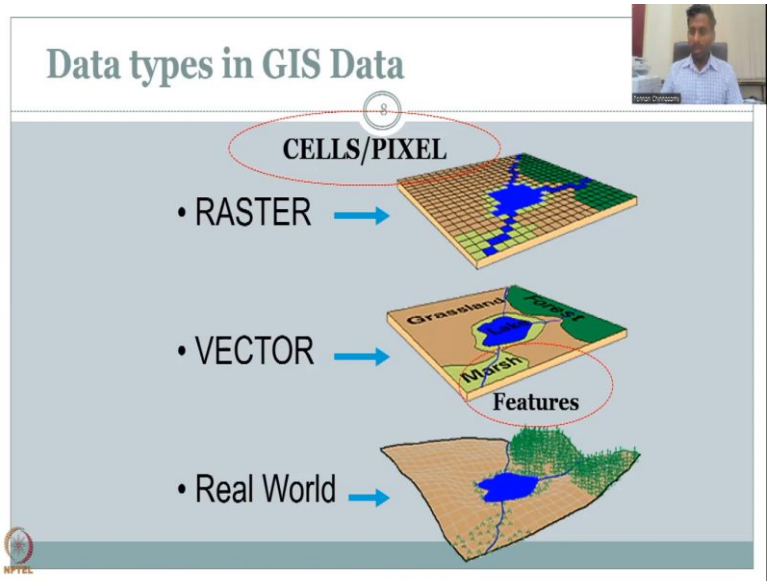
You could see that it is not like three trees are here, four grids are colored. No, it is the space. The space occupied by trees is represented by four grids here and then 1 1 there and then a

house. So, all these are on the location of where the data is being collected. So, this is the raster representation of the real world wherein we have the real world painted as a whole canvas in the raster and but in the vector only the location where exactly the house is.

Here also there is a location but the whole grid is colored as house whereas here the location, exact location only a point is kept for houses. The trees are combined as a forest as a polygon. So, all this area is taken as a polygon and then the river is created as a line. So, here you see the river going through as a line and the data points line and polygon has been covered.

So, this is very important for understanding the conceptualization of a real world into raster and vector. And then we move on to one more example. In the next example I will also show you how the 3D effects are going to be captured or not captured in a raster.

(Refer Slide Time: 24:41)



So, here going from bottom up, we have the real world. You could see that the real world has some imperfections and undulating surfaces, not as flat surface but GIS is about converting this onto a 2D frame, your computer screen is a 2D frame. So, you are going to put this image on a 2D map and then each data point can be converted for the 3D image.

So, in a real world, you have a land mass, some water body, a lake and then streams that get into the lake, a forest and a marshy land. So, this is the real world. Imagine that you are having a real world where you have different types of attributes or data. In a vector world how it looks like is you are converting that into a 2D surface first and not all parts of the surface needs data but still the data is represented as grassland which is the brown.

And then this part is the marshy land where some green is there. So, light green is given. Forest comes on the top north part of your study area. So, that part is a polygon also and forest is being mapped. And then the streams are lines that drain into the another water body polygon which is the lake. So, you have the lake, the forest and the grassland the marshy land as a polygon.

And then you have lines also in this vector representation. You do not have points but that is also okay because point can be used for marshes or forest. But it is not necessary that all vector data should be represented in your image. It can be purely in a map, it can be purely points, it can be purely lines or polygons or a mixture or combination of either, you do not have to have specific all of them in or a partners and stuff. Any number of data as per needed can be used.

Now, come to the raster world. In the raster world, it is not as straightforward. You will have to convert the entire 2D canvas into grids and all each grid has to have a data. There is no part in the grid which says no data. So, you could see that all of them are colored brown and on top of it, there is marshland which is colored yellowish green and then you have the blue which is the river and the water bodies and then you have dark green for forest, the green color is is looking at the forest which is being reflected back as green.

So, what do you see here very carefully is the grid size dominates for these kind of analysis for raster because the thickness of the river can be small, may be small, but if the grid size is big, the entire grid is going to be colored as blue. So, you could see here that the majority is blue. So, all of the grids are showing blue but this line might be smaller, this line might be smaller than the top line and that cannot be differentiated in a raster because the raster will still think that majority of my grid is blue. So, I will color it blue. So, this is the only issue and it is not as clean.

So, you could see that the grid is just connected 1 1 1 1 grid to the lake. So, it is not as as natural as a river flows. So, it does not flow as blocks, correct. So, in raster, you will see that the data flows as blocks whereas in a vector it is smooth as lines and stuff. However, it is not continuous as raster. So, that is where there are benefits of using either data set. So, one data set is not better than the other data type, it depends on your use, how you are going to use the data and what you are going to use the data for.

So, moving on the differences here we see are the raster data has cells and pixels. The grids as I said is called cells or pixel or you can interchangeably use them across your

understanding. Grids is equal to cells and pixels and each of them have a property that across the grid, across the canvas, across the image, the grid size is the same and it is a square grid. You cannot have rectangle on the top and then square on the bottom.

Also the square should be of same size, the size again is divided and estimated by the cameras and the data acquisition system. So, we will not get into that but understand that whenever you download a raster data, you should understand what is the grid size which is a spatial resolution and the grid camera or whatever has been used to form the grids.

In the next part which is vector, you have features. So, data is divided as cells, pixels, grids in the raster whereas data is divided as features in the vector type. So, here you could see marsh, lake, forest, those are names but those are also features under which data are stored as attributes.

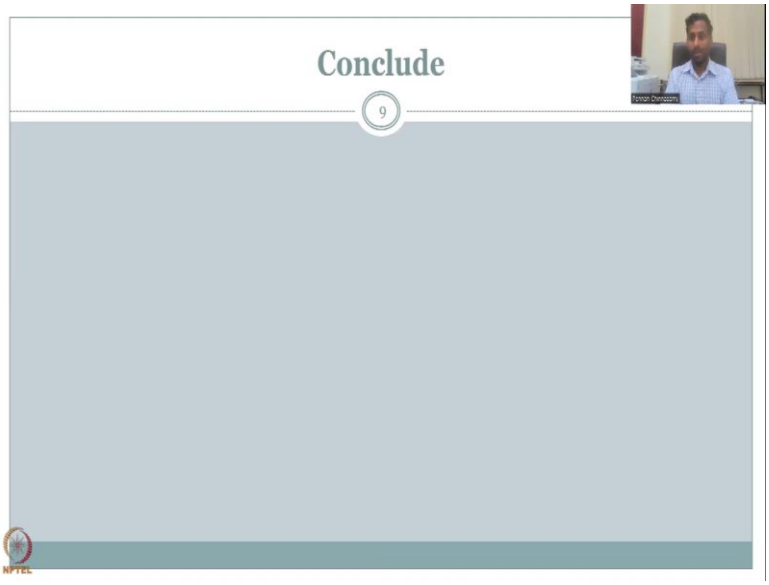
So, you have marsh. So, in the marsh you can have type of marsh, size of marsh, length, breadth, all these things can be added as data. The major item is your feature and within the feature you have attributes. So, it is like your column name could be your feature class whereas your rows in a table could be your attributes.

The grassland is a type of feature, the lake is a type of feature. The entire table can be called as land use, land cover classification and that is being used to represent the real world scenario. Again, the real world is very complex but you are trying your best to capture it in a raster or a vector format.

Please, understand that because there is a lot of grids and more representation continuous data, raster is always bigger in size. It consumes more memory storage compared to vector. Vector you can show a lake as a point, a line or a polygon only part of the map area is going to be used whereas for raster even though a small lake is used the entire image has to be classified. You cannot just classify the lake grid area, you will have to classify the entire area.

So, these are the pros and cons of using the data types. There are two data types in GIS, raster and vector and raster is where a continuous data set is created by discretizing the image, forming grids whereas in vector the important features that are needed to be met are identified, they are converted to point, line or polygon and then they are input into the system.

(Refer Slide Time: 32:32)



With this I will stop today's lecture. I will see you in the next lecture where we will specifically look into the vector data set and how it can be used for rural development whereas in the next week, we will look at the raster data set and how it is used for rural development.

Again, those who would like to learn more on these type of data sets and GIS, please, take the recommended NPTEL courses and or any courses that talk about GIS data types. It is the same, it is not different because different people teach it. It is a very basic, simple entity. We do not teach it here because then the load on GIS will be more rather than rural development.

The understanding is you will learn this or you had already learned GIS data types before you came here. But it is fine, if you do not, please, learn it in the site. We will be using that in the rural development concepts. Thank you.