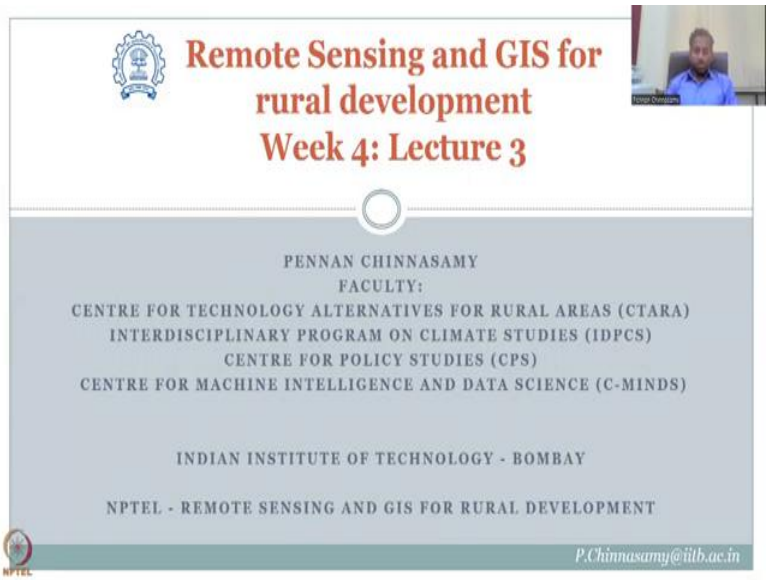
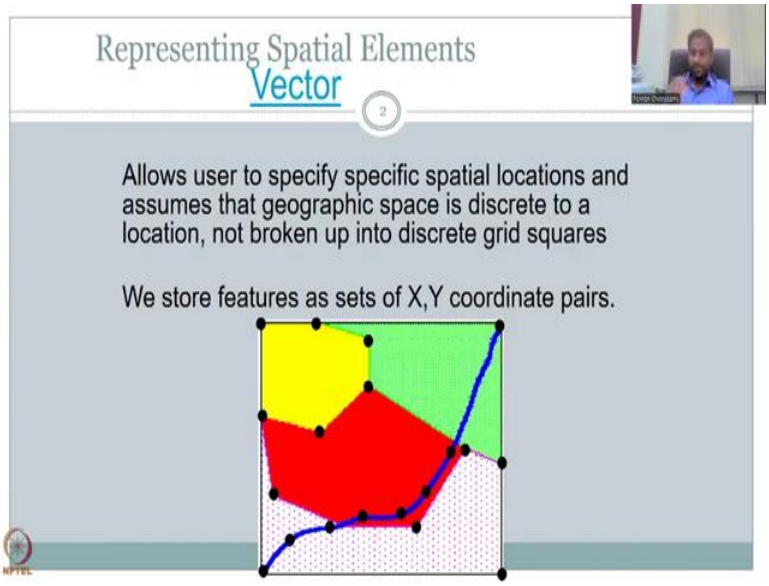
Remote Sensing and GIS for Rural Development
Professor. Pennan Chinnasamy
Centre for Technology Alternatives for Rural Areas (CTARA)
Indian Institute of Technology, Bombay
Week 4 - Lecture 03
Intro to GIS vector data type and QGIS panel

(Refer Slide Time: 0:19)



Hello everyone. Welcome to Remote Sensing and GIS for Rural Development NPTEL Course, this is week 4 lecture 3. In this week we have been focusing on using QGIS for rural development especially using vector datasets. So, we have introduced the concept of vector and raster which are the two data types in GIS and now we will continue the focus on representing spatial elements as a vector.

(Refer Slide Time: 1:02)



There are two data sites, data types but we will be looking at only vector in this section, whereas raster will be done in the next week. This is important to understand that there are very important differences between the dataset and the steps involved in processing this data step in QGIS is also different, so please follow on what are the basic differences.

In GIS always there is a spatial location associated with the data, we have already discussed this when we did GIS. In the vector format, the data is stored as features as sets of X, Y coordinate pairs. So, it assumes, GIS assumes that geographic space is discrete to a location not broken up into discrete grid squares like your raster, here discrete means it is specific to a location you are not putting it in a grid but it is a location along the planet we have different unique X, Y coordinates X, Y coordinate is not going to be the same, inside the grid space the X, Y coordinate is assumed same because we are taking an average value of the grid when a vector notation each object will have a unique discrete X, Y location.

Let us have a look at it. Here you have an image a data and in the data you have multiple types of features, objects. So, you have a point object, you have a line object which is the blue line the point can be any point black dot that is on the screen and you have polygons the red, green and yellow polygons you can see.

So, what vector does is it breaks a polygon into lines, lines into points so because point is the very basic data entry type that is required in the vector. So, you have a point and for the point there is X, Y location stored as a pair and for the point whatever data comes in as attributes is being added.

Now, the line is also a set of points however, the points are connected, so the points are connected and these are called vertex points in along the line and they determine the shape, the turning of the line and the length. So, along the point, along the line you have points where X, Y coordinates are stored and the attribute is only one for the entire line, however, the shape and the location of the line is controlled by vector points.

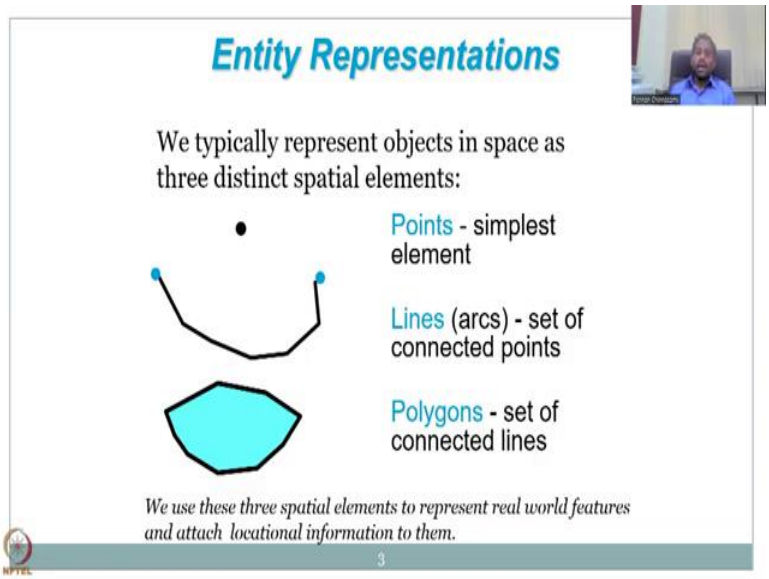
Similarly, a polygon is just different lines connected together and closed and these lines are made up of multiple points. So, you see a lot of points here that have been spatially arranged along the lines of the polygon and each line as said earlier, each line will have unique X, Y pairs but the attributes that go in is common.

For example, in this red polygon you have 1, 2, 3, 4, 5, 6, 7 points, all the 7 points will not have the same X, Y pair because they are spatially different, so they will have different pairs.

Let us say this is the red box or red polygon, in the red polygon we have 7 points that control the shape of the polygon and they have X, Y coordinates stored separately.

Now, the data for the red polygon is only one, it could be land use land cover, multiple multiple different types of attributes but not all 7 points get the data, the points are just used to preserve the shape, the size and the location, whereas the overall feature, the polygon will take attributes which is common to the entire area.

(Refer Slide Time: 5:58)



Entity representation as I said earlier is done as three distinct spatial elements a point which is the simplest element, the point will have an X, Y pair and which is stored as a location specific dataset and then there are projections and coordinates that we will see in the next phase.

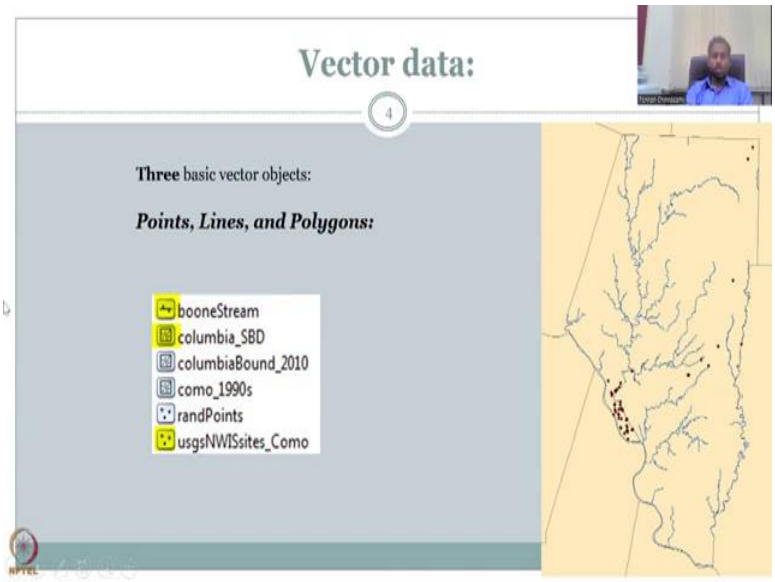
Then we have lines which are basically set of points connected to each other by a line, the shape of the line, the length of the line is determined by points. Again,

in a real world you do not get straight lines between points, you will have to break it into multiple points, so you will have a point here, point here, point here which I can draw for you in this.

So, you can have a point here, you have a point here, you have a point here and point here, these are called kind of vertexes in your lines and these hold the location, hold the location, shape and size of the line. So, the line is anchored on the points and then the length and the shape is determined.

Then we move on to the next which is the polygon which is a set of connected lines and each line has its own points as discussed earlier. So, we use these three spatial elements to represent real world features, features or objects and these objects have location information which are stored as X, Y coordinates X, Y pairs and they are stored separately in the files.

(Refer Slide Time: 7:42)



One more example we will take on to see how the dashboard or your desktop or laptop shows these kind of shapefiles. So, we have three different shapefiles colored here, the bottom one is the usgsNWISsites_como which is the points that we have, these are basically location of rainfall gauges, water discharge monitoring stations.

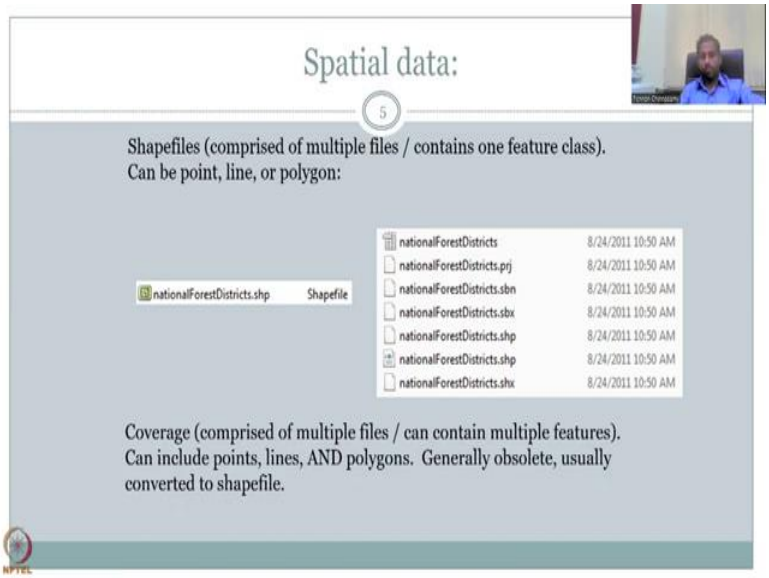
So, this is your study area and you can see that the points where the monitoring stations are placed are anchored as points X, Y points and their point shapefiles. The icon of the shapefile also tells this information because you see points not lines or polygons, so here is a line, here is a polygon.

Then the River or the stream is called boone, so boomeStream is represented by a blue line and that blue line has a line icon in the shapefile, so each shapefile has an icon and the icon is definitely easier to visualize because of the icon which shows that is at a point a polygon or a line, polyline.

Then we have Colombia_SBD and the Columbia_SBD is bounded by a black line you could see here and these are polygon, you have a polygon and then the others are there which we are not highlighting but here the highlighted black line which is here as a polygon you can see.

So, you could see a polygon inside the icon and some dots, the dots are to anchor the polygon but as we discussed in the previous slides these are also made of lines which in turn are anchored by points. So, in this vector data there are three types points, lines and polygons, all of them are anchored on the map by X, Y coordinates and the X, Y coordinates are given to points not the line but the point and the polygon, because they are broken into smaller lines and smaller points data is stored as points, the location files. And then each file has its own attribute which we will be seeing soon on how is the data is stored. So, let us see how the data is being stored or visualized on the screen.

(Refer Slide Time: 10:37)



So, normally when you download a shapefile for rural development, so here I am going to use porous data as per districts. Shapefiles are not single files by themselves they are comprised of multiple files, they may have multiple feature classes inside them can be point, line or polygon.

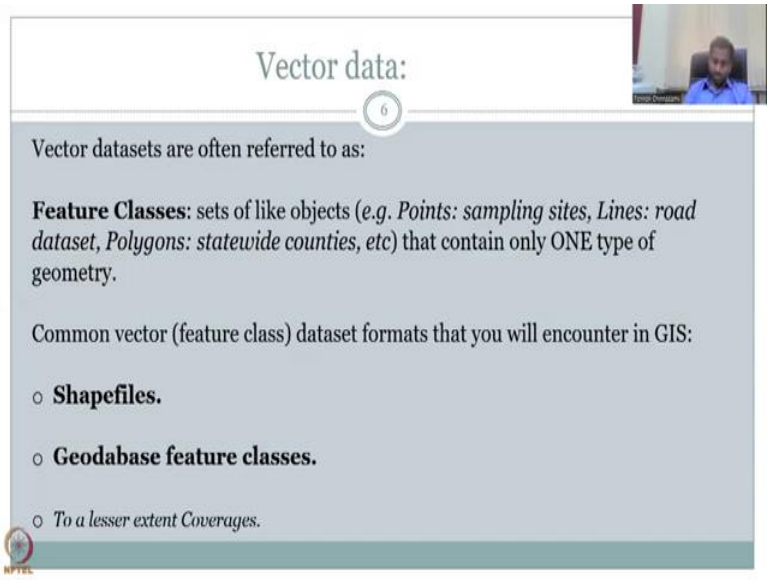
But more importantly each file has 6 to 7 associated files. Here in this example you have 3, 6, 7. So, you have 7 files associated the number changes based on the size of the data and also the orientation of the data but here in this example the national forest district shapefile has 7 associated files. You could see that the naming is same, the name national forest districts is the same, only the extension changes, so dot shp is the shapefile and then there is another dot shp HTML file but more importantly there is a dot shx, there is a dot prj, the dot prj stores the projection and then the coordinates and then of the X, Y location is also stored.

So, you have all these stored in multiple formats, however, all of them open when you click open this national forest district shapefile. This shapefile holds the point, line or polygon and

it is the same name given because it holds the shape it is called a shapefile and it has associated points, if you are sharing data, QGIS data or GIS data please make sure that all the seven files are sent along with the shapefile, you cannot just send the shapefile and assume all the data goes through it will not because the associated files are not going which means the data you open on national forest districts will not open.

So, be careful in sending all the dataset and or make a zip folder, put all the data into one folder, zip it and then send it. So, that is the way you should be sharing data or even downloading the data when you download the data make sure you download all the extensions not just the national forest district dot shp. So, the coverage comprised of multiple files, contain multiple features, can include points, lines and polygons, these are converted to shared files in the GIS model.

(Refer Slide Time: 13:37)



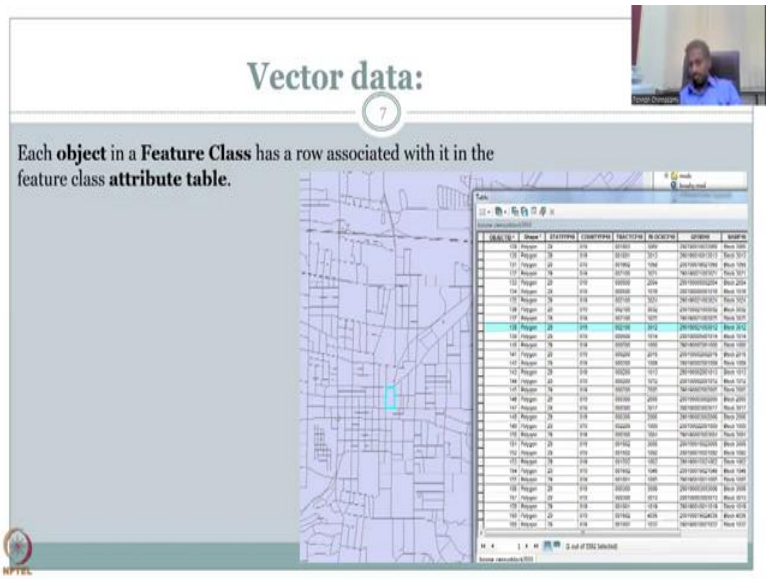
Moving on, vector datasets are often referred to as feature class, sets of like objects, example, points, sampling sites, lines, roads, dataset, in polygons just in statewide countries that contain only one type of geometry. So, you cannot mix and match if there is one shapefile you can only have polygon point or lines, you cannot have one shapefile name and within it points plus polygon plus shape lines.

So, that is why you saw here that it is distinct the icon is distinct, you cannot mix and match. Within the shapefile there are feature classes sets of like objects, for example, if you have points file you can have sampling site, one every district, one every state etc and then say site name is different, state name is different, that is fine but you cannot have different geometries.

Then you have lines, shapefile with a road dataset and all the road names the features are different and then we can have a polygon state-wide counties or districts map where all the district names are different. The common vector feature class datasets that we will encounter in GIS include shapefiles which is what we have been seeing or a collection of the shapefiles in a geodatabase which is called geodatabase feature class, it is kind of advanced we will not get into most of that. And then the last one is to a lesser extent coverages, it is called coverage and when you open the coverage you will have multiple shapefiles put in together.

This is important for looking at long term research work if your project is small you can use shapefiles if you are looking at multiple datasets and you want to organize them you can have geodatabase and if it is very large you have coverages.

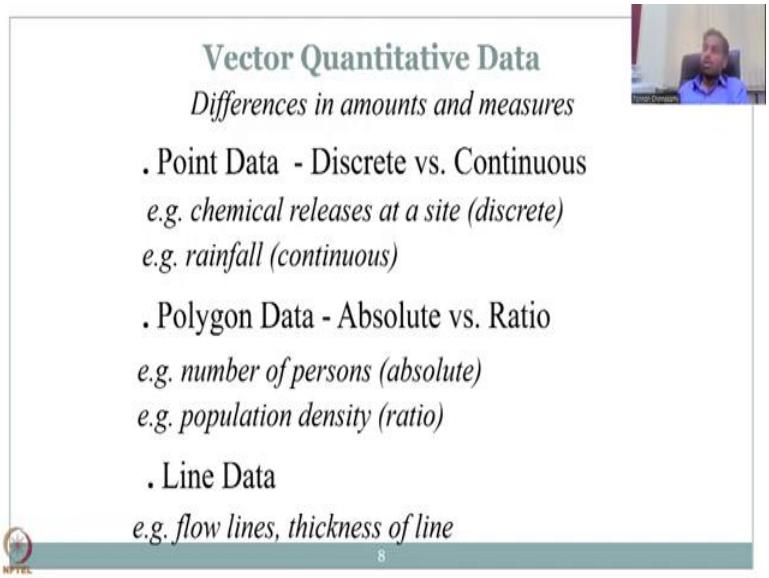
(Refer Slide Time: 16:00)



Now, we have discussed about what is the vector and how it is being stored etc what is stored inside it, when you click a vector database a vector shapefile on a GIS platform each object in a feature class has a row associated with it, so you have a feature class and have objects. So, here this is a district boundaries and land boundaries, in the land boundaries you have different objects so this is one object, this is one object, this is multiple objects.

So, you have one object very unique and each object has a row and this row is called the attributes, so that is why this entire table is called attribute table. This is important to understand that all data are not the same and it is uniquely stored for each object in a row. So, the row might have multiple columns but for simplicity and for the model to work you should not go above 250 columns. So, each object can have up to 250 columns but if you go above a certain limit based on your computing speed and power, your GIS software will lag.

(Refer Slide Time: 17:28)



So, now let us look at the quantitative data that we have. So, vector data is quantitative it has numbers and it has measurements and amounts and measures. So, let us see how we could represent vector data and the differences in amounts and measures. Point data can be discrete or continuous. Discrete means for a point location one sample data taken at one point is called discrete, one point in time and one point in space is called discrete, whereas continuous is the data is collected at multiple time intervals and a continuity is set, even though daily, then it says the daily continuous data if it is weekly, it is a weekly continuous data and so on.

Example of these kind of discrete is chemical releases at a site. So, you have a chemical industry and if the industry is polluting the Rural Water Resources every evening say 9 o'clock in the night where people cannot see the pollutant that is called discrete because at 9 o'clock they open it not every day but on some days where the load is enough to be released.

Then you have rainfall, continuous where the data is being collected at continuous intervals and at a particular location the data is collected, the location is not moved because government is paying certain kind of subsidies and lease amount so you will see that the data at a point can be both discrete for water quality and other measurements or it could be continuous like rainfall.

Then you have polygon data which is absolute versus ratio. In this example you will have data as a particular value for a polygon or a ratio which is a comparison of different polygons, let us say absolute, you have a district map and in the district map which is a polygon if the population is given then it is absolute value let us say 1.1 million people in that district.

But then if you say that I have a ratio it could be a comparison between the districts and that is also available as polygon data. So, number of persons, number of people living in a particular location X, Y or a polygon a particular location people living is called absolute whereas ratio is the population density wherein you take the population and divide it by the number of people or put an average that is population density.

So, the vector data can also be discrete continuous absolute versus ratio that is what we are trying to establish here. Inline data you have multiple thickness of line data example flow lines thickness of the line and you have multiple set of rules that can come for having different line segments in a particular shapefile.

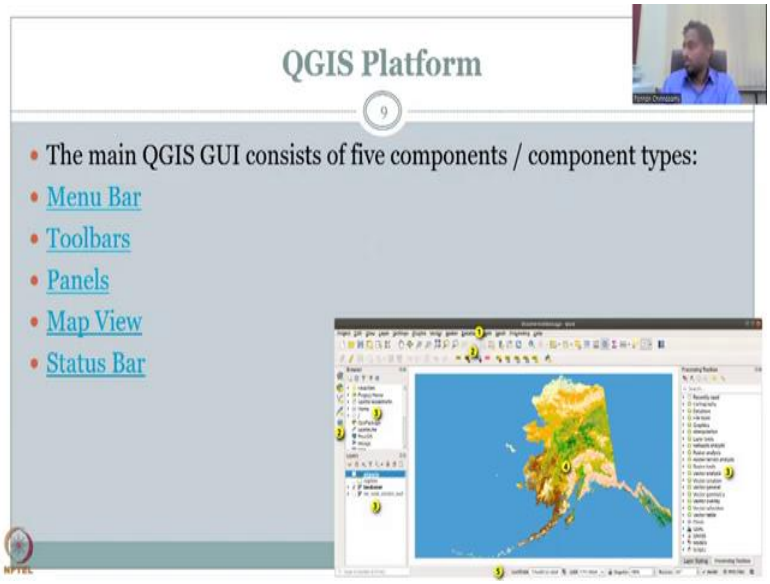
So, normally lines are singular so if we have a river line in the GIS database we will have only one river across the space but if you are adding multiple river layers you can adjust the thickness so that you have differences between a stream going into a big river let us say the Ganges this river is there you have multiple straw smaller streams that flow into the Ganges and in the map it should not be the same thickness because if it is the same thickness we cannot differentiate between a long stream river which has multiple inlet points and large discharge volume compared to a small stream that collects and brings water to the river.

So, these kind of attribute, changes, thicknesses you can do in GIS and it is necessary for visualization because all data is not the same, so there is a need to differentiate between the continuous data and non-continuous, there is a need to differentiate between absolute value and ratio and the flow thickness.

Also note that the absolute versus ratio is kind of used to see who is contributing to the data and if the vote let us say vote is a data and people are not voting in a certain location then quickly people can look at absolute values which is number of person and the population density which is a ratio, so the data allows you to shuffle back and forth.

So, now we have completed the discussion a very short introduction about vector data, we will get into the setup of QGIS which is going to be common for both week 3 and week 4 in terms of GIS presentation and week 5 where we will be looking at rasters. So, I am going to share my other screen where we have set up the QGIS platform. So, let me open it open it.

(Refer Slide Time: 23:37)



So, what we will do is we will be looking at the QGIS software where we will look into uploading of a particular dataset and we will look into the menu bar that we had earlier discussed. You will be seeing the GIS startup let me pull it back again so that you could see also the startup of the screen.

So, my QGIS is starting up where we will be looking at the different components of the QGIS and how it is being set up. So, what we will be looking at is in a QGIS if you open and add data so just this documentation, I have taken from QGIS website, you could see that the drawing board, the real estate area on a QGIS canvas can be divided into menu bar which is on the top and this is the bar where we will go and look at vector tools for this week and you have a browser box and layers box and then a processing tools.

Tools are extra tools that are not put on the toolbar or menu bar because menu bar will mostly have details that bring in from the major major setup of QGIS. For example, saving the file, printing a file, sharing a file, zipping, editing, all these are allowed in the menu bar whereas the processing toolbox may be different, so please do not confuse between these two setup of bars. So, we will have the menu bar on the top where it has the project, edit, view, layer, settings, plugins, vector, etc in the next class we will look into the vector box and see which are the tools that are available and what are they used for.

Then we have the toolbars which is below the project edit view, there is a set of tools and you could place these tools as per your need if you think that now too many lines to so here you have two lines of toolbars if you think that it is too much for your image and it is taking a lot

of space then you can limit it to only one a tool bar line, you can see here there is one line and then you can see here there is the second line.

So, there are two lines but you can always pick and choose how many you want, in my example you have labels on this toolbar and then you have the basic toolbar of save, pinch, zoom in, zoom out, search, attribute table and then other things summations, totalling, areas, etcetera. You can also pull and push this up and down let me see if I could open it along with the presentation.

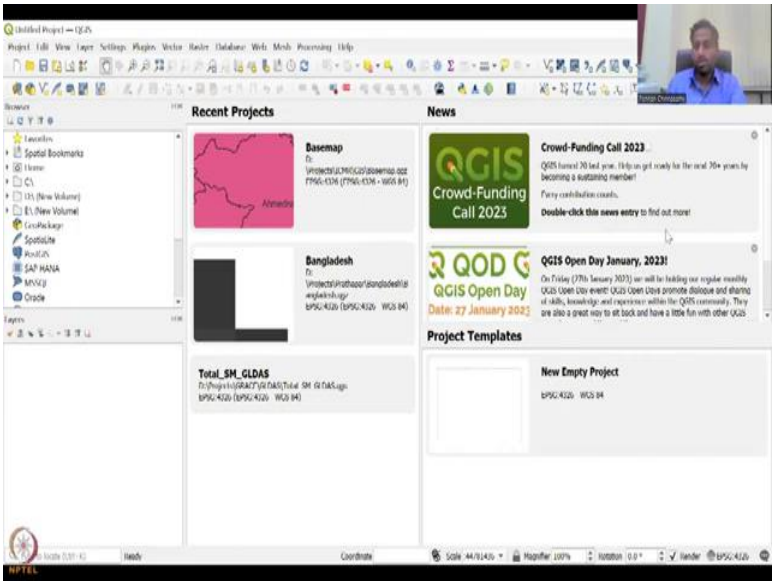
Then you have number three which are the panels, in the panels you have the top panel and the bottom panel the top panels you have a browser where you can use and access multiple two sets but we will keep it away and number three normally I do not use a browser I only use these layers because I want to see what layers are added in my mapping document so that is very important you need to bring data and then put it in the browser in the layer panel and from the layer panel it gets projected into the mapping area.

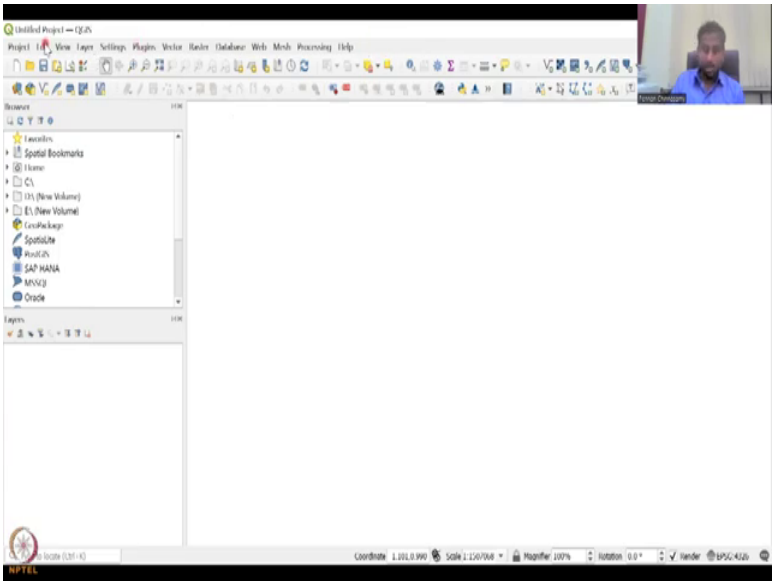
Also note that along the side some of you will have a line here where tools are there these tools are the same as these tools and these tools are the same as file view tools. Please note that you are going to use only one toolbar so it will be good to bring this out and put it here so that you save a lot of space.

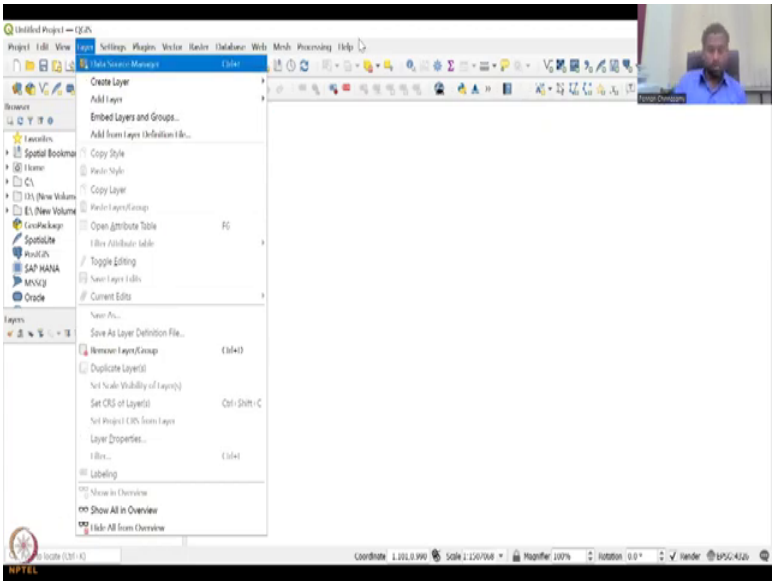
Then we have the map view area or the real estate we call in putting the map and doing the analysis this is the space where you hold for doing and visualizing your analysis so you have to have as much space as possible to bring the map. If you clutter it on the top on the sides with unwanted windows, then your mapping view would be very, very small.

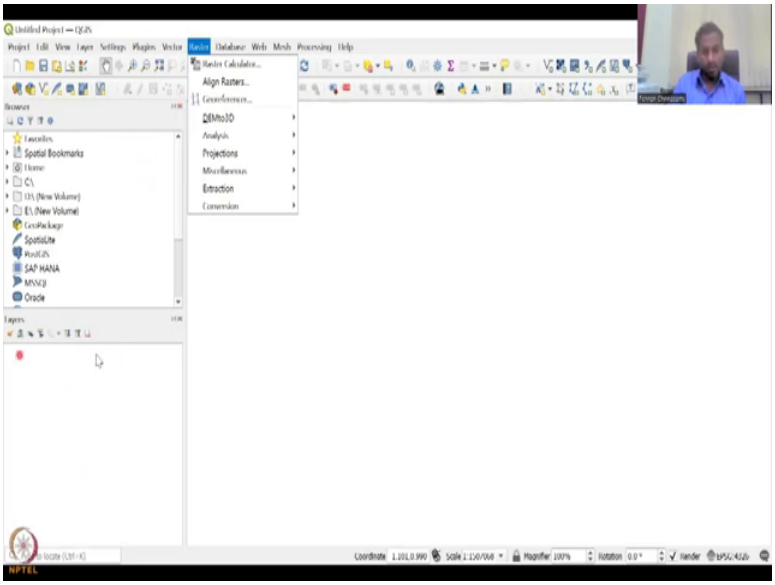
So, basically remove those by pushing the side windows out if they are not necessary just close it, keep the map view document high so that you have more data that you can visualize and then you have the status bar on the bottom these are default so number 1 and number 5 are default, number 2 number 3 you can play with so that number 4 expands, so number 5 is the status bar which gives you the coordinate system the projection and also the process if you are running a process it will update here if the process is running slow or not. So, I am going to see if my QGIS can also be open for you, yes, it can be open I will have to re-share my screen, so let me reshare the QGIS window.

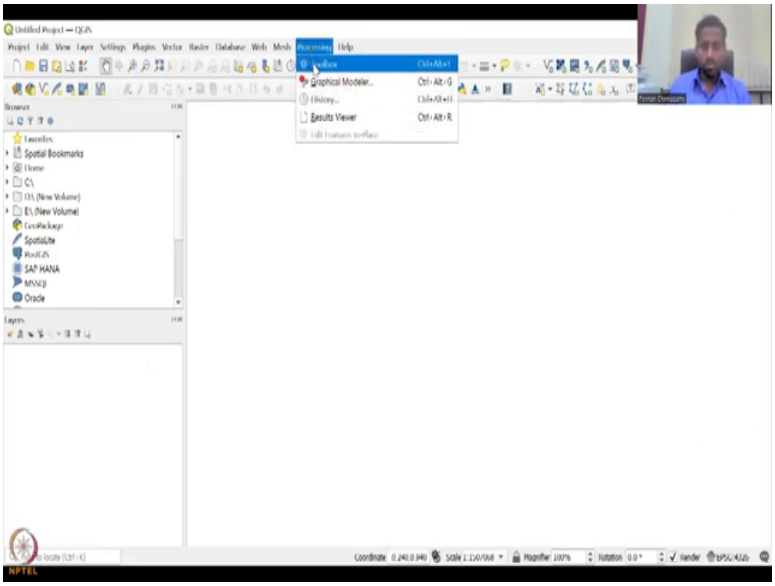
(Refer Slide Time: 29:35)

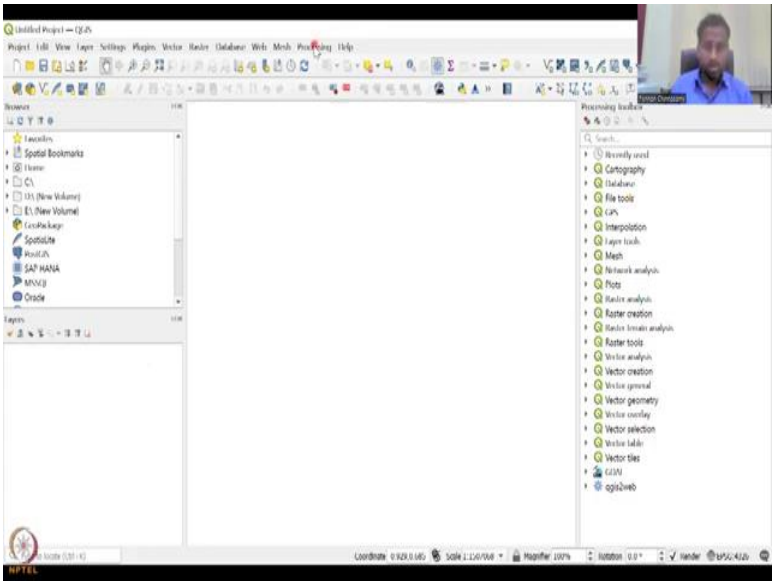







Untitled Project — QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Raster Calculator...
Align Rasters...
Georeferencer...
DEMto3D
Analysis
Projections
Miscellaneous
Extraction
Conversion
Browser
Favorites
Spatial Bookmarks
Home
C:\
D:\ (New Volume)
E:\ (New Volume)
GeoPackage
SpatiaLite
PostGIS
SAP HANA
MSSQL
Oracle
Layers
NPTEL


Untitled Project — QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Toolbox Ctrl+Alt+T
Graphical Modeler... Ctrl+Alt+G
History... Ctrl+Alt+H
Results Viewer Ctrl+Alt+R
Edit Features Toolbar
Browser
Favorites
Spatial Bookmarks
Home
C:\
D:\ (New Volume)
E:\ (New Volume)
GeoPackage
SpatiaLite
PostGIS
SAP HANA
MSSQL
Oracle
Layers
NPTEL


Untitled Project — QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Browser
Favorites
Spatial Bookmarks
Home
C:\
D:\ (New Volume)
E:\ (New Volume)
GeoPackage
SpatiaLite
PostGIS
SAP HANA
MSSQL
Oracle
Layers
Processing Toolbox
Recently used
Cartography
Database
File tools
GPS
Interpolation
Layer tools
Mesh
Network analysis
Plots
Raster analysis
Raster creation
Raster terrain analysis
Raster tools
Vector analysis
Vector creation
Vector general
Vector geometry
Vector overlay
Vector selection
Vector table
Vector tiles
GDAL
ogis2web
NPTEL

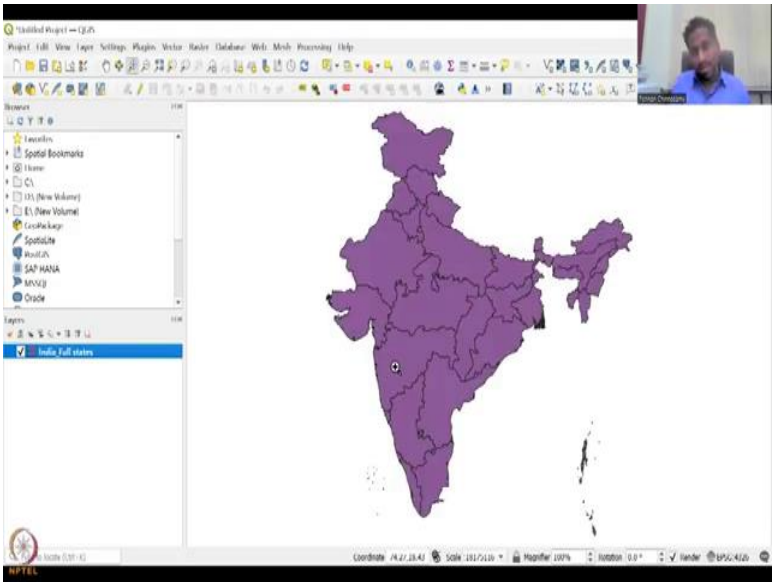
So, when it opens a new project, it will ask and you want to support our crowdfunding, base maps that you have used or you can start a new project. As I said I do not have any disturbances on the site, you do not see my toolbar on the side, all my toolbar is on the top and you can pull and push by just moving the mouse and the mouse turns into a four crosses and then you can pull the four crosses and then put it up or down.

So, these are very flexible to move you can easily move these toolbars, the menu bar is standing there and it is very important so do not use or close all these. The browser bar you can close by closing the button normally I close it when I have too much of layers, so layers is where the data is being added and once the data is added you can use your toolbars to manoeuvre around it.

I do not have a processing box, but you can also see the processing in this toolbox if I click it, it opens so I go on the top I click processing and then toolbox it comes if I do not want it I just close it, it goes and now you could see that I have a bigger area for my mapping document. So, this is important as is because you are going to look at different datasets, you are going to add data and then look at how these data can be used for rural development.

So, what I am going to do is I am going to look at different datasets in GIS format and see if we could use it for downloading datasets, etc okay so these are very important because you will need to see if you have access to all the data in your modelling sphere and performance. And what happens is you will also look into different data requirements in your project and see if you can have different modelling availabilities and venues here.

(Refer Slide Time: 32:10)



So, I have added the data here, I paused the screen share for a second so that you do not get distracted on the datasets I have but this one dataset is very important the boundary of India, please make sure that India has to be represented full, you can and because in many of the downloadable datasets you will see part of India's head marked as a different connotation not as part of India or it will be saying as under dispute and then you will see this part or not mapped properly, this part not mapped properly or maybe this part is given to some other location.

So, be careful as per law for the Indian government you cannot have an map without the India head currently depicted as like here so all the datasets that I use are from the government websites where they take care in bringing these data together. So, with this I will bring back our representation slide as I said this is a different country, so do not worry about accuracies in Alaska US but this was used to look at the status bar and stuff so which I showed in the QGIS explanation, I will continue the explanation in the next class. I will stop here, thank you.