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

Lecture-59
Errors in GIS and Key Elements of Maps

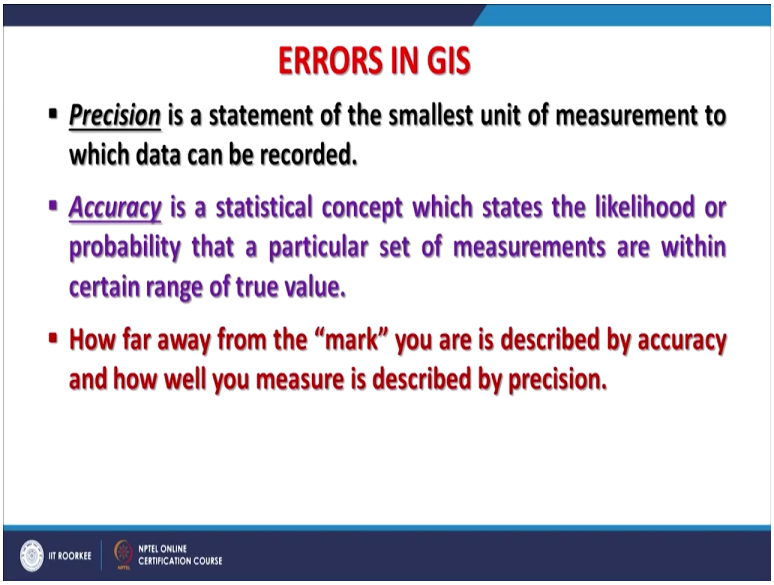
Hello everyone! and welcome to a new discussion. This is the second last discussion of this 60th lectures, 30-hour course. And in this one, we are going to discuss two main topics which generally they do not find space in Literature, neither in softwares but very important one. The one is errors in the GIS and the other one is the key elements of maps. Because ultimately, one has to create outputs and outputs in GIS generally, in form of maps or charts.

May be sometimes we have to create some figures for tables but mainly maps. So, what are the key elements of maps? What are the things which we have to take care while creating these outputs? This is what the discussion we will have. So let us start first with the errors in GIS. This topic is the least discussed in literature. And therefore, people do not take it sometimes seriously, especially the beginners who just started the using of GIS.

So, for them also, it is very-2 important. Nevertheless, 2 important terms which are here, which we must understand along with the errors in GIS. One; the precision and the precision basically is a statement of smallest unit of measurement to which data can be recorded. So, precision depends on the scale which we are using or the instrument which we are using. I can give you an example from our day-to-day life.

Like if we are having one foot scale. We are having generally this is what the practice is, at least in India that on one side you would have inches, another side you would have centimeters. So, you know that against 1 inch, there are only ten divisions. But against one inch, it is 2.5 cm and 25 divisions, because it is 25 millimeters. So obviously the centimeter scale on the same foot or scale, we are having which is having more precise or precision.

(Refer Slide Time: 02:58)



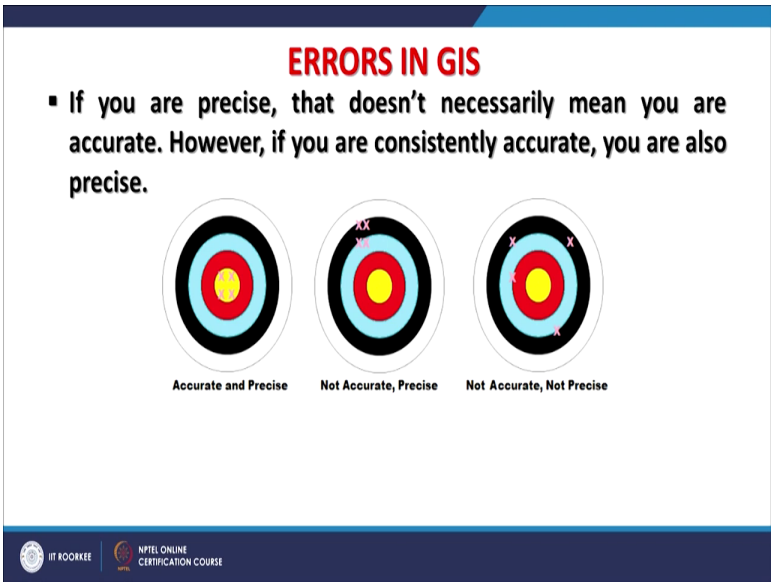
So precision is a treatment of smallest unit of measurement which can be done or recorded by an instrument. And another term which is accuracy. People many times take it synonymously which is completely wrong because these two terms are completely different. As you can see that accuracy is a statistical concept. Basically, it states the likelihood or the probability that a particular set of measurements are within certain range of true value.

If you recall your schooling days and especially the physics experiments, our teachers used to say you know measure or you do this experiment 5 times then take an average. So, an average that is the statistical and sometimes the teacher will say no, your values are very far from the real one because they have measured by using some other techniques. So, this is a statistical term and which basically indirectly tells us that how accurately, we are doing the measurement.

But the measurement which we are using a scale or a sensor or a system will have its own procedure. So, 2 terms; precision and accuracy very relevant in errors and discussion in GIS but completely different. So, should not be interchangeable or should not be used synonymously. Both are different terms. Now as you know that in case of like accuracy that we say that it is acquired by this % or maybe with some units also So, it may fall within 5% so that how far away the mark which I have determined after doing 5 experiments or 5 measurements, that you are describing by accuracy.

And how well you measure, is described by precision. So, our measurement because we are using a scale or a sensor so that will control my precision and how accurately I will do that I think that will decide my accuracy.

(Refer Slide Time: 05:26)



For example, many of us must have played this dart game and in this you know, there are circles, sometimes for the shooting also. Or in other sports, these kinds of targets are used. And as you can see that in the first one, that it is accurate and precise when we are hitting in the inner circle. And that means in that red part or red circle whereas in the second example, the one which is not accurate but precise because precise is that because all our markings are all almost in the same area.

But not accurate because we are very far from the centre point or bulls' eye we also say in short. Whereas the third example is not accurate and neither precise because all our hits are scattered all along these circles and they are quite far from the centre point of bulls' eye. So, sometimes we may be accurate but not precise. And many times, we may be not even accurate and not precise. So, one has to be very-2 careful that if you are precise, that does not necessarily mean you are accurate.

However, if you are consistently accurate, you are also precise like in the first example. Now in GIS, if we talk in terms of GIS that maps and other data which be used in GIS are obtained by

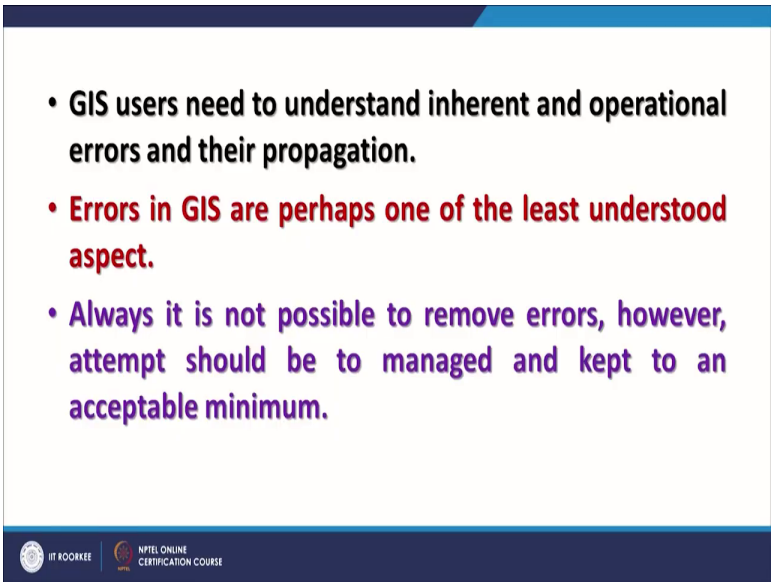
measurement in the many times in the field or may be employing some sensor like Remote Sensing or maybe some laser-based sensors. And therefore, inevitably these data sets which we will be using will have errors.

And these errors as you know also in GIS which we have discussed several times that error propagates in GIS. So, even no matter how small error is, we must understand one and we try to manage that error so that it does not propagate because a small error in the beginning may end up in a large way and our modeling ultimately or our output may have large errors so that people will not accept.

Our level of confidence will also be quite low if our data or our analysis having errors. So, we must take care about even an error is small that we know. And this is why I have I mentioned that the GIS users need to understand inherent. So, errors can be of two types that inherent errors that they are inbuilt basically with the data. And sometimes we have to use the data because we ourselves cannot go in the field and recollect the data as some organizations have already done.

For example, if I use the survey toposheets like in India, Survey of India is the organisation which prepare the toposheets for us.

(Refer Slide Time: 08:58)

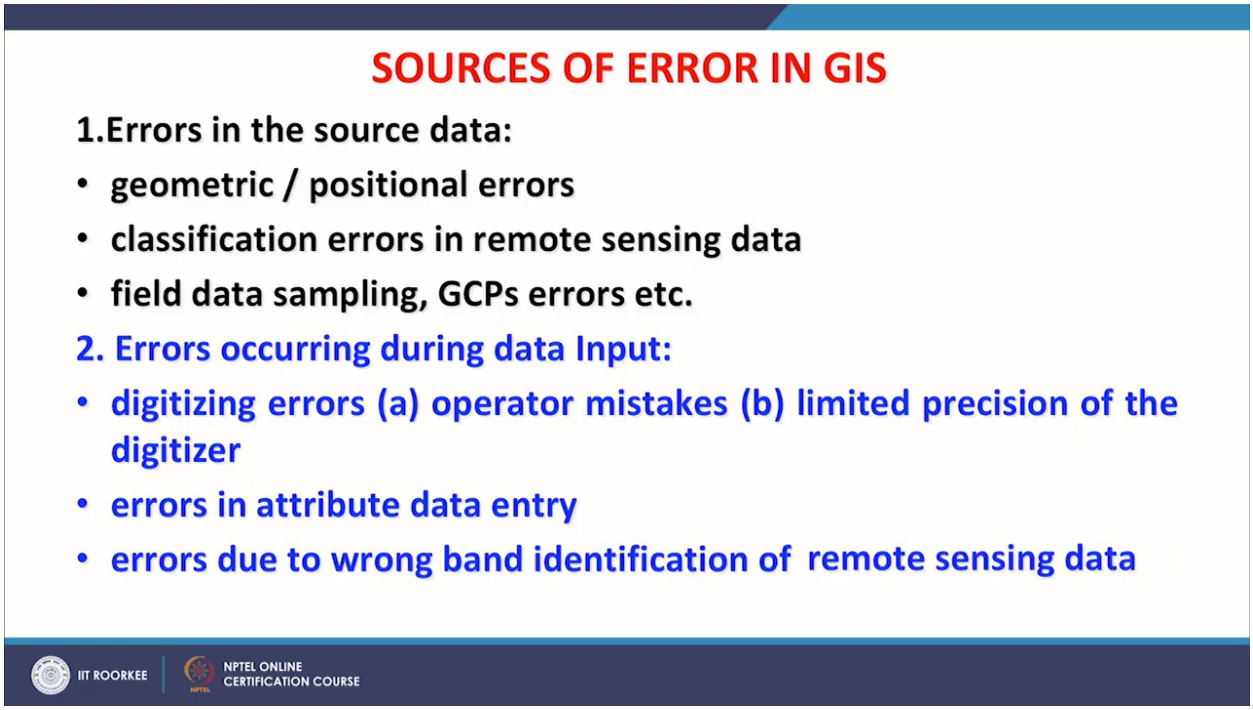


But if it itself is having errors which is having, inherent areas must be there so that we should have the knowledge of first inherent errors and the errors which are operational errors, which are under our control. And of course, if these errors are there and they are not contained then these will propagate in our output. So, this point I have already mentioned that errors in GIS are perhaps one of the least understood aspect and also least discussed aspect.

Many books on GIS you would find, they do not have a chapter or a paragraph for errors in GIS. And of course, the software manuals or online help will also not discuss about the errors. So how to detect errors especially the operational errors that will come through our experience and focusing on that also. That we should not ignore even there is a small error and we should keep checking. That is why I have said short of one of the rules of the GIS is that after each and every step, check for errors because error propagates in GIS.

Now, sometimes you have to live especially with errors that it is not possible to remove all errors. However, attempt; a GIS user must try to manage them and should kept at the acceptable minimum because if it becomes large, people will not accept our output. Further, we have to also understand which are the loopholes or leaking points about the errors; sources of errors in GIS?

(Refer Slide Time: 10:53)



So, first one is Errors in source data. These we will put in under this category of inherent errors because we take lot of data from outside. And nowadays, lot of organizations are preparing data and providing which are completely compatible with GIS platforms or GIS software. If they do not declare that what kind of inherent errors are there then definitely, they are keeping users in dark and we may not be able to tell that how accurate our analysis is?

So, we must explore that possibility. That is why when while discussing the sources of digital elevation models, I also discuss again each digital elevation model which are available free of cost, the associated errors. And most of the good portals will always declare in their metadata files that this kind of error is there. So, we must be aware about that. So, these errors in source data can be positional errors or can be geometrical errors or may be errors in case of digital elevation model above the z-axis.

Now, sometimes we get an already prepared output of remote sensing data. May be a vegetation index maps or may be land use maps prepared by some organisation. They might have used some automatic methods for classification. So, classification will always have some errors. No matter if I am doing, I will create but if I am using somebody else classified image data, I must know what kind of errors are there in that classification.

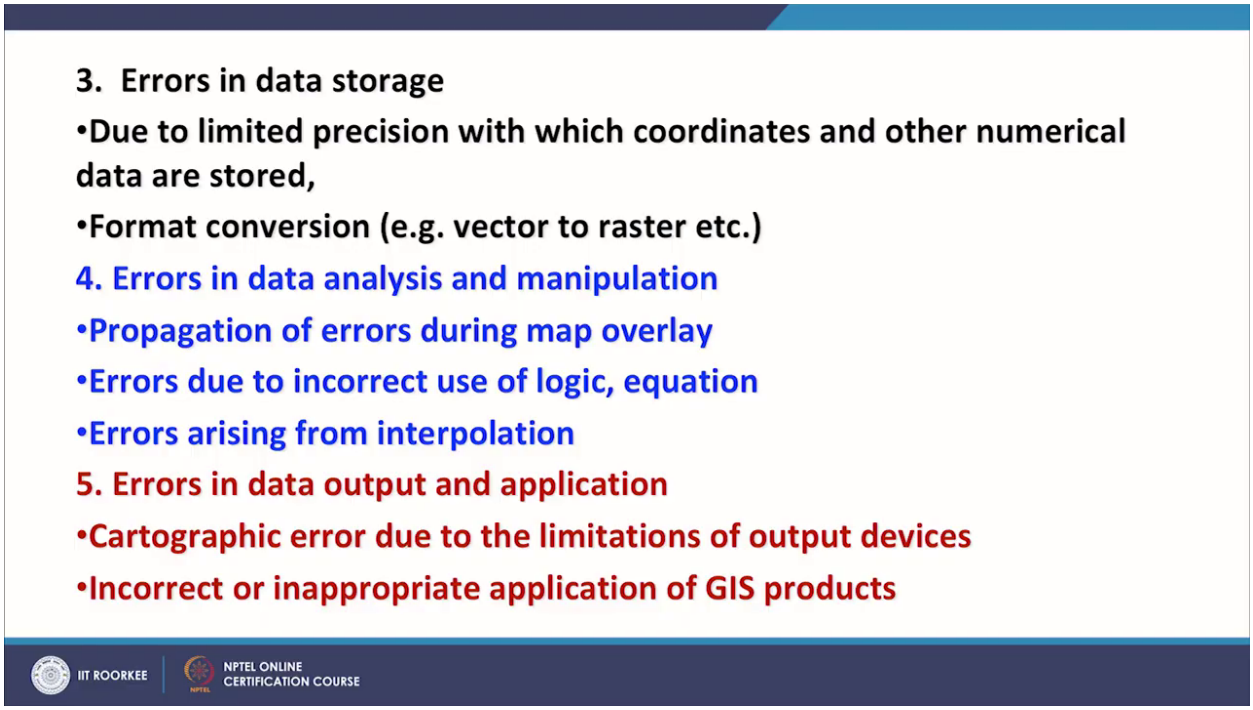
In the field also; field data sampling, measurements, ground control point errors, geo-referencing errors, they will be always there; no matter how small are. Now, second source of errors in GIS is occurring during data input. You know in earlier time when lot of data we have to be key in using our keyboard; type the data. And during that time, lot of errors can come. Also, we use to digitize lot of vector datasets.

So, digitizing will also bring errors. Even today we have to do it many times so if I do the digitization then, there can be mistakes by the operator or there can be you know mistakes ultimately in form of errors because of my screen is not having very high resolution. That will affect the procedure. Or when we used to have digitize assistance, some organization use the digitizer. So, if a digitizer itself is not very high resolution or not precise then it will wrongly record the data.

So, during input stage, errors can come. Errors in attribute data entry; when we have to key in the data using keyboard, at that time also instead of you know just typing say 1000, we just left with hundred or may be 10000 and that will be error. Now errors due to wrong band identification of remote sensing data. This happens many times because now we are not talking about 3 bands or 4 bands.

We are talking about 36 bands, 38 bands and even in hyper spectral remote sensing, we are talking about 256 bands. So, if we have to handle so many bands of remote sensing data then we must take care about that we should know wrongly identified bands. Because each band is representing different part of EM spectrum. So, therefore the identification of band should be also very careful while bringing into our GIS database. Now, third types of errors in GIS are data in storage. Errors, that is due to limited precision.

(Refer Slide Time: 15:06)



I was supposed to typed say 100.58 and I just left 100.5 so with wrong precision if I am not providing, it will carry the error. So, due to limited precision with which coordinates and other numerical data are stored especially this precision procedure will come with the co-ordinate. So, precision means here that to how many places after decimal, we will keep the data. Format conversion because any conversion from one format to another.

While discussing vector to raster and raster to vector, I also mentioned that these conversions are not fully transparent. That means whenever you convert from vector to raster or raster to vector, you will introduce some errors. And these errors may create a wrong output for you or erroneous output. So, whenever we are doing format conversions, we have to be very-2 careful that we keep the errors at the minimum.

So, fourth source of errors that is during the data analysis and manipulation. So, when we are doing the analysis, if we are not taking care, we may introduce some errors in our results. For example, propagation of errors during map overlay. Because when you are doing map overlay of multiple vector polygon maps for example then result map may have hundreds of polygons and then if something is wrong in one of the input maps then the error will propagate in my output map.

Errors due to incorrect use of logic or equation. When we use these map calculators and try to perform the overlay operations even on the raster data, if we have not put this logic or equations correctly, obviously we have wrong results. Mow errors arising from interpolation because interpolation will always introduce some errors especially when it is doing extrapolation where observation inputs are not there.

So, we have to be careful that how much extrapolation a map surface is having. And the final one is the errors in data output and application. This is the 5th source of error. Now that means the cartographic errors due to the limitations of output devices. That my output device does not have a high resolution and therefore, it may create, though up to analysis, I have done everything on a very high resolution but output is on the course resolution.

Incorrect or inappropriate application of the GIS products: There cannot be an application of GIS but unnecessarily we are applying GIS to create certain products and that may also bring errors in our outputs. So, there are various sources of errors. Now here as I also touched that this detection of errors will come only when we start using GIS extensively; that means through experience. But if you follow that rule that after each and every operation in GIS, I will be checking errors.

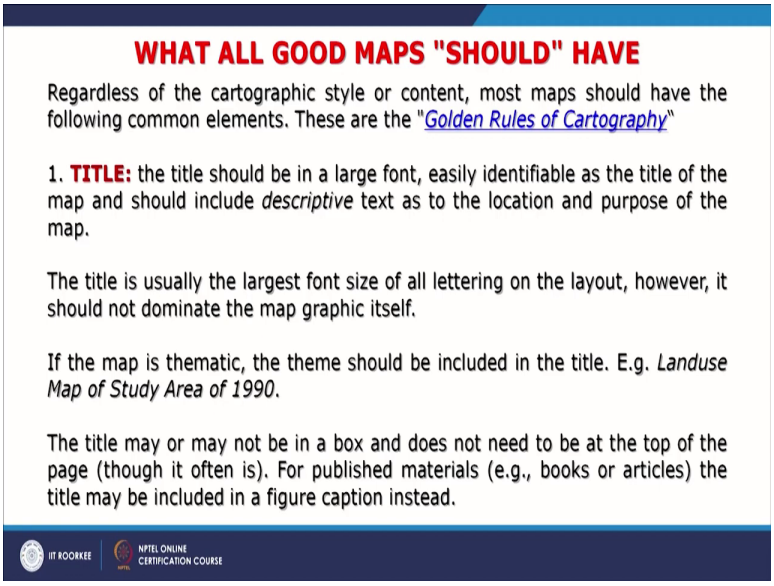
And if found, I will try to minimize or remove errors at that stage and will not leave for others to remove. That will allow us to maintain the errors to their minimum level. Secondly, we must learn how to detect errors. And whenever you get a data set from any source, how reliable that source? We must apply our intelligence and try to check the data. For example, if I have to check my georeferencing then I can check now using Google Earth.

And assuming that Google Earth is the standard one so georeferencing errors can very easily be checked with Google Earth that I will export my output file whether it is a vector or raster into kml or kmlz and then will try to overlay over Google Earth and if it fits where it is supposed to be then that means my georeferencing is accurate. Similarly, for many-2 other operations, we have to invent our own way of checking error and managing errors.

Therefore, it is very-2 important because the reliability of output will tell that how best the GIS works and what device GIS can produce. If outputs are having lot of errors, people will not accept easily. Now the second part of our discussion is that what all good maps should have? Very basic things but again, these are not discussed in the books neither by the software neither helpline; generally, it is not discussed at all.

But these are very important thing because everyone does not have the training of cartography. So, basic concepts of cartography if they are applying in GIS then we will not have that problem. So, regardless of cartographic style or content, most maps should have the following common elements and I call them as map elements. Map elements that these are the golden rules of cartography.

(Refer Slide Time: 21:23)



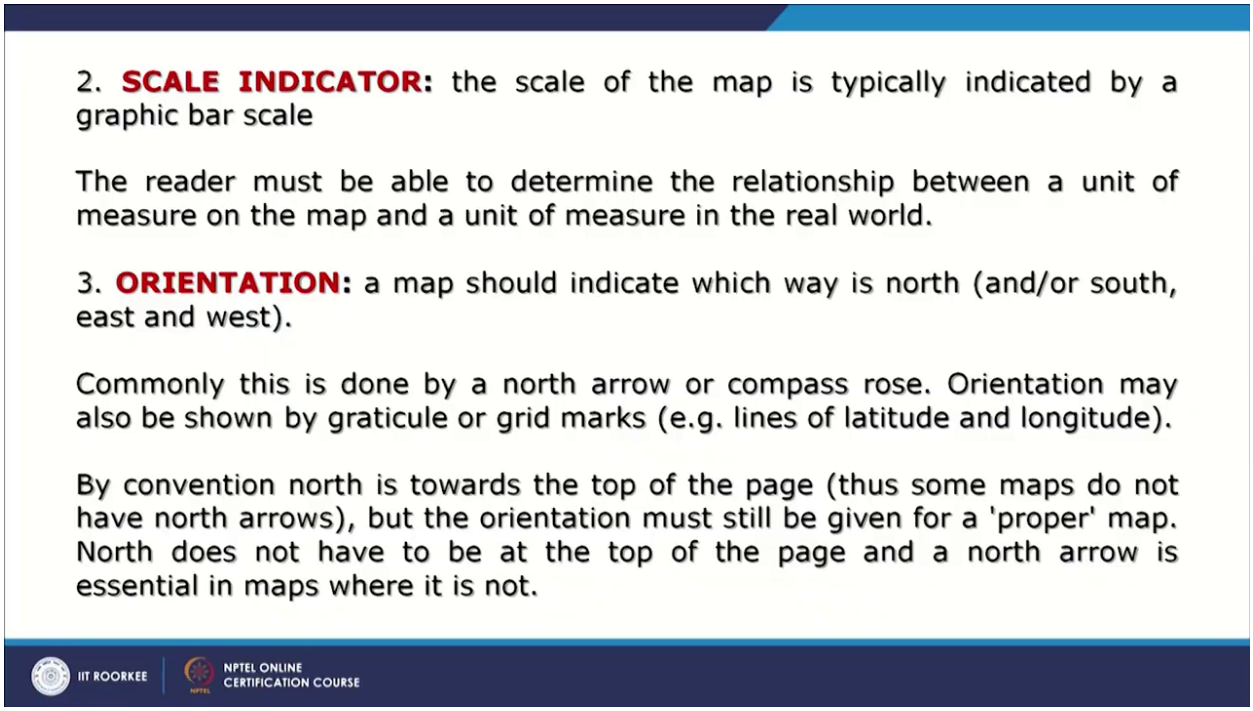
The first one is the title. Each map should contain a title. Generally, we bring the title at the top and that title should be easily identified as what is the map containing. It should also include a little descriptive text as the location and purpose of the map. And that may go as a caption of the figure or map. So, title of map and captions; both are required. Then it will be a complete map.

Now this title usually is the largest font size of all lettering on the layout. However, it should not dominate the map graphic itself. That means the size of the text which we are using for title

should not be very large. So, one has to be optimized the size and if the map is thematic, generally which are in GIS then themes should be included in the title. For example, land use map of the study area of this 1990.

So, we have declared in the title this is the land use map of that area and the year also, it is inbuilt in the title. And the title may or may not be in a box. Because that does not need to be at the top of the page though generally, this is the practice which people follow. But if space is not there; sufficient space for the title, we can bring the title in other positions but generally, it is kept in the top. Now the second map element is the scale.

(Refer Slide Time: 23:17)



And scale indicator is very-2 important and one should not use the ratio's scale because if I use a graphic bar or scale bar and saying that this is zero, this is my 10 kilometers, if that map is reduced by the publisher or by anybody during photocopy, the scale bar will also reduce accordingly. There will not be any error in the scales. But if I write that 1: 20000 and map is reduced that the number 1: 20,000 will not change so that will depict completely wrong scale.

So, ratio scale should never be used in GIS. Always use the graphic scale bar because the reader or user of the map should determine the relationship between a unit of measure on the map and

the unit of measure in the real world. For that purpose, we are having a scale. That for example if the scale is you know like for a toposheets 1: 50,000 so here what it tells that 50000 units of your ground are equivalent to 1 unit in map.

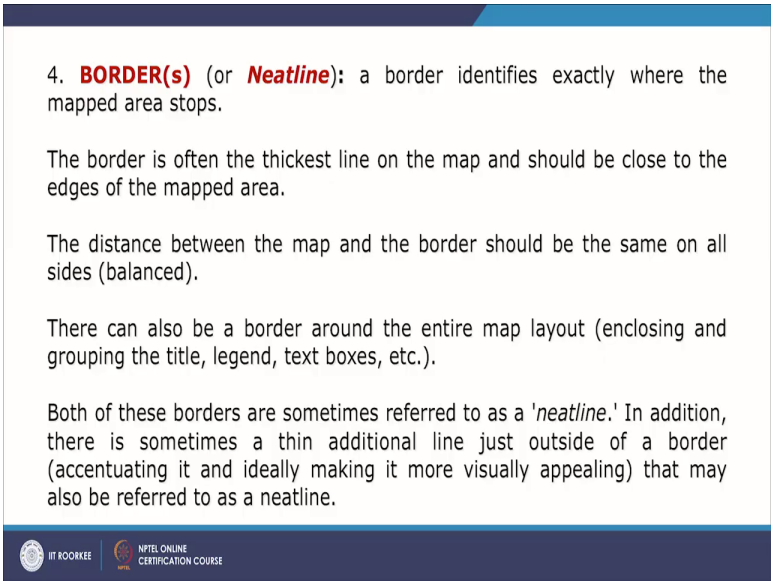
So that the real world has been reduced by 50000 times; something like that. But in case of scale bar, there will not be any problem. So as the map is reduced or enlarge or projected, the scale will also change accordingly. Now, 3rd element may not be that important but it should be discussed; the orientation. Orientation of a map should indicate which way in the North is, that is most important.

Because sometimes if my study area is in a way that if I put in a portrait then it becomes very small in my printout. So, I would like to make it in landscape and therefore I might be requiring to rotate the north. So, we must take care about this orientation part of my map. If north has been rotated, it must be mentioned. Even if it has not been rotated, it must be mentioned. So, putting a North symbol is always an essential part of a map.

So, this is done as you know by using a North Arrow or a compass rose and orientation may also be shown by graticules or grid bars; through a grid which is latitude, longitude grid. Even if you are having latitude, longitude grid or marking on the margins of the map, still you should provide the North symbol because as I have just mentioned that the convention that the north is always on the top side.

And what if space is not there, does not matter. It can be put elsewhere. But normally people will put in the top side of the map, may be on the left side or right side, does not matter.

(Refer Slide Time: 26:52)

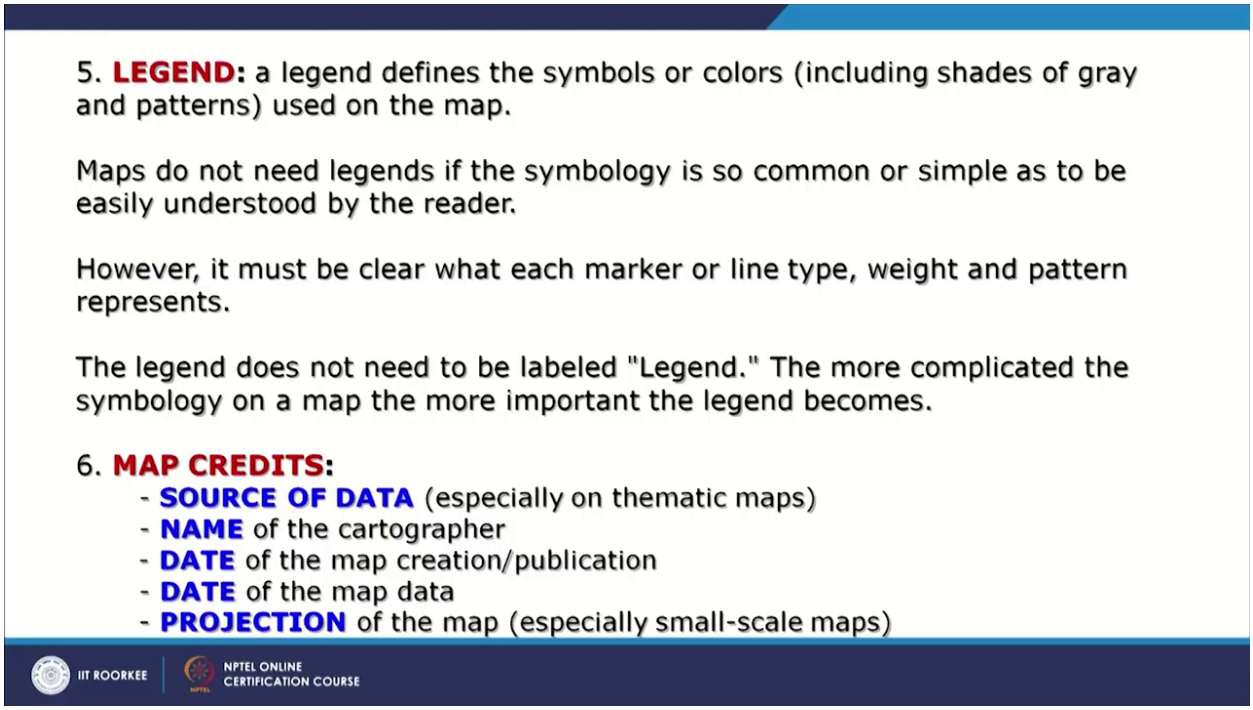


Now the fourth key element of map is the borders. Or also in some software or literature, you may find a different term which is called Neatline. So, border basically identifies exactly where the mapped area stops. That decides the boundary of the work or the study area. And the border is often the thickest line like title is having the largest font size, same with the border. Border is having thickest line on the map and should be closed on the edges on the map perimeter.

So, you are having basically a box. Now this distance between map and the border should be the same on all sides; that means the border should be in a balanced position. And there can also be border around the entire map layout. That means enclosing all groupings as the title, legend, textbox etcetera. And both of these borders, generally it is always that only one border should be put that define this is what the areas is and inside that, everything can come title, legend and other things even north symbol, scale bar and other things.

So, both of these borders are sometimes referred to as a Neatline. And in addition, there is sometimes a thin additional line just outside the border. And this may be accounting and ideally making it more visually appearing so that is part of the decoration. I would prefer that one should not spend much time on decoration. But one must spend time to keep all essential elements of the map before printing and also should always focus on the accuracy.

(Refer Slide Time: 28:58)



Now the 5th key element of a map is the legend. Now there are two terms that people use. Sometimes in certainly; one is legend and another one is index. So, a legend defines the symbols or colours including shades of Grey and patterns used on a map. So, index is used generally, there is no sort of written explanation rule is but legend is used when you use the boxes or symbols or colours for depicting certain area especially in polygon maps.

But index is used when you are having line data, polyline data, point data; there we use the word index. But even if it is interchangeably used, it is not a big problem because sometimes both things are together; legend and index. So why do create two different titles; that one is legend and index. Basically, legend is also index. So, maps do not need legends, if the symbology is common and simple as to be easily understood by the reader or by the user.

See, though there is a lot of standardization of the symbols or symbology which is used in maps. Especially many softwares are implemented symbols of USGS; United States Geological Survey which includes all kinds of symbols. But those symbols may not be used in different countries. So, basically there is no standardization or there is no word agreement for symbols. Nevertheless, one should try to use symbols which are easily understandable to the users and should be logical as well.

And however, it must be clear that each marker or line type, weight and pattern represent because somebody may be using different line type; dash line, dash point dash lines, a thick line, continuous line and how thick and thin, it is and what is the pattern? Because for many geological representations like thrust or faults, we use different symbols along the line. So, these things must be taken care whether if it is north depicting thrust then the arrow or the triangle should be indicating towards the north.

And legend does not lead to be labeled legend. There is no need because it is already there. So, the more complicated the symbology on a map, the more important the legend becomes. How people will read? They first have to check like if you go through a survey of India toposheets, in the bottom right and bottom left, it is full of legends. So, it tells all the symbols which are being used in toposheets.

Though sometimes those symbols, no feature you may find which is shown in the legend but these legends in say 50000 scale toposheets are common throughout all toposheets. But here in GIS, that may be a different scenario. What the purpose here is whatever present, a red line present, a dash line present, a circle is present, a triangle is present or a colour filled polygon is present or a symbol filled polygon is present.

The Reader must know what these things are being represented; for what, these were there so that one should be there. Colours, generally we should use again carefully judiciously that for example, if I am showing a dense forest then I should not show in red colour or grey colour or brown colour. Then a dense forest should always be soon with the green colour. So, this basic concept one must remember.

Another and important one is the map credits of any map element or any map. If I have taken map of somebody else, I must provide the credit. If I have used the data of somebody else then I must mention that one also, may be in the caption. Or preferably if there is a space and the map is not getting cluttered, within the boundary or within the Neatline, we must mentioned that what is the source and to whom you want to give credit for that map?

Because many times we ourselves are not creating maps. We are using somebody else created maps through some publications, Atlas or even from websites. So, we must provide that source from where this thing is coming. If it had been created by ourselves, no problem that means the data has been collected by us. Everything analysis has been done by us and final maps also been prepared then probably you do not require giving credits.

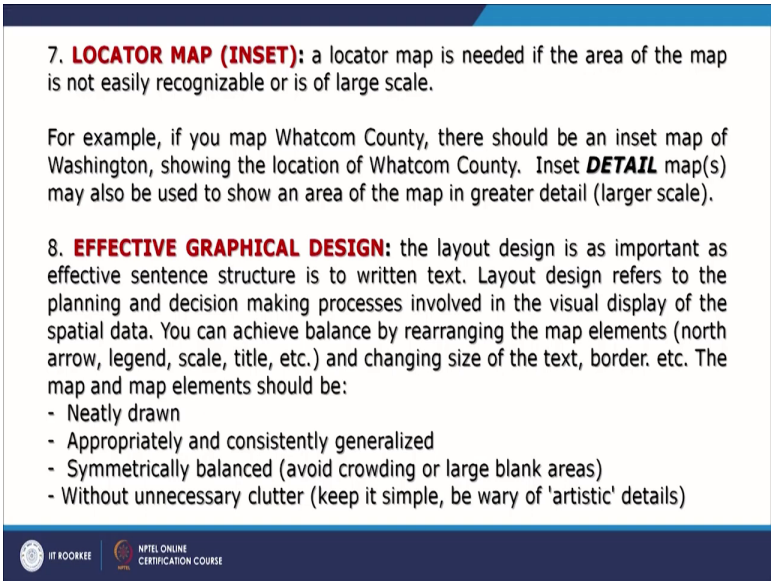
But if many things have come from other sources the credit must score. Otherwise, it will fall in the plagiarism; case of cheating or copying. So, source of data, that will also save you that especially on thematic maps. If I have taken the say contour lines then I must mention that these control lines are coming from survey toposheets. If I am using some forest cover information, I must tell that this is coming from there.

And for large maps generally, this is also a practice that who has prepared that map; the cartographer, name of the cartography is also mentioned. And other good practice is to put the date. So, we would know that when it has been created because these are though may sound very small-2 suggestions but sometimes this becomes very-2 useful. And therefore, one should keep remembering all these things. Not only when map has been created but the age of the data if possible.

That will also help many people those who would use your outputs that what was the time of the data collection. And then projection because this is one part or one information which is generally missing. And in a which projection, your map is because again there is no is standard one and each country is having their own map projections. So, if I am preparing a map which is covering many countries, I must mention. Even if it is of my own country, I must mention.

Because in GIS, we can change from one position to another quite easily. So, the project information should also be there that this map has been projected to this projection. Now another important one that if I have prepared a report then each map should not have an inset map or locator map.

(Refer Slide Time: 36:47)



But at least the first map should be having a locator. It will tell us that which is my study area say within India. So, you just put the boundary of India and boundary of states and within that, you may in form a box and that locator map can go inside that unit area within the main map in one of the corners. So, locater map or insert map is required. If a map like Whatcom country should be an inset map of Washington, showing the location of Whatcom country.

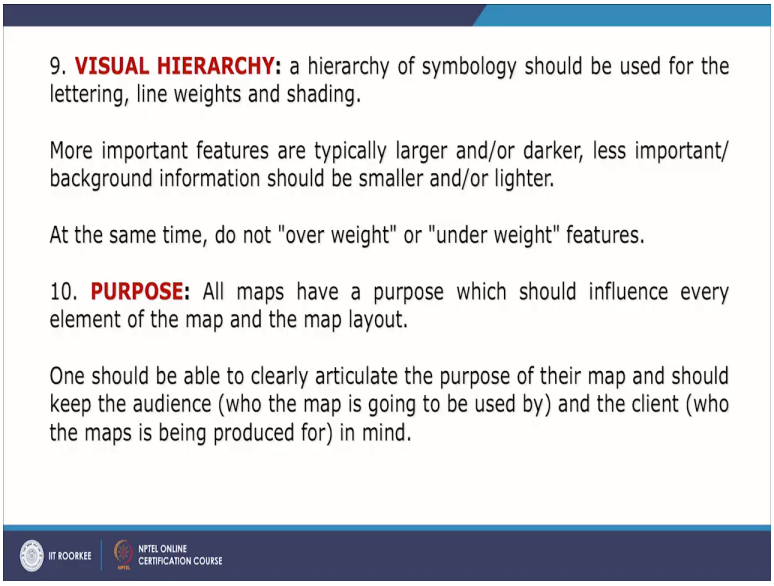
So, this is what is that if I am working say in part of Uttarakhand then a map of India in which boundary of Uttarakhand are also or all state boundaries and then it through a box, I should say that this is my study area. Another important key element of my map is this effective graphical design. Many times, it is on individual choice but I would suggest that whatever the design you follow or whatever layout you used, one must take care that they should not be cluttering. Unnecessary information should not be put.

But essential information or essential elements should not be deemphasized or avoided because people will accept or appreciate if map is having accuracy and all required information. So, a person may have produced a very nice-looking map but it might not be having say North symbol. It might be not having scale bar or legend. So, what is the use of that map? So, a good map in my opinion always should have all essential key elements and should not be cluttered with the information.

It should be neatly drawn. Appropriately and consistently generalized and of course, symmetrical balance; that means on one side within the Neatline or between that box, it should not be cluttered in one corner. So, there should be balance; balance of font sizes, balance of colour, balance of shading. Everything one should use that means while preparing your output, no matter how much time you have put while doing analysis of field collection or data collection, if your output is wrong or output is not good, no one would appreciate your work.

So, everything; your data collection, your data analysis and finally maps or your output should be also up to the mark. And as I have already said that without unnecessary cluttering of the map. Keep it simple; the way artistic details so we wary of the artistic details. Do not make your map cluttered with lot of information, lot of colours, lot of you know fancy symbols. This will not be appreciated at the end of the day.

(Refer Slide Time: 40:21)



Now, the ninth one in this key element is the visual hierarchy means whatever we wanted to convey through our maps, that should have the more weight. So, obviously the user should see that one first. Say for example I am keeping very large or red colour title for my map. But whereas my map is in a very small size and index map is very big size; that means the balance is missing so visual hierarchy is also not followed.

So, hierarchy of symbology should be used for the lettering; that means writing, line weights and shading. And more important features are typically larger. Whatever you want to emphasize or whatever the important thing, you should emphasize. And the other information's which supposed to be there or should be there but they should not be emphasized too much by creating there you know large size or having some fluorescent colour or attractive colours.

And at the same time, do not overweight or underweight features. Therefore, that means the balance is required and that one will learn only through experience and seeing others work also, especially the published one. So, those who are in the research domain or education domain, they can see publications of people; how they publish. And one more point here is the purpose; for what purpose that maps has been prepared.

So, all map should have a purpose and should influence every element of the map and the map layout. And one should be able to clearly articulate the purpose of their map and should keep the audience or users or the client, whom the maps is being produced for. So, one should keep the target, who is going to use my map? What is their background? And you know how good my map should look? So, if it is not looking good for myself then you can assume that others may not appreciate it.

So very neatly designed, very balanced one and emphasize the point which is your real output. Rest you can keep because they are supposed to be kept. So, if you take care about errors in GIS as we have discussed and also take care about the map element, one can really produce very reliable and very good output. So, with this, I end this discussion. Thank you very much.